ADVANCED SAMPLING AND MODELING IN MOLECULAR SIMULATIONS FOR SLOW AND LARGE-SCALE BIOMOLECULAR DYNAMICS

EDITED BY: Xiakun Chu, Yong Wang, Pengfei Tian, Wenfei Li and Davide Mercadante

PUBLISHED IN: Frontiers in Molecular Biosciences

frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88971-992-1
DOI 10.3389/978-2-88971-992-1

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

ADVANCED SAMPLING AND MODELING IN MOLECULAR SIMULATIONS FOR SLOW AND LARGE-SCALE BIOMOLECULAR DYNAMICS

Topic Editors:
Xiakun Chu, Stony Brook University, United States
Yong Wang, Zhejiang University, China
Pengfei Tian, Novozymes (Denmark), Denmark
Wenfei Li, Nanjing University, China
Davide Mercadante, University of Auckland, New Zealand

Citation: Chu, X., Wang, Y., Tian, P., Li, W., Mercadante, D., eds. (2021). Advanced Sampling and Modeling in Molecular Simulations for Slow and Large-Scale Biomolecular Dynamics. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-88971-992-1

Table of Contents

frontiers in Molecular Biosciences

EDITORIAL
published: 12 November 2021
doi: 10.3389/fmolb.2021.795991

Editorial: Advanced Sampling and Modeling in Molecular Simulations for Slow and Large-Scale Biomolecular Dynamics

Xiakun Chu[1], Yong Wang[2]*, Pengfei Tian[3]*, Wenfei Li[4,5]* and Davide Mercadante[6]**

[1]Department of Chemistry, State University of New York, Stony Brook, NY, United States, [2]College of Life Sciences, Shanghai Institute for Advanced Study, Institute of Quantitative Biology, Zhejiang University, Hangzhou, China, [3]Enzyme Research, Novozymes, Lyngby, Denmark, [4]National Laboratory of Solid State Microstructure, Department of Physics, Nanjing University, Nanjing, China, [5]Wenzhou Institute, University of Chinese Academy of Sciences, Wenzhou, China, [6]School of Chemical Sciences, The University of Auckland, Auckland, New Zealand

Keywords: molecular dynamics, enhanced sampling, coarse-grained model, conformational dynamics, free energy landscape

OPEN ACCESS

Edited and reviewed by:
Massimiliano Bonomi, Institut Pasteur, France

****Correspondence:***
Xiakun Chu
xkchu2008@gmail.com
Yong Wang
isb@zju.edu.cn
Pengfei Tian
tianpengfei09@gmail.com
Wenfei Li
wfli@nju.edu.cn
Davide Mercadante
davide.mercadante@auckland.ac.nz

Specialty section:
This article was submitted to Biological Modeling and Simulation, a section of the journal Frontiers in Molecular Biosciences

Received: *15 October 2021*
Accepted: *23 October 2021*
Published: *12 November 2021*

Citation:
Chu X, Wang Y, Tian P, Li W and Mercadante D (2021) Editorial: Advanced Sampling and Modeling in Molecular Simulations for Slow and Large-Scale Biomolecular Dynamics. Front. Mol. Biosci. 8:795991. doi: 10.3389/fmolb.2021.795991

Editorial on the Research Topic

Advanced Sampling and Modeling in Molecular Simulations for Slow and Large-Scale Biomolecular Dynamics

From conception, the impact of molecular dynamics (MD) simulations has grown dramatically (Karplus and McCammon, 2002), even though MD simulations are still affected by a critical timescale issue. Currently, timescales accessible to MD simulations are, on average, shorter than those of the investigated events, often resulting in insufficient sampling. Numerous efforts have been spent on accelerating MD simulations in order to ease the timescale problem. This Research Topic collects contributions focusing on developing and using advanced sampling techniques and modeling strategies to promote applications of MD simulations to study a diverse range of large-scale biomolecular systems.

Exhaustive sampling is especially important for intrinsically disordered proteins (IDPs), which show a more shallow and rugged energy landscape when compared to folded proteins (Papoian, 2008). Ding et al. presented case studies on two IDPs by an iterative screening-after-sampling strategy. In their study accelerated molecular dynamics was used to enhance the sampling of highly diverse conformational ensembles of IDPs, with Small-angle X-ray scattering (SAXS) to guide the sampling iteratively and obtain ensembles in good agreement with experimental data. Such integrative modeling might be more broadly useful for modeling IDPs ensembles. In general, the conformational dynamics of IDPs are strongly affected by the binding to other molecular partners or by aggregation. There are three excellent studies in this Research Topic, focusing on the conformational dynamics and aggregation properties of two Alzheimer's disease-related IDPs: the Aβ42 peptide and the tau protein (Selkoe and Hardy, 2016). In the work by Xie and Guo, replica exchange with solute tempering (REST) (Liu et al., 2005) has been adopted to sample the binding of the intrinsically disordered Aβ42 peptide to the Human serum albumin (HSA), elucidating the molecular mechanism of amyloid inhibition by HSA. Their results suggest that Aβ42 binds to multiple sites on HSA, which shifts the conformational propensity of the peptide towards a more disordered state altering its aggregation propensity altogether. The reward behind the quest to mechanistically characterize fibrillar nucleation in proteopathies is enormous as it would

suggest effective strategies for therapeutic approaches to neurodegeneration. Ma et al. employed coarse-grained simulations to energetically define the fibril growth of the Aβ-peptide: discovering the binding site of new filaments to the protofibril and a downhill mechanism of filament addition. Together with the identification of an emerging mechanical property of the Aβ-peptide protofibril, their research adds valuable insights to our understanding of the nucleation of Aβ-peptide fibrils for the development of strategies to pharmaceutically tackle fibril growth. McCarty and co-workers performed non-equilibrium steered MD simulations to investigate the structural changes of the tau paired-helical filament (PHF) and straight filament (SF) under mechanical force. In particular, the authors identified weak spots of interchain interactions additionally providing the dissociation pathway of a single tau peptide from the protofibril, through metadynamics simulations (Laio and Parrinello, 2002). In addition, the free energy profile for tau dissociation was obtained by umbrella sampling simulations.

One useful strategy to overcome sampling limitations is to introduce coarse-graining into biomolecular models. MARTINI is among the most widely used coarse-grained (CG) models for biomolecules (Monticelli et al., 2008). Mahmood et al. introduced a simple cutoff scheme to improve the definition of native contacts in the structure-based protein model of MARTINI (G$\bar{o}$-MARTINI). By tuning the interaction strengths and cutoff distances, the MARTINI CG simulations can well reproduce the structural fluctuations from atomistic simulations for the membrane proteins investigated in the published study. The refined model has been successfully used to simulate the key steps leading to the assembly of the F-BAR protein involved in membrane remodeling. Using the CG UNited-RESidue (UNRES) model (Liwo et al., 1997), Stevens and He studied the large-scale conformational changes within the multidomain scaffolding protein PICK1. Considering the large size of PICK1, associated with extensive conformational flexibility, brute-force atomistic simulations on this system would inevitably lead to insufficient sampling. The physics-based CG model allowed the authors to reliably characterize the detailed interactions at the residual level and eventually uncover the forces driving the association of PICK1 subdomains. CG modeling has also been suggested to study and understand chromosomal organization and dynamics (Lin et al., 2021). However, determining the potential energy function suitably reproducing the behaviour of chromosomal and multi-chromosomal systems has been historically challenging. Zhang and Huang introduced a "bottom-up" approach, aiming to derive the CG potential from all-atom MD simulations. They studied dynamics and interactions in two-nucleosome systems, providing useful information to determine functional forms and parameters to sample the interaction between nucleosomes.

Coarse-graining the MD simulation trajectories is another useful strategy to reduce system complexity, while elucidating molecular dynamics. In addition to the sampling problem, extracting key structural and dynamic features from high-dimensional MD trajectories is a challenge in biomolecular modeling. Such a task relies on an appropriate definition of feature space, within which the metastable states involved in biologically relevant conformational dynamics can be identified *via* clustering and dimensionality reduction analyses. The humanly understandable thermodynamic and kinetic information can then be reconstructed based on Markov State Models (MSM) (Chodera and Noé, 2014; Lane et al., 2011). In the work by Wang et al., an effective energy rescaling space trajectory mapping method has been developed to detect metastable states and construct kinetic transition networks. In their study the authors are able to successfully describe the major metastable states and the interstate transition kinetics involved in the folding of a dodecapeptide. MSM can be combined with enhanced sampling methods to further improve the performance of structural and dynamic characterizations as done by Fernández-Quintero et al. In this work, the authors demonstrated the correlation between the rigidification of the CDR-H3 loops of antibody fragments and the enhanced antigen specificity in different stages of affinity maturation by using metadynamics simulations in combination with MD simulations and MSM analyses.

This Research Topic also includes two binding case studies. Do et al. employed all-atom MD simulations with the aid of an enhanced sampling method called Gaussian-accelerated molecular dynamics (GaMD) (Wang et al., 2021) to determine the pathways and binding mechanism of caffeine to the human adenosine A2A receptor. By adding a harmonic boost potential, GaMD simulations allowed to capture the spontaneous ligand binding and release in the μs time scale through smoothening the potential energy surface so as to reduce the energy barriers for slow conformational changes. This work provided a good example on how to implement enhanced sampling methods to study a protein-ligand binding mechanism. In investigating the binding between nanoparticles and the clathrin-associated protein adaptin-2 (AP2), Zhu et al. applied molecular modeling and simulations to understand the impact of nanoparticle morphology on binding specificity. They found that binding specificity is majorly dictated by electrostatic interactions as well as nanoparticle morphology. They also observed that nanoparticle binding significantly induces conformational changes in AP2. Overall, the authors provided a microscopic explanation for cargo recognition in clathrin-mediated endocytosis and possible mechanisms to design high-efficiency nano-biomaterials.

As an essential element in MD simulations, the adopted force field is key to affect the precision of the simulation results. Wang and Li developed and tested force field parameters for some noncanonical amino acids (NAAs). NAAs have been widely applied in protein engineering, virus vaccine development, and medical therapeutics due to their strong site specificity, without the need to introduce significant perturbations to a protein structure. Based on quantum mechanics (QM) calculations and experimental data as a benchmark, the authors determined force field parameters for phenylalanine and tyrosine derivatives showing that the newly identified parameters well describe protein-ligand interactions with NAAs as substrates. Finally, to aid structural modeling useful for MD simulations, Xian et al. developed a structure manipulation (StructureMan) tool that proved to be

comprehensive and efficient when studying interactions in large biomolecular systems.

Overall, we believe that this Research Topic provides a well-rounded picture of the latest state-of-the-art developments useful to overcome historical limitations in modeling and sampling of large biomolecular systems and slow processes using classical molecular dynamics simulations.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

YW thanks the financial support from the start-up funds of Zhejiang University. WL thanks the support from the National Natural Science Foundation of China (11974173). DM thanks the financial support from the School of Chemical Sciences at The University of Auckland.

ACKNOWLEDGMENTS

We thank all the contributors and reviewers of this Research Topic.

REFERENCES

Chodera, J. D., and Noé, F. (2014). Markov State Models of Biomolecular Conformational Dynamics. *Curr. Opin. Struct. Biol.* 25, 135–144. doi:10.1016/j.sbi.2014.04.002

Karplus, M., and McCammon, J. A. (2002). Molecular Dynamics Simulations of Biomolecules. *Nat. Struct. Biol.* 9, 646–652. doi:10.1038/nsb0902-646

Laio, A., and Parrinello, M. (2002). Escaping Free-Energy Minima. *Proc. Natl. Acad. Sci.* 99, 12562–12566. doi:10.1073/pnas.202427399

Lane, T. J., Bowman, G. R., Beauchamp, K., Voelz, V. A., and Pande, V. S. (2011). Markov State Model Reveals Folding and Functional Dynamics in Ultra-Long Md Trajectories. *J. Am. Chem. Soc.* 133, 18413–18419. doi:10.1021/ja207470h

Lin, X., Qi, Y., Latham, A. P., and Zhang, B. (2021). Multiscale Modeling of Genome Organization With Maximum Entropy Optimization. *J. Chem. Phys.* 155, 010901. doi:10.1063/5.0044150

Liu, P., Kim, B., Friesner, R. A., and Berne, B. J. (2005). Replica Exchange With Solute Tempering: A Method for Sampling Biological Systems in Explicit Water. *Proc. Natl. Acad. Sci.* 102, 13749–13754. doi:10.1073/pnas.0506346102

Liwo, A., Ołdziej, S., Pincus, M. R., Wawak, R. J., Rackovsky, S., and Scheraga, H. A. (1997). A United-Residue Force Field for Off-Lattice Protein-Structure Simulations. I. Functional Forms and Parameters of Long-Range Side-Chain Interaction Potentials from Protein crystal Data. *J. Comput. Chem.* 18, 849–873. doi:10.1002/(sici)1096-987x(199705)18:7<849::aid-jcc1>3.0.co;2-r

Monticelli, L., Kandasamy, S. K., Periole, X., Larson, R. G., Tieleman, D. P., and Marrink, S.-J. (2008). The Martini Coarse-Grained Force Field: Extension to Proteins. *J. Chem. Theor. Comput.* 4, 819–834. doi:10.1021/ct700324x

Papoian, G. A. (2008). Proteins With Weakly Funneled Energy Landscapes Challenge the Classical Structure-Function Paradigm. *Proc. Natl. Acad. Sci.* 105, 14237–14238. doi:10.1073/pnas.0807977105

Selkoe, D. J., and Hardy, J. (2016). The Amyloid Hypothesis of Alzheimer's Disease at 25 Years. *EMBO Mol. Med.* 8, 595–608. doi:10.15252/emmm.201606210

Wang, J., Arantes, P. R., Bhattarai, A., Hsu, R. V., Pawnikar, S., Huang, Y.-m. M., et al. (2021). Gaussian Accelerated Molecular Dynamics: Principles and Applications. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 11, e1521. doi:10.1002/wcms.1521

Conflict of Interest: Author PT was employed by company Novozymes, Denmark.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Molecular Biosciences

ORIGINAL RESEARCH
published: 07 August 2020
doi: 10.3389/fmolb.2020.00182

Local and Global Rigidification Upon Antibody Affinity Maturation

*Monica L. Fernández-Quintero†, Johannes R. Loeffler†, Lisa M. Bacher, Franz Waibl, Clarissa A. Seidler and Klaus R. Liedl**

Center for Molecular Biosciences Innsbruck, Institute of General, Inorganic and Theoretical Chemistry, University of Innsbruck, Innsbruck, Austria

During the affinity maturation process the immune system produces antibodies with higher specificity and activity through various rounds of somatic hypermutations in response to an antigen. Elucidating the affinity maturation process is fundamental in understanding immunity and in the development of biotherapeutics. Therefore, we analyzed 10 pairs of antibody fragments differing in their specificity and in distinct stages of affinity maturation using metadynamics in combination with molecular dynamics (MD) simulations. We investigated differences in flexibility of the CDR-H3 loop and global changes in plasticity upon affinity maturation. Among all antibody pairs we observed a substantial rigidification in flexibility and plasticity reflected in a substantial decrease of conformational diversity. To visualize and characterize these findings we used Markov-states models to reconstruct the kinetics of CDR-H3 loop dynamics and for the first time provide a method to define and localize surface plasticity upon affinity maturation.

Keywords: antibodies, CDR-H3 loop, affinity maturation, rigidification, localizing plasticity, kinetics, Markov-state models

OPEN ACCESS

Edited by:
Wenfei Li,
Nanjing University, China

Reviewed by:
Prithviraj Nandigrami,
University of Illinois at Urbana-Champaign, United States
Huilin Ma,
Johns Hopkins University, United States

***Correspondence:**
Klaus R. Liedl
Klaus.Liedl@uibk.ac.at

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Biological Modeling and Simulation, a section of the journal Frontiers in Molecular Biosciences

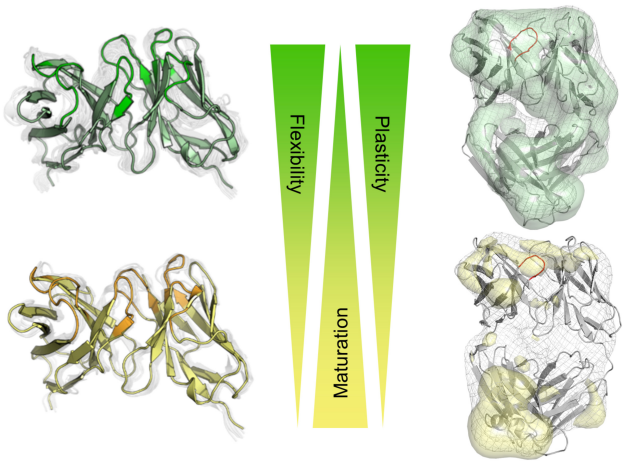
Received: *20 May 2020*
Accepted: *13 July 2020*
Published: *07 August 2020*

Citation:
Fernández-Quintero ML, Loeffler JR, Bacher LM, Waibl F, Seidler CA and Liedl KR (2020) Local and Global Rigidification Upon Antibody Affinity Maturation. Front. Mol. Biosci. 7:182. doi: 10.3389/fmolb.2020.00182

INTRODUCTION

Since the identification of antibodies in the 19th century, the rise and importance of monoclonal antibodies as biotherapeutics over the past 30 years has been extraordinary (Carter, 2006, 2011; Reichert, 2017; Kaplon and Reichert, 2019). Antibodies are composed of two polypepide chains, called V_H and V_L (Edelman, 1973). Each chain consists of a variable and a constant region. The variable domain contains six hypervariable loops, referred to as the complementarity determining regions (CDRs), which shape the antigen-binding site, the paratope (Nguyen et al., 2017). The specificity of an antibody is mainly influenced by the CDR loops and therefore characterization of the paratope is essential for understanding the function of the antibody (James et al., 2003). Five of the six CDR loops, except the CDR-H3 loop, can adopt a limited number of main-chain conformations and have been classified into canonical structures according to their length and sequence composition (Chothia and Lesk, 1987; Al-Lazikani et al., 1997). The highest variability in sequence, length and structure of an antibody can be observed in the CDRs, especially in the CDR-H3 loop, while antibody frameworks are fairly well conserved (∼150 human germline framework sequences). The CDR-H3 loop plays a central role in antigen recognition and has on average the highest counts of contacts with antigens (Marks and Deane, 2017; Regep et al., 2017). Structure

Abbreviations: CDR, complementary determining region; Fab, antigen binding fragment; Fv, antibody variable fragment; MD, molecular dynamics; PCCA, perron cluster cluster analysis; RMSD, root mean square deviation; tICA, time-lagged independent component analysis; V_H, heavy chain; V_L, light chain.



GRAPHICAL ABSTRACT | Visualization of the local and global rigidification as a consequence of affinity maturation.

prediction of the CDR-H3 loop due to its exceptional diversity of both structure and sequence and the ability to adopt various different conformations during V(D)J recombination and somatic hypermutation remains challenging (Bassing et al., 2002; Market and Papavasiliou, 2003; Clark et al., 2006; Burkovitz et al., 2014). Additionally, the CDR-H3 loop length and structure have an effect on the antigen-binding patterns of the CDR loops and influence the specificity of the paratope for target antigens. Thus, to elucidate the role of B cells in adaptive immunity and the evolution of antibodies binding specific antigens, the understanding of the affinity maturation process and its effects on the CDR loops, especially on the CDR-H3 loop, are crucial. The binding site of polyreactive monoclonal antibodies, which bind with low affinity to various structurally unrelated antigens, has been discussed to be significantly more flexible compared to matured antibodies (Zhou et al., 2007; Gunti and Notkins, 2015). Depending on the antigen present, polyreactive antibodies show a broader and shallower free energy surface, reflected in various different binding site conformations and higher conformational diversity of the paratope (Schmidt et al., 2013; Fernández-Quintero et al., 2019b). Especially the CDR-H3 loop substantially influences the shape of the paratope and thus plays a central role in antigen-binding. The correlation between rigidification and enhanced specificity has been discussed in terms of conformational selection (Ma et al., 1999; Tsai et al., 1999). Antibody-antigen binding can be interpreted to follow the paradigm of conformational selection. This implies an ensemble of pre-existing conformations with different probabilities, in which the binding-competent state is selected (Tsai et al., 1999; Csermely et al., 2010). Repeated exposure to the same antigen leads to mutations in the sequences which can result in a rigidification of the antigen binding site. Various studies focused on the effects of affinity maturation on the CDRs suggesting that structural rigidification and less conformational diversity are a consequence of affinity maturation (Wedemayer et al., 1997; Manivel et al., 2000; Yin et al., 2001, 2003; Li et al., 2003; Thielges et al., 2008; Adhikary et al., 2012, 2015; Schmidt et al., 2013; Jeliazkov et al., 2018). Additionally, 3-pulse photon echo peak shift (3PEPS) spectroscopy has been used to quantify antibody dynamics on the femto-to nanosecond timescale. A direct comparison between naïve with mature antibodies showed that mature antibodies can be characterized by a higher rigidity, reflected in smaller motions and conformational changes than naïve antibodies (Jimenez et al., 2003; Adhikary et al., 2012, 2015). Additionally, numerous MD studies investigated and showed the rigidification of the CDR-H3 loop as a consequence of affinity maturation (Thorpe and Brooks, 2007; Wong et al., 2011; Schmidt et al., 2013). Recently, it has been reported that antibody CDR-H3 loops does not result in a rigidification (Jeliazkov et al., 2018), but it has also been shown that on a significantly longer timescale the CDR-H3 loop rigidifies upon affinity maturation (Fernández-Quintero et al., 2019b). Thus, the affinity maturation process represents a direct connection between an enhanced specificity and rigidification. However, rigidification is only one of numerous biophysical mechanisms responsible for the increase in affinity (Jeliazkov et al., 2018).

In this study, we focus on characterizing the conformational diversity of the CDR-H3 loop including transition probabilities and changes in surface plasticity of 10 pairs of antibody fragments upon affinity maturation. We based our investigation on strong experimental structural information and compared naïve (before exposure to an antigen) and matured (after repeated exposure to an antigen) antibodies crystallized with and without the presence of the antigen.

MATERIALS AND METHODS

A previously published method characterizing the CDR-H3 loop ensemble upon antigen-binding in solution (Fernández-Quintero et al., 2019a,b, 2020a,b,c) was used to investigate the conformational diversity of CDR-H3 loop upon affinity maturation. Experimental structural information was available for all considered antibody fragments (Fabs and Fvs).

To avoid repetition, we only discuss three pairs of antibody fragments upon affinity maturation in detail, while the results for the other antibodies are summarized in **Figure 1**. The structural changes upon affinity maturation for all ten antibody pairs are visualized and described in more detail in the **Supplementary Material** (**Supplementary Tables 4–12**). This 10 pairs of antibodies undergoing affinity maturation were chosen as they have been part of previous work considering the effects of affinity maturation on antibody flexibility (Wedemayer et al., 1997; Yin et al., 2001; Jimenez et al., 2003; Li et al., 2003, 2015; Zimmermann et al., 2006; Thorpe and Brooks, 2007; Thielges et al., 2008; Babor and Kortemme, 2009; Wong et al., 2011; Schmidt et al., 2013; Willis et al., 2013; Adhikary et al., 2015; Schiele et al., 2015; Jeliazkov et al., 2018; Fernández-Quintero et al., 2019b).

The first affinity maturation pair analyzed is the D44.1 (naïve) and the F10.6.6 (matured) anti-lysozyme antibody Fab crystallized with and without the antigen lysozyme (Braden et al., 1994). Both antibody Fabs are murine monoclonal antibodies

Antibodies	Naïve bound (PDB Codes)	Naïve AGless (PDB Codes)	Matured bound (PDB Codes)	Matured AGless (PDB Codes)	CDR-H3 Loop Length	Number of CDR-H3 loop clusters naïve	Number of CDR-H3 loop clusters matured
D44.1/F10.6.6	1MLC	1MLB	1P2C	2Q76	7	183	65
28B4	1FL6	1FL5	1KEL	1KEM	8	135	50
48G7	1AJ7	2RCS	1GAF	1HKL	5	210	120
S25	1Q9Q	1Q9K	1Q9W	1Q9O	11	263	150
AZ-28	1D6V	1D5I	1AXS	1DB5	11	123	80
H63/H8	1DQJ		1NDG		5	151	32
aDabiFab	4YHI	4YGV	4YHO	4YHN	11	160	114
Anti-MPTS	4NJA		4NJ9		11	99	59
4-4-20	1T66		4FAB		7	123	68
7G12	1N7M	1NGZ	1NGW	1NGY	5	114	19

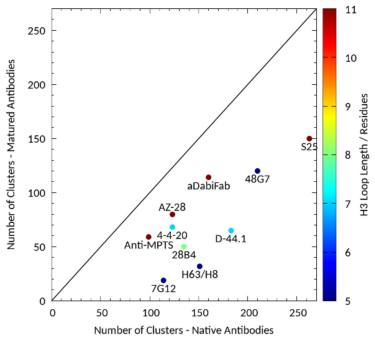


FIGURE 1 | Overview of all antibody fragments analyzed with the available PDB accession codes crystallized with and without antigen and in the naive and matured state. Additionally, the resulting number of clusters of the CDR-H3 loop of the naïve and matured antibody fragments by using the same distance cut-off of 1.2 Å is shown on the left and visualized on the right. The plot on the right shows the number of clusters of the naïve antibody against the number of clusters of the matured antibody color-coded according to their loop length.

which are related in sequence and structure as they originate from the same germline gene rearrangement. The available PDB accession codes of the naïve and matured antibody fragments crystallized with and without antigen are 1MLC, 1MLB, and 2Q76, 1P2C, respectively (Braden et al., 1994; Cauerhff et al., 2004; Acierno et al., 2007). D44.1 Fab differs from the affinity matured variant F10.6.6 in 20 mutations, seven of them located in the CDR loops. Due to the occurring mutations, structural changes yielded in a stabilized V_H–V_L interface with an increase in the affinity toward the antigen. A significant increase in the number of non-covalent bonds between the antibody and the antigen from the naïve complex 93 to the matured complex 129 as well as closer and stronger bonds were observed. The second studied affinity maturation pair is the 28B4 Fab, which catalyzes a periodate-dependent oxidation of sulfide to sulfoxide, whereby the hapten (1-[N-4′-nitrobenzyl-N-4′-carboxybutylamino] methylphosphonic acid) was generated to mimic the transition state of this reaction (Hsieh-Wilson et al., 1996; Yin et al., 2001). The available experimental structures are germline Fabs crystallized with and without the hapten present (PDB codes: 1FL6 and 1FL5) and the respective affinity matured Fab variants (1KEL and 1KEM). Nine mutations, two in the V_L and seven in the V_H were introduced during affinity maturation. Three of these mutated residues of the matured antibody, Asn35H, Lys56H, and Trp101H directly interact with the hapten. A decrease in flexibility and changes in the binding geometry of the antigen due to these mutations led to an increased complementarity and affinity between the antibody and the hapten. The third pair of affinity maturation antibodies is the esterolytic antibody 48G7, which catalyzes an ester and carbonate hydrolysis reaction, whereby the hapten 5-(para-nitrophenyl phosphonate)-pentanoic acid portrays the transition state (Wedemayer et al., 1997). Available crystal structures that were used as starting structures for MD simulations are the germline Fab fragment in complex with and without the antigen present (PDB codes: 1AJ7 and 2RCS) as well as the corresponding affinity matured structures (PDB codes: 1GAF and 1HKL). During the process of affinity maturation, nine mutations were introduced, three in the V_L and six in the V_H.

The starting structures for simulations were prepared in MOE (Molecular Operating Environment, Chemical Computing Group, version 2018.01) using the Protonate3D tool (Labute, 2009; Molecular Operating Environment [MOE], 2018). To neutralize the charges we used the uniform background charge (Roe and Cheatham, 2013; Hub et al., 2014; Case et al., 2016). Using the tleap tool of the AmberTools16 (Roe and Cheatham, 2013; Case et al., 2016) package, the crystal structures were soaked with cubic water boxes of TIP3P water molecules with a minimum wall distance of 10 Å to the protein (Jorgensen et al., 1983). For all crystal structures parameters of the AMBER force field 14SB were used (Maier et al., 2015). The antibody fragments were carefully equilibrated using a multistep equilibration protocol (Wallnoefer et al., 2011).

Metadynamics Simulations

To enhance the sampling of the conformational space well-tempered metadynamics (Barducci et al., 2008, 2011; Biswas et al., 2018) simulations were performed in GROMACS (Pronk et al., 2013; Abraham et al., 2015) with the PLUMED 2 implementation (Tribello et al., 2014). We used a linear combination of sine and cosine of the ψ torsion angles of the CDR-H3 and CDR-L3 loop as collective variables, calculated with functions MATHEVAL and COMBINE implemented in PLUMED 2 (Tribello et al., 2014). As discussed previously the ψ torsion angle captures conformational transitions comprehensively (Ramachandran et al., 1963; Wood and Hirst, 2005; Fernández-Quintero et al., 2019b). The decision to include the CDR-L3 loop ψ torsion angles is based on the structural correlation of the CDR-L3 and CDR-H3 loop and the observed improved sampling efficiency (James and Tawfik, 2005). The simulations were performed at 300 K in an NpT ensemble. We used a Gaussian height of 10.0 kcal/mol. Gaussian deposition occurred every 1,000 steps and a biasfactor of 10 was used. 1 μs metadynamics simulations were performed for each available antibody fragment crystal structure. The resulting trajectories were clustered by using the average linkage hierarchical clustering algorithm in CPPTRAJ (Shao et al., 2007; Roe and Cheatham, 2013) with a distance cut-off criterion of 1.2 Å resulting in a large number of clusters. The cluster

representatives for the antibody fragments were equilibrated and simulated for 100 ns using the AMBER18 (Case et al., 2016) simulation package.

Molecular Dynamics Simulations

Molecular dynamics simulations were performed in an NpT ensemble using pmemd.cuda (Salomon-Ferrer et al., 2013). Bonds involving hydrogen atoms were restrained by applying the SHAKE algorithm (Miyamoto and Kollman, 1992), allowing a time step of 2.0 fs. Atmospheric pressure of the system was preserved by weak coupling to an external bath using the Berendsen algorithm (Berendsen et al., 1984). The Langevin thermostat (Adelman and Doll, 1976) was used to maintain the temperature during simulations at 300 K.

For the obtained trajectories a tICA was performed using the python library PyEMMA 2 employing a lag time of 10 ns (Scherer et al., 2015). Thereby, a dimensionality reduction is obtained by transforming the trajectories into an intuitive measure, e.g., backbone torsions, which represent the slowest coordinates of the system (Pérez-Hernández and Noé, 2016; Wu and Noé, 2017). To construct the tICA we chose as input variables the backbone torsions of the CDR-H3 loop. The first two tICs (time-lagged independent components) describe the two slowest components of the CDR-H3 loop movements. Thermodynamics and kinetics were calculated with a Markov-state model (Chodera and Noé, 2014) by using PyEMMA 2, which uses the k-means clustering algorithm (Likas et al., 2003) to define microstates and the PCCA + clustering algorithm to coarse grain the microstates to macrostates. PCCA + is a spectral clustering method, which discretizes the sampled conformational space based on the eigenvectors of the transition matrix (Röblitz and Weber, 2013). Markov-state models allow to identify significant structural changes during the simulation and reconstruct thermodynamics and kinetics. The sampling efficiency and the reliability of the Markov-state model (e.g., defining optimal feature mappings) can be evaluated with the Chapman–Kolmogorov test (Karush, 1961; Miroshin, 2016), by using the variational approach for Markov processes (Wu and Noé, 2017) and by taking into account the fraction of states used, as the network states must be fully connected to calculate probabilities of transitions and the relative equilibrium probabilities. To build the Markov-state model we used the backbone torsions of the CDR-H3 loop, defined 150 microstates using the k-means clustering algorithm and applied a lag time of 10 ns.

Characterization of Surface Plasticity

Conformational plasticity of proteins has been shown to play key role in molecular mechanisms such as catalytic activity, biomolecular recognition and allosteric regulation (Daberdaku and Ferrari, 2018; Jespersen et al., 2019). Differences of the antibody surface were calculated by using the average surface of the simulation and the respective standard deviations of each frame. To visualize the differences in plasticity upon affinity maturation, we calculated the per-voxel average and standard deviation of the reconstructed grid. The standard deviation is useful to highlight regions that are sometimes occupied by the protein and sometimes solvent-accessible. Flexible regions are characterized by large volumes with high standard deviation. However, the resulting grid is difficult to interpret because even very rigid regions can have a few partially occupied voxels. To emphasize regions with large structural differences, we applied a Gauss filter to smooth the average and the standard deviation grid. To test our method, we used the anti-MPTS Fv, previously analyzed to address the influence of the affinity maturation on the CDR-H3 loop (Fernández-Quintero et al., 2019b), to compare experimentally measured plasticity via 3PEPS spectroscopy (Adhikary et al., 2015) with our calculated plasticity. 3PEPS has been successfully used to characterize protein dynamics such as side chain rotations and loop rearrangements (Oh et al., 2011; Adhikary et al., 2012). In line with the experiment we observe a decrease in plasticity and flexibility for the further matured 8B10 Fv (**Supplementary Figure S1**).

RESULTS

Various studies have discussed the effect of affinity maturation on structural and dynamic properties (James and Tawfik, 2003; Cauerhff et al., 2004; Schmidt et al., 2013; Adhikary et al., 2015; Jeliazkov et al., 2018; Shehata et al., 2019).

We analyzed 10 pairs of antibody fragments supported by strong experimental structural information upon affinity maturation and a summary of the resulting CDR-H3 loop flexibilities of the respective antibody pairs is illustrated in **Figure 1**. On the left the PDB accession codes, the CDR-H3 loop lengths and the resulting numbers of clusters by using the same distance cut-off criterion of 1.2 Å, are displayed for all studied antibody fragments. On the right the number of CDR-H3 loop clusters of the naïve and matured antibody fragments are plotted against each other to visualize the substantial rigidification upon affinity maturation. The clustering also been performed using different cut-off criteria to see if the results presented in **Figure 1** are stable under variation of the cut-off and in all cases the native antibodies reveal a higher number of clusters, indicating a higher flexibility of the CDR-H3 loop before maturation.

As described in the "Materials and Methods" section, we used the cluster representatives as starting structures for each 100 ns MD simulations to be able to reconstruct and characterize thermodynamics and kinetics. **Figure 2A** displays the resulting free energy surface of 18.3 μs of the naïve D44.1 Fab and 6.5 μs of the matured F10.6.6 Fab in the same coordinate system. Upon affinity maturation a substantial rigidification of the CDR-H3 loop dynamics combined with a population shift toward the global minimum in solution could be observed. **Figure 2C** shows the resulting CDR-H3 loop ensemble in solution color-coded according to **Figure 1** and emphasizes the significant decrease in conformational diversity. **Supplementary Figure S2** illustrates the 2D-RMSD plots (based on the Cα coordinates) and the B-factors of the CDR-H3 loop. In line with the decrease in conformational space of the CDR-H3 loop, which can be seen in **Figure 2**, the rigidification of the CDR-H3 loop is reflected in both the 2D-RMSD and the B-factors. Besides, characterizing flexibility by the resulting number of clusters,

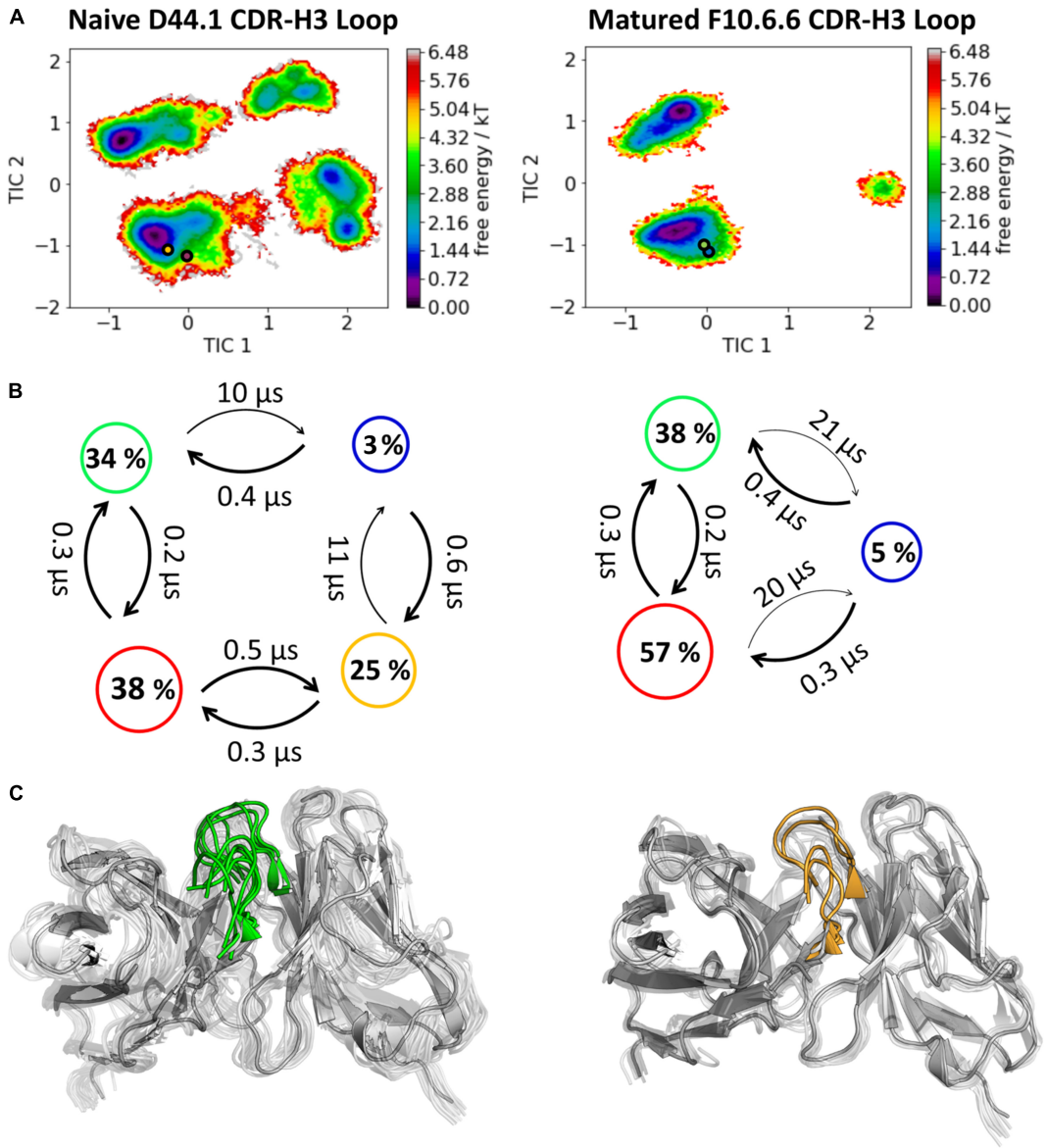


FIGURE 2 | Kinetic, thermodynamic and structural analyses of the CDR-H3 loop ensemble in solution. **(A)** Free energy surface of the naive D44.1 and the matured F10.6.6 Fab in the same coordinate system, including the respective X-ray structures crystallized with and without antigen. The orange and green dots show the bound X-ray structures of the naïve and the matured Fab, respectively, while magenta and blue display the X-ray structures crystallized without antigen. **(B)** Transition timescales between the different macrostates orientated according to the tICA space including the state probabilities. **(C)** Substantial decrease of the conformational ensemble of the D44.1 and F10.6.6 antibody upon affinity maturation.

RMSF or 2D-RMSD plots, we developed a method to analyze and localize surface plasticity of antibody fragments (**Figure 3**). **Figure 3** shows the projection of the calculated plasticity of the naïve and the matured antibody Fab onto a representative ensemble structure. The intensity of the colors reflects regions with higher plasticity. As surface plasticity is an essential aspect of biomolecular recognition, we find that characterization of protein plasticity allows a better shape-based interpretation of the antigen binding site, compared to other flexibility measures such as RMSD and B-factors. Upon affinity maturation we observe a significant decrease in surface plasticity. Also, the CDR-H3 loop reveals substantially less plasticity in the matured F10.6.6 Fab. This observation is in line with the decrease in conformational diversity, in particular of the CDR-H3 loop. The 2D-RMSD plots of both the paratope and the whole variable fragment are illustrated in the **Supplementary Figure S3** and clearly show a global rigidification upon affinity maturation.

The second studied affinity maturation pair is the hapten-binding 28B4 antibody Fab. **Figure 4A** shows the resulting tICA plots of the resulting 13.5 µs trajectories of the naïve and 5.0 µs trajectories of matured 28B4 antibody in the same coordinate system. The available crystal structures are

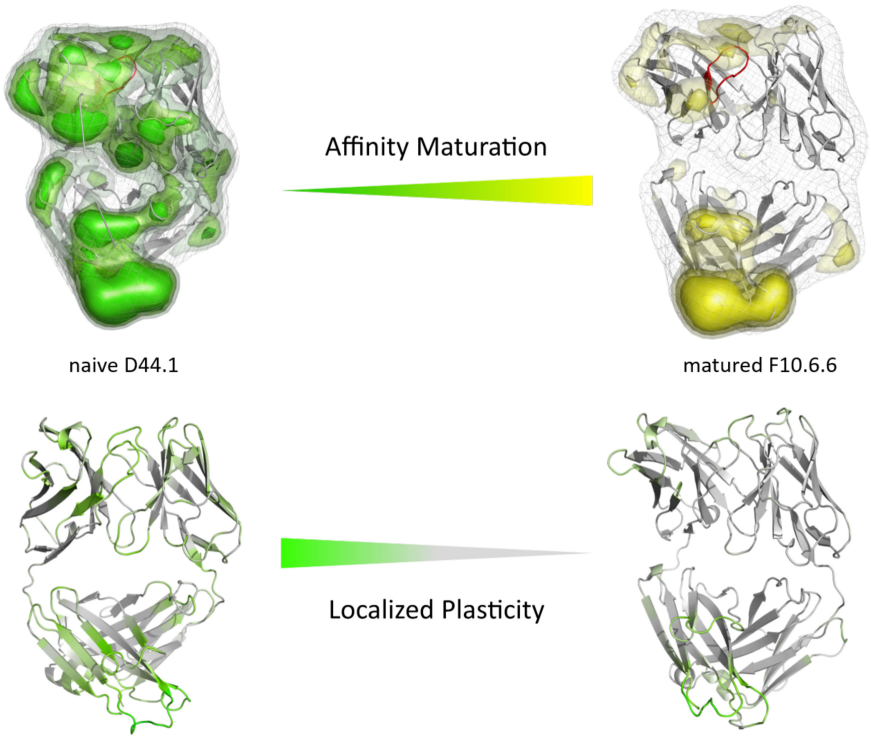


FIGURE 3 | Global **(top)** and localized **(bottom)** surface plasticity of the naive D44.1 and matured F10.6.6 antibody Fab mapped onto a representative ensemble structure. **(Top)** Global plasticity of the naïve and the matured D44.1 and F10.6.6 antibody fragments, highlighting the CDR-H3 loop in red. The intensity of the colors reflects the regions with higher plasticity and thus allows localization. **(Bottom)** Localized plasticity for both the naïve and the matured D44.1 and F10.6.6 antibody fragment Fab, respectively.

projected into the free energy landscape and color-coded respectively. The conformational ensemble of the CDR-H3 loop in solution reveals a substantial rigidification upon affinity maturation, reflected in a substantial decrease in conformational diversity (**Figure 4**). This significant rigidification of the CDR-H3 loop is also shown in **Supplementary Figure S4**. The 2D-RMSD plot clearly depict this decrease in flexibility upon affinity maturation. This finding is supported by the B-factors calculated for the CDR-H3 loop, as always higher values are obtained for the naïve antibodies. **Figure 4C** visualizes the substantial rigidification in the observed conformational diversity, which agrees with previous results. The effect of affinity maturation on the plasticity of the 28B4 antibody is visualized in **Figure 5**. Again, in line with the first analyzed pair we observe a decrease in plasticity upon affinity maturation, especially in the region of the CDR-H3 and CDR-L3 loop. This observation is confirmed by the 2D-RMSD plots of the paratope and the variable fragment illustrated in **Supplementary Figure S5**.

The third in detail discussed affinity maturation pair is the 48G7 hapten binding antibody. **Figure 6A** reflects in agreement in with the substantial decrease in the number of CDR-H3 loop clusters (210 to 120), as a metric of quantifying flexibility, a substantial reduction in conformational space of the CDR-H3 loop. A representative conformational ensemble of the resulting 21 μs (naïve) and 12 μs (matured) trajectories revealing this significant decrease in conformational diversity is illustrated in **Figure 6C**. This finding is in line with localized flexibility metrics, such as the B-factors and the 2D-RMSD of the CDR-H3 loop shown in **Supplementary Figure S6**. **Figure 6B** illustrates the transition probabilities between the obtained macrostates for both the naive and the matured antibody fragment and shows the populations of the respective states. We clearly see that upon affinity maturation the dominant minimum in solution is shifted and the binding competent state becomes the most dominant state in solution (76%). Analysis of the resulting plasticity in **Figure 7** displays significant reduction, especially in the CDR-H3 loop upon affinity maturation, which is highlighted by the localized plasticity in **Figure 7** (bottom). 2D-RMSD plots for the paratope and the variable fragment of the 48G7 antibody are depicted in **Supplementary Figure S7** and confirm this overall rigidification upon affinity maturation.

DISCUSSION

In this present study, we characterize the conformational diversity and the kinetic and thermodynamic properties of the CDR-H3 loop of 10 affinity maturation antibody pairs and present a method to visualize, localize and describe plasticity of antibodies upon affinity maturation. The affinity maturation process comprises the introduction of

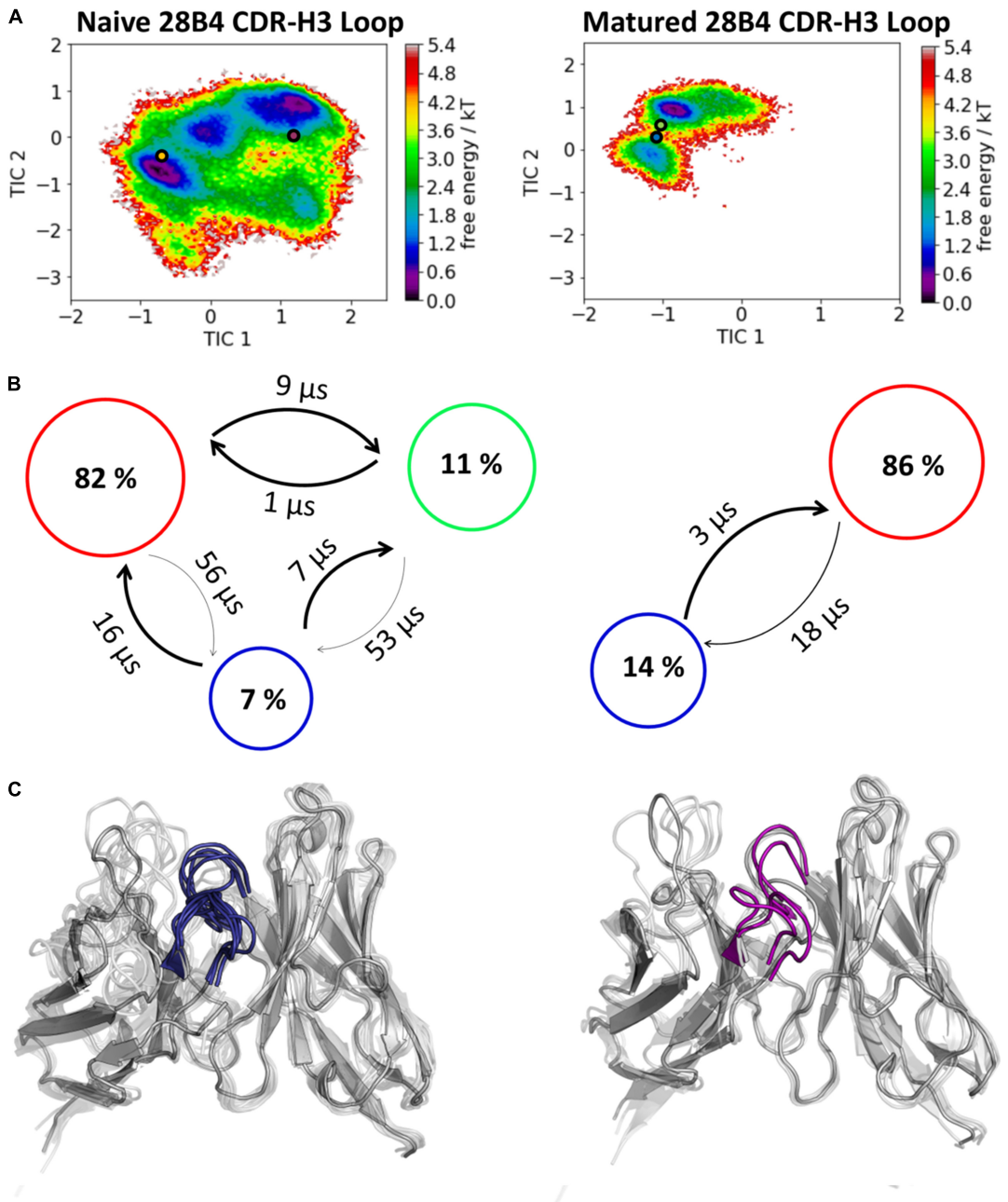


FIGURE 4 | Kinetic, thermodynamic and structural analyses of the CDR-H3 loop ensemble in solution upon affinity maturation. **(A)** Free energy surface of the naive and matured 28B4 Fab in the same coordinate system, including the respective X-ray structures crystallized with and without antigen. The orange and green dots show the bound X-ray structures of the naïve and the matured Fab, respectively, while magenta and blue display the X-ray structures crystallized without antigen. **(B)** Transition timescales between the different macrostates orientated according to the tICA space including the state probabilities. **(C)** Substantial decrease of the conformational ensemble of the 28B4 antibody upon affinity maturation.

combinatorial mutations that increase the binding affinity of the antibody to the antigen and lead to a more effective immune response (French et al., 1989). Numerous structural studies, involving small molecules (haptens) were carried out comparing affinity-matured antibodies and their germline precursor binding to the same antigen (Alzari et al., 1990; Hsieh-Wilson et al., 1996; Chong et al., 1999; Yin et al., 2001; Mishra and Mariuzza, 2018). Thereby, somatic hypermutations in the CDR loops lead to a higher number of hydrogen bonds, electrostatic interactions, van der Waals contacts and an improved shape complementarity (Fernández-Quintero et al., 2019a, 2020b,c). Large conformational preorganization of the paratope in combination with a decrease in flexibility upon affinity maturation has been discussed to increase specificity

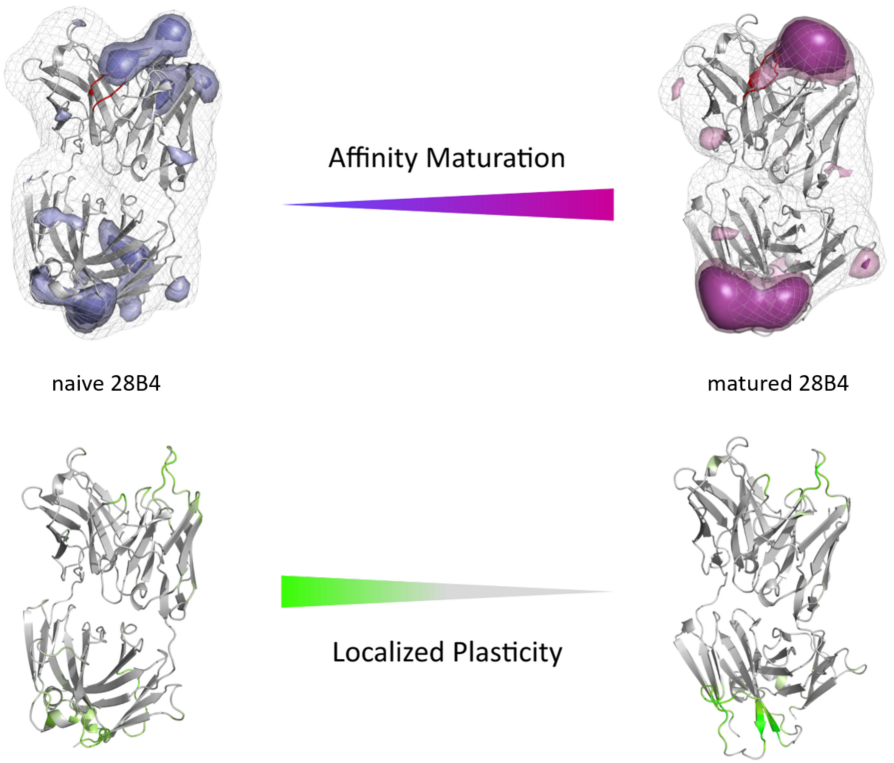


FIGURE 5 | Global **(top)** and localized **(bottom)** surface plasticity of the naive and matured 28B4 antibody Fab mapped onto a representative ensemble structure. **(Top)** Global plasticity of the naïve and the matured 28B4 antibody, highlighting the CDR-H3 loop in red. The intensity of the colors reflects the regions with higher plasticity and allows localization. **(Bottom)** Localized plasticity for both the naïve and the matured 28B4 antibody, respectively.

for the target antigen while reducing the possibility of cross-reactivity with other antigens (Wedemayer et al., 1997; Manivel et al., 2000). Compared to affinity maturation studies focusing on haptens, structural affinity maturation of an antibody in response to a protein, i.e., hen egg white lysozyme, could not be attributed to a higher number of formed hydrogen bonds or salt bridges, but to an improved shape complementarity at the V_H-binding interface accompanied by an increase of hydrophobic interactions (Braden et al., 1994; Li et al., 2003; DeKosky et al., 2016). In order to understand the mechanism of antigen-recognition, characterization of the thermodynamic and kinetics pathway of the affinity maturation process in combination with experimental structural information is crucial (Foote and Milstein, 1991, 1994; Milstein, 1991; Akiba and Tsumoto, 2015). Thus, the results presented in this study highlight that static structural information alone might not be sufficient to describe antibody binding properties as specificity and promiscuity (Akiba and Tsumoto, 2015; Fernández-Quintero et al., 2019b; Alba et al., 2020). Long timescale dynamics from enhanced and classic MD simulations complement experimental structural information with reliable estimations of flexibilities, state probabilities, binding mechanisms, and localization of plasticity. **Figure 1** displays an overview of all studied affinity maturation antibody fragments including the resulting number of CDR-H3 loop clusters by using the same distance cut-off criterion, as a quantification of rigidification upon affinity maturation. We also investigated the stability of the results in dependence of the clustering cut-off and observed the same trend, that upon affinity maturation the flexibility of the CDR-H3 loop decreases substantially. **Figures 2A,B** show the free energy surface of the naïve D44.1 and the matured F10.6.6 Fab in the same tICA coordinate system and reveal a substantial decrease in conformational space of the CDR-H3 loop. The naïve D44.1 Fab displays a broader free energy landscape, compared to the deeper and narrower minima observed for the matured F10.6.6 Fab. Besides, we identified that even without the antigen present within the pre-existing ensemble of conformations, the binding competent state lies in the dominant minimum in solution. This indicates that the D44.1 Fab follows the paradigm of conformational selection. The two highest populated states of the CDR-H3 loop in solution of the naïve D44.1 Fab are the dominant conformations of the matured F10.6.6 Fab. The binding competent state in the naïve antibody becomes the highest populated state upon affinity maturation (38 → 58% state population). **Figure 2C** illustrates the conformational ensemble of the CDR-H3 loop and emphasizes the substantial reduction in conformational diversity upon affinity maturation. This substantial rigidification upon affinity maturation is supported by the 2D-RMSD plots and the B-factors illustrated in **Supplementary Figures S2, S3**. **Figure 3** visualizes and localizes differences and regions with high plasticity. We did not only observe an overall decrease in plasticity, but we could also identify

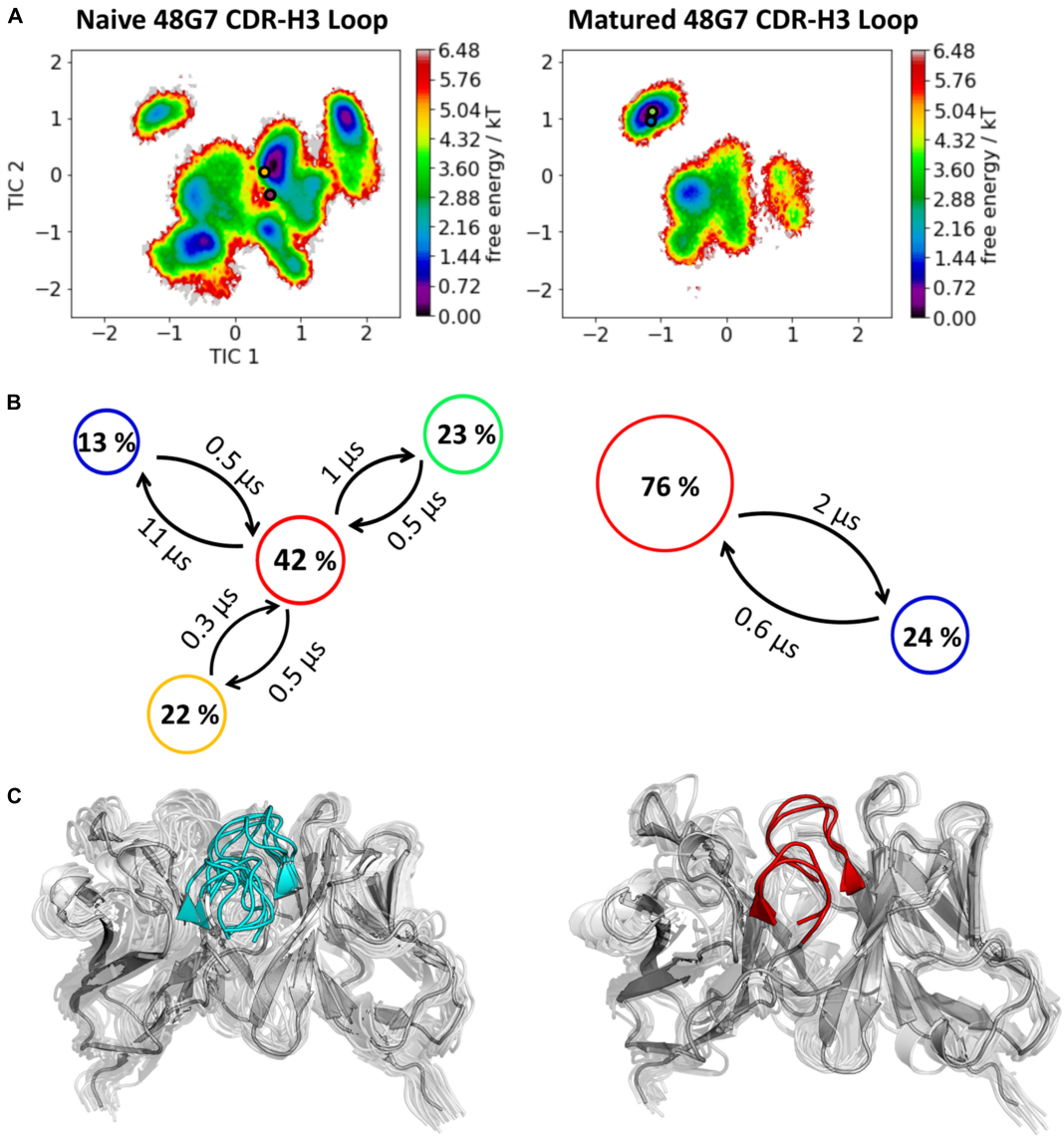


FIGURE 6 | Kinetic, thermodynamic and structural analyses of the CDR-H3 loop ensemble in solution upon affinity maturation. **(A)** Free energy surface of the naive and matured 48G7 Fab in the same coordinate system, including the respective X-ray structures crystallized with and without antigen. The orange and green dots show the bound X-ray structures of the naïve and the matured Fab, respectively, while magenta and blue display the X-ray structures crystallized without antigen. **(B)** Transition timescales between the different macrostates orientated according to the tICA space including the state probabilities. **(C)** Substantial decrease of the conformational ensemble of the 48G7 antibody upon affinity maturation.

a substantial reduction in the CDR-H3 loop surface plasticity. **Supplementary Figure S8** depicts the localized surface plasticity for the CDR-H3 loop for all in detail investigated antibody fragments and **Supplementary Table 13** summarizes the overall reduction in plasticity upon affinity maturation. **Figures 4A,B** illustrate in line with the observations of the D44.1.1/F10.6.6 affinity maturation study, a substantial rigidification of the CDR-H3 loop conformational space of the 28B4 Fab upon affinity maturation. Besides the substantial rigidification we identified that the dominant structure in solution was optimized to bind the antigen, while the Fab X-ray structure crystallized without antigen lies in a local shallow side-minimum, because of the distortion of the loop due to crystal contacts with the tail region of a symmetry mate Fab. The transition kinetics of the CDR-H3 loop for both the naïve and the matured Fab occur in the nano-to microsecond timescale. Additionally, we also observe a strong population shift upon affinity maturation. Again, the significantly reduced conformational ensemble is illustrated in **Figure 4C** and supports in line with all other observations the rigidification upon affinity maturation. **Figure 5** visualizes differences in plasticity of the 28B4 affinity maturation study and clearly shows in particular for the CDR-H3 and CDR-L3 loop a substantial decrease in surface plasticity. In agreement with these results **Figures 6A,B** show not only a decrease in

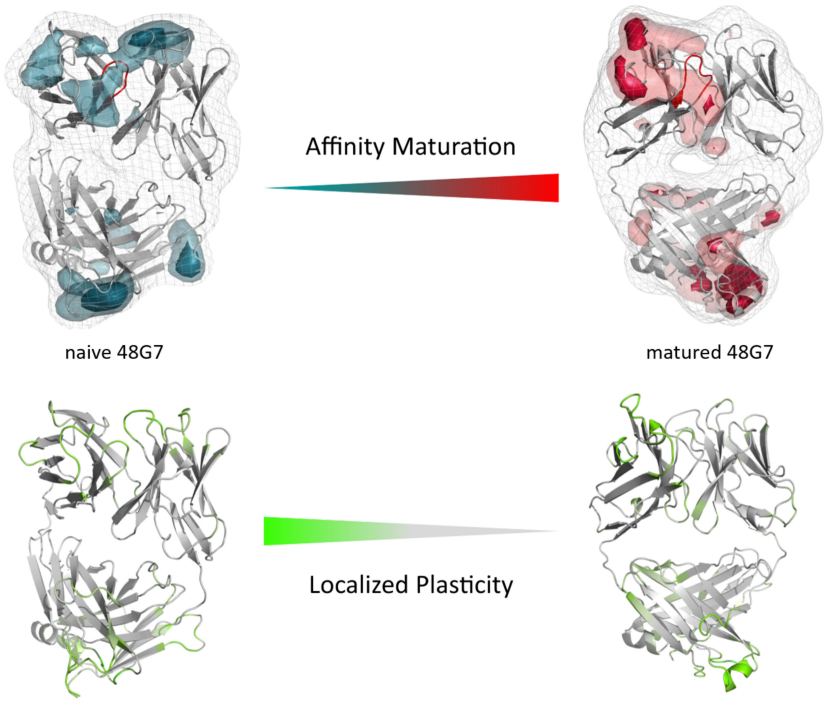


FIGURE 7 | Global **(top)** and localized **(bottom)** surface plasticity of the naive and matured 4G87 antibody Fab mapped onto a representative ensemble structure. **(Top)** Global plasticity of the naïve and the matured 48G7 antibody, highlighting the CDR-H3 loop in red. The intensity of the colors reflects the regions with higher plasticity and allows localization. **(Bottom)** Localized plasticity for both the naïve and the matured 48G7 antibody, respectively.

conformational diversity of the CDR-H3 loop of the 48G7 Fab upon affinity maturation, but clearly reveals a population shift toward the dominant solution structure in the affinity matured Fab. Astonishingly, the dominant CDR-H3 loop conformation of the matured 48G7 Fab is present as a local shallow side-minimum in the free energy surface of the naïve 48G7 Fab. These findings supported the hypothesis that promiscuity might arise from numerous weakly populated conformations each of which is able to bind different binding partners (Zhou et al., 2007; Adhikary et al., 2015; Gunti and Notkins, 2015; Fernández-Quintero et al., 2019b). These probabilities are then shifted toward a smaller number of states which results in a reduction of possible binding partners.

Again, **Figure 6C** shows the reduced structural ensemble of the CDR-H3 loop upon affinity maturation. **Figure 7** characterizes the plasticity of the naïve and matured 48G7 Fabs and reflects the substantial rigidification of the CDR-H3 loop in the matured Fab.

In all in detailed investigated antibody fragments we observe a significant decrease in flexibility and plasticity upon affinity maturation accompanied by strong population shifts toward the binding competent state. The free energy surfaces of the CDR-H3 loop do not only show a reduction in conformational space, but also reveal a smaller number CDR-H3 loop conformational states in solution. This is reduction in conformational diversity is reflected by narrower and deeper minima, while the naïve antibodies have broader and shallower free energy landscapes. Thus, germline antibodies–before maturation–are able to still adopt various distinct conformations, each of which is able to recognize different antigens.

CONCLUSION

For 10 pairs of antibodies we observed a substantial rigidification in flexibility and plasticity upon affinity maturation, in particular for the CDR-H3 loop. Molecular plasticity plays a crucial role in all processes involving molecular recognition. In our manuscript we present for the first time a method to quantify and localize plasticity on an atomistic level. We show that this method is in excellent agreement with 3PEPS spectroscopy. Additionally, we employ this new method to affinity maturation of antibodies, showing for 10 pairs of antibodies, that affinity maturation goes hand in hand with a reduction of plasticity and flexibility. As our method allows for localization, we are even able to identify the areas of reduced plasticity. Consequently, we are able to show that for all 10 pairs of antibodies of different CDR-H3 loop lengths specificity is linked to rigidity. For all affinity maturation studies kinetics and thermodynamics were reconstructed and revealed for the naïve Fabs broader and shallower free energy surfaces, while the matured Fabs showed small and distinct minima. All studied affinity maturation Fabs follow the paradigm of conformational selection, because even without the antigen present the binding competent state is present in solution. In summary, we do not only demonstrate a generalizable method to characterize and localize molecular

plasticity in detail, but we also strongly link it to a general principle in antibody-antigen recognition.

DATA AVAILABILITY STATEMENT

All datasets presented in this study are included in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

MF-Q, JL, LB, and KL conceived the study. MF-Q, JL, LB, and FW performed the research under the supervision of KL, and curated and analyzed the data. CS contributed to the study in visualizing and analyzing the data. MF-Q drafted the manuscript. All authors critically reviewed the manuscript.

FUNDING

This work was supported by the Austrian Science Fund (FWF) via the grants P30565 and P30737.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2020.00182/full#supplementary-material

REFERENCES

Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., et al. (2015). GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1–2, 19–25. doi: 10.1016/j.softx.2015.06.001

Acierno, J. P., Braden, B. C., Klinke, S., Goldbaum, F. A., and Cauerhff, A. (2007). Affinity maturation increases the stability and plasticity of the Fv domain of anti-protein antibodies. *J. Mol. Biol.* 374, 130–146. doi: 10.1016/j.jmb.2007.09.005

Adelman, S. A., and Doll, J. D. (1976). Generalized langevin equation approach for atom/solid-surface scattering: general formulation for classical scattering off harmonic solids. *J. Chem. Phys.* 64, 2375–2388.

Adhikary, R., Yu, W., Oda, M., Walker, R. C., Chen, T., Stanfield, R. L., et al. (2015). Adaptive mutations alter antibody structure and dynamics during affinity maturation. *Biochemistry* 54, 2085–2093. doi: 10.1021/bi501417q

Adhikary, R., Yu, W., Oda, M., Zimmermann, J., and Romesberg, F. E. (2012). Protein dynamics and the diversity of an antibody response. *J. Biol. Chem.* 287, 27139–27147. doi: 10.1074/jbc.m112.372698

Akiba, H., and Tsumoto, K. (2015). Thermodynamics of antibody–antigen interaction revealed by mutation analysis of antibody variable regions. *J. Biochem. (Tokyo)* 158, 1–13. doi: 10.1093/jb/mvv049

Alba, J., Rienzo, L. D., Milanetti, E., Acuto, O., and D'Abramo, M. (2020). Molecular dynamics simulations reveal canonical conformations in different pMHC/TCR interactions. *Cells* 9:942. doi: 10.3390/cells9040942

Al-Lazikani, B., Lesk, A. M., and Chothia, C. (1997). Standard conformations for the canonical structures of immunoglobulins1. *J. Mol. Biol.* 273, 927–948. doi: 10.1006/jmbi.1997.1354

Alzari, P. M., Spinelli, S., Mariuzza, R. A., Boulot, G., Poljak, R. J., Jarvis, J. M., et al. (1990). Three-dimensional structure determination of an anti-2-phenyloxazolone antibody: the role of somatic mutation and heavy/light chain pairing in the maturation of an immune response. *EMBO J.* 9, 3807–3814. doi: 10.1002/j.1460-2075.1990.tb07598.x

Babor, M., and Kortemme, T. (2009). Multi-constraint computational design suggests that native sequences of germline antibody H3 loops are nearly optimal for conformational flexibility. *Proteins* 75, 846–858. doi: 10.1002/prot.22293

Barducci, A., Bonomi, M., and Parrinello, M. (2011). Metadynamics. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 1, 826–843.

Barducci, A., Bussi, G., and Parrinello, M. (2008). Well-tempered metadynamics: a smoothly converging and tunable free-energy method. *Phys. Rev. Lett.* 100:020603.

Bassing, C. H., Swat, W., and Alt, F. W. (2002). The mechanism and regulation of chromosomal V(D)J recombination. *Cell* 109, S45–S55.

Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., and Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81, 3684–3690. doi: 10.1063/1.448118

Biswas, M., Lickert, B., and Stock, G. (2018). Metadynamics enhanced markov modeling of protein dynamics. *J. Phys. Chem. B* 122, 5508–5514. doi: 10.1021/acs.jpcb.7b11800

Braden, B. C., Souchon, H., Eiselé, J.-L., Bentley, G. A., Bhat, T. N., Navaza, J., et al. (1994). Three-dimensional structures of the free and the antigen-complexed Fab from monoclonal anti-lysozyme antibody D44.1. *J. Mol. Biol.* 243, 767–781. doi: 10.1016/0022-2836(94)90046-9

Burkovitz, A., Sela-Culang, I., and Ofran, Y. (2014). Large-scale analysis of somatic hypermutations in antibodies reveals which structural regions, positions and amino acids are modified to improve affinity. *FEBS J.* 281, 306–319. doi: 10.1111/febs.12597

Carter, P. J. (2006). Potent antibody therapeutics by design. *Nat. Rev. Immunol.* 6:nri1837.

Carter, P. J. (2011). Introduction to current and future protein therapeutics: a protein engineering perspective. *Exp Cell Res.* 317, 1261–1269. doi: 10.1016/j.yexcr.2011.02.013

Case, D. A., Betz, R. M., Cerutti, D. S., Cheatham, T. E., Darden, T. A., and Duke, R. E. (2016). *AMBER 2016*. San Francisco, CA: University of California, 2016.

Cauerhff, A., Goldbaum, F. A., and Braden, B. C. (2004). Structural mechanism for affinity maturation of an anti-lysozyme antibody. *Proc. Natl. Acad. Sci. U.S.A.* 101:3539. doi: 10.1073/pnas.0400060101

Chodera, J. D., and Noé, F. (2014). Markov state models of biomolecular conformational dynamics. *Curr. Opin. Struct. Biol.* 25, 135–144. doi: 10.1016/j.sbi.2014.04.002

Chong, L. T., Duan, Y., Wang, L., Massova, I., and Kollman, P. A. (1999). Molecular dynamics and free-energy calculations applied to affinity maturation in antibody 48G7. *Proc. Natl. Acad. Sci. U.S.A.* 96, 14330–14335. doi: 10.1073/pnas.96.25.14330

Chothia, C., and Lesk, A. M. (1987). Canonical structures for the hypervariable regions of immunoglobulins. *J. Mol. Biol.* 196, 901–917. doi: 10.1016/0022-2836(87)90412-8

Clark, L. A., Ganesan, S., Papp, S., and van Vlijmen, H. W. T. (2006). Trends in antibody sequence changes during the somatic hypermutation process. *J. Immunol.* 177, 333. doi: 10.4049/jimmunol.177.1.333

Csermely, P., Palotai, R., and Nussinov, R. (2010). Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. *Trends Biochem. Sci.* 35, 539–546. doi: 10.1016/j.tibs.2010.04.009

Daberdaku, S., and Ferrari, C. (2018). Antibody interface prediction with 3D Zernike descriptors and SVM. *Bioinformatics* 35, 1870–1876. doi: 10.1093/bioinformatics/bty918

DeKosky, B. J., Lungu, O. I., Park, D., Johnson, E. L., Charab, W., Chrysostomou, C., et al. (2016). Large-scale sequence and structural comparisons of human naive and antigen-experienced antibody repertoires. *Proc. Natl. Acad. Sci. U.S.A.* 113, E2636.

Edelman, G. M. (1973). Antibody structure and molecular immunology. *Science* 180, 830–840. doi: 10.1126/science.180.4088.830

Fernández-Quintero, M. L., Heiss, M. C., and Liedl, K. R. (2020a). Antibody humanization—the Influence of the antibody framework on the CDR-H3 loop

ensemble in solution. *Protein Eng. Des. Sel.* 32, 411–422. doi: 10.1093/protein/gzaa004

Fernández-Quintero, M. L., Heiss, M. C., Pomarici, N. D., Math, B. A., and Liedl, K. R. (2020b). Antibody CDR loops as ensembles in solution vs. canonical clusters from X-ray structures. *mAbs* 12, 1744328. doi: 10.1080/19420862.2020.1744328

Fernández-Quintero, M. L., Seidler, C. A., and Liedl, K. R. (2020c). T-Cell receptor variable β domains rigidify during affinity maturation. *Sci. Rep.* 10:4472.

Fernández-Quintero, M. L., Kraml, J., Georges, G., and Liedl, K. R. (2019a). CDR-H3 loop ensemble in solution – Conformational selection upon antibody binding. *mAbs* 11, 1077–1088. doi: 10.1080/19420862.2019.1618676

Fernández-Quintero, M. L., Loeffler, J. R., Kraml, J., Kahler, U., Kamenik, A. S., and Liedl, K. R. (2019b). Characterizing the diversity of the CDR-H3 loop conformational ensembles in relationship to antibody binding properties. *Front. Immunol.* 9:3065. doi: 10.3389/fimmu.2018.03065

Foote, J., and Milstein, C. (1991). Kinetic maturation of an immune response. *Nature* 352, 530–532. doi: 10.1038/352530a0

Foote, J., and Milstein, C. (1994). Conformational isomerism and the diversity of antibodies. *Proc. Natl. Acad. Sci. U.S.A.* 91, 10370–10374. doi: 10.1073/pnas.91.22.10370

French, D., Laskov, R., and Scharff, M. (1989). The role of somatic hypermutation in the generation of antibody diversity. *Science* 244, 1152. doi: 10.1126/science.2658060

Gunti, S., and Notkins, A. L. (2015). Polyreactive antibodies: function and quantification. *J. Infect. Dis.* 212, 42–46.

Hsieh-Wilson, L. C., Schultz, P. G., and Stevens, R. C. (1996). Insights into antibody catalysis: structure of an oxygenation catalyst at 1.9-angstrom resolution. *Proc. Natl. Acad. Sci. U.S.A.* 93, 5363. doi: 10.1073/pnas.93.11.5363

Hub, J. S., de Groot, B. L., Grubmüller, H., and Groenhof, G. (2014). Quantifying artifacts in ewald simulations of inhomogeneous systems with a net charge. *J. Chem. Theory Comput.* 10, 381–390. doi: 10.1021/ct400626b

James, L. C., Roversi, P., and Tawfik, D. S. (2003). Antibody multispecificity mediated by conformational diversity. *Science* 299, 1362–1367. doi: 10.1126/science.1079731

James, L. C., and Tawfik, D. S. (2003). Conformational diversity and protein evolution – a 60-year-old hypothesis revisited. *Trends Biochem. Sci.* 28, 361–368. doi: 10.1016/s0968-0004(03)00135-x

James, L. C., and Tawfik, D. S. (2005). Structure and kinetics of a transient antibody binding intermediate reveal a kinetic discrimination mechanism in antigen recognition. *Proc. Natl. Acad. Sci. U.S.A.* 102, 12730. doi: 10.1073/pnas.0500909102

Jeliazkov, J. R., Sljoka, A., Kuroda, D., Tsuchimura, N., Katoh, N., Tsumoto, K., et al. (2018). Repertoire analysis of antibody CDR-H3 loops suggests affinity maturation does not typically result in rigidification. *Front. Immunol.* 9:413. doi: 10.3389/fimmu.2018.00413

Jespersen, M. C., Mahajan, S., Peters, B., Nielsen, M., and Marcatili, P. (2019). Antibody specific B-Cell epitope predictions: leveraging information from antibody-antigen protein complexes. *Front. Immunol.* 10:298. doi: 10.3389/fimmu.2019.00298

Jimenez, R., Salazar, G., Baldridge, K. K., and Romesberg, F. E. (2003). Flexibility and molecular recognition in the immune system. *Proc. Natl. Acad. Sci. U.S.A.* 100, 92–97.

Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79, 926–935. doi: 10.1063/1.445869

Kaplon, H., and Reichert, J. M. (2019). Antibodies to watch in 2019. *mAbs* 11, 219–238. doi: 10.1080/19420862.2018.1556465

Karush, J. (1961). On the chapman-kolmogorov equation. *Ann. Math. Stat.* 32, 1333–1337.

Labute, P. (2009). Protonate3D: assignment of ionization states and hydrogen coordinates to macromolecular structures. *Proteins* 75, 187–205. doi: 10.1002/prot.22234

Li, T., Tracka, M. B., Uddin, S., Casas-Finet, J., Jacobs, D. J., and Livesay, D. R. (2015). Rigidity emerges during antibody evolution in three distinct antibody systems: evidence from QSFR analysis of Fab fragments. *PLoS Comput. Biol.* 11:e1004327. doi: 10.1371/journal.pcbi.1004327

Li, Y., Li, H., Yang, F., Smith-Gill, S. J., and Mariuzza, R. A. (2003). X-ray snapshots of the maturation of an antibody response to a protein antigen. *Nat. Struct. Mol. Biol.* 10, 482–488. doi: 10.1038/nsb930

Likas, A., Vlassis, N., and Verbeek, J. (2003). The global k-means clustering algorithm. *Biometrics* 36, 451–461. doi: 10.1016/s0031-3203(02)00060-2

Ma, B., Kumar, S., Tsai, C.-J., and Nussinov, R. (1999). Folding funnels and binding mechanisms. *Protein Eng. Des. Sel.* 12, 713–720. doi: 10.1093/protein/12.9.713

Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., and Simmerling, C. (2015). ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* 11, 3696–3713. doi: 10.1021/acs.jctc.5b00255

Manivel, V., Sahoo, N. C., Salunke, D. M., and Rao, K. V. S. (2000). Maturation of an antibody response is governed by modulations in flexibility of the antigen-combining site. *Immunity* 13, 611–620. doi: 10.1016/s1074-7613(00)00061-3

Market, E., and Papavasiliou, F. N. (2003). V(D)J recombination and the evolution of the adaptive immune system. *PLoS Biol.* 1:e16. doi: 10.1371/journal.pbio.0000016

Marks, C., and Deane, C. M. (2017). Antibody H3 structure prediction. *Comput. Struct. Biotechnol. J.* 15, 222–231. doi: 10.1016/j.csbj.2017.01.010

Milstein, C. (1991). Affinity maturation of antibodies. *Immunol. Today* 12, 93. doi: 10.1016/0167-5699(91)90164-o

Miroshin, R. N. (2016). Special solutions of the Chapman–Kolmogorov equation for multidimensional-state Markov processes with continuous time. *Vestn. St Petersburg Univ. Math* 49, 122–129. doi: 10.3103/s1063454116020114

Mishra, A. K., and Mariuzza, R. A. (2018). Insights into the structural basis of antibody affinity maturation from next-generation sequencing. *Front. Immunol.* 9:117. doi: 10.3389/fimmu.2018.00117

Miyamoto, S., and Kollman, P. A. (1992). Settle: an analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* 13, 952–962. doi: 10.1002/jcc.540130805

Molecular Operating Environment [MOE] (2018). *Chemical Computing Group, Version 2018.01*. Montreal, QC: Chemical Computing Group.

Nguyen, M. N., Pradhan, M. R., Verma, C., and Zhong, P. (2017). The interfacial character of antibody paratopes: analysis of antibody–antigen structures. *Bioinformatics* 33, 2971–2976. doi: 10.1093/bioinformatics/btx389

Oh, M. H. J., Salvador, M. R., Wong, C. Y., and Scholes, G. D. (2011). Three-pulse photon-echo peak shift spectroscopy and its application for the study of solvation and nanoscale excitons. *ChemPhysChem* 12, 88–100. doi: 10.1002/cphc.201000712

Pérez-Hernández, G., and Noé, F. (2016). Hierarchical time-lagged independent component analysis: computing slow modes and reaction coordinates for large molecular systems. *J. Chem. Theory Comput.* 12, 6118–6129. doi: 10.1021/acs.jctc.6b00738

Pronk, S., Páll, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., et al. (2013). GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* 29, 845–854. doi: 10.1093/bioinformatics/btt055

Ramachandran, G. N., Ramakrishnan, C., and Sasisekharan, V. (1963). Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.* 7, 95–99. doi: 10.1016/s0022-2836(63)80023-6

Regep, C., Georges, G., Shi, J., Popovic, B., and Deane, C. M. (2017). The H3 loop of antibodies shows unique structural characteristics. *Proteins Struct. Funct. Bioinform.* 85, 1311–1318. doi: 10.1002/prot.25291

Reichert, J. M. (2017). Antibodies to watch in. *mAbs* 2016, 167–181. doi: 10.1080/19420862.2016.1269580

Röblitz, S., and Weber, M. (2013). Fuzzy spectral clustering by PCCA+: application to Markov state models and data classification. *Adv. Data Anal. Classif.* 7, 147–179. doi: 10.1007/s11634-013-0134-6

Roe, D. R., and Cheatham, T. E. (2013). PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.* 9, 3084–3095. doi: 10.1021/ct400341p

Salomon-Ferrer, R., Götz, A. W., Poole, D., Le Grand, S., and Walker, R. C. (2013). Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh ewald. *J. Chem. Theory Comput.* 9, 3878–3888. doi: 10.1021/ct400314y

Scherer, M. K., Trendelkamp-Schroer, B., Paul, F., Pérez-Hernández, G., Hoffmann, M., Plattner, N., et al. (2015). PyEMMA 2: a software

package for estimation, validation, and analysis of markov models. *J. Chem. Theory Comput.* 11, 5525–5542. doi: 10.1021/acs.jctc.5b00743

Schiele, F., van Ryn, J., Litzenburger, T., Ritter, M., Seeliger, D., and Nar, H. (2015). Structure-guided residence time optimization of a dabigatran reversal agent. *mAbs* 7, 871–880. doi: 10.1080/19420862.2015.1057364

Schmidt, A. G., Xu, H., Khan, A. R., O'Donnell, T., Khurana, S., King, L. R., et al. (2013). Preconfiguration of the antigen-binding site during affinity maturation of a broadly neutralizing influenza virus antibody. *Proc. Natl. Acad. Sci. U.S.A.* 110:264. doi: 10.1073/pnas.1218256109

Shao, J., Tanner, S. W., Thompson, N., and Cheatham, T. E. (2007). Clustering molecular dynamics trajectories: 1. Characterizing the performance of different clustering algorithms. *J. Chem. Theory Comput.* 3, 2312–2334. doi: 10.1021/ct700119m

Shehata, L., Maurer, D. P., Wec, A. Z., Lilov, A., Champney, E., Sun, T., et al. (2019). Affinity maturation enhances antibody specificity but compromises conformational stability. *Cell Rep.* 28, 3300–3308.e4.

Thielges, M. C., Zimmermann, J., Yu, W., Oda, M., and Romesberg, F. E. (2008). Exploring the Energy landscape of antibody-antigen complexes: protein dynamics, flexibility, and molecular recognition. *Biochemistry* 47, 7237–7247. doi: 10.1021/bi800374q

Thorpe, I. F., and Brooks, C. L. (2007). Molecular evolution of affinity and flexibility in the immune system. *Proc. Natl. Acad. Sci. U.S.A.* 104, 8821. doi: 10.1073/pnas.0610064104

Tribello, G. A., Bonomi, M., Branduardi, D., Camilloni, C., and Bussi, G. (2014). PLUMED 2: new feathers for an old bird. *Comput. Phys. Commun.* 185, 604–613. doi: 10.1016/j.cpc.2013.09.018

Tsai, C.-J., Kumar, S., Ma, B., and Nussinov, R. (1999). Folding funnels, binding funnels, and protein function. *Protein Sci.* 8, 1181–1190. doi: 10.1110/ps.8.6.1181

Wallnoefer, H. G., Liedl, K. R., and Fox, T. (2011). A challenging system: free energy prediction for factor Xa. *J. Comput. Chem.* 32, 1743–1752. doi: 10.1002/jcc.21758

Wedemayer, G. J., Patten, P. A., Wang, L. H., Schultz, P. G., and Stevens, R. C. (1997). Structural insights into the evolution of an antibody combining site. *Science* 276:1665. doi: 10.1126/science.276.5319.1665

Willis, J. R., Briney, B. S., DeLuca, S. L., Crowe, J. E. Jr., and Meiler, J. (2013). Human germline antibody gene segments encode polyspecific antibodies. *PLoS Comput. Biol.* 9:e1003045. doi: 10.1371/journal.pcbi.1003045

Wong, S. E., Sellers, B. D., and Jacobson, M. P. (2011). Effects of somatic mutations on CDR loop flexibility during affinity maturation. *Proteins Struct. Funct. Bioinform.* 79, 821–829. doi: 10.1002/prot.22920

Wood, M. J., and Hirst, J. D. (2005). Protein secondary structure prediction with dihedral angles. *Proteins Struct. Funct. Bioinform.* 59, 476–481. doi: 10.1002/prot.20435

Wu, H., and Noé, F. (2017). Variational approach for learning Markov processes from time series data. *J. Nonlinear Sci.* 30, 23–66. doi: 10.1007/s00332-019-09567-y

Yin, J., Beuscher, A. E., Andryski, S. E., Stevens, R. C., and Schultz, P. G. (2003). Structural plasticity and the evolution of antibody affinity and specificity. *J. Mol. Biol.* 330, 651–656. doi: 10.1016/s0022-2836(03)00631-4

Yin, J., Mundorff, E. C., Yang, P. L., Wendt, K. U., Hanway, D., Stevens, R. C., et al. (2001). A comparative analysis of the immunological evolution of antibody 28B4. *Biochemistry* 40, 10764–10773. doi: 10.1021/bi010536c

Zhou, Z.-H., Tzioufas, A. G., and Notkins, A. L. (2007). Properties and function of polyreactive antibodies and polyreactive antigen-binding B cells. *J. Autoimmun.* 29, 219–228. doi: 10.1016/j.jaut.2007.07.015

Zimmermann, J., Oakman, E. L., Thorpe, I. F., Shi, X., Abbyad, P., Brooks, C. L., et al. (2006). Antibody evolution constrains conformational heterogeneity by tailoring protein dynamics. *Proc. Natl. Acad. Sci. U.S.A.* 103, 13722–13727. doi: 10.1073/pnas.0603282103

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Molecular Biosciences

ORIGINAL RESEARCH
published: 27 October 2020
doi: 10.3389/fmolb.2020.589718

EspcTM: Kinetic Transition Network Based on Trajectory Mapping in Effective Energy Rescaling Space

*Zhenyu Wang[1], Xin Zhou[2] and Guanghong Zuo[1]**

[1] T-Life Research Center, State Key Laboratory of Surface Physics, Department of Physics, Fudan University, Shanghai, China, [2] School of Physical Sciences, University of Chinese Academy of Sciences, Beijing, China

The transition network provides a key to reveal the thermodynamic and kinetic properties of biomolecular systems. In this paper, we introduce a new method, named effective energy rescaling space trajectory mapping (EspcTM), to detect metastable states and construct transition networks based on the simulation trajectories of the complex biomolecular system. It mapped simulation trajectories into an orthogonal function space, whose bases were rescaled by effective energy, and clustered the interrelation between these trajectories to locate metastable states. By using the EspcTM method, we identified the metastable states and elucidated interstate transition kinetics of a Brownian particle and a dodecapeptide. It was found that the scaling parameters of effective energy also provided a clue to the dominating factors in dynamics. We believe that the EspcTM method is a useful tool for the studies of dynamics of the complex system and may provide new insight into the understanding of thermodynamics and kinetics of biomolecular systems.

Keywords: effective energy, molecular dynamics, trajectory mapping, Markov models, alanine dodecapeptide, transition network

OPEN ACCESS

Edited by:
Wenfei Li,
Nanjing University, China

Reviewed by:
Naoto Hori,
University of Nottingham, United Kingdom
Xingcheng Lin,
Massachusetts Institute of Technology, United States
Zhang Zhi Yong,
University of Science and Technology of China, China

****Correspondence:***
Guanghong Zuo
ghzuo@fudan.edu.cn

Specialty section:
This article was submitted to Biological Modeling and Simulation, a section of the journal Frontiers in Molecular Biosciences

Received: *31 July 2020*
Accepted: *24 September 2020*
Published: *27 October 2020*

Citation:
Wang Z, Zhou X and Zuo G (2020) EspcTM: Kinetic Transition Network Based on Trajectory Mapping in Effective Energy Rescaling Space. Front. Mol. Biosci. 7:589718. doi: 10.3389/fmolb.2020.589718

INTRODUCTION

The biomolecules are fundamentally dynamic in nature (Chodera et al., 2007). Protein folding, for example, involves the conformation change from polypeptide chain to a particular tertiary topology over microseconds to seconds, a process that can go awry and lead to misfolding and cause disease (Chiti and Dobson, 2006; Gregersen et al., 2006; Chodera et al., 2007; Guo et al., 2012; Wei et al., 2016; Zhou et al., 2019). Allosteric enzyme catalysis involves transitions between multiple conformational substates, only a few of which may allow substate access or catalysis (Eisenmesser et al., 2002; Boehr et al., 2006; Buch et al., 2011). Protein–ligand binding may alter the transition kinetics among multiple conformational states; for example, intrinsically disordered protein may have structured and unstructured binding pathways (Ithuralde et al., 2016; Paul et al., 2017; Li et al., 2019; Pan et al., 2019; Weng and Wang, 2020). Understanding of biomolecular dynamics is pivotal to reveal the function of biomolecules. Computer simulations of biomolecules, which made the biomolecular dynamics visible *in silico*, provide valuable insight for understanding how the dynamics of biomolecules drives biology processes (Cheatham and Kollman, 2000;

Mirny and Shakhnovich, 2001; Norberg and Nilsson, 2002; Moraitakis et al., 2003; Levy et al., 2004; Zhou et al., 2004; Gao et al., 2005; Zuo et al., 2006, 2009; Li et al., 2008, 2013; Miyashita et al., 2009; Yang et al., 2014; Yan and Wang, 2019; Wu et al., 2020). In particular, molecular dynamics (MD) simulations can provide atomic-level details that are not always accessible in experiments and make this technique inevitable (Karplus and McCammon, 2002; Adcock and McCammon, 2006; Wang et al., 2009; Zuo et al., 2013). However, too many details will disguise the meaningful information. In most cases, the functional processes of biomolecules, the most interesting or important processes, correspond to slow dynamical processes. To extract these processes from numerous MD simulation trajectories, much effort has been involved in the development of methods for massive high-dimensional simulation data analysis. It was now well established from a variety of studies that an intelligible picture of the dynamics of biomolecules can be described as a transition network between several metastable states based on the simulation trajectories (Zwanzig, 1983; Kampen, 2007).

Markov state model (MSM) provides a powerful framework for analyzing dynamics of biosystems, such as MD simulations, to construct a transition network of metastable states. It has gained widespread use over the past several decades (Chodera et al., 2007; Gfeller et al., 2007; Noe et al., 2007; Bowman and Pande, 2010; Pande et al., 2010; Rao and Karplus, 2010; Bowman et al., 2013; Deng et al., 2013; Weber et al., 2013; Husic and Pande, 2018; Wang et al., 2018; Sengupta et al., 2019). In the analyzing process of MSM, the simulation conformations were first classified into thousands of small groups, named as microstates, by a geometric clustering method wherein these conformations were similar in geometry (Bowman et al., 2009; Pande et al., 2010). These microstates would be further clustered into several macrostates by standard spectral clustering method based on their transition frequency (Deuflhard and Weber, 2005; Chodera et al., 2007; Gfeller et al., 2007; Noe et al., 2007; Noe, 2008; Bowman and Pande, 2010; Pande et al., 2010; Rao and Karplus, 2010; Zuo et al., 2010; Bowman et al., 2013; Deng et al., 2013; Roblitz and Weber, 2013; Weber et al., 2013; Husic and Pande, 2018; Wang et al., 2018; Sengupta et al., 2019). Then, the transition network between the macrostates was reconstructed accordingly (Jayachandran et al., 2006; Buchete and Hummer, 2008; Prinz et al., 2011). Gong and Zhou (2010) presented the trajectory mapping (TM) method to construct a kinetic transition network of metastable states. Compared with MSM, TM grouped simulation trajectory pieces rather than individual conformations. They mapped the averaged conformation of each MD trajectory segment as a vector and calculate the principal components (PCs) of the trajectory-mapped vectors by the principal component analysis (PCA). The similar trajectory-mapped vectors were then grouped as metastable states by spectral clustering method, and transition events in simulation trajectories were further identified (Gong et al., 2015; Zhang et al., 2017; Zhang et al., 2019a; Zhang et al., 2019b).

In both MSM and TM methods, the discretization of MD trajectories, i.e., clustering of structures, plays a vital role in the analysis of MD trajectories. To make clustering of structures as accurate as possible, a variety of structural metrics and their functions were employed in analysis, for example, the torsion angles of backbone, the proportion of native contacts, root mean square deviation, and solvated energy (Gong et al., 2015). These analyses can be effective when all input coordinates are sufficient and irrelevant to each other. Thus, PCA was used to find orthogonal collective coordinates, which are linear combinations of the input coordinates and covered most of variances with only the first several eigenvectors (Lever et al., 2017). However, as mentioned above, the slow dynamical process is the concerned part in most cases. It is not always true that the high variance directions correspond to the kinetically slow-motion mode. Thus, some methods have been developed to obtain slow-motion directions. In the MSM, time-structure based Independent Correlation Analysis (tICA) was used (Naritomi and Fuchigami, 2011, 2013; Perez-Hernandez et al., 2013; Schwantes and Pande, 2013). It finds the slow collective coordinates by eigen-decomposition of a Δt-interval autocorrelation matrix. In the TM, the averaged conformation of every τ-length MD trajectory segment was mapped as a vector in feature space to compose samples for the PCA method. It was argued that fast conformational fluctuations were suppressed after the segment averaging, and the PCs mainly involve slow motions (Zhang et al., 2017). In both tICA-MSM and TM methods, a hyper-parameter, Δt for tICA-MSM and τ for TM, is required. It is difficult for inexperienced users. It is possible to obtain the optimized model by an automated process instead of a process of trial and error. For example, one might consider weighting the input coordination by an order parameter relevant to the functional processes of biomolecules, so that the input coordinates with high correlation contribute the most to the distance calculation and make the clustering effective and efficient to catch the functional processes, i.e., slow-motion patterns of the biomolecular system.

In this paper, we will present a new method, named effective energy rescaling space trajectory mapping (EspcTM), for detecting metastable states and constructing transition networks. It is a parameter-free analysis framework based on the previous TM method. In the EspcTM method, every snapshot of the trajectories was described by a high-dimensional vector and mapped into an orthogonal functional space. Different from the TM method, the features were rescaled by the effective energy of the dynamics to make the space effective to describe the slow processes of the system, and no hyperparameter was required. Here, the effective energy, which was filtered from the total potential energy of simulation trajectories by fast Fourier transform (FFT) and multiple linear regression, is an efficacious order parameter to describe the slow conformational change of complex system. The PCA method was also employed for dimensionality reduction and orthogonalization of the functional space. The metastable states were assigned by a spectral clustering method based on projections of the trajectories in this feature space. Then, the Markov transition matrix is constructed based on the transitions between these metastable states. We show application of this method by the movement of a Brownian particle and conformational dynamics of an alanine dodecapeptide (Ala_{12}). It revealed their metastable states and

kinetic transition network, as well as provided additional insight into the dynamics of these two systems.

THEORY AND METHOD

The EspcTM method is an analysis framework to identify metastable states from simulation data in the effective energy rescaling space and construct the transition network between the states based on the theory of Markov chain. In the EspcTM, an ordered parameter, named effective energy, was introduced to rescale feature space of the system. The simulation trajectories were mapped into the space and discretized to obtain the kinetic transition network of the system based on Markov chain theory. **Figure 1** shows the flow chart of the EspcTM method, and details of the key steps are followed.

Feature Extraction

In our study, there were N_t frames in every trajectory. They were mapped into a space consisting of N_b basis functions $\{\hat{A}^{\mu}(\vec{q})\}_{\mu=1,\ldots,N_b}$. To eliminate the effect of various units of basis function, normalization was performed on every dimension. Then, every trajectory was described as an $N_t \times N_b$-dimension matrix in the feature space, i.e., feature matrix

$$V = \left(\hat{A}^1(\vec{q}), \hat{A}^2(\vec{q}), \hat{A}^3(\vec{q}), \ldots, \hat{A}^{N_b}(\vec{q})\right) \tag{1}$$

where $\vec{q}$ denotes the structural metrics, such as the torsion angle of backbone in peptide. Here, the basis functions $\{\hat{A}^{\mu}(\vec{q})\}_{\mu=1,\ldots,N_b}$ should be chosen to identify typical conformational motions of systems. In this work, we used the sine and cosine of structural metrics as the feature space (Gong and Zhou, 2010; Gong et al., 2015).

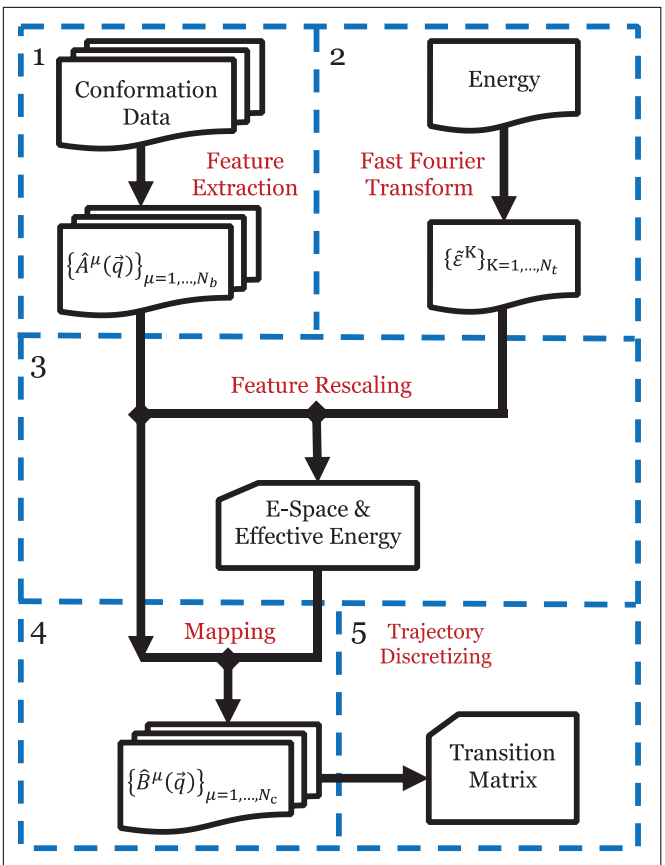


FIGURE 1 | Flow chart of EspcTM method. Step 1: Extracting the conformational metrics with a set of basis functions for all simulation trajectories. Step 2: Extracting the potential energy to $\{\tilde{\varepsilon}^{K}\}_{K=1,\ldots,N_t}$ by fast Fourier transform. Step 3: Multiple linear regression $\tilde{\varepsilon}^{K}$ and features, obtaining effective energy and E-space. Step 4: Mapping all trajectories to E-space. Step 5: Discretizing the trajectories based on the projections in E-space, and calculating the Markov transition matrix.

Noise Reduction

It is obvious that every basis possesses different weight on describing the dynamics of complex system. It was argued that dynamics of complex systems, such as protein folding, can resemble a diffusive process on a rugged landscape of free energy (Onuchic et al., 1997). Thus, energy is an appropriate measure to rescale their coordinates. Most studies of complex system focus on the dynamics of a part of the system, and the rest of the system was regarded as the environment of the study object. For example, studies on protein folding focus on protein molecules. The conformational change of protein in protein folding is the interesting part, instead of the fluctuation of water molecules. However, the atoms of the system interacted with each other in a complicated way. The energy variation caused by the dynamics of the studied object is coupled with the energy caused by the fluctuation of the remaining part. It is difficult to isolate the meaningful energy in a frame without additional hypotheses. On the other hand, as mentioned above, the kinetic slowness is the main character of the interesting processes. Therefore, the dynamics of the important processes can be separated from the fluctuation in the frequency domain, where slow motion is treated as low-frequency signal and fluctuation can be filtered out as high-frequency noise.

In this work, FFT (Cochran et al., 1967) was applied to transform the energy of trajectories into frequency space. For every trajectory, the coefficients of frequencies were obtained by

$$\tilde{\omega}_k = \sum_{n=0}^{N_t-1} \varepsilon_n \cdot e^{-in\omega_k} \tag{2}$$

Here, $i = \sqrt{-1}$ is the imaginary mark, n is the index of frames for the trajectory, ε_n is the total potential energy of the nth frame obtained from the simulation data, N_t is the number of frames of a trajectory, and $\omega_k = 2\pi k/N_t$ corresponds to a frequency. To reduce the false edge, even extension was used before FFT for every trajectory. Then, a reverse FFT was performed on the first K frequencies for every trajectory to obtain the $\tilde{\varepsilon}^{K}$ of every frame:

$$\tilde{\varepsilon}_n^{K} = \sum_{k=0}^{K-1} \tilde{\omega}_k \cdot e^{in\omega_k} \tag{3}$$

The fluctuation whose $\omega \geq \omega_K$ was excluded in $\tilde{\varepsilon}^{K}$. To determine the number K, we performed multiple linear regression

(Schneider et al., 2010) between K-energy vector $\tilde{\varepsilon}^{\mathrm{K}}$ and feature matrix V for all trajectories:

$$\tilde{\varepsilon}^{\mathrm{K}} = a_0^{\mathrm{K}} + V \cdot \hat{a}^{\mathrm{K}} + \epsilon^{\mathrm{K}} \tag{4}$$

Here, a_0^{K} (scalar) and $\hat{a}^{\mathrm{K}}$ (N_b-dimensional vector) are the fitting parameters, and ϵ^{K} is the error for the multiple linear regression. The effective energy $\tilde{\varepsilon} = \tilde{\varepsilon}^{\mathrm{K}^*} - \epsilon^{\mathrm{K}^*}$ with the $\mathrm{K}^* = \arg\max\ r(K)$. Here, $r(K) = \sqrt{1 - \overline{(\epsilon^{\mathrm{K}})^2}/(\sigma^{\mathrm{K}})^2}$ is the multiple correlation coefficient, $(\sigma^{\mathrm{K}})^2$ is the variance of $\tilde{\varepsilon}^{\mathrm{K}}$, and $r = 0$ for the case $(\sigma^{\mathrm{K}})^2 = \overline{(\epsilon^{\mathrm{K}})^2} = 0$. For multiple trajectories, the FFT was performed on every trajectory separately. Due to same length and time interval of all trajectories in our study, all trajectories were mapped into the same frequency space $\{\omega_k\}_{k=1,\ldots,N_t}$. Thus, in the revised FFT, the K-energies of all trajectories are the summary of the same frequencies for every K. Before multiple linear regression, K-energy vectors $\tilde{\varepsilon}^{\mathrm{K}}$ and feature matrixes V of all trajectories were joined into a vector and a feature matrix for equation (4).

Feature Rescaling and Mapping

The regression coefficients $\hat{a}^{\mathrm{K}}$ were used as the weight factors on features. Every trajectory was described as a new $N_t \times N_b$-dimension matrix:

$$\tilde{V} = \mathrm{V} \cdot diag\left(\hat{a}^{\mathrm{K}}\right) \tag{5}$$

Here, $diag(\hat{a}^{\mathrm{K}})$ is an $N_b \times N_b$ diagonal matrix with the elements of $\hat{a}^{\mathrm{K}}$ on its main diagonal. A PCA (Sims et al., 2005) was applied to reduce the dimension and orthogonalize the components of all trajectories $\tilde{V}$. Descending according to eigenvalues, the first N_c eigenvectors were selected to consist of an $N_b \times N_c$ matrix M. Here, $N_c \ll N_b$, and M is the mapping operator, which reduced the $N_b - dimension$ vectors into $N_c - dimension$, given top N_c eigenvalues whose sum has over 90% fraction of the sum of all eigenvalues. Here, we named this $N_c - dimension$ space as E-space since its input coordinates were weighted by the regression coefficients. By using the mapping operator M, we mapped all original feature matrixes V_j into the E-space. Therefore, every frame of the trajectories was described as an $N_c - dimension$ vector $\{\hat{B}^{\mu}(\vec{q})\}_{\mu=1,\ldots,N_c}$.

Trajectory Discretizing

The clustering of conformations was performed in the E-space, i.e., based on the analysis of the projection vectors $\{\hat{B}^{\mu}(\vec{q})\}_{\mu=1,\ldots,N_c}$. Similar to the TM method (Gong and Zhou, 2010; Zhang et al., 2017), every trajectory was divided into a lot of isometric pieces, and the similarity between each two pieces was defined by their average vectors:

$$S\left(t, t'\right) = \frac{\sum_i [B_i(t)B_i(t') + 1]}{\sqrt{\sum_i [B_i(t)B_i(t) + 1]} \times \sqrt{\sum_i [B_i(t')B_i(t') + 1]}} \tag{6}$$

Here, we replaced the vectors of frames by the average vectors of trajectory pieces. It reduced the size of the similarity matrix and cost of computation resource. In practice, the length of the trajectory pieces can be varied in a reasonable range. The Robust Perron Cluster Analysis (PCCA+) method (Roblitz and Weber, 2013), implemented in pyEMMA (Scherer et al., 2015), was used to classify all pieces into N_s states based on the similarity matrix. Here, the number of states N_s was determined by the distribution of the eigenvalues of the similarity matrix (Roblitz and Weber, 2013). The Markov transition matrix P was obtained based on the discretized trajectories (Prinz et al., 2011). Since P is a row stochastic matrix, its largest left eigenvalue is 1. If there is a unique stationary distribution, it is true for our case, then the largest eigenvalue and the corresponding eigenvector is unique too. As the theory of stochastic process, the stationary distribution of the Markov process corresponds to the distribution of equilibrium state. More interestingly, the Markov transition matrix can also be used to reveal the dynamics of the system in non-equilibration conditions (Reuter et al., 2018).

Brownian Dynamic Simulation

For Brownian dynamic simulation, Brownian particles in the presence of a potential, U, are described by the Langevin equation

$$m\frac{dv\,(t)}{dt} = -\nabla U\,(x) - \gamma v\,(t) + \xi\,(t) \tag{7}$$

where $\xi(t)$ is a delta-correlated stationary Gaussian process with zero-mean. A two-dimensional Brownian particle was simulated on the surface with three potential wells in the toy model (see **Figure 2A**). Here, the potential $U(x)$ was defined as:

$$U\,(x) = -\varepsilon\left\{\cos\,(x) + \sin\,(x) + \frac{1}{2}\cos\left(y\right) + 2\cos\,(3x) + 2\exp\left[-20\left(x + \frac{2}{3}\pi\right)^2 - 2y^2\right]\right\} \tag{8}$$

with scaling parameter $\varepsilon = 40$. Multiple trajectories were generated from different initial sites randomly with extensive long simulations.

MD Simulation

In the MD simulation, the termini of Ala_{12} were charged, which leads to versatile metastable structures (Noe et al., 2007). All atoms were modeled by using Amber03 force field. The molecule was solvated in a rhombic dodecahedral periodic box with the distance between the solutes and box boundary at least 10 Å. The SPC water model was used for solvation (see **Figure 3A**). The MD simulations were performed using the Gromacs package 4.6.5 (Hess et al., 2008). In the simulations, the covalent bonds involving H atoms were constrained by the LINCS algorithm, which allowed a time step of 2 fs. The long-range electrostatic interactions were treated with the particle-mesh Ewald method (Darden et al., 1993) with a grid spacing of 1.6 Å. The cutoff for the van der Waals interaction was set to 10 Å. The previous trajectory performed at high temperature was equilibrated by MD simulations for 100 ps at a constant pressure of 1 bar and a temperature of 500 K using Berendsen coupling (Berendsen et al., 1984). Then, the production simulations were performed in NVT ensemble at 500 K for 100 ns. All 50 systems extracted from high-temperature simulation had been iterated 100 ns in NVT

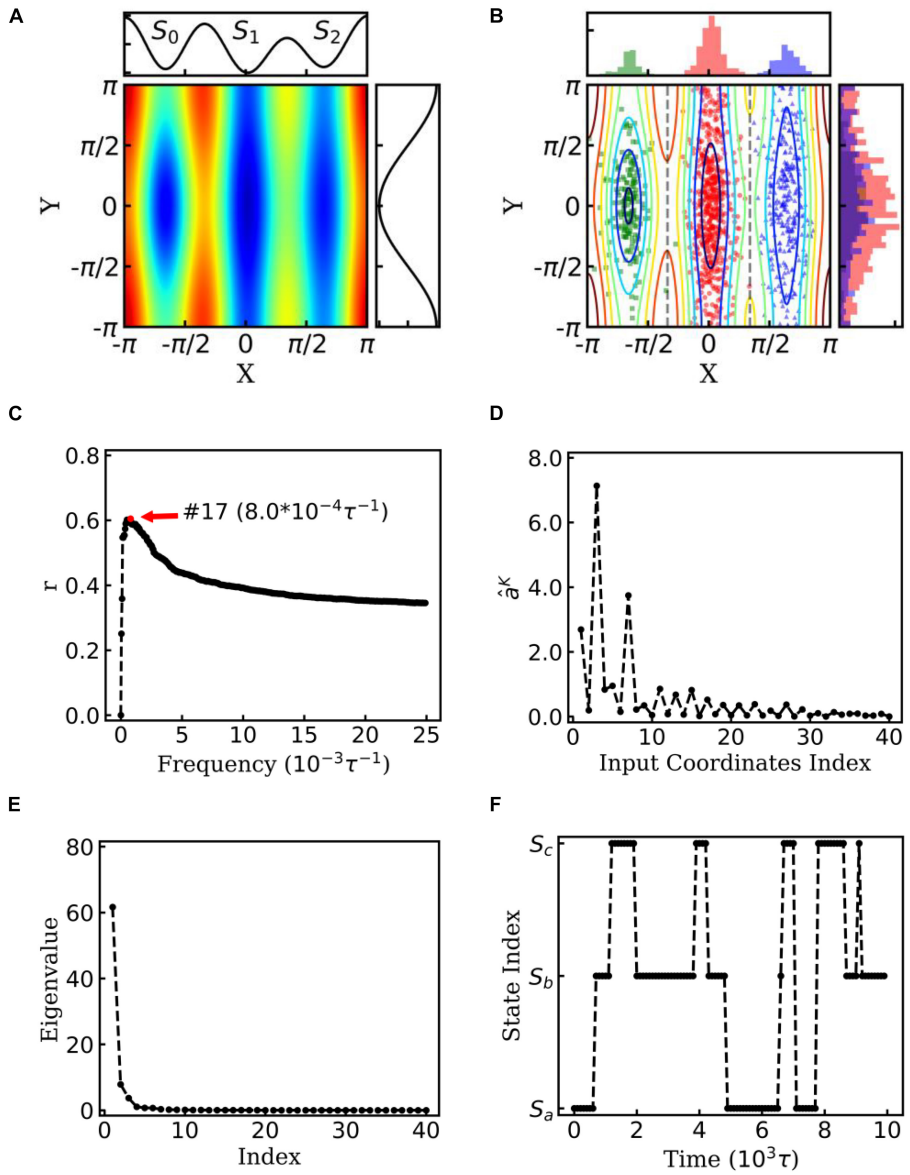


FIGURE 2 | EspcTM on dynamic of Brownian particle. **(A)** The energetic landscape of the toy model. Here, the potential function of the landscape was $-\varepsilon\{\cos(x) + \sin(x) + 2\cos(3x) + \frac{1}{2}\cos(y) + 2\exp[-20(x + \frac{2}{3}\pi)^2 - 2y^2]\}$. Three potential wells from left to right were S_0, S_1, and S_2. The well of S_1 was deeper than that of the other two states, and the barrier between S_0 and S_1 was much higher than that between S_1 and S_2. The black line on the top and right panel represents the potential along line $y = 0$ and $x = 0$, respectively. **(B)** Red, green, and blue dots represent three states of the snapshots of trajectories. The histograms of each state were shown on the top and right panel in different colors. **(C)** Multiple correlation coefficients of $\hat{\varepsilon}^K$ and all 40 conformational coordinates as a function of cutoff frequencies. Here, the maximum of the multiple correlation coefficient located at cutoff frequency equaling $8.0 \times 10^{-4}\tau^{-1}$. **(D)** The regression coefficients for all 40 features. The coordinates corresponding to basis functions $\sin(x)$, $\cos(x)$, and $\cos(2x)$ possessed large weights in the rescaling. **(E)** The eigenvalues in the PCA of trajectory-mapped vector. **(F)** A typical discretized trajectory.

ensemble at 300 K and recorded with time interval $\tau = 5$ ps. There are 20,000 × 50 frames in the analysis.

RESULTS AND DISCUSSION

The EspcTM method was first illustrated with a toy model, i.e., the dynamics of a Brownian particle on a two-dimensional surface. Then, it was applied to investigate the conformational dynamics of alanine dodecapeptide (Ala_{12}), and a transition network between metastable states of Ala_{12} was constructed.

Toy Model

In the toy model, a two-dimensional Brownian particle was moving in the field with three potential wells (see **Figure 2A**). Ten extensive long simulations, which started from different sites

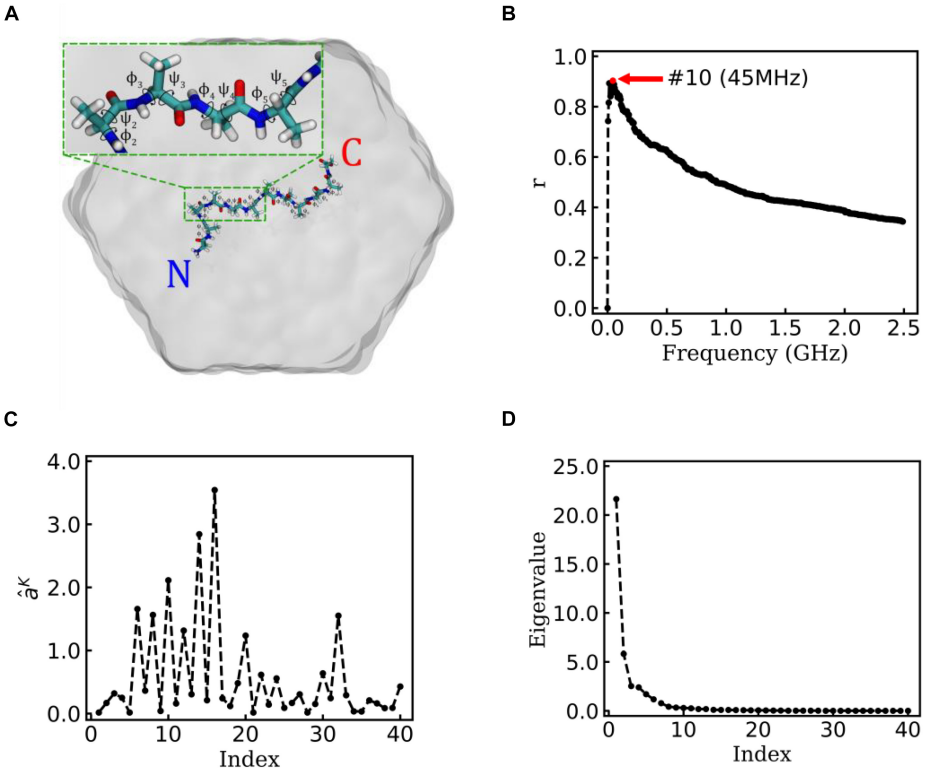


FIGURE 3 | EspcTM on a typical trajectory of Ala_{12}. **(A)** The typical conformation of Ala_{12} represented in sticks with labels of the 10 pairs of dihedral angles φ and ψ, solvated in SPC water molecules represented in gray surface. The inset figure shows the zoom-in of the segment contenting $\varphi_{2\sim5}$ and $\psi_{2\sim5}$. **(B)** Multiple correlation coefficients of $\tilde{\varepsilon}^{\mathrm{K}}$ and all 40 conformational coordinates as a function of the cutoff frequency. The maximum of the multiple correlation coefficient located at a cutoff frequency equaling 45 MHz. **(C)** The regression coefficients for all 40 features. Most coefficients with large value correspond to the basis functions (sine and cosine) of $\varphi_{2\sim5}$. **(D)** The eigenvalues in the PCA of trajectory-mapped vector.

randomly, were performed to make the distribution of samples close to the theoretical values. **Figure 2B** shows the positions and distribution of the samples of these trajectories. In the analysis, $\sin(n\theta)$ *and* $\cos(n\theta)$ were selected as the basis functions. θ indicates the coordinate x *or* y, and $n = 1, \ldots, 10$ for every coordinate in the EspcTM analysis of the toy model. Hence, the trajectories were mapped into a 40-dimensional functional space, e.g.,

$$\begin{aligned} &\sin(x), \sin(y), \cos(x), \cos(y), \ldots \\ &\sin(10x), \sin(10y), \cos(10x), \cos(10y) \end{aligned} \tag{9}$$

All values of the trajectories were normalized in every dimension before they were fitted with $\tilde{\varepsilon}^{\mathrm{K}}$.

Figure 2C shows the multiple correlation coefficient between $\tilde{\varepsilon}^{\mathrm{K}}$ and the values of these 40 features as a function of the cutoff frequency. There was a maximum multiple correlation coefficient at $\mathrm{K} = 17$, and $\tilde{\varepsilon} = \tilde{\varepsilon}^{17} - \epsilon^{17}$ was selected as the effective energy. **Figure 2D** shows regression coefficients between the energy $\tilde{\varepsilon}^{17}$ and features. As shown in **Figure 2D**, the basis functions $\sin(x)$, $\cos(x)$, and $\cos(2x)$ possessed large weight in the rescaling. It should be noted that to consider the effect of the random force by solvation in Brown dynamics, additional energies with Gaussian distribution were added into the energies of the Brownian particle, so that information of potential was mixed with white noise in linear regression. PCA was performed on these effective energy rescaled samples. **Figure 2E** shows the eigenvalues in descending order. It is obvious that apart from the first two eigenvalues, other eigenvalues were very small. The first two eigenvectors were selected to compose the E-space of the toy model, as well as the mapping operator. By using the mapping operator M, composed by these two eigenvectors, all samples were mapped into the E-space.

By using the PCCA+ algorithm, all samples had been grouped into three states (shown by colored dots in **Figure 2B**). As shown in **Figure 2B**, these three states corresponded to the three wells in the potential. A discretized trajectory who visited all three states is shown in **Figure 2F**. The Markov transition matrix P was obtained based on the discretized trajectories (see **Table 1**). The stationary distribution, which corresponds to the distribution of the thermodynamic equilibrium, was obtained by the eigen-decomposition of the Markov transition matrix and shown in **Table 1**. As a benchmark, the distribution of equilibrium state predicted by the theory of statistical physics is shown in **Table 1** as well. It is obvious that the result obtained by the EspcTM method is similar to the theoretical values. Furthermore, the Markov transition matrix contains kinetic information about the system as well. The lifetime of these states,

TABLE 1 | Transition matrix and stationary distribution of the Markov model, distribution obtained by theory of equilibrium statistical physics, and lifetime of states for the dynamics of a Brownian particle.

	Transition matrix P			Stationary distribution	Theory	Lifetime (100τ)
	S0	S1	S2			
S0	0.882	0.069	0.049	0.184	0.186	8.5
S1	0.024	0.858	0.118	0.532	0.538	7.08
S2	0.032	0.221	0.747	0.284	0.276	3.99

which were calculated by the diagonal elements of the transition matrix, is also shown in **Table 1**. It was found that the state S_0 possessed the lowest occurring probability but the longest lifetime. This indicated that the kinetically stable state was not the thermodynamically stable state for this dynamic system.

Dynamics of Alanine Dodecapeptide

Alanine dodecapeptide (Ala_{12}), consisting of 12 alanine residues, is a typical model molecule for MD study (Noe et al., 2007). The MD trajectories of an Ala_{12} was used as an example to test the EspcTM method. According to the previous study (Gong and Zhou, 2010; Gong et al., 2015), sine and cosine of backbone dihedral angles (φ, ψ) were used as basis functions in the analysis of the MD trajectories of Ala_{12}. Here, φ is defined as the backbone dihedral angle around the bond connecting C_α and N atoms and ψ is defined as the backbone dihedral angle around the bond connecting C_α and carbonyl carbon atoms (Hovmoller et al., 2002). There are 10 pairs of dihedral angles φ and ψ for Ala_{12} (see **Figure 3A**), and 40 basis functions were finally included in the analysis, e.g.,

$$\sin(\psi_i), \sin(\varphi_i), \cos(\psi_i), \cos(\varphi_i) \tag{10}$$

Here, $i = 1, \ldots, 10$ indicates the index of dihedrals of Ala_{12} from N-terminal to C-terminal. Based on these basis functions, the EspcTM was first applied on a typical trajectory and then on all the 50 trajectories.

State Transition of a Typical Trajectory

Figure 3B shows the result of the multiple linear regression between $\tilde{\varepsilon}^{\mathrm{K}}$ and functions of the dihedral angles of Ala_{12} for a typical trajectory. There is a maximum of the multiple correlation coefficient, similar to the case of movement of Brownian particle, at 45 MHz (see **Figure 3B**). Therefore, the summary of the first 10 lowest frequencies of energy $\tilde{\varepsilon}^{10}$ was used in the analysis. The regression coefficients between the energy $\tilde{\varepsilon}^{10}$ and functions of dihedral angles are shown in **Figure 3C**. It was found that most factors with large weight corresponded to the basis function (sine and cosine) of $\varphi_{2\sim5}$ (see the inset figure of **Figure 3A**). This indicates that the structure change near N-terminal contributes more to large-scale conformational change than C-terminal in this typical simulation trajectory.

Figure 3D shows the eigenvalues of weighted samples of this trajectory. As shown in **Figure 3D**, the following analysis on this trajectory was performed in the space made up of the first six eigenvectors. **Figure 4A** shows the similarity matrix and the representative structure of the trajectory. It was obvious that there were four metastable states in the trajectory. The discretized trajectory is shown in the middle panel of **Figure 4B**. The secondary structure of the peptide was analyzed by DSSP (Kabsch and Sander, 1983; Touw et al., 2015) and shown in the top panel of **Figure 4B**. The simulation started from a structure with some of the N-terminal α-helix formed (also see the representative structure), i.e., the state S_b. This state was unstable and only existed about 6.4 ns in the 100-ns trajectory. The α-helix formed in this state acted as a nucleus that promoted the formation of the α-helix of the C-terminal of the Ala_{12}. Then, the trajectory transited to the S_a state, in which most of the residues of the peptide formed the α-helix structure. State S_a was more stable than state S_b. It appeared two times in this trajectory and existed about 58.0 ns in total. However, between the two occasions of the state S_a, the α-helix of two termini had been temporally uncoiled and interacted with the α-helix in the middle of the peptide, i.e., the state S_c. This state is unstable and existed only for 16.4 ns in this trajectory. After the state S_c, the peptide folded to the state S_a again. Finally, the peptide unfolded into a random coil, i.e., state S_d, with low structural similarity.

The bottom panel of **Figure 4B** shows effective energy as a function of time for this trajectory. It was calculated from the total energy of the whole biosystem, including the peptide and water molecules. Initially, the energy caused by the conformational change of the peptide was concealed by the noise of the dynamics of water molecules as well as the fluctuation of itself. It seemed that the total energy (shown in gray) varied randomly and dramatically. However, by using the FFT and regression, we obtained the effective energy (shown in red). It was synchronous with conformational change and state transition of the peptide. More interestingly, the effective energy of stable state, state S_a, was much lower than the other three states, in which most of the α-helix was formed. This implied that the stability of this state was supported by energy. On the other hand, the state S_d possessed the highest energy and large conformational variations. This implied that the unfolded coil structure was stabled by the entropy.

Transition Network of Ala_{12}

To obtain statistically significant conclusions, we performed the analysis of EspcTM method on 50 MD trajectories. **Figure 5A** shows the result of the multiple linear regression between $\tilde{\varepsilon}^{\mathrm{K}}$ and functions of the dihedral angles of Ala_{12} for these 50 trajectories. The maximum of the multiple correlation coefficient was found at the frequency equal to 15 MHz. The summation of the first four lowest frequencies of energy $\tilde{\varepsilon}^{4}$ was used in the analysis. **Figure 5B** shows the regression coefficients between the energy $\tilde{\varepsilon}^{4}$ and features. It consistently showed that $\varphi_{2\sim5}$ played important roles in the dynamics of the Ala_{12} though there was a phase shift on $\varphi_{2\sim5}$ caused small weights on the cosine of $\varphi_{2\sim5}$. This indicates that local structure changes near the N-terminal, especially the $\varphi_{2\sim5}$, were the major contributors to the slow conformational change of the Ala_{12}. According to the result of the PCA on the weighted feature space, the clustering algorithm was performed in the space made up of the first 10 eigenvectors, whose sum was over 90% sum of variation (see **Figure 5C**). Every

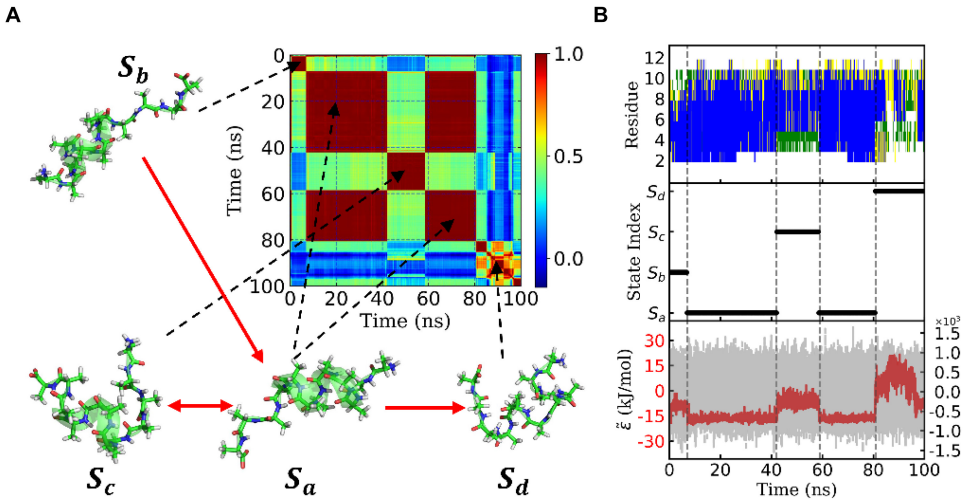


FIGURE 4 | State transition of a typical trajectory of Ala_{12}. **(A)** Similarity matrix and typical conformations in the metastable states and their transitions. The color indicated the degree of similarity. Red means high similarity. The transitions were implied from the transition probability matrix. **(B)** Secondary structure analysis of the typical trajectory by DSSP was shown in the upper panel. The blue, green, yellow, and white patterns represented α-helix, bend, turn, and coil, respectively. The discretized trajectory was shown in the middle panel. The states corresponded to the similarity matrix in panel **(A)**. In the lower panel, the effective energy for this typical trajectory was exhibited in the red dashed curve and the original potential energy was in the gray curve as background. Both curves shared the same *x*-axis but with *y*-axis in different scales. The effective energy's *y*-axis was on the left with an amplitude of about 20 kJ/mol, while the original potential energy's *y*-axis was on the right with an amplitude of about 1.2×10^3 kJ/mol. Here, both effective energy and original potential energy had been zero-centered.

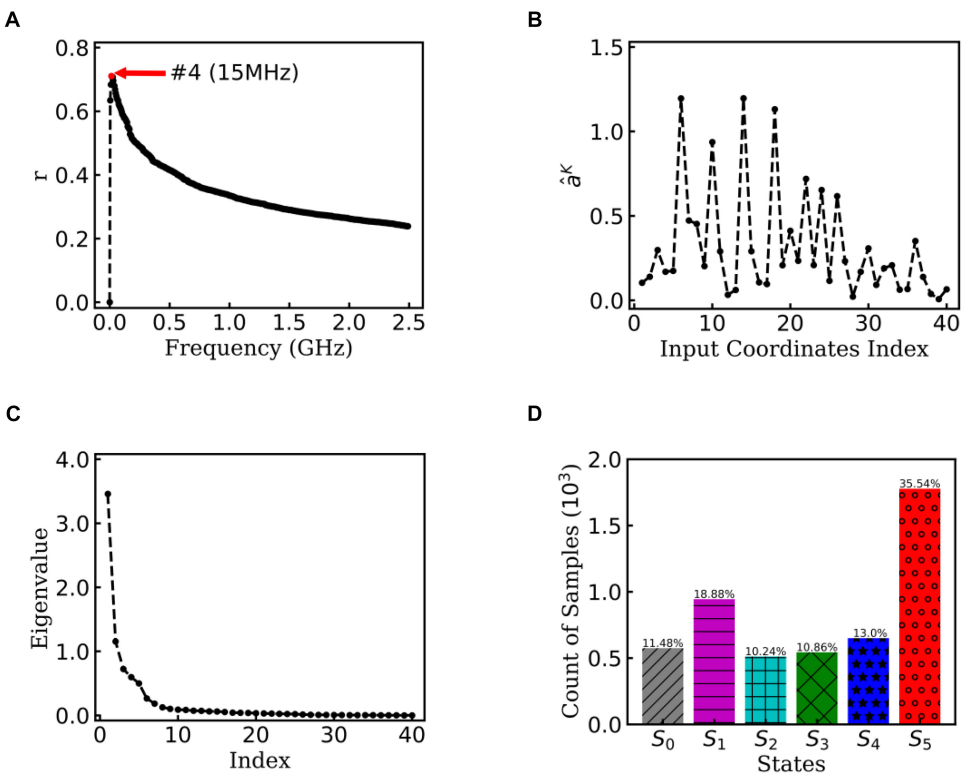


FIGURE 5 | EspcTM on 50 trajectories of Ala_{12}. **(A)** Multiple correlation coefficients of regression between $\tilde{\varepsilon}^K$ and features as a function of cutoff frequencies. The maximum was at 15 MHz. **(B)** The regression coefficients for all 40 features. **(C)** The eigenvalues in the PCA of trajectory-mapped vector. **(D)** The observed probability for each state in all 50 trajectories.

TABLE 2 | Transition matrix and stationary distribution of the Markov model and lifetime of states for the dynamics of Ala_{12}.

	Transition matrix						Stationary distribution	Lifetime (ns)
	S0	S1	S2	S3	S4	S5		
S0	0.610	0.276	0.038	0.031	0.036	0.009	0.114	2.61
S1	0.170	0.693	0.021	0.037	0.026	0.053	0.187	3.30
S2	0.042	0.037	0.758	0.115	0.030	0.018	0.103	4.16
S3	0.032	0.064	0.109	0.725	0.050	0.020	0.109	3.67
S4	0.031	0.038	0.024	0.042	0.798	0.067	0.131	4.97
S5	0.003	0.027	0.005	0.006	0.025	0.934	0.356	15.15

trajectory was divided into 100 pieces. Thus, there were 5,000 vectors, which represent 100×50 trajectory pieces. Six states were identified from these 50 trajectories.

Figure 5D shows the histogram of these six states. Here, the state transitions were obtained from the 50 trajectories with the lag time 1.0 ns. The transition matrix and stationary distribution are shown in **Table 2**. It was found that the stationary distribution obtained by the transition matrix was consistent with the histogram. The state S_5 had a much higher occurring probability than that of other states in the equilibrium state. **Figure 6** displays these six states, represented by their typical structures in cartoons, along with their average effective energy in vertical. The unfolded states S_0, in which peptide unfolded into a random coil, possessed the highest energy and located at the top of the figure. The folded state S_5, in which the peptide folded into α-helices, possessed the lowest effective energy and located at the bottom of the figure. Between these two states, the peptide was half-folded. In the state S_1, a helix was formed in the N-terminal of the peptide. In states S_2, S_3, and S_4, some helices were formed in the C-terminal. A remarkable gap between the effective energy of state S_4 and state S_5 separated the folded state from the other five states. This implied that the energy is the reason for the stability of the folded state.

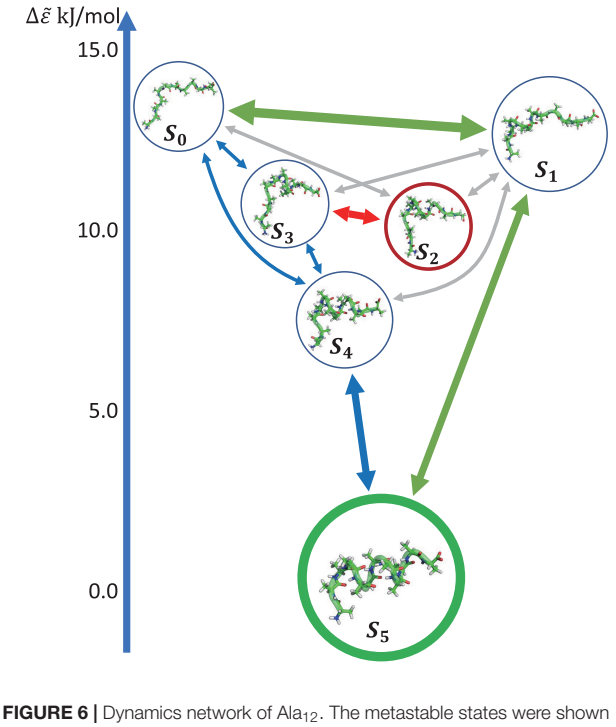


FIGURE 6 | Dynamics network of Ala_{12}. The metastable states were shown by the cartoon structures of their typical conformations and rearranged by their average effective energy in vertical. The size of circles around the pictures indicated the occurring probability in the stationary mode. The arrows showed the main transitions between six states in the equilibrium state. The pathway was indicated by the color of the arrows. The transitions were shown in three levels according to the transition frequency, i.e., ~30, ~10, and ~5μs^{-1}, and indicated by the linewidth of the arrows.

Furthermore, we obtain the dynamics and kinetics of the system based on the transition matrix. **Figure 6** shows the main transition between six states in lines with arrows. The most frequent transition, about 32 μs^{-1}, occurred between the state S_0 and S_1 due to the high flexibility of the peptide in these two states. This high transition frequency made the lifetime of these two states lower than that of states S_2, S_3, and S_4, though the occurring probabilities of these two states were a little higher than the other three states. In the transition network, there were two main folding pathways from the unfolded state to the folded state. The fast folding pathway, which passed through state S_1 and was shown by green arrows, formed the α-helices from the N-terminal to the C-terminal directly. The slow folding pathway, which involved states S_2, S_3, and S_4, was shown by blue and red arrows and was more complex than the fast one. In this pathway, the α-helices formed from the C-terminal to the N-terminal, i.e., passed through states S_3 and S_4 sequentially. The misfolded state S_2 connected with state S_3. A detailed structural study showed that the structures of states S_2 and S_4 were very similar. However, some misfolded residues hindered the formation of the N-terminal helix in the state S_2. To reach the folded state, it must unfold into state S_3. These results indicated that the N-terminal helix plays a vital role in the folding of the peptide in kinetics. It is consistent with the aforementioned result of linear regression, that the $\varphi_{2\sim5}$ of the peptide possessed large rescaling factors, as well as the results by other experimental groups, that alanine-rich peptides folded into the α-helix in the N-terminal at first (Millhauser et al., 1997; Yoder et al., 1997). It must be noted that, as we mentioned before, the biomolecules are intrinsically dynamic (Chodera et al., 2007) and the unfolded states of the peptide were transferred to each other frequently. These two pathways only described the major folding process of Ala_{12}. Some minor branches in the folding pathways also existed.

CONCLUSION

In this work, we introduced our EspcTM method by applying it to investigate the movement of Brownian particle and conformational dynamics of Ala_{12} in this work. In the study of Brownian particle, by using the EspcTM method, we obtained three states from simulation trajectories. The regions of the states given by EspcTM are in accordance with the potential wells of the landscape. In addition, the equilibrium distribution obtained by the kinetic transition network-based Markov chain theory was consistent with the theoretical result. In the study of Ala_{12}, a meaningful kinetic transition network was obtained to describe the folding behavior of Ala_{12}. The effective energy, which was filtered from the total potential energy of simulation trajectories by FFT and multiple linear regression, was shown to be an efficacious order parameter to describe the conformational change of Ala_{12}. We showed that the folding process of Ala_{12} was synchronous with the change of effective energy. The folded state, in which most of the residues were in helices, possessed the lowest effective energy and was most stable in thermodynamics. Two major folding pathways were also found in the kinetic network. The N-terminal helix of the Ala_{12} was found to play an important role in the folding of Ala_{12} in both thermodynamics and kinetics. This is consistent with previous experimental result. Thus, the EspcTM is expected to be a powerful tool for studies of dynamics of complex systems and should be applied to studies of dynamics of large biomolecule systems to improve our understanding of the thermodynamics and kinetics of biomolecular systems.

Technically, the EspcTM method is an analysis framework based on the TM method. It identifies metastable states from simulation data and constructs the transition network between the states based on the theory of Markov chain. Different from the TM method, we provided a *de novo* solution to obtain an analysis space, named as E-space, to describe the slow processes in the EspcTM method. This solution is based on a parameter-free optimization approach. Thus, the EspcTM method is friendly to inexperienced users. The E-space is independent from the TM method. It is convenient to use it in the MSM method. For the experienced users, especially those with knowledge on the dynamics of system, they can set cutoff frequency manually as well. Furthermore, as an extension of the EspcTM method, some new transfer functions, such as logistic function and ReLU, can also be used in the energy filter process. The wavelet analysis method can be used in transforming the energy between time domain and frequency domain.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

GZ and ZW designed the study. ZW collected data and carried out the calculation. ZW, XZ, and GZ wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Natural Science Foundation of China (Grant No. 11474068) and the National Basic Research Program of China (973 Project; Grant No. 2013CB834100).

ACKNOWLEDGMENTS

The authors thank the support of the State Key Laboratory of Applied Surface Physics and the Department of Physics, Fudan University, China.

REFERENCES

Adcock, S. A., and McCammon, J. A. (2006). Molecular dynamics: survey of methods for simulating the activity of proteins. *Chem. Rev.* 106, 1589–1615. doi: 10.1021/cr040426m

Berendsen, H. J. C., Postma, J. P. M., Vangunsteren, W. F., Dinola, A., and Haak, J. R. (1984). Molecular-dynamics with coupling to an external bath. *J. Chem. Phys.* 81, 3684–3690. doi: 10.1063/1.448118

Boehr, D. D., Mcelheny, D., Dyson, H. J., and Wright, P. E. (2006). The dynamic energy landscape of dihydrofolate reductase catalysis. *Science* 313, 1638–1642. doi: 10.1126/science.1130258

Bowman, G. R., Beauchamp, K. A., Boxer, G., and Pande, V. S. (2009). Progress and challenges in the automated construction of markov state models for full protein systems. *J. Chem. Phys.* 131:124101. doi: 10.1063/1.3216567

Bowman, G. R., Meng, L. M., and Huang, X. H. (2013). Quantitative comparison of alternative methods for coarse-graining biological networks. *J. Chem. Phys.* 139:121905. doi: 10.1063/1.4812768

Bowman, G. R., and Pande, V. S. (2010). Protein folded states are kinetic hubs. *Proc. Natl. Acad. Sci. U.S.A.* 107, 10890–10895. doi: 10.1073/pnas.1003962107

Buch, I., Giorgino, T., and De Fabritiis, G. (2011). Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations. *Proc. Natl. Acad. Sci. U.S.A.* 108, 10184–10189. doi: 10.1073/pnas.1103547108

Buchete, N. V., and Hummer, G. (2008). Coarse master equations for peptide folding dynamics. *J. Phys. Chem. B* 112, 6057–6069. doi: 10.1021/jp076 1665

Cheatham, T. E., and Kollman, P. A. (2000). Molecular dynamics simulation of nucleic acids. *Annu. Rev. Phys. Chem.* 51, 435–471. doi: 10.1146/annurev. physchem.51.1.435

Chiti, F., and Dobson, C. M. (2006). Protein misfolding, functional amyloid, and human disease. *Annu. Rev. Biochem.* 75, 333–366. doi: 10.1146/annurev. biochem.75.101304.123901

Chodera, J. D., Singhal, N., Pande, V. S., Dill, K. A., and Swope, W. C. (2007). Automatic discovery of metastable states for the construction of Markov models of macromolecular conformational dynamics. *J. Chem. Phys.* 126:155101. doi: 10.1063/1.2714538

Cochran, W. T., Cooley, J. W., Favin, D. L., Helms, H. D., Kaenel, R. A., Lang, W. W., et al. (1967). What is the fast Fourier transform? *Proc.IEEE* 55, 1664–1674. doi: 10.1109/PROC.1967.5957

Darden, T., York, D., and Pedersen, L. (1993). Particle mesh ewald - an N.Log(N) method for ewald sums in large systems. *J. Chem. Phys* 98, 10089–10092. doi: 10.1063/1.464397

Deng, N. J., Dai, W., and Levy, R. M. (2013). How kinetics within the unfolded state affects protein folding: an analysis based on markov state models and an ultra-long MD trajectory. *J. Phys. Chem. B* 117, 12787–12799. doi: 10.1021/jp401962k

Deuflhard, P., and Weber, M. (2005). Robust perron cluster analysis in conformation dynamics. *Linear Alg. Appl.* 398, 161–184. doi: 10.1016/j.laa.2004.10.026

Eisenmesser, E. Z., Bosco, D. A., Akke, M., and Kern, D. (2002). Enzyme dynamics during catalysis. *Science* 295, 1520–1523. doi: 10.1126/science.1066176

Gao, Y. Q., Yang, W., and Karplus, M. (2005). A structure-based model for the synthesis and hydrolysis of ATP by F-1-ATPase. *Cell* 123, 195–205. doi: 10.1016/j.cell.2005.10.001

Gfeller, D., De Los Rios, P., Caflisch, A., and Rao, F. (2007). Complex network analysis of free-energy landscapes. *Proc. Natl. Acad. Sci. U.S.A.* 104, 1817–1822. doi: 10.1073/pnas.0608099104

Gong, L. C., and Zhou, X. (2010). Kinetic transition network based on trajectory mapping. *J. Phys. Chem. B* 114, 10266–10276. doi: 10.1021/jp100737g

Gong, L. C., Zhou, X., and Ouyang, Z. C. (2015). Systematically constructing kinetic transition network in polypeptide from top to down: trajectory mapping. *PLoS One* 10:e0125932. doi: 10.1371/journal.pone.0125932

Gregersen, N., Bross, P., Vang, S., and Christensen, J. H. (2006). Protein misfolding and human disease. *Annu. Rev. Genomics Hum. Genet.* 7, 103–124. doi: 10.1146/annurev.genom.7.080505.115737

Guo, C., Luo, Y., Zhou, R. H., and Wei, G. H. (2012). Probing the self-assembly mechanism of diphenylalanine-based peptide nanovesicles and nanotubes. *Acs Nano* 6, 3907–3918. doi: 10.1021/nn300015g

Hess, B., Kutzner, C., Van Der Spoel, D., and Lindahl, E. (2008). GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* 4, 435–447. doi: 10.1021/ct700301q

Hovmoller, S., Zhou, T., and Ohlson, T. (2002). Conformations of amino acids in proteins. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 58, 768–776. doi: 10.1107/s0907444902003359

Husic, B. E., and Pande, V. S. (2018). Markov state models: from an art to a science. *J. Am. Chem. Soc.* 140, 2386–2396. doi: 10.1021/jacs.7b12191

Ithuralde, R. E., Roitberg, A. E., and Turjanski, A. G. (2016). Structured and unstructured binding of an intrinsically disordered protein as revealed by atomistic simulations. *J. Am. Chem. Soc.* 138, 8742–8751. doi: 10.1021/jacs.6b02016

Jayachandran, G., Vishal, V., and Pande, V. S. (2006). Using massively parallel simulation and Markovian models to study protein folding: examining the dynamics of the villin headpiece. *J. Chem. Phys.* 124:164902. doi: 10.1063/1.2186317

Kabsch, W., and Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22, 2577–2637. doi: 10.1002/bip.360221211

Kampen, N. G. V. (2007). *Stochastic Processes in Physics and Chemistry*. Amsterdam: North Holland.

Karplus, M., and McCammon, J. A. (2002). Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.* 9, 646–652. doi: 10.1038/nsb0902-646

Lever, J., Krzywinski, M., and Atman, N. (2017). Principal component analysis. *Nat. Methods* 14, 641–642. doi: 10.1038/nmeth.4346

Levy, Y., Wolynes, P. G., and Onuchic, J. N. (2004). Protein topology determines binding mechanism. *Proc. Natl. Acad. Sci. U.S.A.* 101, 511–516. doi: 10.1073/pnas.2534828100

Li, C. X., Yang, D. M., Ma, P. A., Chen, Y. Y., Wu, Y., Hou, Z. Y., et al. (2013). Multifunctional upconversion mesoporous silica nanostructures for dual modal imaging and in vivo drug delivery. *Small* 9, 4150–4159. doi: 10.1002/smll.201301093

Li, W. F., Wang, J., Zhang, J., Takada, S., and Wang, W. (2019). Overcoming the bottleneck of the enzymatic cycle by steric frustration. *Phys. Rev. Lett.* 122:238102. doi: 10.1103/PhysRevLett.122.238102

Li, W. F., Zhang, J., Wang, J., and Wang, W. (2008). Metal-coupled folding of Cys(2)His(2) zinc-finger. *J. Am. Chem. Soc.* 130, 892–900. doi: 10.1021/ja075302g

Millhauser, G. L., Stenland, C. J., Hanson, P., Bolin, K. A., and Vandeven, F. J. M. (1997). Estimating the relative populations of 3(10)-helix and alpha-helix in Ala-rich peptides: a hydrogen exchange and high field NMR study. *J. Mol. Biol.* 267, 963–974. doi: 10.1006/jmbi.1997.0923

Mirny, L., and Shakhnovich, E. (2001). Protein folding theory: from lattice to all-atom models. *Annu. Rev. Biophys. Biomol. Struct.* 30, 361–396. doi: 10.1146/annurev.biophys.30.1.361

Miyashita, N., Straub, J. E., and Thirumalai, D. (2009). Structures of beta-Amyloid Peptide 1-40, 1-42, and 1-55-the 672-726 Fragment of APP-in a Membrane Environment with Implications for Interactions with gamma-Secretase. *J. Am. Chem. Soc.* 131, 17843–17852. doi: 10.1021/ja905457d

Moraitakis, G., Purkiss, A. G., and Goodfellow, J. M. (2003). Simulated dynamics and biological macromolecules. *Rep. Prog. Phys.* 66, 383–406. doi: 10.1088/0034-4885/66/3/203

Naritomi, Y., and Fuchigami, S. (2011). Slow dynamics in protein fluctuations revealed by time-structure based independent component analysis: the case of domain motions. *J. Chem. Phys.* 134:065101. doi: 10.1063/1.3554380

Naritomi, Y., and Fuchigami, S. (2013). Slow dynamics of a protein backbone in molecular dynamics simulation revealed by time-structure based independent component analysis. *J. Chem. Phys.* 139:215102. doi: 10.1063/1.4834695

Noe, F. (2008). Probability distributions of molecular observables computed from Markov models. *J. Chem. Phys.* 128:244103. doi: 10.1063/1.2916718

Noe, F., Horenko, I., Schutte, C., and Smith, J. C. (2007). Hierarchical analysis of conformational dynamics in biomolecules: transition networks of metastable states. *J. Chem. Phys.* 126:155102. doi: 10.1063/1.2714539

Norberg, J., and Nilsson, L. (2002). Molecular dynamics applied to nucleic acids. *Accounts Chem. Res.* 35, 465–472. doi: 10.1021/ar010026a

Onuchic, J. N., Lutheyschulten, Z., and Wolynes, P. G. (1997). Theory of protein folding: the energy landscape perspective. *Annu. Rev. Phys. Chem.* 48, 545–600. doi: 10.1146/annurev.physchem.48.1.545

Pan, A. C., Jacobson, D., Yatsenko, K., Sritharan, D., Weinreich, T. M., and Shaw, D. E. (2019). Atomic-level characterization of protein-protein association. *Proc. Natl. Acad. Sci. U.S.A.* 116, 4244–4249. doi: 10.1073/pnas.1815431116

Pande, V. S., Beauchamp, K., and Bowman, G. R. (2010). Everything you wanted to know about markov state models but were afraid to ask. *Methods* 52, 99–105. doi: 10.1016/j.ymeth.2010.06.002

Paul, F., Wehmeyer, C., Abualrous, E. T., Wu, H., Crabtree, M. D., Schoneberg, J., et al. (2017). Protein-peptide association kinetics beyond the seconds timescale from atomistic simulations. *Nat. Commun.* 8:1095. doi: 10.1038/s41467-017-01163-6

Perez-Hernandez, G., Paul, F., Giorgino, T., De Fabritiis, G., and Noe, F. (2013). Identification of slow molecular order parameters for Markov model construction. *J. Chem. Phys.* 139:015102. doi: 10.1063/1.4811489

Prinz, J. H., Wu, H., Sarich, M., Keller, B., Senne, M., Held, M., et al. (2011). Markov models of molecular kinetics: generation and validation. *J. Chem. Phys.* 134:190401. doi: 10.1063/1.3565032

Rao, F., and Karplus, M. (2010). Protein dynamics investigated by inherent structure analysis. *Proc. Natl. Acad. Sci. U.S.A.* 107, 9152–9157. doi: 10.1073/pnas.0915087107

Reuter, B., Weber, M., Fackeldey, K., Roblitz, S., and Garcia, M. E. (2018). Generalized markov state modeling method for nonequilibrium biomolecular dynamics: exemplified on amyloid beta conformational dynamics driven by an oscillating electric field. *J. Chem. Theory Comput.* 14, 3579–3594. doi: 10.1021/acs.jctc.8b00079

Roblitz, S., and Weber, M. (2013). Fuzzy spectral clustering by PCCA plus : application to Markov state models and data classification. *Adv. Data Anal. Classif.* 7, 147–179. doi: 10.1007/s11634-013-0134-6

Scherer, M. K., Trendelkamp-Schroer, B., Paul, F., Perez-Hernandez, G., Hoffmann, M., Plattner, N., et al. (2015). PyEMMA 2: a software package for estimation, validation, and analysis of markov models. *J. Chem. Theory Comput.* 11, 5525–5542. doi: 10.1021/acs.jctc.5b00743

Schneider, A., Hommel, G., and Blettner, M. (2010). Linear regression analysis part 14 of a series on evaluation of scientific publications. *Dtsch. Arztebl. Int.* 107, 776–782. doi: 10.3238/arztebl.2010.0776

Schwantes, C. R., and Pande, V. S. (2013). Improvements in markov state model construction reveal many non-native interactions in the folding of NTL9. *J. Chem. Theory Comput.* 9, 2000–2009. doi: 10.1021/ct300878a

Sengupta, U., Carballo-Pacheco, M., and Strodel, B. (2019). Automated Markov state models for molecular dynamics simulations of aggregation and self-assembly. *J. Chem. Phys.* 150:115101. doi: 10.1063/1.5083915

Sims, G. E., Choi, I. G., and Kim, S. H. (2005). Protein conformational space in higher order phi-psi maps. *Proc. Natl. Acad. Sci. U.S.A.* 102, 618–621. doi: 10.1073/pnas.0408746102

Touw, W. G., Baakman, C., Black, J., Te Beek, T. A. H., Krieger, E., Joosten, R. P., et al. (2015). A series of PDB-related databanks for everyday needs. *Nucleic Acids Res.* 43, D364–D368. doi: 10.1093/nar/gku1028

Wang, C. L., Lu, H. J., Wang, Z. G., Xiu, P., Zhou, B., Zuo, G. H., et al. (2009). Stable liquid water droplet on a water monolayer formed at room temperature on ionic model substrates. *Phys. Rev. Lett.* 103:137801. doi: 10.1103/PhysRevLett.103.137801

Wang, W., Cao, S. Q., Zhu, L. Z., and Huang, X. H. (2018). Constructing markov state models to elucidate the functional conformational changes of complex biomolecules. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 8:e1343. doi: 10.1002/wcms.1343

Weber, J. K., Jack, R. L., and Pande, V. S. (2013). Emergence of glass-like behavior in markov state models of protein folding dynamics. *J. Am. Chem. Soc.* 135, 5501–5504. doi: 10.1021/ja4002663

Wei, G. H., Xi, W. H., Nussinov, R., and Ma, B. Y. (2016).). Protein ensembles: how does nature harness thermodynamic fluctuations for life? the diverse functional roles of conformational ensembles in the cell. *Chem. Rev.* 116, 6516–6551. doi: 10.1021/acs.chemrev.5b00562

Weng, J. W., and Wang, W. N. (2020). Dynamic multivalent interactions of intrinsically disordered proteins. *Curr. Opin. Struc. Biol.* 62, 9–13. doi: 10.1016/j.sbi.2019.11.001

Wu, K., Xu, S., Wan, B., Xiu, P., and Zhou, X. (2020). A novel multiscale scheme to accelerate atomistic simulations of bio-macromolecules by adaptively driving coarse-grained coordinates. *J. Chem. Phys.* 152:114115. doi: 10.1063/1.5135309

Yan, Z. Q., and Wang, J. (2019). Superfunneled energy landscape of protein evolution unifies the principles of protein evolution, folding, and design. *Phys. Rev. Lett.* 122:018103. doi: 10.1103/PhysRevLett.122.018103

Yang, J. R., Shi, G. S., Tu, Y. S., and Fang, H. P. (2014). High correlation between oxidation loci on graphene Oxide. *Angew. Chem. Int. Ed.* 53, 10190–10194. doi: 10.1002/anie.201404144

Yoder, G., Pancoska, P., and Keiderling, T. A. (1997). Characterization of alanine-rich peptides, Ac-(AAKAA)(n)-GY-NH2 (n=1-4), using vibrational circular dichroism and Fourier transform infrared. Conformational determination and thermal unfolding. *Biochemistry* 36, 15123–15133. doi: 10.1021/bi971460g

Zhang, C. B., Xu, S., and Zhou, X. (2019a). Identifying metastable states of biomolecules by trajectory mapping and density peak clustering. *Phys. Rev. E* 100:033301. doi: 10.1103/PhysRevE.100.033301

Zhang, C. B., Ye, F. F., Li, M., and Zhou, X. (2019b). Enhanced sampling based on slow variables of trajectory mapping. *Sci. China Phys. Mech. Astron.* 62:62. doi: 10.1007/s11433-018-9313-1

Zhang, C. B., Yu, J., and Zhou, X. (2017). Imaging metastable states and transitions in proteins by trajectory map. *J. Phys. Chem. B* 121, 4678–4686. doi: 10.1021/acs.jpcb.7b00664

Zhou, H., Yang, Z. X., Tian, X., Chen, L., Lee, S., Huynh, T., et al. (2019). Lanosterol disrupts the aggregation of amyloid-beta peptides. *ACS Chem. Neurosci.* 10, 4051–4060. doi: 10.1021/acschemneuro.9b00285

Zhou, R. H., Huang, X. H., Margulis, C. J., and Berne, B. J. (2004). Hydrophobic collapse in multidomain protein folding. *Science* 305, 1605–1609. doi: 10.1126/science.1101176

Zuo, G. H., Hu, J., and Fang, H. P. (2009). Effect of the ordered water on protein folding: an off-lattice Go-like model study. *Phys. Rev. E* 79(Pt 1):031925. doi: 10.1103/PhysRevE.79.031925

Zuo, G. H., Kang, S. G., Xiu, P., Zhao, Y. L., and Zhou, R. H. (2013). Interactions between proteins and carbon-based nanoparticles: exploring the origin of nanotoxicity at the molecular level. *Small* 9, 1546–1556. doi: 10.1002/smll.201201381

Zuo, G. H., Li, W. F., Zhang, J., Wang, J., and Wang, W. (2010). Folding of a small rna hairpin based on simulation with replica exchange molecular dynamics. *J. Phys. Chem. B* 114, 5835–5839. doi: 10.1021/jp904573r

Zuo, G. H., Wang, J., and Wang, W. (2006). Folding with downhill behavior and low cooperativity of proteins. *Proteins Struct. Funct. Bioinformat.* 63, 165–173. doi: 10.1002/prot.20857

Zwanzig, R. (1983). From classical dynamics to continuous-time random-walks. *J. Stat. Phys.* 30, 255–262. doi: 10.1007/bf01012300

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Molecular Biosciences

ORIGINAL RESEARCH
published: 15 December 2020
doi: 10.3389/fmolb.2020.608931

Development and Testing of Force Field Parameters for Phenylalanine and Tyrosine Derivatives

Xiaowen Wang[1,2] *and Wenjin Li*[1]*

[1] Institute for Advanced Study, Shenzhen University, Shenzhen, China, [2] College of Physics and Optoelectronic Engineering, Shenzhen University, Shenzhen, China

Theoretical analyses are valuable for the exploration of the effects of unnatural amino acids on enzyme functions; however, many necessary parameters for unnatural amino acids remain lacking. In this study, we developed and tested force field parameters compatible with Amber ff14SB for 18 phenylalanine and tyrosine derivatives. The charge parameters were derived from ab initio calculations using the RESP fitting approach and then adjusted to reproduce the benchmark relative energies (at the MP2/TZ level) of the α- and β-backbones for each unnatural amino acid dipeptide. The structures optimized under the proposed force field parameters for the 18 unnatural amino acid dipeptides in both the α- and β-backbone forms were in good agreement with their QM structures, as the average RMSD was as small as 0.1 Å. The force field parameters were then tested in their application to seven proteins containing unnatural amino acids. The RMSDs of the simulated configurations of these unnatural amino acids were approximately 1.0 Å compared with those of the crystal structures. The vital interactions between proteins and unnatural amino acids in five protein–ligand complexes were also predicted using MM/PBSA analysis, and they were largely consistent with experimental observations. This work will provide theoretical aid for drug design involving unnatural amino acids.

Keywords: unnatural amino acids, charge parameters, Amber ff14SB, relative energy, molecular dynamics, MM/PBSA

OPEN ACCESS

Edited by:
Yong Wang,
University of Copenhagen, Denmark

Reviewed by:
Yifei Qi,
East China Normal University, China
Andrea Spitaleri,
San Raffaele Hospital (IRCCS), Italy
Huiyong Sun,
China Pharmaceutical University, China

***Correspondence:**
Wenjin Li
liwenjin@szu.edu.cn

Specialty section:
This article was submitted to Biological Modeling and Simulation, a section of the journal Frontiers in Molecular Biosciences

Received: *22 September 2020*
Accepted: *23 October 2020*
Published: *15 December 2020*

Citation:
Wang X and Li W (2020) Development and Testing of Force Field Parameters for Phenylalanine and Tyrosine Derivatives. Front. Mol. Biosci. 7:608931. doi: 10.3389/fmolb.2020.608931

INTRODUCTION

As is well known, 20 natural amino acids are the main building blocks of proteins, the macromolecules that perform a broad spectrum of functions within organisms (Qin et al., 2015). Unnatural amino acids (UAAs) also called noncanonical amino acids are analogs or metabolic intermediates of the 20 natural amino acids with only minor structural differences—often just a chemical functional group—which is beneficial for analyzing their effects on enzyme functions (Zhao et al., 2020). Since UAAs are of high chemical diversity, possess strong site specificity, and introduce little disturbance to the protein structure, it is widely applied in protein engineering, virus vaccine development, and medical therapeutics (Minnihan et al., 2011; Si et al., 2016; Young and Schultz, 2018). For instant, biological catalysis and reaction mechanism of tyrosine in aminoacyl-tRNA synthetases (aaRS) were investigated through the incorporation of UAA fluorotyrosine, whose pKa was tuned by changing the number and the site of fluoro-substitution (Minnihan et al., 2011). Si and co-workers employed the UAA N^{ε}-2-azidoethyloxycarbonyl-l-lysine to produce replication-incompetent viral vaccines by introducing premature termination codon into the genome of influenza A virus, and these viral vaccines prevented further damage

inside conventional cells via immune response (Si et al., 2016). In addition, UAAs are utilized in the bio-orthogonal reactivity. For example, UAA-incorporated proteins (such as antibodies, growth factors, and cytokines) specifically interacted with diverse moieties to form bispecific antibodies, antibody-drug conjugates, and pegylated proteins, which provided effective treatments for various clinical testing (Young and Schultz, 2018).

The incorporation of UAAs into canonical proteins expanded significantly the genetic code library (Xiao et al., 2015). Natural UAAs occur commonly in plants, microorganisms, and animals, while those in human organisms must be chemically synthesized (Zou et al., 2018). Typically, an orthogonal amber suppressor aaRS/tRNA pair has been utilized to guide the incorporation of UAAs in response to a unique nonsense codon (Santoro et al., 2002; Liu and Schultz, 2010). Experimentally, many studies have reported the incorporation of UAAs into the designated sites of target proteins by means of popular residue-specific and site-directed mutagenesis approaches (Sakamoto et al., 2002; Fleissner et al., 2009; Xiao et al., 2015; Yuet et al., 2015). For example, Yuet et al. described a method for residue-specific labeling that enabled the use of the UAA *p*-azido-l-phenylalanine (AzF) to tag and analyze protein metabolism in specific cells based on the phenylalanyl-tRNA synthetase (Yuet et al., 2015). Schultz et al. utilized site-directed mutagenesis to mutate Val216 of TEM-1 β-lactamase into *p*-acrylamido-phenylalanine (AcrF), which enhanced the catalytic activity of the enzyme (Xiao et al., 2015). Although the two complementary methods involved in residue-specific and site-directed mutagenesis are widely used to incorporate UAAs into proteins, they often contend with certain technical difficulties.

To compensate for experimental obstacles, theoretical computational methods validated by experimental data offer a novel way to screen potential analogs for natural amino acids. A number of computational methods to study proteins containing UAAs have been successively reported by other groups in recent years (Renfrew et al., 2012; Petrov et al., 2013; Khoury et al., 2014a,b). For example, Renfrew et al. constructed a rotamer library containing 114 UAAs to study the interface of calpain and calpastatin, which was evaluated using a scoring function based on the Rosetta program (Leaverfay et al., 2011; Renfrew et al., 2012). New GROMOS54a7 force field parameters were developed by the Zagrovic group for processing post-translationally modified amino acids by means of molecular dynamics (MD) simulations executed by the GROMACS package (Petrov et al., 2013). In addition, a tool called "Forcefield_NCAA" created by the Floudas lab is now available for generating UAA parameters related to a library of 147 noncanonical amino acids compatible with the Amber ff03 parameters (Khoury et al., 2014a,b).

The aim of our work was to develop and test force field parameters for phenylalanine and tyrosine derivatives, most of which are not included in the reported literature. The structures of the involved UAAs in this study are displayed in **Figure 1**. The newly developed parameters were then applied to mutant proteins or protein–ligand interactions involving UAAs, as listed in **Supplementary Table 1**, by MD simulations and molecular mechanics–Poisson Boltzmann solvent accessible surface area (MM/PBSA) calculations. Based on comparison with experimental data as the benchmark, the simulation results indicate that the new force field parameters can predict protein structures with incorporated UAAs well and generally describe the exact interplay that occurs in the binding pockets of proteins with UAAs as substrates.

SIMULATION STRATEGIES

We first constructed dipeptides of the α- and β-conformer of each UAA in the form of Ace-XXX-NMe using GaussView 6 (Dennington et al., 2016) (Step 1 in **Figure 2**). Here, XXX represents the analogs of phenylalanine and tyrosine shown in **Figure 1**. It is a popular way to employ α- and β-backbones of amino acids to fit parameters in current classical force fields, such as AMBER, CHARMM, and OPLS (Hornak et al., 2006; Best et al., 2012; Robertson et al., 2015), as these backbones dominate in the sterically allowed structural regions of the Ramachandran plot (Ramachandran et al., 1963). For the constructed dipeptides, structural optimization was performed at the B3LYP/6-31G* level, and single-point energy calculations were executed at the MP2/cc-pVTZ level using the Gaussian 16 program (Step 2 in **Figure 2**) (Frisch et al., 2016). Based on the optimized structures obtained at the B3LYP/6-31G* level, the electrostatic potential (ESP) charges at the HF/6-31G* level were further evaluated; this is a popular method to produce ESP charges because of the accurate reproduction of free energies of solvation and liquid enthalpies (Cornell et al., 1993; Wang et al., 2000). In Step 3, restrained electrostatic potential (RESP) charges were generated based on the ESP-fit charge model (Cornell et al., 1995). The general Amber force field (GAFF) is a useful molecular mechanics and is designed to be suitable for organic molecules, especially drug-like small molecules (Wang et al., 2004). In the following stage, we thus produced bonded and non-bonded parameters using GAFF based on the Antechamber tool (Case et al., 2020). The newly generated parameters can be transferred into the GMX format using the ACPYPE.py script for subsequent MD simulations in the GROMACS software package (Sousa da Silva and Vranken, 2012). In Step 5, the initial parameters of the structures of the α- and β-conformers of each UAA were tested. Accordingly, we optimized the charge parameters of the 18 analogs by estimating the relative energies of each UAA pair compared with the benchmark of quantum mechanics (QM) data at the MP2/cc-pVTZ level. In the final step, MD simulations and MM/PBSA calculations on the proteins or protein–ligand complexes involving UAAs were further performed to test the new parameters determined in this work. The complete workflow of parametrization is shown in **Figure 2**, and the detailed methodology for producing the parameters is described below.

QM Calculations

The 18 UAAs shown in **Figure 1** are analogs derived from the amino acids of phenylalanine (F) and tyrosine (Y). Based on the 18 UAAs, we constructed dipeptides of two backbone conformers for each UAA blocked with *N*-methyl and acetyl groups in the form of Ace-XXX-NMe in GaussView 6 (Dennington et al., 2016). The two backbone conformers were designed in the forms

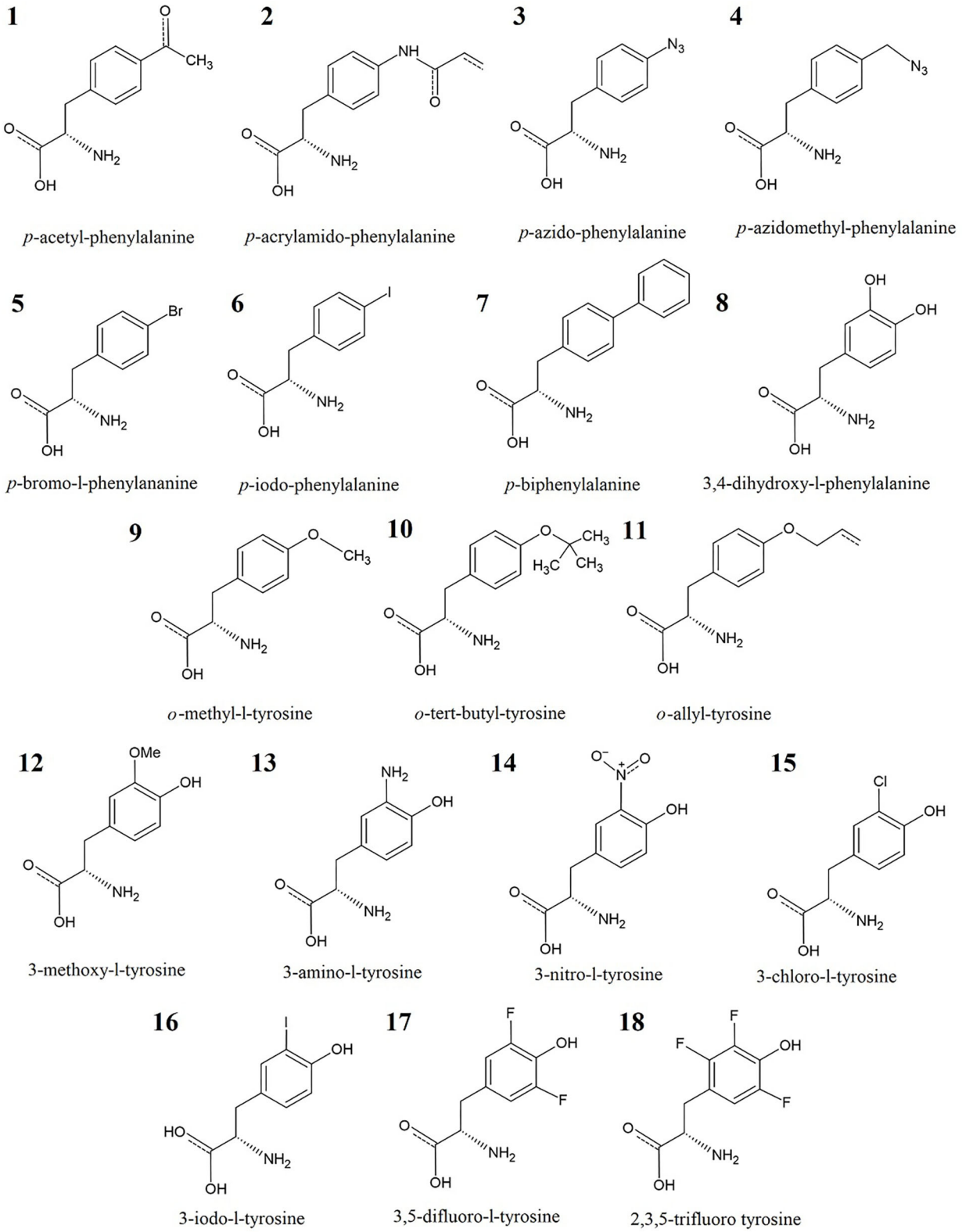


FIGURE 1 | Training set of 18 UAAs: analogs of phenylalanine (No. 1–8) and tyrosine (No. 9–18).

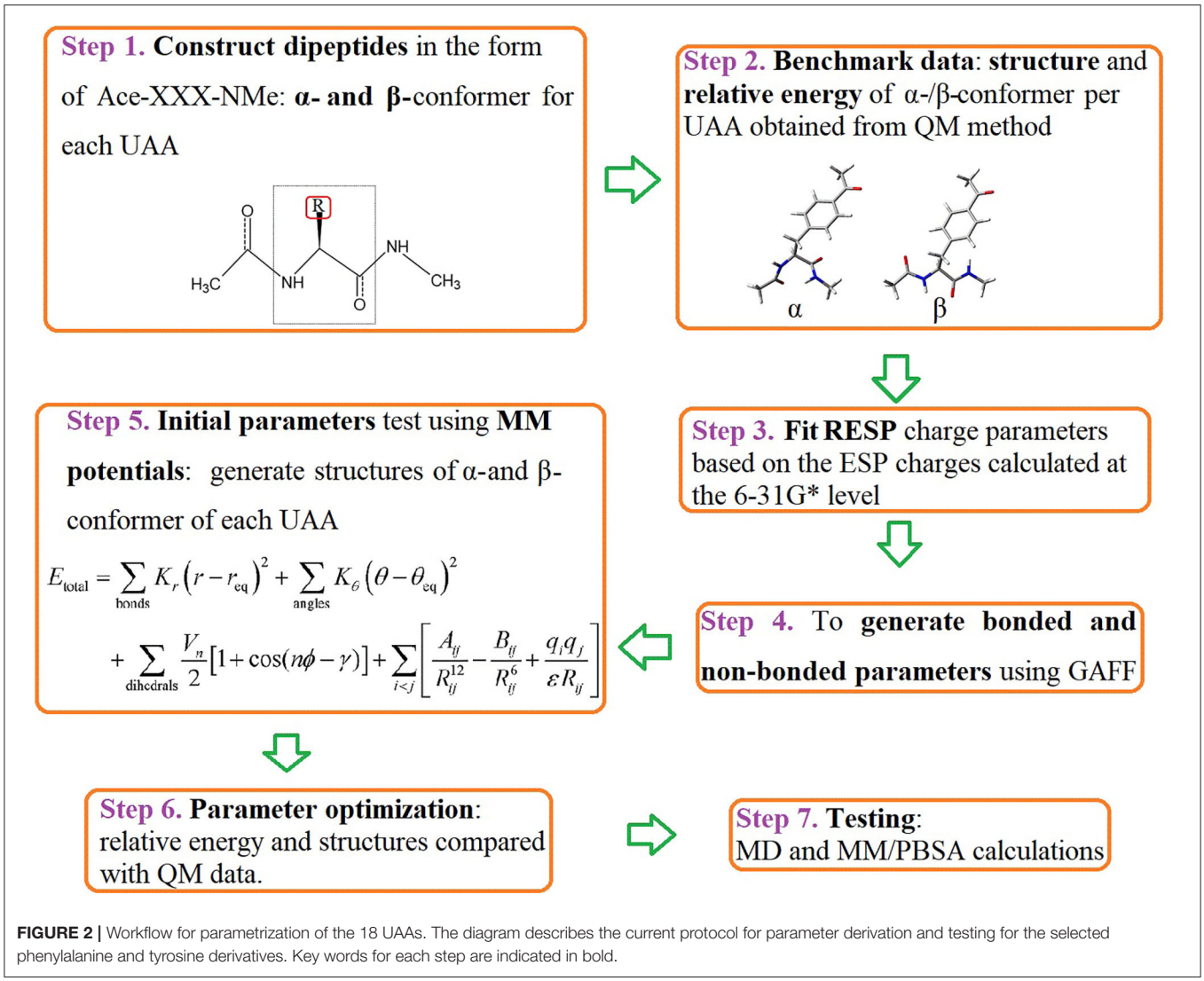


FIGURE 2 | Workflow for parametrization of the 18 UAAs. The diagram describes the current protocol for parameter derivation and testing for the selected phenylalanine and tyrosine derivatives. Key words for each step are indicated in bold.

of an α-helix ($\phi = -60°$, $\psi = -40°$) and β-strand ($\phi = -180°$, $\psi = 180°$). Structural optimizations were performed at the B3LYP/6-31G* level (Mondal et al., 2007), followed by single-point energy calculations at the MP2/cc-pVTZ level (Harder et al., 2016). For comparison, an additional method for the structure and energy calculations was performed at the M06-2X level (Robertson et al., 2015). The pseudopotential for iodine-containing systems was assigned as the SDD basis set in this work (Yurieva et al., 2008). The missing van der Waals (vdW) radius for iodine atoms was chosen as the Pauling radius (2.15 Å) (Pauling, 1939). The QM calculations were performed using the Gaussian 16 program (Frisch et al., 2016).

Energy Model

The total pair potential energy used in this work is written as a sum of terms as follows:

$$E_{total} = E_{bond} + E_{angle} + E_{dihedral} + E_{es} + E_{vdW} \quad (1)$$

Our goal is to develop UAA charge parameters that are compatible with the Amber ff14SB parameter set for the 20 natural amino acids. The energy function (Equation 1) from the Amber force field is thus employed here (Maier et al., 2015). Generally, the vdW radius and epsilon parameters are derived from experimental data (Weiner et al., 1984). The charge parameters were adjusted using the protocol described below.

Parameter Optimization

The partial charges were fitted using RESP charges obtained at the HF/6-31G* level (Cornell et al., 1993, 1995; Wang et al., 2000). The initial bonding and vdW parameters were generated from GAFF using the Antechamber module in AmberTools20 (Case et al., 2020). The charge sets of the Ace and NMe groups are identical to the Amber ff14sb force fields. Together, we used bonded and non-bonded parameters to calculate the structures and relative energies of the α- and β-conformers of the 18 UAAs. By comparing the QM structures and relative energies, we adjusted the charge parameters of the UAA backbones and

side chains until good accordance was achieved with the QM data in terms of the root-mean-square (RMS) deviation, as shown in Equation 2.

$$\mathrm{RMS} = \sqrt{\frac{\sum_{i=1}^{N}\left(RE_{QM}(i) - RE_{our}(i)\right)^2}{N}} \quad (2)$$

where $RE_{\mathrm{QM}}(i)$ and $RE_{\mathrm{our}}(i)$ are the relative energies calculated by QM and the new parameters developed in our work for the ith training set, respectively, and N is the total number of training sets. We minimized the RMS values to obtain the charge parameters.

MD Simulations

MD simulations were performed using the 2019 version of the GROMACS program (Abraham et al., 2015). We chose the Amber ff14SB force field for proteins composed of natural amino acids (Maier et al., 2015). For the UAA components, the new parameters developed in this work were used. We placed the initial systems in the center of a cubic box 10 Å from the box edge. The box was then filled with a water solvent using the TIP3P water model (Jorgensen et al., 1983). The water molecules were randomly replaced by Na^+ and Cl^- ions to a 0.1 M concentration. For each model, energy minimization with a maximum of 5,000 steps was carried out without any restraints. After optimization, two short 200 ps MD simulations in the NVT and NPT ensembles were successively performed with the heavy-atom position restraint at a force constant of 500 kcal/(mol·Å^2). The position restraints were gradually released via four steps of 100 ps NPT simulations with force constants of 250, 100, 50, and 10 kcal/(mol·Å^2) for the heavy atoms. Finally, 20 ns production MD simulations were performed in the NPT ensemble. The time step was set to 2 fs, and the temperature and pressure were kept constant at 300 K and 1 bar, respectively. In the production runs, the velocity-rescaling thermostat was applied for temperature coupling (Berendsen et al., 1984; Bussi et al., 2007), while the Parrinello–Rahman approach was applied for constant pressure control (Parrinello and Rahman, 1981; Nosé and Klein, 1983). The SHAKE algorithm was used to constrain covalent bonds involving hydrogen atoms (Andersen, 1983; Miyamoto and Kollman, 1992). The particle mesh Ewald method was applied to the calculation of long-range electrostatic interactions (Darden et al., 1993). The cutoff values for vdW and electrostatic forces were set to 12 Å, and the simulation structures were saved every 100 ps to obtain the trajectories for analysis.

MM/PBSA Estimation

In general, the binding free energy for protein–ligand interactions can be expressed as

$$\Delta G_{\mathrm{bind}} = \Delta E_{\mathrm{vdW}} + \Delta E_{\mathrm{ele}} + \Delta G_{\mathrm{solv}} - T\Delta S \quad (3)$$

where ΔE_{vdW} and ΔE_{ele} are the non-bonded terms of the system total energy (ΔE_{MM}) due to vdW and electrostatic interactions, respectively. The bonded terms of ΔE_{MM} were assumed to be zero in the single-trajectory setup used in this procedure because of its simplicity and accuracy similar to those of a multi-trajectory setup (Genheden and Ryde, 2015; Wang et al., 2018). ΔG_{solv} is the solvation free energy required to move the solute from a vacuum (dielectric constant of 1) into the solvent (dielectric constant of 80). It can be further decomposed into polar ($\Delta G_{\mathrm{pb/solv}}$) and nonpolar ($\Delta G_{\mathrm{np/solv}}$) contributions to solvation. T and ΔS are the absolute temperature and entropy, respectively. However, the entropy term was ignored in this study because of the significant time consumption, uncertainty of the contributions to the total free energy, and small improvement by comparison with the experimental results (Yang et al., 2011; Kumari et al., 2014).

Furthermore, the binding free energy decomposition of each residue was analyzed to understand the key residue impact at the activation region of the protein–inhibitor interaction. Hence, the free energy of each residue ($\Delta G_{\mathrm{res}}^{\mathrm{bind}}$) can be divided into three terms:

$$\Delta G_{\mathrm{res}}^{\mathrm{bind}} = \Delta E_{\mathrm{res}}^{\mathrm{MM}} + \Delta G_{\mathrm{res}}^{\mathrm{pb/solv}} + \Delta G_{\mathrm{res}}^{\mathrm{np/solv}} \quad (4)$$

where $\Delta E_{\mathrm{res}}^{\mathrm{MM}}$ is the sum of the electrostatic and vdW interactions per residue in a vacuum, and $\Delta G_{\mathrm{res}}^{\mathrm{pb/solv}}$ and $\Delta G_{\mathrm{res}}^{\mathrm{np/solv}}$ are the polar and nonpolar parts of the per-residue solvation free energy, respectively.

In this work, the successive 20 ns trajectories produced were used to perform MM/PBSA calculations on the free energies using the g_mmpbsa tool (Kumari et al., 2014). Here, the system coordinates were saved for every 1 ns used for MM/PBSA analysis such that 20 snapshots for each trajectory were considered to calculate the binding free energies of the protein–inhibitor interactions. The Poisson–Boltzmann (PB) equation was applied to calculate $\Delta G_{\mathrm{pb/solv}}$ (Honig and Nicholls, 1995). The temperature and grid spacing were set to 300 K and 0.5 Å, respectively, and the concentration of charged ions was 0.1 M with radii of 0.95 and 1.81 Å for Na^+ and Cl^-, respectively. The solvent accessible surface area (SASA) model was employed to estimate the nonpolar contributions ($\Delta G_{\mathrm{np/solv}}$) from the function γSASA + b (Sitkoff et al., 1994). The radius value for SASA was 1.4 Å, and the constants γ and b were set to default values of 0.00542 kcal/(mol·Å^2) and 0.92 kcal/mol, respectively.

RESULTS AND DISCUSSION

Initial Parameters Applied to α-/β-Conformer Optimization

After the initial parameters (hereafter referred to as cycle-1 parameters) involved in the bonded and non-bonded terms were generated, we performed structural optimizations for the α-and β-conformers of each UAA. For comparison with the B3LYP/6-31G* structures, we depict the optimized structures of the 18 UAA dipeptides in the α-state from the initial parameters in **Figure 3**; the minimized structures for the β-state are shown in **Supplementary Figure 1**. As shown, the two backbone conformations in the α- and β-states of the training set are in good agreement with the QM structures. The initial parameters also performed well for the side-chain structures. Additionally, the determined heavy-atom and all-atom RMS displacements (RMSDs) for the 18 training

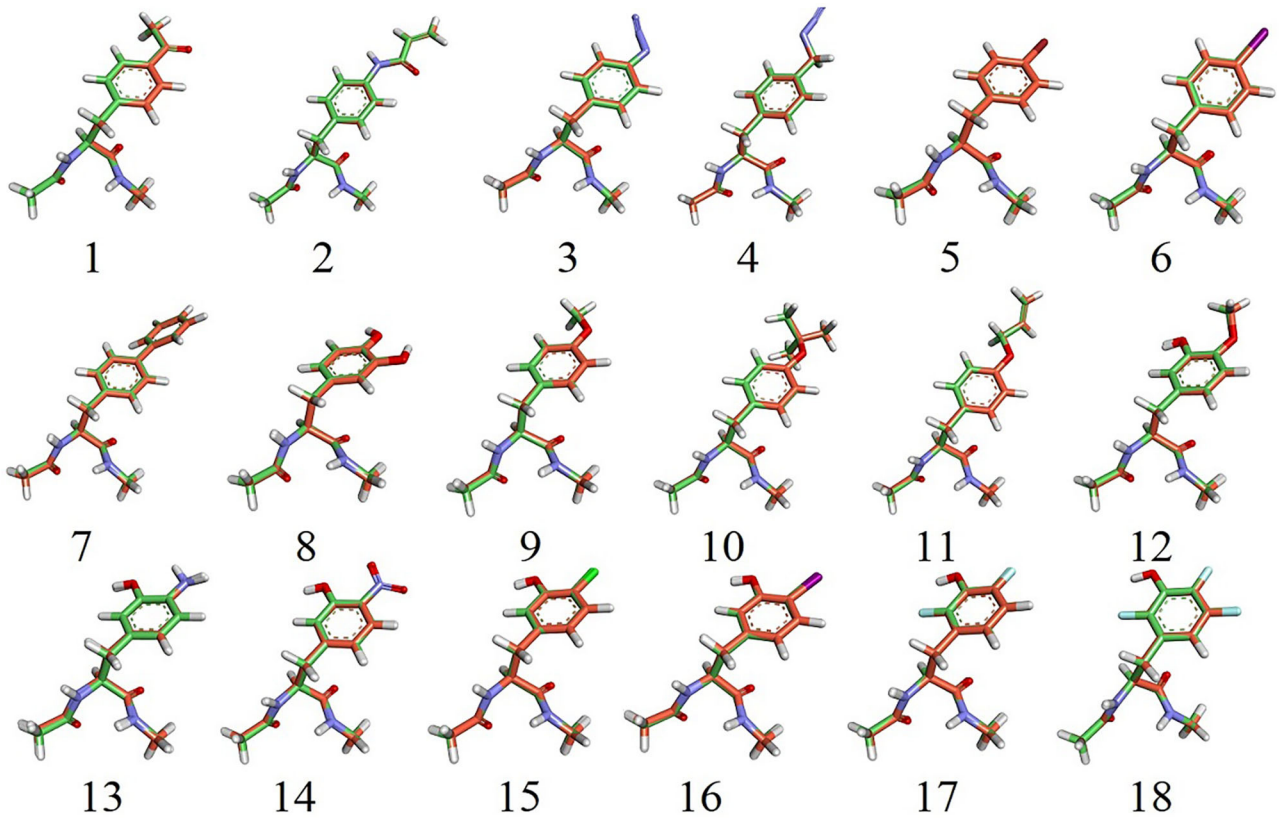


FIGURE 3 | Overlap of 18 α-backbone conformations after energy minimization of the QM (B3LYP/6-31G*) structures. N, O, and H atoms are shown in blue, red, and white, respectively. C atoms from the simulation and QM structures are shown in green and orange, respectively. F(17 and 18), Cl(15), Br(5), and I(6, 16) atoms are shown in cyan, green, red, and magenta, respectively.

sets from **Table 1** are nearly <0.1 Å (refer also to the RMSD distributions in **Supplementary Figure 2**). Among the systems, system 13 has the greatest RMSDs of 0.083–0.116 Å. Meanwhile, **Supplementary Figure 2** shows that the all-atom RMSDs are comparable to the heavy-atom RMSDs but fluctuate to slightly higher values. Overall, the initial parameters yield good results for the 18 training sets, especially the bonded connections, but further improvements to the energies are necessary.

Testing of Optimized Parameters

Displayed in **Table 2** are the relative energies for the 18 training sets. We selected the relative energies evaluated at the MP2/cc-pVTZ//B3LYP/6-31G* level of theory as a benchmark (Mondal et al., 2007; Harder et al., 2016). For comparison, one density functional theory (DFT) method with a small basis set at the M06-2X/6-311++G**//M06-2X/6-31+G* level was used in this work (Robertson et al., 2015). For the parameter optimization process, four cycles were performed. First, we fixed the charges of the Ace and NMe groups in the 18 UAA dipeptides to remain the same as the corresponding Amber ff14sb force field parameter sets and made minor adjustments to the backbone RESP charges. As shown in **Table 2**, the relative energies from the cycle-1 parameters show a correlation of 0.8212 compared with the MP2 energies, with a larger RMS deviation of 4.86 kcal/mol. In the next two cycles, we chose to treat the backbones and side chains as α-helical RESP charges and averaged RESP charges in the α- and β-states, respectively. In the third cycle adjustment, the RMS decreased to 2.33 kcal/mol with a 0.8072 correlation. At this point, we noted that the relative energies of most systems were comparable to the benchmark data except for those of systems 4, 9–12, 17, and 18. Therefore, the parameters for these systems were further optimized. In the final procedure, we chose the β-conformational charges from the first cycle as the determined parameters for systems 4 and 18. For systems 9–12 and 17, different proportions between the α- and β-conformational RESP charges were ultimately treated (see the footnotes in **Table 2**). For the remaining systems, we employed the averaged charges in the α- and β-states. Eventually, we observed a strong correlation between our work and the QM data, with $R^2 = 0.9407$ (**Supplementary Figure 3**). Therefore, the parameters from the fourth cycle were employed in subsequent calculations. Although the partial charges were obtained by fitting to the RESP of independent conformations for each UAA, the partial charges of the atoms in their common structures are quite close to each other (see **Supplementary Material**). Note that these UAAs are phenylalanine and tyrosine derivatives and share a common structure. The observation of such small differences

TABLE 1 | Initial parameter test for the 18 training sets evaluated by heavy-atom and all-atom RMSDs.

Training set	Heavy-atom RMSD (Å)		All-atom RMSD (Å)	
	α-Backbone	β-Backbone	α-Backbone	β-Backbone
1	0.041	0.052	0.045	0.059
2	0.072	0.055	0.083	0.060
3	0.054	0.052	0.051	0.053
4	0.046	0.039	0.043	0.046
5	0.068	0.052	0.067	0.058
6	0.080	0.042	0.079	0.048
7	0.084	0.041	0.082	0.053
8	0.084	0.045	0.082	0.057
9	0.046	0.046	0.048	0.053
10	0.044	0.042	0.048	0.048
11	0.054	0.045	0.060	0.051
12	0.064	0.048	0.068	0.059
13	0.096	0.083	0.116	0.094
14	0.056	0.047	0.058	0.055
15	0.073	0.055	0.074	0.063
16	0.072	0.058	0.075	0.065
17	0.067	0.040	0.069	0.046
18	0.046	0.039	0.047	0.045

indicates that the obtained RESPs for these UAAs are reliable and the charge parameters are well converged.

In addition, we noted that the β-backbone conformation of each UAA is more stable than the α-backbone as predicted by all employed methods. Here, our work shows a more favorable RMS deviation of 0.33 kcal/mol compared with M06-2X with an RMS deviation of 1.08 kcal/mol. Additionally, existing charge parameters from a reference were also tested on reported systems 7, 9, 11, and 13 (Khoury et al., 2014b). The relative energies of these four systems are 5.74, 8.52, 8.05, and 4.40 kcal/mol obtained from the reported parameters, which are comparable to those from the MP2 data of 6.76, 6.86, 6.79, and 8.26 kcal/mol, respectively, but produce absolute errors of approximately 1.5 kcal/mol or higher (**Supplementary Table 2**). Compared with the MP2 energies, as also shown in **Supplementary Table 2**, the RMS deviations obtained from the reference and our work were 2.25 and 0.38 kcal/mol, respectively, for these four systems. Therefore, the energetic performance of the new parameters determined in our work results in more satisfactory predictions. In addition, the structural optimizations from the cycle-4 parameters were again tested on the 18 dipeptides in the α- and β-states. Comparisons of heavy-atom RMS distributions between cycles 1 and 4 are provided in **Supplementary Figure 4**, which clearly shows that the new parameters produce smaller heavy-atom RMS deviations than the initial parameters. Overall, the new parameters show a good performance in terms of structural optimization and relative energy calculations for the 18 UAA models based on comparison with QM results, indicating that the new parameters determined in this work are appropriate for performing further tests via MD simulations.

TABLE 2 | Relative energies (kcal/mol) for the α- and β-conformers of the 18 UAA dipeptides obtained from QM calculations and our work.

Training set	MP2[a]	M06-2X[b]	Our work			
			Cycle 1	Cycle 2	Cycle 3	Cycle 4[c]
1	6.50	7.39	11.07	2.61	6.72	6.72
2	6.93	7.72	10.78	3.34	7.12	7.12
3	6.35	7.13	11.14	1.96	6.70	6.70
4	6.05	7.30	6.63	0.01	3.52	6.63
5	6.45	7.28	10.11	2.01	6.31	6.31
6	6.53	7.38	10.56	2.44	6.55	6.55
7	6.76	7.58	10.36	2.77	6.76	6.76 (5.74)
8	6.77	7.54	10.06	3.18	7.06	7.06
9	6.86	7.62	12.78	5.56	9.57	7.00 (8.52)
10	6.89	9.74	12.40	5.54	9.06	6.77
11	6.79	7.54	12.32	4.33	8.22	7.10 (8.05)
12	8.02	8.90	15.12	6.34	11.08	7.93
13	8.26	8.99	13.78	4.05	8.94	8.94 (4.40)
14	7.12	8.08	14.38	2.44	7.70	7.70
15	7.44	8.37	12.76	1.80	7.58	7.58
16	7.53	8.42	12.88	2.69	7.86	7.86
17	5.07	5.54	6.70	−6.75	0.47	4.96
18	4.80	5.34	4.45	−7.52	−1.53	4.45
[d]RMS MP2	–	1.08	4.86	5.68	2.33	0.33
[e]R^2 (MP2)		0.7831	0.8212	0.6922	0.8072	0.9407

[a] *MP2/cc-pVTZ//B3LYP/6-31G*.*
[b] *M06-2X/6-311++G**//M06-2X/6-31+G*.*
[c] *Values in parentheses were obtained using charge parameters taken from the literature (Khoury et al., 2014b). The different proportions of charge parameters in the final cycle are β/6 + α*5/6 for system 9, β/5 + α*4/5 for system 10, β/3 + α*2/3 for system 11, β/6 + α*5/6 for system 12, β*7/8 + α/8 for system 17, β for systems 4 and 18, and β/2 + α/2 for the other systems.*
[d] *All units for RMS deviations are kcal/mol.*
[e] *Correlation between MP2 and the other methods involved in M06-2X and our work. The relative energy (RE) is defined as* $E_\alpha - E_\beta$.

Testing

MD Simulations of Proteins Containing UAAs

Seven isolated protein systems containing UAAs were selected to identify the new parameters as the testing set. At present, crystals composed of noncanonical amino acids have rarely been recorded in the PDB. We attempted to search for the protein structures covering UAAs related to phenylalanine and tyrosine, which are T4 lysozyme (PDB ID: 3HWL) (Fleissner et al., 2009), CaM-peptide (PDB ID: 6HCS) (Creon et al., 2018), modified threonyl-tRNA synthetase (PDB ID: 4S0I) (Pearson et al., 2015), sphingosine-1-phosphate lyase (PDB ID: 3MBB) (Bourquin et al., 2010), birch pollen allergen Bet v 1.0101 (PDB ID: 4B9R) (Ackaert et al., 2014), ketosteroid isomerase (PDB ID: 5D82) (Wu et al., 2015), and acetyltransferase (PDB ID: 2Z10) (Sakamoto et al., 2009). Each protein mainly contains one UAA, dominated by secondary structures of α-helices, β-sheets, and γ-turns made of natural amino acids. Among them, the UAAs ACF131, AZF108, NIY150, CHY16, and IOY111 incorporated in the T4 lysozyme, CaM-peptide, Bet v 1.0101, ketosteroid isomerase,

and acetyltransferase, respectively, are mainly located at the α-helices. BFA11 and NIY5, 66, and 83 of the threonyl-tRNA synthetase and Bet v 1.0101 are distributed in the β-sheet regions, and AMY249 of sphingosine-1-phosphate lyase is located at the γ-turn. The remaining five systems involving UAAs recorded in the PDB are in regard to the protein–ligand interactions (**Supplementary Table 1**) and will be discussed in the next section, "MM/PBSA analysis of protein–UAA interactions."

Here, we show each UAA fragment from the final MD structure compared with the crystal structure in **Figure 4**, and **Table 3** displays the averaged heavy-atom RMSD values of the UAAs and corresponding proteins in isolated systems. As shown in **Figure 4**, the backbone and side chains of the UAAs in the isolated proteins are generally well overlapped with the experimental structures, although there are slight structural derivations in the fragments of the ACF131 backbone and NIY66 side chain. **Table 3** also shows that the averaged RMSDs for ACF131 and NIY66 are the largest at 1.31 ± 0.11 and 0.95 ± 0.20 Å, respectively, corresponding to moderate RMSDs of 1.94 ± 0.24 and 2.43 ± 0.29 Å for their whole proteins. Simultaneously, we inspected the UAA motions by referring to one equilibrium structure during MD simulations, as listed in **Table 3** (column 3). Almost all the RMSDs of the UAAs are under 0.5 Å, with only NIY5 showing a larger RMSD of 0.66 ± 0.30 Å. In addition, we plotted the RMSD distributions for each UAA in the isolated protein systems compared with the crystal structures over time in **Supplementary Figure 5**. As shown, each trajectory reaches a balance after 20 ns MD simulations. However, the RMSD of NIY5 decreased by <0.5 Å between 5 and 10 ns. After 10 ns, all the UAAs reached equilibrium with RMSDs under 1.5 Å.

Additionally, the backbone conformations of seven UAAs in their isolated proteins obtained from the new parameters during the MD simulations were further investigated (**Figure 5**). As shown in **Figure 5E**, only NIY5 of birch pollen allergen Bet v 1.0101 (PDB ID: 4B9R) inclines toward the more stretched β-sheet backbone conformation during the MD simulation (black symbols). Compared to the crystal structures, the calculated backbone torsions of the remaining UAAs are generally well consistent. The backbones of ACF131, AZF108, AMY249, NIY150, CHY16, and IOY111 are in the form of α-helices, while those of BFA and NIY5, 66, and 83 are formed by β-strands.

MM/PBSA Analysis of Protein–UAA Interactions

Aside from the isolated proteins containing UAAs found in the PDB search, the UAAs were resolved as a ligand role in protein–UAA interactions (Turner et al., 2006; Moor et al., 2011; Takimoto et al., 2011; Li et al., 2013). To evaluate the quality of the new parameters determined in this work, five systems of protein–UAA interactions were studied by MM/PBSA analysis. The complexes are *p*-bromo-l-phenylalanine (BRF) bound to aaRS (PDB ID: 2AG6) (Turner et al., 2006); $tRNA^{Phe}$ with 3,4-dihydroxy-l-phenylalanine (DHF) (PDB ID: 3TEG) (Moor et al., 2011); evolved PylRS charged with *o*-methyl-l-tyrosine (OMY) (PDB ID: 3QTC) (Takimoto et al., 2011); tyrosine-tRNA ligase mutant complexed with 3-methyl-tyrosine (MEY) (PDB ID: 4HPW); and 3,5-difluoro-l-tyrosine (DFY) incorporated into

TABLE 3 | Averaged heavy-atom RMSDs (Å) with standard errors of the mean for the UAAs and corresponding proteins in isolated systems.

UAA	RMSD (Å)			PDB ID
	UAA fit to crystal UAA	UAA fit to MD UAA	Protein fit to backbone	
ACF131	1.31 ± 0.11	0.39 ± 0.39	1.94 ± 0.24	3HWL
AZF108	0.75 ± 0.11	0.28 ± 0.13	3.44 ± 0.31	6HCS
BFA11	0.22 ± 0.06	0.23 ± 0.07	1.75 ± 0.15	4S0I
AMY249	0.42 ± 0.12	0.23 ± 0.09	4.36 ± 0.48	3MBB
NIY5	0.72 ± 0.29	0.66 ± 0.30	2.43 ± 0.29	4B9R
NIY66	0.95 ± 0.20	0.40 ± 0.25	2.43 ± 0.29	4B9R
NIY83	0.30 ± 0.09	0.23 ± 0.08	2.43 ± 0.29	4B9R
NIY150	0.28 ± 0.07	0.31 ± 0.09	2.43 ± 0.29	4B9R
CHY16	0.26 ± 0.01	0.24 ± 0.10	2.80 ± 0.25	5D82
IOY111	0.15 ± 0.04	0.14 ± 0.04	2.02 ± 0.20	2Z10

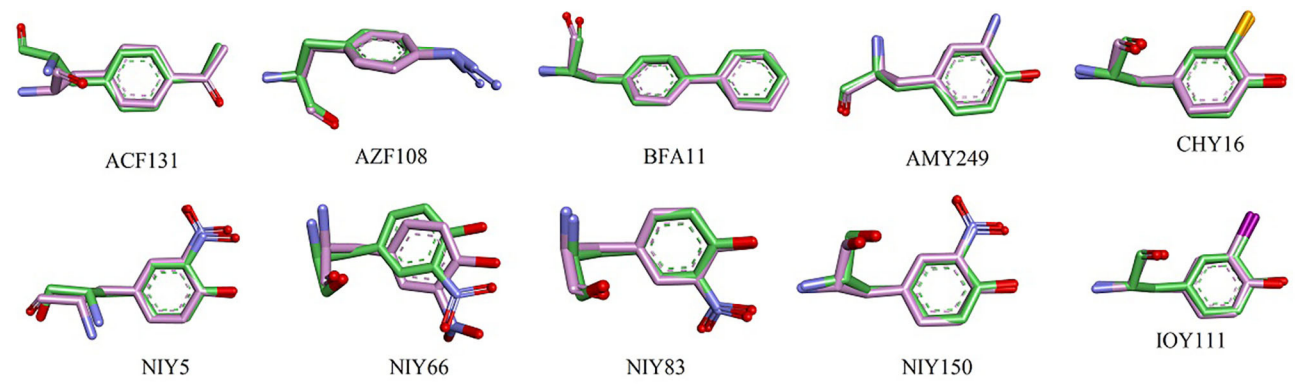


FIGURE 4 | Structural alignment between crystal and MD stable structures of single UAAs in isolated protein systems. C atoms from the crystal and MD stable structures are shown in green and pink, respectively. All N and O atoms are blue and red, and Cl and I atoms from CHY16 and IOY111 are orange and magenta, respectively.

tyrosine phosphorylation (PDB ID: 4HJX) (Li et al., 2013). In addition, we added H and OH groups to the -NH and -C=O termini, respectively, to achieve neutral UAA ligands. We made minor modifications to the charge parameters of the terminal H and OH groups; the modified terminal charges for the H and OH groups of the five ligands BRF, DHF, OMY, MEY, and DFY are listed in **Supplementary Table 3**. These charge parameters should be more appropriate for UAAs when they are treated as ligands.

Compared with the UAAs incorporated into isolated proteins, the UAAs involved as substrates in protein–ligand interactions seem to shift more obviously, particularly the backbone structures (**Figure 6**). This may be due to the flexible UAA structures acting as ligands to bind with the proteins. The average RMSD values were also calculated to be larger, around 1.3 Å, as listed in **Table 4**. After choosing one equilibrium MD structure as the reference, the averaged RMSDs for all UAA ligands decreased to below 1.0 Å, suggesting good stability in the simulation process. **Supplementary Figure 6** plots the RMSD distributions as a function of time for the five ligands BRF, DHF, OMY, MEY, and DFY during the MD simulation starting from the experimental structure set. As shown, the DHF, OMY, MEY, and DFY ligands were well-balanced after 3 ns, whereas BRF reached another stable state after 10 ns. The RMSD values of all the UAA ligands are in the vicinity of 1.5 Å, showing stable movements over the initial structures. The final whole structures also overlap well with the crystal structures, as depicted in **Supplementary Figures 7H–L**, indicating that our new parameters can reproduce the experimental structures of these protein–UAA interactions.

TABLE 4 | Averaged heavy-atom RMSDs (Å) with standard errors of the mean for UAAs and corresponding proteins in protein–ligand complexes.

UAA	RMSD			PDB ID
	UAA fit to crystal UAA	UAA fit to MD UAA	Protein fit to backbone	
BRF	0.96 ± 0.37	0.83 ± 0.51	2.23 ± 0.15	2AG6
DHF	1.33 ± 0.16	0.81 ± 0.49	2.42 ± 0.15	3TEG
OMY	1.33 ± 0.27	0.87 ± 0.46	2.38 ± 0.20	3QTC
MEY	1.31 ± 0.38	0.61 ± 0.44	2.44 ± 0.22	4HPW
DFY	1.11 ± 0.27	0.47 ± 0.31	3.29 ± 0.30	4HJX

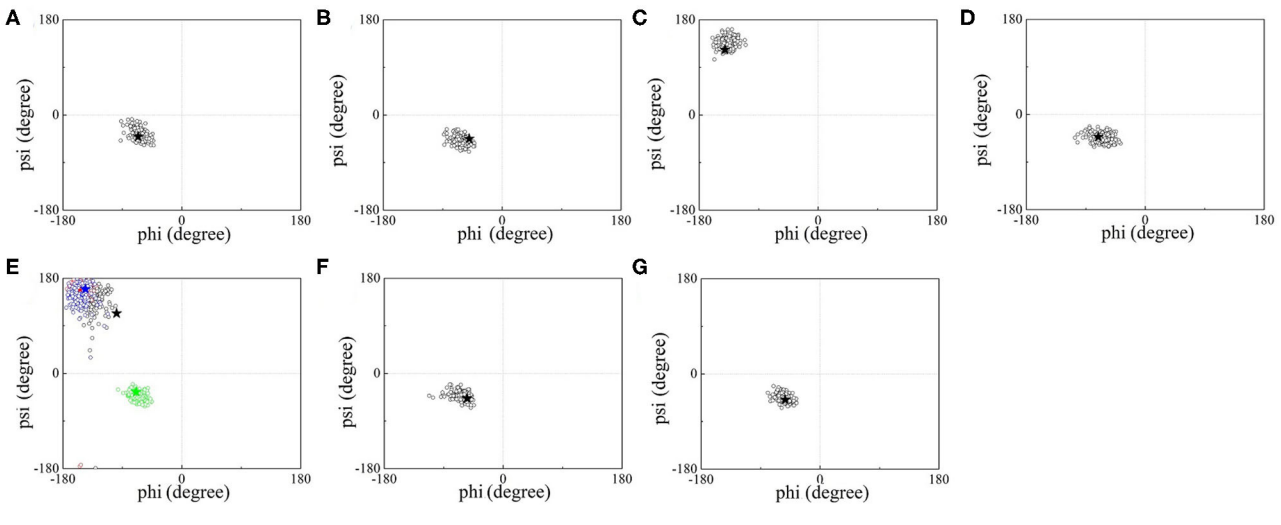


FIGURE 5 | ϕ/ψ backbone torsional statics for **(A)** ACF, **(B)** AZF, **(C)** BFA, **(D)** AMY, **(E)** NIY, **(F)** CHY, and **(G)** IOY during the MD simulations. Black hollow circles describe the torsional distributions of ϕ/ψ over time, and black stars indicate the crystal data for the corresponding UAAs. Four NIY structures are contained in **(E)**, where black, red, blue, and green circles correspond to the backbone torsional distributions of NIY5, 66, 83, and 150, respectively, and the four colored stars indicate the corresponding crystal structures.

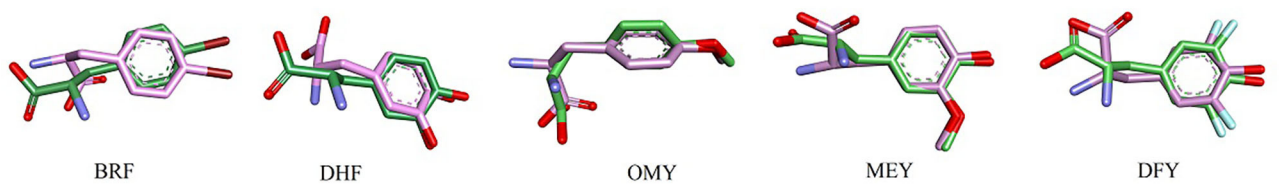


FIGURE 6 | Structural alignment between crystal and MD stable structures of single UAAs in protein–ligand interactions. F and Br atoms from DFY and BRF are light blue and dark red, respectively. The colors of other atoms are the same as in **Figure 4**.

Further, we used the MD structures obtained from the new parameters to calculate the five protein–UAA complexes. **Table 5** shows the binding free energies of aaRS–BRF (PDB ID: 2AG6), $tRNA^{Phe}$-DHF (PDB ID: 3TEG), PylRS–OMY (PDB ID: 3QTC), $tRNA^{Tyr}$-MEY (PDB ID: 4HPW), and tyrosine phosphorylation (F2YRS)–DFY (PDB ID: 4HJX) based on MM/PBSA analysis. The binding free energy of 3TEG is the highest at −21.3 kcal/mol, while the weakest binding affinity of −6.4 kcal/mol corresponds to 2AG6. As mentioned above, the RMSD of BRF reaches another local equilibrium within a period of 12–20 ns. To check the convergence of the MM/PBSA calculation, the results were usually estimated from different time intervals (Spiliotopoulos et al., 2012). As shown in **Supplementary Table 4**, the binding free energy of 2AG6 estimated using the 12–20 ns trajectory is −7.1 kcal/mol, which is largely consistent with the one obtained using entire trajectory and the one in an early stage. We also predicted the binding free energies of $tRNA^{Tyr}$-MEY (4HPW) and tyrosine phosphorylation–DFY (4HJX) as −17.3 and −13.6 kcal/mol, respectively. The energy decomposition analysis also indicates that vdW and electrostatic interactions are the dominate factors contributing to the total binding free energy. The polar energies contribute positively to the solvation. Overall, the binding free energies for the five systems were well stabilized by the MM contributions.

In addition, the binding affinities from experimental data for two of the complexes are available. One is the reported Michaelis constant K_M for the system $tRNA^{Phe}$-DHF (3TEG) of 380 ± 40 μM (Moor et al., 2011), which determines the performance of the catalytic reaction and positively correlates with the dissociation constant K_d (Johnson and Goody, 2011). The other is for OMY as one of the compstatin variants reported as $K_d = 118$ nM and $\Delta G = -9.5 \pm 1.2$ kcal/mol (Magotti et al., 2009). The binding free energy of PylRS and OMY interaction is predicted to be −15.7 ± 0.7 kcal/mol by MM/PBSA, which is in satisfying agreement with the experimental one. No experimental data of the binding affinity is available to date for the other three protein-UAA systems (PDB IDs: 2AG6, 4HPW, and 4HJX). Nevertheless, MM/PBSA analysis has been demonstrated to be an effective approach to estimate qualitatively the relative binding free energy of protein-ligand interaction (Homeyer and Gohlke, 2012; Kumari et al., 2014; Genheden and Ryde, 2015; Wang et al., 2019).

TABLE 5 | Binding free energies (kcal/mol) with standard deviation[a] for the systems 2AG6, 3TEG, 3QTC, 4HPW, and 4HJX obtained from MM/PBSA calculations and various energy components.

Component	2AG6	3TEG	3QTC	4HPW	4HJX
ΔE_{vdW}	−25.2 (0.4)	−24.3 (0.5)	−20.6 (1.4)	−25.4 (0.9)	−26.4 (0.6)
ΔE_{ele}	−11.8 (1.5)	−35.6 (1.7)	−12.7 (1.5)	−20.2 (1.9)	−18.8 (1.5)
$\Delta G_{pb/solv}$	33.6 (1.5)	41.4 (1.4)	20.0 (1.1)	31.1 (1.6)	34.7 (1.5)
$\Delta G_{np/solv}$	−3.0 (0.0)	−2.7 (0.0)	−2.4 (0.1)	−2.9 (0.0)	−3.1 (0.0)
ΔG_{pb}	21.8 (0.0)	5.8 (0.2)	7.3 (0.2)	10.9 (0.2)	15.9 (0.0)
ΔG_{np}	−28.2 (0.2)	−27.0 (0.3)	−23.0 (0.8)	−28.3 (0.6)	−29.5 (0.3)
ΔE_{MM}	−37.0 (0.7)	−59.9 (0.8)	−33.3 (0.0)	−45.6 (0.6)	−45.2 (0.5)
ΔG_{solv}	30.6 (0.9)	38.7 (0.8)	17.6 (0.6)	28.2 (0.9)	31.6 (0.9)
ΔG_{bind}	−6.4 (0.8)	−21.2 (1.1)	−15.7 (1.0)	−17.4 (1.1)	−13.6 (0.8)

[a] *The standard deviations are calculated by the equation* $SD = \sqrt{\sum_{i=1}^{N} (x_i - \overline{x_i})^2 / N}$.

Here, $\Delta G_{pb} = \Delta G_{pb/solv} + \Delta E_{ele}$; $\Delta G_{np} = \Delta G_{np/solv} + \Delta E_{vdW}$; $\Delta E_{MM} = \Delta E_{vdW} + \Delta E_{ele}$; $\Delta G_{solv} = \Delta G_{pb/solv} + \Delta G_{np/solv}$.

Per-Residue Energy Decomposition Analysis of Protein–UAA Interactions

The structural interaction modes between UAAs and proteins have been established by experimental reports (Turner et al., 2006; Moor et al., 2011; Takimoto et al., 2011; Li et al., 2013). We show the interaction details of the UAAs BRF, DHF, OMY, MEY, and DFY as substrates bound to the respective proteins in **Figure 7**. The per-residue binding free energies of the major contacts involved in the interactions are provided in **Table 6**. The interactions of the UAAs as substrates are discussed in the following sections.

aaRS and BRF interactions

In the 2AG6 system, our parameters predicted several direct connections of C–halogen-bonding interactions, which are consistent with the experimental results (**Figure 7A**). For example, the bromine of BRF forms a C-Br · · · π interaction with WT H160, which has been extensively reported in the crystal structures of protein–small molecules (Saraogi et al., 2003; Turner et al., 2006). One crystal structure report showed that the mutant L32 is a key mutant residue providing binding room for the bromine without vdW contributions (Turner et al., 2006). Here, the small contribution is −0.45 kcal/mol of free energy as predicted by our new parameters (see **Table 6**). In addition, we did not observe obvious contact between WT Y161 and BRF, and the predicted binding free energy was 0.47 kcal/mol, with weak contributions from MM, polar, and nonpolar interactions of −0.84, 1.47, and −0.17 kcal/mol, respectively. This is consistent with the experimental finding that the O atom of Y161 is too far (4.6 Å) to form H-bonded contact with the Br-atom of BRF in the active loop (Turner et al., 2006). In addition, two potential H-bonded contacts that have not been anticipated experimentally are predicted here. In particular, WT E36 and WT Q173 of 2AG6 use side chains to combine with the amide group of BRF in the form of H-bonds. As shown in **Table 6**, Q173 produces stronger polar interactions than the MM component, leading to a positive contribution of 5.46 kcal/mol. Strong electrostatic and vdW interactions of −6.29 and −4.31 kcal/mol were calculated for both E36 and Q173, respectively, which provide important conditions for H-bond formation (Li et al., 2014; Hao and Wang, 2015).

$tRNA^{Phe}$ and DHF interactions

We provide the structural basis of the reported 3TEG ($tRNA^{Phe}$ binding with DHF) in **Figure 7B**. F232 and F234 located at the FPF loop maintain major contacts with the phenyl ring of the ligand DHF (Moor et al., 2011). This was also observed from our predictions between F232/F234 and DHF in the form of π · · · π interactions. Simultaneously, F276 has a novel predicted role

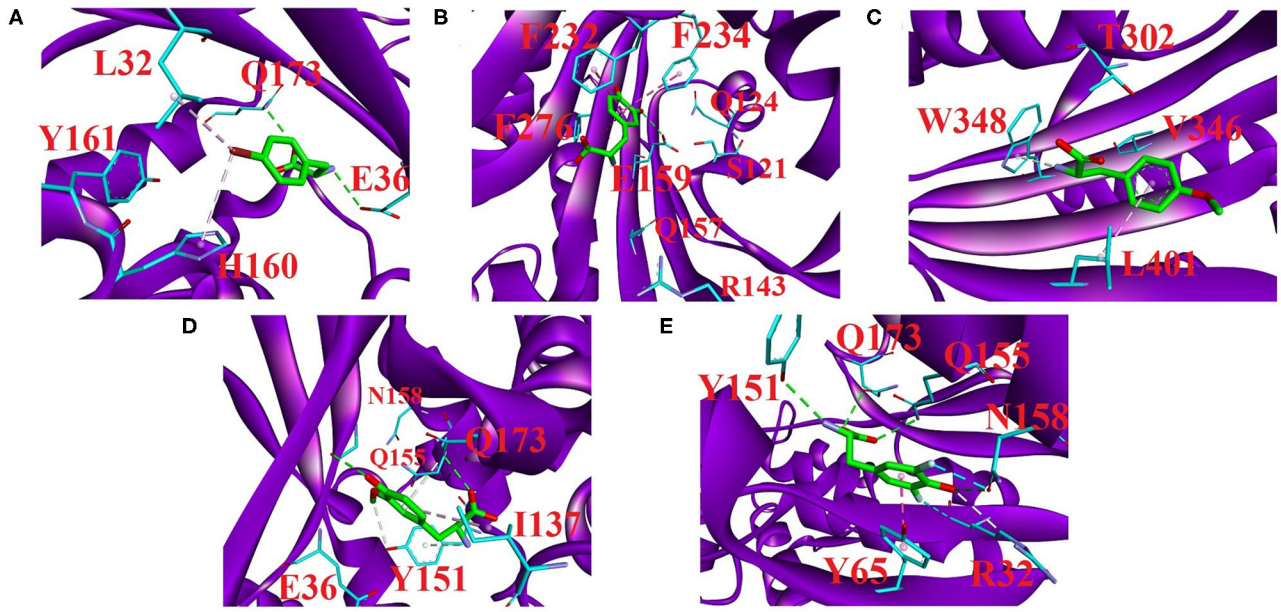


FIGURE 7 | Major residue contacts with the substrates **(A)** BRF, **(B)** DHF, **(C)** OMY, **(D)** MEY, and **(E)** DFY in the active sites obtained by our predictions. The interaction analysis was completed using Discovery Studio 4.5 (BIOvIA, 2015). All C atoms in the substrates are shown as green sticks. The major interaction residues in the proteins (purple cartoon) are displayed as sticks with cyan C atoms. Red letters label the major residue names. Green dashed lines represent hydrogen bonding, and pink dashed lines represent the π-interaction involved in ligand recognition. The light pink in **(A)** and cyan in **(E)** dashed lines represent the halogen bonding occurring in the active regions.

involved in DHF binding through an amide · · · π interaction (see **Supplementary Figure 8A**). These π-interaction modes formed by F232, F234, and F276 are similar to the reported "edge-to-face" contact (Fishman et al., 2001), an interaction network formed by three phenylalanine in $tRNA^{Phe}$ binding with the phenyl moiety of the DHF ligand. The π · · · π and amide · · · π interactions mainly originate from vdW contributions (Gao et al., 2017). As shown in **Table 6**, the total vdW and electrostatic contributions of F232, F234, and F276 are all more negative than −4.0 kcal/mol. Furthermore, the binding free energy of E159 is a remarkable −16.57 kcal/mol, with surprisingly large non-bonded and polar contributions of −55.39 and 39.20 kcal/mol, respectively. As shown in **Figure 7B**, one hydrogen bonding connection occurs through the side-chain O atom of E159 with the negative charge binding to the OH group of DHF. Additionally, H-bonded connections have been reported between S121, Q124, R143, and Q157 in the protein and DHF shown in **Supplementary Figure 8B** (Moor et al., 2011). However, we failed to observe these hydrogen bonding contacts. Per-residue energy decomposition analysis further indicates that only R143 and Q157 provide dispensable non-bonded interactions of −3.66 and −8.86 kcal/mol, respectively. The contributions of S121 and Q124 are almost too weak for binding.

PylRS and OMY interactions

The four residue mutations in PylRS are A302T, N346V, C348W, and V401L, which play a vital role in the OMY selectivity (Takimoto et al., 2011). We also predicted these four important residue contacts with OMY based on the new parameters and per-residue binding free energy analysis. **Figure 7C** shows the interaction modes between OMY and the four residues T302, V346, W348, and LV401, and the binding free energy of each residue (PDB ID: 3QTC) is listed in **Table 6**. As shown, W348 uses a side-chain 5-membered ring as a π-donor to form hydrogen bonds with the N-atom in the amide group of OMY. This results in one quadrupole–dipole interaction formed by the indole plane of W348 being vertical to the O-methyl moiety of OMY (Takimoto et al., 2011). The binding free energy of W348 is −2.99 kcal/mol, providing strong vdW and electrostatic interactions of −4.53 kcal/mol. In the activation region, alkyl · · · π interactions occur by the methylene group of L401 binding with OMY, with the highest binding affinity contribution of −3.91 kcal/mol. Even though no direct connection forms between T302 and OMY, a moderate impact with a −2.11 kcal/mol binding free energy was evaluated, which also provides strong electrostatic and vdW interactions of −7.32 kcal/mol. In addition, the binding contribution of V346 is mainly derived from electrostatic and vdW contributions at −2.03 kcal/mol, but we did not observe hydrogen bonding between them. This is in agreement with the experimental observation that the H-bonds formed by WT N346 and OMY are abolished after the N346V mutation in PylRS (Takimoto et al., 2011).

$tRNA^{Tyr}$ and MEY interactions

The structural basis for MEY recognition to $tRNA^{Tyr}$ has not been reported to date, but the binding modes of the

TABLE 6 | Energy decomposition analysis of 2AG6, 3TEG, 3QTC, 4HPW, and 4HJX for major residues.

PDB ID	#Residue	Energy component (kcal/mol)				Standard deviation (kcal/mol)			
		MM	Polar	Non polar	Total	MM	Polar	Non polar	Total
2AG6	**L32**	−1.26	1.15	−0.34	−0.45	0.38	0.31	0.03	0.48
	E36	−6.29	5.77	−0.33	−0.78	2.98	2.70	0.05	0.99
	H160	−2.24	4.85	−0.21	2.39	0.32	0.83	0.04	0.68
	Y161	−0.84	1.47	−0.17	0.47	0.12	0.20	0.02	0.19
	Q173	−4.31	10.22	−0.44	5.46	0.73	1.43	0.05	0.82
3TEG	**S121**	0.10	−0.40	−0.07	−0.36	0.21	0.38	0.02	0.25
	Q124	1.96	−0.39	−0.02	1.53	0.38	0.58	0.01	0.30
	R143	−3.66	7.75	−0.08	3.91	1.74	1.60	0.03	0.69
	Q157	−8.86	10.66	−0.58	1.18	1.12	0.85	0.04	0.75
	E159	−55.39	39.20	−0.38	−16.57	1.94	2.80	0.02	1.69
	F232	−4.07	2.03	−0.47	−2.53	0.60	0.38	0.03	0.35
	F234	−4.48	3.39	−0.55	−1.66	0.35	0.24	0.04	0.23
	F276	−6.78	3.94	−0.09	−2.91	0.40	0.40	0.02	0.45
3QTC	**T302**	−7.32	5.55	−0.39	−2.11	1.31	0.94	0.06	0.47
	V346	−2.03	0.26	−0.08	−1.84	0.34	0.09	0.02	0.38
	W348	−4.53	1.68	−0.15	−2.99	1.16	0.44	0.04	0.75
	L401	−4.10	0.68	−0.50	−3.91	0.77	0.19	0.07	0.67
4HPW	**E36**	−14.75	9.85	−0.52	−5.34	1.79	1.69	0.07	1.21
	I137	−4.56	0.75	−0.62	−4.43	0.37	0.35	0.04	0.25
	Y151	−13.06	7.57	−0.68	−6.17	0.95	0.78	0.06	1.09
	Q155	−9.69	8.13	−0.32	−1.90	1.19	0.96	0.05	0.45
	N158	−0.17	0.49	−0.02	0.31	0.21	0.49	0.02	0.34
	Q173	−17.58	16.48	−0.81	−1.97	1.55	0.87	0.06	1.02
4HJX	**R32**	−7.32	8.48	−0.27	0.82	0.71	1.19	0.03	0.98
	Y65	−7.78	4.41	−0.58	−3.91	0.68	0.47	0.05	0.56
	Y151	−6.95	6.62	−0.45	−0.76	0.65	0.62	0.05	0.50
	Q155	−23.04	21.54	−1.07	−2.59	1.59	1.62	0.06	0.67
	N158	−7.46	5.45	−0.28	−2.30	1.41	0.68	0.02	0.96
	Q173	−12.13	12.83	−0.50	0.20	1.49	1.17	0.03	0.71

Bold letters represent residue contacts from experimental reports.

$tRNA^{Tyr}$-MEY interaction can be analyzed and determined using the Mol* tool provided in the PDB (Sehnal et al., 2018). Accordingly, E36, Y151, Q155, N158, and Q173 are the main residue contacts with MEY. **Table 6** shows that the binding free energies of these residues provide negative contributions of −1.90 to −6.17 kcal/mol, except for N158 with 0.31 kcal/mol. **Figure 7D** displays the hydrogen bonding and alkyl · · · π interaction network between Y151, Q155, Q173, and MEY. Even though E36 does not form hydrogen bonding with MEY, a −5.34 kcal/mol strong affinity is derived from vdW and electrostatic attractions of −14.75 kcal/mol. Furthermore, I137 shows a new potential contact with MEY via an alkyl · · · π interaction with a −4.34 kcal/mol binding free energy.

F2YRS and DFY interactions

F2YRS shares approximately identical sequences with $tRNA^{Tyr}$ except for the asparagine and cysteine at positions 108 and 109, respectively, corresponding to F108 and G109 in $tRNA^{Tyr}$. The complex of F2YRS–DFY was obtained after Y32R, L65Y, H70G, F108N, Q109C, D158N, and L162S mutations by an experimental technique (Li et al., 2013). We assumed DFY to be in a neutral state due to the reported pKa value close to 7.0 (Seyedsayamdost et al., 2006). **Figure 7E** shows the six key residues binding to the DFY substrate. R32 and N158 form halogen bonding with the two different fluorine atoms of DFY; meanwhile, hydrogen bonding of R32 and N158 occurs with the OH group of DFY. This is consistent with experimental findings (Li et al., 2013). Experiments have also shown that there are strong dipolar interactions between the fluorine atoms and amide/guanidine groups. Notable polar contributions of 8.48 and 5.45 kcal/mol are estimated by our predictions for R32 and N158, respectively. In addition, Y65 forms π · · · π stacking interactions with the phenyl group of DFY, and Y151, Q155, and Q173 form hydrogen bonds with the amide and carbonyl groups of DFY. Among them, Q155 provides the largest MM contribution of −23.04 kcal/mol with 21.54 kcal/mol of polar energy. Y65 and Q155 with −3.91 and −2.59 kcal/mol free energies, respectively, contribute moderately to the observed binding.

CONCLUSION

This work presents the charge parameters of 18 UAAs related to phenylalanine and tyrosine that are compatible with the use of the Amber ff14SB force field included in the GROMACS package. The newly derived charge parameters initially fitted by the RESP protocol were tested on structural optimizations and relative energies of the 18 UAAs in α-/β-backbone conformations, with an RMS deviation of 0.33 kcal/mol compared with the QM dataset, whereas the M06-2X method produces an RMS deviation of 1.08 kcal/mol. After the parameters were determined, the energy function was further applied to MD simulations of the UAA-mutated proteins and protein–UAA complexes. The motifs containing UAAs and their respective backbone torsions generally overlapped well with the initial coordinates, with an average RMSD of approximately 1.5 Å. The MM/PBSA approach showed that the binding free energy of $tRNA^{Phe}$-DHF is higher than that of PylRS–OMY, which is consistent with experimental data. Comparisons with crystal residue contacts and satisfactory treatments for the interaction modes between proteins and UAAs by substrate binding are presented from the analysis of the per-residue energy decomposition.

Nevertheless, the development of force field is too far from only the development of charge parameters. To increase the transferability and compatibility to the standard Amber force field, the atoms in the common structure of these UAAs should be optimized to be of identical partial charges by applying restraints/constraints in the fitting to RESP in a future study. The bonded parameters, especially the torsional terms related to the gas-phase QM conformational potential energy scan, require further adjustment. The current testing concentrated

on conformational and energetic investigations is also limited, and thus more extensive studies focusing on the dynamic and thermodynamic properties of polypeptides and proteins should be explored.

DATA AVAILABILITY STATEMENT

The original contributions generated for the study are included in the article/**Supplementary Materials**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

WL proposed the core idea of the study. XW designed the project, performed the research, analyzed the data and drafted the manuscript. Both authors critically reviewed the manuscript.

FUNDING

This work was financially supported by the National Natural Science Foundation of China under grant number 31770777, Start-up Foundation for Peacock Talents (827-000365), and the Start-up Grant for Young Scientists (860-000002110384), Shenzhen University.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2020.608931/full#supplementary-material

REFERENCES

Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., et al. (2015). GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1–2, 19–25. doi: 10.1016/j.softx.2015.06.001

Ackaert, C., Kofler, S., Horejs-Hoeck, J., Zulehner, N., Asam, C., von Grafenstein, S., et al. (2014). The impact of nitration on the structure and immunogenicity of the major birch pollen allergen Bet v 1.0101. *PLoS ONE* 9:e104520. doi: 10.1371/journal.pone.0104520

Andersen, H. C. (1983). Rattle: a "velocity" version of the shake algorithm for molecular dynamics calculations. *J. Comput. Phys.* 52, 24–34. doi: 10.1016/0021-9991(83)90014-1

Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., and Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81, 3684–3690. doi: 10.1063/1.448118

Best, R. B., Zhu, X., Shim, J., Lopes, P. E. M., Mittal, J., Feig, M., et al. (2012). Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone φ, ψ and side-chain χ1 and χ2 dihedral angles. *J. Chem. Theory Comput.* 8, 3257–3273. doi: 10.1021/ct300400x

BIOvIA, D. S. (2015). *Discovery Studio Modeling Environment*. San Diego: Dassault Systemes. Release. 4.

Bourquin, F., Riezman, H., Capitani, G., and Grütter, M. G. (2010). Structure and function of sphingosine-1-phosphate lyase, a key enzyme of sphingolipid metabolism. *Structure* 18, 1054–1065. doi: 10.1016/j.str.2010.05.011

Bussi, G., Donadio, D., and Parrinello, M. (2007). Canonical sampling through velocity rescaling. *J. Chem. Phys.* 126:014101. doi: 10.1063/1.2408420

Case, D. A., Belfon, K., Ben-Shalom, I. Y., Brozell, S. R., Cerutti, D. S., Cheatham, T.E., et al. (2020). *AMBER 2020*. San Francisco: University of California.

Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., et al. (1995). A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* 117, 5179–5197. doi: 10.1021/ja00124a002

Cornell, W. D., Cieplak, P., Bayly, C. I., and Kollman, P. A. (1993). Application of RESP charges to calculate conformational energies, hydrogen bond energies, and free energies of solvation. *J. Am. Chem. Soc.* 115, 9620–9631. doi: 10.1021/ja00074a030

Creon, A., Josts, I., Niebling, S., Huse, N., and Tidow, H. (2018). Conformation-specific detection of calmodulin binding using the unnatural amino acid p-azido-phenylalanine (AzF) as an IR-sensor. *Struct. Dyn.* 5:064701. doi: 10.1063/1.5053466

Darden, T., York, D., and Pedersen, L. (1993). Particle mesh Ewald: an N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* 98, 10089–10092. doi: 10.1063/1.464397

Dennington, R., Keith, T. A., and Millam, J. M. (2016). *GaussView 6*. Shawnee Mission, KS Semichem Inc.

Fishman, R., Ankilova, V., Moor, N., and Safro, M. (2001). Structure at 2.6 Å resolution of phenylalanyl-tRNA synthetase complexed with phenylalanyl-adenylate in the presence of manganese. *Acta Crystallogr. D* 57, 1534–1544. doi: 10.1107/S090744490101321X

Fleissner, M. R., Brustad, E. M., Kálai, T., Altenbach, C., Cascio, D., Peters, F. B., et al. (2009). Site-directed spin labeling of a genetically encoded unnatural amino acid. *Proc. Natl. Acad. Sci.* 106:21637. doi: 10.1073/pnas.0912009106

Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., et al. (2016). *Gaussian 16 Rev. C.01*. Wallingford, CT: Gaussian, Inc.

Gao, X.-C., Hao, Q., and Wang, C.-S. (2017). Improved polarizable dipole–dipole interaction model for hydrogen bonding, stacking, T-shaped, and X–H···π interactions. *J. Chem. Theory Comput.* 13, 2730–2741. doi: 10.1021/acs.jctc.6b00936

Genheden, S., and Ryde, U. (2015). The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin. Drug Discov.* 10, 449–461. doi: 10.1517/17460441.2015.1032936

Hao, J.-J., and Wang, C.-S. (2015). Rapid evaluation of the interaction energies for carbohydrate-containing hydrogen-bonded complexes via the polarizable dipole–dipole interaction model combined with NBO or AM1 charge. *RSC Adv.* 5, 6452–6461. doi: 10.1039/C4RA12814A

Harder, E., Damm, W., Maple, J., Wu, C., Reboul, M., Xiang, J. Y., et al. (2016). OPLS3: a force field providing broad coverage of drug-like small molecules and proteins. *J. Chem. Theory Comput.* 12, 281–296. doi: 10.1021/acs.jctc.5b00864

Homeyer, N., and Gohlke, H. (2012). Free energy calculations by the molecular mechanics Poisson–Boltzmann surface area method. *Mol. Inform.* 31, 114–122. doi: 10.1002/minf.201100135

Honig, B., and Nicholls, A. (1995). Classical electrostatics in biology and chemistry. *Science* 268:1144. doi: 10.1126/science.7761829

Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. (2006). Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* 65, 712–725. doi: 10.1002/prot.21123

Johnson, K. A., and Goody, R. S. (2011). The original michaelis constant: translation of the 1913 Michaelis–Menten paper. *Biochemistry* 50, 8264–8269. doi: 10.1021/bi201284u

Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79, 926–935. doi: 10.1063/1.445869

Khoury, G. A., Bhatia, N., and Floudas, C. A. (2014a). Hydration free energies calculated using the AMBER ff03 charge model for natural and unnatural amino acids and multiple water models. *Comput. Chem. Eng.* 71, 745–752. doi: 10.1016/j.compchemeng.2014.07.017

Khoury, G. A., Smadbeck, J., Tamamis, P., Vandris, A. C., Kieslich, C. A., and Floudas, C. A. (2014b). Forcefield_NCAA: Ab initio charge parameters to aid in the discovery and design of therapeutic proteins and peptides with unnatural amino acids and their application to complement inhibitors of the compstatin family. *ACS Synth. Biol.* 3, 855–869. doi: 10.1021/sb400168u

Kumari, R., Kumar, R., and Lynn, A. (2014). g_mmpbsa—A GROMACS tool for high-throughput MM-PBSA calculations. *J. Chem. Inf. Model.* 54, 1951–1962. doi: 10.1021/ci500020m

Leaverfay, A., Tyka, M. D., Lewis, S. M., Lange, O. F., Thompson, J., Jacak, R., et al. (2011). Rosetta3: an object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol.* 487, 545–574. doi: 10.1016/B978-0-12-381270-4.00019-6

Li, F., Shi, P., Li, J., Yang, F., Wang, T., Zhang, W., et al. (2013). A genetically encoded 19F NMR probe for tyrosine phosphorylation. *Angew. Chem. Int. Ed.* 52, 3958–3962. doi: 10.1002/anie.201300463

Li, S.-S., Huang, C.-Y., Hao, J.-J., and Wang, C.-S. (2014). A polarizable dipole–dipole interaction model for evaluation of the interaction energies for N-H···O=C and C-H···O=C hydrogen-bonded complexes. *J. Comput. Chem.* 35, 415–426. doi: 10.1002/jcc.23473

Liu, C. C., and Schultz, P. G. (2010). Adding new chemistries to the genetic code. *Annu. Rev. Biochem.* 79, 413–444. doi: 10.1146/annurev.biochem.052308.105824

Magotti, P., Ricklin, D., Qu, H., Wu, Y.-Q., Kaznessis, Y. N., and Lambris, J. D. (2009). Structure-kinetic relationship analysis of the therapeutic complement inhibitor compstatin. *J. Mol. Recognit.* 22, 495–505. doi: 10.1002/jmr.972

Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., and Simmerling, C. (2015). ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* 11, 3696–3713. doi: 10.1021/acs.jctc.5b00255

Minnihan, E. C., Young, D. D., Schultz, P. G., and Stubbe, J. (2011). Incorporation of fluorotyrosines into ribonucleotide reductase using an evolved, polyspecific aminoacyl-tRNA synthetase. *J. Am. Chem. Soc.* 133, 15942–15945. doi: 10.1021/ja207719f

Miyamoto, S., and Kollman, P. A. (1992). Settle: an analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* 13, 952–962. doi: 10.1002/jcc.540130805

Mondal, S., Chowdhuri, D. S., Ghosh, S., Misra, A., and Dalai, S. (2007). Conformational study on dipeptides containing phenylalanine: a DFT approach. *J. Mol. Struct. THEOCHEM* 810, 81–89. doi: 10.1016/j.theochem.2007.02.006

Moor, N., Klipcan, L., and Safro, Mark, G. (2011). Bacterial and eukaryotic phenylalanyl-tRNA synthetases catalyze misaminoacylation of tRNAPhe with 3,4-dihydroxy-L-phenylalanine. *Chem. Biol.* 18, 1221–1229. doi: 10.1016/j.chembiol.2011.08.008

Nosé, S., and Klein, M. L. (1983). Constant pressure molecular dynamics for molecular systems. *Mol. Phys.* 50, 1055–1076. doi: 10.1080/00268978300102851

Parrinello, M., and Rahman, A. (1981). Polymorphic transitions in single crystals: a new molecular dynamics method. *J. Appl. Phys.* 52, 7182–7190. doi: 10.1063/1.328693

Pauling, L. (1939). *The Nature of the Chemical Bond*. Ithaca, NY: Cornell University Press.

Pearson, A. D., Mills, J. H., Song, Y., Nasertorabi, F., Han, G. W., Baker, D., et al. (2015). Trapping a transition state in a computationally designed protein bottle. *Science* 347:863. doi: 10.1126/science.aaa2424

Petrov, D., Margreitter, C., Grandits, M., Oostenbrink, C., and Zagrovic, B. (2013). A systematic framework for molecular dynamics simulations of protein post-translational modifications. *PLoS Comput. Biol.* 9:e1003154. doi: 10.1371/journal.pcbi.1003154

Qin, M., Li, F., Huang, Y., Ran, W., Han, D., and Song, Y. (2015). Twenty natural amino acids identification by a photochromic sensor chip. *Anal. Chem.* 87, 837–842. doi: 10.1021/ac504121d

Ramachandran, G. N., Ramakrishnan, C., and Sasisekharan, V. (1963). Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.* 7, 95–99. doi: 10.1016/S0022-2836(63)80023-6

Renfrew, P. D., Choi, E. J., Bonneau, R., and Kuhlman, B. (2012). Incorporation of noncanonical amino acids into Rosetta and use in computational protein-peptide interface design. *PLoS ONE* 7:e32637. doi: 10.1371/journal.pone.0032637

Robertson, M. J., Tirado-Rives, J., and Jorgensen, W. L. (2015). Improved peptide and protein torsional energetics with the OPLS-AA force field. *J. Chem. Theory Comput.* 11, 3499–3509. doi: 10.1021/acs.jctc.5b00356

Sakamoto, K., Hayashi, A., Sakamoto, A., Kiga, D., Nakayama, H., Soma, A., et al. (2002). Site-specific incorporation of an unnatural amino acid into proteins in mammalian cells. *Nucleic Acids Res.* 30, 4692–4699. doi: 10.1093/nar/gkf589

Sakamoto, K., Murayama, K., Oki, K., Iraha, F., Kato-Murayama, M., Takahashi, M., et al. (2009). Genetic encoding of 3-iodo-l-tyrosine in *Escherichia coli* for single-wavelength anomalous dispersion phasing in protein crystallography. *Structure* 17, 335–344. doi: 10.1016/j.str.2009.01.008

Santoro, S. W., Wang, L., Herberich, B., King, D. S., and Schultz, P. G. (2002). An efficient system for the evolution of aminoacyl-tRNA synthetase specificity. *Nat. Biotechnol.* 20, 1044–1048. doi: 10.1038/nbt742

Saraogi, I., Vijay, V. G., Das, S., Sekar, K., and Guru Row, T. N. (2003). C–halogen…π interactions in proteins: a database study. *Cryst. Eng.* 6, 69–77. doi: 10.1016/S1463-0184(03)00068-6

Sehnal, D., Rose, A. S., Koča, J., Burley, S. K., and Velankar, S. (2018). "Mol*: towards a common library and tools for web molecular graphics," in *Paper Presented at the Proceedings of the Workshop on Molecular Graphics and Visual Analysis of Molecular Data, Brno, Czech Republic*.

Seyedsayamdost, M. R., Reece, S. Y., Nocera, D. G., and Stubbe, J. (2006). Mono-, di-, tri-, and tetra-substituted fluorotyrosines: new probes for enzymes that use tyrosyl radicals in catalysis. *J. Am. Chem. Soc.* 128, 1569–1579. doi: 10.1021/ja055926r

Si, L., Xu, H., Zhou, X., Zhang, Z., Tian, Z., Wang, Y., et al. (2016). Generation of influenza A viruses as live but replication-incompetent virus vaccines. *Science* 354:1170. doi: 10.1126/science.aah5869

Sitkoff, D., Sharp, K. A., and Honig, B. (1994). Accurate calculation of hydration free energies using macroscopic solvent models. *J. Phys. Chem.* 98, 1978–1988. doi: 10.1021/j100058a043

Sousa da Silva, A. W., and Vranken, W. F. (2012). ACPYPE - AnteChamber PYthon Parser interfacE. *BMC Res. Notes* 5:367. doi: 10.1186/1756-0500-5-367

Spiliotopoulos, D., Spitaleri, A., and Musco, G. (2012). Exploring PHD fingers and H3K4me0 interactions with molecular dynamics simulations and binding free energy calculations: AIRE-PHD1, a comparative study. *PLoS ONE* 7:e46902. doi: 10.1371/journal.pone.0046902

Takimoto, J. K., Dellas, N., Noel, J. P., and Wang, L. (2011). Stereochemical basis for engineered pyrrolysyl-tRNA synthetase and the efficient *in vivo* incorporation of structurally divergent non-native amino acids. *ACS Chem. Biol.* 6, 733–743. doi: 10.1021/cb200057a

Turner, J. M., Graziano, J., Spraggon, G., and Schultz, P. G. (2006). Structural plasticity of an aminoacyl-tRNA synthetase active site. *Proc. Natl. Acad. Sci.* 103:6483. doi: 10.1073/pnas.0601756103

Wang, C., Greene, D. A., Xiao, L., Qi, R., and Luo, R. (2018). Recent developments and applications of the MMPBSA method. *Front. Mol. Biosci.* 4:87. doi: 10.3389/fmolb.2017.00087

Wang, E., Sun, H., Wang, J., Wang, Z., Liu, H., Zhang, J. Z. H., et al. (2019). End-point binding free energy calculation with MM/PBSA and MM/GBSA: strategies and applications in drug design. *Chem. Rev.* 119, 9478–9508. doi: 10.1021/acs.chemrev.9b00055

Wang, J., Cieplak, P., and Kollman, P. A. (2000). How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* 21, 1049–1074. doi: 10.1002/1096-987X(200009)21:12<1049::AID-JCC3>3.0.CO;2-F

Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004). Development and testing of a general amber force field. *J. Comput. Chem.* 25, 1157–1174. doi: 10.1002/jcc.20035

Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., et al. (1984). A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.* 106, 765–784. doi: 10.1021/ja00315a051

Wu, Y., Fried, S. D., and Boxer, S. G. (2015). Dissecting proton delocalization in an enzyme's hydrogen bond network with unnatural amino acids. *Biochemistry* 54, 7110–7119. doi: 10.1021/acs.biochem.5b00958

Xiao, H., Nasertorabi, F., Choi, S.-H., Han, G. W., Reed, S. A., Stevens, R. C., et al. (2015). Exploring the potential impact of an expanded genetic code on protein function. *Proc. Natl. Acad. Sci.* 112:6961. doi: 10.1073/pnas.1507741112

Yang, T., Wu, J. C., Yan, C., Wang, Y., Luo, R., Gonzales, M. B., et al. (2011). Virtual screening using molecular simulations. *Proteins* 79, 1940–1951. doi: 10.1002/prot.23018

Young, D. D., and Schultz, P. G. (2018). Playing with the molecules of life. *ACS Chem. Biol.* 13, 854–870. doi: 10.1021/acschembio.7b00974

Yuet, K. P., Doma, M. K., Ngo, J. T., Sweredoski, M. J., Graham, R. L. J., Moradian, A., et al. (2015). Cell-specific proteomic analysis in *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci.* 112:2705. doi: 10.1073/pnas.1421567112

Yurieva, A. G., Poleshchuk, O. K., and Filimonov, V. D. (2008). Comparative analysis of a full-electron basis set and pseudopotential for the iodine atom in DFT quantum-chemical calculations of iodine-containing compounds. *J. Struct. Chem.* 49, 548–552. doi: 10.1007/s10947-008-0073-9

Zhao, J., Burke, A. J., and Green, A. P. (2020). Enzymes with noncanonical amino acids. *Curr. Opin. Chem. Biol.* 55, 136–144. doi: 10.1016/j.cbpa.2020.01.006

Zou, H., Li, L., Zhang, T., Shi, M., Zhang, N., Huang, J., et al. (2018). Biosynthesis and biotechnological application of non-canonical amino acids: complex and unclear. *Biotechnol. Adv.* 36, 1917–1927. doi: 10.1016/j.biotechadv.2018.07.008

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Molecular Biosciences

METHODS
published: 11 January 2021
doi: 10.3389/fmolb.2020.627087

StructureMan: A Structure Manipulation Tool to Study Large Scale Biomolecular Interactions

Yuejiao Xian[1], Yixin Xie[2], Sebastian Miki Silva[3], Chitra B. Karki[2], Weihong Qiu[4,5] and Lin Li[2,3]**

[1] Department of Chemistry and Biochemistry, University of Texas at El Paso, El Paso, TX, United States, [2] Computational Science Program, University of Texas at El Paso, El Paso, TX, United States, [3] Department of Physics, University of Texas at El Paso, El Paso, TX, United States, [4] Department of Physics, Oregon State University, Corvallis, OR, United States, [5] Department of Biochemistry & Biophysics, Oregon State University, Corvallis, OR, United States

Studying biomolecular interactions is a crucial but challenging task. Due to their large scales, many biomolecular interactions are difficult to be simulated via all atom models. An effective approach to investigate the biomolecular interactions is highly demanded in many areas. Here we introduce a Structure Manipulation (StructureMan) program to operate the structures when studying the large-scale biomolecular interactions. This novel StructureMan tool provides comprehensive operations which can be utilized to study the interactions in various large biological systems. Combining with electrostatic calculation programs such as DelPhi and DelPhiForce, StructureMan was implemented to reveal the detailed electrostatic features in two large biological examples, the viral capsid and molecular motor-microtubule complexes. Applications on these two examples revealed interesting binding mechanisms in the viral capsid and molecular motor. Such applications demonstrated that the StructureMan can be widely used when studying the biomolecular interactions in large scale biological problems. This novel tool provides an alternative approach to efficiently study the biomolecular interactions, especially for large scale biology systems. The StructureMan tool is available at our website: http://compbio.utep.edu/static/downloads/script-for-munipulation2.zip.

Keywords: protein-protein interactions, protein-RNA/DNA interactions, electrostatic force, viral capsid assembly, molecular motor, kinesin, DelPhi, DelPhiForce

OPEN ACCESS

Edited by:
Xiakun Chu,
Stony Brook University, United States

Reviewed by:
Yunjie Zhao,
Central China Normal University, China
Yunhui Peng,
National Institutes of Health (NIH), United States

***Correspondence:**
Weihong Qiu
weihong.qiu@oregonstate.edu
Lin Li
lli5@utep.edu

Specialty section:
This article was submitted to Biological Modeling and Simulation, a section of the journal Frontiers in Molecular Biosciences

Received: *08 November 2020*
Accepted: *10 December 2020*
Published: *11 January 2021*

Citation:
Xian Y, Xie Y, Silva SM, Karki CB, Qiu W and Li L (2021) StructureMan: A Structure Manipulation Tool to Study Large Scale Biomolecular Interactions. Front. Mol. Biosci. 7:627087. doi: 10.3389/fmolb.2020.627087

INTRODUCTION

Studying interactions between biomolecules is an important but challenging task. In recent decades, many efforts and progresses have been made to study the biomolecule interactions (Jones and Thornton, 1996; von Mering et al., 2002; Li et al., 2015; Zhou, 2015). Such studies are in two categories: Predicting biomolecule complex structures (Pagadala et al., 2017); and revealing the biomolecule interaction mechanisms (Jones and Thornton, 1996).

To predict the complex structures of biomolecules such as proteins, RNAs/DNAs, many algorithms have been developed based on some physics principals and statistic functions. Some of them are protein-protein docking algorithms (Gabb et al., 1997; Chen et al., 2003; Dominguez et al., 2003; Li et al., 2011), protein-DNA/RNA docking algorithms (Tuszynska and Bujnicki, 2011; Huang et al., 2013; Yan et al., 2017), scoring functions (Chen and Weng, 2003; Jain, 2006; Huang et al., 2010; Li et al., 2013a), etc. To reveal the mechanisms of biomolecular interactions, numerous methods have been developed to simulate the biomolecular binding processes. The two

most challenging issues in studying the biomolecular interactions are that the size scale of the biomolecules and time scale of the binding processes. Traditional all atom molecular dynamic simulations can hardly simulate the binding processes of large biomolecular systems, such as capsid proteins binding to a viral capsid. In order to accelerate the large-scale biomolecule simulations, many successful coarse-grained models have been developed (Liwo et al., 1997; Marrink et al., 2007). Such coarse-grained models are in several categories: elastic network models, Go-like models, beads-based models (Tozzini, 2005). Besides coarse-grained models, some multiscale methods have also been developed (Wang et al., 2014; Li et al., 2016b). We have developed a DelPhiForce steered Molecular Dynamic (DFMD) method (Li et al., 2017a; Peng et al., 2019) to speed up the Molecular Dynamic (MD) simulation. The advantage of DFMD is utilizing the long-range electrostatic interactions in the MD simulations to accelerate the binding process. This DFMD method has been proven very successful in protein-biomolecule binding processes. Therefore, studying electrostatic interactions is crucial to investigate large scale biomolecular interactions.

To study the interactions between two biomolecules in various perspectives, the ligand structure needs to be manipulated with respect to the receptor, such as shifted, spun, rotated around the receptor. For some large biomolecules such as viruses, many proteins are required to assembled a complete viral capsid. Studying such large complex structures need more comprehensive manipulations on individual biomolecules. Therefore, we developed a Structure Manipulation (StructureMan) program to manipulate the biomolecule structures. Four basic and two advanced structural operations were developed to manipulate the structures of biomolecules. These basic operations are developed for two biomolecules (a receptor and a ligand), which include separation, spin, rotation and perpendicular shifting between a pair of receptor and ligand. Furthermore, two advanced operations were developed to study the assembly of multiple biomolecules in pseudo spherical or pseudo cylindrical symmetry. The pseudo spherical operations, including capsid generation, capsid expansion and capsid detachment tools, can be widely used to study the viral capsid assembly problems. With StructureMan, users can easily manipulate the structures of a complex to study the electrostatic interactions for large systems, such as protein-protein interactions in a whole virus capsid or a large piece of microtubule. In this work, we applied the StructureMan on a viral capsid and a molecular motor, which demonstrated that this novel tool is very useful when studying large scale biomolecular interactions.

Turnip crinkle virus (TCV) is a plant pathogenic virus which is composed of ∼4.0 kb plus-sense RNA and 180 copies of capsid protein subunits (Hogle et al., 1986; Wei et al., 1990). These capsid proteins assemble into an icosahedral capsid with a diameter of ∼330 Å. For the purpose of this work, the quasi-three-fold symmetry related subunits are grouped together and referred as one capsomer (**Figure 1A**). The viral capsid has been shown to have multiple functions in stabilizing the genomic RNA materials during viral assembly and protecting RNA and host-defense machinery (Cao et al., 2010). In 2012, an expanded form of TCV was captured and considered to be a putative RNA uncoating intermediate (Bakker et al., 2012). The expanded capsid is resulted from the separation of the capsid proteins. Having its multiple functions and dynamic nature, the TCV capsid is an interesting target for protein-protein interaction studies, especially in the studies of capsid assembly and viral infection (Sorger et al., 1986; Wei et al., 1990; Saunders and Lomonossoff, 2015). As suggested by transmission electron microscopy, the assembly of the TCV capsid is a progressive process where the capsid protein units continuously assemble onto the initiating structure until the viral capsid is completed (Sorger et al., 1986). This assembly process is guided by interactions among the capsid proteins as well as their RNA genome (Sorger et al., 1986; Wei et al., 1990; Bakker et al., 2012; Saunders and Lomonossoff, 2015). In a recent studies, the wild type capsid proteins of TCV expressed in Cowpea Mosaic Virus-Hyper Translatable Expression system self-assembled into TCV-like particles (Saunders and Lomonossoff, 2015). These results suggested that the ability of these capsid protein assemble into the viral capsid is fundamentally essential in TCV's life cycle. Therefore, this study implemented the StructureMan tool to manipulate the structure of TCV capsid, which then facilitated our analyses that reveal the binding mechanisms among capsomers in the TCV capsid. Many interesting features are discovered and shown in the results and discussion section.

Kinesins are a superfamily of molecular motors. Kinesins have vital cellular functions (Mandelkow and Mandelkow, 2002; Endow et al., 2010; Lee et al., 2015; Tseng et al., 2018) in mitosis and become ideal anti-mitotic drug targets for cancer treatment (DeBonis et al., 2004; Tao et al., 2005; Nakai et al., 2009). Traditional anti-mitotic drugs face two significant problems: (1) serious side effects (Jordan and Wilson, 2004; Schmidt and Bastians, 2007); (2) Strong drug resistance for some types of cancers (Kavallaris, 2010). Recent works found that another promising direction of cancer drug design is targeting kinesins (Jackson et al., 2007; Sarli and Giannis, 2008; Huszar et al., 2009). Interrupting the binding or motility of specific kinesins can block the mitosis and kill the cancer cells. Due to the variety types of kinesins (Vale et al., 1985), kinesin targeting drugs will be more selective and also alternative to solve the drugs resistance compared to microtubule targeting drugs. Therefore, discovering and designing drugs targeting certain types of kinesins become a very promising direction for cancer treatment. Efficient drug design approaches highly demand the systematic understanding of binding and motility mechanisms of kinesins. Therefore, many computational studies have been conducted to study the molecular motors including kinesins (Li et al., 2016a,b; Li et al., 2017; Tajielyato et al., 2018). This work utilized StructureMan to study the binding mechanisms between kinesin and microtubule, which sheds light on the drug design targeting the kinesins.

METHODS

Four basic and two advanced structural operations were developed to manipulate the structures of biomolecules. The basic operations are developed for two biomolecules (a receptor

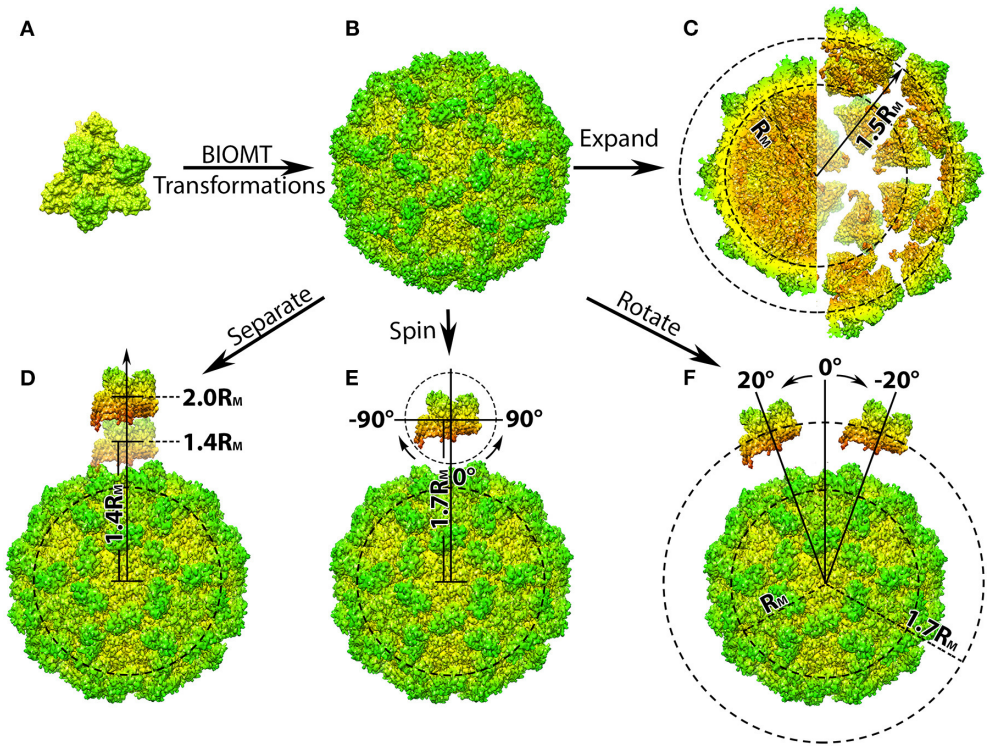


FIGURE 1 | Structure manipulation of the capsid of Turnip Crinkle Virus (TCV). **(A)** Individual capsomer of TCV with the three type of protein subunits labeled on the side (PDB ID 3ZX8); **(B)** the whole TCV capsid generated using the capsid generation tool; Penal **(C–F)** are demonstration of structure manipulation where: **(C)** the native TCV capsid is expended by $0.7R_M$. The central section of the native capsid (left) and the expended capsid (right) are shown as their radius labeled in black, respectively; **(D)** one capsomer is detached from the rest of the capsid by 0.4 to $1.0R_M$, as its distance to the mass center of the whole capsid increases to $1.4R_M$ and $2.0R_M$; **(E)** one capsomer is first detached by $0.7R_M$ and then spun from −90Å to 90Å in the xy-plane (around z-axis); **(F)** after detached by $0.7R_M$, the capsomer is rotated around the capsid from −20Å to 20Å with respect to the mass center of the whole capsid. In all panels, the capsomers and capsids are shown in their density map generated using Chimera (Pettersen et al., 2004; Goddard et al., 2007) and colored by radius from red to green.

and a ligand), which include separation, spin, rotation and perpendicular shifting between a pair of receptor and ligand. Furthermore, two advanced operations were developed to study the assembly of multiple biomolecules in pseudo spherical or pseudo cylindrical symmetry. These pseudo spherical operations, including capsid generation, capsid expansion, and capsid detachment tools, can be widely used to study the viral capsid assembly problems. For the purpose of demonstration of the advanced operations, the protein capsid of the TCV, and the kinesin-microtubule complex were chosen in this work because of their representative pseudo spherical (icosahedral) and pseudo cylindrical symmetry.

Basic Manipulations

Separation

With two separated coordinates files of the protein units as inputs, this tool would displace one of the two units in a user given distance away from the other. For clarification, one of the protein units would be fixed in its original coordinates and is referred as the fixed unit. The other protein unit would be manipulated to result in different positions and orientations and is hence referred as the manipulated unit.

The tool first calculate the mass center of both proteins, C_{fixed}, and $C_{manipulated}$, by averaging the coordinates of each individual atom after weighted by their corresponding atomic mass (**Supplementary Equations 1, 2**). With the obtained mass centers, a vector $\overrightarrow{M}$, form C_{fixed} to $C_{manipulated}$ is calculated (**Supplementary Equation 3**). This vector $\overrightarrow{M}$ can then be normalized with its magnitude to obtain the vector $\overrightarrow{U}$ that defines the direction of the separation (**Supplementary Equations 4, 5**). With the vector $\overrightarrow{U}$ and the user-defined separation distance, d, a separation vector, $\overrightarrow{S}$, would then be generated (**Supplementary Equation 6**), which is then applied to the coordinates of the manipulated unit and create a new structure that is separated from the fixed unit by the user-defined distance d (**Supplementary Equation 7**).

Rotation

The rotation tool would rotate the manipulated unit around the fixed one by a user-given angle. This rotation operation can be carried out in xy plane (around z-axis), xz-plane or yz-plane as users prefer. Rotation in xy-plane is discussed here for a simplified demonstration.

The tool will start by calculating the mass center C_{fixed} using the method demonstrated in the section above. A vector, $\overrightarrow{M_A}$, from C_{fixed} to a randomly chosen atom A in the manipulated unit can be created (**Supplementary Equation 8**). The rotation vector, $\overrightarrow{R_A}$, would then be generated via multiplying vector $\overrightarrow{M_A}$ vector by a rotation matrix that included the user-defined angle for the desired rotation (**Supplementary Equation 9**). This obtained $\overrightarrow{R_A}$ is then applied on the x, y, z coordinates of the manipulated unit, generating a modified structure with the user-defined degree of rotation (**Supplementary Equations 10, 11**).

It is important to notice that the rotation of manipulated unit around the fixed one may introduce clashes if the atoms are closed to each other. Therefore, it is recommended to separate the manipulated unit from the fixed one to a proper distance upon using the rotation tool.

Spin

This tool allows the spinning of the manipulated unit with the respect to its own mass center. Similar to the rotation tool, the spinning can be performed in any of the *xy-*, *yx-*, or *xz-*plane. Spinning the protein unit in xy-plane is discussed here as a demonstration.

With the coordinate file of the unit to be manipulated as the input, this tool first calculates its mass center, $C_{manipulated,}$ using method discussed in section Separation. A vector, $\overrightarrow{M_A}$, from $C_{manipulated}$ to a randomly chosen atom would then be generated and multiplied by a rotation matrix to generate the final spinning vector, $\overrightarrow{S_A}$, using the method demonstrated in section Rotation. The final coordinates of the atom will be calculated using the spinning vector. As the operation being carried out in the xy-plane, the z coordinate of each atom remind the same as original. This process would be repeated on each individual atom within the protein unit and output their spun coordinates into a separate file. To avoid clashes, it is recommended to separate the manipulated unit from the fixed one to a proper distance upon using the spinning tool.

Perpendicular Translation

The perpendicular translation tool shifts the manipulated unit along the line that is perpendicular to the vector of mass centers in the selected plane. The translation in xy-plane is shown as an example.

This tool calculates the mass centers of both protein units and the vector of mass centers, $\overrightarrow{M}$, the normalized vector $\overrightarrow{U}$, as well as the separation vector, $\overrightarrow{S}$, in a similar manner to that in section Separation (**Supplementary Equations 12–15**). The separation vector $\overrightarrow{S}$ would then be rotated 90Å or −90Å to generate the final translation vector $\overrightarrow{T}$ that contains information of the user-defined distance (**Supplementary Equation 16**). Finally, this tool modifies the coordinates of the manipulated unit using translation vector $\overrightarrow{T}$ to create a new structure which is translated along the line perpendicular to the mass center vector by a given distance d (**Supplementary Equation 17**). Similar to the rotation and spinning tool, it is also recommended to separate the manipulated unit from the fixed one to a proper distance upon using this tool in order to avoid any clashes.

Capsid Structure Manipulation

Capsid Generation Tool

Many pdb files of multi-protein complexes deposited in Protein Data Bank (PDB) do not actually contain the coordinates of all the protein units within the complexes, making it inconvenient for researchers who study protein-protein interactions among multiple protein units. However, instructions on how to construct the missing units from the given units are given as BIOMT matrices (**Table 1**). Within the BIOMT matrices, the numbers of biomolecule to be constructed (**Table 1**, Column3), as well as the corresponding transformation matrices are provided (**Table 1**, Column4–7). Therefore, in order to generate the structure of all the individual biomolecule unit within the pseudo spherical (icosahedral) viral capsid, an input coordinate file containing BIOMT matrices information is required.

As shown in **Table 1**, each transformation matric contains BIOMT1, BIOMT2, BIOMT3, which would be apply on x, y, and z coordinates, respectively, using the following equations:

$$\begin{cases} x_f = ax_0 + by_0 + cz_0 + d \\ y_f = ex_0 + fy_0 + gz_0 + h \\ z_f = ix_0 + jy_0 + kz_0 + l \end{cases} \quad (1)$$

Where the coefficients a to l are provided by the BIOMT matrices (**Table 1**), and x_0, y_0, and z_0 represent the original coordinates of individual atom in the given molecule. The calculation would be performed on all other atoms until the structure of protein unit is completed and output as a separated file. This process will then repeat with the next BIOMT matrix until all the required protein units are generated (**Figures 1A,B**).

This tool can be applied in generating structures of individual protein units form any multi-protein complex as long as the BIOMT matrices are provided. As a demonstration, the initial structure of the TCV capsid was downloaded from PDB (ID 3ZX8), by which the structure of a capsomer and the BIOMT matrices were provided. Using the capsid generation tool, 60 copies of capsomer structures were generated and assembled into the native structure of TCV capsid (**Figure 1B**). These capsomer structures can then be collected for further studies, where the interactions among capsomers are investigated.

Capsid Expansion Tool

The next tool allows the shifting of all capsomers away from the mass center of the whole capsid resulting in a viral capsid expended by a user desired distance (**Figure 1C**).

The first step in this operation is to find the capsid's mass center, C_{capsid}, which would be done in a similar manner to the capsid generation tool. When determining the mass center of TCV capsid, this tool first calculates the mass center of the primary capsomer, and then transformed the obtained coordinates according to the given BIOMT matrices. Sixty copies of coordinates would be generated and presenting the mass center of corresponding capsomers in the TCV capsid. With these coordinates, the coordinates of the mass center of the

TABLE 1 | Demonstration of the BIOMT matrices provided in PDB files.

Column1	Column 2	Column 3	Column 4	Column 5	Column 6	Column 7
REMARK 350	BIOMT1	N	*a*	*b*	*c*	*d*
REMARK 350	BIOMT2	N	*e*	*f*	*g*	*h*
REMARK 350	BIOMT3	N	*i*	*j*	*k*	*l*

whole TCV capsid can then calculated by averaging the mass center coordinates of its individual capsomer, as shown in the following equation:

$$\begin{cases} C_{capsid}(x) = \frac{\sum_{n=1}^{N} x_n}{N} \\ C_{capsid}(y) = \frac{\sum_{n=1}^{N} y_n}{N} \\ C_{capsid}(y) = \frac{\sum_{n=1}^{N} z_n}{N} \end{cases} \quad (2)$$

Where N is the total number of capsomers, x_n, y_n, and z_n are the coordinates of the mass center of the capsomer n.

Next, the tool generates the 60 copies of vectors, $\overrightarrow{S}_n$, from the mass center of the whole capsid to the mass center of individual capsomer by the following equation; The vectors, $\overrightarrow{S}_n$, would define the direction of the shifting the individual capsomers.

$$\begin{cases} \overrightarrow{s}_{n,x} = C_{capsid}(x) - C_N(x) \\ \overrightarrow{s}_{n,y} = C_{capsid}(y) - C_N(y) \\ \overrightarrow{s}_{n,z} = C_{capsid}(z) - C_N(z) \end{cases} \quad (3)$$

where $C_{capsid}(x)$, $C_{capsid}(y)$, and $C_{capsid}(z)$ are given in Equation (2), $C_N(x)$, $C_N(y)$, and $C_N(z)$ are the coordinates of mass center of the capsomer n.

The shifting (expansion) distance can again be defined by user. To make it more convenient of users who don't have direct measurement of the desired distance, we introduce the concept of a mean radius of mass distribution, the mean mass radius R_M, which is defined using the following equation:

$$\begin{cases} r_i = \sqrt{[C_{capsid}(x) - x_i]^2 + [C_{capsid}(y) - y_i]^2 + [C_{capsid}(z) - z_i]^2} \\ R_M = \frac{\sum_{i=1}^{l} m_i r_i}{|M_T|} \end{cases} \quad (4)$$

Where x_i, y_i, and z_i are the coordinates of atom i, r_i is the distance between a single atom i and the mass center of the capsid C_{capsid}, m_i is the atomic mass of the corresponding atom, and M_T is the total atomic mass of all atoms within the capsid.

Thanks to the icosahedral symmetry of the vial capsid, the distances from C_{capsid} to the mass center of individual capsomer are equal. The calculation of the R_M can be simplified to one step using following equation:

$$R_M = |\overrightarrow{S}_n| = \sqrt{\left[C_{capsid}(x) - C_N(x)\right]^2 + \left[C_{capsid}(y) - C_N(y)\right]^2 + \left[C_{capsid}(z) - C_N(z)\right]^2} \quad (5)$$

where only the coordinates of the capsid mass center $C_{capsid}(x, y, z)$, and that of the one capsomer n is needed.

With the given distance, d, a expansion vector, $\overrightarrow{E}$, would then be generated using by the following expressions:

$$\begin{cases} \overrightarrow{E}_x = d \cdot R_M \cdot \overrightarrow{S}_{n,x} \\ \overrightarrow{E}_y = d \cdot R_M \cdot \overrightarrow{S}_{n,y} \\ \overrightarrow{E}_z = d \cdot R_M \cdot \overrightarrow{S}_{n,z} \end{cases} \quad (6)$$

Finally, the tool generates the expanded structures based on the given primary capsomer, BIOMT matrices and the calculated expansion vector $\overrightarrow{E}$, using the following equation:

$$\begin{cases} x_f = ax_0 + by_0 + cz_0 + d + \overrightarrow{E}_x \\ y_f = ex_0 + fy_0 + gz_0 + h + \overrightarrow{E}_y \\ z_f = ix_0 + jy_0 + kz_0 + l + \overrightarrow{E}_z \end{cases} \quad (7)$$

Where the coefficients a to l are provided by the BIOMT matrices (**Table 1**), and x_0, y_0, and z_0 are the coordinates of individual atom in the primary capsomer. The calculation would be repeat on all atoms within the capsomer and output the expanded coordinates into a separated file. This process then goes on with the next BIOMT matrix until the expanded capsid is generated (**Figures 1B,C**).

As a demonstration, a TCV capsid expended by $0.5R_M$ was generated by shifting the individual capsomers $0.5R_M$ away from the mass center of TCV capsid (**Figure 1C**). With the structures of expended capsid and its individual capsomer, investigation that aims to determine the driving force of intact viral capsid assembly, can be carried out as discussed in the later section.

Capsomer Detachment Tool

This tool detaches a single capsomer from the viral capsid by a user defined distance. Compare to previous tools, this one will output a structural file of the shifted capsomer and the rest of the capsid in two separated files. The work flow of this tool is very similar to that of the expansion tool, except the expansion vector would only be applied on the chosen capsomer. The detachment distance can be user defined relatively to the mean mass radius R_M. As a demonstration, one of the capsomer from TCV capsid was separated from the rest of the capsid from $0.4R_M$ to $1.0R_M$ in $0.1R_M$ intervals (**Figure 1D**).

Once the structures of the detached capsomer and the rest of the capsid are obtained, operations including spinning and

rotation, can then be carried out for the purpose of investigating the interactions between the one capsomer and rest of the capsid during viral capsid assembly. After detached from the TCV capsid by $0.7R_M$, the capsomer was spun in the xy-plane (around z-axis) from −90Å to 90Å in 2Å interval (**Figure 1E**) using the spinning tool described in section Spin. The capsomer were also rotated around the rest of the capsid from −20Å to 20Å in 2Å interval (**Figure 1F**) using the rotation tool described in section Rotation.

By detaching, spinning and rotating a single capsomer, different orientations and distances of capsomer with respect to the rest of the capsid were obtained, which can be used for the capsid assembly studies where the driving force can be investigated in a manner of mimicking the dynamic assembly process.

Cylindrical Structure Manipulation

The application of the manipulating tool can be extended in complexes where cylindrical structures are involved. One example is to mimic the kinesin motor's movement on the microtubule filament, which can be subsequently used in the investigation of kinesin-microtubule interactions during the cargo transportation. In this work, the complex structure of kinesin binding with microtubule is generated and described in our previous paper (Li et al., 2016b). However, the StructureMan tool can be used to any other microtubule and cylindrical biomolecules. With the tools described above, 4 different operations of the kinesin motor domain were performed: (1) shifted away from the microtubule by 5Å to 50Å in 2Å interval using the separating tool (**Figure 2A**); (2) rotated around one chosen microtubule unit from −20Å to 50Å in 2Å interval using the rotation tool (**Figure 2B**). This rotation range was limited by the steric effects of the neighboring microtubule units; (3) spun from −180Å to 180Å in 2Å interval around the Z axis using the spin tool (**Figure 2C**). (4) translated along the microtubule from −80Å to 80Å in 2Å interval using the perpendicular translation tool (**Figure 2D**); In each operation, the structures of kinesin motor domain with various distances and orientations were output separately and collected for DelPhi (Li et al., 2012a,b, 2013b) and DelPhiForce (Li et al., 2017a,b,c) calculation.

Electrostatic Potential Calculations by DelPhi

Electrostatic calculations were performed on the complex of detached capsomer and incomplete TCV capsid as well as the expended capsid collected from previous sections using method described in our previous paper (Xian et al., 2019). The electrostatic potentials as well as the interactions among the capsomers are visualized in Visual Molecular Dynamics (VMD) (Humphrey et al., 1996; **Figure 3**). The surfaces of the capsid and capsomers are generated using the "Quicksurf" method in VMD and colored from red to blue in a scale range of −3.0 to 3.0 kT/Å. More information on DelPhi analysis can be accessed through this tutorial: http://compbio.clemson.edu/delphi.

Electrostatic Binding Forces Calculation by DelPhiForce

To examine the roles of electrostatic interactions in the process of viral capsid assembly, 115 structures of viral capsomers in various orientations and distances were collected from previous sections and prepared for DelPhiForce calculations using the method mentioned in our previous paper (Xian et al., 2019). The calculated electrostatic forces between the manipulated capsomer and the rest of the capsid were visualized and in VMD (Humphrey et al., 1996), where they are represented by arrows. The rest of the capsids are shown using the "Quicksurf" method and colored from red to blue in a scale range of −3.0 to 3.0 kT/Å (**Figure 4**).

In order to underline the significances of electrostatic interactions in driving kinesin's movement along the microtubule, 342 structures of microtubule and kinesin motor domain in different orientations and distances were collected for DelPhiForce calculations. The parameters for these calculations were set as the same as those of TCV capsid. The visualization of the electrostatic forces was also done in VMD using the same method, except the surface of microtubule were obtained by the "Surf" option in VMD (**Figure 5**). More information on DelPhiForce analysis can be accessed through this tutorial: http://compbio.clemson.edu/delphi-force-web.

RESULTS AND DISCUSSION

Electrostatic Potential of the TCV Capsomers and Its Capsid

The charge distribution on capsomers mainly has two functions: First, the electrostatic interactions among capsomers play significant roles in assembling and stabilizing the whole capsid structure (Li et al., 2012a; Salas et al., 2019; Xian et al., 2019). Second, electrostatic interactions between capsid and DNA/RNA stabilize the encapsidated genomic materials by neutralizing the repulsive forces between the DNA/RNAs (Bakker et al., 2012). In this work, we mainly focused our investigations on electrostatic interactions among capsomers of TCV using the StructureMan.

The electrostatic potential calculations from DelPhi demonstrated the charge distribution on the inner and outer surface of the viral capsid. The inner surface of the viral capsid is dominated by positive charges as shown in **Figures 3c,d**, which explains why the capsomers play crucial roles in stabilizing the packed genomic RNA in previous studies.

On the outer surface of the viral capsid, negatively charged residues are distributed rather evenly through the whole capsid, while the positively charged residues are mostly located at 5-fold axis and 3-fold axes (**Figure 3a**). After one chosen capsomer is detached from the rest of the capsid by $0.5R_M$ (50 percent of the particle mean mass radius), strong attractive electric field lines are present between the detached capsomer and the rest of the capsid (**Figure 3b**), suggesting that the electrostatic interaction guides capsomers to build the viral capsid.

To study the overall electric field lines among the capsomers for a whole capsid, the StructureMan was utilized to expand the capsid. This "expand" operation shifted each of the capsomers

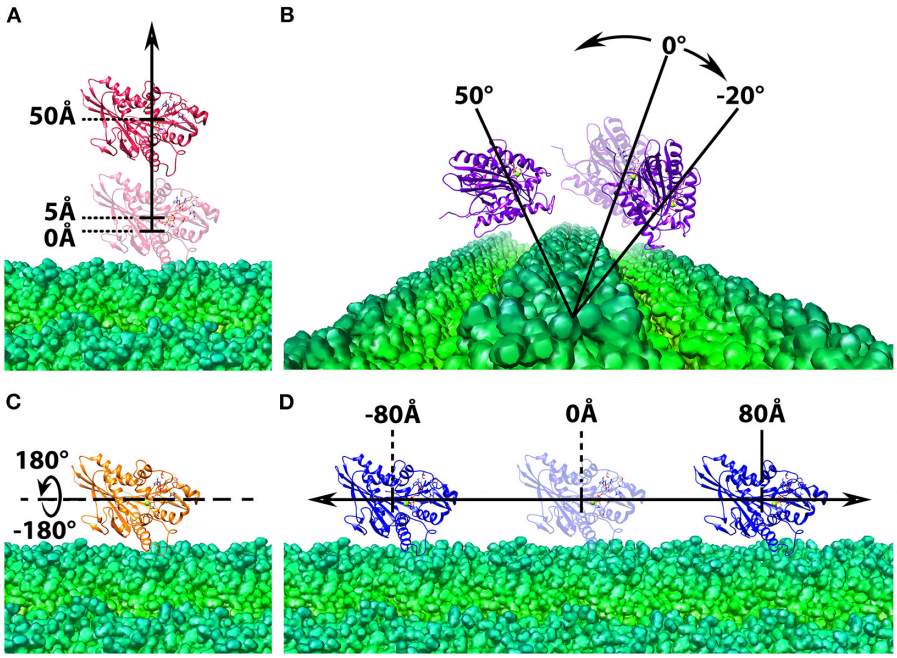


FIGURE 2 | Structure manipulation of the kinesin with respect to microtubule. In each of the four operations, the microtubule is shown at the bottom in green and shown in density map generated by chimera (Pettersen et al., 2004; Goddard et al., 2007), whereas the motor domain of kinesin is shown on the top in ribbon with various color, red, purple, orange, and blue. The motor domain is manipulated by **(A)** shifted from the microtubule from 5Å to 50Å in 2Å interval; **(B)** rotated around one chosen microtubule unit form −20Å to 50Å in 2Å interval; **(C)** spun from −180Å to 180Å in 2Å interval around the z-axis; **(D)** translated along the microtubule from −80Å to 80Å in 2Å interval.

away from the rest of the capsid (**Figures 3c,d**). Electrostatic calculation of the expended capsid shown the electric field lines located at the interfaces of capsomers (**Figures 3c,d**). High density of electric field lines were found among individual capsomers, indicating the present of strong attractive interactions (**Figure 3d**). These attractive interactions were found throughout the viral capsid, which demonstrated the role of electrostatic interactions in stabilizing the whole viral capsid.

Electrostatic Binding Force Between One Detached Capsomer and the Rest of the Capsid

To further characterize the role of electrostatic interaction in the vial capsid assemble process, the electrostatic binding forces were calculated using DelPhiForce (Li et al., 2017a,b). With the structure generated from the StructureMan, the electrostatic binding forces were studied in various orientation and distances of the capsomers (**Figure 4**).

While one chosen capsomer was detached from the rest of the capsid, DelPhiForce was utilized to calculate the electrostatic binding forces, which were represented by orange arrows in **Figure 4A**. Note that all the force arrows are normalized to the same size in order to demonstrate their directions clearly. To compare the strengths of these binding forces, the magnitudes of these binding forces were plotted against the distance between the detached capsomer and its native position (**Figure 5A**), where the more negative value represents the stronger attractive binding force. Binding forces data at $0.3R_M$ or less were not considered because of the possibility of clashes between the atoms. The binding force became neglectable after the capsomer was detached by $0.8R_M$ or further (**Figure 5A**). The forces within the range of $0.4R_M$ to $0.7R_M$ are all attractive as the arrows point toward the rest of the capsid (**Figure 4A**). This indicated that the effective range of electrostatic forces between the capsomer and the rest of the capsid is about $0.7R_M$(∼97Å), which suggests that the electrostatic binding forces guide capsomers from long distance during the viral capsid assembling process.

The binding forces between the detached capsomer and the rest of the capsid were also calculated while the detached capsomer was spun (visualized in **Figure 4B**). If the capsomer was rotated around z-axis within the range from −45Å to 90Å, the binding forces remained attractive. When the capsomer was rotated from −90Å to −45Å, the binding force became repulsive due to the effect introduced by putting the wrong-orientated bulky S domain too close to the rest of the capsid, which resulted in strong electrostatic repulsive forces. This suggests that the electrostatic forces contribute in adjusting the orientations of the capsomers to the native orientations, which were more electrostatically favorable compared to the non-native orientations.

Previous studies on viral capsids have demonstrated the interactions between an individual capsomer and its adjacent capsomers are crucial in the capsid assembly process (Salas et al.,

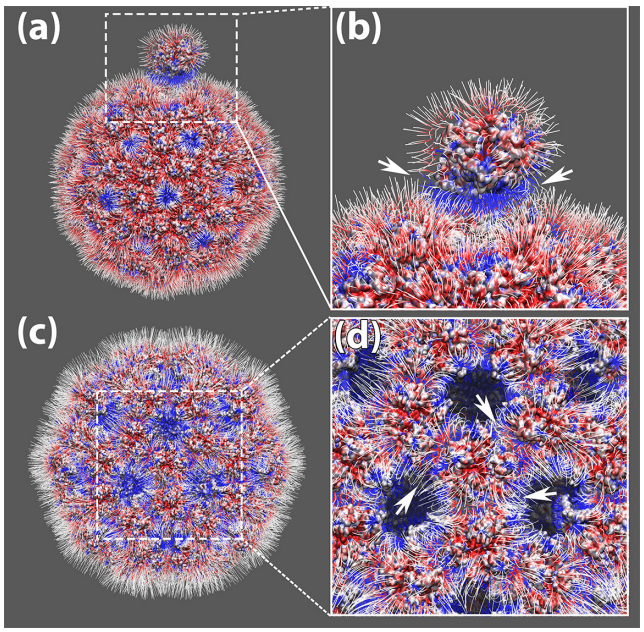


FIGURE 3 | The electrostatic potential field lines among capsomers of TCV capsid. **(a)** The electrostatic potential field lines of a capsomer detached away by 0.5R_M and the rest of the capsid. **(b)** Zoom-in area of the electrostatic potential field lines between the detached capsomer and the rest of the capsid, where their attractive interaction ware pointed out by white arrows; **(c)** The electrostatic potential field lines of the expended capsid obtained by shifting individual capsomers 0.5RM away from the mass center of the whole capsid. **(d)** Zoom-in area of the electrostatic potential field lines among the expanded capsid. The attractive interaction around one capsomer are pointed out by white arrows; All four panels were rendered by VMD (Humphrey et al., 1996). All capsomer surfaces are generated using the "Quicksurf" method. Negatively and positively charged capsomer surface areas are colored from red to blue with a scale of −3.0 to 3.0 kT/Å. The electric field lines were also colored using the same color scheme.

2019; Xian et al., 2019). Here we focused on the interaction between one capsomer and the rest of the capsid. As the detached capsomer was rotated around the rest of the capsid using the StructureMan, the binding forces were analyzed (**Figure 4C**). The rotation was carried out within a range that the detached capsomer was still relatively close to the cavity created by capsomer detachment. As the arrows are all orientated toward the capsid, we conclude that the electrostatic interaction is again attractive between the detached capsomer and the rest of the capsid.

Electrostatic Binding Force Between Kinesin Motor Domain and Microtubule

Similar analyses of the electrostatic binding forces were performed with the kinesin-microtubule complex in which the kinesin motor domain was manipulated by various orientations and distances (**Figures 5B**, **6**). As the kinesin motor domain was separated from the microtubule, the strengths of the attractive binding forces reduced and became insignificant when the distance reached 25Å (**Figures 5B**, **6A**). When the kinesin was separated from the microtubule less than 15Å, the electrostatic binding forces were exerted toward the native binding site on the microtubule. When the separation was in the range of 15Å−25Å, the binding forces were orientated to the neighboring binding site. This suggests that as the distance between the kinesin motor domain and the microtubule increases, the binding force toward the neighboring binding site becomes competitive to that toward the native binding site.

While the kinesin was rotated around one microtubule within the range of −16Å to 10Å, the binding force was exerted toward the native binding site (**Figure 6B**). If the kinesin was rotated further than −16Å, the force became repulsive due to the steric effect from the neighboring microtubule units. When the kinesin was rotated to the top of the microtubule units, the binding force first became repulsive and then changed to attractive as it traveled close enough to the neighboring native binding site (**Figure 6B**). When the kinesin motor domain was spun from −10Å to 10Å (the motor is still in near-native orientation), the binding forces remain attractive. However, when the orientation of kinesin motor domain is significantly changed, the binding forces became randomly directed (**Figure 6C**). The results from both rotation and spin operations reveal that the electrostatic interaction favors the native or near-native orientations of kinesin motor domain, and when its orientation is altered, the binding force reorient it to the native orientation by providing a repulsive binding force.

In the operation where the kinesin motor domain was translated along the microtubule, the binding forces were consistently exerted toward the native binding side within the range of −26Å−12Å (**Figure 6D**). While traveling between the native binding site to the neighboring binding site, the binding forces were shown to be mostly repulsive. However, as the kinesin motor domain traveled closer to the neighboring binding site (from −70Å to −80Å, and 60Å to 80Å), the binding forces were again orientated toward the neighboring binding sites. These results demonstrate that the electrostatic interactions make a significant contribution in guiding the kinesin by favoring the binding on the native binding sites and rejecting the non-native binding positions. Our discovery on the electrostatic interaction between kinesin and microtubule can explain and support the thermal ratchets model for kinesin's motility (Magnasco, 1993; Hwang and Karplus, 2019). The electrostatic analyses in this study also match the previous results from Brownian simulations for kinesins (Grant et al., 2011).

CONCLUSION

Understanding biomolecular interactions is crucial and fundamental to study the biology problems. Due to their large scales, many biomolecular interactions are difficult to be studied via all atom simulations. Here we introduce a Structure Manipulation tool (StructureMan) to offer comprehensive operations for the structures in large scale biomolecular interactions, such as interactions in the viral capsid and molecular motor-microtubule complexes. This StructureMan tool contains operations which can be utilized to study the

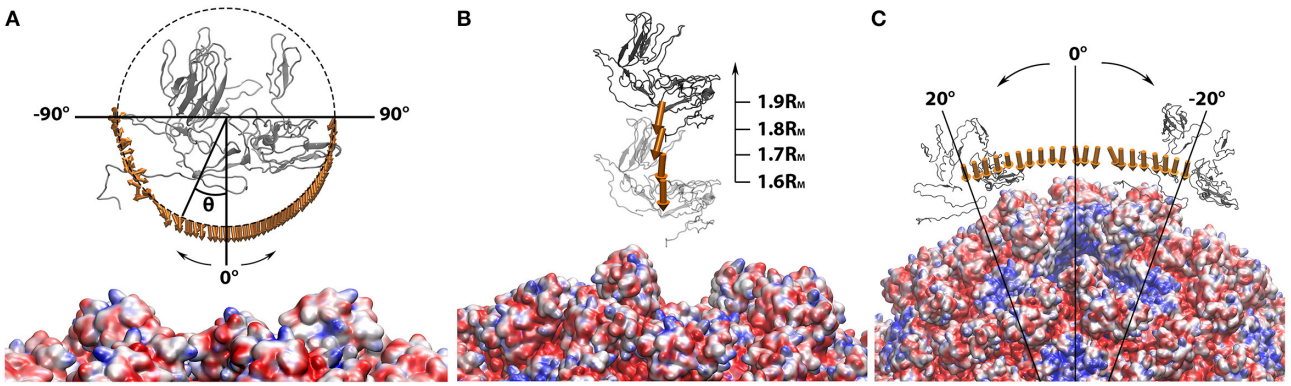


FIGURE 4 | Electrostatic forces between the manipulated capsomer and the rest of the capsid while the capsomer was **(A)** separated from the rest of the capsid by $0.4R_M$ to $0.7R_M$ in $0.1R_M$ interval; **(B)** spun around the z-axis from −90Å to 90Å in 2Å interval; **(C)** rotated around the rest of the capsid from −20Å to 20Å in 2Å interval. In all three panels, the manipulated capsomer was shown in gray ribbon. The rest of the capsid shown in "Quicksurf" colored from red to blue in a scale of −3.0 to 3.0 kT/Å. The electrostatic forces are represented by arrows. In order to clearly show the directions of all binding forces, the arrows in each panel ware normalized to the same size. The tails of arrows in **(B,C)** ware placed at the mass centers of the manipulated capsomer. In **(A)**, the arrow tails are place on a circle where the spinning degrees can be differentiated by the angle theta (θ). All images are rendered by VMD.

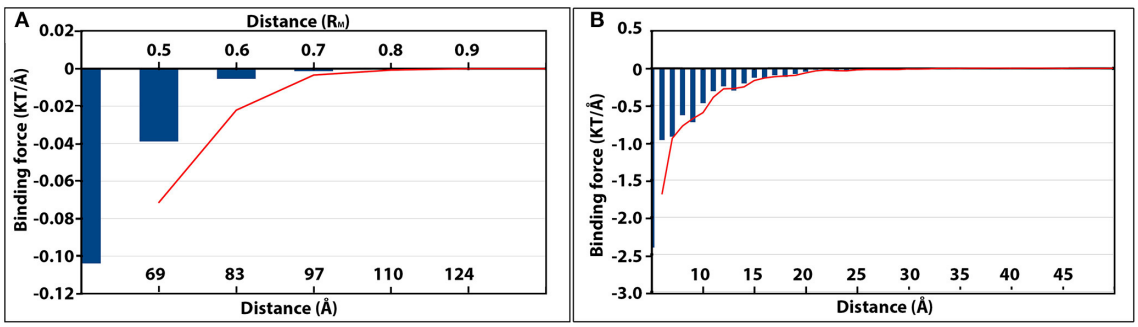


FIGURE 5 | The magnitudes of electrostatic binding forces between the detached capsomer and incomplete TCV capsid **(A)**, and the kinesin motor domain and microtubule **(B)**. In both panels, the strength of the binding forces is represented by the blue vertical histogram bars. The more negative value of binding force (KT/Å), the stronger attractive electrostatic binding force. The changes of binding force as the separation distance increased were shown by the red moving average trendlines in both panels.

interactions in large biological systems. Combining with electrostatic calculation tools such as DelPhi and DelPhiForce, StructureMan can be used to reveal the detailed biomolecular interactions. Two examples are demonstrated in the results and discussion section, which show that the StructureMan is beneficial when studying the biomolecular interactions in large scale biomolecular complexes.

There are four basic and two advanced operations developed in the StructureMan. Note that these basic operations in StructureMan are different from those in existing visualization programs. Those tools in the existing visualization programs can shift or rotate a single biomolecule; However, these four basic operations of StructureMan take two biomolecules as inputs and do the operations between the two biomolecules. For examples, the separation tool in StructureMan shifts the ligand from the receptor in the direction of their mass center connection line, while the existing tools can only shift a single protein. If users want to shift a ligand from the receptor in the direction of their mass center connections, users need to calculate the shifting vector first and normalize the vector, then use the existing tools to shift the ligand. The rotation operation in StructureMan rotates the ligand around the mass center of the receptor. This operation cannot be easily achieved by the existing visualization programs. Instead, users need to write some script to complete such an operation. Advanced operations in StructureMan are even more comprehensive. For example, the capsid expansion operation takes one capsomere as input structure and generates a structure of expanded capsid in which every two adjacent capsomers are separated by a distance defined by the users. Such an operation is not in any of the existing tools. Another advantage is that the StrucrueMan is written in shell script, which can be easily used to handle large number of structures (such as a big number of frames from MD simulations).

In this work, we first focused on investigations on electrostatic interactions among capsomers of TCV using the StructureMan. The charge distribution on the inner surface of the viral capsid

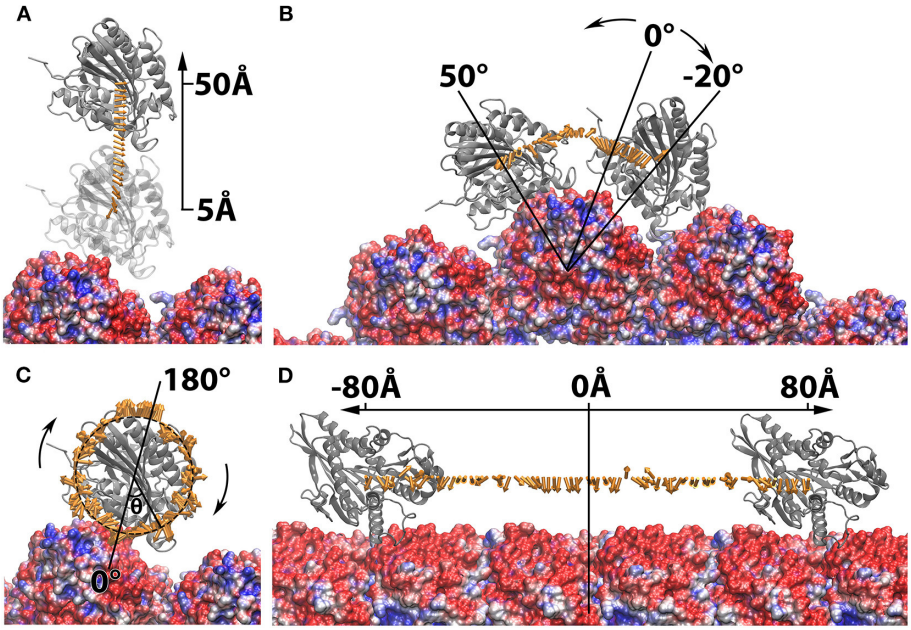


FIGURE 6 | Electrostatic forces between the manipulated kinesin motor domain and microtubule while the kinesin motor domain is **(A)** separated from the microtubule from 5Å to 50Å in 2Å interval; **(B)** rotated around one chosen microtubule unit form −20Å to 50Å in 2Å interval; **(C)** spun from −180Å to 180Å in 2Å intervals around the z-axis. **(D)** Translated along the microtubule from −80Å to 80Å in 2Å interval. The kinesin motor domains are shown in ribbon. The microtubule are shown in surface colored from red to blue in a scale of −3.0 to 3.0 kT/Å. The electrostatic forces are represented by arrows. In **(C)**, the arrow tails are place on a circle where the spinning degrees can be differentiated by the angle theta (θ), while the tails of arrows in all other panels are placed at the mass centers of the manipulated capsomers. The force arrows in each panel are normalized to the same size. All images are rendered by VMD.

indicates that the capsomers stabilize the packed genomic RNA, as observed in previous studies. On the outer surface of the viral capsid, strong attractive electric field lines imply that the electrostatic interactions guide capsomers to build the viral capsid. The "expand" operation shifted each of the capsomers away from the rest of the capsid, which reveals that the attractive interaction among the capsomers is a key factor to stabilize the whole viral capsid.

The StructureMan and DelPhiForce were utilized to further characterize the electrostatic binding forces in the viral capsid. Results demonstrate that the effective range of electrostatic forces between the capsomer and the rest of the capsid is about $0.7R_M$ (∼97Å), which suggests that the electrostatic binding forces guide capsomers from long distances in the viral capsid assembling process. The spin and rotation operations in the StructureMan show that the electrostatic forces contribute in adjusting the orientations of the capsomers to the native orientations.

Similar analyses of the electrostatic binding forces were performed to the kinesin-microtubule complex, where the kinesin motor domain was manipulated by various orientations and distances. The results suggest that when the distance between the kinesin motor domain and the microtubule increases, the binding force toward the neighboring binding site becomes competitive to that toward the native binding site. The calculations from both rotation and spin operations reveal that the electrostatic interaction favors the native or near-native orientations of kinesin motor domain. When the orientation of kinesin motor domain is altered, the binding force reorients it to the native orientation by providing a repulsive binding force. In the operation where the kinesin motor domain was translated along the microtubule, the calculations demonstrate that the electrostatic interactions make significant contributions in guiding the kinesin by favoring the binding on the native binding sites and rejecting the non-native binding positions.

Besides the two examples demonstrated in this work, the StructureMan program is able to help the researchers to study many other large-scale biomolecular interactions. We expect the StructureMan to be combined with DFMD method (Peng et al., 2019) to investigate the biomolecular interactions in the perspective of molecular dynamic simulations in our future work. This novel tool provides an alternative approach to study the biomolecular interactions, especially for large scale biology problems. The StructureMan tool is available at our website: http://compbio.utep.edu/static/downloads/script-for-munipulation2.zip.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

LL and WQ proposed the idea, conducted entire research, and wrote the manuscript. YXia and YXie run the calculations,

collected the data, analyzed the results, and wrote the manuscript. SS developed and tested the programing code. CK tested the program and revised the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This research is funded by Grant SC1GM132043-01 from National Institutes of Health (NIH) and Grant 5U54MD007592 from the National Institutes on Minority Health and Health Disparities (NIMHD), a component of the NIH.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2020.627087/full#supplementary-material

REFERENCES

Bakker, S. E., Ford, R. J., Barker, A. M., Robottom, J., Saunders, K., Pearson, A. R., et al. (2012). Isolation of an asymmetric RNA uncoating intermediate for a single-stranded RNA plant virus. *J. Mol. Biol.* 417, 65–78. doi: 10.1016/j.jmb.2012.01.017

Cao, M., Ye, X., Willie, K., Lin, J., Zhang, X., Redinbaugh, M. G., et al. (2010). The capsid protein of Turnip crinkle virus overcomes two separate defense barriers to facilitate systemic movement of the virus in Arabidopsis. *J. Virol.* 84, 7793–7802. doi: 10.1128/JVI.02643-09

Chen, R., Li, L., and Weng, Z. (2003). ZDOCK: an initial-stage protein-docking algorithm. *Proteins* 52, 80–87. doi: 10.1002/prot.10389

Chen, R., and Weng, Z. (2003). A novel shape complementarity scoring function for protein-protein docking. *Proteins* 51, 397–408. doi: 10.1002/prot.10334

DeBonis, S., Skoufias, D. A., Lebeau, L., Lopez, R., Robin, G., Margolis, R. L., et al. (2004). *In vitro* screening for inhibitors of the human mitotic kinesin Eg5 with antimitotic and antitumor activities. *Mol. Cancer Ther.* 3, 1079–1090.

Dominguez, C., Boelens, R., and Bonvin, M. A. (2003). HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.* 125, 1731–1737. doi: 10.1021/ja026939x

Endow, S. A., Kull, F. J., and Liu, H. (2010). Kinesins at a glance. *J. Cell Sci.* 123, 3420–3424. doi: 10.1242/jcs.064113

Gabb, H. A., Jackson, R. M., and Sternberg, E. M. J. (1997). Modelling protein docking using shape complementarity, electrostatics and biochemical information. *J. Mol. Biol.* 272, 106–120. doi: 10.1006/jmbi.1997.1203

Goddard, T. D., Huang, C. C., and Ferrin, E. T. (2007). Visualizing density maps with UCSF chimera. *J. Struct. Biol.* 157, 281–287. doi: 10.1016/j.jsb.2006.06.010

Grant, B. J., Gheorghe, D. M., Zheng, W., Alonso, M., Huber, G., Dlugosz, M., et al. (2011). Electrostatically biased binding of kinesin to microtubules. *PLoS Biol.* 9:e1001207. doi: 10.1371/journal.pbio.1001207

Hogle, J. M., Maeda, A., and Harrison, C. S. (1986). Structure and assembly of turnip crinkle virus. I. x-ray crystallographic structure analysis at 3.2 A resolution. *J. Mol. Biol.* 191, 625–638. doi: 10.1016/0022-2836(86)90450-X

Huang, S. Y., Grinter, S. Z., and Zou, X. (2010). Scoring functions and their evaluation methods for protein–ligand docking: recent advances and future directions. *Phys. Chem. Chem. Phys.* 12, 12899–12908. doi: 10.1039/c0cp00151a

Huang, Y., Liu, S., Guo, D., Li, L., and Xiao, Y. (2013). A novel protocol for three-dimensional structure prediction of RNA-protein complexes. *Sci. Rep.* 3:1887. doi: 10.1038/srep01887

Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. *J. Mol. Graph.* 14, 33–38. doi: 10.1016/0263-7855(96)00018-5

Huszar, D., Theoclitou, M.-E., Skolnik, J., and Herbst, R. (2009). Kinesin motor proteins as targets for cancer therapy. *Cancer Metastasis Rev.* 28, 197–208. doi: 10.1007/s10555-009-9185-8

Hwang, W., and Karplus, M. (2019). Structural basis for power stroke vs. brownian ratchet mechanisms of motor proteins. *Proc. Natl. Acad. Sci. U.S.A.* 116, 19777–19785. doi: 10.1073/pnas.1818589116

Jackson, J. R., Patrick, D. R., Dar, M. M., and Huang, S. P. (2007). Targeted anti-mitotic therapies: can we improve on tubulin agents? *Nat. Rev. Cancer* 7, 107–117. doi: 10.1038/nrc2049

Jain, A. N. (2006). Scoring functions for protein-ligand docking. *Curr. Protein Pept. Sci.* 7, 407–420. doi: 10.2174/138920306778559395

Jones, S., and Thornton, J. M. (1996). Principles of protein-protein interactions. *Proc. Natl. Acad. Sci. U.S.A.* 93, 13–20. doi: 10.1073/pnas.93.1.13

Jordan, M. A., and Wilson, L. (2004). Microtubules as a target for anticancer drugs. *Nature Reviews Cancer* 4, 253–265. doi: 10.1038/nrc1317

Kavallaris, M. (2010). Microtubules and resistance to tubulin-binding agents. *Nat. Rev. Cancer* 10, 194–204. doi: 10.1038/nrc2803

Lee, Y. R. J., Qiu, W., and Liu, B. (2015). Kinesin motors in plants: from subcellular dynamics to motility regulation. *Curr. Opin. Plant Biol.* 28, 120–126. doi: 10.1016/j.pbi.2015.10.003

Li, C., Li, L., Zhang, J., and Alexov, E. (2012a). Highly efficient and exact method for parallelization of grid-based algorithms and its implementation in DelPhi. *J. Comput. Chem.* 33, 1960–1966. doi: 10.1002/jcc.23033

Li, L., Alper, J., and Alexov, E. (2016a). Cytoplasmic dynein binding, run length, and velocity are guided by long-range electrostatic interactions. *Sci. Rep.* 6:31523. doi: 10.1038/srep31523

Li, L., Alper, J., and Alexov, E. (2016b). Multiscale method for modeling binding phenomena involving large objects: application to kinesin motor domains motion along microtubules. *Sci. Rep.* 6:23249. doi: 10.1038/srep23249

Li, L., Chakravorty, A., and Alexov, E. (2017a). DelPhiForce, a tool for electrostatic force calculations: applications to macromolecular binding. *J. Comput. Chem.* 38, 584–593. doi: 10.1002/jcc.24715

Li, L., Guo, D., Huang, Y., Liu, S., and Xiao, Y. (2011). ASPDock: protein-protein docking algorithm using atomic solvation parameters model. *BMC Bioinformatics* 12:36. doi: 10.1186/1471-2105-12-36

Li, L., Huang, Y., and Xiao, Y. (2013a). How to use not-always-reliable binding site information in protein-protein docking prediction. *PLoS ONE* 8:e75936. doi: 10.1371/journal.pone.0075936

Li, L., Jia, Z., Peng, Y., Chakravorty, A., Sun, L., and Alexov, E. (2017b). DelPhiForce web server: electrostatic forces and energy calculations and visualization. *Bioinformatics* 33, 3661–3663. doi: 10.1093/bioinformatics/btx495

Li, L., Jia, Z., Peng, Y., Godar, S., Getov, I., Teng, S., et al. (2017c). Forces and Disease: Electrostatic force differences caused by mutations in kinesin motor domains can distinguish between disease-causing and non-disease-causing mutations. *Sci. Rep.* 7:8237. doi: 10.1038/s41598-017-08419-7

Li, L., Li, C., Sarkar, S., Zhang, J., Witham, S., Zhang, Z., et al. (2012b). DelPhi: a comprehensive suite for DelPhi software and associated resources. *BMC Biophys.* 5:9. doi: 10.1186/2046-1682-5-9

Li, L., Li, C., Zhang, Z., and Alexov, E. (2013b). On the dielectric constant of proteins: smooth dielectric function for macromolecular modeling and its implementation in DelPhi. *J. Chem. Theory Comput.* 9, 2126–2136. doi: 10.1021/ct400065j

Li, L., Wang, L., and Alexov, E. (2015). On the energy components governing molecular recognition in the framework of continuum approaches. *Front. Mol. Biosci.* 2:5. doi: 10.3389/fmolb.2015.00005

Liwo, A., Ołdziej, S., Pincus, M. R., Wawak, R. J., Rackovsky, S., and Scheraga, A. H. (1997). A united-residue force field for off-lattice protein-structure simulations. I. functional forms and parameters of long-range side-chain interaction potentials from protein crystal data. *J. Comp. Chem.* 18, 849–873. doi: 10.1002/(SICI)1096-987X(199705)18:7<849::AID-JCC1>3.0.CO;2-R

Magnasco, M. O. (1993). Forced thermal ratchets. *Phys. Rev. Lett.* 71, 1477–1481. doi: 10.1103/PhysRevLett.71.1477

Mandelkow, E., and Mandelkow, E. M. (2002). Kinesin motors and disease. *Trends Cell Biol.* 12, 585–591. doi: 10.1016/S0962-8924(02)02400-5

Marrink, S. J., Risselada, H. J., Yefimov, S., Tieleman, D. P., and de Vries, H. A. (2007). The MARTINI force field: coarse grained model for

biomolecular simulations. *J. Phys. Chem. B* 111, 7812–7824. doi: 10.1021/jp071097f

Nakai, R., Iida, S. I, Takahashi, T., Tsujita, T., Okamoto, S., Takada, C., Akasaka, K., et al. (2009). K858, a novel inhibitor of mitotic kinesin Eg5 and antitumor agent, induces cell death in cancer cells. *Cancer Res.* 69, 3901–3909. doi: 10.1158/0008-5472.CAN-08-4373

Pagadala, N. S., Syed, K., and Tuszynski, J. (2017). Software for molecular docking: a review. *Biophys. Rev.* 9, 91–102. doi: 10.1007/s12551-016-0247-1

Peng, Y., Yang, Y., Li, L., Jia, Z., Cao, W., and Alexov, E. (2019). DFMD: fast and effective DelPhiForce steered molecular dynamics approach to model ligand approach toward a receptor: application to spermine synthase enzyme. *Front. Mol. Biosci.* 6:74. doi: 10.3389/fmolb.2019.00074

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., et al. (2004). UCSF chimera–a visualization system for exploratory research and analysis. *J. Comput. Chem.* 25, 1605–1612. doi: 10.1002/jcc.20084

Salas, G. G. S., Hernandez, A. E. L., He, J., Karki, C., Xie, Y., Sun, S., et al. (2019). Using computational approaches to study dengue virus capsid assembly. *Comp. Math. Biophys.* 7, 64–72. doi: 10.1515/cmb-2019-0005

Sarli, V., and Giannis, A. (2008). Targeting the kinesin spindle protein: basic principles and clinical implications. *Clin. Cancer Res.* 14, 7583–7587. doi: 10.1158/1078-0432.CCR-08-0120

Saunders, K., and Lomonossoff, G. P. (2015). The generation of turnip crinkle virus-like particles in plants by the transient expression of wild-type and modified forms of its coat protein. *Front. Plant Sci.* 6:1138. doi: 10.3389/fpls.2015.01138

Schmidt, M., and Bastians, H. (2007). Mitotic drug targets and the development of novel anti-mitotic anticancer drugs. *Drug Resist. Updates* 10, 162–181. doi: 10.1016/j.drup.2007.06.003

Sorger, P. K., Stockley, P. G., and Harrison, C. S. (1986). Structure and assembly of turnip crinkle virus: II. mechanism of reassembly *in vitro*. *J. Mol. Biol.* 191, 639–658. doi: 10.1016/0022-2836(86)90451-1

Tajielyato, N., Li, L., Peng, Y., Alper, J., and Alexov, E. (2018). E-hooks provide guidance and a soft landing for the microtubule binding domain of dynein. *Sci. Rep.* 8:13266. doi: 10.1038/s41598-018-31480-9

Tao, W., South, V. J., Zhang, Y., Davide, J. P., Farrell, L., Kohl, N. E., et al. (2005). Induction of apoptosis by an inhibitor of the mitotic kinesin KSP requires both activation of the spindle assembly checkpoint and mitotic slippage. *Cancer Cell* 8, 49–59. doi: 10.1016/j.ccr.2005.06.003

Tozzini, V. (2005). Coarse-grained models for proteins. *Curr. Opin. Struct. Biol.* 15, 144–150. doi: 10.1016/j.sbi.2005.02.005

Tseng, K. F., Wang, P., Y., Lee, R. J., Bowen, J., Gicking, A. M., et al. (2018). The preprophase band-associated kinesin-14 OsKCH2 is a processive minus-end-directed microtubule motor. *Nat. Commun.* 9:1067. doi: 10.1038/s41467-018-03480-w

Tuszynska, I., and Bujnicki, J. M. (2011). DARS-RNP and QUASI-RNP: New statistical potentials for protein-RNA docking. *BMC Bioinformatics* 12:348. doi: 10.1186/1471-2105-12-348

Vale, R. D., Reese, T. S., and Sheetz, P. M. (1985). Identification of a novel force-generating protein, kinesin, involved in microtubule-based motility. *Cell* 42, 39–50. doi: 10.1016/S0092-8674(85)80099-4

von Mering, C., Krause, R., Snel, B., Cornell, M., Oliver, S. G., Fields, S., et al. (2002). Comparative assessment of large-scale data sets of protein–protein interactions. *Nature* 417, 399–403. doi: 10.1038/nature750

Wang, Y., Chu, X., and Wang, J. (2014). *Multiscale Simulations of Large Conformational Changes of Disordered and Ordered Proteins induced by their Partners in Computational Approaches to Protein Dynamics: from Quantum to Coarse-Grained Methods*. Boca Raton: CRC Press.

Wei, N., Heaton, L. A., Morris, T. J., and Harrison, C. S. (1990). Structure and assembly of turnip crinkle virus. VI. identification of coat protein binding sites on the RNA. *J. Mol. Biol.* 214, 85–95. doi: 10.1016/0022-2836(90)90148-F

Xian, Y., Karki, C. B., Silva, S. M., Li, L., and Xiao, C. (2019). The roles of electrostatic interactions in capsid assembly mechanisms of giant viruses. *Int. J. Mol. Sci.* 20:1876. doi: 10.3390/ijms20081876

Yan, Y., Zhang, D., Zhou, P., Li, B., and Huang, S. Y. (2017). HDOCK: a web server for protein–protein and protein–DNA/RNA docking based on a hybrid strategy. *Nucleic Acids Res.* 45, W365–W373. doi: 10.1093/nar/gkx407

Zhou, G. P. (2015). Editorial: current progress in structural bioinformatics of protein-biomolecule interactions. *Med. Chem.* 11, 216–217. doi: 10.2174/1573406411666141229162618

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Molecular Biosciences

ORIGINAL RESEARCH
published: 27 January 2021
doi: 10.3389/fmolb.2020.616135

Residue-Level Contact Reveals Modular Domain Interactions of PICK1 Are Driven by Both Electrostatic and Hydrophobic Forces

*Amy O. Stevens and Yi He**

Department of Chemistry and Chemical Biology, The University of New Mexico, Albuquerque, NM, United States

PICK1 is a multi-domain scaffolding protein that is uniquely comprised of both a PDZ domain and a BAR domain. While previous experiments have shown that the PDZ domain and the linker positively regulate the BAR domain and the C-terminus negatively regulates the BAR domain, the details of internal regulation mechanisms are unknown. Molecular dynamics (MD) simulations have been proven to be a useful tool in revealing the intramolecular interactions at atomic-level resolution. PICK1 performs its biological functions in a dimeric form which is extremely computationally demanding to simulate with an all-atom force field. Here, we use coarse-grained MD simulations to expose the key residues and driving forces in the internal regulations of PICK1. While the PDZ and BAR domains do not form a stable complex, our simulations show the PDZ domain preferentially interacting with the concave surface of the BAR domain over other BAR domain regions. Furthermore, our simulations show that the short helix in the linker region can form interactions with the PDZ domain. Our results reveal that the surface of the βB-βC loop, βC strand, and αA-βD loop of the PDZ domain can form a group of hydrophobic interactions surrounding the linker helix. These interactions are driven by hydrophobic forces. In contrast, our simulations reveal a very dynamic C-terminus that most often resides on the convex surface of the BAR domain rather than the previously suspected concave surface. These interactions are driven by a combination of electrostatic and hydrophobic interactions.

Keywords: PICK1, inter-domain dynamics, coarse-grained simulations, key residues, physical forces

OPEN ACCESS

Edited by:
Xiakun Chu,
Stony Brook University, United States

Reviewed by:
Liu Fei,
Nankai University, China
Lin Li,
The University of Texas at El Paso, United States
Yongqi Huang,
Hubei University of Technology, China

***Correspondence:**
Yi He
yihe@unm.edu

Specialty section:
This article was submitted to Biological Modeling and Simulation, a section of the journal Frontiers in Molecular Biosciences

Received: *11 October 2020*
Accepted: *15 December 2020*
Published: *27 January 2021*

Citation:
Stevens AO and He Y (2021) Residue-Level Contact Reveals Modular Domain Interactions of PICK1 Are Driven by Both Electrostatic and Hydrophobic Forces. Front. Mol. Biosci. 7:616135. doi: 10.3389/fmolb.2020.616135

INTRODUCTION

Protein Interacting with C Kinase-1 (PICK1) is a multi-domain mammalian membrane protein (Staudinger et al., 1995). In the monomeric form, PICK1 is comprised of one PDZ (PSD-95/Dlg1/ZO-1) domain (Sheng and Sala, 2001; Hung and Sheng, 2002) and one BAR (Bin/amphiphysin/Rvs) domain (Takei et al., 1999). While each is a common modular domain, PICK1 is unique as it is the only known protein that contains both a PDZ and a BAR domain. The domains are connected via an intrinsically disordered linker that allows the PDZ domain to have a wide range of motion around the BAR domain. This range of motion increases the effective concentration of PDZ domain so that it can form protein-protein interactions with a variety of

cellular proteins. Furthermore, the N- and C-termini are intrinsically disordered regions that may be involved in the regulation mechanism of PICK1. The short N-terminus (∼18 residues) sits before the PDZ domain and is enriched with many acidic residues. The lengthy C-terminus (∼60 residues) follows the central BAR domain and is characterized by a stretch of acidic residues. The structure of PICK1 is shown in **Figure 1**. Functionally, PICK1 is involved in the trafficking of a variety of proteins, including receptors, transporters, and ionic channels (Staudinger et al., 1997; Torres et al., 1998, 2001; Dev et al., 1999; Boudin et al., 2000; Cowan et al., 2000; El Far et al., 2000; Takeya et al., 2000; Jaulin-Bastard et al., 2001; Lin et al., 2001a,b; Penzes et al., 2001; Duggan et al., 2002; Hruska-Hageman et al., 2002; Perroy et al., 2002; Enz and Croci, 2003; Hirbec et al., 2003; Leonard et al., 2003; Williams et al., 2003; Meyer et al., 2004; Reymond et al., 2005). Its wide range of functions in regulating membrane proteins has drawn attention as a possible drug target. PICK1 has been identified as a possible target in ischemia (Dixon et al., 2009), Alzheimer's disease (Alfonso et al., 2014), Parkinson's disease (He et al., 2018), chronic pain (Garry et al., 2003), and cocaine addiction (Jensen et al., 2018). If PICK1 is to be targeted with the necessary affinity and specificity, an in-depth understanding of the activation mechanism and protein-protein interactions of PICK1 are vital.

PICK1 interacts with the final C-terminal residues of receptors, transporters and transmembrane channels via its PDZ domain (Hanley, 2008). The PICK1 PDZ domain has a well-defined binding pocket with canonical Class II ligand-PDZ interactions (Madsen et al., 2005). PICK1 regulates the trafficking of membrane proteins via electrostatic interactions between the membrane and the dimeric BAR domain. The family of BAR domain proteins is one of the largest groups of membrane curving proteins in the cell. The amphiphysin BAR domain binds to the negatively charged lipid membrane via two pairs of positively charged residues (Peter et al., 2004). Sequence alignment with the amphiphysin BAR domain suggests that five positively charged residues (K251, K252, K257, K266, and K268) on the PICK1 BAR domain are responsible for its interactions with the lipid membrane (Xu and Xia, 2007). Point mutation analysis further confirms the importance of these residues in lipid membrane binding (Jin et al., 2006). While an atomic-level understanding of these processes remains unclear, detailed hypotheses of auto-inhibition exist. Jin et al. used truncated mutants of PICK1 to test their lipid-binding capabilities (Jin et al., 2006). It was shown that the deletion of the C-terminus promotes BAR interactions with the lipid membrane. Results affirmed that PICK1 is negatively regulated by its C-terminus and positively regulated by its linker and PDZ domain (Jin et al., 2006). Furthermore, it is suggested that the negatively charged region of the C-terminus negatively regulates the function of PICK1 by interacting with and thus covering the critical positively charged residues on the concave surface of the BAR domain.

Our previous work has shown that the PDZ domain forms interactions with the BAR domain, which may prevent the binding between the BAR domain to the lipid membrane (He et al., 2011). These results support the hypothesis of an inactivated state of PICK1 in which ligand binding results in activation via a conformational change to expose the BAR domain to the membrane (Lu and Ziff, 2005; Rocca et al., 2008). A more recent experiment has revealed a more dynamic pattern for the interactions between the BAR and PDZ domains (Karlsen et al., 2015). Small-angle X-ray scattering (SAXS) analysis revealed the wide range of flexibility of the PDZ domain via the intrinsically disordered linker. Higher-order oligomeric structures of PICK1 further enable the dynamic positioning of the PDZ domains (Karlsen et al., 2015). Moreover, several

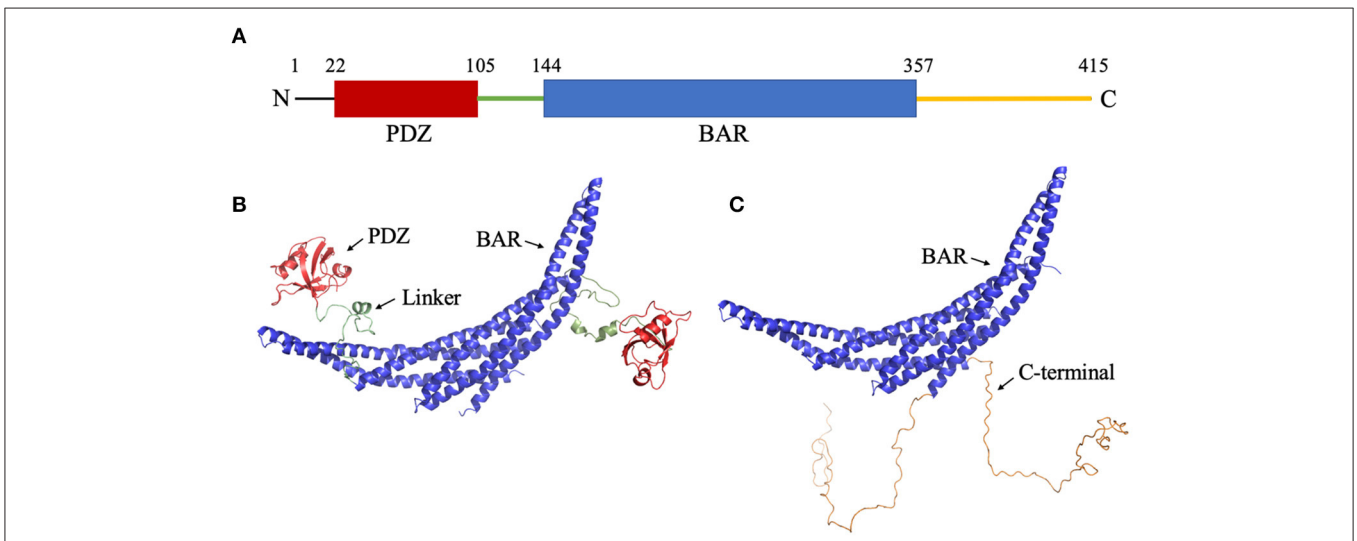


FIGURE 1 | Structure of PICK1. **(A)** Sequence of PICK1. Monomeric PICK1 is comprised of two modular domains, PDZ (red) and BAR (blue), and three intrinsically disordered regions, N-terminal, linker (green), and C-terminal (yellow). **(B)** Structure of PICK1 in the absence of N- and C-termini. **(C)** Dimeric BAR domain and intrinsically disordered C-terminal.

experiments done by different groups show that the linker of the PICK1 protein may play a key role in promoting BAR interactions with the lipid membrane (Jin et al., 2006; Herlo et al., 2018). To understand the interplay between different parts of the PICK1 protein in its biological dimeric form, dynamics information at residue resolution and a very fine time resolution (picosecond or nanosecond time scale) is essential. However, such dynamics information is difficult to obtain from experiments since PICK1 is inside of the cell and forms aggregates with itself.

PICK1 is a large protein and performs its biological function in dimeric form. Such a system consists of more than 800 residues and may have a dimension over 20 nm because of its flexibility (Karlsen et al., 2015). A system of this size is extremely computationally demanding to simulate with all-atom force fields. Physics-based coarse-grained models have a long history of helping scientists to tackle systems of this size with reasonable computational resources. The physics-based UNited-RESidue (UNRES) (Liwo et al., 1997a) force field, which was originally proposed by Liwo and Scheraga, is one of the extensively tested coarse-grained models that can be used to predict protein structure (He et al., 2009, 2019) and probe large protein dynamics (He et al., 2011; Gołaś et al., 2012; Mozolewska et al., 2015). With several generations of optimization, UNRES is a reliable tool to explore the inter-domain dynamics of PICK1.

Here, we present the needed structural and dynamics information that is responsible for the auto-inhibition of PICK1 and provides a complete picture of the inter-domain dynamics of PICK1. We implemented coarse-grained UNRES molecular dynamics simulations to model two systems: (1) BAR domain with PDZ domain and linker and (2) BAR domain with C-termini. These truncations are modeled after experimental work (Jin et al., 2006) that describes systems (1) and (2) as the two extreme cases of the enhanced and reduced biological function of PICK1, respectively. The truncated systems allow us to more readily isolate the key interactions in each of these extreme cases. Our results show that the PDZ domain and linker form dynamic interactions on the concave surface and side of the BAR domain dimer. The PDZ domain interacts with the BAR domain dimer via residues that are located in the regions which are regulated by the electrostatic allosteric effects upon the formation of the PDZ-ligand complex. Surprisingly, our results do not show the C-termini interacting with the concave surface of the BAR domain via electrostatic interactions as previously expected. Rather, the movements of the C-termini are vastly dynamic and generally reside at the central region of the convex surface of the BAR domain.

METHODS

Though the experimental structures of PICK1 and the PICK1 dimer have not yet been determined, the PICK1 BAR domain has a high sequence identity with Arfaptin-2, a N-BAR domain protein (Nakamura et al., 2012). The dimer structures of N-BAR domains have been well-established. The starting structures used in the simulations were created using the BAR dimer in Arfaptin-2 as a template to create the PICK1 dimeric BAR domains using MODELER (Šali and Blundell, 1993; Fiser et al., 2000; Martí-Renom et al., 2000; Webb and Sali, 2016). With the BAR dimer, the structure of the C-termini of PICK1 was randomly generated and attached to the BAR domain. The structure of the PDZ domain has been previously experimentally determined (Pan et al., 2007) and was used as a structural template in our protocol. After the PDZ domains were randomly placed with respect to the BAR domain, the intrinsically disordered linker was added to connect the PDZ domain and BAR domain. The initial structures are shown in **Figures 1B,C**.

UNRES (Liwo et al., 1997a,b, 1998, 2011; He et al., 2009; Sieradzan et al., 2014) uses a simplified representation in which a protein chain is composed of a sequence of α-carbon atoms connected by virtual bonds with attached side chains. To reduce computational cost and maintain residue-level resolution, each residue is represented by two interaction sites. One interaction site is centered between two consecutive Cα atoms, and the other is located at the center of the mass of the corresponding side chain. As a physics-based coarse-grained force field, the UNRES energy function has been averaged over the lost degree of freedom when simplifying from all-atom to coarse-grained representations. Recently, UNRES has been expanded to include both nucleic acids and lipid membranes (He et al., 2013; Sieradzan et al., 2018; Zięba et al., 2019). Canonical MD simulations (13 trajectories) were carried out for each complex to explore the interplay between the different parts of PICK1 and the crescent BAR domains. The most recently parameterized UNRES force field (Sieradzan et al., 2017; Lubecka et al., 2019), which has been evaluated based on CASP 13 targets, was used in this work.

The input files (including all input parameters) were generated using the UNRES server at http://unres-server.chem.ug.edu.pl. While the UNRES server was used to generate input files for each system, simulations were performed locally because the UNRES server has a size limit that is smaller than the system sizes explored in this work. The simulations used the latest UNRES source code that can be downloaded at https://unres.pl/downloads. The input files generated by the UNRES server used the most recent UNRES force field, namely "NEWCT-9P = JCP 150 155104 (2019)." Users must click the "advanced" button (after selecting "MD" option) located at the top-right of the web page to use this force field. Both systems started from the PDB structures described above with periodic boundary conditions set at 10,000.0 Angtroms. No secondary structure restraints were applied. Distance restraints were manually added to the input files generated by the UNRES server to maintain the structure of the BAR and PDZ domains but not the linker or the C-terminus. It should be noted that the UNRES server does not include keywords to add distance restraints. Since PICK1 is much larger than the proteins used to parametrize the force field, a higher temperature (350K) was used for all canonical MD simulations of the two systems simulated. It should be noted that the 350K used here does not directly correspond to 350K in a biological system. Rather, a temperature of 350K is used to estimate a temperature between 300K and 350K in a biological system based on the evaluation of the previous work (Liwo et al., 2019). The

time increment for integrating the equations of motion δt was 9.78 fs. Thirteen trajectories were carried out, and each trajectory has 80,000,000 steps. Snapshots of structures are outputted every 10,000 steps. Our input files of all systems have been included in the supporting information. All other parameters are default values provided by the UNRES server.

RESULTS

Root mean square deviation (RMSD) and radius of gyration (Rg) analysis were performed to quantify the flexibility of the dynamic system. Frequency refers to the proportion of frames with the given distance. **Figure 2** shows the RMSD and radius of gyration calculated using all the trajectories of System 1 (the BAR domain with the PDZ domain and linker). The RMSD plot (**Figure 2A**) shows a median RMSD at ∼28 Å. This is a significant variation from the initial structure. Furthermore, the radius of gyration analysis supports these results as the size of the protein fluctuates between 30 and 60 Å with relatively significant frequencies. This analysis reveals the wide range of motion of the PDZ domain about the BAR domain as a result of the flexible linker.

For each of the two systems, contact maps were used to reveal the major interactions between any pair of residues. Contact was defined as any two Cα atoms at least five residues apart with a distance separation of 8Å or less. In System 1, the PICK1 complex is in the proposed inactivated state as the protein was neither in complex with ligand nor in proximity with the lipid membrane. As expected for inactivated PICK1, the PDZ domains formed contact with a wide range of residues located on the concave face and side surface of the BAR domains, as seen in **Figure 3**. **Figure 3A** describes the contact between the BAR domain and the PDZ domain and the BAR domain and the intrinsically disordered linker. Both the PDZ domain and the linker form the majority of interactions with residues 150–200 and 250–300 of the BAR domain. **Figure 3B** highlights these regions of residues on the BAR domain dimer. The PDZ domain and linker region reside near the concave surface of the BAR domain dimer in the inactivated state of PICK1. Though there are extensive interactions between PDZ and BAR domains, none of the interactions appear in >10% of the frames in the combined trajectories. This agrees with previous experimental observations that suggest a dynamic interaction pattern between the PDZ and BAR domains.

Detailed residue-residue interaction analysis revealed that the short helical portion of the linker region forms significant interactions with the PDZ domain. Key residues of PDZ-linker interaction were elucidated by identifying the most prevalent contacts, in this case, forming a contact in 13% of the frames. Ten key interaction pairs were identified between the PDZ domain and the linker, as shown in **Table 1**. It is not surprising that the linker can form significant contact with the PDZ domain of PICK1 as the linker and PDZ domain are next to each other in sequence. It should be noted that all listed contacts in **Table 1** are formed between the helical fragment of the linker and the PDZ domain. Previous work has highlighted the importance of the helical fragment in the linker region in assisting the alignment of the BAR domain to the membrane (Herlo et al., 2018). The

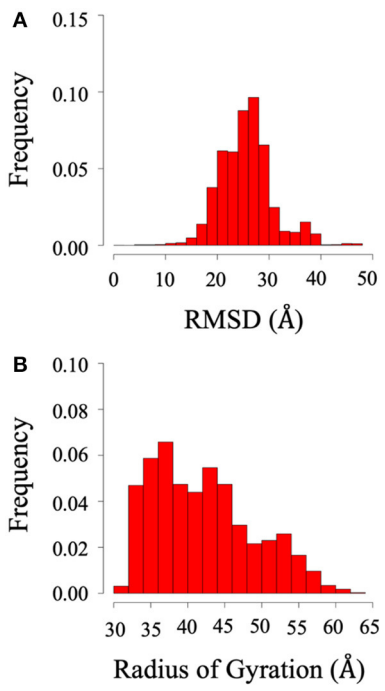


FIGURE 2 | RMSD and radius of gyration of System 1 (BAR domain with PDZ domain and linker). **(A)** RMSD. **(B)** Radius of gyration. The wide range of frequency signifies the system is very dynamic.

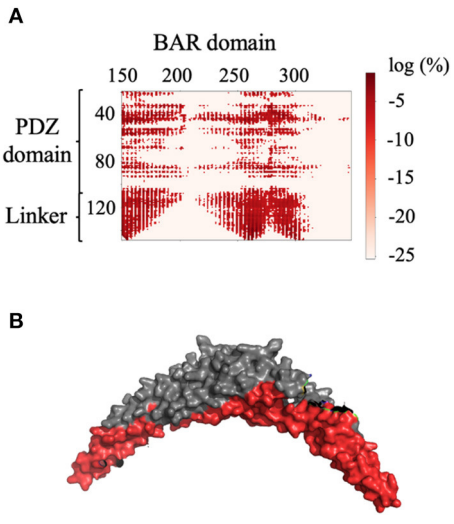


FIGURE 3 | Contact between the BAR domain dimer and the PDZ domain and linker. **(A)** Contact map of BAR domain with PDZ domain and linker region. PDZ domain and linker region interact with approximate residues 150–200 and 250–300 of the BAR domain. **(B)** Dimeric BAR domain with residues 150–200 and 250–300 colored red. The color bar describes the probability of the contact as the log of the percentage of frames that the contact occurs.

linker may compete with the BAR domain to interact with the PDZ domain in the inactivated PICK1 dimer.

These ten key pairs are hydrophobic interactions between the βB-βC loop, βC strand, and αA-βD loop of the PDZ domain and the short helix fragment of the linker region, as shown in **Figure 4**. It can be seen that the PDZ domain forms a group of hydrophobic interactions surrounding the hydrophobic helical fragment in the linker region. It has been shown that this short helical region in the linker is critical for the biological function of the BAR domain (Herlo et al., 2018). Our results suggest that this linker may mediate and/or regulate the interactions between the PDZ and the BAR domains. In addition to the frequency of each contact pair, the lifetime of each pair has also been investigated. The lifetime was calculated based on the lasting time of each contact. Since it is difficult to directly connect UNRES simulation steps to the real world time scale, lifetime is defined directly using UNRES steps. While all contacts shown in **Table 1** have a probability larger than 13%, their lifetime is rather short compared to PDZ and BAR domain interactions. This may be due to the flexible nature of the linker region.

The interaction pattern between the PDZ and the BAR domains is quite different than the interactions between the PDZ domain and the linker. The top ten contact residue pairs are shown in **Table 2**. The three most probable interaction pairs (probability >7%) are between the βB-βC loop of the PDZ domain and the BAR domain, as shown in **Figure 5**. Though the probability of each of the ten pairs is below 10%, the lifetime of these interactions is much longer than the lifetime of the PDZ-linker interactions. These results suggest that the PDZ and

TABLE 1 | Interacting residue pairs between PDZ and Linker.

Residue 1	Residue type	Residue 2	Residue type	Probability (%)	Lifetime*
50	VAL	114	LEU	21.1	1.70 ± 1.31
66	ALA	114	LEU	17.2	1.80 ± 1.65
43	TYR	113	SER	16.6	1.59 ± 1.14
41	ALA	117	VAL	13.7	1.67 ± 1.19
43	TYR	117	VAL	13.6	1.69 ± 1.16
66	ALA	112	MET	13.5	1.68 ± 1.38
43	TYR	112	MET	13.1	1.55 ± 1.27
50	VAL	118	LEU	13.1	1.67 ± 1.34
66	ALA	113	SER	13.1	1.64 ± 1.19
50	VAL	113	SER	13.0	1.58 ± 1.18

**unit is 100,000 UNRES simulation steps.*

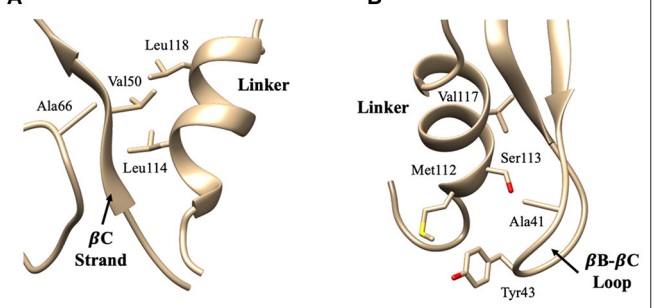


FIGURE 4 | Key interaction pairs between the PDZ domain and the linker. **(A)** shows the hydrophobic core forming between the PDZ βC-strand/αA-βD loop and the short helix of the linker. Key interaction pairs Val50-Leu114 and Ala66-Leu114 listed in **Table 1** can be visualized in **(A)**. **(B)** shows the hydrophobic core forming between the PDZ βB-βC loop and the short helix of the linker. Key interaction pairs Tyr43-Ser113 and Ala41-Val117 listed in β can be visualized in **(B)**.

TABLE 2 | Interacting residue pairs between PDZ and BAR.

Residue 1	Residue type	Residue 2	Residue type	Probability (%)	Lifetime*
44	CYS	156	LEU	8.5808	4.28 ± 5.65
42	GLN	156	LEU	7.490471	3.95 ± 5.44
43	TYR	156	LEU	7.446148	6.46 ± 11.89
130	SER	608	SER	6.905416	3.01 ± 3.25
44	CYS	152	ARG	6.205124	3.68 ± 4.63
43	TYR	153	LEU	5.983512	3.90 ± 4.73
42	GLN	160	ALA	5.983512	2.06 ± 1.88
54	ASP	156	LEU	5.921461	4.61 ± 9.70
42	GLN	258	PHE	5.673256	3.79 ± 4.38
84	VAL	265	LEU	5.575747	3.84 ± 3.89

**unit is 100,000 UNRES simulation steps.*

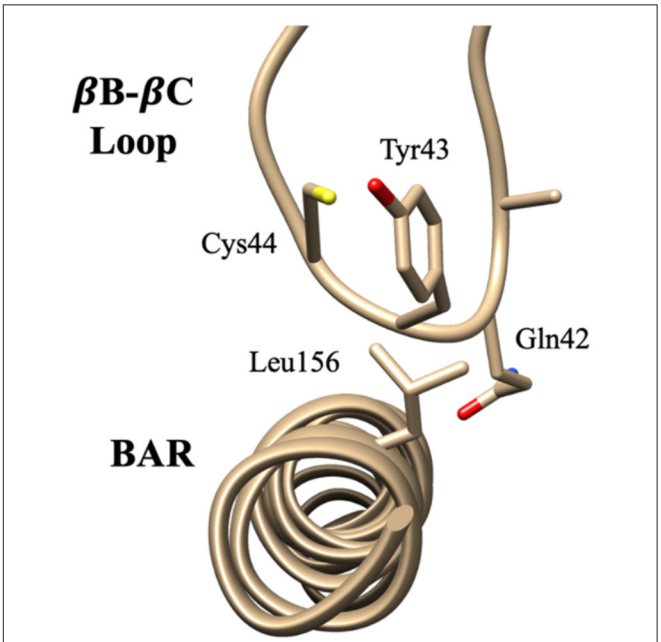


FIGURE 5 | Key interaction pairs between the PDZ domain and the BAR domain. Visualization of key interaction pairs between the PDZ βB-βC loop and the BAR domain (Gln42-Leu156, Tyr43-Leu156, and Cys44-Leu15) listed in **Table 2**.

BAR domain interactions are more stable than the PDZ-linker interactions despite lower probabilities.

To identify the preferred regions on the dimeric BAR domain which interact with the PDZ domains, cluster analysis revealed the five most probable positions of the PDZ domains in space. **Figure 6** portrays an overlay of these five clusters, where the dimeric BAR domain is shown in gray and each cluster is represented by a unique color of the PDZ domain. Furthermore, the five key positively charged residues on the concave surface of the BAR domain that readily interact with the surface of the lipid membrane are colored red. The most probable positions of the PDZ domain could physically block these key residues on the BAR domain from interacting with the membrane. While the most probable positions of the PDZ domains are on the concave surface of the BAR domain, the movement of the PDZ domains remains very dynamic. When the PDZ domains depart from the concave surface of BAR dimer, it may interact with the C-terminus of its binding partners and pull the BAR domain closer to the lipid membrane.

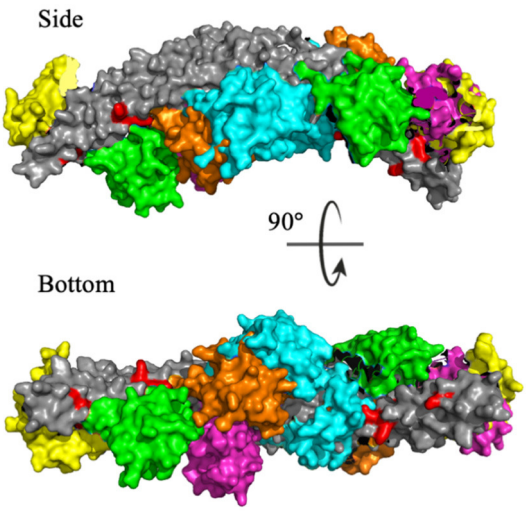


FIGURE 6 | Cluster analysis reveals the most probable positions of the PDZ domains. The dimeric BAR domain is shown in gray and each cluster of the PDZ domains is shown in a unique color. Cluster 1 (yellow) represents 30.9% of the frames, Cluster 2 (pink) represents 20.8% of the frames, Cluster 3 (green) represents 18.9% of the frames, Cluster 4 (orange) represents 18.8% of the frames, and Cluster 5 (cyan) represent 10.5% of the frames. K251, K252, K257, K266, and K268 are colored red.

While cluster analysis reveals the most probable positions of the PDZ domain in respect to the BAR domain, RMSD and radius of gyration analysis reveal that the system has widely dynamic movements. In efforts to capture this range of motion and make a direct comparison to data reported by previous experiments (Karlsen et al., 2015), we performed centroid distance analysis as shown in **Figure 7**. Overall, our results agree with experimental data. The peak of the wide range of distances demonstrates the wide range of motion of the PDZ domain about the BAR domain. The major peaks of the distance distributions are for PDZ to BAR-Linker and PDZ to BAR-Tip reported by experiments was 20*A* to 40*A*, which agrees with our simulation data. In contrast, the distance between PDZ and BAR-Center does not precisely agree with experiments. Experimental data report the distance to be 60*A* to 100*A* while our simulations have shown a much broader distribution for this pair. For PDZ-PDZ distance, our simulations were able to capture the range corresponding to the range reported by experiments. It should be noted that while our simulations did produce a minor peak near 120*A* that directly agrees with experiments, the overall distance distribution is shifted slightly to the left.

Root mean square deviation (RMSD) and radius of gyration (Rg) analysis was performed to quantify the flexibility of System

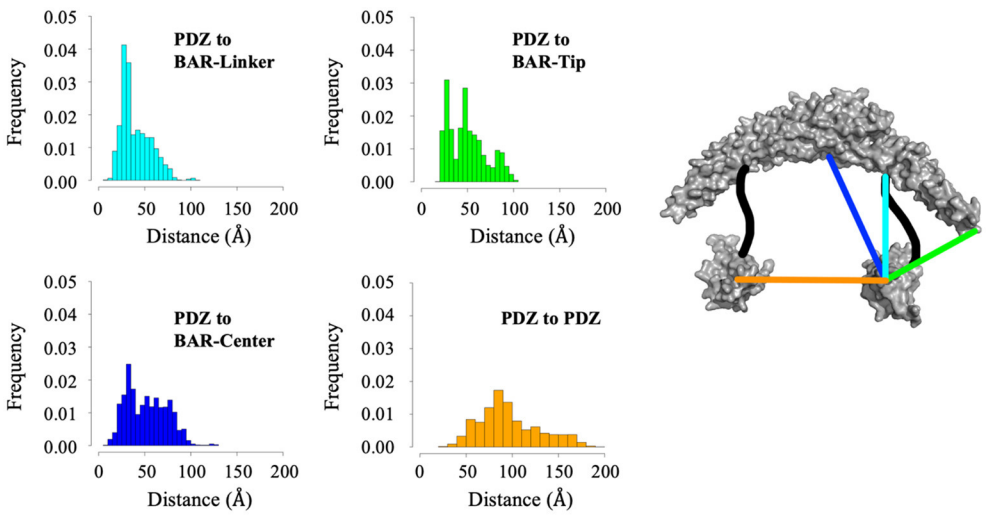


FIGURE 7 | Distance analysis between the PDZ domain and the BAR domain. Distance between the PDZ domain and the BAR-Linker was defined by residues L60 and S130. Distance between the PDZ domain and the BAR-tip was defined by residues L60 and S262. Distance between the PDZ domain the BAR-Center was defined by residues L60 and T167. Distance between the two PDZ domains was defined by residues L60 and D390.

2 (the BAR domain with the C-termini), as shown in **Figure 8**. The RMSD plot (**Figure 8A**) shows the RMSD ranging from ∼25 to 45 Å. This is a significant variation from the initial structure and demonstrates the wide range of motion of the C-termini. Furthermore, the radius of gyration analysis (**Figure 8B**) supports these results as the size of the protein fluctuates between 30 and 40 Å with relatively significant frequencies. This basic analysis reveals the wide range of motion of the C-terminal.

The previous literature (Jin et al., 2006) hypothesizes that the C-terminus negatively regulates the function of PICK1 by interacting with the key positively charged residues (K251, K252, K257, K266, and K268) on the concave surface of the BAR domain dimer that are critical to forming interactions with the lipid membrane. Interestingly, our results do not support these hypotheses. The C-termini are very dynamic and have a wide range of interactions with both with each other and the dimeric BAR domain as shown in **Figure 9**. The black boxes indicate contact between negatively charged stretch of residues that comprise the C-terminus (D380-D389) may form electrostatic interactions with the positively charged residues (K251, K252, K257, K266, and K268) on the BAR domains. These interactions formed contact in <1% of the frames with a separation of <8.0 Å. The two C-termini formed contact with each other as well. Most notably, the contact dissipates at the stretch of negatively charged residues (D380-D389).

While the C-terminus contains many charged residues, the driving forces guiding the interactions between the BAR and the C-terminus is unknown. Our work identifies the top ten pairs of residues forming interactions between the BAR domain and the C-terminus of PICK1. While the majority of identified pairs are driven by hydrophobic interactions, we also identified electrostatic interactions such as K209–D347. The high prevalence of hydrophobic interactions that are entropy driven may be due to the flexibility of the C-termini. All the contacts between the BAR domain and the C-terminus have a short average lifetime that is similar to the PDZ and linker interactions. Since the C-termini are flexible, the contacts between BAR and C-terminus form and break continuously.

All residues identified in the top ten interaction pairs listed in **Table 3** are highlighted in red in the BAR-C-termini structure shown in **Figure 10**. Potentially, these residues form the most

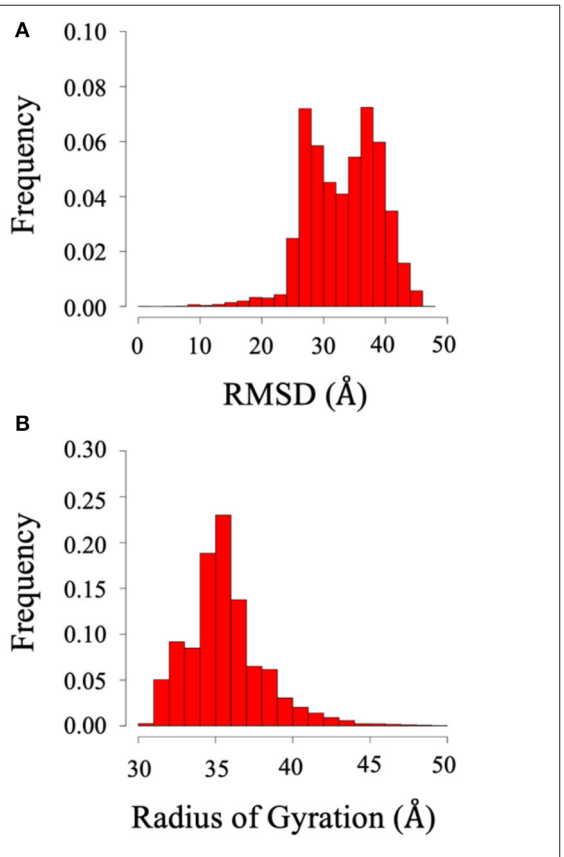


FIGURE 8 | RMSD and radius of gyration of System 2 (BAR domain with C-termini). **(A)** RMSD. **(B)** Radius of gyration. The wide range of frequency signifies the system is very dynamic.

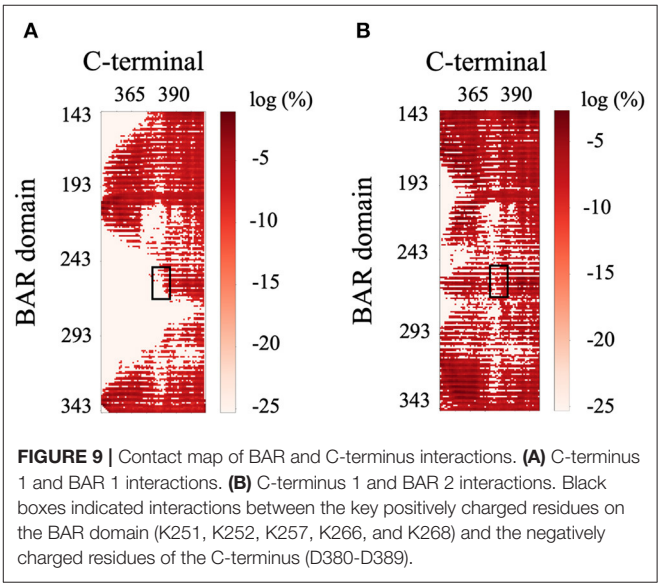


FIGURE 9 | Contact map of BAR and C-terminus interactions. **(A)** C-terminus 1 and BAR 1 interactions. **(B)** C-terminus 1 and BAR 2 interactions. Black boxes indicated interactions between the key positively charged residues on the BAR domain (K251, K252, K257, K266, and K268) and the negatively charged residues of the C-terminus (D380-D389).

TABLE 3 | Interacting residue pairs between C-terminus and BAR domains.

Residue 1	Residue type	Residue 2	Residue type	Probability (%)	Lifetime*
341	MET	348	CYS	21.4	1.60 ± 1.10
342	SER	348	CYS	17.2	1.40 ± 0.80
209	LYS	347	ASP	17.0	1.31 ± 0.70
342	SER	347	ASP	16.6	1.26 ± 0.63
206	ALA	348	CYS	16.6	1.41 ± 0.89
341	MET	349	TYR	14.5	1.46 ± 0.85
342	SER	349	TYR	13.6	1.44 ± 0.84
210	PHE	348	CYS	13.5	1.38 ± 0.80
206	ALA	351	VAL	13.4	1.55 ± 1.03
340	THR	345	TYR	13.0	1.18 ± 0.53

**unit is 100,000 UNRES simulation steps.*

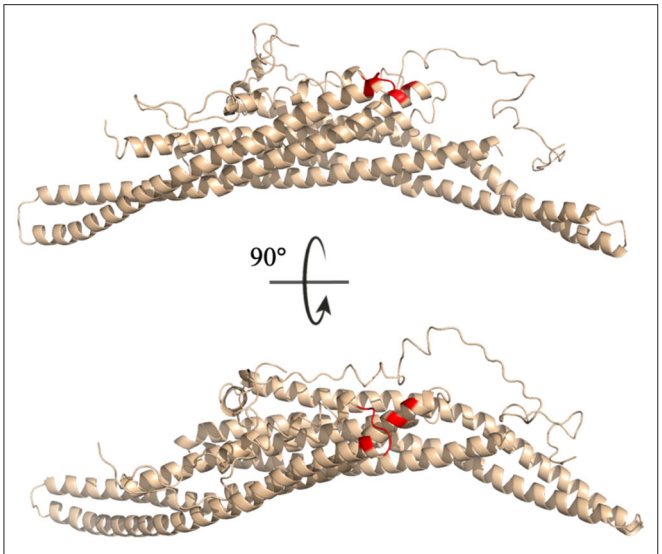


FIGURE 10 | Key interaction pairs between the C-terminal and the BAR domain. All key residues listed in **Table 3** have been colored red.

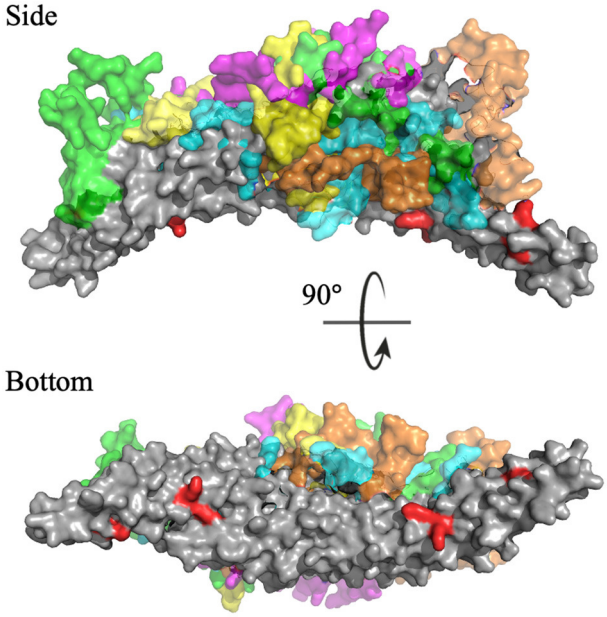


FIGURE 11 | Cluster analysis reveals the most probable positions of the C-termini. The dimeric BAR domain is shown in gray and each cluster of the C-termini is shown in a unique color. Cluster 1 (purple) represents 66.6% of the frames, Cluster 2 (green) represents 38.9% of the frames, Cluster 3 (cyan) represents 22.3% of the frames, Cluster 4 (orange) represent 7.5% of the frames, and Cluster 5 (yellow) represents 4.6% of the frames. K251, K252, K257, K266, and K268 are colored red.

probable interactions because of simple proximity. The BAR and C-termini most readily interacting at their connection site reinforces the notion of significantly flexibile C-termini.

Cluster analysis revealed the five most probable positions of the C-termini in space. **Figure 11** portrays an overlay of these five clusters, where the dimeric BAR domain is shown in gray and each cluster is represented by a unique color of the C-termini. Furthermore, the key positively charged residues on the concave surface of the BAR domain that readily interact with the surface of the lipid membrane are colored red. The most probable positions of the C-termini are centered on the convex surface of the dimeric BAR domain. The C-termini do not readily cover the key positively charged residues on the concave surface of the dimeric BAR domain as previously suspected. These results are in agreement with previous MD simulations of the PICK1 system (Salzer et al., 2017).

DISCUSSION

Our results demonstrate that the interdomain dynamics of PICK1 are driven by both electrostatic and hydrophobic interactions. Here, we identified key interaction pairs between the PDZ domain, linker, and dimeric BAR domain that are primarily hydrophobic interactions. While our results agree with previous experimental observations which suggest dynamic PDZ and BAR interaction patterns, the PDZ domain does have preferences on regions of interactions on the BAR domain. Interestingly, key residue interactions do not include the previously suspected positively charged residues (K251, K252, K257, K266, and K268) of the BAR domain but rather include neighboring residues. Surprisingly, the short helical fragment in the linker can form extensive interactions with the PDZ domain, potentially outcompeting the BAR domain. The biological function of the helical fragment may be more than just help to align to the BAR domain on the lipid membrane.

The interaction pairs demonstrate the significance of the βB-βC loop (Ala41, Gln42, and Tyr43) of the PDZ domain in initiating PDZ-BAR and PDZ-linker contact. Previous structural prediction via small-angle X-ray scattering (SAXS) analysis was unable to determine the orientation of the PDZ domain in PDZ-BAR interactions, but made the prediction that the βB-βC loop of the PDZ domain would orient toward the concave surface of the BAR domain (Madasu et al., 2015) Our simulations support this early hypothesis. Furthermore, previous literature reports the importance of the βB-βC loop in complex formation between the PDZ domain and activation ligand. Our previous work demonstrates the uniqueness of the PICK1 PDZ βB-βC loop (Stevens and He, 2020). A recent publication demonstrated a small-molecule inhibitor of the PICK1 PDZ domain with both strong affinity and specificity via targeting both the binding pocket and βB-βC loop of the PDZ domain (Christensen et al., 2020) Additionally, the βB-βC loop has been identified as an important player in PDZ-membrane interactions (Pan et al., 2007; Erlendsson and Madsen, 2015). Here, we show the relevance of the βB-βC loop in PDZ-BAR contact in the absence of an activating ligand. Key hydrophobic and electrostatic interactions between the PDZ domain and the BAR domain are initiated by residues that comprised the βB-βC loop. Furthermore, the interaction pairs

reveal the significance of the βB-βC loop in initiating PDZ-BAR contact.

Previous experimental results (Jin et al., 2006) suggest that the C-terminus negatively regulates the function of PICK1 by physically covering the concave surface of the BAR domain dimer that interacts with the lipid membrane. The negatively charged stretch of residues that comprise the C-terminus (D380–D389) may form electrostatic interactions with the positively charged residues on the BAR domains that are critical in interactions with the negatively charged lipid bilayer. Interestingly, our results do not support these hypotheses. Our results demonstrate that the C-termini of PICK1 could directly interact with the positively charged residues (K251, K252, K257, K266, and K268) on the BAR domain, but actual interactions between these residues observed in our simulations are rare. We suspect that the C-termini may inhibit the higher-order aggregates of PICK1. PICK1 performs its biological function by forming clusters at the cell surface. Rather than covering key positively charged residues on the concave surface of the BAR domain, the C-termini may negatively inhibit the function of PICK1 by preventing scaffolding.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

AS performed data analysis, carried out simulations, and wrote the manuscript. YH designed research strategy, performed data analysis, and wrote the manuscript.

FUNDING

This work was supported by the Substance Use Disorders Grand Challenge Pilot Research Award, the Research Allocations Committee (RAC) Award, and the startup fund from the University of New Mexico. Molecular graphics and analyses performed with UCSF Chimera (Pettersen et al., 2004), developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from NIH P41-GM103311.

REFERENCES

Alfonso, S., Kessels, H. W., Banos, C. C., Chan, T. R., Lin, E. T., Kumaravel, G., et al. (2014). Synapto-depressive effects of amyloid beta require PICK1. *Eur. J. Neurosci.* 39, 1225–1233. doi: 10.1111/ejn.12499

Boudin, H., Doan, A., Xia, J., Shigemoto, R., Huganir, R. L., Worley, P., et al. (2000). Presynaptic clustering of MGluR7a requires the PICK1 PDZ domain binding site. *Neuron* 28, 483–497. doi: 10.1016/S0896-6273(00)00127-6

Christensen, N. R., De Luca, M., Lever, M. B., Richner, M., Hansen, A. B., Noes-Holt, G., et al. (2020). A high-affinity, bivalent PDZ domain inhibitor complexes PICK 1 to alleviate neuropathic pain. *EMBO Mol. Med.* 12:11248. doi: 10.15252/emmm.201911248

Cowan, C. A., Yokoyama, N., Bianchi, L. M., Henkemeyer, M., and Fritzsch, B. (2000). EphB2 guides axons at the midline and is necessary for normal vestibular function. *Neuron* 26, 417–430. doi: 10.1016/S0896-6273(00)81174-5

Dev, K. K., Nishimune, A., Henley, J. M., and Nakanishi, S. (1999). The protein kinase Cα binding protein PICK1 interacts with short but not long form alternative splice variants of AMPA receptor subunits. *Neuropharmacology* 34, 15415–15424. doi: 10.1016/S0028-3908(98)00230-5

Dixon, R. M., Mellor, J. R., and Hanley, J. G. (2009). PICK1-mediated glutamate receptor subunit 2 (GluR2) trafficking contributes to cell death in oxygen/glucose-deprivedhippocampal neurons. *J. Biol. Chem.* 284, 14230–14235. doi: 10.1074/jbc.M901203200

Duggan, A., García-Añoveros, J., and Corey, D. P. (2002). The PDZ domain protein PICK1 and the sodium channel BNaC1 interact and localize at mechanosensory terminals of dorsal root ganglion neurons and dendrites of central neurons. *J. Biol. Chem.* 277, 5203–5208. doi: 10.1074/jbc.M104748200

El Far, O., Airas, J., Wischmeyer, E., Nehring, R. B., Karschin, A., and Betz, H. (2000). Interaction of the C-terminal tail region of the metabotropic glutamate receptor 7 with the protein kinase C substrate PICK1. *Eur. J. Neurosci.* 12, 4215–4221. doi: 10.1046/j.1460-9568.2000.01309.x

Enz, R., and Croci, C. (2003). Different binding motifs in metabotropic glutamate receptor type 7b for filamin A, protein phosphatase 1C, protein interacting with protein kinase C (PICK) 1 and syntenin allow the formation of multimeric protein complexes. *Biochem. J.* 372, 183–191. doi: 10.1042/bj20021750

Erlendsson, S., and Madsen, K. L. (2015). Membrane binding and modulation of the PDZ domain of PICK1. *Membranes* 5, 597–615. doi: 10.3390/membranes5040597

Fiser, A., Do, R. K. G., and Šali, A. (2000). Modeling of loops in protein structures. *Protein Sci.* 9, 1753–1773. doi: 10.1110/ps.9.9.1753

Garry, E. M., Moss, A., Rosie, R., Delaney, A., Mitchell, R., and Fleetwood-Walker, S. M. (2003). Specific involvement in neuropathic pain of AMPA receptors and adapter proteins for the GluR2 subunit. *Mol. Cell. Neurosci.* 24, 10–22. doi: 10.1016/S1044-7431(03)00134-9

Gołaś, E., Maisuradze, G. G., Senet, P., Ołdziej, S., Czaplewski, C., Scheraga, H. A., et al. (2012). Simulation of the opening and closing of Hsp70 chaperones by coarse-grained molecular dynamics. *J. Chem. Theory Comput.* 8, 1750–1764. doi: 10.1021/ct200680g

Hanley, J. G. (2008). PICK1: a multi-talented modulator of AMPA receptor trafficking. *Pharmacol. Therapeutics* 118, 152–160. doi: 10.1016/j.pharmthera.2008.02.002

He, J., Xia, M., Yeung, P. K. K., Li, J., Li, Z., Chung, K. K., et al. (2018). PICK1 inhibits the E3 ubiquitin ligase activity of parkin and reduces its neuronal protective effect. *Proc. Natl. Acad. Sci. U.S.A.* 115, 7193–7201. doi: 10.1073/pnas.1716506115

He, Y., Liwo, A., Weinstein, H., and Scheraga, H. A. (2011). PDZ binding to the BAR domain of PICK1 is elucidated by coarse-grained molecular dynamics. *J. Mol. Biol.* 405, 298–314. doi: 10.1016/j.jmb.2010.10.051

He, Y., Maciejczyk, M., Ołdziej, S., Scheraga, H. A., and Liwo, A. (2013). Mean-Field Interactions between nucleic-acid-base dipoles can drive the formation of a double helix. *Phys. Rev. Lett.* 110:098101. doi: 10.1103/PhysRevLett.110.098101

He, Y., Mozolewska, M. A., Krupa, P., Sieradzan, A. K., Wirecki, T. K., Liwo, A., et al. (2019). Lessons from application of the UNRES force field to predictions of structures of CASP10 targets. *Proc. Natl. Acad. Sci. U.S.A.* 110, 14936–14941. doi: 10.1073/pnas.1313316110

He, Y. Y., Xiao, Y., Liwo, A., and Scheraga, H. A. (2009). Exploring the parameter space of the coarse-grained UNRES force field by random search: selecting a transferable medium-resolution force field. *J. Comput. Chem.* 30, 2127–2135. doi: 10.1002/jcc.21215

Herlo, R., Lund, V. K., Lycas, M. D., Jansen, A. M., Khelashvili, G., Andersen, R. C., et al. (2018). An amphipathic helix directs cellular membrane curvature sensing and function of the BAR domain protein PICK1. *Cell Rep.* 23, 2056–2069. doi: 10.1016/j.celrep.2018.04.074

Hirbec, H., Francis, J. C., Lauri, S. E., Braithwaite, S. P., Coussen, F., Mulle, C., et al. (2003). Rapid and differential regulation of AMPA and kainate receptors at hippocampal mossy fibre synapses by PICK1 and GRIP. *Neuron* 37, 625–638. doi: 10.1016/S0896-6273(02)01191-1

Hruska-Hageman, A. M., Wemmie, J. A., Price, M. P., and Welsh, M. J. (2002). Interaction of the synaptic protein PICK1 (Protein Interacting with

C Kinase 1) with the non-voltage gated sodium channels BNC1 (Brain Na+ Channel 1) and ASIC (Acid-Sensing Ion Channel). *Biochem. J.* 361, 443–450. doi: 10.1042/bj3610443

Hung, A. Y., and Sheng, M. (2002). PDZ domains: structural modules for protein complex assembly. *J. Biol. Chem.* 277, 5699–5700. doi: 10.1074/jbc.R100065200

Jaulin-Bastard, F., Saito, H., Le Bivic, A., Ollendorff, V., Marchetto, S., Birnbaum, D., et al. (2001). The ERBB2/HER2 receptor differentially interacts with ERBIN and PICK1 PSD-95/DLG/ZO-1 domain proteins. *J. Biol. Chem.* 276, 15256–15263. doi: 10.1074/jbc.M010032200

Jensen, K. L., Sørensen, G., Dencker, D., Owens, W. A., Rahbek-Clemmensen, T., Lever, M. B., et al. (2018). Pick1-deficient mice exhibit impaired response to cocaine and dysregulated dopamine homeostasis. *eNeuro* 5, 0422–0417. doi: 10.1523/ENEURO.0422-17.2018

Jin, W., Ge, W. P., Xu, J., Cao, M., Peng, L., Yung, W., et al. (2006). Lipid binding regulates synaptic targeting of PICK1, AMPA receptor trafficking, and synaptic plasticity. *J. Neurosci.* 29, 2380–2390. doi: 10.1523/JNEUROSCI.3503-05.2006

Karlsen, M. L., Thorsen, T. S., Johner, N., Ammendrup-Johnsen, I., Erlendsson, S., Tian, X., et al. (2015). Structure of dimeric and tetrameric complexes of the BAR domain protein PICK1 determined by small-angle X-Ray scattering. *Structure* 23, 1258–1270. doi: 10.1016/j.str.2015.04.020

Leonard, A. S., Yermolaieva, O., Hruska-Hageman, A., Askwith, C. C., Price, M. P., Wemmie, J. A., et al. (2003). CAMP-dependent protein kinase phosphorylation of the acid-sensing ion channel-1 regulates its binding to the protein interacting with C-Kinase-1. *Proc. Natl. Acad. Sci. U.S.A.* 100, 2029–2034. doi: 10.1073/pnas.252782799

Lin, S. H. S., Arai, A. C., Wang, Z., Nothacker, H. P., and Civelli, O. (2001b). The carboxyl terminus of the prolactin-releasing peptide receptor interacts with PDZ domain proteins involved in α-Amino-3-Hydroxy-5-Methylisoxazole-4-Propionic acid receptor clustering. *Mol. Pharmacol.* 60, 916–923. doi: 10.1124/mol.60.5.916

Lin, W. J., Chang, Y. F., Wang, W. L., and Huang, C. Y. F. (2001a). Mitogen-stimulated TIS21 protein interacts with a protein-kinase-Cα-binding protein RPICK1. *Biochem. J.* 354, 635–643.. doi: 10.1042/bj3540635

Liwo, A., He, Y., and Scheraga, H. A. (2011). Coarse-grained force field: general folding theory. *Phys. Chem. Chem. Phys.* 13, 16890–16901. doi: 10.1039/c1cp20752k

Liwo, A., Kamierkiewicz, R., Czaplewski, C., Groth, M., Oldziej, S., Wawak, R. J., et al. (1998). United-Residue force field for off-lattice protein-structure simulations: III. Origin of backbone hydrogen-bonding cooperativity in United-Residue Potentials. *J. Comput. Chem.* 19, 259–276. doi: 10.1002/(SICI)1096-987X(199802)19:3<259::AID-JCC1>3.0.CO;2-S

Liwo, A., Pincus, M. R., Wawak, R. J., Rackovsky, S., Ołdziej, S., and Scheraga, H. A. (1997b). A United-Residue force field for off-lattice protein-structure simulations. II. Parameterization of short-range interactions and determination of weights of energy terms by Z-score optimization. *J. Comput. Chem.* 18, 874–887. doi: 10.1002/(SICI)1096-987X(199705)18:7<874::AID-JCC2>3.0.CO;2-O

Liwo, A., Sieradzan, A. K., Lipska, A. G., Czaplewski, C., Joung, I., Zmudzińska, W., et al. (2019). A general method for the derivation of the functional forms of the effective energy terms in coarse-grained energy functions of polymers. III. Determination of scale-consistent backbone-local and correlation potentials in the UNRES force field and force-F. *J. Chem. Phys.* 150:155104. doi: 10.1063/1.5093015

Liwo, A. Odziej, S., Pincus, M. R., Wawak, R. J., Rackovsky, S., and Scheraga, H. A. (1997a). A United-Residue force field for off-lattice protein-structure simulations. I. Functional forms and parameters of long-range side-chain interaction potentials from protein crystal data. *J. Comput. Chem.* 18, 849–873. doi: 10.1002/(SICI)1096-987X(199705)18:7<849::AID-JCC1>3.0.CO;2-R

Lu, W., and Ziff, E. B. (2005). PICK1 interacts with ABP/GRIP to regulate AMPA receptor trafficking. *Neuron* 47, 407–421. doi: 10.1016/j.neuron.2005.07.006

Lubecka, E. A., Karczyńska, A. S., Lipska, A. G., Sieradzan, A. K., Zi?ba, K., Sikorska, C., et al. (2019). Evaluation of the scale-consistent UNRES force field in template-free prediction of protein structures in the CASP13 experiment. *J. Mol. Graph. Model.* 92, 54–166. doi: 10.1016/j.jmgm.2019.07.013

Madasu, Y., Yang, C., Boczkowska, M., Bethoney, K. A., Zwolak, A., Rebowski, G., et al. (2015). PICK1 is implicated in organelle motility in an Arp2/3 complex-independent manner. *Mol. Biol. Cell* 26, 1308–1322. doi: 10.1091/mbc.E14-10-1448

Madsen, K. L., Beuming, T., Niv, M. Y., Chang, C. W., Dev, K. K., Weinstein, H., et al. (2005). Molecular determinants for the complex binding specificity of the PDZ domain in PICK1. *J. Biol. Chem.* 280, 20539–20548. doi: 10.1074/jbc.M500577200

Martí-Renom, M. A., Stuart, A. C., Fiser, A., Sánchez, R., Melo, F., and Šali, A. (2000). Comparative protein structure modeling of genes and genomes. *Annual Rev. Biophys. Biomol. Structure* 29, 291–325. doi: 10.1146/annurev.biophys.29.1.291

Meyer, G., Varoqueaux, F., Neeb, A., Oschlies, M., and Brose, N. (2004). The complexity of PDZ domain-mediated interactions at glutamatergic synapses: a case study on neuroligin. *Neuropharmacology* 47, 724–733. doi: 10.1016/j.neuropharm.2004.06.023

Mozolewska, M. A., Krupa, P., Scheraga, H. A., and Liwo, A. (2015). Molecular modeling of the binding modes of the iron-sulfur protein to the Jac1 Co-Chaperone from *Saccharomyces cerevisiae* by all-atom and coarse-grained approaches. *Proteins* 83, 1414–1426. doi: 10.1002/prot.24824

Nakamura, K., Man, Z., Xie, Y., Hanai, A., Makyio, H., Kawasaki, M., et al. (2012). Structural basis for membrane binding specificity of the Bin/Amphiphysin/Rvs (BAR) domain of arfaptin-2 determined by Arl1 GTPase. *J. Biol. Chem.* 287, 25478–25489. doi: 10.1074/jbc.M112.365783

Pan, L., Wu, H., Shen, C., Shi, Y., Jin, W., Xia, J., et al. (2007). Clustering and synaptic targeting of PICK1 requires direct interaction between the PDZ domain and lipid membranes. *EMBO J.* 26, 4576–4587. doi: 10.1038/sj.emboj.7601860

Penzes, P., Johnson, R. C., Sattler, R., Zhang, X., Huganir, R. L., Kambampati, V., et al. (2001). The neuronal Rho-GEF kalirin-7 interacts with PDZ domain-containing proteins and regulates dendritic morphogenesis. *Neuron* 29, 229–242. doi: 10.1016/S0896-6273(01)00193-3

Perroy, J., El Far, O., Bertaso, F., Pin, J. P., Betz, H., Bockaert, J., et al. (2002). PICK1 is required for the control of synaptic transmission by the metabotropic glutamate receptor 7. *EMBO J.* 21, 2990–2999. doi: 10.1093/emboj/cdf313

Peter, B. J., Kent, H. M., Mills, I. G., Vallis, Y., Butler, P. J. G., Evans, P. R., et al. (2004). BAR domains as sensors of membrane curvature: the amphiphysin BAR structure. *Science* 303, 495–499. doi: 10.1126/science.1092586

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., et al. (2004). UCSF chimera - a visualization system for exploratory research and analysis. *J. Comput. Chem.* 25,1605–1612. doi: 10.1002/jcc.20084

Reymond, N., Garrido-Urbani, S., Borg, J. P., Dubreuil, P., and Lopez, M. (2005). PICK-1: a scaffold protein that interacts with nectins and JAMs at cell junctions. *FEBS Lett.* 579, 2243–2249. doi: 10.1016/j.febslet.2005.03.010

Rocca, D. L., Martin, S., Jenkins, E. L., and Hanley, J. G. (2008). Inhibition of Arp2/3-mediated actin polymerization by PICK1 regulates neuronal morphology and AMPA receptor endocytosis. *Nat. Cell Biol.* 10, 259–271. doi: 10.1038/ncb1688

Šali, A., and Blundell, T. L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* 234, 779–815. doi: 10.1006/jmbi.1993.1626

Salzer, U., Kostan, J., and Djinović-Carugo, K. (2017). Deciphering the BAR code of membrane modulators. *Cell. Mol. Life Sci.* 74, 2413–2438. doi: 10.1007/s00018-017-2478-0

Sheng, M., and Sala, C. (2001). PDZ Domains and the organization of supramolecular complexes. *Annu. Rev. Neurosci.* 24, 1–29. doi: 10.1146/annurev.neuro.24.1.1

Sieradzan, A. K., Giełdoń, A., Yin, Y., He, Y.; Scheraga, H. A., and Liwo, A. (2018). A new protein nucleic-acid coarse-grained force field based on the UNRES and NARES-2P force fields. *J. Comput. Chem.* 39, 2360–2370. doi: 10.1002/jc c.25571

Sieradzan, A. K., Makowski, M., Augustynowicz, A., and Liwo, A. (2017). A general method for the derivation of the functional forms of the effective energy terms in coarse-grained energy functions of polymers. I. Backbone potentials of coarse-grained polypeptide chains. *J. Chem. Phys.* 146:124106. doi: 10.1063/1.4978680

Sieradzan, A. K., Niadzvedtski, A., Scheraga, H. A., and Liwo, A. (2014). Revised backbone-virtual-bond-angle potentials to treat the l- and d-amino acid residues in the coarse-grained united residue (UNRES) force field. *J. Chem. Theory Comput.* 10, 2194–2203. doi: 10.1021/ct500119r

Staudinger, J., Lu, J., and Olson, E. N. (1997). Specific interaction of the PDZ domain protein PICK1 with the COOH terminus of protein kinase C-α. *J. Biol. Chem.* 272, 32019–32024. doi: 10.1074/jbc.272.51.32019

Staudinger, J., Zhou, J., Burgess, R., Elledge, S. J., and Olson, E. N. (1995). PICK1: A Perinuclear binding protein and substrate for protein kinase C isolated by the yeast two-hybrid system. *J. Cell Biol.* 128, 263–271. doi: 10.1083/jcb.128.3.263

Stevens, A. O., and He, Y. (2020). The electrostatic allostery could be the trigger for the changes in dynamics for the PDZ domain of PICK1. *bioRxiv [preprint]*. doi: 10.1101/2020.10.06.328617

Takei, K., Slepnev, V. I., Haucke, V., and De Camilli, P. (1999). Functional partnership between amphiphysin and dynamin in clathrin-mediated endocytosis. *Nat. Cell Biol.* 1, 33–39. doi: 10.1038/9004

Takeya, R., Takeshige, K., and Sumimoto, H. (2000). Interaction of the PDZ domain of human PICK1 with class I ADP-ribosylation factors. *Biochem. Biophys. Res. Commun.* 267, 149–155. doi: 10.1006/bbrc.1999.1932

Torres, G. E., Yao, W. D., Mohn, A. R., Quan, H., Kim, K. M., Levey, A. I., Staudinger, J., Caron, M. G. (2001). Functional interaction between monoamine plasma membrane transporters and the synaptic PDZ domain-containing protein PICK1. *Neuron* 30, 121–134. doi: 10.1016/S0896-6273(01)00267-7

Torres, R., Firestein, B. L., Dong, H., Staudinger, J., Olson, E. N., Huganir, R. L., et al. (1998). PDZ proteins bind, cluster, and synaptically colocalize with Eph receptors and their ephrin ligands. *Neuron* 21, 1453–1463. doi: 10.1016/S0896-6273(00)80663-7

Webb, B., and Sali, A. (2016). Comparative protein structure modeling using MODELLER. *Curr. Protoc. Bioinforma.* 54, 5.6.1-5.5.37. doi: 10.1002/cpbi.3

Williams, M. E., Wu, S. C. Y., McKenna, W. L., and Hinck, L. (2003). Surface expression of the netrin receptor UNC5H1 is regulated through a protein kinase C-interacting protein/protein kinase-dependent mechanism. *J. Neurosci.* 23, 11279–11288. doi: 10.1523/JNEUROSCI.23-36-11279.2003

Xu, J., and Xia, J. (2007). Structure and function of PICK1. *NeuroSignals* 15, 190–201. doi: 10.1159/000098482

Zięba, K., Slusarz, M., Slusarz, R., Liwo, A., Czaplewski, C., and Sieradzan, A. K. (2019). Extension of the UNRES coarse-grained force field to membrane proteins in the lipid bilayer. *J. Phys. Chem. B.* 123, 7820–7839. doi: 10.1021/acs.jpcb.9b06700

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Molecular Biosciences

ORIGINAL RESEARCH
published: 22 February 2021
doi: 10.3389/fmolb.2021.619381

Optimizing Gō-MARTINI Coarse-Grained Model for F-BAR Protein on Lipid Membrane

Md. Iqbal Mahmood[1], *Adolfo B. Poma*[2] *and Kei-ichi Okazaki*[1]*

[1]Department of Theoretical and Computational Molecular Science, Institute for Molecular Science, National Institutes of Natural Sciences, Okazaki, Japan, [2]Institute of Fundamental Technological Research, Polish Academy of Sciences, Warsaw, Poland

Coarse-grained (CG) molecular dynamics (MD) simulations allow us to access much larger length and time scales than atomistic MD simulations, providing an attractive alternative to the conventional simulations. Based on the well-known MARTINI CG force field, the recently developed Gō-MARTINI model for proteins describes large-amplitude structural dynamics, which has not been possible with the commonly used elastic network model. Using the Gō-MARTINI model, we conduct MD simulations of the F-BAR Pacsin1 protein on lipid membrane. We observe that structural changes of the non-globular protein are largely dependent on the definition of the native contacts in the Gō model. To address this issue, we introduced a simple cutoff scheme and tuned the cutoff distance of the native contacts and the interaction strength of the Lennard-Jones potentials in the Gō-MARTINI model. With the optimized Gō-MARTINI model, we show that it reproduces structural fluctuations of the Pacsin1 dimer from atomistic simulations. We also show that two Pacsin1 dimers properly assemble through lateral interaction on the lipid membrane. Our work presents a first step towards describing membrane remodeling processes in the Gō-MARTINI CG framework by simulating a crucial step of protein assembly on the membrane.

Keywords: molecular dynamics simulation, MARTINI force field, Gō model, membrane remodeling, Pacsin

OPEN ACCESS

Edited by:
Wenfei Li,
Nanjing University, China

Reviewed by:
Xubo Lin,
Beihang University, China
Xinqiu Yao,
Georgia State University, United States
Francesco Pesce,
University of Copenhagen, Denmark

***Correspondence:**
Kei-ichi Okazaki
keokazaki@ims.ac.jp

Specialty section:
This article was submitted to Biological Modeling and Simulation, a section of the journal Frontiers in Molecular Biosciences

Received: *20 October 2020*
Accepted: *14 January 2021*
Published: *22 February 2021*

Citation:
Mahmood MI, Poma AB and Okazaki K (2021) Optimizing Gō-MARTINI Coarse-Grained Model for F-BAR Protein on Lipid Membrane. Front. Mol. Biosci. 8:619381. doi: 10.3389/fmolb.2021.619381

INTRODUCTION

Large-scale shape changes of membrane structures in the cell are important in many biological processes such as endocytosis, exocytosis and vesicle trafficking (McMahon and Gallop, 2005). These membrane remodeling processes emerge from the interplay between lipids and proteins (McMahon and Gallop, 2005; Suetsugu et al., 2014; Bassereau et al., 2018). Because of dynamic nature of these processes, molecular dynamics needs to be clarified to understand their mechanisms. The molecular dynamics (MD) simulation is a powerful tool to study the dynamic processes at molecular level (Marrink et al., 2019). However, the conventional all-atom (AA) MD has limitations in size and time scales. It is too costly to simulate a large system of membrane remodeling that contains large lipid membrane, large number of proteins and solvent molecules with a time scale longer than microseconds by AA MD. Thus, the coarse-grained (CG) model that represents a group of atoms by a single bead, offers a good alternative to study large membrane remodeling processes (Marrink et al., 2019).

Various CG models of lipids and proteins have been developed previously (Tozzini, 2005; Ayton et al., 2007; Klein and Shinoda, 2008; Takada, 2012; Marrink et al., 2019). For lipids, there are reasonably accurate and transferable CG models such as MARTINI and SPICA (Marrink and

Tieleman, 2013; Marrink et al., 2019; Seo and Shinoda, 2019). For proteins, there are structure-based models such as elastic network (EN) and Gō models (Tozzini, 2005; Takada et al., 2015). However, relatively less effort has been made on CG models of the combined protein-membrane system, which should be important for describing the membrane remodeling processes. For example, the popular MARTINI model introduces the EN model to proteins (denoted as EN-MARTINI) (Periole et al., 2009), which assumes unbreakable harmonic bonds, and thus, is unable to describe large-scale motions such as protein unfolding or conformational changes between two stable conformations. These large-scale motions should be important to describe realistic dynamics of the membrane remodeling. The recently developed Gō-MARTINI addressed this issue by replacing the harmonic potential with the Lennard-Jones (LJ) potential based on the contact map of the native protein structure (Poma et al., 2017). The Gō-MARTINI model combines the flexibility of the Cα-based Gō-like model for the sampling of large conformational changes in proteins (Okazaki et al., 2006, 2012; Okazaki and Takada, 2008; Poma et al., 2018, 2019; Senapati et al., 2019) and the versatility of the MARTINI force field that allows the description of different biomolecules, (e.g. lipids, polysaccharides, polymers and nucleic acids) at almost atomistic resolution (Marrink and Tieleman, 2013; Uusitalo et al., 2015; Souza et al., 2020). At the moment, some studies including the original developmental work have used Gō-MARTINI for protein-only systems (Poma et al., 2017; Souza et al., 2019), and not much has been done for protein-membrane systems. Only a few studies have used Gō-MARTINI for protein-membrane systems (Thallmair et al., 2019).

In this study, we apply the Gō-MARTINI model to the F-Bin/Amphiphysin/Rvs (F-BAR) protein Pacsin1 as a model protein that is involved in the membrane remodeling. Pacsin proteins are involved in clathrin-mediated endocytosis, actin polymerization and neuronal development. In the previous study, we showed that Pacsin1 induces and senses the membrane curvature in the EN-MARTINI framework (Mahmood et al., 2019). However, it was found that structural fluctuations of Pacsin1 in the EN-MARTINI model are underestimated, which can affect the stability of the protein complex (Baaden and Marrink, 2013; Stark et al., 2013). Since the association and dissociation of protein complexes play a crucial role in membrane remodeling processes, the underestimated fluctuations can lead to an incorrect description of the processes. Here, in order to overcome the limitations of the EN model, we introduced a simple cutoff scheme of the Gō-MARTINI and tuned the parameters to reproduce structural fluctuations of Pacsin1 on the lipid membrane observed in the AA simulations. We further show that Pacsin1 properly assembles on the membrane with the optimized parameters. This study is a first step toward describing realistic dynamics of the membrane remodeling in the Gō-MARTINI framework.

MATERIALS AND METHODS

All-atom MD Simulations

For our study, we have chosen the human Pacsin1 F-BAR domain crystal structure with the PDB ID 3HAH (Wang et al., 2009). The structure consists of two monomers with some missing residues. MODELLER (Martí-Renom et al., 2000; Webb and Sali, 2016) was employed for modeling the Pacsin1 dimer missing residues (first monomer: T172-L191, second monomer: T172-K194) without referring to a homologous structure. The missing residues at the N- and C-terminal parts were not considered in the simulations. The N-terminal part consists of 15 residues with four negatively charged amino acids. Although the role of the N-terminal part remains unclear, it is unlikely that this highly negatively charged region is involved in interaction with the negatively charged lipid head groups of the membrane. The C-terminal part consists of the central linker and SH3 domain, which have been experimentally shown to decrease the membrane transformation activity (Wang et al., 2009). First, the coordinates of mixed lipid bilayer (POPC 20%, POPE 20%, POPS 60%) (Wang et al., 2009) were generated by the membrane builder tool of CHARMM-GUI (Sunhwan et al., 2008; Wu et al., 2014). Then, Pacsin1 structure was placed on the lipid bilayer using VMD (Humphrey et al., 1996). TIP3P water molecules and neutralizing ions of 0.15 M Na^+ and Cl^- were added to the system, making a periodic boundary box (x:23 nm, y: 23 nm and z:18 nm) with the total number of atoms 909109. The CHARMM36 force field was used for lipid bilayers and protein (Venable et al., 2010). The simulation procedure was the same as that of our previous work (Mahmood et al., 2019). The 500 ns production runs were conducted at a temperature of 310 K and a pressure of 1 atm.

Conventional EN-MARTINI Simulations

The MARTINI coarse-grained (CG) molecular dynamics (MD) simulations described in this paper were performed with the GROMACS-2018 simulation package (Abraham et al., 2015) (www.gromacs.org). The CG model of the Pacsin1-membrane system was constructed using the MARTINI force field version 2.2 (Marrink et al., 2007; Monticelli et al., 2008; De Jong et al., 2013; Marrink and Tieleman, 2013) with additional EN potential for the protein. The EN model was used to maintain the secondary and tertiary structures of proteins based on definition by the DSSP algorithm (version 2.2.1) (Kabsch and Sander, 1983). The spring constant of 500 kJ $mol^{-1}nm^{-2}$, the lower and upper elastic bond cut-off to 0.5 and 1.2 nm, respectively (Periole et al., 2009; Mahmood et al., 2019) were applied to the Pacsin1 crystal structure (PDB ID 3HAH) (Wang et al., 2009). The numbers of the elastic bonds from this definition were 2688 for chain A and 2646 for chain B. The numbers are different between the two chains, reflecting a slight difference in their structures. A possible approach to improve the definition of the elastic bonds, as well as the Gō native contacts, is mentioned in DISCUSSION. The protein CG structure and topology were generated using the script "martinize.py" (De Jong et al., 2013). Then, we used a script "insane.py" (Wassenaar et al., 2015) for constructing the flat lipid membrane, aligning proteins on the membrane, generating water and ions. The lipid membrane consists of mixed lipids POPC, POPE and POPS (20%:20%:60%). The systems were hydrated using CG water beads and made charge neutral by addition of an appropriate number of ions with 0.15 M Na^+ and Cl^-. The total number of beads in the system was about 292743 CG beads and

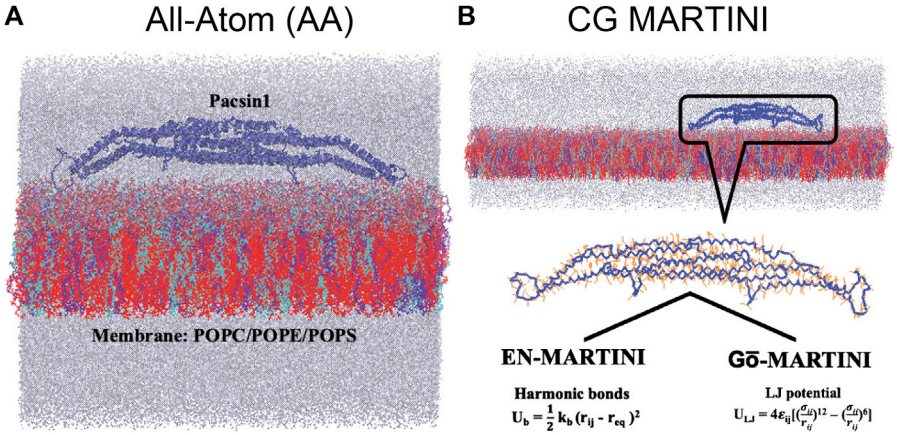


FIGURE 1 | All-atom (AA), coarse-grained (CG) EN-MARTINI and Gō-MARTINI molecular dynamics systems of the protein and membrane are shown **(A)** and **(B)** respectively. Blue color carton represents Pacsin1 protein. Three kind of lipids composition (cyan: 20% POPC, blue: 20% POPE and red: 60% POPS). Water molecules and ions are represented in gray color.

the system box size was x:60 nm, y:30 nm and z:20 nm. Energy minimization of the system was performed with 5,000 steps of the steepest descent method. After minimization, the system was equilibrated for 0.5 ns in the NPT ensemble using the Berendsen pressure coupling (Berendsen et al., 1984). The following production simulations were run at 300 K with separate temperature coupling for the solvent, lipids and protein using the stochastic rescaling scheme (Bussi et al., 2007) (τ = 1 ps) and the Parrinello-Rahman (Parrinello and Rahman, 1981) semiisotropic pressure coupling at 1 bar. A time step of dt = 20 fs was used. The reaction field electrostatics and LJ potentials were shifted to zero at the cut-off distance of 1.2 nm.

Gō-MARTINI Simulations

In the Gō-MARTINI simulations, we have replaced the harmonic bonds of the commonly used EN model with the LJ potential based on the contact map of the native protein structure as in Gō models (Poma et al., 2017). There are several types of contact maps with different definitions of native contacts (Clementi et al., 2000; Koga and Takada, 2001; Sułkowska and Cieplak, 2008; Noel et al., 2012). The original Gō-MARTINI adopts the atomic overlap criterion (OV) and chemistry-based rCSU for definition of the native contacts (Sułkowska and Cieplak, 2008; Wołek et al., 2015; Poma et al., 2017). With this definition, the Pacsin1 conformation became distorted during the simulations with respect to the conformations observed in the all-atom simulations (**Supplementary Figure S1**). Although it worked for globular proteins (Poma et al., 2017), the OV + rCSU definition of the native contacts might result in an unnatural conformation for extended structures like Pacsin1 (**Figure 1A**) in a balance between the native contacts and the non-native interactions of the MARTINI force field. Thus, we adopt a simpler cutoff scheme for the native contact definition as described in the following. First, all i th and $(i + 3)$ th amino-acid pairs in the sequence are considered as the native contact, providing a similar interaction as the dihedral term in the typical Gō models. Then, for i th and $j>i + 3$ th amino acid pairs, if the residue-residue minimum distance considering all non-hydrogen atoms is below a cutoff distance, the pair is considered as the native contact. The cutoff distance of 4.5 Å, 5.0 Å and 5.5 Å were tested. As shown in **Supplementary Table S1**, the number of the native contacts significantly increased with the new definition, while keeping almost all contacts from the OV + rCSU definition. The number of the native contacts, however, are less than the number of the elastic bonds used in the EN-MARTINI (see the previous section). The numbers of the native contacts are different between the two chains, reflecting a slight difference in their crystal structures. For the cutoff distance of 5.0 Å, the number of the common contacts shared between the two chains is 793, which is 93% and 94% of the total contacts of the chain A and chain B, respectively. The rest of the contacts is unique to each chain. Note that the native contacts were defined for each chain of Pacsin1 and no native contact was defined between the two chains. Thus, it would be interesting to see if the interface structure is maintained only with the MARTINI force field. To check the interface structure between the two chains, we calculated the fraction of the "virtual" (that is, not considered in the model potential) native contacts at the interface present during the Gō-MARTINI simulations (Q_{AB}) (Poma et al., 2017). The virtual native contacts were defined in the same way as the intra-chain contacts with the cutoff distance 5.0 Å. The native contact between residues i, j is considered to be present when its distance satisfies $r_{ij} < 1.5\sigma_{ij} \approx 1.34r_{ij}^0$ (see below for definitions of σ_{ij} and r_{ij}^0). The backbone beads (BB), that is, Cα positions, were used for the interaction sites. In the LJ potential, the parameter ε_{ij} controls the strength of the native contact interaction in unit of ε, where ε = 6.276 kJ mol^{-1}. This value corresponds to the typical energy scale of hydrogen bonds in proteins (Poma et al., 2015) and λ in the native contact energy, $\varepsilon_{ij} = \lambda\,\varepsilon$, is a tunable parameter. In this study, λ = 1.0, 1.5 were tested. The LJ potential of the native contacts is defined as,

$$U_{LJ} = \sum_{i,\,j\ \in \text{Native contacts}} 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right], \qquad (1)$$

where $\sigma_{ij} = r_{ij}^0/2^{1/6}$ with r_{ij}^0 being the Cα-Cα distance of the native-contact pair in the native structure. The bond angle term is another factor in Gō models, biasing towards the native structure.

The bond angle force constants for helices and the other secondary structures were set to K_{BBB} = 700 kJ mol^{-1} and K_{BBB} = 20 kJ mol^{-1}, respectively (Monticelli et al., 2008). For proline residues in the helix kink region (Pro 145 and Pro221), the force constants were set to K_{BBB} = 20 kJ mol^{-1}. Our modified version of "go_martinize.py" script was used to generate the protein coarse-grained structure and topology. The script is available on GitHub (https://github.com/OkazakiLab/Go-MARTINI). The following system setup and simulations were done in the same way as the conventional EN-MARTINI described in the previous section.

Principal Component Analysis

The principal component analysis (PCA) was performed to identify large-amplitude conformational changes of Pacsin1 from simulation trajectories. We only considered Cα positions of the AA or backbone-bead (BB) positions of the MARTINI simulations, after Pacsin1 structure was superimposed in the trajectories. Then, a covariance matrix was calculated and diagonalized to obtain eigenvalues and eigenvectors in the order of their contributions to the conformational changes. First, PCA was performed for each simulation: AA, EN-MARTINI, and Gō-MARTINI. In order to compare the PCA results, we calculated the root mean square inner product (RMSIP) (Amadei et al., 1999),

$$\mathrm{RMSIP} = \sqrt{\frac{1}{10}\sum_{i=1}^{10}\sum_{j=1}^{10}\left(u_i \cdot v_j\right)^2}, \qquad (2)$$

where $\boldsymbol{u}_i$ and $\boldsymbol{v}_j$ represent eigenvectors obtained by two different PCAs, and the first 10 eigenvectors were considered. The RMSIP quantifies how much two simulation trajectories are overlapped in a subspace described by the first 10 eigenvectors. Another way to quantify an overlap among multiple trajectories is to perform a single PCA using all trajectories and project them onto common principal components (Martín-García et al., 2015). We performed the PCA using all three simulations, after superimposing Pacsin1 structure in all three trajectories.

RESULTS

Structural Flexibility of Pacsin1 With EN-MARTINI and Gō-MARTINI

The structural flexibility of a single Pacsin1 dimer on the lipid membrane was investigated through the AA and CG MD simulations (**Figure 1**). We carried out ~500 ns AA simulation and ~1000 ns CG simulations with the conventional EN-MARTINI and Gō-MARTINI (**Figures 1A**,B). First, we calculated the root mean squared fluctuation (RMSF) of the Pacsin1 dimer and compared it between the AA and CG simulations (see **Figure 2**). In this analysis, only the coordinates for backbone atoms of Pacsin1 were used. The RMSF represents the extent of amino acid residue fluctuation around their average positions. A comparison among simulations suggested that the fluctuation in the tip-loop region from the EN-MARTINI simulation is significantly underestimated compared to the AA simulation result (**Figure 2**). The underestimation of the RMSF in the tip-loop region is due to a limitation of the elastic network potential. To address this issue, we employed the Gō-MARTINI model, which can describe large-scale unfolding motions. We introduced a simple cutoff scheme to define the native contacts in the Gō model (see Methods). After exploring the Gō-MARTINI parameters, we found that the RMSF from the Gō-MARTINI simulation with the native contact cutoff 5.0 Å and interaction strength of the LJ potential λ = 1.0 is well fitted with the AA simulation result, including the tip-loop region residues (**Figure 2**). The RMSFs from the Gō-MARTINI simulations with the native contact cutoff values 4.5 and 5.5 Å or λ = 1.5 are slightly suppressed (see **Supplementary Figures S2,3**). In addition, the principal component analysis, Pacsin1 binding and assembly on the membrane support that the native contact cutoff 5.0 Å and λ = 1.0 is a best set of parameters, as we see below. These results indicate that choice of the force field parameters influence structural dynamics of the protein.

Second, we performed the principal component analysis (PCA). The PCA identifies the axes of maximal variance of global structural fluctuations. The PCA was performed for trajectories from the AA and CG MARTINI MD simulations. In our analysis, we consider only Cα atoms of the protein. **Figure 3** shows a visualization of the structural fluctuations from the first principal component mode and the eigenvalues along the principal component modes. The AA MD simulation shows that the tip-loop regions of the protein have high magnitude of fluctuations, which can be seen in the PC1 eigenvector. The PCA result from the Gō-MARTINI simulation with cutoff 5.0 Å, λ = 1.0 is in good agreement with the AA result, regarding not only the PC1 vector but also the eigenvalue profile along the PC modes. In contrast, the EN-MARTINI result shows an underestimated fluctuation, which is evident from the PC1 eigenvector and the eigenvalue profile. In order to compare the PCA results of the Gō-MARTINI and EN-MARTINI simulations to the reference AA result, we calculated RMSIP (see MATERIALS and METHODS) between the Gō-MARTINI and AA results, as well as between the EN-MARTINI and AA results. The RMSIP quantifies how much two simulation trajectories are overlapped in a subspace described by the first 10 eigenvectors. It was found that the RMSIP (Gō-MARTINI, AA) of 0.691 is higher than the RMSIP (EN-MARTINI, AA) of 0.652, indicating that the overlap between Gō-MARTINI and AA is better than that of EN-MARTINI and AA. We note that the time scale of 500 ns for the AA simulation might not be enough to fully cover slow conformational dynamics of the tip loops that contribute significantly to the global conformational changes. We also performed a single PCA using all three trajectories: AA, EN-MARTINI, and Gō-MARTINI (cutoff 5.0 Å, λ = 1.0), and projected each trajectory onto the common PC1 and PC2 (**Figure 3D**). The plot shows that conformations sampled in the AA and Gō-MARTINI overlap at the edges to some extent, while the EN-MARTINI samples an isolated, restricted region. The common PC1 and PC2 involve motions of the flexible tip loops (**Supplementary Figure S4**), which are expected to be

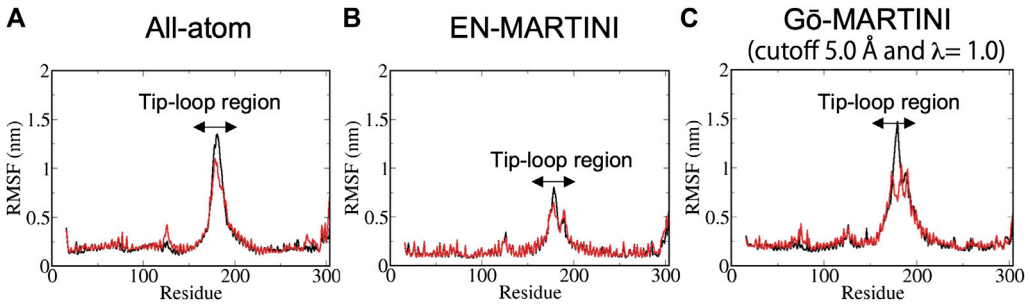


FIGURE 2 | The RMSF results from **(A)** AA **(B)** EN-MARTINI **(C)** Gō-MARTINI (cutoff 5.0 Å and λ = 1.0) simulations are shown. For the RMSF calculation, the last half of the trajectories was used. Black and red lines represent chain A and B, respectively. Arrows indicates the tip-loop region in chain A and B.

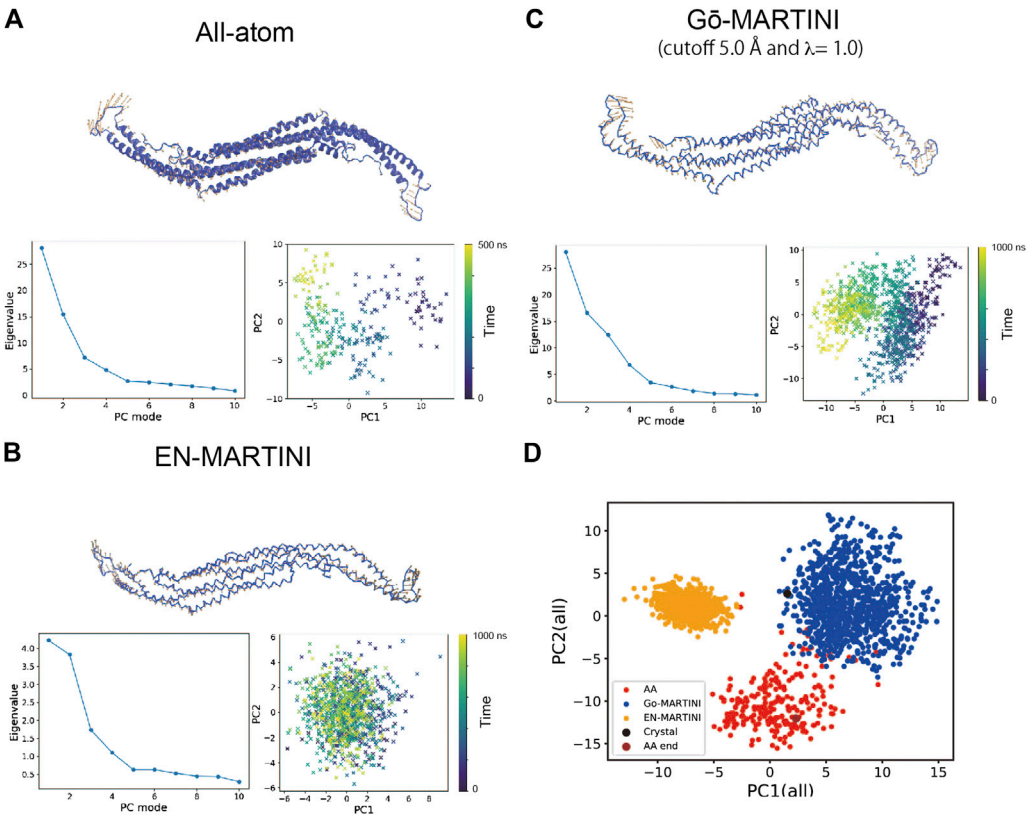


FIGURE 3 | Principal component analysis (PCA) of the Pacsin1 structural fluctuations. For **(A)** All-atom **(B)** EN-MARTINI **(C)** Gō-MARTINI (cutoff 5.0 Å and λ = 1.0) results, the first principal component (PC1) eigenvector on the Pacsin1 structure, the eigenvalue profile along the principal component modes, and mapping of the Pacsin1 conformations on the PC1-PC2 surface are shown. Colorbars in the PC1-PC2 mapping represent the time progress in nanoseconds. In **(D)**, mapping of the Pacsin1 conformations on the PC1-PC2 surface obtained from PCA using all three simulations is shown.

slow and might not be fully covered by the 500 ns AA simulation. As the native contact cutoff of the Gō-MARTINI increases, magnitude of fluctuations decreases as seen from smaller eigenvalues (**Supplementary Figure S5**). Thus, the Gō-MARTINI with the native contact cutoff 5.0 Å and λ = 1.0 reproduces both local and global structural fluctuations of Pacsin1.

We also analyzed the interface structure between two chains of Pacsin1 during the simulations, because the native contacts (elastic bonds) were not considered for the interface in the current Gō-MARTINI (EN-MARTINI) simulations. We calculated the fraction of the virtual native contacts at the interface present during the simulations (Q_{AB}, see **Materials and Methods**). The time courses of Q_{AB} for the Gō-MARTINI

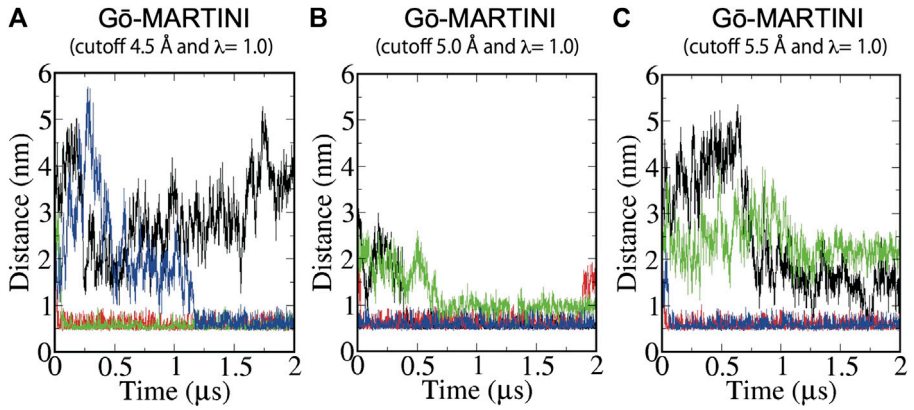


FIGURE 4 | Binding of Pacsin1 to the membrane with the wedge loops inserted into the membrane. Distance between positively charged K123 of the wedge loop and negatively charged phosphate of the lipid head group during MD simulations of **(A)** Gō-MARTINI (cutoff 4.5 Å and λ = 1.0), **(B)** Gō-MARTINI (cutoff 5.0 Å and λ = 1.0), and **(C)** Gō-MARTINI (cutoff 5.5 Å and λ = 1.0) are shown. Black, red, green and blue lines represent the wedge loop 1, 2 of the first Pacsin1 and that of the second Pacsin1, respectively.

and EN-MARTINI simulations as well as the AA simulation are shown in **Supplementary Figure S6**. For the Gō-MARTINI models, the average value of Q_{AB} from the last half of the trajectory is 0.83 or higher, which indicates that the interface structure is basically maintained only with the MARTINI force field at the interface. For the EN-MARTINI model, we observed a similar average value of 0.86. These values are lower than the average value of 0.94 observed in the AA simulation. Note that the time scale of the AA simulation is much shorter than the Gō-MARTINI or EN-MARTINI simulations. This is ~ 8 times shorter if we consider that MARTINI dynamics is faster than AA dynamics with the speed-up factor of ~ 4 (Marrink et al., 2004). Using this factor and comparing all simulations in the same time scale, we can report a higher Q_{AB} value above 0.9 during the first 200 ns of the Gō-MARTINI (cutoff 5.0 Å and λ = 1.0) simulation, which would match the AA result.

Pacsin1 Binding on the Lipid Membrane

The structure of the F-BAR domain of Pacsin1 revealed distinctive wedge loops that are involved in the membrane binding and insertion (Wang et al., 2009). The wedge loop is a signature of Pacsin proteins and possibly affects their assembly (Bai et al., 2012). In our analysis, we found that positively charged residue Lys (K123) of the wedge loop interacts with negatively charged phosphate of the lipid head group during MD simulations. Thus, we calculated a minimum distance between K123 and the lipid phosphate as a measure of Pacsin1 binding. For the Gō-MARTINI (cutoff 4.5 Å, λ = 1.0), after a few nanoseconds, two wedge loops from different Pacsin1 dimers are inserted in the membrane (distance ~0.5 nm) throughout the simulations. One of the remaining wedge loops is inserted in the membrane after 1.2 μs. The last one is not inserted in the membrane during the simulations. For the optimized Gō-MARTINI (cutoff 5.0 Å, λ = 1.0), we observed a clear interaction between the wedge loop and the membrane (**Figure 4B**). That is, the distance between the wedge loop (residue K123) and the membrane stayed close for all wedge-loops. In contrast, for the other Gō-MARTINI (cutoff 5.5 Å, λ = 1.0), two wedge loops from different Pacsin1 dimers are inserted in the membrane from the early stage of MD simulations (**Figure 4C**). But other two wedge loops are not inserted into the membrane and the distances stay larger than 1 nm.

Pacsin1 Assembly Process on the Lipid Membrane

Assembly of Pacsin1 on the lipid membrane is one of the key features involved in the membrane remodeling. We carried out the Gō-MARTINI simulations with two Pacsin1 dimers on a flat tensionless membrane. During the 2 μs long simulation, stable Pacsin1-Pacsin1 lateral interaction was observed for the optimized Gō-MARTINI (cutoff 5.0 Å, λ = 1.0), while improper interactions were observed for the other Gō-MARTINIs (**Figure 5**). The lateral interaction observed in the optimized Gō-MARTINI was formed within a few nanoseconds and maintained throughout the simulation (**Figure 5B**). The similar lateral interaction was observed in the crystal structure of Pacsin1 (PDB entry, 3HAI) (Wang et al., 2009). Our previous study also revealed the similar lateral interaction of Pacsin1-Pacsin1 with the EN-MARTINI simulations (Mahmood et al., 2019). The inter-protein interaction is due to the physico-chemical interactions of the MARTINI force field. Thus, our results confirm that protein-protein interactions are well described by MARTINI (Baaden and Marrink, 2013). More importantly, it was also demonstrated that the observed inter-protein interactions, with the same MARTINI force field describing them, are strictly dependent on the definition of the intra-protein potentials.

DISCUSSION

In this study, we have adapted the Gō-MARTINI model to describe structural dynamics and assembly of the F-BAR protein Pacsin1. We

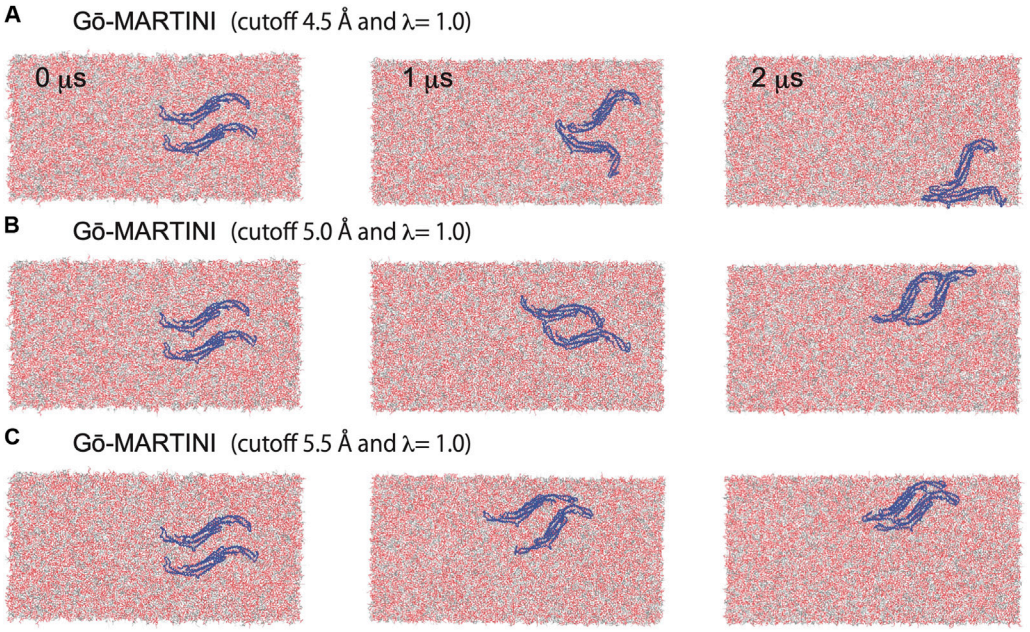


FIGURE 5 | Pacsin1 assemblies on the lipid membrane simulated with the Gō-MARTINI (cutoff 4.5 Å and λ = 1.0), (cutoff 5.0 Å and λ = 1.0) and (cutoff 5.5 Å and λ = 1.0) are shown in **(A)**, **(B)** and **(C)**, respectively.

introduced a simple cutoff scheme for definition of the native contacts instead of the OV + rCSU approach used in the original Gō-MARTINI (Poma et al., 2017). The cutoff scheme is more flexible and allows us to explore parameters such as the cutoff distance of the native contacts. The optimized Gō-MARTINI simulations reproduce global and local structural fluctuations from the AA simulation. The transferability of the current scheme including the cutoff distance of the native contacts should be tested with other systems to build a universal model. It was also shown that Pacsin1 binding and assembly on the membrane were reproduced properly by the optimized Gō-MARTINI. These results suggest that protein-lipid and protein-protein interactions are well described by the physico-chemical MARTINI force field, once proper intra-protein structures are prepared. The earlier success of the EN-MARTINI model for protein-protein interactions supports this notion (Baaden and Marrink, 2013). However, large conformational changes of intra-protein structures are beyond the scope of the EN-MARTINI model. Our results show that EN-MARTINI can be replaced by Gō-MARTINI, and the Gō-MARTINI model performs better than the EN-MARTINI model in terms of intra-protein structural fluctuations. The Gō-MARTINI model also offers advantages over the commonly used bond-angle restrained MARTINI, which maintains the local secondary structures. The bond-angle restrained MARTINI has been used for rather small or flexible proteins, such as helical peptides (Monticelli et al., 2008) or α-synuclein (Braun et al., 2012). However, this model is not applicable to proteins that have specific native structures more complicated than a single helix. Thus, the Gō-MARTINI model has advantages in simulating conformational dynamics of proteins with the specific native structures.

We note that the protein model in Gō-MARTINI is not a pure "Gō model", because it has non-native (that is, physico-chemical) interactions from the MARTINI force field. Previous works on protein-protein interactions showed that the MARTINI force field tends to overestimate protein-protein interactions, and thus, down-scaling of the interactions is necessary to reproduce experimental results (Stark et al., 2013; Javanainen et al., 2017; Benayad et al., 2020). By optimizing both the structure-based Gō interactions (Li et al., 2011, 2012) and physico-chemical MARTINI interactions (Alessandri et al., 2019), we have a unique opportunity to properly describe intra and inter protein structural stability and dynamics with the Gō-MARTINI model. Possible improvements of the structure-based Gō interactions include refinement of contact energy in a residue-pair specific manner. The previously developed methods such as atomic-interaction-based coarse-grained (AICG) model (Li et al., 2011) or Miyazawa-Jernigan statistical contact energy (Karanicolas and Brooks, 2002) can be used. The definition of the native contacts itself can be improved by symmetrizing between homodimers or analyzing the contacts in the all-atom simulations instead of the static experimental structure. The dynamic contact analysis of the all-atom simulations discerns between stable and transient contacts (Moreira et al., 2020), where the latter can be excluded from the native contacts. Another improvement would be an extension of the current single-basin Gō model to a multiple-basin Gō model to explore conformational changes between different stable conformations such as ligand-free and bound conformations. The previous methods such as the multiple-basin energy landscape model (Okazaki et al., 2006) or the

double-well ultra-coarse-grained model (Zhang et al., 2020) can be introduced.

To describe realistic dynamics of membrane remodeling in the Gō-MARTINI framework, we might need a further reduction of the dimension of the model. The simulation system can become very large when whole membrane remodeling processes are considered with large membranes and many proteins involved. One possible way to reduce the dimension in Gō-MARTINI is to replace normal MARTINI with Dry MARTINI, an implicit solvent version of MARTINI (Arnarez et al., 2015). This is a highly effective approach because solvent beads dominate the total number of beads as the simulation system becomes large. Although there are some modifications of the force field parameters in Dry MARTINI, our Gō-MARTINI framework is expected to apply with possible minor changes. In addition, an extreme reduction to a continuum membrane model and backmapping to the MARTINI model has been explored recently to simulate membrane transformation of an entire mitochondrion (Pezeshkian et al., 2020). When this type of multiscale approach is combined with an accurate description of membrane-protein system at the molecular level by Gō-MARTINI, it will be a powerful tool to simulate large systems of membrane remodeling processes.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

MM, AP, and KO conceived the study. MM, AP, and KO performed the computations and analyzed the data. MM, AP, and KO discussed the results. MM and KO wrote the manuscript with contributions from AP.

FUNDING

This work was supported by the Building of Consortia for the Development of Human Resources in Science and Technology, MEXT, Japan. KO was supported by JSPS KAKENHI Grant number JP18H02415. AP thanks the financial support from the National Science Centre, Poland, under grant number 2017/26/D/NZ1/00466.

ACKNOWLEDGMENTS

Computations were partially performed using the Research Center for Computational Science, Okazaki, Japan. This research was supported in part by PLGrid Infrastructure.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2021.619381/full#supplementary-material.

REFERENCES

Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., et al. (2015). Gromacs: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *Software* 1 (2), 19–25. doi:10.1016/j.softx.2015.06.001

Alessandri, R., Souza, P. C. T., Thallmair, S., Melo, M. N., De Vries, A. H., and Marrink, S. J. (2019). Pitfalls of the martini model. *J. Chem. Theor. Comput.* 15, 5448–5460. doi:10.1021/acs.jctc.9b00473

Amadei, A., Ceruso, M. A., and Di Nola, A. (1999). On the convergence of the conformational coordinates basis set obtained by the essential dynamics analysis of proteins' molecular dynamics simulations. *Proteins* 36, 419–424. doi:10.1002/(SICI)1097-0134(19990901)36:4<419::AID-PROT5>3.0.CO%3B2-U

Arnarez, C., Uusitalo, J. J., Masman, M. F., Ingólfsson, H. I., De Jong, D. H., Melo, M. N., et al. (2015). Dry martini, a coarse-grained force field for lipid membrane simulations with implicit solvent. *J. Chem. Theor. Comput.* 11, 260–275. doi:10.1021/ct500477k

Ayton, G. S., Noid, W. G., and Voth, G. A. (2007). Multiscale modeling of biomolecular systems: in serial and in parallel. *Curr. Opin. Struct. Biol.* 17, 192–198. doi:10.1016/j.sbi.2007.03.004

Baaden, M., and Marrink, S. J. (2013). Coarse-grain modelling of protein-protein interactions. *Curr. Opin. Struct. Biol.* 23, 878–886. doi:10.1016/j.sbi.2013.09.004

Bai, X., Meng, G., Luo, M., and Zheng, X. (2012). Rigidity of wedge loop in PACSIN 3 protein is a key factor in dictating diameters of tubules. *J. Biol. Chem.* 287, 22387–22396. doi:10.1074/jbc.M112.358960

Bassereau, P., Jin, R., Baumgart, T., Deserno, M., Dimova, R., Frolov, V. A., et al. (2018). The 2018 biomembrane curvature and remodeling roadmap. *J. Phys. D Appl. Phys.* 51, 343001. doi:10.1088/1361-6463/aacb98

Benayad, Z., von Bülow, S., Stelzl, L. S., and Hummer, G. (2020). Simulation of FUS protein condensates with an adapted coarse-grained model. *J. Chem. Theory Comput., acs.jctc* 17(1), 525–537. doi:10.1021/acs.jctc.0c01064

Berendsen, H. J. C., Postma, J. P. M., Van Gunsteren, W. F., Dinola, A., and Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81, 3684–3690. doi:10.1063/1.448118

Braun, A. R., Sevcsik, E., Chin, P., Rhoades, E., Tristram-Nagle, S., and Sachs, J. N. (2012). α-Synuclein induces both positive mean curvature and negative Gaussian curvature in membranes. *J. Am. Chem. Soc.* 134, 2613–2620. doi:10.1021/ja208316h

Bussi, G., Donadio, D., and Parrinello, M. (2007). Canonical sampling through velocity rescaling. *J. Chem. Phys.* 126, 014101. doi:10.1063/1.2408420

Clementi, C., Nymeyer, H., and Onuchic, J. N. (2000). Topological and energetic factors: what determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? An investigation for small globular proteins. *J. Mol. Biol.* 298, 937–953. doi:10.1006/jmbi.2000.3693

De Jong, D. H., Singh, G., Bennett, W. F., Arnarez, C., Wassenaar, T. A., Schäfer, L. V., et al. (2013). Improved parameters for the martini coarse-grained protein force field. *J. Chem. Theor. Comput.* 9, 687–697. doi:10.1021/ct300646g

Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. *J. Mol. Graph.* 14, 33–38. doi:10.1016/0263-7855(96)00018-5

Javanainen, M., Martinez-Seara, H., and Vattulainen, I. (2017). Excessive aggregation of membrane proteins in the Martini model. *PLoS ONE* 12 (11), e0187936. doi:10.1371/journal.pone.0187936

Kabsch, W., and Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22, 2577–2637. doi:10.1002/bip.360221211

Karanicolas, J., and Brooks, C. L. (2002). The origins of asymmetry in the folding transition states of protein L and protein G. *Protein Sci.* 11, 2351–2361. doi:10.1110/ps.0205402

Klein, M. L., and Shinoda, W. (2008). Large-scale molecular dynamics simulations of self-assembling systems. *Science* 321, 798–800. doi:10.1126/science.1157834

Koga, N., and Takada, S. (2001). Roles of native topology and chain-length scaling in protein folding: a simulation study with a Go-like model. *J. Mol. Biol.* 313, 171–180. doi:10.1006/jmbi.2001.5037

Li, W., Terakawa, T., Wang, W., and Takada, S. (2012). Energy landscape and multiroute folding of topologically complex proteins adenylate kinase and 2ouf-knot. *Proc. Natl. Acad. Sci. United States* 109, 17789–17794. doi:10.1073/pnas.1201807109

Li, W., Wolynes, P. G., and Takada, S. (2011). Frustration, specific sequence dependence, and nonlinearity in large-amplitude fluctuations of allosteric proteins. *Proc. Natl. Acad. Sci. United States* 108, 3504–3509. doi:10.1073/pnas.1018983108

Mahmood, M. I., Noguchi, H., and Okazaki, K. (2019). Curvature induction and sensing of the F-BAR protein Pacsin1 on lipid membranes via molecular dynamics simulations. *Sci. Rep.* 9, 14557. doi:10.1038/s41598-019-51202-z

Marrink, S. J., Corradi, V., Souza, P. C. T., Ingólfsson, H. I., Tieleman, D. P., and Sansom, M. S. P. (2019). Computational modeling of realistic cell membranes. *Chem. Rev.* 119, 6184–6226. doi:10.1021/acs.chemrev.8b00460

Marrink, S. J., Risselada, H. J., Yefimov, S., Tieleman, D. P., and De Vries, A. H. (2007). The MARTINI force field: coarse grained model for biomolecular simulations. *J. Phys. Chem. B* 111, 7812–7824. doi:10.1021/jp071097f

Marrink, S. J., and Tieleman, D. P. (2013). Perspective on the martini model. *Chem. Soc. Rev.* 42, 6801–6822. doi:10.1039/c3cs60093a

Marrink, S. J., de Vries, A. H., and Mark, A. E. (2004). Coarse grained model for semiquantitative lipid simulations. *J. Phys. Chem. B* 108, 750–760. doi:10.1021/jp036508g

Martí-Renom, M. A., Stuart, A. C., Fiser, A., Sánchez, R., Melo, F., and Šali, A. (2000). Comparative protein structure modeling of genes and genomes. *Annu. Rev. Biophys. Biomol. Struct.* 29, 291–325. doi:10.1146/annurev.biophys.29.1.291

Martín-García, F., Papaleo, E., Gomez-Puertas, P., Boomsma, W., and Lindorff-Larsen, K. (2015). Comparing molecular dynamics force fields in the essential subspace. *PloS One* 10, e0121114–16. doi:10.1371/journal.pone.0121114

McMahon, H. T., and Gallop, J. L. (2005). Membrane curvature and mechanisms of dynamic cell membrane remodelling. *Nature* 438, 590–596. doi:10.1038/nature04396

Monticelli, L., Kandasamy, S. K., Periole, X., Larson, R. G., Tieleman, D., and Marrink, S. J. (2008). The MARTINI coarse-grained force field: extension to proteins. *J. Chem. Theor. Comput.* 4, 819–834. doi:10.1021/ct700324x

Moreira, R. A., Guzman, H. V., Boopathi, S., Baker, J. L., and Poma, A. B. (2020). Characterization of structural and energetic differences between conformations of the SARS-CoV-2 spike protein. *Materials.* 13, 5362. doi:10.3390/ma13235362

Noel, J. K., Whitford, P. C., and Onuchic, J. N. (2012). The shadow map: a general contact definition for capturing the dynamics of biomolecular folding and function. *J. Phys. Chem. B* 116, 8692–8702. doi:10.1021/jp300852d

Okazaki, K., Koga, N., Takada, S., Onuchic, J. N., and Wolynes, P. G. (2006). Multiple-basin energy landscapes for large-amplitude conformational motions of proteins: structure-based molecular dynamics simulations. *Proc. Natl. Acad. Sci. United States* 103, 11844–11849. doi:10.1073/pnas.0604375103

Okazaki, K., Sato, T., and Takano, M. (2012). Temperature-enhanced association of proteins due to electrostatic interaction: a coarse-grained simulation of actin-myosin binding. *J. Am. Chem. Soc.* 134, 8918–8925. doi:10.1021/ja301447j

Okazaki, K., and Takada, S. (2008). Dynamic energy landscape view of coupled binding and protein conformational change: induced-fit versus population-shift mechanisms. *Proc. Natl. Acad. Sci. UNITED STATES* 105, 11182–11187. doi:10.1073/pnas.0802524105

Parrinello, M., and Rahman, A. (1981). Polymorphic transitions in single crystals: a new molecular dynamics method. *J. Appl. Phys.* 52, 7182–7190. doi:10.1063/1.328693

Periole, X., Cavalli, M., Marrink, S. J., and Ceruso, M. A. (2009). Combining an elastic network with a coarse-grained molecular force field: structure, dynamics, and intermolecular recognition. *J. Chem. Theor. Comput.* 5, 2531–2543. doi:10.1021/ct9002114

Pezeshkian, W., König, M., Wassenaar, T. A., and Marrink, S. J. (2020). Backmapping triangulated surfaces to coarse-grained membrane models. *Nat. Commun.* 11, 2296. doi:10.1038/s41467-020-16094-y

Poma, A. B., Chwastyk, M., and Cieplak, M. (2015). Polysaccharide-protein complexes in a coarse-grained model. *J. Phys. Chem. B* 119, 12028–12041. doi:10.1021/acs.jpcb.5b06141

Poma, A. B., Cieplak, M., and Theodorakis, P. E. (2017). Combining the MARTINI and structure-based coarse-grained approaches for the molecular dynamics studies of conformational transitions in proteins. *J. Chem. Theor. Comput.* 13, 1366–1374. doi:10.1021/acs.jctc.6b00986

Poma, A. B., Guzman, H. V., Li, M. S., and Theodorakis, P. E. (2019). Mechanical and thermodynamic properties of Aβ42, Aβ40, and α-synuclein fibrils: a coarse-grained method to complement experimental studies. *Beilstein J. Nanotechnol.* 10, 500–513. doi:10.3762/BJNANO.10.51

Poma, A. B., Li, M. S., and Theodorakis, P. E. (2018). Generalization of the elastic network model for the study of large conformational changes in biomolecules. *Phys. Chem. Chem. Phys.* 20, 17020–17028. doi:10.1039/C8CP03086C

Senapati, S., Poma, A. B., Cieplak, M., Filipek, S., and Park, P. S. H. (2019). Differentiating between inactive and active states of rhodopsin by atomic force microscopy in native membranes. *Anal. Chem.* 91, 7226–7235. doi:10.1021/acs.analchem.9b00546

Seo, S., and Shinoda, W. (2019). SPICA force field for lipid membranes: domain formation induced by cholesterol. *J. Chem. Theor. Comput.* 15, 762–774. doi:10.1021/acs.jctc.8b00987

Souza, P. C. T., Thallmair, S., Marrink, S. J., and Mera-Adasme, R. (2019). An allosteric pathway in copper, zinc superoxide dismutase unravels the molecular mechanism of the G93A amyotrophic lateral sclerosis-linked mutation. *J. Phys. Chem. Lett.* 10, 7740–7744. doi:10.1021/acs.jpclett.9b02868

Souza, P. C. T., Thallmair, S., Conflitti, P., Ramírez-Palacios, C., Alessandri, R., Raniolo, S., et al. (2020). Protein–ligand binding with the coarse-grained Martini model. *Nat. Commun* 11, 1–11. doi:10.1038/s41467-020-17437-5

Stark, A. C., Andrews, C. T., and Elcock, A. H. (2013). Toward optimized potential functions for protein-protein interactions in aqueous solutions: osmotic second virial coefficient calculations using the MARTINI coarse-grained force field. *J. Chem. Theor. Comput.* 9, 4176–4185. doi:10.1021/ct400008p

Suetsugu, S., Kurisu, S., and Takenawa, T. (2014). Dynamic shaping of cellular membranes by phospholipids and membrane-deforming proteins. *Physiol. Rev.* 94, 1219–1248. doi:10.1152/physrev.00040.2013

Sunhwan, J., Taehoon, K., G., I. V., and Wonpil, I. (2008). CHARMM-GUI: a web-based graphical user interface for CHARMM. *J. Comput. Chem.* 29, 1859–1865. doi:10.1002/jcc.20945

Sułkowska, J. I., and Cieplak, M. (2008). Selection of optimal variants of Gō-like models of proteins through studies of stretching. *Biophys. J* 95, 3174–3191. doi:10.1529/biophysj.107.127233

Takada, S., Kanada, R., Tan, C., Terakawa, T., Li, W., and Kenzaki, H. (2015). Modeling structural dynamics of biomolecular complexes by coarse-grained molecular simulations. *Acc. Chem. Res.* 48, 3026–3035. doi:10.1021/acs.accounts.5b00338

Takada, S. (2012). Coarse-grained molecular simulations of large biomolecules. *Curr. Opin. Struct. Biol.* 22, 130–137. doi:10.1016/j.sbi.2012.01.010

Thallmair, S., Vainikka, P. A., and Marrink, S. J. (2019). Lipid fingerprints and cofactor dynamics of light-harvesting complex II in different membranes. *Biophys. J* 116, 1446–1455. doi:10.1016/j.bpj.2019.03.009

Tozzini, V. (2005). Coarse-grained models for proteins. *Curr. Opin. Struct. Biol.* 15, 144–150. doi:10.1016/j.sbi.2005.02.005

Uusitalo, J. J., Ingólfsson, H. I., Akhshi, P., Tieleman, D. P., and Marrink, S. J. (2015). Martini coarse-grained force field: extension to DNA. *J. Chem. Theor. Comput.* 11, 3932–3945. doi:10.1021/acs.jctc.5b00286

Venable, K. J., Venable, M., R., Freites, A., O'Connor, W., J., Tobias, J., D., Ramirez-Mondragon, C., et al. (2010). Update of the CHARMM all-atom additive force field for lipids: validation on Six lipid types. *J. Phys. Chem. B* 114, 7830–7843. doi:10.1021/jp101759q

Wang, Q., Navarro, M. V., Peng, G., Molinelli, E., Goh, S., Judson, B. L., et al. (2009). Molecular mechanism of membrane constriction and tubulation mediated by the F-BAR protein Pacsin/Syndapin. *Proc Natl Acad Sci UNITED STATES* 106, 12700–12705. doi:10.1073/pnas.0902974106

Wassenaar, T. A., Ingólfsson, H. I., Böckmann, R. A., Tieleman, D. P., and Marrink, S. J. (2015). Computational lipidomics with insane: a versatile tool for generating custom membranes for molecular simulations. *J. Chem. Theor. Comput.* 11, 2144–2155. doi:10.1021/acs.jctc.5b00209

Webb, B., and Sali, A. (2016). Comparative protein structure modeling using MODELLER. *Curr. Protoc. Bioinformatics* 54, 5.6.1–5.6.37. doi:10.1002/cpbi.3

Wołek, K., Gómez-Sicilia, À., and Cieplak, M. (2015). Determination of contact maps in proteins: a combination of structural and chemical approaches. *J. Chem. Phys.* 143, 243105. doi:10.1063/1.4929599

Wu, E. L., Cheng, X., Jo, S., Rui, H., Song, K. C., Dávila-Contreras, E. M., et al. (2014). CHARMM-GUI membrane builder toward realistic biological membrane simulations. *J. Comput. Chem.* 35, 1997–2004. doi:10.1002/jcc.23702

Zhang, Y., Cao, Z., Zhang, J. Z., and Xia, F. (2020). Double-well ultra-coarse-grained model to describe protein conformational transitions. *J. Chem. Theor. Comput.* 16, 6678–6689. doi:10.1021/acs.jctc.0c00551

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Molecular Biosciences

ORIGINAL RESEARCH
published: 23 February 2021
doi: 10.3389/fmolb.2020.629520

Albumin Alters the Conformational Ensemble of Amyloid-β by Promiscuous Interactions: Implications for Amyloid Inhibition

*Huisi Xie and Cong Guo**

Department of Physics and International Centre for Quantum and Molecular Structures, College of Sciences, Shanghai University, Shanghai, China

Human serum albumin (HSA) is a key endogenous inhibitor of amyloid-β (Aβ) aggregation. In vitro HSA inhibits Aβ fibrillization and targets multiple species along the aggregation pathway including monomers, oligomers, and protofibrils. Amyloid inhibition by HSA has both pathological implications and therapeutic potential, but the underlying molecular mechanism remains elusive. As a first step towards addressing this complex question, we studied the interactions of an Aβ42 monomer with HSA by molecular dynamics simulations. To adequately sample the conformational space, we adapted the replica exchange with solute tempering (REST2) method to selectively heat the Aβ42 peptide in the absence and presence of HSA. Aβ42 binds to multiple sites on HSA with a preference to domain III and adopts various conformations that all differ from the free state. The β-sheet abundances of H14-E22 and A30-M33 regions are significantly reduced by HSA, so are the β-sheet lengths. HSA shifts the conformational ensemble towards more disordered states and alters the β-sheet association patterns. In particular, the frequent association of Q15-V24 and N27-V36 regions into β-hairpin which is critical for aggregation is impeded. HSA primarily interacts with the latter β-region and the N-terminal charged residues. They form promiscuous interactions characterized by salt bridges at the edge of the peptide-protein interface and hydrophobic cores at the center. Consequently, intrapeptide interactions crucial for β-sheet formation are disrupted. Our work builds the bridge between the modification of Aβ conformational ensemble and amyloid inhibition by HSA. It also illustrates the potential of the REST2 method in studying interactions between intrinsically disordered peptides and globular proteins.

Keywords: alzheimer's disease, amyloid-beta, serum albumin, conformational ensemble, solute tempering, promiscuous interactions

OPEN ACCESS

Edited by:
Wenfei Li,
Nanjing University, China

Reviewed by:
Philippe Derreumaux,
UPR9080 Laboratoire de Biochimie Théorique (LBT), France
Wenhui Xi,
Chinese Academy of Sciences (CAS), China

****Correspondence:***
Cong Guo
congguo@shu.edu.cn

Specialty section:
This article was submitted to Biological Modeling and Simulation, a section of the journal Frontiers in Molecular Biosciences

Received: *15 November 2020*
Accepted: *30 December 2020*
Published: *23 February 2021*

Citation:
Xie H and Guo C (2021) Albumin Alters the Conformational Ensemble of Amyloid-β by Promiscuous Interactions: Implications for Amyloid Inhibition. Front. Mol. Biosci. 7:629520. doi: 10.3389/fmolb.2020.629520

INTRODUCTION

The pathogenesis of Alzheimer's disease (AD) is tightly correlated with the abnormal aggregation of amyloid-β (Aβ) in the central nervous system (CNS). Numerous endogenous proteins interacting with Aβ can modulate its amyloidogenic process (Bohrmann et al., 1999; Han et al., 2016). Human serum albumin (HSA), the most abundant protein in blood, has been recognized as an inhibitor of Aβ aggregation (Biere et al., 1996; Bohrmann et al., 1999; Kuo et al., 2000; Ezra et al., 2016). It binds Aβ and facilitates Aβ efflux from the cerebrospinal fluid (CSF) to plasma (Boada et al., 2020).

Reduced serum albumin levels are associated with increasing cognitive impairment in AD patients (Yamamoto et al., 2014). Moreover, a phase IIb/III trial using plasma exchange with albumin replacement has presented initial encouraging results (Boada et al., 2020). In vitro, substantial evidence suggests that HSA inhibits Aβ aggregation and binds multiple species along the aggregation pathway which include monomers, oligomers, and protofibrils (Milojevic et al., 2007; Milojevic et al., 2009; Milojevic and Melacini, 2011; Stanyon and Viles, 2012; Algamal et al., 2013; Milojevic et al., 2014; Wang et al., 2016; Choi et al., 2017; Algamal et al., 2017; Bode et al., 2018). Despite the biological and therapeutic significance of HSA-Aβ interactions, the underlying mechanism is not fully understood. Molecular dynamics (MD) simulations hold great potential to contribute to solving the puzzle. However, with conventional MD, it is challenging to adequately sample the conformational space of the Aβ-HSA complex due to the intrinsic disorder of Aβ and the large system size. The present work reports the adaption of an enhanced sampling method called replica exchange with solute tempering (REST2) (Wang et al., 2011) to study the interactions of monomeric Aβ with HSA.

Aβ is a 36-43-residue peptide derived from the amyloid precursor protein (Nasica-Labouze et al., 2015). The two common isoforms are the 40-residue Aβ40 and 42-residue Aβ42, with the latter having two extra residues (I41-A42). Although Aβ40 is more abundantly produced, Aβ42 is more disease relevant as it is more abundant in amyloid plaques and shows a greater tendency to aggregate in vitro (Nasica-Labouze et al., 2015). The amino acid sequence of Aβ42 can be divided into four regions according to hydrophobicity: the hydrophilic N-terminal D1-K16 region that is comprised of 6 charged residues and 3 histidines, the central hydrophobic core (CHC) region L17-A21, the hydrophilic central region E22-G29, and the hydrophobic C-terminal region A30-A42. Monomeric Aβ is classified as an intrinsically disordered peptide (IDP), but solution nuclear magnetic resonance (NMR) experiments have detected transient β-sheet structures, especially in the CHC, I31-V36, and V39-I41 regions (Hou et al., 2004). β-hairpin conformation with two legs at residues L17-D23 and A30-V36 was stabilized by the amyloid inhibitor protein $Z_{Aβ3}$, indicating an important role of the β-hairpin structure in fibrillization (Hoyer et al., 2008). Different from monomers, Aβ fibrils are featured by in-register parallel cross-β sheet structures. Recently, several groups have solved atomic resolution structures of Aβ42 fibrils with advanced solid-state NMR and cryo-electron microscopy (cryo-EM) techniques (Xiao et al., 2015; Colvin et al., 2016; Wälti et al., 2016; Gremer et al., 2017). In these structures, the N-terminal region is disordered or partially ordered while the other regions are arranged into 3 or 4 β-strands linked by loops, which results in an overall S-shape. Especially, residues in the CHC region and the C-terminal region constitute the cross-β structures in all structures, reinforcing their critical roles in aggregation as have been established by many studies (Liu et al., 2004; Williams et al., 2004; Bernstein et al., 2005).

The aggregation process of Aβ is described by a nucleation-condensation polymerization model, which involves a lag phase for nucleation, a subsequent elongation phase for the rapid growth of oligomers and protofibrils into fibrils, and a final plateau phase. Though Aβ peptides circulate in CSF and in blood at similar concentrations of 0.1–0.5 nM (Stanyon and Viles, 2012), amyloid plaques were only found in CNS. It is primarily attributed to the fact that ~90% plasma Aβ is sequestered by HSA which has a concentration of 640 μM in plasma as opposed to a remarkably low level of 3 μM in CSF (Biere et al., 1996). In vitro, HSA at physiological concentrations significantly increased the lag phase time and decreased the total amount of amyloid fibers (Stanyon and Viles, 2012). A 35-residue segment in domain III retained the inhibitory effect of HSA (Picón-Pagès et al., 2019) while natural HSA ligands negated such effect (Bode et al., 2018). HSA interfered with different stages of aggregation and targeted multiple species including monomers, oligomers, and protofibrils with increasing affinities (Wang et al., 2016; Algamal et al., 2017). Although the molecular mechanism underlying the protective inhibition of Aβ aggregation by HSA has not been fully elucidated, these studies consistently indicate a role of monomeric Aβ-HSA interactions in the process, which also lay the foundation for high-order interactions between Aβ oligomers/protofibrils and HSA. Therefore, revealing the interaction mechanism of monomeric Aβ with HSA is essential for understanding the amyloid regulation by HSA.

Many experiments have been devoted to studying the monomeric Aβ-HSA interactions but current understanding of this issue is still limited due to certain inconsistency in the literature. HSA was found to bind monomeric Aβ at a stoichiometric ratio of 1:1 (Kuo et al., 2000). It is agreed that the monomeric Aβ-HSA interactions are weak. However, very different disassociation constants (K_d) ranging from submicromolar to submillimolar have been reported (Rózga et al., 2007; Costa et al., 2012; Wang et al., 2016; Algamal et al., 2017). Aβ40 and Aβ42 have different affinities to HSA whereas the order of the two is a subject of debate (Algamal et al., 2017; Litus et al., 2019). Molecular-level characterization of Aβ binding to HSA has also been provided. Saturation transfer difference NMR experiments by Algamal et al. have identified the C-terminal region of Aβ as the primary interaction site with HSA (Algamal et al., 2017). With mass spectrometry and small-angle X-ray scattering, Choi and coworkers found that HSA predominantly captured a single Aβ monomer at the groove between domains I and III, resulting in a structural change of Aβ from a random coil to an α-helix but no structural variations of HSA (Choi et al., 2017). Contradictorily, a more recent study reported that domain II contained the primary binding sites for Aβ monomers (Ishima et al., 2020). The above discrepancies could be due to different Aβ sample preparation procedures and buffer conditions which are shown to influence the Aβ-HSA interactions (Litus et al., 2019) and the presence of Aβ oligomers in the sample resulting from the intrinsic propensity of Aβ to aggregate. These factors bring challenges to experimental measurements on the monomeric Aβ-HSA interactions. Several questions remain open: 1) a comprehensive characterization of the Aβ conformations and binding sites in the complex with HSA is still lacking, which is essential for

understanding the interaction mechanism; 2) it is unknown how such information is related to amyloid inhibition.

MD is a powerful tool to probe the molecular mechanisms at the atomic level through investigating conformational ensembles of biomolecules. Previously using conventional MD simulations, we found that domain III was the primary target for Aβ binding and that fatty acids interfered with Aβ binding to HSA by quenching the conformational flexibility of the latter (Guo and Zhou, 2019). However, we failed to capture any possible Aβ conformational transitions upon binding to HSA, probably due to the relatively short simulation time and Aβ as an IDP possessing a flat free energy surface. On this issue, enhanced sampling methods are needed, among which replica exchange molecular dynamics (REMD) (Sugita and Okamoto, 1999) has been widely used to study Aβ peptides (Rosenman et al., 2016; Man et al., 2017) and other IDPs (Guo et al., 2015). In REMD, multiple replicas of a system are simulated at different temperatures simultaneously and neighboring replicas are attempted to exchange periodically using the Metropolis criterion. A random walk of replicas in the temperature space allows them to escape local minimum. However, the use of REMD to large systems such as the Aβ-HSA complex (>620 residues) is computationally restricted by the large number of replicas required to cover a wide temperature range with reasonable exchange probabilities.

As an alternative, the replica exchange with solute tempering (REST) method has been developed (Liu et al., 2005) and later modified in REST2 (Wang et al., 2011) to improve sampling efficiency. It has been successfully applied to the conformational sampling of IDPs (Côté et al., 2015; Rossetti et al., 2016; Smith et al., 2016; Han et al., 2017; Lee and Chen, 2017; Hicks and Zhou, 2018). REST2 is a new form of Hamiltonian replica exchange method wherein all replicas are simulated at the same temperature T_0 albeit on different deformed potential energy surfaces. With delicate energy scaling, exchange probability between two replicas is exclusively determined by protein-related energy terms that involve a small number of atoms, not by the energy of a large number of solvent molecules. Consequently, the number of replicas can be reduced four to five times without changes in the temperature range (Smith et al., 2016). Another important consequence is that part of the solute instead of all solute atoms can be chosen for scaling to achieve enhanced sampling. For example, it has been used to sample the conformations of a disordered loop in a globular protein (Pang and Zhou, 2015). This feature is perfectly suited for exploring the conformational ensemble of Aβ in the large complex with HSA, whereby Aβ is highly dynamic while HSA experiences little conformational changes (Choi et al., 2017).

Herein, we have employed the REST2 protocols to study the interactions of the more toxic Aβ42 monomer with HSA. By choosing Aβ42 atoms for scaling, we can use the same number of replicas to achieve enhanced sampling of Aβ42 with and without HSA. Simulations of the isolated Aβ42 peptide yield consistent secondary structure contents with previous REMD studies (Rosenman et al., 2016), demonstrating the applicability of the REST2 protocols. Aβ42 binds to five major sites on the HSA surface with a preference to domain III, consistent with our previous work (Guo and Zhou, 2019). The binding site at the cleft of domains I and III is similar to the one reported by ion mobility mass spectrometry (Choi et al., 2017). Aβ42 adopts different conformations at different binding sites, which in general are less β-sheet-rich and contain shorter β-strands than the free state. HSA significantly suppresses the β-sheet propensities of the H14-E22 and A30-M33 regions and alters the intrapeptide interaction patterns as well. Particularly the interactions between the Q15-V24 region and the N27-V36 region which are dominant in the free state are disrupted by HSA. Aβ42 interacts with HSA primarily via the N-terminal charged residues and the K28-M35 segment. An interaction mechanism is proposed wherein Aβ42 promotes promiscuous interactions with HSA that conflict with intrapeptide interactions curial for β-sheet formation. Implications of our findings in amyloid inhibition are also discussed.

MATERIALS AND METHODS

System Preparation

The sequence of Aβ42 is DAEFRHDSGY^{10} EVHHQKLVFF^{20} AEDVGSNKGA^{30} IIGLMVGGVV^{40} IA. The starting structure of Aβ42 was built upon the NMR structure of Aβ40 in aqueous solution (PDB 2LFM) (Vivekanandan et al., 2011) by adding the two C-terminal residues (I41-A42) with PyMol (DeLano, 2002). HSA is a 585-residue protein and consists of three homologous domains I to III (**Figure 1A**). Each domain can be further divided into subdomains a and b. The initial coordinates of HSA were taken from its crystal structure (PDB 1AO6) (Sugio et al., 1999). Two systems were simulated, the Aβ42 monomer alone (Aβ42) and in the presence of HSA (Aβ42 + HSA). The Aβ42 + HSA system contained one Aβ42 molecule and one HSA molecule, for which 8 different initial configurations (**Figure 1A**) were generated by randomly placing the Aβ42 peptide at different positions 10 Å away from HSA. Each initial configuration seeded two replica simulations.

Simulation Setup

We performed all simulations using the GROMACS 2018.1 software package (Abraham et al., 2015) patched with the PLUMED plug-in (version 2.4.2) for REST simulations (Bussi, 2014; Tribello et al., 2014). GPU acceleration (Páll and Hess, 2013) was used to increase computation performance. The Amber99sb-ILDN (Lindorff-Larsen et al., 2010) force field and the TIP3P water model were used. For both Aβ42 and Aβ42 + HSA, the solute was energy minimized in vacuum first and then solvated in a dodecahedron box with a minimal distance of 10 Å from the box boundaries. Counterions were added to neutralize the net charge of proteins and generate a salt concentration of 150 mM. The whole system was heated gradually to 300 K in 200 ps. Then, it was equilibrated for 200 ps under an NVT ensemble and for another 200 ps under an NPT ensemble. During the whole equilibration process, protein heavy atoms were restrained. In the final production runs, these restraints were removed and all protein bonds were restrained by LINCS (Hess et al., 1997). The Particle Mesh Ewald method (Darden et al.,

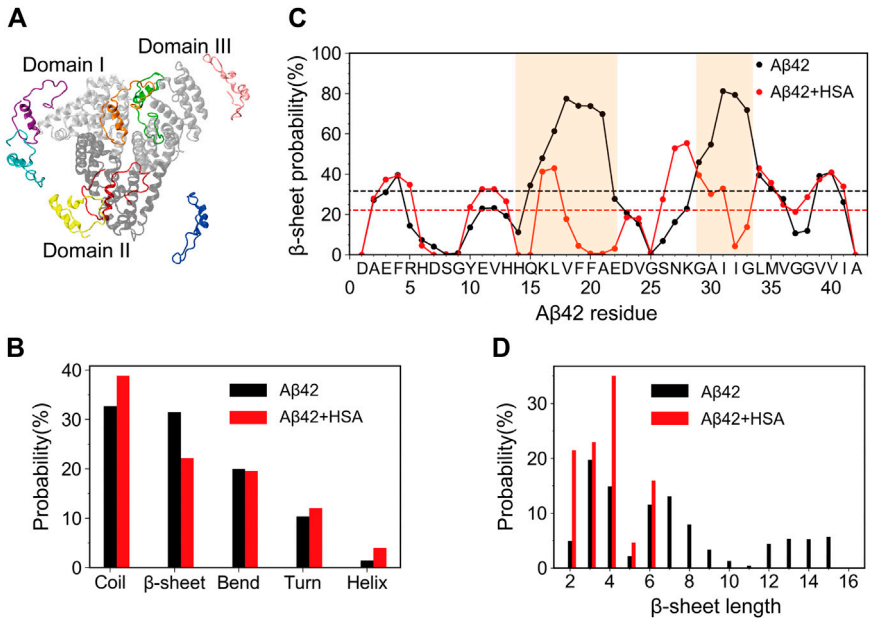


FIGURE 1 | Starting structures of simulations for Aβ42 with HSA and secondary structure changes of Aβ42 upon binding to HSA. **(A)** Superimposition of 8 Aβ42 (in color) starting positions around HSA. Domains I, II, and III of HSA are shown in silver, gray, and light gray, respectively. **(B)** The average probability of each secondary structure content. **(C)** Residue-specific β-sheet probability. The average value of each curve is shown as a horizontal dashed line. Regions (H14-E22 and G29-G34) that display significant changes are highlighted by brown shading. **(D)** Histograms of β-sheet lengths of Aβ42 in the two systems.

1993) with a real-space cut-off of 10 Å was used to calculate long-range electrostatic interactions. Temperature was maintained at 300 K by the velocity rescaling method (Bussi et al., 2007). Pressure was maintained at 1 bar by the Parrinello-Rahman coupling method (Parrinello and Rahman, 1981; Nosé and Klein, 1983). The simulation time step was 2 fs. Snapshots were saved every 10 ps. More details about REST2 simulation parameters are given below.

Details of REST2 Protocol

The REST2 method was used to enhance the sampling of the Aβ42 peptide. In REST2, the total potential energy of a system is decomposed into three components: the protein intramolecular energy E_{pp}, the interaction energy between protein and solvent E_{pw}, and the self-interaction energy between solvent molecules E_{ww}. For each replica, its potential energy is

$$E = \lambda E_{pp} + \sqrt{\lambda} E_{pw} + E_{ww}.$$

Scaling is limited to the first two terms and all replicas are assigned different scaling factors λ ranging from 0 to 1. Enhanced sampling is achieved by equivalently heating protein to a higher effective temperature T_0/λ while the solvent molecules remain cold at T_0. For both Aβ42 and Aβ42 + HSA, all atoms of the Aβ42 peptide were selected as the "hot" solute region; the other atoms were kept unperturbed which were equivalently treated as the "solvent" region. In different replicas, Aβ42-Aβ42 and Aβ42-other interactions were scaled to generate an effective temperature ladder for the "hot" region, while the "solvent" temperature remained a constant. We used 16 replicas at the effective temperatures exponentially spaced between 300 and 600 K. The effective temperature ladder was 300.0, 314.1, 328.9, 344.8, 361.0, 377.8, 395.8, 414.4, 434.2, 454.5, 476.2, 498.3, 522.6, 547.4, 572.5, and 600.0. Exchange between neighboring replicas was attempted every 2 ps. The average exchange rates for the two systems are the same, 33.2% for Aβ42 and 32.8% for Aβ42 + HSA. Each replica simulation lasted 800 ns for Aβ42 and 500 ns for Aβ42 + HSA. For both systems, the last 200 ns from the unscaled replica (i.e., at 300 K) was used for analysis.

Analysis

All analyses were carried out with built-in tools in GROMACS and our in-house-developed codes. Secondary structures of Aβ42 were determined by the DSSP (Kabsch and Sander, 1983) program. The cluster analysis of Aβ42 conformations was performed with gmx cluster in GROMACS using a backbone root-mean-square deviation (RMSD) cut-off of 0.2 nm. The binding propensity of one residue in one protein was defined as the percentage of snapshots in which it was in contact with the partner protein. A contact was defined when two heavy atoms lie within 5.4 Å. For each snapshot, the Aβ42 binding pose was characterized by the position of Aβ42 relative to HSA, which was calculated as the center-of-mass coordinates of Aβ42 after superimposing HSA to the starting structure using backbone atoms. All poses sampled in the last 200 ns were partitioned into clusters by the DBSCAN algorithm (Ester et al., 1996). A salt bridge is considered to be formed if the distance between any of the oxygen atoms of acidic residues and the nitrogen atoms of

basic residues is within 4 Å. All structure figures were prepared in VMD (Humphrey et al., 1996).

RESULTS

Convergence of Simulations

We carried out comparative REST2 simulations of Aβ42 with and without HSA so as to provide atomic-level insight on Aβ42-HSA interactions, with a focus on the effect of HSA on Aβ42 conformational ensemble and binding properties of Aβ42. Two systems are denoted by Aβ42 + HSA and Aβ42, respectively. By taking the advantage that the REST2 method can heat a part of the system, we selectively enhanced the sampling of Aβ42 conformational ensemble with affordable computation cost. A 700/500 ns REST2 simulation was performed for Aβ42/Aβ42 + HSA, which led to an accumulative simulation time of 11.2/9 μs. Throughout each of the two simulations, each of the 16 replicas visited all of the 16 effective temperatures. The percentages of dwell time of 16 replicas at each effective temperature fluctuate around 6.25% with standard deviations at 1~4% for Aβ42 and at 2~6% for Aβ42 + HSA (**Supplementary Figures S1A,B**). It indicates sufficient exchanges between replicas and thus verifies the sampling efficiency. Furthermore, the convergency of simulations was checked by comparing the radius of gyration (Rg) and the secondary structure probabilities of Aβ42 in different time intervals from the unscaled replica (i.e., 300 K). For both systems, the distribution curves of Rg in two independent 100 ns time intervals of the last 200 ns overlap well with each other (**Supplementary Figures S1C,D**); the probabilities of each secondary structure content in two different time intervals are the same (**Supplementary Figures S1E,F**). Moreover, secondary structures of Aβ42 are consistent with previous REMD simulations which started from extended coils (Rosenman et al, 2016), evidencing the insensitivity of simulation results to the initial conformation. These results demonstrate that two REST2 simulations have reasonably converged in the last 200 ns.

Initially, the Aβ42 peptide was randomly placed at 8 different positions 10 Å away from HSA. At the effective temperature of 300 K, Aβ42 diffuses onto the surface of HSA within 50 ns and basically remains in a bound state until 500 ns. Disassociation of Aβ42 from HSA is observed but the frequency is extremely low. Especially in the last 200 ns, Aβ42 is disassociated from HSA in only 4% of total frames. With the increase of effective temperature, the binding probability of Aβ42 to HSA decreases. Above 414 K, Aβ42 is bound to HSA in 34~76% of total frames. During simulations, HSA displayed an average backbone RMSD at 3.5 Å at both low and high temperatures, justifying our assumption that HSA has little conformational changes upon Aβ42 binding. Root-mean-square fluctuations (RMSFs) of HSA residues do not change with temperatures (**Supplementary Figure S2**). Large conformational changes of HSA are not accessible by our simulations. Structural stabilities of HSA probably account for the high binding probabilities of Aβ42 at high temperatures. Only data from the unscaled replica (i.e., 300 K) are meaningful for analysis, because in the other replicas, the system evolves on a deformed energy surface. Unless specified, all results below are based on data of the last 200 ns at 300 K, during which it is fair to consider that Aβ42 remains bound to HSA.

HSA Reduces the β-Sheet Abundance of the H14-E22 and A30-G33 Regions of Aβ42 and Prevents Formation of Long β-Strands

We first analyzed the influence of HSA on the secondary structures of Aβ42. The average probability of each secondary structure (including coil, β-sheet, bend, turn, and helix) was calculated. As shown in **Figure 1B**, the isolated Aβ42 peptide mainly adopts random coil (32.7%) and β-sheet (31.5%) structures, in accordance with its intrinsically disordered nature. Bend and turn contents are a little lower (20.1% and 10.3%) while the helix content (1.4%) can be neglected. These results are similar to those obtained by circular dichroism (CD) spectroscopy (27% β-sheet and 6% helix) (Fezoui and Teplow, 2002) and previous REMD simulations using the same force field (~36% coil, ~26% β-sheet, ~19% bend, ~16% turn, and <3% helix) (Rosenman et al., 2016). Upon binding to HSA, the β-sheet content is significantly reduced to 20.7%, whereas the coil content increases to 40.5% and the helix content slightly increases to 4.0%. The increase of helix propensities upon complexation with HSA was also detected by previous CD experiments (Choi et al., 2017). The bend and turn contents do not change much, which are 19.9% and 12.0%, respectively.

To elaborate the apparent changes of the β-sheet abundance, we show the residue-specific β-sheet probabilities of Aβ42 with and without HSA in **Figure 1C**. For each system, the average β-sheet probability is indicated by a horizontal dashed line. For the isolated Aβ42 peptide, three continuous segments form β-sheets, which include two long stretches spanning residues Y10-V24 and S26-I41 and a short stretch covering the N-terminal residues A2-H6. Residues Q15-A21 and G29-M35 exhibit relatively high β-sheet propensities. The former covers the CHC region and the latter belongs to the C-terminal region. We recall that both regions are critical for fibrillization (Liu et al., 2004; Bernstein et al., 2005). Besides, residues E3-F4 and V39-V40 display above-average β-sheet probabilities. Similar β-sheet profiles were reported by previous REMD simulations of the Aβ42 monomer (Rosenman et al., 2016) and dimer (Man et al., 2017). Our results are also consistent with NMR experiments which detected β-strands in the CHC region, residues I31-V36 and V39-I41 (Hou et al., 2004).

In the presence of HSA, the above-mentioned β-regions are preserved to some extent, but pronounced changes occur to two continuous β-segments spanning Y10-V24 and S26-I41. The first region splits into three short ones, Y10-H13, H16-F19, and E22-V24. Particularly, residues H14-E22 suffer the greatest reduction of β-sheets with all β-sheet probabilities falling below the average. The S26-I41 region splits as well at I32. The β-sheet probabilities of residues A30-G33 are also significantly decreased. The discontinuous β-regions in the presence of HSA imply that the length of β-sheets should vary from that without HSA. Therefore, we plotted the histograms of β-sheet lengths of Aβ42 in the two

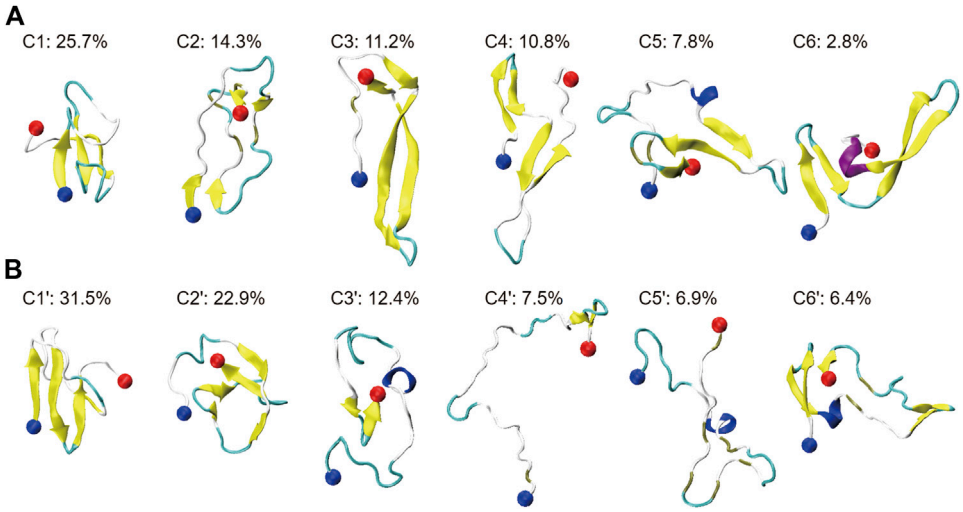


FIGURE 2 | Cluster analysis on the conformational ensemble of the Aβ42 peptide in two systems: **(A)** Aβ42 and **(B)** Aβ42 + HSA. For two systems, representative conformations of the Aβ42 molecule in the top six most-populated clusters are shown as well as the corresponding population of each cluster. The blue and red balls refer to the Cα atoms of the N- and C-terminal residues (D1 and A42), respectively.

systems (**Figure 1D**). Without HSA, the β-sheet length of Aβ42 ranges from 2 to 15. Both short β-stands (3–4 residues) and long β-strands (6–8 or 12–15 residues) have relatively high probabilities. With HSA, Aβ42 is more prone to form short β-strands composed of 2–4 and 6 residues; longer β-strands disappear.

Overall, HSA suppresses the β-sheet formation of Aβ42, in line with its inhibitory effect on Aβ fibrillization (Stanyon and Viles, 2012). Not only are the β-sheet propensities at residues H14-E22 and A30-G33 significantly reduced, but also the β-sheet length is much shorter in the presence of HSA. Changes in secondary structures hint at different tertiary structures of Aβ42 in two systems.

HSA Shifts the Conformational Ensemble of Aβ42 Towards Less β-Sheet-Rich States and Modifies the β-Sheet Associations

To investigate the influence of HSA on the conformational ensemble of Aβ42, we clustered Aβ42 conformations using a backbone RMSD cut-off of 0.2 nm. For Aβ42 and Aβ42 + HSA, 679 and 181 clusters are found, respectively. Representative conformations of the top six most-populated clusters and the corresponding populations are shown in **Figure 2** (C1–C6 for Aβ42, C1′–C6′ for Aβ42 + HSA). These clusters account for 73 and 88% of the total snapshots of Aβ42 and Aβ42 + HSA, respectively. For both systems, conformations in the remaining clusters resemble those in the top 6 clusters, as judged by the residue-specific β-sheet probabilities (**Supplementary Figure S3**). The β-sheet structures in the G29–G37 region are further suppressed by HSA in the remaining clusters of Aβ42 + HSA. Thus, the remaining clusters which all have populations below 1% are omitted here. The smaller number of clusters and the larger proportion of the top 6 clusters reflect that the structural diversity of Aβ42 in the presence of HSA is less pronounced than the isolated form.

Without HSA, the conformational ensemble of Aβ42 is featured by β-sheet-rich structures. The most populated conformation contains a three-stranded β-sheet structure, which also appears in C5. Meanwhile, β-hairpin structures are frequently observed in C3, C4, and C6. Disordered structures are only observed in C2, which contain two short β-hairpins at the N- and C-terminus. In the presence of HSA, even though C1′, C2′, and C6′ are still β-sheet-rich, the conformational ensemble is shifted towards more disordered states. Conformations in C3′, C4′, and C5′ are dominated by extended coils. Among them, the conformation of C4′ is the most extended. In addition, short helices are observed in C3′, C5′, and C6′, accounting for 25.7% of total snapshots. For the isolated Aβ42, helical structures are observed in C5 and C6 with a total percentage of 10.6%. It is consistent with the slight increase of helix content.

To better characterize the tertiary structures, we illustrate the β-strands and β-sheet associations in each representative conformation in **Figure 3**. β-strands are represented with strips and those assembling into one β-sheet are paired with the same color. We partitioned the whole sequence into 5 conserved β-regions by grouping residues that form β-strands in more than 2 clusters of Aβ42 (C1 to C6) or Aβ42 + HSA (C1′ to C6′). They are identified as follows: A2-H6 (β1), Y10-H13 (β2), Q15-V24 (β3), N27-V36 (β4), and G38-I41 (β5). As have been reported by previous simulations (Song et al., 2015; Man et al., 2017), the extra two C-terminal residues of Aβ42 stabilize an additional β-strand spanning G38 to I41 that is absent in Aβ40. One conformation differs from another in β-regions and the way they assemble into β-sheets. Therefore, we listed the composition of β-regions for each conformation and used a dash character to represent the hydrogen bonding connection. Interestingly, the five β-regions, consistent with previous MD simulations of the Aβ42 monomer (Song et al., 2015; Rosenman et al., 2016) and dimer (Man et al., 2017),

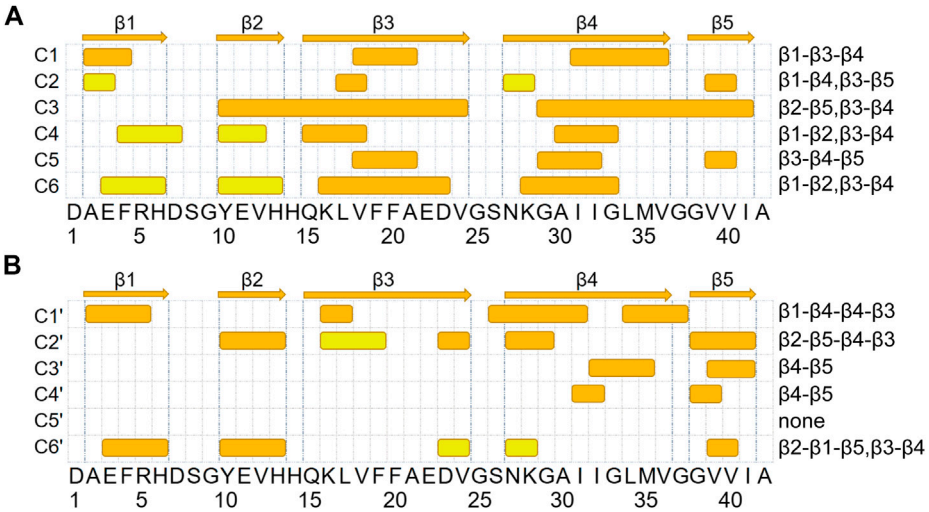


FIGURE 3 | β-sheet associations for Aβ42 and Aβ42 + HSA are displayed in **(A, B)**, respectively. β-strands formed in the representative conformations of the top six most-populated clusters are shown with colored strips. The corresponding snapshot of each conformation is shown in **Figure 2**. β-strands that associate into the same β-sheet are paired with the same color. At the top, five β-segments are represented by arrows.

overlap well with those in Aβ42 fibrils (**Supplementary Figure S5**) (Xiao et al., 2015; Colvin et al., 2016; Wälti et al., 2016; Gremer et al., 2017). The best match is with the fibril structures determined by cryo-EM (Gremer et al., 2017), wherein four β-segments are at A2-G9, E11-A21, N27-L34, and V39-I41, respectively, and the second segment combines β2 and β3 here. It implies that β-sheet motifs of the Aβ42 monomer are closely related to the fibrillization process.

For the isolated Aβ42 peptide, the most frequently occurring β-regions are Q15-V24 (β3) and N27-V36 (β4), consistent with the fact that the two regions have the highest β-sheet propensities (**Figure 1C**). The other three β-regions are also observed in multiple clusters, but the corresponding β-sheet lengths are much shorter (2–4 residues). Furthermore, the most frequent association is between Q15-V24 (β3) and N27-V36 (β4) as well, which appears in all clusters except C2. In C1 and C5, β3 and β4 form a three-stranded β-sheet together with an additional β-strand. REMD simulations of the Aβ42 dimer also reported the similar β-hairpin (CHC and A30-V36) and the three-stranded β-sheet in C5 (L17-A21, A30-V36, and V39-I41) (Man et al., 2017). In C3, C4, and C6, β3 and β4 form a β-hairpin. The A2–H6 (β1) region primarily associates with Q15-V24 (β3) as in C1 and also has certain probabilities to associate with Y10-H13 (β2) as in C4 and C6. The V39-I41 (β5) region at the C-terminus mainly associates with Y10-H13 (β2) as in C3 and with N27-V36 (β4) as in C4 and C6. Note that C4 and C6 share the same β-sheet association pattern (β1-β2, β3-β4) but the relative orientations of the resulting two β-hairpins are different. From these data, we conclude that the association of Q15-V24 with N27-V36 (i.e., β3-β4) serves as a core of β-sheet-rich conformations. Consistently, the β-hairpin formed by residues K16-E22 and G29-M35 has been suggested as a basic monomeric unit for the aggregation process (Abelein et al., 2014).

Upon binding to HSA, the β-sheet association of each cluster is different from any of clusters C1 to C6. Though the β3-β4 association is frequently observed in C1′, C2′, and C6′, the β3 strands are much shorter. Moreover, the association between β4 and β5 is also frequent, which occurs in C2′, C3′, and C4′. Meanwhile, two new associations emerge. The first is between two β-segments within the β4 region (i.e., β4-β4) in C1′. The second is between β1 and β5 in C6′. In contrast, the associations of β3 with β1 and β5 observed in C1 and C2, respectively, disappear. To sum up, HSA impairs associations of the Q15-V24 region with the rest, promotes associations of the N27-V36 region with the C-terminus, and induces new associations within the N27-V36 region and between N- and C-terminal β-regions.

HSA Modifies Intrapeptide Interaction Patterns of Aβ42

The decrease in the abundance and lengths of β-strands together with the changes in β-sheet associations suggests that the intrapeptide interactions of Aβ42 would be changed by HSA. To validate this conjecture, we calculated the contact probabilities of all the residue pairs of Aβ42 with and without HSA and showed the results in **Figure 4**. For the isolated Aβ42, the matrix elements with high contact probabilities are away from the diagonal, indicating that long-range interactions are dominated. The strongest interactions are observed between Q15-V24 (β3) and N27-V36 (β4), consistent with the highest β-sheet propensities of β3 and β4 (**Figure 1C**) and the frequent association between the two (**Figure 3B**). The corresponding antidiagonal submatrix signifies an antiparallel arrangement of two β-strands, as observed in C1 and C3 to C6 (**Figure 2A**). The submatrix constituted by A2-H6 and Y10-H13 regions has the second highest contact probabilities, corresponding to the formation of β-hairpin by β1 and β2 in C4 and C6. The antidiagonal submatrix constituted by β1 and β3 regions also shows high contact probabilities, corresponding to antiparallel β-sheets in C1.

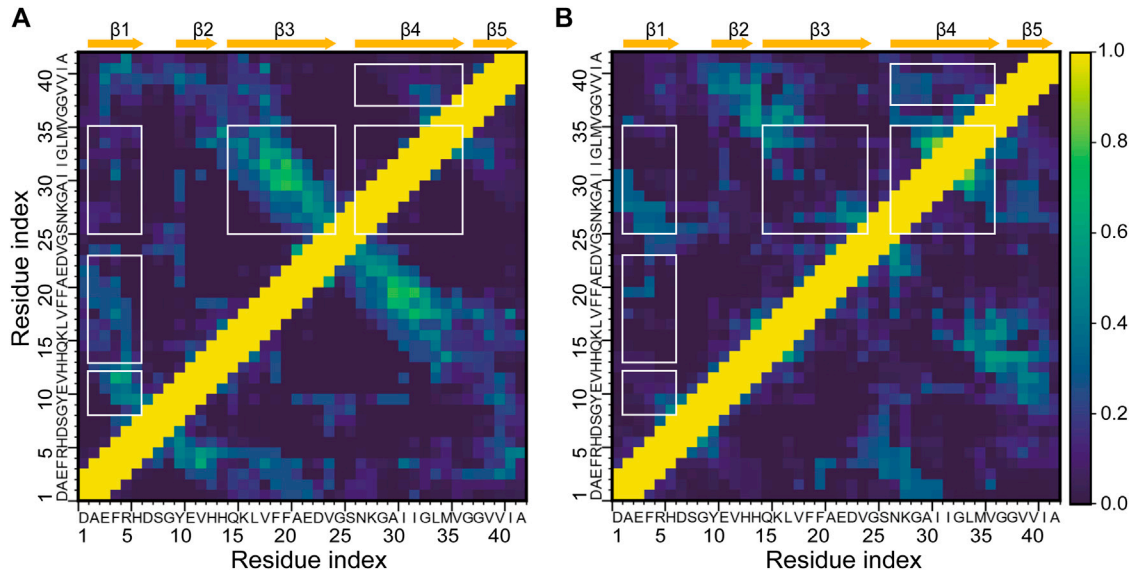


FIGURE 4 | Intrapeptide interaction maps of the Aβ42 molecule **(A)** in the absence and **(B)** in the presence of HSA. Contact probabilities are displayed in a color scale from navy to yellow. Submatrices with distinct changes upon binding to HSA are highlighted by boxes in white.

In the presence of HSA, interactions between β3 and β4 are greatly weakened, consistent with the decrease of β-sheet propensities at the two regions. Instead, β4-β5 interactions are enhanced as β4 frequently associates with β5 as well. Interactions of β1 with β2 and β3 are much weaker, too. The former results from the suppression of the β1-β2 associations, which only appears in C6′ with a population of 6.4%, as opposed to appearing in C4 and C6 of the isolated Aβ42 with a total population of 13.6%. The latter can be attributed to the disappearance of the β1-β3 association. In contrast, β1 is paired with β4 in C1′ and contributes to forming an antiparallel β-sheet. Consistently, the antidiagonal elements of the submatrix constituted by A2-R5 in β1 and S26-I31 in β4 display high contact probabilities. Lastly, local interactions within β4 are stronger, consistent with the β-sheet associations in C1′, where residues S26-I31 and L34-G37 within β4 are arranged into an antiparallel β-sheet.

The above results manifest that HSA interferes with the interactions of the Q15-V24 region with the A2-H6 and N27-V36 fragments, which are dominant in the isolated Aβ42 system and are essential for β-sheet formation. While such long-range interactions are prevented, local interactions within β4 and those between β4 and β5 are enhanced instead.

Charged and Polar Residues in the N-Terminal Region and the K28-M35 Segment are More Likely to Interact With HSA

To explain the effect of HSA on Aβ conformations, next we analyzed the binding properties of Aβ42 with HSA. Clustering of the Aβ42 positions in all snapshots identifies five major binding poses (**Supplementary Figure S4**). Poses 1 and 4 are within domain III; pose 2 is at the cleft between domains I and III; poses 3 and 5 are within domain II. Obviously, domain III is the most populated binding site among the three HSA domains. Our previous work reported similar results (Guo and Zhou, 2019), wherein we attributed high binding propensities of domain III to its high conformational flexibility (**Supplementary Figure S2B**) which was essential for HSA to adapt Aβ binding. Here, we focus on the Aβ side.

Interestingly the residue-specific HSA-binding probabilities of Aβ42 (**Figure 5**) show a dependence of residue types. In total, 20 residues have above-average binding propensities, among which 5 residues are charged, 9 are polar, and 6 are hydrophobic. The opposite trend is observed for the other 22 residues with below-average binding propensities. The number of charged, hydrophilic, and hydrophobic residues are 4, 5, and 13, respectively. It suggests that electrostatic interactions play an important role in Aβ42 binding to HSA.

The K28-M35 fragment has the highest binding propensities and is the primary interaction site with HSA. The central residues H13-Q15 also exhibit relatively high binding probabilities. These results are consistent with recent NMR data (Algamal et al., 2017) and our previous MD results (Guo and Zhou, 2019). Both studies have identified the C-terminal region as the primary binding site of HSA. It is noteworthy that the fragment K28-M35 is at the center of β4 region, which frequently associates with the Q15-V24 (β3) region into β-sheets in the absence of HSA. Although the β3 segment displays below-average HSA-binding probabilities, binding of the β4 segment to HSA would interfere with β3-β4 interactions and result in the loss of hydrogen bond partners for both. This result is reconciled with the decrease of β-sheet propensities at H14-E22 and A30-G33 regions and weaker interactions between Q15-V24 and N27-V36 regions. The C-terminal β-region V39-A42 exhibits below-average binding

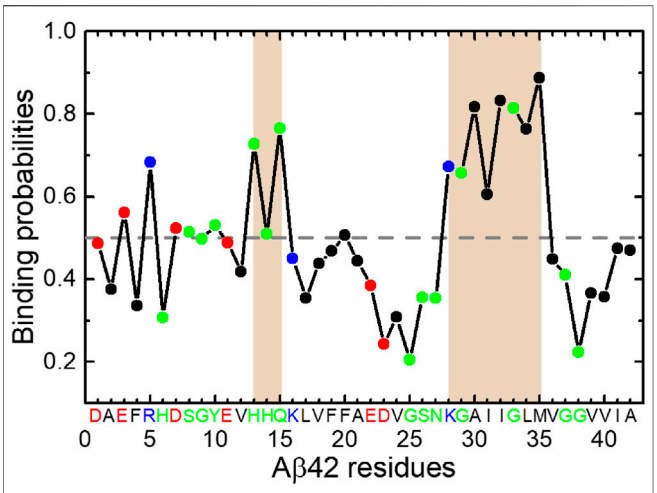


FIGURE 5 | Binding probabilities of Aβ42 residues with HSA. Data points are colored according to residue types (acidic: red; basic: blue; polar: green; hydrophobic: black). Regions displaying relatively high binding probabilities are highlighted by brown shading. The average binding probability is displayed as a horizontal dashed line.

propensities, consistent with the NMR data (Algamal et al., 2017) which have shown that interaction of the C-terminal β-strand with HSA is reduced in Aβ42 monomer but promoted in protofibrils, possibly due to the stabilization of a C-terminal turn at G37 and G38 by the last two residues.

Promiscuous Interactions Between Aβ42 and HSA Facilitate Optimal Binding but Disrupt Intramolecular Interactions Crucial for β-Sheet Formation

To further reveal the interaction mechanism of Aβ42 with HSA, structural characterizations of the Aβ42-HSA complex are necessary. By visual inspection of the snapshots, we found that the binding positions of Aβ42 are approximately the same among conformations of each cluster. Clusters C1′ to C6′ correspond to five binding poses, which are virtually identical to those shown in **Supplementary Figure S4**. The mapping relations between clusters and binding poses are as follows: C1′ to pose 1, C2′ to pose 3, C3′ and C5′ to pose 2, C4′ to pose 4, and C6′ to pose 5. Therefore, the corresponding complex conformations of cluster centers serve as good representations of all snapshots. Below we provide the structural details of each complex (**Figure 6**), paying special attention to potential conflict with intrapeptide interactions.

In C1′, the N-terminal (i.e., β1-β2) region and the K28-M35 fragment (i.e., β4) of Aβ42 bind to the pose enclosed by IIIa-h1, IIIa-h2, and the h2-h3 loop of IIIb (**Figure 6A**). These HSA-binding residues belong to a three-stranded β-sheet, which is lidded at the periphery of the complex by a random coil in the β3 region. At the edges of the binding interface, residues R5, D7, and H13 form salt bridges or hydrogen bonds (H-bonds) with HSA residues E492, K538, and E393, respectively, anchoring the Aβ42 peptide to HSA surface. Embedded within the complex, K28 forms a salt bridge with E492 of HSA, which positions the K28-M35 fragment in proximities of HSA. As a result, the β-sheet-rich structure of Aβ42 is trapped by HSA via direct interactions.

The binding pose of C2′ is constituted by the IIa-IIb loop, IIb-h3, IIb-h4, IIa-h1, IIa-h3, and IIa-h4 (**Figure 6B**). Aβ42 interacts with HSA mainly via residues H13-D23 (i.e., β3) and A30-V36 (i.e., β4). Just like in C1′, Aβ42 is anchored to HSA by two salt bridges (E11-HSA:K378 and K16-HSA:D301) at the edge of the interface. Interestingly, residues K16-F19 in the CHC region form an intermolecular β-sheet with HSA residues D301-S304. The hydrophobic loop (A30-V36) between two intramolecular β-sheets inserts into the hydrophobic groove between IIb-h3 and IIIa-h1, confining the β-sheets at the near side of HSA. At the far side, these β-sheets are covered by the disordered N-terminal residues. As can be seen, Aβ42 achieves optimal binding on the HSA surface via multipronged interactions including salt bridges, H-bonds, and hydrophobic stacking. The interaction pattern is independent of the binding sites and conformations of Aβ as manifested by preceding results and as detailed next. Intermolecular salt bridges and the corresponding probabilities are summarized in **Supplementary Table S1**.

C3′ and C5′ share a similar binding pose at the cleft between domains I and III, which involves Ia-h1, Ib-h1, Ib-h2, the Ia-Ib loop, IIIa-h3, IIIa-h4, IIIb-h1, IIIb-h2, IIIb-h3, and IIIb-h4 (**Figure 6C,E**). This binding site is similar to the one detected by mass spectroscopy (Choi et al., 2017). In two clusters, Aβ42 interacts with HSA via different residues but forms similar types of interactions. In C3′, several charged residues in the N-terminal region, S26-M35 (i.e., β4) and V39-A42 (i.e., β5) fragments are bound to HSA, whereas in C5′, all residues are in contact with HSA except the N27-L34 fragment (i.e., β4). In both clusters, hydrophobic residues (β4 and β5 for C3′ and β5 for C5′) are embedded into the groove surrounded by Ia-h3, Ib-h1, and Ib-h2, forming hydrophobic stacking with the Ia-Ib loop; charged residues form salt bridges at the interface boundaries, which involve D1, R5, K16, and K28 in C3′ and E11, K16, and E22 in C5′.

In C4′, Aβ42 binds to the backside of domains I and III, which involves the Ia-Ib loop, IIb-h3, IIb-h4, the entire IIIa, and the IIIa-IIIb loop (**Figure 6D**). It interacts with HSA extensively via residues D1-E11, L17-A21 (i.e., β3), D23-I31 (i.e., β4), and V39-A42 (i.e., β5). The N-terminal and central regions of Aβ42 are anchored to the HSA surface by salt bridges (D1-HSA:R114 and K28-HSA:E376) and H-bonds (e.g., H6-HSA:E531). Lastly, the binding pose of C6' is within domain II surrounded by IIa-h2, IIa-h3, IIa-h4, the IIa-IIb loop, IIb-h1, and IIb-h2 (**Figure 6F**). Residues D1-F4 (i.e., β1), V18-E22 (i.e., β3), S26-M35 (i.e., β4), and I41-A42 directly interact with HSA. Aβ42 adapts to the HSA surface via the K28-HSA:D308 salt bridge, H-bonds (e.g., H14-HSA:E227), and hydrophobic stacking of β3 and β4 regions with IIb-h2 and the IIa-IIb loop.

Both electrostatic and van der Waals interactions are at play in Aβ42 binding. The N-terminal residues, K16 and K28, contribute to forming salt bridges or H-bonds at the rim. Intermolecular H-bonds are especially prominent in C3′ and C4′ as listed in **Table 1**. Hydrophobic stacking via β4 or β5 regions is observed in most clusters. Generally, the electrostatic interactions are significantly stronger than the van der Waals interactions,

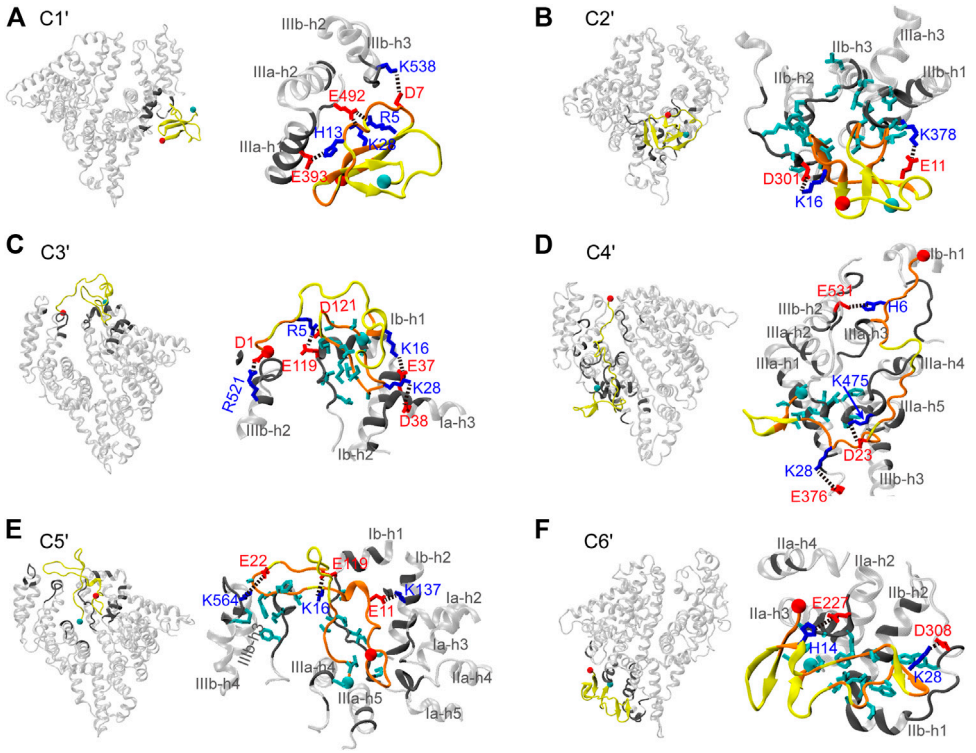


FIGURE 6 | (A–F) Representative structures of the Aβ42-HSA complexes in the six most-populated clusters. In each panel, an overview of the complex is shown on the left and an enlarged view of the binding surface is shown on the right. HSA is transparent and residues in contact with Aβ42 are highlighted in gray. Aβ42 is in yellow and residues in contact with HSA are in orange. Side chains are colored according to the residue types (acidic: red; basic: blue; hydrophobic: cyan). The Cα atoms of residues D1 and A42 are indicated by red and cyan balls, respectively.

except in C2′ and C5′ for which two terms are comparable to each other. In short, the Aβ42 peptide takes the advantage of intrinsic flexibilities to form promiscuous interactions with HSA at different binding sites. HSA usually directly targets the most occurring β-regions (i.e., β3 and β4), or it traps the β-sheet-rich conformation by protecting β-sheets from water. Either way, HSA interferes with the interaction determinants of Aβ42 aggregation.

DISCUSSION

We have applied the REST2 method to gain mechanistic insights into the interactions of Aβ42 with HSA through selectively enhanced sampling of the Aβ42 peptide. HSA dramatically changes the conformational ensemble of Aβ42 in several aspects. First, the suppression of overall β-sheet structures by HSA demonstrates the inhibitory effect on Aβ fibrillization. Second, conformations of Aβ42 are more disordered in the complex; long continuous β-strands (>6 residues) that are highly populated in the free state are completely impeded. Third, HSA weakens intrapeptide interactions and alters the patterns of remnant interactions as well. For the isolated Aβ42, the two most occurring β-regions Q15-V24 and N27-V36 assemble into β-sheets, serving as a core of β-sheet-rich structures. Residues A2-H6 interact strongly with residues Y10-H13 and Q15-V24. In the complex, all these interactions are impaired and new interaction pairs are formed. Residues Q15-V24 interact weakly with the rest of the peptide; residues N27-V36 switch to interact internally and with residues A2-H6 and V39-I41. For the other β-regions, β-sheet propensities are not affected by HSA, but interaction partners are different in the two systems.

Conformational changes of Aβ42 result from promiscuous interactions, which conflict with intramolecular β-sheet associations. HSA simultaneously interacts with both hydrophilic and hydrophobic regions, which mainly include the N-terminal charged and polar residues and the hydrophobic K28-M35 fragment. Two additional

TABLE 1 | Hydrogen bonds and interaction potential energies between Aβ42 and HSA calculated for each cluster. Interaction energies are decomposed into the electrostatic (E_{elec}) and van der Waals (E_{vdw}) terms. Standard deviations are given in parentheses.

Cluster	C1′	C2′	C3′	C4′	C5′	C6′
Hydrogen bond	4.7 (2.1)	7.1 (2.3)	15.0 (3.1)	15.0 (3.1)	9.6 (2.4)	7.3 (2.1)
E_{elec} (kJ/mol)	−296 (97)	−342 (97)	−638 (146)	−782 (124)	−510 (93)	−494 (93)
E_{vdw} (kJ/mol)	−166 (37)	−374 (37)	−227 (37)	−560 (64)	−480 (44)	−283 (46)

hydrophobic regions (CHC and C-terminus) directly interact with HSA as well but with lower probabilities. Electrostatic and van der Waals interactions cooperate to optimize the binding interface with the former being more dominant. The Aβ42-HSA interface is characterized by salt bridges or H-bonds primarily between the N-terminal region and HSA residues at the rim and stacking of hydrophobic regions at the center. Residues K16 or K28 adjacent to the hydrophobic core also form salt bridges with HSA in all six clusters. Consequently, interactions of the A2-H6 region with Y10-H13 and Q15-V24 regions are impaired; β-sheet probabilities and associations of Q15-V24 and N27-V36 regions are suppressed. In addition, such extensive interactions with HSA are incompatible with distal interactions. Instead, local intrapeptide interactions are preferable, such as interactions of the N27-V36 region with itself and the C-terminus.

Our findings provide atomistic insights into the role HSA played at the initial stage of Aβ aggregation. HSA could interfere with Aβ nucleation in several ways, which explains why HSA lengthens the lag phase of Aβ fibrillization (Stanyon and Viles, 2012). First, interactions with HSA hinder the β-sheet formation and eliminate structural characteristics resembling Aβ42 fibrils. In the free state, residues Q15-D23 and N27-V36 frequently associate into β-sheets as a core of β-sheet-rich conformations. Consistently, residues Q15-V18 and A30-I32 always formed β-sheets in all Aβ42 fibril structures (**Supplementary Figure S5**) (Xiao et al., 2015; Colvin et al., 2016; Wälti et al., 2016; Gremer et al., 2017). Residues H15-V24 formed a hydrophobic cluster with N27-L34, stabilizing a disease-relevant amyloid fibril (Wälti et al., 2016). However, HSA directly targets the second region (specifically K28-M35), significantly decreases the β-sheet abundance of both regions (especially H14-E22 and G30-G33), and impairs intrapeptide interactions between them. In additions, direct interactions of Aβ42 charged residues with HSA conflict with several salt bridges stabilizing fibril structures, including K28-D1 (Gremer et al., 2017), K28-A42 (Xiao et al., 2015; Colvin et al., 2016; Wälti et al., 2016), R5-D7 (Gremer et al., 2017), E11-H6 (Gremer et al., 2017), and E11-H13 (Gremer et al., 2017). Our findings are consistent with experiments by Stanyon and coworkers which have shown that Aβ bound to HSA is trapped in a nonfibrillar form (Stanyon and Viles, 2012). Second, although ordered β-sheet structures can be formed on the HSA surface, they are protected from exposure to water by HSA and disordered regions of Aβ42, which potentially prevents further β-sheet growth upon addition of monomers. Lastly, as Aβ42 binds to multiple sites on the HSA, it is possible that HSA concurrently traps several Aβ monomers, effectively decreasing the concentrations of monomers for nucleation. It is conceivable that HSA would interfere with Aβ42 dimerization by disrupting the common structural features shared by Aβ42 monomer and dimer (Man et al., 2017), which include similar β-sheet profiles, the β-hairpin spanning CHC and A30-V36 regions, and the three-stranded β-sheet involving L17-A21, A30-V36, and V39-I41. NMR data have shown that two terminal residues of Aβ42 extend direct interactions of protofibrils with HSA to the very C-terminal residues as compared to Aβ40 (Algamal et al., 2017). It would be interesting to carry out comparative simulations of multiple Aβ42 or Aβ40 monomers binding to HSA.

The promiscuity-centered interaction mechanism proposed here has important biological implication in the context of IDPs. Aβ42 and many other amyloid peptides (e.g., tau, amylin, and α-synuclein) belong to the family of IDPs. Accumulative evidence suggests that interacting partners of these IDPs can modulate the amyloidogenic process. In addition to diverse partners that interfere with Aβ fibrillization (Han et al., 2016; Wallin et al., 2017; Sun and Ding, 2020), amyloidosis of amylin is affected by 7B2, proSAAS (Peinado et al., 2013), lysozyme, and alpha-lactalbumin (Pilkington et al., 2017). IDPs usually promote nonspecific and dynamics multivalent interactions with targets (Weng and Wang, 2020). Some transition from disorder to order upon binding to protein partners while some keep various degrees of disorder. Aβ42-HSA interactions are typical of IDP-protein interaction regime. First, Aβ42 binds to multiple sites on the HSA surface and adopts multiple conformations including ordered β-sheet structures and extended random coils. Second, their interactions are promiscuous and do not rely on specific residue sequences. The binding interfaces commonly have salt bridges at the rim and hydrophobic cores at the center. In addition, electrostatic interactions known to enhance the binding rates and the formation of IDP-protein complexes (Zhou and Pang, 2018) play an important role in Aβ42 binding to HSA. The proposed Aβ42-HSA interaction mechanism reinforces the important role of promiscuous interactions in regulating amyloidosis. It might apply to other modulators of Aβ aggregation and probably is prevalent in the amyloid regulation by endogenous proteins.

Our study demonstrates that the REST2 method is suitable for studying IDPs, as other studies have done (Pang and Zhou, 2015; Rossetti et al., 2016; Smith et al., 2016; Han et al., 2017; Lee and Chen, 2017; Hicks and Zhou, 2018). More intriguingly, we present an example of using it to achieve efficient sampling of the IDP-protein complex, given that simulations of such systems are generally resource demanding. Our work has confirmed the theoretical expectation that the REST method can be readily used to only heat part of the system with affordable computational cost (Han et al., 2017) as the replica exchange probabilities exclusively depend on the degrees of freedom related to the hot region. For the isolated Aβ42 peptide, we obtained converged sampling with 16 replicas covering an effective temperature range from 300 to 600 K. Our results are consistent with previous REMD simulations, which in comparison used much more replicas for a similar temperature range (52 replicas spanning 270.0–601.2 K) (Rosenman et al., 2016). In the complex system, only Aβ42 was still chosen as the hot region as HSA experiences little conformational changes. Compared with the free monomer system, though the total number of atoms increases by an order of magnitude (~16000 vs. ~132000), the same number of replicas were used within the same temperature range. We hope that our study would spur future applications of REST2 on similar occasions, such as the binding and coupled folding of IDP to its globular protein partners whereby large-scale conformational changes occur to the IDP only.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

HX performed the research, analyzed the data, and wrote the manuscript. CG designed research, analyzed data, and wrote the manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China (no. 11804218) and the Young Eastern Scholar Program of the Shanghai Municipal Education Commission (no. QD2017011).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2020.629520/full#supplementary-material.

REFERENCES

Abelein, A., Abrahams, J. P., Danielsson, J., Gräslund, A., Jarvet, J., Luo, J., et al. (2014). The hairpin conformation of the amyloid β peptide is an important structural motif along the aggregation pathway. *J. Biol. Inorg. Chem.* 19, 623–634. doi:10.1007/s00775-014-1131-8

Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., et al. (2015). Gromacs: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1-2 (2), 19–25. doi:10.1016/j.softx.2015.06.001

Algamal, M., Ahmed, R., Jafari, N., Ahsan, B., Ortega, J., and Melacini, G. (2017). Atomic-resolution map of the interactions between an amyloid inhibitor protein and amyloid β (Aβ) peptides in the monomer and protofibril states. *J. Biol. Chem.* 292, 17158–17168. doi:10.1074/jbc.M117.792853

Algamal, M., Milojevic, J., Jafari, N., Zhang, W., and Melacini, G. (2013). Mapping the interactions between the Alzheimer's Aβ-peptide and human serum albumin beyond domain resolution. *Biophys. J.* 105, 1700–1709. doi:10.1016/j.bpj.2013.08.025

Bernstein, S. L., Wyttenbach, T., Baumketner, A., Shea, J. E., Bitan, G., Teplow, D. B., et al. (2005). Amyloid beta-protein: monomer structure and early aggregation states of Abeta42 and its Pro19 alloform. *J. Am. Chem. Soc.* 127, 2075–2084. doi:10.1021/ja044531p

Biere, A. L., Ostaszewski, B. L., Stimson, E. R., Maggio, J. E., and Selkoe, D. J. (1996). 762 Aβ is transported on lipoproteins and albumin in human biological fluids. *Neurobiol. Aging* 17, S189. doi:10.1016/S0197-4580(96)80764-0

Boada, M., López, O. L., Olazarán, J., Núñez, L., Pfeffer, M., Paricio, M., et al. (2020). A randomized, controlled clinical trial of plasma exchange with albumin replacement for Alzheimer's disease: primary results of the AMBAR study. *Alzheimers. Dement.* 16, 1412–1425. doi:10.1002/alz.12137

Bode, D. C., Stanyon, H. F., Hirani, T., Baker, M. D., Nield, J., and Viles, J. H. (2018). Serum albumin's protective inhibition of amyloid-β fiber formation is suppressed by cholesterol, fatty acids and warfarin. *J. Mol. Biol.* 430, 919–934. doi:10.1016/j.jmb.2018.01.008

Bohrmann, B., Tjernberg, L., Kuner, P., Poli, S., Levet-Trafit, B., Näslund, J., et al. (1999). Endogenous proteins controlling amyloid beta-peptide polymerization. Possible implications for beta-amyloid formation in the central nervous system and in peripheral tissues. *J. Biol. Chem.* 274, 15990–15995. doi:10.1074/jbc.274.23.15990

Bussi, G. (2014). Hamiltonian replica exchange in GROMACS: a flexible implementation. *Mol. Phys.* 112, 379–384. doi:10.1080/00268976.2013.824126

Bussi, G., Donadio, D., and Parrinello, M. (2007). Canonical sampling through velocity rescaling. *J. Chem. Phys.* 126, 014101. doi:10.1063/1.2408420

Choi, T. S., Lee, H. J., Han, J. Y., Lim, M. H., and Kim, H. I. (2017). Molecular insights into human serum albumin as a receptor of amyloid-β in the extracellular region. *J. Am. Chem. Soc.* 139, 15437–15445. doi:10.1021/jacs.7b08584

Colvin, M. T., Silvers, R., Ni, Q. Z., Can, T. V., Sergeyev, I., Rosay, M., et al. (2016). Atomic resolution structure of monomorphic Aβ42 amyloid fibrils. *J. Am. Chem. Soc.* 138, 9663–9674. doi:10.1021/jacs.6b05129

Costa, M., Ortiz, A. M., and Jorquera, J. I. (2012). Therapeutic albumin binding to remove amyloid-β. *J. Alzheimer's Dis.* 29, 159–170. doi:10.3233/JAD-2012-111139

Côté, S., Binette, V., Salnikov, E. S., Bechinger, B., and Mousseau, N. (2015). Probing the Huntingtin 1–17 membrane anchor on a phospholipid bilayer by using all-atom simulations. *Biophys. J.* 108, 1187–1198. doi:10.1016/j.bpj.2015.02.001

Darden, T., York, D., and Pedersen, L. (1993). Particle mesh Ewald: AnN·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* 98, 10089–10092. doi:10.1063/1.464397

DeLano, W. (2002). *Pymol: An open-source molecular graphics tool*. San Carlos, CA: DeLano Scientific.

Ester, M., Kriegel, H.-P., Sander, J., and Xu, X. (1996). "A density-based algorithm for discovering clusters in large spatial databases with noise,"in Proceedings of the 2nd international conference on knowledge discovery and data mining, Menlo Park, CA, August 1996, 226–231.

Ezra, A., Rabinovich-Nikitin, I., Rabinovich-Toidman, P., and Solomon, B. (2016). Multifunctional effect of human serum albumin reduces alzheimer's disease related pathologies in the 3xTg mouse model. *J. Alzheimers Dis.* 50, 175–188. doi:10.3233/JAD-150694

Fezoui, Y., and Teplow, D. B. (2002). Kinetic studies of amyloid beta-protein fibril assembly. Differential effects of alpha-helix stabilization. *J. Biol. Chem.* 277, 36948–36954. doi:10.1074/jbc.M204168200

Gremer, L., Schölzel, D., Schenk, C., Reinartz, E., Labahn, J., Ravelli, R. B. G., et al. (2017). Fibril structure of amyloid-β(1–42) by cryo-electron microscopy. *Science* 358, 116–119. doi:10.1126/science.aao2825

Guo, C., Côté, S., Mousseau, N., and Wei, G. (2015). Distinct helix propensities and membrane interactions of human and rat IAPP(1–19) monomers in anionic lipid bilayers. *J. Phys. Chem. B* 119, 3366–3376. doi:10.1021/jp5111357

Guo, C., and Zhou, H.-X. (2019). Fatty acids compete with Aβ in binding to serum albumin by quenching its conformational flexibility. *Biophys. J.* 116, 248–257. doi:10.1016/j.bpj.2018.11.3133

Han, M., Xu, J., and Ren, Y. (2017). Sampling conformational space of intrinsically disordered proteins in explicit solvent: comparison between well-tempered ensemble approach and solute tempering method. *J. Mol. Graph. Model.* 72, 136–147. doi:10.1016/j.jmgm.2016.12.014

Han, S. H., Park, J. C., and Mook-Jung, I. (2016). Amyloid β-interacting partners in alzheimer's disease: from accomplices to possible therapeutic targets. *Prog. Neurobiol.* 137, 17–38. doi:10.1016/j.pneurobio.2015.12.004

Hess, B., Bekker, H., Berendsen, H. J. C., and Fraaije, J. G. E. M. (1997). LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.* 18, 1463–1472. doi:10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H

Hicks, A., and Zhou, H.-X. (2018). Temperature-induced collapse of a disordered peptide observed by three sampling methods in molecular dynamics simulations. *J. Chem. Phys.* 149, 072313. doi:10.1063/1.5027409

Hou, L., Shao, H., Zhang, Y., Li, H., Menon, N. K., Neuhaus, E. B., et al. (2004). Solution NMR studies of the A beta(1–40) and A beta(1–42) peptides establish that the Met35 oxidation state affects the mechanism of amyloid formation. *J. Am. Chem. Soc.* 126, 1992–2005. doi:10.1021/ja036813f

Hoyer, W., Grönwall, C., Jonsson, A., Ståhl, S., and Härd, T. (2008). Stabilization of a beta-hairpin in monomeric alzheimer's amyloid-beta peptide inhibits amyloid

formation. *Proc. Natl. Acad. Sci. USA* 105, 5099–5104. doi:10.1073/pnas.0711731105

Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. *J. Mol. Graph.* 14, 33–38. doi:10.1016/0263-7855(96)00018-5

Ishima, Y., Mimono, A., Tuan Giam Chuang, V., Fukuda, T., Kusumoto, K., Okuhira, K., et al. (2020). Albumin domain mutants with enhanced Aβ binding capacity identified by phage display analysis for application in various peripheral Aβ elimination approaches of alzheimer's disease treatment. *IUBMB Life* 72, 641–651. doi:10.1002/iub.2203

Kabsch, W., and Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22, 2577–2637. doi:10.1002/bip.360221211

Kuo, Y. M., Kokjohn, T. A., Kalback, W., Luehrs, D., Galasko, D. R., Chevallier, N., et al. (2000). Amyloid-beta peptides interact with plasma proteins and erythrocytes: implications for their quantitation in plasma. *Biochem. Biophys. Res. Commun.* 268, 750–756. doi:10.1006/bbrc.2000.2222

Lee, K. H., and Chen, J. (2017). Efficacy of independence sampling in replica exchange simulations of ordered and disordered proteins. *J. Comput. Chem.* 38, 2632–2640. doi:10.1002/jcc.24923

Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J. L., Dror, R. O., et al. (2010). Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* 78, 1950–1958. doi:10.1002/prot.22711

Litus, E. A., Kazakov, A. S., Sokolov, A. S., Nemashkalova, E. L., Galushko, E. I., Dzhus, U. F., et al. (2019). The binding of monomeric amyloid β peptide to serum albumin is affected by major plasma unsaturated fatty acids. *Biochem. Biophys. Res. Commun.* 510, 248–253. doi:10.1016/j.bbrc.2019.01.081

Liu, P., Kim, B., Friesner, R. A., and Berne, B. J. (2005). Replica exchange with solute tempering: a method for sampling biological systems in explicit water. *Proc. Natl. Acad. Sci. USA* 102, 13749–13754. doi:10.1073/pnas.0506346102

Liu, R., McAllister, C., Lyubchenko, Y., and Sierks, M. R. (2004). Residues 17–20 and 30–35 of beta-amyloid play critical roles in aggregation. *J. Neurosci. Res.* 75, 162–171. doi:10.1002/jnr.10859

Man, V. H., Nguyen, P. H., and Derreumaux, P. (2017). Conformational ensembles of the wild-type and S8C Aβ1–42 dimers. *J. Phys. Chem. B* 121, 2434–2442. doi:10.1021/acs.jpcb.7b00267

Milojevic, J., Costa, M., Ortiz, A. M., Jorquera, J. I., and Melacini, G. (2014). In vitro amyloid-β binding and inhibition of amyloid-β self-association by therapeutic albumin. *J. Alzheimers Dis.* 38, 753–765. doi:10.3233/JAD-131169

Milojevic, J., Esposito, V., Das, R., and Melacini, G. (2007). Understanding the molecular basis for the inhibition of the alzheimer's abeta-peptide oligomerization by human serum albumin using saturation transfer difference and off-resonance relaxation NMR spectroscopy. *J. Am. Chem. Soc.* 129, 4282–4290. doi:10.1021/ja067367+

Milojevic, J., and Melacini, G. (2011). Stoichiometry and affinity of the human serum albumin-alzheimer's Aβ peptide interactions. *Biophys. J.* 100, 183–192. doi:10.1016/j.bpj.2010.11.037

Milojevic, J., Raditsis, A., and Melacini, G. (2009). Human serum albumin inhibits Abeta fibrillization through a "monomer-competitor" mechanism. *Biophys. J.* 97, 2585–2594. doi:10.1016/j.bpj.2009.08.028

Nasica-Labouze, J., Nguyen, P. H., Sterpone, F., Berthoumieu, O., Buchete, N. V., Coté, S., et al. (2015). Amyloid β protein and alzheimer's disease: when computer simulations complement experimental studies. *Chem. Rev.* 115, 3518–3563. doi:10.1021/cr500638n

Nosé, S., and Klein, M. L. (1983). Constant pressure molecular dynamics for molecular systems. *Mol. Phys.* 50, 1055–1076. doi:10.1080/00268978300102851

Páll, S., and Hess, B. (2013). A flexible algorithm for calculating pair interactions on SIMD architectures. *Comput. Phys. Commun.* 184, 2641–2650. doi:10.1016/j.cpc.2013.06.003

Pang, X., and Zhou, H. X. (2015). Disorder-to-order transition of an active-site loop mediates the allosteric activation of sortase a. *Biophys. J.* 109, 1706–1715. doi:10.1016/j.bpj.2015.08.039

Parrinello, M., and Rahman, A. (1981). Polymorphic transitions in single crystals: a new molecular dynamics method. *J. Appl. Phys.* 52, 7182–7190. doi:10.1063/1.328693

Peinado, J. R., Sami, F., Rajpurohit, N., and Lindberg, I. (2013). Blockade of islet amyloid polypeptide fibrillation and cytotoxicity by the secretory chaperones 7B2 and proSAAS. *FEBS Lett.* 587, 3406–3411. doi:10.1016/j.febslet.2013.09.006

Picón-Pagès, P., Bonet, J., García-García, J., Garcia-Buendia, J., Gutierrez, D., Valle, J., et al. (2019). Human albumin impairs amyloid β-peptide Fibrillation through its C-terminus: from docking modeling to protection against neurotoxicity in alzheimer's disease. *Comput. Struct. Biotechnol. J.* 17, 963–971. doi:10.1016/j.csbj.2019.06.017

Pilkington, E. H., Xing, Y., Wang, B., Kakinen, A., Wang, M., Davis, T. P., et al. (2017). Effects of protein corona on IAPP amyloid aggregation, fibril remodelling, and cytotoxicity. *Sci. Rep.* 7, 2455–2513. doi:10.1038/s41598-017-02597-0

Rosenman, D. J., Wang, C., and García, A. E. (2016). Characterization of Aβ monomers through the convergence of ensemble properties among simulations with multiple force fields. *J. Phys. Chem. B* 120, 259–277. doi:10.1021/acs.jpcb.5b09379

Rossetti, G., Musiani, F., Abad, E., Dibenedetto, D., Mouhib, H., Fernandez, C. O., et al. (2016). Conformational ensemble of human α-synuclein physiological form predicted by molecular simulations. *Phys. Chem. Chem. Phys.* 18, 5702–5706. doi:10.1039/c5cp04549e

Rózga, M., Kłoniecki, M., Jabłonowska, A., Dadlez, M., and Bal, W. (2007). The binding constant for amyloid Aβ40 peptide interaction with human serum albumin. *Biochem. Biophys. Res. Commun.* 364, 714–718. doi:10.1016/j.bbrc.2007.10.080

Smith, A. K., Lockhart, C., and Klimov, D. K. (2016). Does replica exchange with solute tempering efficiently sample Aβ peptide conformational ensembles? *J. Chem. Theory Comput.* 12, 5201–5214. doi:10.1021/acs.jctc.6b00660

Song, W., Wang, Y., Colletier, J. P., Yang, H., and Xu, Y. (2015). Varied probability of staying collapsed/extended at the conformational equilibrium of monomeric Aβ40 and Aβ42. *Sci. Rep.* 5, 11024–11013. doi:10.1038/srep11024

Stanyon, H. F., and Viles, J. H. (2012). Human serum albumin can regulate amyloid-β peptide fiber growth in the brain interstitium: implications for alzheimer disease. *J. Biol. Chem.* 287, 28163–28168. doi:10.1074/jbc.C112.360800

Sugio, S., Kashima, A., Mochizuki, S., Noda, M., and Kobayashi, K. (1999). Crystal structure of human serum albumin at 2.5 A resolution. *Protein Eng.* 12, 439–446. doi:10.1093/protein/12.6.439

Sugita, Y., and Okamoto, Y. (1999). Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* 314, 141–151. doi:10.1016/S0009-2614(99)01123-9

Sun, Y., and Ding, F. (2020). αB-crystallin chaperone inhibits Aβ aggregation by capping the β-sheet-rich oligomers and fibrils, *J. Phys. Chem. B.* 124, 10138–10146. doi:10.1021/acs.jpcb.0c07256

Tribello, G. A., Bonomi, M., Branduardi, D., Camilloni, C., and Bussi, G. (2014). PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun.* 185, 604–613. doi:10.1016/j.cpc.2013.09.018

Vivekanandan, S., Brender, J. R., Lee, S. Y., and Ramamoorthy, A. (2011). A partially folded structure of amyloid-beta(1–40) in an aqueous environment. *Biochem. Biophys. Res. Commun.* 411, 312–316. doi:10.1016/j.bbrc.2011.06.133

Wallin, C., Luo, J., Jarvet, J., Wärmländer, S. K. T. S., and Gräslund, A. (2017). The amyloid-β peptide in amyloid formation processes: interactions with blood proteins and naturally occurring metal ions. *Isr. J. Chem.* 57, 674–685. doi:10.1002/ijch.201600105

Wälti, M. A., Ravotti, F., Arai, H., Glabe, C. G., Wall, J. S., Böckmann, A., et al. (2016). Atomic-resolution structure of a disease-relevant Aβ(1–42) amyloid fibril. *Proc. Natl. Acad. Sci. USA* 113, E4976–E4984. doi:10.1073/pnas.1600749113

Wang, C., Cheng, F., Xu, L., and Jia, L. (2016). HSA targets multiple Aβ42 species and inhibits the seeding-mediated aggregation and cytotoxicity of Aβ42 aggregates. *RSC Adv.* 6, 71165–71175. doi:10.1039/c6ra14590f

Wang, L., Friesner, R. A., and Berne, B. J. (2011). Replica exchange with solute scaling: a more efficient version of replica exchange with solute tempering (REST2). *J. Phys. Chem. B* 115, 9431–9438. doi:10.1021/jp204407d

Weng, J., and Wang, W. (2020). Dynamic multivalent interactions of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* 62, 9–13. doi:10.1016/j.sbi.2019.11.001

Williams, A. D., Portelius, E., Kheterpal, I., Guo, J., Cook, K. D., Xu, Y., et al. (2004). Mapping abeta amyloid fibril secondary structure using scanning proline mutagenesis. *J. Mol. Biol.* 335, 833–42. doi:10.1016/j.jmb.2003.11.008

Xiao, Y., Ma, B., McElheny, D., Parthasarathy, S., Long, F., Hoshi, M., et al. (2015). Aβ(1–42) fibril structure illuminates self-recognition and replication of amyloid in alzheimer's disease. *Nat. Struct. Mol. Biol.* 22, 499–505. doi:10.1038/nsmb.2991

Yamamoto, K., Shimada, H., Koh, H., Ataka, S., and Miki, T. (2014). Serum levels of albumin-amyloid beta complexes are decreased in alzheimer's disease. *Geriatr. Gerontol. Int.* 14, 716–723. doi:10.1111/ggi.12147

Zhou, H. X., and Pang, X. (2018). Electrostatic interactions in protein structure, folding, binding, and condensation. *Chem. Rev.* 118, 1691–1741. doi:10.1021/acs.chemrev.7b00305

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Molecular Biosciences

ORIGINAL RESEARCH
published: 04 March 2021
doi: 10.3389/fmolb.2021.627015

The Morphology of Hydroxyapatite Nanoparticles Regulates Cargo Recognition in Clathrin-Mediated Endocytosis

Cheng Zhu[1,2†‡¶], Xuejie Zhou[1†], Ziteng Liu[3], Hongwei Chen[3], Hongfeng Wu[4], Xiao Yang[4], Xiangdong Zhu[4], Jing Ma[3,5‡] and Hao Dong[1,5,6*‡]*

[1]Kuang Yaming Honors School, Nanjing University, Nanjing, China, [2]Tianjin Key Laboratory of Function and Application of Biological Macromolecular Structures, School of Life Sciences, Tianjin University, Tianjin, China, [3]Key Laboratory of Mesoscopic Chemistry of Ministry of Education, Institute of Theoretical and Computational Chemistry, School of Chemistry and Chemical Engineering, Nanjing University, Nanjing, China, [4]National Engineering Research Center for Biomaterials, Sichuan University, Chengdu, China, [5]Nanxin Pharm. Co., Ltd., Nanjing, China, [6]Institute for Brain Sciences, Nanjing University, Nanjing, China

The clathrin-associated protein adaptin-2 (AP2) is a distinctive member of the hetero-tetrameric clathrin adaptor complex family. It plays a crucial role in many intracellular vesicle transport pathways. The hydroxyapatite (HAp) nanoparticles can enter cells through clathrin-dependent endocytosis, induce apoptosis, and ultimately inhibit tumor metastasis. Exploring the micro process of the binding of AP2 and HAp is of great significance for understanding the molecular mechanism of HAp's anti-cancer ability. In this work, we used molecular modeling to study the binding of spherical, rod-shaped, and needle-shaped HAps toward AP2 protein at the atomic level and found that different nanoparticles' morphology can determine their binding specificity through electrostatic interactions. Our results show that globular HAp significantly changes AP2 protein conformation, while needle-shaped HAP has more substantial binding energy with AP2. Therefore, this work offers a microscopic picture for cargo recognition in clathrin-mediated endocytosis, clarifies the design principles and possible mechanisms of high-efficiency nano-biomaterials, and provides a basis for their potential anti-tumor therapeutic effects.

Keywords: nanoparticles, conformational change, endocytosis, cancer, molecular modeling and simulation, adaptin

OPEN ACCESS

Edited by:
Yong Wang,
University of Copenhagen, Denmark

Reviewed by:
Yinghao Wu,
Albert Einstein College of Medicine, United States
Tong Zhu,
East China Normal University, China

****Correspondence:***
Jing Ma
majing@nju.edu.cn
Hao Dong
donghao@nju.edu.cn

[†]These authors have contributed equally to this work

[‡]ORCID:
Cheng Zhu
orcid.org/0000-0003-0260-6287
Jing Ma
orcid.org/0000-0001-5848-9775
Hao Dang
orcid.org/0000-0001-7280-7506

[¶]Cheng Zhu was a short-term visiting scholar in Kuang Yaming Honors School, Nanjing University.

Specialty section:
This article was submitted to Biological Modeling and Simulation, a section of the journal Frontiers in Molecular Biosciences

Received: *07 November 2020*
Accepted: *15 January 2021*
Published: *04 March 2021*

Citation:
Zhu C, Zhou X, Liu Z, Chen H, Wu H, Yang X, Zhu X, Ma J and Dong H (2021) The Morphology of Hydroxyapatite Nanoparticles Regulates Cargo Recognition in Clathrin-Mediated Endocytosis. Front. Mol. Biosci. 8:627015. doi: 10.3389/fmolb.2021.627015

INTRODUCTION

The bioactivity and biocompatibility of nanomaterials, notably their antitumor therapeutic effects, have been the focus of recent medical investigations. Various models systems of nanoparticles have been evaluated for their clinical potentials and applied to cancer cells (Brigger et al., 2012). The hydroxyapatite (HAp) nanoparticles could enter cells through the clathrin-dependent endocytosis process, stimulating mitochondria-dependent apoptosis and eventually suppressing tumor proliferation (Chu et al., 2012). Nanosheets of metal dichalcogenides (WS_2 and MoS_2) were also internalized by epithelial and macrophage cells, colocalized with lysosomes, and induced ferroptotic cell death in the mouse lung tissues (Hao et al., 2017). Calcium peroxide (CaO_2) nanoparticles exhibited dual functions of calcium overload and oxidative stress under tumor microenvironment, inhibiting the tumor growth *in vivo* (Zhang et al., 2019b). Hence, the ability to understand and

manipulate the interplay between nanoparticles and the cellular environment is vital to innovative ways of developing nano-medicines.

Despite the recent advancements in applying nanoparticles against tumor growth or metastasis, the underlying molecular mechanisms, specifically the interactions between nanoparticles and endocytosis systems (clathrin, adaptor proteins, membrane receptors, etc.) have not been fully explored. The key factors that affect nanoparticles' functions as biomedicines, such as particle size, morphology, chemical compositions, etc., remain largely elusive. Nanoparticles enter the cells by endocytosis, and cells regulated the movements of extracellular molecules (including nanomaterials) with coated vesicular carriers. Specifically, clathrin-coated vesicles mediate multiple trafficking routes, including internalization from the plasma membrane (Kovtun et al., 2020). Among the clathrin-associated proteins, adaptin-2 (AP2) is the most abundant endocytic clathrin adaptor and a functional hub linking the cargo molecules and the clathrin cage. AP2 consists of four subunits (α, β2, μ2, and σ2), (Jackson et al., 2010) which can adopt a range of conformations under different conditions. To interact with cargos, the AP2-μ2 subunit undergoes a conformational change from a "locked" state to an "open" state, exposing the YxxΦ motif in the C-terminal region and facilitating recruitment of proteins, lipids or nanoparticles (Traub and Bonifacino, 2013).

We previously characterized the HAp nanoparticles and their inhibitory activities against tumor cells or promoting activities on normal tissue cells, and elucidated the downstream signaling pathways evoked by HAp-internalization and others (Chen et al., 2013; Zhao et al., 2017; Wang et al., 2018a; Zhang et al., 2019a; Liu et al., 2019; Wang et al., 2019; Wu et al., 2019; Li et al., 2020). If the cells were pretreated with chlorpromazine (clathrin-pathway inhibitor) (Shi et al., 2017) or NaN_3 (ATP inhibitor), (Ryan et al., 2015) the uptake of HAp were significantly prohibited, indicating clathrin-mediated and ATP-dependent endocytosis for HAp nanoparticles. The physicochemical properties of HAp as well as its interactions with proteins (such as collagen) or substrates have also been extensively studied at atomic level (Cheng et al., 2017; Wang et al., 2018b; Wang et al., 2018c; Liu et al., 2018; Xie et al., 2018; Gu et al., 2019; Xue et al., 2019; Liu et al., 2020; Ma et al., 2020; Tan et al., 2020; Wang et al., 2020). However, only a few studies have reported the clathrin-mediated adhesion and endocytosis of HAp, (Shi et al., 2017; Shi et al., 2018; Huang et al., 2020) while the molecular details remain elusive. Modeling the dynamic process of HAp binding with AP2 at the atomic level is crucial for a detailed understanding of the driving forces, especially the early events at the nano-biological interface. In the current study, we investigated the interplay of HAp with the upstream pathway of endocytosis to uncover the main factors governing the HAp-AP2 interactions, and to illuminate the design principle of highly efficient nano-biomaterials.

Computational Details

Construction of nanoparticle structural models. The models of hydroxyapatite (sphere, rod, and needle morphology) were built by Materials Studio (Accelrys, 2006). All the structures were optimized using DFT methods at the level of PBE/DND4.4 in the $DMol^3$ package. SwissParam (http://www.swissparam.ch/) was used to generate the CHARMM force field parameters and topology files for each nanoparticle (Zoete et al., 2011). The Mulliken charge was used to generate the electrostatic potential surface. As shown later, all the HAp structures could be well maintained in the following molecular dynamics (MD) simulations, and therefore validates that the parameters can well describe the HAps.

Molecular Dynamics Simulations. The all-atom MD simulations were performed on Gromacs 5.1.1 (Pronk et al., 2013) with CHARMM27 force field (Best et al., 2012) and explicit solvent model TIP3P for water (Jorgensen et al., 1983). The crystal structure of AP2-μ2 subunit (PDB entry: 2XA7. pdb) (Jackson et al., 2010) was adopted as the initial structural model. The missing residues, Q136-Q141 and K224-K235 (both are loops), were built by MODELLER (Fiser et al., 2000).

The initial binding conformations of AP2-HAp complexes were constructed on HADDOCK2.2.(van Zundert et al., 2016). For the initial poses, we also considered the following criteria: 1) biological-relevance. The AP2-μ2 domain was known to mediate cargo binding, while α, β2 domains mostly register μ2 domain in place or interact with the plasma membrane. Within the μ2 domain, the C-terminal region undergoes a significant conformational change upon ligand binding, hence, the C-terminal region is likely enriched with ligand-interacting sites; and 2) compatible with electrostatic potential. As shown in **Supplementary Figure S3**, docking on the full-length AP2 protein or the AP2-μ2 show consistent results. Then, we performed MD simulations using the aforementioned three initial poses, and only the site one gives meaningful results, as described later.

The system was then solvated, neutralized with 150 mM NaCl, minimized using steepest descent method, and then equilibrated for 1 ns at 300 K. Simulated annealing was applied to accelerate the sampling and a typical annealing procedure was: starting from 300 K, the system was heated to 500 K within 5 ns, and then gradually cooled to 300 K within 40 ns, and then to 200 K within another 80 ns. After obtaining a stable configuration of the nanoparticle, MD simulations were carried out to accumulate another 100 ns trajectory. Throughout the simulations, velocity-rescale thermostat and constant pressure (1 bar, Parrinello-Rahman NPT ensemble) were adopted. The nonbonded interaction cut-off for electrostatics calculations was set as 10 Å and the particle mesh Ewald (PME) method was used in the calculation of long-range electrostatic interactions. For each system, two independent simulations were carried out to improve the statistic.

For systems with some residues mutated, we started from the aforementioned systems that have been well equilibrated. Another 10 ns MD simulations were carried out after mutation, and 200 frames were evenly extracted from the trajectory to do the following binding free energy calculations with MM/PBSA.

The principal component calculations. To detect possible transition between the open and the locked configurations of the AP2 μ2 domain, we calculated a principal component, λ,

along the vector connecting the two states (**Eq. 1**). $\lambda = 0$ refers to the open state configuration, and $\lambda = 1$ refers to the locked state. For each sampled structure, the value represents the specific state of the structure.

$$\lambda = \frac{\left|v(x,y,z) - v1(x,y,z)\right|}{\left|v2(x,y,z) - v1(x,y,z)\right|} \quad (1)$$

λ is the value of the principal component; v is the position of the sampled structure, $v1$ and $v2$ are the positions of the reference structures 1 (the open state) and 2 (the locked state), respectively.

The matrix of distance fluctuation**.** To monitor the structural dynamics and plasticity of the protein, we calculated the matrix of distance fluctuation according to **Eq. 2**:

$$\tau_{ij} = \left(d_{ij} - d_{ij}\right)^2 \quad (2)$$

where d_{ij} is the time dependent distance between atoms i and j, and bracket represents the time average. Clearly, τ_{ij} is independent of the reference structure.

Binding free energy calculations. The Gromacs tool "g_mmpbsa" (Kumari et al., 2014) was used for calculations. For each system, 101 representative frames were evenly extracted from the 100 ns MD trajectory in the production phase. The MM/PBSA method calculates the binding free energy of the protein with ligand in solvent according with **Eq. 3**:

$$\Delta G_{binding} = G_{complex} - \left(G_{protein} + G_{ligand}\right) \quad (3)$$

$G_{complex}$, $G_{protein}$, and G_{ligand} are total free energies of the protein-ligand complex, isolated protein and isolated ligand in solvent, respectively. These terms were calculated according to **Eq. 4**, individually:

$$G_x = \langle E_{MM} \rangle - TS + \langle G_{solvation} \rangle \quad (4)$$

where G_x represents the free energy of isolated protein, or isolated ligand, or protein-ligand complex. T is the temperature, and S represents the entropy in vacuum. $\langle G_{solvation} \rangle$ is the solvation free energy. $\langle E_{MM} \rangle$ is the average molecular mechanics (MM) potential energy in vacuum, including the contributions from bonded and non-bonded interactions determined by the force field parameters (**Eq. 5**):

$$E_{MM} = E_{bonded} + E_{vdW} + E_{elec} \quad (5)$$

where E_{bonded} is the bonded interaction energies consisting of bond, angle, dihedral and improper interactions. E_{vdW} and E_{elec} are modeled using a Coulomb and Lennard-Jones (LJ) potential function, respectively. $\langle G_{solvation} \rangle$ is the energy required to transfer a solute from vacuum into the solvent, which is expressed as **Eq. 6**:

$$G_{solvation} = G_{polar} + G_{nonpolar} \quad (6)$$

where, G_{polar} is estimated by solving the Poisson-Boltzmann (PB) equation, which is the electrostatic contribution. $G_{nonpolar}$ is estimated through solvent accessible surface area (SASA) model to get the non-electrostatic contribution. Through g_mmpbsa, E_{MM}, G_{polar}, and $G_{nonpolar}$ are calculated individually, which means the binding energy could be decomposed on a per residue basis as **Eq. 7**:

$$\Delta R_x^{BE} = \sum_{i=0}^{n}\left(A_i^{complex} - A_i^{free}\right) \quad (7)$$

ΔR_x^{BE} is the contribution to the binding energy of residue x. $A_i^{complex}$ is the energy of the ith atom on x residue in bound state and A_i^{free} is the energy in the unbound state.

RESULTS AND DISCUSSION

Morphology of HAp

During the adsorption of proteins on a solid surface, there are several main driving forces, including electrostatic and hydrophobic interactions, as well as the structural rearrangement of the adsorbed proteins (Dee et al., 2003). Due to the highly charged nature of HAps, the electrostatic interaction has been confirmed to strongly affect their affinities toward proteins through charge-charge and charge-dipole interactions (Zhu et al., 2007; Zhu et al., 2010; Chen et al., 2014; Wang et al., 2014). For the HAps with a similar shape (for example, the spherical one) at a given ratio between Ca and P atoms, our calculations show that the difference in their sizes results in certain differences in the electrostatic potential distribution of the surface (**Figure 1A**). However, the surface charge density of different spheres is close to a constant value, implying a unique electrostatic property at the surface (**Figure 1B**). Therefore, the HAp-S with the diameter of 1.926 nm was used in the following calculations. On the other hand, changes of the nanoparticles' geometrical parameters, such as the Ca:P ratio or the aspect ratio, may lead to variable net-charge and the electrostatic potential surface (**Supplementary Figure S1**), though it was reported that the zeta potentials of HAp nanoparticles of various shape that have been characterized in experiments are all negative (Wu et al., 2019); besides, changes in the size may also lead to a change in the geometric matching of the interface between the two. Therefore, we will focus on the interactions between the fixed-size nanoparticles (**Figure 1C,D**) and the protein in this work.

Three different morphologies of HAps were examined in this work, which were named as HAp-S (sphere), HAp-R (rod), and HAp-N (needle), respectively. The three systems have distinct electrostatic potential surfaces (EPS) due to different total charges and the charged groups exposed on the surfaces (**Figure 1**): HAp-S and HAp-N are dominated by negatively charged distributions, while HAp-R is mostly positively charged. More interestingly, the EPSs are not evenly distributed among the three: on HSP-S, the weak positive and strong negative charges are distributed alternately on the surface; on HSP-N, both ends of the needle-like structure are positively charged, whereas the remaining part is mainly negatively charged; the HSP-R features a weak positively charged EPS. Seemingly, the difference in electrostatic potential distribution on HAps' surface indicates their different abilities to interact with the target protein AP2.

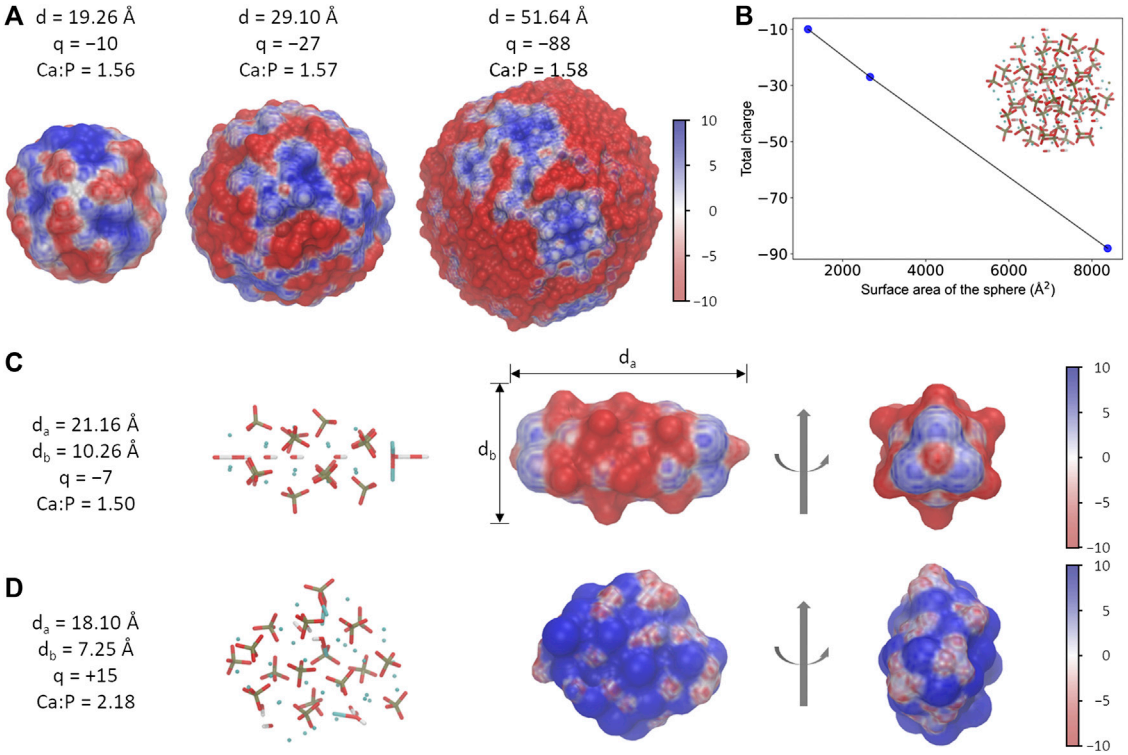


FIGURE 1 | Different morphology of HAp nanoparticles. **(A)** The spherical-shaped HAps with different sizes show slightly different electrostatic potential surfaces. Particles of various sizes are only schematic diagrams and do not correspond to the real scale. **(B)** The ratio between the total charge and the surface area is nearly a constant in HAps. **(C, D)** Information of the needle **(C)** and the rod **(D)** shaped HAps. The unit for electrostatic potential is k_BT/e.

Binding of HAp on AP2-μ2

We then examined the effects of particle morphologies (HAp-S, -R, and -N) on the interactions between endocytic adaptor protein AP2 and HAp nanoparticles. Due to the relatively large size of the AP2 complex and the major role of the μ2 subunit as a cargo-binding region, we focus on the protein structural preservation and alteration of the μ2 subunit (denoted as AP2-μ2) upon the binding of HAp in the following simulations.

To our best knowledge, no structural information of the complex formed by the HAp and AP2-μ2 is available. Thus, the initial guess about the interface between the two components was derived from rigid-body docking and the matching of electrostatic potential surfaces. Presumably, polar residues K311, E313, K315, E380, and E382 (**Figures 2**, **3**) at the β-sheet-rich C-terminal side of the AP2-μ2 (represented as AP2-μ2-C, residue ID 172-443) mediates the binding, which was designated as binding sites I. It should be noted that this segment is originally deeply buried in the bowl of AP2 core arranged by the α, β2, μ2, and σ2 subunits when it is in a locked state, but is released from the bowl and rotates roughly about its long axis when it is in an open state. Molecular dynamics (MD) simulations show that HAp-S and HAp-N have specific binding sites on AP2-μ2 and consequently cause different conformational changes of the protein.

HAp-S bound to AP2-μ2 tightly at the binding site I throughout the 100 ns simulations. The positively charged residues K311 on the β-sheet-rich C-terminal side of AP2-μ2 retains close interactions with the exposed phosphate groups on the surface of HAp-S. E380 forms stable interactions with an exposed Ca^{2+} on the surface of HAp-S (**Figure 2**). The HAps well maintain the spherical structure without significant change of its morphology (**Supplementary Figure S2**), and no release of Ca^{2+} ions was observed within the simulated time scale. Notably, the binding site I is close to the experimentally-identified binding pocket of the μ2-subunit-specific YxxΦ-type-binding motif, which is functionally conserved for AP2 (Kittler et al., 2008).

In contrast, the HAp-N migrated away from the initial binding site I and settled on a pocket (binding site II, **Figure 2**) formed by α-helices and a number of spatial surrounding amino acids at the α-helix-rich N-terminal region (represented as AP2-μ2-N, residue ID 1-121). One possible reason for the detachment of HAp-N from the binding site I could be attributed to the relatively flat surface at the β-sheet-rich domain. Notably, the binding site II is not fully exposed in the AP2 adaptor complex formed by the α, β2, μ2, and σ2 subunits, but has some overlap with the interface between the μ2 and β2 subunits. In the case of HAp-R, it rapidly detached from AP2-μ2 (5ns after constraint release in the simulations) and remained distant from the protein.

To explore the AP2 conformation and HAp binding sampled in the MD simulations with further details, we clustered the

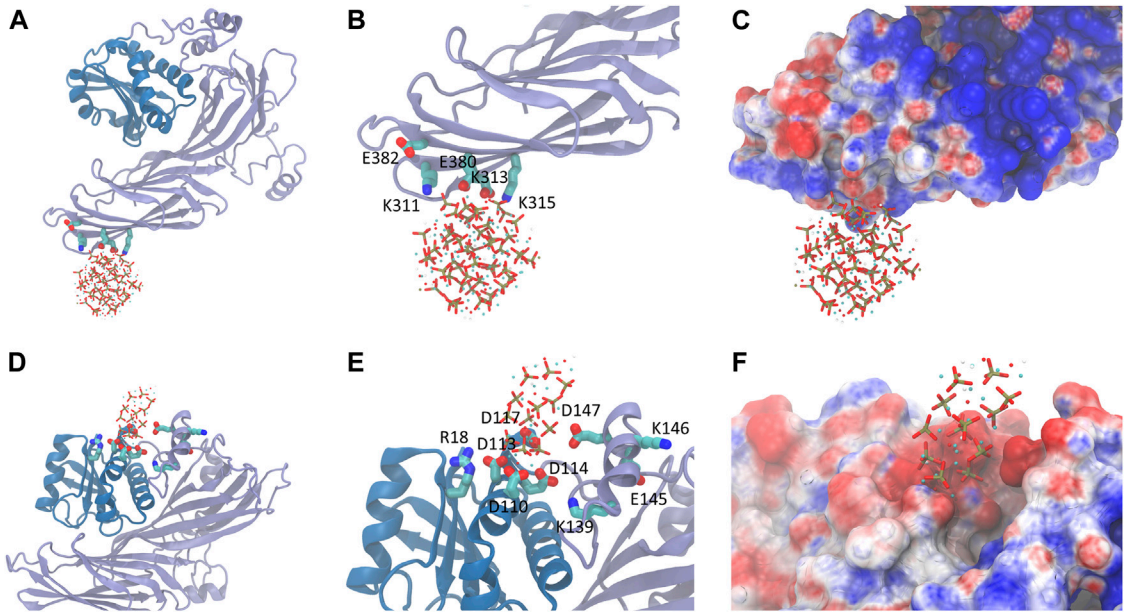


FIGURE 2 | The bound state of HAp-S **(top panel)** or HAp-N **(bottom panel)** on AP2-μ2. **(A)** The binding of HAp-S at the β-sheet-rich C-terminal side of AP2. **(B)** The key residues involved in HAp-S binding. **(C)** The electrostatic potential distribution at the binding site of the HAp-S/AP2-μ2 complex. **(D–F)** Information for the binding of HAp-N. The binding of HAp-N is at the α-helix -rich N-terminal side of AP2.

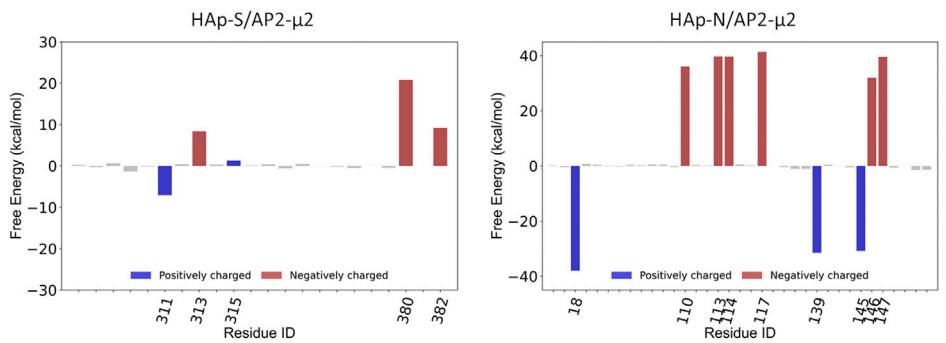


FIGURE 3 | The key residues on AP2-μ2 contributed to the binding of HAp-S **(left panel)** and HAp-N **(right panel)**.

trajectories into six groups by using 1.5 Å root-mean-square deviation (RMSD) as the criterion in each system. Notably, in both AP2-μ2/HAp-S and AP2-μ2/HAp-N complexes, the first three ethnic groups account for more than 85% of all the MD trajectories. Therefore, we calculated the binding free energy with an implicit solvation model, the molecular mechanics/Poisson-Boltzmann surface area (MM/PBSA) method (Kollman et al., 2000; Baker et al., 2001).

Compared with the AP2-μ2/HAp-S complex, the AP2-μ2/HAp-N has a much larger contact area. The total binding free energy indicates that the AP2-μ2 has a much stronger binding affinity with HAp-N than that with HAP-S. Among different interactions, the electrostatic interaction is dominant in both systems, evidenced by several conserved key residues mediating the majority of contributions toward binding.

To further validate the contributions from the charged residues on HAps toward the binding affinity, we also studied the AP2-μ2 with some key residues mutated. We carried out two sets of mutations on the AP2-μ2/HAp-X (X = N or S) complex (**Table 1**): in the first set, only a single mutation on AP2-μ2 was made, where the residue contributing most to the binding was replaced with a neutral residue alanine. The resulting mutants are AP2-μ2-D117A/HAp-N and AP2-μ2-E380A/HAp-S; in the second set, simultaneous mutations at four important sites toward binding were made. The resulting mutants are AP2-μ2-E110A-D113A-Q143A-E147A/HAp-N and AP2-μ2-R305A-K311A-K315A-E380A/HAp-S. Our data show that

TABLE 1 | The binding free energies of the AP2-μ2/HAp-X (X = N or S) complex calculated with MM/PBSA (in kcal/mol).

	HAp-N/AP2-μ2			HAp-S/AP2-μ2		
	Wild type	**D117A**	**E110A, D113A, Q143A, E147A**	**Wild type**	**E380A**	**R305A, K311A, K315A, E380A**
E_{vdw}	9.26 ± 5.20	12.35 ± 4.62	15.60 ± 5.56	−3.31 ± 4.39	−1.32 ± 3.70	0.53 ± 3.61
E_{elec}	−209.75 ± 18.67	−304.22 ± 10.57	−375.56 ± 36.64	−109.95 ± 14.99	744.56 ± 36.84	582.17 ± 20.70
G_{polar}	−54.22 ± 15.17	211.44 ± 12.45	224.03 ± 27.80	99.48 ± 39.49	88.49 ± 15.39	52.67 ± 25.81
$G_{nonpolar}$	−2.49 ± 0.20	−1.66 ± 0.11	−1.29 ± 0.28	−2.00 ± 0.47	−1.50 ± 0.48	−1.16 ± 0.36
$G_{binding}$	−257.20 ± 22.54	−82.07 ± 13.15	−137.22 ± 14.42	−15.78 ± 29.13	830.23 ± 37.76	635.64 ± 27.14

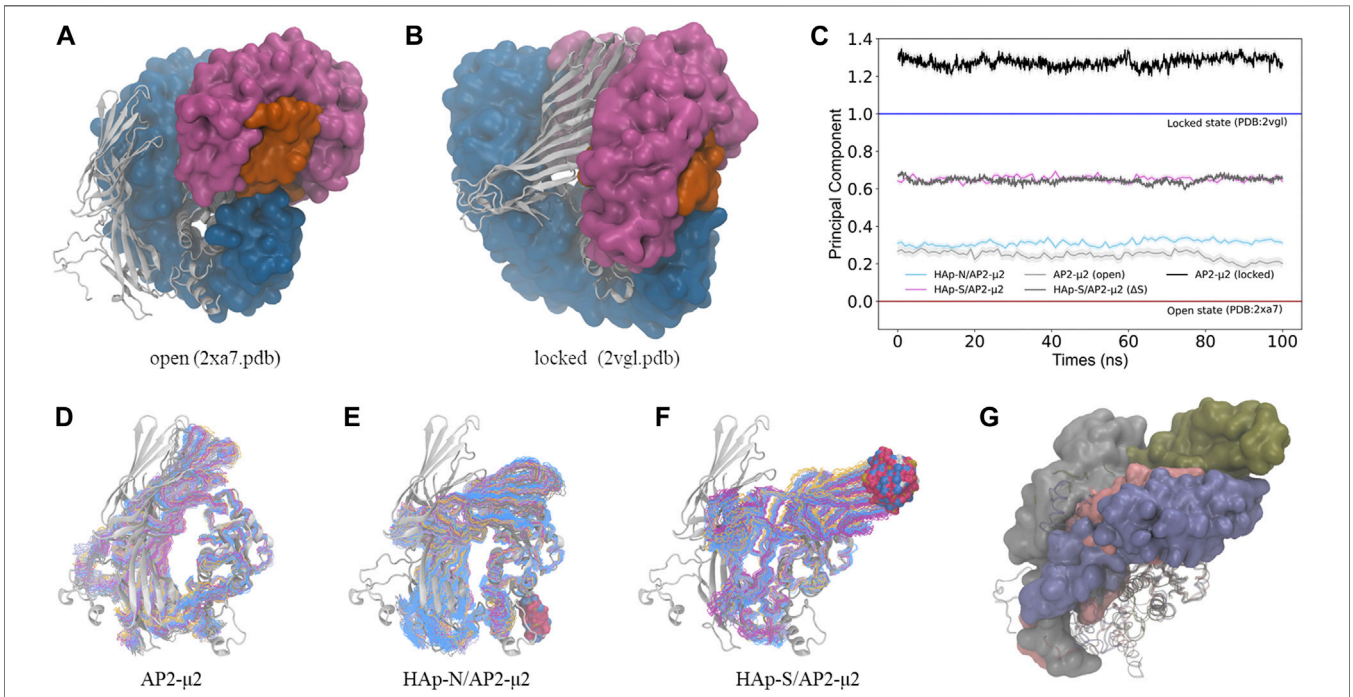


FIGURE 4 | The conformational change of the AP2-μ2 protein upon binding of HAp nanoparticles. **(A, B)** The open **(A)** and the locked **(B)** states of the AP2 protein. The α **(magenta)**, β2 **(blue)**, and σ2 **(orange)** subunits are in the surface mode, and the μ2 subunit **(white)** is in the cartoon mode. **(C)** The calculated principal component of the conformations between the open (λ = 0) and the locked (λ = 1) states of the AP2-μ2 protein sampled with MD simulations. **(D–F)** The sampled conformations (in the line mode) of the AP2-μ2 alone **(D)**, in the bound state with HAp-N **(E)**, or with HAp-S **(F)**, where the AP2-μ2 in the open state is taken from the crystal structure (in the cartoon mode) and shown as the reference. In each ensemble, different colors are used for different clusters of the sampled structures, and the bound HAp particles are shown in the space-filling mode. **(G)** The representative structure of AP2-μ2 in the HAp-N/AP2-μ2 **(pink)**, HAp-S/AP2-μ2 **(blue)** complexes, where the open **(gray)** and locked **(green)** structures are also shown for comparison. The AP2-μ2-C subunit is shown in the surface mode to illustrate the direction of conformational change, and the AP2-μ2-N subunit is shown in the cartoon mode. All the structures are in a view pre-aligned with the AP2-μ2-N subunit. For clarity, the bound HAps are not shown.

mutation of the key residue(s) dramatically reduces the binding between the AP2-μ2 and HAp-X, further illustrating the electrostatic interaction nature between the two.

Seemingly, HAps with different morphologies have binding specificity toward AP2-μ2. It should be noted that the inherent limitations of MM/PBSA, as in the lack of implicit information about solvent water around the binding site and large fluctuations of conformational entropy (Hou et al., 2011) hinders the precise calculation of the absolute binding energies. However, given the computational efficiency of this method, the estimated relative values of binding provide meaningful information to distinguish the binding of different HAps.

HAp-Binding Induced Conformational Change of AP2-μ2

To monitor the conformational change of AP2-μ2 upon binding of HAp, we calculated the principal component, λ, of the complex trajectory along a vector connecting the open (Jackson et al., 2010) (PDB entry: 2xa7. pdb, λ = 0) and the locked (Collins et al., 2002) (PDB entry: 2vgl.pdb, λ = 1) states of AP2-μ2 (**Figure 4**) determined by x-ray crystallography. In the absence of HAp, the conformations of AP2-μ2 generated with MD simulations resemble the open state crystal structure, as indicated by the λ values fluctuating just above 0, showing the stability of open state conformation. Presumably, the small deviation (with the mean

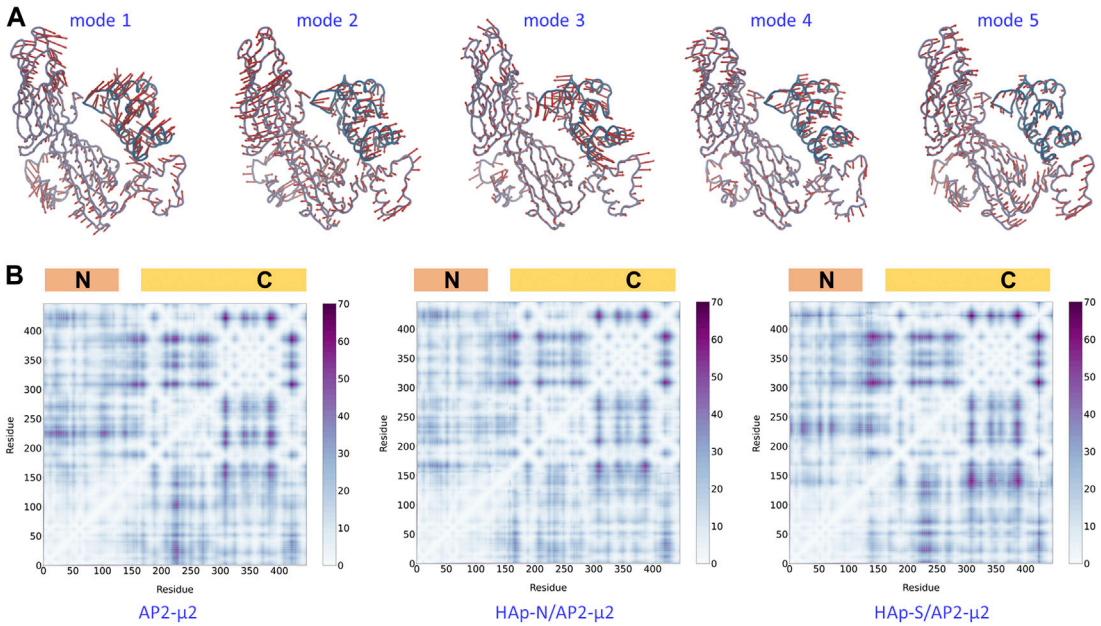


FIGURE 5 | The dynamics and plasticity of AP2-μ2. Top panel: the first five low-frequency modes of collective motion generated with the normal mode analysis on the open state crystal structure. Bottom panel: the matrix of distance fluctuations of the AP2-μ2 protein in the absence **(left panel)** and the presence of HAp-N **(middle panel)** or HAp-S **(right panel)** in the MD simulations. The N- and C-terminal domains are labeled on the top of each panel. The color bar shows the magnitude of fluctuations. Only Cα atoms were used in both calculations.

value of λ = 0.24) could be attributed to the lack of anchoring subunits α, β2, and σ2 in the present MD simulations.

In contrast, the binding of HAp leads to a certain degree of conformational changes of AP2-μ2. Specifically, AP2-μ2 with both nanoparticles tends to transform to the locked state, as indicated by the value of λ deviating from 0 toward the direction of 1. For the AP2-μ2/HAp-S complex, the average value of λ is ~0.65, much larger than that of the AP2-μ2/HAp-N complex (~0.32), demonstrating a more significant conformational change induced by the bound of spherical nanoparticle (HAp-S) than the needle one (HAp-N). In other words, the HAp-N bound AP2-μ2 protein is closer to its open state, while the HPA-S bound one is more comparable to its locked state (**Figure 4**).

By comparing the locked and the open state structures (**Figure 4**), it is clear that the domain rearrangement from the open to the locked state could be described as the rotation of the AP2-μ2-C along its long axis, as well as the approaching of the AP2-μ2-C toward the AP2-μ2-N, as the AP2-μ2-N is relatively fixed in the bowl. We then studied the collective motion of the open state structure of AP2-μ2 with normal mode analysis (**Figure 5A**), assuming that the system is stabilized by harmonic potentials. The first five low-frequency modes show that relative motions between the C- and N-terminal segments within AP2-μ2 are dominant, indicating the functional role of these intrinsic motions.

The AP2-μ2 experiences different conformational changes in the presence of different HAPs (**Figure 4**). Presumably, the synergistic effect of the protein plasticity has an impact on the binding affinity of HAp. Therefore, we monitored the matrix of distance fluctuations of the AP2-μ2 protein (**Figure 5B**). Taking the two domains in the AP2-μ2 as the reference, the binding of HAp-N induces less fluctuations at the N-domain, which is likely to be attributed to its strong binding affinities toward the N-terminal domain (**Table 1**). Seemingly, the presence of HAp-N locks AP2-μ2 at a certain configuration. In contrast, the binding of HAp-S triggers much larger fluctuations at the C-domain, which is likely to be related to its weak binding affinity. Therefore, HAp-S leads to a more significant conformational change on AP2-μ2 than that of HAp-N (**Figure 4**).

In the meantime, the binding of the negatively charged HAp partially neutralizes the net charge on AP2-μ2, which is likely to reduce the electrostatic repulsion between the AP2-μ2-N and AP2-μ2-C segments, and therefore facilitate the relative motion between the two. Consequently, HAp-S favors AP2-μ2-C surfaces, while HAp-N is attracted to AP2-μ2-N. Compared to HAp-N, HAp-S possess higher-densities of surface charges and interacts with a relatively exposed protein area (the binding site I). Hence, HAp-S leads to a more significant conformational change on AP2-μ2 than that of HAp-N. In literature, it was reported that the sphere-shaped HAp nanoparticles effectively inhibited the growth of A375 melanoma cells (34.90% viability); in contrast, the rod or needle-like HAp nanoparticles moderately affected the viabilities of melanoma cells (60.43%–74.90%) (Wu et al., 2019). Presumably, HAp-S is likely to facilitate its transportation in the cellular environment by shifting AP2-μ2 to a locked conformation, and therefore resulting in a more profound tumor-suppressive effect.

CONCLUSION

In this work, we used molecular dynamics simulations to study the regulation of nanoparticles HAp with the clathrin adaptor AP2. We found that the different morphologies of HAps feature distinct binding affinities toward AP2; the binding of HAps with different morphology leads to structurally and functionally distinct configurations of AP2, which is likely to affect cargo recognition in clathrin-mediated endocytosis profoundly. Our work offers a microscopic explanation for cargo recognition in clathrin-mediated endocytosis and possible mechanisms of designing high-efficiency nano-biomaterials, thus providing a basis for understanding their specificity and potential as intracellular agents.

DATA AVAILABILITY STATEMTENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

HD and JM designed and directed the project. CZ and XJZ performed computations. ZTL and HWC built the nanostructures. HFW, XY and XDZ provided critical feedback and helped shape the research. All authors discussed the results and contributed to the final manuscript.

FUNDING

This work was supported by the National Key Research and Development Program of China (Grand No. 2017YFB0702600, 2017YFB0702601), the National Natural Science Foundation of China (Grant No. 21773115, 21833002, 22077094, 22007071, 22033004, 21873045), the Natural Science Foundation of Jiangsu Province (Grant No. BK20190056), Jiangsu Graduate Research and Innovation Program (KYCX20_0057), and the "Fundamental Research Funds for the Central Universities" (Grand No. 021514380014). Parts of the calculations were performed using computational resources on an IBM Blade cluster system from the High-Performance Computing Center (HPCC) of Nanjing University.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2021.627015/full#supplementary-material.

REFERENCES

Accelrys, I. (2006). *Materials Studio version 4.0*, San Diego, CA: Biovia.

Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U.S.A.* 98 (18), 10037–10041. doi:10.1073/pnas.181342398

Best, R. B., Zhu, X., Shim, J., Lopes, P. E., Mittal, J., Feig, M., et al. (2012). Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone φ, ψ and side-chain χ(1) and χ(2) dihedral angles. *J. Chem. Theory. Comput.* 8 (9), 3257–3273. doi:10.1021/ct300400x

Brigger, I., Dubernet, C., and Couvreur, P. (2012). Nanoparticles in cancer therapy and diagnosis. *Adv. Drug Delivery Rev.* 64, 24–36. doi:10.1016/j.addr.2012.09.006

Chen, X.-Q., Chen, X.-N., Zhu, X.-D., Cai, B., Fan, H.-S., and Zhang, X.-D. (2013). Effect of surface topography of hydroxyapatite on human osteosarcoma MG-63 cell. *J. Inorg. Mater.* 28 (8), 901–906. doi:10.3724/sp.J.1077.2013.13058

Chen, Y., Wang, J., Zhu, X., Fan, Y., and Zhang, X. (2014). Adsorption and release behaviors of vascular endothelial growth factor on porous hydroxyapatite ceramic under competitive conditions. *J. Biomater. Tissue Eng.* 4 (2), 155–161. doi:10.1166/jbt.2014.1144

Cheng, X., Wu, H., Zhang, L., Ma, X., Zhang, X., and Yang, M. (2017). Hydroxyl migration disorders the surface structure of hydroxyapatite nanoparticles. *Appl. Surf. Sci.* 416, 901–910. doi:10.1016/j.apsusc.2017.04.250

Chu, S. H., Feng, D. F., Ma, Y. B., and Li, Z. Q. (2012). Hydroxyapatite nanoparticles inhibit the growth of human glioma cells *in vitro* and *in vivo*. *Int. J. Nanomed.* 7, 3659–3666. doi:10.2147/ijn.S33584

Collins, B. M., McCoy, A. J., Kent, H. M., Evans, P. R., and Owen, D. J. (2002). Molecular architecture and functional model of the endocytic AP2 complex. *Cell* 109 (4), 523–535. doi:10.1016/s0092-8674(02)00735-3

Dee, K. C., Puleo, D. A., and Bizios, R. (2003). *An introduction to tissue-biomaterial interactions.* New York, NY: John Wiley and Sons.

Fiser, A., Do, R. K., and Sali, A. (2000). Modeling of loops in protein structures. *Protein Sci.* 9 (9), 1753–1773. doi:10.1110/ps.9.9.1753

Gu, H., Xue, Z., Wang, M., Yang, M., Wang, K., and Xu, D. (2019). Effect of hydroxyapatite surface on BMP-2 biological properties by docking and molecular simulation approaches. *J. Phys. Chem. B* 123 (15), 3372–3382. doi:10.1021/acs.jpcb.9b01982

Hao, J., Song, G., Liu, T., Yi, X., Yang, K., Cheng, L., et al. (2017). *In Vivo* long-term biodistribution, excretion, and toxicology of PEGylated transition-metal dichalcogenides MS2 (M = Mo, W, Ti) nanosheets. *Adv. Sci.* 4 (1). doi:10.1002/advs.201600160

Hou, T., Wang, J., Li, Y., and Wang, W. (2011). Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. *J. Chem. Inf. Model.* 51 (1), 69–82. doi:10.1021/ci100275a

Huang, L. H., Han, J., Ouyang, J. M., and Gui, B. S. (2020). Shape-dependent adhesion and endocytosis of hydroxyapatite nanoparticles on A7R5 aortic smooth muscle cells. *J. Cell. Physiol.* 235 (1), 465–479. doi:10.1002/jcp.28987

Jackson, L. P., Kelly, B. T., McCoy, A. J., Gaffry, T., James, L. C., Collins, B. M., et al. (2010). A large-scale conformational change couples membrane recruitment to cargo binding in the AP2 clathrin adaptor complex. *Cell* 141 (7), 1220–1229. doi:10.1016/j.cell.2010.05.006

Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79 (2), 926–935. doi:10.1063/1.445869

Kittler, J. T., Chen, G., Kukhtina, V., Vahedi-Faridi, A., Gu, Z., Tretter, V., et al. (2008). Regulation of synaptic inhibition by phospho-dependent binding of the AP2 complex to a YECL motif in the GABAA receptor gamma2 subunit. *Proc. Natl. Acad. Sci. U.S.A.* 105 (9), 3616–3621. doi:10.1073/pnas.0707920105

Kollman, P. A., Massova, I., Reyes, C., Kuhn, B., Huo, S., Chong, L., et al. (2000). Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res.* 33 (12), 889–897. doi:10.1021/ar000033j

Kovtun, O., Dickson, V. K., Kelly, B. T., Owen, D. J., and Briggs, J. A. G. (2020). Architecture of the AP2/clathrin coat on the membranes of clathrin-coated vesicles. *Sci Adv* 6 (30), eaba8381. doi:10.1126/sciadv.aba8381

Kumari, R., Kumar, R., Lynn, A., and Lynn, A. (2014). g_mmpbsa--a GROMACS tool for high-throughput MM-PBSA calculations. *J. Chem. Inf. Model.* 54 (7), 1951–1962. doi:10.1021/ci500020m

Li, Z., Tang, J., Wu, H., Ling, Z., Chen, S., Zhou, Y., et al. (2020). A systematic assessment of hydroxyapatite nanoparticles used in the treatment of melanoma. *Nano Res.* 13 (8), 2106–2117. doi:10.1007/s12274-020-2817-6

Liu, D., Wu, Y., Wu, H., Li, X., Yang, X., Zhu, X., et al. (2019). Effect of process parameters on the microstructure and property of hydroxyapatite precursor powders and resultant sintered bodies. *Int. J. Appl. Ceram. Technol.* 16 (2), 444–454. doi:10.1111/ijac.13097

Liu, Q., Xue, Z., and Xu, D. (2020). Molecular dynamics characterization of Sr-doped biomimetic hydroxyapatite nanoparticles. *J. Phys. Chem. C* 124 (36), 19704–19715. doi:10.1021/acs.jpcc.0c06391

Liu, Z., Zhang, L., Wang, X., Jiang, G., and Yang, M. (2018). A comparative study of the dissolubility of pure and silicon substituted hydroxyapatite from density functional theory calculations. *J. Mol. Model.* 24 (7). doi:10.1007/s00894-018-3708-6

Ma, X., Zhang, L., and Yang, M. (2020). Interfacial interaction and its influence on the mechanical performances of hydroxyapatite through a polycrystalline model. *Phys. B Condens. Matter* 594, 412338. doi:10.1016/j.physb.2020.412338

Pronk, S., Páll, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., et al. (2013). Gromacs 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* 29 (7), 845–854. doi:10.1093/bioinformatics/btt055

Ryan, A. J., Gleeson, J. P., Matsiko, A., Thompson, E. M., and O'Brien, F. J. (2015). Effect of different hydroxyapatite incorporation methods on the structural and biological properties of porous collagen scaffolds for bone repair. *J. Anat.* 227 (6), 732–745. doi:10.1111/joa.12262

Shi, X., Zhou, K., Huang, F., and Wang, C. (2017). Interaction of hydroxyapatite nanoparticles with endothelial cells: internalization and inhibition of angiogenesis *in vitro* through the PI3K/Akt pathway. *Int. J. Nanomedicine* 12, 5781–5795. doi:10.2147/ijn.S140179

Shi, X., Zhou, K., Huang, F., Zhang, J., and Wang, C. (2018). Endocytic mechanisms and osteoinductive profile of hydroxyapatite nanoparticles in human umbilical cord Wharton's jelly-derived mesenchymal stem cells. *Int. J. Nanomedicine* 13, 1457–1470. doi:10.2147/ijn.S155814

Tan, X., Xue, Z., Zhu, H., Wang, X., and Xu, D. (2020). How charged amino acids regulate nucleation of biomimetic hydroxyapatite nanoparticles on the surface of collagen mimetic peptides: molecular dynamics and free energy investigations. *Cryst. Growth Des.* 20 (7), 4561–4572. doi:10.1021/acs.cgd.0c00353

Traub, L. M., and Bonifacino, J. S. (2013). Cargo recognition in clathrin-mediated endocytosis. *Cold Spring Harb. Perspect. Biol.* 5 (11), a016790. doi:10.1101/cshperspect.a016790

van Zundert, G. C. P., Rodrigues, J. P. G. L. M., Trellet, M., Schmitz, C., Kastritis, P. L., Karaca, E., et al. (2016). The HADDOCK2.2 web server: user-friendly integrative modeling of biomolecular complexes. *J. Mol. Biol.* 428 (4), 720–725. doi:10.1016/j.jmb.2015.09.014

Wang, J., Chen, Y., Zhu, X., Yuan, T., Tan, Y., Fan, Y., et al. (2014). Effect of phase composition on protein adsorption and osteoinduction of porous calcium phosphate ceramics in mice. *J. Biomed. Mater. Res. A* 102 (12), 4234–4243. doi:10.1002/jbm.a.35102

Wang, J., Zhu, Y., Wang, M., Liu, D., Chen, X., Zhu, X., et al. (2018a). Fabrication and preliminary biological evaluation of a highly porous biphasic calcium phosphate scaffold with nano-hydroxyapatite surface coating. *Ceram. Int.* 44 (2), 1304–1311. doi:10.1016/j.ceramint.2017.08.053

Wang, X., Zhang, L., Liu, Z., Zeng, Q., Jiang, G., and Yang, M. (2018b). Probing the surface structure of hydroxyapatite through its interaction with hydroxyl: a first-principles study. *RSC Adv.* 8 (7), 3716–3722. doi:10.1039/c7ra13121f

Wang, X., Zhang, L., Zeng, Q., Jiang, G., and Yang, M. (2018c). First-principles study on the hydroxyl migration from inner to surface in hydroxyapatite. *Appl. Surf. Sci.* 452, 381–388. doi:10.1016/j.apsusc.2018.05.050

Wang, J., Wang, M., Chen, F., Wei, Y., Chen, X., Zhou, Y., et al. (2019). Nano-hydroxyapatite coating promotes porous calcium phosphate ceramic-induced osteogenesis via BMP/smad signaling pathway. *Int. J. Nanomedicine* 14, 7987–8000. doi:10.2147/ijn.S216182

Wang, X., Wu, H., Cheng, X., Yang, M., and Zhang, L. (2020). Probing the surface activity of hydroxyapatite nanoparticles through their interaction with water molecules. *AIP Adv.* 10 (6), 065217. doi:10.1063/5.0010750

Wu, H., Li, Z., Tang, J., Yang, X., Zhou, Y., Guo, B., et al. (2019). The *in vitro* and *in vivo* anti-melanoma effects of hydroxyapatite nanoparticles: influences of material factors. *Int. J. Nanomedicine* 14, 1177–1191. doi:10.2147/ijn.S184792

Xie, Q., Xue, Z., Gu, H., Hu, C., Yang, M., Wang, X., et al. (2018). Molecular dynamics exploration of ordered-to-disordered surface structures of biomimetic hydroxyapatite nanoparticles. *J. Phys. Chem. C* 122 (12), 6691–6703. doi:10.1021/acs.jpcc.8b00178

Xue, Z., Yang, M., and Xu, D. (2019). Nucleation of biomimetic hydroxyapatite nanoparticles on the surface of type I collagen: molecular dynamics investigations. *J. Phys. Chem. C* 123 (4), 2533–2543. doi:10.1021/acs.jpcc.8b10342

Zhang, K., Zhou, Y., Xiao, C., Zhao, W. L., Wu, H. F., Tang, J. Q., et al. (2019a). Application of hydroxyapatite nanoparticles in tumor-associated bone segmental defect. *Science Advances* 5 (8), 16. doi:10.1126/sciadv.aax6946

Zhang, M., Song, R., Liu, Y., Yi, Z., Meng, X., Zhang, J., et al. (2019b). Calcium-overload-mediated tumor therapy by calcium peroxide nanoparticles. *Inside Chem.* 5 (8), 2171–2182. doi:10.1016/j.chempr.2019.06.003

Zhao, R., Xie, P., Zhang, K., Tang, Z., Chen, X., Zhu, X., et al. (2017). Selective effect of hydroxyapatite nanoparticles on osteoporotic and healthy bone formation correlates with intracellular calcium homeostasis regulation. *Acta Biomater.* 59, 338–350. doi:10.1016/j.actbio.2017.07.009

Zhu, X. D., Fan, H. S., Zhao, C. Y., Lu, J., Ikoma, T., Tanaka, J., et al. (2007). Competitive adsorption of bovine serum albumin and lysozyme on characterized calcium phosphates by polyacrylamide gel electrophoresis method. *J. Mater. Sci. Mater. Med.* 18 (11), 2243–2249. doi:10.1007/s10856-006-0057-2

Zhu, X. D., Zhang, H. J., Fan, H. S., Li, W., and Zhang, X. D. (2010). Effect of phase composition and microstructure of calcium phosphate ceramic particles on protein adsorption. *Acta Biomater.* 6 (4), 1536–1541. doi:10.1016/j.actbio.2009.10.032

Zoete, V., Cuendet, M. A., Grosdidier, A., and Michielin, O. (2011). SwissParam: a fast force field generation tool for small organic molecules. *J. Comput. Chem.* 32 (11), 2359–2368. doi:10.1002/jcc.21816

Conflict of Interest: JM is the founder of Nanxin Pharm. Co., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Molecular Biosciences

ORIGINAL RESEARCH
published: 19 March 2021
doi: 10.3389/fmolb.2021.624302

Insight Into Seeded Tau Fibril Growth From Molecular Dynamics Simulation of the Alzheimer's Disease Protofibril Core

*Cass Leonard, Christian Phillips and James McCarty**

Department of Chemistry, Western Washington University, Bellingham, WA, United States

Aggregates of the microtubule associated tau protein are a major constituent of neurofibrillary lesions that define Alzheimer's disease (AD) pathology. Increasing experimental evidence suggests that the spread of tau neurofibrillary tangles results from a prion-like seeding mechanism in which small oligomeric tau fibrils template the conversion of native, intrinsically disordered, tau proteins into their pathological form. By using atomistic molecular dynamics (MD) simulations, we investigate the stability and dissociation thermodynamics of high-resolution cryo-electron microscopy (cryo-EM) structures of both the AD paired-helical filament (PHF) and straight filament (SF). Non-equilibrium steered MD (SMD) center-of-mass pulling simulations are used to probe the stability of the protofibril structure and identify intermolecular contacts that must be broken before a single tau peptide can dissociate from the protofibril end. Using a combination of exploratory metadynamics and umbrella sampling, we investigate the complete dissociation pathway and compute a free energy profile for the dissociation of a single tau peptide from the fibril end. Different features of the free energy surface between the PHF and SF protofibril result from a different mechanism of tau unfolding. Comparison of wild-type tau PHF and post-translationally modified pSer356 tau shows that phosphorylation at this site changes the dissociation free energy surface of the terminal peptide. These results demonstrate how different protofibril morphologies template the folding of endogenous tau in distinct ways, and how post-translational modification can perturb the folding mechanism.

Keywords: tau, Alzheimer's disease, neurofibrillary tangles, paired-helical filament, straight filament, umbrella sampling, molecular dynamics

OPEN ACCESS

Edited by:
Pengfei Tian,
Novozymes (Denmark), Denmark

Reviewed by:
Xingcheng Lin,
Massachusetts Institute of Technology, Cambridge, MA, United States
Fanjie Meng,
National Institutes of Health Clinical Center (NIH), Maryland, MD, United States

****Correspondence:***
James McCarty
jay.mccarty@wwu.edu

Specialty section:
This article was submitted to Biological Modeling and Simulation, a section of the journal Frontiers in Molecular Biosciences

Received: *31 October 2020*
Accepted: *09 February 2021*
Published: *19 March 2021*

Citation:
Leonard C, Phillips C and McCarty J (2021) Insight Into Seeded Tau Fibril Growth From Molecular Dynamics Simulation of the Alzheimer's Disease Protofibril Core. Front. Mol. Biosci. 8:624302. doi: 10.3389/fmolb.2021.624302

1 INTRODUCTION

Alzheimer's disease (AD) is characterized by extracellular plaque deposits of amyloid-β (Aβ) peptides and intracellular neurofibrillary tangles (NFTs) of the microtubule associated protein tau (Selkoe and Hardy, 2016; Polanco et al., 2018). Both Aβ and tau contribute to neuroinflammation and neurodegeneration (Bolós et al., 2017). Evidence suggests that Aβ and tau interact synergistically in AD pathogenesis (Rhein et al., 2009; Ittner and Götz, 2011), and amyloid plaques have been shown to facilitate the seeding of tau fibrils (He et al., 2018). As a complement to amyloid-β-based drugs for AD, tau is a potential target for therapeutics aimed at blocking tau aggregation (Noble et al., 2011; Li and Götz, 2017).

Tau is a soluble, intrinsically disordered protein (IDP) predominantly found in axons (Konzack et al., 2007). Under physiological conditions, tau binds to microtubules and plays an important role in microtubule stabilization, the regulation of active axonal transport, and neuronal polarity (Gonzalez-Billault et al., 2002; Götz et al., 2006). Full-length tau (2N4R) consists of 441 amino acids with a N-terminal region, a proline-rich domain, four (R1-R4) microtubule binding repeat (MBR) domains, and a C-terminal region. Alternative mRNA splicing produces six isoforms of tau in the human brain with either three (R1, R3, and R4) or four (R1-R4) microtubule binding repeat (MBR) domains. In AD NFTs, both the three- and four-repeat isoforms are present (Goedert and Spillantini, 2006). Tau has 84 available serine (S), threonine (T), and tyrosine (Y) phosphorylation sites, located primarily in the proline-rich and C-terminal domains (Goedert et al., 2017a). Hyperphosphorylation of tau at both physiological and pathological phosphorylation sites causes tau to dissociate from microtubules and is observed in both AD patients and transgenic mouse models (Ballatore et al., 2007).

It is hypothesized that the spread of tau pathology in the brain progresses via a prion-like mechanism, in which oligomeric tau or other aberrant pre-fibrillar species induces other tau molecules to adopt a particular pathological structure (Mudher et al., 2017; Goedert et al., 2017a; Goedert et al., 2017b; Ayers et al., 2018; Jucker and Walker, 2013). Both *in vitro* and *in vivo* experiments show that small, oligomeric tau complexes seed the growth of tau fibrils (Strang et al., 2018). This seed can be a small tau protofibril isolated from mouse or human brain tissue. Different structural seeds induce different tau fibril morphologies by recruiting tau in solution to polymerize onto the protofibril end, consistent with a prion-like hypothesis (Ayers et al., 2018). The precise mechanism by which endogenous tau is converted into a particular fibril structure is not well-understood, but likely depends on subtle differences in chemical environment, signaling, environmental stress, or mutations.

The predominant component of tau NFTs is a paired helical filament (PHF) structure formed by a twisted, double helical stack of C-shaped subunits (Kidd, 1963). A second structural polymorph, called the straight filament (SF), is also found in tau inclusions and consists of a similar C-shaped unit forming different lateral contacts between filament subunits (Crowther, 1991). Recent high-resolution cryo-EM structures of the tau fibril core, isolated from the brain of an individual with AD, reveal the detailed C-shaped core structure formed by residues 306–378 that are part of the R3-R4 repeat domain (Fitzpatrick et al., 2017). The C-shaped fibril core has a combined cross-β/β-helix structure typical of amyloid fibrils and prion structures (Sunde et al., 1997; Govaerts et al., 2004). The core residues include the PHF6 hexapeptide ($_{306}$VQIVYK$_{311}$ in R3) that has been identified as a minimal interaction motif for tau aggregation and amyloid formation (Friedhoff et al., 2000; Luo et al., 2014; Ganguly et al., 2015; Xie et al., 2015).

Molecular dynamics (MD) simulations can provide atomic-resolution information about the stability and thermodynamics of tau fibril elongation. For example, Li et al. performed all-atom, 100 ns MD simulations of the C-shaped motif, demonstrating that the form is stable only for the R3-R4 repeat domains while the R1-R2 adopts a linear shape (Li et al., 2018). MD simulations of full-length tau in solution reveal that tau samples both extended and compact conformations and can transiently form secondary structures resembling the fibril state (Battisti et al., 2012). All-atom replica exchange MD (REMD) simulations (Larini et al., 2013; Ganguly et al., 2015; Levine et al., 2015) and coarse-grained (Smit et al., 2017) simulations of important nucleating fragments of tau have provided information about the early stages of tau aggregation and which factors stabilize either parallel or antiparallel β-sheet structures (Ganguly et al., 2015). REMD simulations performed by Derreumaux et al. of the R3-R4 domain dimer identified elongated, U-shaped, V-shaped, and globular configurations, but not the C-shaped structure characteristic of AD NFTs (Derreumaux et al., 2020). Recent steered molecular dynamics (SMD) simulations assessed the stability and dissociation of tau from an isolated protofibril pentamer, suggesting that the PHF and SF protofibrils induce a different pathway for misfolding of tau (Liu et al., 2019).

Despite advances in processing power and designated custom hardware (Shaw et al., 2014), conventional atomistic MD simulations in explicit solvent of protofibril nucleation and elongation remain particularly challenging due to the large time scale and system sizes characteristic of protein aggregation. Enhanced sampling methods can overcome this challenge by accelerating the exploration of configurational state space through the use of an applied bias potential. Umbrella sampling (Torrie and Valleau, 1977) is particularly suited to compute the free energy surface (FES) along a pre-defined reaction coordinate (Roux, 1995). Umbrella sampling has been applied to study both the thermodynamics and kinetics of Aβ fibril growth, providing insight into the stability and formation of the Aβ fibril (Lemkul and Bevan, 2010; Schwierz et al., 2016).

In this work, we use all-atom molecular dynamics simulation in explicit solvent to study the paired PHF and SF protofibril structures and a post-translationally modified PHF, phosphorylated at residue Ser356. We assess the stability of the protofibril structure from SMD simulations, identifying the structural changes that occur in response to the applied force at its maximum value. The structural changes that result from the applied force reveal interchain interactions that impart stability to the fibril. We then perform an exploratory metadynamics simulation to determine the dissociation pathway of a single tau peptide "monomer" from the protofibril end. To obtain the free energy surface (FES) for tau dissociation, we perform umbrella sampling simulations of configurations sampled along the dissociation pathway, using a harmonic restraining potential to sample the configurational space along the COM distance reaction coordinate. The FES along the dissociation coordinate, obtained using the weighted histogram analysis method (WHAM), provides thermodynamic and mechanistic insight into the stability of tau protofibrils and the dissociation mechanism of a single tau peptide from the fibril end. This work can provide a foundation for designing and interpreting tau seeding experiments with different fibril morphologies and for designing small molecule inhibitors that destabilize tau NFTs.

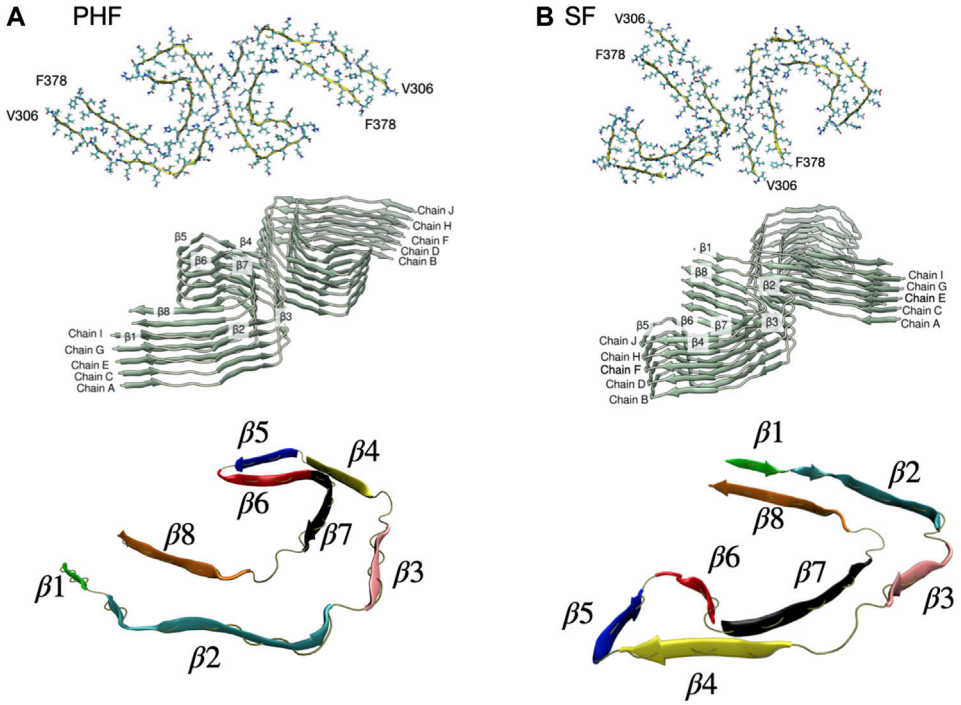


FIGURE 1 | (A) PHF protofilament based on the cryo-EM R3-R4 structure of residues V306-F378 (PDB ID entry 503 L). We used ten total chains labelled A-J arranged as five paired filaments. The C-shaped core structure has eight β sheets labelled β1-β8. **(B)** SF protofilament based on the cryo-EM R3-R4 structure of residues V306-F378 (PDB ID entry 5O3T). The SF structure has the same eight β sheets labelled β1-β8 but a different orientation of paired filaments. A rendered view of the secondary structure is shown at the bottom with the β sheets individually colored.

2 METHODS

The complete cryo-EM structures of both the PHF (PDB ID entry 503 L) and SF (PDB ID entry 5O3T) fibril contain 14 pairs of chains arranged in a helical stack with a C-shaped cross-section. As a minimal stable protofibril starting structure, we used ten protofilament core chains (chains A-J) resulting in a structure with five stacked and paired C-shaped subunits, as shown in **Figure 1**. The N-terminus of each peptide was capped with an acetyl group (ACE) and the C-terminus was capped with a N-methionine group (NME) to give uncharged terminal ends. A post-translationally modified PHF protofibril was created using the PyTMs plugin for PyMOL (Warnecke et al., 2014; Schrödinger, LLL, 2015). We added a phosphate group (−2 charge) onto Ser 356 for chains labelled G and I in **Figure 1A**. All titratable amino acids were assigned a charge based on physiological pH. We used the CHARMM36m force field (Huang et al., 2017) with the TIP3P water model. CHARMM36 parameters are available for phosphorylated serine amino acids. Each protofibril chain was solvated in a box with periodic boundary conditions. The system was neutralized with counter ions to achieve a final salt concentration of 150 mM.

All simulations were performed using the Gromacs 2019.4 MD code (Hess et al., 2008; Lindahl et al., 2019) with an integration time step of 2 fs. Long-range electrostatics were calculated using the PME method (Essmann et al., 1995), and we used a cutoff radius of 1.0 nm for both real-space Coulombic and Lennard-Jones interactions. Equilibrium MD simulations of both the SF and PHF protofibril in solution show that the structure is stable for at least 10 ns (see **Supplementary Figure S1**).

The protofibril was placed in an elongated box with dimensions 15.8 × 8.7 × 25.7 nm for PHF and 12.4 × 12.5 × 25.7 nm for SF, as determined by the minimum image convention for periodic boundary conditions. After solvating the box with TIP3P water and counter ions, the PHF system contained 358,966 atoms and the SF had 402,600 atoms. Following a steepest descent energy minimization step, the system was equilibrated for 100 ps in an NVT ensemble at a temperature of 310 K using the velocity rescaling thermostat (Bussi et al., 2007) and position restraints on all heavy atoms. This was followed by a 100 ps NPT equilibration using a Berendsen weak-coupling barostat (Berendsen et al., 1984) to maintain a pressure isotropically at 1 bar. Position restraints were removed from all heavy atoms except for peptides G and H, which were used as an immobile reference (See **Figure 1** for a definition of chains and secondary structure). This restraint mimics the effect of the larger fibril structure (Takeda and Klimov, 2009a; Takeda and Klimov, 2009b). All production runs were performed in the NPT ensemble at a temperature of 310 K and pressure of 1 bar, using the velocity rescaling thermostat and the Parrinello-Rahman barrostat (Parrinello and Rahman, 1981).

To quantify structural changes that occur during a simulation, the distance root mean square deviation (dRMSD) of the backbone atoms with respect to a reference β sheet structure was monitored for each inter-chain β sheet formed between chain G and chain I (β1–β8 in **Figure 1**). The dRMSD was computed during the simulation using the PLUMED2.6 plugin (Tribello et al., 2014). The dRMSD is a measure of the distance between two structures $\mathbf{X}^a$ and $\mathbf{X}^b$, defined as

$$d(\mathbf{X}^A, \mathbf{X}^B) = \sqrt{\frac{1}{N(N-1)} \sum_{i \neq j} \left[d\left(\mathbf{x}_i^a, \mathbf{x}_j^a\right) - d\left(\mathbf{x}_i^b, \mathbf{x}_j^b\right) \right]^2} \quad (1)$$

where N is the number of backbone atoms in the reference structure, and $d(\mathbf{x}_i, \mathbf{x}_j)$ is the distance between atoms i and j. For each of the β sheets (β1–β8) the cryo-EM structure after a short 200 ps equilibration was used as the reference structure in the calculation.

2.1 Steered MD Simulations

For each SMD (COM pulling) simulation, the COM of peptide chain I was pulled away from the COM of peptide chain G of the fibril core along the z-axis for 1,200 ps using a constant pull rate of 0.01 nm ps^{-1} and a spring constant of 1,000 kJ mol^{-1} nm^{-2}. The simulation box, placement of the chain, and definition of the elongation axis are shown in the **Supplementary Figure S2**. A slower pulling rate of 0.0025 nm ps^{-1} resulted in similar force curves (see **Supplementary Figures S3–S5**). A final COM distance between peptide chain I and the fibril end of approximately 10–12 nm was achieved.

2.2 Metadynamics Simulations

The dissociation path of a single tau peptide chain from the protofibril structure was investigated with metadynamics (Laio and Parrinello, 2002). In this work, the history-dependent metadyanimcs bias is applied in order to explore the dissociation pathway and identify partially folded intermediates. However, due to the slow convergence of the metadynamics bias, we do not compute the free energy surface from reweighting. Instead, we take representative configurations along the dissociation path sampled via metadynamics and perform umbrella sampling along the COM separation distance. Metadynamics enhances the exploration of phase space by applying a history-dependent bias potential along a chosen set of collective variables (CVs) (Valsson et al., 2016). Analogous to protein folding (Best et al., 2013), we compute the fraction of native contacts Q between chain I and the adjacent chains H, F, and G (see **Figure 1**). Here, Q includes all the native contact pairs i, j between heavy atoms i and j, considered in contact if the distance between i and j is less than 0.45 nm. The CV Q is then computed as

$$Q(\mathbf{X}) = \frac{1}{N} \sum_{(i,j)} \frac{1}{1 + \exp\left[\beta\left(r_{ij}(\mathbf{X}) - \lambda r_{ij}^{\circ}\right)\right]} \quad (2)$$

where $r_{ij}(\mathbf{X})$ is the distance between atom i and j in configuration X, r_{ij}° is the reference distance in the cryo-EM structure, β is a smoothing parameter set to 50 nm^{-1} and λ is a tolerance distance set to 1.5 nm. We identify 1,260 contacts between the terminal chain I and adjacent chains (G, H, and J) in the PHF structure and 1,149 contacts for the SF protofibril. The fewer native contacts in the SF structure is due to the looser packing of this structure and less contact between the adjacent paired helical structures shown in **Figure 1**.

Metadynamics simulations were carried out using the open-source, community-developed PLUMED library (Bonomi et al., 2019), version PLUMED2.6 plugin (Tribello et al., 2014). In addition to the fraction of native contacts CV Q, we also biased the COM distance between chains I and G (see **Figure 1** for chain definitions). The two-dimensional bias was deposited every picosecond with a Gaussian hill height of 1.0 kJ/mol and a width of σ = 0.01 for Q and σ = 0.1 nm for the COM distance. Finally, a ratchet-and-pawl like restraint was placed on the COM distance to evolve the system toward further separation distances and dampen fluctuations back towards the protofibril end. A definition of this restraint is presented in the **Supplementary Material Section 1**.

2.3 Umbrella Sampling

From representative frames of the exploratory metadynamics trajectories, configurations were selected to generate starting configurations for umbrella sampling (Lemkul and Bevan, 2010). Frames for the umbrella sampling windows were selected every 0.1 nm up to 2 nm COM separation distance between chains I and G and every 0.2 nm beyond up to 11.0 nm. This resulted in a total of 62 umbrella windows for each protofibril. Since each umbrella window is taken from a snapshot of the metadynamics simulation, each window has the same box size and number of particles. For each umbrella window a harmonic restraint was employed centered at the reaction coordinate of the initial COM distance for that window. We used a force constant of 1,000 kJ mol^{-1} nm^{-2}. After a short 200 ps equilibration, a 10 ns long production MD simulation was run for each umbrella window. The total production simulation time for the set of umbrella sampling simulations is 620 ns for each protofibril. A histogram of the sampled reaction coordinate for each system is shown in the **Supplementary Figure S6**, showing the overlap of the sampled distance distribution between adjacent windows. The free energy surface from the umbrella sampling simulations was computed using the weighted histogram analysis method (WHAM) as implemented in GROMACS 2019.4 (Hub et al., 2010). All umbrella sampling simulations were preformed on a GPU workstation (8 CPU threads and 1 Nvidia RTX 2080 GPU). We also made use of the SDSC Comet Supercomputer available through the Extreme Science and Engineering Discovery Environment (XSEDE) (Towns et al., 2014).

3 RESULTS

In the present work we focus on the fibril stability of the two AD tau cryo-EM structures: the PHF filament and the SF filament. We also investigate a post-translationally modified PHF fibril with a phosphate group at Ser356, located in the MTB repeat domain. The PHF and SF filaments are structural polymorphs

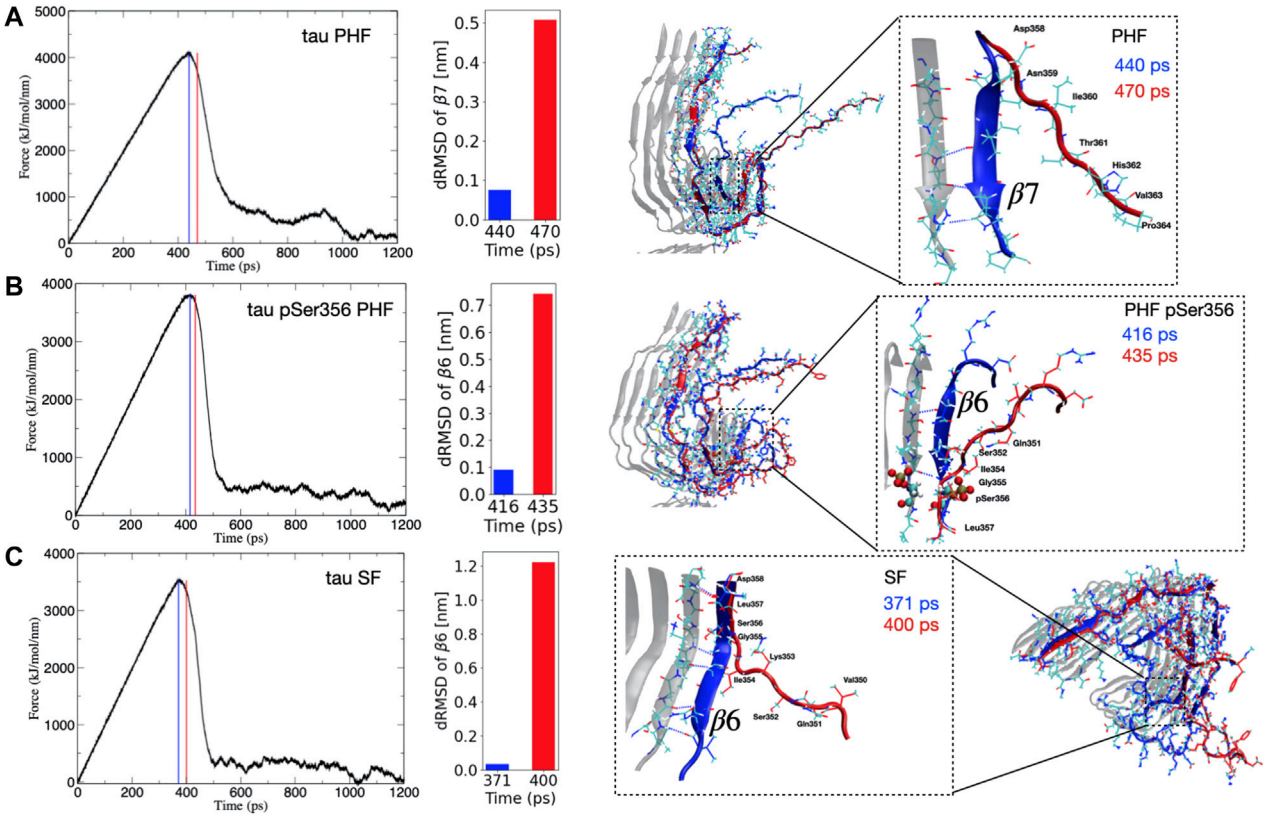


FIGURE 2 | (Left) Time evolution of the force during SMD simulations for the **(A)** PHF, **(B)** phosphorylated pSer356 PHF, and **(C)** SF tau protofibrils. The blue vertical line (and depicted blue peptide structure on the right) corresponds to the structure at the maximum force. The vertical red line (and depicted red peptide structure on the right) correspond to a representative frame just after this maximum force has been reached and the force curve begins to decrease. This corresponds to a major structural transition where important hydrogen bonds have been broken during dissociation. **(Right)** The inset highlights the region of secondary structure loss that occurs between the point of maximum force (blue) and immediately after the force curve begins to decrease (red). For PHF **(A)** this event correlates with loss of the β7 secondary structure as indicated by the increase in the dRMSD value for this structure during this time window. For the PHF-pSer356 **(B)** and SF **(C)** the adjacent β6 region unfolds during the time window with corresponding increase in the dRMSD with respect to the reference β6 structure.

with the same number of amino acids but different packing and relative orientation of paired filaments. These structural differences are expected to lead to different relative stabilities of the two fibril types. As shown in **Figure 1**, the structural core unit for both fibrils consists of eight β-strands (labelled β1–β8) adopting a C-shaped structure.

3.1 SMD Simulations Identify That Interchain Contacts Formed by Residues Within the β6 and β7 Regions Impart Critical Structural Stability to the Protofibril

SMD simulations can be used to identify important interactions between subunits that are broken during the non-equilibrium trajectory (Izrailev et al., 1999). SMD has been applied in the context of protein-ligand binding (Grubmüller et al., 1996), DNA-binding proteins (Jakubec and Vondrášek, 2020), Aβ fibril growth (Lemkul and Bevan, 2010), and on tau fibril dissociation (Liu et al., 2019). During the non-equilibrium simulation, the force increases as a result of the applied bias until a breaking point is reached, at which time critical interactions are disrupted, allowing the peptide to dissociate from the core protofibril structure. The point of maximum force corresponds to the instant just before these key interactions are broken. Because the work performed during a SMD simulation is path-dependent, a single SMD pulling trajectory is insufficient to determine the free energy surface, and different force-time curves will produce different dissociation pathways. For this reason, the precise order of events leading to dissociation cannot reliably be determined from SMD simulations at the high force values used here.

Despite the fact that tau dissociates from each protofibril end through different pathways, we consistently observe that the point of maximum force for each protofibril architecture corresponds to the breaking of hydrogen bonds between the parallel β-sheets, formed largely by residues Arg349 to Val363 that make up the β6–β7 region (see **Figure 1**). The observation that these hydrogen bonds break at the point of maximum force along the SMD trajectory suggests that these interactions impart critical stability to the protofibril. **Figure 2** shows the force applied during the pulling of a single tau peptide chain from the protofibril tip during a 1,200 ps SMD simulation. A snapshot

along the dissociation pathway at the point of maximum force is shown with the dissociating peptide chain colored blue, and the subsequent loss of structure immediately after the point of maximum force is shown with the dissociating peptide colored red. For the PHF protofibril (**Figure 2A**), the point of maximum force occurs at 440 ps and involves breaking the interchain hydrogen bonds between residues Ile360 and Val363 within the β7 region. A snapshot highlighting this region of the protofibril just before the point of maximum force at 440 ps is shown in blue and is compared to the same chain when the force-time curve is decreasing at 470 ps shown in red. We observe that during this 30 ps window, the *β*-sheet formed by residues Asp358-His362 (β7 in **Figure 1**) is lost and hydrogens bonds between parallel *β*-sheets in this region are broken. This is confirmed by the large increases in the distance dRMSD for the β7 region during this time window.

For the post-translationally modified PHF at Ser356 (PHF-pSer356), the force-time curve (**Figure 2B**) exhibits a maximum at 416 ps and involves loss of structure and breaking of interchain hydrogen bonds between the *β* sheet formed by residues Gln351 to Gly355 in the β6 region. The middle structure in **Figure 2B** compares a snapshot of the PHF-pSer356 structure just before the maximum force at 416 ps and at 435 ps when the force is decreasing, showing the interchain hydrogen bonds that are broken during this window. This region is directly adjacent to the phosphorylated Ser356. The dRMSD of the β6 sheet shows an abrupt increase during this time window indicating a loss of the β6 structure at the point of maximum pulling force.

For the SF protofibril (**Figure 2C**), the point of maximum force occurs earlier, at 371 ps. The structure just before the point of maximum force at 371 ps and at 400 ps (where the force is decreasing) is compared in bottom structure in **Figure 2C**, showing loss of the β-sheet structure between residues Gln351 and Asp358 that make up the β6 region and beginning of the β7 region. Interchain hydrogen bonds within this region are broken during this time window, and the dRMSD of the β6 sheet increases sharply as the β6 structure is lost.

In all SMD simulations, it is clear that a single main structural transition corresponds to a destabilization of the fibril structure, leading to dissociation at the protofibril tip. These structural transitions involve the parallel *β*-sheets that define the β6 and β7 region.

3.2 The Free Energy Surface From Umbrella Sampling Reveals Different Template-Induced Folding Mechanisms of Tau

Because the large pulling force may result in an unfolding mechanism that does not resemble the true dissociation path, we produce configurations along the dissociation path using metadynamics. Metadynamics builds a history-dependent bias during the simulation that allows the system to escape free energy minima, and the dissociation path revealed during a metadynamics simulation should closely follow the true dissociation mechanism. Furthermore, by including the fraction of native contacts Q as a CV, the applied bias should lead to enhanced fluctuations of native contacts leading to tau dissociation along the protein folding pathway. As the metadynamics bias builds during the simulation, weaker contacts should break before stronger contacts; thus the sequence of dissociation events from metadynamics should reflect the dissociation mechanism. We then perform umbrella sampling simulations restrained along the one-dimensional COM distance for frames extracted from metadynamics simulations.

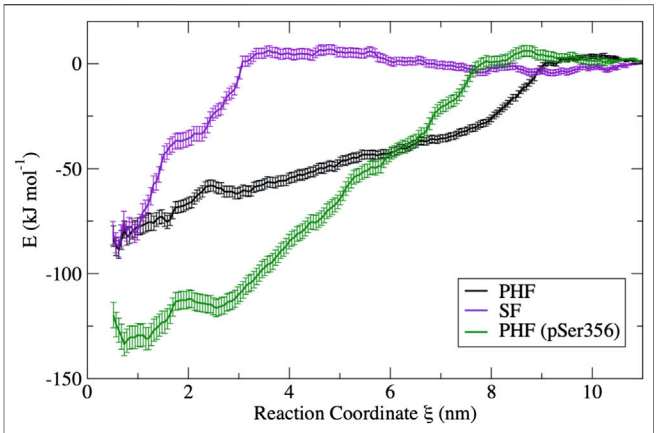


FIGURE 3 | Free energy curves obtained from umbrella sampling simulations for each fibril in this study. The PHF free energy profile (black) and the SF free energy profile (violet) have a similar free energy minimum; however, the different shape reflects a different dissociation mechanism. The PHF fibril is stabilized by phosphorylation at Ser356 (green) despite having a similar mechanism of dissociation. Error bars were determined by bootstrapping.

The free energy surfaces computed from umbrella sampling simulations are shown in **Figure 3**. Error bars are determined using a bootstrap method (Hub and de Groot, 2006). Individual bootstrap profiles are presented in the **Supplementary Figure S7**. To provide a sense of the convergence of the free energy surface from the limited 10 ns umbrella sampling trajectories, the FES was computed separately for different trajectory blocks of 2–4, 4–6, 6–8, and 8–10 ns. A comparison of the FES for these different regions is presented in the **Supplementary Figure S8**. In all cases, the FES profile does not change appreciably after 5 ns. Therefore, the FES shown in **Figure 3** is computed only over the final 5 ns of the 10 ns production simulation.

The difference in the free energy surfaces between the PHF fibril (black line) and SF fibril (purple line) of **Figure 3** reflect differences in the dissociation and unfolding mechanism of tau for these two fibril structures. The PHF FES (black line) has a broader basin, indicating that partially folded tau forms contacts with the protofibril end at longer separation distances. The broad basin for the PHF fibril corresponds to the sequential breaking of interchain interactions, leading to unfolding of the tau peptide through a series of partially unfolded intermediates that will be discussed in more detail below. On the other hand, the FES for the SF fibril (purple line) has a narrower and steeper basin where partially folded tau makes contact with the protofibril end at shorter separation distance. The shape of the FES indicates that

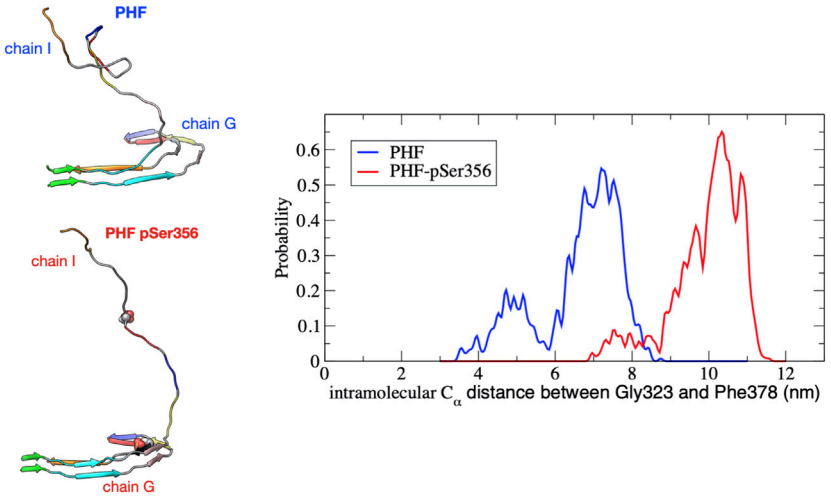


FIGURE 4 | Distribution of the intramolecular Cα distance between residues Gly323 and Phe378 on chain I for the PHF (blue) and PHF-pSer356 (red) protofibril over the umbrella window restrained at a COM distance of 3.6 nm. The distribution is accumulated over the 10 ns production simulation. At this separation distance, the partially folded tau remains in contact with the protofibril at the β1 end (V306). A snapshot taken at 5 ns during the 10 ns trajectory is shown for both structures. The PHF-pSer356 remains predominantly in an extended conformation while the wild-type PHF adopts more compact configurations forming non-native intramolecular contacts.

tau unfolds in a more concerted mechanism from the SF protofibril as compared to the PHF fibril. Interestingly, the difference in free energy between the bound structure (ξ = 0.6 nm) and the free dissociated tau (ξ = 11 nm) of both the PHF and SF protofibril is nearly identical, with a value of −89 kJ/mol ± 5 kJ/mol (35 k_BT) for the PHF and −87 ± 6 kJ/mol (34 k_BT) for the SF. This value compares reasonably well to the experimental value for fibril elongation of Aβ(1–40) of −38 kJ/mol (34 k_BT) (O'Nuallain et al., 2005). However, we note that our calculated value may include errors from limited sampling.

Phosphorylation at pSer356 changes the shape of the FES profile for the PHF protofibril (green line). Interestingly, the free energy of the final folded state of the phosphorylated tau is lower with respect to the unphoshorylated PHF implying that the phosphate actually stabilizes the fibril structure. The origin of this increased stability of the pSer356 PHF protofibril is not obvious. Simulations of native tau in solution suggest that phosphorylation of tau at Ser356 can facilitate aggregation by destabilizing compact configuration and enhancing the distribution of extend conformations that expose residues to the protofibril template (Popov et al., 2019). To investigate this, **Figure 4** shows a representative structure at a separation distances of ξ = 3.6 nm for the PHF and PHF pSer356 along with the intramolecular distribution of the distance between the Cα of residues Gly323 and Phe378 on chain I averaged over the 10 ns trajectory. The wild-type tau adopts partially folded compact intermediate states while the phosphorylated tau remains in a much more extended conformation. As seen in **Figure 3** (black line), after the wild-type tau makes initial contact with the PHF fibril at around 8 nm, the FES exhibits a docking region without a steep folding funnel. This feature agrees with recent models of amyloid aggregation progressing via a random search through multiple, non-productive conformations before the peptide samples an extended configuration that is able to form native contacts with the fibril template (Jia et al., 2017; Jia et al., 2020). In contrast, phosphorylation at Ser356 shifts the conformational ensemble towards more extended conformations (**Figure 4**), and the FES in **Figure 3** (green line) exhibits a steeper folding funnel, along which native contacts form in successive order along the fibril template.

3.2.1 The Dissociation Mechanism of Tau From the PHF Protofibril

We now discuss in more detail the order of partial unfolding events that lead to dissociation of a single tau peptide from the PHF protofibril end. We analyze conformations from different umbrella windows along the reaction coordinate in terms of the dRMSD of the various β sheets that form the folded structure. **Figure 5** shows representative structures from umbrella sampling windows at key intermediate stages of dissociation. Escape from the free energy minima that represents the bound conformation begins with loss of the β6 and β5 structure, followed almost immediately with partial stabilization of the β7 region. This conformational change is shown between structure 1 (ξ = 0.6 nm) and structure 2 (ξ = 0.9 nm) in **Figure 5** and the corresponding increase in dRMSD for the β5-β7 region. Next, the β4 structure is lost between structure 2 and 3 (ξ = 1.4 nm). Between structures 3 and 4 (ξ = 5.8 nm), the partially unfolded tau forms non-native intramolecular contacts like that shown in **Figure 4**. The remaining sequence of dissociation events is loss of β3 between structures 3 and 4, loss of β2 between structures 4 and 5 (ξ = 8.8 nm), and finally, loss of β1 between structures 5 and 6 (ξ = 10.4 nm) that results in complete dissociation (structure 6) of the peptide. This final dissociation step (loss of β1) involves breaking interactions between the nucleating

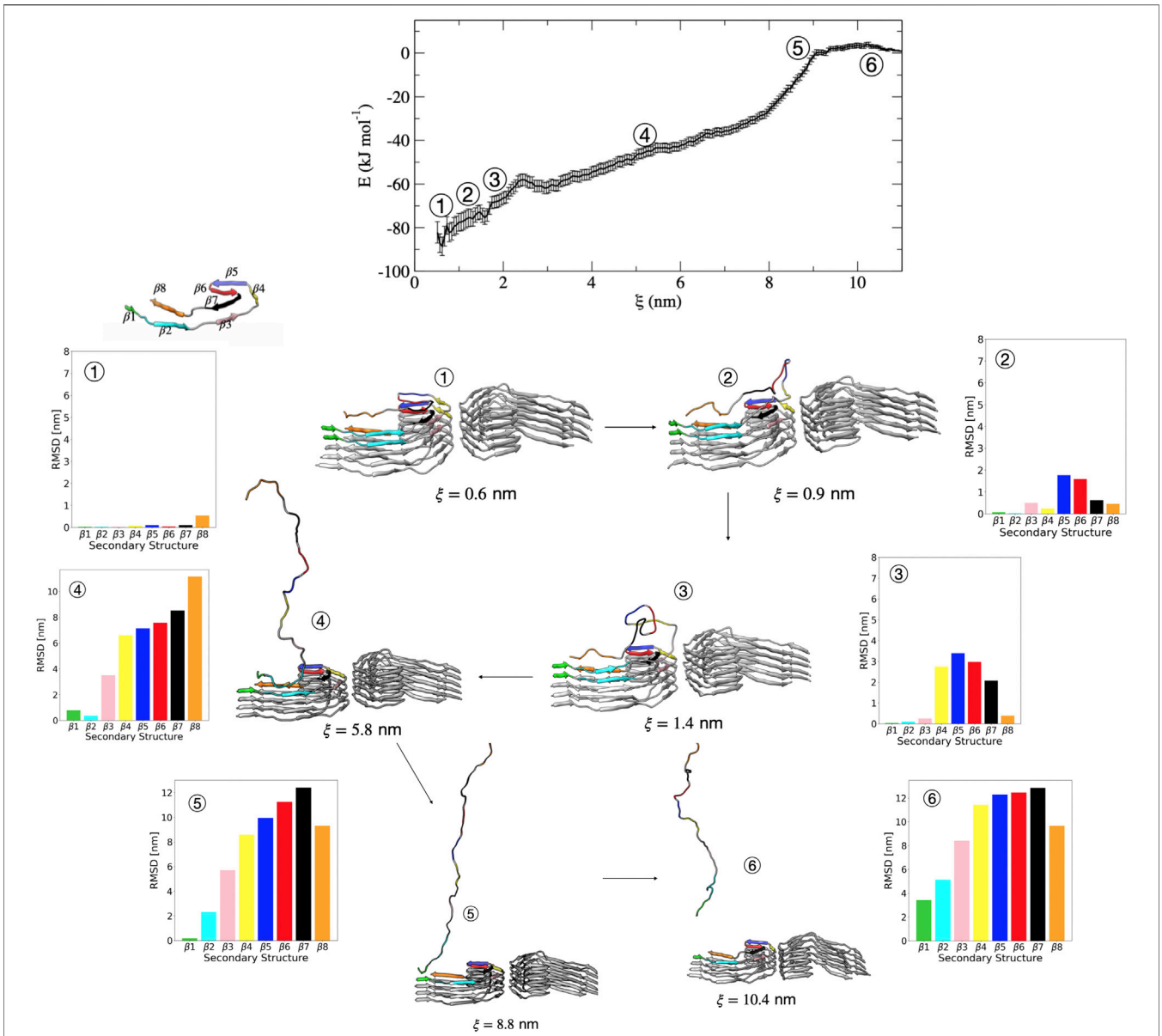


FIGURE 5 | Free energy profile for the association/dissociation of a single tau chain onto the PHF protofibril. Representative structures are shown along with the dRMSD of the secondary structure regions. Between structures 1 and 2 the dRMSD increases for β5 (blue), β6 (red), and β7 (black). Structure 3 shows loss of β4 (yellow), and structure 4 shows loss of the β3 (pink). Structure 5 shows loss of β2 (cyan). The complete dissociation occurs with loss of β1 (green) that includes the PHF6 region. Structures were obtained from umbrella sampling simulation window within the different regions along the reaction coordinate after 5 ns. The dRMSD values in each of the bar graphs are calculated from **Eq. 1** for the depicted representative structures taken at 5 ns of simulation, where the final folded state is the reference structure.

PHF6 hexapeptide ^{306}VQIVYK^{311} region. This supports the hypothesis that the formation of the PHF6 hexapeptide initiates tau misfolding and aggregation.

3.2.2 Effect of Phosphorylation at pSer356 on the Dissociation Mechanism of Tau From the PHF Protofibril

Figure 6 shows representative structures along the reaction coordinate for the pSer356 post-translationally modified tau. The mechanism of dissociation in terms of the order of events is similar to that of the wild-type PHF. In this case, escape from the free energy minima that represents the bound conformation begins with loss of the β7 structure followed immediately by loss of the adjacent β6 and β8 region. This is shown in **Figure 6** by the increase in dRMSD between structure 1 ($\xi = 0.6$ nm) and structure 2 ($\xi = 1.8$ nm) for these regions. Next, the β5 structure is lost, followed by the loss of β4 as shown between structure 2 and structure 3 ($\xi = 5.2$ nm). The remaining unfolding

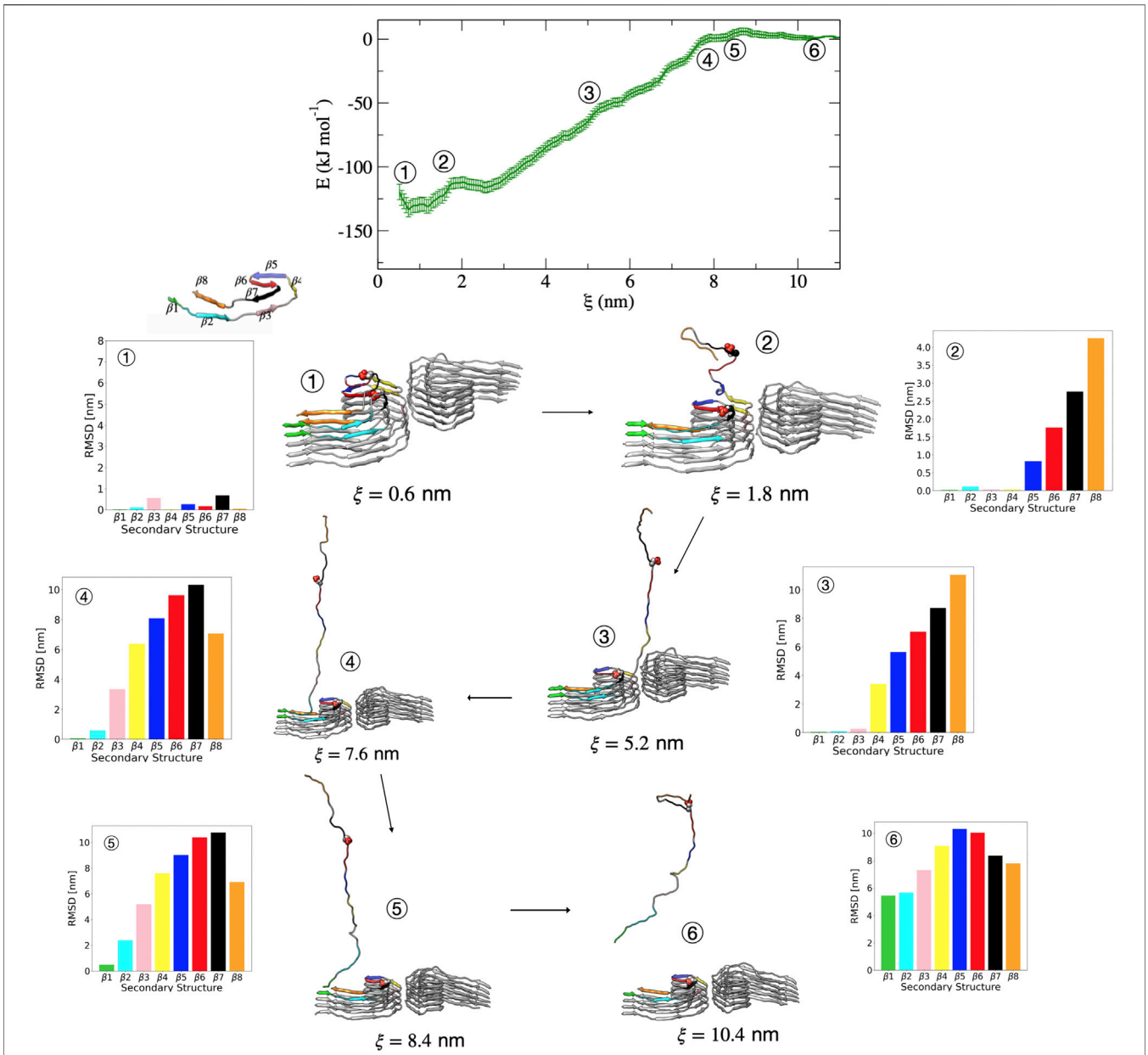


FIGURE 6 | Free energy profile for the association/dissociation of a single pSer356 peptide from the phosphorylated PHF protofibril at Ser356. Representative structures are shown along with the dRMSD of the secondary structure regions. Between structures 1 and 2, the dRMSD increases for β5 (blue), β6 (red), β7 (black), and β8 (orange). Structure 3 shows loss of the β4 (yellow), and structure 4 shows loss of β3 (pink). Structure 5 shows loss of β2 (cyan). The complete dissociation occurs with loss of β1 (green) that includes the PHF6 region. Structures were obtained from umbrella sampling simulation window within the different regions along the reaction coordinate after 5 ns. The dRMSD values in each of the bar graphs are calculated from **Eq. 1** for the depicted representative structures taken at 5 ns of simulation, where the final folded state is the reference structure.

events proceed identically to the wild-type PHF in the order of β4 → β3 → β2 → β1. We concluded that for both PHF and PHF pSer356 the β1-β4 region is involved in nucleation and the initial docking of tau to the protofibril template, occurring at separation distances between 2.0 and 9.0 nm. The subsequent folding of the β6-β8 regions leads to locking of tau into a pathological structure and occurs at separation distances less than 2.0 nm.

3.2.3 The Dissociation Mechanism of Tau From the SF Protofibril

Compared with the PHF protofibril, tau dissociates from the SF protofibril with a different unfolding mechanism, as suggested by the different free energy surface. **Figure 7** shows representative structures from umbrella sampling windows at key intermediate stages of dissociation of a

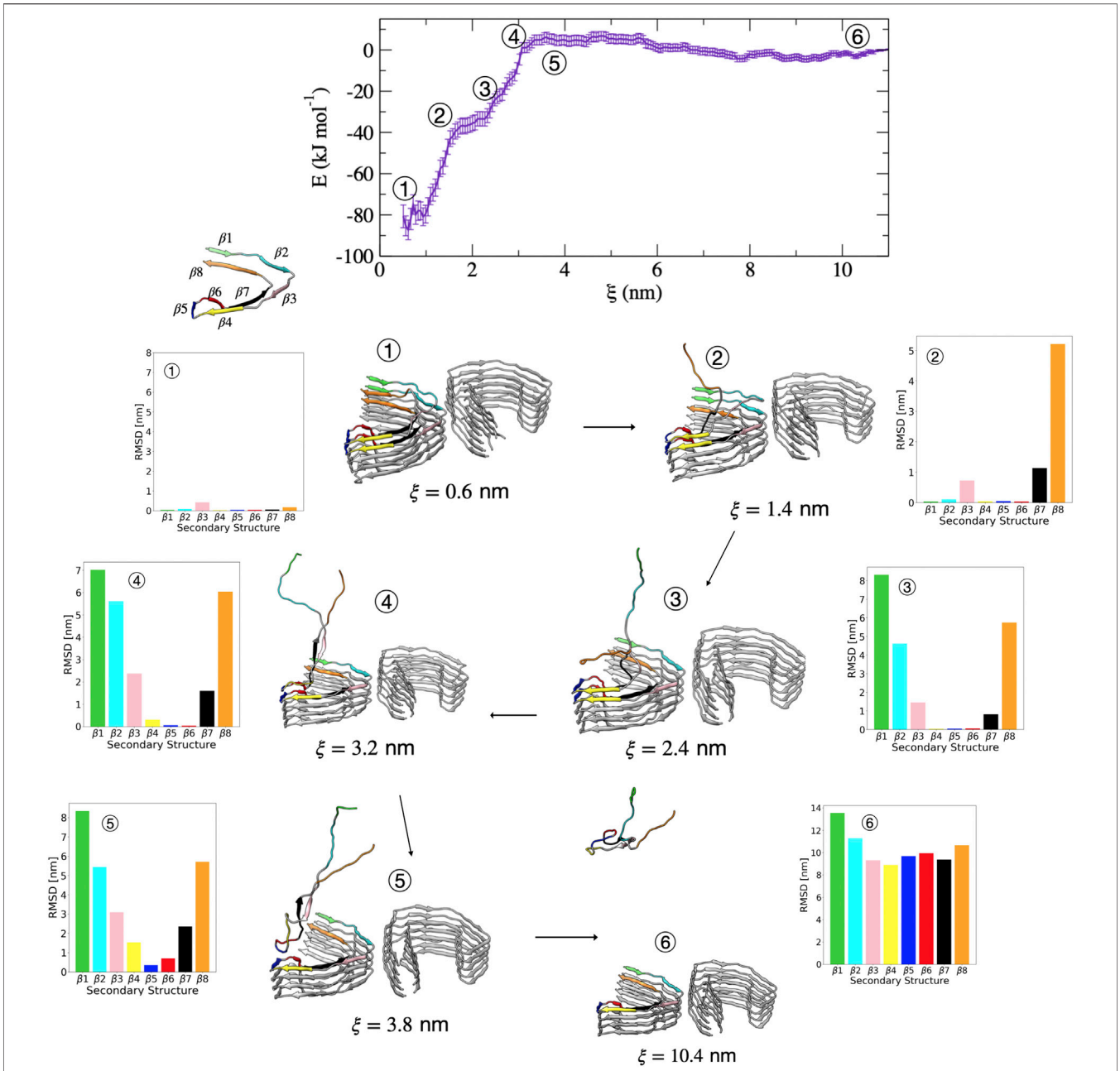


FIGURE 7 | Free energy profile for the association/dissociation of a single tau chain onto the SF protofibril. Representative structures are shown along with the dRMSD of the secondary structure regions. Tau dissociates starting from the β8 end, shown by the increase in dRMSD between structures 1 and 2 for β7 (black) and β8 (orange). Structure 3 shows simultaneous loss of β1 (green) and β2 (cyan). Structure 4 shows simultaneous loss of β3 (pink) and β7 (black). Structure 5 shows loss of β4 (yellow) and an increased dRMSD for β6 (red). The complete dissociation occurs with loss of β5 (blue). Structures were obtained from umbrella sampling simulation window within the different regions along the reaction coordinate after 5 ns. The dRMSD values in each of the bar graphs are calculated from **Eq. 1** for the depicted representative structures taken at 5 ns of simulation, where the final folded state is the reference structure.

single tau peptide from the SF protofibril end. Dissociation begins with loss of the β8 region, as shown by the increase in dRMSD for this region between structures 1 (ξ = 0.6 nm) and 2 (ξ = 1.4 nm) shown in **Figure 7**. In contrast to the PHF, the β1 and β2 regions unfold next, as shown by structures 2 and 3 (ξ = 2.4 nm), followed by the loss of β7 (between structures 3 and 4 (ξ = 3.2 nm)). Between structures 4 and 5 (ξ = 3.8 nm) along the FES, loss of the β4 and β6 region occurs. The final dissociation step involves loss of the β5 region between structures 5 and 6 (ξ = 10.4 nm).

4 DISCUSSION

Using all-atom MD simulations and a combination of enhanced sampling methods including non-equilibrium SMD, metadynamics, and umbrella sampling, we have determined several key factors that are important for the stability of tau AD NFTs and for fibril elongation through induced folding of monomeric tau. Our computational study suggests that the β6–β7 region imparts stability to both the PHF and SF fibril despite differing dissociation mechanisms. SMD simulations indicate that intermolecular contacts formed within this region are energetically strongest. We explore the full dissociation mechanism using metadynamics simulations, enhancing fluctuations along both the COM distance and the fraction of native contacts formed with the fibril end. We observe a different dissociation mechanism between the PHF and SF protofibril. Using umbrella sampling, we compute a free energy surface of tau dissociation from the protofibril end along the COM distance coordinate.

Assuming that fibril elongation proceeds via the reverse process of the observed dissociation, metadynamcis simulations show that elongation of the tau PHF fibril begins with association at the β1 and β2 region. The β1 region, including the PHF6 (^{306}VQIVYK^{311}) motif, is important for nucleation and docking of bulk tau in solution to the PHF fibril. This observation is supported by previous simulations that identified the PHF6 region as a critical fragment for nucleation of amyloid structures (Ganguly et al., 2015). The folding of tau along the PHF fibril template then proceeds with the sequential formation of the β3, β4 and β5 regions. Finally, the formation of native contacts within the β6–β7 region locks the tau peptide at the fibril end. In contrast, tau in solution initiates contact with the SF fibril end at the β5 region, forming the β4–β6 region, followed by β3 and β7 formation, before the templated folding of the β1 and β2 region. Taken together, the free energy surface and corresponding key intermediate structures presents a detailed picture of important steps in AD pathogenesis. Such detailed mechanistic information can give insight into tau seeding experiments in which different protofibril seeds can templet different tau morphologies (Strang et al., 2018).

Post-translational modification or mutations that affect the stability of the fibril might disrupt the formation of toxic NFTs. We identify the β6–β7 region as being important for maintaining both the SF and PHF protofibril stability. To investigate this idea, we have studied a post-translationally modified PHF fibril phosphorylated at Ser356, which is located between the β6 and β7 region. Our results show that pSer356 modifies the FES and alters the order of the β6 and β7 loss of structure in the dissociation mechanism. However, it is not obvious how this subtle difference in the mechanism will manifest during *in vitro* tau seeding experiments. Phosphorylation of Ser356 has been shown experimentally to block tau interactions with Aβ peptide (Guo et al., 2006) and inhibits the seeding activity of the K18 tau construct in the presence of heparin (Haj-Yahya et al., 2020). Meanwhile, REMD simulations of the PHF dimer show that pSer356 modifies the conformational ensemble of a tau dimer in solution (Derreumaux et al., 2020). It has been suggested that pSer356 may lead to increased sampling of extended conformations of disordered tau, thereby exposing residues to the fibril template during binding (Popov et al., 2019). Our umbrella sampling simulations give credence to this idea, showing that partially folded pSer356 tau remains extended while docking and is able to form native contacts with the fibril template without needing to unfold compact conformations or break non-native contacts. Further experiments and simulation work is needed to fully understand the effect of phosphorylation at Ser 356 as well as other possible phosphorylation sites within the MTB region.

It would be interesting in the context of AD targeted therapeutics to investigate small molecule inhibitors or mutations that affect the β5–β7 region. This region is involved in the final folding of tau onto the PHF template, while the formation of initial contacts between tau and the SF fibril involves the β5 region. While the anthraquinone derivative Purpurin molecule has been shown to inhibit tau fibrillization by forming hydrophobic contacts with the PHF6 nucleating hexapeptide region ^{306}VQIVYK^{311}, our results suggest the β5–β7 regions as an alternative target.

This work presents a detailed thermodynamic and mechanistic analysis of tau fibril dissociation for the two structural polymorphs of tau relevant to AD neurodegeneration using recent cryo-EM structures. Differences in the FES for template-induced misfolding of tau by the two AD protofibril structures can provide a more complete understanding of tau seeding from these structures. In addition to further work to understand how mutations and the binding of small molecules might perturb the thermodynamics of fibril elongation through templated folding, simulations of other tau morphologies, such as the widely studied K18 construct, could give additional insights into tau folding mechanisms. While this work presents a picture of tau dissociation, the free energy surface projected along the one dimensional separation distance may hide other relevant conformations or missing slow degrees of freedom that could provide more thermodynamic insight. The present study could be complemented by other enhanced sampling methods that could more completely explore configuration space. Another area for further exploration is in the kinetics of dissociation of tau, which could be elucidated by studying the position-dependent diffusion along the reaction coordinate.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

JM conceived the project and supervised the research. CL and CP performed all the simulations, generated the force-time curves,

and performed the WHAM analysis. JM wrote the manuscript, and all authors edited and reviewed the manuscript.

FUNDING

This work was supported by faculty start-up funds provided by Western Washington University. This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1548562. Simulations were performed using the SDSC Comet Supercomputer allocation ID: TG-MCB200041. Funds to assist in the open-access publication cost were provided by a mini grant from Western Washington University's Office of Research and Sponsored Programs (RSP) Mini-Grant Award #MS0015.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2021.624302/full#supplementary-material.

REFERENCES

Ayers, J. I., Giasson, B. I., and Borchelt, D. R. (2018). Prion-like spreading in tauopathies. *Biol. Psychiatry* 83, 337–346. doi:10.1016/j.biopsych.2017.04.003

Ballatore, C., Lee, V. M.-Y., and Trojanowski, J. Q. (2007). Tau-mediated neurodegeneration in Alzheimer's disease and related disorders. *Nat. Rev. Neurosci.* 8, 663–672. doi:10.1038/nrn2194

Battisti, A., Ciasca, G., Grottesi, A., Bianconi, A., and Tenenbaum, A. (2012). Temporary secondary structures in tau, an intrinsically disordered protein. *Mol. Simul.* 38, 525–533. doi:10.1080/08927022.2011.633347

Berendsen, H. J., Postma, J. v., van Gunsteren, W. F., DiNola, A., and Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81, 3684–3690. doi:10.1063/1.448118

Best, R. B., Hummer, G., and Eaton, W. A. (2013). Native contacts determine protein folding mechanisms in atomistic simulations. *Proc. Natl. Acad. Sci.* 110, 17874–17879. doi:10.1073/pnas.1311599110

Bolós, M., Perea, J. R., and Avila, J. (2017). Alzheimer's disease as an inflammatory disease. *Biomol. Concepts* 8, 37–43. doi:10.1515/bmc-2016-0029

Bonomi, M., Bussi, G., Camilloni, C., Tribello, G. A., Banáš, P., Barducci, A., et al. (2019). Promoting transparency and reproducibility in enhanced molecular simulations. *Nat. Methods* 16, 670–673. doi:10.1038/s41592-019-0506-8

Bussi, G., Donadio, D., and Parrinello, M. (2007). Canonical sampling through velocity rescaling. *J. Chem. Phys.* 126, 014101. doi:10.1063/1.2408420

Crowther, R. (1991). Straight and paired helical filaments in Alzheimer disease have a common structural unit. *Proc. Natl. Acad. Sci.* 88, 2288–2292. doi:10.1073/pnas.88.6.2288

Derreumaux, P., Man, V. H., Wang, J., and Nguyen, P. H. (2020). Tau R3–R4 domain dimer of the wild type and phosphorylated ser356 sequences. *I.* In solution by atomistic simulations. *J. Phys. Chem. B* 124, 2975–2983. doi:10.1021/acs.jpcb.0c00574

Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., and Pedersen, L. G. (1995). A smooth particle mesh ewald method. *J. Chem. Phys.* 103, 8577–8593. doi:10.1063/1.470117

Fitzpatrick, A. W., Falcon, B., He, S., Murzin, A. G., Murshudov, G., Garringer, H. J., et al. (2017). Cryo-em structures of tau filaments from Alzheimer's disease. *Nature* 547, 185–190. doi:10.1038/nature23002

Friedhoff, P., Biernat, J., Heberle, J., Mandelkow, E., Mandelkow, E., et al. (2000). Assembly of tau protein into Alzheimer paired helical filaments depends on a local sequence motif [(306) vqivyk (311)] forming beta structure. *Proc. Natl. Acad. Sci. U.S.A.* 97, 5129–5134. doi:10.1073/pnas.97.10.5129

Götz, J., Ittner, L. M., and Kins, S. (2006). Do axonal defects in tau and amyloid precursor protein transgenic animals model axonopathy in Alzheimer's disease? *J. Neurochem.* 98, 993–1006. doi:10.1111/j.1471-4159.2006.03955.x

Ganguly, P., Do, T. D., Larini, L., LaPointe, N. E., Sercel, A. J., Shade, M. F., et al. (2015). Tau assembly: the dominant role of PHF6 (VQIVYK) in microtubule binding region repeat R3. *J. Phys. Chem. B* 119, 4582–4593. doi:10.1021/acs.jpcb.5b00175

Goedert, M., Eisenberg, D. S., and Crowther, R. A. (2017a). Propagation of tau aggregates and neurodegeneration. *Annu. Rev. Neurosci.* 40, 189–210. doi:10.1146/annurev-neuro-072116-031153

Goedert, M., Masuda-Suzukake, M., and Falcon, B. (2017b). Like prions: the propagation of aggregated tau and α-synuclein in neurodegeneration. *Brain* 140, 266–278. doi:10.1093/brain/aww230

Goedert, M., and Spillantini, M. G. (2006). A century of Alzheimer's disease. *Science* 314, 777–781. doi:10.1126/science.1132814

Gonzalez-Billault, C., Engelke, M., Jimenez-Mateos, E., Wandosell, F., Caceres, A., and Avila, J. (2002). Participation of structural microtubule-associated proteins (maps) in the development of neuronal polarity. *J. Neurosci. Res.* 67, 713–719. doi:10.1002/jnr.10161

Govaerts, C., Wille, H., Prusiner, S. B., and Cohen, F. E. (2004). Evidence for assembly of prions with left-handed β-helices into trimers. *Proc. Natl. Acad. Sci.* 101, 8342–8347. doi:10.1073/pnas.0402254101

Grubmüller, H., Heymann, B., and Tavan, P. (1996). Ligand binding: molecular mechanics calculation of the streptavidin-biotin rupture force. *Science* 271, 997–999. doi:10.1126/science.271.5251.997

Guo, J.-P., Arai, T., Miklossy, J., and McGeer, P. L. (2006). Aβ and tau form soluble complexes that may promote self aggregation of both into the insoluble forms observed in Alzheimer's disease. *Proc. Natl. Acad. Sci.* 103, 1953–1958. doi:10.1073/pnas.0509386103

Haj-Yahya, M., Gopinath, P., Rajasekhar, K., Mirbaha, H., Diamond, M. I., and Lashuel, H. A. (2020). Site-specific hyperphosphorylation inhibits, rather than promotes, tau fibrillization, seeding capacity, and its microtubule binding. *Angew. Chem.* 132, 4088–4096. doi:10.1002/anie.201913001

He, Z., Guo, J. L., McBride, J. D., Narasimhan, S., Kim, H., Changolkar, L., et al. (2018). Amyloid-β plaques enhance Alzheimer's brain tau-seeded pathologies by facilitating neuritic plaque tau aggregation. *Nat. Med.* 24, 29. doi:10.1038/nm.4443

Hess, B., Kutzner, C., Van Der Spoel, D., and Lindahl, E. (2008). Gromacs 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theor. Comput.* 4, 435–447. doi:10.1021/ct700301q

Huang, J., Rauscher, S., Nawrocki, G., Ran, T., Feig, M., de Groot, B. L., et al. (2017). Charmm36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* 14, 71–73. doi:10.1038/nmeth.4067

Hub, J. S., and de Groot, B. L. (2006). Does CO_2 permeate through aquaporin-1? *Biophysical J.* 91, 842–848. doi:10.1529/biophysj.106.081406

Hub, J. S., De Groot, B. L., and Van Der Spoel, D. (2010). g_wham-a free weighted histogram analysis implementation including robust error and autocorrelation estimates. *J. Chem. Theor. Comput.* 6, 3713–3720. doi:10.1021/ct100494z

Ittner, L. M., and Götz, J. (2011). Amyloid-β and tau-a toxic pas de deux in Alzheimer's disease. *Nat. Rev. Neurosci.* 12, 67–72. doi:10.1038/nrn2967

Izrailev, S., Stepaniants, S., Isralewitz, B., Kosztin, D., Lu, H., Molnar, F., et al. (1999). "Steered molecular dynamics," in *Computational molecular dynamics: challenges, methods, ideas*. Editors P. Deuflhard, J. Hermans, B. Leimkuhler, A. E. Mark, S. Reich, and R. D. Skeel (Berlin, Heidelberg: Springer), 39–65. doi:10.1007/978-3-642-58360-5_2

Jakubec, D., and Vondrášek, J. (2020). Efficient estimation of absolute binding free energy for a homeodomain–dna complex from nonequilibrium pulling simulations. *J. Chem. Theor. Comput.* 16, 2034–2041. doi:10.1021/acs.jctc.0c00006

Jia, Z., Beugelsdijk, A., Chen, J., and Schmit, J. D. (2017). The levinthal problem in amyloid aggregation: sampling of a flat reaction space. *J. Phys. Chem. B* 121, 1576–1586. doi:10.1021/acs.jpcb.7b00253

Jia, Z., Schmit, J. D., and Chen, J. (2020). Amyloid assembly is dominated by misregistered kinetic traps on an unbiased energy landscape. *Proc. Natl. Acad. Sci.* 117, 10322–10328. doi:10.1073/pnas.1911153117

Jucker, M., and Walker, L. C. (2013). Self-propagation of pathogenic protein aggregates in neurodegenerative diseases. *Nature* 501, 45–51. doi:10.1038/nature12481

Kidd, M. (1963). Paired helical filaments in electron microscopy of Alzheimer's disease. *Nature* 197, 192–193. doi:10.1038/197192b0

Konzack, S., Thies, E., Marx, A., Mandelkow, E.-M., and Mandelkow, E. (2007). Swimming against the tide: mobility of the microtubule-associated protein tau in neurons. *J. Neurosci.* 27, 9916–9927. doi:10.1523/JNEUROSCI.0927-07.2007

Laio, A., and Parrinello, M. (2002). Escaping free-energy minima. *Proc. Natl. Acad. Sci.* 99, 12562–12566. doi:10.1073/pnas.202427399

Larini, L., Gessel, M. M., LaPointe, N. E., Do, T. D., Bowers, M. T., Feinstein, S. C., et al. (2013). Initiation of assembly of tau (273–284) and its ΔK280 mutant: an experimental and computational study. *Phys. Chem. Chem. Phys.* 15, 8916–8928. doi:10.1039/c3cp00063j

Lemkul, J. A., and Bevan, D. R. (2010). Assessing the stability of Alzheimer's amyloid protofibrils using molecular dynamics. *J. Phys. Chem. B* 114, 1652–1660. doi:10.1021/jp9110794

Levine, Z. A., Larini, L., LaPointe, N. E., Feinstein, S. C., and Shea, J.-E. (2015). Regulation and aggregation of intrinsically disordered peptides. *Proc. Natl. Acad. Sci.* 112, 2758–2763. doi:10.1073/pnas.1418155112

Li, C., and Götz, J. (2017). Tau-based therapies in neurodegeneration: opportunities and challenges. *Nat. Rev. Drug Discov.* 16, 863–883. doi:10.1038/nrd.2017.155

Li, X., Dong, X., Wei, G., Margittai, M., Nussinov, R., and Ma, B. (2018). The distinct structural preferences of tau protein repeat domains. *Chem. Commun.* 54, 5700–5703. doi:10.1039/c8cc01263f

Lindahl, E., Abraham, M., Hess, B., and van der Spoel, D. (2019). GROMACS 2019.4 Manual (Version 2019.4). *Zenodo*. doi:10.5281/zenodo.346041510.5281/zenodo.3460414

Liu, H., Liu, X., Zhou, S., An, X., Liu, H., and Yao, X. (2019). Disclosing the template-induced misfolding mechanism of tau protein by studying the dissociation of the boundary chain from the formed tau fibril based on a steered molecular dynamics simulation. *ACS Chem. Neurosci.* 10, 1854–1865. doi:10.1021/acschemneuro.8b00732

Luo, Y., Ma, B., Nussinov, R., and Wei, G. (2014). Structural insight into tau protein's paradox of intrinsically disordered behavior, self-acetylation activity, and aggregation. *J. Phys. Chem. Lett.* 5, 3026–3031. doi:10.1021/jz501457f

Mudher, A., Colin, M., Dujardin, S., Medina, M., Dewachter, I., Naini, S. M. A., et al. (2017). What is the evidence that tau pathology spreads through prion-like propagation? *Acta Neuropathol. Commun.* 5, 99. doi:10.1186/s40478-017-0488-7

Noble, W., Pooler, A. M., and Hanger, D. P. (2011). Advances in tau-based drug discovery. *Expert Opin. Drug Discov.* 6, 797–810. doi:10.1517/17460441.2011.586690

O'Nuallain, B., Shivaprasad, S., Kheterpal, I., and Wetzel, R. (2005). Thermodynamics of aβ (1–40) amyloid fibril elongation. *Biochemistry* 44, 12709–12718. doi:10.1021/bi050927h

Parrinello, M., and Rahman, A. (1981). Polymorphic transitions in single crystals: a new molecular dynamics method. *J. Appl. Phys.* 52, 7182–7190. doi:10.1063/1.328693

Polanco, J. C., Li, C., Bodea, L.-G., Martinez-Marmol, R., Meunier, F. A., and Götz, J. (2018). Amyloid-β and tau complexity–towards improved biomarkers and targeted therapies. *Nat. Rev. Neurol.* 14, 22. doi:10.1038/nrneurol.2017.162

Popov, K. I., Makepeace, K. A., Petrotchenko, E. V., Dokholyan, N. V., and Borchers, C. H. (2019). Insight into the structure of the "unstructured" tau protein. *Structure* 27, 1710–1715. doi:10.1016/j.str.2019.09.003

Rhein, V., Song, X., Wiesner, A., Ittner, L. M., Baysang, G., Meier, F., et al. (2009). Amyloid-β and tau synergistically impair the oxidative phosphorylation system in triple transgenic Alzheimer's disease mice. *Proc. Natl. Acad. Sci.* 106, 20057–20062. doi:10.1073/pnas.0905529106

Roux, B. (1995). The calculation of the potential of mean force using computer simulations. *Comp. Phys. Commun.* 91, 275–282. doi:10.1016/0010-4655(95)00053-i

Schrödinger, LLC (2015). *The PyMOL molecular graphics system, version 1.8*. Schrödinger, LLC.

Schwierz, N., Frost, C. V., Geissler, P. L., and Zacharias, M. (2016). Dynamics of seeded aβ40-fibril growth from atomistic molecular dynamics simulations: kinetic trapping and reduced water mobility in the locking step. *J. Am. Chem. Soc.* 138, 527–539. doi:10.1021/jacs.5b08717

Selkoe, D. J., and Hardy, J. (2016). The amyloid hypothesis of Alzheimer's disease at 25 years. *EMBO Mol. Med.* 8, 595–608. doi:10.15252/emmm.201606210

Shaw, D. E., Grossman, J., Bank, J. A., Batson, B., Butts, J. A., Chao, J. C., et al. (2014). "Anton 2: raising the bar for performance and programmability in a special-purpose molecular dynamics supercomputer," in SC'14: proceedings of the international conference for high performance computing, networking, storage and analysis, New Orleans, LA, November 6–21, 2014 (IEEE), 41–53. doi:10.1109/SC.2014.9

Smit, F. X., Luiken, J. A., and Bolhuis, P. G. (2017). Primary fibril nucleation of aggregation prone tau fragments PHF6 and PHF6*. *J. Phys. Chem. B* 121, 3250–3261. doi:10.1021/acs.jpcb.6b07045

Strang, K. H., Croft, C. L., Sorrentino, Z. A., Chakrabarty, P., Golde, T. E., and Giasson, B. I. (2018). Distinct differences in prion-like seeding and aggregation between tau protein variants provide mechanistic insights into tauopathies. *J. Biol. Chem.* 293, 2408–2421. doi:10.1074/jbc.M117.815357

Sunde, M., Serpell, L. C., Bartlam, M., Fraser, P. E., Pepys, M. B., and Blake, C. C. (1997). Common core structure of amyloid fibrils by synchrotron x-ray diffraction. *J. Mol. Biol.* 273, 729–739. doi:10.1006/jmbi.1997.1348

Takeda, T., and Klimov, D. K. (2009a). Probing the effect of amino-terminal truncation for aβ1- 40 peptides. *J. Phys. Chem. B* 113, 6692–6702. doi:10.1021/jp9016773

Takeda, T., and Klimov, D. K. (2009b). Replica exchange simulations of the thermodynamics of aβ fibril growth. *Biophys. J.* 96, 442–452. doi:10.1016/j.bpj.2008.10.008

Torrie, G. M., and Valleau, J. P. (1977). Nonphysical sampling distributions in Monte Carlo free-energy estimation: umbrella sampling. *J. Comput. Phys.* 23, 187–199. doi:10.1016/0021-9991(77)90121-8

Towns, J., Cockerill, T., Dahan, M., Foster, I., Gaither, K., Grimshaw, A., et al. (2014). XSEDE: accelerating scientific discovery. *Comput. Sci. Eng.* 16, 62–74. doi:10.1109/MCSE.2014.80

Tribello, G. A., Bonomi, M., Branduardi, D., Camilloni, C., and Bussi, G. (2014). Plumed 2: new feathers for an old bird. *Comp. Phys. Commun.* 185, 604–613. doi:10.1016/j.cpc.2013.09.018

Valsson, O., Tiwary, P., and Parrinello, M. (2016). Enhancing important fluctuations: rare events and metadynamics from a conceptual viewpoint. *Annu. Rev. Phys. Chem.* 67, 159–184. doi:10.1146/annurev-physchem-040215-112229

Warnecke, A., Sandalova, T., Achour, A., and Harris, R. A. (2014). Pytms: a useful pymol plugin for modeling common post-translational modifications. *BMC Bioinformatics* 15, 1–12. doi:10.1186/s12859-014-0370-6

Xie, C., Soeda, Y., and Shinzaki, Y., In, Y., Tomoo, K., Ihara, Y., et al. (2015). Identification of key amino acids responsible for the distinct aggregation properties of microtubule-associated protein 2 and tau. *J. Neurochem.* 135, 19–26. doi:10.1111/jnc.13228

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Molecular Biosciences

ORIGINAL RESEARCH
published: 12 April 2021
doi: 10.3389/fmolb.2021.624679

Interactions Between Nucleosomes: From Atomistic Simulation to Polymer Model

Chengwei Zhang[1,2,3,4] *and Jing Huang*[2,3,4]*

[1]College of Life Sciences, Zhejiang University, Hangzhou, China, [2]Key Laboratory of Structural Biology of Zhejiang Province, School of Life Sciences, Westlake University, Hangzhou, China, [3]Westlake Laboratory of Life Sciences and Biomedicine, Hangzhou, China, [4]Institute of Biology, Westlake Institute for Advanced Study, Hangzhou, China

The organization of genomes in space and time dimension plays an important role in gene expression and regulation. Chromatin folding occurs in a dynamic, structured way that is subject to biophysical rules and biological processes. Nucleosomes are the basic unit of chromatin in living cells, and here we report on the effective interactions between two nucleosomes in physiological conditions using explicit-solvent all-atom simulations. Free energy landscapes derived from umbrella sampling simulations agree well with recent experimental and simulation results. Our simulations reveal the atomistic details of the interactions between nucleosomes in solution and can be used for constructing the coarse-grained model for chromatin in a bottom-up manner.

Keywords: chromatin, nucleosome, molecular dynamics simulation, umbrella sampling, potential of mean force, coarse-grain

OPEN ACCESS

Edited by:
Xiakun Chu,
Stony Brook University, United States

Reviewed by:
Wenjun Xie,
Massachusetts Institute of Technology, United States
Vinicius Contessoto,
Rice University, United States

***Correspondence:**
Jing Huang
huangjing@westlake.edu.cn

Specialty section:
This article was submitted to Biological Modeling and Simulation, a section of the journal Frontiers in Molecular Biosciences

Received: *31 October 2020*
Accepted: *09 February 2021*
Published: *12 April 2021*

Citation:
Zhang C and Huang J (2021) Interactions Between Nucleosomes: From Atomistic Simulation to Polymer Model. Front. Mol. Biosci. 8:624679. doi: 10.3389/fmolb.2021.624679

INTRODUCTION

Chromatin is highly compacted and condensed into the small space of the nucleus (Goloborodko et al., 2016b; Finn and Misteli, 2019). Human diploid genome contains about six billion DNA base pairs, and it will be approximately 2 m long if fully extended in a double helix (Fraser et al., 2015; Dans et al., 2016; Saurabh et al., 2016). In contrast, the size of the nucleus is only a few micrometers, and much remains to be understood that how chromatin fit into the nucleus in such a compacted and condensed way. The organization and dynamics of chromatin in the nucleus are found to be neither random nor stochastic, instead it is well defined and regulates the gene expression intricately.

The organization and folding processes of chromatin can be mainly divided into four layers: 1) The antiparallel double helical and right-handed B-DNA structure with each base pair rising up about 3.4 Å along the helical axis, which makes DNA very stable while it provides potential for the binding of proteins. 2) About 146 base pairs wrap around a histone octamer (also called histone core), including two copies each of the four histone core proteins (H2A, H2B, H3, and H4) in a left-handed superhelix, which constitute a nucleosome. Nucleosomes are the basic units of chromatin and provide controlled accessibility for DNA-binding proteins such as transcription machines and structural maintenance of chromosome (SMC) complexes. The height of a nucleosome is about 55 Å, so the formation of nucleosome compacts DNA by about 9 times as 146 base pairs × 3.4Å ÷ 55Å ≈ 9. Even though the DNA sequences binding to histone core are nonspecific, the binding affinity between some DNA sequences and histone core are higher than others (Field et al., 2008; Teif et al., 2012; Teif and Clarkson, 2019). About 75–90% of the genome are organized in the form of nucleosomes (Field et al., 2008). 3) Under the view of electron microscopes, chromatin appears as "beads on a string," in which beads correspond to nucleosomes and the string between beads

corresponds to the double helical DNA called linker. The string-like chromatin is shaped by loops (Alipour and Marko, 2012; Rao et al., 2014; Goloborodko et al., 2016a, Goloborodko et al., 2016b), topologically associating domains (TADs) (Dixon et al., 2012; Schwarzer et al., 2017; Krietenstein et al., 2020), and A/B compartments (Dekker et al., 2002; Zhao et al., 2006; Lieberman-aiden et al., 2009; Dekker et al., 2013). Because chromatins are fluid and dynamic, it is appropriate to use structural ensembles to describe the chromatin structure. While TADs and compartments fluctuate at a single-cell level, population-averaged TADs and compartments are tissue specific, meaning that their patterns are similar and conserved in one cell line and have high variability across different cell lines (Cheng et al., 2020; Contessoto et al., 2021). The variability of the organization and dynamics in this layer suggests direct impacts on the expression and regulation of genes (Finn and Misteli, 2019). 4) Each chromatin has its own territory in the nucleus, in which locus from the same chromatin have higher probability to localize together and form exclusive subregions (Stam et al., 2019). In the normal cell nucleus, euchromatins, which are gene-rich and transcriptionally active segments, cluster together in the interior, whereas heterochromatins, which are gene-poor and silenced, cluster together in the nuclear envelope. Visualization in real time and simulation in silicon provides a comprehensive understanding of the phase separation and chromosome territory (Liu et al., 2020; Oliveira Junior et al., 2020; Su et al., 2020). Overall, chromatin is organized in a highly hierarchical architecture. Each of these four layers is also highly dynamic, which provide potential to regulate important biological processes, in particular the gene expression.

The human genome is the blueprint of life consisting of more than 20,000 genes and millions of regulatory candidate elements (Dunham et al., 2012). Despite intensive efforts, it is far from complete to understand how these elements function and interact with each other in the spatial and temporal dimensions to regulate gene expression, as cells with identical DNA sequence can function differently. The three-dimensional structure and dynamics of chromatin play critical roles in bringing into physical proximity the regulatory elements with target genes across hundreds of kilobases or even megabase distance. The abnormal chromatin organization leads to the occurrence of diseases, in particular cancer (Goes et al., 2011; Meaburn et al., 2016). 3C-based methods such as Hi-C (Dekker et al., 2002; Lieberman-aiden et al., 2009) and FISH (Fraser et al., 2015; Fudenberg and Imakaev, 2017) are two mainstream experimental techniques to probe the organization and dynamics of chromatin. However, how to integrate and interpret the Hi-C and the FISH measurements remains challenging, and in some cases, they can even lead to contradictory results (Fudenberg and Imakaev, 2017).

Complementary to experiments, computer simulations provide unprecedented resolution to investigate the folding of chromatin. Polymer model theory can be used to study the organization and dynamics of chromatin fibers at genome scale. Coarse-grained (CG) models of nucleosome can be used to study the interaction and dynamics of nucleosomes array which is the local subregion of the chromatin fiber. All-atom molecular dynamics (MD) simulations of nucleosome can provide further detailed information at the atomistic level. Combining these models for multiscale simulations would be useful to enhance our understanding of chromatin folding processes. In the polymer model (Philip et al., 1993; Münkel and Langowski, 1998), chromatin is represented as polymers consisting of different monomers connected by harmonic bonds, and different persistence lengths are chosen according to the compaction ratio, for example, how many base pairs are coarse-grained into one monomer. The models also account for particular interactions related to the biological activities of chromatin, for example, loop extrusion and compartmental segregation. Recently, two sophisticated computational models with slightly different potential energy function formulas have been developed (Di Pierro et al., 2016; Fudenberg et al., 2016), and shed light on the mechanisms underlying the folding and organization of chromatin (Gibcus et al., 2018; Mirny et al., 2019; Stam et al., 2019; Banigan et al., 2020).

The quality of polymer model simulations depends critically on the accuracy of their potential energy functions, which are a summation of different pairwise interaction terms. Previous researches (Sanborn et al., 2015; Fudenberg et al., 2016; Goloborodko et al., 2016a; Goloborodko et al., 2016b) typically used grid search strategy to optimize the parameters in potential energy functions, which means trial-and-error until simulated properties match experiments. If there are four parameters in one energy function and each parameter have 10 grids, 10,000 sets of parameter combination are tried from which one will be selected, the one most consistent with the Hi-C contact map. In addition to grid search strategy, the maximum entropy principle (Zhang and Wolynes, 2015; Di Pierro et al., 2016) was also applied to derive the potential functions using the experimental contact map information as inputs. A prior knowledge is needed to derive models from the maximum entropy principle, where it is used by grid search strategies as the criterion to select the best parameter combination. An alternative way to construct coarse-grained models is using the "bottom-up" or "*ab initio*" approach. One can derive coarse-grained physical potentials from more detailed simulations, for example, explicit-solvent all-atom MD simulations (Noid et al., 2008; Li et al., 2016). Here, we present a first step toward constructing a multiscale polymer model for chromatin based on atomistic simulations of two nucleosomes.

Nucleosomes are fundamental units for chromatin folding and the main carrier of epigenetic marks, so we coarse-grain one nucleosome as one bead in the coarse-grained model. Since the high-resolution X-ray structure of the nucleosome was determined in 1997 (Luger et al., 1997), there have been many studies on the properties of nucleosomes (Portela and Esteller, 2010; Biswas et al., 2013; Feng and Li, 2017; Zhou et al., 2019). Recently, Funke et al. (2016) employed force spectrometer and single-particle electron microscopy to measure the forces and interaction profiles between nucleosomes by placing two nucleosomes close to each other in a variety of defined relative orientations and recording the frequency of their distances.

Multiscale simulations revealed that increased secondary structure resulting from acetylation of H4 tail has an important effect on the rigidification and also impaired the interactions between stacked nucleosomes (Collepardo-Guevara et al., 2015). The effect of H4 tail on stabilizing the stacked nucleosome is also validated by a recent first atomistic simulation of stacked nucleosomes using conventional and steered MD simulations (Saurabh et al., 2016). Ishida and Kono (2017) explored the energy landscape of two stacked nucleosomes using umbrella sampling with nucleosomes restrained at a few distinctive orientations. Moller et al. (2019) evaluated the anisotropic energy landscape of stacked nucleosomes across a variety of parameters in configurational and environmental space using residue coarse-grained simulations. Lequieu et al. (2019) constructed a coarse-grained multiscale model of chromatin by mapping the energy landscape of stacked nucleosomes to a reduced coarse-grained topology. Interestingly, Spakowitz and co-workers investigated the effects of epigenetic modifications, especially methylation of lysine-9 of histone H3, on the organization and dynamics of chromatin using polymer model (MacPherson et al., 2018; Sandholtz et al., 2020).

Nucleosome positioning, referring to the location of the nucleosome dyads in linear DNA, regulates the accessibility of DNA to DNA-bound proteins (Portela and Esteller, 2010). Adjacent nucleosomes in sequence are connected by linkers, whose lengths are about tens of base pairs in eukaryotes. The average radius of folded protein is about a few nanometers (Milo et al., 2009), whereas the compartment structures are across megabases. So the process of compartmentalization is self-organized by the interactions between nucleosomes, rather than mediated by proteins. Modifications of nucleosome, such as DNA methylation and histone modification, change the properties of nucleosomes and thus alter their interaction landscapes. Theoretical modeling of the interactions between two nucleosomes, in particular its dependence on the distance and relative orientation of nucleosomes, are crucial to our understanding of chromatin folding.

In this work, we present all-atom MD simulations of two nucleosomes interacting with each other in physiologically relevant explicit-solvent environment and analyze their interaction landscapes in the context of the 30-nm chromatin model. We consider two simulation systems, one containing two linked nucleosomes, whereas the other one containing two unlinked nucleosomes proximal in space. Atomistic simulation results of these two systems could be used to determine the function forms and parameters in the model, such as the diameter of the beads, the strength of the harmonic bonds connecting beads, and the strength of the weak attraction between beads in different chromatin states. The manuscript is organized as follows: Details of MD simulations and trajectory analysis will be provided in the Methods section. In the Results section, unbiased MD simulation results will first be presented, followed by the potential of mean force (PMF) and coarse-graining calculations. The manuscript ends with a short discussion and conclusion.

METHODS

System Setup

The interaction between two nucleosomes can be naturally divided into two types, one type representing the interactions between nucleosomes that are connected by the linker DNA and the other type representing the interactions between two spatially adjacent nucleosomes that are stacked together with no connecting linker. Accordingly, two simulation systems were set up as shown in **Figure 1**. The initial structures were taken from that of classical 30 nm fiber (pdb id: 6hkt) (Garcia-Saez et al., 2018). Although its resolution is relatively low (9.7 Å), it includes the information about the nucleosomal stacking and packing patterns which are important for building the simulated systems to model the interactions between nucleosomes. We note that the histone tails that are missing in the crystal structure are not modeled in the simulation systems.

For the linked-nucleosomes (LN) system, two nucleosomes and their corresponding linker DNA base pairs were extracted from the experimental structure and solvated in an explicit TIP3P (Jorgensen et al., 1983) water box with dimensions of 463 Å × 142 Å × 110 Å (**Figure 1B**). CHARMM (Brooks et al., 1983; Brooks et al., 2009) was used to build the missing hydrogen atom coordinates and patch protein and nucleic acid terminals. 150 mM KCl was added with additional cations to neutralize the system. In total, the LN system contains 676,742 atoms which are composed of 16 protein chains, two DNA chains, 209,471 water molecules, 1,140 K^+ ions, and 594 Cl^- ions.

The unlinked-nucleosome (ULN) system contains two stacked nucleosomes plus 15 flanking base pairs extracted from the experimental 30 nm fiber structure. A similar procedure was used to solvate the system in a cubic water box with dimensions of 182 Å × 182 Å × 182 Å (**Figure 1C**). In total, the ULN system has 558,630 atoms composed of 16 protein chains, four DNA chains, 170,064 water molecules, 1,039 K^+ ions, and 485 Cl^- ions. Proteins and DNAs were modeled by the CHARMM36m protein force field (Huang et al., 2017) and the CHARMM36 nucleic acid force field (Hart et al., 2012), respectively.

Molecular Dynamics Simulations

MD simulations were performed using OpenMM (Eastman and Pande, 2010; Eastman et al., 2017) with the isothermal–isobaric (NPT) ensemble. Periodic boundary condition (PBC) was applied and particle mesh Ewald (PME) (Essmann et al., 1995) was used to compute all the nonbonded interactions with a real space cutoff at 9 Å. We noted that both electrostatics and van der Waals interactions are fully accounted for with no truncations, as the latter were treated by the recently developed LJ-PME method (Wennberg et al., 2015). All hydrogen-containing bonds were constrained by the SETTLE algorithm (Miyamoto and Kollman, 1992) and the Verlet integrator was used with a time step of 2 fs The temperature was maintained at 303.15 K using the Andersen thermostat (Andersen, 1980) with a damping coefficient of 1 ps^{-1}. A Monte Carlo barostat (Aqvist et al., 2004) was used to maintain the pressure at 1 atm by attempting to change the box dimension

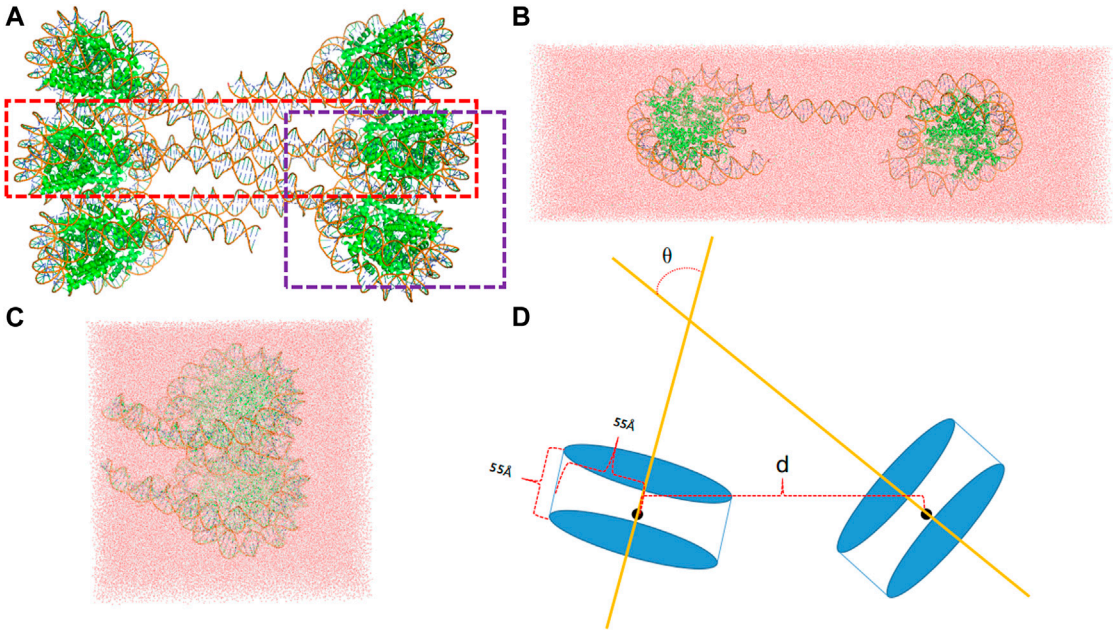


FIGURE 1 | (A) The template structure (pdb id: 6hkt) for the setup of simulation system includes six nucleosomes with a flat two-start helix with uniform nucleosomal stacking interfaces and an uncondensed nucleosome packing density. The LN system is marked by a red rectangle, whereas the ULN by a purple rectangle. **(B)** A snapshot of the LN system after 10 ns MD simulation. **(C)** A snapshot of the ULN system after 10 ns MD simulation **(D)** The distance, *d*, defined using the centers of nucleosomes and the angle, *θ*, defined using the superhelical axes of nucleosomes.

every 25 steps. Both LN and ULN systems were simulated to 1 μs, and frames were saved every 5,000 steps (10 ps).

Umbrella Sampling and Potential of Mean Force Calculations

To investigate the free energy landscape between nucleosomes, advanced simulation techniques need to be employed. Here, we perform umbrella sampling simulations using the distance between the two nucleosomes as the reaction coordinate.

$$H_{umb} = H_0 + \frac{1}{2}k(d - d_0)^2, \quad (1)$$

where the system Hamiltonian H_0 is biased by a harmonic potential that restrains the select reaction coordinate d at a certain value d_0. Biased MD simulations employing **Eq. 1** with different d_0 values (windows) can be carried out to enhance the sampling of events hindered by free energy barriers. The initial configuration for each window was extracted from the unbiased MD simulations with d, most close to the targeted d_0 value. The NVT ensemble was used for umbrella sampling. The system was equilibrated for 1 ns with no biased potential added and then subject to 30 ns umbrella sampling simulations.

Umbrella sampling simulations were carried out using OpenMM with a plugin for PLUMED (Tribello et al., 2014). The center of a nucleosome was defined as the center of mass of the phosphorus atoms of DNA wrapped on the nucleosome histone cores (LN system) or the center of mass of protein C_α atoms of the histone cores (ULN system). The reaction coordinate d was then defined as the distance between the centers of nucleosomes. For the LN system, 127 windows were used in total, d_0 ranging from 175 to 250 Å with an interval of 1 Å and additional windows ranging from 187.5 to 237.5 Å with an interval of 1 Å. For the ULN system, 68 windows were used in total, d_0 ranging from 60 to 91 Å with an interval of 0.5 Å and additional five windows ranging from 55 to 59 Å with an interval of 1 Å. For the LN system, 5 kJ/mol/Å^2 was selected as the value of k, whereas 10 kJ/mol/Å^2 was used in alternative windows for the ULN system. Good phase space overlap between windows was achieved (**Supplementary Figures S1, S2**). The potential of mean force profiles were calculated using the weighted histogram analysis method (WHAM) (Grossfield, 2020–9).

RESULTS

Unbiased Molecular Dynamics Simulations

Unbiased MD simulations were carried out for 1,000 ns for both LN and ULN systems. For individual nucleosome, the binding between DNA and histone core was very stable as indicated by their root mean square deviations (RMSDs) being around 4 Å with respect to the initial structures (**Supplementary Figure S3**). The relative motion between nucleosomes, in contrast, was highly dynamic with respect to both their distance and orientation over the microsecond timescale (**Figure 2**). For nucleosomes connected by the linker, their distance d varied between 165 and 239 Å. The distance increased to 238 Å in the first 200 ns of the simulations and gradually decreased to 170 Å after 400 ns.

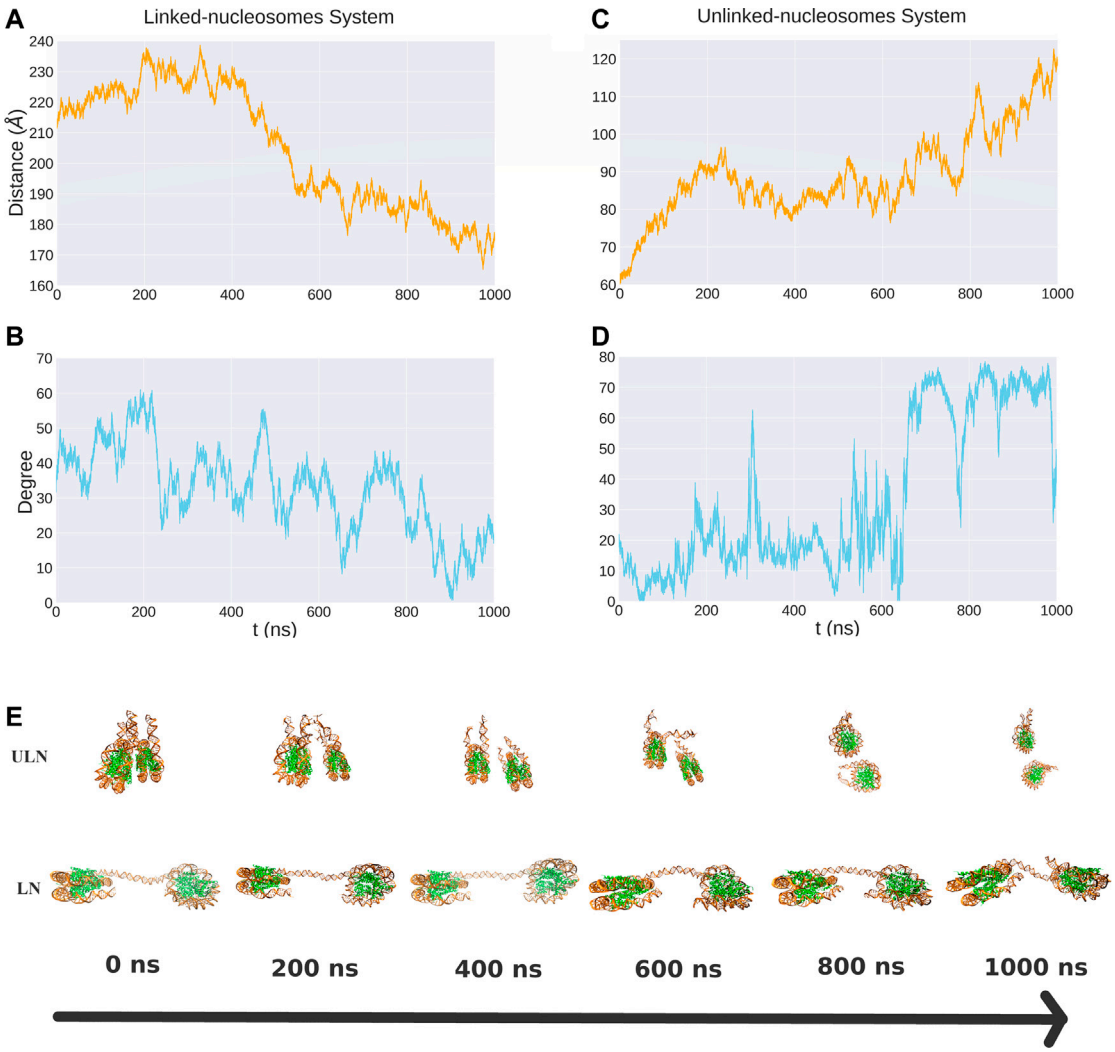


FIGURE 2 | The distance and relative orientation between two nucleosomes in the LN **(A, B)** and ULN **(C, D)** systems from 1 μs unbiased MD simulations. Also shown are snapshots corresponding to different time points **(E)**.

For unlinked nucleosomes, *d* increased almost linearly from the starting value of 60 to about 90 Å at the first 200 ns, and then fluctuated between 75 to 100 Å in the following 600 ns, and rose to 120 Å after 800 ns. This suggests that the equilibrium distance of two free nucleosomes in aqueous environment is much larger than the 62 Å in the crystal structure of 30 nm fiber, which are impacted by the crystal packing and the stacking of nucleosomes. Based on our simulations, the relative velocity between two nucleosomes can be estimated to be about 0.1 Å/ns or 0.01 m/s.

We also analyzed the relative orientation between the two nucleosomes, characterizing it using the angle between the superhelical axes of nucleosomes (**Figure 1D**). There were large variations of angle in 1,000 ns MD simulations for both LN and ULN systems. The length of the linker DNA in the linked-nucleosomes system is 40 base pairs, smaller than the persistence length of double helix DNA (about 150 base pairs). A weak correlation between orientation and distance *d* was observed in both the LN and ULN systems. In the LN system, the relative orientation angle has a tendency to decrease when *d* decreases. In the ULN system, the angle has a tendency to increase with the distance. The frequency of angle variation is significantly higher than that of distance. During 1,000 ns MD simulation, distance went down in the LN or up in the ULN system, whereas the angle vibrated up and down regularly.

Interaction Free Energy Landscapes

To study the effective interactions between nucleosomes in both systems, umbrella sampling simulations were performed to compute the PMFs as a function of distance between the nucleosomes. In general, the free energy profiles are shallow and flat, consistent with the hypothesis that nucleosomes are highly dynamic. The convergence of PMF is relatively good for both LN and ULN systems as indicated by the statistical uncertainties from the Monte Carlo bootstrapping calculations

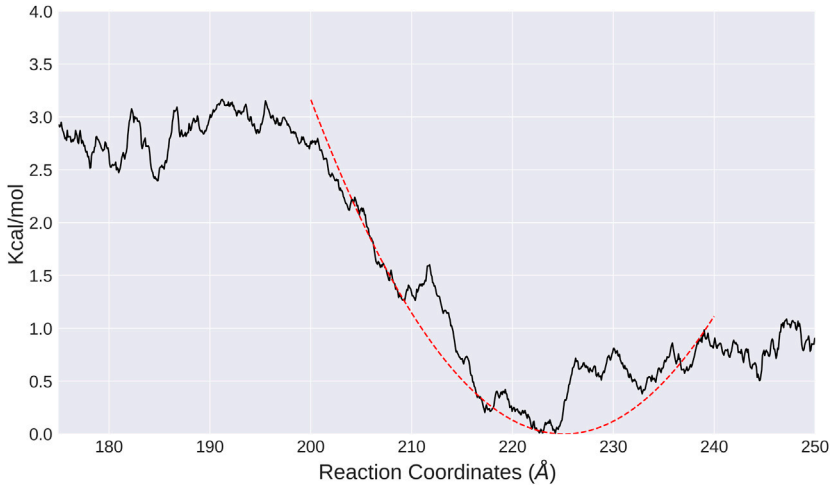


FIGURE 3 | PMF calculated from umbrella sampling in the LN system along the distance between two nucleosomes (black solid line). The fitted harmonic potential function is shown as red dotted line.

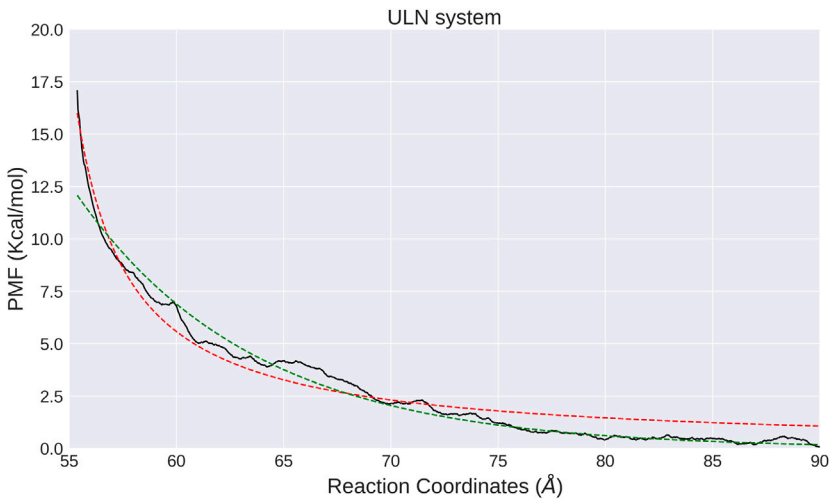


FIGURE 4 | PMF calculated from umbrella sampling in the ULN system along the distance between two nucleosomes (black solid lines). The fitted exponential and shifted Coulomb potential functions are shown as red dotted line and green solid line, respectively.

with WHAM (Grossfield, 2020–9) being low (**Supplementary Figures S8–S11**). For the LN system, the global minimum is found at d = 224 Å. Fitting the PMF around this minimum into a harmonic potential led to an equilibrium distance of 225 Å and a vibrational frequency of 2.41 × 10^9 s^{-1}. As shown in **Figure 3**, the interaction free energy increases harmonically when the two nucleosomes are compressed from the global minima and then becomes flat and rugged below 200 Å, whereas the interaction free energy increases slightly and then becomes flat and rugged. This suggests that the linker DNA is more like a rubber string other than a spring. We also analyzed the relative orientation of nucleosomes in each window. The orientation angle along the time series in each window was stable and vibrated regularly within a small interval (**Supplementary Figure. S4**), whereas the average values were strongly dependent on the reaction coordinate d (**Supplementary Figure. S5**). The umbrella sampling results are consistent with the observations from unbiased MD simulations (**Figure 2A**), which provide more quantitative information on the interaction landscape of linked nucleosomes.

The interaction between two unlinked nucleosomes in solution has a completely different free energy profile, featured by a strong repulsion wall at smaller distances, and a flat curve for larger d (**Figure 4**). It indicates that the interactions between ULN nucleosomes are repulsive. As histone tails are not included in our ULN simulation system, the repulsive interaction is consistent

with previous simulations (Collepardo-Guevara et al., 2015; Saurabh et al., 2016; Ishida and Kono, 2017; Moller et al., 2019) and experiments (Dorigo et al., 2003; Gordon et al., 2005), which showed histone tails are crucial for holding stacked nucleosomes together. We note that the nucleosome has a disc-like shape with a height of 55 Å, so the sharp free energy rise below d = 60 Å is probably not caused by steric clash but instead by the unfavorable electrostatic interactions as each individual nucleosome is highly negatively charged. The asymptotic free energy was not fully resolved in our calculations, due to the limited size of the simulation box (182 Å × 182 Å × 182 Å). When the two nucleosomes are pulled away with d larger than 90 Å, interactions with their PBC images become possible such that the WHAM analysis is no longer valid.

Similar to the LN system, we also analyzed the orientation of nucleosomes in each window. The relative orientation angle along the time series in each window is stable and vibrates regularly at a small interval (**Supplementary Figure. S6**) and correlates with the distance d (**Supplementary Figure. S7**). In the ULN system, orientation angle has a tendency to increase with d and has larger fluctuation at larger d, consistent with unbiased MD simulation in which angle increases with distance (**Figures 2C,D**). The computational PMF for the ULN system can be directly compared with a recent experimental measurement of the association of nucleosomes (Funke et al., 2016). By integrating two nucleosomes into a carefully designed and calibrated DNA origami-based force spectrometer, Funke et al derived the Boltzmann-weighted distance-dependent energy landscape for two nucleosomes interacting with each other. They observed three major features: a strong repulsion at distances smaller than 60 Å; a minimum located somewhere between 60 and 70 Å; and vanishing interactions at distances greater than 130 Å. The computational PMF shown in **Figure 4** agrees well with strong repulsion at small distances and vanishing interactions at distal distance. No global minimum is found in our PMF due to the absence of histone tails in the ULN system, suggesting the importance of histone tails from an indirect perspective.

Constructing Coarse-Grained Potentials

With umbrella sampling, we are able to dissect the interactions between nucleosomes with atomistic details. On the other hand, the human chromatin contains about 28 million nucleosomes, so it is not feasible to use atomistic simulations to study chromatin folding in the foreseeable future. The free energy profiles we obtained can serve as the starting point to construct coarse-grained potentials to study the conformational dynamics of chromatin elements such as 30 nm fibers. Current polymer theory models for chromatin (Di Pierro et al., 2016; Fudenberg et al., 2016) are typically composed of five terms:

$$H_{\text{total}} = \sum H_{\text{bond}} + \sum H_{\text{stiffness}} + \sum H_{\text{loop}} + \sum H_{\text{compartment}} + \sum H_{\text{permeability}}, \quad (2)$$

where the H_{bond} term models the interaction between adjacently connected monomers, $H_{\text{stiffness}}$ describes the rigidity of chromatin fibers, H_{loop} term represents the biological process of loop extrusion, $H_{\text{compartment}}$ term models the compartmentalization, and $H_{\text{permeability}}$ models the biological function of topoisomerase II that makes DNA free of knots. In general, the H_{bond} and $H_{\text{stiffness}}$ are common items that are intrinsic in the conformation of chromatin and conserved over the cells of different sources and species. The mechanism behind the H_{loop} item is complicated because how the involved SMC complexes establish the loop structure at the molecular level remains unknown, especially a cell may have different H_{loop} at different states. $H_{\text{compartment}}$ is highly correlated with the interactions between unlinked nucleosomes and epigenetic modifications on histone cores and can describe the phase separation responsible for the compartmentalization. Here, we derived H_{bond} and $H_{\text{compartment}}$ items using all-atom MD simulations.

If we construct a CG model in which each nucleosome is coarse-grained into one monomer, the H_{bond} term can then be directly inferred from the atomistic simulation results of the LN system. As shown in **Figure 3**, the PMF can be fitted with a harmonic potential:

$$V(d) = \frac{1}{2}k(d - d_0)^2, \quad (3)$$

where k equals 0.01 kcal/mol/Å^2 and d_0 = 225.0 Å.

The free energy profile of the ULN system shows that the interactions between a pair of nucleosomes in distant distance are very weak (**Figure 4**). This suggests that the nucleosomes might not self-aggregate if there are no additional restraints such as histone tail effects, epigenetic modifications or binding of proteins such as cohesins or condensins. $H_{\text{compartment}}$ in **Eq. 2** models compartmentalization, which means that nucleosomes in different compartments (A and B) have different interaction strength (Di Pierro et al., 2016). In general, A compartments match euchromatins and B compartments correspond to heterochromatins. The interaction landscape of two free, unmodified nucleosomes from explicit-solvent all-atom MD simulations would be useful to construct some of the $H_{\text{compartment}}$ terms. If we use an exponential potential to fit the PMF

$$V(d) = e^{-\alpha \cdot d - \beta} \quad (4)$$

we could determine the parameters to be α = 0.1213 Å^{-1} and β = 9.2086 (red line in **Figure 4**).

Another suitable potential energy function form to fit the PMF would be a shifted Coulomb potential.

$$V(d) = \frac{A}{d - B}. \quad (5)$$

However, fitting the PMF with such a shifted Coulomb potential leads to A = 39.53 kcal/mol · Å and B = 52.82 Å (green line in **Figure 4**).

DISCUSSION AND CONCLUSION

A spatially and temporally resolved understanding of chromatin organization is currently one of the central topics in molecular

and cell biology. Here, we used classical MD simulations and enhanced sampling methods to study the basic element of chromatin, nucleosome. In particular, we studied the conformational dynamics and the free energy landscapes between two nucleosomes in aqueous environment. Not only the widely investigated stacked nucleosomes but also the linked nucleosomes are involved in our study. With umbrella sampling calculations, we obtained detailed free energy profiles for nucleosome pairs either free or connected by a linker DNA. Our simulation results compare favorably with a variety of experimental findings and MD simulations, in particular the PMF of the ULN system correlates well with a recent single-molecule force spectrometer measurement.

Considering that the configurations of nucleosomes are highly dynamic, the relative orientations of the stacked nucleosomes are often restrained to achieve quick convergence in previous simulation studies. In this work, the ULN (stacked) system was simulated freely without additional restrains at the cost of longer simulations, which makes the dynamics more natural and nucleosomes free to explore possible relative orientations. Both unbiased and biased simulations of unlinked nucleosomes show that electrostatic repulsion dominates distance distribution in the absence of histone tails. In addition to the ULN system, we also studied the dynamics and free energy profiles of linked nucleosomes. Results of LN and ULN systems could provide new insights into the organization and dynamics of larger chromatin elements such as 30 nm fibers.

Umbrella sampling is a useful technique to investigate the effective interaction between biological macromolecules. Recently, Lai et al (2020) studied the free energy profile between two identical DNA double helices using extensive umbrella sampling. In this study, we carried out 1 μs conventional MD simulations and found out that the relative motion between nucleosomes is very slow. We then performed accumulatively more than 3.8 μs umbrella sampling simulations on systems with more than half a million atoms to understand the slow- and large-scale dynamics between nucleosomes. This is only possible with recent advances in both hardware and software that utilize the graphics processing units (GPUs) for MD simulations. Simulation of the LN system (~ 677,000 atoms) runs about 10 ns per day on a single Tesla V100 GPU card. One advantage of umbrella sampling is that the method is naively parallel so simulations of different windows can be carried out on different GPU cards at the same time.

We note that K^+ ions were used as the cation in our simulation systems, whereas salt concentration and composition are more complicated in physiological environment and will impact the strength of nucleosome interactions. Mg^{2+} ions were known to play a crucial role in directly binding with and stabilizing nucleic acids. However, more accurate polarizable force fields might be needed to model Mg^{2+} ions (Huang et al., 2014; Lemkul and MacKerell, 2016; Walker et al., 2020). It would be interesting to investigate how the interaction landscapes between nucleosomes change with the different ion strength and composition. Another flaw of the current study concerns the limited simulation box size. For the ULN system, we are not able to obtain the asymptotic free energy over distance larger than 90 Å, as nucleosomes will interact through their images due to the PBC conditions. A larger and probably noncubic water box might be needed to overcome these limitations.

We carried out atomistic simulations of nucleosomes with the long-term goal to construct a coarse-grained potential for 30 nm chromatin structure to bridge the gap between all-atom model of nucleosomes and polymer model. This would constitute a "bottom-up" approach to determine the function forms for polymer models and to optimize their parameters. With the analytically fitted CG potentials presented in this work, one would already be able to simulate, for example, the classical 30 nm chromatin fibers to investigate the stability and dynamics of such nucleosome fiber arrays and then optimize the polymer model with this coarse-grained potential. For this purpose, we built a system consisting of 100 nucleosomes which had no open ends and was covalently bonded to itself through periodic boundary conditions. When we propagated the simulation system using derived potential functions (**Eqs 3, 4**) for 10^6 steps (1 μs), the system was unstable and quickly crashed. In contrast, the structure remained stable if a weak attractive interaction between stacked nucleosomes was added mimicking the effect exerted by histone tails (**Supplementary Figure S12**). Unstability of 30-nm chromatin fiber with our CG model is consistent with previous research studies showing that histone tails are critical to hold nucleosomes together (Dorigo et al., 2003; Gordon et al., 2005).

This highlights one of the current limitations of the preliminary CG model presented here. Conversion between different compacted states of 30 nm chromatin significantly depend on the histone tails, especially H4 tails, so simulations with histone tails need to be carried out to derive the corresponding free energy profiles. Epigenetic modifications often occur on the histone tails, changing the properties of nucleosomes such as charge, rigidification, and solvent exchange which make a big difference in the dynamics and organization of chromatin and gene expression. Considering the importance of histone tails for both nucleosome stacking interactions and higher level chromatin structures, we are currently performing similar simulations with full histone tails. However, this involves significantly more extensive MD simulations that are out of the scope of the current study, which represents a first step toward building up a CG model for chromatin with a bottom-up approach. The binding between the linker histone and linker DNA also plays important roles in chromatin compaction (Luque et al., 2014), especially the interactions between linked nucleosomes. Similar linked nucleosomes systems with linker histone should be studied in future work. On the other hand, the model is not applied to any practical biological problem, for example, loop extrusion and compartmentalization. In reality, the disk-like shape of nucleosomes and nonuniform distribution of their charges make interactions highly anisotropic, whereas interactions are assumed to be isotropic in our coarse-grained model.

As for the next step, we will perform similar umbrella sampling simulations to study how epigenetic modifications on nucleosomes change the interaction landscapes and derive the corresponding potential parameters. Modifications of

interests include acetylation of lysines on histone H4, mono-, di-, and trimethylation of lysine 27 on histone H3 and monoubiquitination of lysine 119 on histone H2A. Parameterization for these epigenetic modifications would allow us to construct a transferable computational model for chromatin. We plan to construct a coarse-grained model at nucleosome resolution, aiming to make the model as simple as possible so that it can be used to simulate one whole chromosome even genome which consists of millions of nucleosomes. Four or more beads (monomers) will be needed to intimate the anisotropicity of one nucleosome and the highly negative charge will be included simultaneously in the future. Ultimately, the polymer model will be optimized using "bottom-up" strategy. Such a model is expected to be useful in understanding development-related and disease-related chromatin dynamics, for example, how the binding of Polycomb repressive complex 1 (PRC1) and 2 (PRC2) (Comet et al., 2016) induces the condensation of chromatin.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

JH conceived the project, CZ performed simulations. JH and CZ analyzed the data and wrote the paper together.

FUNDING

The work is supported by Zhejiang Provincial Natural Science Foundation of China (Grant No. LR19B030001) and National Natural Science Foundation of China (Grant No. 21803057), Westlake Education Foundation and Tencent Foundation.

ACKNOWLEDGMENTS

We thank Westlake University Supercomputer Center for computational resource and related assistance.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2021.624679/full#supplementary-material.

REFERENCES

Alipour, E., and Marko, J. F. (2012). Self-organization of domain structures by DNA-loop-extruding enzymes. *Nucleic Acids Res.* 40, 11202–11212. doi:10.1093/nar/gks925

Andersen, H. C. (1980). Molecular dynamics simulations at constant pressure and (or) temperature. *J. Chem. Phys.* 72. doi:10.1063/1.439486

Aqvist, J., Wennerström, P., Nervall, M., Bjelic, S., and Brandsdal, B. O. (2004). Molecular dynamics simulations of water and biomolecules with a Monte Carlo constant pressure algorithm. *Chem. Phys. Lett.* 384, 288–294. doi:10.1016/j.cplett.2003.12.039

Banigan, E. J., van den Berg, A. A., Brandão, H. B., Marko, J. F., and Mirny, L. A. (2020). Chromosome organization by one-sided and two-sided loop extrusion. *eLife* 9, e53558. doi:10.7554/eLife.53558

Biswas, M., Langowski, J., and Bishop, T. C. (2013). Atomistic simulations of nucleosomes. *Wires Comput. Mol. Sci.* 3, 378–392. doi:10.1002/wcms.1139

Brooks, B. R., Brooks, C. L., III, Mackerell, A. D., Jr, Nilsson, L., Petrella, R. J., Roux, B., et al. (2009). Charmm: the biomolecular simulation program. *J. Comput. Chem.* 30, 1545–1614. doi:10.1002/jcc.21287

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. (1983). Charmm: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* 4, 187–217. doi:10.1002/jcc.540040211

Cheng, R. R., Contessoto, V. G., Lieberman Aiden, E., Wolynes, P. G., Di Pierro, M., and Onuchic, J. N. (2020). Exploring chromosomal structural heterogeneity across multiple cell lines. *eLife* 9, 1–21. doi:10.7554/eLife.60312

Collepardo-Guevara, R., Portella, G., Vendruscolo, M., Frenkel, D., Schlick, T., and Orozco, M. (2015). Chromatin unfolding by epigenetic modifications explained by dramatic impairment of internucleosome interactions: a multiscale computational study. *J. Am. Chem. Soc.* 137, 10205–10215. doi:10.1021/jacs.5b04086

Comet, I., Riising, E. M., Leblanc, B., and Helin, K. (2016). Maintaining cell identity: PRC2-mediated regulation of transcription and cancer. *Nat. Rev. Cancer* 16, 803–810. doi:10.1038/nrc.2016.83

Contessoto, V. G., Cheng, R. R., Hajitaheri, A., Dodero-Rojas, E., Mello, M. F., Lieberman-Aiden, E., et al. (2021). The Nucleome Data Bank: web-based resources to simulate and analyze the three-dimensional genome. *Nucleic Acids Res.* 49, D172–D182. doi:10.1093/nar/gkaa818

Dans, P. D., Walther, J., Gómez, H., and Orozco, M. (2016). Multiscale simulation of DNA. *Curr. Opin. Struct. Biol.* 37, 29–45. doi:10.1016/j.sbi.2015.11.011

Dekker, J., Marti-Renom, M. A., and Mirny, L. A. (2013). Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. *Nat. Rev. Genet.* 14, 390–403. doi:10.1038/nrg3454

Dekker, J., Rippe, K., Dekke, M., and Klechner, N. (2002). Capturing chromosome conformation. *Science* 295, 1306–1311. doi:10.1126/science.1067799

Di Pierro, M., Zhang, B., Aiden, E. L., Wolynes, P. G., and Onuchic, J. N. (2016). Transferable model for chromosome architecture. *Proc. Natl. Acad. Sci. U.S.A.* 113, 12168–12173. doi:10.1073/pnas.1613607113

Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., et al. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485, 376–380. doi:10.1038/nature11082

Dorigo, B., Schalch, T., Bystricky, K., and Richmond, T. J. (2003). Chromatin fiber folding: requirement for the histone H4 N-terminal tail. *J. Mol. Biol.* 327, 85–96. doi:10.1016/S0022-2836(03)00025-1

Dunham, I., Kundaje, A., Aldred, S. F., Collins, P. J., Davis, C. A., Doyle, F., et al. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57–74. doi:10.1038/nature11247

Eastman, P., and Pande, V. (2010). OpenMM: a hardware-independent framework for molecular simulations. *Comput. Sci. Eng.* 12, 34–39. doi:10.1109/MCSE.2010.27

Eastman, P., Swails, J., Chodera, J. D., McGibbon, R. T., Zhao, Y., Beauchamp, K. A., et al. (2017). OpenMM 7: rapid development of high performance algorithms for molecular dynamics. *PLoS Comput. Biol.* 13 (7), e1005659. doi:10.1371/journal.pcbi.1005659

Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., and Pedersen, L. G. (1995). A smooth particle mesh Ewald method. *J. Chem. Phys.* 103, 8577–8593. doi:10.1063/1.470117

Feng, Z.-X., and Li, Q.-Z. (2017). Recognition of long-range enhancer-promoter interactions by adding genomic signatures of segmented regulatory regions. *Genomics* 109, 341–352. doi:10.1016/j.ygeno.2017.05.009

Field, Y., Kaplan, N., Fondufe-Mittendorf, Y., Moore, I. K., Sharon, E., Lubling, Y., et al. (2008). Distinct modes of regulation by chromatin encoded through nucleosome positioning signals. *PLoS Comput. Biol.* 4, e1000216. doi:10.1371/journal.pcbi.1000216

Finn, E. H., and Misteli, T. (2019). Molecular basis and biological function of variability in spatial genome organization. *Science* 365 (6457), eaaw9498. doi:10.1126/science.aaw9498

Fraser, J., Williamson, I., Bickmore, W. A., and Dostie, J. (2015). An overview of genome organization and how we got there: from FISH to Hi-C. *Microbiol. Mol. Biol. Rev.* 79, 347–372. doi:10.1128/mmbr.00006-15

Fudenberg, G., and Imakaev, M. (2017). FISH-ing for captured contacts: towards reconciling FISH and 3C. *Nat. Methods* 14, 673–678. doi:10.1038/nmeth.4329

Fudenberg, G., Imakaev, M., Lu, C., Goloborodko, A., Abdennur, N., and Mirny, L. A. (2016). Formation of chromosomal domains by loop extrusion. *Cell Rep.* 15, 2038–2049. doi:10.1016/j.celrep.2016.04.085

Funke, J. J., Ketterer, P., Lieleg, C., Schunter, S., Korber, P., and Dietz, H. (2016). Uncovering the forces between nucleosomes using DNA origami. *Sci. Adv.* 2 (11), e1600974. doi:10.1126/sciadv.1600974

Garcia-Saez, I., Menoni, H., Boopathi, R., Shukla, M. S., Soueidan, L., Noirclerc-Savoye, M., et al. (2018). Structure of an H1-bound 6-nucleosome array reveals an untwisted two-start chromatin fiber conformation. *Mol. Cell* 72, 902–915. doi:10.1016/j.molcel.2018.09.027

Gibcus, J. H., Samejima, K., Goloborodko, A., Samejima, I., Naumova, N., Nuebler, J., et al. (2018). A pathway for mitotic chromosome formation. *Science* 359 (6376), eaao6135. doi:10.1126/science.aao6135

Goes, A. C. S., Cappellen, D., Santos, G. C., Pirozhkova, I., Lipinski, M., Vassetzky, Y., et al. (2011). Loop domain organization of the p53 locus in normal and breast cancer cells correlates with the transcriptional status of the TP53 and the neighboring genes. *J. Cell. Biochem.* 112, 2072–2081. doi:10.1002/jcb.23129

Goloborodko, A., Imakaev, M. V., Marko, J. F., and Mirny, L. (2016a). Compaction and segregation of sister chromatids via active loop extrusion. *eLife* 5, 1–16. doi:10.7554/elife.14864

Goloborodko, A., Marko, J. F., and Mirny, L. A. (2016b). Chromosome compaction by active loop extrusion. *Biophys. J.* 110, 2162–2168. doi:10.1016/j.bpj.2016.02.041

Gordon, F., Luger, K., and Hansen, J. C. (2005). The core histone N-terminal tail domains function independently and additively during salt-dependent oligomerization of nucleosomal arrays. *J. Biol. Chem.* 280, 33701–33706. doi:10.1074/jbc.M507048200

Grossfield, A. (2020). Data from: Wham: the weighted histogram analysis method. Version XXXX. http://membrane.urmc.rochester.edu/wordpress/?page_id=126.

Hart, K., Foloppe, N., Baker, C. M., Denning, E. J., Nilsson, L., and MacKerell, A. D. (2012). Optimization of the CHARMM additive force field for DNA: improved treatment of the BI/BII conformational equilibrium. *J. Chem. Theor. Comput.* 8, 348–362. doi:10.1021/ct200723y

Huang, J., Lopes, P. E. M., Roux, B., and MacKerell, A. D. (2014). Recent advances in polarizable force fields for macromolecules: microsecond simulations of proteins using the classical drude oscillator model. *J. Phys. Chem. Lett.* 5, 3144–3150. doi:10.1021/jz501315h

Huang, J., Rauscher, S., Nawrocki, G., Ran, T., Feig, M., de Groot, B. L., et al. (2017). Charmm36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* 14, 71–73. doi:10.1038/nmeth.4067

Ishida, H., and Kono, H. (2017). H4 tails potentially produce the diversity in the orientation of two nucleosomes. *Biophys. J.* 113, 978–990. doi:10.1016/j.bpj.2017.07.015

Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79, 926–935. doi:10.1063/1.445869

Krietenstein, N., Abraham, S., Venev, S. V., Abdennur, N., Gibcus, J., Hsieh, T.-H. S., et al. (2020). Ultrastructural details of mammalian chromosome architecture. *Mol. Cell* 78, 554–565. doi:10.1016/j.molcel.2020.03.003

Lai, C.-L., Chen, C., Ou, S.-C., Prentiss, M., and Pettitt, B. M. (2020). Interactions between identical DNA double helices. *Phys. Rev. E* 101, 32414. doi:10.1103/PhysRevE.101.032414

Lemkul, J. A., and MacKerell, A. D. (2016). Balancing the interactions of mg2+ in aqueous solution and with nucleic acid moieties for a polarizable force field based on the classical drude oscillator model. *J. Phys. Chem. B* 120, 11436–11448. doi:10.1021/acs.jpcb.6b09262

Lequieu, J., Córdoba, A., Moller, J., and De Pablo, J. J. (2019). 1CPN: a coarse-grained multi-scale model of chromatin. *J. Chem. Phys.* 150, 215102. doi:10.1063/1.5092976

Li, M., Zhang, J. Z., and Xia, F. (2016). Constructing optimal coarse-grained sites of huge biomolecules by fluctuation maximization. *J. Chem. Theor. Comput.* 12, 2091–2100. doi:10.1021/acs.jctc.6b00016

Lieberman-aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326, 289–293. doi:10.1126/science.1181369

Liu, M., Lu, Y., Yang, B., Chen, Y., Radda, J. S. D., Hu, M., et al. (2020). Multiplexed imaging of nucleome architectures in single cells of mammalian tissue. *Nat. Commun.* 11, 1–14. doi:10.1038/s41467-020-16732-5

Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 389, 251–260. doi:10.1038/38444

Luque, A., Collepardo-Guevara, R., Grigoryev, S., and Schlick, T. (2014). Dynamic condensation of linker histone C-terminal domain regulates chromatin structure. *Nucleic Acids Res.* 42, 7553–7560. doi:10.1093/nar/gku491

MacPherson, Q., Beltran, B., and Spakowitz, A. J. (2018). Bottom-up modeling of chromatin segregation due to epigenetic modifications. *Proc. Natl. Acad. Sci. U.S.A.* 115, 12739–12744. doi:10.1073/pnas.1812268115

Meaburn, K. J., Agunloye, O., Devine, M., Leshner, M., Roloff, G. W., True, L. D., et al. (2016). Tissue-of-origin-specific gene repositioning in breast and prostate cancer. *Histochem. Cell Biol.* 145, 433–446. doi:10.1007/s00418-015-1401-8

Milo, R., Jorgensen, P., Moran, U., Weber, G., and Springer, M. (2009). BioNumbers-the database of key numbers in molecular and cell biology. *Nucleic Acids Res.* 38, D750–D753. doi:10.1093/nar/gkp889

Mirny, L. A., Imakaev, M., and Abdennur, N. (2019). Two major mechanisms of chromosome organization. *Curr. Opin. Cell Biol.* 58, 142–152. doi:10.1016/j.ceb.2019.05.001

Miyamoto, S., and Kollman, P. A. (1992). Settle: an analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* 13, 952–962. doi:10.1002/jcc.540130805

Moller, J., Lequieu, J., and De Pablo, J. J. (2019). The free energy landscape of internucleosome interactions and its relation to chromatin fiber structure. *ACS Cent. Sci.* 5, 341–348. doi:10.1021/acscentsci.8b00836

Münkel, C., and Langowski, J. (1998). Chromosome structure predicted by a polymer model. *Phys. Rev. E* 57, 5888–5896. doi:10.1103/physreve.57.5888

Noid, W. G., Chu, J.-W., Ayton, G. S., Krishna, V., Izvekov, S., Voth, G. A., et al. (2008). The multiscale coarse-graining method. I. a rigorous bridge between atomistic and coarse-grained models. *J. Chem. Phys.* 128, 244114. doi:10.1063/1.2938860

Oliveira Junior, A. B., Contessoto, V. G., Mello, M. F., and Onuchic, J. N. (2020). A scalable computational approach for simulating complexes of multiple chromosomes. *J. Mol. Biol.* 433, 166700. doi:10.1016/j.jmb.2020.10.034

Philip, H., John, E. H., David, B. J., Rainer, K. S., and Lynn, R. H. (1993). Polymer models for interphase chromosomes. *PNAS* 90, 7854–7858. doi:10.1073/pnas.90.16.7854

Portela, A., and Esteller, M. (2010). Epigenetic modifications and human disease. *Nat. Biotechnol.* 28, 1057–1068. doi:10.1038/nbt.1685

Rao, S. S. P., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., et al. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159, 1665–1680. doi:10.1016/j.cell.2014.11.021

Sanborn, A. L., Rao, S. S. P., Huang, S.-C., Durand, N. C., Huntley, M. H., Jewett, A. I., et al. (2015). Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc. Natl. Acad. Sci. U.S.A.* 112, E6456–E6465. doi:10.1073/pnas.1518552112

Sandholtz, S. H., MacPherson, Q., and Spakowitz, A. J. (2020). Physical modeling of the heritability and maintenance of epigenetic modifications. *Proc. Natl. Acad. Sci. U.S.A.* 117, 20423–20429. doi:10.1073/pnas.1920499117

Saurabh, S., Glaser, M. A., Lansac, Y., and Maiti, P. K. (2016). Atomistic simulation of stacked nucleosome core particles: tail bridging, the H4 tail, and effect of hydrophobic forces. *J. Phys. Chem. B* 120, 3048–3060. doi:10.1021/acs.jpcb.5b11863

Schwarzer, W., Abdennur, N., Goloborodko, A., Pekowska, A., Fudenberg, G., Loe-Mie, Y., et al. (2017). Two independent modes of chromatin organization revealed by cohesin removal. *Nature* 551, 51–56. doi:10.1038/nature24281

Stam, M., Tark-Dame, M., and Fransz, P. (2019). 3D genome organization: a role for phase separation and loop extrusion? *Curr. Opin. Plant Biol.* 48, 36–46. doi:10.1016/j.pbi.2019.03.008

Su, J.-H., Zheng, P., Kinrot, S. S., Bintu, B., and Zhuang, X. (2020). Genome-scale imaging of the 3D organization and transcriptional activity of chromatin. *Cell* 182, 1641–1659. doi:10.1016/j.cell.2020.07.032

Teif, V. B., and Clarkson, C. T. (2019). Nucleosome positioning. *Ency. Bioinform. Comput. Biol.* 1, 308–317. doi:10.1016/B978-0-12-809633-8.20242-2

Teif, V. B., Vainshtein, Y., Caudron-Herger, M., Mallm, J.-P., Marth, C., Höfer, T., et al. (2012). Genome-wide nucleosome positioning during embryonic stem cell development. *Nat. Struct. Mol. Biol.* 19, 1185–1192. doi:10.1038/nsmb.2419

Tribello, G. A., Bonomi, M., Branduardi, D., Camilloni, C., and Bussi, G. (2014). Plumed 2: new feathers for an old bird. *Comput. Phys. Commun.* 185, 604–613. doi:10.1016/j.cpc.2013.09.018

Walker, B., Jing, Z., and Ren, P. (2020). Molecular dynamics free energy simulations of atp:mg^{2+} and adp:mg^{2+} using the polarisable force field amoeba. *Mol. Simul.* 0, 1–10. doi:10.1080/08927022.2020.1725003

Wennberg, C. L., Murtola, T., Páll, S., Abraham, M. J., Hess, B., and Lindahl, E. (2015). Direct-space corrections enable fast and accurate lorentz-berthelot combination rule Lennard-Jones lattice summation. *J. Chem. Theor. Comput.* 11, 5737–5746. doi:10.1021/acs.jctc.5b00726

Zhang, B., and Wolynes, P. G. (2015). Topology, structures, and energy landscapes of human chromosomes. *Proc. Natl. Acad. Sci. U.S.A.* 112, 6062–6067. doi:10.1073/pnas.1506257112

Zhao, Z., Tavoosidana, G., Sjölinder, M., Göndör, A., Mariano, P., Wang, S., et al. (2006). Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra and interchromosomal interactions. *Nat. Genet.* 38, 1341–1347. doi:10.1038/ng1891

Zhou, K., Gaullier, G., and Luger, K. (2019). Nucleosome structure and dynamics are coming of age. *Nat. Struct. Mol. Biol.* 26, 3–13. doi:10.1038/s41594-018-0166-x

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Molecular Biosciences

ORIGINAL RESEARCH
published: 15 April 2021
doi: 10.3389/fmolb.2021.621128

Integrating an Enhanced Sampling Method and Small-Angle X-Ray Scattering to Study Intrinsically Disordered Proteins

Chengtao Ding[1], *Sheng Wang*[2] *and Zhiyong Zhang*[1]*

[1]MOE Key Laboratory for Membraneless Organelles and Cellular Dynamics, National Science Center for Physical Sciences at Microscale, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, China, [2]Tencent AI Lab, Shenzhen, China

Intrinsically disordered proteins (IDPs) have been paid more and more attention over the past decades because they are involved in a multitude of crucial biological functions. Despite their functional importance, IDPs are generally difficult to investigate because they are very flexible and lack stable structures. Computer simulation may serve as a useful tool in studying IDPs. With the development of computer software and hardware, computational methods, such as molecular dynamics (MD) simulations, are popularly used. However, there is a sampling problem in MD simulations. In this work, this issue is investigated using an IDP called unique long region 11 (UL11), which is the conserved outer tegument component from herpes simplex virus 1. After choosing a proper force field and water model that is suitable for simulating IDPs, integrative modeling by combining an enhanced sampling method and experimental data like small-angle X-ray scattering (SAXS) is utilized to efficiently sample the conformations of UL11. The simulation results are in good agreement with experimental data. This work may provide a general protocol to study structural ensembles of IDPs.

Keywords: IDPs, biological function, MD simulation, sampling, integrative modeling

OPEN ACCESS

Edited by:
Yong Wang,
University of Copenhagen, Denmark

Reviewed by:
Haiguang Liu,
Beijing Computational Science Research Center, China
Andreas Larsen,
University of Oxford, United Kingdom

***Correspondence:**
Zhiyong Zhang
zzyzhang@ustc.edu.cn

Specialty section:
This article was submitted to Biological Modeling and Simulation, a section of the journal Frontiers in Molecular Biosciences

Received: *25 October 2020*
Accepted: *08 February 2021*
Published: *15 April 2021*

Citation:
Ding C, Wang S and Zhang Z (2021) Integrating an Enhanced Sampling Method and Small-Angle X-Ray Scattering to Study Intrinsically Disordered Proteins. Front. Mol. Biosci. 8:621128. doi: 10.3389/fmolb.2021.621128

INTRODUCTION

It has been recognized that a large segment of the human proteome comprises intrinsically disordered proteins (IDPs) that lack stable secondary and tertiary structures under physiological conditions (Colak et al., 2013; Kulkarni and Uversky, 2019). IDPs play important roles in a multitude of crucial biological functions despite their lack of a stable structure, such as cell cycle regulation, molecular recognition, and signal transduction (Dunker et al., 2005; Uversky et al., 2005). According to previous work, IDPs are involved in the majority of human cancer (Iakoucheva et al., 2002) and many chronic diseases like cardiovascular disease (Cheng et al., 2006), neurodegenerative diseases (Uversky, 2009; Uversky, 2014), and type 2 diabetes (Du and Uversky, 2017).

Although researchers continue to discover the functional importance of IDPs, it remains difficult to explore the structure-function relationship because getting the high-resolution structures of IDPs remains elusive. Since an IDP is generally not stable in one conformational state, these classical technologies of structural biology, including X-ray crystallography and cryo-EM, cannot determine its atomic-resolution structure. Alternatively, structural information on the ensemble average of the

IDP is available by techniques like nuclear magnetic resonance (NMR) (Dunker and Oldfield, 2015), small-angle X-ray scattering (SAXS) (Bernado and Svergun, 2012), and Förster resonance energy transfer (FRET) (LeBlanc et al., 2018).

In order to obtain structural details of IDPs, atomistic molecular dynamics simulation is a useful and complementary method for illuminating the molecular nature of IDPs' conformational ensembles because it can provide spatial and temporal resolution unavailable from experiments (Potoyan and Papoian, 2011; Burger et al., 2014; Granata et al., 2015; Bhowmick et al., 2016). Despite the significant progress made, a sampling problem remains in MD simulations of IDPs. The conformational space of an IDP is generally very large, so conventional MD simulations at a timescale of microseconds (μs) cannot capture all the states adequately. To tackle this problem, many enhanced sampling methods have been developed, which achieve good sampling by modifying potential energy function (Hamelberg et al., 2004) or increasing the temperature of barrier regions (Zhang et al., 2003; Hu et al., 2012). In recent years, a new kind of sampling techniques has been proposed, which are built on iterative multiple independent MD (MIMD) simulations (Harada and Kitao, 2013, Harada and Kitao, 2015; Shkurti et al., 2019; Yuan et al., 2020; Zhang and Gong, 2020). Such a method generally contains many cycles, and each cycle consists of a number of short MIMD simulations starting from selected seed conformations. The sampling efficiency would depend on the strategy of selecting seeds, and different criteria have been tried (Harada and Shigeta, 2018).

Many studies have shown the possibility of combining experimental data and computational simulations to interpret structural dynamics of large biomolecules in a solution that is called integrative modeling (Braitbard et al., 2019). There are various integrative modeling techniques for the interpretation of different structural data (Bonomi et al., 2017; Saltzberg et al., 2019; Orioli et al., 2020), which can be divided into two categories: refining-while-sampling and the screening-after-sampling (Zhang et al., 2015). A refining-while-sampling method directly adds an extra pseudo energy term based on the experimental data and then a conformation or an ensemble is simulated by optimizing the energy (Zheng and Tekpinar, 2011; Bjorling et al., 2015). In a screening-after-sampling method, a structure pool of the biomolecule is firstly sampled without experimental restraints, and then a reweighting method acts on these conformations to optimize their weights in order to fit the experimental data well (Bottaro et al., 2020). An ensemble containing a small number of conformations selected from the pool could be determined (Bernado et al., 2007; Curtis et al., 2012).

In this work, we propose a general strategy to study the conformations of IDPs. After choosing a suitable force field and water model for simulating IDPs, an integrative modeling procedure combining an enhanced sampling method based on iterative MIMD and SAXS data is used to sample conformations of IDPs efficiently. We present a case study on an IDP called unique long region 11 (UL11), an RNA-binding protein that is one of the conserved outer tegument components from herpes simplex virus 1 (HSV-1) (Bowzard et al., 2000; Metrick et al., 2020).

HSV-1 contains a unique tegument layer sandwiched between the capsid and lipid envelope, including 24 tegument proteins (McLauchlan and Rixon, 1992). UL11 is the smallest tegument protein with only 96 amino-acid residues (MacLean et al., 1989; Bowzard et al., 2000). UL11 and its homologs have been found to play crucial roles in efficient viral replication (MacLean et al., 1992; Baird et al., 2010) and tegument assembly (Owen et al., 2015). However, the mechanistic understanding of its role in these processes is limited due to the lack of knowledge of its biochemical and structural properties. A recent article (Metrick et al., 2020) has suggested that UL11 is an IDP in solution, which can undergo liquid–liquid phase separation (LLPS) *in vitro*. Analysis of experimental SAXS data showed that the protein is highly dynamic. Here, we aim to construct an atomic structural ensemble of UL11 that is in agreement with the available experimental data.

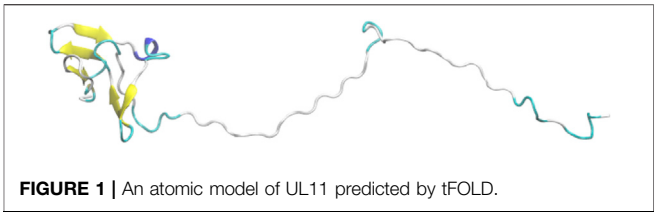


FIGURE 1 | An atomic model of UL11 predicted by tFOLD.

MATERIALS AND METHODS

An Initial Atomic Model of UL11

The UL11 construct used in this work is called UL11-Stll (Metrick et al., 2020), which is the UL11 sequence (96 residues) plus a small C-terminal Strep-tag II (Stll) including eight residues (WSHPQFEK). We used this 104-residue construct, on which the SAXS experiment was conducted. In the following, we call this construct UL11 for simplicity.

According to a prediction from the FoldUnfold server (http://bioinfo.protres.ru/ogu), many residues of UL11 are predicted to be disordered, except for some N-terminal residues that are natively folded (Metrick et al., 2020). We predicted an atomic model of UL11 using the tFOLD server (https://drug.ai.tencent.com/console/cn/tfold) (**Figure 1**). There are some β-strands at the N-terminus (residues 11–14, 17–20, 24–27, 39–41, and 44–47), while the other regions are disordered till the C-terminal end. The tFOLD model is consistent to the prediction of the disorder, so we used it as a starting structure for simulations.

Simulation Details

In this work, all-atom conventional MD (cMD) simulations and accelerated MD (aMD) simulations were conducted using the Amber20 package.

Conventional MD (cMD) Simulation

It has been recognized that, in MD simulations using those traditional force fields and water models, IDPs may become over-compact. Therefore, combinations of new force fields and water models have been proposed to address this issue (Kuzmanic et al., 2019). In this work, we used the A99SB force field in combination with a 4-point OPC water model (Izadi et al., 2014).

It has been reported that this A99SB/OPC combination is suitable for simulating conformations of IDPs (Shabane et al., 2019).

The system was built via the LEaP module (Case et al., 2005). The OPC waters (Izadi et al., 2014) were added to a truncated octahedral box with a minimal distance of 10.0 Å between the solute and the box boundary. 102 Na^+ and 98 Cl^- ions were added by replacing water molecules to balance the charge on the system and bring the salt concentration to about 100 mM NaCl. The box size is 1.66×10^6 Å^3, with 205,909 atoms in total. To remove bad contacts, the waters and ions were initially minimized for 2,000 steps using the steepest descent method for the first 1,500 steps and then the conjugate gradient for the last 500 steps, with the position of protein fixed (force constant was 500 kcal mol^{-1} Å^{-2}). In the second energy minimization, the restraints on the protein were removed. This stage was conducted for 2,500 steps, using the steepest descent method in the first 1,000 steps and then the conjugate gradient algorithm for the last 1,500 steps. After that, a heat-up MD was run at a constant volume. The system was heated from 0 to 300 K for 100 ps with a weak restraint of 10 kcal mol^{-1} Å^{-2} on the solute. A free MD simulation of 150 ns was carried out under the NPT condition utilizing the GPU-accelerated pmemd.cuda code. The temperature was regulated using the Langevin dynamics with a collision frequency of 1.0 ps^{-1} (Pastor et al., 1988). Pressure was controlled with isotropic position scaling at 1 bar with a relaxation time of 2.0 ps. All the bonds involving hydrogen atoms were constrained using the SHAKE algorithm (Ryckaert et al., 1977). A 2 fs integration step was used. Van der Waals interactions outside the cutoff distance were approximated via a continuum model (vdwmeth = 1) (Izadi et al., 2014; Izadi and Onufriev, 2016). The long-range electrostatic interaction was calculated using the PME method (Muller et al., 1996) with a 10 Å cutoff for the range-limited nonbonded interaction.

Accelerated MD (aMD) Simulation

The aMD (Muller et al., 1996) introduces a boost potential, ΔV(r), to the original potential energy V(r) when the latter is below a threshold energy E:

$$\Delta V(r) = \begin{cases} 0, & V(r) \geq E, \\ \left[\frac{(E - V(r))}{\alpha + (E - V(r))}\right]^2, & V(r) < E. \end{cases} \quad (1)$$

where α is a factor that tunes the depth of the modified energy basins. Boosting potentials were applied to both the total potential and the individual dihedral energy term. The aforementioned 150 ns cMD simulation was used to estimate the aMD parameters. In the cMD trajectory, the average total potential energy was −641,138 kcal mol^{-1} and the average dihedral energy was 1,068 kcal mol^{-1}. UL11 has 104 residues and the simulated system consists of 205,909 atoms. The following parameters were set based on the above information:

E (tot) = −641,138 kcal mol^{-1} + (0.2 kcal mol^{-1} atom^{-1} × 205,909 atoms)≈−599,956 kcal mol^{-1}
α (tot) = 205,909 atoms × 0.2 kcal mol^{-1} atom^{-1} ≈41,182 kcal mol^{-1}
E (dih) = 1,068 kcal mol^{-1} + (3.5 kcal mol^{-1} residue^{-1} × 104 residues)≈ 1,432 kcal mol^{-1}
α (dih) = 0.2 × (3.5 kcal mol^{-1} residue^{-1} × 104 residues)≈ 73 kcal mol^{-1}

With these parameters, a 150 ns aMD simulation was conducted. All the other parameters were the same to the aforementioned cMD simulation.

The Strategy of Integrative Modeling

We have previously developed a method called SAXS-oriented ensemble refinement (SAXS-ER) (Cheng et al., 2017), and the flowchart is as follows (**Figure 2**). The code is available at https://github.com/pcheng27/SAXS-ER/tree/v1.1.

1) Set up the system starting from an initial structure of the biomolecule, and perform a preliminary simulation. Any simulation method can be utilized, such as atomistic MD simulations, enhanced sampling techniques, or coarse-grained modeling. In this work, we are studying an IDP, and the sampling is challenging. Therefore, aMD simulations are carried out using the most updated code of pmemd.cuda in the Amber20 package.
2) Calculate the scoring function and obtain an ensemble of conformers with the best score. The number of conformers in the ensemble is N_{es}. In this work, the scoring function is χ^2 between the calculated SAXS profile of the ensemble and the experimental SAXS profile. More details will be introduced in the "Ensemble Optimization Method" section.
3) Starting from the N_{es} conformers selected by scoring function, N_{sim} (=N_{es})-independent simulations are carried out. Multiple independent short-time simulations may achieve a better sampling than a single long-time simulation. All the trajectories are combined.
4) Repeat steps 2 and 3 for N cycles. Analyze all those cycles with the saturated scoring function.

SAXS Data

The SAXS data of UL11 were taken from SASBDB (www.sasbdb.org) with the ID SASDEX4. All the experimental details and analyzed results can be found in the database and the published article (Metrick et al., 2020). In this work, we took the data points with q from 0.009 to 0.206 Å^{-1} ($q = 4\pi \sin\theta/\lambda$, where 2θ is the scattering angle and λ is the wavelength of 1.246 Å), and the signal–noise ratios in this range are essentially larger than 2.0 (**Figure 3A**). The radius of gyration (R_g) of the protein was estimated to be 24.1 ± 1.7 Å by Guinier analysis using the autoRg program in the ATSAS package (Franke et al., 2017). The pair distance distribution function (PDDF) was calculated by GNOM (Semenyuk and Svergun, 1991) using the maximum dimension (D_{max}) of 89.0 Å as input. The normalized PDDF is asymmetrical and tailed off to a large distance (**Figure 3B**), which resembles the shape of an elongated ellipsoid (Mertens and Svergun, 2010). Therefore, the protein should be able to take extended conformations in the solution that can be disordered. The Kratky plot (**Figure 3C**) also supports that the protein is an IDP with partially folded regions.

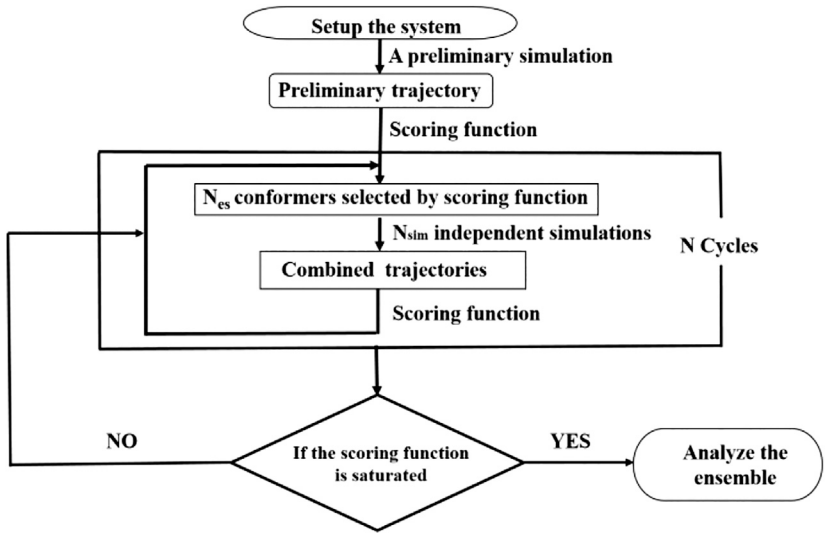


FIGURE 2 | Flowchart of integrative modeling that is a modification from **Figure 1** in (Cheng et al., 2017).

CRYSOL (Svergun et al., 1995) was used to compute the theoretical SAXS profile of a known atomic structure in PDB format, and then autoRg was run on the SAXS profile to estimate the R_g of the structure. The CaPP software, available at github.com/Niels-Bohr-Institute-XNS-StructBiophys/CaPP, was used to calculate PDDF from these PDB files.

Ensemble Optimization Method

A structural ensemble was obtained by the ensemble optimization method (EOM) (Bernado et al., 2007). EOM was used to select a small number of representative conformations from a pool containing lots of conformations of UL11 in order to fit the experimental SAXS data. The scoring function of EOM is as follows:

$$\chi^2 = \frac{1}{K-1}\sum_{i=1}^{K}\left[\frac{\mu I(q_i) - I_{\exp}(q_i)}{\sigma(q_i)}\right]^2, \quad (2)$$

where K is the number of data points in the SAXS profile and σ(q) are experimental errors. For every conformation in the ensemble, its theoretical scattering profile is computed. I(q) is the average of them, and μ is a scaling factor.

A new version of EOM called EOM2 (Tria et al., 2015) was used to compute the scoring function (**Eq. 2**) and pick the ensembles. In the original SAXS-ER using EOM2 (Cheng et al., 2017), the program automatically determined the ensemble size in each cycle that was generally small. An IDP should be represented by an ensemble containing more conformers than folded proteins. Therefore, in this work, we used an option of fixing the ensemble size to a relatively large number like 24 when running EOM2 in each cycle.

RESULTS AND DISCUSSION

aMD of UL11 without Integrating the SAXS Data. Three independent aMD simulations, each of 150 ns, were conducted. We converted a trajectory into sequentially individual PDB files; then CRYSOL and autoRg were run to obtain R_g of each atomic structure as described in the "SAXS Data" section. The initial structure of UL11 (**Figure 1**) is extended with R_g of 35.2 Å. In the first 70 ns of the aMD simulations, the protein is equilibrating with a clear tendency of R_g decrease (**Figure 4A**), and then the R_g values essentially fluctuate between 21.0 and 27.5 Å in the remaining simulations. According to the R_g distribution of the conformations in the last 80 ns (**Figure 4B**), they seem to show agreement with the experimental R_g of 24.1 ± 1.7 Å. We calculated the PDDF of each conformation in the last 80 ns of one trajectory and then plotted the ensemble-averaged PDDF (**Figure 4C**). The shape of the three ensemble-averaged PDDF curves is obviously not similar to that of the experimental PDDF (**Figure 3B**). That is to say, the aMD simulations at the time scale of 150 ns cannot adequately sample solution conformations of the IDP, which is the cause for the discrepancy between the simulated and the experimental PDDF. A straightforward way is to simply run longer simulations so that the protein could expand again and sample diverse conformations. However, it is not sure how long would be long enough to give a representative picture of the IDP. Therefore, we performed integrative modeling of UL11.

Integrative Modeling of UL11. Starting from the same structural model (**Figure 1**), we conducted integrative modeling of UL11 using the protocol introduced in **Figure 2**. A cycle consisted of N_{sim} = 24 independent 200 ps aMD simulations using A99SB/OPC. In each aMD simulation, a conformation was recorded every 1 ps, so a structural pool containing 4,800 conformations was generated in one cycle. By fitting the experimental SAXS data of UL11, EOM2 selected an ensemble with the size of N_{es} = 24 from the pool. Starting from these conformations, the next cycle of multiple independent simulations was run. We carried out 30 cycles, so the total simulation time was 144 ns (200 ps × 24 aMD × 30 cycles).

The χ^2 and the average R_g ($<R_g>$) of the ensemble are plotted against the cycle number (**Figure 5A**). The initial model of UL11

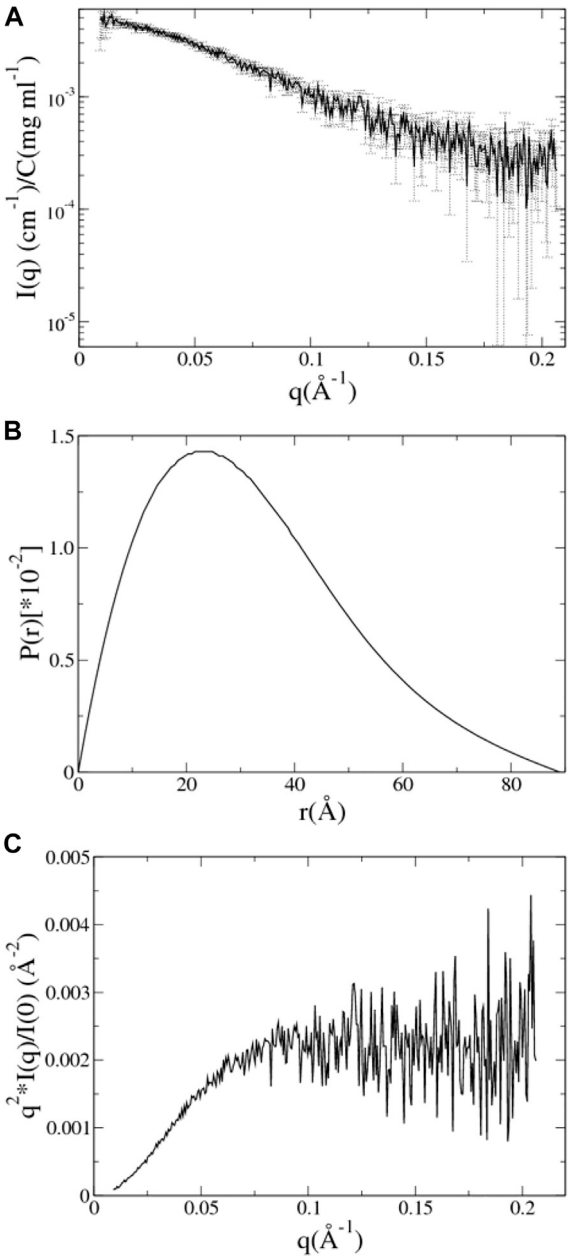


FIGURE 3 | SAXS data analysis of UL11. **(A)** The experimental SAXS profile of UL11 is shown with errors. **(B)** The pair distance distribution function (PDDF) is normalized so that the sum under the curve is 1. **(C)** Kratky plot.

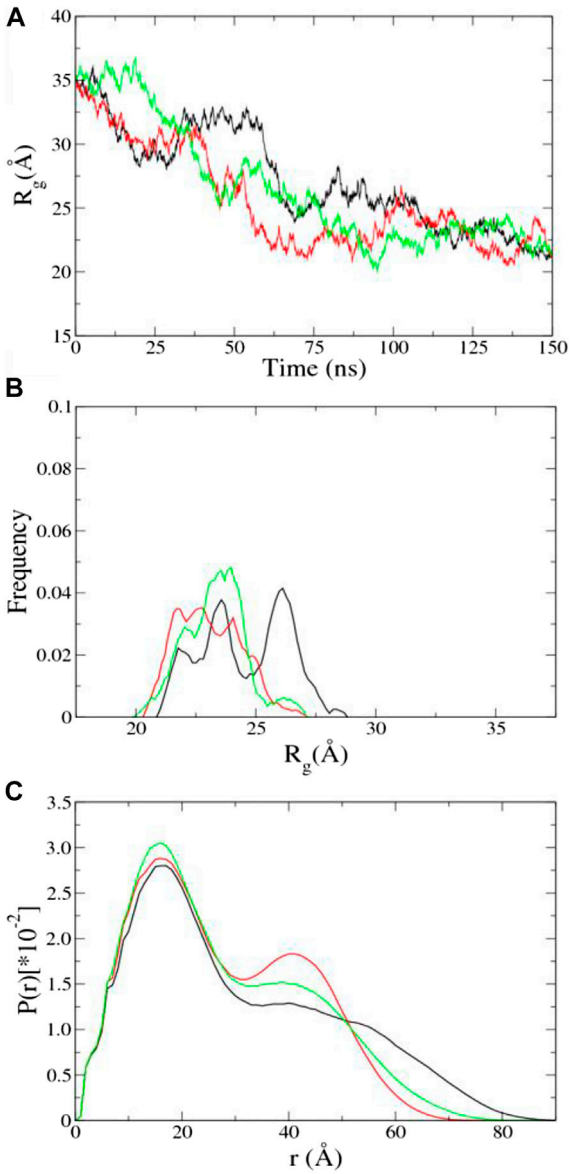


FIGURE 4 | Results of aMD using A99SB/OPC. **(A)** Time evolution of R_g. **(B)** R_g distribution in the last 80 ns aMD simulations. **(C)** Ensemble-averaged PDDF in the last 80 ns aMD simulations. The three independent simulations are shown in different colors.

is very extended (**Figure 1**); the EOM ensemble generated at cycle 0 cannot fit the experimental SAXS data well, with a χ^2 of 2.3. It is found that χ^2 decreases relatively fast in the first eight cycles (from 2.3 to 1.0), and then it slowly converges to about 0.9 after the 10th cycle (**Figure 5A**, circle). When looking at the time evolution of the $<R_g>$ (**Figure 5A**, up-triangle), it converges to 25.5 Å after 12 cycles, that is in good agreement with the estimated R_g (24.1 ± 1.7 Å) from the experimental SAXS data (**Figure 3A**). Therefore, we plotted the calculated SAXS profile of the ensemble at the 12th cycle and its error-weighted residual (**Figure 5B**). The residuals are defined as $(I_{exp}(q) - I_{calc}(q))/\sigma_{exp}(q)$, corresponding to the difference between the experimental and the computed intensities weighted by the experimental uncertainty (Carter et al., 2015; Trewhella et al., 2017). The residual difference plot is flat, which indicates that the results are in good agreement with the data. The inset is the normalized average PDDF of the ensemble, which has a similar shape to the experimental PDDF (**Figure 3B**).

To characterize conformations consistent with the SAXS data, we analyzed the R_g distribution of all the ensembles after the 11th cycle (**Figure 5C**). There is a major peak with the R_g value around 24.6 Å, a minor peak located between 27.5 and 30.0 Å, and two

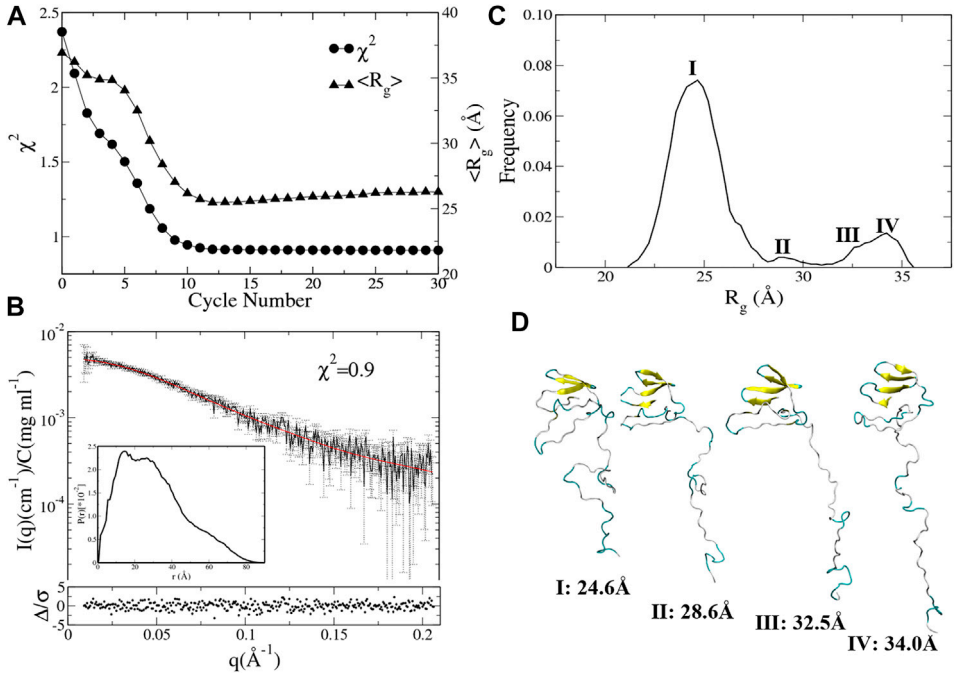


FIGURE 5 | Integrative modeling from an extended structure of UL11 (**Figure 1**). **(A)** The minimal χ^2 (circle) and the corresponding $<R_g>$ (up-triangle) at each cycle. **(B)** The back-calculated SAXS profile of the selected ensemble (red line) is fitted to the experimental data (black line with errors). The lower plot shows the error-weighted residual of the model fitting. The inset is the normalized ensemble-averaged PDDF. **(C)** The distribution of R_g values calculated from the ensembles after the 11th cycle. **(D)** Representative structures according to the R_g distribution.

more peaks with the R_g values larger than 30.0 Å that do not appear in the 150 ns aMD simulations (**Figure 4B**). A representative structure of each peak is shown in **Figure 5D**. One can clearly see several states of UL11, which correspond to relatively compact, intermediate, and extended conformations, respectively.

To test the reproducibility of the results, we also conducted the integrative modeling starting from a relatively compact structure of UL11 (inset in **Figure 6A**) taken from the 150 ns cMD simulation using A99SB/OPC. χ^2 and $<R_g>$ of the ensemble are plotted against the cycle number (**Figure 6A**). χ^2 of the ensemble at cycle 0 is 1.8, and only after seven cycles, it converges to 0.9 (**Figure 6A**, circle). $< R_g>$ of the ensemble at cycle 0 is 23.6 Å, and it converges to 25.8 Å after 11th cycles (**Figure 6A**, up-triangle). We plotted the calculated SAXS profile of the ensemble at the 12th cycle and its error-weighted residual (**Figure 6B**). The residual difference plot between the experimental and the computed I(q) is flat, which indicates that the results fit with the data. The normalized ensemble-averaged PDDF is in agreement with the experimental curve (**Figure 3B**). The R_g distribution of all the ensembles after the 12th cycle also indicates a major peak around 24.1 Å, a minor one between 27.5 and 30.0 Å, and two more peaks with the R_g values larger than 30.0 Å (**Figure 6C**). The representative structures of the peaks (**Figure 6D**) correspond to states of UL11 from the relatively compact, the intermediate, and to the extended conformations. It has been found that the two independent integrative models of UL11 starting from the different structures show fairly consistent results.

It is worth noting that the total time scale of the integrative modeling is only 144 ns, but it can achieve a more efficient sampling and better convergence than the 150 ns aMD simulations (**Figure 4**).

In a previous work (Metrick et al., 2020), the authors ran RANCH, an internal program of EOM2, to generate a coarse-grained structural pool using a simple exclusion energy term. Then EOM was applied to the pool to pick an ensemble by fitting the SAXS data. The ensemble also included states from compact to extended. Our results of integrative modeling support their study. However, our ensembles consist of atomic models generated by fine Amber force field and explicit water model, which should be physically more reasonable than those generated by RANCH. However, more experimental data would be needed to further validate these models.

CONCLUSION

This work integrates an enhanced sampling method and experimental SAXS data to study IDPs. In our strategy, we first need to choose a combination of the force field and water model, such as A99SB/OPC, that is suitable for simulating IDPs, and then an enhanced sampling technique like aMD is taken. After that, integrative modeling is conducted based on iterative multiple independent simulations. Experimental data like SAXS are used to design a scoring function for screening conformations and thus guide the simulations toward an ensemble that fits the experimental data well. Therefore, we think this strategy of integrative modeling is well suited for investigating conformational ensembles of IDPs.

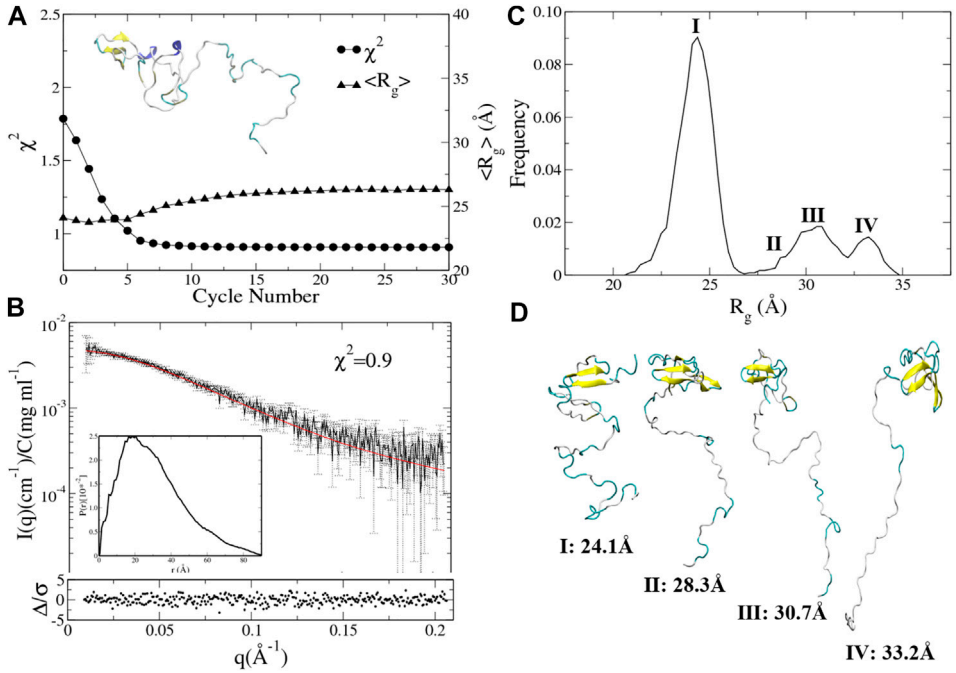


FIGURE 6 | Integrative modeling from a relatively compact structure of UL11. **(A)** The minimal χ^2 (circle) and the corresponding <R_g> (up-triangle) at each cycle. The inset is the starting structure. **(B)** The back-calculated SAXS profile of the selected ensemble (red line) is fitted to the experimental data (black line with errors). The lower plot shows the error-weighted residual of the model fitting. The inset is the ensemble-averaged PDDF. **(C)** Distribution of R_g values calculated from the ensembles after the 11th cycle. **(D)** Representative structures according to the R_g distribution.

We have carried out the integrative modeling of UL11, which is important for efficient viral replication and tegument assembly. To the best of our knowledge, the understanding of its biochemical structure and mechanism is still limited, except for some coarse-grained structural information (Metrick et al., 2020). In this work, we have predicted an ensemble of atomic structures, which includes both the relatively compact and extended conformations of UL11. This ensemble is in agreement with the available experimental data and may provide information on the functional mechanism of UL11. It has been said that UL11 undergoes LLPS *in vitro* (Metrick et al., 2020). Our study on the monomer and the integrative modeling strategy may be helpful for future research on LLPS.

There are various tools for integrative modeling (Bonomi et al., 2017; Orioli et al., 2020), which use either the refining-while-sampling or the screening-after-sampling strategy. A refining-while-sampling method is efficient, but one needs to modify complicated simulation code to add an energy term for experimental restraints. In a screening-after-sampling method, although there is no need to change the simulation code, the postprocessing reweighting procedure would rely on adequately sampling conformations of the biomolecule, which is, however, a nontrivial issue for IDPs. Our method can be regarded as an iterative screening-after-sampling strategy, so we do not change the MD code. However, the sampling is still efficient because it is guided by the experimental data.

Our integrative modeling method has some other characteristics. The first is that the iterative multiple independent simulations are very suitable for parallel computing. In this work, 24 independent simulations are run simultaneously, but one can use more CPU/GPU if they are available. The second is the high adaptability. Any sampling methods and ensemble optimization methods can be easily implemented with minor modifications to the scripts. Last but not least, many experimental data may be integrated simultaneously as long as a proper scoring function is designed. One of the future improvements is to input multiple initial models at the beginning of the integrative modeling in order to sample the conformations of IDPs as adequately as possible.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

ZZ and CD designed the study. CD collected data and carried out the calculation. SW did the structure prediction. ZZ and CD wrote the manuscript.

FUNDING

This work is supported by the National Natural Science Foundation of China (91953101, 21573205), the Strategic

Priority Research Program of the Chinese Academy of Science (XDB37040202), the Hefei National Science Center Pilot Project Funds, and the New Concept Medical Research Fund of USTC. The Supercomputing Center of USTC provides computer resources for this project.

REFERENCES

Baird, N. L., Starkey, J. L., Hughes, D. J., and Wills, J. W. (2010). Myristylation and palmitylation of HSV-1 UL11 are not essential for its function. *Virology* 397 (1), 80–88. doi:10.1016/j.virol.2009.10.046

Bernadó, P., Mylonas, E., Petoukhov, M. V., Blackledge, M., and Svergun, D. I. (2007). Structural characterization of flexible proteins using small-angle X-ray scattering. *J. Am. Chem. Soc.* 129 (17), 5656–5664. doi:10.1021/ja069124n

Bernadó, P., and Svergun, D. I. (2012). Structural analysis of intrinsically disordered proteins by small-angle X-ray scattering. *Mol. Biosyst.* 8 (1), 151–167. doi:10.1039/c1mb05275f

Bhowmick, A., Brookes, D. H., Yost, S. R., Dyson, H. J., Forman-Kay, J. D., Gunter, D., et al. (2016). Finding our way in the dark proteome. *J. Am. Chem. Soc.* 138 (31), 9730–9742. doi:10.1021/jacs.6b06543

Björling, A., Niebling, S., Marcellini, M., van der Spoel, D., and Westenhoff, S. (2015). Deciphering solution scattering data with experimentally guided molecular dynamics simulations. *J. Chem. Theor. Comput.* 11 (2), 780–787. doi:10.1021/ct5009735

Bonomi, M., Heller, G. T., Camilloni, C., and Vendruscolo, M. (2017). Principles of protein structural ensemble determination. *Curr. Opin. Struct. Biol.* 42, 106–116. doi:10.1016/j.sbi.2016.12.004

Bottaro, S., Bengtsen, T., and Lindorff-Larsen, K. (2020). Integrating molecular simulation and experimental data: a bayesian/maximum entropy reweighting approach. *Methods Mol. Biol.* 2112, 219–240. doi:10.1007/978-1-0716-0270-6_15

Bowzard, J. B., Visalli, R. J., Wilson, C. B., Loomis, J. S., Callahan, E. M., Courtney, R. J., et al. (2000). Membrane targeting properties of a herpesvirus tegument protein-retrovirus Gag chimera. *J. Virol.* 74 (18), 8692–8699. doi:10.1128/jvi.74.18.8692-8699.2000

Braitbard, M., Schneidman-Duhovny, D., and Kalisman, N. (2019). Integrative structure modeling: overview and assessment. *Annu. Rev. Biochem.* 88, 113–135. doi:10.1146/annurev-biochem-013118-111429

Burger, V., Gurry, T., and Stultz, C. (2014). Intrinsically disordered proteins: where computation meets experiment. *Polymers* 6 (10), 2684–2719. doi:10.3390/polym6102684

Carter, L., Kim, S. J., Schneidman-Duhovny, D., Stöhr, J., Poncet-Montange, G., Weiss, T. M., et al. (2015). Prion protein-antibody complexes characterized by chromatography-coupled small-angle X-ray scattering. *Biophys. J.* 109 (4), 793–805. doi:10.1016/j.bpj.2015.06.065

Case, D. A., Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., et al. (2005). The Amber biomolecular simulation programs. *J. Comput. Chem.* 26 (16), 1668–1688. doi:10.1002/jcc.20290

Cheng, P., Peng, J., and Zhang, Z. (2017). SAXS-oriented ensemble refinement of flexible biomolecules. *Biophys. J.* 112 (7), 1295–1301. doi:10.1016/j.bpj.2017.02.024

Cheng, Y., LeGall, T., Oldfield, C. J., Dunker, A. K., and Uversky, V. N. (2006). Abundance of intrinsic disorder in protein associated with cardiovascular disease. *Biochemistry* 45 (35), 10448–10460. doi:10.1021/bi060981d

Colak, R., Kim, T., Michaut, M., Sun, M., Irimia, M., Bellay, J., et al. (2013). Distinct types of disorder in the human proteome: functional implications for alternative splicing. *Plos Comput. Biol.* 9 (4), e1003030. doi:10.1371/journal.pcbi.1003030

Curtis, J. E., Raghunandan, S., Nanda, H., and Krueger, S. (2012). SASSIE: a program to study intrinsically disordered biological molecules and macromolecular ensembles using experimental scattering restraints. *Comp. Phys. Commun.* 183 (2), 382–389. doi:10.1016/j.cpc.2011.09.010

Du, Z., and Uversky, V. N. (2017). A comprehensive survey of the roles of highly disordered proteins in type 2 diabetes. *Int. J. Mol. Sci.* 18 (10), 2010. doi:10.3390/ijms18102010

Dunker, A. K., Cortese, M. S., Romero, P., Iakoucheva, L. M., and Uversky, V. N. (2005). Flexible nets. The roles of intrinsic disorder in protein interaction networks. *FEBS J.* 272 (20), 5129–5148. doi:10.1111/j.1742-4658.2005.04948.x

Dunker, A. K., and Oldfield, C. J. (2015). Back to the future: nuclear magnetic resonance and bioinformatics studies on intrinsically disordered proteins. *Adv. Exp. Med. Biol.* 870, 1–34. doi:10.1007/978-3-319-20164-1_1

Franke, D., Petoukhov, M. V., Konarev, P. V., Panjkovich, A., Tuukkanen, A., Mertens, H. D. T., et al. (2017). : a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J. Appl. Crystallogr.* 50 (Pt 4), 1212–1225. doi:10.1107/S1600576717007786

Granata, D., Baftizadeh, F., Habchi, J., Galvagnion, C., De Simone, A., Camilloni, C., et al. (2015). The inverted free energy landscape of an intrinsically disordered peptide by simulations and experiments. *Sci. Rep.* 5, 15449. doi:10.1038/srep15449

Hamelberg, D., Mongan, J., and McCammon, J. A. (2004). Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. *J. Chem. Phys.* 120 (24), 11919–11929. doi:10.1063/1.1755656

Harada, R., and Kitao, A. (2015). Nontargeted parallel cascade selection molecular dynamics for enhancing the conformational sampling of proteins. *J. Chem. Theor. Comput.* 11 (11), 5493–5502. doi:10.1021/acs.jctc.5b00723

Harada, R., and Kitao, A. (2013). Parallel cascade selection molecular dynamics (PaCS-MD) to generate conformational transition pathway. *J. Chem. Phys.* 139 (3), 035103. doi:10.1063/1.4813023

Harada, R., and Shigeta, Y. (2018). Selection rules on initial structures in parallel cascade selection molecular dynamics affect conformational sampling efficiency. *J. Mol. Graph. Model.* 85, 153–159. doi:10.1016/j.jmgm.2018.08.014

Hu, Y., Hong, W., Shi, Y., and Liu, H. (2012). Temperature-accelerated sampling and amplified collective motion with adiabatic reweighting to obtain canonical distributions and ensemble averages. *J. Chem. Theor. Comput.* 8 (10), 3777–3792. doi:10.1021/ct300061g

Iakoucheva, L. M., Brown, C. J., Lawson, J. D., Obradović, Z., and Dunker, A. K. (2002). Intrinsic disorder in cell-signaling and cancer-associated proteins. *J. Mol. Biol.* 323 (3), 573–584. doi:10.1016/s0022-2836(02)00969-5

Izadi, S., Anandakrishnan, R., and Onufriev, A. V. (2014). Building water models: a different approach. *J. Phys. Chem. Lett.* 5 (21), 3863–3871. doi:10.1021/jz501780a

Izadi, S., and Onufriev, A. V. (2016). Accuracy limit of rigid 3-point water models. *J. Chem. Phys.* 145 (7), 074501. doi:10.1063/1.4960175

Kulkarni, P., and Uversky, V. N. (2019). Intrinsically disordered proteins in chronic diseases. *Biomolecules* 9 (4), 147. doi:10.3390/biom9040147

Kuzmanic, A., Pritchard, R. B., Hansen, D. F., and Gervasio, F. L. (2019). Importance of the force field choice in capturing functionally relevant dynamics in the von willebrand factor. *J. Phys. Chem. Lett.* 10 (8), 1928–1934. doi:10.1021/acs.jpclett.9b00517

LeBlanc, S. J., Kulkarni, P., and Weninger, K. R. (2018). Single molecule FRET: a powerful tool to study intrinsically disordered proteins. *Biomolecules* 8 (4), 140. doi:10.3390/biom8040140

MacLean, C. A., Clark, B., and McGeoch, D. J. (1989). Gene UL11 of herpes simplex virus type 1 encodes a virion protein which is myristylated. *J. Gen. Virol.* 70 (Pt 12), 3147–3157. doi:10.1099/0022-1317-70-12-3147

MacLean, C. A., Dolan, A., Jamieson, F. E., and McGeoch, D. J. (1992). The myristylated virion proteins of herpes simplex virus type 1: investigation of their role in the virus life cycle. *J. Gen. Virol.* 73 (Pt 3), 539–547. doi:10.1099/0022-1317-73-3-539

McLauchlan, J., and Rixon, F. J. (1992). Characterization of enveloped tegument structures (L particles) produced by alphaherpesviruses: integrity of the tegument does not depend on the presence of capsid or envelope. *J. Gen. Virol.* 73 (Pt 2), 269–276. doi:10.1099/0022-1317-73-2-269

Mertens, H. D., and Svergun, D. I. (2010). Structural characterization of proteins and complexes using small-angle X-ray solution scattering. *J. Struct. Biol.* 172 (1), 128–141. doi:10.1016/j.jsb.2010.06.012

Metrick, C. M., Koenigsberg, A. L., and Heldwein, E. E. (2020). Conserved outer tegument component UL11 from herpes simplex virus 1 is an intrinsically disordered, RNA-binding protein. *mBio* 11 (3). doi:10.1128/mBio.00810-20

Müller, C. W., Schlauderer, G. J., Reinstein, J., and Schulz, G. E. (1996). Adenylate kinase motions during catalysis: an energetic counterweight balancing substrate binding. *Structure* 4 (2), 147–156. doi:10.1016/s0969-2126(96)00018-4

Orioli, S., Larsen, A. H., Bottaro, S., and Lindorff-Larsen, K. (2020). How to learn from inconsistencies: integrating molecular simulations with experimental data. *Prog. Mol. Biol. Transl. Sci.* 170, 123–176. doi:10.1016/bs.pmbts.2019.12.006

Owen, D. J., Crump, C. M., and Graham, S. C. (2015). Tegument assembly and secondary envelopment of alphaherpesviruses. *Viruses* 7 (9), 5084–5114. doi:10.3390/v7092861

Pastor, R. W., Brooks, B. R., and Szabo, A. (1988). An analysis of the accuracy of Langevin and molecular dynamics algorithms. *Mol. Phys.* 65(6), 1409–1419. doi:10.1080/00268978800101881

Potoyan, D. A., and Papoian, G. A. (2011). Energy landscape analyses of disordered histone tails reveal special organization of their conformational dynamics. *J. Am. Chem. Soc.* 133 (19), 7405–7415. doi:10.1021/ja1111964

Ryckaert, J.-P., Ciccotti, G., and Berendsen, H. J. C. (1977). Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* 23(3), 327–341. doi:10.1016/0021-9991(77)90098-5

Saltzberg, D., Greenberg, C. H., Viswanath, S., Chemmama, I., Webb, B., Pellarin, R., et al. (2019). Modeling biological complexes using integrative modeling platform. *Methods Mol. Biol.* 2022, 353–377. doi:10.1007/978-1-4939-9608-7_15

Semenyuk, A. V., and Svergun, D. I. (1991). Gnom—a program package for small-angle scattering data processing. *J. Appl. Cryst.* 24, 537–540. doi:10.1107/S002188989100081x

Shabane, P. S., Izadi, S., and Onufriev, A. V. (2019). General purpose water model can improve atomistic simulations of intrinsically disordered proteins. *J. Chem. Theor. Comput.* 15 (4), 2620–2634. doi:10.1021/acs.jctc.8b01123

Shkurti, A., Styliari, I. D., Balasubramanian, V., Bethune, I., Pedebos, C., Jha, S., et al. (2019). CoCo-MD: a simple and effective method for the enhanced sampling of conformational space. *J. Chem. Theor. Comput.* 15 (4), 2587–2596. doi:10.1021/acs.jctc.8b00657

Svergun, D., Barberato, C., and Koch, M. H. J. (1995). CRYSOL-a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Cryst.* 28, 768–773. doi:10.1107/s0021889895007047

Trewhella, J., Duff, A. P., Durand, D., Gabel, F., Guss, J. M., Hendrickson, W. A., et al. (2017). 2017 publication guidelines for structural modelling of small-angle scattering data from biomolecules in solution: an update. *Acta Crystallogr. D Struct. Biol.* 73 (Pt 9), 710–728. doi:10.1107/S2059798317011597

Tria, G., Mertens, H. D., Kachala, M., and Svergun, D. I. (2015). Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. *IUCrJ* 2 (Pt 2), 207–217. doi:10.1107/S205225251500202X

Uversky, V. N., Oldfield, C. J., and Dunker, A. K. (2005). Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling. *J. Mol. Recognit.* 18 (5), 343–384. doi:10.1002/jmr.747

Uversky, V. N. (2009). Intrinsic disorder in proteins associated with neurodegenerative diseases. *Front. Biosci.* 14, 5188–5238. doi:10.2741/3594

Uversky, V. N. (2014). The triple power of D^3: protein intrinsic disorder in degenerative diseases. *Front. Biosci.* 19, 181–258. doi:10.2741/4204

Yuan, Y., Zhu, Q., Song, R., Ma, J., and Dong, H. (2020). A two-ended data-driven accelerated sampling method for exploring the transition pathways between two known states of protein. *J. Chem. Theor. Comput.* 16 (7), 4631–4640. doi:10.1021/acs.jctc.9b01184

Zhang, J., and Gong, H. (2020). Frontier expansion sampling: a method to accelerate conformational search by identifying novel seed structures for restart. *J. Chem. Theor. Comput.* 16 (8), 4813–4821. doi:10.1021/acs.jctc.0c00064

Zhang, Y.-H., Peng, J.-H., and Zhang, Z.-Y. (2015). Structural modeling of proteins by integrating small-angle x-ray scattering data. *Chin. Phys. B* 24 (12), 126101. doi:10.1088/1674-1056/24/12/126101

Zhang, Z., Shi, Y., and Liu, H. (2003). Molecular dynamics simulations of peptides and proteins with amplified collective motions. *Biophys. J.* 84 (6), 3583–3593. doi:10.1016/S0006-3495(03)75090-5

Zheng, W., and Tekpinar, M. (2011). Accurate flexible fitting of high-resolution protein structures to small-angle x-ray scattering data using a coarse-grained model with implicit hydration shell. *Biophys. J.* 101 (12), 2981–2991. doi:10.1016/j.bpj.2011.11.003

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Molecular Biosciences

BRIEF RESEARCH REPORT
published: 27 April 2021
doi: 10.3389/fmolb.2021.673170

Pathways and Mechanism of Caffeine Binding to Human Adenosine A_{2A} Receptor

*Hung N. Do, Sana Akhter and Yinglong Miao**

Center for Computational Biology and Department of Molecular Biosciences, University of Kansas, Lawrence, KS, United States

Caffeine (CFF) is a common antagonist to the four subtypes of adenosine G-protein-coupled receptors (GPCRs), which are critical drug targets for treating heart failure, cancer, and neurological diseases. However, the pathways and mechanism of CFF binding to the target receptors remain unclear. In this study, we have performed all-atom-enhanced sampling simulations using a robust Gaussian-accelerated molecular dynamics (GaMD) method to elucidate the binding mechanism of CFF to human adenosine A_{2A} receptor (A_{2A}AR). Multiple 500–1,000 ns GaMD simulations captured both binding and dissociation of CFF in the A_{2A}AR. The GaMD-predicted binding poses of CFF were highly consistent with the x-ray crystal conformations with a characteristic hydrogen bond formed between CFF and residue N6.55 in the receptor. In addition, a low-energy intermediate binding conformation was revealed for CFF at the receptor extracellular mouth between ECL2 and TM1. While the ligand-binding pathways of the A_{2A}AR were found similar to those of other class A GPCRs identified from previous studies, the ECL2 with high sequence divergence serves as an attractive target site for designing allosteric modulators as selective drugs of the A_{2A}AR.

Keywords: adenosine A_{2A} receptor, caffeine, Gaussian accelerated molecular dynamics, ligand binding, mechanism, pathways

OPEN ACCESS

Edited by:
Yong Wang,
University of Copenhagen, Denmark

Reviewed by:
Lei Fu,
Beijing Normal University, China
Haohao Fu,
Nankai University, China
James Joseph McCarty,
Western Washington University, United States

****Correspondence:***
Yinglong Miao
miao@ku.edu

Specialty section:
This article was submitted to Biological Modeling and Simulation, a section of the journal Frontiers in Molecular Biosciences

Received: *26 February 2021*
Accepted: *24 March 2021*
Published: *27 April 2021*

Citation:
Do HN, Akhter S and Miao Y (2021) Pathways and Mechanism of Caffeine Binding to Human Adenosine A_{2A} Receptor. Front. Mol. Biosci. 8:673170. doi: 10.3389/fmolb.2021.673170

INTRODUCTION

Adenosine receptors (ARs) are a subfamily of G-protein-coupled receptors (GPCRs) with adenosine as the endogenous ligands (Fredholm et al., 1997). They belong to class A GPCRs and consist of four known subtypes: A_1AR, A_{2A}AR, A_{2B}AR, and A_3AR (Jacobson and Gao, 2006). Despite their broad distribution in human tissues and functional differences, ARs share common antagonists of caffeine (CFF) and theophylline, both of which antagonize the receptors upon binding. The sequence alignment by *MultiSeq* in VMD (Humphrey et al., 1996) showed that the seven transmembrane (TM) helix bundles of the A_{1A}AR shares high similarity with A_{2A}AR by 71%, A_{2B}AR by 70%, and A_{3A}AR by 77%. The sequence similarity is significantly reduced in the three extracellular loops (ECLs), being 43% for A_{2A}AR, 45% for A_{2B}AR, and 35% for A_3AR when compared with A_1AR. The CFF antagonist binds to all four subtypes of ARs, but with different binding affinities (Porkka-Heiskanen et al., 1997). Understanding the binding mechanism of CFF is expected to facilitate drug design targeting the functionally important ARs.

The human A_{2A}AR is one of the best structurally characterized GPCRs at the atomic level, with more than 30 x-ray and cryo-EM structures published to date (Carpenter and Lebon, 2017).

Structures of its three distinct conformational states have been reported, including the inactive conformation bound to an antagonist or inverse agonist, the intermediate conformation bound to an agonist, and the active conformation bound to both an agonist and engineered G protein. The orthosteric binding pocket of the A_{2A}AR is defined by the residues interacting with both the endogenous adenosine agonist and antagonist such as CFF (Carpenter and Lebon, 2017). The residues include M5.38, M7.35, I7.39, and F45.52^{ECL2} (Doré et al., 2011; Lebon et al., 2011; Cheng et al., 2017). The GPCR residues are numbered according to the Ballesteros–Weinstein scheme (Ballesteros and Weinstein, 1995). Receptor residue N6.55 can form hydrogen bonds with the N7 atom of adenosine (Lebon et al., 2011) and the O11 or O13 atom of CFF, which results in two distinct binding orientations referred to as CFF A and B (Cheng et al., 2017). Other interacting residues include V3.32, L3.33, T3.36, W6.48, L6.51, S7.42, and H7.43 (Lebon et al., 2012; Cheng et al., 2017).

Molecular dynamics (MD) simulations have been previously carried out to characterize binding of the CFF antagonist to the human A_{2A}AR. Cao et al. (2015) performed 800 ns MD simulations to elucidate the effect of membrane composition on the CFF-bound A_{2A}AR. They discovered that the seven TM helix folds were maintained across the systems over the course of their simulations. CFF was flexible and exhibited multiple binding poses in the receptor orthosteric binding pocket. The four most populated binding poses of CFF were extracted with the interacting residues, including A2.61, I2.64, S2.65, V3.32, L3.33, T3.36, F45.52, E169^{ECL2}, M5.38, N5.42, L6.51, H6.52, N6.55, H264^{ECL3}, M7.35, I7.39, and H7.43. In particular, CFF forms a hydrogen bond with receptor residue N6.55 and water-bridge contact with residue H7.43 (Cao et al., 2015). Guo et al. (2016) performed 10 temperature-accelerated MD (TAMD) simulations starting from the 4EIY PDB structure to investigate the dissociation pathway of the ZM241385 antagonist from the A_{2A}AR. The method specifically accelerated the center of mass of the ligand, and thus the A_{2A}AR was almost rigid. They found 16 residues that could potentially interact with ZM241385 during the ligand dissociation process, including G1^{TM1}, I2.63, S2.64, T2.65, Q148^{ECL2}, G152^{ECL2}, K153^{ECL2}, S156^{ECL2}, Q157^{ECL2}, E169^{ECL2}, T6.58, H7.29, A7.30, P7.31, L7.32, and Y7.36. Specifically, the residues E169^{ECL2}, T6.58, and H7.29 along with the structural water of 4EIY formed a hydrogen bond network interacting with the ligand ZM241385 (Guo et al., 2016). Caliman et al. (2018) applied the FTMap fragment-based mapping algorithm on the four distinct conformers obtained from MD simulations of two ligand free receptor conformations of the A_{2A}AR (PDBs: 3QAK and 3EML). They uncovered five non-orthosteric binding sites that were located in the intracellular region of the TM helices TM3/TM4, the G-protein-binding site in the intracellular region between TM2/TM3/TM6/TM7, the lipid interface of TM5/TM6, the intracellular region of TM1/TM7, and the extracellular region of TM3/TM4 of the A_{2A}AR. Their analysis also revealed residues in the orthosteric binding site, including I2.64, V3.32, L3.33, T3.36, Q3.37, I3.40, L45.51^{ECL2}, F45.52^{ECL2}, E169^{ECL2}, M5.38, N5.42, W6.48, L6.51, H6.52, N6.55, T6.58, H264^{ECL3}, L7.32, M7.35, Y7.36, I7.39, S7.42, and H7.43 (Caliman et al., 2018).

Gaussian-accelerated MD (GaMD) is a computational method that allows for simultaneous unconstrained enhanced sampling and free energy calculations of large biomolecules (Miao et al., 2015). By adding a harmonic boost potential, GaMD smooths the potential energy surface of biomolecules to reduce the system energy barriers (Miao et al., 2015). The harmonic boost potential mostly exhibits a Gaussian distribution. Cumulant expansion to the second order ("Gaussian approximation") can thus be applied to achieve proper energetic reweighting. GaMD resolves the energetic noise problem encountered in the previous accelerated MD (aMD) method (Hamelberg et al., 2004; Shen and Hamelberg, 2008), thereby allowing us to recover the original free energy profiles of biomolecules (Miao et al., 2015). Even though it is exceedingly difficult to obtain convergent free energy profiles for large biomolecular systems, "semiquantitative" low-energy conformational states of biomolecules can be identified from the GaMD-reweighted free energy profiles. GaMD does not require carefully predefined collective variables and as such it is advantageous to study complex biological processes. GaMD has been demonstrated on enhanced sampling and free energy calculations of ligand binding (Miao et al., 2015; Pang et al., 2017), protein folding (Miao et al., 2015; Pang et al., 2017), GPCR activation (Miao and McCammon, 2016), and protein–membrane (Bhattarai et al., 2020), protein–protein (Miao and McCammon, 2018; Wang and Miao, 2019), and protein–nucleic acid (Ricci et al., 2019; East et al., 2020) interactions. Of relevance to studies of GPCRs, GaMD simulations have successfully revealed the mechanisms of GPCR activation, ligand binding, and GPCR–G-protein interactions, which were consistent with experimental data and/or long timescale conventional MD (cMD) simulations (Miao and McCammon, 2016, Miao et al., 2018; Pawnikar and Miao, 2020).

In this study, we have performed all-atom GaMD simulations to determine the pathways and mechanism of CFF binding to the human A_{2A}AR. The GaMD simulations have captured both binding and dissociation of CFF in the A_{2A}AR. The simulation-predicted binding poses were consistent with x-ray crystal conformations of CFF in the 5MZP PDB structure (Cheng et al., 2017). An important intermediate binding site of CFF was also revealed from the GaMD simulations. The simulation findings could provide a molecule basis for rational computer-aided drug design targeting the A_{2A}AR and other ARs.

METHODS

System Setup

The x-ray crystal structure of the human A_{2A}AR in complex with CFF at 2.1 Å resolution (PDB: 5MZP) (Cheng et al., 2017) was used for setting up the simulation system. The structure included 296 out of the total 306 residues of the A_{2A}AR, with 10 missing residues (209–218). The T4-lysozyme, lipid molecules, CFF, water, and heteroatom molecules were removed. A total of 10 CFF ligand molecules were placed randomly at a distance >15 Å from the extracellular surface of the A_{2A}AR (**Figure 1A**). The simulation system was then prepared using the CHARMM-GUI webserver with the membrane input generator (Wang et al., 2006;

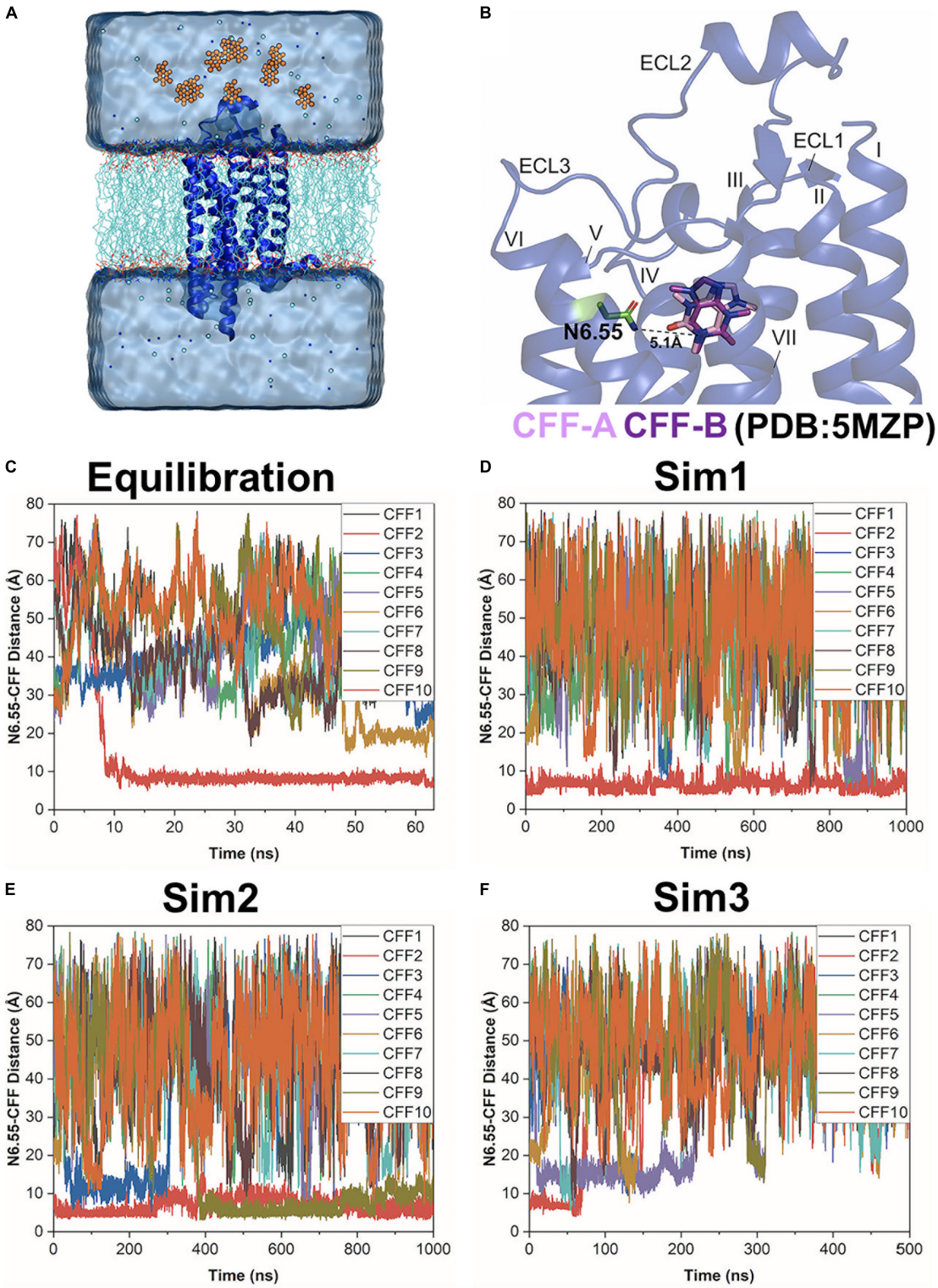


FIGURE 1 | Gaussian-accelerated molecular dynamics (GaMD) simulations successfully captured both binding and dissociation of caffeine (CFF) in the A_{2A}AR. **(A)** Computational model used for simulations of the A_{2A}AR (blue ribbons) with 10 CFF molecules (orange spheres) placed far away in the solvent. The receptor was inserted in a POPC lipid bilayer (cyan sticks) and solvated in an aqueous solution (cyan) of 0.15 M NaCl. **(B)** X-ray structure of CFF-bound A_{2A}AR (PDB: 5MZP). A hydrogen bond is formed between either O11 or O13 atom of CFF with the ND2 atom of the receptor residue N6.55 in two X-ray conformations of the ligand (CFF-A and CFF-B), in which the distance between the N1 atom that connects atoms O11 and O13 in CFF and the ND2 atom of residue N6.55 stays at 5.1 Å. The seven transmembrane (TM) helices I–VII and three extracellular loops (ECL) 1–3 are labeled in the A_{2A}AR. **(C–F)** Time courses of the N6.55:ND2–CFF:N1 distance calculated from 63 ns GaMD equilibration and three independent 500–1,000 ns GaMD simulations.

Jo et al., 2007, 2008; Wu et al., 2014; Lee et al., 2016, 2020). The system dimension was 81.18 × 81.18 × 115.11 Å. It included 161 POPC lipid molecules, with 81 molecules on the upper leaflet and 80 molecules on the lower leaflet, and 14,627 water molecules. All chain termini were capped with neutral patches (acetyl and methylamide). The system was solvated in 0.15 M NaCl solution at temperature 310 K. The AMBER FF19SB (Tian et al., 2019) parameter set was used for the receptor, LIPID17 (Gould et al., in preparation) for the POPC lipids, TIP3P (Jorgensen et al., 1983) for water, and GAFF2 (Wang et al., 2004; He et al., 2020) for CFF. The output files from CHARMM-GUI were used to perform GaMD simulations with AMBER 20 (Case et al., 2020).

Simulation Protocol

The output files of CHARMM-GUI webserver were used for initial energy minimization, equilibration, and cMD to prepare the system for GaMD simulations. The system was energetically minimized for 5,000 steps using the steepest-descent algorithm and equilibrated with the constant number, volume, and temperature (NVT) ensemble at 310 K using default parameters given by CHARMM-GUI. It was further equilibrated for 375 ps at 310 K with the constant number, pressure, and temperature (NPT) ensemble. The cMD simulation was then performed for 10 ns using the NPT ensemble with constant surface tension at 1 atm pressure and 310 K temperature.

Gaussian-accelerated MD implemented in GPU version of AMBER 20 (Salomon-Ferrer et al., 2013; Miao et al., 2015; Case et al., 2020) was applied to simulate the A_{2A}AR system. The simulations involved an initial short cMD of 3.0 ns to calculate GaMD acceleration parameters and GaMD equilibration of added boost potential for 60 ns. Three independent 500–1,000 ns GaMD production simulations with randomized initial atomic velocities were performed on the A_{2A}AR with 10 unbound CFF molecules. All GaMD simulations were run at the "dual-boost" level by setting the reference energy to the lower bound. One boost potential was applied to the dihedral energetic term and the other to the total potential energetic term. The average and SD of the system potential energies were calculated every 300,000 steps (0.6 ns) for all simulation systems. The upper limit of the boost potential SD, σ_0 was set to 6.0 kcal/mol for both the dihedral and the total potential energetic terms. The simulation frames were saved every 1.0 ps for analysis. The GaMD simulations are summarized in **Supplementary Table 1**.

The NPT ensemble with constant surface tension was used in the short cMD and GaMD simulations. The input files for GaMD equilibration and GaMD simulations have been attached as the Supporting Information. The standard protocol for MD simulations of membrane proteins was followed using notably the system configuration files generated from CHARMM-GUI (Wang et al., 2006; Jo et al., 2007, 2008; Wu et al., 2014; Lee et al., 2016, 2020). Using the MEMBPLUGIN 1.1 plugin of VMD (Guixà-González et al., 2014), we calculated the area per lipid to be 81.58 ± 8.92 $Å^2$ and the membrane thickness to be 70.74 ± 2.09 Å from the GaMD production simulations. The area per lipid was consistent with the initial value of 81.68 Å from CHARMM-GUI. The density of the entire system was calculated to be 1.008 ± 0.001 g/cm^3 from the GaMD production simulation outputs, being similar to the value of 1.020 g/cm^3 in the cMD simulation. Therefore, the system was expected to behave normally as in other simulation studies.

Simulation Analysis

Simulation analysis was carried out using CPPTRAJ (Roe and Cheatham, 2013) and VMD (Humphrey et al., 1996). The software tools were applied to track the binding and dissociation of CFF from the A_{2A}AR. A hydrogen bond could be formed between O11 or O13 of CFF with atom ND2 in residue N6.55 of the A_{2A}AR, so the distance between atom N1 that connects atoms O11 and O13 in CFF and atom ND2 of residue N6.55 was calculated to monitor ligand binding (**Figure 1B**). The distance between atom ND2 of receptor residue N6.55 and atom N1 of CFF and the distance of important interactions between receptor residues of TM helices (TM) III, VI, and VII were identified to calculate 2D potential mean force (PMF) free energy profiles using the *PyReweighting* toolkit (Miao et al., 2014). A bin size of 1 Å was used for the distances. The cutoff was set to 500 frames in one bin for reweighting.

The hierarchical agglomerative clustering algorithm was used to cluster the snapshots of protein conformations with all GaMD production simulations combined. The combined GaMD simulations of CFF binding to the A_{2A}AR were clustered to obtain clusters that corresponded to the low-energy states in the 2D PMF free energy profiles.

RESULTS

Gaussian-Accelerated MD Simulations Captured Both Binding and Dissociation of CFF in the A_{2A}AR

Three independent dual-boost GaMD simulations showed similar averages and SDs of the added boost potentials: 16.21 ± 4.50 kcal/mol for Sim1, 16.20 ± 4.49 kcal/mol for Sim2, and 16.32 ± 4.52 kcal/mol for Sim3, respectively (**Supplementary Table 1**). Spontaneous binding of CFF to the orthosteric site of the A_{2A}AR was detected at ∼9 ns into the GaMD equilibration (**Figure 1C**). The first two independent 1,000 ns GaMD simulations (Sim1 and Sim2) captured binding of the CFF in the receptor orthosteric pocket (**Figures 1D,E**). Remarkably, at ∼400 ns into GaMD Sim2, a second CFF bound to the orthosteric pocket of the A_{2A}AR, while the first CFF remained bound (**Figure 1E**). The complete dissociation of CFF from the orthosteric pocket of the A_{2A}AR was observed at ∼60 ns in the last independent 500 ns GaMD simulation (Sim3). The traces of CFF binding and dissociation were then analyzed in detail using CPPTRAJ and VMD. The representative CFF poses were selected at distances between receptor residue N6.55 atom ND2 and CFF atom N1 of ∼15, 10, and 5 Å to calculate the interacting residues from the A_{2A}AR in the binding and dissociation pathways using *LigPlot* (Wallace et al., 1995; **Supplementary Figures 1, 2**).

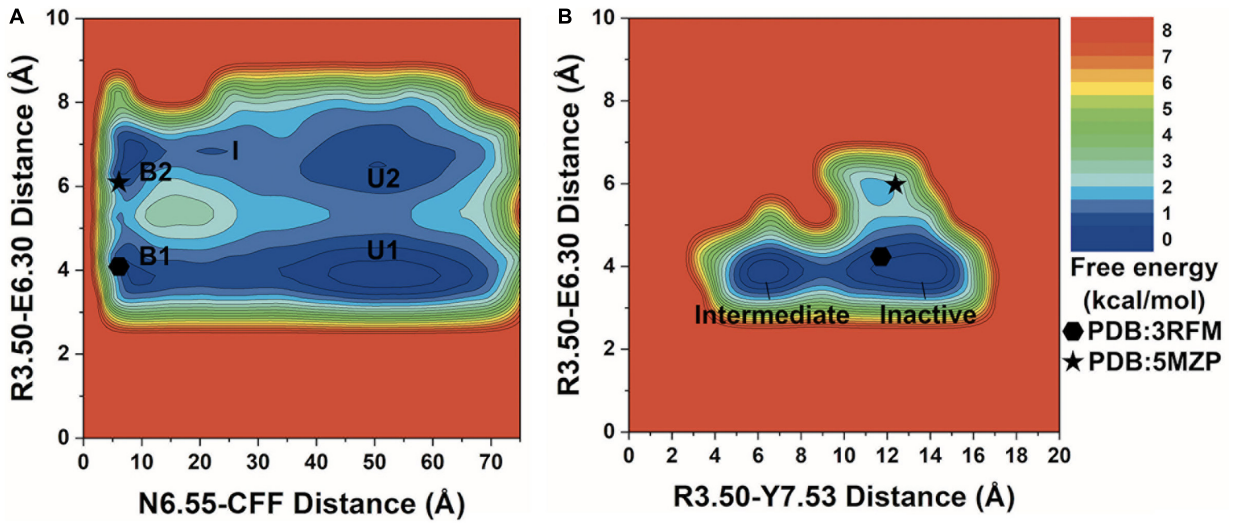


FIGURE 2 | 2D potential of mean force (PMF) free energy profiles of the A_{2A}AR–caffeine (CFF) interactions. **(A)** Two-dimensional (2D) PMF of the distance between receptor residue N6.55 atom ND2 and CFF atom N1 and the ionic lock distance between charge centers of receptor residues R3.50 and E6.30. The low-energy states are labeled in the PMF profile, including the unbound (U1 and U2), intermediate (I), and bound (B1 and B2). **(B)** 2D PMF of the distance between atom CZ in receptor residue R3.50 and the hydroxyl oxygen atom of residue Y7.53 and the ionic lock distance between charge centers of receptor residues R3.50 and E6.30. The low-energy inactive and intermediate conformational states are labeled. The 3RFM and 5MZP PDB structures of inactive A_{2A}AR are mapped to the free energy surface as hexagons and stars.

Free Energy Profiles of CFF Binding to the A_{2A}AR Receptor

We combined all three GaMD production simulations to calculate reweighted free energy profiles to characterize the binding of CFF to the A_{2A}AR. The distance between the ND2 atom of receptor residue N6.55 and the N1 atom of CFF, ionic lock distance between the CZ atom of residue R3.50 and the CD atom of residue E6.30, and the distance between the CZ atom of residue R3.50 and the OH atom of residue Y7.53 were selected as reaction coordinates to calculate the one-dimensional (1D) (**Supplementary Figure 3**) and two-dimensional (2D) (**Figure 2**) PMF free energy profiles. The 1D PMF free energy profiles with variations were obtained by averaging the three GaMD production simulations. Despite the free energy variations, relatively low-energy wells could be identified from 1D PMF profiles of the distances of CFF–residue N6.55, residues R3.50–E6.30, and residues R3.50–Y7.53 (**Supplementary Figure 3**). The corresponding distances from representative PDB structures (3RFM and 5MZP) were mapped to the 2D free energy profiles for comparison. In the 3RFM PDB structure, the distances between CFF and residue N6.55, residues R3.50 and E6.30, and residues R3.50 and Y7.53 are 5.5, 4.6, and 12.0 Å, respectively. In the 5MZP PDB structure, the distances between CFF and residue N6.55, residues R3.50 and E6.30, and residues R3.50 and Y7.53 are 5.2–5.5, 6.0, and 12.9 Å, respectively.

In the 2D free energy profile of the distances between residue N6.55 and CFF and residues R3.50 and E6.30 (**Figure 2A**), we identified five low-energy conformational states: unbound (U1, U2), intermediate (I), and bound (B1, B2). In the unbound states (U1, U2), the distance between receptor residue N6.55 and CFF exhibited a broad energy well from ∼35 Å to ∼65 Å, illustrating CFF diffusion in the bulk solvent. The ionic lock distance between residues R3.50 and E6.30 increased from ∼3.5–4.5 Å in U1 to ∼6–7.5 Å in U2. The intermediate state (I) was identified at ∼20–25 Å distance between receptor residue N6.55 and CFF and at ∼7 Å distance between residues R3.50 and E6.30, suggesting that CFF was located at the extracellular mouth of the A_{2A}AR. Both the bound states (B1, B2) were observed at ∼5–10 Å distance between receptor residue N6.55 and CFF. CFF was located in the orthosteric pocket in these states. Similar to the unbound states, the ionic lock distance was ∼3.5–4.5 Å in B1 and ∼6.5–7 Å in B2.

We identified two low-energy conformational states from the free energy profile of the R3.50–Y7.53 and R3.50–E6.30 distances in **Figure 2B**, labeled as the inactive and intermediate states. In the inactive state, the ionic lock distance between receptor residues R3.50 and E6.30 was ∼3.5–4.5 Å and the distance between receptor residues R3.50 and Y7.53 was ∼10–15 Å. In the intermediate state, the ionic lock distance remained the same, but the distance between receptor residues R3.50 and Y7.53 decreased to ∼5–8 Å.

Binding and Dissociation Pathways of CFF in the A_{2A}AR

In the equilibration trajectory of GaMD simulation, 1 out of the 10 CFF molecules (CFF2) that freely diffused in the solvent bound to the A_{2A}AR through a pathway connecting ECL2, the extracellular mouth between ECL2 and ECL3, and finally the receptor orthosteric site (**Figure 3A** and **Supplementary Figure 1A**). At ∼15 Å distance between CFF and the receptor residue N6.55, CFF interacted with the receptor N-terminus and ECL2 (**Supplementary Figure 1B**). At ∼10 Å distance between CFF and the receptor residue N6.55, CFF was located at the

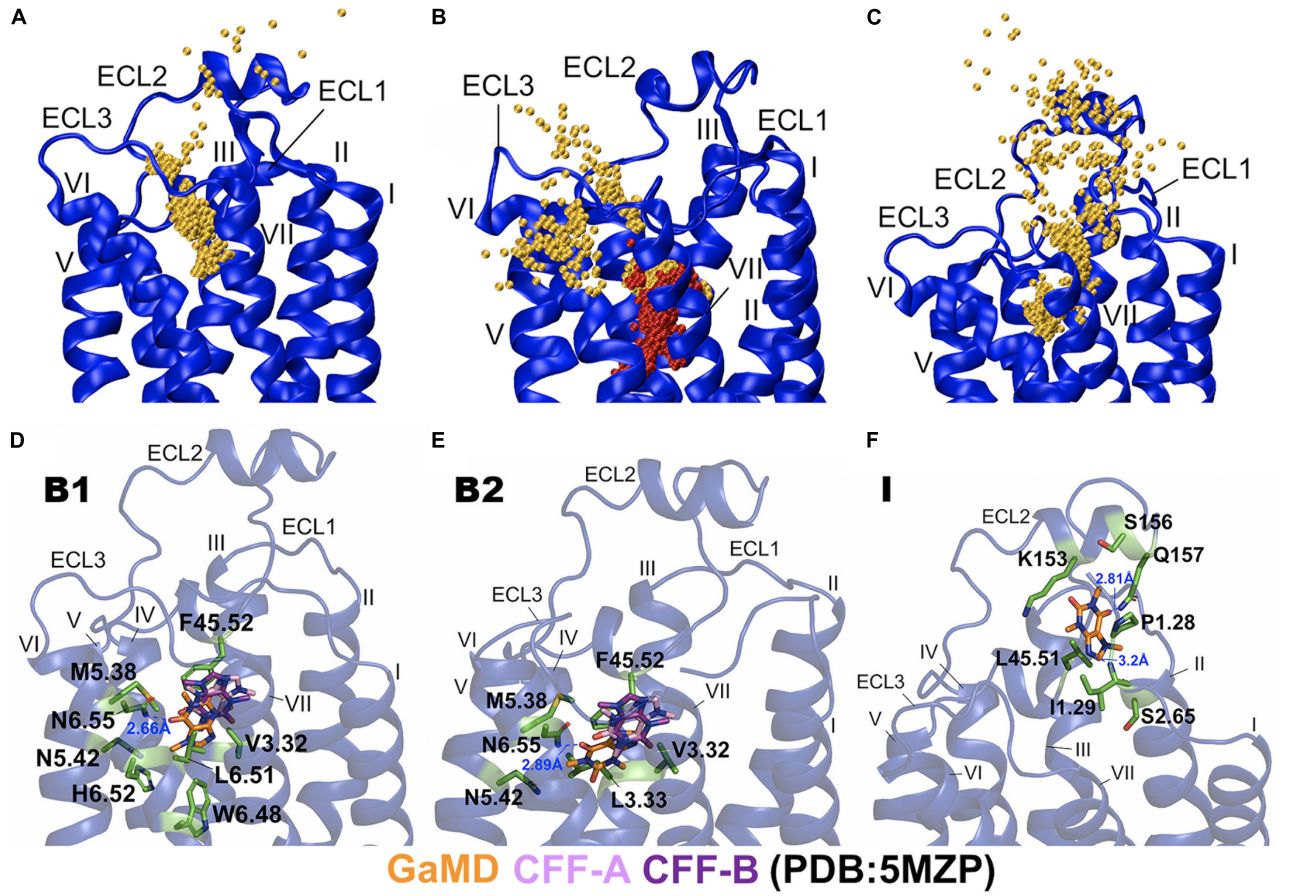


FIGURE 3 | Binding and dissociation pathways of caffeine (CFF) in the A_{2A}AR revealed from the Gaussian-accelerated molecular dynamics (GaMD) simulations. **(A)** Trace of CFF (orange) binding to the A_{2A}AR observed in the GaMD equilibration. Starting from free diffusion in the solvent, CFF bound to the orthosteric site of the A_{2A}AR receptor. **(B)** Binding of the second CFF (orange) to orthosteric pocket of the A_{2A}AR observed in GaMD Sim2. The first bound CFF is shown in red. **(C)** Pathway of CFF that dissociated from orthosteric site of the A_{2A}AR to the bulk solvent observed in GaMD Sim3. The A_{2A}AR receptor is shown in blue ribbons and the CFF traces are shown as orange beads. **(D)** The B1-bound conformational state of CFF was located between ECL2, TM3, TM5, and TM6 with interacting residues F45.52^{ECL2}, V3.32, M5.38, N5.42, W6.48, L6.51, H6.52, and N6.55. **(E)** The B2-bound conformational state of CFF was located between ECL2, TM3, TM5, and TM6 with interacting residues V3.32, L3.33, F45.52^{ECL2}, M5.38, N5.42, and N6.55. CFF formed a hydrogen bond with the receptor residue N6.55 in both the B1 and B2 states. **(F)** The intermediate (I) conformational state of CFF that was located between ECL2, N-terminus of TM1, and TM2 with interacting residues P1.28, I1.29, S2.65, K153^{ECL2}, S156^{ECL2}, Q157^{ECL2}, and L45.51^{ECL2}. CFF formed hydrogen bonds with both receptor residues I1.29 and Q157^{ECL2}. The CFF ligand is represented by sticks with carbon atoms colored in orange for simulation-derived low-energy conformations and pink and purple for two x-ray conformations in the 5MZP PDB structure. The receptor-interacting residues are highlighted in green.

extracellular mouth of the A_{2A}AR between ECL2, ECL3, and TM6, interacting with residues L45.51^{ECL2}, E169^{ECL2}, S263^{ECL3}, and T6.58 (**Supplementary Figure 1C**). At ∼5 Å distance between CFF and receptor residue N6.55, CFF bound to the receptor orthosteric site. In Sim2 of GaMD simulation trajectory, out of the nine remaining CFF molecules that freely diffused in the solvent, another CFF (CFF9) bound to the orthosteric pocket of the A_{2A}AR, while CFF2 remained bound (**Figures 1E**, **3B**). The binding pathway of CFF9 was mostly similar to that of CFF2, except a slight difference that CFF9 explored a region between ECL2 and TM6 after entry into the receptor (**Figure 3B**).

In Sim3 of GaMD simulation trajectory, CFF dissociated from the orthosteric site of the A_{2A}AR to the bulk solvent through a pathway connecting the receptor orthosteric pocket and the extracellular mouth between ECL2 and TM7 (**Figure 3C** and **Supplementary Figure 2A**). At ∼5 Å distance between CFF and the receptor residue N6.55, CFF bound to the orthosteric site of the A_{2A}AR (**Supplementary Figure 2B**). At ∼10 Å distance between CFF and the receptor residue N6.55, CFF was located at the extracellular mouth of the A_{2A}AR between ECL2 and TM7, interacting with residues L45.51^{ECL2}, I1.29, I2.64, L7.32, and Y7.36 (**Supplementary Figure 2C**). At ∼15 Å distance between CFF and the receptor residue N6.55, CFF moved near ECL2–TM1 and interacted with receptor residues L45.51^{ECL2}, A1.26, P1.27, P1.28, and I1.29 (**Supplementary Figure 2D**). At ∼20 Å distance between CFF and the receptor residue N6.55, CFF is in the intermediate (I) conformational state. While the GaMD simulations were not sufficiently converged with only a few ligand-binding events captured, the binding and dissociation pathways of CFF characterized using the GaMD energetically

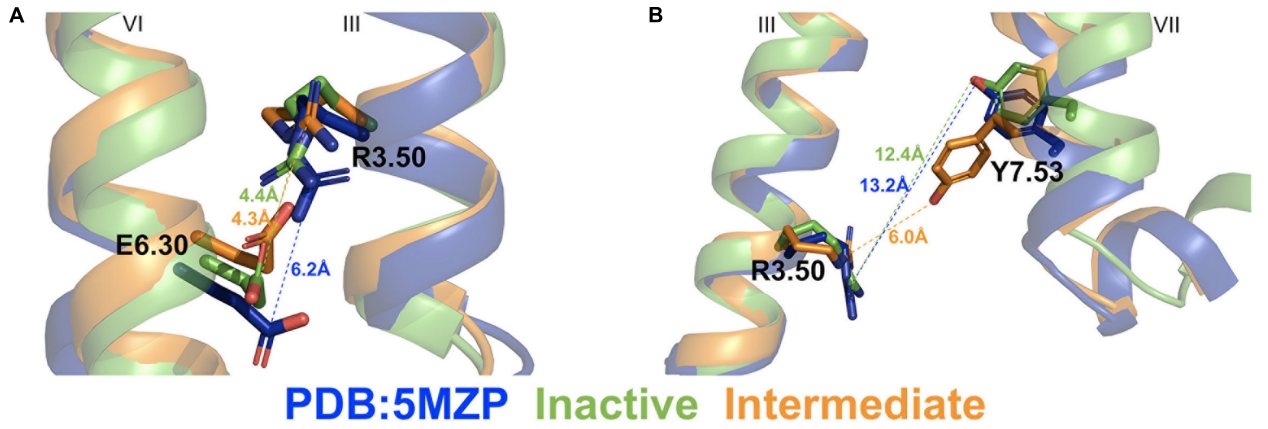


FIGURE 4 | Representative "inactive" (green) and "intermediate" (orange) low-energy conformations of the A_{2A}AR compared with the 5MZP PDB structure (blue). **(A)** The ionic lock distance between charge centers of residues R3.50 and E6.30 in the 5MZP, inactive, and intermediate conformations are 6.0, 4.4, and 4.3 Å, respectively. **(B)** The distance between atom CZ in residue R3.50 and hydroxyl oxygen atom of Y7.53 in the 5MZP, inactive, and intermediate conformations are 13.2, 12.4, and 6.0 Å, respectively.

reweighted structural clusters of the ligand (**Supplementary Figure 4**) were similar to those as shown in **Figure 2**.

Low-Energy Binding Poses of CFF in the A_{2A}AR

Next, we combined GaMD simulations of CFF binding to the A_{2A}AR and clustered the simulation snapshots of CFF to obtain representative structural clusters that corresponded to the low-energy states in the 2D PMF free energy profiles (**Figure 2A**). In the B1-bound state, CFF bound to the orthosteric pocket of the A_{2A}AR and interacted with residues F45.52^{ECL2}, V3.32, M5.38, N5.42, W6.48, L6.51, H6.52, and N6.55. In particular, a hydrogen bond was formed between the ND2 atom of receptor residue N6.55 and O13 atom of CFF at a distance of 2.7 Å (**Figure 3D** and **Supplementary Figure 1D**). In the B2-bound state, CFF bound to the orthosteric pocket of the A_{2A}AR in the presence of another CFF molecule in the pocket. The orthosteric pocket was located within the receptor TM bundle between ECL2, TM3, TM5, and TM6. CFF interacted with residues F45.52^{ECL2}, V3.32, L3.33, M5.38, N5.42, and N6.55. In particular, a hydrogen bond was formed between the ND2 atom of receptor residue N6.55 and O13 atom of CFF at a distance of 2.9 Å (**Figure 3E**).

In the intermediate (I) conformational state (**Figure 3F**), CFF was located at the extracellular mouth of the A_{2A}AR between ECL2 and TM1, interacting with residues P1.28, I1.29, S2.65, K153^{ECL2}, S156^{ECL2}, Q157^{ECL2}, and L45.51^{ECL2}. In particular, a hydrogen bond was formed between the N atom of receptor residue I1.29 and N9 atom of CFF and another hydrogen bond was formed between the NE2 atom of receptor residue Q157^{ECL2} and O13 atom of CFF (**Figure 3F**).

Conformational Changes of the A_{2A}AR During CFF Binding

Two different low-energy conformational states were identified from GaMD simulations of the A_{2A}AR during CFF binding, including the inactive and intermediate states (**Figure 2B**). The hierarchical agglomerative clustering algorithm was used to cluster snapshots of the A_{2A}AR conformations with all the GaMD production simulations combined. The combined GaMD simulation trajectories were clustered to identify representative low-energy conformational states of the receptor (**Figure 4**). The ionic lock distance between residues R3.50 and E6.30 changed from 6 Å in the 5MZP PDB structure to 4.3–4.4 Å in the inactive and intermediate conformations (**Figure 4A**). The distance between the atom CZ of residue R3.50 and the OH atom of residue Y7.53 decreased from 12.4 Å in the inactive conformation (similar to 13.2 Å in the 5MZP PDB structure) to 6.0 Å in the intermediate cluster (**Figure 4B**). Therefore, the NPxxY motif, a highly conserved motif in the intracellular end of TM7 of class A GPCRs, moved inward during the conformational transition of the A_{2A}AR from the inactive to the intermediate state.

DISCUSSION

In this study, all-atom GaMD simulations have been applied to elucidate the pathways and mechanism of CFF binding to the human A_{2A}AR. The GaMD simulations have successfully captured both spontaneous binding and dissociation of CFF in the receptor. With GaMD-enhanced sampling, we were able to simulate the complete binding of the CFF antagonist with the final orthosteric pocket deeply buried in the receptor TM domain. However, it is important to note that only two ligand-binding events and one dissociation event were observed in the presented GaMD simulations (**Figure 1**). Quantitative characterization of the ligand-binding free energy and kinetics would require sampling of significantly more ligand-binding events, which will be investigated in the future using a very recently developed, potentially more efficient Ligand GaMD (LiGaMD) method (Miao et al., 2020) and other applicable algorithms. Nevertheless, energetic reweighting of

the GaMD simulations enabled us to identify relatively low-energy conformational states of CFF binding to the A_{2A}AR. Our results were consistent with the experimental data from 3RFM PDB (Doré et al., 2011) and 5MZP PDB (Cheng et al., 2017). In the 3RFM PDB structure, the residues interacting with CFF include F45.52^{ECL2}, M5.38, L6.51, N6.55, M7.35, and I7.39. In the 5MZP PDB structure, the residues interacting with CFF include I2.64, V3.32, F45.52^{ECL2}, M5.38, L6.51, N6.55, and I7.39. The B1- and B2-bound conformations identified from the GaMD free energy profiles were comparable with the experimental 3RFM and 5MZP PDB structures in terms of the ligand interacting residues in the orthosteric pocket and distances between residues R3.50 and E6.30 and residues R3.50 and Y7.53.

We identified a dominant pathway of CFF binding to the A_{2A}AR from the GaMD simulations. CFF approached the A_{2A}AR through interactions with ECL2, extracellular mouth between ECL2, ECL3, and TM7, and finally the receptor orthosteric site located deeply within the receptor TM bundle (**Figure 3A** and **Supplementary Figure 1**). A slightly different binding pathway was observed when two CFF molecules bound to the orthosteric pocket of the A_{2A}AR. In this pathway, the second CFF explored a region between ECL3 and TM7 during the binding process (**Figure 3B**). The dissociation pathway of CFF observed from the GaMD simulation was mostly the reverse of the dominant binding pathway (**Figure 2** and **Supplementary Figure 4**). CFF moved from the receptor orthosteric site to the extracellular mouth between ECL2 and TM7 and then ECL2 and TM1 before dissociating to the bulk solvent (**Figure 3C** and **Supplementary Figure 2**). Two low-energy conformational states were identified from the GaMD simulations of the A_{2A}AR during CFF binding, i.e., the inactive and intermediate states. In the inactive state, the distances between residues R3.50 and E6.30 and R3.50 and Y7.53 were 4.4 and 12.4 Å, respectively. In this context, the average distances between residues R3.50 and E6.30 and R3.50 and Y7.53 in 46 experimental structures of the inactive A_{2A}AR (**Supplementary Table 2**; Jaakola et al., 2008; Doré et al., 2011; Congreve et al., 2012; Hino et al., 2012; Liu et al., 2012; Batyuk et al., 2016; Segala et al., 2016; Cheng et al., 2017; Martin-Garcia et al., 2017, 2019; Melnikov et al., 2017; Sun et al., 2017; Weinert et al., 2017; Broecker et al., 2018; Eddy et al., 2018; Rucktooa et al., 2018; Ishchenko et al., 2019; Shimazu et al., 2019; Borodovsky et al., 2020; Ihara et al., 2020; Jespers et al., 2020; Lee et al., 2020; Nass et al., 2020) were calculated to be 6.5 $\pm$ 1.0 Å and 12.7 $\pm$ 0.4 Å, respectively. Therefore, the highly conserved residues R3.50 and E6.30 ionic lock became fully closed in the GaMD simulations of the inactive A_{2A}AR during binding of the CFF antagonist. In comparison, the average distances between residues R3.50 and E6.30 and residues R3.50 and Y7.53 in the nine available structures of active A_{2A}AR (**Supplementary Table 2**; Lebon et al., 2011, 2015; Xu et al., 2011; Carpenter et al., 2016; Garcia-Nafria et al., 2018; White et al., 2018) were calculated as 11.1 $\pm$ 0.4 Å and 4.4 $\pm$ 0.2 Å, respectively. No intermediate structure is currently available for the A_{2A}AR (Pándy-Szekeres et al., 2017). In the GaMD-predicted intermediate conformational state of the A_{2A}AR, the ionic lock distance between residues R3.50 and E6.30 was 4.3 Å, similar to that in the inactive receptor. The distance between residues R3.50 and Y7.53, however, decreased to 6.0 Å, comparable with the average of active A_{2A}AR structures. Therefore, while the ionic lock remained closed, the conserved NPxxY motif in the intracellular end of TM7 was able to move inward in the intermediate state of the A_{2A}AR, being consistent with the pathway and mechanism of GPCR activation revealed from earlier studies (Dror et al., 2011; Miao et al., 2013).

The binding and dissociation of CFF antagonist in our GaMD simulations of the A_{2A}AR involved receptor residues P1.28, I1.29, S2.65, L45.51^{ECL2}, F45.52^{ECL2}, K153^{ECL2}, S156^{ECL2}, Q157^{ECL2}, E169^{ECL2}, A259^{ECL3}, S263^{ECL3}, H264^{ECL3}, N6.55, T6.58, F6.59, P7.31, L7.32, L7.34, M7.35, and Y7.36 (**Figure 3** and **Supplementary Figures 1, 2**). The orthosteric pocket was located within the receptor TM bundle and made of receptor residues F45.52^{ECL2}, V3.32, L3.33, M5.38, N5.42, W6.48, L6.51, H6.52, and N6.55. Notably, CFF formed a hydrogen bond with receptor residue N6.55. The four most populated CFF binding poses in the A_{2A}AR found by Cao et al. consisted of residues A2.61, I2.64, S2.65, V3.32, L3.33, T3.36, F45.52, E169^{ECL2}, M5.38, N5.42, L6.51, H6.52, N6.55, H264^{ECL3}, M7.35, I7.39, and H7.43. Furthermore, CFF formed a hydrogen bond with receptor residue N6.55 and water-bridge contact with residue H7.43 (Cao et al., 2015). The ligand dissociation pathway in the A_{2A}AR discovered by Guo et al. involved 16 receptor residues: G1^{TM1}, I2.63, S2.64, T2.65, Q148^{ECL2}, G152^{ECL2}, K153^{ECL2}, S156^{ECL2}, Q157^{ECL2}, E169^{ECL2}, T6.58, H7.29, A7.30, P7.31, L7.32, and Y7.36 (Guo et al., 2016). The residues in the orthosteric binding site of the A_{2A}AR revealed by Caliman et al. (2018) were I2.64, V3.32, L3.33, T3.36, Q3.37, I3.40, L45.51^{ECL2}, F45.52^{ECL2}, E169^{ECL2}, M5.38, N5.42, W6.48, L6.51, H6.52, N6.55, T6.58, H264^{ECL3}, L7.32, M7.35, Y7.36, I7.39, S7.42, and H7.43. Overall, our results were in good agreement with previous studies of the A_{2A}AR, in terms of the receptor residues involved in the ligand dissociation and binding.

An intermediate ligand-binding site was also revealed from the GaMD simulations of CFF binding and dissociation in the A_{2A}AR. It was located at the extracellular mouth between ECL2 and TM1 of the A_{2A}AR. This region has been identified as an allosteric site of many class A GPCRs (Dror et al., 2013; Kruse et al., 2013; Miao and McCammon, 2016; Miao et al., 2018; Pawnikar and Miao, 2020). Taken together, our simulations suggest that CFF binds to the orthosteric pocket of A_{2A}AR *via* an intermediate site located at the receptor extracellular mouth. The ECL2 with high sequence divergence could serve as an attractive target site for designing allosteric modulators as selective drugs of the A_{2A}AR and other ARs (Miao et al., 2018).

DATA AVAILABILITY STATEMENT

The authors acknowledge that the data presented in this study must be deposited and made publicly available in an acceptable repository, prior to publication. Frontiers cannot accept a manuscript that does not adhere to our open data policies.

AUTHOR CONTRIBUTIONS

YM designed the research. HD performed the research. HD, SA, and YM analyzed the data. HD and YM wrote the manuscript. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

We appreciate the preliminary simulations of ARs by Keeley Collins and Amar Kumar. This work used supercomputing resources with allocation award TG-MCB180049 through the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1548562, and project M2874 through the National Energy Research Scientific Computing Center (NERSC), which is a U.S. Department of Energy Office of Science User Facility operated under Contract No. DE-AC02-05CH11231, and the Research Computing Cluster at the University of Kansas. This work was supported in part by the American Heart Association (Award 17SDG33370094), the National Institutes of Health (R01GM132572), and the startup funding in the College of Liberal Arts and Sciences at the University of Kansas.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2021.673170/full#supplementary-material

REFERENCES

Ballesteros, J. A., and Weinstein, H. (1995). Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors. *Methods Neurosci.* 25, 366–428. doi: 10.1016/s1043-9471(05)80049-7

Batyuk, A., Galli, L., Ishchenko, A., Han, G. W., Gati, C., Popov, P. A., et al. (2016). Native phasing of x-ray free-electron laser data for a G protein-coupled receptor. *Sci. Adv.* 2:e1600292. doi: 10.1126/sciadv.1600292

Bhattarai, A., Wang, J., and Miao, Y. (2020). G-protein-coupled receptor-membrane interactions depend on the receptor activation state. *J. Comput. Chem.* 41, 460–471. doi: 10.1002/jcc.26082

Borodovsky, A., Barbon, C. M., Wang, Y., Ye, M., Prickett, L., Chandra, D., et al. (2020). Small molecule AZD4635 inhibitor of A2AR signaling rescues immune cell function including CD103+ dendritic cells enhancing anti-tumor immunity. *J. Immunother. Cancer* 8:e000417. doi: 10.1136/jitc-2019-000417

Broecker, J., Morizumi, T., Ou, W.-L., Klingel, V., Kuo, A., Kissick, D. J., et al. (2018). High-throughput in situ X-ray screening of and data collection from protein crystals at room temperature and under cryogenic conditions. *Nat. Protoc.* 13, 260–292. doi: 10.1038/nprot.2017.135

Caliman, A. D., Miao, Y., and McCammon, J. A. (2018). Mapping the allosteric sites of the A2A adenosine receptor. *Chem. Biol. Drug Des.* 91, 5–16.

Cao, R. Y., Rossetti, G., Bauer, A., and Carloni, P. (2015). Binding of the antagonist CFF to the human adenosine receptor hA2AR in nearly physiological conditions. *PLoS One* 10:e0126833. doi: 10.1371/journal.pone.0126833

Carpenter, B., and Lebon, G. (2017). Human adenosine A2A receptor: molecular mechanism of ligand binding and activation. *Front. Pharmacol.* 8:898. doi: 10.3389/fphar.2017.00898

Carpenter, B., Nehmé, R., Warne, T., Leslie, A. G. W., and Tate, C. G. (2016). Structure of the adenosine A2A receptor bound to an engineered G protein. *Nature* 536:104. doi: 10.1038/nature18966

Case, D., Belfon, K., Ben-Shalom, I. Y., Brozell, S. R., Cerutti, D. S., Cheatham, V. W. D. III, et al. (2020). *AMBER 2020*. San Francisco, CA: University of California.

Cheng, R. K. Y., Segala, E., Robertson, N., Deflorian, F., Doré, A. S., Errey, J. C., et al. (2017). Structures of human A1 and A2A adenosine receptors with xanthines reveal determinants of selectivity. *Structure* 25, 1275–1285.e4.

Congreve, M., Andrews, S. P., Doré, A. S., Hollenstein, K., Hurrell, E., Langmead, C. J., et al. (2012). Discovery of 1,2,4-Triazine derivatives as adenosine A(2A) antagonists using structure based drug design. *J. Med. Chem.* 55, 1898–1903. doi: 10.1021/jm201376w

Doré, A. S., Robertson, N., Errey, J. C., Ng, I., Hollenstein, K., Tehan, B., et al. (2011). Structure of the adenosine A2A receptor in complex with ZM241385 and the Xanthines XAC and CFF. *Structure* 19, 1283–1293. doi: 10.1016/j.str.2011.06.014

Dror, R. O., Arlow, D. H., Maragakis, P., Mildorf, T. J., Pan, A. C., Xu, H., et al. (2011). Activation mechanism of the 2-adrenergic receptor. *Proc. Natl. Acad. Sci. U.S.A.* 108, 18684–18689.

Dror, R. O., Green, H. F., Valant, C., Borhani, D. W., Valcourt, J. R., Pan, A. C., et al. (2013). Structural basis for modulation of a G-protein coupled receptor by allosteric drugs. *Nature* 503, 295–299. doi: 10.1038/nature12595

East, K. W., Newton, J. C., Morzan, U. N., Narkhede, Y. B., Acharya, A., Skeens, E., et al. (2020). Allosteric motions of the CRISPR-Cas9 HNH nuclease probed by NMR and molecular dynamics. *J. Am. Chem. Soc.* 142, 1348–1358. doi: 10.1021/jacs.9b10521

Eddy, M., Lee, M.-Y., Gao, Z.-G., White, K. L., Didenko, T., Horst, R., et al. (2018). Allosteric coupling of drug binding and intracellular signaling in the A2A adenosine receptor. *Cell* 172, 68–80.e12.

Fredholm, B., Abbracchio, M. P., Burnstock, G., Dubyak, G. R., Harden, T. K., Jacobson, K. A., et al. (1997). Towards a revised nomenclature for P1 and P2 receptors. *Trends Pharmacol. Sci.* 18, 79–82. doi: 10.1016/s0165-6147(96)01038-3

Garcia-Nafria, J., Lee, Y., Bai, X., Carpenter, B., and Tate, C. G. (2018). Cryo-EM structure of the adenosine A2A receptor coupled to an engineered heterotrimeric G protein. *eLife* 7:e35946.

Guixà-González, R., Rodriguez-Espigares, I., Ramíez-Anguita, J. M., Carrió-Gaspar, P., Martinez-Seara, H., Giorgino, T., et al. (2014). MEMBPLUGIN: studying membrane complexity in VMD. *Bioinformatics* 30, 1478–1480. doi: 10.1093/bioinformatics/btu037

Guo, D., Pan, A. C., Dror, R. O., Mocking, T., Liu, R., Heitman, L. H., et al. (2016). Molecular basis of ligand dissociation from the adenosine A2A receptor. *Mol. Pharmacol.* 89, 485–491. doi: 10.1124/mol.115.102657

Hamelberg, D., Mongan, J., and McCammon, J. (2004). Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. *J. Chem. Phys.* 120, 11919–11929. doi: 10.1063/1.1755656

He, X., Man, V. H., Yang, W., Lee, T.-S., and Wang, J. (2020). A fast and high-quality charge model for the next generation general AMBER force field. *J. Chem. Phys.* 153:114502. doi: 10.1063/5.0019056

Hino, T., Arakawa, T., Iwanari, H., Yurugi-Kobayashi, T., Ikeda-Suno, C., Nakada-Nakura, Y., et al. (2012). G-protein-coupled receptor inactivation by an allosteric inverse-agonist antibody. *Nature* 482, 237–240. doi: 10.1038/nature10750

Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. *J. Mol. Graph.* 14, 33–38. doi: 10.1016/0263-7855(96)00018-5

Ihara, K., Hato, M., Nakane, T., Yamashita, K., Kimura-Someya, T., Hosaka, T., et al. (2020). Isoprenoid-chained lipid EROCOC 17+4 : a new matrix for membrane protein crystallization and a crystal delivery medium in serial femtosecond crystallography. *Sci. Rep.* 10:19305.

Ishchenko, A., Stauch, B., Han, G. W., Batyuk, A., Shiriaeva, A., Li, C., et al. (2019). Toward G protein-coupled receptor structure-based drug design using X-ray lasers. *IUCrJ* 6, 1106–1119. doi: 10.1107/s2052252519013137

Jaakola, V.-P., Griffith, M. T., Hanson, M. A., Cherezov, V., Chien, E. Y. T., Lane, J. R., et al. (2008). The 2.6 angstrom crystal structure of a human A2A adenosine receptor bound to an antagonist. *Science* 322, 1211–1217. doi: 10.1126/science.1164772

Jacobson, K. A., and Gao, Z.-G. (2006). Adenosine receptors as therapeutic targets. *Nat. Rev. Drug Discov.* 5, 247–264.

Jespers, W., Verdon, G., Azuaje, J., Majellaro, M., Keranen, H., García-Mera, X., et al. (2020). X-Ray crystallography and free energy calculations reveal the binding mechanism of A 2A adenosine receptor antagonists. *Angew. Chem. Int. Ed. Engl.* 59, 16536–16543. doi: 10.1002/anie.202003788

Jo, S., Kim, T., and Im, W. (2007). Automated builder and database of protein/membrane complexes for molecular dynamics simulations. *PLoS One* 2:e880. doi: 10.1371/journal.pone.0000880

Jo, S., Kim, T., Iyer, V., and Im, W. (2008). CHARMM-GUI: a web-based graphical user interface for CHARMM. *J. Comput. Chem.* 29, 1859–1865. doi: 10.1002/jcc.20945

Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79, 926–935. doi: 10.1063/1.445869

Kruse, A. C., Ring, A. M., Manglik, A., Hu, J., Hu, K., Eitel, K., et al. (2013). Activation and allosteric modulation of a muscarinic acetylcholine receptor. *Nature* 504, 101–106.

Lebon, G., Edwards, P., Leslie, A., and Tate, C. (2015). Molecular determinants of Cgs21680 binding to the human adenosine A2A receptor. *Mol. Pharmacol.* 87:907. doi: 10.1124/mol.114.097360

Lebon, G., Warne, T., Edwards, P. C., Bennett, K., Langmead, C. J., Leslie, A. G. W., et al. (2011). Agonist-bound adenosine A2A receptor structures reveal common features of GPCR activation. *Nature* 474, 521–525. doi: 10.1038/nature10136

Lebon, G., Warne, T., and Tate, C. (2012). Agonist-bound structures of G protein-coupled receptors. *Curr. Opin. Struct. Biol.* 22, 482–490. doi: 10.1016/j.sbi.2012.03.007

Lee, J., Cheng, X., Swalis, J. M., Yeom, M. S., Eastman, P. K., Lemkul, J. A., et al. (2016). CHARMM-GUI input generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM simulations using the CHARMM36 additive force field. *J. Chem. Theory Comput.* 12, 405–413. doi: 10.1021/acs.jctc.5b00935

Lee, J., Hitzenberger, M., Rieger, M., Kern, N. R., Zacharias, M., and Wonpil, I. (2020). CHARMM-GUI supports the Amber force fields. *J. Chem. Phys.* 153:035103. doi: 10.1063/5.0012280

Lee, M.-Y., Geiger, J., Ishchenko, A., Han, G. W., Barty, A., White, T. A., et al. (2020). Harnessing the power of an X-ray laser for serial crystallography of membrane proteins crystallized in lipidic cubic phase. *IUCrJ* 7, 976–984. doi: 10.1107/s2052252520012701

Liu, W., Chun, E., Thompson, A. A., Chubukov, P., Xu, F., Katritch, V., et al. (2012). Structural basis for allosteric regulation of GPCRs by sodium ions. *Science* 337, 232–236. doi: 10.1126/science.1219218

Martin-Garcia, J., Conrad, C. E., Nelson, G., Stander, N., Zatsepin, N. A., Zook, J., et al. (2017). Serial millisecond crystallography of membrane and soluble protein microcrystals using synchrotron radiation. *IUCrJ* 4, 439–454.

Martin-Garcia, J., Zhu, L., Mendez, D., Lee, M.-Y., Chun, E., Li, C., et al. (2019). High-viscosity injector-based pink-beam serial crystallography of microcrystals at a synchrotron radiation source. *IUCrJ* 6, 412–425. doi: 10.1107/s205225251900263x

Melnikov, I., Polovinkin, V., Kovalev, K., Gushchin, I., Shevtsov, M., Shevchenko, V., et al. (2017). Fast iodide-SAD phasing for high-throughput membrane protein structure determination. *Sci. Adv.* 3:e1602952. doi: 10.1126/sciadv.1602952

Miao, Y., Bhattarai, A., Nguyen, A. T. N., Christopoulos, A., and May, L. T. (2018). Structural basis for binding of allosteric drug leads in the adenosine A1 receptor. *Sci. Rep.* 8:16836.

Miao, Y., Bhattarai, A., and Wang, J. (2020). Ligand gaussian accelerated molecular dynamics (LiGaMD): characterization of ligand binding and thermodynamics and kinetics. *J. Chem. Theory Comput.* 16, 5526–5547.

Miao, Y., Feher, V. A., and McCammon, J. A. (2015). Gaussian accelerated molecular dynamics: unconstrained enhanced sampling and free energy calculation. *J. Chem. Theory Comput.* 11, 3584–3595. doi: 10.1021/acs.jctc.5b00436

Miao, Y., and McCammon, J. A. (2016). Graded activation and free energy landscapes of a muscarinic G-protein-coupled receptor. *Proc. Natl. Acad. Sci. U.S.A.* 113, 12162–12167. doi: 10.1073/pnas.1614538113

Miao, Y., and McCammon, J. A. (2018). Mechanism of the G-protein mimetic nanobody binding to a muscarinic G-protein-coupled receptor. *Proc. Natl. Acad. Sci. U.S.A.* 115, 3036–3041. doi: 10.1073/pnas.1800756115

Miao, Y., Nichols, S. E., Gasper, P. M., Metzger, V. T., and McCammon, J. A. (2013). Activation and dynamic network of the M2 muscarinic receptor. *Proc. Natl. Acad. Sci. U.S.A.* 110, 10982–10987.

Miao, Y., Sinko, W., Pierce, L., Bucher, D., Walker, R. C., and McCammon, J. A. (2014). Improved reweighting of accelerated molecular dynamics simulations for free energy calculation. *J. Chem. Theory Comput.* 10, 2677–2689. doi: 10.1021/ct500090q

Nass, K., Cheng, R., Vera, L., Mozzanica, A., Redford, S., Ozerov, D., et al. (2020). Advances in long-wavelength native phasing at X-ray free-electron lasers. *IUCrJ* 7, 965–975.

Pándy-Szekeres, G., Munk, C., Tsonkov, T. M., Mordalski, S., Harpsoe, K., Hauser, A. S., et al. (2017). GPCRdb in 2018: adding GPCR structure models and ligands. *Nucleic Acids Res.* 46, D440–D446.

Pang, Y., Miao, Y., and McCammon, J. (2017). Gaussian accelerated molecular dynamics in NAMD. *J. Chem. Theory Comput.* 13, 9–19. doi: 10.1021/acs.jctc.6b00931

Pawnikar, S., and Miao, Y. (2020). Pathway and mechanism of drug binding to chemokine receptors revealed by accelerated molecular simulations. *Future Med. Chem.* 12, 1213–1225. doi: 10.4155/fmc-2020-0044

Porkka-Heiskanen, T., Strecker, R. E., Thakkar, M., Bjorkum, A. A., Greene, R. W., and McCarley, R. W. (1997). Adenosine: a mediator of the sleep-inducing effects of prolonged wakefulness. *Science* 276, 1265–1268. doi: 10.1126/science.276.5316.1265

Ricci, C. G., Chen, J. S., Miao, Y., Jinek, M., Doudna, J. A., McCammon, J. A., et al. (2019). Deciphering Off-target effects in CRISPR-Cas9 through accelerated molecular dynamics. *ACS Cent. Sci.* 5, 651–662. doi: 10.1021/acscentsci.9b00020

Roe, D. R., and Cheatham, I. T. E. (2013). PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.* 9, 3084–3095. doi: 10.1021/ct400341p

Rucktooa, P., Cheng, R. K. Y., Segala, E., Geng, T., Errey, J. C., Brown, G. A., et al. (2018). Towards high throughput GPCR crystallography: In Meso soaking of Adenosine A2A receptor crystals. *Sci. Rep.* 8:41.

Salomon-Ferrer, R., Gotz, A. W., Poole, D., Le Grand, S., and Walker, R. C. (2013). Routined microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* 9, 3878–3888. doi: 10.1021/ct400314y

Segala, E., Guo, D., Cheng, R. K. Y., Bortolato, A., Deflorian, F., Doré, A. S., et al. (2016). Controlling the dissociation of ligands from the adenosine A2A receptor through modulation of salt bridge strength. *J. Med. Chem.* 59, 6470–6479. doi: 10.1021/acs.jmedchem.6b00653

Shen, T., and Hamelberg, D. (2008). A statistical analysis of the precision reweighting-based simulations. *Journal of Chem. Phys.* 129:034103. doi: 10.1063/1.2944250

Shimazu, Y., Tono, K., Tanaka, T., Yamanaka, Y., Nakane, T., Mori, C., et al. (2019). High-viscosity sample-injection device for serial femtosecond crystallography at atmospheric pressure. *J. Appl. Crystallogr.* 52, 1280–1288. doi: 10.1107/s1600576719012846

Sun, B., Bachhawat, P., Chu, M. L.-H., Wood, M., Ceska, T., Sands, Z. A., et al. (2017). Crystal structure of the adenosine A2A receptor bound to an antagonist reveals a potential allosteric pocket. *Proc. Natl. Acad. Sci. U.S.A.* 114, 2066–2071. doi: 10.1073/pnas.1621423114

Tian, C., Kasavajhala, K., Belfon, K. A. A., Raguette, L., Huang, H., Migues, A. N., et al. (2019). ff19SB: amino-acid specific protein backbone parameters trained against quantum mechanics energy surfaces in solution. *J. Chem. Theory Comput.* 16, 528–552. doi: 10.1021/acs.jctc.9b00591

Wallace, A., Laskowski, R., and Thornton, J. (1995). LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng. Des. Select.* 8, 127–134. doi: 10.1093/protein/8.2.127

Wang, J., and Miao, Y. (2019). Mechanistic insights into specific G protein interactions with adenosine receptors. *J. Phys. Chem. B* 123, 6462–6473. doi: 10.1021/acs.jpcb.9b04867

Wang, J., Wang, W., Kollman, P., and Case, D. (2006). Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model.* 25:247260.

Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. (2004). Development and testing of a general AMBER force field. *J. Comput. Chem.* 25, 1157–1174. doi: 10.1002/jcc.20035

Weinert, T., Olieric, N., Cheng, R., Brunle, S., James, D., Ozerov, D., et al. (2017). Serial millisecond crystallography for routine room-temperature structure determination at synchrotrons. *Nat. Commun.* 8: 542.

White, K., Eddy, M., Gao, Z.-G., Han, G. W., Lian, T., Deary, A., et al. (2018). Structural connection between activation microswitch and allosteric sodium site in GPCR signaling. *Structure* 26, 259–269.e5.

Wu, E., Cheng, X., Jo, S., Rui, H., Song, K. C., Dávilia-Contreras, E. M., et al. (2014). CHARMM-GUI membrane builder toward realistic biological membrane simulations. *J. Comput. Chem.* 35, 1997–2004.

Xu, F., Wu, H., Katritch, V., Han, G. W., Jacobson, K. A., Gao, Z.-G., et al. (2011). Structure of an agonist-bound human A2A adenosine receptor. *Science* 332, 322–327.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Molecular Biosciences

ORIGINAL RESEARCH
published: 06 August 2021
doi: 10.3389/fmolb.2021.719320

Fibril Surface-Dependent Amyloid Precursors Revealed by Coarse-Grained Molecular Dynamics Simulation

*Yuan-Wei Ma, Tong-You Lin and Min-Yeh Tsai**

Department of Chemistry, Tamkang University, New Taipei City, Taiwan

Amyloid peptides are known to self-assemble into larger aggregates that are linked to the pathogenesis of many neurodegenerative disorders. In contrast to primary nucleation, recent experimental and theoretical studies have shown that many toxic oligomeric species are generated through secondary processes on a pre-existing fibrillar surface. Nucleation, for example, can also occur along the surface of a pre-existing fibril—secondary nucleation—as opposed to the primary one. However, explicit pathways are still not clear. In this study, we use molecular dynamics simulation to explore the free energy landscape of a free Abeta monomer binding to an existing fibrillar surface. We specifically look into several potential Abeta structural precursors that might precede some secondary events, including elongation and secondary nucleation. We find that the overall process of surface-dependent events can be described at least by the following three stages: 1. Free diffusion 2. Downhill guiding 3. Dock and lock. And we show that the outcome of adding a new monomer onto a pre-existing fibril is pathway-dependent, which leads to different secondary processes. To understand structural details, we have identified several monomeric amyloid precursors over the fibrillar surfaces and characterize their heterogeneity using a probability contact map analysis. Using the frustration analysis (a bioinformatics tool), we show that surface heterogeneity correlates with the energy frustration of specific local residues that form binding sites on the fibrillar structure. We further investigate the helical twisting of protofilaments of different sizes and observe a length dependence on the filament twisting. This work presents a comprehensive survey over the properties of fibril growth using a combination of several openMM-based platforms, including the GPU-enabled openAWSEM package for coarse-grained modeling, MDTraj for trajectory analysis, and pyEMMA for free energy calculation. This combined approach makes long-timescale simulation for aggregation systems as well as all-in-one analysis feasible. We show that this protocol allows us to explore fibril stability, surface binding affinity/heterogeneity, as well as fibrillar twisting. All these properties are important for understanding the molecular mechanism of surface-catalyzed secondary processes of fibril growth.

Keywords: abeta, MD simulation, coarse-grained model, fibril surface, secondary nucleation, fibrillar twisting, binding sites, elongation (growth)

OPEN ACCESS

Edited by:
Davide Mercadante,
University of Auckland, New Zealand

Reviewed by:
Wei Chen,
Independent researcher, Austin, TX, United States
Sandro Bottaro,
Humanitas University, Italy

****Correspondence:***
Min-Yeh Tsai
mytsai@mail.tku.edu.tw

Specialty section:
This article was submitted to Biological Modeling and Simulation, a section of the journal Frontiers in Molecular Biosciences

Received: *02 June 2021*
Accepted: *26 July 2021*
Published: *06 August 2021*

Citation:
Ma Y-W, Lin T-Y and Tsai M-Y (2021) Fibril Surface-Dependent Amyloid Precursors Revealed by Coarse-Grained Molecular Dynamics Simulation. Front. Mol. Biosci. 8:719320. doi: 10.3389/fmolb.2021.719320

1 INTRODUCTION

The formation of oligomeric species of Abeta protein and subsequent amyloid deposition are implicated in causing the pathogenesis of Alzhaimer's Disease (AD) (Chen and Mobley, 2019). As the hallmark of AD, amyloid fibrils display a range of structural variations called fibril polymorphism (Tycko, 2015; Riek and Eisenberg, 2016), which challenges the developments for molecular imaging and therapeutic strategies (Fändrich et al., 2018). The fibrillar structure of amyloid-beta peptides (Abeta40 and Abeta42), for example, are quite different (Colvin et al., 2016; Wälti et al., 2016; Gremer et al., 2017), but they share similar protofilament structures and both display primarily left-handed twisted filament architecture *in vitro* (Schmidt et al., 2009; Zhang et al., 2009). Brain-derived amyloid fibrils, however, are right-handed (Kollmer et al., 2019). These studies demonstrate how amyloid proteins show their structural plasticity under different contexts, and thus are important in determining the pathogenesis of neurodegenerative disorders. Despite many fibrillar morphologies being available, the molecular mechanism underlying the aggregation process of amyloid proteins is still not fully understood. In the process of fibril growth, cross-seeding experiments have revealed some correlation of fibril growth over fibrillar surfaces with selected amyloid peptides. Abeta40 and Abeta42 can cross-seed their constituent fibrils, however, the growth rate displays a different profile for different original fibril seeds (Thacker et al., 2020). Different amyloid peptides can also mutually seed each other. These results suggest a common structural feature of the fibrillar surface that exhibit physicochemical similarity at the molecular level, though identical amyloid backbone virtually is not sufficient for cross-seeding (Daskalov et al., 2021).

Amyloids form by a sequence of chemical reactions. Protein monomers first need to oligomerize into critical nuclei through primary nucleation (Tsai, 2019). These nuclei then may transform into active oligomer species for subsequent secondary processes to occur, for example, elongation, fragmentation, and secondary nucleation. These processes all together make aggregation itself much more complex than descriptions using simple mass-action kinetics (Xue et al., 2008; Tsai et al., 2015). Recent advances in exploring the aggregation free energy landscapes of Abeta peptides have shown the complex paths of interconversion between different but structurally similar states of oligomers and have demonstrated the structural diversity for conformational conversion between pre-fibrillar to fibrillar oligomers (Zheng et al., 2016; Zheng et al., 2017). The detailed molecular interactions such as salt bridges, intercalation of water molecules, and hydrophobic clusters formed in different fibril polymorphic forms were found to significantly affect the capacity for cross-seeding as well as secondary nucleation (e.g., speeding up the aggregation of Abeta40 with Abeta42 fibril) (Xiao et al., 2015; Colvin et al., 2016). These results suggest the role of early stages of aggregation in modulating the chemical properties of the fibrillar surfaces—surface heterogeneity—that provide different specific or nonspecific chemical screening capacity. Surface heterogeneity, perhaps, plays a pivotal role in the later stages of aggregation that involves a variety of secondary processes.

To understand the molecular mechanism underlying surface-dependent aggregation events, in this work, we focus on the formation of monomeric precursors on the existing fibrillar surfaces as it can potentially initiate subsequent secondary processes. There are several possibilities when a protein monomer interacts with an existing fibrillar surface. During the interaction process, the monomer can either bind and stay at a particular surface patch, forming a new surface nucleation site or the monomer ends up localizing at both ends and elongates the fibril as a result. The former process, secondary nucleation, describes protein monomers being nucleated on the fibrillar surface through a two-dimensional search. This newly formed critical nucleus then plays a role of seeding new fibrils. The latter describes the elongation of the existing fibrillar structure which may take place at both fibril ends along the fibril axis. Although "elongation" and "secondary nucleation" are two seemingly distinct aggregation processes, it is now a popular view that they should be investigated with caution at the molecular level, as these two processes can be mutually correlated from a broader view of aggregation energy landscapes Cohen et al. (2018); Scheidt et al. (2019).

Elongation is the major process of aggregation when growing a fibril. There exist considerable experimental studies on the elongation process, in particular in measurements of the rate of fibril elongation (Xu et al., 2019). The experimental work offers an opportunity for theorists to construct models to understand molecular mechanisms of fibril growth. Some have proposed models concerning structural rearrangements and intermediates, while others address the molecular species from an energy perspective. Wei et al. first carried out atomistic molecular dynamics simulation to study thermodynamics and kinetics of fibril elongation of Abeta17–42. They used a kinetic network model to reveal detailed pathways for fibril elongation (Han and Schulten, 2014). Although elongation primarily concerns monomer addition at fibril ends, kinetic analyses have suggested multiple steps are involved, from solution free monomers all the way to final elongated fibril. Crespo et al. showed that elongation includes lateral migration of attached monomers towards the fibril ends and this process is not rate-limiting (Crespo et al., 2012). Since there exist several polymorphs of fibrils, some are disease-relevant, dissociation of Abeta monomers from such fibrillar structures have also been investigated in silico. For example, S-shaped fibrillar structure showed the stop-and-go mechanism at fibril ends due to the structural flexibility of the N-terminal monomer (Ilie and Caflisch, 2018). Simulation study over a three-fold protofibril from human tissue, however, supported the two-step dock-and-lock mechanism, where Abeta monomer interacts with fibril surface by direct docking onto it, and then, the docked peptide undergoes conformational arrangements on the surface in order to fit the fibril template over the ends for elongation (Sasmal et al., 2016). According to their result, docking is faster than the locking process by about an order or so, depending on the type of monomer ensemble. In contrast to the stop-and-go kinetics, recent experimental work showed

relatively steady Abeta40 fibril growth and dissolution rates (Xu et al., 2019). Structural and dynamics difference between the two ends (even/odd) of amyloid fibril adds more kinetic complexity for the understanding of the mechanism of fibril growth. The even end grows faster as has been shown in experiment (Konno et al., 2020), independently verified in another simulation work (Okumura and Itoh, 2016). All these results raise an issue about the pathway-dependent nature for elongation, in particular, due to their different mechanistic details. For example, there might be multiple pathways of elongation channels that can potentially contribute to fibril growth. In this regard, the most probable path has been discussed (Rodriguez et al., 2018). From a kinetic perspective, indeed, multiple pathways would lead to a variation in predicting kinetic properties. Regarding the binding thermodynamics, several approaches are available in the calculation of the binding affinities, for example, alchemical free energy perturbation method or potential-of-mean force (PMF) approach (Deng and Roux, 2009). Depending on the simulation type, different approaches in general yield at least qualitative agreement across different all-atom and coarse-grained force fields.

In contrast to fibril elongation, secondary nucleation describes a surface-catalyzed nucleation process where new fibril seeds emerge. Recently, an experimental study that combines theoretical analysis showed that secondary nucleation and elongation occur at different sites, suggesting a potential dynamic interplay for a protein monomer searching over fibrillar surfaces (Scheidt et al., 2019). In the propagation of amyloid fibrils, researchers have shown the role of hydrophobic patches in growing fibrils *via* secondary nucleation (Thacker et al., 2020). All these studies have clearly pointed out the importance of surface heterogeneity for different secondary processes.

There are many existing studies that focus on amyloid aggregation, either from a nucleated-polymerization perspective or a templated fibril growth perspective. However, little is known about how fibril polymorphic surfaces affect secondary processes. Specifically, how existing fibril seeds catalyze the process of fibril growth by recruiting new monomers onto the fibrillar surfaces. As we have mentioned above, the mechanistic details of aggregation, in particular the kinetically relevant events, are significantly limited by the experimental means and physical/chemical parameters we used to probe them. Theoretical models and simulation techniques, in this regard, are very useful for probing complex processes and therefore allow us to explore parameter space that is difficult to achieve *via* experimental methods alone.

Fully atomistic modeling has shown great promise in tackling many important problems in protein biophysics. This technique is particularly useful for exploring full dynamics of Abeta monomers to necessary atomistic details, and is an ideal approach for exploring conformational ensembles for various calibration purposes (Grazioli et al., 2019). However, modern pressing biological problems involve molecular assemblies having thousands of amino acid residues and functional dynamic motions taking place in timescales more than milliseconds, seconds and beyond. As a result, exploring the biologically relevant timescales using fully atomistic simulation makes such realization a daunting task. Coarse-graining therefore becomes a conceptual prerequisite for addressing the major problems of modern biology. Another important motivation for developing coarse-grained modeling is that many large-scale protein motions concern emergent properties caused by collective organizing principles (Laughlin et al., 2000), in which details of intra/intermolecular forces are averaged out. Folding, binding, and functional transition in proteins are examples of emergent phenomena that can be fully understood at an appropriately coarser resolution. In this study, we use one such coarse-grained model, Associative-memory, Water-mediated, Structure and Energy Model (AWSEM) (Davtyan et al., 2012; Tsai et al., 2016b), to study the process of Abeta peptides interacting with the surfaces of a protofilament. AWSEM is a transferable, coarse-grained, and non-additive protein force field that incorporates physically motivated energy terms and knowledge-based information using the principle of minimal frustration (Ferreiro et al., 2018). AWSEM has been proven useful for exploring many of the important biological processes, such as folding, binding (Tsai et al., 2016b; Zheng et al., 2012), aggregation (Zheng et al., 2016, 2017; Chen et al., 2016), protein-DNA interaction (Tsai et al., 2016a; Tsai et al., 2019; Potoyan et al., 2016a; Potoyan et al., 2016b), and chromosome remodeling (Zhang et al., 2016), and has continued to be a suitable coarse-grained model for studying the aggregation system of interest here. Here, we employ a GPU-enabled AWSEM code, openAWSEM (Lu et al., 2020). This new version includes a recent advance in GPU acceleration built on the openMM platform (Eastman et al., 2017).

In this work, we explore the experimentally determined fibrillar structure of Abeta11-42 (ssNMR) using openAWSEM. We first study the stability of the fibrillar surface structure using the AWSEM coarse-grained force field and confirm the structural integrity of the S-shaped polymorphic fibrillar structure. We then investigate fibrillar surface heterogeneity by exploring the binding free energy landscapes of a free Abeta monomer to a short fibrillar surface. For a broader surface sampling purpose, we choose to use the biasing coordinate that allows efficient sampling over the surface with a simple distance restraint. This biasing strategy is somewhat different from the conventional approach, where the fibril ends usually are chosen to bias with (Han and Schulten, 2014) or some positional restraints at the filament tips are applied in order to prevent twisting motions (Schwierz et al., 2016). These approaches are important for obtaining structurally stable binding sites for elongation. We nevertheless explore the PMF along the coordinates that allow more sampling over amyloid structural precursors that can lead to different aggregation pathways. To ensure this, different initial spatial orientations are used. The biasing coordinates are determined by a C-alpha in the free monomer and a C-alpha of a fibrillar monomer in the fibril (chain in the middle of the fibril, not the fibril ends). We aim to explore the configurations of monomeric amyloid precursors that precede elongation and other secondary processes. Our overall hypothesis is that secondary processes, such as elongation and secondary nucleation, share similar monomeric amyloid precursors that drive different aggregation pathways. We

have identified several structural precursors and potential surface binding sites accordingly. To understand the intrinsic energetics of the local residues forming binding pockets, we have carried out energy frustration analysis for a series of fibril polymorphic structures and have predicted the preferred surface binding sites. The simulation protocol used allows us to characterize some key steps in the process of aggregation and allows efficient sampling for binding sites that are specific to secondary processes. Finally, we model the fibrillar twist polymorphism of a protofilament with different filament sizes. We find a filament size dependent effect on filament twisting, which potentially can modulate the surface heterogeneity of a fibril.

Understanding the molecular mechanism of elongation and secondary nucleation can help predict how mutations and external factors affect fibril growth and how antibody drugs intervene in the processes of elongation and secondary nucleation. It allows us to predict the effect of drugging in promotion or inhibition of fibril growth. Our results support that secondary nucleation and elongation occur at different binding sites, confirming their independent inhibitory effects by molecular chaperones follow a different pathway (Scheidt et al., 2019). We anticipate that our study will be useful for developing rational design principles for new therapeutic drugs.

2 METHODS

2.1 Models and Simulation

2.1.1 Modeling of Fibrillar Structure

Human amyloid-beta (Abeta) peptide (pdbID 2MXU [Xiao et al., 2015)] was used to model the aggregation system in the present study. The solid-state NMR structure features a S-shaped, three-layer fibrillar architecture with 12 Abeta peptides (11–42, length = 32), in-register and parallel aligned, labeled as chain A, B, . . . , L, respectively. In simulation, we prepared one such 12-monomer fibril structure and one additional free monomer having its initial structure as it is in the fibril structure. For the central fibril, "single fragment memory" was used in order to strongly bias the aggregation energy landscape towards the native fibril structure. For the additional free monomer, we adopted two different monomer structural ensembles: 1. Relaxed ensemble (no biasing fragment memory is used) 2. Fibril-like ensemble (single fragment memory as is in the fibril structure). The descriptions about the "fragment memory" library can be found elsewhere (Davtyan et al., 2012; Tsai et al., 2016b). The free monomer is initially positioned in six different orientations with respect to the central fibril: up, down, left, right, front, and back. These six independent simulations ensure a better sampling quality while leveraging the availability of our computation resources. The system is built and visualized using VMD (Humphrey et al., 1996).

2.1.2 Molecular Dynamics Simulation Using OpenAWSEM

In this study, we use openAWSEM, a python version of the AWSEM protein coarse-grained force field developed by Wolynes and his coworkers Lu et al. (2020). This new simulation platform is built on openMM (Eastman et al., 2017) for a fast (GPU-enabled), flexible, easy-to-use purpose. OpenAWSEM inherits from the lineage of the Associative-memory, Water-mediated, Structure and Energy Model (AWSEM), for molecular dynamics (MD) simulation (Davtyan et al., 2012). In AWSEM, each amino acid residue is represented by three atoms: C_α, C_β, and O (glycine is an exception). The physicochemical properties of different types of side chains are reflected on C_β atoms. The AWSEM-MD simulation protocol has been used to address a variety of biological questions, such as protein structure prediction (Davtyan et al., 2012; Tsai et al., 2016b; Sirovetz et al., 2017; Chen et al., 2018a), protein binding prediction (Tsai et al., 2016b), protein aggregation (Zheng et al., 2016, 2017), as well as complex protein-DNA assemblies and remodeling (Tsai et al., 2016a, 2019; Potoyan et al., 2016b; Zhang et al., 2016). Interested readers are encouraged to test the online web-server version, AWSEM-Suite for structure prediction (Jin et al., 2020). Because of openMM's extensibility in python scripting, openAWSEM benefits from such flexibility for interfacing with other post-processed analysis and visualization toolkits, such as MDTraj, pyEMMA and NGLviewer.

For the fibril stability test, three independent simulations were carried out, with each simulation trajectory running for 10 million simulation time steps (=10,000 frames).

2.1.3 Importance Sampling and WHAM

Free energy calculation, more precisely potential of mean force (PMF), is carried out using the pyEMMA python package, developed by Noe and his coworkers (Scherer et al., 2015). In the calculation of PMFs, one needs to choose a specific progress coordinate of interest to sample along. Because high energy configuration space is not easily accessed through thermal activation, a biasing force is required in order to "bias" the sampling route towards the configurational space of interest-a procedure termed "importance sampling". In practical use, all the sampling tasks were carried out by running molecular dynamics simulation on the openAWSEM platform (Lu et al., 2020). To sample along the route of a free monomer diffusing towards a designated position on the surface of a fibril, a large number of sampling windows were prepared. Different sampling windows were deployed by applying a series of harmonic biasing restraints between two atoms (one from the free monomer and the other from the fibril); they are centered at a distance in a range of 10–100 Å (with an interval of 1 Å). A total of 91 independent simulations were generated as a result. The biasing coordinate is defined as the distance between the C-beta atom of residue N27 of chain F in the fibril and the C-beta atom of residue N27 of the free monomer. The biasing force constant k is set to be 2.4 kcal/mol (10 kJ/mol) for all the 91 simulations (sampling windows) using the harmonic biasing form $k(r - r_0)^2$, with r_0 = 10, 11, 12, . . . 100 Å. Each simulation is run for 5 million time steps; 5,000 frames were outputted for analysis purposes (output frequency is every 1,000 time steps). After collecting data from the simulation trajectories (from a total of 91 sampling windows), the data were reweighted using the WHAM technique (Kumar et al., 1992) implemented in pyEMMA package (thermo.wham) (Scherer

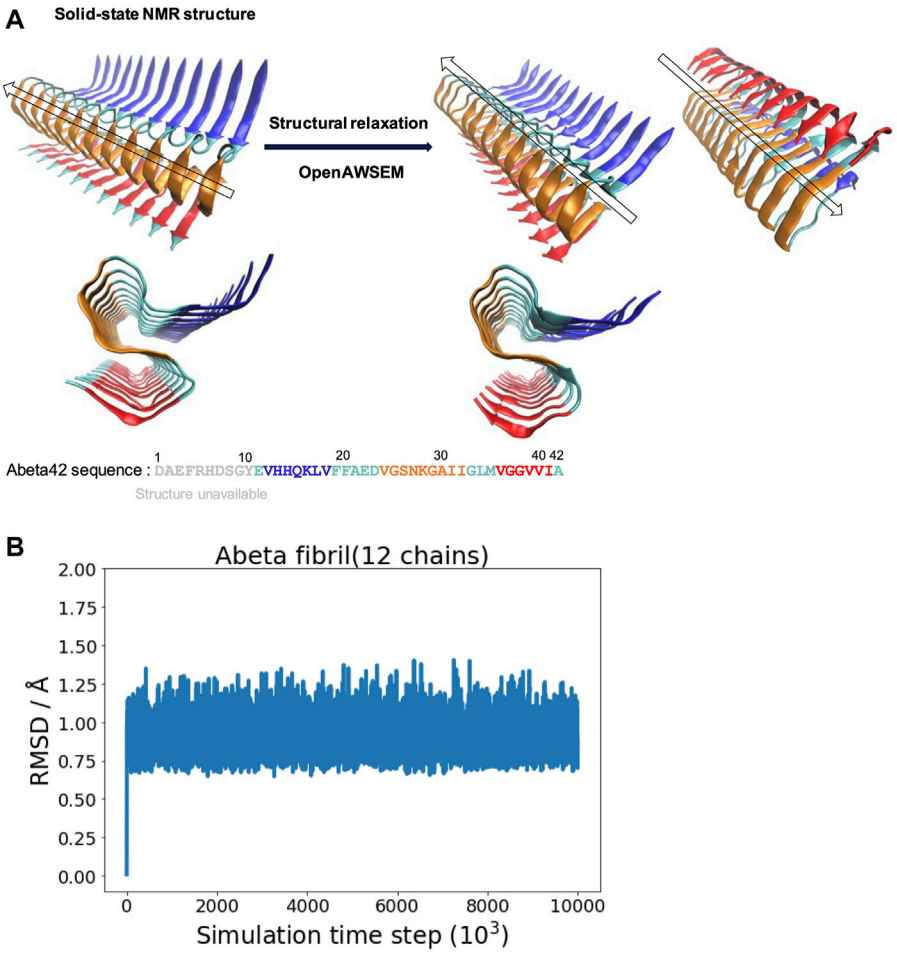


FIGURE 1 | The overall structure of the experimentally determined fibrillar structure and the simulated fibrillar structure are compared. **(A) Left:** The fibrillar structure of Abeta42 is determined by solid-state NMR [PDB ID: 2MXU (Xiao et al., 2015)]. The actual sequence length of the individual peptides in the structure is 32 (spanning from 11 to 42). The "S" shaped triple parallel-beta-sheet architecture can be seen from the side view below. **Right:** The simulated fibrillar structure. The experimental structure is structurally relaxed *via* molecular dynamics simulation using openAWSEM. Two different orientations of the relaxed structure are shown, with an arrow showing the direction of the fibril axis. The "S" shaped triple parallel-beta-sheet remains in the simulation. Different colors represent different local structural features of the Abeta42 peptide, defined in the original PDB file, illustrated using the sequence below. Blue strand (V12-V18), Cyan loops (E11,F19-D23,G33-M35,A42), Orange strand (V24-I32) and Red strand (V36-I41). **(B)** The RMSD is calculated as a function of simulation time step (only one of the three independent simulation trajectories is shown).

et al., 2015), discarding the first 500 thousands time steps (500 frames) for equilibration. The distribution of all the sampling windows is shown in the Supporting Information (**Supplementary Figures S1, S5**).

2.2 Analyses

2.2.1. Thermodynamic Binding Affinity

In the calculation of binding affinity, we use the formula that is based on one-dimensional radial potential of mean force (PMF) given in the literature (Deng and Roux, 2009). One simple equation is shown below.

$$K_d = 4\pi \int_{site} r^2 e^{-\beta(w(r)-w(r^*))} dr \tag{1}$$

where w refers to the PMF as a function of the distance r; r^* refers to a reference position set to be far away in the bulk ($w(r^*) = 0$ at $r^* = 80$ Å); $\beta = 1/k_BT$. In practical use, to enhance the fluctuations of the orientation of the Abeta monomer (ligand) with respect to the central fibril, we adopted six independent simulations with each of them representing a different initial orientation (front, back, up, down, even, odd). See the subplot in **Figure 1A**. The thermodynamic binding affinity thus can be calculated

$$\Delta G^\circ = -RT \ln(c_0 K_b), \tag{2}$$

where $c_0 = 1M$ refers to the standard state (1 mol/L = (1,660 Å^3)$^{-1}$). See **Table.1** for the calculated value as well as the experimental values obtained from literature.

TABLE 1 | Thermodynamic binding affinity of Abeta binding to a fibril.

ΔG_b (kcal/mol)	Topology	Length	Solubility (μM)	Condition	Type
−8.7	Twofold	1–40	0.44	<10 μM (27°C)	Exp. Xu et al. (2019)
−8.7	Twofold	1–40	0.3–0.4	<75 μM (24°C)	Exp. Qiang et al. (2013)
−9	-	1–40	0.8–1.0	<30 μM (37°C)	Exp. O'Nuallain et al. (2005)
−12 (even) −11.3 (odd)	Twofold	17–42	-	37°C	Sim. Han and Schulten, (2014)
−15.6	Twofold	9–40	-	37°C	Sim. Schwierz et al. (2017)
−20.7	Single	11–42	-	27°C	This work

2.2.2 Probability Contact Maps

The detailed monomer-fibril interactions are quantified using the Contact Map Explorer python module, which is based on tools implemented in MDTraj (McGibbon et al., 2015). The formation of a contact is defined by setting a cutoff distance between any of two C-alpha atoms in the residues. A contact forms if the distance is smaller/equal to the cutoff value; the cutoff is set to be 6.5 Å in the present study. The step-by-step tutorial (Jupyter notebook) is provided in the Github.

2.2.3 Nematic/Polar Order Parameter

To characterize the structural difference of fibril polymorphism, we have used the nematic order parameter (P_2 value) as the structural orderness of fibrillar surface structure. Originally designed for describing the structural order of liquid crystals, nematic order parameter was first introduced to study protein aggregation by Caflisch and co-workers (Cecchini et al., 2004). This order parameter was further applied to describe elongation of Amyloid-beta fibrils (Schwierz et al., 2017) and recently being compared with neural network learned order parameters Charest et al. (2020).

2.2.4 Probing Fibrillar Twisting

The twist angle is calculated using the protocol described in the literature (Ilie and Caflisch, 2018). The angle, θ_i, is defined by the two vectors: one represents the vector of the reference chain i and the other denotes the vector of the neighboring chain $i + 1$, respectively (see **Figure 6A**). Similarly, θ_{i+1}, θ_{i+2} and so forth can also be calculated by propagating the current chain pair to the next neighboring chain pairs. The vector is defined by the two C-alpha atoms of Q15 and F19 of individual chains as indicated in the reference (Ilie and Caflisch, 2018). To avoid the effect of simulated fibril structural fluctuation on angle calculation, we defined a fibril axis vector and a normal plane perpendicular to this vector for angle correction purpose. We report two different angular properties in order to quantify the twist morphology of a fibril. 1. The averaged θ ($\bar{\theta}$, the twist angle is averaged over all chain pairs of the same fibrillar structure) 2. The accumulated total twist angle, θ_{tot} (the twist angles from all of the chain pairs are added up). θ_{tot} represents the extent of global twisting morphology of the fibril.

In simulating fibrillar helical twisting, five protofilament models of different sizes (12, 24, 36, 48, 62 chains) were prepared. Long protofilament models (24–62 chains) were made using the 12-chain model (pdbID 2MXU). To ensure the strands at the joint boundary are properly connected between the 12-chain model fragments, we also, if necessary, constrain the resulting elongated protofilament by applying a mild harmonic biasing force along the fibril long axis. The harmonic biasing form is $\frac{1}{2}k'(r - r_0)^2$.The magnitude of the force constant/center distance pairs (k', r_0) were set to be (6 kcal/mol, 10 Å). The center of mass of the first two chains and the center of mass of the last two chains of the model fragment were the constrained objects to which the biasing force is applied. As a result, we carried out 1 million simulation time steps for all the five protofilament models for a pre-equilibration/relaxation purpose. The relaxed structure of each protofilament model along with the structure's θ angle distribution can be found in the Supporting Information (**Supplementary Figure S10**). After that, 10 million simulation time steps were performed and the data (each with 10,000 frames) were collected for fibrillar twisting analysis.

2.2.5 Frustration Profiles for Polymorphic Fibrillar Surfaces

According to the energy landscape theory of protein folding, the evolutionarily conserved protein structure is energetically minimally frustrated while protein functional activities emerge through frustration (Ferreiro et al., 2018). The corresponding energy landscape for robust folding is manifested as a funneled shape. However, a recent study has pointed out functional roles of energetically frustrating areas in binding protein-DNA partners (Tsai et al., 2016b,a; Marcovitz and Levy, 2013; Potoyan et al., 2017), forming assemblies, and ligand binding (Chen et al., 2020). Energy frustration of proteins involves the statistical energy survey over a series of decoy states, which can be generated through pairwise residue substitution, direct mutation, and position shifting. Different decoy settings correspond with different physical contexts. The frustration is defined using the standard scores (z-scores) in statistics. Three different scales are classified accordingly: Frustrated (<−1), neutral (>−1 and <0.78), and minimally frustrated (>0.78). Interested readers should refer to the reference provided for details (Parra et al., 2016).

The frustration calculation for a series of amyloid fibrillar surfaces was conducted using the frustratometer server Parra et al. (2016), http://frustratometer.qb.fcen.uba.ar/. This frustration computation protocol utilized the same AWSEM-MD energy function (sequence separation is set to be 3) along with the electrostatics enhanced feature (optional) to compute the frustration profile. The frustration profiles for a variety of polymorphic fibrillar structures are shown in **Figure 5**. The PDBIDs used in the analysis include 5KK3 5OQV 2MXU 2M4J 2LMQ 2LMN. The results are summarized in **Table 2**. The protein residues in purple indicate they are conserved

TABLE 2 | Polymorphic properties of aggregation.

	Polymor.1	Polymor.2	Polymor.3	Polymor.4	Polymor.5	Polymor.6	This work
PdbID	5KK3	5OQV	2MXU	2M4J	2LMQ	2LMN	2MXU
# of patches	4	6	4	6	4	4	2
Frustrated residues	E11 H14,Q15 K16 E22,D23 N27,K28	D1,E3 E11 Q15,K16 E22,D23 N27,K28	E11 K16 E22,D23 N27,K28	D1,E3 E11 K16 E22,D23 K28	Y10,E11 K16 E22,D23 G25,S26 N27,K28	Y10,E11 K16 E22,D23 S26,N27 K28	E22, D23 N27, K28
P_2 value	0.95	0.93	0.99	0.25	0.26	0.93	0.99
Method	ssNMR	cryoEM	ssNMR	ssNMR	ssNMR	ssNMR	Sim

residues and therefore energetically minimally frustrated, while the residues in red represent highly frustrated residues. This means they may have functional significance in interacting with its partner, such as protein, DNA/RNA, or even membranes.

3 RESULTS AND DISCUSSION

3.1 The Same "S" Shaped, Triple Parallel-Beta-Sheet Architecture Remains in the Simulation

Starting with the experimentally determined fibrillar structure, we examine the stability of the fibrillar structure in the simulation using the AWSEM force field. We find that the overall fibrillar structure along with its surface architecture is well maintained in the simulation. **Figure 1A** shows the structural refinement of Abeta42 fibrillar structure. The NMR structure exhibits a S-shaped triple parallel-beta-sheet architecture. This fibrillar architecture shows a common cross-beta structure, which contains specific cooperative residue interactions along the 2D plane that is perpendicular to the fibril axis. **Figure 1B** shows one of the three RMSD trajectories (based on the Root Mean Square Deviation of C_α atoms with respect to the reference structure of 2MXU); the rest of two are not shown. The RMSD value is saturated at an averaged value of 0.90 Å; the average is taken over the last 1,000 frames. The resulting polymorph has been described as "ribbon folding" by Wolynes and his coworkers (Chen et al., 2018b). Because of the cooperative coupling among different fibril dimensions, a variety of fibrillar polymorphs become possible when different pairs of residues are preferred using different force fields. According to the ribbon-folding landscape schemes, different fibrillar architectures can be seen at the energy local minimum along the polymorph energy landscape (Chen et al., 2018b). In this study, we show that openAWSEM is suitable for exploring the ideal cross-beta fibrillar architecture. The cross-beta architecture along with its fibrillar surface later will be used to study surface heterogeneity and helical twisting of a protofilament. Here, we show that the openAWSEM-refined fibrillar structure retains the same S-shaped, triple parallel-beta-sheet architecture observed from the NMR structure.

3.2 A Monomer Binding to Fibrillar Surfaces can be Characterized at Least by Three Different Stages: Free Diffusion, Downhill Guiding, and Dock and Lock

We explore the free energy landscape along a distance separation between a free Abeta monomer and a fibril surface. To enhance sampling over different spatial orientations, we carry out several independent simulations with different initial positions of the monomer with respect to the central fibril. A total of six different positions were chosen to address the fluctuations of orientation. The six simulations, having the monomer being put in different orientations: front, back, up, down, even, and odd, respectively, were performed (see the subplot in **Figure 2A** for a schematic description). **Figure 2A** presents a representative free energy profile with the monomer being positioned in the "front" position. From the free energy profile, several features can be observed. They are classified into three different stages accordingly: I. Free diffusion. II. Downhill guiding. III. Dock and lock. When the Abeta monomer is far from the central fibril ($r > 80$ Å), the dynamics of the free monomer is primarily diffusive and that the free energy profile is nearly flat in the plateau (Stage I). As the distance between the fibril and the monomer decreases, the monomer is subject to a long-range guiding force due to electrostatics, and therefore, the monomer begins to approach the fibril. This long-range guidance yields an energetically downhill profile (Stage II). The downhill free energy continues until its slope significantly changes at $r \approx 35$ Å where the free energy profile displays a curvature. After that, the monomer begins to have physical contacts with the fibril (Stage III). In stage III, there are many ways for the monomer to dock the fibril. The biasing strategy used allows spatially orientational flexibility for the monomer to dock the fibril. As a result, the monomer is able to dock the fibrillar surface through different sites. All the resulting binding configurations lead to a clear free energy basin at $r \approx 33$ Å. **Figures 2B–F** show example configurations of several key binding configurations whose population is significant and that their interaction pattern is well characterized. The result shows that the monomer can interact with the fibril's C-terminal surface (red), N-terminal surface (blue), cleft interface, even-end, and odd-end. We will look into their structural features more carefully in the next section. The rest of the free energy profiles, with the monomer

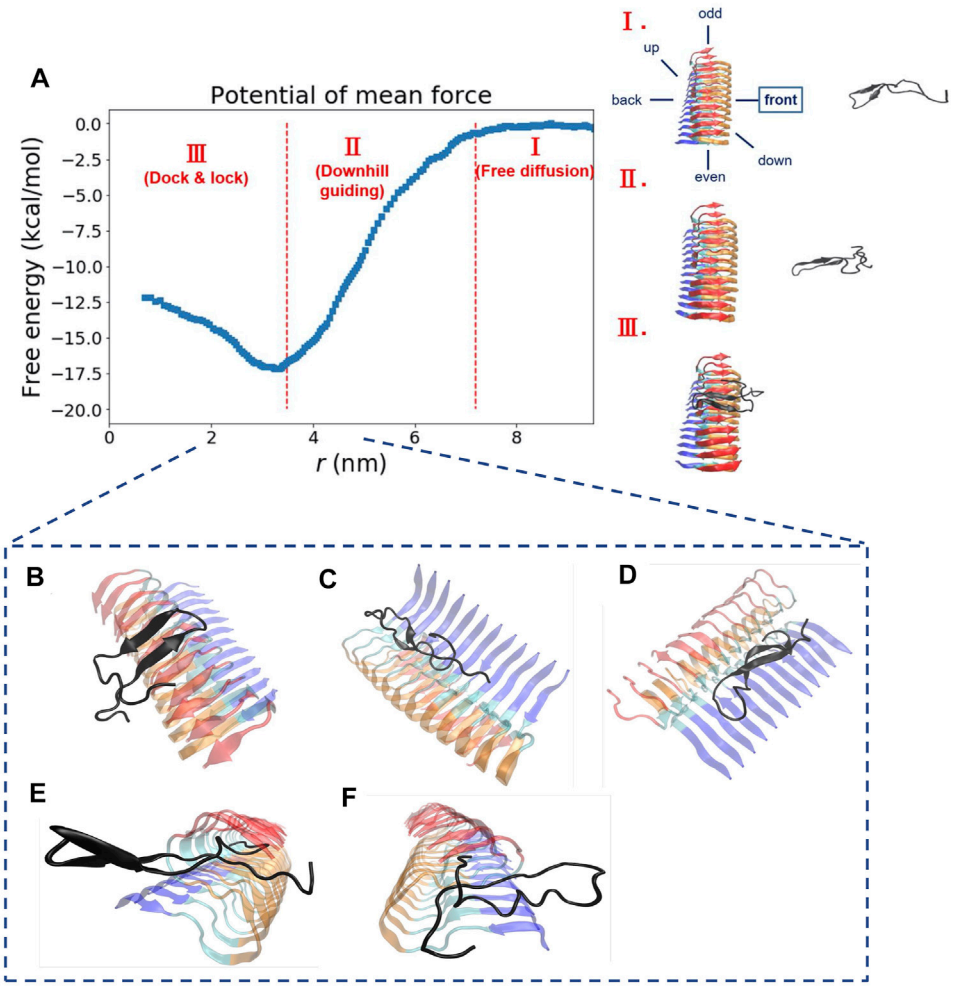


FIGURE 2 | The free energy profile for a single Abeta11-42 monomer binding to the Abeta42 fibrillar surface (12 chains) is shown. **(A)** The free energy profile features three different aggregation stages, labeled as I. Free diffusion. II. Downhill guiding. III. Dock and lock. *r* is defined as the distance between the C-beta of residue 27th in the free monomer and the C-beta of residue 27th of chain F in the fibril. A representative configuration at each aggregation stage is schematically shown on the right. Simulations were prepared with six different monomer positions with respect to the central pre-existing fibril. Six orientations: front, back, up, down, even, odd are schematically shown in the diagram. The free energy profile shown refers to the result obtained from the simulation setup with the monomer positioned in the "front" orientation. The free energy profiles for the rest of the orientations are shown in the Supporting Information (**Supplementary Figure S2**). (B–F) The representative structures taken from stage III in which different binding configurations are formed upon the monomer landing and searching over the fibrillar surface. These binding configurations are potentially surface-catalyzed precursors for fibril growth. **(B)** C-ter surface precursor **(C)** N-ter surface precursor **(D)** Cleft-gate precursor **(E)** Even-end precursor **(F)** Odd-end precursor.

initially positioned in a different orientation, can be found in the Supporting Information (**Supplementary Figure S2**).

Next, we look into the thermodynamic binding affinity, defined by the potential of mean force (PMF). In determining the free energy of binding, multiple free energy calculations have shown variation in r_b (r_b refers to the distance at which the global free energy basin is found), suggesting that the Abeta monomer binds to the fibril surface through a pathway-dependent manner. This pathway dependence very likely causes some variations in the binding free energy profiles since the monomer might interact with the fibril surfaces through different "dock" sites. Here, we do not assume any specific binding site *a priori* for the monomer to bind with. Instead, we aim to sample different binding trajectories and then combine these trajectories to determine the standard binding affinity (with c_0 = 1 M, see **2.2.1**). The value is computed to be −20.7 ± 2.8 kcal/mol if we use the data of all the six orientations to ensure the orientational fluctuations. The binding affinity, determined by the simulation trajectory of individual single orientation, ranges from −17 to −23 kcal/mol. This energy variation is due to a different PMF obtained from individual orientation (see Supporting Information for details). The experimental values of −8.7 kcal/mol (Xu et al., 2019; Qiang et al., 2013) and −9 kcal/mol (O'Nuallain et al., 2005). have been exclusively reported for the process of fibril elongation. Their corresponding binding affinity was also calculated using computational methods, which are summarized in **Table 1** as well. Although the reported values for binding affinity are rather diverse, these values are within the same order of magnitude.

Fibrillar surface heterogeneity, presumably, plays an important role in the process of monomer binding. As a result, finding a proper reaction coordinate is a non-trivial task. In other words, monomer binding may undergo different pathways; the overall process can be under a kinetic control. For example, when the monomer binds to a specific surface site, the monomer may undergo a conformational conversion, searching for the right conformation or the position for subsequent secondary events. Indeed, several surface-dependent aggregation mechanisms have been discussed, such as conformational rearrangement (Xu et al., 2019), lateral migration (Crespo et al., 2012), or other surface-based events. These all together may have significant influence on the thermodynamic interpretation of binding and thus determine the kinetics of fibril growth. A similar multi-pathway issue using a complex collective variable for describing the loop interaction in adenine riboswitch has been discussed in the literature (Di Palma et al., 2015). It is important to know if the progress coordinate of interest is sufficient to drive the system through the appropriate transition states. Here, we recall the importance of the biasing protocol used for the interpretation of results.

One interesting result is worth noting. Free energy calculation using a different monomeric structural ensemble (fibril-like) shows a somewhat similar free energy profile (with the same three stages as described above) but now with a rather different pattern of contacts (see next section for further discussion). This result can be attributed to the structural rigidity of the specific fibril-like conformation used for the monomer. This finding suggests that the specificity for binding energy are encoded in the sequence, irrespective to the monomer conformation adopted.

3.3 Surface Binding Heterogeneity: Several Binding Sites are Identified Over the Fibrillar Surfaces

From our simulations, we have identified several Abeta binding configurations that potentially can be structural precursors for subsequent surface-dependent processes, e.g., fibril elongation, secondary nucleation. These structural precursors are named with the preferable binding region along the fibrillar surfaces to which the single monomer binds. These binding regions include sites located on the C-terminal (C-ter) surface, the N-terminal (N-ter) surface, the cleft interface, and the two fibril ends (even and odd). Here we would like to characterize their structural features and quantify their contacts with the fibrillar surfaces. The structure of the monomer on the fibrillar surfaces shows primarily a beta-hairpin conformation with their strand vector either parallel (N-ter and cleft-gate precursors) or orthogonal (C-ter, even-end, and odd-end precursors) to the fibril long axis. They are binding configurations of the free energy basin (area III, see **Figure 2A**). **Figure 3** shows the probability contact maps for these structural precursors. The C-ter surface consists of an alignment of hydrophobic sequence segments (36-VGGVVIA-42) along the direction of the fibril axis. Abeta monomer interacts with the C-ter surface primarily through the same VGGVVIA hydrophobic sequence motif of its C-terminus and thus facilitating C-ter/C-ter hydrophobic clustering. In contrast to the C-terminus, the N-ter surface shows a different pattern of contacts. The binding region involves some charged residues in the sequence (11-EVHHQKLVFFAEDVGS-26) along the fibrillar surface. The contact pattern therefore is more diverse, with additional charge-charge interactions (D, E, H and K) that participate in the stabilization of monomer binding. E22 and D23 are particularly important since these two residues are also predicted to be the most highly frustrated areas in the frustration analysis (see below). **Figures 3A,B** show the C-ter surface contact map and the N-ter surface contact map, respectively.

A somewhat unconventional binding site is identified on the other side of the N-ter surface. Because of its gate-like shape with a cleft at the interface, we name such a monomer binding configuration "cleft-gate" precursor (see **Figure 3C**). The pattern of the contact map of the cleft-gate precursor looks very similar to the N-ter surface precursor (both display a "S" shape), except that the overall profile is shifted towards the even end. One signature of the cleft-gate precursor is that the contacts formed over the individual chains of the fibril is long-range along the sequence. Therefore, the micro profile of individual chains is different from that of the N-ter surface precursor (see **Supplementary Figure S3** in the Supporting Information). **Figures 3D,E** show the probability contact maps of the even-end and odd-end precursors, respectively. In both cases, the monomer moves to a fibril end and localize onto it. These binding configurations presumably correspond to a fibril being elongating. However, we do not observe the monomer conformation being in the fibril form throughout the simulation. It is very likely that the monomer structural ensemble also plays a role in the process of elongation.

3.4 An Elongating Fibril Requires the Added Monomers Being in the "Activated" Conformation

On the other hand, we have also carried out similar probability contact map analysis for simulations using a fibril-like monomer ensemble ("activated" conformation). We find that the structural precursors identified share similar binding sites with those found in the case of relaxed monomer ensemble. This finding implies that some interactions between peptide and fibril on the surface are well conserved, e.g., C-ter/N-ter hydrophobic patches. The overall contact maps of the fibril-like and the relaxed monomer ensembles, however, exhibit quite different features due to the conformational dynamics and rigidity that the monomer intrinsically has. **Figure 4** shows the structure of the monomer being elongating on both fibril ends. The simulated structures were taken from the simulation trajectories with the monomer being in the fibril-like conformation. These elongated species, either formed at the even-end (**Figure 4A**) or the odd-end (**Figure 4B**), are not observed in the simulations using a structurally unbiased monomer. This result suggests that the monomer conformation being "activated" (conforms to the same shape as in the fibril) plays a determining role in the elongation process while monomer in non-activated form does not significantly contribute to fibril elongation. This result also echoes the two-step dock-and-lock mechanism where the second locking step involves an "activated monomer" that irreversibly

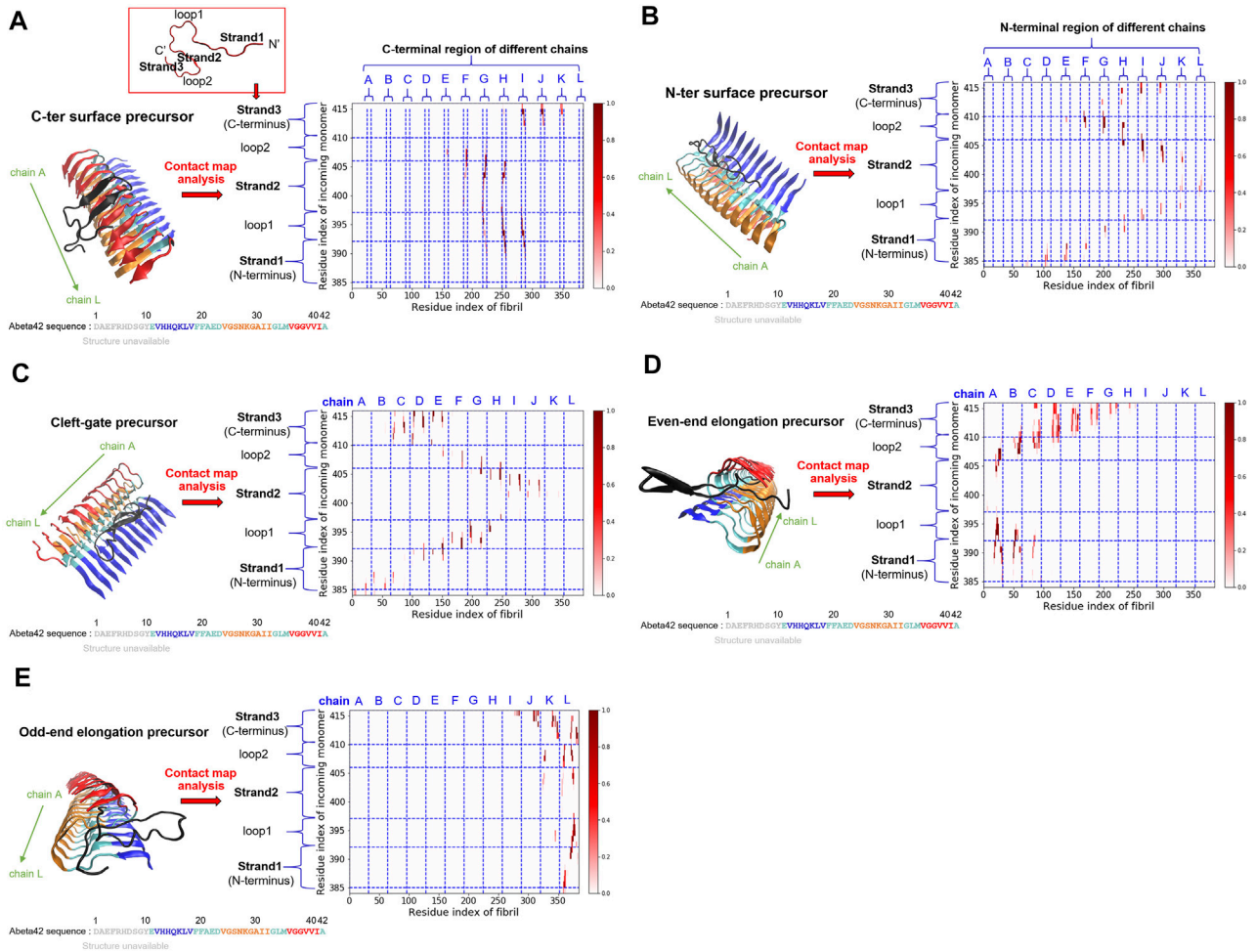


FIGURE 3 | Five key structural precursors on the fibrillar surface and their probability contact maps are shown. **(A)** C-ter surface precursor **(B)** N-ter surface precursor **(C)** Cleft-gate precursor **(D)** Even-end elongation precursor **(E)** Odd-end elongation precursor. The probability contact map next to each structural precursor presents the contacts formed between the monomer and the fibril. The colorbar, scaled by probability, is shown on the right. The horizontal axis uses the fibril index (1–384; 32 × 12 = 384), which sequentially renumbers the 12 monomers in the fibril. The vertical axis describes the residue index of the free monomer by adding up the existing fibril index, 385–416 (384 + 32 = 416). Different structural features are labeled as "strand" or "loop". Note that in **(A)**, a schematic monomer structure is shown in a red box, with its structure in the fibrillar form. Three strands (strand 1, 2, and 3) and two loops (loop 1 and 2) are indicated. The color scheme for the structure is the same as that used in **Figure 1**.

binds to the fibril end and elongate (Sasmal et al., 2016). The contact maps for the rest of surface structural precursors can be found in the Supporting Information (**Supplementary Figure S7**). **Supplementary Figures. S4–S7** show the results for the fibril-like monomer binding to the fibrillar surface.

Our simulation study shows that the conformational ensemble of single Abeta plays a key role in determining the kinetic pathways of elongation. The structural rearrangement of the monomer from the "dock" state into the "lock" state involves activated fibril-like conformation that irreversibly binds to the fibril end. This result suggests that the elongation free energy landscape in general can be reduced into a few dimensions: 1. The dimension of the monomer that reversibly searches the landing site over the fibril surface (dock). 2. The dimension of the conformational ensemble of the single monomer on the surface (lock). Once the monomer is in the activated form, the "docking" state merges into the "locking" state. This conversion irreversibly leads to a one-step fibril elongation.

3.5 Fibril Surface Binding Site Prediction Using Frustration Analysis

We have shown fibril surface heterogeneity of Abeta protofibril by identifying several binding interfaces. To further our understanding of those predicted sites, we carry out a series of frustration analyses over fibrils of different polymorphs. We aim to compare their results with the results from our MD simulation and provide insight into the predicted sites from an energy perspective. Frustration analysis uses the AWSEM energy functions to access the extent of frustration in the spatially localized interactions in proteins at a residue level (Ferreiro et al., 2007; Parra et al., 2016). If a residue and its neighboring residues are predicted to be highly

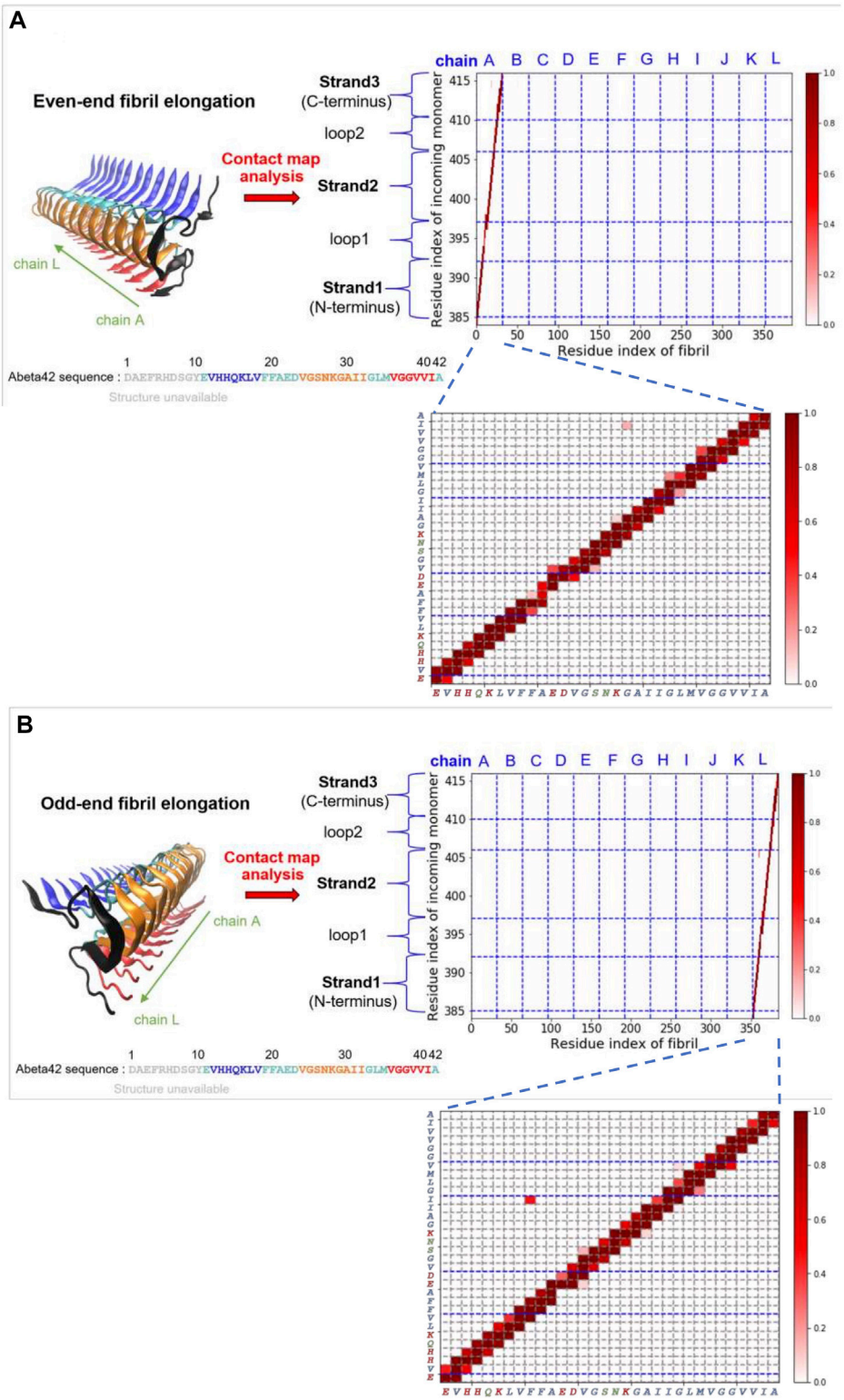


FIGURE 4 | Fibril elongation with the monomer being in the activated fibril-like conformation and their contact maps are shown. **(A)** Even-end fibril elongation. **(B)** Odd-end fibril elongation.

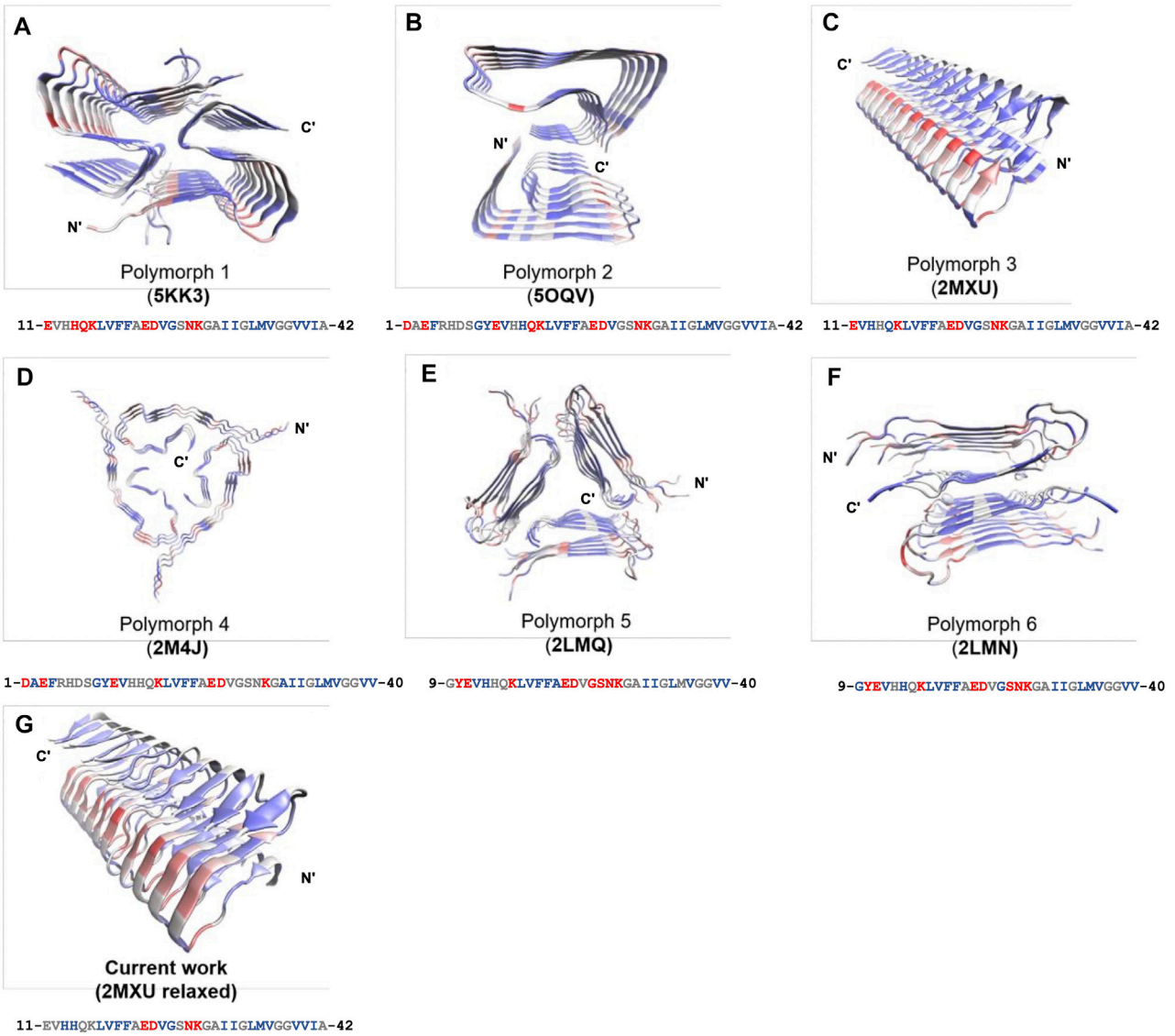


FIGURE 5 | Frustration profiles for a variety of polymorphic fibrillar structures of Abeta peptides available to date. **(A)** 5KK3. **(B)** 5OQV. **(C)** 2MXU. **(D)** 2M4J. **(E)** 2LMQ. **(F)** 2LMN. **(G)** MD-relaxed structure using 2MXU. The magnitude of frustration is quantified using different colors with red (highly frustrated), grey (neutral), and purple (minimally frustrated). The frustration profile is obtained using the frustratometer server (http://frustratometer.qb.fcen.uba.ar/). Single-residue frustration mode is adopted throughout the analysis. Note that for each panel, a 1D sequence representation is shown below (residues are colored using the same frustration color code).

frustrated, they form a cluster of residues and this cluster may play a role in binding its partners (ex., protein, DNA, RNA, ligands) or serving as an allosteric site. **Figure 5** presents the frustration profile for different types of fibrillar polymorphic structures. The highly frustrated residues are shown in red while minimally frustrated residues are shown in purple. We can see that the predicted frustrated areas are not unique but are distributed over the fibrillar surfaces. The exact location of the frustrated area is associated with the structure of the backbone in the given fibrillar architecture. **Table 2** summarizes the results of different fibrillar polymorphs. Interestingly, we find that the frustration profiles of different polymorphic structures share several common residues that are predicted to be highly frustrated, although their fibrillar structures are quite different. Charged residues, E22, D23, and K28, for example, are predicted to be highly frustrated across all the fibrillar structures studied. In this work, we also specifically look into the structure of 2MXU. The frustration profile of the 2MXU structure and that of the simulated one are quite similar. This result suggests that structural relaxation due to geometric packing of residues does not significantly affect the frustration profile. For the relaxed fibrillar structure, the most frustrated residues contain E22, D23, N27, K28, which are located in two separate areas in space: (E22, D23) and (N27, K28). The former includes primarily charged residues and is apparently electrostatically driven. The predicted E22-D23 site here agrees well with the binding sites of the N-ter surface, obtained from our

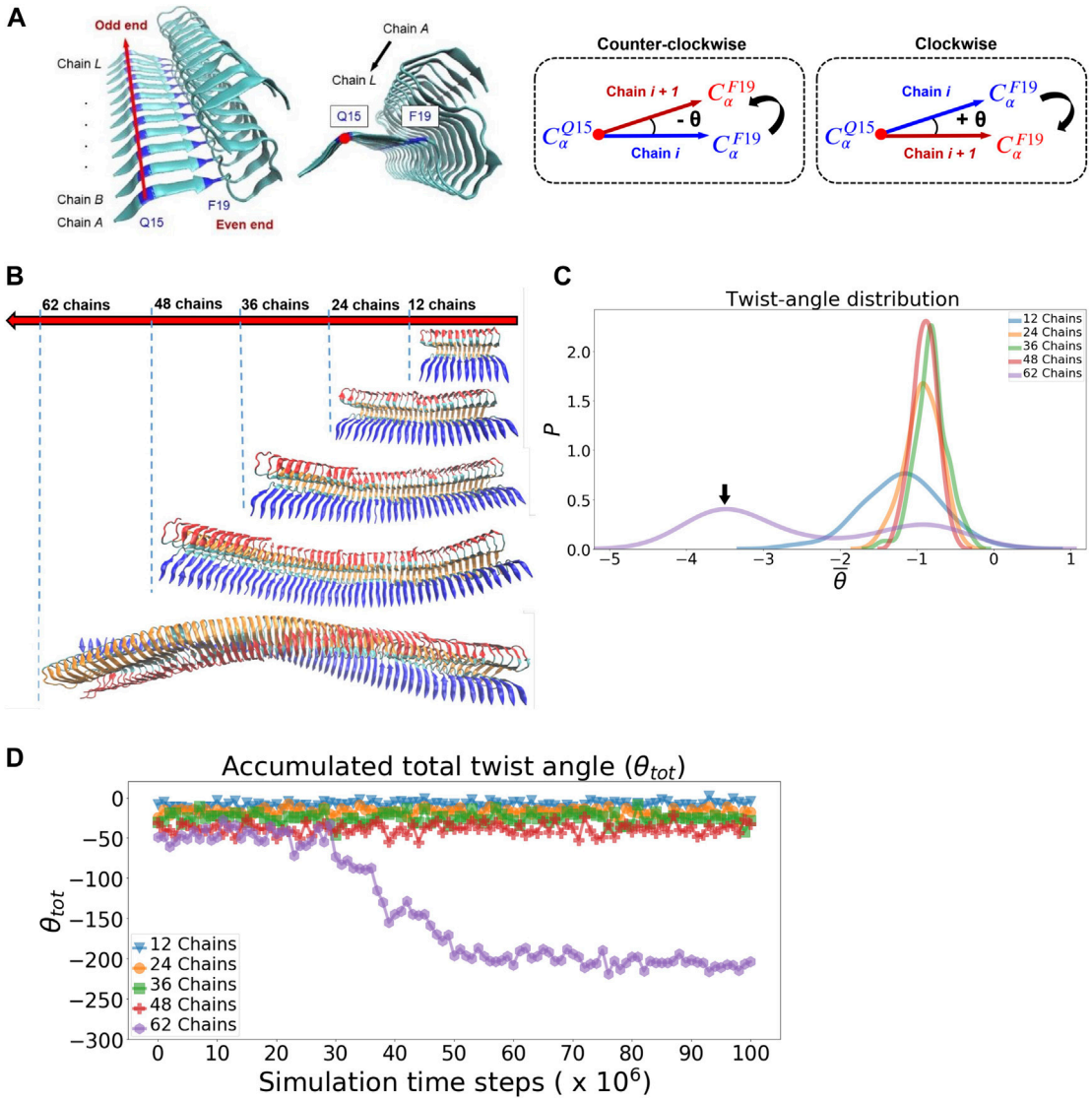


FIGURE 6 | Fibrillar twisting of protofilaments of different sizes is analyzed and compared. **(A)** The definition of the fibrillar twist angle θ is shown. **(B)** Representative fibrillar structures of different sizes are displayed (from top, 12, 24, 36, 48, and 62 chains). **(C)** Distribution of the averaged twist angle per chain ($\bar{\theta}$) for the protofilaments of different sizes. The black arrow indicates the second peak of the 62-chain protofilament, which represents the twisting structure of the 62-chain protofilament seen in **(B)**. **(D)** Accumulated total twist angle (θ_{tot}) for protofilaments of different sizes. For each of the filament species, the twist angle of each chain is added up to yield a final total twist angle. Note that the initial structure of the protofilament model is taken from the last frame of a pre-equilibration simulation (see **Supplementary Figure S10** in the Supporting Information for details).

MD simulation, to which the monomer binds (see **Figure 3B**). Here, we have shown an agreement in predicting fibrillar binding sites between the coarse-grained MD simulation and the bioinformatics tool (frustratometer2).

3.6 The Helical Twisting Around the Fibril Axis is an Emergent Mechanical Property of a Long Protofilament

A variety of protofilament morphologies of antiparallel beta-sheets have been reported (Stroud et al., 2012). In addition to cross-beta structure, one common feature across different fibrillar polymorphs is the helical topology that arises from the degree of overall filament twisting. The overall helical twisting accumulated from the twist angles of individual neighboring pairs of peptide chains. We carry out molecular dynamics simulation for protofilaments of different sizes and examine their twisting features. Starting with the relaxed fibrillar structure that was previously obtained, we used it to build models for the protofilaments with their sizes: 24, 36, 48, up to 62 chains. The calculation of the twist angle along the fibril axis is detailed in Methods, illustrated in **Figure 6A**. **Figure 6B** compares the final simulated fibrillar structures of different sizes in the simulation trajectory. The twist angle of individual neighboring

chain pairs are recorded in time series. **Figure 6C** shows the distribution of the averaged twist angle per chain pair $\bar{\theta}$. The distribution is primarily a single gaussian-like curve with its peak centered around −1° for protofilaments of size 12 to 48 chains; their corresponding structures are shown in **Figure 6B**. This averaged twist angle $\bar{\theta} \approx -1°$ represents the twisted morphology initially obtained from the solid-state NMR structure (2MXU). Interestingly, there exists a small second peak at −3° to −4°, exclusively for the 62-chain protofilament. This small shoulder observed indicates a further filament twisting to a larger degree, as shown in **Figure 6B** (62-chain protofilament is the last one). The alternative filament twisting observed is consistent with the existing literature on the structure of amyloid fibrils (Bedrood et al., 2012). **Figure 6D** shows the accumulated total twist angle of protofilament of all sizes as a function of simulation time steps. We find that the alternative twisting feature is not significant for those short filaments. In contrast, $\bar{\theta} = -3°$ to −4° is not observed until the filament size increases to the number of 62 chains. The result of the 62-chain protofilament clearly shows that the fibrillar structure starts to transform into a left-handed twisted fibrillar form at the time step $\approx 30 \times 10^6$. In other words, filament twisting becomes more significant for protofilaments in a large size. This result suggests that fibrillar twisting is associated with the propagation of localized interactions between neighboring pairs along the fibril axis—*via* cooperative effect. Since the energetics for filament twisting is primarily enthalpy-driven (Periole et al., 2018), all these results support that the filament twisting polymorphic structure is an emergent mechanical property, driven by the size effect of the filament. Such a large-scale mechanical coupling overall contributes to the helical twisting polymorphism of filaments. We show that the simulation protocol used can accurately simulate the mechanical feature of fibrillar twisting as well as the global structural rearrangement of amyloid protofilaments.

4 CONCLUSION

The interaction of a free amyloid protein monomer with a pre-existing fibrillar surface is an essential process which initiates subsequent fibril growth. Efficient fibril growth is mediated by several secondary processes such as elongation and surface-catalyzed nucleation. Understanding their kinetic pathways can provide mechanistic insight into the molecular mechanism of fibril growth. In this study, we have constructed a simulation platform for studying the early stage of fibril growth using a new GPU-enabled coarse-grained protein force field (openAWSEM). This simulation platform allows us to carry out long time simulations for protein aggregation over a fibrillar surface. We have investigated the thermodynamic binding affinity for a single monomer binding to fibrillar surfaces and find out that surface heterogeneity can significantly influence the predicted binding affinity. Accordingly, we have also identified several key surface binding sites: C-ter, N-ter, cleft, even-end, odd-end. Our study reveals several monomer-fibril binding configurations which potentially are amyloid precursors for subsequent elongation and secondary nucleation. This finding suggests that surface heterogeneity, entailed by the protein sequence and the resulting self-assembly, plays a key role in determining the aggregation pathways and, more importantly, it inevitably leads to variation in the thermodynamic binding affinity. In addition, we have used a bioinformatics tool to predict binding sites over different polymorphic fibrillar surfaces. For the fibril structure of interest, the frustration analysis predicts several potential functional sites, including residue E22 and D23 (mostly frustrated). These residues belong to the N-ter surface identified from the simulation. This binding site presumably can be modulated electrostatically (e.g., pH, ionic strength) to reflect its binding plasticity. To understand surface properties of fibrils in response to the global fibrillar twisting, we have simulated fibrillar twisting of single protofilaments with different sizes. Our result shows that fibrillar twisting is an emergent, collective property that correlates with the number of monomers participating in the fibril. We propose that the length-dependent fibrillar twisting may influence the population distribution of the amyloid precursors and thus drive different aggregation pathways for fibril growth. This work demonstrates the capability of the current simulation protocol for a comprehensive survey over fibril stability, binding affinity, surface heterogeneity, and mechanical twisting of polymorphic protofilaments. All these properties are prerequisite for understanding the molecular mechanism of surface-catalyzed secondary processes. We leave that task for future work.

DATA AVAILABILITY STATEMENT

The datasets (generated/ANALYZED) for this study can be found in the code repository of MYTLab/pyAggreg on Github. (https://github.com/MYTLab/pyAggreg).

AUTHOR CONTRIBUTIONS

Y-WM performed all the simulations. Y-WM, T-YL, and M-YT analyzed the data. M-YT designed the research, supervised it, and wrote the paper.

FUNDING

This work was supported by the Ministry of Science and Technology (MOST), Taiwan (R.O.C.), Grant No. 108-2113-M-032-003-MY2.

ACKNOWLEDGMENTS

We would like to thank Wei Lu for technical support of openAWSEM.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2021.719320/full#supplementary-material

REFERENCES

Bedrood, S., Li, Y., Isas, J. M., Hegde, B. G., Baxa, U., Haworth, I. S., et al. (2012). Fibril Structure of Human Islet Amyloid Polypeptide. *J. Biol. Chem.* 287, 5235–5241. doi:10.1074/jbc.m111.327817

Cecchini, M., Rao, F., Seeber, M., and Caflisch, A. (2004). Replica Exchange Molecular Dynamics Simulations of Amyloid Peptide Aggregation. *J. Chem. Phys.* 121, 10748–10756. doi:10.1063/1.1809588

Charest, N., Tro, M., Bowers, M. T., and Shea, J.-E. (2020). Latent Models of Molecular Dynamics Data: Automatic Order Parameter Generation for Peptide Fibrillization. *J. Phys. Chem. B* 124, 8012–8022. doi:10.1021/acs.jpcb.0c05763

Chen, M., Chen, X., Schafer, N. P., Clementi, C., Komives, E. A., Ferreiro, D. U., et al. (2020). Surveying Biomolecular Frustration at Atomic Resolution. *Nat. Commun.* 11, 5944. doi:10.1038/s41467-020-19560-9

Chen, M., Lin, X., Lu, W., Schafer, N. P., Onuchic, J. N., and Wolynes, P. G. (2018a). Template-Guided Protein Structure Prediction and Refinement Using Optimized Folding Landscape Force fields. *J. Chem. Theor. Comput.* 14, 6102–6116. doi:10.1021/acs.jctc.8b00683

Chen, M., Schafer, N. P., and Wolynes, P. G. (2018b). Surveying the Energy Landscapes of Aβ Fibril Polymorphism Fibril Polymorphism. *J. Phys. Chem. B* 122, 11414–11430. doi:10.1021/acs.jpcb.8b07364

Chen, M., Tsai, M., Zheng, W., and Wolynes, P. G. (2016). The Aggregation Free Energy Landscapes of Polyglutamine Repeats. *J. Am. Chem. Soc.* 138, 15197–15203. doi:10.1021/jacs.6b08665

Chen, X.-Q., and Mobley, W. C. (2019). Alzheimer Disease Pathogenesis: Insights from Molecular and Cellular Biology Studies of Oligomeric Aβ and Tau Species and Tau Species. *Front. Neurosci.* 13, 659. doi:10.3389/fnins.2019.00659

Cohen, S. I. A., Cukalevski, R., Michaels, T. C. T., Šarić, A., Törnquist, M., Vendruscolo, M., et al. (2018). Distinct Thermodynamic Signatures of Oligomer Generation in the Aggregation of the Amyloid-β Peptide Peptide. *Nat. Chem.* 10, 523–531. doi:10.1038/s41557-018-0023-x

Colvin, M. T., Silvers, R., Ni, Q. Z., Can, T. V., Sergeyev, I., Rosay, M., et al. (2016). Atomic Resolution Structure of Monomorphic Aβ42 Amyloid Fibrils42 Amyloid Fibrils. *J. Am. Chem. Soc.* 138, 9663–9674. doi:10.1021/jacs.6b05129

Crespo, R., Rocha, F. A., Damas, A. M., and Martins, P. M. (2012). A Generic Crystallization-like Model that Describes the Kinetics of Amyloid Fibril Formation. *J. Biol. Chem.* 287, 30585–30594. doi:10.1074/jbc.m112.375345

Daskalov, A., Martinez, D., Coustou, V., El Mammeri, N., Berbon, M., Andreas, L. B., et al. (2021). Structural and Molecular Basis of Cross-Seeding Barriers in Amyloids. *Proc. Natl. Acad. Sci. U. S. A.* 118, e2014085118. doi:10.1073/pnas.2014085118

Davtyan, A., Schafer, N. P., Zheng, W., Clementi, C., Wolynes, P. G., and Papoian, G. A. (2012). AWSEM-MD: Protein Structure Prediction Using Coarse-Grained Physical Potentials and Bioinformatically Based Local Structure Biasing. *J. Phys. Chem. B.* 116, 8494–8503. doi:10.1021/jp212541y

Deng, Y., and Roux, B. (2009). Computations of Standard Binding Free Energies with Molecular Dynamics Simulations. *J. Phys. Chem. B.* 113, 2234–2246. doi:10.1021/jp807701h

Di Palma, F., Bottaro, S., and Bussi, G. (2015). Kissing Loop Interaction in Adenine Riboswitch: Insights from Umbrella Sampling Simulations. *BMC Bioinformatics.* 16 (Suppl. 9), S6. doi:10.1186/1471-2105-16-S9-S6

[Dataset] Eastman, P., Swails, J., Chodera, J. D., McGibbon, R. T., Zhao, Y., Beauchamp, K. A., et al. (2017). OpenMM 7: Rapid Development of High Performance Algorithms for Molecular Dynamics. *Plos Comput. Biol.* 13, e1005659. doi:10.1371/journal.pcbi.1005659

Fändrich, M., Nyström, S., Nilsson, K. P. R., Böckmann, A., LeVine, H., 3rd, and Hammarström, P. (2018). Amyloid Fibril Polymorphism: a challenge for Molecular Imaging and Therapy. *J. Intern. Med.* 283, 218–237. doi:10.1111/joim.12732

Ferreiro, D. U., Hegler, J. A., Komives, E. A., and Wolynes, P. G. (2007). Localizing Frustration in Native Proteins and Protein Assemblies. *Proc. Natl. Acad. Sci.* 104, 19819–19824. doi:10.1073/pnas.0709915104

Ferreiro, D. U., Komives, E. A., and Wolynes, P. G. (2018). Frustration, Function and Folding. *Curr. Opin. Struct. Biol.* 48, 68–73. doi:10.1016/j.sbi.2017.09.006

Grazioli, G., Martin, R. W., and Butts, C. T. (2019). Comparative Exploratory Analysis of Intrinsically Disordered Protein Dynamics Using Machine Learning and Network Analytic Methods. *Front. Mol. Biosci.* 6, 42. doi:10.3389/fmolb.2019.00042

Gremer, L., Schölzel, D., Schenk, C., Reinartz, E., Labahn, J., Ravelli, R. B. G., et al. (2017). Fibril Structure of Amyloid-B(1-42) by Cryo-Electron Microscopy(1-42) by Cryo-Electron Microscopy. *Science* 358, 116–119. doi:10.1126/science.aao2825

Han, W., and Schulten, K. (2014). Fibril Elongation by Aβ17-42: Kinetic Network Analysis of Hybrid-Resolution Molecular Dynamics Simulations17–42: Kinetic Network Analysis of Hybrid-Resolution Molecular Dynamics Simulations. *J. Am. Chem. Soc.* 136, 12450–12460. doi:10.1021/ja507002p

Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: Visual Molecular Dynamics. *J. Mol. Graph.* 14, 33–38. doi:10.1016/0263-7855(96)00018-5

Ilie, I. M., and Caflisch, A. (2018). Disorder at the Tips of a Disease-Relevant Aβ42 Amyloid Fibril: A Molecular Dynamics Study42 Amyloid Fibril: A Molecular Dynamics Study. *J. Phys. Chem. B* 122, 11072–11082. doi:10.1021/acs.jpcb.8b05236

Jin, S., Contessoto, V. G., Chen, M., Schafer, N. P., Lu, W., Chen, X., et al. (2020). AWSEM-suite: a Protein Structure Prediction Server Based on Template-Guided, Coevolutionary-Enhanced Optimized Folding Landscapes. *Nucleic Acids Res.* 48, W25–W30. doi:10.1093/nar/gkaa356

Kollmer, M., Close, W., Funk, L., Rasmussen, J., Bsoul, A., Schierhorn, A., et al. (2019). Cryo-EM Structure and Polymorphism of Aβ Amyloid Fibrils Purified from Alzheimer's Brain Tissue Amyloid Fibrils Purified from Alzheimer's Brain Tissue. *Nat. Commun.* 10, 4760. doi:10.1038/s41467-019-12683-8

Konno, H., Watanabe-Nakayama, T., Uchihashi, T., Okuda, M., Zhu, L., Kodera, N., et al. (2020). Dynamics of Oligomer and Amyloid Fibril Formation by Yeast Prion Sup35 Observed by High-Speed Atomic Force Microscopy. *Proc. Natl. Acad. Sci. USA* 117, 7831–7836. doi:10.1073/pnas.1916452117

[Dataset] Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H., and Kollman, P. A. (1992). THE Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules. I. The Method. *J. Comput. Chem.* 13, 1011–1021. doi:10.1002/jcc.540130812

Laughlin, R. B., Pines, D., Schmalian, J., Stojkovic, B. P., and Wolynes, P. (2000). The Middle Way. *Proc. Natl. Acad. Sci. U S A.* 97, 32–37. doi:10.1073/pnas.97.1.32

Lu, W., Bueno, C., Schafer, N. P., Moller, J., Jin, S., Chen, M., et al. (2020). OpenAWSEM with Open3SPN2: a Fast, Flexible, and Accessible Framework for Large-Scale Coarse-Grained Biomolecular Simulations. *Plos Comput. Biol.* 17, e1008308. doi:10.1371/journal.pcbi.1008308

[Dataset] Marcovitz, A., and Levy, Y. (2013). Weak Frustration Regulates Sliding and Binding Kinetics on Rugged Protein–DNA Landscapes. *J. Phys. Chem. B.* 117, 13005–13014. doi:10.1021/jp402296d

McGibbon, R. T., Beauchamp, K. A., Harrigan, M. P., Klein, C., Swails, J. M., Hernández, C. X., et al. (2015). MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophysical J.* 109, 1528–1532. doi:10.1016/j.bpj.2015.08.015

Okumura, H., and Itoh, S. G. (2016). Structural and Fluctuational Difference between Two Ends of Aβ Amyloid Fibril: MD Simulations Predict Only One End Has Open Conformations Amyloid Fibril: MD Simulations Predict Only One End Has Open Conformations. *Sci. Rep.* 6, 38422. doi:10.1038/srep38422

[Dataset] O'Nuallain, B., Shivaprasad, S., Kheterpal, I., and Wetzel, R. (2005). Thermodynamics of *Aβ*(1–40) Amyloid Fibril Elongation. *Biochemistry* 44, 12709–12718. doi:10.1021/bi050927h

Parra, R. G., Schafer, N. P., Radusky, L. G., Tsai, M.-Y., Guzovsky, A. B., Wolynes, P. G., et al. (2016). Protein Frustratometer 2: a Tool to Localize Energetic Frustration in Protein Molecules, Now with Electrostatics. *Nucleic Acids Res.* 44, W356–W360. doi:10.1093/nar/gkw304

Periole, X., Huber, T., Bonito-Oliva, A., Aberg, K. C., van der Wel, P. C. A., Sakmar, T. P., et al. (2018). Energetics Underlying Twist Polymorphisms in Amyloid Fibrils. *J. Phys. Chem. B* 122, 1081–1091. doi:10.1021/acs.jpcb.7b10233

Potoyan, D. A., Bueno, C., Zheng, W., Komives, E. A., and Wolynes, P. G. (2017). Resolving the NFκB Heterodimer Binding Paradox: Strain and Frustration Guide the Binding of Dimeric Transcription FactorsB Heterodimer Binding Paradox: Strain and Frustration Guide the Binding of Dimeric Transcription Factors. *J. Am. Chem. Soc.* 139, 18558–18566. doi:10.1021/jacs.7b08741

Potoyan, D. A., Zheng, W., Ferreiro, D. U., Wolynes, P. G., and Komives, E. A. (2016a). PEST Control of Molecular Stripping of NFκB from DNA

Transcription SitesB from DNA Transcription Sites. *J. Phys. Chem. B* 120, 8532–8538. doi:10.1021/acs.jpcb.6b02359

Potoyan, D. A., Zheng, W., Komives, E. A., and Wolynes, P. G. (2016b). Molecular Stripping in theNF-κB/IκB/DNAgenetic Regulatory networkB/IB/DNA Genetic Regulatory Network. *Proc. Natl. Acad. Sci. USA* 113, 110–115. doi:10.1073/pnas.1520483112

Qiang, W., Kelley, K., and Tycko, R. (2013). Polymorph-Specific Kinetics and Thermodynamics of β-Amyloid Fibril Growth-Amyloid Fibril Growth. *J. Am. Chem. Soc.* 135, 6860–6871. doi:10.1021/ja311963f

Riek, R., and Eisenberg, D. S. (2016). The Activities of Amyloids from a Structural Perspective. *Nature* 539, 227–235. doi:10.1038/nature20416

Rodriguez, R. A., Chen, L. Y., Plascencia-Villa, G., and Perry, G. (2018). Thermodynamics of Amyloid-β Fibril Elongation: Atomistic Details of the Transition State Fibril Elongation: Atomistic Details of the Transition State. *ACS Chem. Neurosci.* 9, 783–789. doi:10.1021/acschemneuro.7b00409

Sasmal, S., Schwierz, N., and Head-Gordon, T. (2016). Mechanism of Nucleation and Growth of Aβ40 Fibrils from All-Atom and Coarse-Grained Simulations40 Fibrils from All-Atom and Coarse-Grained Simulations. *J. Phys. Chem. B* 120, 12088–12097. doi:10.1021/acs.jpcb.6b09655

Scheidt, T., Łapińska, U., Kumita, J. R., Whiten, D. R., Klenerman, D., Wilson, M. R., et al. (2019). Secondary Nucleation and Elongation Occur at Different Sites on Alzheimer's Amyloid- Aggregates. *Sci. Adv.* 5, eaau3112. doi:10.1126/sciadv.aau3112

Scherer, M. K., Trendelkamp-Schroer, B., Paul, F., Pérez-Hernández, G., Hoffmann, M., Plattner, N., et al. (2015). PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models. *J. Chem. Theor. Comput.* 11, 5525–5542. doi:10.1021/acs.jctc.5b00743

[Dataset] Schmidt, M., Sachse, C., Richter, W., Xu, C., Fandrich, M., and Grigorieff, N.. (2009). Comparison of Alzheimer a (1-40) and a (1-42) Amyloid Fibrils Reveals Similar Protofilament Structuresdoi:10.1073/pnas.0905007106

Schwierz, N., Frost, C. V., Geissler, P. L., and Zacharias, M. (2016). Dynamics of Seeded Aβ40-Fibril Growth from Atomistic Molecular Dynamics Simulations: Kinetic Trapping and Reduced Water Mobility in the Locking Step40-Fibril Growth from Atomistic Molecular Dynamics Simulations: Kinetic Trapping and Reduced Water Mobility in the Locking Step. *J. Am. Chem. Soc.* 138, 527–539. doi:10.1021/jacs.5b08717

Schwierz, N., Frost, C. V., Geissler, P. L., and Zacharias, M. (2017). From Aβ Filament to Fibril: Molecular Mechanism of Surface-Activated Secondary Nucleation from All-Atom MD Simulations Filament to Fibril: Molecular Mechanism of Surface-Activated Secondary Nucleation from All-Atom MD Simulations. *J. Phys. Chem. B* 121, 671–682. doi:10.1021/acs.jpcb.6b10189

Sirovetz, B. J., Schafer, N. P., and Wolynes, P. G. (2017). Protein Structure Prediction: Making AWSEM AWSEM-ER by Adding Evolutionary Restraints. *Proteins* 85, 2127–2142. doi:10.1002/prot.25367

Stroud, J. C., Liu, C., Teng, P. K., and Eisenberg, D. (2012). Toxic Fibrillar Oligomers of Amyloid- Have Cross- Structure Have Cross- Structure. *Proc. Natl. Acad. Sci.* 109, 7717–7722. doi:10.1073/pnas.1203193109

Thacker, D., Sanagavarapu, K., Frohm, B., Meisl, G., Knowles, T. P. J., and Linse, S. (2020). The Role of Fibril Structure and Surface Hydrophobicity in Secondary Nucleation of Amyloid Fibrils. *Proc. Natl. Acad. Sci. USA* 117, 25272–25283. doi:10.1073/pnas.2002956117

Tsai, M.-Y. (2019). Role of Physical Nucleation Theory in Understanding Conformational Conversion between Pathogenic and Nonpathogenic Aggregates of Low-Complexity Amyloid Peptides. *Res. Chem. Intermed.* 45, 5357–5373. doi:10.1007/s11164-019-03974-2

Tsai, M.-Y., Yuan, J.-M., and Lin, S.-H. (2015). Thermodynamic Insight into Protein Aggregation Using a Kinetic Ising Model. *Jnl Chin. Chem. Soc* 62, 21–25. doi:10.1002/jccs.201400272

Tsai, M.-Y., Zhang, B., Zheng, W., and Wolynes, P. G. (2016a). Molecular Mechanism of Facilitated Dissociation of Fis Protein from DNA. *J. Am. Chem. Soc.* 138, 13497–13500. doi:10.1021/jacs.6b08416

Tsai, M.-Y., Zheng, W., Balamurugan, D., Schafer, N. P., Kim, B. L., Cheung, M. S., et al. (2016b). Electrostatics, Structure Prediction, and the Energy Landscapes for Protein Folding and Binding. *Protein Sci.* 25, 255–269. doi:10.1002/pro.2751

Tsai, M.-Y., Zheng, W., Chen, M., and Wolynes, P. G. (2019). Multiple Binding Configurations of Fis Protein Pairs on DNA: Facilitated Dissociation versus Cooperative Dissociation. *J. Am. Chem. Soc.* 141, 18113–18126. doi:10.1021/jacs.9b08287

Tycko, R. (2015). Amyloid Polymorphism: Structural Basis and Neurobiological Relevance. *Neuron* 86, 632–645. doi:10.1016/j.neuron.2015.03.017

Wälti, M. A., Ravotti, F., Arai, H., Glabe, C. G., Wall, J. S., Böckmann, A., et al. (2016). Atomic-resolution Structure of a Disease-Relevant Aβ(1-42) Amyloid Fibril(1-42) Amyloid Fibril. *Proc. Natl. Acad. Sci. USA* 113, E4976–E4984. doi:10.1073/pnas.1600749113

Xiao, Y., Ma, B., McElheny, D., Parthasarathy, S., Long, F., Hoshi, M., et al. (2015). Aβ(1-42) Fibril Structure Illuminates Self-Recognition and Replication of Amyloid in Alzheimer's Disease. *Nat. Struct. Mol. Biol.* 22, 499–505. doi:10.1038/nsmb.2991

Xu, Y., Safari, M. S., Ma, W., Schafer, N. P., Wolynes, P. G., and Vekilov, P. G. (2019). Steady, Symmetric, and Reversible Growth and Dissolution of Individual Amyloid-β Fibrils Fibrils. *ACS Chem. Neurosci.* 10, 2967–2976. doi:10.1021/acschemneuro.9b00179

Xue, W.-F., Homans, S. W., and Radford, S. E. (2008). Systematic Analysis of Nucleation-dependent Polymerization Reveals New Insights into the Mechanism of Amyloid Self-Assembly. *Pnas* 105, 8926–8931. doi:10.1073/pnas.0711664105

Zhang, B., Zheng, W., Papoian, G. A., and Wolynes, P. G. (2016). Exploring the Free Energy Landscape of Nucleosomes. *J. Am. Chem. Soc.* 138, 8126–8133. doi:10.1021/jacs.6b02893

Zhang, R., Hu, X., Khant, H., Ludtke, S. J., Chiu, W., Schmid, M. F., et al. (2009). Interprotofilament Interactions between Alzheimer's A 1-42 Peptides in Amyloid Fibrils Revealed by cryoEM. *Proc. Natl. Acad. Sci.* 106, 4653–4658. doi:10.1073/pnas.0901085106

Zheng, W., Schafer, N. P., Davtyan, A., Papoian, G. A., and Wolynes, P. G. (2012). Predictive Energy Landscapes for Protein-Protein Association. *Proc. Natl. Acad. Sci.* 109, 19244–19249. doi:10.1073/pnas.1216215109

Zheng, W., Tsai, M.-Y., Chen, M., and Wolynes, P. G. (2016). Exploring the Aggregation Free Energy Landscape of the Amyloid-β Protein (1-40) Protein (1-40). *Proc. Natl. Acad. Sci. USA* 113, 11835–11840. doi:10.1073/pnas.1612362113

Zheng, W., Tsai, M.-Y., and Wolynes, P. G. (2017). Comparing the Aggregation Free Energy Landscapes of Amyloid Beta(1-42) and Amyloid Beta(1-40). *J. Am. Chem. Soc.* 139, 16666–16676. Accepted. doi:10.1021/jacs.7b08089

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership