MOLECULAR DYNAMICS AND MACHINE LEARNING IN DRUG DISCOVERY

EDITED BY: Sergio Decherchi, Andrea Cavalli, Pratyush Tiwary and Francesca Grisoni
PUBLISHED IN: Frontiers in Molecular Biosciences

frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-88966-863-2
DOI 10.3389/978-2-88966-863-2

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

MOLECULAR DYNAMICS AND MACHINE LEARNING IN DRUG DISCOVERY

Topic Editors:
Sergio Decherchi, Italian Institute of Technology (IIT), Italy
Andrea Cavalli, Italian Institute of Technology (IIT), Italy
Pratyush Tiwary, University of Maryland, College Park, United States
Francesca Grisoni, ETH Zürich, Switzerland

Dr. Sergio Decherchi and Dr. Andrea Cavalli are co-founders of BiKi Technologies s.r.l. - a company that commercializes a Molecular Dynamics-based software suite for drug discovery. All other Topic Editors declare no competing interests with regards to the Research Topic subject.

Citation: Decherchi, S., Cavalli, A., Tiwary, P., Grisoni, F., eds. (2021). Molecular Dynamics and Machine Learning in Drug Discovery. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88966-863-2

Table of Contents

frontiers in Molecular Biosciences

EDITORIAL
published: 13 April 2021
doi: 10.3389/fmolb.2021.673773

Editorial: Molecular Dynamics and Machine Learning in Drug Discovery

Sergio Decherchi[1,2]*, *Francesca Grisoni*[3], *Pratyush Tiwary*[4] and *Andrea Cavalli*[1,5]

[1] Fondazione Istituto Italiano di Tecnologia, Genoa, Italy, [2] BiKi Technologies s.r.l., Genoa, Italy, [3] Department of Biomedical Engineering, Eindhoven University of Technology, Eindhoven, Netherlands, [4] Department of Chemistry and Biochemistry, Institute for Physical Science and Technology, University of Maryland, College Park, MD, United States, [5] Department of Pharmacy and Biotechnology (FaBiT), Alma Mater Studiorum – University of Bologna, Bologna, Italy

Keywords: machine learning, molecular dynamics, computational chemistry (molecular design), drug discovery, physics-based models

Editorial on the Research Topic

Molecular Dynamics and Machine Learning in Drug Discovery

The drug discovery process is very long and expensive, and many factors hamper its final success. In the attempt to accelerate a drug candidate's progress along the discovery pipeline, computational modeling represents a key tool to address the design and optimization of lead compounds. While physics-based white-box modeling (e.g., docking and molecular dynamics), has represented the standard de facto for many years in the computational chemistry community, nowadays machine learning methodologies represent a powerful modeling alternative. The deep learning paradigm in particular can be considered a black box methodology as it can be difficult to extract rules or laws from the trained model.

This Research Topic collects selected contributions that deal with both types of modeling approaches, some of which lie at the interface between the two. This "gray box" hybrid approach should not surprise as machine learning and statistical mechanics share several theoretical principles (Ferrarotti et al., 2019; Noé et al., 2019; Agliari et al., 2020; Decherchi and Cavalli, 2020; Ferraro et al., 2020; Tsai et al., 2020) as they both deal with distributions, manifolds, and hence free energies.

Molecular dynamics (MD) is based on statistical mechanics. Setting up a MD run for complex systems can be still a not trivial task, requiring continuous automation tools to allow for a wider exploitation in academic and industrial settings. In this regard, the contribution from Schneider et al. discusses the implementation of a webserver for the setup of hybrid molecular mechanics and coarse-grained simulations for Human G-Protein Coupled Receptors (GPCRs) and ligands complexes. GPCRs represent the most important class of druggable targets, hence the importance of having handy tools to setup their systematic simulations. Analyzing and understanding MD outcomes can then be rather complicated, mainly because of the large amount of raw data. Bunker and Rog present a review on the mechanistic understanding of MD generated data for drug delivery in pharmaceutical research. Mechanistic interpretations can be supported by proper machine learning tools: it is often convenient to devise/use clustering, projections or feature extraction algorithms to extract actionable knowledge. This greatly facilitates the interpretation of results and can also allow to define order parameters in some cases, often dubbed collective variables in the MD realm. In the contribution from Arthur et al. Authors devise a combination of MD simulations of proteins and hierarchical pharmacophore features extraction. This strategy represents a smart and widely applicable paradigm (Spyrakis et al., 2015) which combines MD sampling (to recover some of the target flexibility) with a non-dynamical tool (e.g., virtual screening, static docking etc.). While this paper derives features directly for drug discovery, Spiwok and Kriz propose a more

Edited and reviewed by:
Francesco Luigi Gervasio, University College London, United Kingdom

***Correspondence:**
Sergio Decherchi
sergio.decherchi@iit.it

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

Received: *28 February 2021*
Accepted: *19 March 2021*
Published: *13 April 2021*

Citation:
Decherchi S, Grisoni F, Tiwary P and Cavalli A (2021) Editorial: Molecular Dynamics and Machine Learning in Drug Discovery. Front. Mol. Biosci. 8:673773. doi: 10.3389/fmolb.2021.673773

general approach. The Authors present a new machine learning algorithm, named time lagged t-SNE, which is able to explicitly take into consideration different time scales in the simulation. Such detection (and acceleration) of slow and fast time scales is very important in drug discovery; in protein ligand binding, for instance, several phenomena of interest (e.g., unbinding) happen at very large time scales as they are rare events.

At variance of qualitative analysis for devising mechanistic hypothesis, quantitatively converged estimations are the only possible path to try predicting physical observables. Statistical mechanics research has developed powerful theories for quantifying observables of interest and, in the drug discovery realm, binding free energy is a key physicochemical quantity. Despite the great improvements in the last 20 years both in theoretical and technological terms, predictive free energy computations still remain partially elusive for many reasons, such as the massive computing power needed for convergence, the force field accuracy, possible numerical instabilities in some cases, and the partial disconnection between experimental observables and what is effectively estimated by computations. Hall et al. discuss this important aspect, namely the relationship between kinetics estimated via the weighted ensemble method and the experimental affinity. While the general relationship between experimental kinetics and affinity is known, when it comes to simulations the situation becomes subtler. Authors show that some correction terms (for instance finite-size effects) whose energetic contribution is not negligible arise. These corrections allow to get much more accurate free energy estimations derived from kinetics rates estimated via the weighted ensemble. Free energy (or kinetics) simulations can be quite expensive, hence approximate methods can be devised. Fully data driven or approximate physics-based models have proved more or less effective in granting a compromise between accuracy and efficiency. The Linear Interaction Energy method is one of such approximate physics-based strategy and Rifai et al. discuss recent advances of this methodology. Interestingly, from a machine learning perspective, this kind of methodologies could be ascribed to the previously mentioned "gray box" approaches. They start from a physically sound *ansatz* and then switch to a data-driven style to tune the remaining parameters to save computing time. End-to-end data-driven attempts are also possible, often based on *ad-hoc* engineered features to describe the ligand and the protein, on which machine learning can be applied. This is what happens in the contributions of Holderbach et al. and in Parks et al. where physicochemical features are first devised and then used to predict affinity.

This collection of articles has dealt with many, and often interconnected, algorithmic approaches to speed-up the discovery of new drugs and the estimation of key observables such as free energy. We believe this collection will be useful to computational and medicinal chemists willing to apply recent *in-silico* methodologies, ranging from pure MD to fully data-driven approaches. We thank all Authors, co-Authors, and Reviewers for their contribution to this Research Topic and acknowledge Frontiers Team members'support.

AUTHOR CONTRIBUTIONS

SD conceived the topic, managed the review process, and wrote the editorial. FG, PT, and AC contributed to the review process and to the editorial. All authors contributed to the article and approved the submitted version.

REFERENCES

Agliari, E., Barra, A., Sollich, P., and Zdeborová, L. (2020). Machine learning and statistical physics: preface. *J. Phys. A Math. Theoretic.* 53:500401. doi: 10.1088/1751-8121/abca75

Decherchi, S., and Cavalli, A. (2020). Thermodynamics and kinetics of drug-target binding by molecular simulation. *Chem. Rev.* 120, 12788–12833. doi: 10.1021/acs.chemrev.0c00534

Ferraro, M., Decherchi, S., De Simone, A., Recanatini, M., Cavalli, A., and Bottegoni, G. (2020). Multi-target dopamine D3 receptor modulators: actionable knowledge for drug design from molecular dynamics and machine learning. *Eur. J. Med. Chem.* 188:111975. doi: 10.1016/j.ejmech.2019.111975

Ferrarotti, M. J., Rocchia, W., and Decherchi, S. (2019). Finding principal paths in data space. *IEEE Transac. Neural Netw. Learn. Syst.* 30, 2449–2462. doi: 10.1109/TNNLS.2018.2884792

Noé, F., Olsson, S., Köhler, J., and Wu, H. (2019). Boltzmann generators: sampling equilibrium states of many-body systems with deep learning. *Science* 365:eaaw1147. doi: 10.1126/science.aaw1147

Spyrakis, F., Benedetti, P., Decherchi, S., Rocchia, W., Cavalli, A., Alcaro, S., et al. (2015). A pipeline to enhance ligand virtual screening: integrating molecular dynamics and fingerprints for ligand and proteins. *J. Chem. Inf. Model.* 55, 2256–2274. doi: 10.1021/acs.jcim.5b00169

Tsai, S.-T., Kuo, E.-J., and Tiwary, P. (2020). Learning molecular dynamics with simple language model built upon long short-term memory neural network. *Nat. Commun.* 11:5115. doi: 10.1038/s41467-020-18959-8

Conflict of Interest: SD and AC are co-founders of BiKi Technologies s.r.l. a company which commercializes computational chemistry tools for drug discovery.

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: 05 June 2020
doi: 10.3389/fmolb.2020.00106

On Calculating Free Energy Differences Using Ensembles of Transition Paths

Robert Hall[1], *Tom Dixon*[1,2] *and Alex Dickson*[1,2]*

[1] Department of Biochemistry & Molecular Biology, Michigan State University, East Lansing, MI, United States, [2] Department of Computational Mathematics, Science and Engineering, Michigan State University, East Lansing, MI, United States

The free energy of a process is the fundamental quantity that determines its spontaneity or propensity at a given temperature. In particular, the binding free energy of a drug candidate to its biomolecular target is used as an objective quantity in drug design. Recently, binding kinetics—rates of association (k_{on}) and dissociation (k_{off})—have also demonstrated utility for their ability to predict efficacy and in some cases have been shown to be more predictive than the binding free energy alone. Some methods exist to calculate binding kinetics from molecular simulations, although these are typically more difficult to calculate than the binding affinity as they depend on details of the transition path ensemble. Assessing these rate constants can be difficult, due to uncertainty in the definition of the bound and unbound states, large error bars and the lack of experimental data. As an additional consistency check, rate constants from simulation can be used to calculate free energies (using the log of their ratio) which can then be compared to free energies obtained experimentally or using alchemical free energy perturbation. However, in this calculation it is not straightforward to account for common, practical details such as the finite simulation volume or the particular definition of the "bound" and "unbound" states. Here we derive a set of correction terms that can be applied to calculations of binding free energies using full reactive trajectories. We apply these correction terms to revisit the calculation of binding free energies from rate constants for a host-guest system that was part of a blind prediction challenge, where significant deviations were observed between free energies calculated with rate ratios and those calculated from alchemical perturbation. The correction terms combine to significantly decrease the error with respect to computational benchmarks, from 3.4 to 0.76 kcal/mol. Although these terms were derived with weighted ensemble simulations in mind, some of the correction terms are generally applicable to free energies calculated using physical pathways via methods such as Markov state modeling, metadynamics, milestoning, or umbrella sampling.

Keywords: free energy, molecular dynamics, enhanced sampling, binding kinetics, statistical mechanics, nonequilibrium

OPEN ACCESS

Edited by:
Pratyush Tiwary,
University of Maryland, College Park, United States

Reviewed by:
Steffen Wolf,
University of Freiburg, Germany
Wei Yang,
State College of Florida, Manatee-Sarasota, United States

***Correspondence:**
Alex Dickson
alexrd@msu.edu

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

Received: *26 February 2020*
Accepted: *06 May 2020*
Published: *05 June 2020*

Citation:
Hall R, Dixon T and Dickson A (2020) On Calculating Free Energy Differences Using Ensembles of Transition Paths. Front. Mol. Biosci. 7:106. doi: 10.3389/fmolb.2020.00106

1. INTRODUCTION

In recent years there is a growing appreciation for the utility of binding kinetics in the prediction of drug efficacy (Lu and Tonge, 2010; Carroll et al., 2012; Vauquelin et al., 2012; Pei et al., 2014; Ayaz et al., 2016; Copeland, 2016; Costa et al., 2016; Guo et al., 2016; Tonge, 2017; Bruce et al., 2018; Lee et al., 2019; Nunes-Alves et al., 2020). Pharmacokinetic and pharmacodynamic models of drug activity in the body are inherently out of equilibrium: a drug is administered, it is absorbed, distributed to different tissues, metabolized and eliminated from the body. As such, kinetic constants of binding and release—beyond just the equilibrium constants of binding—are required to model drug action when the timescales of binding and release cannot be separated from the other competing processes (Bernetti et al., 2017). The relationship between molecular structure and the kinetics of binding (also called "structure-kinetic relationships" or SKR) is complicated, as small changes to structure can change kinetic constants by orders of magnitude (Ayaz et al., 2016). It is important to note that changes in kinetics are not always tied to changes in affinity (Guo et al., 2014), and that to accurately predict changes in kinetics, models of the ligand-binding transition state are needed to estimate transition-state stabilization or destabilization (Spagnuolo et al., 2017).

Computational methods that reveal structures of transition states and calculate binding (k_{on}) and unbinding (k_{off}) rate constants for real compounds are in their infancy, but are quickly developing (Dickson et al., 2017). It is a tremendous challenge to obtain reliable values for these quantities, as (1) they depend on the entire (un)binding pathway, not just its endpoints, and (2) the timescales of ligand binding and release often exceed the capabilities of molecular dynamics simulations by orders of magnitude. Specialized computing platforms have been applied to generate continuous binding pathways (Dror et al., 2011), although the unbinding process is typically beyond the reach of molecular dynamics simulation for compounds beyond millimolar drug fragments (Guo et al., 2016; Pan et al., 2017). Recent studies have used enhanced sampling methods in molecular dynamics to simulate ligand (un)binding pathways and determine mechanisms and rate constants k_{on} and k_{off} (Casasnovas et al., 2017; Tiwary et al., 2017; Dickson, 2018; Kokh et al., 2018; Lotz and Dickson, 2018; Bruno et al., 2019; Deb and Frank, 2019; Kirberger et al., 2019; Dixon et al., 2020). Some of these rate constants have shown surprisingly good agreement with experiment—given the extraordinarily long timescales involved—however these have the confounding uncertainty of force field accuracy (Yin et al., 2017; Camilloni and Pietrucci, 2018), there is a possibility for fortuitous cancelation of error. Unfortunately, the computational cost required to predict these quantities is typically massive (Camilloni and Pietrucci, 2018), especially for large protein systems and ligands with extremely long residence times, precluding the study of these events under a series of different simulation conditions (e.g., forcefields, water models, polarizability).

In the field of biomolecular modeling, blind challenges—where a series of objectives are released by the organizers, and participants entries are directly judged by their agreement with experiment—have been useful catalysts for the development of predictive algorithms (Lensink et al., 2017; Synapse, 2018; Croll et al., 2019; Parks et al., 2020). Although no blind challenge currently exists for the prediction of k_{on} and k_{off}, we recently participated in the SAMPL6 SAMPLing challenge, which required participants to compute free energies as a function of simulation time and to compare the computational cost of different free energy calculation methods (Dixon et al., 2018; Rizzi et al., 2018, 2020). This challenge allows sampling methods to be assessed independently of force field accuracy, as all entries used the same initial coordinates, force field parameters and partial charges. Importantly, the challenge makes use of very small model systems (host-guest) that require considerably less computational resources to simulate, which allowed us to efficiently simulate binding and release for a number of systems, determine k_{on} and k_{off}, and predict values for the binding free energy (ΔG) that would then be compared to experimental observables, as well as results from alchemical free energy perturbation methods (Gilson et al., 1997; Shirts and Chodera, 2008).

The standard free energy of binding was determined as a function of rate constants:

$$\Delta G = -k_B T \ln \frac{C^0 k_{on}}{k_{off}} \tag{1}$$

where C^0 is a reference concentration of 1 mol/L. In this paper, we revisit this equation in detail and explicitly examine the assumptions made when the rate constants used in Equation (1) are computed through typical simulations with finite box-size and periodic boundary conditions. In section 3.1, we derive three correction terms that can be easily computed and facilitate a better connection with both experiment and alchemical computational free energy calculations. One term accounts for the particular definitions of the bound and unbound states. The second term accounts for residual electrostatic interactions that might still be present between the molecules, which is especially useful if one or both of the molecules carry an explicit charge. The third term accounts for the volume of the unbound state in the simulation box, which is useful to keep the simulated volume as small as possible during rate calculations. These terms were derived particularly with weighted ensemble simulations in mind, where rates are computed using the trajectory flux between two non-equilibrium ensembles. However, the second and third term can be directly applied to other simulation methods which employ physical simulation of the transition path ensemble, such as Markov state modeling (Singhal et al., 2004; Gu et al., 2014), metadynamics (Laio and Parrinello, 2002; Tiwary et al., 2017), milestoning (Faradjian and Elber, 2004; Votapka et al., 2017), and umbrella sampling (Torrie and Valleau, 1977; Nishikawa et al., 2018).

To examine questions of convergence, we reproduce our binding and unbinding simulations for a host-guest system with larger numbers of replicas and longer simulation times. We also explore the effects of the Langevin integrator on the prediction of unbinding and binding rates; in particular, how altering the friction coefficient (γ), defined in the Langevin integrator,

impacts the binding and release processes. Although γ does not appear in the internal energy function, and hence cannot affect thermodynamic properties such as the binding free energy, we examine whether lower friction coefficients can accelerate the convergence of unbinding simulations.

2. METHODS

2.1. Host-Guest Systems

The host-guest system utilized in this study is referred to as "OA-G6" (**Figure 1**), where the host is a Gibb deep cavity cavitand, referred to as an "octa acid" or "OA" (Gan et al., 2011). OA forms a basket-like structure with 4-fold symmetry, functionalized with four benzoic-acid substituents on the top rim of the basket and four more on the bottom. The guest ligand we study here is 4-methyl pentanoic acid (referred to as "G6"). This ligand harbors a negative charge at the carboxyl end of the alkyl chain.

2.2. Molecular Dynamics

The OA-G6 configuration was obtained from the organizers of the SAMPLing challenge (Rizzi et al., 2020). The system was solvated in a (roughly) cubic box with box length 4.28, 4.33, and 4.33 nm in the x, y and z dimensions, respectively. The system provided had a total of 7,976 atoms: 2,586 water molecules to solvate the system, 12 sodium and 3 chloride ions to neutralize the system, and the remaining atoms belonging to either the host or the guest. Forcefield parameters for the system are as provided by the original organizers of the SAMPLing challenge (Rizzi et al., 2018). The system was parameterized using GAFF (Wang et al., 2004) and then converted into Gromacs format. The conversion was done using ParmEd version 2.7.3. OpenMM v7.2.1 (Eastman et al., 2017) was used to run dynamics with the CUDA v9.0.176 platform. A Monte Carlo barostat is used to maintain a constant pressure of 1 atm. A timestep of 2 fs was used across all simulations.

We utilize the Langevin integrator, which uses a drag term and a noise term to account for the friction of solvent molecules and high velocity collisions that perturb the system. Langevin dynamics allows for the temperature to be controlled and can be used as a thermostat; we run all dynamics here at 300 K. Our host-guest system is propagated with the Langevin equation, shown below:

$$\boldsymbol{F} = m\boldsymbol{a} = -\nabla U(\boldsymbol{r}) - m\gamma\boldsymbol{v} + \sqrt{\frac{2m\gamma k_B T}{\tau}}\boldsymbol{R}(t) \tag{2}$$

where $U(\boldsymbol{r})$ is the particle interaction potential, $\boldsymbol{R}(t)$ is a random Gaussian noise term evaluated every timestep, T is the temperature, k_B is the Boltzmann constant, τ is the timestep and γ is the friction coefficient in units of inverse time. The friction term plays two different roles here, both modulating the second "drag" term, and the Gaussian noise. As γ approaches zero, the noise gets weaker and the dynamics becomes more deterministic. Here we run binding and unbinding simulations with γ values of 1.0, 0.1, and 0.01 ps^{-1}.

2.3. Reweighting of Ensembles by Variance Optimization

To generate an ensemble of ligand unbinding events, we need to employ enhanced sampling as the timescale of ligand unbinding events in this system is prohibitively long: we found in previous studies a mean first passage time of 2.1 s (Dixon et al., 2018), which is six orders of magnitude longer than the reach of conventional MD sampling. In this work, we implement the REVO resampling method, based on weighted ensemble (WE) framework, to encourage the sampling of rare unbinding/rebinding events. WE accelerates the sampling of rare events using an ensemble of trajectories that are each assigned a statistical weight (Huber and Kim, 1996). The ensemble is integrated forward in time in a parallel fashion, and periodically "resampled" by cloning certain trajectories and merging others. When a trajectory is cloned, its weight is divided amongst the clones, but the multiple copies of the trajectory go on to evolve independently. By repeatedly cloning trajectories that are in undersampled regions of space we can obtain statistics on very long-timescale events using only short-timescale simulations.

The REVO resampling method (Resampling Ensembles by Variation Optimization) was designed to efficiently perform cloning and merging operations on small ensembles of trajectories that are evolving in high-dimensional spaces (Donyapour et al., 2019). This is valuable in situations where it is difficult to define one or two progress variables that capture the long-timescale events of interest. In REVO, coupled cloning and merging operations are proposed (e.g., clone trajectory i, and merge trajectories j and k) and are accepted or rejected based on an objective function called the "trajectory variation":

$$V = \sum_i V_i = \sum_i \sum_j (d_{ij}/d_0)^\alpha \phi_i \phi_j \tag{3}$$

where d_{ij} is the distance between trajectories i and j, α and d_0 are parameters, and ϕ_x is a function that measures the importance, or "novelty" of a trajectory x, which in our work here is strictly a function of the weight of the trajectory: $\phi_i = \log w_i - C$, where w_i is the weight of trajectory i and C is a constant. Trajectories with the highest V_i values in Equation (3) are chosen for cloning, and those with the lowest V_i are chosen for merging. More information about the algorithm can be found in previous work (Donyapour et al., 2019).

We run separate simulations for the binding and unbinding processes. In our unbinding simulations, the ligands start in the bound state and are terminated as they unbind. In the rebinding simulations, the ligands start in the unbound state and are terminated as they bind. The distance function (d_{ij}) we use in Equation (3) is different for these two simulation types. For the unbinding simulations, we superimpose the hosts from trajectories i and j, and then compute the root mean squared distance (RMSD) between the guest molecules, without any further alignment (Dickson and Lotz, 2016, 2017),. As there is 4-fold symmetry in this system, we perform the alignment four times (once for each symmetrically-equivalent mapping) and use the smallest such distance as d_{ij}. For the rebinding simulations, we calculate the distance to the native state for each trajectory

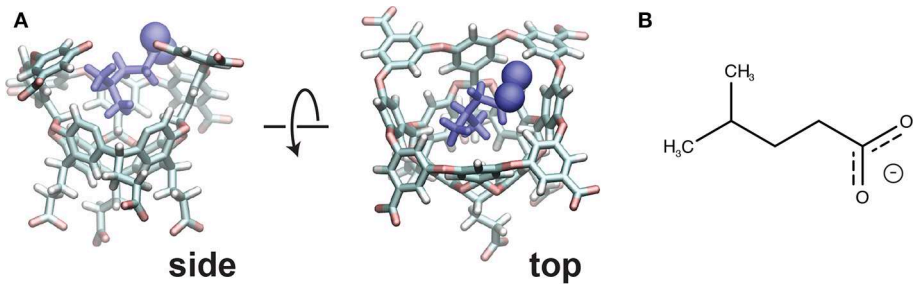


FIGURE 1 | (A) The initial pose for the OA-G6 system (side view: left, top view: right). Note that some atoms from the host are removed in the side view for clarity. The carboxyl oxygens are shown in sphere representation. **(B)** The chemical structure of the G6 ligand in the deprotonated form.

($d_{native}(\boldsymbol{X}_i)$), which again takes into account the four symmetry mappings, using the lowest such distance. The distance between trajectories i and j is then calculated as $d_{ij} = |1/d_{native}(\boldsymbol{X}_i) - 1/d_{native}(\boldsymbol{X}_j)|$, where the inverse is used to prioritize differences between small values of d_{native}.

2.4. Calculating Rates by Ensemble Splitting

A major advantage of the REVO method, much like other weighted ensemble methods, is that it can calculate kinetic parameters in real time as the simulation progresses. This is achieved by running separate simulations for the binding and unbinding processes, and in each case, measuring the trajectory flux into the opposite basin (Dickson et al., 2009, 2011; Vanden-Eijnden and Venturoli, 2009; Costaouec et al., 2013; Suárez et al., 2014). The unbound basin is defined as the set of structures where the closest host-guest interatomic distance is $>$ 1 nm, following previous work (Dickson and Lotz, 2016, 2017; Lotz and Dickson, 2018). The bound basin is defined as the set of structures where the guest RMSD (compared to the native structure) is $<$ 0.1 nm after aligning to the host. Again, this RMSD measurement takes into account the four symmetry-equivalent mappings of OA.

In our studies, the binding and rebinding REVO simulations are conducted separately. However, the methodology of obtaining on and off rates is essentially the same. After each dynamics step, if a walker has entered the opposite basin, as described above, its weight is recorded and its structure is "warped" back to the starting structure at the beginning of the simulation. The atomic coordinates are set to the starting structure and the velocities are reinitialized; however, the weight of the trajectory remains the same. Before the warping event to the starting structure, the structure of the walker is recorded and is referred to as an "exit point." In our unbinding simulations, the initial starting structure is the initial bound pose provided. In our rebinding simulations, the initial starting structure is chosen from a set of exit points generated from the unbinding simulations. Therefore, the unbinding analyses were performed prior to initialization and the subsequent running of our rebinding simulations.

The off and on rates are calculated by using the flux of trajectories into either the unbound or bound state, respectively.

$$k_{\text{off}}(t) = \frac{\sum_i w_i}{t} \tag{4}$$

$$k_{\text{on}}(t) = \frac{\sum_i w_i}{Ct} \tag{5}$$

where the sum is over the set of "warped" trajectories, t is the elapsed simulation time, and C is the concentration of ligand, computed as $1/V$ where V is the box volume. The box volume was approximately 80.2 nm^3, corresponding to a concentration of ligand of 0.0207 M.

There are a few key differences between the REVO simulations discussed here and our previous studies (Dixon et al., 2018). For both the unbinding and rebinding simulations in this study, the total simulation time is 2.25 times longer compared to our previous study, as our current unbinding and rebinding simulations were run for 4,500 and 450 cycles, respectively. Additionally, ten independent unbinding simulations were run for each of the four friction coefficients, whereas our previous study only ran five independent simulations for each starting pose. However, only five independent rebinding simulations were run for each of the coefficients, as we observe much less variation in the k_{on} estimates. Finally, 48 walkers were used in both studies and the time per cycle is consistent, where the unbinding simulations are 20 ps/cycle and the rebinding simulations are 200 ps/cycle.

2.5. Calculating Electrostatic Interaction Energies

The electrostatic energy between the host and guest molecules for use in the second correction term was calculated as: $E_{\text{int}} = \frac{1}{4\pi\epsilon_w}\frac{Q_iQ_j}{r_{ij}}$ where Q_a is the partial charge of atom a used in the force field during simulation. r_{ij} is the interatomic distance between atoms i and j calculated by using the minimum image convention. $\epsilon_w = 6.88 \times 10^{-10}$ F/m is the permittivity of water at 300 K calculated by linear interpolation of the water dielectric constant at 298.15 and 303.15 K (Archer and Wang, 1990).

3. RESULTS

3.1. Derivation of Correction Terms

The binding free energy can be calculated using the rate constants k_{on} and k_{off} as $\Delta G = G_{\text{bound}} - G_{\text{unbound}} = -kT\ln K_{eq}C_0 = -kT\ln\frac{C_0k_{\text{on}}}{k_{\text{off}}}$, where K_{eq} is the binding equilibrium constant, C_0 is the reference concentration of 1 mol/L, k is Boltzmann's constant and T is the temperature in Kelvin. While this relationship is correct in the macroscopic limit, it fails to account for the box

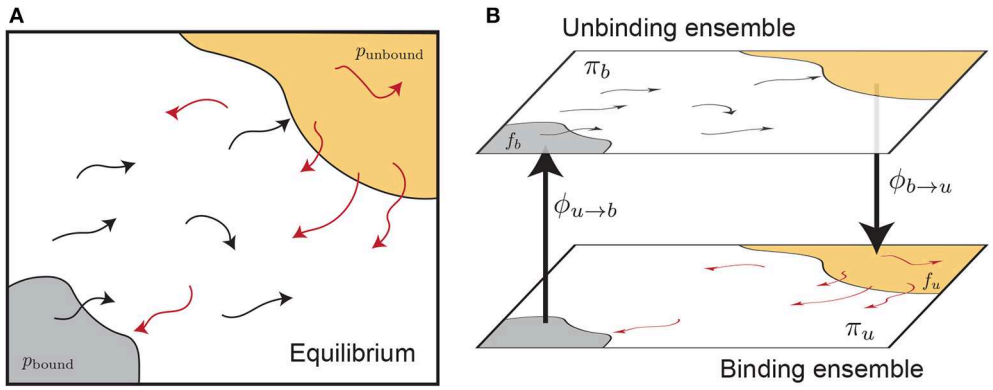


FIGURE 2 | Splitting an equilibrium ensemble **(A)** into two history-dependent ensembles using basins. The bound and unbound basins are shown in gray and light orange, respectively. The unbinding ensemble (**B**, top) contains all trajectories that last visited the bound basin, which are shown in black. The binding ensemble (**B**, bottom, also referred to as the "rebinding" ensemble) contains all trajectories that last visited the unbound basin, shown in red. Simulations in a given ensemble are terminated once they reach the destination basin and thus switch ensembles. The trajectory flux between ensembles is denoted by $\phi_{u\to b}$ and $\phi_{b\to u}$. The quantity π_b refers to the probability of the entire top ensemble, and the quantity f_b denotes the probability of the bound basin within the unbinding ensemble.

size and the volume of the unbound state in finite simulation environments with periodic boundary conditions. Here we derive a more accurate expression for the binding free energy that accounts for the finite box size in a typical MD simulation.

Our starting point is an expression for K_{eq}, which is valid for a dilute solution in thermodynamic equilibrium. We use the notation of Woo and Roux (see Equation 4 from Woo and Roux, 2005):

$$K_{eq} = \frac{\int_{\text{bound}} d\mathbf{1} \int d\mathbf{X} e^{-\beta U}}{\int_{\text{bulk}} d\mathbf{1} \delta(\mathbf{r}_1 - \mathbf{r}_1^*) \int d\mathbf{X} e^{-\beta U}} \tag{6}$$

where U is the internal energy of the system, $\beta = 1/kT$ is the inverse temperature, $\mathbf{r}_1$ is the center of mass of the ligand (referred to as a "guest" molecule) and $\mathbf{r}_1^*$ is an arbitrary position of the guest in the bulk. Note that $d\mathbf{1}$ integrates over the guest positions, and $d\mathbf{X}$ integrates over everything else: the host and the solvent degrees of freedom. Note also that K_{eq} has units of volume, as the delta function constraining the center of mass in the denominator removes three spatial degrees of freedom.

Here we examine the calculation of free energies using rates determined from split ensemble calculations (**Figure 2**, see section 2.4 for more details). We denote the probability of these two ensembles as π_b and π_u, where $\pi_b + \pi_u = 1$, and:

$$\frac{\pi_b}{\pi_u} = \frac{\phi_{u\to b}}{\phi_{b\to u}} \tag{7}$$

where $\phi_{a\to b}$ is the time-averaged flux from the a ensemble to the b ensemble (i.e., across the dotted lines in **Figure 2**). The equilibrium probability of a position $\mathbf{X}$ can be obtained by combining estimates from both ensembles:

$$p(\mathbf{X}) = p_u(\mathbf{X})\pi_u + p_b(\mathbf{X})\pi_b \tag{8}$$

where $p_a(\mathbf{X})$ is the probability of conformation $\mathbf{X}$ in ensemble a, which is normalized such that $\int p_a(\mathbf{X}) d\mathbf{X} = 1$.

Let us define the bound state as the domain of the integral in the numerator of Equation (6), and the unbound state as a set of structures considered unbound in simulation (not the same as the bulk state in Equation 6). These states are shown as shaded regions in **Figure 2**. The ratio of the probabilities of these two states, at equilibrium, is given by:

$$\frac{p_{\text{bound}}}{p_{\text{unbound}}} = \frac{\int_{\text{bound}} d\mathbf{1} \int d\mathbf{X} e^{-\beta U}}{\int_{\text{unbound}} d\mathbf{1} \int d\mathbf{X} e^{-\beta U}} \tag{9}$$

which can also be calculated in our ensemble splitting simulations:

$$\frac{p_{\text{bound}}}{p_{\text{unbound}}} = \frac{\pi_b \int_{\text{bound}} p_b(\mathbf{X}) d\mathbf{X}}{\pi_u \int_{\text{unbound}} p_u(\mathbf{X}) d\mathbf{X}} = \frac{\pi_b f_b}{\pi_u f_u} \tag{10}$$

where f_a is the probability of the basin state within ensemble a.

Expanding Equation (6) we have:

$$\begin{aligned} K_{eq} &= \frac{\int_{\text{bound}} d\mathbf{1} \int d\mathbf{X} e^{-\beta U}}{\int_{\text{unbound}} d\mathbf{1} \int d\mathbf{X} e^{-\beta U}} \frac{\int_{\text{unbound}} d\mathbf{1} \int d\mathbf{X} e^{-\beta U}}{\int_{\text{bulk}} d\mathbf{1} \delta(\mathbf{r}_1 - \mathbf{r}_1^*) \int d\mathbf{X} e^{-\beta U}} \\ &= \frac{\pi_b f_b}{\pi_u f_u} \frac{\int_{\text{unbound}} d\mathbf{1} \int d\mathbf{X} e^{-\beta U}}{\int_{\text{bulk}} d\mathbf{1} \delta(\mathbf{r}_1 - \mathbf{r}_1^*) \int d\mathbf{X} e^{-\beta U}}. \end{aligned} \tag{11}$$

The unbound state in simulation is far enough that the host and guest do not interact directly through van der Waals interactions, although if both molecules carry an explicit charge—as in the example considered here—there could still be significant host-guest electrostatic interactions. To account for these, we introduce another intermediate state with an altered energy function (U^*) which is the same as U except that it does not

include electrostatic interactions between the host and the guest:

$$K_{eq} = \frac{\pi_b f_b}{\pi_u f_u} \frac{\int_{\text{unbound}} d\mathbf{1} \int d\mathbf{X} e^{-\beta U}}{\int_{\text{unbound}} d\mathbf{1} \int d\mathbf{X} e^{-\beta U^*}} \frac{\int_{\text{unbound}} d\mathbf{1} \int d\mathbf{X} e^{-\beta U^*}}{\int_{\text{bulk}} d\mathbf{1} \delta(\mathbf{r}_1 - \mathbf{r}_1^*) \int d\mathbf{X} e^{-\beta U}} \tag{12}$$

$$= \frac{\pi_b f_b}{\pi_u f_u} \left\langle e^{\beta E_{\text{int}}} \right\rangle_{\text{unb}}^{-1} \frac{\int_{\text{unbound}} d\mathbf{1} \int d\mathbf{X} e^{-\beta U^*}}{\int_{\text{bulk}} d\mathbf{1} \delta(\mathbf{r}_1 - \mathbf{r}_1^*) \int d\mathbf{X} e^{-\beta U}} \tag{13}$$

where $E_{\text{int}} = U - U^*$ and the subscript "unb" indicates an ensemble average over structures in the unbound state obtained with the normal energy function U. Note the final step used the relation:

$$\frac{\int_{\text{unbound}} d\mathbf{1} \int d\mathbf{X} e^{-\beta U^*}}{\int_{\text{unbound}} d\mathbf{1} \int d\mathbf{X} e^{-\beta U}} = \frac{\int_{\text{unbound}} d\mathbf{1} \int d\mathbf{X} e^{\beta E_{\text{int}}} e^{-\beta U}}{\int_{\text{unbound}} d\mathbf{1} \int d\mathbf{X} e^{-\beta U}} = \left\langle e^{\beta E_{\text{int}}} \right\rangle_{\text{unb}}. \tag{14}$$

We can now reasonably assume that the guest in the unbound state is non-interacting with the host. This allows us to write $e^{-\beta U}$ as $e^{-\beta U_G} e^{-\beta U_{HS}}$, where U_G are the terms in the energy function that depend only on the coordinates of the guest, and U_{HS} are terms that only depend on the host and the solvent. We can then pull the integral $\int d\mathbf{X} e^{-\beta U_{HS}}$ out of the numerator and denominator of the last term of Equation (11):

$$\frac{\int_{\text{unbound}} d\mathbf{1} \int d\mathbf{X} e^{-\beta U^*}}{\int_{\text{bulk}} d\mathbf{1} \delta(\mathbf{r}_1 - \mathbf{r}_1^*) \int d\mathbf{X} e^{-\beta U}} = \frac{\int_{\text{unbound}} d\mathbf{1} e^{-\beta U_G}}{\int_{\text{bulk}} d\mathbf{1} \delta(\mathbf{r}_1 - \mathbf{r}_1^*) e^{-\beta U_G}}. \tag{15}$$

The bottom integral has the center of mass of the ligand fixed and is only over internal and rotational degrees of freedom of the ligand. This can also be separated and removed from the numerator, which simplifies the ratio to be the volume of the unbound state, defined as:

$$V_{\text{unbound}} = \frac{\int_{\text{unbound}} d\mathbf{1} e^{-\beta U_G}}{\int_{\text{guest}} d\mathbf{G}_1 e^{-\beta U_G}} = \int_{\text{box}} d\boldsymbol{R} \phi_u(\boldsymbol{R}) \tag{16}$$

where we use $\mathbf{G}_1$ to denote the internal and rotational degrees of freedom of the guest that remain after specification of $\mathbf{r}_1$. The quantity $\phi_u(\mathbf{R})$ is the fraction of conformers with center of mass $\mathbf{R}$ that satisfy the unbound boundary conditions: here, that the guest atoms are all farther than a cutoff distance of 1 nm away from the host. This integral can be calculated by Monte Carlo, where a center of mass position and orientation of the ligand is randomly generated, and the number of successful unbound conformers is recorded:

$$V_{\text{unbound}} = V_{\text{box}} \frac{N_{\text{unbound}}}{N_{\text{trials}}}. \tag{17}$$

Note that for large boxes $V_{\text{unbound}} \approx V_{\text{box}}$.

Putting this all together we have:

$$K_{eq} = \frac{\pi_b f_b}{\pi_u f_u} \left\langle e^{\beta E_{\text{int}}} \right\rangle_{\text{unb}}^{-1} V_{\text{unbound}}, \tag{18}$$

which differs from the straightforward interpretation used in our previous work (Dixon et al., 2018):

$$K_{eq}^0 = \frac{\pi_b}{\pi_u [L]} = \frac{\pi_b}{\pi_u} V_{\text{box}} \tag{19}$$

Using $\Delta G = -kT \ln(K_{eq} C_0)$, we have:

$$\Delta G = \Delta G^0 - kT \ln\left(\frac{f_b}{f_u}\right) + kT \ln \left\langle e^{\beta E_{\text{int}}} \right\rangle_{\text{unb}} - kT \ln\left(\frac{V_{\text{unbound}}}{V_{\text{box}}}\right) \tag{20}$$

which explicitly shows ΔG as the sum of $\Delta G^0 = -kT \ln(K_{eq}^0 C_0)$ and the three newly derived correction terms. The first term will go to zero in the limit that the basin states are chosen to represent the vast majority of the probability in both the binding and unbinding ensembles. In other words, this term goes to zero when both f_b and f_u approach one. The second term is likely to only be non-negligible in the case of explicitly charged host and guest molecules and regardless would go to zero as the definition of the unbound state is moved to farther and farther distances. The third term would also go to zero for large simulation boxes, but in practice this is often not feasible due to computational constraints. Consequently, $V_{\text{unbound}}/V_{\text{box}}$ could be much less than one, introducing a correction in the positive direction. Below we calculate these three correction terms and apply them to free energy calculations.

3.2. Extended Trajectory Ensembles With Lower Friction Coefficients

In previous work, we used a Langevin integrator with a value of $\gamma = 1$ ps^{-1} for the friction coefficient. As the simulations already have explicit solvent, this adds extra friction into the system that is not physical. Here we investigate whether reducing γ to values less than one will significantly affect our rate calculations. We thus run a set of trajectory ensembles at multiple values of γ and extend each ensemble to be larger and longer than those published in our prior study (Dixon et al., 2018) to more fully examine questions of convergence.

As γ governs the coupling to the Langevin thermostat, we determine the minimum value of γ where our target temperature (300 K) is maintained. We first ran a series of short simulations (one 10 ns trajectory for each γ) and find that temperature control is completely lost for friction coefficients less than $\gamma = 0.001$ (**Figure 3A**). We then ran longer simulations for $\gamma = 1$, 0.1, 0.01, and 0.001, examining not only the mean temperature, but the probability of significant temperature fluctuations, which could spur anomalous results in our ligand dissociation simulations. **Figure 3B** shows the probability distribution of observed temperatures over an ensemble of 240 trajectories run for 90 ns each. For $\gamma = 0.01$, 0.1 and 1 ps^{-1}, the temperature distribution is normally distributed around the mean (300 K) as seen by the parabolic curves on a log scale. Temperature control is not fully maintained for $\gamma = 0.001$ ps^{-1}, as shown by a rightward shift and slight widening of the parabolic distribution. We thus

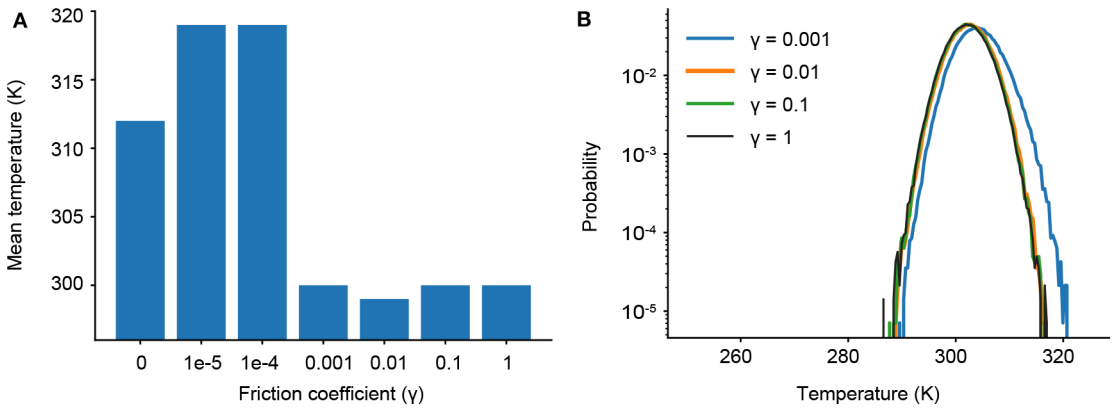


FIGURE 3 | (A) Average temperatures observed in short simulations for different friction coefficients (γ). **(B)** Probability distributions of observed temperatures from ensembles of longer simulations with different γ.

restrict our analysis to three values of the friction coefficient: $\gamma = 0.01, 0.1$, and 1 ps^{-1}.

We run both unbinding and rebinding REVO simulations for the OAG6 system. For unbinding, we ran 10 simulations for each of the three friction coefficients; for rebinding, we ran five simulations for each coefficient, yielding a total of 30 simulations for unbinding and 15 simulations for rebinding. A set of binding and unbinding simulations were also run for $\gamma = 0.001$—despite the impaired temperature control—which are reported in the **Supplemental Information**. The estimates for the unbinding and binding fluxes are depicted in **Figure 4**, where each curve represents an individual REVO simulation. The averages, illustrated with a bolded line, are calculated by averaging the trajectory flux over the entire set of simulations for that value of γ. The upward jumps on these plots indicate that an exit point was recorded that has a higher weight than was previously observed.

By reducing γ to values <1, we observed no change in the binding rates, and small changes to the unbinding rates which are on the border of significance. With regard to unbinding rates, the two largest friction coefficients yielded the smallest error and similar k_{off} values, where $\gamma = 1$ yielded an average off rate of 16.4 s^{-1} and $\gamma = 0.1$ yielded an off rate of 11.5 s^{-1}. The off-rate increased by 10-fold for $\gamma = 0.01$, although this is mostly driven by exit points observed in a single simulation. In our previous OA-G6 results using $\gamma = 1$, we calculated an unbinding rate of 0.48 s^{-1} which slightly differs from the value calculated in this study using $\gamma = 1$ (**Table 1**). Unbinding rates for $\gamma = 0.001$ ps^{-1} were approximately 1000-fold higher, although these are known to be affected by a higher average temperature (**Supplemental Information**). Taking a closer look at the binding rates, we saw no discernible difference across the friction coefficients. The binding rate was approximately 10^9 s^{-1} M^{-1}, for all friction coefficients, which was about 5-fold larger when compared to our previous study using $\gamma = 1$. For both binding and unbinding rates we have more confidence in the results obtained here, as they are based on more extensive simulation data.

TABLE 1 | Binding and unbinding rates as a function of friction coefficient (γ).

	k_{on} (10^8 M^{-1} s^{-1})	k_{off} (s^{-1})
$\gamma = 0.01$	17 ± 1	122 ± 94
$\gamma = 0.1$	16 ± 2	22 ± 12
$\gamma = 1$	13 ± 1	16.4 ± 9.4
Dixon et al. (2018) ($\gamma = 1$)	2.8 ± 1.0	0.48 ± 0.11

The uncertainties shown use the standard error of the mean calculated from 5 and 10 independent REVO runs for binding and unbinding, respectively. The quantities from Dixon et al. (2018) were obtained with 5 REVO runs that used different initial conformations, each of which were 2,000 cycles in length.

For both the unbinding and rebinding simulations, across all friction coefficients, we observed at least 1,000 warping events (**Figure S4**). As expected, we observe that rebinding occurs with a much higher probability when compared to unbinding, by several orders of magnitude. The unbinding walker weights are limited at the low end by the minimum walker probability (p_{min}), which is set to 10^{-12}. The rebinding walker weights are limited at the high end by the maximum walker probability (p_{max}), which is set to 10^{-1}. respectively. **Figure S4** shows that the 10-fold larger unbinding rate fro $\gamma = 0.01$ was largely due to a single unbinding point in a single simulation, which underscores the sensitivity and uncertainty of rate calculations using trajectory fluxes. **Figure S2** shows unbinding fluxes for $\gamma = 0.001$, which is known to have elevated temperatures. There we see a large number of high-weight unbinding events in two different simulations, leading to the 1,000-fold increase in k_{off}.

3.3. Free Energy Estimates, Correction Terms, and Comparison With Previous Benchmarks

As the friction coefficient unevenly affected the rates of binding and unbinding, there was a net effect on the binding free energies. As shown in **Figure 5** and **Table 2**, the binding free energy increases as the friction coefficient is lowered, independent of the

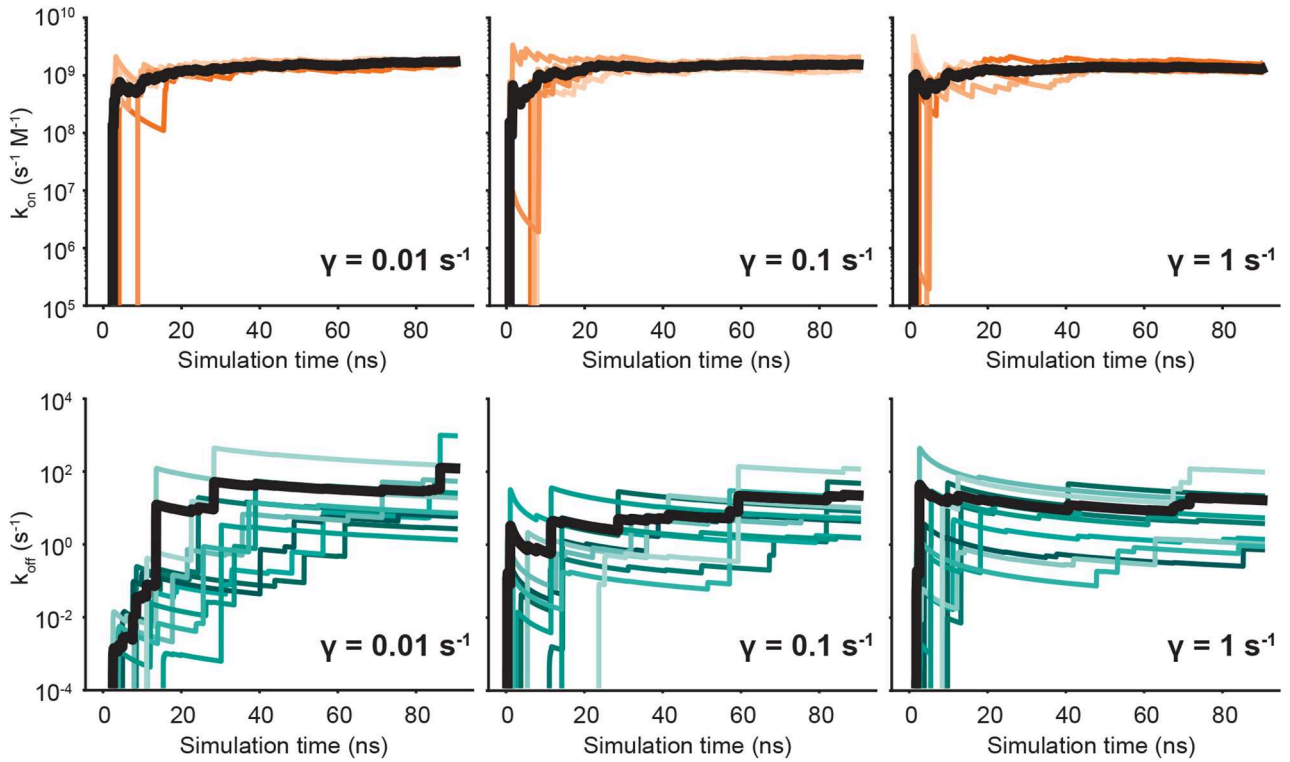


FIGURE 4 | Predicted on- **(top)** and off-rates **(bottom)** as a function of simulation time. Each panel is labeled according to the friction coefficient used for that set of simulations. The independent simulations are shown in shades of orange (k_{on}) and blue (k_{off}), and the averages are depicted by bold black lines.

free energy correction terms derived in section 3.1. **Table 2** shows the free energies computed using the averaged fluxes across all simulations at each γ value. For all friction coefficients, the calculated free energy was always higher than that from our previous study (−12.1 kcal/mol; red line), even for $\gamma = 1$, signifying that extending the simulation time aided in predicting experimentally determined binding free energies.

The correction terms are calculated using data obtained from the simulations, but they are mostly functions of geometric properties of the simulation box and boundary conditions, and are not expected to change as a function of γ. The first term, $-kT \ln f_b/f_u$, was calculated to be 0.74 ± 0.10 kcal/mol, with f_b and f_u taking on values of 0.157 and 0.54, respectively. As described in section 3.1, f_b is the probability of the being in the bound basin given that you are in the unbinding ensemble, which is calculated using the sum of the weights of trajectories in the bound basin, divided by the total sum of the weights of the trajectories considered. The f_b value in particular was lower than expected, indicating that our definition of the bound state might be too restrictive, even though we did account for all symmetry-equivalent conformations in our calculation of f_b.

The second term, $+kT \ln \left\langle e^{\beta E_{\text{int}}} \right\rangle_{\text{unb}}$, was calculated to be 1.64 ± 0.002 kcal/mol. This was calculated by determining the electrostatic interaction energies (see section 2.5) for the set of unbound states observed in the rebinding simulations. The expectation value in the correction term again accounted for trajectory weights and was computed using 71,428 interaction energy measurements that were selected from the unbound ensemble. The uncertainty was computed as the standard error of the mean of this set of energies. To calculate the third correction term, $-kT \ln \left(\frac{V_{\text{unbound}}}{V_{\text{box}}} \right)$, we directly estimated $V_{\text{unbound}}/V_{\text{box}}$ using the Monte Carlo procedure described in section 3.1. The ratio was computed as 0.56 ± 0.0037 using five batches of 10,000 trials each, where the uncertainty is the standard error of the mean across the sets of trials.

Together these three terms sum to 2.72 kcal/mol, which is a significant correction to the binding free energies computed here. Over half of this comes from the residual electrostatic interaction energy between the host and the guest. Note that both the host and the guest have negative charges, and the residual interaction between the two molecules is repulsive. Turning this interaction off *releases* 1.64 kcal/mol of energy, which lowers the free energy gap between the bound and unbound states. The corrected and uncorrected free energies are shown as a function of γ in **Figure 5**. For $\gamma \geq 0.01$ the calculated free energies are almost equal to within standard error and the correction terms significantly reduce the error with respect to the computational reference value (Rizzi et al., 2018, 2020).

4. DISCUSSION AND CONCLUSION

In this study, we sought to better connect the calculation of binding and unbinding rates with the calculation of binding

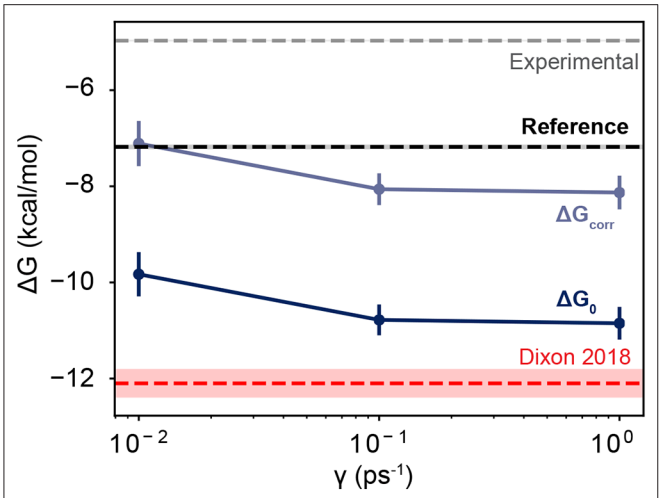


FIGURE 5 | Free energies as a function of friction coefficient. The dark blue line shows the uncorrected free energies calculated at three different γ values. The light blue line shows the corrected values, which are shifted upwards by 2.72 kcal/mol. The thin red line shows the value reported in Dixon et al. (2018), which employed a friction coefficient of 1.0 ps^{-1} and used a smaller dataset than is reported here. The black horizontal line shows the value of a computational reference computed using alchemical perturbation, reported in Rizzi et al. (2020). The dashed gray line shows the experimental measurement, reported in Sullivan et al. (2019). The shaded area for each line shows its associated uncertainty, which is less than the line thickness for the computational reference and the experimental measurement.

TABLE 2 | Raw (ΔG^0) and corrected (ΔG_{corr}) free energy values using simulation data from three different friction coefficients.

	ΔG^0 (kcal/mol)	ΔG_{corr} (kcal/mol)
$\gamma = 0.01$	-9.83 ± 0.46	-7.11 ± 0.47
$\gamma = 0.1$	-10.78 ± 0.32	-8.06 ± 0.33
$\gamma = 1$	-10.85 ± 0.34	-8.13 ± 0.36
Dixon et al. (2018) ($\gamma = 1$)	-12.1 ± 1.0	-9.38 ± 1.0
Comp. (Rizzi et al., 2020)	–	-7.0 ± 0.1
Exp. (Sullivan et al., 2019)	–	-4.97 ± 0.02

Values are in kcal/mol and uncertainties are calculated using propagation of the standard error of the mean.

free energies. The rate calculations measured the microscopic fluxes of trajectories from one basin to another. These fluxes can be visualized in an extended history-dependent conformation space, where trajectories change their "color" based on which basin ("bound" or "unbound") they have most recently visited (Dickson et al., 2009, 2011; Vanden-Eijnden and Venturoli, 2009; Costaouec et al., 2013; Suárez et al., 2014). The ratio of these rates gives a ratio of two populations: the trajectories that have most recently visited the "bound" basin and the trajectories that have most recently visited the "unbound" basin. The first correction term adjusts this ratio to instead only account for the probability contained within the basins themselves and is particular to rates that are calculated using this history-dependent formalism. The third term can be seen as a volume correction term, which is used to accurately account for the volume in the unbound state. This is done in other approaches where restraints are used, such as methods based on calculation of the potential of mean force (Deng and Roux, 2009). In our case the unbound state cannot be easily approximated by a geometric object, such as the volume of a spherical shell.

The second term accounts for residual interactions in the unbound ensemble. This could be used by other approaches that directly determine free energy differences between bound and unbound conformations, such as Markov state modeling, metadynamics, milestoning, and umbrella sampling. The conventional approach is to define a simulation box that is large enough such that the interactions between the host and guest are negligible in the unbound state. However, this can significantly increase the cost of the simulation. It is worth noting that umbrella sampling results for this system (OA-G6) obtained by Song et al. (2018), −8.50 kcal/mol, were also below both the computational benchmark and the experimental value. Their unbound state was defined as a 20 Å distance between an atom in the guest and a dummy atom in the center of the host, which is roughly comparable to our unbound basin of 10 Å of clearance between the host and the guest. Assuming a similar value for the electrostatic correction term, it would have brought their prediction to −6.86 kcal/mol, which is in line with the computational benchmarks (Rizzi et al., 2020).

The electrostatic term can also be viewed as a sort of "decoupling" between the host and the guest, and it is warranted to discuss similarities and differences with similar procedures in alchemical free energy methods. They are similar in that we are computing a free energy between two Hamiltonians, one in which an interaction is turned off. We could thus use similar techniques for computing these free energy differences, such as thermodynamic integration (Kirkwood, 1935; Bhati et al., 2019), BAR (Gutiérrez et al., 2019), MBAR (Shirts and Chodera, 2008; Bhati et al., 2019), or MM/PBSA (Rifai et al., 2019), although here we effectively use a simple free energy perturbation (FEP) expression (Zwanzig, 1954; Jorgensen and Thomas, 2008). The approaches are different in that we are only considering ensembles of structures where the interactions being turned off are relatively weak. We are assuming here—as is always the case with FEP—that the conformational ensembles of both the host and the guest are highly overlapping between the two Hamiltonians, which considerably simplifies the problem. We also note that although we employ electrostatic decoupling to compute free energies, our simulations still reveal important information about the (un)binding kinetics and mechanism.

Given these correction terms for the binding affinity, it is reasonable to ask if and how the rate constants should be modified. The correction terms each have the effect of "loosening" the interaction, indicating that either the corrected off-rate should increase, or the corrected on-rate should decrease. It is reasonable to assume that lowering the on-rate should account for the vast majority of this correction, as the on-rate measured starts from an unbound conformation that is much closer (a clearance of 1 nm between the host and guest) than is likely in experimental conditions. More accurate calculations of the binding rate can be achieved with better sampling of the unbound state, for instance using the Northrup-Allison-McCammon method (Northrup et al., 1984). It would be

interesting to see whether such calculations can recapitulate part of the free energy differences observed here.

We also examined the role that the Langevin integrator plays in the prediction of kinetic and thermodynamic quantities. In particular, we adjusted the friction coefficient (γ), defined in the Langevin integrator, while maintaining the stability of temperature at 300 K. We did not expect that altering the friction coefficient would have an impact on the calculation of equilibrium quantities. As γ does not appear in the Hamiltonian of the system, it should not affect the probability of a given microstate $P(\mathbf{X})$, which is given by the Canonical probability density $\exp(-\beta U(\mathbf{X}))$. While we did expect it to affect rates, we expected that these effects would offset: that if unbinding was accelerated 10-fold, we would observe the binding process to be sped up by the same factor. However, we observe that the on-rate was very stable as a function of γ, while the off-rate changed slightly. One explanation is that unbinding is a much more rare event when compared to rebinding, and estimates of k_{off} were not converged. Lower friction coefficients could be accelerating sampling of these events and making it easier to observe higher probability walkers unbind in our simulations.

Convergence is of utmost priority in weighted ensemble simulations that calculate kinetic quantities. In our previous study, we hypothesized that it was possible that extending the time of the unbinding simulations could capture more high weight walkers exiting from the bound state. Indeed, we observe a higher unbinding flux in this study across all friction coefficients. In **Figure 4**, we observe large upward jumps, for all γ values, even after 40 ns of simulation time per walker, which was sampling limit in our previous study. These upward jumps, as previously described, signify that an exit point was recorded that has a higher weight than previously observed. This highlights the challenges involved in accurate determination of rate fluxes for rare events. It is worth noting that by using our correction terms to account for small unbound volumes and persistent but small electrostatic interactions in the unbound state, we can keep box sizes small, allowing for better convergence of rate fluxes at fixed computational cost.

Of course the binding free energy alone is still an important quantity for drug design (Homeyer et al., 2014). If one is only interested in the absolute binding free energy, calculating it through the ratio of rates is needlessly complicated; free energy is a state function and thus only depends on the endpoints of the binding pathway. The prediction of k_{off} and k_{on} themselves is challenging, since they are not state functions: they depend on the transition path ensemble between the bound and unbound state. Sampling of these physical pathways is a large challenge for molecular dynamics, largely due to the long timescales of the binding and release processes. Ensuring that the ratio of rates is consistent with binding free energy calculations—as done here—provides an additional, powerful consistency check. In particular, comparing to well-converged computational benchmarks is more useful than experimental quantities, as we avoid an additional layer of uncertainty associated with the force field used to describe the system.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

AUTHOR CONTRIBUTIONS

AD designed the project. RH, TD, and AD conducted the research, analysis, and wrote the manuscript.

FUNDING

This work was funded by the National Institutes of Health (R01GM130794) and the National Science Foundation (DMS 1761320).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2020.00106/full#supplementary-material

REFERENCES

Archer, D., and Wang, P. (1990). The dielectric constant of water and debye-hückel limiting law slopes. *J. Phys. Chem. Refer. Data* 19, 371–411. doi: 10.1063/1.555853

Ayaz, P., Andres, D., Kwiatkowski, D. A., Kolbe, C. C., Lienau, P., Siemeister, G., et al. (2016). Conformational adaption may explain the slow dissociation kinetics of roniciclib (BAY 1000394), a type I CDK inhibitor with kinetic selectivity for CDK2 and CDK9. *ACS Chem. Biol.* 11, 1710–1719. doi: 10.1021/acschembio.6b00074

Bernetti, M., Cavalli, A., and Mollica, L. (2017). Protein-Ligand (un)binding kinetics as a new paradigm for drug discovery at the crossroad between experiments and modelling. *Med. Chem. Commun.* 8, 534–550. doi: 10.1039/C6MD00581K

Bhati, A. P., Wan, S., and Coveney, P. V. (2019). Ensemble-based replica exchange alchemical free energy methods: the effect of protein mutations on inhibitor binding. *J. Chem. Theory Comput*. 15, 1265–1277. doi: 10.1021/acs.jctc.8b01118

Bruce, N. J., Ganotra, G. K., Kokh, D. B., Sadiq, S. K., and Wade, R. C. (2018). New approaches for computing ligand-receptor binding kinetics. *Curr. Opin. Struct. Biol.* 49, 1–10. doi: 10.1016/j.sbi.2017.10.001

Bruno, A., Barresi, E., Simola, N., Da Pozzo, E., Costa, B., Novellino, E., et al. (2019). Unbinding of translocator protein 18 KDa (TSPO) ligands: from *in vitro* residence time to *in vivo* efficacy via *in silico* simulations. *ACS Chem. Neurosci.* 10, 3805–3814. doi: 10.1021/acschemneuro.9b00300

Camilloni, C., and Pietrucci, F. (2018). Advanced simulation techniques for the thermodynamic and kinetic characterization of biological systems. *Adv. Phys.* 2018:1477531. doi: 10.1080/23746149.2018.1477531

Carroll, M. J., Mauldin, R. V., Gromova, A. V., Singleton, S. F., Edward, J., and Lee, A. L. (2012). Evidence for dynamics in proteins as a mechanism for ligand dissociation. *Nat. Chem. Biol.* 8, 246–252. doi: 10.1038/nchembio.769

Casasnovas, R., Limongelli, V., Tiwary, P., Carloni, P., and Parrinello, M. (2017). Unbinding kinetics of a P38 MAP kinase type II inhibitor from metadynamics simulations. *J. Am. Chem. Soc.* 139, 4780–4788. doi: 10.1021/jacs.6b12950

Copeland, R. A. (2016). The drug-target residence time model: a 10-year retrospective. *Nat. Rev. Drug Discov.* 15, 87–95. doi: 10.1038/nrd.2015.18

Costa, B., Da Pozzo, E., Giacomelli, C., Barresi, E., Taliani, S., Da Settimo, F., et al. (2016). TSPO ligand residence time: a new parameter to predict compound neurosteroidogenic efficacy. *Sci. Rep*. 6:18164. doi: 10.1038/srep18164

Costaouec, R., Feng, H., Izaguirre, J., and Darve, E. (2013). Analysis of the accelerated weighted ensemble methodology. *Discrete Cont. Dyn. Syst*. 2013, 171–181. doi: 10.3934/proc.2013.2013.171

Croll, T. I., Sammito, M. D., Kryshtafovych, A., and Read, R. J. (2019). Evaluation of template-based modeling in CASP13. *Proteins* 87, 1113–1127. doi: 10.1002/prot.25800

Deb, I., and Frank, A. T. (2019). Accelerating rare dissociative processes in biomolecules using selectively scaled MD simulations. *J. Chem. Theory Comput*. 15, 5817–5828. doi: 10.1021/acs.jctc.9b00262

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

Dickson, A. (2018). Mapping the ligand binding landscape. *Biophys. J*. 115, 1707–1719. doi: 10.1016/j.bpj.2018.09.021

Dickson, A., and Lotz, S. (2017). Multiple ligand unbinding pathways and ligand-induced destabilization revealed by WExplore. *Biophys. J*. 112, 620–629. doi: 10.1016/j.bpj.2017.01.006

Dickson, A., and Lotz, S. D. (2016). Ligand release pathways obtained with WExplore: residence times and mechanisms. *J. Phys. Chem. B* 120, 5377–5385. doi: 10.1021/acs.jpcb.6b04012

Dickson, A., Maienschein-Cline, M., Tovo-Dwyer, A., Hammond, J. R., and Dinner, A. R. (2011). Flow-dependent unfolding and refolding of an RNA by nonequilibrium umbrella sampling. *J. Chem. Theory Comput*. 7, 2710–2720. doi: 10.1021/ct200371n

Dickson, A., Tiwary, P., and Vashisth, H. (2017). Kinetics of ligand binding through advanced computational approaches: a review. *Curr. Top. Med. Chem*. 17, 2626–2641. doi: 10.2174/1568026617666170414142908

Dickson, A., Warmflash, A., and Dinner, A. R. (2009). Separating forward and backward pathways in nonequilibrium umbrella sampling. *J. Chem. Phys*. 136:154104. doi: 10.1063/1.3244561

Dixon, T., Lotz, S. D., and Dickson, A. (2018). Predicting ligand binding affinity using on and off-rates for the SAMPL6 SAMPLing challenge. *J. Comput. Aided Mol. Design* 32, 1001–1012. doi: 10.1007/s10822-018-0149-3

Dixon, T., Uyar, A., Ferguson-miller, S., and Dickson, A. (2020). Membrane-mediated ligand unbinding of the PK-11195 ligand from the translocator protein (TSPO). *bioRxiv*. doi: 10.1101/2020.01.21.914127

Donyapour, N., Roussey, N. M., and Dickson, A. (2019). REVO: resampling of ensembles by variation optimization. *J. Chem. Phys*. 150:244112. doi: 10.1063/1.5100521

Dror, R. O., Pan, A. C., Arlow, D. H., Borhani, D. W., Maragakis, P., Shan, Y., et al. (2011). Pathway and mechanism of drug binding to G-protein-coupled receptors. *Proc. Natl. Acad. Sci. U.S.A*. 108, 13118–13123. doi: 10.1073/pnas.1104614108

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

Faradjian, A. K., and Elber, R. (2004). Computing time scales from reaction coordinates by milestoning. *J. Chem. Phys*. 120:10880. doi: 10.1063/1.1738640

Gan, H., Benjamin, C. J., and Gibb, B. C. (2011). Nonmonotonic assembly of a deep-cavity cavitand. *J. Am. Chem. Soc*. 133, 4770–4773. doi: 10.1021/ja200633d

Gilson, M. K., Given, J. A., Bush, B. L., and McCammon, J. A. (1997). The statistical-thermodynamic basis for computation of binding affinities: a critical review. *Biophys. J*. 72, 1047–1069. doi: 10.1016/S0006-3495(97)78756-3

Gu, S., Silva, D. A., Meng, L., Yue, A., and Huang, X. (2014). Quantitatively characterizing the ligand binding mechanisms of choline binding protein using markov state model analysis. *PLoS Comput. Biol*. 10:e1003767. doi: 10.1371/journal.pcbi.1003767

Guo, D., Heitman, L. H., and Ijzerman, A. P. (2016). The added value of assessing ligand-receptor binding kinetics in drug discovery. *ACS Med. Chem. Lett*. 7, 819–821. doi: 10.1021/acsmedchemlett.6b00273

Guo, D., Xia, L., van Veldhoven, J. P. D., Hazeu, M., Mocking, T., Brussee, J., et al. (2014). Binding kinetics of ZM241385 derivatives at the human adenosine A 2A receptor. *ChemMedChem* 9, 752–761. doi: 10.1002/cmdc.201300474

Gutiérrez, M., Vallejos, G. A., Cortés, M. P., and Bustos, C. (2019). Bennett acceptance ratio method to calculate the binding free energy of bace1 inhibitors: theoretical model and design of new ligands of the enzyme. *Chem. Biol. Drug Des*. 93, 1117–1128. doi: 10.1111/cbdd.13456

Homeyer, N., Stoll, F., Hillisch, A., and Gohlke, H. (2014). Binding free energy calculations for lead optimization: assessment of their accuracy in an industrial drug design context. *J. Chem. Theory Comput*. 10, 3331–3344. doi: 10.1021/ct5000296

Huber, G. G. A., and Kim, S. (1996). Weighted-ensemble brownian dynamics simulations for protein association reactions. *Biophys. J*. 70, 97–110. doi: 10.1016/S0006-3495(96)79552-8

Jorgensen, W. L., and Thomas, L. L. (2008). Perspective on free-energy perturbation calculations for chemical equilibria. *J. Chem. Theory Comput*. 6, 869–876. doi: 10.1021/ct800011m

Kirberger, S. E., Ycas, P. D., Johnson, J. A., Chen, C., Ciccone, M. F., Woo, R. W., et al. (2019). Selectivity, ligand deconstruction, and cellular activity analysis of a BPTF bromodomain inhibitor. *Organ. Biomol. Chem*. 17, 2020–2027. doi: 10.1039/C8OB02599A

Kirkwood, J. G. (1935). Statistical mechanics of fluid mixtures. *J. Chem. Phys*. 3, 300–313. doi: 10.1063/1.1749657

Kokh, D. B., Amaral, M., Bomke, J., Grädler, U., Musil, D., Buchstaller, H. P., et al. (2018). Estimation of drug-target residence times by τ-random acceleration molecular dynamics simulations. *J. Chem. Theory Comput*. 14, 3859–3869. doi: 10.1021/acs.jctc.8b00230

Laio, A., and Parrinello, M. (2002). Escaping free-energy minima. *Proc. Natl. Acad. Sci. U.S.A*. 99, 12562–12566. doi: 10.1073/pnas.202427399

Lee, K. S. S., Yang, J., Niu, J., Ng, C. J., Wagner, K. M., Dong, H., et al. (2019). Drug-target residence time affects *in vivo* target occupancy through multiple pathways. *ACS Central Sci*. 5, 1614–1624. doi: 10.1021/acscentsci.9b00770

Lensink, M. F., Velankar, S., and Wodak, S. J. (2017). Modeling protein-protein and protein-peptide complexes: CAPRI 6th edition. *Proteins* 85, 359–377. doi: 10.1002/prot.25215

Lotz, S. D., and Dickson, A. (2018). Unbiased molecular dynamics of 11 min timescale drug unbinding reveals transition state stabilizing interactions. *J. Am. Chem. Soc*. 140, 618–628. doi: 10.1021/jacs.7b08572

Lu, H., and Tonge, P. J. (2010). Drug-target residence time: critical information for lead optimization. *Curr. Opin. Chem. Biol*. 14, 467–474. doi: 10.1016/j.cbpa.2010.06.176

Nishikawa, N., Han, K., Wu, X., Tofoleanu, F., and Brooks, B. R. (2018). Comparison of the umbrella sampling and the double decoupling method in binding free energy predictions for SAMPL6 octa-acid host-guest challenges. *J. Comput. Aided Mol. Design* 32, 1075–1086. doi: 10.1007/s10822-018-0166-2

Northrup, S. H., Allison, S. A., and McCammon, J. A. (1984). Brownian dynamics simulation of diffusion-influenced bimolecular reactions. *J. Chem. Phys*. 80:1517. doi: 10.1063/1.446900

Nunes-Alves, A., Kokh, D. B., and Wade, R. C. (2020). Recent progress in molecular simulation methods for drug binding kinetics. *arXiv*. arXiv: 2002.08983.

Pan, A. C., Xu, H., Palpant, T., and Shaw, D. E. (2017). Quantitative characterization of the binding and unbinding of millimolar drug fragments with molecular dynamics simulations. *J. Chem. Theory Comput*. 13, 3372–3377. doi: 10.1021/acs.jctc.7b00172

Parks, C. D., Gaieb, Z., Chiu, M., Yang, H., Shao, C., Walters, W. P., et al. (2020). D3R grand challenge 4: blind prediction of protein-ligand poses, affinity rankings, and relative binding free energies. *J. Comput. Aided Mol. Des*. 34, 99–119. doi: 10.1007/s10822-020-00289-y

Pei, J., Yin, N., Ma, X., and Lai, L. (2014). Systems biology brings new dimensions for structure-based drug design. *J. Am. Chem. Soc*. 136, 11556–11565. doi: 10.1021/ja504810z

Rifai, E. A., Dijk, M., Vermeulen, N. P. E., Yanuar, A., and Geerke, D. P. (2019). A comparative linear interaction energy and MM/PBSA study on SIRT1-ligand binding free energy calculation. *J. Chem. Inform. Model*. 59, 4018–4033. doi: 10.1021/acs.jcim.9b00609

Rizzi, A., Jensen, T., Slochower, D. R., Aldeghi, M., Gapsys, V., Ntekoumes, D., et al. (2020). The SAMPL6 SAMPLing challenge: assessing the reliability and efficiency of binding free energy calculations. *J. Comput. Aided Mol. Des*. 34, 601–633. doi: 10.1007/s10822-020-00290-5

Rizzi, A., Murkli, S., McNeill, J. N., Yao, W., Sullivan, M., Gilson, M. K., et al. (2018). Overview of the SAMPL6 host-guest binding affinity prediction challenge. *J. Comput. Aided Mol. Des.* 32, 937–963. doi: 10.1007/s10822-018-0170-6

Shirts, M. R., and Chodera, J. D. (2008). Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.* 129:124105. doi: 10.1063/1.2978177

Singhal, N., Snow, C. D., and Pande, V. S. (2004). Using path sampling to build better Markovian state models: predicting the folding rate and mechanism of a tryptophan zipper beta hairpin. *J. Chem. Phys.* 121, 415–425. doi: 10.1063/1.1738647

Song, L. F., Bansal, N., Zheng, Z., and Merz, K. M. (2018). Detailed potential of mean force studies on host-guest systems from the SAMPL6 challenge. *J. Comput. Aided Mol. Des.* 32, 1013–1026. doi: 10.1007/s10822-018-0153-7

Spagnuolo, L. A., Eltschkner, S., Yu, W., Daryaee, F., Davoodi, S., Knudson, S. E., et al. (2017). Evaluating the contribution of transition-state destabilization to changes in the residence time of triazole-based InHa inhibitors. *J. Am. Chem. Soc.* 139, 3417–3429. doi: 10.1021/jacs.6b11148

Suárez, E., Lettieri, S., Zwier, M. C., Stringer, C. A., Subramanian, S. R., Chong, L. T., et al. (2014). Simultaneous computation of dynamical and equilibrium information using a weighted ensemble of trajectories. *J. Chem. Theory Comput.* 10, 2658–2667. doi: 10.1021/ct401065r

Sullivan, M. R., Yao, W., and Gibb, B. C. (2019). The thermodynamics of guest complexation to octa-acid and tetra-endo-methyl octa-acid: reference data for the sixth statistical assessment of modeling of proteins and ligands (SAMPL6). *Supramol. Chem.* 31, 184–189. doi: 10.1080/10610278.2018.1549327

Synapse (2018). *IDG-DREAM Drug-Kinase Binding Prediction Challenge*.

Tiwary, P., Mondal, J., and Berne, B. J. (2017). How and when does an anticancer drug leave its binding site? *Sci. Adv.* 3:e1700014. doi: 10.1126/sciadv.1700014

Tonge, P. J. (2017). Drug-target kinetics in drug discovery. *ACS Chem. Neurosci.* 9, 29–39. doi: 10.1021/acschemneuro.7b00185

Torrie, J. M., and Valleau, J. P. (1977). Non-physical sampling distributions in Monte-Carlo free-energy estimation umbrella sampling. *J. Comput. Phys.* 23, 187–199. doi: 10.1016/0021-9991(77)90121-8

Vanden-Eijnden, E., and Venturoli, M. (2009). Exact rate calculations by trajectory parallelization and tilting. *J. Chem. Phys.* 131:044120. doi: 10.1063/1.3180821

Vauquelin, G., Bostoen, S., Vanderheyden, P., and Seeman, P. (2012). *Clozapine, Atypical Antipsychotics, and the Benefits of Fast-Off D2 Dopamine Receptor Antagonism, Vol. 385*. Brussels: Springer. doi: 10.1007/s00210-012-0734-2

Votapka, L. W., Jagger, B. R., Heyneman, A. L., and Amaro, R. E. (2017). SEEKR: Simulation Enabled Estimation of Kinetic Rates, a computational tool to estimate molecular kinetics and its application to trypsin-benzamidine binding. *J. Phys. Chem. B* 121, 3597–3606. doi: 10.1021/acs.jpcb.6b09388

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

Woo, H. J., and Roux, B. (2005). Calculation of absolute protein-ligand binding free energy from computer simulations. *Proc. Natl. Acad. Sci. U.S.A.* 102, 6825–6830. doi: 10.1073/pnas.0409005102

Yin, J., Henriksen, N. M., Slochower, D. R., Shirts, M. R., Chiu, M. W., Mobley, D. L., et al. (2017). Overview of the SAMPL5 host-guest challenge: are we doing better? *J. Comput. Aided Mol. Des.* 31, 1–19. doi: 10.1007/s10822-016-9974-4

Zwanzig, R. W. (1954). High-temperature equation of state by a perturbation method. I. Nonpolar gases. *J. Chem. Phys.* 22, 1420–1426. doi: 10.1063/1.1740409

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

PERSPECTIVE
published: 17 June 2020
doi: 10.3389/fmolb.2020.00114

Recent Developments in Linear Interaction Energy Based Binding Free Energy Calculations

*Eko Aditya Rifai, Marc van Dijk and Daan P. Geerke**

AIMMS Division of Molecular and Computational Toxicology, Department of Chemistry and Pharmaceutical Sciences, Vrije Universiteit Amsterdam, Amsterdam, Netherlands

The linear interaction energy (LIE) approach is an end–point method to compute binding affinities. As such it combines explicit conformational sampling (of the protein-bound and unbound-ligand states) with efficiency in calculating values for the protein-ligand binding free energy ΔG_{bind}. This perspective summarizes our recent efforts to use molecular simulation and empirically calibrated LIE models for accurate and efficient calculation of ΔG_{bind} for diverse sets of compounds binding to flexible proteins (e.g., Cytochrome P450s and other proteins of direct pharmaceutical or biochemical interest). Such proteins pose challenges on ΔG_{bind} computation, which we tackle using a previously introduced statistically weighted LIE scheme. Because calibrated LIE models require empirical fitting of scaling parameters, they need to be accompanied with an applicability domain (AD) definition to provide a measure of confidence for predictions for arbitrary query compounds within a reference frame defined by a collective chemical and interaction space. To enable AD assessment of LIE predictions (or other protein-structure and -dynamic based ΔG_{bind} calculations) we recently introduced strategies for AD assignment of LIE models, based on simulation and training data only. These strategies are reviewed here as well, together with available tools to facilitate and/or automate LIE computation (including software for combined statistically-weighted LIE calculations and AD assessment).

Keywords: binding affinity computation, free energy calculation, molecular simulation, linear interaction energy, protein flexibility, binding promiscuity, applicability domain, reliability estimation

OPEN ACCESS

Edited by:
Sergio Decherchi, Italian Institute of Technology (IIT), Italy

Reviewed by:
Enrico Purisima, National Research Council Canada (NRC-CNRC), Canada
Hugo Gutiérrez De Teran, Uppsala University, Sweden

****Correspondence:***
Daan P. Geerke
d.p.geerke@vu.nl

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

Received: *11 December 2019*
Accepted: *14 May 2020*
Published: *17 June 2020*

Citation:
Rifai EA, van Dijk M and Geerke DP (2020) Recent Developments in Linear Interaction Energy Based Binding Free Energy Calculations. Front. Mol. Biosci. 7:114. doi: 10.3389/fmolb.2020.00114

1. END-POINT METHODS AND LINEAR INTERACTION ENERGY

Mutual molecular recognition is the starting point for a wide variety of biological processes (Gohlke and Klebe, 2002). Binding affinity governs ligand binding to target proteins, and being able to quantitatively understand and predict affinity in terms of binding free energy (ΔG_{bind}) can greatly support lead finding and/or optimization in the drug discovery process (Pohorille et al., 2010). Hence, improved efficiency and accuracy of computer–aided protein–ligand affinity methods play a pivotal role in accelerating and increasing success rates of drug discovery and design. ΔG_{bind} computation is still challenging, considering that virtual screening based on docking and scoring typically lacks sufficient accuracy, whereas use of rigorous alchemical methods can be too compute intensive for high–throughput scenarios, especially in case of flexible proteins that may bind ligands in multiple different orientations. As an alternative, end–point methods aim to provide a balance between accuracy and efficiency in ΔG_{bind} computation, and position themselves between

fast docking/scoring approaches and rigorous alchemical strategies for ΔG_{bind} calculation. They combine explicit conformational sampling (typically in molecular dynamics (MD) simulation) with a relatively fast scoring approach. By definition, end–point methods only require initial and final states to be simulated, i.e., the ligand free in solution and bound to the target protein, respectively, and/or interactions being either turned off and on (Wang et al., 2019).

Linear interaction energy (LIE) is an end–point method that was introduced in 1994 by Åqvist and coworkers (Åqvist et al., 1994). It is derived from the Linear Response Approximation (LRA) (Lee et al., 1992) to compute the electrostatic contributions to the binding affinity. As such, LIE is directly derived from the Zwanzig expression for free–energy perturbation (Leach, 2001). The non-polar contribution to ΔG_{bind} is in LIE also represented by calculating differences in average non-bonded (i.e., van der Waals) interaction energies between the ligand and its environment in either the protein-bound or unbound state (Åqvist et al., 1994). To compute ΔG_{bind} from the simulations of the ligand either bound to the protein or free in solvent, the obtained average van der Waals (*vdw*) and electrostatic (*ele*) interaction energies of the ligand with its environment are scaled by LIE parameters α and β:

$$\Delta G_{bind} = \alpha\left(\left\langle V^{vdw}_{lig-surr}\right\rangle_{bound} - \left\langle V^{vdw}_{lig-surr}\right\rangle_{unbound}\right) + \beta\left(\left\langle V^{ele}_{lig-surr}\right\rangle_{bound} - \left\langle V^{ele}_{lig-surr}\right\rangle_{unbound}\right) \quad (1)$$

Originally LRA was followed and β was set to 0.5 (Åqvist et al., 1994). In subsequent studies (Åqvist and Hansson, 1996; Hansson et al., 1998) Åqvist and co-workers assigned values to β based on electrostatic properties and chemical composition of the compounds of interest. From free energy perturbation studies on the electrostatic contribution to solvation free energies ΔG_{solv} (Åqvist and Hansson, 1996) and binding affinity prediction for 18 protein-ligand complexes (Hansson et al., 1998) it was concluded that for charged compounds β = 0.5 can be used, while lower values for different types of neutral ligands were found to best describe the electrostatic contribution to ΔG_{bind} and ΔG_{solv} (β = 0.43, 0.37, and 0.33 for neutral molecules with 0, 1, or (more than) 2 hydroxyl groups, respectively). The deviation from linear response for neutral compounds and the decrease in assigned β values with the number of hydroxyl groups were explained to originate from variations in solvent reordering around and interactions with the ligands (Åqvist and Hansson, 1996). Accordingly pre-assigned values for β have been used since then in various LIE binding affinity studies; see e.g., Shamsudin Khan et al., 2014 for a recent example in which these values were successfully used, in efforts to automate LIE binding free energy prediction within a drug-design context. Assignment of β based on the chemical nature of the compound of interest has also been extended toward other (hydrogen-bonding) functional groups in a large-scale (solvation) free energy perturbation study by Almlöf et al. (2007). They used a set of hundreds of small organic molecules to derive a model in which β values are assigned based on the number and types of functional groups of the compounds of interest, where each functional group adds a pre-defined perturbation to the base value for β (of 0.43).

We and others (see e.g., Carlson and Jorgensen, 1995; Wall et al., 1999) have chosen to incorporate β as an effective parameter in LIE binding free energy models and (together with α) train it based on experimentally available affinity data. In such cases, separate local models (with different values for α and β) may well be needed to accurately describe binding affinities for complete sets of binders for a given protein of interest, as shown e.g., in van Dijk et al., 2017 for 132 inhibitors of Cytochrome P450 19A1 (CYP19A1). Note that the meaning of empirically calibrated values of the parameters in trained LIE models is not always obvious and/or discussed. One of the exceptions is work of Kollman and co-workers (Wang et al., 1999) who found a correlation between α and the hydrophobicity of the binding site of the system of interest (with a larger number of hydrophobic groups buried after binding resulting in higher affinity and α values). After α and β are pre-assigned and/or calibrated based on experimental data, Equation (1) can be used to predict binding affinities of ligands with unknown experimental data. An optional offset parameter (often denoted as γ) can be added to Equation (1) as a fitting parameter. Fitted values of γ are typically system dependent and have been related to the hydrophobicity of the binding site (Almlöf et al., 2004). Optimal values for an offset parameter in calibrated models may also depend on the compounds of interest, as we illustrated by deriving local LIE models to predict binding affinities for a (diverse) set of 132 CYP19A1 binders, for which inclusion of an offset parameter would have led to different calibrated γ values in the three obtained local models (van Dijk et al., 2017). The use of additional LIE terms and associated scaling parameters has also been proposed such as the introduction of a γ parameter for the scaling and explicit inclusion of a surface-area term (Carlson and Jorgensen, 1995).

From the above, LIE assumes that intramolecular energies, entropic terms, desolvation effects, or other factors contributing to ΔG_{bind} can be handled and canceled out by fitting and scaling of the model parameters, as it is assumed to correlate linearly with the intermolecular interactions (Åqvist and Marelius, 2001). This scaling and fitting allows for the calculation of "absolute" (direct) values for ΔG_{bind}. For that purpose it can be critical to include and derive an offset γ parameter for the system under consideration (Almlöf et al., 2004). Having direct ΔG_{bind} values available makes it straightforward to use a Boltzmann–like statistical weighting scheme to include multiple binding poses of ligands combined into a single prediction of ΔG_{bind} (Stjernschantz and Oostenbrink, 2010). This is relevant for flexible proteins such as Cytochrome P450s that may bind their ligands in different orientations or that may adopt multiple (partial) conformations upon complexation (Stjernschantz et al., 2008). LIE can also handle diverse ligands in the dataset that may involve too large perturbations to be simulated (which may become impractical for alchemical free energy calculations), while simultaneously accounting for the unbound state of the ligand that is not considered by most empirical scoring functions (Brooijmans and Kuntz, 2003). Section 2 summarizes our recent progress in calibrating (statistically–weighted) LIE models for

diverse sets of binders of Cytochrome P450s or other flexible and promiscuous proteins.

When fitting parameters in Equation (1) based on experimental data, it should be realized that use of the resulting LIE model(s) asks for the definition of its (their) domain of applicability in order to be able to assess the reliability of predictions for arbitrary query compounds (Carrió et al., 2014). This is especially relevant when using LIE models in industrial or other applied settings, considering e.g., that some years ago the Organisation for Economic Cooperation and Development (OECD) formalized applicability domain (AD) assessment as principle to evaluate model validity (Jaworska et al., 2005). To enable reliability estimation of LIE predictions based on simulation and training data only, we recently introduced strategies to assign the AD of LIE or other protein-structure and -dynamic based models (as reviewed in section 3). Section 4 lists several software tools that have come available to facilitate (semi-)automated LIE modeling. These include our software for automated (statistically–weighted) LIE computation and associated AD assessment, and the availability of these and other tools may well be an important next step for applied use of LIE.

2. STATISTICAL WEIGHTING OF MULTIPLE PROTEIN-LIGAND BINDING CONFORMATIONS

Some years ago, Stjernschantz and Oostenbrink (2010) introduced an extended version of the LIE method in which results from multiple MD simulations starting from different protein conformations and/or binding orientations are combined into a single ΔG_{bind} calculation. With this method, protein–conformational sampling and the description of ligand–binding promiscuity can be improved when computing ΔG_{bind} for e.g., Cytochrome P450s or other flexible proteins that may bind their ligand in different binding orientations. The contribution of an individual simulation i that starts from a given protein conformation and ligand–docking pose is scaled via a Boltzmann–like statistical weighting scheme as follows (Hritz and Oostenbrink, 2009):

$$W_i = \frac{e^{-\Delta G_{bind,i}/k_B T}}{\sum_i e^{-\Delta G_{bind,i}/k_B T}} \tag{2}$$

with $\Delta G_{bind,i}$ the binding free energy calculated from simulation i according to Equation (1). The individual weights are then used to calculate ΔG_{bind} for a compound from N different simulations via:

$$\begin{aligned}\Delta G_{bind} =& \alpha \sum_i^N W_i \left(\left\langle V_{lig-surr}^{vdw} \right\rangle_{bound,i} - \left\langle V_{lig-surr}^{vdw} \right\rangle_{unbound} \right) \\ &+ \beta \sum_i^N W_i \left(\left\langle V_{lig-surr}^{ele} \right\rangle_{bound,i} - \left\langle V_{lig-surr}^{ele} \right\rangle_{unbound} \right)\end{aligned} \tag{3}$$

Because the weights W_i are directly dependent on the values of α and β, model calibration based on experimental data has now to be performed using an iterative fitting scheme (Stjernschantz and Oostenbrink, 2010). Hence, this extended version of LIE is also referred to as *iterative* LIE.

This approach was first tested for thiourea binding to Cytochrome P450 2C9 and provided a model with high accuracy when including simulations starting from multiple ligand–binding poses, whereas experimental accuracy could not be obtained when using a single MD simulation per compound (Stjernschantz and Oostenbrink, 2010). Subsequently, model improvement was also shown for thiourea binding to Cytochrome P450 2D6 by using not only different ligand poses but also multiple protein starting structures for MD (Perić–Hassler et al., 2013). The method was further extended by using multiple replicates per docking poses to further increase accuracy (Perić–Hassler et al., 2013). Later, our group has successfully used this Boltzmann–weighting LIE scheme for binding affinity prediction to e.g., CYP1A2 (Capoferri et al., 2015), CYP19A1 (van Dijk et al., 2017), JAK2 kinase (Capoferri et al., 2017), and FXR (Rifai et al., 2018), and it has been implemented in an automatic way in the *eTOX ALLIES* (Capoferri et al., 2017) and *MDStudio* platforms (van Dijk, 2017) (section 4). As part of these efforts, Vosmeer et al. proposed a Fourier–transform filtering strategy to detect stable parts of MD time series of the interaction energy terms (Vosmeer et al., 2016). Only segments with fluctuations smaller than a pre–defined cut–off were subsequently used to average ligand–surrounding interaction energies over. Using previously calculated ΔG_{bind} data of Cytochrome P450 2D6 (Vosmeer et al., 2014), this filtering strategy was able to make LIE calculation slightly more accurate while potentially greatly improving compute efficiency (Vosmeer et al., 2016). The reason that such filtering does not only improve efficiency but can also enhance accuracy is that the weighting scheme of Equations (2) and (3) is only valid when using results from individual simulations that cover well-separated parts of the potential energy surface of the system of interest (Hritz and Oostenbrink, 2009). Note that Nunes–Alves and Arantes (Nunes–Alves and Arantes, 2014) used a similar Boltzmann–weighting approach to incorporate multiple binding modes into their binding affinity prediction using an implicit solvent model and they tested it on four different receptors.

Besides calculating ΔG_{bind} with the inclusion of several binding poses in LIE, it was also shown that the probability of a binding pose to occur can be predicted by inspecting the weighting values obtained from Equation (2) (Rifai et al., 2019). We verified this recently for a system of SIRT1–ligands and found for the considered compounds a correlation between the simulations of the protein–bound state with highest weight W_i and information from a co–crystallization study, in terms of the observed protein–ligand interactions and/or the starting poses used in simulation (as compared to the co-crystallized binding poses) (Rifai et al., 2019). Thus, when being able to generate and select appropriate binding poses from docking and/or experimental information on protein–ligand interactions for (a vast majority of the) training compounds, iterative LIE training may well be subsequently performed in an unsupervised manner.

3. APPLICABILITY DOMAIN ANALYSIS FOR LIE

Provided that we use the LIE framework as a purely empirical method, i.e., not considering categories of the β parameter based on the chemical nature of the ligand (Hansson et al., 1998; Almlöf et al., 2007), the need for training and the use of fitted parameters (α and β) may raise the question of how reliable the predicted value will be for an arbitrary query compound. Hence there is a need to provide a measure to estimate the reliability of a prediction for a new compound with unknown binding affinity, and to evaluate if the query compound of interest is sufficiently represented by the employed set of model training compounds. This can be expressed in terms of the *applicability domain* (AD) of a given LIE model. The AD is a set of knowledge or information on the training set of the model and can give a measure for the confidence in a given prediction, in a similar vein as commonly applied in ligand–based empirical approaches (Carrió et al., 2014). A few years ago our group introduced an approach to allow AD assignment of LIE models, based on simulation and training data only (Capoferri et al., 2015, 2017; van Dijk et al., 2017). To our knowledge, this is the first method to analyze the AD of protein–structure (and –dynamic) based models such as LIE.

Inspired by a previous applicability domain analysis (ADAN) approach of Pastor and co–workers to define the domain of applicability of ligand–based QSAR models (Carrió et al., 2014), we have proposed an AD analysis strategy in a LIE study on Cytochrome P450 1A2 binding (Capoferri et al., 2015). In this study a relatively large set of (57) structurally–diverse training and test compounds were employed to explore the possibility to define the AD of calibrated LIE models in terms of five metrics. To estimate the reliability of a given prediction, these metrics or *confidence indices* are used to evaluate the similarity of an arbitrary query ligand (for which ΔG_{bind} is to be predicted) with the model's training set. This is not only evaluated in terms of structural similarity (according to Tanimoto scores) and computed ΔG_{bind} (as compared to the spread in experimental data used for calibration), but also in terms of the characteristics of the protein–ligand interactions. For the latter purpose, Mahalanobis–distance and (two) principal–component analyses are performed to enable a quantitative comparison between the averaged and the most relevant per–residue van der Waals and electrostatic interactions during simulation of either the protein–bound query or training compounds (Vosmeer et al., 2014; Capoferri et al., 2015). With these metrics in hand and after splitting the set of ligands with known binding affinity into a training and test set (of 35 and 22 compounds, respectively), a distinction could be successfully made between accurate and inaccurate ΔG_{bind} predictions for the test set compounds by looking at how many of the confidence metrics were violated per prediction (Capoferri et al., 2015).

An important conclusion from Capoferri's AD analysis was that the nature of the protein–ligand interactions (in terms of averaged non-bonded energies and the involved interacting protein residues) were more relevant descriptors for the AD of the LIE model than the molecular structure of the ligands alone. In the LIE study of van Dijk et al. mentioned in section 1 (van Dijk et al., 2017), this finding was confirmed for local models that were inferred for a set of 132 structurally–diverse CYP19A1 binders. By profiling and comparing per–residue interactions as observed for the protein–ligand simulations used for training, van Dijk showed differences in protein–ligand interactions among the three local models inferred, while structurally related compounds were not necessarily part of the same local model, indicating that protein–ligand interactions are a better measure to quantify if a given compound falls within the AD of a LIE model when compared to the molecular structure or other properties of the ligand alone.

The performances of LIE and the above mentioned AD analysis approach were evaluated in a real–life scenario of a community blind affinity prediction challenge organized by Drug Design Data Resource (D3R) during phase 2 of Grand Challenge 2 (GC2) (Gaieb et al., 2018). In D3R GC2, the challenge was to predict binding affinities of (102) agonists with different scaffolds for nuclear receptor FXR. For a subset of benzimidazole compounds (n = 9), a predictive accuracy (with a deviation from experiment of less than 5 kJ mol^{-1}) was obtained. Importantly, we showed that our AD analysis can yield representative metrics (in terms of an index for the level of confidence) to quantify the reliability of the binding affinity predictions based on simulation data only. It should also be noted that LIE might fail to predict the binding affinity of compounds with different structural properties and protein–ligand interaction profiles and/or domain of applicability from the ligands used for model training, or when the number of compounds constituting the training set cannot cover the range of experimental data of the test set. However, this can be estimated by the confidence level retrieved from AD analysis, to indicate possible limitations of the obtained LIE model. To enrich the interpretation of the applicability domain, we incorporated protein–ligand interaction profiling to evaluate the interaction of FXR with its ligands per obtained confidence level. We found that the confidence levels of the AD analysis were in line with the frequencies of ligand interactions with hotspot residues in the protein and with the model deviation and correlation metrics obtained from the predictions (Rifai et al., 2018).

4. (SEMI-)AUTOMATED LIE MODELING AND ANALYSIS TOOLS

Several software modules or packages (**Table 1**) are available that can be used to facilitate LIE modeling, such as the built–in package *gmx lie* within GROMACS (van der Spoel et al., 2005; Abraham et al., 2015) which can be used to directly obtain free energies of binding from interaction energy term analyses. *Q* (Marelius et al., 1998; Bauer et al., 2018) can semi–automatically perform LIE and FEP calculations, and is assisted with a Graphical User Interface (GUI) (Isaksen et al., 2015). The *Free Energy Workflow (FEW)* tool (Homeyer and

TABLE 1 | Selection of available tools to perform, facilitate and/or automate LIE calculations.

Software	Operating system	Free/Commercial	Requirement	Type
gmx lie	Windows, Linux, MacOS	Free	GROMACS	Program in simulation software
Q	Windows, Linux, MacOS	Free	–	Simulation software
FEW	Windows, Linux, MacOS	Free*	Amber, AmberTools	Perl script
CaFE	Windows, Linux, MacOS	Free	VMD	Tcl scripts
eTOX ALLIES	Windows, Linux, MacOS	Free	– (all required softwares are in the virtual machine)	Python scripts in virtual-machine environment
MDStudio	Windows, Linux, MacOS	Free*	Open Babel, PLANTS, AmberTools and GROMACS	Python scripts in docker and microservice environment
FESetup	Windows, Linux, MacOS	Free*	GROMACS, Amber, Sire, NAMD	Shell script
Desmond	Windows, Linux, MacOS	Commercial	Schrödinger suite	Software

**These tools (may) require Amber or PLANTS, which are commercial softwares.*

Gohlke, 2013, 2015) also enables LIE (as well as other free energy) calculations by facilitating setup and execution within the Amber suite (Case et al., 2005; Salomon–Ferrer et al., 2013). *CaFE* (Liu and Hou, 2016) specializes in calculating ΔG_{bind} by using end–point methods including LIE. *FESetup* (Loeffler et al., 2015) can facilitate alchemical free energy simulations and provides the ability to perform end-point calculations as well. *Desmond* (Bowers et al., 2006; Gao et al., 2012) from Schrödinger can also be used to extract ligand–surrounding interaction energies from MD simulations. Our *eTOX ALLIES* pipeline (Capoferri et al., 2017) enables automated molecular docking, MD simulation, iterative LIE and associated AD analysis, which can be based on inclusion of multiple binding modes and/or protein conformations as input for the MD simulations (Capoferri et al., 2017). Recently, we have made such LIE workflow also available within our modular and flexible *MDStudio* workflow management system (van Dijk, 2017).

5. CONCLUSIONS

The current perspective summarizes how we have explored the use of (statistically-weighted) LIE to predict binding affinity for challenging flexible (off-)target proteins such as Cytochrome P450s and nuclear receptor FXR. In addition we reviewed possibilities to evaluate the confidence in LIE predictions with an AD assessment approach for LIE or other protein-structure and -dynamic based free energy methods. Especially when the AD of a LIE model can be defined, LIE can treat sets of ligands that may involve too large perturbations to be simulated and become impractical for alchemical free energy perturbation or thermodynamic integration, while simultaneously accounting for the unbound state of the ligand that is not considered by most combined docking/scoring approaches. Thus, calibrated LIE models can be viewed as a combination of (4D–)QSAR and sampling approaches to estimate protein–binding affinities. Combined with the possibility to employ tools that facilitate and automate LIE calculations and (AD) analysis, the potential of addressing protein flexibility and promiscuity in statistically-weighted models and the availability of metrics for applicability domain analysis show direct promises for use of LIE in applied settings.

AUTHOR CONTRIBUTIONS

ER, MD, and DG contributed to drafting, writing and editing of the manuscript.

FUNDING

ER received financial support from Indonesia Endowment Fund for Education, Ministry of Finance, Republic of Indonesia (LPDP). DG gratefully acknowledges financial support by the Netherlands Organisation for Scientific Research (NWO, VIDI Grant 723.012.105).

REFERENCES

Åqvist, J., and Hansson, T. (1996). On the validity of electrostatic linear response in polar solvents. *J. Phys. Chem.* 100, 9512–9521. doi: 10.1021/jp953640a

Åqvist, J., and Marelius, J. (2001). The linear interaction energy method for predicting ligand binding free energies. *Comb. Chem. High Throughput Screen.* 4, 613–626. doi: 10.2174/1386207013330661

Åqvist, J., Medina, C., and Samuelsson, J.-E. (1994). A new method for predicting binding affinity in computer-aided drug design. *Protein Eng. Des. Sel.* 7, 385–391. doi: 10.1093/protein/7.3.385

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, 19–25. doi: 10.1016/j.softx.2015.06.001

Almlöf, M., Brandsdal, B. O., and Åqvist, J. (2004). Binding affinity prediction with different force fields: examination of the linear interaction energy method. *J. Comput. Chem.* 25, 1242–1254. doi: 10.1002/jcc.20047

Almlöf, M., Brandsdal, B. O., and Åqvist, J. (2007). Improving the accuracy of the linear interaction energy method for solvation free energies. *J. Chem. Theory Comput.* 3, 2162–2175. doi: 10.1021/ct700106b

Bauer, P., Barrozo, A., Purg, M., Amrein, B. A., Esguerra, M., Wilson, P. B., et al. (2018). Q6: a comprehensive toolkit for empirical valence bond and related free energy calculations. *SoftwareX*. 7, 388–395. doi: 10.1016/j.softx.2017.12.001

Bowers, K. J., Chow, D. E., Xu, H., Dror, R. O., Eastwood, M. P., Gregersen, B. A., et al. (2006). Scalable algorithms for molecular dynamics simulations on commodity clusters. In *SC'06: Proceedings of the 2006 ACM/IEEE Conference on Supercomputing* (Tampa, FL: IEEE), 43. doi: 10.1145/1188455.1188544

Brooijmans, N., and Kuntz, I. D. (2003). Molecular recognition and docking algorithms. *Annu. Rev. Biophys. Biomol. Struct*. 32, 335–373. doi: 10.1146/annurev.biophys.32.110601.142532

Capoferri, L., van Dijk, M., Rustenburg, A. S., Wassenaar, T. A., Kooi, D. P., Rifai, E. A., et al. (2017). eTOX ALLIES: an automated pipeLine for linear interaction energy-based simulations. *J. Chem. Inf.* 9:58. doi: 10.1186/s13321-017-0243-x

Capoferri, L., Verkade-Vreeker, M. C., Buitenhuis, D., Commandeur, J. N. M., Pastor, M., Vermeulen, N. P. E., et al. (2015). Linear interaction energy based prediction of Cytochrome P450 1A2 binding affinities with reliability estimation. *PLoS ONE* 10:e0142232. doi: 10.1371/journal.pone.0142232

Carlson, H. A., and Jorgensen, W. L. (1995). An extended linear response method for determining free energies of hydration. *J. Phys. Chem*. 99, 10667–10673. doi: 10.1021/j100026a034

Carrió, P., Pinto, M., Ecker, G., Sanz, F., and Pastor, M. (2014). Applicability domain analysis (ADAN): a robust method for assessing the reliability of drug property predictions. *J. Chem. Inf. Model*. 54, 1500–1511. doi: 10.1021/ci500172z

Case, D. A., Cheatham, T. E. III., Darden, T., Gohlke, H., Luo, R., Merz, K. M. Jr., et al. (2005). The Amber biomolecular simulation programs. *J. Comput. Chem*. 26, 1668–1688. doi: 10.1002/jcc.20290

Gaieb, Z., Liu, S., Gathiaka, S., Chiu, M., Yang, H., Shao, C., et al. (2018). D3R Grand Challenge 2: blind prediction of protein-ligand poses, affinity rankings, and relative binding free energies. *J. Comput. Aided Mol. Des*. 32, 1–20. doi: 10.1007/s10822-017-0088-4

Gao, C., Herold, J. M., and Kireev, D. (2012). Assessment of free energy predictors for ligand binding to a methyllysine histone code reader. *J. Comput. Chem*. 33, 659–665. doi: 10.1002/jcc.22888

Gohlke, H., and Klebe, G. (2002). Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. *Angew. Chem. Int. Ed*. 41, 2644–2676. doi: 10.1002/1521-3773(20020802)41:15<2644::AID-ANIE2644>3.0.CO;2-O

Hansson, T., Marelius, J., and Åqvist, J. (1998). Ligand binding affinity prediction by linear interaction energy methods. *J. Comput. Aided Mol. Des*. 12, 27–35. doi: 10.1023/A:1007930623000

Homeyer, N., and Gohlke, H. (2013). FEW: a workflow tool for free energy calculations of ligand binding. *J. Comput. Chem*. 34, 965–973. doi: 10.1002/jcc.23218

Homeyer, N., and Gohlke, H. (2015). Extension of the free energy workflow FEW towards implicit solvent/implicit membrane MM-PBSA calculations. *Biochim. Biophys. Acta Gen*. 1850, 972–982. doi: 10.1016/j.bbagen.2014.10.013

Hritz, J., and Oostenbrink, C. (2009). Efficient free energy calculations for compounds with multiple stable conformations separated by high energy barriers. *J. Phys. Chem. B* 113, 12711–12720. doi: 10.1021/jp902968m

Isaksen, G. V., Andberg, T. A. H., Åqvist, J., and Brandsdal, B. O. (2015). QGUI: a high-throughput interface for automated setup and analysis of free energy calculations and empirical valence bond simulations in biological systems. *J. Mol. Graph. Modell*. 60, 15–23. doi: 10.1016/j.jmgm.2015.05.007

Jaworska, J., Nikolova-Jeliazkova, N., and Aldenberg, T. (2005). QSAR applicability domain estimation by projection of the training set in descriptor space: a review. *Alternat. Lab. Anim*. 33, 445–459. doi: 10.1177/026119290503300508

Leach, A. R. (2001). *Molecular Modelling: Principles and Applications, 2nd ed.* Upper Saddle River, NJ: Prentice Hall.

Lee, F. S., Chu, Z.-T., Bolger, M. B., and Warshel, A. (1992). Calculations of antibody-antigen interactions: microscopic and semi-microscopic evaluation of the free energies of binding of phosphorylcholine analogs to McPC603. *Protein Eng. Des. Sel*. 5, 215–228. doi: 10.1093/protein/5.3.215

Liu, H., and Hou, T. (2016). CaFE: a tool for binding affinity prediction using end-point free energy methods. *Bioinformatics* 32, 2216–2218. doi: 10.1093/bioinformatics/btw215

Loeffler, H. H., Michel, J., and Woods, C. (2015). FESetup: automating setup for alchemical free energy simulations. *J. Chem. Inf. Model*. 55, 2485–2490. doi: 10.1021/acs.jcim.5b00368

Marelius, J., Kolmodin, K., Feierberg, I., and Åqvist, J. (1998). Q: a molecular dynamics program for free energy calculations and empirical valence bond simulations in biomolecular systems1. *J. Mol. Graph. Modell*. 16, 213–225. doi: 10.1016/S1093-3263(98)80006-5

Nunes-Alves, A., and Arantes, G. M. (2014). Ligand-receptor affinities computed by an adapted linear interaction model for continuum electrostatics and by protein conformational averaging. *J. Chem. Inf. Model*. 54, 2309–2319. doi: 10.1021/ci500301s

Perić-Hassler, L., Stjernschantz, E., Oostenbrink, C., and Geerke, D. P. (2013). CYP 2D6 binding affinity predictions using multiple ligand and protein conformations. *Int. J. Mol. Sci*. 14, 24514–24530. doi: 10.3390/ijms141224514

Pohorille, A., Jarzynski, C., and Chipot, C. (2010). Good practices in free-energy calculations. *J. Phys. Chem. B* 114, 10235–10253. doi: 10.1021/jp102971x

Rifai, E. A., van Dijk, M., Vermeulen, N. P. E., and Geerke, D. P. (2018). Binding free energy predictions of farnesoid X receptor (FXR) agonists using a linear interaction energy (LIE) approach with reliability estimation: application to the D3R Grand Challenge 2. *J. Comput. Aided Mol. Des*. 32, 239–249. doi: 10.1007/s10822-017-0055-0

Rifai, E. A., van Dijk, M., Vermeulen, N. P. E., Yanuar, A., and Geerke, D. P. (2019). A comparative linear interaction energy and MM/PBSA study on SIRT1-ligand binding free energy calculation. *J. Chem. Inf. Model*. 59, 4018–4033. doi: 10.1021/acs.jcim.9b00609

Salomon-Ferrer, R., Case, D. A., and Walker, R. C. (2013). An overview of the Amber biomolecular simulation package. *Wiley Interdiscip. Rev. Comput. Mol. Sci*. 3, 198–210. doi: 10.1002/wcms.1121

Shamsudin Khan, Y., Gutierrez-de Teran, H., Boukharta, L., and Åqvist, J. (2014). Toward an optimal docking and free energy calculation scheme in ligand design with application to COX-1 inhibitors. *J. Chem. Inform. Model*. 54, 1488–1499. doi: 10.1021/ci500151f

Stjernschantz, E., and Oostenbrink, C. (2010). Improved ligand-protein binding affinity predictions using multiple binding modes. *Biophys. J*. 98, 2682–2691. doi: 10.1016/j.bpj.2010.02.034

Stjernschantz, E., Vermeulen, N. P., and Oostenbrink, C. (2008). Computational prediction of drug binding and rationalisation of selectivity towards cytochromes P450. *Expert Opin. Drug Metab. Toxicol*. 4, 513–527. doi: 10.1517/17425255.4.5.513

van der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., and Berendsen, H. J. (2005). GROMACS: fast, flexible, and free. *J. Comput. Chem*. 26, 1701–1718. doi: 10.1002/jcc.20291

van Dijk, M. (2017). https://bioexcel.eu/webinar-mdstudio-microservice-based-molecular-dynamics-workflows-2017-10-11/

van Dijk, M., ter Laak, A. M., Wichard, J. D., Capoferri, L., Vermeulen, N. P. E., and Geerke, D. P. (2017). Comprehensive and automated linear interaction energy based binding-affinity prediction for multifarious Cytochrome P450 aromatase inhibitors. *J. Chem. Inf. Model*. 57, 2294–2308. doi: 10.1021/acs.jcim.7b00222

Vosmeer, C. R., Kooi, D. P., Capoferri, L., Terpstra, M. M., Vermeulen, N. P. E., and Geerke, D. P. (2016). Improving the iterative linear interaction energy approach using automated recognition of configurational transitions. *J. Mol. Model*. 22:31. doi: 10.1007/s00894-015-2883-y

Vosmeer, C. R., Pool, R., van Stee, M. F., Perić-Hassler, L., Vermeulen, N. P. E., and Geerke, D. P. (2014). Towards automated binding affinity prediction using an iterative linear interaction energy approach. *Int. J. Mol. Sci*. 15, 798–816. doi: 10.3390/ijms15010798

Wall, I. D., Leach, A. R., Salt, D. W., Ford, M. G., and Essex, J. W. (1999). Binding constants of neuraminidase inhibitors: an investigation of the linear

interaction energy method. *J. Med. Chem*. 42, 5142–5152. doi: 10.1021/jm990105g

Wang, E., Sun, H., Wang, J., Wang, Z., Liu, H., Zhang, J. Z., 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, W., Wang, J., and Kollman, P. A. (1999). What determines the van der Waals coefficient β in the LIE (linear interaction energy) method to estimate binding free energies using molecular dynamics simulations? *Proteins Struct. Funct. Bioinf*. 34, 395–402. doi: 10.1002/(SICI)1097-0134(19990215)34:3<395::AID-PROT11>3.0.CO;2-4

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: 24 June 2020
doi: 10.3389/fmolb.2020.00093

An Analysis of Proteochemometric and Conformal Prediction Machine Learning Protein-Ligand Binding Affinity Models

*Conor Parks, Zied Gaieb and Rommie E. Amaro**

Department of Chemistry and Biochemistry, University of California, San Diego, La Jolla, CA, United States

Protein-ligand binding affinity is a key pharmacodynamic endpoint in drug discovery. Sole reliance on experimental design, make, and test cycles is costly and time consuming, providing an opportunity for computational methods to assist. Herein, we present results comparing random forest and feed-forward neural network proteochemometric models for their ability to predict pIC50 measurements for held out generic Bemis-Murcko scaffolds. In addition, we assess the ability of conformal prediction to provide calibrated prediction intervals in both a retrospective and semi-prospective test using the recently released Grand Challenge 4 data set as an external test set. In total, random forest and deep neural network proteochemometric models show quality retrospective performance but suffer in the semi-prospective setting. However, the conformal predictor prediction intervals prove to be well-calibrated both retrospectively and semi-prospectively showing that they can be used to guide hit discovery and lead optimization campaigns.

Keywords: conformal prediction, proteochemometric, protein-ligand binding affinity, bemis-murcko scaffolding, random forest, deep neural net (DNN)

OPEN ACCESS

Edited by:
Sergio Decherchi,
Italian Institute of Technology (IIT), Italy

Reviewed by:
Oscar Mendez Lucio,
Bayer, France
Zhili Zuo,
Kunming Institute of Botany (CAS), China

****Correspondence:***
Rommie E. Amaro
ramaro@ucsd.edu

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

Received: *28 January 2020*
Accepted: *22 April 2020*
Published: *24 June 2020*

Citation:
Parks C, Gaieb Z and Amaro RE (2020) An Analysis of Proteochemometric and Conformal Prediction Machine Learning Protein-Ligand Binding Affinity Models. Front. Mol. Biosci. 7:93. doi: 10.3389/fmolb.2020.00093

INTRODUCTION

One of the most important phases of a drug discovery campaign is the discovery of a potent inhibitor to a target driving the disease phenotype. Experimental design, make, test cycles seek to optimize initial hits to lead compounds by optimizing the protein-ligand binding affinity. However, this process is frequently slow and costly, adding to the large cost of drug discovery. As such, computational methods that can accelerate this optimization phase by predicting protein-ligand binding affinity values are readily sought. Fully atomistic simulation approaches model protein-ligand binding physics through time integrating Newton's equations of motion in molecular dynamics simulations (Jorgensen and Thomas, 2008; Chodera et al., 2011; Mobley and Klimovich, 2012; Christ and Fox, 2014; Abel et al., 2017; Cournia et al., 2017; Mobley and Gilson, 2017). However, molecular dynamics approaches can suffer from large computational costs, insufficient sampling, and variably accurate force fields. As an alternative, quantitative structure activity modeling (QSAR) uses machine learning (ML) as a stand in for physically rigorous simulations by seeking to model statistical correlations between ligand information and protein-ligand binding affinity (Cherkasov et al., 2014). Traditional QSAR models do not model the protein directly, and hence do not allow learning from related protein family members during training. In contrast proteochemometric (PCM) models combine both protein and ligand information to create a composite feature vector that allows the model to learn mappings between all protein-ligand

pairs in a training set (Cortés-Ciriano et al., 2015a). PCM models have been applied to a diverse number of protein families including G-protein coupled receptors (Gao et al., 2013), HDACs (Tresadern et al., 2017), kinases (Subramanian et al., 2013), Cytochrome P450s, HIV proteases (Lapins et al., 2008), Poly(ADP-ribose) polymerases (Cortés-Ciriano et al., 2015b), and bromodomains (Giblin et al., 2018). Recently, PCM and multi-task neural networks have also been benchmarked using ChEMBL data where the utility of PCM modeling for binder/non-binder classification was demonstrated (Lenselink et al., 2017).

We first compare the performance of random forest (RF) and feed-forward neural network (FFN) PCM models trained using the recently released ChEMBL25 data set (Gaulton et al., 2017) with full length protein sequence features on a generic Bemis-Murcko scaffold split. It is shown that the RF and FFN models achieve comparable overall performance suggesting that both models approach the upper levels of performance possible given the heterogeneous IC50 measurements in ChEMBL25 (Kalliokoski et al., 2013). Feature analysis shows that both models leverage the protein sequence features extensively, albeit in different ways. We demonstrate that the optimization of entity embeddings for ECFP6 categorical variables allows FFN models to perform feature engineering in a data driven manner (Guo and Berkhahn, 2016). In addition, we compare the validity and efficiency of regression conformal predictors first in a retrospective test using ChEMBL25 data and subsequently in semi-prospective test using the recent Drug Design Data (D3R) Grand Challenge 4 (GC4) data set (Parks et al., 2019a). The D3R issues blinded prediction challenges to the computer aided drug design (CADD) community to assess method performance in truly blinded scenarios. Although the data has been since released to the community, we use the data in the most recent GC4 (Parks et al., 2019a) data set as a test set external to the data in ChEMBL25 for what we refer to as a semi-prospective test. These results show that the performance of both models suffer on the GC4 dataset relative to the ChEMBL25 validation set. However, the performance is in line with the top performing ML models in prior D3R Grand Challenges (Gathiaka et al., 2016; Gaieb et al., 2018, 2019; Parks et al., 2019a). Finally, the prediction intervals from the conformal predictors are shown to be valid on the GC4 data set, demonstrating the validity of conformal prediction confidence intervals on a high-quality external test set.

MATERIALS AND METHODS

Data Set Source and Preparation

The recently released ChEMBL25 (Gaulton et al., 2017) database was used for ML model training. Only molecules with specified canonical SMILES strings, standard units of nM, no potential duplicates, confidence score of 9, activity comment not equal to inconclusive, and against protein targets with specified gene ids, and protein sequences from Swiss-Prot were kept. Ligands with PAINS patterns identified via RDKit[1] were removed. Only ligands with a molecular weight in the range of 75 to 800 Da were retained. We replaced multiple IC50 values for the same protein-ligand pair by the median IC50 value. SMILES strings were standardized and canonicalized using the charge parent function in MolVS[2]. SMILES strings were either featurized using 4,096 bit length ECFP6 fingerprints, molecular weight, topological surface area, number of hydrogen donors, number of hydrogen acceptors, LogP, heavy atom count, number of rotatable bonds, and ring count with RDKit or CDDD descriptors (Winter et al., 2019). Protein sequences were featurized using amino acid, dipeptide, composition, transition, and distribution descriptors. Ligand and protein descriptors were then concatenated to create the full feature vector. Non ECFP6 bit values were scaled using the standard scaler function in Scikit-learn (Pedregosa et al., 2011). All IC50 values were converted to pIC50 values and scaled using the standard scaler function. The final data set contained 302,325 data points consisting of 213,502 unique SMILES strings across 940 unique UniProt IDs.

Machine Learning Model Training and Conformal Prediction

RF models were trained using the Scikit-learn (Pedregosa et al., 2011) library via a grid search hyperparameter optimization strategy. The following hyperparameter values were explored: 100, 500, and 1,000 for the number of estimators; sqrt, log2, 0.3, and 0.5 for max features; and 1,3,5,10,25 for min samples leaf. FFN models were trained using Fast.ai (Howard and Gugger, 2020). We investigated treating ECFP6 bit vectors as categorical variables whose embeddings were optimized via backpropagation during model training (Guo and Berkhahn, 2016). A 3-layer model was employed with 2,000 nodes in the first layer, 1,000 in the second, and 500 in the third. Linear layer outputs were then passed through an activation and then a batch norm and dropout layer. A ReLU function were used for activations, with the exception of the last layer, where a Sigmoid function was used. This was done to facilitate training by scaling outputs from the last linear layer to a range of values between the max and minimum scaled pIC50 values in the training set multiplied by a scaling factor of 1.2. Weight decay was set to 0.01. Dropout of 0.25 was used in each layer, except for the embedding layers, where a dropout of 0.01 was used. All other Fast.ai tabular model defaults were used. The FFN model was trained with the fit_one_cycle (Smith, 2018) method. All models were then analyzed using mean squared error (MSE), Pearson correlation coefficient, and Kendall's Tau metrics.

Rigorous quantification of model confidence is essential in fields such as drug discovery where chemical space is essentially infinitely vast and models are trained on only a small fraction of possible compounds. Conformal prediction is a state of the art method to provide confidence intervals, i.e., a region where the true value is predicted to be, and whose size is determined in part by a user defined confidence level (Shafer and Vovk, 2008; Norinder et al., 2014; Cortés-Ciriano et al., 2015a; Sun et al., 2017; Svensson et al., 2017). Here, we define validity as

[1]RDKit Avaliable Online at: https://www.rdkit.org/ (accessed August 13, 2019).

[2]MolVS: Molecule Validation and Standardization — MolVS 0.1.1 documentation Avaliable Online at: https://molvs.readthedocs.io/en/latest/ (accessed August 13, 2019).

the frequency at which the confidence interval contains the true value. For example, the 95% confidence prediction intervals for a well-calibrated conformal predictor would contain the true value 95% of the time. The efficiency of a conformal predictor is a reflection of the size of the confidence internals produced. Here, a model that produces smaller confidence intervals than another model would be considered to have higher efficiency. These two variables, validity and efficiency, quantify the performance of conformal prediction models. Conformal prediction requires a way to gauge the similarity of a new piece of data to training data. Recent literature has shown that the standard deviation across the trees of a RF model (Svensson et al., 2018), and the use of test-time dropout in the case of FFN models (Cortés-Ciriano and Bender, 2019a), provide valid and efficient conformal predictors. However, these methods have not been analyzed in prospective settings extensively. We use these methods to assess non-conformity herein. For an overview of conformal prediction in the field of drug discovery, we direct the reader to a recent review (Cortés-Ciriano and Bender, 2019b).

It has been shown that overly simplistic (i.e., random) training/validation/test splits lead to overestimates in the accuracy of machine learning models (Wallach and Heifets, 2018). This is one explanation for the discrepancy between performance metrics seen in retrospective settings and those seen in true prospective tests (Gathiaka et al., 2016; Gaieb et al., 2018, 2019; Parks et al., 2019a). In an attempt to mitigate this, we elected to perform hyperparameter optimization using an 80/20 generic Bemis-Murcko scaffold split. This leads to a more difficult split, as the validation set does not contain any compounds with a generic Bemis-Murcko scaffold already present in the training set. Optimal hyperparameters and performance metrics on the ChEMBL25 data set for all models were first determined with this split. In all subsequent models, these sets of hyperparameters were using for training. The training set was then further divided using a random 80/20 split into a new training and calibration set. Models were then retrained and the calibration set was used to calibrate the confidence intervals of the conformal predictors. These models were used to assess the validity and efficiency of the conformal predictors on the ChEMBL25 validation data set. Finally, the entire ChEMBL25 dataset was randomly split into another 80/20 training and calibration set. All models were again trained and calibrated. These final models were used to assess performance on the external GC4 dataset in a semi-prospective test.

RESULTS

Retrospective Analysis

In **Table 1** we provide the performance metrics and model architectures for the best performing models across the full validation set.

Table 1 demonstrates that both model types (RF/FFN) perform equally well on the applied generic Bemis-Murcko scaffold split. We found the RF model with ECFP6 fingerprints, and the FFN model with entity embeddings to be the best performing models. The following sets of hyperparameters were found to be optimal for the RF models: {n_estimators=1000, max_features=sqrt, min_samples_leaf=1} using ECFP6 fingerprints, and {n_estimators=1000, max_features=log2, min_samples_leaf=1} using CDDD features. The scaled pIC50 MSE values in **Table 1** translate to a root mean squared error of approximately 0.8 pIC50 units. These values are in the range of expected errors for ML models trained on heterogeneous ChEMBL25 data (Kalliokoski et al., 2013), and are in agreement with prior literature that also demonstrated that RF and FFN models approached the upper limit of overall accuracy across the dataset, given the heterogeneous IC50 measurements in ChEMBL25 (Cortés-Ciriano and Bender, 2019a).

TABLE 1 | ChEMBL25 validation set performance metrics for both the RF and FFN models as well as the SMILES featurization method used.

Model type	MSE (scaled pIC50)	Pearson correlation	Kendall's Tau
RF (ECFP6)	0.38	0.80	0.61
RF (CDDD)	0.47	0.75	0.55
FFN (entity embeddings)	0.39	0.79	0.60
FFN (ECFP6)	0.41	0.78	0.58
FFN (CDDD)	0.42	0.78	0.58

TABLE 2 | Performance metrics averaged across individual UniProt IDs.

Model type	MSE (scaled pIC50)	Pearson correlation	Kendall's Tau
RF	0.37 +/– 0.17	0.65 +/– 0.18	0.46 +/– 0.15
FFN	0.38 +/– 0.15	0.61 +/– 0.22	0.43 +/– 0.17

Ensemble averaging is a strategy to improve prediction performance by averaging the individual predictions of multiple models. This wisdom of crowd approach works best when individual models are uncorrelated, allowing errors to be averaged out. The residuals of the RF and FFN model are correlated with an R^2 metric of 0.88. As such, only a modest improvement at best is noted when the FFN and RF model predictions are averaged to yield the following performance metrics: 0.36 MSE, 0.81 Pearson correlation, and 0.61 Kendall's Tau.

The number of data points per gene is heterogeneous in ChEMBL25, allowing a few genes to contribute more to performance metrics than others. To remove this bias, we calculated performance metrics across each individual gene with more than 100 data points in the validation set (**Table 2**). Analysis of **Table 2** shows that model performances vary moderately across the various genes with metric fluctuations (standard deviation/value) on the order of 30–50%. We found the amount of training data for a given UniProt ID to be a poor predictor for future successful predictions. We calculated the Kendall's Tau correlation between amount of data in the training set and the MSEs for each gene and found no correlation (tau = 0.05 both the RF and FFN model). Performance is spread across protein families as well, with the top 30 performing UniProt IDs for the RF model containing proteins from the kinase, protease, and RNA polymerase families. This suggests that chemical similarity

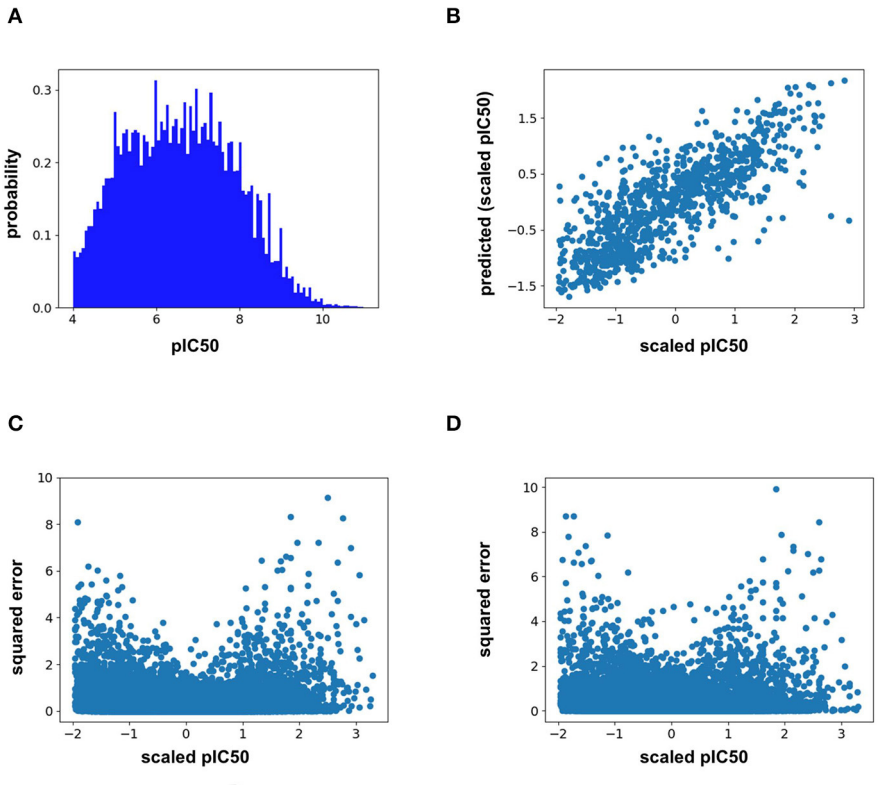


FIGURE 1 | (A) ChEMBL25 pIC50 probability distribution, **(B)** validation set scaled pIC50 vs. RF predictions, **(C)** validation set scaled pIC50 vs. RF prediction squared error, and **(D)** validation set scaled pIC50 vs. FFN prediction squared error.

and bias between the training and validation set are the most import variables determining performance.

Analysis of the training distribution of the pIC50 values shows that a strong amount of publication bias exists in the ChEMBL25 dataset (**Figure 1A**). Active compounds are over represented as shown by the average pIC50 value of the whole distribution being 6.57. In true prospective unbiased chemical library screens, it is common for 95% of the compounds to have pIC50 < 5 (i.e., non-binders). This demonstrates that the pIC50 predictions from the ML models herein should be used with caution for prospective virtual screening of chemical libraries, as large test-time distribution shift is certain. In addition, the distribution of training pIC50s is non-uniform (**Figure 1A**). This can lead to heterogeneous model performance across differing ranges of pIC50 values. For example, the RF model predictions correlate well with the measured values for the given split overall (**Figure 1B**). However, **Figures 1C,D** demonstrates the accuracy of both the RF and FFN models suffer in the tail of the pIC50 distribution, with quality performance obtained in a range of +/− 1 standard deviation from the mean.

Feature importance analysis allows us to determine whether the protein sequence component of the feature vector is pertinent to the model performance. This analysis indicates that the feature importance for the RF model plateaus at approximately the 1500th ranked feature out of the total 4,104 features (**Figure 2A**). All 567 protein sequence features, and all 8 physiochemical property features, fall in the top 1500 features. For the FFN model, the feature importance plateaus at approximately the 2000th ranked feature out of the total 4,104 features for the FFN model (**Figure 2B**). The top 2,000 features consist of 268 of 567 protein sequence features, 1,725 of 4,096 Morgan fingerprint bits, and 7 of 8 physiochemical properties. This demonstrates that both model types rank protein sequence features among the most important features. Both models find the physiochemical properties features to be the most import features overall, including MW. The Kendall's Tau ranking of the validation set scaled pIC50 values using solely MW alone is 0.15 demonstrating the trend of lead optimization campaigns to result in increasingly larger MW molecules. To further interrogate the protein sequence features, new training and validation sets were generated via a random UniProt ID split and models were retrained. This resulted in a much tougher split with the FFNs being the clearly best performing model by MSE (**Table 3**). This suggests that the FFN models are able to leverage the protein sequence features more effectively.

Molecular fingerprints are the most commonly used technique to encode molecules for ML model training. This technique hashes atomic neighborhoods for each atom to bits to represent molecules with 1D dimensional vectors. This approach can suffer from bit collision. In addition, there is no meaning of

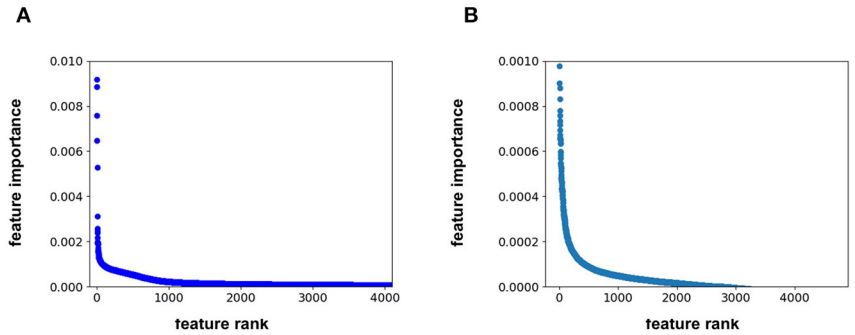


FIGURE 2 | (A) feature rank vs. feature importance for the **(A)** RF model and **(B)** FFN model. In **(B)**, the y-axis was capped at 0.001 to maintain resolution.

TABLE 3 | ChEMBL25 validation set performance metrics for both the RF and FFN models as well as the SMILES featurization method used.

Model type	MSE (scaled pIC50)	Pearson correlation	Kendall's Tau
RF (ECFP6)	0.96	0.41	0.27
RF (CDDD)	0.89	0.46	0.31
FFN (entity embeddings)	0.80	0.51	0.35
FFN (ECFP6)	0.79	0.49	0.34
FFN (CDDD)	0.78	0.48	0.33

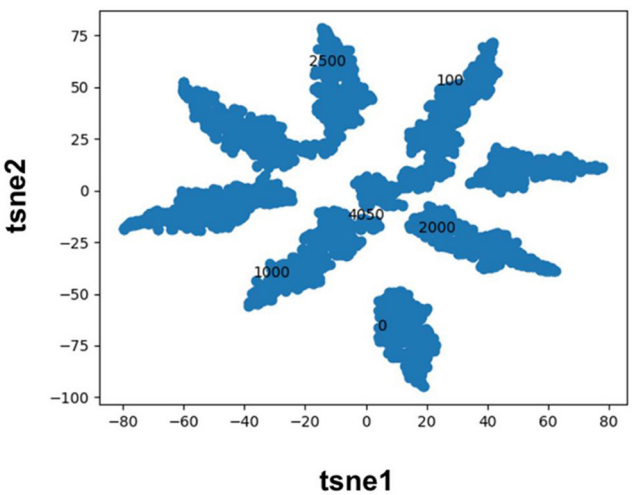


FIGURE 3 | TSNE plot of the FFN entity embeddings for the category 1 variables of the morgan fingerprint vector. Category 1 was selected for plotting as this denotes the presence of a chemical fragment. The category 1 weights of the 0, 100, 1,000, 2,000, 2,500, and 4,050 bits are plotted in black to illustrate how the FFN model groups bits into distinct clusters.

proximity of bits (Feinberg et al., 2019). As an alternative, prior literature has sought to extract features directly from images (Ragoza et al., 2017; Ciriano and Bender, 2018; Jiménez et al., 2018; Parks et al., 2019b) or graphs (Kearnes et al., 2016; Kipf and Welling, 2017; Feinberg et al., 2019; Torng and Altman, 2019) using convolution neural networks. Here, we pursue a complementary approach where we map categorical variables into Euclidean space using entity embeddings (Guo and Berkhahn, 2016). Entity embeddings are sets of weights in the neural network that represent possible categories of a feature. These weights are then optimized by the neural network during training. Hence, each category (0/1) of each bit in the Morgan fingerprint is now represented by a unique set of weights in the neural network. This allows the network to learn relationships between chemical fragments in a data driven manner. Our results demonstrate that the FFN learns to group Morgan fingerprint bits into multiple clusters through the optimization of embedding weights (**Figure 3**) providing a novel compound fingerprint. For the generic Murcko scaffold split used herein, the use of entity embeddings led to a 5% reduction in MSE for the FFN model. However, no performance variation was seen between the CDDD/ECFP6/and entity embedding features for the FFN models on the random protein split. Future work will be needed to interrogate the utility of the entity embeddings as a supplement to the tradition ECFP fingerprint.

The ChEMBL25 validation set was also used to assess the retrospective performance of conformal prediction. Analysis of the RF prediction interval sizes shows that they span a larger range of values than those from the FFN (**Figure 4A**), and hence are less efficient. However, the RF model has better validities than the FFN model (**Figure 4B**), but both models still achieve quality validities overall. Interestingly, there is very little variation in the size of the FFN confidence intervals across all predictions on the validation set (**Figure 4A**) but this is still sufficient for the FFN to generate valid prediction intervals (**Figure 4B**). In total, conformal prediction is able to accurately gauge both RF and FFN model confidence for predictions on held-out validation data, in agreement with prior literature (Svensson et al., 2018; Cortés-Ciriano and Bender, 2019a).

Semi-prospective Test

The true test of model performance is a prospective one as this is how ML models are used in practice. Here, we use the recently released GC4 dataset (Parks et al., 2019a) as an external, semi-prospective test for the best performing models [RF (ECFP6) and FFN (entity embeddings)] from the generic Murcko split. GC4

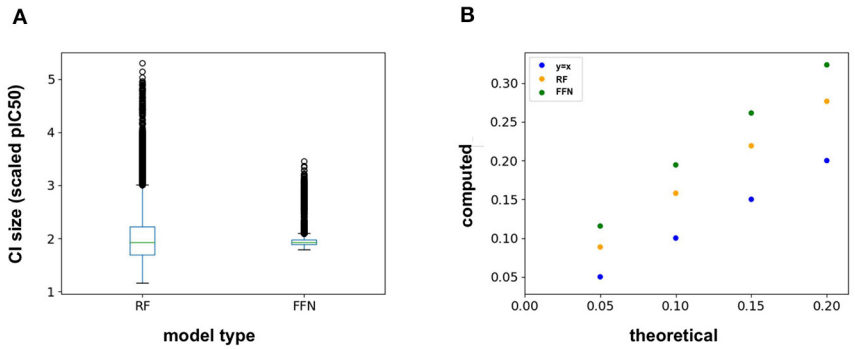


FIGURE 4 | (A) Box-whisker plot of the 95% confidence interval (CI) sizes for both the RF and FFN model computed on the validation set, and **(B)** theoretical vs. computed error rates (1-validities) on the validation set for varying error thresholds (1-confidence threshold).

TABLE 4 | GC4 BACE-1 and CatS performance metrics.

Model	BACE-1 Pearson correlation	BACE-1 Kendall's Tau	CatS Pearson correlation	CatS Kendall's Tau
RF	0.25 (0.71)	0.19 (0.51)	0.25 (0.68)	0.38 (0.49)
FFN	0.5 (0.77)	0.29 (0.59)	0.19 (0.8)	0.26 (0.61)

The corresponding performance metrics on the ChEMBL25 validation sets for BACE-1 and CatS are contained in parenthesis for comparison.

provided IC50 data to two targets: Cathepsin S (CatS), and Beta-amyloid secretase 1 (BACE-1). Analysis of **Table 4** shows that the ranking of model performances is target dependent, with the FFN performing best for BACE-1 and the RF model performing best for CatS. By comparing the metrics for both models on the same targets in the retrospective validation set, we see that both models suffer heavily in the prospective setting (**Table 4**). This is in agreement with prior literature showing data set split bias inflates model performance metrics (Wallach and Heifets, 2018), despite our attempts to mitigate this via a generic Bemis-Murcko scaffold split. However, these metrics are in line with the performance seen during GC4 with ML models (Parks et al., 2019a). Based on the Kendall's Tau value of the RF model for the CatS dataset, the RF model would have ranked in the top 10 performing methods for the CatS affinity prediction challenge. Similarly, the FFN would have placed in the top 10 for the BACE-1 affinity prediction challenge.

In addition to the model predictions, this semi-prospective test allows us to analyze the validity and efficiency of the conformal prediction intervals for both the RF and FFN. Both models achieve excellent validities on the CatS data set. Here, the prediction intervals contained the true measured value 95 and 97% of the time for the RF/FFN models, respectively. The validities of both models suffer slightly on the BACE-1 data set relative to CatS, but overall still achieve excellent performance with 81 and 80% validities. As a possible explanation for the validity performance degradation between CatS and BACE1, we note that the distribution of nearest neighbor Tanimoto coefficients demonstrates that the CatS GC4 compounds are more similar to the ChEMBL25 training data, and hence providing an easier test for the conformal predictor, than the BACE1 GC4 compounds (**Supplementary Figure 1**). For both CatS and BACE-1, we find the FFN provides more efficient prediction intervals (**Figures 5A,B**). In total, these results demonstrate the ability of the conformal predictors to generate valid prediction intervals in a semi-prospective setting.

We next sought to compare the performance of the PCM models against a standard QSAR target model trained using only CatS or BACE-1 ChEMBL25 data, respectively. Here, a RF QSAR model achieved approximately the same performance metrics on both targets as the RF PCM model with a 0.26 Pearson Correlation and 0.37 Kendall's Tau for CatS and 0.26 Pearson Correlation and 0.18 Kendall's Tau for BACE-1. This indicates that the RF model may benefit from additional protein sequence descriptors such as those from unsupervised training (Kim et al., 2019). The conformal prediction validities remained approximately equivalent at 93% on the CatS data set and 83% on the BACE-1 dataset. However, the data augmentation from PCM training improved the efficiencies of the confidence intervals relative to those from the QSAR RF model for BACE-1 (**Figures 5A,C**). The FFN QSAR model performance degraded relative to the PCM model with a 0.23 Pearson Correlation and 0.25 Kendall's Tau for CatS and a 0.46 Pearson Correlation and 0.22 Kendall's Tau for BACE-1. This suggests that the FFN is able to leverage the information of other protein sequences in the data set more effectively than the RF model during training. The validities of the FFN QSAR conformal predictor for BACE-1 degraded to 74%, but remained roughly the same at 98% for CatS. We find the PCM FFN model confidence interval to be more efficient than those of the QSAR FFN model (**Figures 5B,D**). This is most vividly captured in the case of CatS.

Finally, we sought to test the impact of deleting all CatS and BACE1 data and retraining the models using the same procedure. As shown in **Supplementary Table 1**, this had the expected negative impact on performance metrics for both model types with the FFN performance suffering the least. The only exception was the Pearson's correlation between FFN model predictions and target values for CatS where the metric remained

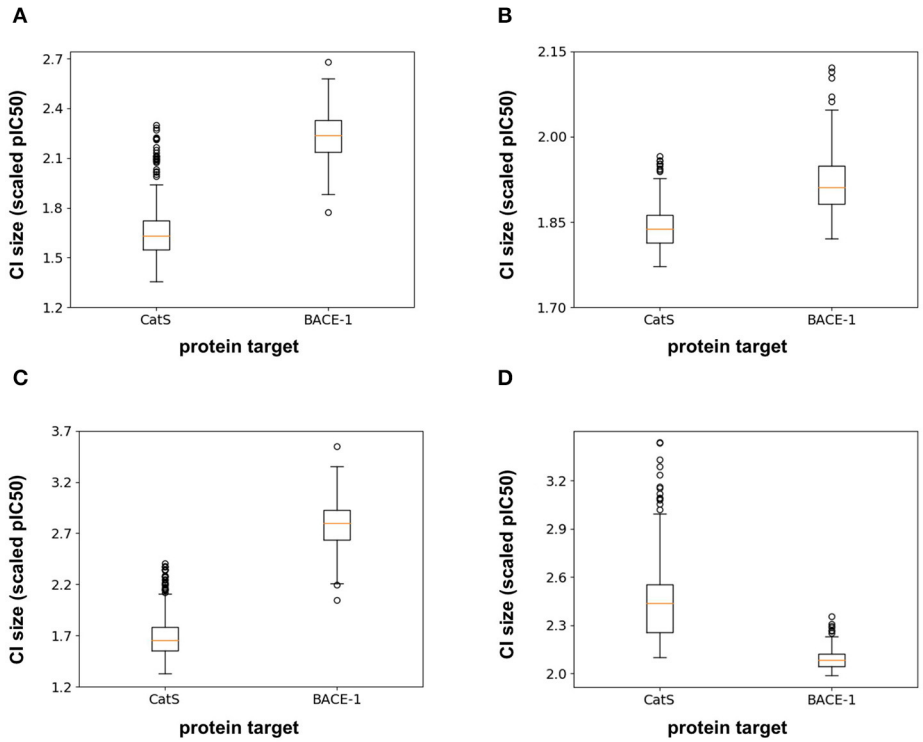


FIGURE 5 | Box-whisker plots of the 95% confidence interval (CI) sizes (scaled pIC50) for CatS/BACE-1 predictions for **(A)** RF PCM model, **(B)** PCM FFN model, **(C)** QSAR RF model, and **(D)** QSAR FFN model.

statistically equivalent. Notably, the Kendall's Tau ranking of model predictions is now only as good as a LogP ranking of the data for both model types. However, the conformal predictor remains well-calibrated with a 99 and 80% validity for the RF model, and 80 and 72% validity for the FFN model for CatS and BACE1, respectively. **Supplementary Figure 2** demonstrates that there is a marked reduction in the efficiency of the prediction intervals from the RF model. However, the FFN prediction interval efficiencies only degrade slightly. Despite the reduction in the performance of the point prediction values, we conclude conformal prediction remains an impactful method to gauge model confidence even on never before seen protein targets.

CONCLUSION

Protein-ligand binding affinity is a key variable during hit discovery and lead optimization in drug discovery. Experimental design, make, test cycles that seek to optimize this property are costly and time consuming, and hence limit the rate of entry of novel therapies to the clinic. Computational methods seek to accelerate these cycles by producing reliable protein-ligand binding affinity predictions. Traditional QSAR models train using only chemical compound data for a specific target. Alternatively, PCM models featurize both protein sequence and ligand to create the final feature vector. This allows ML models to be trained on protein-ligand binding affinity data from multiple proteins at once, hence augmenting the size of the training set, and potentially allowing the model to learn from related proteins (Lapins et al., 2008; Gao et al., 2013; Subramanian et al., 2013; Cortés-Ciriano et al., 2015a,b; Tresadern et al., 2017; Giblin et al., 2018).

Here we first analyze the performance of PCM models trained using the most recent ChEMBL25 database (Gaulton et al., 2017). The results above show that a RF and FFN model achieve comparable performance on the generic Bemis-Murcko scaffold split of ChEMBL25 data. The root mean squared error of the models were approximately 0.8 pIC50 units suggesting both models are approaching the limit of accuracy given the heterogenous IC50 measurements in ChEMBL25 (Kalliokoski et al., 2013). Feature importance analysis of both models demonstrated that protein sequence features were among the most important features overall. We show that entity embeddings for the categorical ECFP6 Morgan fingerprints can be optimized during FFN training and provide quality performance for drug discovery applications (Guo and Berkhahn, 2016). This allows for feature engineering in a data driven manner and provides an alternative to other methods that seek to derive novel chemical features using convolutional (Ragoza et al., 2017; Ciriano and Bender, 2018; Jiménez et al., 2018; Parks et al., 2019b) and graph-convolutional neural networks (Kearnes et al., 2016; Kipf and Welling, 2017; Torng and Altman, 2019).

Finally, we analyze the utility of conformal prediction to provide prediction intervals to assess model confidence. Conformal prediction was implemented using the standard deviation across the trees of the RF model and Bayesian dropout (Cortés-Ciriano and Bender, 2019a) for the FFN model. Both models generated well-calibrated and efficient confidence intervals on the ChEMBL25 validation set. In addition, we assessed the performance of the RF model in a semi-prospective setting using the recently released GC4 CatS and BACE-1 datasets. Here, we find that performance of the models is in line with the top performing machine learning methods from previous Grand Challenges (Gathiaka et al., 2016; Gaieb et al., 2018, 2019; Parks et al., 2019a), but significantly below the performance on the original validation set. This occurred despite the use of a generic Bemis-Murcko scaffold split to assess model performance retrospectively. However, the prediction intervals from the conformal predictor on the GC4 dataset are well-calibrated both retrospectively and semi-prospectively and thus can serve as a reliable tool to mitigate false positives in hit discovery campaigns and aid compound selection for synthesis during lead optimization.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in the https://drugdesigndata.org/about/datasets, https://chembl.gitbook.io/chembl-interface-documentation/downloads.

AUTHOR CONTRIBUTIONS

CP, ZG, and RA planned the research, performed the analysis, and wrote the paper. CP and ZG performed the research.

FUNDING

This work was supported in part by NIH U01 GM111528 and NIH R01 GM132826 to RA.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2020.00093/full#supplementary-material

REFERENCES

Abel, R., Wang, L., Mobley, D. L., and Friesner, R. A. (2017). A critical review of validation, blind testing, and real- world use of alchemical protein-ligand binding free energy calculations. *Curr. Top. Med. Chem.* 17, 2577–2585. doi: 10.2174/1568026617666170414142131

Cherkasov, A., Muratov, E. N., Fourches, D., Varnek, A., Baskin, I. I., Cronin, M., et al. (2014). QSAR modeling: where have you been?. Where are you going to? *J. Med. Chem.* 57, 4977–5010. doi: 10.1021/jm4004285

Chodera, J. D., Mobley, D. L., Shirts, M. R., Dixon, R. W., Branson, K., and Pande, V. S. (2011). Alchemical free energy methods for drug discovery: progress and challenges. *Curr. Opin. Struct. Biol.* 21, 150–160. doi: 10.1016/j.sbi.2011.01.011

Christ, C. D., and Fox, T. (2014). Accuracy assessment and automation of free energy calculations for drug design. *J. Chem. Inf. Model.* 54, 108–120. doi: 10.1021/ci4004199

Ciriano, I. C., and Bender, A. (2018). kekulescope: improved prediction of cancer cell line sensitivity using convolutional neural networks trained on compound images. arXiv:1811.09036 [cs].

Cortés-Ciriano, I., Ain, Q. U., Subramanian, V., Lenselink, E. B., Méndez-Lucio, O., IJzerman, A. P., et al. (2015a). Polypharmacology modelling using proteochemometrics (PCM): recent methodological developments, applications to target families, and future prospects. *Med. Chem. Commun.* 6, 24–50. doi: 10.1039/C4MD00216D

Cortés-Ciriano, I., and Bender, A. (2019a). Reliable prediction errors for deep neural networks using test-time dropout. *J. Chem. Inf. Model.* 59, 3330–3339. doi: 10.1021/acs.jcim.9b00297

Cortés-Ciriano, I., and Bender, A. (2019b). Concepts and applications of conformal prediction in computational drug discovery. arXiv:1908.03569 [cs, q-bio].

Cortés-Ciriano, I., Bender, A., and Malliavin, T. (2015b). Prediction of PARP inhibition with proteochemometric modelling and conformal prediction. *Mol. Inform.* 34, 357–366. doi: 10.1002/minf.201400165

Cournia, Z., Allen, B., and Sherman, W. (2017). Relative binding free energy calculations in drug discovery: recent advances and practical considerations. *J. Chem. Inf. Model.* 57, 2911–2937. doi: 10.1021/acs.jcim.7b00564

Feinberg, E. N., Sheridan, R., Joshi, E., Pande, V. S., and Cheng, A. C. (2019). Step change improvement in admet prediction with potentialnet deep featurization. arXiv:1903.11789 [cs, stat]. doi: 10.1021/acs.jmedchem.9b02187

Gaieb, Z., Liu, S., Gathiaka, S., Chiu, M., Yang, H., Shao, C., et al. (2018). D3R grand challenge 2: blind prediction of protein–ligand poses, affinity rankings, and relative binding free energies. *J. Comput. Aided Mol. Des.* 32, 1–20. doi: 10.1007/s10822-017-0088-4

Gaieb, Z., Parks, C. D., Chiu, M., Yang, H., Shao, C., Walters, W. P., et al. (2019). D3R grand challenge 3: blind prediction of protein–ligand poses and affinity rankings. *J. Comput. Aided Mol. Des.* 33, 1–18. doi: 10.1007/s10822-018-0180-4

Gao, J., Huang, Q., Wu, D., Zhang, Q., Zhang, Y., Chen, T., et al. (2013). Study on human GPCR–inhibitor interactions by proteochemometric modeling. *Gene* 518, 124–131. doi: 10.1016/j.gene.2012.11.061

Gathiaka, S., Liu, S., Chiu, M., Yang, H., Stuckey, J. A., Kang, Y. N., et al. (2016). D3R grand challenge 2015: evaluation of protein–ligand pose and affinity predictions. *J. Computer-Aided Mol. Design.* 30, 651–668. doi: 10.1007/s10822-016-9946-8

Gaulton, A., Hersey, A., Nowotka, M., Bento, A. P., Chambers, J., Mendez, D., et al. (2017). The ChEMBL database in 2017. *Nucleic Acids Res.* 45, D945–D954. doi: 10.1093/nar/gkw1074

Giblin, K. A., Hughes, S. J., Boyd, H., Hansson, P., and Bender, A. (2018). Prospectively validated proteochemometric models for the prediction of small-molecule binding to bromodomain proteins. *J. Chem. Inf. Model.* 58, 1870–1888. doi: 10.1021/acs.jcim.8b00400

Guo, C., and Berkhahn, F. (2016). Entity embeddings of categorical variables. arXiv:1604.06737 [cs].

Howard, J., and Gugger, S. (2020). Fastai: a layered API for deep learning. *Information.* 11:108. doi: 10.3390/info11020108

Jiménez, J., Škalič, M., and Martínez-Rosell, G., De Fabritiis, G. (2018). KDEEP: protein–ligand absolute binding affinity prediction via 3d-convolutional neural networks. *J. Chem. Inf. Model.* 58, 287–296. doi: 10.1021/acs.jcim.7b00650

Jorgensen, W. L., and Thomas, L. L. (2008). Perspective on free-energy perturbation calculations for chemical equilibria. *J. Chem. Theory Comput.* 4, 869–876. doi: 10.1021/ct800011m

Kalliokoski, T., Kramer, C., Vulpetti, A., and Gedeck, P. (2013). Comparability of mixed IC50 data – a statistical analysis. *PLoS ONE.* 8:e61007. doi: 10.1371/journal.pone.0061007

Kearnes, S., McCloskey, K., Berndl, M., Pande, V., and Riley, P. (2016). Molecular graph convolutions: moving beyond fingerprints. *J. Comput Aided Mol. Des.* 30, 595–608. doi: 10.1007/s10822-016-9938-8

Kim, P., Winter, R., and Clevert, D.-A. (2019). Deep protein-ligand binding prediction using unsupervised learned representations. *ChemRxiv*. doi: 10.26434/chemrxiv.11523117.v1

Kipf, T. N., and Welling, M. (2017). Semi-supervised classification with graph convolutional networks. arXiv:1609.02907 [cs, stat].

Lapins, M., Eklund, M., Spjuth, O., Prusis, P., and Wikberg, J. E. S. (2008). Proteochemometric modeling of HIV protease susceptibility. *BMC Bioinform.* 9:181. doi: 10.1186/1471-2105-9-181

Lenselink, E. B., ten Dijke, N., Bongers, B., Papadatos, G., van Vlijmen, H. W. T., Kowalczyk, W., et al. (2017). Beyond the hype: deep neural networks outperform established methods using a ChEMBL bioactivity benchmark set. *J. Cheminform*. 9:45. doi: 10.1186/s13321-017-0232-0

Mobley, D. L., and Gilson, M. K. (2017). Predicting binding free energies: frontiers and benchmarks. *Annu. Rev. Biophys*. 46, 531–558. doi: 10.1146/annurev-biophys-070816-033654

Mobley, D. L., and Klimovich, P. V. (2012). Perspective: alchemical free energy calculations for drug discovery. *J. Chem. Phys*. 137:230901. doi: 10.1063/1.4769292

Norinder, U., Carlsson, L., Boyer, S., and Eklund, M. (2014). Introducing conformal prediction in predictive modeling. a transparent and flexible alternative to applicability domain determination. *J. Chem. Inf. Model*. 54, 1596–1603. doi: 10.1021/ci5001168

Parks, C., Gaieb, Z., and Amaro, R. (2019b). Machine learning for acute oral system toxicity regression and classification. *ChemRxiv*. doi: 10.26434/chemrxiv.9733973.v1

Parks, C., Gaieb, Z., Chiu, M., Yang, H., Shao, C., Walters, W. P., et al. (2019a). D3R grand challenge 4: blind prediction of protein-ligand poses, affinity rankings, and relative binding free energies. *ChemRxiv*. doi: 10.26434/chemrxiv.11363006

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-Learn: machine learning in python. *J. Machine Learn. Res.* 12, 2825–2830.

Ragoza, M., Hochuli, J., Idrobo, E., Sunseri, J., and Koes, D. R. (2017). Protein–ligand scoring with convolutional neural networks. *J. Chem. Inf. Model*. 57, 942–957. doi: 10.1021/acs.jcim.6b00740

Shafer, G., and Vovk, V. (2008). A tutorial on conformal prediction. *J. Mach Learn Res.* 9, 371–421.

Smith, L. N. (2018). A disciplined approach to neural network hyper-parameters: part 1 – learning rate, batch size, momentum, and weight decay. arXiv:1803.09820 [cs, stat].

Subramanian, V., Prusis, P., Pietilä, L.-O., Xhaard, H., and Wohlfahrt, G. (2013). Visually interpretable models of kinase selectivity related features derived from field-based proteochemometrics. *J. Chem. Inf. Model* 53, 3021–3030. doi: 10.1021/ci400369z

Sun, J., Carlsson, L., Ahlberg, E., Norinder, U., Engkvist, O., and Chen, H. (2017). Applying mondrian cross-conformal prediction to estimate prediction confidence on large imbalanced bioactivity data sets. *J. Chem. Inf. Model*. 57, 1591–1598. doi: 10.1021/acs.jcim.7b00159

Svensson, F., Aniceto, N., Norinder, U., Cortes-Ciriano, I., Spjuth, O., Carlsson, L., et al. (2018). Conformal regression for quantitative structure–activity relationship modeling—quantifying prediction uncertainty. *J. Chem. Inf. Model.* 58, 1132–1140. doi: 10.1021/acs.jcim.8b00054

Svensson, F., Norinder, U., and Bender, A. (2017). Improving screening efficiency through iterative screening using docking and conformal prediction. *J. Chem. Inf. Model*. 57, 439–444. doi: 10.1021/acs.jcim.6b00532

Torng, W., and Altman, R. B. (2019). Graph convolutional neural networks for predicting drug-target interactions. *J. Chem. Inf. Model.* 59, 4131–4149. doi: 10.1021/acs.jcim.9b00628

Tresadern, G., Trabanco, A. A., Pérez-Benito, L., Overington, J. P., van Vlijmen, H. W. T., and van Westen, G. J. P. (2017). Identification of allosteric modulators of metabotropic glutamate 7 receptor using proteochemometric modeling. *J. Chem. Inf. Model* 57, 2976–2985. doi: 10.1021/acs.jcim.7b00338

Wallach, I., and Heifets, A. (2018). Most ligand-based classification benchmarks reward memorization rather than generalization. *J. Chem. Inf. Model.* 58, 916–932. doi: 10.1021/acs.jcim.7b00403

Winter, R., Montanari, F., Noé, F., and Clevert, D.-A. (2019). Learning continuous and data-driven molecular descriptors by translating equivalent chemical representations. *Chem. Sci*. 10, 1692–1701. doi: 10.1039/C8SC04175J

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: 30 June 2020
doi: 10.3389/fmolb.2020.00132

Time-Lagged t-Distributed Stochastic Neighbor Embedding (t-SNE) of Molecular Simulation Trajectories

Vojtěch Spiwok[1]* *and Pavel Kříž*[2]

[1] Department of Biochemistry and Microbiology, University of Chemistry and Technology, Prague, Czechia, [2] Department of Mathematics, University of Chemistry and Technology, Prague, Czechia

Molecular simulation trajectories represent high-dimensional data. Such data can be visualized by methods of dimensionality reduction. Non-linear dimensionality reduction methods are likely to be more efficient than linear ones due to the fact that motions of atoms are non-linear. Here we test a popular non-linear t-distributed Stochastic Neighbor Embedding (t-SNE) method on analysis of trajectories of 200 ns alanine dipeptide dynamics and 208 μs Trp-cage folding and unfolding. Furthermore, we introduce a time-lagged variant of t-SNE in order to focus on rarely occurring transitions in the molecular system. This time-lagged t-SNE efficiently separates states according to distance in time. Using this method it is possible to visualize key states of studied systems (e.g., unfolded and folded protein) as well as possible kinetic traps using a two-dimensional plot. Time-lagged t-SNE is a visualization method and other applications, such as clustering and free energy modeling, must be done with caution.

Keywords: molecular dynamics, dimensionality reduction, trajectory analysis, tSNE, Time-lagged Independent Component Analysis

OPEN ACCESS

Edited by:
Pratyush Tiwary, University of Maryland, College Park, United States

Reviewed by:
Carlo Camilloni, University of Milan, Italy
Steffen Wolf, University of Freiburg, Germany
James Joseph McCarty, Western Washington University, United States

***Correspondence:**
Vojtěch Spiwok
spiwokv@vscht.cz

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

Received: *06 March 2020*
Accepted: *03 June 2020*
Published: *30 June 2020*

Citation:
Spiwok V and Kříž P (2020) Time-Lagged t-Distributed Stochastic Neighbor Embedding (t-SNE) of Molecular Simulation Trajectories. Front. Mol. Biosci. 7:132. doi: 10.3389/fmolb.2020.00132

1. INTRODUCTION

The main goal of molecular simulations is identification of key states of studied systems and building thermodynamic and kinetic models of transitions between these states. Identification of key states is often based on some numerical descriptors known as collective variables. Distance between two atoms can be seen as one of the simplest collective variables. It can be used, for example, to distinguish between the bound and unbound state in a simulation of protein-ligand interaction. For some more complex processes it is necessary to use more complex collective variables.

Collective variables are in fact dimensionality reduction methods because they represent high dimensional structure of a molecular system using few numerical descriptors. It is therefore no surprise that general linear and non-linear dimensionality reduction methods have been applied on molecular simulation trajectories. Namely, principal component analysis (Amadei et al., 1993; Spiwok et al., 2007; Sutto et al., 2010) and its dihedral version (Mu et al., 2005), diffusion maps (Ferguson et al., 2010, 2011), sketch map (Ceriotti et al., 2011; Tribello et al., 2012), Isomap (Das et al., 2006; Brown et al., 2008; Spiwok and Králová, 2011), autoencoders (Chen and Ferguson, 2018), t-SNE (van der Maaten and Hinton, 2008; Duan et al., 2013; Tribello and Gasparotto, 2019) and others (Plaku et al., 2007; Stamati et al., 2010; Noé and Clementi, 2015) have been tested in analysis of trajectories, data compression or sampling enhancement.

Advantage of non-linear dimensionality reduction methods is their ability to describe more variance in data compared to linear methods with the same number of dimensions. This is especially true for t-distributed Stochastic Neighbor Embedding (t-SNE) (van der Maaten and Hinton, 2008). This method became highly popular in many fields, including data science, bioinformatics, and computational linguistics.

There are two features of t-SNE that contributed to its success. First, t-SNE converts high-dimensional points into low-dimensional points in a way to reproduce their proximity rather than distance. For example, for a bioinformatician analyzing genomic data to develop genomics-based diagnosis it is important that samples with the same diagnosis are close to each other after dimensionality reduction. It is unimportant how distant are samples with different diagnosis, provided that they are distant enough. In t-SNE the distances in the high-dimensional space $D_{ij} = \|X_i - X_j\|$ are converted into proximities p_{ij} as:

$$p_{ij} = \frac{\exp(-D_{ij}^2/2\sigma_i^2)}{\sum_{k\neq i}\exp(-D_{ik}^2/2\sigma_i^2)}, \quad (1)$$

where σ_i^2 is the variance of a Gaussian centered on a datapoint X_i (discussed later). The matrix of proximities is then symmetrized. Next, proximities in the low-dimensional space q_{ij} are calculated from distances in the low-dimensional space d_{ij} as:

$$q_{ij} = \frac{(1+d_{ij}^2)^{-1}}{\sum_{k\neq i}(1+d_{ik}^2)^{-1}}. \quad (2)$$

Finally, positions of points in the low-dimensional space are optimized to minimize Kullback-Leibler divergences of p_{ij} and q_{ij} (a sort of a distance between proximities p and q).

The second advantage of t-SNE lies in the fact that it unifies density of low-dimensional points in the output space. This feature, which can be controlled by a parameter called perplexity, makes visual representation of points more effective. Perplexity is related to variances σ_i^2 of Gaussians centered on datapoints X_i. Unification of densities is done by different variances σ_i^2. The user can specify the value of perplexity. t-SNE searches for optimal values of σ_i^2 in order to produce values of $2^{-\sum_j p_{ji}\log_2 p_{ji}}$ to match the predefined perplexity. Low perplexity (e.g., 5) forces focus on local structure of the input data whereas larger perplexity (e.g., 50) takes more global structure into the account. As discussed later, this feature improves visualization by t-SNE but at the same time it complicates application in situations when preservation of densities is required.

Disadvantage of application of general dimensionality reduction methods on molecular simulation trajectories is that these methods pick the most intensive (in terms of changes of atomic coordinates) motions in the system. However, such motions are often not interesting. For example, such intensive motions may represent motions of disordered loops or terminal chains in proteins.

Instead, for building of thermodynamic and kinetic models or to enhance sampling it is useful to extract motions that occur most rarely, i.e., those with the highest barriers. This can be done by Time-lagged Independent Component Analysis (TICA) (Molgedey and Schuster, 1994; Perez-Hernandez et al., 2013; Schwantes and Pande, 2013). TICA extracts the most rarely occurring transitions in the molecular system because it correlates the state of the system with the state of the same system after a short delay (lag). This lag can be controlled.

Here we attempt to join the advantages of t-SNE and TICA into a single method of time-lagged t-SNE. The method was tested on two molecular trajectories—on 200 ns simulation of alanine dipeptide and 208.8 μs simulation of Trp-cage mini-protein folding and unfolding (trajectory kindly provided by DE Shaw Research) (Lindorff-Larsen et al., 2011).

2. METHODS

Time-lagged t-SNE is inspired by implementation of TICA using the AMUSE algorithm (Hyvarinen et al., 2001). We start with atomic coordinates $\mathbf{X}(t)$ recorded over time t. First, coordinates are superimposed to reference coordinates of the system to eliminate translational and rotational motions. After that, time-averaged coordinates are subtracted, leading to atomic displacements $\mathbf{X}'(t)$. Next, its covariance matrix is calculated as:

$$C_{ij}^{\mathbf{X}'} = \langle X_i'(t)X_j'(t)\rangle, \quad (3)$$

where i and j are indexes of atomic coordinates and $\langle\rangle$ denotes time-averaging. Next, covariance matrix is decomposed to a diagonal matrix with eigenvalues $\lambda^{\mathbf{X}'}$ (the square matrix with eigenvalues on diagonal and zeros elsewhere) and eigenvectors $\mathbf{W}^{\mathbf{X}'}$ (the matrix with eigenvectors as columns):

$$\mathbf{C}^{\mathbf{X}'}\mathbf{W}^{\mathbf{X}'} = \mathbf{W}^{\mathbf{X}'}\lambda^{\mathbf{X}'}. \quad (4)$$

Coordinates $\mathbf{X}'(t)$ are transformed onto principal components and normalized by roots of eigenvalues (space-whitening the signal) to get flattened normalized projections:

$$\mathbf{Y}(t) = (\lambda^{\mathbf{X}'})^{-1/2}((\mathbf{W}^{\mathbf{X}'})^T\mathbf{X}'(t)). \quad (5)$$

A time-lagged covariance matrix is calculated as:

$$C_{ij}^{\mathbf{Y}} = \langle Y_i(t)Y_j(t+\tau)\rangle, \quad (6)$$

where τ is an adjustable time lag. Because the matrix C is non-symmetric it must be symmetrized as:

$$\mathbf{C}_{sym}^{\mathbf{Y}} = 1/2(\mathbf{C}^{\mathbf{Y}} + (\mathbf{C}^{\mathbf{Y}})^T). \quad (7)$$

Next, this symmetric matrix is decomposed to eigenvalues $\lambda^{\mathbf{Y}}$ and eigenvectors $\mathbf{W}^{\mathbf{Y}}$:

$$\mathbf{C}_{sym}^{\mathbf{Y}}\mathbf{W}^{\mathbf{Y}} = \mathbf{W}^{\mathbf{Y}}\lambda^{\mathbf{Y}}. \quad (8)$$

Finally, $\mathbf{Y}(\mathbf{t})$ are transformed onto principal components and expanded by eigenvalues:

$$\mathbf{Z} = (\lambda^{\mathbf{Y}})^{1/2}((\mathbf{W}^{\mathbf{Y}})^T\mathbf{Y}). \quad (9)$$

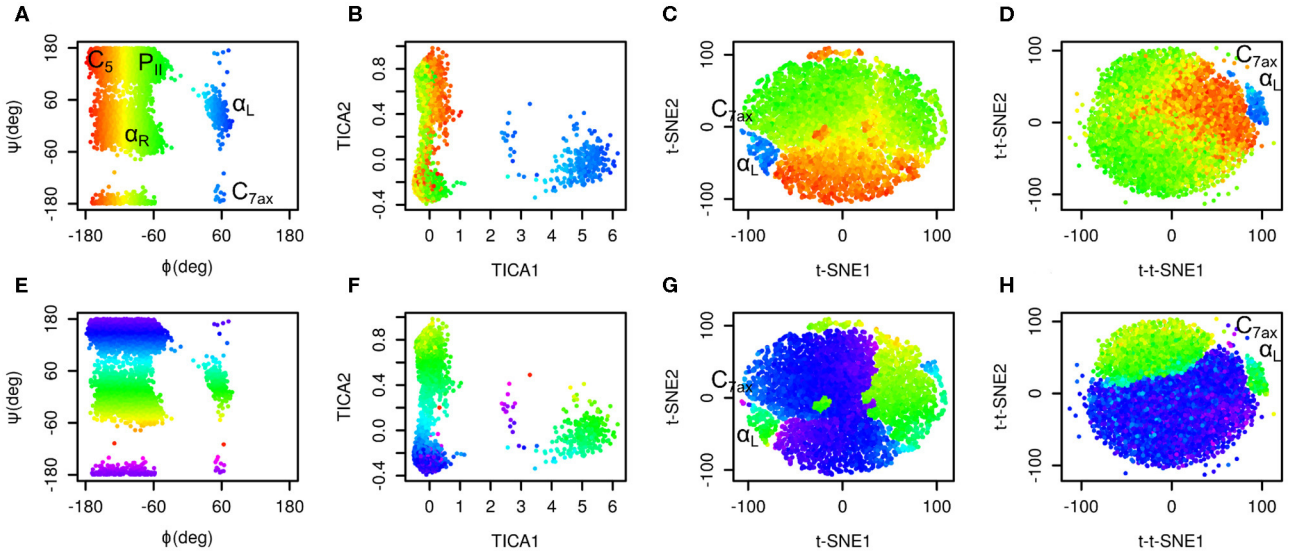


FIGURE 1 | Time-lagged t-SNE (t-t-SNE) applied on 200 ns simulation of alanine dipeptide in water. Conformations sampled in the simulations were projected into the space of Ramachandran torsions ϕ and ψ **(A,E)**, TICA coordinates **(B,F)**, t-SNE **(C,G)** and time-lagged t-SNE **(D,H)**. Points are colored by Ramachandran torsion ϕ **(A–D)** and ψ **(E–H)**.

This step expands distances in directions with highest autocorrelations, which represent directions of rarely occurring transitions.

It is possible to use certain number of eigenvectors with highest eigenvalues instead of all eigenvectors. This selection may be driven by relaxation time decays (see Wehmeyer et al., 2019) but this is out of scope of this article.

t-SNE can be applied on distances between simulation snapshots calculated in the space of **Z** as:

$$D_{t,t'} = \|\mathbf{Z}(t) - \mathbf{Z}(t')\|. \tag{10}$$

Low-dimensional embeddings obtained in this step are further referred to as time-lagged t-SNE coordinates (t-t-SNE). For the sake of comparison, low-dimensional embedding obtained by standard TICA (without t-SNE step) and standard t-SNE (without TICA step) were also calculated and are further referred to as TICA coordinates and t-SNE coordinates, respectively. t-SNE and t-t-SNE coordinates are unit-free because they are set in order to fit the corresponding unit-free proximities (both D and σ in Equation 1 are measured in the same units). It must be kept in mind that t-SNE and t-t-SNE use random initiation of low-dimensional points, so recalculation leads to a different plot.

All analyses were done by programs written in Python with MDtraj (McGibbon et al., 2015) (for reading trajectories), PyEMMA (Wehmeyer et al., 2019) (for testing of algorithms), numpy (Oliphant, 2006) (to implement AMUSE algorithm) and scikit-learn (Pedregosa et al., 2011) (to run t-SNE) libraries. It is available at GitHub (https://github.com/spiwokv/tltsne) and using PyPI.

The trajectory of alanine dipeptide was obtained by unbiased 200 ns molecular dynamics simulation of a system containing alanine dipeptide and 874 TIP3P (Jorgensen et al., 1983) water molecules in Gromacs (Abraham et al., 2015). It was modeled by Amber99SB-ILDN force field (Lindorff-Larsen et al., 2010). Simulation step was set to 2 fs and all bonds were constrained by LINCS algorithm (Hess et al., 1997). Electrostatic interactions were treated by particle-mesh Ewald method (Darden et al., 1998). Temperature was kept constant (NVT ensemble) at 300 K by V-rescale thermostat (Bussi et al., 2007).

The trajectory of Trp-cage folding and unfolding was kindly provided by DE Shaw Research.

3. RESULTS

The method was tested on two molecular systems—on alanine dipeptide and Trp-cage. In order to test time-lagged t-SNE we compare time-lagged t-SNE with standard t-SNE and TICA.

3.1. Alanine Dipeptide

Time-lagged t-SNE was first applied on a trajectory of alanine dipeptide without water and hydrogen atoms. It is important to remove hydrogen atoms because rotamers of methyl groups by approx. 120 deg are mathematically distinguishable but chemically identical. The trajectory was sampled every 20 ps (10,001 snapshots). Time lag τ was set to 3 frames (60 ps). The value of perplexity was set to 3.0 and Euclidean space was used to calculate the distance matrix **D**.

The value of lag time was chosen based on TICA results. Similar calculations with lag time set to 1 to 12 steps show that lag time set to 1–7 works well on a simple system such as alanine dipeptide (see **Supplementary Material**). All eigenvectors $\mathbf{W}^{\mathbf{Y}}$ were used in Equation (9).

The results are depicted in **Figure 1**. Plots in the space of Ramachandran torsions show that all relevant conformations of alanine dipeptide were sampled. Plots in the space of TICA coordinates show that rotation around ϕ is the slowest and

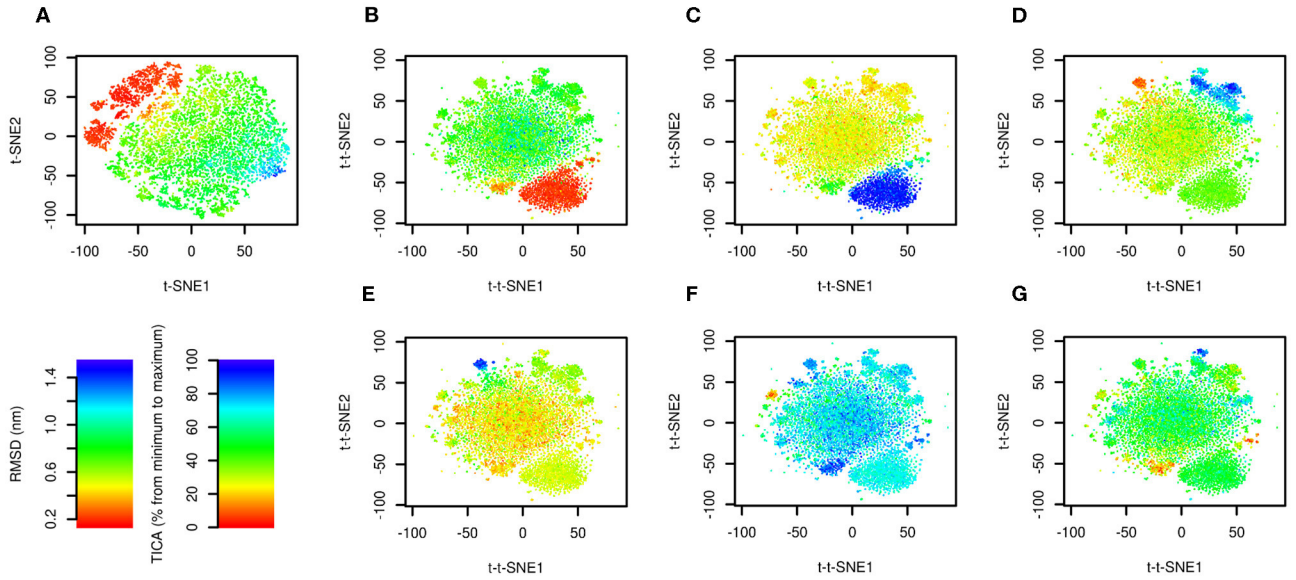


FIGURE 2 | Time-lagged t-SNE (t-t-SNE) applied on 208.8 μs of Trp-cage folding and unfolding. The trajectory was analyzed by t-SNE **(A)** and time-lagged t-SNE **(B–G)**. Points are colored by RMSD from the native structure **(A,B)** and by the first **(C)**, second **(D)**, third **(E)**, fourth **(F)**, and fifth **(G)** TICA coordinate.

rotation around ψ is the second slowest motion in the studied system (slowest in terms of number of occurrences).

Plots in the space of t-SNE coordinates have a circular shape cut into multiple pieces by borders between different conformers. These plots show a limitation of conventional t-SNE, which is an improper resolution of conformations. Namely, there is a green island in the blue area of the plot colored by ϕ values (G).

Time-lagged t-SNE (t-t-SNE) does not suffer this problem. The blue area in the plot generated by time-lagged tSNE is continuous and does not contain any islands of conformations with positive ϕ values (H). This can be explained by the fact that introduction of a time lag into t-SNE causes higher separation of key conformations of alanine dipeptide.

One feature is common to the original t-SNE as well as our time-lagged variant. This is the fact that t-SNE flattens the distribution of points in the output space. This results in an almost uniform distribution of points in each minimum.

It is possible to calculate a histogram of some molecular collective variable or collective variables and convert it into a free energy surface. Most common interpretation of such free energy surfaces is that deep minima correspond to stable states, whereas shallow minima correspond to unstable states. This approach can be applied for conventional descriptors, such as Ramachandran angles of alanine dipeptide. However, due to flattening of distribution of points by t-SNE or by time-lagged t-SNE such free energy surface is relatively flat. Populations of different states can be estimated from areas of free energy minima rather than from their depths. In general, time-lagged t-SNE (as well as t-SNE) must be used with caution when applied to identify metastable states and to calculate free energy surfaces.

3.2. Trp-Cage

t-SNE and time-lagged t-SNE analysis was performed on the trajectory of Trp-cage folding and unfolding sampled every 20 ns (10,440 snapshots). Lag time was set to three frames (60 ns). Similarly to alanine dipeptide, lag time was chosen based on TICA analysis. Comparison of embeddings calculated for lag time set to 1, 2, 3, 4, 5, 10, 15, and 20 (in number of frames) shows that lag time 1–5 works well (see **Supplementary Material**).

Perplexity was set to 10.0. Several values were tested and perplexity set to 10 performs well in terms of the focus on local vs. global structure of data. **Supplementary Material** contains the results obtained for perplexity 5, 10, 20, 50, and 100. These results indicate that time-lagged t-SNE is relatively robust in terms of choice of perplexity and perplexity 10 and higher perform well.

Initial analysis by time-lagged t-SNE resulted in a circular plot with multiple points located outside clusters on the edges of the circle. This indicates that there are many points with high distances $D_{t,t'}$. In order to eliminate these points we reduced the number of eigenvectors $\mathbf{W}^{\mathbf{Y}}$ to top 50 eigenvectors (option `-maxpcs` in the code).

The results are depicted in **Figure 2**. **Figure 2A** shows the trajectory analyzed by conventional t-SNE colored by RMSD from the native structure (PDB ID: 1l2y, Neidigh et al., 2002). There is a clear relationship between t-SNE coordinates, in particular t-SNE1, and RMSD. The native structure (in red) forms a cluster in the top left corner of the plot. Structures with high RMSD (in blue) are characterized by highest values of t-SNE1.

The trajectory analyzed by time-lagged t-SNE colored by RMSD is depicted in **Figure 2B**. Similarly to **Figure 2A** the native structure forms a distinct cluster. In contrast to the conventional t-SNE, structures with high values of RMSD are scattered in the

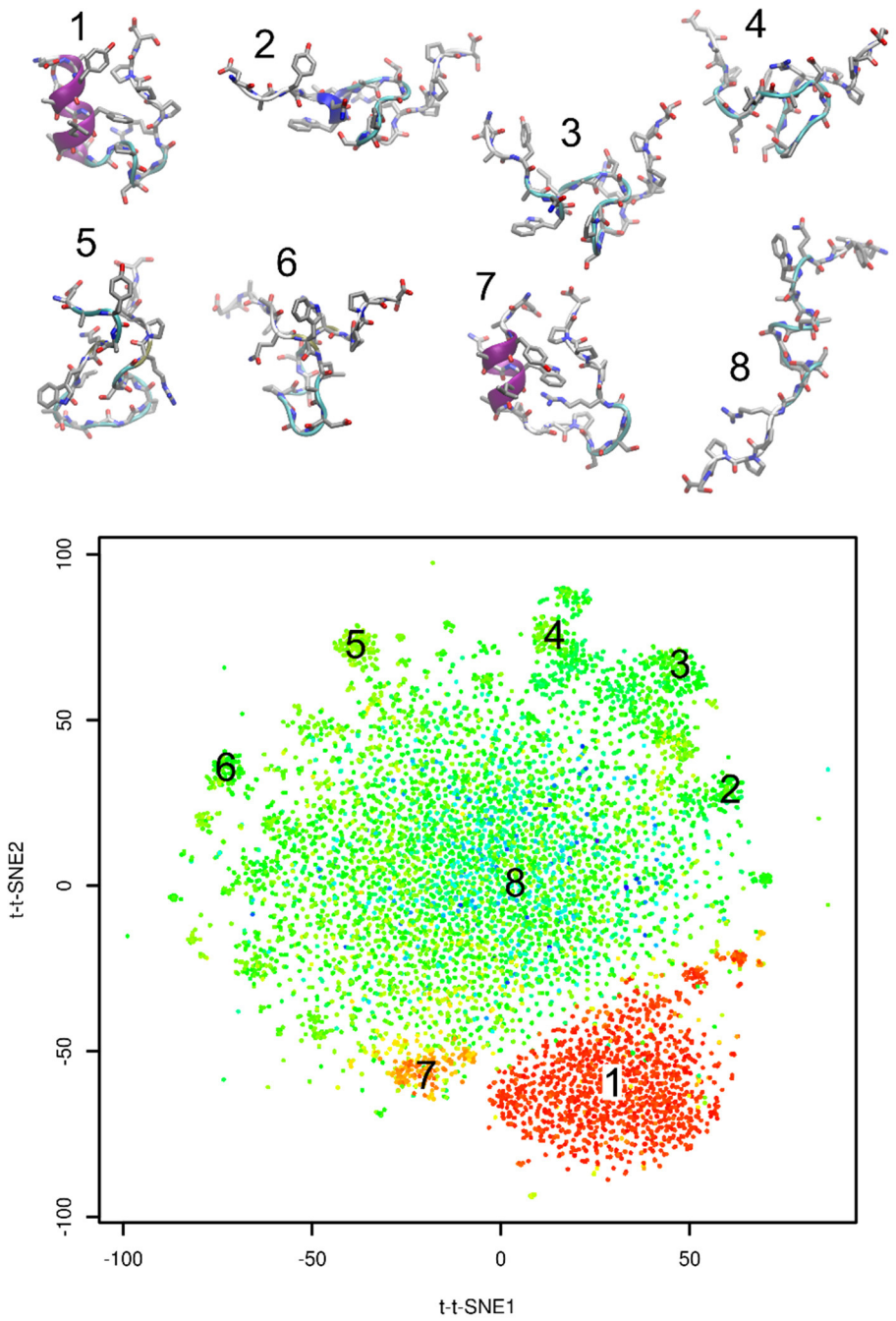


FIGURE 3 | Representative structures projected onto time-lagged t-SNE embeddings. Plot is colored by RMSD from the native structure (as in **Figure 2B**).

large cluster in the center. This indicates that transitions between high-RMSD structures are fast.

Figures 2C–G show the same plots colored by TICA coordinates. The first TICA coordinate (**Figure 2C**) distinguishes folded and unfolded structures. Plots colored by other TICA coordinates (**Figures 2D–G**) in most cases show a red or blue clusters on edges of the plot. This shows that time-lagged t-SNE captures rarely occurring transitions characterized by TICA, but more efficiently than TICA itself, because these motions can be depicted in a single plot.

Figure 3 shows representative structures of Trp-cage from the simulation trajectory projected onto time-lagged t-SNE embeddings. Structure 1 is the native structure. Structure 7 is a known near-native structure. Structures 2–6 were sampled from clusters on peripheral areas of time-lagged t-SNE embeddings. Finally, structure 8 was taken from the origin of the plot. Visual

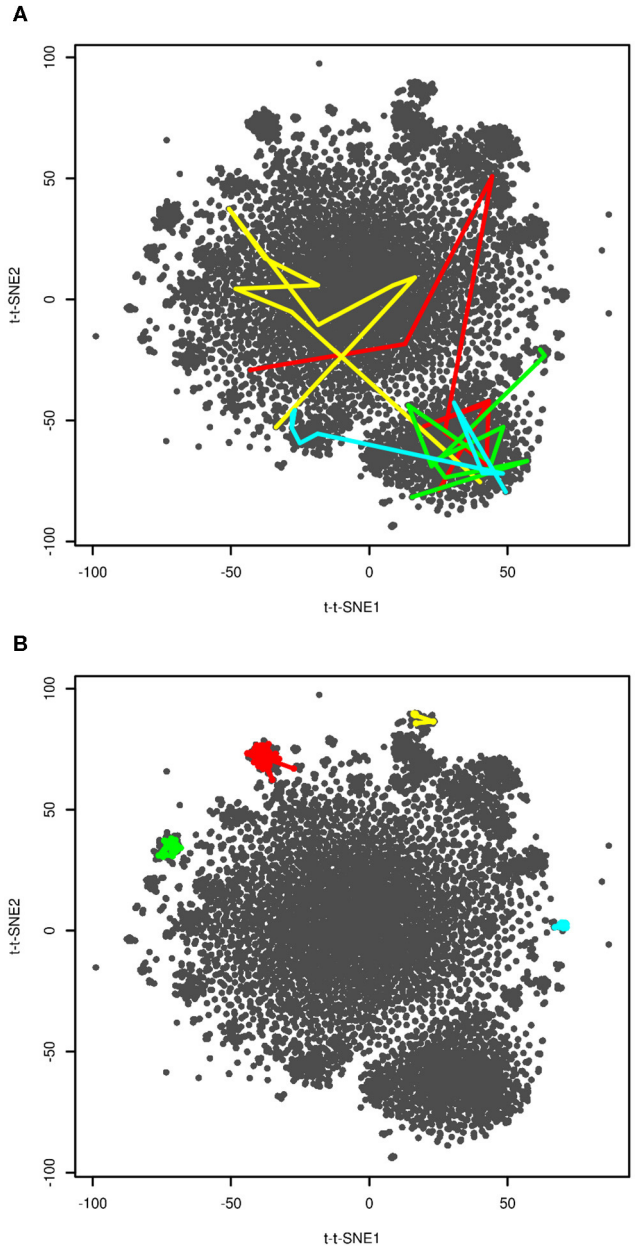


FIGURE 4 | Visualization of folding events **(A)** and kinetic traps **(B)**. Four selected folding events (at approx. 20, 40, 104 and 206 μs in red, yellow, green, and blue, respectively) are depicted as last 10 frames (200 ns) before reaching t-t-SNE2 < -75. Four selected kinetic traps on edges of the plot are depicted as connected series of snapshots. These regions were sampled for 800 (red), 140 (yellow), 400 (green), and 460 (blue) ns.

inspection indicates that structures 2–6 may be kinetic traps of Trp-cage folding, because these structures are characterized by formation of numerous non-native hydrogen bonds and other interactions. Also the near-native structure 7 is likely to be a kinetic trap of Trp-cage folding.

In order to further interpret the plot we visualized four selected folding events. They are depicted in **Figure 4A**. These plots show snapshots sampled last 200 ns (10 snapshots) before folding. Unfortunately, we were not able to provide higher resolution of time, because this would require either analysis of a higher number of snapshots or recalculation of time-lagged t-SNE. The former was not possible due to computational costs, the latter due to impossibility of calculation of time-lagged t-SNE on out-of-sample structures (discussed later). Despite limited resolution of time, the plot shows that unfolded and folded structures are clearly separated. The fact that some folding processes passed clusters on edges of the plot close to the native structure may indicate that these clusters are near-native metastable states.

In previous paragraphs we interpreted clusters on edges of the plot (structures 2–6 in **Figure 3**). We investigated how long the system stayed in these regions. The results are shown in **Figure 4B**. The system stayed in these regions for 140–800 ns. This supports our interpretation of these regions as kinetic traps. Interestingly, all four regions depicted in **Figure 4B** were sampled multiple times in the simulation.

4. DISCUSSION

Dimensionality reduction methods are frequently used to analyze data from biomolecular simulations. Linear methods such as PCA have been used for decades, whereas application of non-linear methods is relatively new. Various linear and non-linear dimensionality reduction methods have various advantages and disadvantages.

PCA and other linear methods are easy to use (no additional parameters have to be set), they realistically map densities of states from the high-dimensional to low-dimensional space and it is straightforward to calculate low-dimensional embedding for a new out-of-sample structure. On the other hand, their performance in visualization is low because they usually require three or more dimensions to separate key states of the studied system.

Non-linear methods perform much better in dimensionality reduction but mapping of densities may be distorted (this is the case of t-SNE and its time-lagged variant, which tend to flatten the output densities) and calculation of low-dimensional embeddings for a new out-of-sample structure is complicated. t-SNE is useful specially for visualization purposes.

Comparison of t-SNE and time-lagged t-SNE shows a great advantage of our variant. **Figure 2A** shows that t-SNE coordinates correlate with RMSD from the native structure. The yellow-green-blue cloud of non-native structures in this plot represents a pool of non-native conformations in which short-living and long-living states overlap. On the other hand, in the time-lagged t-SNE there are short-living states in the center and long-living states, including the native state, are located on the edges of the plot. In a single plot it is possible to distinguish multiple key long-living states.

There is a disadvantage of time-lagged methods in their dependence on the choice of lag time. Choice of lag time for time-lagged t-SNE was driven by TICA analysis. Values of 3 frames (60 ps, 0.03% of the whole trajectory) for alanine dipeptide and

3 frames (60 ns, 0.029% of the whole trajectory) for Trp-cage led to visually plausible low dimensional embeddings. This indicates that 0.03% of trajectory size is a good initial choice of lag time.

Another disadvantage of time-lagged t-SNE is in distortion of densities and impossibility to easily calculate low-dimensional embeddings for a new out-of-sample structure. As an alternative to time-lagged t-SNE it is possible to use time-lagged autoencoders recently reported by Wehmeyer and Noé (2018). Autoencoders are feed-forward neural networks with an hourglass-like architecture. The input signal (atomic coordinates or other features) from the input layer are transformed via hidden layers into the central bottleneck layer. Next, the signal from the bottleneck layer is transformed via hidden layers into the output layers. Parameters of the network are trained to obtain agreement between the input and output signal. The signal in the bottleneck layer represents a non-linear low-dimensional representation of the input signal. Unlike classical autoencoders, time-lagged autoencoders focus on the most rarely occurring transitions, not on the most intensive motions (Wehmeyer and Noé, 2018).

The clear advantage of autoencoders and their time-lagged variant is the possibility to calculate low-dimensional embeddings for a new out-of-sample structure. Extensive testing of time-lagged autoencoders in the original article (Wehmeyer and Noé, 2018) was possible owing to this fact. Time-lagged autoencoders can be trained on a training set and tested on a validation set, i.e., they can be evaluated by cross-validation. Furthermore, they can be trained on a small training set and then applied on a large set of input data. This is efficient since the training part is in general significantly more expensive than the calculation of embeddings on out-of-sample structures. Time-lagged autoencoders are useful for pre-processing of structural data for building of Markov state models.

There are limited options for calculation of t-SNE low-dimensional embeddings for out-of-sample structures. Therefore, t-SNE and time-lagged t-SNE are not suitable for pre-processing of the structural data. We see the advantage of time-lagged t-SNE (similarly to t-SNE) in visualization.

Time-lagged t-SNE in the current implementation also cannot be used as collective variables in simulations using bias force or bias potential because these methods require on-the-fly calculation of low-dimensional embeddings and their derivatives with respect to atomic coordinates. However, there are tools to approximate such low-dimensional embeddings (Spiwok and Králová, 2011; Sultan and Pande, 2018; Trapl et al., 2019).

One of key features of t-SNE is that it can reconstruct proximities and not distances in the low-dimensional output space. In time-lagged t-SNE this means that states separated by low energy barriers are close to each other. States separated by large energy barriers are far from each other, but time-lagged t-SNE does not attempt to preserve their distances accurately. This means that two close points in the time-lagged t-SNE plot can be connected by an energetically favorable path.

Another key feature of t-SNE is perplexity and the fact that t-SNE flattens the distribution of points in the output space. This is useful for visualization. For this reason t-SNE (as well as time-lagged t-SNE) must be used with caution as a pre-processing for calculation of free energy surfaces and for clustering. t-SNE can also create artificial clusters when perplexity is not set properly.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in the https://github.com/spiwokv/tltsne.

AUTHOR CONTRIBUTIONS

Both authors developed the method and wrote the manuscript. VS wrote codes and run simulations and analysis.

FUNDING

This work was funded by COST action OpenMultiMed (CA15120, Ministry of Education, Youth and Sports of the Czech Republic LTC18074), Czech Science Foundation (19-16857S), and Czech National Infrastructure for Biological Data (ELIXIR CZ, Ministry of Education, Youth and Sports of the Czech Republic LM2015047).

ACKNOWLEDGMENTS

Authors would like to thank D. E. Shaw Research for data used in this work.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2020.00132/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

Amadei, A., Linssen, A. B., and Berendsen, J. (1993). H. Essential dynamics of proteins. *Prot. Struct. Funct. Bioinform.* 17, 412–425. doi: 10.1002/prot.340170408

Brown, W. M., Martin, S., Pollock, S. N., Coutsias, E. A., and Watson, J.-P. (2008). Algorithmic dimensionality reduction for molecular structure analysis. *J. Chem. Phys.* 129:064118. doi: 10.1063/1.2968610

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

Ceriotti, M., Tribello, G. A., and Parrinello, M. (2011). Simplifying the representation of complex free-energy landscapes using sketch-map. *Proc. Natl. Acad. Sci. U.S.A.* 108, 13023–13028. doi: 10.1073/pnas.1108486108

Chen, W., and Ferguson, L. A. (2018). Molecular enhanced sampling with autoencoders: on-the-fly collective variable discovery and accelerated

free energy landscape exploration. *J. Comput. Chem*. 39, 2079–2102. doi: 10.1002/jcc.25520

Darden, T., York, D., and Pedersen, L. (1998). Particle mesh Ewald: An N · log(N) method for Ewald sums in large systems. J. Chem. Phys. 98:10089. doi: 10.1063/1.464397

Das, P., Moll, M., Stamati, H., Kavraki, L. E., and Clementi, C. (2006). Low-dimensional, free-energy landscapes of protein-folding reactions by nonlinear dimensionality reduction. *Proc. Natl. Acad. Sci. U.S.A*. 103, 9885–9890. doi: 10.1073/pnas.0603553103

Duan, M., Fan, J., Li, M., Han, L., and Huo, S. (2013). Evaluation of dimensionality-reduction methods from peptide folding-unfolding simulations. *J. Chem. Theory Comput*. 9, 2490–2497. doi: 10.1021/ct400052y

Ferguson, A. L., Panagiotopoulos, A. Z., Debenedetti, P. G., and Kevrekidis, G. I. (2010). Systematic determination of order parameters for chain dynamics using diffusion maps. *Proc. Natl. Acad. Sci. U.S.A*. 107, 13597–13602. doi: 10.1073/pnas.1003293107

Ferguson, A. L., Panagiotopoulos, A. Z., Kevrekidis, I. G., and Debenedetti, P. G. (2011). Nonlinear dimensionality reduction in molecular simulation: the diffusion map approach. *Chem. Phys. Lett*. 509, 1–11. doi: 10.1016/j.cplett.2011.04.066

Hess, B., Bekker, H., Berendsen, H. J. C., and Fraaije, J. G. E. M. (1997). LINCS: a linear constraint solver for molecular simulations. *J. Comp. Chem*. 18, 1463–1472. doi: 10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H

Hyvarinen, A., Karhunen, J., and Oja, E. (2001). *Independent Component Analysis*. New York, NY: John Wiley & Sons Int. doi: 10.1002/0471221317

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

Lindorff-Larsen, K., Piana, S., Dror, R. O., and Shaw, A. D. (2011). How fast-folding proteins fold. *Science* 334, 517–520. doi: 10.1126/science.1208351

Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J. L., O Dror, R., et al. (2010). Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* 78, 1950–1958. doi: 10.1002/prot.22711

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. *Biophys. J*. 109, 1528–1532. doi: 10.1016/j.bpj.2015.08.015

Molgedey, L., and Schuster, G. H. (1994). Separation of a mixture of independent signals using time delayed correlations. *Phys. Rev. Lett*. 72, 3634–3637. doi: 10.1103/PhysRevLett.72.3634

Mu, Y., Nguyen, P. H., and Stock, G. (2005). Energy landscape of a small peptide revealed by dihedral angle principal component analysis. *Prot. Struct. Funct. Bioinform*. 58, 45–52. doi: 10.1002/prot.20310

Neidigh, J. W., Fesinmeyer, R. M., and Andersen, N. H. (2002). Designing a 20-residue protein. *Nat. Struct. Biol*. 9, 425–430. doi: 10.1038/nsb798

Noé, F., and Clementi, C. (2015). Kinetic distance and kinetic maps from molecular dynamics simulation. *J. Chem. Theory Comput*. 11, 5002–5011. doi: 10.1021/acs.jctc.5b00553

Oliphant, T. E. (2006). *A Guide to NumPy*. Spanish Fork, UT: Trelgol Publishing.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res*. 12, 2825–2830.

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

Plaku, E., Stamati, H., Clementi, C., and Kavraki, L. E. (2007). Fast and reliable analysis of molecular motion using proximity relations and dimensionality reduction. *Prot. Struct. Funct. Bioinform*. 67, 897–907. doi: 10.1002/prot.21337

Schwantes, C. R., and Pande, S. V. (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

Spiwok, V., and Králová, B. (2011). Metadynamics in the conformational space nonlinearly dimensionally reduced by Isomap. *J. Chem. Phys*. 135, 224504 doi: 10.1063/1.3660208

Spiwok, V., Lipovová, P, and Králová, B. (2007). Metadynamics in essential coordinates: free energy simulation of conformational changes. *J. Phys. Chem. B* 111, 3073–3076. doi: 10.1021/jp068587c

Stamati, H., Clementi, C., and Kavraki, L. E. (2010). Application of nonlinear dimensionality reduction to characterize the conformational landscape of small peptides. *Prot. Struct. Funct. Bioinform*. 78, 223–235. doi: 10.1002/prot.22526

Sultan, M. M., and Pande, S. V. (2018). Automated design of collective variables using supervised machine learning. *J. Chem. Phys*. 149:094106. doi: 10.1063/1.5029972

Sutto, L., Dabramo, M., and Gervasio, F. L. (2010). Comparing the efficiency of biased and unbiased molecular dynamics in reconstructing the free energy landscape of met-enkephalin. *J. Chem. Theory Comput*. 6, 3640–3646. doi: 10.1021/ct100413b

Trapl, D., Horvacanin, I., Mareska, V., Ozcelik, F., Unal, G., and Spiwok, V. (2019). Anncolvar: approximation of complex collective variables by artificial neural networks for analysis and biasing of molecular simulations. *Front. Mol. Biosci*. 6:25. doi: 10.3389/fmolb.2019.00025

Tribello, G. A., Ceriotti, M., and Parrinello, M. (2012). Using sketch-map coordinates to analyze and bias molecular dynamics simulations. *Proc. Natl. Acad. Sci. U.S.A*. 109, 5196–5201 doi: 10.1073/pnas.1201152109

Tribello, G. A., and Gasparotto, P. (2019). Using dimensionality reduction to analyze protein trajectories *Front. Mol. Biosci*. 6:46 doi: 10.3389/fmolb.2019.00046

van der Maaten, L. J. P., and Hinton, G. E. (2008). Visualizing high-dimensional data using t-SNE. *J. Mach. Learn. Res*. 9, 2579–2605.

Wehmeyer, C., and Noé, F. (2018). Time-lagged autoencoders: deep learning of slow collective variables for molecular kinetics. *J. Chem. Phys*. 148:241703. doi: 10.1063/1.5011399

Wehmeyer, C., Scherer, M. K., Hempel, T., Husic, B. E., Olsson, S., Noé, F. (2019). Introduction to Markov state modeling with the PyEMMA software. *Living J. Comp. Mol. Sci*. 1:5965. doi: 10.33011/livecoms.1.1.5965

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 September 2020
doi: 10.3389/fmolb.2020.576689

Hybrid MM/CG Webserver: Automatic Set Up of Molecular Mechanics/Coarse-Grained Simulations for Human G Protein-Coupled Receptor/Ligand Complexes

Jakob Schneider[1,2,3], Rui Ribeiro[4], Mercedes Alfonso-Prieto[1,5], Paolo Carloni[1,2,3] and Alejandro Giorgetti[1,4]**

[1] Computational Biomedicine, Institute for Advanced Simulation IAS-5/Institute for Neuroscience and Medicine INM-9, Forschungszentrum Jülich GmbH, Jülich, Germany, [2] JARA-Institute: Molecular Neuroscience and Neuroimaging, Institute for Neuroscience and Medicine INM-11/JARA-BRAIN Institute JBI-2, Forschungszentrum Jülich GmbH, Jülich, Germany, [3] Department of Physics, RWTH Aachen University, Aachen, Germany, [4] Department of Biotechnology, University of Verona, Verona, Italy, [5] Medical Faculty, Cécile and Oskar Vogt Institute for Brain Research, Heinrich Heine University Düsseldorf, Düsseldorf, Germany

OPEN ACCESS

Edited by:
Sergio Decherchi,
Italian Institute of Technology (IIT), Italy

Reviewed by:
Yong Wang,
University of Copenhagen, Denmark
Giovanni Grazioso,
University of Milan, Italy

***Correspondence:**
Paolo Carloni
p.carloni@fz-juelich.de
Alejandro Giorgetti
a.giorgetti@fz-juelich.de

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

Received: 26 June 2020
Accepted: 13 August 2020
Published: 04 September 2020

Citation:
Schneider J, Ribeiro R, Alfonso-Prieto M, Carloni P and Giorgetti A (2020) Hybrid MM/CG Webserver: Automatic Set Up of Molecular Mechanics/Coarse-Grained Simulations for Human G Protein-Coupled Receptor/Ligand Complexes. Front. Mol. Biosci. 7:576689. doi: 10.3389/fmolb.2020.576689

Hybrid Molecular Mechanics/Coarse-Grained (MM/CG) simulations help predict ligand poses in human G protein-coupled receptors (hGPCRs), the most important protein superfamily for pharmacological applications. This approach allows the description of the ligand, the binding cavity, and the surrounding water molecules at atomistic resolution, while coarse-graining the rest of the receptor. Here, we present the Hybrid MM/CG Webserver (mmcg.grs.kfa-juelich.de) that automatizes and speeds up the MM/CG simulation setup of hGPCR/ligand complexes. Initial structures for such complexes can be easily and efficiently generated with other webservers. The Hybrid MM/CG server also allows for equilibration of the systems, either fully automatically or interactively. The results are visualized online (using both interactive 3D visualizations and analysis plots), helping the user identify possible issues and modify the setup parameters accordingly. Furthermore, the prepared system can be downloaded and the simulation continued locally.

Keywords: MM/CG, molecular mechanics, coarse-grained, hybrid methods, webserver, G protein-coupled receptor, ligand, molecular dynamics simulation

INTRODUCTION

Human G protein-coupled receptors (hGPCRs) are the largest drug target superfamily (Hauser et al., 2017). One third of FDA-approved drugs target ∼14% hGPCRs (Hauser et al., 2018) and this protein class has a further, untapped pharmacological potential. Unfortunately, rational drug design is hampered by the lack of experimental structures for more than 90%

Abbreviations: AA, all-atom; CFF, caffeine; CG, coarse-grained; FDA, United States Food and Drug Administration; $hA_{2A}R$, human adenosine 2A receptor; hGPCRs, human G protein-coupled receptors; MD, molecular dynamics; MM, molecular mechanics; MM/CG, molecular mechanics/coarse-grained.

hGPCRs[1] (Munk et al., 2019; Qu et al., 2020). Structural insights into ligand poses for these cases can be obtained by computational modeling. Reliable predictions can be made by docking approaches on homology models, based on templates sharing overall sequence identity >35–40% and/or high conservation of binding site residues (Beuming and Sherman, 2012; Kufareva et al., 2014). For lower resolution models, however, the uncertainty in the structure (and particularly in the orientation of side chains) decreases the accuracy of the docking predictions and, thus, follow-up molecular dynamics (MD) simulations are recommended (Kufareva et al., 2014; Cavasotto and Palomba, 2015; Esguerra et al., 2016; Heifetz et al., 2016; Fierro et al., 2017; Lupala et al., 2018; Rodríguez-Espigares et al., 2020). In an effort at addressing this issue, we have developed a Hybrid Molecular Mechanics/Coarse-Grained (MM/CG) simulation approach (Neri et al., 2005, 2008; Leguèbe et al., 2012; Marchiori et al., 2013; Sandal et al., 2015; Capaldi et al., 2018; Alfonso-Prieto et al., 2019; Fierro et al., 2019). The receptor/ligand interactions are described in atomistic detail, including explicit water molecules in the binding site (MM region), while the rest of the receptor is coarse-grained (CG region) (Schneider et al., 2018). The all-atom force fields used (Schneider et al., 2020) for the MM part of the protein and water are the Amber14SB (Maier et al., 2015) and TIP3P (Jorgensen et al., 1983), respectively, whereas the ligand can be described using either GAFF or GAFF2 (Wang et al., 2004; Case et al., 2020). The CG region is described by a Gō-like (Gō and Abe, 1981) potential. A region at the interface between the MM and CG parts couples the two levels of resolution (**Figure 1**). The membrane is described implicitly by introducing five potential walls (Leguèbe et al., 2012; Schneider et al., 2018). Two planar walls coincide with the height of the head groups of the membrane lipids, two hemispheric walls cap the extracellular and intracellular ends of the protein and prevent water evaporation, and the last wall follows the initial shape of the interface between protein and membrane, mimicking the effect of the lipid acyl tails (**Figure 1**). The scheme can be used also for GPCRs other than those from *Homo sapiens*. This approach turned out to be able to reproduce the ligand poses for four different hGPCRs (Schneider et al., 2020). These include the adenosine 2A receptor in complex with caffeine (i.e., the example case of the webserver), the human bitter receptor 16 in complex with phenyl-β-D-glucopyranoside, the β2-adrenergic receptor with adrenaline, and the dopamine D3 receptor with eticlopride. The structures of these hGPCR/ligand complexes were either experimentally determined, taken from all-atom MD trajectories or predicted based on templates with decreasing resolution, up to extremely low sequence identity (<15%) (Schneider et al., 2020). Retrospective validation against available X-ray structures and mutagenesis data confirmed that the MM/CG approach can predict correct ligand poses and identify experimentally determined binding residues[2] (Schneider et al., 2020), regardless of the model resolution. In addition, the MM/CG simulations can provide insights into the flexibility of receptor–ligand interactions and hydration of the binding cavity, at a lower computational cost than all-atom molecular dynamics simulations.

The increasing number of applications and requests to use the MM/CG approach has prompted us to develop the Hybrid MM/CG webserver, a publicly accessible web interface aimed at preparing and running short MM/CG simulations of hGPCR/ligand complexes (available since December 2019). To the best of our knowledge, this is the only webserver dedicated to this task. It complements other excellent online resources (**Table 1**) aimed at preparing all-atom or coarse-grained molecular dynamics (MD) simulations of GPCRs and GPCR/ligand complexes (or in general membrane proteins).

The Hybrid MM/CG webserver requires only the coordinates of the receptor/ligand complex (as PDB file). These may come from experimental structures, simulation snapshots, or computational models generated with other webservers, such as the ones listed in the "Input" section. Furthermore, we have linked our GOMoDo webserver[3] for modeling and docking of GPCRs to the Hybrid MM/CG webserver, so that the complexes generated with GOMoDo can be directly transferred. The user is first guided through a set of user-friendly forms for preparing the setup files. The procedure requires only a few minutes. Then, the user runs the initial MM/CG simulation steps (up to 10 ns) directly on the server. The resulting files can be downloaded and the MM/CG simulation can be continued using local resources.

MATERIALS AND METHODS

Input

The input PDB file can be obtained from experimental databases, such as the Protein Data Bank (Berman et al., 2000; Rose et al., 2016), GPCRdb (Pándy-Szekeres et al., 2018) or GPCRmd (Rodríguez-Espigares et al., 2020), as well as computational services like GOMoDo (Sandal et al., 2013), GPCR-ModSim (Esguerra et al., 2016), GPCR-SSFE (Worth et al., 2017), GPCRM (Miszta et al., 2018), Galaxy7TM (Lee and Seok, 2016), GPCRautomodel (Launay et al., 2012), @TOME (Pons and Labesse, 2009) and others (reviewed in Busato and Giorgetti, 2016). The GOMoDo webserver (Sandal et al., 2013), which can be used for homology modeling of hGPCRs and subsequent docking of ligands, is linked directly with the Hybrid MM/CG webserver. Several structures of the same hGPCR/ligand complex obtained under different membrane compositions and/or in different activation states can be funneled into the webserver to indirectly (and very approximately) account for the influence of explicit lipids and of large conformational changes of the receptor, respectively.

Setup Modes

The Hybrid MM/CG webserver offers two system setup modes:

Automatic Preparation

Here, the user just needs to upload a PDB file of the hGPCR in complex with its ligand or transfer it directly from the GOMoDo

[1] https://gpcrdb.org/structure/statistics as of June 2020.

[2] the residues whose mutation is known experimentally to affect binding

[3] gomodo.grs.kfa-juelich.de

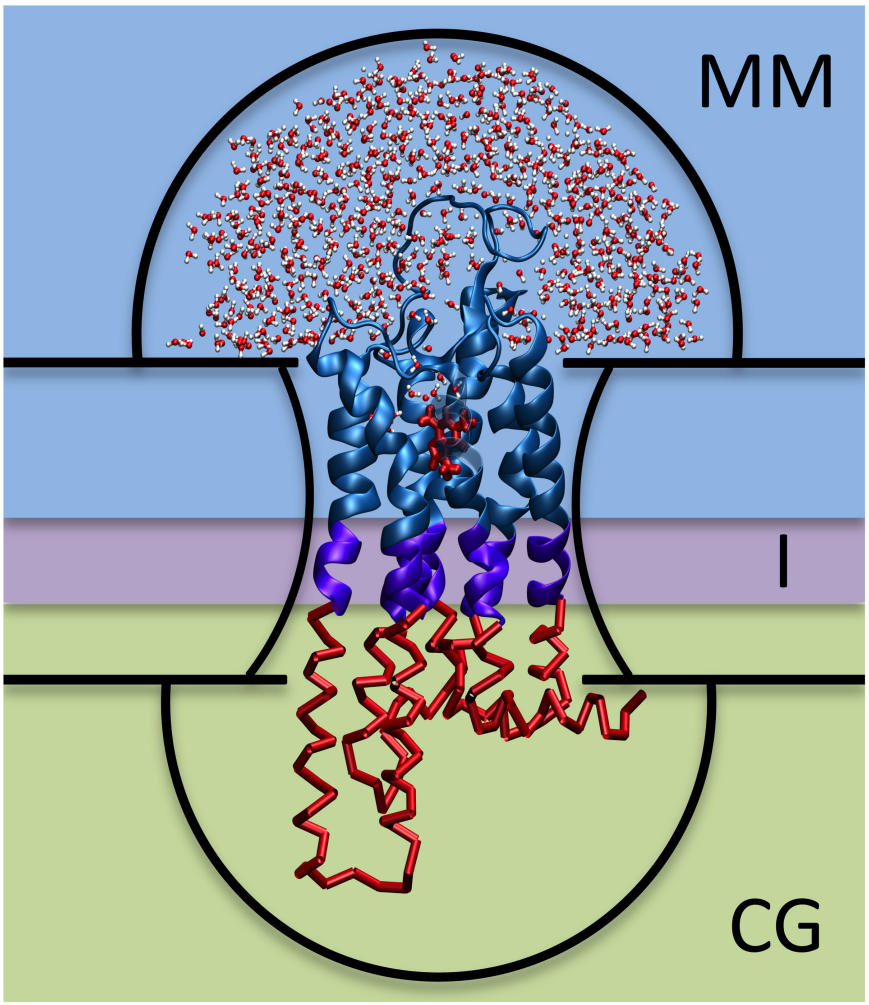


FIGURE 1 | Hybrid MM/CG scheme. The three regions (MM, I, and CG) are framed with different background colors and the five potential walls (upper and lower hemispheres, upper and lower membrane planes, and membrane surface) are indicated with black lines.

webserver[3] (Sandal et al., 2013). All simulation parameters are set according to default values that are defined in the documentation[4].

Interactive Preparation

Using the interactive preparation method, the same steps as in the automatic preparation are carried out to set up the system. The advantage is that several parameters, such as the position of the interface between MM and CG regions and the size of the hemispheric potential walls (see **Figure 1**), can be adjusted. Furthermore, intermediate results can be visualized, such as the input structure, aligned orientation in membrane, solvation (water drop), level of coarse-graining, and position of the wall potentials. More details about the individual preparation steps and parameters are explained in the Documentation section of the webserver[4].

[4]mmcg.grs.kfa-juelich.de/documentation

Workflow

The steps carried out by the webserver for MM/CG system preparation and short simulation (**Figure 2**) are the following:

- *File upload*. The preparation starts with a PDB file of the hGPCR/ligand complex. This file can be obtained from one of the databases and webservers mentioned in the "Input" section and uploaded by the user into the Hybrid MM/CG webserver. Alternatively, it can be transferred directly from GOMoDo (Sandal et al., 2013).
- *Check Input*. The input PDB file is checked for missing residues and the numbering is corrected if possible (i.e., no residues are missing). The ligand name is determined by comparing the list of residue names in the input PDB file and in the Amber14SB force field. Known amino acids and capping groups are discarded and the remaining residue name is considered to be the ligand.
- *Alignment*. The orientation and position of the receptor with respect to the hydrophobic core of the lipid bilayer

TABLE 1 | Available online resources for MD simulations of GPCRs.

Name	Functionality	System Setup				MD simulations
		Structural modeling	Membrane inclusion	Ligand inclusion	Force field used	
GPCR-ModSim (Esguerra et al., 2016)	Web-interactive system setup/simulation	✓[a]	✓	∼[b]	MM	✓ (AA, 5ns)
CHARMM-GUI (Jo et al., 2008; Lee et al., 2016)	Web-interactive system setup	×	✓	∼[c]	MM, CG	×
MERMAID (Damre et al., 2019)	Web-interactive system setup/simulation	×	✓	×	CG	✓ (CG, 100ns)
Hybrid MM/CG (This work)	Web-interactive system setup/simulation	∼[d]	✓ (implicit)	✓	MM/CG	✓ (MM/CG, 10ns)
PACKMOL-Memgen (Schott-Verdugo and Gohlke, 2019)	Local command line system setup	×	✓	∼[e]	MM	×
GPCRmd (Rodríguez-Espigares et al., 2020)	Database of precomputed MD simulations	×	✓	✓	MM	✓ (AA)[f]
MemProtMD (Newport et al., 2019)	Database of precomputed MD simulations	×	✓	✓	MM, CG	∼ (CG; AA)[g]

AA, all-atom; CG, coarse-grained; MM, molecular mechanics. [a]Homology modeling of GPCRs only; docking must be performed externally. [b]Ligands should already be included in the OPLS-AA force field or parameterized externally. [c]Although the Membrane Builder tool does not offer the option to include ligands, ligand parameterization can be accomplished using the Ligand Reader and Modeler tool also available in the CHARMM-GUI webserver. [d]Homology modeling and ligand docking can be performed with other available webservers, including the linked GOMoDo webserver. [e]Packmol-Memgen allows to keep ligands of interest but they have to be parameterized externally using other AmberTools. [f]All set up and trajectory files can be downloaded. Simulations of the GPCRmd dataset typically include three replicas, 0.5 μs each, whereas simulations of the so-called "Individual contributions" section can vary in the number of replicas and simulation length. [g]Trajectory files of the 1 μs CG assembly simulations cannot be downloaded. However, an equilibrated CG snapshot and the corresponding backmapped AA structure, together with the files needed to run subsequent CG or AA simulations, respectively, are available for download.

are determined using the PPM tool (Lomize et al., 2012). The initial PDB file of the hGPCR/ligand complex is then superimposed to the PPM-aligned structure of the receptor using lovoalign (Martínez et al., 2007).

- *Build topology: Protein*. The program pdb2gmx (Berendsen et al., 1995; Hess et al., 2008) is used to build the receptor topology using the Amber14SB force field (Maier et al., 2015). Protonation states of receptor titratable residues are determined automatically by pdb2gmx. In case the user wants to bypass the automatic assignment, different protonation states can be enforced with minimal manual editing of the input PDB file by changing the corresponding residue name (e.g., from "HIS" to "HIP" for a doubly protonated histidine).
- *Build topology: Ligand*. Protonation of the ligand at pH = 7 is determined using OpenBabel (O'Boyle et al., 2011). The ligand topology is built with ACPYPE (Sousa da Silva and Vranken, 2012) and Antechamber (Wang et al., 2001), using AM1-BCC charges (Jakalian et al., 2004) and the general Amber force field. GAFF (Wang et al., 2004) is used in the automatic preparation, but the user can choose between GAFF and GAFF2 (Case et al., 2020) in the interactive preparation mode.
- *Solvation*. The simulation box is created and the system is solvated using the TIP3P (Jorgensen et al., 1983) force field. Water molecules below the upper membrane plane are deleted. Hemisphere sizes are defined (default radii: 50 Å) and water outside the upper hemisphere is deleted, so that only a water "drop" solvating the extracellular part of the receptor is kept (see **Figure 1**).
- *Coarse-graining*. The regions of different resolutions are defined by two cutoff values, one defining the boundary between the MM and interface regions (default is centered between the two membrane planes) and the other between the interface and CG regions (6 Å below the first cutoff). Coarse-graining is performed on residues below the second cutoff level using a Gō-like (Gō and Abe, 1981) model.
- *Minimization*. A simple minimization of the system is done using the steepest descent method until a maximum force of 1,000 kJ mol^{-1} nm^{-1} is reached.
- *Add wall potentials*. Wall potentials are added to the system according to the aforementioned heights and radii. Wall grid points that are too close to the ligand are deleted according to the chosen cutoff (default 7 Å).
- *Visualization*. The user is redirected to the "Results" section where preparation results can be inspected using interactive 3D visualizations and graphs.
- *Equilibration*. An optional short equilibration (2 ns) and short MD simulation (2 ns) can be started from the "Results" section as well.
- *MD*. The simulation can be extended up to 10 ns, in 2 ns increments, in order to offer the possibility to check the intermediate results.

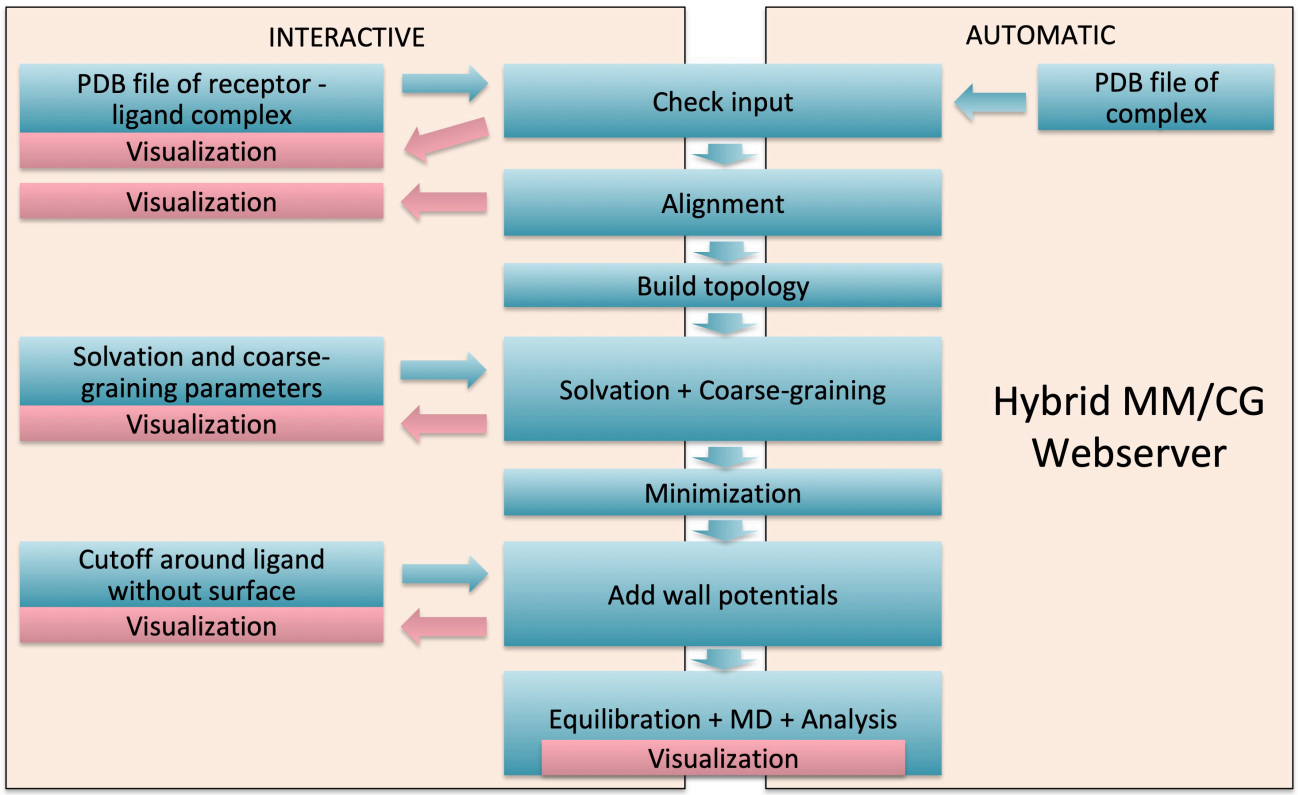


FIGURE 2 | Webserver workflow. Interactive mode with manual intervention is presented on the left and fast automatic mode on the right.

- *Visualization.* The minimization, equilibration, and MD are analyzed automatically in terms of potential energy and temperature after completion of the individual step (see **Supporting Information**). Interactive 3D visualizations are available for the results of the preparation, equilibration, and MD (see **Figure 3**).

Output

Results can be viewed and downloaded for 2 weeks (renewal possible) by bookmarking the link or alternatively by using the corresponding ID[5]. The full output of the preparation can be downloaded as a compressed archive file including the input, output, and log files of all preparation and simulation steps. The downloaded files can be used to continue the MM/CG simulations locally. Experienced users have the possibility to download the prepared system and tune the partial charges, as well as other parameters, before running the simulation on their local computer. The Download Code section of the server provides an installation script, which contains the links to the source code for both GROMACS 4.5.1 (from the GROMACS website) and the Hybrid MM/CG patch (from our server website).

[5]mmcg.grs.kfa-juelich.de/results/+ID

Server Architecture

Front-End

The web interface was developed with the DJANGO 3.0 Web framework[6] and designed with the Bootstrap 4 front-end open-source toolkit[7]. Interactive 3D visualizations were implemented with the NGL library for molecular visualization (Rose and Hildebrand, 2015) and the interactive graphs were created using the Highcharts SVG-based JavaScript charting library[8].

Back-End

For historical reasons, different programming and scripting languages are used for the workflow such as Bash, Python, Perl, Tcl, AWK, and JavaScript. Furthermore, different software packages are used in the preparation: a patched GROMACS 4.5.1 version (Berendsen et al., 1995; Hess et al., 2008), VMD 1.9.3 (Humphrey et al., 1996), PPM (Lomize et al., 2012), lovoalign (Martínez et al., 2007), ACPYPE Rev: 10101 (Sousa da Silva and Vranken, 2012), Antechamber (Wang et al., 2001) from AmberTools16 and OpenBabel 2.3.2 (O'Boyle et al., 2011). All parts were combined using the DJANGO 3.0 Web framework[6].

[6]djangoproject.com
[7]getbootstrap.com
[8]highcharts.com

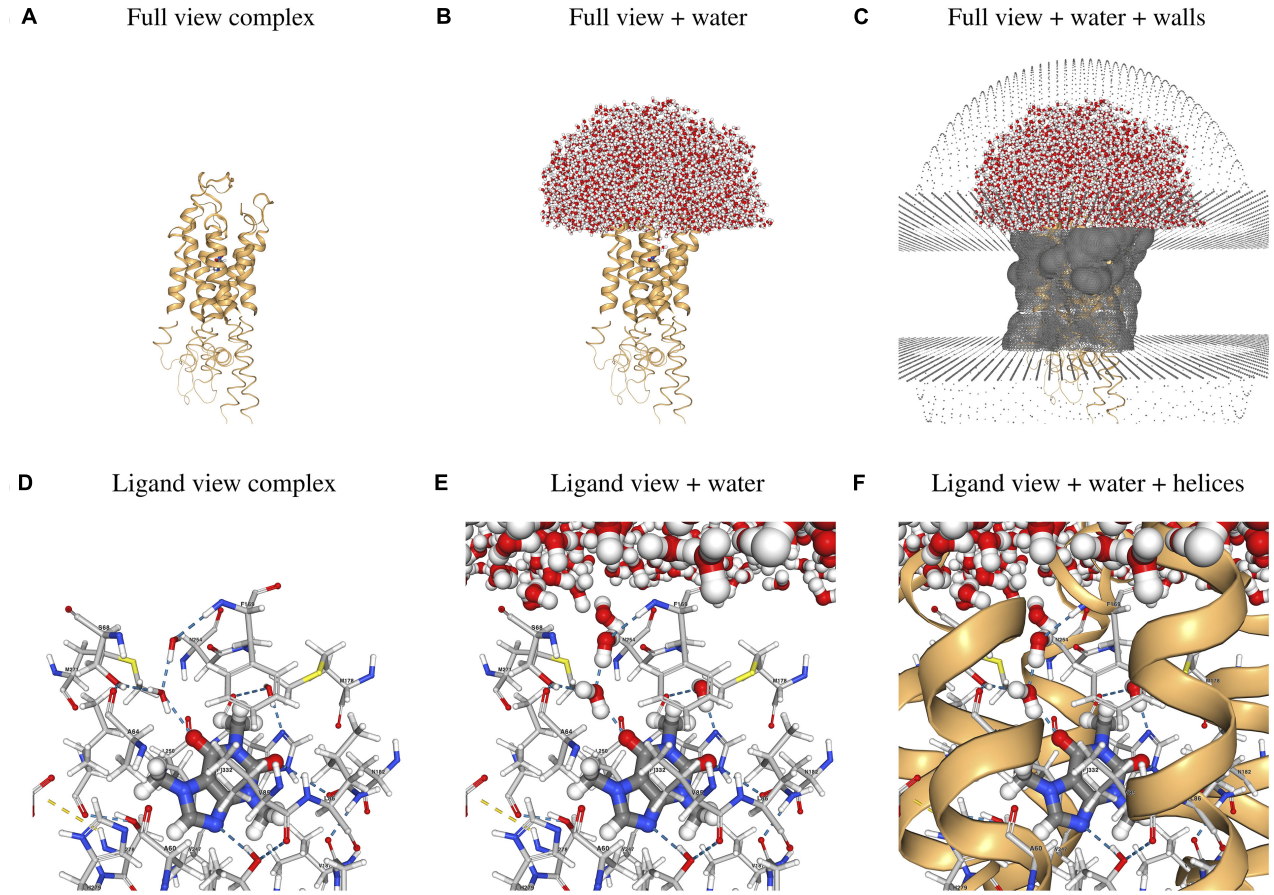


FIGURE 3 | Sample 3D visualizations created by the webserver using as example the $hA_{2A}R$/CFF complex. **(A–C)** Full view of the hGPCR/ligand complex and **(D–F)** close-up view of the ligand and its interactions with protein and water molecules. Images were generated with the screenshot tool in the NGL viewer (Jakalian et al., 2004) used in the "Results" section of the webserver. Atoms are colored using the default NGL color code and ligand interactions are determined using the default NGL definition.

RESULTS

As an application case, we considered the human adenosine 2A receptor ($hA_{2A}R$) in complex with its antagonist caffeine (CFF). CFF was proposed to have a protective and therapeutic effect against Parkinson's disease (Chen et al., 2001; Sonsalla et al., 2012; Petzer and Petzer, 2015; Nazario et al., 2017). The input PDB file of the $hA_{2A}R$/CFF complex used for the Example can be downloaded from the same page. The demonstration can be run by simply starting the preparation workflow[9] without uploading any PDB file. Precomputed results for this example case can be found on the server[10] and selected parts of these results can be seen in **Figure 3**. **Figures 3A–C** show the whole hA_{2A} receptor in complex with CFF, with and without water solvating the binding site and the extracellular loops, as well as the walls mimicking the membrane and preventing water evaporation. **Figures 3D–F** show the ligand and its direct and water-mediated interactions with residues in the binding cavity, as determined by the NGL viewer (Jakalian et al., 2004). The system can be inspected in 3D after the preparation, short equilibration, and production runs. In addition, plots showing the time evolution of the potential energy and temperature during the different simulation steps are generated (see **Supplementary Figures 1–3**).

CONCLUSION

The hybrid MM/CG protocol has been successfully used to predict ligand poses in a variety of hGPCRs (Leguèbe et al., 2012; Marchiori et al., 2013; Sandal et al., 2015; Fierro et al., 2017; Capaldi et al., 2018; Fierro et al., 2019). Moreover, the Amber-based MM/CG poses improved significantly relative to those obtained by simple docking, especially for low resolution starting models (Schneider et al., 2020). However, the setup of such simulations has been time-consuming and system-dependent.

[9]https://mmcg.grs.kfa-juelich.de/
[10]mmcg.grs.kfa-juelich.de/example_results/

The webserver presented here automatically prepares all files needed to start the equilibration, in a short time: the $hA_{2A}R$/CFF example requires only 6 min. An additional advantage of the Hybrid MM/CG webserver over most of the online services for simulation of GPCRs (**Table 1**) is the automatic ligand parameterization. Although other webservers can also include ligands (see **Table 1**), their parameterization usually has to be done externally. The results of the preparation steps can be checked on-the-fly (e.g., ligand structure and full solvation of extracellular loops). No additional software (except the patch for the GROMACS code, which is available for download[11]) is needed to extend the simulations locally. This makes the method useful for people acquainted with MD simulations and not necessarily familiar with every single MM/CG preparation step.

The server could prospectively be used for setting up MM/CG simulations of different ligands in complex with the same hGPCR. Capitalizing on the limited computational cost of the MM/CG approach, this could pave the way for low-throughput virtual screening efforts. Furthermore, comparative studies of hGPCRs in multiple states, e.g., active/inactive, can be accomplished by setting up several simulations starting from different initial structures. Modeling of allosteric effects is currently not possible, because of the use of the Gō-like model for the part of the receptor pointing toward the intracellular side. We also plan to integrate the Amber-based MM/CG approach used by the webserver with the recently developed open boundary MM/CG for grand canonical simulations (Tarenzi et al., 2017, 2019), aiming at calculating ligand binding free energies. Further planned developments include the extension to other MD codes, the implementation of the CHARMM36m force field (Huang et al., 2017) for the MM/I regions, as well as the adaptation of the scheme to other membrane protein families, such as ion channels.

[11] mmcg.grs.kfa-juelich.de/download/install

DATA AVAILABILITY STATEMENT

The Hybrid MM/CG Webserver, available at https://mmcg.grs.kfa-juelich.de/, is free and open to all. There is no login requirement.

AUTHOR CONTRIBUTIONS

JS, MA-P, PC, and AG contributed to workflow design and documentation. JS developed the server. RR designed the interface on the GOMoDo webserver. JS, RR, MA-P, and AG tested the webserver. All authors participated in writing the article.

FUNDING

This work was supported by the "Ernesto Illy Foundation" (Trieste, Italy) and by the EBRAINS research infrastructure, funded from the European Union's Horizon 2020 Framework Programme for Research and Innovation under the Specific Grant Agreement No. 945539 (Human Brain Project SGA3). Funding for the open access charge was provided by the Central Library of Forschungszentrum Jülich.

ACKNOWLEDGMENTS

We thank Emiliano Ippoliti for the setup of the server and Jonas Goßen for programming advice.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2020.576689/full#supplementary-material

REFERENCES

Alfonso-Prieto, M., Navarini, L., and Carloni, P. (2019). Understanding ligand binding to G-protein coupled receptors using multiscale simulations. *Front. Mol. Biosci.* 6:29. doi: 10.3389/fmolb.2019.00029

Berendsen, H. J. C., van der Spoel, D., and van Drunen, R. (1995). GROMACS: a message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* 91, 43–56. doi: 10.1016/0010-4655(95)00042-e

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., et al. (2000). The Protein Data Bank. *Nucl. Acids Res.* 28, 235–242. doi: 10.1093/nar/28.1.235

Beuming, T., and Sherman, W. (2012). Current assessment of docking into GPCR crystal structures and homology models: successes, challenges, and guidelines. *J. Chem. Inf. Model* 52, 3263–3277. doi: 10.1021/ci300411b

Busato, M., and Giorgetti, A. (2016). Structural modeling of G-protein coupled receptors: An overview on automatic web-servers. *Int. J. Biochem. Cell Biol.* 77(Pt. B), 264–274. doi: 10.1016/j.biocel.2016.04.004

Capaldi, S., Suku, E., Antolini, M., Di Giacobbe, M., Giorgetti, A., and Buffelli, M. (2018). Allosteric sodium binding cavity in GPR3: a novel player in modulation of Aβ production. *Sci. Rep.* 8:11102. doi: 10.1038/s41598-018-29475-7

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

Cavasotto, C. N., and Palomba, D. (2015). Expanding the horizons of G protein-coupled receptor structure-based ligand discovery and optimization using homology models. *Chem Commun.* 51, 13576–13594. doi: 10.1039/c5cc05050b

Chen, J.-F., Xu, K., Petzer, J. P., Staal, R., Xu, Y.-H., Beilstein, M., et al. (2001). Neuroprotection by caffeine and A2A adenosine receptor inactivation in a model of Parkinson's disease. *J. Neurosci.* 21, RC143–RC143. doi: 10.1523/JNEUROSCI.21-10-j0001.2001

Damre, M., Marchetto, A., and Giorgetti, A. (2019). MERMAID: dedicated web server to prepare and run coarse-grained membrane protein dynamics. *Nucleic Acids Res.* 47, W456–W461. doi: 10.1093/nar/gkz416

Esguerra, M., Siretskiy, A., Bello, X., Sallander, J., and Gutiérrez-de-Terán, H. (2016). GPCR-ModSim: a comprehensive web based solution for modeling G-protein coupled receptors. *Nucleic Acids Res.* 44, W455–W462. doi: 10.1093/nar/gkw403

Fierro, F., Giorgetti, A., Carloni, P., Meyerhof, W., and Alfonso-Prieto, M. (2019). Dual binding mode of "bitter sugars" to their human bitter taste receptor target. *Sci. Rep.* 9:8437. doi: 10.1038/s41598-019-44805-z

Fierro, F., Suku, E., Alfonso-Prieto, M., Giorgetti, A., Cichon, S., and Carloni, P. (2017). Agonist binding to chemosensory receptors: a systematic bioinformatics analysis. *Front. Mol. Biosci.* 4:63. doi: 10.3389/fmolb.2017.00063

Gō, N., and Abe, H. (1981). Noninteracting local-structure model of folding and unfolding transition in globular proteins. I. Formulation. *Biopolymers* 20, 991–1011. doi: 10.1002/bip.1981.360200511

Hauser, A. S., Attwood, M. M., Rask-Andersen, M., Schiöth, H. B., and Gloriam, D. E. (2017). Trends in GPCR drug discovery: new agents, targets and indications. *Nat. Rev. Drug Discov.* 16, 829–842. doi: 10.1038/nrd.2017.178

Hauser, A. S., Chavali, S., Masuho, I., Jahn, L. J., Martemyanov, K. A., Gloriam, D. E., et al. (2018). Pharmacogenomics of GPCR drug targets. *Cell* 172, 41-54.e19. doi: 10.1016/j.cell.2017.11.033

Heifetz, A., James, T., Morao, I., Bodkin, M. J., and Biggin, P. C. (2016). Guiding lead optimization with GPCR structure modeling and molecular dynamics. *Curr. Opin. Pharmacol.* 30, 14–21. doi: 10.1016/j.coph.2016.06.004

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

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

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

Jakalian, A., Jack, D. B., and Bayly, C. I. (2004). Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J. Comput. Chem.* 23, 1623–1641. doi: 10.1002/jcc.10128

Jo, S., Kim, T., Iyer, V. G., 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

Kufareva, I., Katritch, V., Stevens, R. C., and Abagyan, R. (2014). Advances in GPCR modeling evaluated by the GPCR Dock 2013 assessment: meeting new challenges. *Structure* 22, 1120–1139. doi: 10.1016/j.str.2014.06.012

Launay, G., Téletchéa, S., Wade, F., Pajot-Augy, E., Gibrat, J.-F., and Sanz, G. (2012). Automatic modeling of mammalian olfactory receptors and docking of odorants. *Protein Eng. Des. Sel.* 25, 377–386. doi: 10.1093/protein/gzs037

Lee, G. R., and Seok, C. (2016). Galaxy7TM: flexible GPCR–ligand docking by structure refinement. *Nucleic Acids Res.* 44, W502–W506. doi: 10.1093/nar/gkw360

Lee, J., Cheng, X., Swails, 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

Leguèbe, M., Nguyen, C., Capece, L., Hoang, Z., Giorgetti, A., and Carloni, P. (2012). Hybrid molecular mechanics/coarse-grained simulations for structural prediction of G-protein coupled receptor/ligand complexes. *PLoS One* 7:e47332. doi: 10.1371/journal.pone.0047332

Lomize, M. A., Pogozheva, I. D., Joo, H., Mosberg, H. I., and Lomize, A. L. (2012). OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic Acids Res.* 40, D370–D376. doi: 10.1093/nar/gkr703

Lupala, C. S., Rasaeifar, B., Gomez-Gutierrez, P., and Perez, J. J. (2018). Using molecular dynamics for the refinement of atomistic models of GPCRs by homology modeling. *J. Biomol. Struct. Dyn.* 36, 2436–2448. doi: 10.1080/07391102.2017.1357503

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

Marchiori, A., Capece, L., Giorgetti, A., Gasparini, P., Behrens, M., Carloni, P., et al. (2013). Coarse-grained/molecular mechanics of the TAS2R38 bitter taste receptor: experimentally-validated detailed structural prediction of agonist binding. *PLoS One* 8:e64675. doi: 10.1371/journal.pone.0064675

Martínez, L., Andreani, R., and Martínez, J. M. (2007). Convergent algorithms for protein structural alignment. *BMC Bioinformatics* 8:306. doi: 10.1186/1471-2105-8-306

Miszta, P., Pasznik, P., Jakowiecki, J., Sztyler, A., Latek, D., and Filipek, S. (2018). GPCRM: a homology modeling web service with triple membrane-fitted quality assessment of GPCR models. *Nucleic Acids Res.* 46, W387–W395. doi: 10.1093/nar/gky429

Munk, C., Mutt, E., Isberg, V., Nikolajsen, L. F., Bibbe, J. M., Flock, T., et al. (2019). An online resource for GPCR structure determination and analysis. *Nat. Methods* 16, 151–162. doi: 10.1038/s41592-018-0302-x

Nazario, L. R., da Silva, R. S., and Bonan, C. D. (2017). Targeting adenosine signaling in parkinson's disease: from pharmacological to non-pharmacological approaches. *Front. Neurosci.* 11:658. doi: 10.3389/fnins.2017.00658

Neri, M., Anselmi, C., Cascella, M., Maritan, A., and Carloni, P. (2005). Coarse-grained model of proteins incorporating atomistic detail of the active site. *Phys. Rev. Lett.* 95:218102. doi: 10.1103/PhysRevLett.95.218102

Neri, M., Baaden, M., Carnevale, V., Anselmi, C., Maritan, A., and Carloni, P. (2008). Microseconds dynamics simulations of the outer-membrane protease T. *Biophys. J.* 94, 71–78. doi: 10.1529/biophysj.107.116301

Newport, T. D., Sansom, M. S. P., and Stansfeld, P. J. (2019). The MemProtMD database: a resource for membrane-embedded protein structures and their lipid interactions. *Nucleic Acids Res.* 47, D390–D397. doi: 10.1093/nar/gky1047

O'Boyle, N. M., Banck, M., James, C. A., Morley, C., Vandermeersch, T., and Hutchison, G. R. (2011). Open babel: an open chemical toolbox. *J. Cheminform.* 3:33. doi: 10.1186/1758-2946-3-33

Pándy-Szekeres, G., Munk, C., Tsonkov, T. M., Mordalski, S., Harpsøe, K., Hauser, A. S., et al. (2018). GPCRdb in 2018: adding GPCR structure models and ligands. *Nucleic Acids Res.* 46, D440–D446. doi: 10.1093/nar/gkx1109

Petzer, J. P., and Petzer, A. (2015). Caffeine as a lead compound for the design of therapeutic agents for the treatment of Parkinson's disease. *Curr. Med. Chem.* 22, 975–988. doi: 10.2174/0929867322666141215160015

Pons, J.-L., and Labesse, G. (2009). @ TOME-2: a new pipeline for comparative modeling of protein–ligand complexes. *Nucleic Acids Res.* 37(Suppl._2), W485–W491. doi: 10.1093/nar/gkp368

Qu, X., Wang, D., and Wu, B. (2020). "Progress in GPCR structure determination," in *GPCRs: Structure, Function, and Drug Discovery*, ed. S.-H. Beata Jastrzebska Paul (Cambridge, MA: Academic Press). doi: 10.1016/C2018-0-00025-0

Rodríguez-Espigares, I., Torrens-Fontanals, M., Tiemann, J. K., Aranda-García, D., Ramírez-Anguita, J. M., Stepniewski, T. M., et al. (2020). GPCRmd uncovers the dynamics of the 3D-GPCRome. *Nat. Methods* 17, 777–787. doi: 10.1038/s41592-020-0884-y

Rose, A. S., and Hildebrand, P. W. (2015). NGL Viewer: a web application for molecular visualization. *Nucleic Acids Res.* 43, W576–W579. doi: 10.1093/nar/gkv402

Rose, P. W., Prlić, A., Altunkaya, A., Bi, C., Bradley, A. R., Christie, C. H., et al. (2016). The RCSB protein data bank: integrative view of protein, gene and 3D structural information. *Nucleic Acids Res.* 45:gkw1000. doi: 10.1093/nar/gkw1000

Sandal, M., Behrens, M., Brockhoff, A., Musiani, F., Giorgetti, A., Carloni, P., et al. (2015). Evidence for a transient additional ligand binding site in the TAS2R46 bitter taste receptor. *J. Chem. Theory Comput.* 11, 4439–4449. doi: 10.1021/acs.jctc.5b00472

Sandal, M., Duy, T. P., Cona, M., Zung, H., Carloni, P., Musiani, F., et al. (2013). GOMoDo: a GPCRs online modeling and docking webserver. *PLoS One* 8:e74092. doi: 10.1371/journal.pone.0074092

Schneider, J., Korshunova, K., Musiani, F., Alfonso-Prieto, M., Giorgetti, A., and Carloni, P. (2018). Predicting ligand binding poses for low-resolution membrane protein models: perspectives from multiscale simulations. *Biochem. Biophys. Res. Commun.* 498, 366–374. doi: 10.1016/j.bbrc.2018.01.160

Schneider, J., Korshunova, K., Si Chaib, Z., Giorgetti, A., Alfonso-Prieto, M., and Carloni, P. (2020). Ligand pose predictions for human G protein-coupled receptors: insights from the amber-based hybrid molecular mechanics/coarse-grained approach. *J. Chem. Inf. Model.* doi: 10.1021/acs.jcim.0c00661 [Epub ahead of print].

Schott-Verdugo, S., and Gohlke, H. (2019). PACKMOL-Memgen: a simple-to-use, generalized workflow for membrane-protein-lipid-bilayer system

building. *J. Chem. Inf. Model.* 59, 2522–2528. doi: 10.1021/acs.jcim.9b00269

Sonsalla, P. K., Wong, L.-Y., Harris, S. L., Richardson, J. R., Khobahy, I., Li, W., et al. (2012). Delayed caffeine treatment prevents nigral dopamine neuron loss in a progressive rat model of Parkinson's disease. *Exp. Neurol.* 234, 482–487. doi: 10.1016/j.expneurol.2012.01.022

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

Tarenzi, T., Calandrini, V., Potestio, R., and Carloni, P. (2019). Open-boundary molecular mechanics/coarse-grained framework for simulations of low-resolution G-protein-coupled receptor–ligand complexes. *J. Chem. Theory Comput.* 15, 2101–2109. doi: 10.1021/acs.jctc.9b00040

Tarenzi, T., Calandrini, V., Potestio, R., Giorgetti, A., and Carloni, P. (2017). Open boundary simulations of proteins and their hydration shells by Hamiltonian adaptive resolution scheme. *J. Chem. Theory Comput.* 13, 5647–5657. doi: 10.1021/acs.jctc.7b00508

Wang, J., Wang, W., Kollman, P. A., and Case, D. A. (2001). Antechamber: an accessory software package for molecular mechanical calculations. *J. Am. Chem. Soc.* 222:U403.

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

Worth, C. L., Kreuchwig, F., Tiemann, J. K. S., Kreuchwig, A., Ritschel, M., Kleinau, G., et al. (2017). GPCR-SSFE 2.0-a fragment-based molecular modeling web tool for Class A G-protein coupled receptors. *Nucleic Acids Res.* 45, W408–W415. doi: 10.1093/nar/gkx399

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

REVIEW
published: 25 November 2020
doi: 10.3389/fmolb.2020.604770

Mechanistic Understanding From Molecular Dynamics Simulation in Pharmaceutical Research 1: Drug Delivery

Alex Bunker[1]* *and Tomasz Róg*[2]*

[1] Division of Pharmaceutical Biosciences, Drug Research Program, Faculty of Pharmacy, University of Helsinki, Helsinki, Finland, [2] Department of Physics, University of Helsinki, Helsinki, Finland

In this review, we outline the growing role that molecular dynamics simulation is able to play as a design tool in drug delivery. We cover both the pharmaceutical and computational backgrounds, in a pedagogical fashion, as this review is designed to be equally accessible to pharmaceutical researchers interested in what this new computational tool is capable of and experts in molecular modeling who wish to pursue pharmaceutical applications as a context for their research. The field has become too broad for us to concisely describe all work that has been carried out; many comprehensive reviews on subtopics of this area are cited. We discuss the insight molecular dynamics modeling has provided in dissolution and solubility, however, the majority of the discussion is focused on nanomedicine: the development of nanoscale drug delivery vehicles. Here we focus on three areas where molecular dynamics modeling has had a particularly strong impact: (1) behavior in the bloodstream and protective polymer corona, (2) Drug loading and controlled release, and (3) Nanoparticle interaction with both model and biological membranes. We conclude with some thoughts on the role that molecular dynamics simulation can grow to play in the development of new drug delivery systems.

Keywords: pharmaceutics, nanomedicine, molecular dynamics, drug delivery, nanoparticle

OPEN ACCESS

Edited by:
Sergio Decherchi,
Italian Institute of Technology (IIT), Italy

Reviewed by:
Luca Monticelli,
Centre National de la Recherche Scientifique (CNRS), France
Sophie Sacquin-Mora,
UPR 9080 Laboratoire de Biochimie Théorique (LBT), France

***Correspondence:**
Tomasz Róg
tomasz.rog@helsinki.fi;
tomasz.rog@gmail.com
Alex Bunker
alex.bunker@helsinki.fi

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

Received: *10 September 2020*
Accepted: *02 November 2020*
Published: *25 November 2020*

Citation:
Bunker A and Róg T (2020) Mechanistic Understanding From Molecular Dynamics Simulation in Pharmaceutical Research 1: Drug Delivery. Front. Mol. Biosci. 7:604770. doi: 10.3389/fmolb.2020.604770

INTRODUCTION

The exponential advance of the computational power available to us has led to related approaches attaining a prominent, one can argue now dominant, position within pharmaceutical research. The majority of this toolkit, as we will elaborate below, are methodologies that fit experimental data to a mathematical model that provides a numerical answer, for example a specific drug molecule structure or delivery system formulation. A subset of computational methodologies provide something further: mechanistic understanding; in place of just an answer, i.e., an optimum value or set of values, mechanistic understanding means an elucidation of what is actually occurring in the system that produces the results: in simple terms, a model of the system, expressed as a cartoon in three dimensions, of what is happening. Such an output, often referred to as a simulation, has power far beyond that provided by a mere result of what is optimal for the specific application

sought; it can lead to an informed design process that is more efficient, allows for broader intuitive leaps from its interpretation and provides insight that transcends the specific application studied.

An extremely successful computational scheme for attaining mechanistic understanding is molecular dynamics simulation (MD), a methodology that models the system as a set of particles that interact through classical mechanics. An intuitive choice for these particles, particularly for those with a background in chemistry, is for them to represent atoms, with interactions between the atoms producing the intramolecular forces that govern the structure of molecules and the intermolecular forces that govern interactions between molecules. This is, however, not the only choice that can be made, as particles can be chosen to represent larger structures than single atoms; they can represent groups of atoms, whole molecules, or even groups of molecules. Such models can obtain insight into the system on a larger length and time scale than can be achieved through a model with atomistic resolution and are known as coarse grained (Ingólfsson et al., 2014).

In this review we will highlight the growing role that MD has played and will continue to play in drug delivery, what has been referred to as *computational pharmaceutics* by Ouyang and Smith (2015), using computational methods to address issues related to drug delivery including dissolution, solubility, protection from the bodies defense mechanisms, controlled release and targeted delivery. The development of advanced mechanisms for drug delivery based on nanoscale drug delivery vehicles, a field known as *nanomedicine* (Riehemann et al., 2009; Lammers and Ferrari, 2020; Moghimi et al., 2020), is a particular area where MD methods have borne fruit. This review paper has two target audiences: (1) pharmaceutical researchers, intrigued by the rapid rise of computational methods applicable to their research, who are interested in learning what kind of insight MD can provide and (2) computational physicists and chemists, with a background in MD methods, atomistic and coarse grained, who, for reasons I most probably do not need to inform the reader of, realize that at this point in history pharmaceutical applications are an extremely desirable context for their research. Both of the target audiences will find certain elements of this review to be trivially basic and may even bristle at some oversimplifications; one should keep in mind the dual audience focused nature of this review. As the subject matter is extremely broad, with several areas covered by comprehensive reviews themselves, this publication can, to some extent, be seen as a meta-review, to be used as an initial jumping off point leading to many further review papers, in addition to original work.

At its core, pharmaceutical science is roughly (1) the search for small molecules that, over the scale of the entire organism, do more good than harm under certain conditions: drug design and (2) development of the means by which these molecules can enter the body and reach their target tissue intact: drug delivery or pharmaceutics. Pharmaceutical science begins with Paracelsus, the man who is to pharmacy what Isaac Newton is to physics and his maxim “the dose makes the poison" (Rozman and Doull, 2001); substances exist that, at too high a dose are a poison that will kill you, but when taken at a certain dose can actually help you. The substance enters the body, dissolution occurs and the drug molecules within the substance are freed and diffuse through the body and enough reaches, intact, the desired location in sufficient quantity to induce the desired effect. Any drug molecule will reach other parts of the body and have different effects which are undesirable: the toxicity, i.e., side effects, of the drug. The conventional drug design paradigm is thus a balancing act between efficacy, toxicity and solubility. A very efficacious drug can be found that either has intolerable toxicity or too poor solubility to be carried through the bloodstream or, due to the nature of the target tissue, insufficient quantities of the drug reach it to have the desired effect.

Initially drugs were found through trial and error, however, the search space is gigantic: the number of different small organic molecules that are theoretically possible to synthesize is $\sim 10^{63}$ (Bohacek et al., 1996; Hoffmann and Gastreich, 2019) a number that dwarfs such quantities as Avogadro’s number and the number of stars in the universe; drug design can be seen as searching this discrete "drug structure space." The latter half of the twentieth century saw the onset of a systematic approach to searching this space based on the "lock and key" paradigm: drug molecules were designed to fit a certain active site on a certain protein to either inhibit or activate them. This was propelled by advances in three areas (1) robotics to enable massive simultaneous parallel screening experiments, (2) increasing numbers of high resolution protein structures, determined first through X-ray crystallography, but now increasingly through cryo-EM, and (3) the computational power and advanced algorithms to analyze the massive data sets produced. The computational component of this, computational drug design, can be divided into two methodologies: (1) ligand-based (Acharya et al., 2011) where the target protein structure is not known and (2) structure based (Alonso et al., 2006; Sousa et al., 2006; Sliwakosky et al., 2014; Ferreira et al., 2015), where the binding free energy of potential drug molecules is calculated, using the experimentally determined high resolution protein structure, a calculation known as "ligand docking and scoring." Ligand based methods use pattern recognition, now trendily referred to as "machine learning," algorithms where elements of the structural properties are mapped to either (1) high throughput screening results for activity, i.e., efficacy and other desirable properties, e.g., solubility parameters: Quantitative Structure Activity Relationship/Quantitative Structure Property Relationship (QSAR/QSPR) (Liu and Long, 2009; Nantasenamat et al., 2009; Ghasemi et al., 2018; Toporov and Toporova, 2020) or (2) elements of three dimensional structure of the molecule: pharmacophore modeling (Acharya et al., 2011).

Apart from the pharmacological research to determine appropriate target protein active sites, the above mentioned methodologies for drug discovery together are a fixed, simplified, purely empirical, paradigm: fitting data without insight. As is the case with research carried out using a fixed paradigm, metaphorically speaking continuing to turn the crank on the same machine, one reaches a point of diminishing returns; this is exactly what has occurred for the case of pharmaceutical research: as the resources spent globally on pharmaceutical research increase exponentially, the number of new drugs approved each

year remains constant, a phenomenon referred to as "Eroom's law" (Scanell et al., 2012) the reverse of the famous Moore's law regarding the exponential increase in computational technology we have witnessed over the past half century: pharmaceutical research is slowing down exponentially; moving forward requires moving beyond this oversimplified model.

The situation for drug delivery, i.e., pharmaceutics, is similar. When a given molecule is designed, using the above methodology, a set of rules of thumb are applied regarding its properties, known as "Lipinski's rule of 5" (Lipinski et al., 2001; Lipinski, 2004). This determines whether the molecule is "drug-like," i.e., a molecular structure likely to have a sufficiently optimal solubility profile, or not. Behavior of the drug in the body apart from its drug action, known as its Absorption, Distribution, Metabolism, and Excretion (ADME) properties, is a critical aspect that partially determines both efficacy and toxicity. This is modeled using numerical solutions to complex sets of coupled differential equations that represent the interactions of drugs and drug metabolites, as their distribution varies in time in the different tissues of the organism; this form of numerical computational modeling is known as pharmacokinetic/pharmacodynamics modeling (Craig, 1998; Ruiz-Garcia et al., 2008; Belfo and Lemos, 2021). While this form of modeling is not entirely empirical, as it is dependent on known metabolic relations, it still remains a method to calculate a quantitative result from experimentally measured parameters.

Given that the global pharmaceutical industry is estimated to have a turnover in excess of 1 trillion USD, there is obviously a substantial continuing effort to break out of the rut of diminishing returns it finds itself in. Regarding pharmaceutics, the last 30 years has seen the development of increasingly sophisticated mechanisms for enhancing the solubility profile, carrying/protecting drugs in the bloodstream and targeting them to the desired tissue (Zhang W. et al., 2017): the aforementioned nanomedicine (Moghimi et al., 2020). These involve either covalently bonding the drug to a molecule or packaging the drug into a nanoscale (diameter 100nm or less) vehicle that performs this function. As this field has developed, these means have become increasingly complex and intricate and, as a result, this avenue has also become stuck (Park, 2016): while increasingly complex devices make for engaging narratives leading to well cited publications, the greater the complexity the more that can go wrong, resulting in a field of research that is far better at producing publications than real approved therapies; as Venditto and Szoka have put it "so many papers and so few drugs!" (Venditto and Szoka, 2013); the resulting system, coupled to the human physiological environment, is far too complex to be developed through the above described limited, mostly empirical, paradigm.

It can be argued that what is missing is mechanistic understanding: insight into what is actually physically happening, i.e., what are the molecules actually doing? The above described computational methods do not provide this; what they provide is a numerical answer. Mechanistic understanding is obtained by a computational method that can, given knowledge of the structure of molecules, provide insight into how the molecules interact, i.e., what structures they form and how they move with respect to each other with time: a three dimensional movie of what is happening on the molecular length scale. A molecule, or system of molecules, is a set of nuclei and electrons interacting in a specific way. How this interaction affects the motion of the atoms, i.e., the physics of the system, is quantum mechanics. Exact calculation is impossible, however, the discipline of theoretical quantum chemistry has developed many methods for approximating the behavior of molecules governed by quantum mechanics (Cramer, 2002). While these calculations can be simplified through the use of semi-empirical methods (Thiel, 2014), we are still left with a calculation that is too computationally intensive to simulate the length and time scales that are of interest to us. Making a set of approximations and accepting certain limitations of the variety of phenomena that can be observed, we arrive at the molecular mechanics paradigm: the molecule modeled as a set of particles with their interactions governed by classical mechanics.

THE MOLECULAR MECHANICS PARADIGM AND MOLECULAR DYNAMICS SIMULATION

The molecular mechanics paradigm is based on a combination of insight from the quantum mechanical interactions of atoms and empirical physical chemistry. The resulting model, illustrated in **Figure 1**, can be intuitively pictured as a set of sticky rubber balls (the short range attractive van der Waals (Israelachvili, 1985) and repulsive Pauli exclusion forces modeled through what is known as the Lennard-Jones potential term) that are charged (electronegativity of atoms and H-bonding behavior modeled through partial charges) connected by springs (the bond forces) with hinges (angular interactions), axels (proper dihedral potentials) and other 4-body interactions to produce correct structure (improper dihedral potentials); the atoms and molecules follow Newton's equations of motions, knocking into each other and rattling about in response to these forces; the result is a three dimensional movie of the system with atomistic resolution: molecular dynamics simulation (Allen and Tildesley, 1989; Frenkel and Smit, 2001). This has been referred to as a "computational microscope" by Lee E. H. et al. (2009), however, we feel this analogy is misleading as this is not a visualization of a piece of a real system but rather the isolation and study of a specific aspect of the system that we have assembled the appropriate set of models of molecules to study. Discussion of the methods used to determine the parameters of this model can be found elsewhere (Plimpton, 1995; Karplus and McCammon, 2002; Case et al., 2005; Phillips et al., 2005; Hess et al., 2008; van Gunsteren et al., 2008; Brooks et al., 2009; Abraham et al., 2015).

Several competing potential sets exist and for simulating any system with new molecules that have never been simulated before, often the case in pharmaceutical as opposed to biological research since we deal with unique man-made molecules, quantum chemistry calculations must be performed; choosing and building potential sets for the model to obtain the correct result requires significant expertise. While molecular dynamics simulation with an all atom model has seen significant success

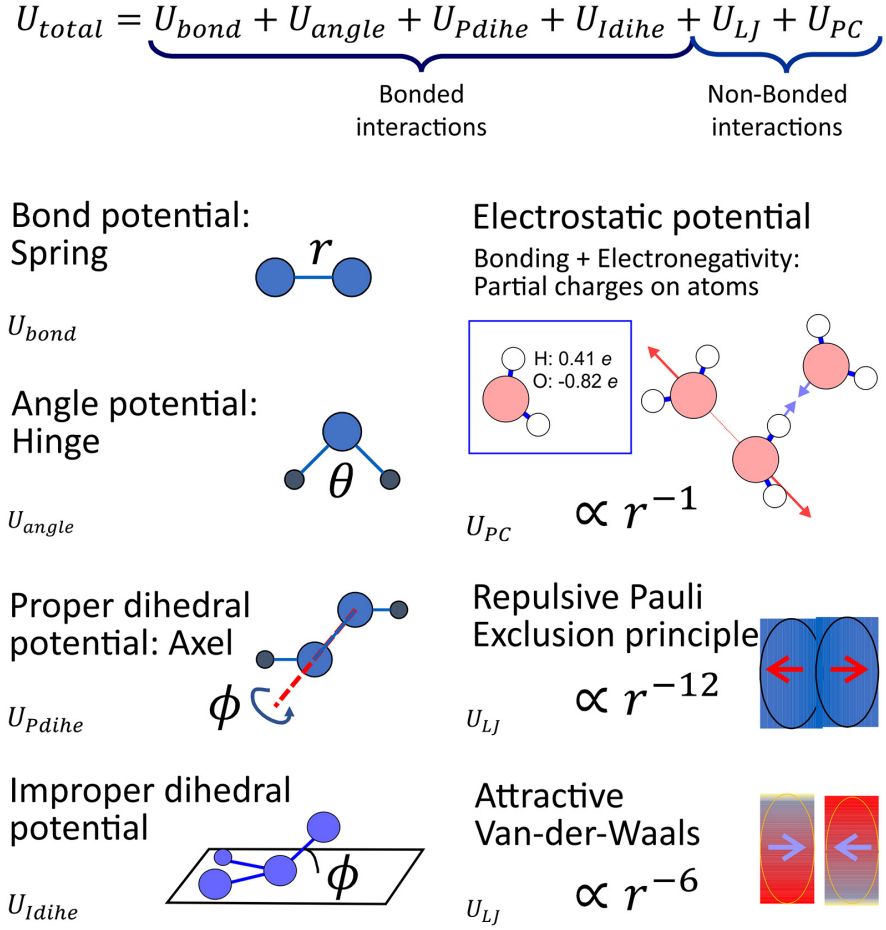
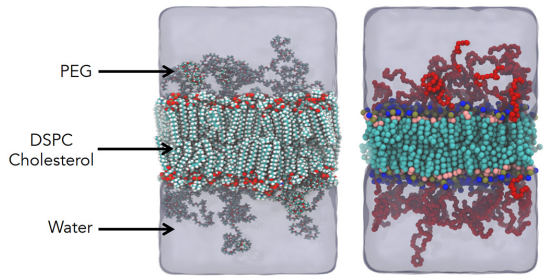


FIGURE 1 | The set of interactions that are the molecular mechanics paradigm that defines the forces that drive the motion of atoms in a molecular dynamics simulation with all atom resolution. The bond, angle and dihedral potentials are the intramolecular interactions that define molecule structure and the interactions between covalently bound atoms. Each atom has a partial charge that interacts with all other atoms through electrostatic forces to model electronegativity and H-bonding behavior; short range attractive and repulsive forces due to the Van-der-Waals and Pauli Exclusion Principle, respectively, are modeled through the Lennard-Jones interaction.

in the study of a wide range of biophysical systems, it is still limited to a length scales of ∼15 nm and time scales of ∼1–2 μs: too small to obtain insight into several phenomena we wish to study. Here we return to the aforementioned coarse grained models (**Figure 2**). While several phenomena cannot be observed as they are dependent on specific interatomic interactions, e.g., salt bridges and H-bonds, with good judgment such models allow for the metaphorical camera to zoom out and study behavior on larger length and timescales, however, with reduced resolution.

While several schemes for the development of coarse grained models have been proposed (Miyazaki et al., 2020) the two that have been most frequently used are the MARTINI potential set (Marrink et al., 2007), where the coarse grained particles are groups of ∼3 atoms with the potential sets developed based on the solubility parameters of these groups and Dissipative Particle Dynamics (DPD) (Groot and Warren, 1997; Español and Warren, 2017) where the degree of coarse graining is greater still, where the particles are soft "momentum carriers" and temperature is controlled through a thermostat designed to conserve local momentum as the effects of hydrodynamics become important at this larger length and time scale. Another scheme is incorporating the effect of the solvent through adjustment to the interactions between particles in the molecules of interest, i.e., simulating with adjusted potentials in a vacuum; this is known as the "implicit solvent" model (Murtola et al., 2009). An ideal that is often sought and discussed is "multiscale simulation"—combining the insight from simulations carried out with different methodologies on different length and time scales (Haddish-Berhane et al., 2007; Murtola et al., 2009; Meier et al., 2013); in 2013 The Nobel Prize in Chemistry was awarded to Arieh Warshel, Martin Karplus and Michael Levitt for "development of multiscale models for complex chemical systems" (The Nobel Prize in Chemistry 2013, 2013). From the literature search for this review, it can, however, be surmised that this ideal, for the most part, remains an ideal: for the recent original research found, in our literature search for this review, that applied MD simulation in the field of drug delivery, the number of publications that use more than one methodology



Resolution	all atom	coarse grained (3-4 atoms -> 1 bead)
Computing performance	100ns/day (256 cores)	1 μs/day (64 cores)
Number of particles	~110000	~25000

FIGURE 2 | Illustration of coarse graining: the same system, a PEGylated membrane, modeled with an all atom and a MARTINI model is shown along with the decrease in number of particles and acceleration in the simulation time. Figure taken from Bunker: (Bunker et al., 2016) with permission.

remain a small minority. Several reviews cover the use of coarse grained methods for the simulation of systems composed of lipids, polymers and proteins (Bennun et al., 2009; Loverde, 2014; Cascella and Vanni, 2016; Brancolini and Tozzini, 2019).

Now that we have this three-dimensional movie of our system, known as a trajectory, beyond just visualization there are several techniques to analyze this result and obtain useful insight into the system. Here we provide a few examples of frequently calculated properties from the trajectory. Considering pharmaceutical applications of MD simulations, a description of a binding mode (hydrogen bonds, salt bridges, stacking interaction, and hydrophobic interactions) of a drug in the protein binding cavity is the first key information to examine. Unlike binding modes obtained from experimental structural studies or docking predictions, MD simulations provide a dynamic description of the interaction between drug and protein (e.g., Kaszuba et al., 2012; Chen J. et al., 2019); this allows additional insight regarding the importance of individual interactions. Moreover, simulations provide explicit information concerning water participation in the binding mode (e.g., Kaszuba et al., 2010; Postila et al., 2013; Aguayo-Ortiz and Dominguez, 2019; **Figure 3A**), typically not resolved in structural studies and not considered in docking calculations. Analysis of intermolecular interactions is not limited to drug-protein interactions but can also be performed for any type of molecule/macromolecule studied, e.g., drug-lipid interactions are frequently studied (Cramariuc et al., 2012; Mayne et al., 2016; Pasenkiewicz-Gierula et al., 2016; Postila and Róg, 2020).

Additional observable properties used to evaluate intermolecular interactions are the radial distribution function (RDF) (**Figure 3C**) and the number of contacts. The RDF for the pairs of particles P1 and P2 gives us the normalized density of particle P2 at a given distance from particle P1. For the shortest distances the RDF value is 0 due to steric repulsion and converges to a constant value in the limit of infinite distance; for homogenous systems this value will always be 1. For an interacting pair of two particles, the RDF value initially rises with increasing distance to a maximum followed by a subsequent minimum (**Figure 3C**). E.g., for a pair of heavy atoms that form an H-bond, the maximum position is at ∼0.25 nm, and the minimum at ∼0.325 nm (Pasenkiewicz-Gierula et al., 1997). The number of contacts is the number of pairs of heavy atoms of two molecules located at a distance shorter than the selected cutoff. The most frequent choice for a cutoff length is the position of maximum or minimum at the RDF for carbon atoms in the liquid hydrocarbons. Calculations of numbers of contact are useful to evaluate equilibration in the simulations where self-assembly is studied. When a stable number of contacts is reached one can assume the end of the self-assembly process.

For interactions of larger molecules, MD simulations provide an area of contact (A_{cont}). To obtain this, the solvent accessible surface area (SASA) (Connolly, 1983) for the considered molecule is first calculated separately (*Amol1* and *Amol2*), and next, the same calculations are performed for the dimer (*Adimer*); this results in an area of contact:

$$A_{cont} = (A_{mol1} + A_{mol}2) - A_{dimer})/2 \quad (1)$$

Extensive MD simulation, either performed over a long time (Hurst et al., 2010; Dror et al., 2011) or as many multiple parallel simulations (Lolicato et al., 2020), are capable of elucidating the process of ligand entry into the binding pocket, however, the most frequent steered MD simulation methods (Izrailev et al., 1997) or randomly accelerated MD (RAMD) simulations (Lüdemann et al., 2000; Kokh et al., 2018) are used to reveal the entry/exit patch as they are more computationally efficient. For the case of functionalized proteins, their stability can be evaluated via calculations of secondary protein structure (**Figure 3B**). Other standard measurable properties provided by MD simulations include root mean square deviation (RMSD) and root mean square fluctuations (RMSF). The RMSD describes the similarity between the structures at the given time with the initial structure; thus, a large increase of this parameter can indicate a lack of protein stability. In studies of the interaction of drugs and nanoparticles with lipid bilayers, one can obtain insight into the xenobiotic degree of membrane perturbation.

The most frequently used tools to study lipid bilayer properties are surface area per lipid molecule, bilayer thickness and the order parameter. The most frequently calculated order parameters are the deuterium order parameter, S_{CD} and molecular order parameter S_{mol} (Vermeer et al., 2007; **Figure 3D**). The position in the membrane of any given xenobiotic molecule is quantitatively described by so-called density plots, which show the density of selected atoms, atom groups or whole molecules, along the bilayer normal. As a reference point, selected atoms of lipid molecules can be used, e.g., headgroups, glycerol moiety, or the last carbon of the acyl tails (**Figure 3E**). The next parameter describing drug behavior in the lipid bilayer is the drug molecule orientation with respect to the bilayer normal. Location and orientation of the drug in the lipid bilayer can be important for the entry of the drug into a protein binding cavity (e.g., Magarkar et al., 2018). Next, simulations describe the physicochemical properties of nanoparticles, including their size (quantitatively measured as the

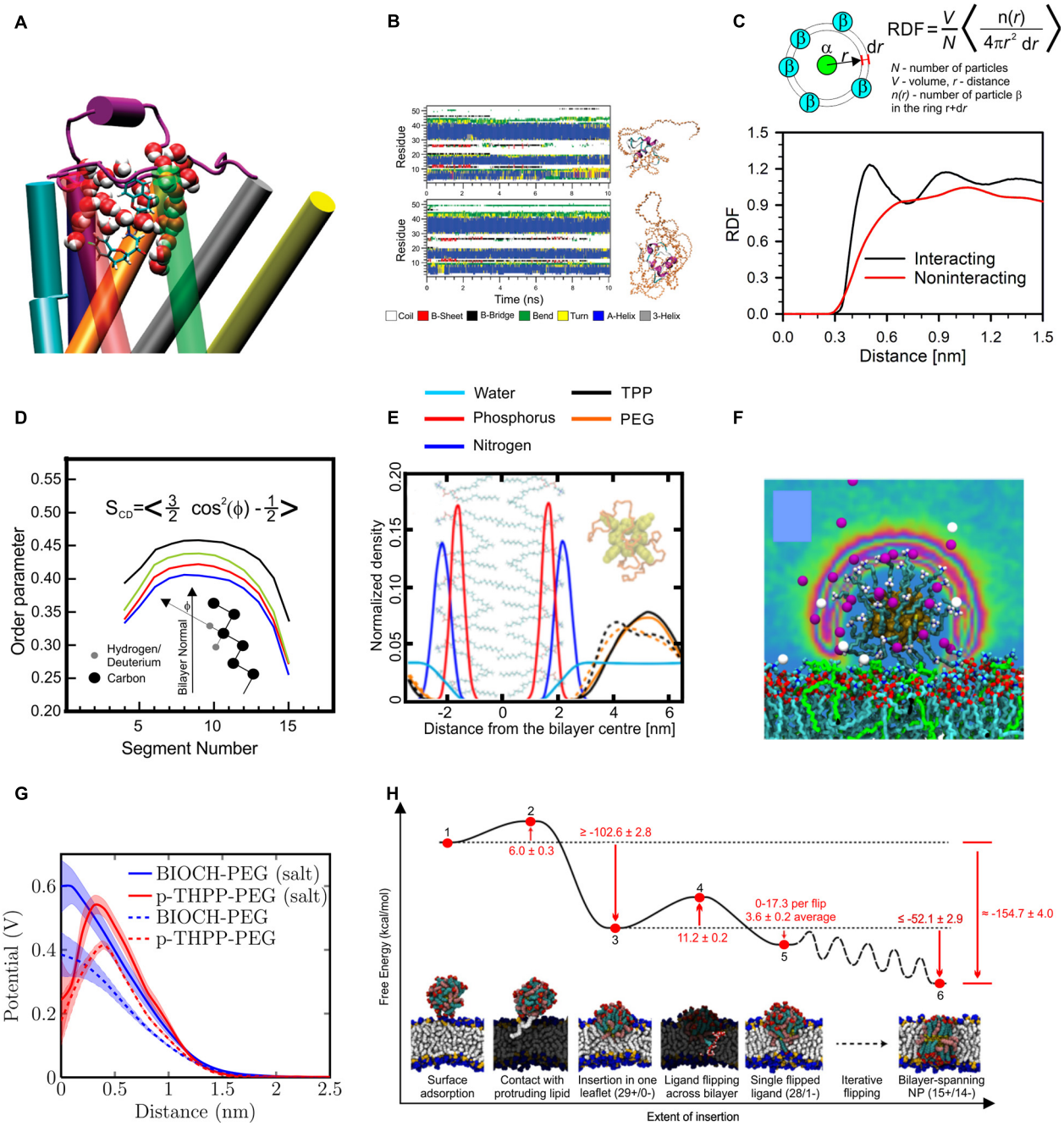


FIGURE 3 | Results of MD simulations. **(A)** Snapshot showing involvement of water in binding mode of nebivolol to β2-adrenergic receptor, reproduced with permission from Kaszuba et al. (2010), Copyright (2010) American Chemical Society. **(B)** Time evolution of secondary structure of PEGylated insulin molecules, reproduced with permission from Yang et al. (2011), Copyright (2011) American Chemical Society. **(C)** An example of radial distribution functions (RDF) for interacting particles (black line) and non-interacting particles, data taken from Roig and Pasenkiewicz-Gierula (2004). **(D)** Example of order parameter profile along the lipids acyl chain, reproduced from Mobarak et al. (2018) (CC BY 4.0). **(E)** An example of density profile showing position of atoms of lipid headgroups (phosphorus and nitrogen) and PEGylated tetra-phenyl-porphyrin (PEG and porphyrin densities are shown separately), at the presence (dashed line) and absence (solid line) of salt in solution, reproduced with permission from Rissanen et al. (2014), Copyright (2014) American Chemical Society. **(F)** Distribution of counter ions around gold nanoparticle functionalized with hydrocarbons capped with amine group, reproduced with permission from Heikkilä et al. (2014a), Copyright (2014) American Chemical Society, **(G)** electrostatic potential profile around PEGylated Biochanin (BIOH) and tetra-phenyl-porphyrin (p-THPP) in the presence and absence of salt in solution, reproduced with permission from Rissanen et al. (2014), Copyright (2014) American Chemical Society; **(H)** free energy landscape for the process of insertion of dendrimer into lipid bilayer, reproduced from Van Lehn and Alexander-Katz (2019), Copyright: 2019 Van Lehn, Alexander-Katz.

radius of gyration), nanoparticle hydration, interaction with ions (see **Figure 3F**) and electrostatic potential at the given distance from the nanoparticle center (**Figure 3G**). Finally, one should consider the statistical significance of results to avoid over-interpretation (Gapsys and Groot, 2020), carefully validate results against experimental data (Botan et al., 2015; Ollila and Pabst, 2016), and be critical as simulations are prone to methodological artifacts (Wong-ekkabut and Karttunen, 2016).

Sometimes the unbiased trajectory is not sufficient to obtain the insight we seek. The phenomenon we wish to study may occur in a region that is not sampled so frequently or we wish to calculate the free energy difference between two separate conformations of the system. For this we need the ability to apply a bias to the simulation to push it artificially toward a certain region of conformation space that we wish to examine. From calculating the bias needed along a path between two conformations one can obtain the free energy difference between then, an important measure of such quantities as the binding affinity of a drug for a specific active site of a protein (Michel and Essex, 2010). Two methods to calculate this free energy are umbrella sampling (Roux, 1995; Frenkel and Smit, 2001; Neale and Pomès, 2016; **Figure 3H**), where the path taken is through conformation space and what is known as a potential of mean force (PMF) (Roux, 1995) is calculated along this path and thermodynamic integration (Matos et al., 2017), an analogous calculation but where the path is through parameter space. The free energy calculations are computationally demanding and sensitive to force field details. Also one should consider possible artifacts due to a bias force, e.g., deformation of the lipid bilayers was observed in a few studies during umbrella sampling calculation of the profile of PMF of the studied compound along the bilayer normal (Neale et al., 2011; Filipe et al., 2014; Neale and Pomès, 2016). Metadynamics (Bussi and Laio, 2020) is an adaptive means to explore conformation space in an enhanced fashion by constantly biasing the system away from the regions of conformation space that have already been explored.

The remainder of this review paper will cover examples of how this tool, molecular dynamics simulation, has been and can continue to be, used in the context of drug delivery research (pharmaceutics). We will discuss applications across the breadth of the field, including obtaining insight relevant to dissolution and solubility, however, the majority of the discussion will cover the recent explosion in publications that use molecular dynamics simulation to study the more advanced drug delivery mechanisms, collectively known as nanomedicine.

MECHANISTIC INSIGHT INTO DRUG DISSOLUTION AND SOLUBILITY FROM MD SIMULATION

The simplest application of MD simulation in drug delivery is gaining mechanistic insight into the universal processes of dissolution and solvation (**Figure 4**). Drugs often enter the body in crystalline form and dissolution of these crystals is the first step. Larsen et al. (2017a,b, 2019) have used MD simulation to study alteration to the crystal structure with varying levels of hydration. For systems with long range order such as this, a more accurate and computationally intensive COMPASS force field (Sun, 1998) is required, instead of the potential sets normally used for simulations of systems in the liquid state. The Ouyang group has studied the dissolution of drug molecules complexed with solid dispersions as a remedy for poor solubility using MD (Chen and Ouyang, 2017; Chan and Ouyang, 2018; Han et al., 2019a) in addition to machine learning techniques (Han et al., 2019b). Coarse grained simulations using the DPD protocol have been used by Otto et al. to study the release of the drug quercetin from poly(ethylene-glycol) (PEG) solid dispersions (Otto et al., 2013).

As stated above, simplified QSAR/QSPR (Mathieu, 2020) or related machine learning models (Hutchinson and Kobayashi, 2019) are generally used to correlate drug structure to solubility using pattern recognition to relate structure to experimental solubility data; MD simulation can, however, be used to obtain both a more accurate result and, additionally, provide mechanistic understanding. The partition coefficient between water and octanol can be calculated for the specific molecule through MD simulation (Bannan et al., 2016) using the aforementioned techniques for free energy calculation, either by (1) using umbrella sampling to physically pull the candidate drug molecule structure through the boundary between a water and an octanol phase and calculate the free energy change along this path, the aforementioned PMF (example of a PMF calculation shown in **Figure 3H**) or (2) performing thermodynamic integration between the drug solvated in water and the drug solvated in octanol. Such a calculation is not the mechanistic insight advertised in the introduction; here we are using MD simulation as a tool to obtain a numerical estimate of a quantitative result. It is possible, however, to examine the MD simulation output further to obtain mechanistic insight regarding the relation between the structure of the molecule and the solvent; for example, Zhang et al. have investigated the H-bond network of the drug ibuprofen in water and ethanol (Zhang M. et al., 2020). Erlebach et al. (2020) have used a different technique combining simulations with atomistic resolution with solubility calculations based on Flory-Huggins theory. Other examples of MD used for solubility prediction also exist (Lüder et al., 2007, 2009; Westergren et al., 2007; Patel et al., 2010a; Gupta et al., 2011; Paluch et al., 2015; Matos et al., 2017; Matos and Mobley, 2019; Dasari and Mallik, 2020). To aid in the delivery of drugs that are otherwise too lipophilic, they are administered not alone but in a formulation with other molecules, known as excipients. Optimizing this drug formulation can be performed through combining screening experiments with pattern recognition and optimization algorithms, however, here too, MD simulation can play a powerful role in complementing other computational methods (Mehta et al., 2019), for example MD simulations of cyclodextrin-drug complexes (Zhao et al., 2018; Huang et al., 2019); cyclodextrin is a common agent for assisting the delivery of poorly soluble drugs. Persson et al. (2013) have used MD simulation to study drug solubility in excipient formulations and MD has been used to study polymeric excipients. Benson and Pleiss (2014) have used MD to study self-emulsifying drug delivery systems and Hathout et al. have modeled drug loading in the gelatin matrix (Ahmad et al., 2010;

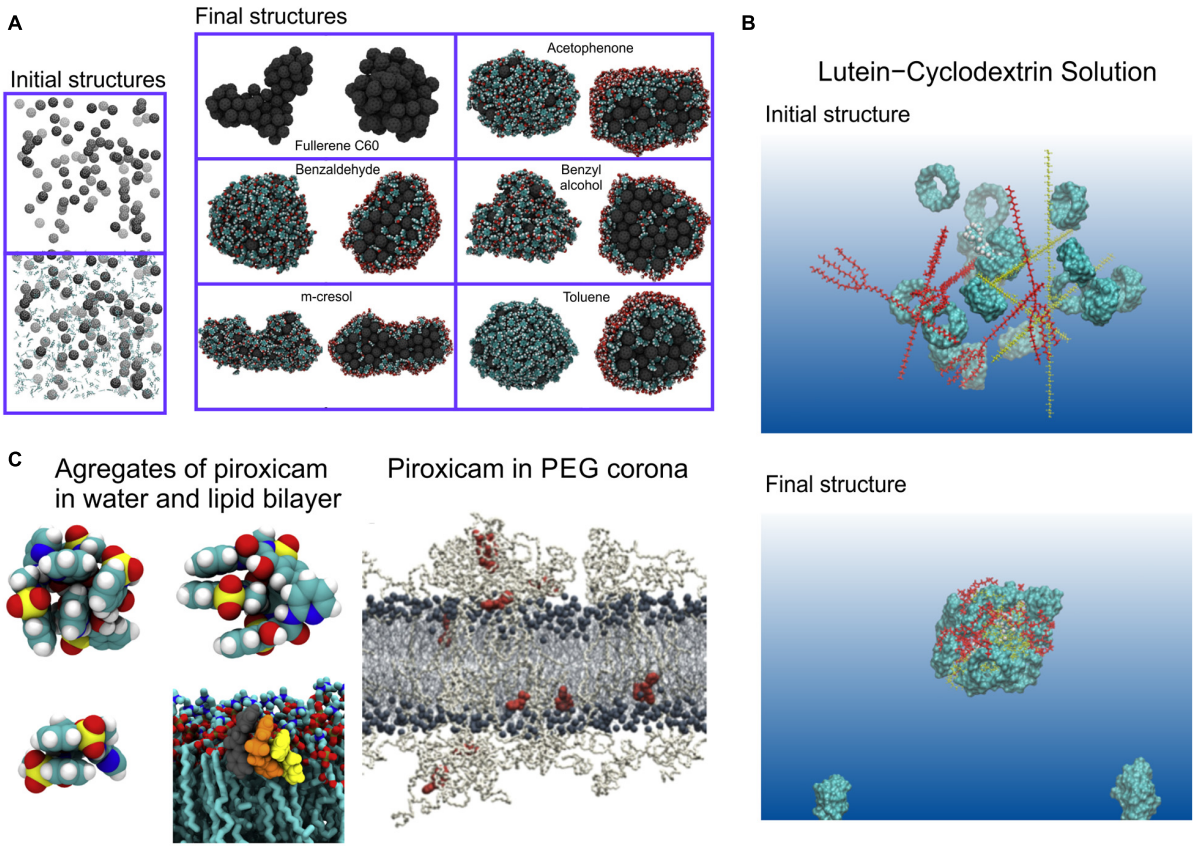


FIGURE 4 | Dissolution and solubility. **(A)** Formation of clusters of fullerene with organic solvents, for the final structures clusters full view (left side) and its cross section (right side) are shown, organic solvent covers fullerene from outside and are present in small quantities inside the cluster, reproduced from Lehto et al. (2014), Copyright: 2014; **(B)** formation of lutein and cyclodextrin complexes (Zhao et al., 2018), Copyright (2018) American Chemical Society; **(C)** Aggregates of piroxicam formed in water and lipid bilayer, and piroxicam molecules dispersed in PEG corona of PEGylated lipid bilayer, reproduced with permission from Wilkosz et al. (2017).

Warren et al., 2013; Jha and Larson, 2014; Hathout et al., 2020). Several comprehensive review papers have been written on the synergistic use of MD with other computational techniques to determine the solubility and dissolution characteristics of drugs and drug formulations (Johnson and Zheng, 2006; Bergström and Larsson, 2018; Li et al., 2018; Hossain et al., 2019; Das et al., 2020).

Describing the ease with which a drug travels through the body to reach its target through this one parameter, solubility, alone, is of course an extreme oversimplification: in addition to dissolving in the blood, drugs must traverse a variety of biological barriers, in particular cell membranes and perfect solubility will not insure this (Smith et al., 2018). Building systems to deliver drugs through these barriers requires an extra level of complexity; we now cross from simple formulation with the goal to optimize solubility into nanomedicine: nanoscale vectors designed to transport the drug through the bloodstream while protecting it from the body's defense mechanisms and targeting the desired tissue.

NANOMEDICINE

Nanomedicine is officially defined as pharmaceutical applications of nanotechnology. Since "nanotechnology" is a meaningless buzzword quickly fading from fashion (Park, 2019) this is not a concise definition; in practical terms this encompasses all drug delivery systems that involve packaging the drug in structures with diameters =100 nm but larger than a single drug molecule: one or more drug molecules combined with one or more carrier molecules. For example, even merely grinding a crystal of the drug into pieces smaller than this size officially fits this definition, the result known as "nanocrystals" (Song et al., 2011) and recognized as the simplest form of nanomedicine. A very broad range of mechanisms have been developed that fit this definition and the nomenclature is cluttered, i.e., the language used to define different varieties, and how components are described is inconsistent; we will now describe the nomenclature and definitions we intend to use, but be warned: when you read the cited publications, the nomenclature may not be consistent.

When the drug and carrier are combined, the result is referred to as a nanoparticle. Nanoparticles are formed in one of two ways: (1) directly functionalizing a molecule to the drug, i.e., chemically bonding a molecule to the drug to alter its behavior in the bloodstream (Ekladious et al., 2019) or (2) combining one or more drug molecules with one or more carrier molecules that self-assemble to form the nanoparticle; I will refer to this as the functionalization and self-assembly routes of nanoparticle

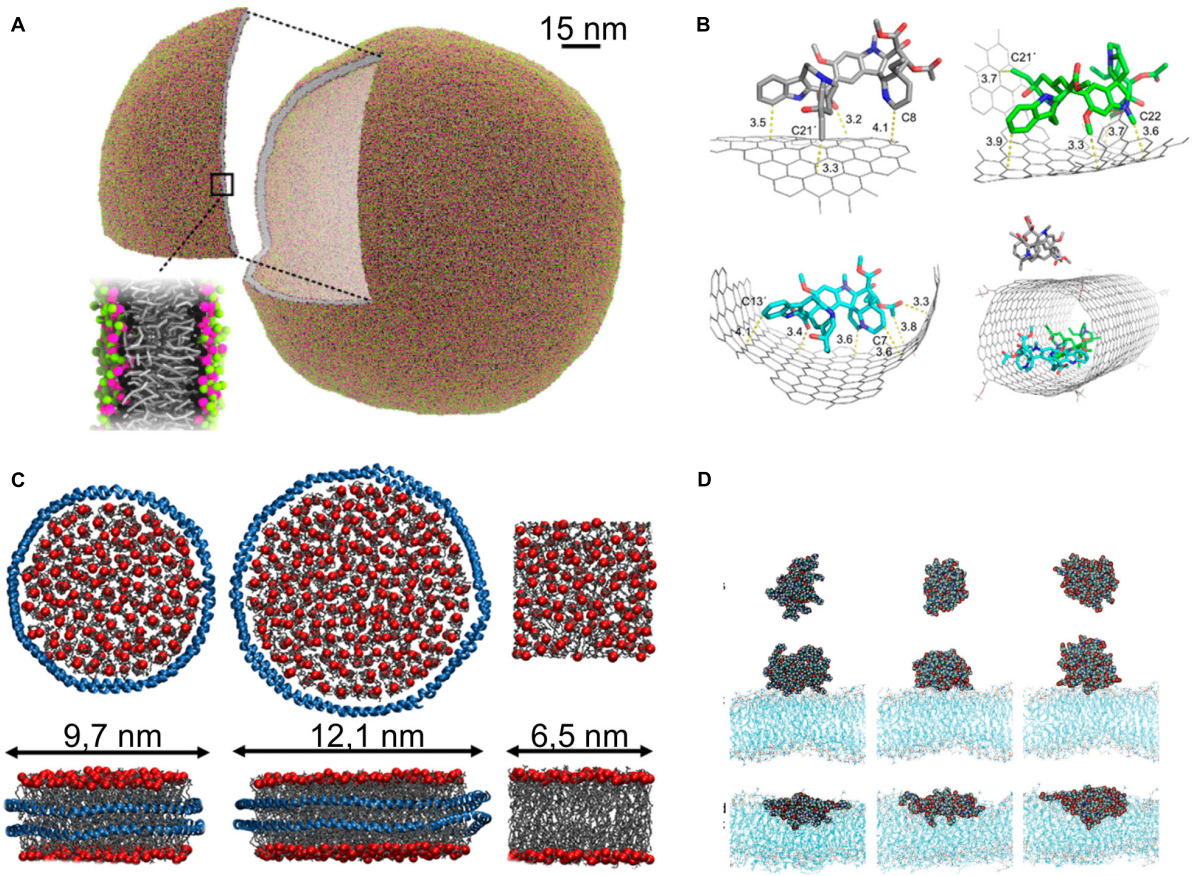


FIGURE 5 | Nanoparticles. **(A)** Liposome simulated with dry MARTINI model, reproduced with permission from Arnarez et al. (2015), Copyright (2015) American Chemical Society; **(B)** carbon nanotube used for delivery of vinblastine (Li et al., 2016a), Copyright (2016) American Chemical Society; **(C)** nanodiscs formed of POPC and membrane scaffold protein MSP1D1 (Left), MSP1E3D1 (middle), and lipid bilayer (right), protein is shown as blue ribbon, phosphate groups of lipids shown as red sphere, and acyl tail as gray sticks, reproduced with permission from Stepien et al. (2020); **(D)** PAMAM dendrimer in water phase (top), at the lipid bilayer in gel phase (middle), and at the lipid bilayer in fluid phase (down), reproduced with permission from Kelly et al. (2008), Copyright (2008) American Chemical Society.

formation. The functionalization route to nanoparticle creation can also lead to the formation of a nanoparticle composed of more than one drug molecule, for example functionalizing a hydrophobic drug with a polymer could result in the formation of micelles with the drugs at the core. In most cases the direct functionalization is to a polymer, a long unstructured molecule, that, as a result, forms a protective sheath around the drug molecule in the bloodstream, however functionalization to a smaller molecule is also possible, for example, folic acid (Wolski et al., 2018; Alinejad et al., 2020) or glycine (Ghadri et al., 2020). A particularly ingenious idea is functionalization to amphiphilic "molecular umbrellas" that aid the transfection of hydrophilic drugs through the hydrophobic core of cell membranes (Janout et al., 2001, 2002, 2005, 2014; Jing et al., 2003; Janout and Regen, 2005, 2009; Ge et al., 2009). Drugs functionalized to polymers where the drug is activated by enzyme cleavage of the polymer are also referred to as "prodrugs" (Luo et al., 2019). Functionalization to peptides or small proteins can result in very specific fine tuning of the behavior of the drug as it interacts with its environment (Lu et al., 2015). Functionalization of lipids for a variety of applications is reviewed by Kepczynski and Róg (2016) and specifically for drug delivery by Kohli et al. (2014). Regarding nanoparticles formed via the self-assembly route, a rigorous literature search leads to a subdivision of the majority according to topology and choice of carrier molecule into roughly the following 9 categories: (1) solid inorganic, (2) micelles, (3) vesicles (**Figure 5A**), (4) lipoprotein based structures (**Figure 5C**), (5) other lipid-polymer structures, (6) carbon architectures (**Figure 5B**), (7) dendrimers (**Figure 5D**), (8) protein/peptide, and (9) the aforementioned nanocrystals. Bobo et al. (2016) have compiled the list of FDA-approved forms of nanomedicine, as of 2016.

Solid inorganic nanoparticles are rigid structures formed from inorganic substances. These include gold (Ghosh et al., 2008; Charchar et al., 2016; Rossi and Monticelli, 2016), silver (Eckhardt et al., 2013; BurduŞel et al., 2018), titanium dioxide (Aranha et al., 2020), silica (Santos et al., 2014) nanoparticles, and boron nitride oxide nanoflakes (Duverger and Picaud, 2020). Gold and silver nanoparticles are solid structures that can be associated with drugs, or can be functionalized themselves to

perform a specific function: the nanoparticle itself is a drug. For the case of silica nanoparticles they can be porous and contain drugs and can even have complex multi-compartment structure, carrying many different drug molecules (Torchilin, 2007; Pattni et al., 2015; Bulbake et al., 2017; El-Hammadi and Arias, 2019; Bhardwaj et al., 2020; Crommelin et al., 2020), for example for applications like theragnostics (Janib et al., 2010). In the same fashion as solid inorganic nanoparticles, carbon architectures are contiguous solid structures, however, due to its unique chemistry, composition from carbon allows for a wide variety of forms, including carbon dots (Peng et al., 2017; Ghosal and Ghosh, 2019), nanotubes (Sun et al., 2014), nanodiamonds (Barnard, 2016; Ge and Wang, 2017), nanographene (Zhang L. et al., 2013; Sun et al., 2014; Sgarlata et al., 2016; Ghadari and Kashefi, 2017; Hasanzade and Raissi, 2017; Moradi et al., 2018; Alinejad et al., 2020; Mahdavi et al., 2020), and graphene oxide (Duverger and Picaud, 2020; Shahabi and Raissi, 2020).

Micelles and vesicles are both formed from amphiphilic organic molecules but differ in topology: micelles have a hydrophobic core surrounded by a hydrophilic shell while in vesicles the amphiphilic molecules form a bilayer that itself forms into an enclosed pocket. In both cases they can be formed from a wide range of molecules, usually surfactants, lipids or diblock copolymers, however, other amphiphilic molecules are possible, for example, janus dendrimers (Nummelin et al., 2017; Yang Y.-L. et al., 2019). The most common micellar nanoparticle is the polymeric micelle (Cagel et al., 2017), composed of diblock co-polymers with hydrophobic drugs carried in the micelle core. The most common form of vesicular nanoparticle is the liposome (Bunker et al., 2016), a vesicle formed from naturally occurring phospholipids. Other amphiphilic molecules formed into vesicles are, however, also used in drug delivery, including ethosomes (Touitou et al., 2000), niosomes (Marianecci et al., 2014; Khan and Irchhaiya, 2016; Chen S. et al., 2019; Kapoor et al., 2019; Khalkhali et al., 2019; Inglut et al., 2020), polymersomes (Aibani et al., 2020; Khan et al., 2020), exosomes (Antimisiaris et al., 2018; Villa et al., 2019; Chung et al., 2020; Rahmati et al., 2020), ufasomes (Han, 2013), and drimersomes (Nummelin et al., 2017), comprehensive reviews have been written about vesicle formation (Šegota and Durdica, 2006) and application in drug delivery (Kapoor et al., 2019) in a general context. Polymers and lipids can be formed into other structures than micelles or vesicles, for example two different polymers can be used to form core-shell structures (Ramli et al., 2013; Abbott et al., 2017; Chen G. et al., 2018), for example a solid outer shell with a liquid polymer with drug encapsulated inside; solid lipid nanoparticles (Beloqui et al., 2016; Gordillo-Galeano and Mora-Huertas, 2018; Subramaniam et al., 2020), chitosan (Bernkop-Schnürch and Dünnhaupt, 2012), lipoplex (Scheideler et al., 2020) and other lipid-polymer nanoparticles (Date et al., 2018) have also been proposed. Another form of polymer based nanoparticle is dendrimers (Tomalia et al., 1990; Fatemi et al., 2020) and pseudodendrimers (Ghadari and Sabri, 2019), hyper-branched polymers with a fractal structure that results in a molecule that is, qualitatively, in the form of a fuzzy ball and can store molecules in their interior or bind nucleic acids to form a dendrimerplex. A particularly common form of dendrimer that has been proposed for drug delivery is poly(amidoamine) (PAMAM) dendrimers (Xiao et al., 2020).

Lipoproteins are used as the body's mechanism for lipid transport. These are structures of several different lipids with proteins that control the form of the structure and the composition of the lipid types within the structure. As they transport lipids they undergo structural change upon deposition of their cargo from a spherical structure to a disk-like structure. Taking these structures as a starting point and modifying them to work as drug carriers, or building structures inspired by lipoproteins, is a novel avenue of nanomedicine that is currently being explored (Bricarello et al., 2011; Huang et al., 2015; Kuai et al., 2016a; Simonsen, 2016; Aranda-Lara et al., 2020; Chuang et al., 2020). The disk-like form of lipoprotein, known as nanodiscs have proven to be an extremely useful structure for a variety of applications, including nanomedicine (Denisov and Sligar, 2017). Nanodiscs were successfully used as a drug delivery vehicles to treat viral lung infections (Numata et al., 2013) and were used as a platform accommodating antigens and adjuvants in personalized cancer vaccines (Kuai et al., 2016b). Use of nanodiscs for simultaneous delivery of antigen and adjuvant has been found to increase the response of the immunological system by orders of magnitude in comparison to traditional vaccines. Due to the variety of possible applications of nanodiscs, their properties are the subject of intensive study (Debnath and Schäfer, 2015; Siuda and Tieleman, 2015; Stepien et al., 2015, 2020; Martinez et al., 2017; Pourmousa and Pastor, 2018; Bengtsen et al., 2020; Schachter et al., 2020); they are tuned via modification of their lipid composition (Augustyn et al., 2019) or alterations to the sequence, thus structure, of the scaffold proteins (Denisov et al., 2004; Nasr et al., 2016).

All of these structures can have their properties fine-tuned by being functionalized to polymers or smaller molecules themselves, in the same fashion as described above for the drug molecule itself. For example functionalizing poly(ethylene glycol) (PEG) (Israelachvili, 1997), a process known as "PEGylation" (Bunker, 2012, 2015; Pasut and Veronese, 2012; Bunker et al., 2016; Zhang Z. et al., 2020) has been proposed and studied for virtually all of these nanoparticle forms and, as we will discuss in further detail in the next section, alternate polymers to PEG are under investigation. The extent to which these systems can be fine-tuned is limitless, for example formulation alteration of liposomes offer an extremely broad pallet (Bunker et al., 2016; Li et al., 2019). We are thus left with several variables for their formulation in addition to the extremely complex environment of human physiology with which they interact, the topic that we will now discuss.

NANOPARTICLE DESIGN AND FUNCTION

Nanoparticles have been developed to assist in drug delivery in a very broad range of pharmaceutical contexts, for example treating atherosclerosis (Lobatto et al., 2011; Chen J. et al., 2020; Ramalho et al., 2020) and other neurodegenerative diseases (Goldsmith et al., 2014), cardiovascular disease (Godin et al., 2010), diabetes

(Veiseh et al., 2015), infections disease (Zhu et al., 2014; Zazo et al., 2016), protein drugs (Qin et al., 2019), and vaccine delivery (Pison et al., 2006; Zhao et al., 2014) in fact vaccine adjuvant development involves many of the same mechanisms as nanomedicine (Copland et al., 2005; Wang et al., 2019); it can be argued that it is only for historical reasons that it is not referred to as nanomedicine. The main application of nanomedicine is, however, cancer therapy (Tong and Kohane, 2016; Youn and Bae, 2018), particularly chemotherapy agent delivery, as this involves drugs with extremely high toxicity; targeted delivery, where the drug is kept from the rest of the body and the greatest possible fraction is delivered to the target tissue, in this case the tumor, is extremely desirable. The nanoparticle is designed to have features that protect the drug, in the context of nanomedicine commonly referred to as the "payload" of the nanoparticle. Targeting is achieved through either active or passive means. Active targeting (Nag and Delehanty, 2019) involves a specific ligand functionalized to the nanoparticle exterior that binds to receptors that are overexpressed in the outer cell membrane of cells of the target tissue and passive targeting involves global properties (Ogawara et al., 2013) of the nanoparticle that lead to a greater percentage becoming lodged in the target tissue in comparison to other tissues. An example of passive targeting is what is referred to as the enhanced permeability and retention (EPR) effect (Maeda et al., 2013); liposomes can be designed to take advantage of the leaky vasculature of tumor tissue to become preferentially lodged there; PEGylation is a common means to achieve this. It must, however, be stated that whether or not the EPR effect is an effective passive targeting strategy in practical nanomedicine applications, has recently been brought into question (Danhier, 2016).

The nanoparticle thus carries and protects its drug payload through the bloodstream and preferentially delivers it to its target tissue. In the bloodstream, foreign particles in the size range of nanoparticles are removed (uptaken) by the mononuclear phagocyte system (MPS) (Chow et al., 2011); this involves an extremely complex and specific cascade of proteins: complement activation (Ricklin et al., 2010; Sarma and Ward, 2011). The efficiency with which a nanoparticle is removed through complement activation is determined by its surface properties. The nanoparticle can be designed to have a surface that inhibits uptake, thus prolonging circulation in the bloodstream and, as a result, the amount of the drug that reaches the target tissue per administered dose; such a nanoparticle surface is referred to as a "stealth sheath" and the aforementioned PEGylation is the gold standard to achieve a this (Pasut and Veronese, 2012; Bunker, 2015; Parray et al., 2020). While PEGylation is an extremely successful strategy, it is not perfect and the investigation of alternate polymers to PEG is an active field of research (Knop et al., 2010). Alternatives that have been proposed and studied include polyoxazolines (Sedlacek et al., 2012; Lorson et al., 2018), PASylation® (Viegas et al., 2011; Schlapschy et al., 2013; García et al., 2014; Binder and Skerra, 2017; Gebauer and Skerra, 2018), zwitterionic polymers (García et al., 2014), hydroxyethyl starch (Liebner et al., 2014), and polypeptides (Hou and Lu, 2019).

PEGylation, or the creation of an alternate polymer stealth sheath, is achieved though functionalizing the polymer to a component of the nanoparticle. For the case of the functionalization route to nanoparticle creation, functionalization to the protective polymer itself can be the nanoparticle. It is also possible to functionalize the drug to a copolymer where one of the copolymer constituents is the hydrophilic stealth sheath and the other performs another function, e.g., a hydrophobic polymer that encapsulates the drug. Examples of this include PEGylated boron nitride (Farzad and Hashemzadeh, 2020), folic acid (Wolski et al., 2017b; Alinejad et al., 2020), interferon (Xu et al., 2018), insulin (Yang et al., 2011; **Figure 6E**), other PEGylated peptides (Xue et al., 2011; Hamed et al., 2015; Ma et al., 2016; **Figure 6F**) and protein drugs (Katre, 1993; Jevševar et al., 2010; Yang et al., 2011; Zhang et al., 2012; Mu et al., 2013; Wu et al., 2014; Lawrence et al., 2014; Nischan and Hackenberger, 2014; Lawrence and Price, 2016; Xu et al., 2018; Wilding et al., 2018; Gupta et al., 2019; Zaghmi et al., 2019; Munasinghe et al., 2019; Kaupbayeva and Russell, 2020; **Figure 6B**); the broader context of polymer-protein drug molecules is covered in several reviews (Pelegri-Oday et al., 2014; Wang et al., 2019). As far back as 1977, long before "nano" was a word, functionalizing PEG to proteins was proposed to alter their immunological properties (Abuchowski et al., 1977). For the self-assembly route to nanoparticle creation the polymers are functionalized to constituent molecules of the nanoparticle. PEGylation has been proposed for virtually every one of the nanoparticle types described in the previous section. This includes PEGylated carbon nanotubes (Pennetta et al., 2020), gold nanoparticles (Oroskar et al., 2016; Lin et al., 2017; Sun et al., 2019), silver nanoparticle (Pinzaru et al., 2018), silver-graphene nanoparticles (Habiba et al., 2015), nano-graphene (Zhang et al., 2014; Zhang Z. et al., 2020; Mahdavi et al., 2020), lipid micelles (Arleth et al., 2005; Viitala et al., 2019; **Figure 6D**), nanodiscs (Zhang et al., 2014), dendrimers (Kojima et al., 2000; Lee and Larson, 2009, 2011; Zhang et al., 2014), and a topic covered comprehensively in our previous review, liposomes (Bunker et al., 2016; **Figures 6A,C**).

For the case of inorganic nanoparticles, in particular gold nanoparticles, various alternatives to PEG coatings have been considered. Gold nanoparticles can be functionalized via a thiol group with hydrocarbons capped with a methyl group (Bolintineanu et al., 2014; Potdar and Sammalkorpi, 2015; Giri and Spohr, 2018), hydroxyl group (Potdar and Sammalkorpi, 2015; Villarreal et al., 2016; Yamanaka et al., 2019), carboxylic group (Heikkilä et al., 2014b; Giri and Spohr, 2018; **Figure 7A**), amine group (Heikkilä et al., 2014a,b; Giri and Spohr, 2018; Das et al., 2019; Lolicato et al., 2019), choline sulfate (Yamanaka et al., 2019), or a para-mercaptobenzoic acid (**Figure 7B**; Salorinne et al., 2016). Also, bulky branched coatings have been used to functionalize gold nanoparticles (Giri and Spohr, 2018; Yamanaka et al., 2019). The alternative coating can also be used to direct the nanoparticle to a selected environment, e.g., Potdar and Sammalkorpi proposed using a hydrophobic coating to cause the particle to locate to the hydrophobic core of the bilayer and a coating ended with a hydroxyl group to anchor the particle to the lipid headgroups (Potdar and Sammalkorpi, 2015). A coating composed of two types of moieties one a hydrophobic 1-octanethiol and the other a negatively charged

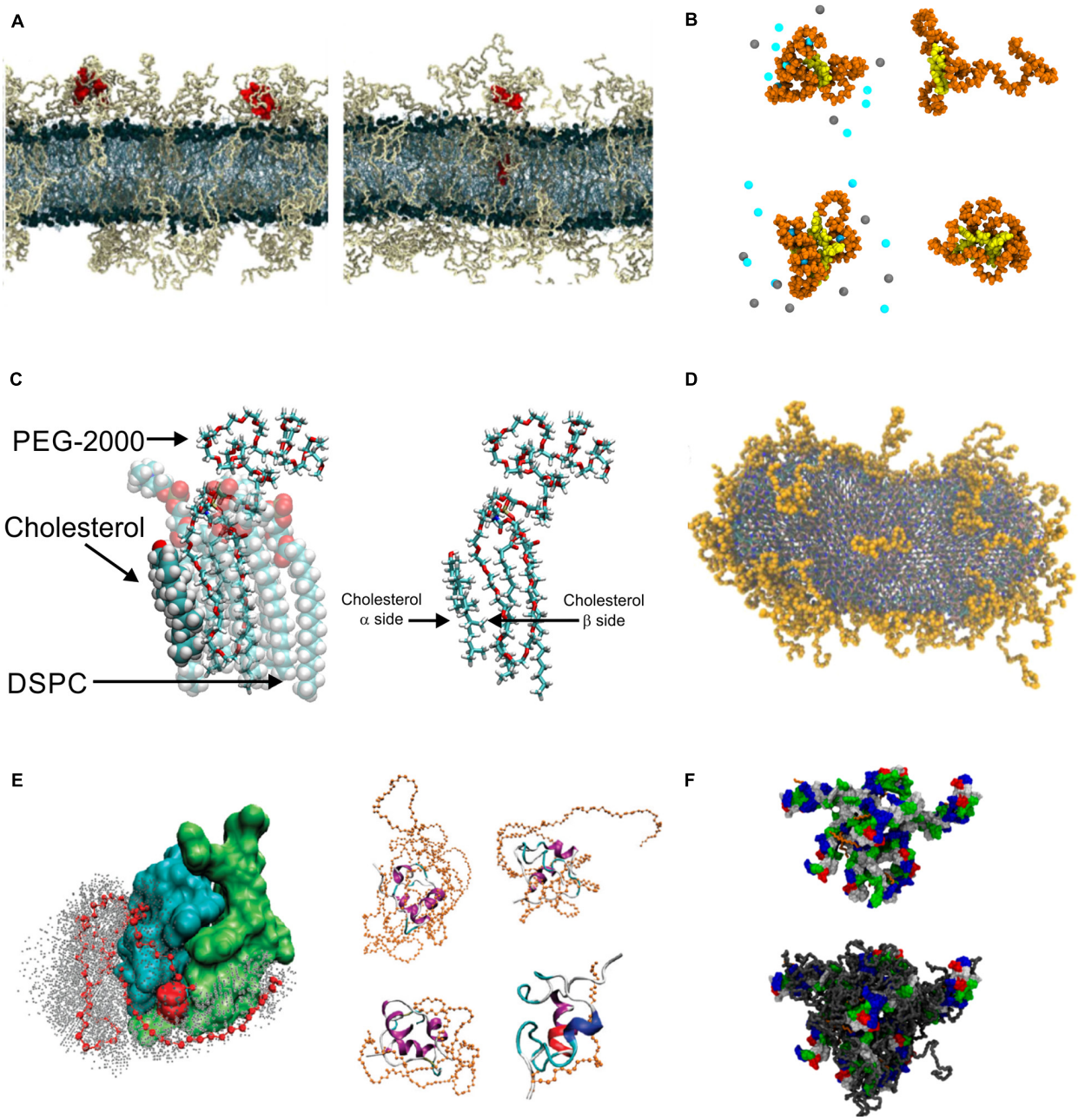


FIGURE 6 | PEGylation. **(A)** Snapshot showing PEGylated lipid bilayer, reproduced with permission from Dzieciuch et al. (2015), Copyright (2015) American Chemical Society; **(B)** PEGylated biochanin (upper) and tetra-phenyl-porphyrin (lower), with salt (left) and without salt (right), reproduced with permission from Rissanen et al. (2014), Copyright (2014) American Chemical Society; **(C)** Snapshots showing DSPC, cholesterol and DSPE-PEG molecules, reproduced with permission from Magarkar et al. (2014), Copyright (2014) American Chemical Society; **(D)** PEGylated bicelle containing 10.5 mol % DSPE-PEG, reproduced from Viitala et al. (2019), Copyright: 2019; **(E)** PEGylated insulin, left panel shows position of PEG atoms during simulations, right panel shows snapshots of insulin PEGylated with PEG of various length, reproduced with permission from Yang et al. (2011), Copyright (2011) American Chemical Society.

11-mercapto-1-undecanesulfonate causes the nanoparticle to locate to the center of the bilayer with its polar sulfonate groups exposed to the water at both membrane interfaces; this induces a local thinning of the bilayer (Van Lehn et al., 2013; Van Lehn and Alexander-Katz, 2014a, 2019; Simonelli et al., 2015) (Fi), or possibly even large scale deformation (Salassi et al., 2017). With the same coating moieties with polar coating placed on one-half of the particle and non-polar on the other, one can form an amphiphilic gold nanoparticle that will locate to the boundary between the water phase and the hydrophobic membrane core, *i.e.,* the position of the lipid headgroups (Ou et al., 2020; **Figure 7D**). Such

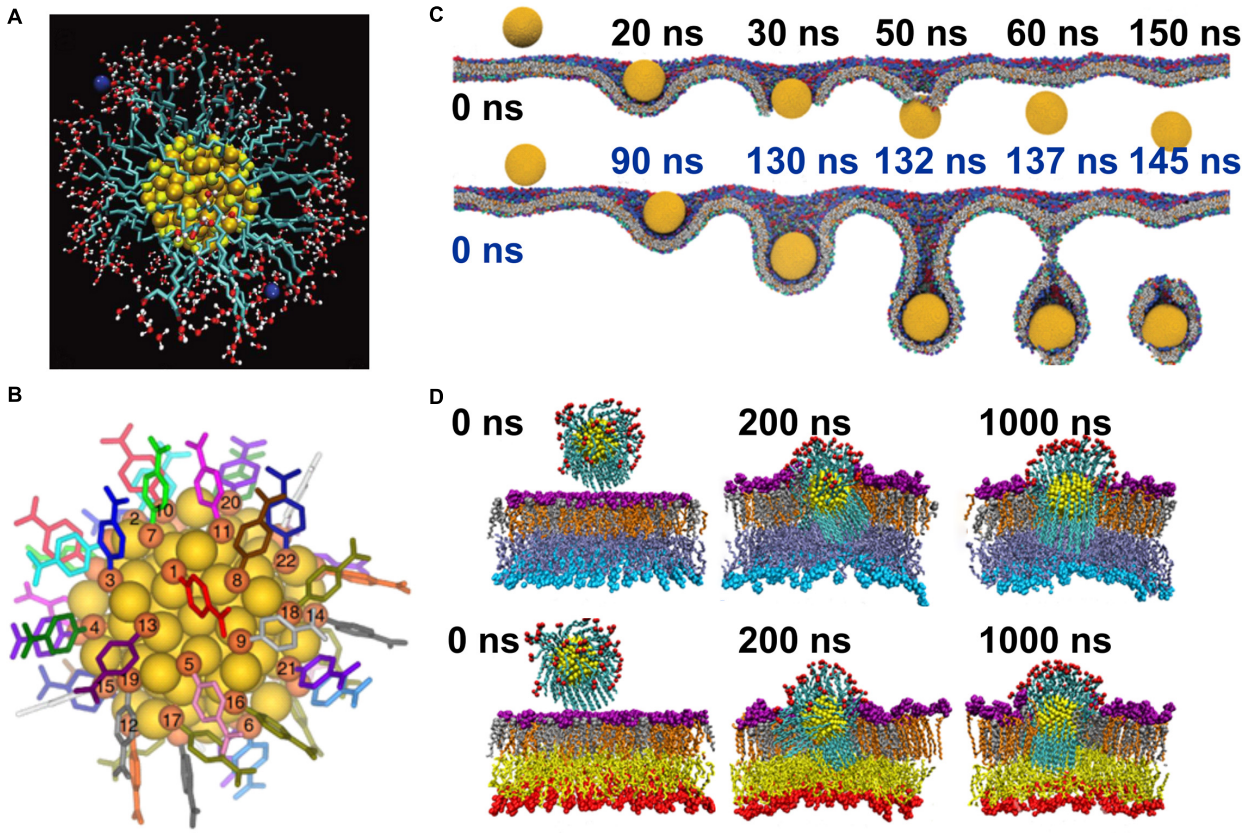


FIGURE 7 | Gold nanoparticles. **(A)** Gold nanoparticle coated with dodecanoic acid, hydrating water and ions are shown, reproduced with permission from Heikkilä et al. (2012), Copyright (2012) American Chemical Society; **(B)** gold nanoparticle coated with para-mercaptobenzoic acid, gold is shown in yellow and sulfur in orange, reproduced from Salorinne et al. (2016), Copyright: 2016; **(C)** snapshots of (upper panel) internalizations of a neutral gold nanoparticle and (lower panel) uptake of a positively charged gold nanoparticle, reproduced with permission from Lunnoo et al. (2019), Copyright (2019) American Chemical Society; **(D)** snapshots of the MD trajectory of the insertion of amphipathic janus nanoparticle into lipid bilayers, reproduced with permission from Ou et al. (2020), Copyright (2020) American Chemical Society.

coating of other solid inorganic nanoparticles has also been considered, e.g., silver nanoparticles were coated with hydrophilic polymer poly(N-vinyl-2-pyrrolidone) (Kyrychenko et al., 2015), graphene nanoflakes with ssDNA (Moore et al., 2019), and silica nanoparticle with hydrocarbons (Peters et al., 2012).

Proteins will agglomerate to any foreign particle in the bloodstream in the approximate size range of a nanoparticle resulting in a shell of proteins surrounding them, known as the "protein corona" (Walkey and Chan, 2012; Xiao et al., 2013; del Pino et al., 2014; Kharazian et al., 2016; Mahmoudi, 2016; Hadjidemetriou and Kostarelos, 2017; Pederzoli et al., 2017; Brancolini and Tozzini, 2019; Casalini et al., 2019; Nienhaus and Nienhaus, 2019; Zhadanov, 2019; Berrecoso et al., 2020; Gupta and Roy, 2020). The stealth sheath modulates the formation of this corona in a fashion that is not completely understood and has been a point of contention in the field for several decades. Regarding PEGylation, it was originally thought that it inhibits protein adhesion (Du et al., 1997; Bradley et al., 1998) then others found evidence that it actually accelerates protein corona formation (Szebeni et al., 2002) and yet others argued that they found evidence it had no effect (Price et al., 2001). It has been argued that the PEG sheath preferentially binds the common bloodstream protein albumin (Vert and Domurado, 2000) creating an albumin protein corona that, itself, acts as the stealth sheath that inhibits complement activation (Caracciolo, 2015). Alternate protective mechanisms unrelated to the protein corona have also been proposed, including direct inhibition of absorption by macrophages (Price et al., 2001). Most recently, evidence has been found that the formation of the protein corona is essential for the stealth properties of PEG (Schöttler et al., 2016). The most recent reviews of this much discussed topic are found here (Nienhaus and Nienhaus, 2019; Zhadanov, 2019; Li Z. et al., 2020).

Complement activation and the formation of a protein corona is only one aspect of the environment that the nanoparticle must traverse; in addition to the body's defenses the nanoparticle must navigate the hydrodynamic environment of the bloodstream and, in most cases, deliver the payload drug through the cell membrane. While the surface properties of the nanoparticle play a role, both of these are heavily influenced by its size, shape (Truong et al., 2015) and rigidity/elasticity (Geng et al., 2007; Lee S.-Y. et al., 2009; Toy et al., 2011). Once the nanoparticle

reaches the bloodstream its environment can be approximated as laminar flow in a cylinder. In this environment, in addition being pushed in the direction of the flow, a particle is subject to a force perpendicular to the flow that causes the particle to move toward the cylinder wall, a phenomenon known as margination (Gentile et al., 2008). Evolution has taken advantage of this: red blood cells are relatively rigid and have a disk-like form in order to minimize margination, as disk-like and more rigid particles experience a lesser extent of this force in comparison to spherical and more elastic particles; leukocytes have evolved to have the opposite structure, spherical and elastic, as margination to the blood vessel wall plays an essential role in their function (Lee S.-Y. et al., 2009). This is one of the reasons why the previously mentioned nanodiscs are a very promising form of nanoparticle, however, it is not the only reason: size, shape and elasticity of the particle also affect the interaction between nanoparticle and cell membranes (Lin X. et al., 2010; Zhang et al., 2012; Banerjee et al., 2016). Several design features of the nanoparticle are involved in tuning its properties to deliver the payload through biological barriers (Blanco et al., 2015) like the cell membrane, into the target cell and in some cases to a specific organelle within the cell; both surface properties and the size (Lin X. et al., 2010; Lunnoo et al., 2019), shape (Lin X. et al., 2010; Zhang et al., 2012; Lunnoo et al., 2019), and elasticity of the nanoparticles play a role in this. Also presence of negatively charged lipids affects intake of functionalized, cationic gold nanoparticles (Lolicato et al., 2019). There are several mechanisms through which this is possible; nanoparticles can directly permeate the membrane (Song et al., 2011), in many cases disrupting its structure (**Figure 7C**). For the case of liposomes and micelles (De Nicola et al., 2014), the payload can be delivered through membrane fusion and the nanoparticle can also be designed to induce endocytosis (Vácha et al., 2011). As mentioned previously, nanoparticles can be functionalized with targeting ligands that trigger preferential uptake by target cells (Bazak et al., 2015). It, however, must be said that active targeting, while a popular topic for research, has so far seen limited success; as far as the authors are aware there is only one approved therapy that features active targeting: Denileukin Diftox (Turturro, 2014). Finally, nanoparticles can be designed to release their drug payload when there is a certain external trigger, a scheme known as controlled release. This trigger can be pH change that occurs during endocytosis or an externally applied trigger used to cause the drug to release in the tissue to which this trigger is applied, for example a locally applied optical magnetic or thermal trigger (**Table 1**).

Altogether, we see that the landscape of nanomedicine is extremely complex, both with a wide range of directions that nanoparticle design can take and the extremely complex environment of human physiology and the body's natural defenses. While *in vitro* experimental insight and clinical studies can make some progress, one quickly reaches a dead end in a sea of complexity without the rational design approach made possible by a mechanistic understanding. The next section shows how molecular dynamics simulation, alongside complementary experimental analysis techniques, to some extent provide this.

MOLECULAR DYNAMICS SIMULATION APPLIED TO NANOMEDICINE

Now that we have outlined the different forms of nanomedicine and the issues encountered by nanoparticles in their context as drug delivery agents, we can proceed to showcase many examples where molecular dynamics simulation, using different degrees of coarse graining, have provided mechanistic insight that complements the research program to develop new nanoparticle based drug delivery mechanisms. The amount of work carried out in this area has exploded in the past decade, with molecular dynamics studies being applied to virtually every variety of nanoparticle discussed above in their context as drug delivery vehicles, including dendrimers, gel nanoparticles, polymeric micelles, other polymeric forms of nanoparticles, solid lipid nanoparticles, other micelles, nanocrystals, carbon dots, carbon nanotubes, nanographene, DNA nanotubes, nanodiamonds, peptide nanoparticles, gold nanoparticles, silver nanoparticles, silica nanoparticles, latex nanoparticle and vesicles, of which the application of molecular modeling to liposome based drug delivery systems is covered comprehensively in our previous review (Bunker et al., 2016); there has, however, been a significant amount of work performed since its publication, and molecular modeling has now been applied to the study of other vesicle based drug delivery systems including niosomes, ufasomes, polymeric vesicles (polymersomes), and glyceryl monostearate vesicles. A list of publications that feature the use of molecular dynamics modeling to study each of these systems if found in **Table 2**. One intriguing omission by the scientific community is lipoprotein based nanoparticles, including nanodiscs. Nanodiscs have been studied in the context of their possible use as a drug delivery mechanism and have been studied, in a general context, using molecular dynamics simulation, however, molecular dynamics simulation has never been applied in the context of their possible use in drug delivery.

Regarding the functionalization route to nanoparticle development, there has also been a considerable amount of computational study carried out using molecular dynamics modeling. Protein structures can be downloaded and their potentials have already been parameterized; attach a polymer to the protein, solvate in water and you can study its behavior. Both PASylated (Hedayati et al., 2017) and PEGylated (Cohan et al., 2011) human recombinant erythropoietin have been simulated; Munasinghe et al. (2019) used molecular dynamics

TABLE 1 | Triggers used to release drug payload.

Trigger type and references
pH change (Guo et al., 2010; Zheng et al., 2011; Nie et al., 2013, 2014; Wang et al., 2015a, 2016, Luo Z. et al., 2016; Rungrotmongkol and Poo-arporn, 2016; Min et al., 2017; Wang Y. et al., 2017, Wang Z. et al., 2017; Wolski et al., 2017b, 2018; Quan et al., 2017; Gao et al., 2019; Wu W. et al., 2019; Wu Z. et al., 2019; Maleki et al., 2020)
Optical (Lajunen et al., 2016, 2018; Massiot et al., 2017)
Magnetic (Panczyk et al., 2013; Yang C. et al., 2020; Zhang X. et al., 2020)
Thermal (Dhawan et al., 2004; Pérez-Sánchez et al., 2020)

TABLE 2 | Drug delivery vehicles studied using molecular modeling methods.

Carbon Dots - (Erimban and Daschakraborty, 2020)
Carbon Nanotubes - (Panczyk et al., 2013, 2020; Izadyar et al., 2016; Li et al., 2016a; Rungrotmongkol and Poo-arporn, 2016; Hashemzadeh and Raissi, 2017; Kamel et al., 2017; Wolski et al., 2017a, 2018, 2020, 2019; Zaboli and Raissi, 2017; Karnati and Wang, 2018; Kavyani et al., 2018a,b; Zhang et al., 2018; Contreras et al., 2019; Dehneshin et al., 2019; Mortazavifar et al., 2019; Kordzadeh et al., 2019; Ghadri et al., 2020; Maleki et al., 2020; Pakdel et al., 2020; Pennetta et al., 2020)
Dendrimers - (Kojima et al., 2000; Lee et al., 2002, 2011; Maiti and Bagchi, 2006; Lee and Larson, 2008, 2009, 2011; Vasumathi and Maiti, 2010; Nandy and Maiti, 2011; Huynh et al., 2012; Nandy et al., 2012, 2013; Jain et al., 2013, 2016; Kłos and Sommer, 2013; Tian and Ma, 2013; Tu et al., 2013; Martinho et al., 2014; Wen et al., 2014; Jiang et al., 2015; Kavyani et al., 2016, 2018b,a; Smeijers et al., 2016a,b; Badalkhani-Khamseh et al., 2017, 2019; Yang et al., 2017; Farmanzadeh and Ghaderi, 2018; Ghadari and Mohammedzadeh, 2018; Gupta and Biswas, 2018; Su et al., 2018; Ghadari and Sabri, 2019; Ramos et al., 2019; He et al., 2020; Kłos and Paturej, 2020)
DNA Nanotubes - (Liang et al., 2017)
Gel Nanoparticles - (Kasomova et al., 2012; Smith et al., 2020)
Glyceryl Monostearate Vesicles - (Marwah et al., 2018)
Gold Nanoparticles - (Sun and Xia, 2003; Lin J. et al., 2010; Lin et al., 2011; Kyrychenko et al., 2011; Mhashal and Roy, 2014; Gupta and Rai, 2016, 2017; Mhasal and Roy, 2016; Oroskar et al., 2016; Gupta et al., 2017, 2018; Quan et al., 2017; Yang et al., 2017; Sridhar et al., 2018; Xie et al., 2018; Lunnoo et al., 2019, 2020; Tavanti et al., 2019; Yamanaka et al., 2019; Exner and Ivanova, 2020)
Latex Nanoparticle - (Li et al., 2016b)
Liposomes - (Dhawan et al., 2016; Lajunen et al., 2016; Pathak et al., 2016; Dzieciuch-Rojek et al., 2017; Laudadido et al., 2017; Magarkar et al., 2017; Wilkosz et al., 2017; Belubbi et al., 2018; Monpara et al., 2018; Poojari et al., 2020)
Nanocrystals - (Song et al., 2011)
Nanodiamonds - (Chen et al., 2009; Adnan et al., 2011)
Nanodiscs - (Ghosh et al., 2011, 2014; Koivuniemi and Vattulainen, 2012; Zhang et al., 2012, 2014; Pan and Segrest, 2016; Denisov and Sligar, 2017; Pourmousa and Pastor, 2018; Augustyn et al., 2019; Damiati et al., 2019; Chen Q. et al., 2020; Lundsten et al., 2020; Stepien et al., 2020)
Nanographene - (Zhang L. et al., 2013; Sgarlata et al., 2016; Ghadari and Kashefi, 2017; Hasanzade and Raissi, 2017; Moradi et al., 2018; Alinejad et al., 2020)
Niosomes - (Myung et al., 2016; Ritwiset et al., 2016; Somjid et al., 2018)
Oher Micelles - (De Nicola et al., 2014; Chun et al., 2015; Johnston et al., 2016)
Other Polymeric Forms of Nanoparticles - (Guo et al., 2009a,b; Durbin and Buxton, 2010; Rodríguez-Hidalgo et al., 2011; Macháèková et al., 2013; Buxton, 2014; Loverde, 2014; Razmimanesh et al., 2015; Esalmi et al., 2016; Ghitman et al., 2019; Mazloom-Jalali and Shariatinia, 2019; Shadrack and Swai, 2019; Golda-Cepa et al., 2020)
Peptide Nanoparticles - (Lu et al., 2015; Miller et al., 2019; Nikfar and Shariatinia, 2019)
Polymeric Micelles - (Ghosh et al., 2008; Kuramochi et al., 2009; Guo et al., 2010, 2012a; Loverde et al., 2011; Vukoviæ et al., 2011; Zheng et al., 2011; Kasomova et al., 2012; Luo and Jiang, 2012; Yang et al., 2012, Yang C. et al., 2019, Yang C. et al., 2020; Nie et al., 2013, 2014; Srinivas et al., 2013; Lin et al., 2014, 2019; Wang et al., 2015b; Luo S. et al., 2016; Luo et al., 2019; Myint et al., 2016; Prhashanna et al., 2016; Ramezani and Shamsara, 2016; Shi et al., 2016; Aziz et al., 2017; Min et al., 2017; Chang et al., 2017; Hu et al., 2017; Mousavi et al., 2018; Raman et al., 2018; Albano et al., 2019; Alves et al., 2019; Wu W. et al., 2019; Wu Z. et al., 2019; Gao et al., 2019; Hao et al., 2019; Kacar, 2020; Koochaki et al., 2020)
Polymeric Vesicles (Polymersomes) - (Luo Z. et al., 2016; Wang Z. et al., 2017; Grillo et al., 2018)
Silica Nanoparticles - (Soltani et al., 2010; Mousavi et al., 2019)
Silver Nanoparticles - (Sun and Xia, 2003; Kyrychenko et al., 2015; Blazhynska et al., 2018)
Solid Lipid Nanoparticles - (Hathout and Metwally, 2016)
Ufasomes - (Han, 2013; Csongradi et al., 2017; Bolla et al., 2019)

simulation to study conjugation of PEG to a hydrophobic pocket of bovine serum albumin using a model with atomistic resolution and Wilding et al. (2018) used a coarse grained model to study site specific PEGylation of the protein lysozyme. Atomistic MD has been used to study the effect of PEGylation on the stability and potency of interferon (Xu et al., 2018) and insulin (Yang et al., 2011) and the steric shielding effect that results from the PEGylation of Staphlokinase (Mu et al., 2013). A recent comprehensive overview of the application of molecular simulation to the study of protein-polymer conjugation has been written by Lin and Colina (2019).

In terms of the delivery of specific drugs using nanomedicine, a very large number have been simulated incorporated into a wide variety of nanoparticle types. These drugs include Alzheimer's medication, anti-worm drugs, antibiotics, anti-cancer drugs, including chemotherapy agents, anti-viral agents, antifungal drugs, anti-inflammatory drugs, antimicrobial peptides, drug used for diabetes treatment, immunomodulators and immunosuppressants, local anesthetics, and others; a list is found, with citations, in **Table 3**. Altogether, it becomes clear that there is simply too much work that has been carried out to concisely summarize in its entirety in this review. We will instead focus on a few key areas where MD modeling has provided important insight and discuss review papers that focus on certain aspects of the use of molecular dynamics in the context of nanomedicine and some key examples of original research that demonstrate the power of the technique. The discussion will include key examples where we show concrete insight gained my molecular dynamics simulation. We will focus on three areas: (1) behavior of the nanoparticle in the bloodstream and the protective polymer corona, (2) drug loading and release and (3) nanoparticle interaction with lipid membranes and entry into the cell. We would like to here alert the reader to the fact that there are other reviews of aspects of the use of computational

TABLE 3 | List of drugs studied with MD simulations in context of drug delivery.

Drugs, theirs applications, and references
5-flouracil - anti-cancer drug (Barraza et al., 2015; Kacar, 2019)
Albendazole - anti-worm drug (Rodríguez-Hidalgo et al., 2011)
Amphotercin B - antifungal drugs (Mobasheri et al., 2016)
Anakinra - used in arthritis therapy (Liebner et al., 2014)
Camptothecin - chemotherapy agent (Ansari et al., 2018; Alinejad et al., 2020)
Carmustine - chemotherapy agent (Wolski et al., 2017a; Mortazavifar et al., 2019)
Chlortetracycline - antibiotic (Dowlatabadi et al., 2019)
Cisplatin - chemotherapy agent (Panczyk et al., 2013)
Curcurbitacin drug families (Patel et al., 2010a)
Cyclosporine - immunosuppressant (Tokarský et al., 2011)
Dicolofenac - anti-inflammatory agents (Karjiban et al., 2012)
Doxorubicin - chemotherapy agent (Guo et al., 2010, 2012b; Yang et al., 2012; Yang C. et al., 2019; Yang Y.-L. et al., 2019; Zhang et al., 2012, 2014, 2018; Nie et al., 2013; Shan et al., 2014; Lin et al., 2014, 2019; Izadyar et al., 2016; Rungrotmongkol and Poo-arporn, 2016; Wolski et al., 2017b, 2018, 2019; Hu et al., 2017; Mousavi et al., 2018; Kordzadeh et al., 2019; Alinejad et al., 2020; Exner and Ivanova, 2020; Maleki et al., 2020; Pakdel et al., 2020; Koochaki et al., 2020; Li J. et al., 2020
Erlotinib - anti-cancer drugs (Hlaváč et al., 2018)
Exemestane - breast cancer drug (Ghadri et al., 2020)
Flavonoid (Myung et al., 2016; Laudadido et al., 2017)
Flutamide - prostate cancer drug (Kamel et al., 2017)
Fluvestrant - breast cancer drug (Ghadri et al., 2020)
Gemcitabine - chemotherapy agent (Razmimanesh et al., 2015; Sgarlata et al., 2016; Ansari et al., 2018; Farzad and Hashemzadeh, 2020)
GF-17 - antimicrobial peptide (Asadzadeh et al., 2020)
Ibuprofen - pain medication and anti-inflammatory (Thota et al., 2016; Kacar, 2020)
Ifofamide - chemotherapy agent (Mazloom-Jalali and Shariatinia, 2019; Shariatinia and Mazloom-Jalali, 2019)
Insulin - diabetes treatment (Yang et al., 2011)
Interferon - immunomodulator (Xu et al., 2018)
Interferon Alpha - anti-cancer and anti-viral agent (Gupta et al., 2018)
Itraconazole - antifungal drugs (Dzieciuch-Rojek et al., 2017; Poojari et al., 2019, 2020)
Letrozole - breast cancer drug (Ghadri et al., 2020),
Metronidazole antibiotic (Kumar et al., 2019)
Nicotine (Zaboli and Raissi, 2017; Li Z. et al., 2020)
Nystatin - antifungal drugs (Mobasheri et al., 2016)
Paclitaxel (taxol) - chemotherapy agent (Guo et al., 2009a, 2012b; Loverde et al., 2011; Wang et al., 2013; Ghadari and Kashefi, 2017; Hasanzade and Raissi, 2017; Hashemzadeh and Raissi, 2017; Monpara et al., 2018)
Piaglitazone (Zaboli and Raissi, 2017)
Picoplatin - colorectal cancer drug (Farmanzadeh and Ghaderi, 2018)
Piroxicam (Wilkosz et al., 2017)
Prilocane - local anesthetic (Grillo et al., 2018)
Sorafenib - kidney cancer drug (Dehneshin et al., 2019)
Streptozotocin - neuendocrine tumors drug (Dehneshin et al., 2019)
Sunitinib - renal carcinoma medication (Dehneshin et al., 2019)
Tacrine - Alzheimer's medication (Esalmi et al., 2016)
Vinblastine - chemotherapy agent (Li et al., 2016a)

modeling for nanoparticle design (Angioletti-Uberti, 2017b; Bouzo et al., 2020).

MD INSIGHT EXAMPLES

Behavior in the Bloodstream and Protective Polymer Corona

As we discussed previously, when the nanoparticle enters the bloodstream it encounters hydrodynamic forces and a corona of bloodstream proteins forms on its surface; a subset of these proteins form the highly specific complement activation reaction that leads to removal by macrophages. Regarding behavior in the bloodstream and the effect of size and shape (Shah et al., 2011), the most suitable method is not MD, but rather a combination of theoretical calculation (Decuzzi et al., 2005) and a discretized continuum model known as computational fluid dynamics (CFD), described and used to model this by Li et al. (2014b), Gupta (2016), and Gao et al. (2020) to model nanoparticle transport in the faulty tumor vasculature (Gao et al.). As we have mentioned, the formation of the protein corona is an extremely complex process that still remains poorly

understood. What is clear, however, is that the surface properties of the nanoparticle affect this and the mechanism through which the protective polymer corona increases the bloodstream lifetime is modulation of the interaction with bloodstream proteins. Schafer et al. (2017) and Settanni et al. (2017a,b, 2018) combined experimental analysis with MD to study the interaction between two protective polymers, PEG and poly-sarcosine, with a set of proteins found in the bloodstream. They found evidence that the interactions are not amino acid specific but rather a general tendency dependent on the charge and polarity of the amino acid and the nature of the interaction between the polymer and water, in addition to the direct polymer-protein interaction. Their methodology, synergistically combined with experimental work, could provide a route to a rational design approach to the development of new polymer materials being developed that may have superior performance as a protective polymer corona. Lee et al. used the coarse grained MARTINI model to directly simulate the effect of PEGylation and PEGylation density on the interaction between the liposome and blood stream proteins; Lee also used MD simulation with atomistic resolution to study the effect of nanoparticle electrostatics in protein corona formation (Lee, 2020a). There are other examples of the use of MD modeling to study the protein corona of nanoparticles (Dell'Orco et al., 2010; Vilaseca et al., 2013; Lopez and Lobaskin, 2015; Shao and Hall, 2016).

In our previous review publication, focused on liposome based delivery systems (Bunker et al., 2016), we surveyed the work that had been carried out using molecular dynamics modeling, particularly with a model with all atom resolution, on the interaction between the protective PEG corona and the lipid bilayer (Stepniewski et al., 2011; Magarkar et al., 2012, 2014). Since this time the methodology has been used to study the effect of exchanging PEG with two different poly-oxazolines, poly-ethoxazoline (PEOZ) and poly-methoxazoline (PMOZ), with the result indicating that several properties of PEG are highly specific and related to its amphiphilic nature and the ease with which it acts as a polymer electrolyte (Magarkar et al., 2017). We also simulated the effect of change in PEG length, branched structures, and functionalizing PEG to the cholesterol or cholane in the membrane rather than phospholipids and our results complemented both *in vivo* and *in vitro* experiments carried out on these novel liposome based delivery systems (Mastrotto et al., 2020).

PEGylation, in the context of other nanoparticle forms than liposomes, has also been studied extensively using MD simulation. Ambrosio et al. (2018) complemented experimental study by demonstrating, using MD simulation with all atom resolution, that a 2:1 ratio or greater of PEG-cholane molecules to the VIP-palm peptide being delivered, is required to form supramolecular assemblies; these assemblies were shown to effectively cover the VIP- peptide with a protective corona of PEG. In previous work we have used MD simulation to study the PEGylation of small drug molecules (Li Y.-C. et al., 2012). Two recent reviews, written by Lee, very comprehensively cover MD simulation work, using coarse grained in addition to all atom models, to study the structure and behavior of PEGylated nanoparticles, one covering PEGylated biomolecules, liposomes and solid nanoparticles (Lee, 2020b) and the other covering PEGylated peptides dendrimers and carbon nanotubes (Lee, 2014). Li et al. carried out a coarse grained MARTINI model simulation to investigate the effect of PEG chain length on the shielding effect of PEGylated nanoparticles (Li and Hu, 2014). A comprehensive review of computational modeling of PEGylation has been written by Souza et al. (2018).

Drug Loading and Controlled Release

The ability of nanoparticles to hold drugs and release them with an external trigger has been studied for several nanoparticle forms by several groups. In most cases the drugs being considered are hydrophobic and sit within a non-polar region. Nanoparticles that have been simulated carrying their drug payload include carbon nanotubes, nanographene, peptide carriers, PAMAM dendrimers, polymeric nanoparticles, polymeric micelles, hydrophobic drugs within the membrane of liposomes, other issues related to drug loading of liposomes (Cern et al., 2014) and polymersomes (Grillo et al., 2018) (further citations found in **Table 4**). Drug cargoes studied include cucurbitacin, carmustine, 5-flouracil (Barraza et al., 2015), chacone, picoplatin, porphyrins, ibuprofen, paclitaxel, and albendazole, however, the most popular drug for these model systems is doxorubicin (see **Table 4** for citations). In many cases these nanoparticles are designed to release their drug payload in response to a pH change trigger (see **Table 1**); MD simulation is able to model the effect of pH change. In an MD simulation pH is modeled through the partial charges on the atoms, so the system can be equilibrated with the partial charges corresponding to neutral pH and then the partial charges can be changed to model pH change and the behavior of the system in response to this, i.e., the drug release, can be modeled. One interesting aspect of the work carried out using MD simulation in this area is that use of all three levels of coarse graining is represented: atomistic MD, MARTINI model and DPD (**Table 4**). Reading this literature with this in mind provides a very good case study of the strengths and weaknesses of each model and the aspects of the system each are most ideally suited to investigate.

Studies of itraconazole in a liposome, combining MD simulation with experiment, provides an example of where MD simulation was able to provide concrete insight not obtainable experimentally. Itraconazole is an antifungal drug characterized by low solubility, which limits its bioavailability. One possible solution to overcome low solubility is incorporating drugs into liposomes, which was achieved in a few independent studies. To optimize the liposome properties, cholesterol is frequently used as a molecule known to increase the stability of lipid bilayers (Róg and Vattulainen, 2014). In fact, cholesterol is used in 9 out of 15 liposome-based formulations approved for clinical use (Bulbake et al., 2017). Thus, the incorporation of cholesterol into the liposome-itraconazole formulation was the next step. MD simulations showed that this is not the right choice because cholesterol and itraconazole do not mix well and separate into small domains (Poojari et al., 2020). This observation was next validated in experimental studies, which

showed decreased affinity of itraconazole toward liposomes containing cholesterol (Poojari et al., 2020). The observed behavior of the itraconazole may be explained by its structure: it is a long rigid molecule with a few weakly polar groups distributed along the molecule backbone. Due to this structure, itraconazole molecules locate to the water membrane interface oriented parallel to the membrane surface and, in turn, orient other molecules in the same fashion (Poojari et al., 2019). On the other hand, cholesterol has strong preferences to adopt an orientation perpendicular to the membrane surface and affect the orientation of neighboring molecules; these opposite preferences lead to the observed demixing of drugs and cholesterol.

Nanoparticle Interaction With the Lipid Membrane

Once the nanoparticle has reached the cell, surviving the journey through the bloodstream with its payload still contained and intact, there is one final barrier that must be crossed for the drug delivery system to be efficacious: the cell membrane

TABLE 4 | Nanoparticles, cargo molecules, and methods used to study drug loading and release.

Nanoparticles
Carbon Nanotubes (Wolski et al., 2017a; Kordzadeh et al., 2019; Ghadri et al., 2020)
Liposomes (Cern et al., 2014; Dzieciuch et al., 2015)
Nanographene (Moradi et al., 2018; Mahdavi et al., 2020; Maleki et al., 2020)
PAMAM dendrimers (Kelly et al., 2009; Wen et al., 2014; Barraza et al., 2015; Badalkhani-Khamseh et al., 2017, 2019; Farmanzadeh and Ghaderi, 2018; Fox et al., 2018)
Peptide Carriers (Thota et al., 2016)
Polymeric Micelles (Patel et al., 2010a,b; Guo et al., 2012a; Kasomova et al., 2012; Nie et al., 2014; Myint et al., 2016; Shi et al., 2016; Gao et al., 2019; Kacar, 2019; Wu W. et al., 2019)
Polymeric Nanoparticles (Shen et al., 2017; Yahyaei et al., 2017; Ghitman et al., 2019; Styliari et al., 2020)
Polymersomes (Grillo et al., 2018)
Cargo Molecules:
5-flouracil (Barraza et al., 2015)
Albendazole (da Silva Costa et al., 2020)
Carmustine (Wolski et al., 2017a)
Chacone (Badalkhani-Khamseh et al., 2019)
Cucurbitacin (Patel et al., 2010a)
Doxorubicin (Yang et al., 2012; Nie et al., 2013; Kordzadeh et al., 2019; Koochaki et al., 2020; Mahdavi et al., 2020; Maleki et al., 2020)
Ibuprofen (Thota et al., 2016)
Paclitaxel (Wang et al., 2013)
Picoplatin (Farmanzadeh and Ghaderi, 2018)
Porphyrins (Stepniewski et al., 2012; Rissanen et al., 2014; Dzieciuch et al., 2015)
Methods:
Atomistic MD (Patel et al., 2010a; Wang et al., 2013; Barraza et al., 2015; Dzieciuch et al., 2015; Shi et al., 2016; Thota et al., 2016; Badalkhani-Khamseh et al., 2017, 2019; Dzieciuch-Rojek et al., 2017; Wolski et al., 2017a; Grillo et al., 2018; Moradi et al., 2018; Kordzadeh et al., 2019; Ghadri et al., 2020; Maleki et al., 2020)
MARTINI model (Grillo et al., 2018; Koochaki et al., 2020)
DPD (Guo et al., 2009a,b, 2010, 2012a; Yang et al., 2012; Nie et al., 2013, 2014; Wen et al., 2014; Myint et al., 2016; Wang et al., 2016, 2015b; Gao et al., 2019; Wu W. et al., 2019; Kacar, 2019)

TABLE 5 | Nanoparticles and methods used to study theirs interactions with membranes.

Nanoparticles:
Carbon Dots (Erimban and Daschakraborty, 2020)
Graphene (Raczyński et al., 2020), dendrimers (Lee and Larson, 2008; Kanchi et al., 2018; He et al., 2020)
Gold Nanoparticles (Lin et al., 2011; Gkeka et al., 2014; Mhashal and Roy, 2014; Mhasal and Roy, 2016; Oroskar et al., 2016; Quan et al., 2017; Das et al., 2019)
Nanocrystals (Song et al., 2011)
Methods:
Atomistic MD (Mhashal and Roy, 2014; Van Lehn and Alexander-Katz, 2014b; Mhasal and Roy, 2016; Erimban and Daschakraborty, 2020)
MARTINI model (Lin X. et al., 2010; Lin X. et al., 2020; Song et al., 2011, 2012; Lin and Gu, 2014; Oroskar et al., 2015; Shimizu et al., 2016; Quan et al., 2017; Su et al., 2017; Zhang Z. et al., 2017; Bai et al., 2018; Das et al., 2019; Salassi et al., 2019; He et al., 2020)
DPD (Lee and Larson, 2008; Yang and Ma, 2010; Ding and Ma, 2012, 2014a; Ding et al., 2012; Tian et al., 2014b; Liu et al., 2016; Bai et al., 2018),
Implicit Solvent (Vácha et al., 2011; Schubertová et al., 2015)

(Smith et al., 2018). It is possible for nanoparticles, particularly if they are hydrophobic, to directly transfect, also referred to as translocation, through the cell membrane and many nanoparticles enter the cell through this route. There is, however, an alternative: the nanoparticle can be designed to cross the membrane via receptor mediated endocytosis (Gao et al., 2005). When a nanoparticle is taken up via endocytosis it is possible to take advantage of pH triggered release due to the lowered pH environment on the interior of the endosome (Hu et al., 2015). Valuable insight in both the context of direct membrane transfection and endocytosis have been obtained through MD simulation. As is the case with drug loading and release, MD simulation of nanoparticle-lipid membrane interactions have been carried out for different nanoparticle forms, including carbon dots, graphene, dendrimers, gold nanoparticles (shown in **Table 5**), and nanocrystals (Song et al., 2011); examples can be found of all levels of model resolution being used including atomistic MD, MARTINI model, DPD, and implicit solvent (**Table 5**). An overview of MD simulation of nanoparticle – lipid membrane interactions has been written by Tian et al. (2014a).

For nanoparticles that enter the cell through direct transfection, the issue is the direct physical reaction between the nanoparticle and the membrane; this phenomenon can be studied directly through an MD simulation of the nanoparticle interacting with the membrane (Yang and Ma, 2010; Ding et al., 2012; Liang, 2013; Van Lehn and Alexander-Katz, 2014b; Shimizu et al., 2016; Zhang Z. et al., 2017; Erimban and Daschakraborty, 2020; Gupta et al., 2020). When nanoparticles translocate through the membrane, the membrane structure can be disrupted and leakage and even pore formation can occur; this has been studied directly using MD (Song et al., 2012; Mhashal and Roy, 2014; Van Lehn and Alexander-Katz, 2014b; Oroskar et al., 2015; Ding and Ma, 2018). The effect of size (Lin X. et al., 2010), shape (Li Y. et al., 2012; Liu et al., 2016; Yang Y. et al., 2019), and surface properties (Ding and Ma, 2016) of the nanoparticle on membrane transfection has also been studied, including effect of PEGylation (Oroskar et al., 2016; Bai et al., 2018), and other polymer coatings (Liang, 2013; Xia et al., 2020) as well as protein (Ding and Ma, 2014a) and, for the study of inhaled nanoparticles, pulmonary surfactant corona (Bai et al., 2018) and other issues related to translocation across the pulmonary surfactant monolayer (Chen P. et al., 2018). Additionally, Gupta et al. used MD simulations to study transdermal delivery of interferon-alpha using gold nanoparticles (Gupta et al., 2018).

Regarding receptor mediated endocytosis, the interaction is more complex; while direct simulation of nanoparticle endocytosis has been performed and gained important insight (Vácha et al., 2011; Ding and Ma, 2012; Li et al., 2014a) this only tells part of the story as many aspects of the specific interaction between the ligand and the receptors are not elucidated by such a simulation. Nanoparticles can be designed to actively target specific cell types through functionalizing targeting ligands onto the nanoparticle surface. These targeting ligands bind to the specific receptors that induce endocytosis. There are two issues that govern the efficacy of this binding: (1) the distribution of the targeting ligands on the surface, i.e., the pattern of where they are located and (2) the orientation and, as a result of orientation, extent of exposure at the nanoparticle surface and thus availability to the target receptors. The effect of ligand distribution has been studied by Liu et al. (2016) through direct MD simulation of nanoparticle-membrane interactions and ligand density has been studied through a different computational modeling technique: Monte Carlo simulation (Martinez-Veracoechea and Frenkel, 2011; Wang and Dormidontova, 2012; Angioletti-Uberti, 2017a).

Regarding the orientation, and thus exposure, of the targeting ligand to the receptor that induces receptor mediated endocytosis, one needs chemically accurate atomistic simulations of the nanoparticle surface to investigate the degree to which the targeting ligand is exposed and available to the receptor. We have performed such simulations for the case of liposome based delivery systems, with targeting ligands, in several previous publications, for example our determination of the cause of failure of the new AETP moiety (Lehtinen et al., 2012). These involved simulating a section of the liposome membrane with the targeting ligands and, in some cases, the protective polymer corona as well. Our study regarding the AETP moiety was another example of a specific topic where we were able to obtain concrete insight, not obtainable experimentally. The AETP moiety was found to be successful, when its binding affinity for the target receptor was tested experimentally, however, when functionalized to a PEGylated liposome the targeting moiety failed to show any effect. In comparison to the more hydrophilic RGD peptide, that has been shown to be an effective targeting moiety for a PEGylated liposome, the AETP moiety is more hydrophobic; it could be hypothesized, from the experiment alone, that the moiety is obscured within the membrane core; our MD simulation, however, showed this not to be the case: it was rather the PEG corona itself that was obscuring the AETP moiety; as PEG is soluble in both polar and non-polar solvents it was thus a more comfortable, i.e., less hydrophilic, environment for the AETP moiety than the polar solvent (Lehtinen et al., 2012). Since it was the PEG corona itself that was the culprit we could propose a solution: replace PEG with a more hydrophilic polymer that has been approved for internal use. Just such a polymer exists: Poly-methoxazoline (PMOZ); in a subsequent study we performed a simulation with the PEG corona replaced by a PMOZ corona and we saw increased exposure of the AETP moiety (Magarkar et al., 2017). We have also studied liposomes functionalized with stearylamine arginine ligands (Pathak et al., 2016). A comprehensive review of the theoretical and computational investigation of nanoparticle interactions with biomembranes has been written by Ding and Ma (2014b).

CONCLUSION

In this review, we have attempted to summarize the role that molecular dynamics modeling can play as a tool in drug delivery research in a fashion that is hopefully comprehensible to both those with an expertise in molecular modeling who wish to pursue pharmaceutical applications of their research and pharmaceutical researchers interested in what insight this new tool can provide. All aspects of the journey that the drug

molecule takes, from dissolution to solvation or transit through the bloodstream inside a nanoparticle, to finally crossing the plasma membrane of the target cell, is a story of molecular interactions. The interactions involved, however, are not all interactions. Any chemist reading this review will have noticed an omission: chemical reactions; these play a very small role in the story, one dominated by intermolecular interactions. For this reason, MD simulation is the perfect tool to obtain molecular level insight as precisely the variety of interaction it is best able to study are those which play the dominant role: formation of structure based on lipophilicity and H-bonding. Whether it is the hydration that occurs during dissolution, interaction between drug molecule and excipients, behavior of molecules at the surface of the nanoparticle in the bloodstream, or the interaction between the nanoparticle and the plasma membrane of the target cell, these are the interactions that determine the most important aspects of behavior.

Molecular dynamics modeling is still a new tool in the design of drug delivery mechanisms; only in the past decade have we seen the explosion in the number of publications that make use of this tool. Widespread adoption is hindered by the fact that, unlike computational drug design tools like ligand docking and QSAR/QSPR, the calculations involved are, as of yet, for the most part too expensive to be carried out anywhere other than national level supercomputing resources. As the widely available computational power continues to grow exponentially, this barrier may dissipate. Looking toward the future and the role that molecular dynamics modeling will be able to play in the development of drug delivery systems, the analogy that we feel is most apt is that of computationally assisted design (CAD) (Narayan et al., 2008), in mechanical and civil engineering. Before the advent of computational technology, engineers were forced to build scale models of systems and experiment with them, testing every aspect with real experimental models and sometimes varying parameters empirically. Now, with widespread computational resources available to all engineers, CAD allows every aspect of a new machine, or structure, to be examined and tested in silico with all aspects of mechanical stress, heat dissipation etc. of the system visible, and the change resulting from any design alteration straightforward to analyze entirely virtually. While we clearly do not mean to imply that human physiology is no more complex than designing a car or a bridge, we foresee that, in the future, drug delivery devices will be designed in an analogous fashion, with molecular dynamics modeling playing the role in pharmaceutics that CAD plays in mechanical and civil engineering. Our studies of the AETP targeting moiety and itraconazole in liposome based delivery systems show clear examples of how the design approach can be applied, using *in silico* modeling to test aspects of the delivery system design in an analogous approach to CAD. Alongside cutting edge experimental techniques that complement it, molecular dynamics simulation has the potential to lead the way to a new era of rational design in the development of drug delivery systems. Finally, other complementary reviews that cover similar material can also be found (Thota and Jiang, 2015; Ramezanpour et al., 2016; Thewalt and Tieleman, 2016; Katiyar and Jha, 2018; Sen et al., 2018; Shamsi et al., 2019).

AUTHOR CONTRIBUTIONS

AB and TR wrote the manuscript. Both authors contributed to the article and approved the submitted version.

REFERENCES

Abbott, A. P., Ahmed, E. I., Prasad, K., Qader, I. B., and Ryder, K. S. (2017). Liquid pharmaceuticals formulation by eutectic formation. *Fluid Phase Equilib.* 448, 2–8. doi: 10.1016/j.fluid.2017.05.009

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

Abuchowski, A., Van Es, T., Palczuk, N. C., and Davis, F. F. (1977). Alteration of immunological properties of bovine serum albumin by covalent attachment of polyethylene glycol. *J. Biol. Chem.* 252, 3578–3581.

Acharya, C., Coop, A., Polli, J. E., and Mackerell, A. D. Jr. (2011). Recent advances in ligand-based drug design: relevance and utility of the conformationally samples pharmacophore approach. *Curr. Comput. Drug Des.* 7, 10–22. doi: 10.2174/157340911793743547

Adnan, A., Lam, R., Chen, H., Lee, J., Schaffer, D. J., Barnard, A. S., et al. (2011). Atomistic simulation and measurement of pH dependent cancer therapeutic interactions with nanodiamond carrier. *Mol. Pharm.* 8, 368–374. doi: 10.1021/mp1002398

Aguayo-Ortiz, R., and Dominguez, L. (2019). APH-1A component of γ-secretase forms an internal water and ion-containing cavity. *ACS Chem. Neurosci.* 10, 2931–2938. doi: 10.1021/acschemneuro.9b00150

Ahmad, S., Johnston, B. F., Mackay, S. P., Schatzlein, A. G., Gellert, P., Sengupta, D., et al. (2010). In silico modelling for drug-polymer interactions for pharmaceutical formulations. *J. R. Soc. Interf.* 7, S423–S433. doi: 10.1098/rsif.2010.0190.focus

Aibani, N., Khan, T. N., and Callan, B. (2020). Liposome mimicking polymersomes; a comparative study of the merits of polymersomes in terms of formulation and stability. *Int. J. Pharm. X* 2:100040. doi: 10.1016/j.ijpx.2019.100040

Albano, J. M. R., Grillo, D., Facelli, J. C., Ferraro, M. B., and Pickholz, M. (2019). Study of the lammellar and micellar phases of pluronic F127: a molecular dynamics approach. *Processes* 7:606. doi: 10.3390/pr7090606

Alinejad, A., Raissi, H., and Hashemzadeh, H. (2020). Understanding co-loading of doxorubicin and camptothecin on graphene and folic acid-conjugated graphene for targeting drug delivery: classical MD simulation and DFT calculation. *J. Biomol. Struct. Dyn.* 38, 2737–2745. doi: 10.1080/07391102.2019.1645044

Allen, M. P., and Tildesley, D. J. (1989). *Computer Simulation of Liquids*. New York, NY: Clarendon Press.

Alonso, H., Bliznyuk, A. A., and Gready, J. E. (2006). Combining docking and molecular dynamics simulations in drug design. *Med. Res. Rev.* 26, 531–568. doi: 10.1002/med.20067

Alves, V. M., Hwang, D., Muratov, E., Sokolsky-Papkov, M., Varlamova, E., Vinod, N., et al. (2019). Cheminformatics-driven discovery of polymeric micelle formulations for poorly soluble drugs. *Sci. Adv.* 5:eaav9784. doi: 10.1126/sciadv.aav9784

Ambrosio, E., Podmore, A., Gomes Dos Santos, A. L., Magarkar, A., Bunker, A., Caliceti, P., et al. (2018). Control of peptide aggregation and fibrillation by physical PEGylation. *Biomacromolecules* 19, 3958–3969. doi: 10.1021/acs.biomac.8b00887

Angioletti-Uberti, S. (2017a). Exploiting receptor competition to enhance nanoparticle binding selectivity. *Phys. Rev. Lett.* 118:68001. doi: 10.1103/PhysRevLett.118.068001

Angioletti-Uberti, S. (2017b). Theory simulations and design of functionalized nanoparticles for biomedical applications: a soft matter perspective. *NPJ Comput. Mater.* 3:48. doi: 10.1038/s41524-017-0050-y

Ansari, M., Moradi, S., and Shahlaei, M. (2018). A molecular dynamics simulation study on the mechanism of loading of gemcitabine and camptothecin in poly lactic-co-glycolic acid as a nano drug delivery system. *J. Mol. Liq.* 269, 110–118. doi: 10.1016/j.molliq.2018.08.032

Antimisiaris, S. G., Mourtas, S., and Marazioti, A. (2018). Exosomes and exosome-driven vesicles for targeted drug delivery. *Pharmaceutics* 10:218. doi: 10.3390/pharmaceutics10040218

Aranda-Lara, L., Morales-Avila, E., Luna-Gutiérrez, M. A., Olivé-Alvarez, E., and Isaac-Olivé, K. (2020). Radiolabeled liposomes and lipoproteins as lipidic nanoparticles for imaging and therapy. *Chem. Phys. Lipids* 230:104934. doi: 10.1016/j.chemphyslip.2020.104934

Aranha, M. P., Mukherjee, D., Petridis, L., and Khomami, B. (2020). An atomistic molecular dynamics study of titanium dioxide adhesion to lipid bilayers. *Langmuir* 36, 1043–1052. doi: 10.1021/acs.langmuir.9b03075

Arleth, L., Ashok, B., Onyuksel, H., Thiyagarajan, P., Jacob, J., and Hjelm, R. P. (2005). Detailed structure of hairy mixed micelles formed by phosphatidylcholine and PEGylated phospholipids in aqueous media. *Langmuir* 21, 3279–3290. doi: 10.1021/la047588y

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. Theory Comput.* 11, 260–275. doi: 10.1021/ct500477k

Asadzadeh, H., Moosavi, A., and Arghavani, J. H. (2020). The effect of chitosan and PEG polymers on stabilization of GF-17 structure: a molecular dynamics study. *Carbohydr. Polym.* 237:116124. doi: 10.1016/j.carbpol.2020.116124

Augustyn, B., Stepien, P., Poojari, C., Mobarak, E., Polit, A., Wisniewska-Becker, A., et al. (2019). Cholesteryl hemisuccinate is not a good replacement for cholesterol in lipid nanodiscs. *J. Phys. Chem. B* 123, 9839–9845. doi: 10.1021/acs.jpcb.9b07853

Aziz, Z. A. B. A., Ahmad, A., Mohd-Setapar, S. H., Hassan, H., Lokhat, D., Kamal, M. A., et al. (2017). Recent advances in drug delivery of polymeric nano-micelles. *Curr. Drug Metab.* 18, 16–29. doi: 10.2174/1389200217666160921143616

Badalkhani-Khamseh, F., Ebrahim-Habibi, A., and Hadipour, N. L. (2017). Atomistic computer simulations on multi-loaded PAMAM denrimers: a comparison of amine- and hydroxyl-terminated dendrimers. *J. Comput. Aid. Mol. Des.* 31, 1097–1111. doi: 10.1007/s10822-017-0091-9

Badalkhani-Khamseh, F., Ebrahim-Habibi, A., and Hadipour, N. L. (2019). Influence of dendrimer surface chemistry and pH on the binding and release pattern of chalcone studied by molecular dynamics simulations. *J. Mol. Recognit.* 32:e2757. doi: 10.1002/jmr.2757

Bai, X., Xu, M., Liu, S., and Hu, G. (2018). Computational investigations of the interaction between the cell membrane and nanoparticles coated with pulmonary surfactant. *ACS Appl. Mater. Interf.* 10, 20368–20376. doi: 10.1021/acsami.8b06764

Banerjee, A., Qi, J., Gogoi, R., Wong, J., and Mitragorti, S. (2016). Role of nanoparticle size, shape and surface chemistry in oral drug delivery. *J. Control. Release* 238, 176–185. doi: 10.1016/j.jconrel.2016.07.051

Bannan, C. C., Calabró, G., Kyu, D. Y., and Mobley, D. L. (2016). Calculating partition coefficients of small molecules in octanol/water and cyclohexane/water. *J. Chem. Theory Comput.* 12, 4015–4024. doi: 10.1021/acs.jctc.6b00449

Barnard, A. S. (2016). Challenges in modelling nanoparticles for drug delivery. *J. Phys. Condens. Matter* 28:23002. doi: 10.1088/0953-8984/28/2/023002

Barraza, L. F., Jiménez, V. A., and Alderete, J. B. (2015). Effect of PEGylation on the structure and drug loading capacity of PAMAM-G4 dendrimers: a molecular modelling approach on the complexation of 5-flourouracil with native and PEGylated PAMAM-G4. *Macromol. Chem. Phys.* 216, 1689–1701. doi: 10.1002/macp.201500179

Bazak, R., Houri, M., Achy, S. E., Kamel, S., and Refaat, T. (2015). Cancer active targeting by nanoparticles: a comprehensive review of literature. *J. Cancer res. Clin. Oncol.* 141, 769–784. doi: 10.1007/s00432-014-1767-3

Belfo, J. P., and Lemos, J. M. (eds). (2021). "Pharmacokinetic and pharmacodynamic models ," in *Optimal Impulsive Control for Cancer Therapy*, (Cham: Springer), 11–18. doi: 10.1007/978-3-030-50488-5_2

Beloqui, A., Solinís, M. Á, Rodríguez-Gascón, A., Almeida, A. J., and Préat, V. (2016). Nanostructured lipid carriers: promising drug delivery systems for future clinics. *Nanomed Nanotechnol.* 12, 143–161. doi: 10.1016/j.nano.2015.09.004

Belubbi, T., Shevade, S., Dhawan, V., Sridhar, V., Majumdar, A., Nunes, R., et al. (2018). Lipid architectonics for superior oral bioavailability of nelfinavir mesylate: comparative it in vitro and in vivo assessment. *AAPS Pharm. Sci. Tech.* 19, 3584–3598. doi: 10.1208/s12249-018-1156-3

Bengtsen, T., Holm, V. L., Kjølbye, L. R., Midtgaard, S. R., Johansen, N. T., Tesei, G., et al. (2020). Structure and dynamics of a nanodisc by integrating NMR, SAXS and SANS experiments with molecular dynamics simulations. *eLife* 9:e56518. doi: 10.7554/elife.56518

Bennun, S. V., Hoopes, M. I., Xing, C., and Faller, R. (2009). Coarse-grained modeling of lipids. *Chem. Phys. Lipids* 159, 59–66. doi: 10.1016/j.chemphyslip.2009.03.003

Benson, S. P., and Pleiss, J. (2014). Molecular dynamics simulations of self-emulsifying drug-delivery systems (SEDDS): influence of excipients on droplet nanostructure and drug localization. *Langmuir* 30, 8471–8480. doi: 10.1021/la501143z

Bergström, C. A. S., and Larsson, P. (2018). Computational prediction of drug solubility in water based systems: qualitative and quantitative approaches used in the current drug discovery and development settings. *Int. J. Pharm.* 540, 185–193. doi: 10.1016/j.ijpharm.2018.01.044

Bernkop-Schnürch, A., and Dünnhaupt, S. (2012). Chitosan-based drug delivery systems. *Eur. J. Pharm. Biopharm.* 81, 463–469. doi: 10.1016/j.ejpb.2012.04.007

Berrecoso, G., Crecente-Campo, J., and Alonso, M. J. (2020). Unveiling the pitfalls of the protein corona of polymeric drug nanocarriers. *Drug Deliv. Transl. Res.* 10, 730–750. doi: 10.1007/s13346-020-00745-0

Bhardwaj, P., Tripathi, P., Gupta, R., and Panday, S. (2020). Niosomes: a review on niosomal research in the last decade. *J. Drug Deliv. Sci. Tech.* 56:101581. doi: 10.1016/j.jddst.2020.101581

Binder, U., and Skerra, A. (2017). PASylation: a versatile technology to extend drug delivery. *Curr. Opin. Colloid Interf. Sci.* 31, 10–17. doi: 10.1016/j.cocis.2017.06.004

Blanco, E., Shen, H., and Ferrari, M. (2015). Principles of nanoparticle design for overcoming biological barriers to drug delivery. *Nat. Biotechnol.* 33, 941–951. doi: 10.1038/nbt.3330

Blazhynska, M. M., Kyrychenko, A., and Kalugin, O. N. (2018). Molecular dynamics simulation of the size-dependent morphological stability of cubic shape silver nanoparticles. *Mol. Simulat.* 44, 981–991. doi: 10.1080/08927022.2018.1469751

Bobo, D., Robinson, K. J., Islam, J., Thurecht, K. J., and Corrie, S. R. (2016). Nanoparticle-based medicines: a review of FDA-approved materials and clinical trials to date. *Pharm. Res.* 33, 2373–2387. doi: 10.1007/s11095-016-1958-5

Bohacek, R. S., McMartin, C., and Guida, W. C. (1996). The art and practice of structure-based drug design: a molecular modeling perspective. *Med. Res. Rev.* 16, 3–50. doi: 10.1002/(sici)1098-1128(199601)16:1<3::aid-med1>3.0.co;2-6

Bolintineanu, D. S., Lane, J. M. D., and Grest, G. S. (2014). Effects of functional groups and ionization on the structure of alkanethiol-coated gold nanoparticles. *Langmuir* 30, 11075–11085. doi: 10.1021/la502795z

Bolla, P. K., Meraz, C. A., Rodriguez, V. A., Deaguero, I., Singh, M., Yellepeddi, V. K., et al. (2019). Clotrimazole loaded ufosomes for topical delivery: formulation development and in-vitro studies. *Molecules* 24:3139. doi: 10.3390/molecules24173139

Botan, A., Favela-Rosales, F., Fuchs, P. F. J., Javanainen, M., Kanduè, M., Kulig, W., et al. (2015). Toward atomistic resolution structure of phosphatidylcholine headgroup and glycerol backbone at different ambient conditions. *J. Phys. Chem. B* 119, 15075–15088. doi: 10.1021/acs.jpcb.5b04878

Bouzo, B. L., Calvelo, M., Martín-Pastor, M., García-Fandiño, R., and De La Fuente, M. (2020). In vitro- in silico modeling approach to rationally designed simple and versatile drug delivery systems. *J. Phys. Chem. B* 124, 5788–5800. doi: 10.1021/acs.jpcb.0c02731

Bradley, A. J., Devine, D. V., Ansell, S. M., Janzen, J., and Brooks, D. E. (1998). Inhibition of liposome-induced complement activation by incorporated poly(ethylene glycol)-lipids. *Arch. Biochem. Biophys.* 357, 185–194. doi: 10.1006/abbi.1998.0798

Brancolini, G., and Tozzini, V. (2019). Multiscale modeling of proteins interaction with functionalized nanoparticles. *Curr. Opin. Colloid Interf. Sci.* 41, 66–73. doi: 10.1016/j.cocis.2018.12.001

Bricarello, D. A., Smilowitz, J. T., Zivkovic, A. M., German, J. B., and Parikh, A. N. (2011). Reconstituted lipoprotein: a versatile nanostructures. *ACS Nano* 5, 42–57. doi: 10.1021/nn103098m

Brooks, B. R., Brooks, C. L. III, Mackerell, A. D., Nilsson, L., Petrella, R. J., Roux, B., et al. (2009). CHARMM: the biomolecular simulation program B. *J. Comput. Chem.* 30, 1545–1614. doi: 10.1002/jcc

Bulbake, U., Doppalapudi, S., Kommineni, N., and Khan, W. (2017). Liposomal formulations in clinical use: an updated review. *Pharmaceutics* 9:12. doi: 10.3390/pharmaceutics9020012

Bunker, A. (2012). Poly(ethylene glycol) in drug delivery, why does it work, and can we do better? All atom molecular dynamics simulation provides some answers. *Phys. Proc.* 34, 24–33. doi: 10.1016/j.phpro.2012.05.004

Bunker, A. (2015). "Molecular modeling as a tool to understand the role of Poly(Ethylene) glycol in drug delivery," in *Computational pharmaceutics: Applications of Molecular modelling in Drug Delivery*, eds D. Ouyang and S. C. Smith (Hoboken, NJ: John Wiley & Sons), 217–234. doi: 10.1002/9781118573983.ch11

Bunker, A., Magarkar, A., and Viitala, T. (2016). Rational design of liposomal drug delivery systems, a review: combined experimental and computational studies of lipid membranes, liposomes and their PEGylation. *Biochim. Biophys. Acta Biomembr.* 1858, 2334–2352. doi: 10.1016/j.bbamem.2016.02.025

BurduŞel, A.-C., Gherasim, O., Grumezescu, A. M., Mogoantă, L., Ficai, A., and Andronescu, E. (2018). Biomedical applications of silver nanoparticles: an up-to-date overview. *Nanomaterials* 8:681. doi: 10.3390/nano8090681

Bussi, G., and Laio, A. (2020). Using metadynamics to explore complex free-energy landscapes. *Nat. Rev. Phys.* 2, 200–212. doi: 10.1038/s42254-020-0153-0

Buxton, G. A. (2014). Simulating the co-encapsulation of drugs in a "Smart" core-shell-shell polymer nanoparticle. *Eur. Phys. J. E* 37:14. doi: 10.1140/epje/i2014-14014-5

Cagel, M., Tesan, F. C., Bernabeu, E., Salgueiro, M. J., Zubillaga, M. B., Moretton, M. A., et al. (2017). Polymeric mixed micelles as nanomedicines: achievements and perspectives. *Eur. J. Pharm. Biopharm.* 113, 211–228. doi: 10.1016/j.ejpb.2016.12.019

Caracciolo, G. (2015). Liposome-protein corona in a physiological environment: challenges and opportunities for targeted delivery of nanomedicines. *Nanomed. Nanotechnol. Biol. Med.* 11, 543–557. doi: 10.1016/j.nano.2014.11.003

Casalini, T., Limongelli, V., Schmutz, M., Som, C., Jordan, O., Wick, P., et al. (2019). Molecular modeling for nanomaterial-biology interactions: opportunities, challenges, and perspectives. *Front. Bioeng. Biotech.* 7:268. doi: 10.3389/fbioe.2019.00268

Cascella, M., and Vanni, S. (2016). Toward accurate coarse-graining approaches for protein and membrane simulations. *Chem. Model.* 12, 1–52. doi: 10.1039/9781782622703-00001

Case, D. A., Cheatham, T. E. III, Darden, T., Gohlke, H., Luo, R., Merz, K. M., et al. (2005). The Amber biomolecular simulation programs. *J. Comput. Chem.* 26, 1668–1688. doi: 10.1002/jcc.20290

Cern, A., Barenholz, Y., Tropsha, A., and Goldblum, A. (2014). Computer-aided design of liposomal drugs: in silico prediction and experimental validation of drug candidates for liposomal remote loading. *J. Control. Release* 173, 125–131. doi: 10.1016/j.jconrel.2013.10.029

Chan, T., and Ouyang, D. (2018). Investigating the molecular dissolution process of binary solid dispersions by molecular dynamics simulations. *Asian J. Pharm. Sci.* 13, 248–254. doi: 10.1016/j.ajps.2017.07.011

Chang, C.-Y., Ju, S.-P., Wang, L.-F., Chen, C.-C., Chuang, Y.-C., Wu, H.-L., et al. (2017). Investigation of the self-assembly of CS and PCL copolymers with different molecular weights in water solution by coarse-grained molecular dynamics simulation. *J. Mol. Model.* 23:151. doi: 10.1007/s00894-017-3294-z

Charchar, P., Christofferson, A. J., Todorova, N., and Yarovsky, I. (2016). Understanding and designing the gold-bio interface: insights from simulations. *Small* 12, 2395–2418. doi: 10.1002/smll.201503585

Chen, G., Wang, Y., Xie, R., and Gong, S. (2018). A review on core-shell structured unimolecular nanoparticles for biomedical applications. *Adv. Drug Deliv. Rev.* 130, 58–72. doi: 10.1016/j.addr.2018.07.008

Chen, P., Zhang, Z., Gu, N., and Ji, M. (2018). Effect of the surface charge density of nanoparticles on their translocation across pulmonary surfactant monolayer: a molecular dynamics simulation. *Mol. Simulat.* 44, 85–93. doi: 10.1080/08927022.2017.1342118

Chen, J., Wang, J., Yin, B., Pang, L., Wang, W., and Zhu, W. (2019). Molecular mechanism of binding selectivity of inhibitors toward BACE1 and BACE2 revealed by multiple short molecular dynamics simulations and free-energy predictions. *ACS Chem. Neurosci.* 10, 4303–4318. doi: 10.1021/acschemneuro.9b00348

Chen, S., Hanning, S., Falconer, J., Locke, M., and Wen, J. (2019). Recent advances in non-ionic surfactant vesicles (niosomes): fabrication, characterization, pharmaceutical and cosmetic applications. *Eur. J. Pharm. Biopharm.* 144, 18–39. doi: 10.1016/j.ejpb.2019.08.015

Chen, J., Zhang, X., Millican, R., Creutzmann, J. E., Martin, S., and Jun, H.-W. (2020). High density lipoprotein mimicking nanoparticles for atherosclerosis. *Nano Converg.* 7:6. doi: 10.1186/s40580-019-0214-1

Chen, Q., Guan, G., Deng, F., Yang, D., Wu, P., Kang, S., et al. (2020). Anisotropic active ligandations in siRN-loaded hybrid nanodisks lead to distinct carcinostatic outcomes by regulating nano-bio interactions. *Biomaterials* 251:120008. doi: 10.1016/j.biomaterials.2020.120008

Chen, M., Pierstorff, E. D., Lam, R., Li, S.-Y., Huang, H., Osawa, E., et al. (2009). Nanodiamond-mediated delivery of water-insoluble therapeutics. *ACS Nano* 3, 2016–2022. doi: 10.1021/nn900480m

Chen, W., and Ouyang, D. (2017). Investigation of molecular dissolution mechanism of ketoprofen binary and ternary solid dispersions by molecular dynamics simulations. *Mol. Simulat.* 43, 13–16. doi: 10.1080/08927022.2017.1321755

Chow, A., Brown, B. D., and Merad, M. (2011). Studying the mononuclear phagocyte system in the molecular age. *Nat. Rev. Immunol.* 11, 788–798. doi: 10.1038/nri3087

Chuang, S. T., Cruz, S., and Narayanaswami, V. (2020). Reconfiguring nature's cholesterol accepting lipoproteins as nanoparticle platforms for transport and delivery of therapeutic and imaging agents. *Nanomaterials* 10:906. doi: 10.3390/nano10050906

Chun, B. J., Choi, J. I., and Jang, S. S. (2015). Molecular dynamics simulation of sodium dodecyl sulfate micelle: water penetration and sodium dodecyl sulfate dissociation. *Colloid Surf. A* 474, 36–43. doi: 10.1021/jp013616z

Chung, I.-M., Rajakumar, G., Venkidasamy, B., Subramanian, U., and Thiruvengadam, M. (2020). Exosomes: current use and future applications. *Clin. Chim. Acta* 500, 226–232. doi: 10.1016/j.cca.2019.10.022

Cohan, R. A., Madadkar-Sobhani, A., Khanahmad, H., Roohvand, F., Aghasadeghi, M. R., Hedayati, M. H., et al. (2011). Design, modeling, expression, and chemoselective PEGylation of a new nanosize cysteine analog of erythropoietin. *Int. J. Nanomed.* 6, 1217–1227. doi: 10.2147/IJN.S19081

Connolly, M. L. (1983). Solvent-accessible surfaces of proteins and nucleic acids. *Science* 221, 709–714. doi: 10.1126/science.6879170

Contreras, M. L., Torres, C., Villarroel, I., and Rozas, R. (2019). Molecular dynamics assessment of doxorubicin-carbon nanotubes molecular interactions for the design of drug delivery systems. *Struct. Chem.* 30, 369–384. doi: 10.1007/s11224-018-1210-5

Copland, M. J., Rades, T., Davies, N. M., and Baird, M. A. (2005). Lipid based particulate formulations for delivery of antigen. *Immunol. Cell Biol.* 83, 97–105. doi: 10.1111/j.1440-1711.2005.01315.x

Craig, W. A. (1998). Pharmacokinetics/pharmacodynamics parameters: rationale for antibacterial dosing of mice and men. *Clin. Infect. Dis.* 26, 1–12. doi: 10.1086/516284

Cramariuc, O., Rogl, T., and Vattulainen, I. (2012). Drug-lipid membrane interaction mechanisms revealed through molecular simulations. *Curr. Phys. Chem.* 2, 379–400. doi: 10.2174/1877946811202040379

Cramer, C. J. (2002). *Essentials of Computational Chemistry*. Hoboken, NJ: John Wiley & Sons Ltd.

Crommelin, D. J. A., van Hoogevest, P., and Storm, G. (2020). The role of liposomes in clinical nanomedicine development. What now? Now what? *J. Control. Release* 318, 256–263. doi: 10.1016/j.jconrel.2019.12.023

Csongradi, C., du Plessis, J., Aucamp, M. E., and Gerber, M. (2017). Topical delivery of roxithromycin solid-state forms entrapped in vesicles. *Eur. J. Pharm. Biopharm.* 114, 96–107. doi: 10.1016/j.ejpb.2017.01.006

da Silva Costa, A. C., Carvalho, S. C., de Farias Silva, N., Do Nascimento-Júnior, A. E. S., Cruz, J. N., and Maia de Jesus Chaves Neto, A. (2020). Effect of

chitosan/albendazole nanocarriers' solvation by molecular dynamics. *Theor. Chem. Acc.* 139:105. doi: 10.1007/s00214-020-02620-0

Damiati, S., Scheberl, A., Zayni, S., Damiati, S. A., Schuster, B., and Kompella, U. B. (2019). Albumin-bound nanodisks as delivery vehicle candidates: development and characterization. *Biophys. Chem.* 251:106178. doi: 10.1016/j.bpc.2019.106178

Danhier, F. (2016). To exploit the tumor microenvironment: since the EPR effect fails in the clinic, what is the future of nanomedicine? *J. Control. Release* 244, 108–121. doi: 10.1016/j.jconrel.2016.11.015

Das, M., Dahal, U., Mesele, O., Liang, D., and Cui, Q. (2019). Molecular dynamics simulation of interaction between functionalized nanoparticles with lipid membranes: analysis of coarse-grained models. *J. Phys. Chem. B* 123, 10547–10561. doi: 10.1021/acs.jpcb.9b08259

Das, T., Mehta, C. H., and Nayak, U. Y. (2020). Multiple approaches for achieving drug solubility: an it in silico perspective. *Drug Discov. Today* 25, 1206–1212. doi: 10.1016/j.drudis.2020.04.016

Dasari, S., and Mallik, B. S. (2020). Solubility and solvation free energy of a cardiovascular drug, LASSBio-294, in ionic liquids: a computational study. *J. Mol. Liq.* 301:112449. doi: 10.1016/j.molliq.2020.112449

Date, T., Nimbalkar, V., Kamat, J., Mittal, A., Mahato, R. I., and Chitkara, D. (2018). Lipid-polymer hybrid nanocarriers for delivering cancer therapeutics. *J. Control. Release* 271, 60–73. doi: 10.1016/j.jconrel.2017.12.016

De Nicola, A., Hezaveh, S., Zhao, Y., Kawakatsu, T., Roccatano, D., and Milano, G. (2014). Micellar drug nanocarriers and biomembranes: how do they interact? *Phys. Chem. Chem. Phys.* 16, 5093–5105. doi: 10.1039/c3cp54242d

Debnath, A., and Schäfer, L. V. (2015). Structure and dynamics of phospholipid nanodiscs from all-atom and coarse-grained simulations. *J. Phys. Chem. B* 119, 6991–7002. doi: 10.1021/acs.jpcb.5b02101

Decuzzi, P., Lee, S., Bhushan, B., and Ferrari, M. (2005). A theoretical model for the margination of particles within blood vessels. *Ann. Biomed. Eng.* 33, 179–190. doi: 10.1007/s10439-005-8976-5

Dehneshin, N., Raissi, H., Hasanzade, Z., and Farzad, F. (2019). Using molecular dynamics simulation to explore the binding of the three potent anticancer drugs sorafenib streptozotocin and sunitinib to functionalized carbon nanotubes. *J. Mol. Model.* 25:159. doi: 10.1007/s00894-019-4024-5

del Pino, P., Pelaz, B., Zhang, Q., Maffre, P., Nienhaus, G. U., and Parak, W. J. (2014). Protein corona formulation around nanoparticles - from the past to the future. *Mater. Horiz.* 1, 301–313. doi: 10.1039/C3MH00106G

Dell'Orco, D., Lundqvist, M., Oslakovic, C., Cedervall, T., and Linse, S. (2010). Modeling the time evolution of the nanoparticle-protein corona in a body fluid. *PLoS One* 5:e10949. doi: 10.1371/journal.pone.0010949

Denisov, I. G., Grinkova, Y. V., Lazarides, A. A., and Sligar, S. G. (2004). Directed self-assembly of monodisperse phospholipid bilayer nanodiscs with controlled size. *J. Am. Chem. Soc.* 126, 3477–3487. doi: 10.1021/ja0393574

Denisov, I. G., and Sligar, S. G. (2017). Nanodiscs in membrane biochemistry and biophysics. *Chem. Rev.* 117, 4669–4713. doi: 10.1021/acs.chemrev.6b00690

Dhawan, M., Price, D., Trump, D., Kanter, P., Shore, N., Needham, D., et al. (2004). New drug delivery approach for the treatment of prostate cancer (preclinical results and phase I study results). *J. Clin. Oncol.* 22:4661. doi: 10.1200/jco.2004.22.90140.4661

Dhawan, V., Magarkar, A., Joshi, G., Makhija, D., Jain, A., Shah, J., et al. (2016). Stearylated cycloargenine nanosystems for intracellular delivery - simulations, formulation and proof of concept. *RSC Adv.* 6, 113538–113550. doi: 10.1039/C6RA16432C

Ding, H. M., and Ma, Y. Q. (2014a). Computer simulation of the role of protein corona in cellular delivery of nanoparticles. *Biomaterials* 35, 8703–8710. doi: 10.1016/j.biomaterials.2014.06.033

Ding, H. M., and Ma, Y. Q. (2014b). Theoretical and computational investigations of nanoparticle-biomembrane interactions in cellular delivery. *Small* 11, 1055–1071. doi: 10.1002/smll.201401943

Ding, H., and Ma, Y. (2012). Role of physicochemical properties of coating ligands in receptor-mediated endocytosis of nanoparticles. *Biomaterials* 33, 5798–5802. doi: 10.1016/j.biomaterials.2012.04.055

Ding, H., and Ma, Y. (2016). Design strategy of surface decoration for efficient delivery of nanoparticles by computer simulation. *Sci. Rep.* 6:26783. doi: 10.1038/srep26783

Ding, H., and Ma, Y. (2018). Computational approaches to cell-nanomaterial interactions: keeping balance between therapeutic efficiency and cytotoxicity. *Nanoscale Horiz.* 3, 6–27. doi: 10.1039/C7NH00138J

Ding, H., Tian, W., and Ma, Y. (2012). Designing nanoparticle translocation through membranes by computer simulations. *ACS Nano* 6, 1230–1238. doi: 10.1021/nn2038862

Dowlatabadi, M., Jahangiri, M., and Farhadian, N. (2019). Prediction of chlortetracycline adsorption on the FE3O4 nanoparticle using molecular dynamics simulation. *J. Biomol. Struct. Dyn.* 37, 3616–3626. doi: 10.1080/07391102.2018.1521746

Dror, R. O., Pan, A. C., Arlow, D. H., Borhani, D. W., Maragakis, P., Shan, Y., et al. (2011). Pathway and mechanism of drug binding to G-protein-coupled receptors. *Proc. Natl. Acad. Sci. U.S.A.* 108, 13118–13123. doi: 10.1073/pnas.1104614108

Du, H., Chandaroy, P., and Hui, S. W. (1997). Grafted poly-(ethylene glycol) on lipid surfaces inhibits protein adsorption and cell adhesion. *Biochim. Biophys. Acta* 1326, 236–248. doi: 10.1016/S0005-2736(97)00027-8

Durbin, E. W., and Buxton, G. (2010). A Coarse-grained model of targeted drug delivery from responsive polymer nanoparticles. *Soft Matter* 6, 762–767. doi: 10.1039/B918476G

Duverger, E., and Picaud, F. (2020). Theoretical study of ciprofloxacin antibiotic trapping on graphene or boron nitride oxide nanoflakes. *J. Mol. Model.* 26:135. doi: 10.1007/s00894-020-04410-8

Dzieciuch, M., Rissanen, S., Szydłowska, N., Bunker, A., Kumorek, M., Jamróz, D., et al. (2015). Pegylated liposomes as carriers of hydrophobic porphyrins. *J. Phys. Chem. B* 119, 6646–6657. doi: 10.1021/acs.jpcb.5b01351

Dzieciuch-Rojek, M., Poojari, C., Bednar, J., Bunker, A., Kozik, B., Nowakowska, M., et al. (2017). Effects of membrane PEGylation on entry and location of antifungal drug itraconazole and their pharmacological implications. *Mol. Pharmacol.* 14, 1057–1070. doi: 10.1021/acs.molpharmaceut.6b00969

Eckhardt, S., Brunetto, P. S., Gagnon, J., Priebe, M., Giese, B., and Fromm, K. M. (2013). Nanobio silver: its interactions with peptides and bacteria, and its uses in medicine. *Chem. Rev.* 113, 4708–4754. doi: 10.1021/cr300288v

Ekladious, I., Colson, Y. L., and Grinstaff, M. W. (2019). Polymer-drug conjugate therapeutics: advances, insights and prospects. *Nat. Rev. Drug Discov.* 18, 273–294. doi: 10.1038/s41573-018-0005-0

El-Hammadi, M. M., and Arias, J. L. (2019). An update on liposomes in drug delivery: a patent review (2014 - 2018). *Expert. Opin. Ther. Pat.* 29, 891–907. doi: 10.1080/13543776.2019.1679767

Erimban, S., and Daschakraborty, S. (2020). Translocation of a hydroxyl functionalized carbon dot across a lipid bilayer: an all-atom molecular dynamics simulation study. *Phys. Chem. Chem. Phys.* 22, 6335–6350. doi: 10.1039/C9CP05999G

Erlebach, A., Muljajew, I., Chi, M., Bückmann, C., Weber, C., Schubert, U. S., et al. (2020). Predicting solubility of small molecules in macromolecular compounds for nanomedicine applications from atomistic simulations. *Adv. Theory Simul.* 3:2000001. doi: 10.1002/adts.202000001

Esalmi, M., Nikkhah, S. J., Hashemianzadeh, S. M., and Sajadi, S. A. S. (2016). The compatibility of tacrine molecule with poly(n-butylcyanoacrylate) and chitosan as efficient carriers for drug delivery: a molecular dynamics study. *Eur. J. Pharm. Sci.* 82, 79–85. doi: 10.1016/j.ejps.2015.11.014

Español, P., and Warren, P. B. (2017). Perspective: dissipative particle dynamics. *J. Chem. Phys* 146:150901. doi: 10.1063/1.4979514

Exner, K. S., and Ivanova, A. (2020). Identifying a gold nanoparticle as a proactive carrier for transport of a doxorubicin-peptide complex. *Colloid Surf. B* 194:111155. doi: 10.1016/j.colsurfb.2020.111155

Farmanzadeh, D., and Ghaderi, M. (2018). A computational study of PAMAM dendrimer interaction with trans isomer of picoplatin anticancer drug. *J. Mol. Graph. Model.* 80, 1–6. doi: 10.1016/j.jmgm.2017.12.010

Farzad, F., and Hashemzadeh, H. (2020). Probing the effect of polyethylene glycol on the adsorption mechanisms of Gem on the hexagonal boron nitride as a highly efficient polymerbased drug delivery system: DFT, classical MD and well-tempered metadynamics simulations. *J. Mol. Graph. Model.* 98:107613. doi: 10.1016/j.jmgm.2020.107613

Fatemi, S. M., Fatemi, S. J., and Abbasi, Z. (2020). PAMAM dendrimer-based macromolecules and their potential applications: recent advances in theoretical studies. *Polym. Bull.* 77, 6671–6691. doi: 10.1007/s00289-019-03076-4

Ferreira, L. G., dos Santos, R. N., Oliva, G., and Andricopulo, A. D. (2015). Molecular docking and structure-based drug design strategies. *Molecules* 20, 13384–13421. doi: 10.3390/molecules200713384

Filipe, H. A. L., Moreno, M. J., Ròg, T., Vattulainen, I., and Loura, L. M. S. (2014). How to tackle the issues in free energy simulations of long amphiphiles interacting with lipid membranes: convergence and local membrane deformations. *J. Phys. Chem. B* 118, 3572–3581. doi: 10.1021/jp501622d

Fox, L. J., Richardson, R. M., and Briscoe, W. H. (2018). PAMAM dendrimer - cell membrane interactions. *Adv. Colloid Interf. Sci.* 257, 1–18. doi: 10.1016/j.cis.2018.06.005

Frenkel, D., and Smit, B. (2001). *Understanding Molecular Simulation: From Algorithms to Applications*. Cambridge, MA: Academic Press.

Gao, H., Shi, W., and Freund, L. B. (2005). Mechanics of receptor-mediated endocytosis. *Proc. Natl. Acad. Sci. U.S.A.* 102, 9469–9474. doi: 10.1073/pnas.0503879102

Gao, J., Wang, P., Wang, Z., Li, C., Sun, S., and Hu, S. (2019). Self-assembly of DCPD-loaded cross-linked micelle from triblock copolymers and its pH-responsive behavior: a dissipative particle dynamics study. *Chem. Eng. Sci.* 195, 325–334. doi: 10.1016/j.ces.2018.09.028

Gao, Y., Shi, Y., Fu, M., Feng, Y., Lin, G., Kong, D., et al. (2020). Simulation study of the effects of interstitial fluid pressure and blood flow velocity on transvascular transport of nanoparticles in tumor microenvironment. *Comput. Meth. Prog. Biol.* 193:105493. doi: 10.1016/j.cmpb.2020.105493

Gapsys, V., and Groot, B. L. D. (2020). On the importance of statistics in molecular simulations for thermodynamics, kinetics and simulation box size. *eLife* 9:e57589. doi: 10.7554/eLife.57589

García, K. P., Zarschler, K., Barbaro, L., Barreto, J. A., O'Malley, W., Spiccia, L., et al. (2014). Zwitterionic-coated "stealth" nanoparticles for biomedical applications: recent advances in countering biomolecular corona formation and uptake by the mononuclear phagocyte system. *Small* 13, 2516–2529. doi: 10.1002/smll.201303540

Ge, D., Wu, D., Wang, Z., Shi, W., Wu, T., Zhang, A., et al. (2009). Cellular uptake mechanism of molecular umbrella. *Bioconjug. Chem.* 20, 2311–2316. doi: 10.1021/bc9003074

Ge, Z., and Wang, Y. (2017). Estimation of nanodiamond surface charge density from zeta potential and molecular dynamics simulations. *J. Phys. Chem. B* 121, 3394–3402. doi: 10.1021/acs.jpcb.6b08589

Gebauer, M., and Skerra, A. (2018). Prospects of PASylation for the design of protein and peptide therapeutics with extended half-life and enhanced action. *Bioorg. Med. Chem.* 26, 2882–2887. doi: 10.1016/j.bmc.2017.09.016

Geng, Y., Dalhaimer, P., Cai, S., Tsai, R., Tewari, M., Minko, T., et al. (2007). Shape effects of filaments versus spherical particles in flow and drug delivery. *Nat. Nanotechnol.* 2, 249–255. doi: 10.1038/nnano.2007.70

Gentile, F., Curcio, A., Indolfi, C., Ferrari, M., and Decuzzi, P. (2008). The margination propensity of spherical particles for vascular targeting in microcirculation. *J. Nanobiotechnol.* 6:9. doi: 10.1186/1477-3155-6-9

Ghadari, R., and Kashefi, A. (2017). A computational study of the usability of amino acid-functionalized nitrogen doped graphene oxides as temperature-responsive drug delivery systems. *Int. J. Hyperth.* 33, 785–795. doi: 10.1080/02656736.2017.1308020

Ghadari, R., and Mohammedzadeh, Y. (2018). MD simulation studies on the effect of the temperature and protonation state on the imide-linked amino acid-based dendrimers. *Comput. Mater. Sci.* 151, 124–131. doi: 10.1016/j.commatsci.2018.05.011

Ghadari, R., and Sabri, A. (2019). In silico study on core-shell pseudodendrimeric glycodise structures in drug delivery related usages. *Polyhedron* 160, 10–19. doi: 10.1016/j.poly.2018.12.013

Ghadri, Z., Raissi, H., Shahabi, M., and Farzad, F. (2020). Molecular dynamics simulation study of glycine tip-functionalisation of single-walled carbon nanotubes as emerging nanovectors for the delivery of anticancer drugs. *Mol. Simulat.* 46, 111–120. doi: 10.1080/08927022.2019.1679363

Ghasemi, F., Mehridehnavi, A., Pérez-Garrado, A., and Pérez-Sánchez, H. (2018). Neural network and deep-learning algorithms used in QSAR studies: merits and drawbacks. *Drug Discov. Today* 23, 1784–1790. doi: 10.1016/j.drudis.2018.06.016

Ghitman, J., Stan, R., Vlasceanu, G., Vasile, E., and Iovu, H. (2019). Predicting the drug loading efficiency into hybrid nanocarriers based on PLGA-vegetable oil using molecular dynamics simulation approach and Flory-Huggins theory. *J. Drug Deliv. Sci. Tech.* 53:101203. doi: 10.1016/j.jddst.2019.101203

Ghosal, K., and Ghosh, A. (2019). Carbon dots: the next generation platform for biomedical applications. *Mater. Sci. Eng. C* 96, 887–903. doi: 10.1016/j.msec.2018.11.060

Ghosh, M., Ren, G., Simonsen, J. B., and Ryan, R. O. (2014). Cationic lipid nanodisks as an siRNA delivery vehicle. *Biochem. Cell Biol.* 92, 200–205. doi: 10.1139/bcb-2014-0027

Ghosh, M., Singh, A. T. K., Xu, W., Sulchek, T., Gordon, L. I., and Ryan, R. O. (2011). Curcumin nanodisks: formulation and characterization. *Nanomed. Nanotechnol.* 7, 162–167. doi: 10.1016/j.nano.2010.08.002

Ghosh, P., Han, G., De, M., Kim, C. K., and Rotello, V. M. (2008). Gold nanparticles in delivery applications. *Adv. Drug Deliv. Rev.* 60, 1307–1315. doi: 10.1016/j.addr.2008.03.016

Giri, A. K., and Spohr, E. (2018). Influence of chain length and branching on the structure of functionalized gold nanoparticles. *J. Phys. Chem. C* 122, 26739–26747. doi: 10.1021/acs.jpcc.8b08590

Gkeka, P., Angelikopoulos, P., Sarkisov, L., and Cournia, Z. (2014). Membrane partitioning of anionic, ligand-coated nanoparticles is accompanied by ligand snorkeling, local disordering, and cholesterol depletion. *PLoS Comput. Biol.* 10:e1003917. doi: 10.1371/journal.pcbi.1003917

Godin, B., Sakamoto, J. H., Serda, R. E., Grattoni, A., Bouamrani, A., and Ferrari, M. (2010). Emerging applications of nanomedicine for the diagnosis and treatment of cardiovascular diseases. *Trends Pharmacol. Sci.* 31, 199–205. doi: 10.1016/j.tips.2010.01.003

Golda-Cepa, M., Riedlová, K., Kulig, W., Cwiklik, L., and Kotarba, A. (2020). Functionalization of the parylene c surface enhances the nucleation of calcium phosphate: combined experimental and molecular dynamics simulations approach. *ACS Appl. Mater. Interf.* 12, 12426–12435. doi: 10.1021/acsami.9b20877

Goldsmith, M., Abramovitz, L., and Peer, D. (2014). Precision nanomedicine in neurodegenerative diseases. *ACS Nano* 8, 1958–1965. doi: 10.1021/nn501292z

Gordillo-Galeano, A., and Mora-Huertas, C. E. (2018). Solid lipid nanoparticles and nanostructured lipid carriers: a review emphasizing on particle structure and drug release. *Eur. J. Pharm. Biopharm.* 133, 285–308. doi: 10.1016/j.ejpb.2018.10.017

Grillo, D. A., Albano, J. M. R., Mocskos, E. E., Facelli, J. C., Pickholz, M., and Ferraro, M. B. (2018). Mechanical properties of drug loaded diblock copolymer bilayers: a molecular dynamics study. *J. Chem. Phys.* 148:214901. doi: 10.1063/1.5028377

Groot, R. D., and Warren, P. B. (1997). Dissipative particle dynamics: bridging the gap between atomistic and mesoscopic simulation. *J. Chem. Phys.* 107, 4423–4435. doi: 10.1063/1.474784

Guo, X. D., Qian, Y., Zhang, C. Y., Nie, S. Y., and Zhang, L. J. (2012a). Can drug molecules diffuse into the core of micelles? *Soft Matter* 8, 9989–9995. doi: 10.1039/C2SM26200B

Guo, X. D., Zhang, L. J., and Qian, Y. (2012b). Systematic multiscale method for studying the structure-performance relationship of drug-delivery systems. *Ind. Eng. Chem. Res.* 51, 4719–4730. doi: 10.1021/ie2014668

Guo, X. D., Tan, J. P. K., Kim, S. H., Zhang, L. J., Zhang, Y., Hedrick, J. L., et al. (2009a). Computational studies on self-assembled paclitaxel structures: templates for heirarchical block copolymer assemblies and sustained drug release. *Biomaterials* 30, 6556–6563. doi: 10.1016/j.biomaterials.2009.08.022

Guo, X. D., Tan, J. P. K., Zhang, L. J., Khan, M., Liu, S. Q., Yang, Y. Y., et al. (2009b). Phase Behavior study of paclitaxel loaded amphiphilic copolymer in two solvents by dissipative particle dynamics simulations. *Chem. Phys. Lett.* 473, 336–342. doi: 10.1016/j.cplett.2009.04.009

Guo, X. D., Zhang, L. J., Wu, Z. M., and Qian, Y. (2010). Dissipative particle dynamics studies on microstructure of pH-sensitive micelles for sustained drug delivery. *Macromolecules* 43, 7839–7844. doi: 10.1021/ma101132n

Gupta, A. S. (2016). Role of particle size, shape, and stiffness in design of intravascular drug delivery systems: insights from computations, experiments, and nature. *WIREs Nanomed. Nanobiotechnol.* 8, 255–270. doi: 10.1002/wnan.1362

Gupta, J., Nunes, C., Vyas, S., and Jonnalagadda, S. (2011). Prediction of solubility parameters and miscibility of pharmaceutical compounds by molecular dynamics simulation. *J. Phys. Chem. B* 115, 2014–2023. doi: 10.1021/jp108540n

Gupta, M. N., and Roy, I. (2020). How corona formation impacts nanomaterials as drug carriers. *Mol. Pharm.* 17, 725–737. doi: 10.1021/acs.molpharmaceut.9b01111

Gupta, R., Badhe, Y., Mitragorti, S., and Rai, B. (2020). Permeation of nanoparticles across the intestinal lipid membrane: dependence on shape and surface chemistry studied through molecular simulations. *Nanoscale* 12, 6318–6333. doi: 10.1039/C9NR09947F

Gupta, R., Kashyap, N., and Rai, B. (2017). Transdermal cellular membrane penetration of proteins with gold nanoparticles: a molecular dynamics study. *Phys. Chem. Chem. Phys.* 19, 7537–7545. doi: 10.1039/C6CP08775B

Gupta, R., Kashyap, N., and Rai, B. (2018). Molecular mechanism of transdermal co-delivery of interferon-alpha protein with gold nanoparticle - a molecular dynamics study. *Mol. Simulat.* 44, 274–284. doi: 10.1080/08927022.2017.1367094

Gupta, R., and Rai, B. (2016). Penetration of gold nanoparticles through human skin: unraveling its mechanisms at the molecular scale. *J. Phys Chem. B* 120, 7133–7142. doi: 10.1021/acs.jpcb.6b03212

Gupta, R., and Rai, B. (2017). Effect of size and surface charge of gold nanoparticles on their skin permeability: a molecular dynamics study. *Sci. Rep.* 7:45292. doi: 10.1038/srep45292

Gupta, S., and Biswas, P. (2018). Effect of pH on size and internal structure of poly(propylene imine) dendrimers: a molecular dynamics simulation study. *J. Phys. Chem. B* 122, 9250–9263. doi: 10.1021/acs.jpcb.8b04653

Gupta, V., Bhavanasi, S., Quadir, M., Singh, K., Ghosh, G., Vasamreddy, K., et al. (2019). Protein PEGylation for cancer therapy: bench to bedside. *J. Cell Commun. Signal.* 13, 319–330. doi: 10.1007/s12079-018-0492-0

Habiba, K., Bracho-Rincon, D. P., Gonzalez-Feliciano, J. A., Villalobos-Santos, J. C., Makarov, V. I., Ortiz, D., et al. (2015). Synergistic antibacterial activity of PEGylated silver-graphene quantum dots nanocomposites. *Appl. Mater. Today* 1, 80–87. doi: 10.1016/j.apmt.2015.10.001

Haddish-Berhane, N., Rickus, J. L., and Haghighi, K. (2007). The role of multiscale computational approaches for rational design of conventional and nanoparticle oral drug delivery systems. *Int. J. Nanomed.* 2, 315–331.

Hadjidemetriou, M., and Kostarelos, K. (2017). Evolution of the nanoparticle corona. *Nat. Nanotechnol.* 12, 288–290. doi: 10.1038/nnano.2017.61

Hamed, E., Ma, D., and Keten, S. (2015). Multiple PEG chains attached onto the surface of a helix bundle: conformations and implications. *ACS Biomater. Sci. Eng.* 1, 79–84. doi: 10.1021/ab500088b

Han, R., Huang, T., Liu, X., Yin, X., Li, H., Lu, J., et al. (2019a). Insight into the dissolution molecular mechanism of ternary solid dispersions by combined experiments and molecular simulations. *AAPS Pharm. Sci. Tech.* 20:274. doi: 10.1208/s12249-019-1486-9

Han, R., Xiong, H., Ye, Z., Yang, Y., Huang, T., Jing, Q., et al. (2019b). Predicting physical stability of solid dispersions by machine learning techniques. *J. Control. Release* 311–312, 16–25. doi: 10.1016/j.jconrel.2019.08.030

Han, S. (2013). Molecular dynamics simulation of oleic acid/oleate bilayers: an atomistic model for ufasome membrane. *Chem. Phys. Lipids* 175–176, 79–83. doi: 10.1016/j.chemphyslip.2013.08.004

Hao, L., Lin, L., and Zhou, J. (2019). pH-responsive zwitterionic copolymer DHA-PBLG-PCB for targeted drug delivery: a computer simulation study. *Langmuir* 35, 1944–1953. doi: 10.1021/acs.langmuir.8b00626

Hasanzade, Z., and Raissi, H. (2017). Investigation of graphene-based nanomaterial as nanocarrier for adsorption of paclitaxel anticancer drug: a molecular dynamics simulation study. *J. Mol. Model.* 23:36. doi: 10.1007/s00894-017-3207-1

Hashemzadeh, H., and Raissi, H. (2017). The functionalization of carbon nanotubes to enhance the efficacy of the anticancer drug paclitaxel: a molecular dynamics simulation study. *J. Mol. Model.* 23:222. doi: 10.1007/s00894-017-3391

Hathout, R. M., and Metwally, A. A. (2016). Towards better modelling of drug loading in solid lipid nanoparticles: molecular dynamics, docking experiments and gaussian processes machine learning. *Eur. J. Pharm. Biopharm.* 108, 262–268. doi: 10.1016/j.ejpb.2016.07.019

Hathout, R. M., Metwally, A. A., Woodman, T. J., and Hardy, J. G. (2020). Prediction of drug loading in the gelatin matrix using computational methods. *ACS Omega* 5, 1549–1556. doi: 10.1021/acsomega.9b03487

He, X., Gu, Z., Wang, L., Qu, Z., and Xu, F. (2020). Coarse-grained molecular dynamics simulation of dendrimer transmembrane transport with temperature-dependent membrane phase states. *Int. J. Heat Mass Transf.* 155:119797. doi: 10.1016/j.ijheatmasstransfer.2020.119797

Hedayati, M. H., Norouzian, D., Aminian, M., Teimourian, S., Ahangari Cohan, R., Sardari, S., et al. (2017). Molecular design, expression and evaluation of PASylated human recombinant erythropoietin with enhanced functional properties. *Protein J.* 36, 36–48. doi: 10.1007/s10930-017-9699-9

Heikkilä, E., Gurtovenko, A. A., Martinez-Seara, H., Häkkinen, H., Vattulainen, I., and Akola, J. (2012). Atomistic simulations of functional Au 144(SR) 60 gold nanoparticles in aqueous environment. *J. Phys. Chem. C* 116, 9805–9811. doi: 10.1021/jp301094m

Heikkilä, E., Martinez-Seara, H., Gurtovenko, A. A., Javanainen, M., Häkkinen, H., Vattulainen, I., et al. (2014a). Cationic Au nanoparticle binding with plasma membrane-like lipid bilayers: potential mechanism for spontaneous permeation to cells revealed by atomistic simulations. *J. Phys. Chem. C* 118, 11131–11141. doi: 10.1021/jp5024026

Heikkilä, E., Martinez-Seara, H., Gurtovenko, A. A., Vattulainen, I., and Akola, J. (2014b). Atomistic simulations of anionic Au144(SR)60 nanoparticles interacting with asymmetric model lipid membranes. *Biochim. Biophys. Acta Biomembr.* 1838, 2852–2860. doi: 10.1016/j.bbamem.2014.07.027

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

Hlaváč, D., Klushina, D., and Tokarský, J. (2018). Interaction of antitumoral drug erlotinib with biodegradable triblock copolymers: a molecular modeling study. *Chem. Pap.* 72, 2023–2034. doi: 10.1007/s11696-018-0413-y

Hoffmann, T., and Gastreich, M. (2019). The next level in chemical space navigation: going far beyond enumerable compound libraries. *Drug Discov. Today* 24, 1148–1156. doi: 10.1016/j.drudis.2019.02.013

Hossain, S., Kabedev, A., Parrow, A., Bergström, C. A. S., and Larsson, P. (2019). Molecular simulation as a computational pharmaceutics tool to predict drug solubility, solubilization processes and partitioning. *Eur. J. Pharm. Biopharm.* 137, 46–55. doi: 10.1016/j.ejpb.2019.02.007

Hou, Y., and Lu, H. (2019). Protein PEPylation: a new paradigm of protein-polymer conjugation. *Bioconjug. Chem.* 30, 1604–1616. doi: 10.1021/acs.bioconjchem.9b00236

Hu, W., Mao, A., Wong, P., Larsen, A., Yazaki, P. J., Wong, J. Y. C., et al. (2017). Characterization of 1,2-Distearol-sn-glycero-3-phosphoethanolamine-N-[methoxy(polyethylene glycerol)-2000] and its complex with doxorubicin using nuclear magnetic resonance spectroscopy and molecular dynamics. *Bioconjug. Chem.* 28, 1777–1790. doi: 10.1021/acs.bioconjchem.7b00238

Hu, Y. B., Dammer, E. B., Ren, R. J., and Wang, G. (2015). The endosomal-lysosomal system: from acidification and cargo sorting to neurodegeneration. *Transl. Neurodegener.* 4, 1–10. doi: 10.1186/s40035-015-0041-1

Huang, H., Cruz, W., Chen, J., and Zheng, G. (2015). Learning from biology: synthetic lipoproteins for drug delivery. *Interdiscip. Rev. Nanomed. Nanobiotechnol.* 7, 298–314. doi: 10.1002/wnan.1308

Huang, T., Zhao, Q., Su, Y., and Ouyang, D. (2019). Investigation of molecular aggregation mechanism of glipizide/cyclodextrin complexation by combined experimental and molecular modeling approaches. *Asian J. Pharm. Sci.* 14, 609–620. doi: 10.1016/j.ajps.2018.10.008

Hurst, D. P., Grossfield, A., Lynch, D. L., Feller, S., Romo, T. D., Gawrisch, K., et al. (2010). A lipid pathway for ligand binding is necessary for a cannabinoid G protein-coupled receptor. *J. Biol. Chem.* 285, 17954–17964. doi: 10.1074/jbc.M109.041590

Hutchinson, S. T., and Kobayashi, R. (2019). Solvent-specific featurization for predicting free energies of solvation through machine learning. *J. Chem. Inf. Model.* 59, 1338–1346. doi: 10.1021/acs.jcim.8b00901

Huynh, L., Neale, C., Pomès, R., and Allen, C. (2012). Computational approaches to the rational design of nanoemulsions, polymeric micelles, and dendrimers for drug delivery. *Nanomed. Nanotechnol.* 8, 20–36. doi: 10.1016/j.nano.2011.05.006

Inglut, C. T., Sorrin, A. J., Kuruppu, T., Vig, S., Cicalo, J., Ahmad, H., et al. (2020). Immunological and toxicological considerations for the design of liposomes. *Nanomaterials* 10:190. doi: 10.3390/nano10020190

Ingólfsson, H. I., Lopez, C. A., Uusitalo, J. J., de Jong, D. H., Gopal, S. M., Periole, X., et al. (2014). The power of coarse graining in biomolecular simulations. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 4, 225–248. doi: 10.1002/wcms.1169

Israelachvili, J. (1997). The different faces of poly(ethylene glycol). *Proc. Natl. Acad. Sci. U.S.A.* 94, 8378–8379. doi: 10.1073/pnas.94.16.8378

Israelachvili, J. N. (1985). *Intermolecular and Surface Forces*. Cambridge, MA: Academic Press.

Izadyar, A., Farhadian, N., and Chenarani, N. (2016). Molecular dynamics simulation of doxorubicin adsorption on a bundle of functionalized CNT. *J. Biomol. Struct. Dyn.* 34, 1797–1805. doi: 10.1080/07391102.2015.1092475

Izrailev, S., Stepaniants, S., Balsera, M., Oono, Y., and Schulten, K. (1997). Molecular dynamics study of unbinding of the avidin-biotin complex. *Biophys. J.* 72, 1568–1581. doi: 10.1016/S0006-3495(97)78804-0

Jain, V., Maingi, V., Maiti, P. K., and Bharatam, P. V. (2013). Molecular dynamics simulations of PPI dendrimer-drug complexes. *Soft Matter* 9, 6482–6496. doi: 10.1039/C3SM50434D

Jain, V., Maiti, P. K., and Bharatam, P. V. (2016). Atomic level insights into realistic molecular models of dendrimer-drug complexes through MD simulations. *J. Chem. Phys.* 145:124902. doi: 10.1063/1.4962582

Janib, S. M., Moses, A. S., and MacKay, J. A. (2010). Imaging and drug delivery using theranostic nanoparticles. *Adv. Drug Deliv. Rev.* 62, 1052–1063. doi: 10.1016/j.addr.2010.08.004

Janout, V., Cline, L. L., Feuston, B. P., Klein, L., O'Brien, A., Tucker, T., et al. (2014). Molecular umbrella conjugate for the ocular delivery of siRNA. *Bioconjug. Chem.* 25, 197–201. doi: 10.1021/bc400506m

Janout, V., Jing, B., and Regen, S. L. (2002). Molecular umbrella-assisted transport of thiolated AMP and ATP across phospholipid bilayers. *Bioconjug. Chem.* 13, 351–356. doi: 10.1021/bc015564m

Janout, V., Jing, B., and Regen, S. L. (2005). Molecular umbrella-assisted transport of an oligonucleotide across cholesterol-rich phospholipid bilayers. *J. Am. Chem. Soc.* 127, 15862–15870. doi: 10.1021/ja053930x

Janout, V., and Regen, S. L. (2005). A needle-and-thread approach to bilayer transport: permeation of a molecular umbrella-oligonucliotide conjugation across a phospholipid membrane. *J. Am. Chem. Soc.* 127, 22–23. doi: 10.1021/ja044257z

Janout, V., and Regen, S. L. (2009). Bioconjugate-based molecular umbrellas. *Bioconjug. Chem.* 20, 183–192. doi: 10.1021/bc800296g

Janout, V., Zhang, L., Staina, I. V., Di Georgio, C., and Regen, S. L. (2001). Molecular umbrella-assisted transport of glutathione across a phospholipid membrane. *J. Am. Chem. Soc.* 123, 5401–5406. doi: 10.1021/ja010124r

Jevševar, S., Kunstelj, M., and Porekar, V. G. (2010). PEGylation of therapeutic proteins. *Biotechnol. J.* 5, 113–128. doi: 10.1002/biot.200900218

Jha, P. K., and Larson, R. G. (2014). Assessing the efficiency of polymeric excipients by atomistic molecular dynamics simulations. *Mol. Pharm.* 11, 1676–1686. doi: 10.1021/mp500068w

Jiang, W., Luo, J., and Nangia, S. (2015). Multiscale approach to investigate self-assembly of telodendrimer based nanocarriers for anticancer drug delivery. *Langmuir* 31, 4270–4280. doi: 10.1021/la503949b

Jing, B., Janout, V., and Regen, S. L. (2003). Fully detachable molecular umbrellas as peptide delivery agents. *Bioconjug. Chem.* 14, 1191–1196. doi: 10.1021/bc034074m

Johnson, S. R., and Zheng, W. (2006). Recent progress in the computational prediction of aqueous solubility and absorption. *AAPS J.* 8, E27–E40. doi: 10.1208/aapsj080104

Johnston, M. A., Swope, W. C., Jordan, K. E., Warren, P. B., Noro, M. G., Bray, D. J., et al. (2016). Toward a standard protocol for micelle simulation. *J. Phys Chem. B* 120, 6337–6351. doi: 10.1021/acs.jpcb.6b03075

Kacar, G. (2019). Molecular understanding of interactions, structure, and drug encapsulation efficiency of pluronic micelles from dissipative particle dynamics simulations. *Colloid Polym. Sci.* 297, 1037–1051. doi: 10.1007/s00396-019-04535-0

Kacar, G. (2020). Thermodynamic stability of ibuprofen loaded poloxamer micelles. *Chem. Phys.* 533:110713. doi: 10.1016/j.chemphys.2020.110713

Kamel, M., Raissi, H., and Morsali, A. (2017). Theoretical study of solvent and co-solvent effects on the interaction of flutamide anticancer drug with carbon nanotube as a drug delivery system. *J. Mol. Liq.* 248, 490–500. doi: 10.1016/j.molliq.2017.10.078

Kanchi, S., Gosika, M., Ayappa, K. G., and Maiti, P. K. (2018). Dendrimer interactions with lipid bilayer: comparison of force field and effect of implicit vs explicit solvation. *J. Chem. Theory Comput.* 14, 3825–3839. doi: 10.1021/acs.jctc.8b00119

Kapoor, B., Gupta, R., Gulati, M., Singh, S. K., Khursheed, R., and Gupta, M. (2019). The why, where, who, how and what of the vesicular delivery systems. *Adv. Colloid Interf.* 271:101985. doi: 10.1016/j.cis.2019.07.006

Karjiban, R. A., Basri, M., Basyaruddin, M., Rahman, A., and Salleh, A. B. (2012). Molecular dynamics simulation of palmitate ester self-assembly with diclofenac. *Int. J. Mol. Sci.* 13, 9572–9583. doi: 10.3390/ijms13089572

Karnati, K. R., and Wang, Y. (2018). Understanding the co-loading and releasing of doxorubicin and paclitaxel using chitosan functionalized single walled carbon nanotubes by molecular dynamics simulations. *Phys. Chem. Chem. Phys.* 20, 9389–9400. doi: 10.1039/C8CP00124C

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

Kasomova, A. O., Pavan, G. M., Danani, A., Mondon, K., Cristiani, A., Scapozza, L., et al. (2012). Validation of a novel molecular dynamics simulation approach for lipophilic drug incorporation into polymer micelles. *J. Phys. Chem. B* 116, 4338–4345. doi: 10.1021/jp2104819

Kaszuba, K., Roìg, T., Bryl, K., Vattulainen, I., and Karttunen, M. (2010). Molecular dynamics simulations reveal fundamental role of water as factor determining affinity of binding of β-blocker nebivolol to β-adrenergic receptor. *J. Phys. Chem. B* 114, 8374–8386. doi: 10.1021/jp909971f

Kaszuba, K., Roìg, T., Danne, R., Canning, P., Fülöp, V., Juhász, T., et al. (2012). Molecular dynamics, crystallography and mutagenesis studies on the substrate gating mechanism of prolyl oligopeptidase. *Biochimie* 94, 1398–1411. doi: 10.1016/j.biochi.2012.03.012

Katiyar, R. S., and Jha, P. K. (2018). Molecular simulations in drug delivery: opportunities and challenges. *WIREs Comput. Mol. Sci.* 8:e1358. doi: 10.1002/wcms.1358

Katre, N. V. (1993). The conjugation of proteins with polyethylene glycol and other polymers altering properties of proteins to enhance their therapeutic potential. *Adv. Drug Deliv. Rev.* 10, 91–114. doi: 10.1016/0169-409X(93)90005-O

Kaupbayeva, B., and Russell, A. J. (2020). Polymer-enhanced biomacromolecules. *Prog. Polym. Sci.* 101:101194. doi: 10.1016/j.progpolymsci.2019.101194

Kavyani, S., Amjad-Iranagh, S., Dadvar, M., and Modarress, H. (2016). Hybrid dendrimers of PPI(core)-PAMAM(shell): a molecular dynamics study. *J. Phys. Chem. B* 120, 9564–9575. doi: 10.1021/acs.jpcb.6b05142

Kavyani, S., Dadvar, M., Modarress, H., and Amjad-Iranagh, S. (2018a). A coarse grained molecular dynamics simulation study on the structural properties of carbon nanotube - dendrimer composites. *Soft Matter* 14, 3151–3163. doi: 10.1039/C8SM00253C

Kavyani, S., Dadvar, M., Modarress, H., and Amjad-Iranagh, S. (2018b). Molecular perspective mechanism for drug loading on carbon nanotube-dendrimer: a coarse-grained molecular dynamics study. *J. Phys. Chem. B* 122, 7956–7969. doi: 10.1021/acs.jpcb.8b04434

Kelly, C. V., Leroueil, P. R., Orr, B. G., Holl, M. M. B., and Andricioaei, L. (2008). Poly(amidoamine) dendrimers on lipid bilayers II: effects of bilayer phase and dendrimer termination. *J. Phys. Chem. B* 112, 9346–9353. doi: 10.1021/jp8013783

Kelly, C. V., Liroff, M. G., Triplett, L. D., Leroueil, P. R., Mullen, D. G., Wallace, J. M., et al. (2009). Stoichiometry and structure of poly(amidoamine) dendrimer- Lipid complexes. *ACS Nano* 3, 1886–1896. doi: 10.1021/nn900173e

Kepczynski, M., and Róg, T. (2016). Functionalized lipids and surfactants for specific applications. *Biochim. Biophys. Acta Biomembr.* 1858, 2362–2379. doi: 10.1016/j.bbamem.2016.02.038

Khalkhali, M., Mohammadinejad, S., Khoeini, F., and Rostamizadeh, K. (2019). Vesicle-like structure of lipid based nanoparticles as drug delivery system revealed by molecular dynamics simulation. *Int. J. Pharm.* 559, 173–181. doi: 10.1016/j.ijpharm.2019.01.036

Khan, R., and Irchhaiya, R. (2016). Niosomes: a potential tool for novel drug delivery. *J. Pharm. Investig.* 46, 165–204. doi: 10.1007/s40005-016-0249-9

Khan, S., McCabe, J., Hill, K., and Beales, P. A. (2020). Biodegradable hybrid block copolymer - lipid vesicles as potential drug delivery systems. *J. Colloid Interf. Sci.* 562, 418–428. doi: 10.1016/j.jcis.2019.11.101

Kharazian, B., Hadipour, N. L., and Ejtehadi, M. R. (2016). Understanding the nanoparticle - protein corona complexes using computational and experimental methods. *Int. J. Biochem. Cell Biol.* 75, 162–174. doi: 10.1016/j.biocel.2016.02.008

Kłos, J. S., and Paturej, J. (2020). Charged dendrimers with finite-size counterions. *J. Phys. Chem. B.* 124, 7957–7968. doi: 10.1021/acs.jpcb.0c05092

Kłos, J. S., and Sommer, J.-U. (2013). Coarse grained simulations of neutral and charged dendrimers. *Polym. Sci. Ser. C* 55, 125–153. doi: 10.1134/S1811238213070023

Knop, K., Hoogenboom, R., Fischer, D., and Schubert, U. (2010). Poly(ethylene glycol) in drug delivery: pros and cons as well as potential alternatives. *Angew. Chem. Int. Edn.* 49, 6288–6308. doi: 10.1002/anie.200902672

Kohli, A. G., Kierstead, P. H., Venditto, V. J., Walsh, C. L., and Szoka, F. C. (2014). Designer lipids for drug delivery: from heads to tails. *J. Control. Release* 190, 274–287. doi: 10.1016/j.jconrel.2014.04.047

Koivuniemi, A., and Vattulainen, I. (2012). Revealing structural and dynamical properties of high density lipoproteins through molecular simulations. *Soft Matter* 8, 1262–1267. doi: 10.1039/C1SM06742G

Kojima, C., Kono, K., Maruyama, K., and Takagishi, T. (2000). Synthesis of polyamidoamine dendrimers having poly(ethylene glycol) grafts and their ability to encapsulate anticancer drugs. *Bioconjug. Chem.* 11, 910–917. doi: 10.1021/bc0000583

Kokh, D. B., Amaral, M., Bomke, J., Grädler, U., Musil, D., Buchstaller, H. P., et al. (2018). Estimation of drug-target residence times by τ-random acceleration molecular dynamics simulations. *J. Chem. Theory Comput.* 14, 3859–3869. doi: 10.1021/acs.jctc.8b00230

Koochaki, A., Moghbeli, M. R., Nikkhah, S. J., Ianiro, A., and Tuinier, R. (2020). Dual responsive PMEEECL-PAE block copolymers: a computational self assembly and doxorubicin uptake study. *RSC. Adv.* 10, 3233–3245. doi: 10.1039/C9RA09066E

Kordzadeh, A., Amjad-Iranagh, S., Zarif, M., and Modarress, H. (2019). Adsorption and encapsulation of the drug doxorubicin on covalent functionalized carbon nanotubes: a scrutinized study by using molecular dynamics simulation and quantum mechanics calculation. *J. Mol. Graph. Model.* 88, 11–22. doi: 10.1016/j.jmgm.2018.12.009

Kuai, R., Li, D., Chen, Y. E., Moon, J. J., and Schwendeman, A. (2016a). High-density lipoproteins: nature's multifunctional nanoparticles. *ACS Nano* 10, 3015–3041. doi: 10.1021/acsnano.5b07522

Kuai, R., Ochyl, L. J., Bahjat, K. S., Schwendeman, A., and Moon, J. J. (2016b). Designer vaccine nanodiscs for personalized cancer immunotherapy. *Nat. Mater.* 16, 489–496. doi: 10.1038/nmat4822

Kumar, G., Kant, A., Kumar, M., and Masram, D. T. (2019). Synthesis, characterizations and kinetic study of metal organic framework nanocomposite excipient used as extended release delivery vehicle for an antibiotic drug. *Inorgan. Chim. Acta* 496:119036. doi: 10.1016/j.ica.2019.119036

Kuramochi, H., Andoh, Y., Yoshii, N., and Okazaki, S. (2009). All-atom molecular dynamics study of a spherical micelle composed of N-acetylated poly(ethylene glycol)-poly(γ-benzyl L-glutamate) block copolymers: a potential carrier of drug delivery systems for cancer. *J. Phys. Chem. B* 113, 15181–15188. doi: 10.1021/jp906155z

Kyrychenko, A., Karpushina, G. V., Bogatyrenko, S. L., Kryshtal, A. P., and Doroshenko, A. O. (2011). Preparation, structure, and a coarse-grained molecular dynamics model for dodecanethiol-stabilized gold nanoparticles. *Comput. Theor. Chem.* 977, 34–39. doi: 10.1016/j.comptc.2011.09.003

Kyrychenko, A., Korsun, O. M., Gubin, I. I., Kovalenko, S. M., and Kagulin, O. N. (2015). Atomistic simulations of coating of silver nanoparticles with poly(vinylpyrrolidone) oligomers: effect of oligomer chain length. *J. Phys. Chem. C* 119, 7888–7899. doi: 10.1021/jp510369a

Lajunen, T., Kontturi, L. S., Viitala, L., Manna, M., Cramariuc, O., and Roìg, T. (2016). Indocyanine green-loaded liposomes for light-triggered drug release. *Mol. Pharm.* 13, 2095–2107. doi: 10.1021/acs.molpharmaceut.6b00207

Lajunen, T., Nurmi, R., Wilbie, D., Ruoslahti, T., Johansson, N. G., Korhonen, O., et al. (2018). The effect of light sensitizer localization on the stability of indocyanine green liposomes. *J. Control Release* 284, 213–223. doi: 10.1016/j.jconrel.2018.06.029

Lammers, T., and Ferrari, M. (2020). The success of nanomedicine. *Nano Today* 31:100853. doi: 10.1016/j.nantod.2020.100853

Larsen, A. S., Olsen, M. A., Moustafa, H., Larsen, F. H., Sauer, S. P. A., Rantanen, J., et al. (2019). Determining short-lived solid forms during phase transformations using molecular dynamics. *Cryst. Eng. Commun.* 21, 4020–4024. doi: 10.1039/C9CE00460B

Larsen, A. S., Rantanen, J., and Johansson, K. E. (2017a). Computational dehydration of crystalline hydrates using molecular dynamics simulations. *J. Pharm. Sci.* 106, 348–355. doi: 10.1016/j.xphs.2016.10.005

Larsen, A. S., Ruggiero, M. T., Johansson, K. E., Zeitler, J. A., and Rantanen, J. (2017b). Tracking dehydration mechanisms in crystalline hydrates with molecular dynamics simulations. *Cryst. Growth Des.* 17, 5017–5022. doi: 10.1021/acs.cgd.7b00889

Laudadido, E., Mobbili, G., Minelli, C., Massaccesi, L., and Galeazzi, R. (2017). Salts influence cathechins and flavonoids encapsulation in liposomes: a molecular dynamics investigation. *Mol. Inform.* 36:1700059. doi: 10.1002/minf.201700059

Lawrence, P. B., Gavrilov, Y., Matthews, S. S., Langlois, M. I., Shental-Bechor, D., Greenblatt, H. M., et al. (2014). Criteria for selecting PEGylation sites of proteins for higher thermodynamic and proteolytic stability. *J. Am. Chem. Soc.* 136, 17547–17560. doi: 10.1021/ja5095183

Lawrence, P. B., and Price, J. L. (2016). How PEGylation influences protein conformational stability. *Curr. Opin. Chem. Biol.* 34, 88–94. doi: 10.1016/j.cbpa.2016.08.006

Lee, E. H., Hsin, J., Sotomayor, M., Comellas, G., and Schulten, K. (2009). Discovery through the computational microscope. *Structure* 17, 1295–1306. doi: 10.1016/j.str.2009.09.001

Lee, S.-Y., Ferrari, M., and Decuzzi, P. (2009). Shaping nano-/micro-particles for enhanced vascular interaction in laminar flows. *Nanotechnology* 20:495101. doi: 10.1088/0957-4484/20/49/495101

Lee, H. (2014). Molecular modeling of PEGylated peptides, dendrimers, and single-walled carbon nanotubes for biomedical applications. *Polymers* 6, 776–798. doi: 10.3390/polym6030776

Lee, H. (2020a). Effects of nanoparticle electrostatics and protein - protein interactions on corona formation: conformation and hydrodynamics. *Small* 16:1906598. doi: 10.1002/smll.201906598

Lee, H. (2020b). Molecular simulations of PEGylated biomolecules, liposomes, and nanoparticles for drug delivery applications. *Pharmaceutics* 12:533. doi: 10.3390/pharmaceutics12060533

Lee, H., Choi, J. S., and Larson, R. G. (2011). Molecular dynamics studies of the size and internal structure of the PAMAM dendrimer grafted with arginine and histidine. *Macromolecules* 44, 8681–8686. doi: 10.1021/ma2019396

Lee, H., and Larson, R. G. (2008). Coarse-grained molecular dynamics simulation of the concentration and size dependence of fifth- and seventh-generation PAMAM dendrimers on pore formation in DMPC bilayer. *J. Phys. Chem. B* 112, 7778–7784. doi: 10.1021/jp802606y

Lee, H., and Larson, R. G. (2009). Molecular dynamics study of the structure and interparticle interactions of polyethylene glycol-conjugated PAMAM dendrimers. *J. Phys. Chem. B* 113, 13202–13207. doi: 10.1021/jp906497e

Lee, H., and Larson, R. G. (2011). Effects of PEGylation on the size and internal structure of dendrimers: self-penetration of long PEG chains into the dendrimer core. *Macromolecules* 44, 2291–2298. doi: 10.1021/ma102482u

Lee, I., Athey, B. D., Wetzel, A. W., Meixner, W., and Baker, J. R. (2002). Structural molecular dynamics studies on polyamidoamine dendrimers for a therapeutic application: effects of pH and generation. *Macromolecules* 35, 4510–4520. doi: 10.1021/ma010354q

Lehtinen, J., Magarkar, A., Stepniewski, M., Hakola, S., Bergman, M., Roìg, T., et al. (2012). Analysis of cause of failure of new targeting peptide in PEGylated liposome: molecular modeling as rational design tool for nanomedicine. *Eur. J. Pharm. Sci.* 46, 121–130. doi: 10.1016/j.ejps.2012.02.009

Lehto, M., Karilainen, T., Roìg, T., Cramariuc, O., Vanhala, E., Tornaeus, J., et al. (2014). Co-exposure with fullerene may strengthen health effects of organic industrial chemicals. *PLoS One* 9:e114490. doi: 10.1371/journal.pone.0114490

Li, J., Ying, S., Ren, H., Dai, J., Zhang, L., Liang, L., et al. (2020). Molecular dynamics study of the encapsulation and release of anticancer drug doxorubicin by chitosan. *Int. J. Pharm.* 580:119241. doi: 10.1016/j.ijpharm.2020.119241

Li, Z., Wang, Y., Zhu, J., Zhang, Y., Zhang, W., Zhou, M., et al. (2020). Emerging well-tailored nanoparticulate delivery system based on in situ regulation of the protein corona. *J. Control Release* 320, 1–18. doi: 10.1016/j.jconrel.2020.01.007

Li, L., Totton, T., and Frenkel, D. (2018). Computational methodology for solubility prediction: application to sparingly soluble organic/inorganic materials. *J. Chem. Phys.* 149:54102. doi: 10.1063/1.5040366

Li, M., Du, C., Guo, N., Teng, Y., Meng, X., Sun, H., et al. (2019). Composition design and medical application of liposomes. *Eur. J. Med. Chem.* 164, 640–653. doi: 10.1016/j.ejmech.2019.01.007

Li, Y., and Hu, Y. (2014). Computational investigation of the influence of chain length on the shielding effect of PEGylated nanoparticles. *RSC Adv.* 4, 51022–51031. doi: 10.1039/c4ra11142g

Li, Y., Kröger, M., and Liu, W. K. (2014a). Endocytosis of PEGylated nanoparticles accompanied by structural and free energy changes of the grafted polyethylene glycol. *Biomaterials* 35, 8467–8478. doi: 10.1016/j.biomaterials.2014.06.032

Li, Y., Stroberg, W., Lee, T.-R., Kim, H. S., Man, H., Ho, D., et al. (2014b). Multiscale modeling and uncertainty quantification in nanoparticle-mediated drug/gene delivery. *Comput. Mech.* 53, 511–537. doi: 10.1007/s00466-013-0953-5

Li, Y.-C., Rissanen, S., Stepniewski, M., Cramariuc, O., Róg, T., Mirza, S., et al. (2012). Study of interaction between PEG carrier and three relevant drug molecules: piroxicam, paclitaxel, and hematoporphyrin. *J. Phys. Chem. B* 116, 7334–7341. doi: 10.1021/jp300301z

Li, Y., Yue, T., Yang, K., and Zhang, X. (2012). Molecular modeling of the relationship between nanoparticle shape anisotropy and endocytosis kinetics. *Biomaterials* 33, 4965–4973. doi: 10.1016/j.biomaterials.2012.03.044

Li, Z., Tozer, T., and Alisaraie, L. (2016a). Molecular dynamics studies for optimization of noncovalent loading of vinblastine on single-walled carbon nanotube. *J. Phys. Chem. C* 120, 4061–4070. doi: 10.1021/acs.jpcc.5b10646

Li, Z., Van Dyk, A. K., Fitzwater, S. J., Fichthorn, K. A., and Milner, S. T. (2016b). Atomistic molecular dynamics simulations of charged latex particle surfaces in aqueous solution. *Langmuir* 32, 428–441. doi: 10.1021/acs.langmuir.5b03942

Liang, L., Shen, J.-W., and Wang, Q. (2017). Molecular dynamic study on DNA nanotubes as drug delivery vehicle for anticancer drugs. *Colloid Surf. B* 153, 168–173. doi: 10.1016/j.colsurfb.2017.02.021

Liang, Q. (2013). Penetration of polymer-grafted nanoparticles through a lipid bilayer. *Soft Matter* 9, 5594–5601. doi: 10.1039/C3SM27254K

Liebner, R., Mathaes, R., Meyer, M., Hey, T., Winter, G., and Besheer, A. (2014). Protein HESylation for half-life extension: synthesis, characterization and pharmacokinetics of HESylated anakinra. *Eur. J. Pharm. Biopharm.* 87, 378–385. doi: 10.1016/j.ejpb.2014.03.010

Lin, J., Zhang, H., Chen, Z., and Zheng, Y. (2010). Penetration of lipid membranes by gold nanoparticles: insights into cellular uptake, cytotoxicity, and their relationship. *ACS Nano* 4, 5421–5429. doi: 10.1021/nn1010792

Lin, X., Li, Y., and Gu, N. (2010). Nanoparticle's size effect on its translocation across a lipid bilayer: a molecular dynamics simulation. *J. Comput. Theor. Nanosci.* 7, 269–276. doi: 10.1166/jctn.2010.1358

Lin, J., Zhang, H., Morovati, V., and Dargazany, R. (2017). PEGylation on mixed monolayer gold nanoparticles: effect of grafting density, chain length, and surface curvature. *J. Colloid Interf. Sci.* 504, 325–333. doi: 10.1016/j.jcis.2017.05.046

Lin, J.-Q., Zheng, Y.-G., Zhang, H.-W., and Chen, Z. (2011). A simulation study on nanoscale holes generated by gold nanoparticles on negative lipid bilayers. *Langmuir* 27, 8323–8332. doi: 10.1021/la201086u

Lin, P., and Colina, C. M. (2019). Molecular simulation of protein-polymer conjugates. *Curr. Opin. Chem. Eng.* 23, 44–50. doi: 10.1016/j.coche.2019.02.006

Lin, W., Xue, Z., Wen, L., Li, Y., Liang, Z., Xu, J., et al. (2019). Mesoscopic simulations of drug loaded diselenide crosslinked micelles: stability, drug loading and release properties. *Colloid Surf. B* 182:110313. doi: 10.1016/j.colsurfb.2019.06.043

Lin, W. J., Nie, S. Y., Chen, Q., Qian, Y., Wen, X. F., and Zhang, L. J. (2014). Structure-property relationship of pH-sensitive (PCL) (PDEA-b-PPEGMA) Micelles: experiment and DPD simulation. *AICHe J.* 60, 3634–3646. doi: 10.1002/aic.14562

Lin, X., and Gu, N. (2014). Surface properties of encapsulating hydrophobic nanoparticles regulate the main phase transition temperature of lipid bilayers: a simulation study. *Nano Res.* 7, 1195–1204. doi: 10.1007/s12274-014-0482-3

Lin, X., Lin, X., and Gu, N. (2020). Optimization of hydrophobic nanoparticles to better target lipid rafts with molecular dynamics simulations. *Nanoscale* 12, 4101–4109. doi: 10.1039/C9NR09226A

Lipinski, C. A. (2004). Lead- and drug-like compounds: the rule-of-five revolution. *Drug Discov. Today Technol.* 1, 337–341. doi: 10.1016/j.ddtec.2004.11.007

Lipinski, C. A., Lombardo, F., Dominy, B. W., and Feeney, P. J. (2001). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* 46, 3–26. doi: 10.1016/S0169-409X(96)00423-1

Liu, P., and Long, W. (2009). Current mathematical methods used in QSAR/QSPR studies. *Int. J. Mol. Sci.* 10, 1978–1998. doi: 10.3390/ijms10051978

Liu, Y. Y., Peng, B., Sohrabi, S., and Liu, Y. Y. (2016). The configuration of copolymer ligands on nanoparticles affects adhesion and uptake. *Langmuir* 32, 10136–10143. doi: 10.1021/acs.langmuir.6b02371

Lobatto, M. E., Fuster, V., Fayad, Z. A., and Mulder, W. J. M. (2011). Perspectives and opportunities for nanomedicine in the management of atherosclerosis. *Nat. Rev. Drug Discov.* 10, 835–852. doi: 10.1038/nrd3578

Lolicato, F., Joly, L., Martinez-Seara, H., Fragneto, G., Scoppola, E., Baldelli Bombelli, F., et al. (2019). The role of temperature and lipid charge on intake/uptake of cationic gold nanoparticles into lipid bilayers. *Small* 15:1805046. doi: 10.1002/smll.201805046

Lolicato, F., Juhola, H., Zak, A., Postila, P. A., Saukko, A., Rissanen, S., et al. (2020). Membrane-dependent binding and entry mechanism of dopamine into its receptor. *ACS Chem. Neurosci.* 11, 1914–1924. doi: 10.1021/acschemneuro.9b00656

Lopez, H., and Lobaskin, V. (2015). Coarse-grained model of adsorption of blood plasma proteins onto nanoparticles. *J. Chem. Phys.* 143:243138. doi: 10.1063/1.4936908

Lorson, T., Lübtow, M. M., Wegener, E., Haider, M. S., Borova, S., Nahm, D., et al. (2018). Poly(2-oxazoline)s based biomaterials: a comprehensive and critical update. *Biomaterials* 178, 204–280. doi: 10.1016/j.biomaterials.2018.05.022

Loverde, S. M. (2014). Computer simulation of polymer and biopolymer self-assembly for drug delivery. *Mol. Simulat.* 40, 794–801. doi: 10.1080/08927022.2014.898118

Loverde, S. M., Klein, M. L., and Discher, D. E. (2011). Nanoparticle shape improves delivery: rational coarse grain molecular dynamics (rCG-MD) of Taxol in worm-like PEG-PCL micelles. *Adv. Mater.* 24, 3823–3830. doi: 10.1002/adma.201103192

Lu, S., Bennett, W. F. D., Ding, Y., Zhang, L., Fan, H. Y., Zhao, D., et al. (2015). Design and characterization of a multi-functional pH-triggered peptide C8 for selective anticancer activity. *Adv. Healthc. Mater.* 4, 2709–2718. doi: 10.1002/adhm.201500636

Lüdemann, S. K., Lounnas, V., and Wade, R. C. (2000). How do substrates enter and products exit the buried active site of cytochrome P450cam? 1. Random expulsion molecular dynamics investigation of ligand access channels and mechanisms. *J. Mol. Biol.* 303, 797–811. doi: 10.1006/jmbi.2000.4154

Lüder, K., Lindfors, L., Westergren, J., Nordholm, S., and Kjellander, R. (2007). In silico prediction of drug solubility: 2. free energy of solvation in pure melts. *J. Phys. Chem. B* 111, 7303–7311. doi: 10.1021/jp0642239

Lüder, K., Lindfors, L., Westergren, J., Nordholm, S., Persson, R., and Pedersen, M. (2009). In silico prediction of drug solubility: 4. will simple potentials suffice? *J. Comput. Chem.* 30, 1859–1871. doi: 10.1002/jcc.21173

Lundsten, S., Hernández, V. A., Gedda, L., Sarén, T., Brown, C. J., Lane, D. P., et al. (2020). Tumor-targeted delivery of the p53-activating VIP116 with PEG-stabilized lipodisks. *Nanomaterials* 10:783. doi: 10.1002/mats.201900026

Lunnoo, T., Assawakhajornasak, J., and Puangmali, T. (2019). In silico study of gold nanoparticle uptake into a mammalian cell: interplay of size, shape, surface charge, and aggregation. *J. Phys. Chem. C* 123, 3801–3810. doi: 10.1021/acs.jpcc.8b07616

Lunnoo, T., Assawakhajornasak, J., Ruangchai, S., and Puangmali, T. (2020). Role of surface functionalization on cellular uptake of AuNPs characterized by computational microscopy. *J. Phys. Chem. B* 124, 1898–1908. doi: 10.1021/acs.jpcb.9b11600

Luo, S., Zhang, Y., Cao, J., He, B., and Li, S. (2016). Arginine modified polymeric micelles as a novel drug delivery system with enhanced endocytosis efficiency. *Colloid Surf. B* 148, 181–192. doi: 10.1016/j.colsurfb.2016.07.023

Luo, Z., Li, Y., Wang, B., and Jiang, J. (2016). pH-sensitive vesicles formed by amphiphilic grafted copolymers with tunable membrane permeability for drug loading/release: a multiscale simulation study. *Macromolecules* 49, 6084–6094. doi: 10.1021/acs.macromol.6b01211

Luo, X., Wang, S., Xu, S., and Lang, M. (2019). Relevance of the polymeric prodrug and its drug loading efficiency: comparison between computer simulation and experiment. *Macromol. Theory Simul.* 28:1900026.

Luo, Z., and Jiang, J. (2012). pH-sensitive drug loading/releasing in amphiphilic copolymer PAE-PEG: integrating molecular dynamics and dissipative particle dynamics simulations. *J. Control Release* 162, 185–193. doi: 10.1016/j.jconrel.2012.06.027

Ma, D., DeBenedictis, E. P., Lund, R., and Keten, S. (2016). Design of polymer conjugated 3-helix micelles as nanocarriers with tunable shapes. *Nanoscale* 8, 19334–19342. doi: 10.1039/C6NR07125B

Macháèková, M., Tokarskı, J., and Èapková, P. (2013). A simple molecular modeling method for the characterization of polymeric drug carriers. *Eur. J. Pharm. Sci.* 48, 316–322. doi: 10.1016/j.ejps.2012.11.010

Maeda, H., Nakamura, H., and Fang, J. (2013). The EPR effect for macromolecular drug delivery to solid tumors: improvement of tumor uptake, lowering of systemic toxicity and distinct tumor imaging in vivo. *Adv. Drug Del. Rev.* 65:71. doi: 10.1016/j.addr.2012.10.002

Magarkar, A., Karakas, E., Stepniewski, M., Roìg, T., and Bunker, A. (2012). Molecular dynamics simulation of PEGylated bilayer interacting with salt ions: a model of the liposome surface in the bloodstream. *J. Phys. Chem. B* 116, 4212–4219. doi: 10.1021/jp300184z

Magarkar, A., Parkkila, P., Viitala, T., Lajunen, T., Mobarak, E., Licari, G., et al. (2018). Membrane bound COMT isoform is an interfacial enzyme: general mechanism and new drug design paradigm. *Chem. Commun.* 54, 3440–3443. doi: 10.1039/c8cc00221e

Magarkar, A., Roìg, T., and Bunker, A. (2014). Molecular dynamics simulation of PEGylated membranes with cholesterol: building toward the DOXIL formulation. *J. Phys. Chem. C* 118, 15541–15549. doi: 10.1021/jp504962m

Magarkar, A., Roìg, T., and Bunker, A. (2017). A computational study suggests that replacing PEG with PMOZ may increase exposure of hydrophobic targeting moiety. *Eur. J. Pharm. Sci.* 103, 128–135. doi: 10.1016/j.ejps.2017.03.008

Mahdavi, M., Fattahi, A., Tajkorshid, E., and Nouranian, S. (2020). Molecular insights into the loading and dynamics of doxorubicin on PEGylated graphene oxide nanocarriers. *ACS Appl. Biol. Mater.* 3, 1354–1363. doi: 10.1021/acsabm.9b00956

Mahmoudi, M. (2016). Protein corona: the golden gate to clinical applications of nanoparticles. *Int. J. Biochem. Cell Biol.* 75, 141–142. doi: 10.1016/j.biocel.2016.03.016

Maiti, P. K., and Bagchi, B. (2006). Structure and dynamics of DNA-dendrimer complexation: role of counterions, water, and base pair. *Nano Lett.* 6, 2478–2485. doi: 10.1021/nl061609m

Maleki, R., Afrouzi, H. H., Hosseini, M., Toghraie, D., Piranfar, A., and Rostami, S. (2020). pH-sensitive loading/releasing of doxorubicin using single-walled carbon nanotube and multi-walled carbon nanotube: a molecular dynamics study. *Comput. Meth. Prog. Biol.* 186:105210. doi: 10.1016/j.cmpb.2019.105210

Marianecci, C., Di Marzio, L., Rinaldi, F., Celia, C., Paolino, D., Alhaique, F., et al. (2014). Niosomes from 80s to present: the state of the art. *Adv. Colloid Interf.* 205, 187–206. doi: 10.1016/j.cis.2013.11.018

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

Martinez, D., Decossas, M., Kowal, J., Frey, L., Stahlberg, H., Dufourc, E. J. E. J., et al. (2017). Lipid internal dynamics probed in nanodiscs. *Chem. Phys. Chem.* 18, 2651–2657. doi: 10.1002/cphc.201700450

Martinez-Veracoechea, F. J., and Frenkel, D. (2011). Designing super selectivity in multivalent nano-particle binding. *Proc. Natl. Acad. Sci. U.S.A.* 108, 10963–10968. doi: 10.1073/pnas.1105351108

Martinho, N., Florindo, H., Silva, L., Brocchini, S., Zloh, M., and Barata, T. (2014). Molecular modeling to study dendrimers for biomedical applications. *Molecules* 19, 20424–20467. doi: 10.3390/molecules191220424

Marwah, M., Magarkar, A., Ray, D., Aswal, V. K., Bunker, A., and Nagarsenker, M. (2018). Glyceryl monostearate: probing the self assembly of a lipid amenable to surface modification for hepatic targeting. *J. Phys. Chem. C* 122, 22160–22169. doi: 10.1021/acs.jpcc.8b05931

Massiot, J., Makky, A., Di Meo, F., Chapron, D., Trouillas, P., and Rosilio, V. (2017). Impact of lipid composition and photosensitizer hydrophobicity on the efficiency of light-triggered liposomal release. *Phys. Chem. Chem. Phys.* 19, 11460–11473. doi: 10.1039/C7CP00983F

Mastrotto, F., Brazzale, C., Bellato, F., De Martin, S., Grange, G., Mahmoudzadeh, M., et al. (2020). In vitro and in vivo behavior of liposomes decorated with PEGs with different chemical features. *Mol. Pharm.* 17, 472–487. doi: 10.1021/acs.molpharmaceut.9b00887

Mathieu, D. (2020). QSPR versus fragment-based methods to predict octanol-air partition coefficients: revisiting a recent comparison of both approaches. *Chemosphere* 245:125584. doi: 10.1016/j.chemosphere.2019.125584

Matos, G. D. R., Kyu, D. Y., Loeffler, H. H., Chodera, J. D., Shirts, M. R., and Mobley, D. L. (2017). Approaches for calculating solvation free energies and enthalpies demonstrated with an update of the FreeSolv database. *J. Chem. Eng. Data* 62, 1559–1569. doi: 10.1021/acs.jced.7b00104

Matos, G. D. R., and Mobley, D. L. (2019). Challenges in the use of atomistic simulations to predict solubilities of drug-like molecules. *F1000Research* 7:686. doi: 10.12688/f1000research.14960.2

Mayne, C. G., Arcario, M. J., Mahinthichaichan, P., Baylon, J. L., Vermaas, J. V., Navidpour, L., et al. (2016). The cellular membrane as a mediator for small molecule interaction with membrane proteins. *Biochim. Biophys. Acta Biomembr.* 1858, 2290–2304. doi: 10.1016/j.bbamem.2016.04.016

Mazloom-Jalali, A., and Shariatinia, Z. (2019). Polycaprilactone nanocomposite systems used to deliver ifosfamide anticancer drug: molecular dynamics simulations. *Struct. Chem.* 30, 863–876. doi: 10.1007/s11224-018-1233-y

Mehta, C. H., Narayan, R., and Nayak, U. Y. (2019). Computational modelling for formulation design. *Drug Discov. Today* 24, 781–788. doi: 10.1016/j.drudis.2018.11.018

Meier, K., Choutko, A., Dolenc, J., Eichenberger, A. P., Riniker, S., and van Gunsteren, W. F. (2013). Multi-resolution simulation of biomolecular systems: a review of methodological issues. *Angew. Chem. Int. Edn.* 52, 2820–2834. doi: 10.1002/anie.201205408

Mhasal, A. R., and Roy, S. (2016). Free energy of bare and capped gold nanoparticles permeating through a lipid bilayer. *Chem. Phys. Chem.* 17, 3504–3514. doi: 10.1002/cphc.201600690

Mhashal, A. R., and Roy, S. (2014). Effect of gold nanoparticle on structure and fluidity of lipid membrane. *PLoS One* 9:e114152. doi: 10.1371/journal.pone.0114152

Michel, J., and Essex, J. W. (2010). Prediction of protein-ligand binding affinity by free energy simulations: assumptions, pitfalls and expectations. *J. Comput. Aid. Mol. Des.* 24, 639–658. doi: 10.1007/s10822-010-9363-3

Miller, S. E., Yamada, Y., Patel, N., Suárez, E., Andrews, C., Tau, S., et al. (2019). Electrostatically driven Guanidinium interaction domains that control hydrogel-mediated protein delivery in vivo. *ACS Cent. Sci.* 5, 1750–1759. doi: 10.1021/acscentsci.9b00501

Min, W., Zhao, D., Quan, X., Sun, D., Li, L., and Zhou, J. (2017). Computer simulations on the pH-sensitive tri-block copolymer containing zwitterionic sulfobetaine as a novel anti-cancer drug carrier. *Colloid Surf. B* 152, 260–268. doi: 10.1016/j.colsurfb.2017.01.033

Miyazaki, Y., Okazaki, S., and Shinoda, W. (2020). pSpica: a coarse-grained force field for lipid membranes based on a polar water model. *J. Chem. Theory Comput.* 16, 782–793. doi: 10.1021/acs.jctc.9b00946

Mobarak, E., Javanainen, M., Kulig, W., Honigmann, A., Sezgin, E., Aho, N., et al. (2018). How to minimize dye-induced perturbations while studying biomembrane structure and dynamics: PEG linkers as a rational alternative. *Biochim. Biophys. Acta Biomembr.* 1860, 2436–2445. doi: 10.1016/j.bbamem.2018.07.003

Mobasheri, M., Attar, H., Sorkhabadi, S. M. R., Khamesipour, A., and Jaafari, M. R. (2016). Solubilization behavior of polyene antibiotics in nanomicellar system: insights from molecular dynamics simulation of the amphotericin B and nystatin interactions with polysorbate 80. *Molecules* 21:6. doi: 10.3390/molecules21010006

Moghimi, S. M., Simberg, D., and Anchordoquy, T. J. (2020). Tuning the engines of nanomedicine. *Mol. Ther.* 28, 693–694. doi: 10.1016/j.ymthe.2020.01.025

Monpara, J., Kanthou, C., Tozer, G. M., and Vavia, P. R. (2018). Rational design of cholesterol derivative for improved stability of paclitaxel cationic liposomes. *Pharm. Res.* 35:90. doi: 10.1007/s11095-018-2367-8

Moore, T. C., Yang, A. H., Ogungbesan, O., Hartkamp, R., Iacovella, C. R., Zhang, Q., et al. (2019). Influence of single-stranded DNA coatings on the Interaction between Graphene Nanoflakes and Lipid Bilayers. *J. Phys. Chem. B* 123, 7711–7721. doi: 10.1021/acs.jpcb.9b04042

Moradi, S., Taran, M., Mohajeri, P., Sadrjavadi, K., Sarrami, F., Karton, A., et al. (2018). Study of dual encapsulation possibility of hydrophobic and hydrophilic drugs into a nanocarrier based on bio-polymer coated graphene oxide using density functional theory, molecular dynamics and experimental methods. *J. Mol. Liq.* 262, 204–217. doi: 10.1016/j.molliq.2018.04.089

Mortazavifar, A., Raissi, H., and Akbari, A. (2019). DFT and MD investigations on the functionalized boron nitride nanotube as an effective drug delivery carrier

for carmustine anticancer drug. *J. Mol. Liq.* 276, 577–587. doi: 10.1016/j.molliq.2018.12.028

Mousavi, M., Hakimian, S., Mustafa, T. A., Aziz, F. M., Salihi, A., Ale-Ebrahim, M., et al. (2019). The interaction of silica nanoparticles with catalase and human mesenchymal stem cells: biophysical, theoretical and cellular studies. *Int. J. Nanomed.* 14, 5355–5368. doi: 10.2147/IJN.S210136

Mousavi, S.-D., Maghsoodi, F., Panahandeh, F., Tazdian-Robati, R., Reisi-Vanani, A., and Tafaghodi, M. (2018). Doxorubicin delivery via magnetic nanomicelles comprising from reduction-responsive poly(ethylene glycol)-b-poly(s-caprilactone) (PEG-SS-PCL) and loaded with superparamagnetic iron oxide (SPIO) nanoparticles: preparation, characterization and loaded with. *Mater. Sci. Eng. C* 92, 631–643. doi: 10.1016/j.msec.2018.06.066

Mu, Q., Hu, T., and Yu, J. (2013). Molecular insight into the steric shielding effect of PEG on the conjugated staphylokinase: biochemical characterization and molecular dynamics simulation. *PLoS One* 8:e68559. doi: 10.1371/journal.pone.0068559

Munasinghe, A., Mathavan, A., Mathavan, A., Lin, P., and Colina, C. M. (2019). PEGylation within a confined hydrophobic cavity of a protein. *Phys. Chem. Chem. Phys.* 21, 25584–25596. doi: 10.1039/C9CP04387J

Murtola, T., Bunker, A., Vattulainen, I., Deserno, M., and Karttunen, M. (2009). Multiscale modeling of emergent materials: biological and soft matter. *Phys. Chem. Chem. Phys.* 11, 1869–1892. doi: 10.1039/B818051B

Myint, K. H., Brown, J. R., Shim, A. R., Wylouzil, B. E., and Hall, L. M. (2016). Encapsulation of nanoparticles during polymer micelle formation: a dissipative particle dynamics study. *J. Phys. Chem. B* 120, 11582–11594. doi: 10.1021/acs.jpcb.6b07324

Myung, Y., Yeom, S., and Han, S. (2016). A niosomal bilayer of sorbitan monostearate in complex with flavones: a molecular dynamics study. *J. Liposome Res.* 26, 336–344. doi: 10.3109/08982104.2016.1144204

Nag, O. K., and Delehanty, J. B. (2019). Active cellular and subcellular targeting of nanoparticles for drug delivery. *Pharmaceutics* 11:543. doi: 10.3390/pharmaceutics11100543

Nandy, B., and Maiti, P. K. (2011). DNA compaction by a dendrimer. *J. Phys. Chem. B* 115, 217–230. doi: 10.1021/jp106776v

Nandy, B., Maiti, P. K., and Bunker, A. (2013). Force biased molecular dynamics simulation study of effect of dendrimer generation on interaction with DNA. *J. Chem. Theory Comput.* 9, 722–729. doi: 10.1021/ct300678r

Nandy, B., Santosh, M., and Maiti, P. K. (2012). Interaction of nucleic acids with carbon nanotubes and dendrimers. *J. Biosci.* 37, 457–474. doi: 10.1007/s12038-012-9220-8

Nantasenamat, C., Isarankura-Na-Ayudhya, C., Naenna, T., and Prachayasittikul, V. (2009). A practical overview of qualitative structure-activity relationship. *EXCLI J.* 8, 74–88. doi: 10.17877/DE290R-690

Narayan, K. L., Rao, K. M., and Sarcar, M. M. M. (2008). *Computer Aided Design and Manufacturing*. New Delhi: Prentice Hall of India.

Nasr, M. L., Baptista, D., Strauss, M., Sun, Z. Y. J., Grigoriu, S., Huser, S., et al. (2016). Covalently circularized nanodiscs for studying membrane proteins and viral entry. *Nat. Methods* 14, 49–52. doi: 10.1038/nmeth.4079

Neale, C., Bennett, W. F. D., Tieleman, D. P., and Pomès, R. (2011). Statistical convergence of equilibrium properties in simulations of molecular solutes embedded in lipid bilayers. *J. Chem. Theory Comput.* 7, 4175–4188. doi: 10.1021/ct200316w

Neale, C., and Pomès, R. (2016). Sampling errors in free energy simulations of small molecules in lipid bilayers. *Biochim. Biophys. Acta Biomembr.* 1858, 2539–2548. doi: 10.1016/j.bbamem.2016.03.006

Nie, S. Y., Lin, W. J., Yao, N., Guo, X. D., and Zhang, L. J. (2014). Drug release from pH-sensitive polymeric micelles with different drug distributions: insight from coarse-grained simulations. *ACS Appl. Mater. Interf.* 6, 17668–17678. doi: 10.1021/am503920m

Nie, S. Y., Sun, Y., Lin, W. J., Wu, W. S., Guo, X. D., Qian, Y., et al. (2013). Dissipative particle dynamics studies of doxorubicin-loaded micelles assembled from four-arm star triblock polymers 4AS-PCL-b-PDEAEMA-b-PPEGMA and their pH-release mechanism. *J. Phys. Chem. B* 117, 13688–13697. doi: 10.1021/jp407529u

Nienhaus, K., and Nienhaus, G. U. (2019). Towards a molecular-level understanding of the protein corona around nanoparticles - recent advances and persisting challenges. *Curr. Opin. Biomed. Eng.* 10, 11–22. doi: 10.1016/j.cobme.2019.01.002

Nikfar, Z., and Shariatinia, Z. (2019). The RGD tripeptide anticancer drug carrier: DFT computations and molecular dynamics simulations. *J. Mol. Liq.* 281, 565–583. doi: 10.1016/j.molliq.2019.02.114

Nischan, N., and Hackenberger, C. P. R. (2014). Site-specific PEGylation of proteins: recent developments. *J. Organ. Chem.* 79, 10727–10733. doi: 10.1021/jo502136n

Numata, M., Grinkova, Y. V., Mitchell, J. R., Chu, H. W., Sligar, S. G., and Voelker, D. R. (2013). Nanodiscs as a therapeutic delivery agent: inhibition of respiratory syncytial virus infection in the lung. *Int. J. Nanomed.* 8, 1417–1427. doi: 10.2147/IJN.S39888

Nummelin, S., Selin, M., Legrand, S., Ropponen, J., Seitsonen, J., Nykänen, A., et al. (2017). Modular synthesis of self-assembling Janus-dendrimers and facule preparation of drug-loaded dendrimersomes. *Nanoscale* 9, 7189–7198. doi: 10.1039/C6NR08102A

Ogawara, K., Yoshizawa, Y., Un, K., Araki, T., Kimura, T., and Higaki, K. (2013). Nanoparticle-based passive drug targeting to tumors: considerations and implications for optimization. *Biol. Pharm. Bull.* 36, 698–702. doi: 10.1248/bpb.b13-00015

Ollila, O. H. S., and Pabst, G. (2016). Atomistic resolution structure and dynamics of lipid bilayers in simulations and experiments. *Biochim. Bbiophys. Acta Biomembr.* 1858, 2512–2528. doi: 10.1016/j.bbamem.2016.01.019

Oroskar, P. A., Jameson, C. J., and Murad, S. (2015). Surface-functionalized nanoparticle permeation triggers lipid displacement and water and ion leakage. *Langmuir* 31, 1074–1085. doi: 10.1021/la503934c

Oroskar, P. A., Jameson, C. J., and Murad, S. (2016). Simulated permeation and characterization of PEGylated gold nanoparticles in a lipid bilayer system. *Langmuir* 32, 7541–7555. doi: 10.1021/acs.langmuir.6b01740

Otto, D. P., Otto, A., and De Villers, M. M. (2013). Experimental and mesoscale computational dynamics studies of the relationship between solubility and release of quercetin from PEG solid dispersions. *Int. J. Pharm.* 456, 282–292. doi: 10.1016/j.ijpharm.2013.08.039

Ou, L., Corradi, V., Tieleman, D. P., and Liang, Q. (2020). Atomistic simulations on interactions between amphiphilic Janus nanoparticles and lipid bilayers: effects of lipid ordering and leaflet asymmetry. *J. Phys. Chem. B* 124, 4466–4475. doi: 10.1021/acs.jpcb.9b11989

Ouyang, D., and Smith, S. C. (2015). "Introduction to computational pharmaceutics," in *Computational Pharmaceutics: Application of Molecular Modeling in Drug Delivery*, eds D. Ouyang and S. C. Smith (Hoboken, NJ: Wiley), 27–31.

Pakdel, M., Raissi, H., and Shahabi, M. (2020). Predicting doxorubicin drug delivery by single-walled carbon nanotube through cell membrane in the absence and presence of nicotine molecules: a molecular dynamics simulation study. *J. Biomol. Struct. Dyn.* 38, 1488–1498. doi: 10.1080/07391102.2019.1611474

Paluch, A. S., Parameswaran, S., Liu, S., Kolavennu, A., and Mobley, D. L. (2015). Predicting the excess solubility of acetanilide, acetaminophen, phenacetin, benzocaine, and caffeine in binary water/ethanol mixtures via molecular simulation. *J. Chem. Phys.* 142:44508. doi: 10.1063/1.4906491

Pan, L., and Segrest, J. P. (2016). Computational studies of plasma lipoprotein lipids. *Biochim. Biophys. Acta Biomembr.* 1858, 2401–2420. doi: 10.1016/j.bbamem.2016.03.010

Panczyk, T., Jagusiak, A., Pastorin, G., Ang, W. H., and Narkiewicz-Michalek, J. (2013). Molecular dynamics study of cisplatin release from carbon nanotubes capped by magnetic nanoparticles. *J. Phys. Chem. C* 117, 17327–17336. doi: 10.1021/jp405593u

Panczyk, T., Wojton, P., and Wolski, P. (2020). Molecular dynamics study of the interaction of carbon nanotubes with telomeric DNA fragment containing noncanonical G-quadruplex and i-motif forms. *Int. J. Mol. Sci.* 21:1925. doi: 10.3390/ijms21061925

Park, K. (2016). Drug delivery of the future: chasing the invisible gorilla. *J. Control Release* 240, 2–8. doi: 10.1016/j.jconrel.2015.10.048

Park, K. (2019). The beginning of the end of nanomedicine. *J. Control Release* 305, 221–222. doi: 10.1016/j.jconrel.2019.05.044

Parray, Z. A., Hassan, M. I., Ahmad, F., and Islam, A. (2020). Amphiphilic nature of polyethylene glycols and their role in medical research. *Polym. Test.* 82:106316. doi: 10.1016/j.polymertesting.2019.106316

Pasenkiewicz-Gierula, M., Baczynski, K., Markiewicz, M., and Murzyn, K. (2016). Computer modelling studies of the bilayer/water interface. *Biochim. Biophys. Acta Biomembr.* 1858, 2305–2321. doi: 10.1016/j.bbamem.2016.01.024

Pasenkiewicz-Gierula, M., Takaoka, Y., Miyagawa, H., Kitamura, K., and Kusumi, A. (1997). Hydrogen bonding of water to phosphatidylcholine in the membrane as studied by a molecular dynamics simulation: location, geometry, and lipid-lipid bridging via hydrogen-bonded water. *J. Phys. Chem. A* 101, 3677–3691. doi: 10.1021/jp962099v

Pasut, G., and Veronese, F. M. (2012). State of the art in PEGylation: the great versatility achieved after 40 years of research. *J. Control Release* 161, 461–472. doi: 10.1016/j.jconrel.2011.10.037

Patel, S. K., Lavasanifar, A. A., and Choi, P. (2010a). Molecular dynamics study of the encapsulation capability of a PCL-PEO based block copolymer for hydrophobic drugs with different spatial distributions of hydrogen bond donors and acceptors. *Biomaterials* 31, 1780–1786.

Patel, S. K., Lavasanifar, A., and Choi, P. (2010b). Prediction of the solubility of cururbitacin drugs in self associationg poly(ethylene oxide)-b-poly(alpha-benzyl carboxylate epsilon-caprolactone) block copolymer with different tacticities using molecular dynamics simulation. *Biomaterials* 31, 345–357.

Pathak, P., Dhawan, V., Magarkar, A., Danne, R., Govindarajan, S., Ghosh, S., et al. (2016). Design of cholesterol arabinogalactan anchored liposomes for asialoglycoprotein receptor mediated targeting to hepatocellular carcinoma: *In silico* modeling, *in vitro* and *in vivo* evaluation. *Int. J. Pharm.* 509, 149–158. doi: 10.1016/j.ijpharm.2016.05.041

Pattni, B. S., Chupin, V. V., and Torchilin, V. P. (2015). New developments in liposomal drug delivery. *Chem. Rev.* 115, 10938–10966. doi: 10.1021/acs.chemrev.5b00046

Pederzoli, F., Tosi, G., Vandelli, M. A., Belletti, D., Forni, F., and Ruozi, B. (2017). Protein corona and nanoparticles: how can we investigate on? *WIREs Nanomed. Nanobiotechnol.* 9:e1467. doi: 10.1002/wnan.1467

Pelegri-Oday, E. M., Lin, E. W., and Maynard, H. D. (2014). Therapeutic protein-polymer conjugates: advancing beyond pegylation. *J. Am. Chem. Soc.* 136, 14323–14332. doi: 10.1021/ja504390x

Peng, Z., Han, X., Li, S., Al-Youbi, A. O., Bashammakh, A. S., El-Shahawi, M. S., et al. (2017). Carbon dots: biomacromolecule interaction, bioimaging and nanomedicine. *Coord. Chem. Rev.* 343, 256–277. doi: 10.1016/j.ccr.2017.06.001

Pennetta, C., Floresta, G., Graziano, A. C.-E., Cardile, V., Rubino, L., Galimberti, M., et al. (2020). Functionalization of single and multi-walled carbon nanotubes with polypropylene glycol decorated pyrrole for the development of doxorubicin nano-conveyors for cancer drug delivery. *Nanomaterials* 10:1073. doi: 10.3390/nano10061073

Pérez-Sánchez, G., Vicente, F. A., Schaeffer, N., Cardoso, I. S., Ventura, S. P. M., Jorge, M., et al. (2020). Unravelling the interactions between surface-active ionic liquids and triblock copolymers for the design of thermal responsive systems. *J. Phys. Chem. B* 124, 7046–7058. doi: 10.1021/acs.jpcb.0c02992

Persson, L. C., Porter, C. J. H., Charman, W. N., and Bergström, C. A. S. (2013). Computational prediction of drug solubility in lipid based formulation excipients. *Pharm. Res.* 30, 3225–3237. doi: 10.1007/s11095-013-1083-7

Peters, B. L., Lane, J. M. D., Ismail, A. E., and Grest, G. S. (2012). Fully atomistic simulations of the response of silica nanoparticle coatings to alkane solvents. *Langmuir* 28, 17443–17449. doi: 10.1021/la3023166

Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., et al. (2005). Scalable molecular dynamics with NAMD. *J. Comput. Chem.* 26, 1781–1802. doi: 10.1002/jcc.20289

Pinzaru, I., Coricovac, D., Dehelean, C., Moacã, E. A., Mioc, M., Baderca, F., et al. (2018). Stable PEG-coated silver nanoparticles-A comprehensive toxicological profile. *Food Chem. Toxicol.* 111, 546–556. doi: 10.1016/j.fct.2017.11.051

Pison, U., Welte, T., Giersing, M., and Groneberg, D. A. (2006). Nanomedicine for respiratory diseases. *Eur. J. Pharmacol.* 533, 341–350. doi: 10.1016/j.ejphar.2005.12.068

Plimpton, S. (1995). Fast parallel algorithms for short-range molecular dynamics. *J. Comp. Phys.* 117, 1–19. doi: 10.1006/jcph.1995.1039

Poojari, C., Wilkosz, N., Lira, R. B., Dimova, R., Jurkiewicz, P., Petka, R., et al. (2019). Behavior of the DPH fluorescence probe in membranes perturbed by drugs. *Chem. Phys. Lipids* 223:104784. doi: 10.1016/j.chemphyslip.2019.104784

Poojari, C., Zak, A., Dzieciuch-Rojek, M., Bunker, A., Kepczynski, M., and Róg, T. (2020). Cholesterol reduces partitioning of antifungal drug itraconazole into lipid bilayers. *J. Phys. Chem. B* 124, 2139–2148. doi: 10.1021/acs.jpcb.9b11005

Postila, P. A., Kaszuba, K., Sarewicz, M., Osyczka, A., Vattulainen, I., and Róg, T. (2013). Key role of water in proton transfer at the Q_o-site of the cytochrome bc_1 complex predicted by atomistic molecular dynamics simulations. *Biochim. Biophys. Acta Bioenerg.* 1827, 761–768. doi: 10.1016/j.bbabio.2013.02.005

Postila, P. A., and Róg, T. (2020). A perspective: active role of lipids in neurotransmitter dynamics. *Mol. Neurobiol.* 57, 910–925. doi: 10.1007/s12035-019-01775-7

Potdar, D., and Sammalkorpi, M. (2015). Asymmetric heat transfer from nanoparticles in lipid bilayers. *Chem. Phys.* 463, 22–29. doi: 10.1016/j.chemphys.2015.09.016

Pourmousa, M., and Pastor, R. W. (2018). Molecular dynamics simulations of lipid nanodiscs. *Biochim. Bbiophys. Acta Biomembr.* 1860, 2094–2107. doi: 10.1016/j.bbamem.2018.04.015

Prhashanna, A., Tan, W. K., Khan, S. A., and Chen, S. B. (2016). Co-micellization behavior of triblock copolymers in the presence of hydrophobic drug molecules: a simulation study. *Colloid Surf. B* 148, 299–307. doi: 10.1016/j.colsurfb.2016.09.004

Price, M. E., Cornelius, R. M., and Brash, J. L. (2001). Protein adsorption to polyethylene glycol modified liposomes from fibrinogen solution and from plasma. *Biochim. Biophys. Acta* 1512, 191–205. doi: 10.1016/S0005-2736(01)00330-3

Qin, X., Yu, C., Wei, J., Li, L., Zhang, C., Wu, Q., et al. (2019). Rational design of nanocarriers for intracellular protein delivery. *Adv. Mater.* 31:1902791. doi: 10.1002/adma.201902791

Quan, X., Zhao, D., Li, L., and Zhou, J. (2017). Understanding the cellular uptake of pH-responsive zwitterionic gold nanoparticles: a computer simulation study. *Langmuir* 33, 14480–14489. doi: 10.1021/acs.langmuir.7b03544

Raczyński, P., Górny, K., Bełdowski, P., Yuvan, S., and Dendzik, Z. (2020). Application of graphene as a nanoindenter interacting with phospholipid membranes-computer simulation study. *J. Phys. Chem. B* 124, 6592–6602. doi: 10.1021/acs.jpcb.0c02319

Rahmati, S., Shojaei, F., Shojaeian, A., Rezakhani, L., and Dehkordi, M. B. (2020). An overview of current knowledge in biological functions and potential theragnostic applications of exosomes. *Chem. Phys. Lipids* 226:104836. doi: 10.1016/j.chemphyslip.2019.104836

Ramalho, M. J., Andrade, S., Loureiro, J. A., and do Carmo Pereira, M. (2020). Nanotechnology to improve the Alzheimer's disease therapy with natural compounds. *Drug Deliv. Transl. Res.* 10, 380–402. doi: 10.1007/s13346-019-00694-3

Raman, A. S., Pajak, J., and Chiew, Y. C. (2018). Interaction of PCL based self-assembled nano-polymeric micelles with model lipid bilayers using coarse-grained molecular dynamics simulations. *Chem. Phys. Lett.* 712, 1–6. doi: 10.1016/j.cplett.2018.09.049

Ramezani, M., and Shamsara, J. (2016). Application of DPD in the design of polymeric nano-micelles as drug carriers. *J. Mol. Graph. Model.* 66, 1–8. doi: 10.1016/j.jmgm.2016.01.010

Ramezanpour, M., Leung, S. S. W., Delgado-Magnero, K. H., Bashe, B. Y. M., Thewalt, J., and Tieleman, D. P. (2016). Computational and experimental approaches for investigating nanoparticle-based drug delivery systems. *Biochim. Biophys. Acta* 1858, 1688–1709. doi: 10.1016/j.bbamem.2016.02.028

Ramli, R. A., Laftah, W. A., and Hashim, S. (2013). Core-shell polymers: a review. *RSC Adv.* 3, 15543–15565. doi: 10.1039/C3RA41296B

Ramos, M. C., Horta, V. A. C., and Horta, B. A. C. (2019). Molecular dynamics simulations of PAMAM and PPI dendrimers using the GROMOS-compatible 2016H66 force field. *J. Chem. Inf. Model.* 59, 1444–1457. doi: 10.1021/acs.jcim.8b00911

Razmimanesh, F., Amjad-Iranagh, S., and Modarress, H. (2015). Molecular dynamics simulation study of chitosan and gemcitabine as a drug delivery system. *J. Mol. Model.* 21:165. doi: 10.1007/s00894-015-2705-2

Ricklin, D., Hajishengallis, G., Yang, K., and Lambris, J. D. (2010). Complement: a key system for immune surveillance and homeostasis. *Nat. Immunol.* 11, 785–797. doi: 10.1038/ni.1923

Riehemann, K., Schneider, S. W., Luger, T. A., Godin, B., Ferrari, M., and Fuchs, H. (2009). Nanomedicine-challenge and perspectives. *Angew. Chem. Int. Edn.* 48, 872–897. doi: 10.1002/anie.200802585

Rissanen, S., Kumorek, M., Martinez-Seara, H., Li, Y. C., Jamróz, D., Bunker, A., et al. (2014). Effect of PEGylation on drug entry into lipid bilayer. *J. Phys. Chem. B* 118, 144–151. doi: 10.1021/jp4105745

Ritwiset, A., Krongsuk, S., and Johns, J. R. (2016). Molecular structure and dynamical properties of niosome bilayers with and without cholesterol incorporation: a molecular dynamics simulation study. *Appl. Surf. Sci.* 380, 23–31. doi: 10.1016/j.apsusc.2016.02.092

Rodríguez-Hidalgo, M.-R., Soto-Figueroa, C., and Vicente, L. (2011). Mesoscopic simulation of the drug release mechanism on the polymeric vehicle P(ST-DVB) in an acid environment. *Soft Matter* 7, 8224–8230.

Rossi, G., and Monticelli, L. (2016). Gold nanoparticles in model biological membranes: a computational perspective. *Biochim. Biophys. Acta Biomembr.* 1858, 2380–2389. doi: 10.1016/j.bbamem.2016.04.001

Roux, B. (1995). The calculation of the potential of mean force using computer simulations. *Comput. Phys. Commun.* 91, 275–282. doi: 10.1016/0010-4655(95)00053-I

Rozman, K. K., and Doull, J. (2001). Paracelsus, Haber and Arndt. *Toxicology* 160, 191–196. doi: 10.1016/s0300-483x(00)00447-9

Roìg, T., and Pasenkiewicz-Gierula, M. (2004). Non-polar interactions between cholesterol and phospholipids: a molecular dynamics simulation study. *Biophys. Chem.* 107, 151–164. doi: 10.1016/j.bpc.2003.09.002

Róg, T., and Vattulainen, I. (2014). Cholesterol, sphingolipids, and glycolipids: what do we know about their role in raft-like membranes? *Chem. Phys. Lipids* 184, 82–104. doi: 10.1016/j.chemphyslip.2014.10.004

Ruiz-Garcia, A., Bermejo, M., Moss, A., and Casabo, V. G. (2008). Pharmacokinetics in drug discovery. *J. Pharm. Sci.* 97, 654–690. doi: 10.1002/jps.21009

Rungrotmongkol, C. R. T., and Poo-arporn, R. P. (2016). pH-controlled doxorubicin anticancer loading and release from carbon nanotube noncovalently modified by chitosan: MD simulations. *J. Mol. Graph. Model.* 70, 70–76. doi: 10.1016/j.jmgm.2016.09.011

Salassi, S., Canepa, E., Ferrando, R., and Rossi, G. (2019). Anionic nanoparticle-lipid membrane interactions: the protonation of anionic ligands at the membrane surface reduces membrane disruption. *RSC. Adv.* 9, 13992–13997. doi: 10.1039/C9RA02462J

Salassi, S., Simonelli, F., Bochicchio, D., Ferrando, R., and Rossi, G. (2017). Au nanoparticles in lipid bilayers: a comparison between atomistic and coarse-grained models. *J. Phys. Chem. C* 121, 10927–10935. doi: 10.1021/acs.jpcc.6b12148

Salorinne, K., Malola, S., Wong, O. A., Rithner, C. D., Chen, X., Ackerson, C. J., et al. (2016). Conformation and dynamics of the ligand shell of a water-soluble Au 102 nanoparticle. *Nat. Commun.* 7:10401. doi: 10.1038/ncomms10401

Santos, H. A., Mäkilä, E., Airaksinen, A. J., Bimbo, L. M., and Hirvonen, J. (2014). Porous silicon nanoparticles for nanomedicine: preparation and biomedical applications. *Nanomedicine* 9, 535–554. doi: 10.2217/nnm.13.223

Sarma, J. V., and Ward, P. A. (2011). The complement system. *Cell Tissue Res.* 343, 227–235. doi: 10.1007/s00441-010-1034-0

Scanell, J. W., Blanckley, A., Boldon, H., and Warrington, B. (2012). Diagnosing the decline in pharmaceutical R&D efficiency. *Nat. Rev. Drug Discov.* 11, 191–200. doi: 10.1038/nrd3681

Schachter, I., Allolio, C., Khelashvili, G., and Harries, D. (2020). Confinement in nanodiscs anisotropically modifies lipid bilayer elastic properties. *J. Phys. Chem. B* 124, 7166–7175. doi: 10.1021/acs.jpcb.0c03374

Schafer, T., Zhou, J., Schmid, F., and Settanni, G. (2017). "Blood proteins and their interactions with nanoparticles investigated using molecular dynamics simulation," in *High Performance Computing in Science and Engineering '17*, eds W. E. Nagel, D. H. Kröner, and M. M. Resch (Berlin: Springer), 5–19. doi: 10.1007/978-3-319-68394-2

Scheideler, M., Vidakovic, I., and Prassl, R. (2020). Lipid nanocarriers for microRNA delivery. *Chem. Phys. Lipids* 226:104837. doi: 10.1016/j.chemphyslip.2019.104837

Schlapschy, M., Binder, U., Börger, C., Theobald, I., Wachinger, K., Kisling, S., et al. (2013). PASylation: a biological alternative to PEGylation for extending the plasma half-life of pharmaceutically active proteins. *Protein. Eng. Des. Sel.* 26, 489–501. doi: 10.1093/protein/gzt023

Schöttler, S., Becker, G., Winzen, S., Steinbach, T., Mohr, K., Landfester, K., et al. (2016). Protein adsorption is required for stealth effect of poly(ethylene glycol)- and poly(phosphoester)- coated nanocarriers. *Nat. Nanotechnol.* 11, 372–377. doi: 10.1038/nnano.2015.330

Schubertová, V., Martinez-Veracoechea, F. J., and Vácha, R. (2015). Influence of ligand distribu- tion on uptake efficiency. *Soft Matter* 11, 2726–2730. doi: 10.1039/C4SM02815E

Sedlacek, O., Monnery, B. D., Filippov, S. K., Hoogenboom, R., and Hruby, M. (2012). Poly(2-oxazoline)s - are they more advantageous for biomedical applications than other polymers? *Macromol. Rapid Commun.* 33, 1648–1662. doi: 10.1002/marc.201200453

Šegota, S., and Durdica, T. (2006). Spontaneous formation of vesicles. *Adv. Coll. Int. Sci.* 121, 51–75. doi: 10.1063/1.465966

Sen, S., Han, Y., Rehak, P., Vuković, L., and Král, P. (2018). Computational studies of micellar and nanoparticle nanomedicines. *Chem. Soc. Rev.* 47, 3849–3860. doi: 10.1039/C8CS00022K

Settanni, G., Schäfer, T., Muhl, C., Barz, M., and Schmid, F. (2018). Poly-sarcosine and poly(ethylene-glycol) interactions with proteins investigated using molecular dynamics simulations. *Comput. Struct. Biotechnol.* 16, 543–550. doi: 10.1016/j.csbj.2018.10.012

Settanni, G., Zhou, J., and Schmid, F. (2017a). Interactions between proteins and poly(ethylene-glycol) investigated using molecular dynamics simulations. *J. Phys. Conf. Ser.* 921:12002. doi: 10.1088/1742-6596/921/1/012002

Settanni, G., Zhou, J., Suo, T., Schöttler, S., Landfester, K., Schmid, F., et al. (2017b). Protein corona composition of poly(ethylene glycol)- and poly(phosphoester)-coated nanoparticles correlates strongly with the amino acid composition of the protein surface. *Nanoscale* 9, 2138–2144. doi: 10.1039/C6NR07022A

Sgarlata, C., D'Urso, L., Consiglio, G., Grasso, G., Satriano, C., and Forte, G. (2016). pH sensitive functionalized graphene oxide as a carrier for delivering Gemcitabine: a computational approach. *Comput. Theor. Chem.* 1096, 1–6. doi: 10.1016/j.comptc.2016.09.026

Shadrack, D. M., and Swai, H. S. (2019). Solvent effects on molecular encapsulation of Toussantine-A by chiotosan nanoparticle: a metadynamics study. *J. Mol. Liq.* 292:111434. doi: 10.1016/j.molliq.2019.111434

Shah, S., Liu, Y., Hu, W., and Gao, J. (2011). Modeling particle shape-dependent dynamics in nanomedicine. *J. Nanosci. Nanotechnol.* 11, 919–928. doi: 10.1166/jnn.2011.3536

Shahabi, M., and Raissi, H. (2020). Payload delivery of anticancer drug Tegafur with the assistance of graphene oxide nanosheet during biomembrane penetration: molecular dynamics simulation survey. *Appl. Surf. Sci.* 517:146186. doi: 10.1016/j.apsusc.2020.146186

Shamsi, M., Mohammadi, A., Manshadi, M. K. D., and Sanati-Nezhad, A. (2019). Mathematical and computational modeling of nano-engineered drug delivery systems. *J. Control Release* 307, 150–165. doi: 10.1016/j.jconrel.2019.06.014

Shan, P., Shen, J.-W., Xu, D.-H., Shi, L.-Y., Gao, J., Lan, Y., et al. (2014). Molecular dynamics study on the interaction between doxorubicin and hydrophobically modified chitosan oligosaccaride. *RSC. Adv.* 4, 23730–23739. doi: 10.1039/C4RA01199F

Shao, Q., and Hall, C. K. (2016). Protein adsorption on nanoparticles: model development using computer simulation. *J. Phys. Condens. Matter* 28:414019. doi: 10.1088/0953-8984/28/41/414019

Shariatinia, Z., and Mazloom-Jalali, A. (2019). Chitosan nanocomposite drug delivery systems designed for the ifosfamide anticancer drug using molecular dynamics simulations. *J. Mol. Liq.* 273, 346–367. doi: 10.1016/j.molliq.2018.10.047

Shen, J.-W., Li, J., Zhao, Z., Zhang, L., Peng, G., and Liang, L. (2017). Molecular dynamics study on the mechanism of polynucleotide encapsulation by chitosan. *Sci. Rep.* 7:5050. doi: 10.1038/s41598-017-05197-0

Shi, C., Sun, Y., Wu, H., Zhu, C., Wei, G., Li, J., et al. (2016). Exploring the effect of hydrophilic and hydrophobic structure of grafted polymeric micelles on drug loading. *Int. J. Pharm.* 512, 282–291. doi: 10.1016/j.ijpharm.2016.08.054

Shimizu, K., Nakamura, H., and Watano, S. (2016). MD simulation study of direct permeation of a nanoparticle across the cell membrane under an external electric field. *Nanoscale* 8:11897. doi: 10.1039/C6NR02051H

Simonelli, F., Bochicchio, D., Ferrando, R., and Rossi, G. (2015). Monolayer-protected anionic au nanoparticles walk into lipid membranes step by step. *J. Phys. Chem. Lett.* 6, 3175–3179. doi: 10.1021/acs.jpclett.5b01469

Simonsen, J. B. (2016). Evaluation of reconstituted high-density lipoprotein (rHDL) as a drug delivery platform - a detailed survey of rHDL particles ranging

from biophysical properties to clinical implications. *Nanomed. Nanotechnol. Biol. Med.* 12, 2161–2179. doi: 10.1016/j.nano.2016.05.009

Siuda, I., and Tieleman, D. P. (2015). Molecular models of nanodiscs. *J. Chem. Theory Comput.* 11, 4923–4932. doi: 10.1021/acs.jctc.5b00668

Sliwakosky, G., Kothiwale, S., Meiler, J., and Lowe, E. W. (2014). Computational methods in drug discovery. *Pharmacol. Rev.* 66, 334–395. doi: 10.1124/pr.112.007336

Smeijers, A. F., Markvoort, A. J., Pieterse, K., and Hilbers, P. A. J. (2016a). Coarse-grained modelling of urea-adamantyl functionalized poly(propylene imine) dendrimers. *Mol. Simulat.* 42, 882–895. doi: 10.1080/08927022.2015.1096359

Smeijers, A. F., Markvoort, A. J., Pieterse, K., and Hilbers, P. A. J. (2016b). Coarse-grained simulations of poly(propylene imine) dendrimers in solution. *J. Chem. Phys.* 144:74903. doi: 10.1063/1.4941379

Smith, C. J., Wagle, D. V., Bhawawet, N., Gehrke, S., Hollóczki, O., Pingali, S. V., et al. (2020). Combined small-angle neutron scattering, diffusion NMR, and Molecular dynamics study of a eutectogel: illuminating the dynamical behavior of glyceline confined in bacterial cellulose gels. *J. Phys. Chem. B* 124, 7647–7658. doi: 10.1021/acs.jpcb.0c04916

Smith, D. J., Leal, L.-G., Mitragorti, S., and Shell, M. S. (2018). Nanoparticle transport across model cellular membranes: when do solubility-diffusion models break down? *J. Phys. D Appl. Phys.* 51:294004. doi: 10.1088/1361-6463/aacac9

Soltani, S., Sardari, S., and Soror, S. A. (2010). Computer simulation of a novel pharmaceutical silicon nanocarrier. *Nanotechnol. Sci. Appl.* 3, 149–157. doi: 10.2147/NSA.S8378

Somjid, S., Krongsuk, S., and Johns, J. R. (2018). Cholesterol concentration effect on the bilayer properties and phase formation of niosome bilayers: a molecular dynamics simulation study. *J. Mol. Liq.* 256, 591–598.

Song, B., Yuan, H., Jameson, C. J., and Murad, S. (2011). Permeation of nanocrystals across lipid membranes. *Mol. Phys.* 109, 1511–1526. doi: 10.1080/00268976.2011.569511

Song, B., Yuan, H., Pham, S. V., Jameson, C. J., and Murad, S. (2012). Nanoparticle permeation induces water penetration, ion transport and lipid flip-flop. *Langmuir* 28, 16989–17000. doi: 10.1021/la302879r

Sousa, S. F., Fernandes, P. A., and Ramos, M. J. (2006). Protein-ligand docking: current status and future challenges. *Proteins* 65, 15–26. doi: 10.1002/prot.21082

Souza, S. F., Peres, J., Coelho, M., and Vieira, T. F. (2018). Analyzing PEGylation through molecular dynamics. *Chem. Select* 3, 8415–8427. doi: 10.1002/slct.201800855

Sridhar, D. B., Gubta, R., and Rai, B. (2018). Effect of surface coverage and chemistry on self-assembly of monolayer protected gold nanoparticles: a molecular dynamics simulations study. *Phys. Chem. Chem. Phys.* 20, 25883–25891. doi: 10.1039/c8cp04044c

Srinivas, G., Mohan, R. V., and Kelkar, A. D. (2013). Polymer micelle assisted transport and delivery of model hydrophilic components inside a biological lipid vesicle: a coarse grained simulation study. *J. Phys. Chem. B* 117, 12095–12104. doi: 10.1021/jp405381k

Stepien, P., Augustyn, B., Poojari, C., Galan, W., Polit, A., Vattulainen, I., et al. (2020). Complexity of seemingly simple lipid nanodiscs. *Biochim. Bbiophys. Acta Biomembr.* 1862:183420. doi: 10.1016/j.bbamem.2020.183420

Stepien, P., Polit, A., and Wisniewska-Becker, A. (2015). Comparative EPR studies on lipid bilayer properties in nanodiscs and liposomes. *Biochim. Biophys. Acta Biomembr.* 1848, 60–66. doi: 10.1016/j.bbamem.2014.10.004

Stepniewski, M., Kepczynski, M., Jamróz, D., Nowakowska, M., Rissanen, S., Vattulainen, I., et al. (2012). Interaction of hematoporphyrin with lipid membranes. *J. Phys. Chem. B* 116, 4889–4897. doi: 10.1021/jp300899b

Stepniewski, M., Pasenkiewicz-Gierula, M., Roìg, T., Danne, R., Orłowski, A., Karttunen, M., et al. (2011). Study of PEGylated lipid layers as a model for PEGylated liposome surfaces: molecular dynamics simulation and langmuir monolayer studies. *Langmuir* 27, 7788–7798. doi: 10.1021/la200003n

Styliari, I. D., Taresco, V., Theophilus, A., Alexander, C., Garnett, M., and Laughton, C. (2020). Nanoformulation-by-design: an experimental and molecular dynamics study for polymer coated drug nanoparticles. *RSC Adv.* 10, 19521–19533. doi: 10.1039/D0RA00408A

Su, C.-F., Merliz, H., Rabbel, H., and Sommer, J.-U. (2017). Nanoparticles of various degrees of hydrophobicity interacting with lipid membranes. *J. Phys. Chem. Lett.* 8, 4069–4076. doi: 10.1021/acs.jpclett.7b01888

Su, Y., Quan, X., Li, L., and Zhao, J. (2018). Computer simulation of DNA condensation by PAMAM dendrimer. *Macromol. Theory Simul.* 27:1700070. doi: 10.1002/mats.201700070

Subramaniam, B., Siddik, Z. H., and Nagoor, N. H. (2020). Optimization of nanostructured lipid carriers: understanding the types, designs and parameters in the process of formulations. *J. Nanopart. Res.* 22:141. doi: 10.1007/s11051-020-04848-0

Sun, H. (1998). COMPASS: an ab initio force-field optimized for condensed-phase applications - overview with details on alkane and benzene compounds. *J. Phys. Chem. B* 102, 7338–7346. doi: 10.1021/jp980939v

Sun, H., She, P., Lu, G., Xu, K., Zhang, W., and Liu, Z. (2014). Recent advances in the development of functionalized carbon nanotubes: a versatile vector for drug delivery. *J. Mater. Sci.* 49, 6845–6854. doi: 10.1007/s10853-014-8436-4

Sun, X., Riccardi, L., De Biasi, F., Rastrelli, F., De Vivo, M., and Mancin, F. (2019). Molecular-dynamics-simulation-directed rational design of nanoreceptors with targeted affinity. *Angew. Chem. Int. Edn.* 58, 7702–7707. doi: 10.1002/anie.201902316

Sun, Y., and Xia, Y. (2003). Gold and silver nanoparticles: a class of chromophores with colors tunable in the range from 400 to 750 nm. *Analyst* 128, 686–691. doi: 10.1039/B212437H

Szebeni, J., Baranyi, L., Savay, S., Milosevits, J., Bunger, R., Laverman, P., et al. (2002). Role of complement activation in hypersensitivity reactions to doxil and hynic PEG liposomes: experimental and clinical studies. *J. Liposome Res.* 12, 165–172. doi: 10.1081/lpr-120004790

Tavanti, F., Pedone, A., and Menziani, M. C. (2019). Multiscale molecular dynamics simulation of multiple protein adsorptin on gold nanoparticles. *Int. J. Mol. Sci.* 20:3539. doi: 10.3390/ijms20143539

The Nobel Prize in Chemistry 2013. (2013). *NobelPrize.org. Nobel Media AB.* Available online at: https://www.nobelprize.org/prizes/chemistry/2013/summary/

Thewalt, J., and Tieleman, D. P. (2016). Biophysical experiments and simulation in nanoparticle based drug delivery systems. *J. Drug Target.* 24, 768–773. doi: 10.1080/1061186X.2016.1221957

Thiel, W. (2014). Semiempirical quantim-chemical methods. *WIREs Comput. Mol. Sci.* 4, 145–157. doi: 10.1021/acs.jctc.8b01265

Thota, N., Hu, Z., and Jiang, J. (2016). Ibuprofen loading and release in amphiphilic peptide FA32 and its derivatives: a coarse-grained molecular dynamics simulation study. *Mol. Simulat.* 42, 679–687. doi: 10.1080/08927022.2015.1079907

Thota, N., and Jiang, J. (2015). Computational amphiphilic materials for drug delivery. *Front. Mater.* 2:64. doi: 10.3389/fmats.2015.00064

Tian, F., Yue, T., Li, Y., and Zhang, X. (2014a). Computer simulation studies on the interactions between nanoparticles and cell membrane. *Sci. China Chem.* 57, 1662–1671. doi: 10.1007/s11426-014-5231-7

Tian, F., Zhang, X., and Dong, W. (2014b). How hydrophobic nanoparticles aggregate in the interior of membranes: a computer simulation. *Phys. Rev. E* 90:52701. doi: 10.1103/PhysRevE.90.052701

Tian, W., and Ma, Y. (2013). Theoretical and computational studies of dendrimers as delivery vectors. *Chem. Soc. Rev.* 42, 705–727. doi: 10.1039/c2cs35306g

Tokarský, J., Andrýsek, T., and ČapkovŁ, P. (2011). Molecular modeling of gel nanoparticles with cyclosporine A for oral drug delivery. *Int. J. Pharm.* 410, 196–205. doi: 10.1016/j.ijpharm.2011.03.026

Tomalia, D. A., Naylor, A. M., and Goddard, W. A. III (1990). Starburst dendrimers: molecular-level control of size, shape, surface chemistry, topology, and flexibility from atoms to macroscopic matter. *Angew. Chem. Int. Edn. Engl.* 29, 138–175. doi: 10.1002/anie.199001381

Tong, R., and Kohane, D. S. (2016). New strategies in cancer nanomedicine. *Annu. Rev. Pharmacol. Toxicol.* 56, 41–57. doi: 10.1146/annurev-pharmtox-010715-103456

Toporov, A. A., and Toporova, A. P. (2020). QSPR/QSAR: state-of-the-art, wierdness, the future. *Molecules* 25:1292. doi: 10.3390/molecules25061292

Torchilin, V. P. (2007). Micellar nanocarriers: pharmaceutical perspectives. *Pharm. Res.* 24, 1–16. doi: 10.1007/s11095-006-9132-0

Touitou, E., Dayan, N., Bergelson, L., Godin, B., and Eliaz, M. (2000). Ethosomes - novel vesicular carriers for enhanced delivery: characterization and skin penetration properties. *J. Control Release* 65, 403–418. doi: 10.1016/s0168-3659(99)00222-9

Toy, R., Hayden, E., Shoup, C., Baskaran, H., and Karathanasis, E. (2011). Effect of particle size, density and shape on margination of nanoparticles in microcirculation. *Nanotechnology* 22:115101. doi: 10.1088/0957-4484/22/11/115101

Truong, N. P., Whittaker, M. R., Mak, C. W., and Davis, T. P. (2015). The importance of nanoparticle shape in cancer drug therapy. *Expert Opin. Drug Deliv.* 12, 129–142. doi: 10.1517/17425247.2014.950564

Tu, C.-K., Chen, K., Tian, W., and Ma, Y. (2013). Computational investigations of a peptide-modified dendrimer interacting with lipid membranes. *Macromol. Rapid Commun.* 34, 1237–1242. doi: 10.1002/marc.201300360

Turturro, F. (2014). Denileukin diftox: a biotherapeutic paradigm shift in the treatment of lymphoid-derived disorders. *Expert. Rev. Anticancer Ther.* 7, 11–17. doi: 10.1586/14737140.7.1.11

Vácha, R., Martinez-Veracoechea, F. J., and Frenkel, D. (2011). Receptor-mediated endocytosis of nanoparticles of various shapes. *Nano Lett.* 11, 5391–5395. doi: 10.1021/nl2030213

van Gunsteren, W. F., Dolenc, J., and Mark, A. E. (2008). Molecular simulation as an aid to experimentalists. *Curr. Opin. Struct. Biol.* 18, 149–153. doi: 10.1016/j.sbi.2007.12.007

Van Lehn, R. C., and Alexander-Katz, A. (2014a). Fusion of ligand-coated nanoparticles with lipid bilayers: effect of ligand flexibility. *J. Phys. Chem. A* 118, 5848–5856. doi: 10.1021/jp411662c

Van Lehn, R. C., and Alexander-Katz, A. (2014b). Membrane-embedded nanoparticles induce lipid rearrangements similar to those exhibited by biological membrane proteins. *J. Phys. Chem. B* 118, 12586–12598. doi: 10.1021/jp506239p

Van Lehn, R. C., and Alexander-Katz, A. (2019). Energy landscape for the insertion of amphiphilic nanoparticles into lipid membranes: a computational study. *PLoS One* 14:e0209492. doi: 10.1371/journal.pone.0209492

Van Lehn, R. C., Atukorale, P. U., Carney, R. P., Yang, Y. S., Stellacci, F., Irvine, D. J., et al. (2013). Effect of particle diameter and surface composition on the spontaneous fusion of monolayer-protected gold nanoparticles with lipid bilayers. *Nano Lett.* 13, 4060–4067. doi: 10.1021/nl401365n

Vasumathi, V., and Maiti, P. K. (2010). Complexation of siRNA with dendrimer: a molecular modeling approach. *Macromolecules* 43, 8264–8274. doi: 10.1021/ma1012495

Veiseh, O., Tang, B. C., Whitehead, K. A., Anderson, D. G., and Langer, R. (2015). Managing diabetes with nanomedicine: challenges and opportunities. *Nat. Rev. Drug Discov.* 14, 45–57. doi: 10.1038/nrd4477

Venditto, V. J., and Szoka, F. C. (2013). Cancer nanomedicines: so many papers and so few drugs! *Adv. Drug Del. Rev.* 65, 80–88. doi: 10.1016/j.addr.2012.09.038

Vermeer, L. S., de Groot, B. L., Réat, V., Milon, A., and Czaplicki, J. (2007). Acyl chain order parameter profiles in phospholipid bilayers: computation from molecular dynamics simulations and comparison with 2H NMR experiments. *Eur. Biophys. J.* 36, 919–931. doi: 10.1007/s00249-007-0192-9

Vert, M., and Domurado, D. (2000). Poly(ethylene glycol): protein-repulsive or albumin-compatible? *J. Biomater. Sci. Polym. Edn.* 11, 1307–1317. doi: 10.1163/156856200744345

Viegas, T. X., Bentley, M. D., Harris, J. M., Fang, Z., Yoon, K., Dizman, B., et al. (2011). Polyoxazoline: chemistry, properties, and applications in drug delivery. *Bioconjug. Chem.* 22, 976–986. doi: 10.1021/bc200049d

Viitala, L., Pajari, S., Gentile, L., Määttä, J., Gubitosi, M., Deska, J., et al. (2019). Shape and phase transitions in a PEGylated phospholipid system. *Langmuir* 35, 3999–4010. doi: 10.1021/acs.langmuir.8b03829

Vilaseca, P., Dawson, K. A., and Franzese, G. (2013). Understanding and modulating the competitive surface-adsorption of proteins through coarse-grained molecular dynamics simulations. *Soft Matter* 9, 6978–6985. doi: 10.1039/C3SM50220A

Villa, F., Quarto, R., and Tasso, R. (2019). Extracellular vesicles as natural, safe and efficient drug delivery systems. *Pharmaceutics* 11:557. doi: 10.3390/pharmaceutics11110557

Villarreal, O. D., Rodriguez, R. A., Yu, L., and Wambo, T. O. (2016). Molecular dynamics simulations on the effect of size and shape on the interactions between negative Au18(SR)14, Au102(SR)44 and Au144(SR)60 nanoparticles in physiological saline. *Colloids Surf. A Physicochem. Eng. Asp.* 503, 70–78. doi: 10.1016/j.colsurfa.2016.05.038

Vukoviæ, L., Khatib, F. A., Drake, S. P., Madriaga, A., Brandenburg, K. S., Král, P., et al. (2011). Structure and dynamics of highly PEG-ylated sterically stabilized micelles in aqueous media. *J. Am. Chem. Soc.* 133, 13481–13488. doi: 10.1021/ja204043b

Walkey, C. D., and Chan, W. C. W. (2012). Understanding and controlling the interaction of nanomaterials with proteins in a physiological environment. *Chem. Soc. Rev.* 41, 2780–2799. doi: 10.1039/C1CS15233E

Wang, N., Chen, M., and Wang, T. (2019). Liposomes used as a vaccine adjuvant-delivery system: from basicis to clinical immunization. *J. Control Release* 303, 130–150. doi: 10.1016/j.jconrel.2019.04.025

Wang, S., and Dormidontova, E. E. (2012). Selectivity of ligand-receptor interactions between nanoparticles and cell surfaces. *Phys. Rev. Lett.* 109:238102. doi: 10.1103/PhysRevLett.109.238102

Wang, X.-Y., Zhang, L., Wei, X.-H., and Wang, Q. (2013). Molecular dynamics of paclitaxel encapsulated by salicyclic acid-grafted chitosan oligosaccharide aggregates. *Biomaterials* 34, 1843–1851. doi: 10.1016/j.biomaterials.2012.11.024

Wang, Y., Chen, B. Z., Liu, Y. J., Wu, Z. M., and Guo, X. D. (2017). Application of mesoscale simulation to explore the aggregate morphology of pH-senstitive nanoparticles used as the oral drug delivery carriers. *Colloid Surf. B* 151, 280–286. doi: 10.1016/j.colsurfb.2016.12.027

Wang, Z., Gao, J., Ustach, V., Li, C., Sun, S., Hu, S., et al. (2017). Tunable permeability of cross-linked microcapsules from pH-responsive amphiphilic diblock copolymers: a dissipative particle dynamics study. *Langmuir* 33, 7288–7297. doi: 10.1021/acs.langmuir.7b01586

Wang, Y., Li, Q. Y., Liu, X. B., Zhang, C. Y., Wu, Z. M., and Guo, X. D. (2015a). Mesoscopic simulations and experimental studies of pH-sensitive micelles for controlled drug delivery. *ACS Appl. Mater. Interf.* 7, 25592–25600. doi: 10.1021/acsami.5b08366

Wang, Y., Zhu, D. D., Zhou, J., Wang, Q. L., Zhang, C. Y., Liu, Y. J., et al. (2015b). Mesoscopic simulation studies on the formation mechanism of drug loaded polymeric micelles. *Colloid Surf. B* 136, 536–544. doi: 10.1016/j.colsurfb.2015.09.049

Wang, Y., Ren, J. W., Zhang, C. Y., He, M. C., Wu, Z. M., and Guo, X. D. (2016). Compatability studies between an amphiphilic pH-sensitive polymer and hydrophobic drug using multiscale simulations. *RSC Adv.* 6:101323. doi: 10.1039/C6RA15950H

Warren, D. B., King, D., Benameur, H., Poulton, C. W., and Chalmers, D. K. (2013). Glyceride lipid formulations: molecular dynamics modeling of phase behavior during dispersion and molecular interactions between drugs and excipients. *Pharm. Res.* 30, 3238–3253. doi: 10.1007/s11095-013-1206-1

Wen, X.-F., Lan, J.-L., Cai, Z.-Q., Pi, P.-H., Xu, S.-P., Zhang, L., et al. (2014). Dissipative particle dynamics simulation on drug loading/release in polyester-PEG dendrimer. *J. Nanopart. Res.* 16:2403. doi: 10.1007/s11051-014-2403-5

Westergren, J., Lindfors, L., Höglund, T., Lüder, K., Nordholm, S., and Kjellander, R. (2007). In silico prediction of drug solubility: 1. free energy of hydration. *J. Phys. Chem. B* 111, 1872–1882. doi: 10.1021/jp064220w

Wilding, K. M., Smith, A. K., Wilkerson, J. W., Bush, D. B., Knotts, T. A. IV, and Bundy, B. C. (2018). The locational impact of site-specific PEGylation: streamlines screening with cell-free protein expression and coarse-grain simulation. *ACS Synth. Biol.* 7, 510–521. doi: 10.1021/acssynbio.7b00316

Wilkosz, N., Rissanen, S., Cyza, M., Szybka, R., Nowakowska, M., Bunker, A., et al. (2017). Effect of piroxicam on lipid membranes: drug encapsulation and gastric toxicity aspects. *Eur. J. Pharm. Sci.* 100, 116–125. doi: 10.1016/j.ejps.2017.01.007

Wolski, P., Narkiewicz-Michalek, J., Panczyk, M., Pastorin, G., and Panczyk, T. (2017a). Molecular dynamics modeling of the encapsulation and de-encapsulation of the carmustine anticancer drug in the inner volume of a carbon nanotube. *J. Phys Chem. C* 121, 18922–18934. doi: 10.1021/acs.jpcc.7b05229

Wolski, P., Nieszporek, K., and Panczyk, T. (2017b). PEGylated and folic acid functionalized carbon nanotubes as pH controlled carriers of doxorubicin. Molecular dynamics analysis of the stability and drug release mechanism. *Phys. Chem. Chem. Phys.* 19, 9300–9312. doi: 10.1039/C7CP00702G

Wolski, P., Nieszporek, K., and Panczyk, T. (2018). Multimodal, pH sensitive, and magnetically assisted carrier of doxorubicin designed and analyzed by means

of computer simulations. *Langmuir* 34, 2543–2550. doi: 10.1021/acs.langmuir.7b04211

Wolski, P., Nieszporek, K., and Panczyk, T. (2020). Carbon nanotubes and short cytosine-rich telomeric DNA oligomers as platforms for controlled release of doxorubicin-a molecular dynamics study. *Int. J. Mol. Sci.* 21:3619. doi: 10.3390/ijms21103619

Wolski, P., Wojton, P., Nieszporek, K., and Panczyk, T. (2019). Interaction of human telomeric i-motif DNA with single-walled carbon nanotubes: insights from molecular dynamics simulations. *J. Phys. Chem. B* 123, 10343–10353. doi: 10.1021/acs.jpcb.9b07292

Wong-ekkabut, J., and Karttunen, M. (2016). The good, the bad and the user in soft matter simulations. *Biochim. Biophys. Acta Biomembr.* 1858, 2529–2538. doi: 10.1016/j.bbamem.2016.02.004

Wu, J., Zhao, C., Lin, W., Hu, R., Wang, Q., Chen, H., et al. (2014). Binding characteristics between polyethylene glycol (PEG) and proteins in aqueous solution. *J. Mater. Chem. B* 2, 2983–2992. doi: 10.1039/C4TB00253A

Wu, W., Yi, P., Zhang, J., Cheng, Y., Li, Z., Hao, X., et al. (2019). 4/6-hetero-arm and 4/6-mikto-arm star shaped block polymeric drug loaded micelles and their pH-responsive controlled release properties: a dissipative particle dynamics simulation. *Phys. Chem. Chem. Phys.* 21, 15222–15232. doi: 10.1039/C9CP02411E

Wu, Z., Duan, M., Xiong, D., and Zhang, C. Y. (2019). Mesoscale simulations of pH-responsive amphiphilic polymeric micelles for oral drug delivery. *Pharmaceutics* 11:620. doi: 10.3390/pharmaceutics11120620

Xia, Q., Zhu, T., Jiang, Z., Ding, H., and Ma, Y. (2020). Enhancing the targeting ability of nanoparticles via protected copolymers. *Nanoscale* 12, 7804–7813. doi: 10.1039/d0nr01176b

Xiao, T., Li, D., Shi, X., and Shen, M. (2020). PAMAM dendrimer-based nanodevices for nuclear medicine applications. *Macromol. Biosci.* 20:1900282. doi: 10.1002/mabi.201900282

Xiao, X.-F., Jiang, X.-Q., and Zhou, L.-J. (2013). Surface modification of poly ethylene glycol to resist nonspecific adsorption of proteins. *Chin. J. Anal. Chem.* 41, 445–453. doi: 10.1016/S1872-2040(13)60638-6

Xie, X., Xu, S., Pi, P., Cheng, J., Wen, X., Liu, X., et al. (2018). Dissipative particle dynamic simulation on the assembly and release of siRNA/polymer/gold nanoparticles based polyplex. *AIChE J.* 64, 810–821. doi: 10.1002/aic.15961

Xu, D., Smolin, N., Shaw, R. K., Battey, S. R., Tao, A., Huang, Y., et al. (2018). Molecular insights into the improved clinical performance of PEGylated interferon therapeutics: a molecular dynamics perspective. *RSC. Adv.* 8, 2315–2322. doi: 10.1039/c7ra12480e

Xue, Y., O'Mara, M. L., Surawski, P. P. T., Trau, M., and Mark, A. E. (2011). Effect of poly(ethylene glycol) (PEG) spacers on the conformational properties of small peptides: a molecular dynamics study. *Langmuir* 27, 296–303. doi: 10.1021/la103800h

Yahyaei, M., Mehrnejad, F., Naderi-manesh, H., and Rezayan, A. H. (2017). Follicle-stimulating hormone encapsulation in the cholesterol-modified chitosan nanoparticles via molecular dynamics simulations and binding free energy calculations. *Eur. J. Pharm. Sci.* 107, 126–137. doi: 10.1016/j.ejps.2017.07.007

Yamanaka, T., De Nicola, A., Munaò, G., and Soares, T. A. (2019). Effect of the ligand's bulkiness on the shape of functionalized gold nanoparticles in aqueous solutions: a molecular dynamics study. *Chem. Phys. Lett.* 731:136576. doi: 10.1016/j.cplett.2019.07.004

Yang, C., Liu, W., Xiao, J., Yuan, C., Chen, Y., Hangbo Yue, J. G., et al. (2020). pH-sensitive mixed micelles assembled from PDEAEMA-PPEGMA and PCL-PPEGMA for doxorubicin delivery: experimental and DPD simulation study. *Pharmaceutics* 12:170. doi: 10.3390/pharmaceutics12020170

Yang, C., Lu, D., and Liu, Z. (2011). How PEGylation enhances the stability and potency of insulin: a molecular dynamics simulation study. *Biochemistry* 50, 2585–2593. doi: 10.1021/bi101926u

Yang, C., Sun, Y., Zhang, L., Zhu, G., Zhang, C., and Qian, Y. (2012). Dissipative particle dynamics study on aggregation of MPEG-PAE-PLA block polymer micelles loading doxorubicin. *Chin. J. Chem.* 30, 1980–1986. doi: 10.1002/cjoc.201200629

Yang, C., Yuan, C., Liu, W., Guo, J., Feng, D., Yin, X., et al. (2019). DPD studies on mixed micelles self-assembled from MPEG-PDEAEMA and MPEG-PCL for controlled doxorubicin release. *Colloid Surf. B* 178, 56–65. doi: 10.1016/j.colsurfb.2019.02.043

Yang, Y.-L., Sheng, Y.-J., and Tsao, H.-K. (2019). Bilayer membranes of Janus dendrimers with hybrid hydrogenated and fluorinated dendrons: microstructures and coassembly with lipids. *Phys. Chem. Chem. Phys.* 21, 15400–15407. doi: 10.1039/c9cp01635j

Yang, Y., Nie, D., Liu, Y., Yu, M., and Gan, Y. (2019). Advances in particle shape engineering for improved drug delivery. *Drug Discov. Today* 24, 575–583. doi: 10.1016/j.drudis.2018.10.006

Yang, K., and Ma, Y.-Q. (2010). Computer simulation of the translocation of nanoparticles with different shapes across a lipid bilayer. *Nat. Nanotechnol.* 5, 579–583. doi: 10.1038/NNANO.2010.14

Yang, P.-Y., Ju, S.-P., Chuang, Y.-C., and Chen, H.-Y. (2017). Molecular dynamics simulations of PAMAM dendrimer-encapsulated Au nanoparticles of different sizes under different pH conditions. *Comput. Mater. Sci.* 137, 144–152. doi: 10.1016/j.commatsci.2017.05.020

Youn, Y. S., and Bae, Y. H. (2018). Perspectives on the past, present, and future of cancer nanomedicine. *Adv. Drug Deliv. Rev.* 130, 3–11. doi: 10.1016/j.addr.2018.05.008

Zaboli, M., and Raissi, H. (2017). The influence of nicotine on pioglitazone encapsulation into carbon nanotube: the investigation of molecular dynamic and density functional theory. *J. Biomol. Struct. Dyn.* 35, 520–534. doi: 10.1080/07391102.2016.1152565

Zaghmi, A., Mendez-Villuendas, E., Greschner, A. A., Liu, J. Y., de Haan, H. W., and Gauthier, M. A. (2019). Mechanisms of activity loss for a multi-PEGylated protein by experiment and simulation. *Mater. Today Chem.* 12, 121–131. doi: 10.1016/j.mtchem.2018.12.007

Zazo, H., Colino, C. I., and Lanao, J. M. (2016). Current applications of nanoparticles in infectious diseases. *J. Control Release* 224, 86–102. doi: 10.1016/j.jconrel.2016.01.008

Zhadanov, V. P. (2019). Formation of a protein corona around nanoparticles. *Curr. Opin. Colloid Int.* 41, 95–103. doi: 10.1016/j.cocis.2018.12.002

Zhang, L., Peng, G., Li, J., Liang, L., Kong, Z., Wang, H., et al. (2018). Molecular dynamics study on the configuration and arrangement of doxorubicin in carbon nanotubes. *J. Mol. Liq.* 262, 295–301. doi: 10.1016/j.molliq.2018.04.097

Zhang, L., Wang, Z., Lu, Z., Shen, H., Huang, J., Zhao, Q., et al. (2013). PEGylated reduced graphene oxide as a superior ssRNA delivery system. *J. Mater. Chem. B* 1, 749–755. doi: 10.1039/C2TB00096B

Zhang, M., Huang, Y., Hao, D., Ji, Y., and Ouyang, D. (2020). Solvation structure and molecular interactions of ibuprofen with ethanol and water: a theoretical study. *Fluid Phase Equilibr.* 510:112454. doi: 10.1016/j.fluid.2019.112454

Zhang, X., Zheng, L., Luo, M., Shu, C., and Wang, E. (2020). Evaluation of particle shape, size and magnetic field intensity for targeted delivery efficiency and plaque injury in treating atherosclerosis. *Powder Technol.* 366, 63–72. doi: 10.1016/j.powtec.2020.02.003

Zhang, Z., Zhang, Y., Song, S., Yin, L., Sun, D., and Gu, J. (2020). Recent advances in the bioanalytical methods of polyethylene glycols and PEGylated pharmaceuticals. *J. Sep. Sci.* 43, 1978–1997. doi: 10.1002/jssc.201901340

Zhang, W., Sun, J., Liu, Y., Tao, M., Ai, X., Su, X., et al. (2014). PEG-stabilized bilayer nanodisks and carriers of doxorubicin delivery. *Mol. Pharm.* 11, 3279–3290. doi: 10.1021/mp400566a

Zhang, W., Zhao, Q., Deng, J., Hu, Y., Wang, Y., and Ouyang, D. (2017). Big data analysis of global advances in pharmaceutics and drug delivery 1980-2014. *Drug Discov. Today* 22, 1201–1208. doi: 10.1016/j.drudis.2017.05.012

Zhang, Z., Lin, X., and Gu, N. (2017). Effects of temperature and PEG grafting density on the translocation of PEGylated nanoparticles across asymmetric lipid membrane. *Colloid Surf. B* 160, 92–100. doi: 10.1016/j.colsurfb.2017.09.013

Zhang, Y., Tekobo, S., Tu, Y., Zhou, Q., Jin, X., Dergunov, S. A., et al. (2012). Permission to enter cell by shape: nanodisk vs. nanosphere. *ACS Appl. Mater. Interf.* 4, 4099–4105. doi: 10.1021/am300840p

Zhao, L., Seth, A., Wibowo, N., Zhao, C.-X., Mitter, N., Yu, C., et al. (2014). Nanoparticle vaccines. *Vaccine* 32, 327–337. doi: 10.1016/j.vaccine.2013.11.069

Zhao, Q., Miriyala, N., Su, Y., Chen, W., Gao, X., Shao, L., et al. (2018). Computer-aided formulation design for a highly soluble lutein-cyclodextrin multiple-component delivery system.

Mol. Pharm. 15, 1664–1674. doi: 10.1021/acs.molpharmaceut.8b00056

Zheng, L. S., Yang, Y. Q., Guo, X. D., Sun, Y., Qian, Y., and Zhang, L. J. (2011). Mesoscopic simulations on the aggregation behavior of pH-responsive polymeric micelles for drug delivery. *J. Colloid Interf. Sci.* 363, 114–121. doi: 10.1016/j.jcis.2011.07.040

Zhu, X., Radovic-Moreno, A. F., Wu, J., Langer, R., and Shi, J. (2014). Nanomedicine in the management of microbial infection - overview and perspectives. *Nano Today* 9, 478–498. doi: 10.1016/j.nantod.2014.06.003

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

METHODS
published: 14 December 2020
doi: 10.3389/fmolb.2020.599059

Hierarchical Graph Representation of Pharmacophore Models

Garon Arthur[1], Wieder Oliver[1], Bareis Klaus[1], Seidel Thomas[1], Ibis Gökhan[2], Bryant Sharon[2], Theret Isabelle[3], Ducrot Pierre[1,3] and Langer Thierry[1]*

[1] Department of Pharmaceutical Chemistry, University of Vienna, Vienna, Austria, [2] Inte:Ligand Software-Entwicklungs und Consulting GmbH, Vienna, Austria, [3] Institut de Recherches Servier (IdRS), Croissy-sur-Seine, France

For the investigation of protein-ligand interaction patterns, the current accessibility of a wide variety of sampling methods allows quick access to large-scale data. The main example is the intensive use of molecular dynamics simulations applied to crystallographic structures which provide dynamic information on the binding interactions in protein-ligand complexes. Chemical feature interaction based pharmacophore models extracted from these simulations, were recently used with consensus scoring approaches to identify potentially active molecules. While this approach is rapid and can be fully automated for virtual screening, additional relevant information from such simulations is still opaque and so far the full potential has not been entirely exploited. To address these aspects, we developed the hierarchical graph representation of pharmacophore models (HGPM). This single graph representation enables an intuitive observation of numerous pharmacophore models from long MD trajectories and further emphasizes their relationship and feature hierarchy. The resulting interactive depiction provides an easy-to-apprehend tool for the selection of sets of pharmacophores as well as visual support for analysis of pharmacophore feature composition and virtual screening results. Furthermore, the representation can be adapted to include information involving interactions between the same protein and multiple different ligands. Herein, we describe the generation, visualization and use of HGPMs generated from MD simulations of two x-ray crystallographic derived structures of the human glucokinase protein in complex with allosteric activators. The results demonstrate that a large number of pharmacophores and their relationships can be visualized in an interactive, efficient manner, unique binding modes identified and a combination of models derived from long MD simulations can be strategically prioritized for VS campaigns.

Keywords: pharmacophore modeling, protein structure, clustering, human glucokinase, hierarchical graph representation, protein ligand binding, molecular dynamic (MD) simulation, virtual screening

OPEN ACCESS

Edited by:
Pratyush Tiwary, University of Maryland, College Park, United States

Reviewed by:
Matteo Salvalaglio, University College London, United Kingdom
Arianna Fornili, Queen Mary University of London, United Kingdom

****Correspondence:***
Seidel Thomas
thomas.seidel@univie.ac.at

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

Received: *26 August 2020*
Accepted: *24 November 2020*
Published: *14 December 2020*

Citation:
Arthur G, Oliver W, Klaus B, Thomas S, Gökhan I, Sharon B, Isabelle T, Pierre D and Thierry L (2020) Hierarchical Graph Representation of Pharmacophore Models. Front. Mol. Biosci. 7:599059. doi: 10.3389/fmolb.2020.599059

INTRODUCTION

Understanding the biomolecular recognition of ligands and their interactions with macromolecular targets is of utmost importance for the successful discovery of novel biologically active compounds (Fenwick et al., 2011). One way to approach this problem in drug design is the modeling of ligand-target interactions as pharmacophores. Pharmacophores are defined as an ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target and to trigger (or block) its biological response (Wermuth et al., 1998).

In general, pharmacophore models are either derived from ligand-target complexes (structure-based) and/or a set of known active molecules (ligand-based) and can then be used as queries for an *in silico* virtual screening (VS) to find compounds with similar stereoelectronic features (Langer, 2010; Leach et al., 2010; Schuster, 2010). One limitation of structure-based (SB) modeling is that all possible interactions between a target-ligand complex may not be captured since they are derived from static representations. The fact that proteins are flexible structures and interactions with ligands are inherently dynamic is well-known and remains to be an important problem with emerging *in silico* solutions in various contexts (Cozzini et al., 2008; Boehr et al., 2009). Molecular dynamics (MD) simulations have recently been used to sample possible protein conformations (Durrant and McCammon, 2011; De Vivo et al., 2016; Liu et al., 2018) which were then used to derive multiple pharmacophore models from an initially static crystallographic structure. Choudhury et al. (2015) generated 3-D pharmacophore models from each snapshot of a MD simulation and selected the best performing model after docking and VS rescoring. The selection of a single "best performing" pharmacophore model was also pursued by means of clustering (Sohn et al., 2013; Spyrakis et al., 2015), providing better VS results than "classical" x-ray crystallographic derived structure-based pharmacophore models. However, to determine the "best performing" model requires datasets of known active and inactive compounds to assess the performance of the models. In cases with new targets during early hit finding stages, this information may be yet not be available and prioritizing pharmacophore models for VS campaigns can be challenging.

To overcome the need to select one unique representative set of pharmacophore models, Wieder et al. (2017) developed the "Common Hits Approach" (CHA) in which multiple 3D pharmacophore models derived from a MD simulation were partitioned according to their feature compositions and used for subsequent VS runs. A single final hit-list was obtained using a consensus scoring function to rank and combine the screening results which were originally obtained for each unique model enabling a prioritization of virtual hits based on a set of MD derived models. Recently, Polishchuk et al. (2019) improved the workflow by adapting the consensus scoring function to consider the number of conformations of each molecule retrieved by the VS runs. Based on these studies, Madzhidov et al. (2020) analyzed the performance of a set of pharmacophore models and developed a probabilistic approach for consensus scoring, leading to a method which is less sensitive to the poor performing models in the pool. Although these consensus-based approaches provided better results than a "classical" pharmacophore approach, they demanded considerable computational resources due to the required multiple VS runs.

Nowadays, MD simulations allow for a thorough sampling of the conformational space—even of large biological systems and the generation of structure-based pharmacophore models is no longer limited to single crystallographic structures. As a direct result of the improvement of modern hardware, most computation laboratories can now perform MD simulations at the nanosecond scale in a few hours. However, performing consecutive VS runs on very large libraries (millions of compounds) is still a crucial time limiting factor. To address this issue, this paper presents a hierarchical graph representation of pharmacophore models "HGPM," which aims at the easy to comprehend visualization of pharmacophore model related information and thus can greatly aid in the prioritization and selection of pharmacophore models for subsequent processing steps. While previous works reduced the number of pharmacophore models by clustering crystallographic structures or 3D pharmacophore information, this graph representation focuses on the view of hierarchical pharmacophore feature information to support the users in the model selection process in order to reduce the number of models for ensuing VS runs. A single representation of multiple pharmacophore models, for example, derived from an MD simulation, has several advantages: (i) The introduction of an easy to comprehend graph-based view of all unique models and their relationship, that were observed (Maggiora and Bajorath, 2014; Métivier et al., 2018). (ii) A simpler, less error-prone selection process of 3D pharmacophore models especially for long MD simulations for virtual screening runs. (iii) The possibility to expand the displayed information by the addition of models generated from other systems or MD simulations.

The following sections will focus on the algorithmic details and the computational procedure for the generation of the hierarchical graph representation of pharmacophore models. Furthermore, the methodology will be demonstrated and discussed in the context of the human hexokinase IV as a case study, illustrating how pharmacophore information derived from MD simulations can be displayed and put to good use with this approach.

MATERIALS AND METHODS

Protein–Ligand Complex Preparation

Two crystal structures of the human glucokinase in complex with activators were downloaded from the RCSB PDB databank (Berman, 2000), with PDB IDs 1v4s (Kamata et al., 2004) and 4no7 (Petit et al., 2011). The sequences of the proteins were aligned and the amino acid subsequently renumbered, using the RCSB PDB comparison tool (Prlić et al., 2010) and the jFATCAT_flexible algorithm (Ye and Godzik, 2003). Amino acids 92–99 were not present in the 4no7 complex. Since they did not impact the protein stability during the simulations and no interactions with the ligand in the 1v4s system could be observed, they were not modeled. A table containing the alignment block is available in **Supplementary Figure 1**. The Maestro software (Schrodinger, 2010) was used to remove water molecules, add hydrogens and minimize the structures. The capping of the termini, the solvation and the addition of ions for the protein complexes had been set up through the CHARM-GUI web interface (Jo et al., 2008). Information about the prepared protein-ligand complexes is available in **Supplementary Table 1**.

Molecular Dynamics Simulations

MD simulations were carried out using Amber 16 (Case et al., 2016). Parameters for the ligands were generated by tleap using the general AMBER force field (GAFF) (Wang et al., 2004). The MD simulation protocol started with an equilibration and thermalization phase of 125 ps with a 1 fs time step. Then each system was simulated for a total of 300 ns composed of 3 replicates of 100 ns with different initial velocities and using Langevin dynamics at a temperature of 303.15 K. The pressure was kept around 1 atm by a Monte Carlo barostat. The SHAKE algorithm (Ryckaert et al., 1977) was used to keep all bonds involving hydrogen atoms rigid. The time step of the production runs was set to 2 fs. Plots of the root-mean-square deviations for the proteins and their ligands are shown in **Supplementary Figure 2**.

Library Generation

Compounds with experimental activities measured on human glucokinase were taken from the ChEMBL database (Gaulton et al., 2017). In total, 756 unique molecules with activity toward the target protein expressed in EC50 were extracted. This set was split based on the activity value threshold of 1.5 μM, resulting in 601 molecules labeled as actives and 155 as decoys. The KNIME Analytics platform (Berthold et al., 2009) was used in combination with the InteLigand Expert KNIME LigandScout Diversity Picker node (InteLigand Expert KNIME Extensions) to extract the 20,000 diverse molecules from the ChEMBL library based on extended connectivity fingerprint (ECFP) similarity, also labeled as decoys. Finally, a library for virtual screening was calculated using the idbgen algorithm in LigandScout 4.4 Expert (LigandScout 4.4 Expert). The procedure included the generation of a maximum of 25 conformations for each of the 20,756 molecules using the icon Fast settings (Poli et al., 2018). The active molecules were clustered in 5 groups based on ECFP similarity. Examples of the molecules present in each cluster is shown in **Supplementary Table 2**. The ligands from the x-ray derived structures PDB codes 1v4s and 4no7 were in cluster numbers 4, and 2, respectively.

Pharmacophore Generation and Virtual Screening

Structure-based pharmacophore models were generated for each frame output from the MD simulations using LigandScout 4.4 Expert (Wolber and Langer, 2005). Models generated by LigandScout support the following chemical feature types: hydrophobic interactions, hydrogen bonds donor/acceptor, positive/negative ionizable area, aromatic ring and halogen bond donor features. In addition, pharmacophore models from the x-ray derived crystallographic structures of 1v4s and 4no7 were created. Water molecules were discarded before the generation of the models. The LigandScout activity profiling KNIME node was used to perform all virtual screening runs of models against the dedicated database of 20,756 molecules. Receiver operating characteristic (ROC) curves were generated for the virtual screening runs and the performance of the models was assessed by the calculation of area under the curve (AUC) values at specific percentages of the number of screened database molecules.

Hierarchical Graph of Pharmacophore Models Generation

Feature Vectors and Graph Nodes

The pharmacophore models derived from the MD simulations were transformed into feature vectors in a related manner as described by the paper of Wieder et al. (2017). Each element of the vector represents a unique pharmacophore feature observed in the system. For this study, pharmacophore features are considered unique if they differ in any of the following components—pharmacophore feature type, ligand identifier, and/or identifier of the interacting environment residue(s). The 3D information is not taken into account in the identification of the unique pharmacophore features. Thus, e.g., a hydrogen bond acceptor feature generated for a ligand nitrogen atom that interacts with Serine will be considered as being different from a corresponding feature which represents the same nitrogen atom interacting with Threonine regardless of their shared feature type or their 3D position. Both pharmacophore features will be considered as being unique. In that, unique features are specific to the set of pharmacophores they were created from, as are the feature vectors. Feature vectors are represented as bit-strings describing the composition of the 3D pharmacophore models: a value of 0 simply denotes that the considered unique feature is absent in the model, and a value of 1 that the feature is present. A bit-string representation allows quick filtering of pharmacophore models with similar feature sets and furthermore enables a fast calculation of feature appearance counts during the simulation. Each node in the hierarchical graph representation is associated with a unique feature vector and contains additional derived information such as related frame number(s), appearance count, and linked pharmacophore models. **Figure 1** shows the feature vector generation process for a set of pharmacophore models and their association to the graph nodes. To limit noise in the initial set of pharmacophore models, unique feature vectors are filtered according to their appearance count. The pharmacophore models were filtered to keep the models which appear at least 2 times as in the paper of Wieder et al. (2017), or 0.001 times the number of initial frames. The first pharmacophore models observed during the MD simulation for every unique feature vector in the hierarchical graphs were considered for VS.

Hierarchical Linkage

The hierarchical linkage of the graph is based on the unique pharmacophore feature composition of the feature vector in each node. Links are created between nodes if their feature vectors are a subset or a superset of each other. **Figure 2** depicts the linkage process. If two feature vectors do not exhibit a subset or superset relation, a new feature vector is temporarily created. This new feature vector represents the intersection set of the unique features for the two considered nodes. If this temporary feature vector is identical to an already existing node, the two considered nodes are linked to this one. If the temporary node is unique a new permanent node is created. The creation of "Artificial" feature vectors associated with a new node has been implemented to allow the generation of a unique hierarchical graph. Therefore, the "Observed" or "Artificial" nature of the

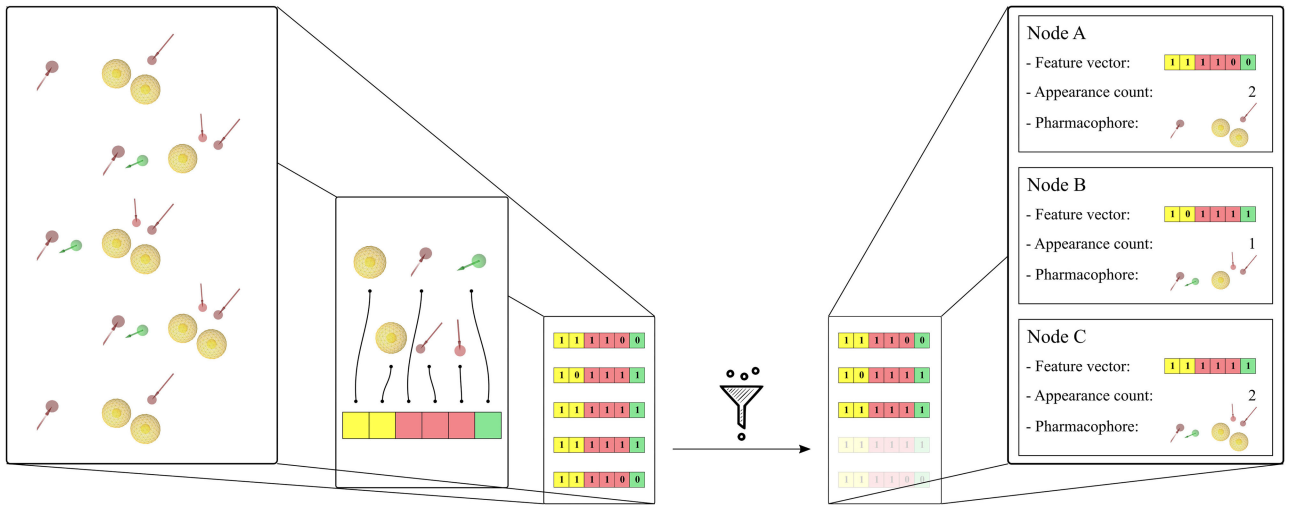


FIGURE 1 | Generation of the feature vectors and their node representation from an initial set of 5 pharmacophore models. The set of pharmacophore models was converted into a corresponding set of feature vectors by first identifying all encountered unique features. Then the feature vector elements are initialized with 1 or 0 depending on the presence or absence of the corresponding unique feature in the pharmacophore model. A filtering step is done in order to remove duplicates with identical feature vectors. Finally, the graph nodes are created for each unique feature vector and the corresponding pharmacophore models and appearance count values are stored. The pharmacophore model features are: yellow spheres (hydrophobic), red and green arrows (hydrogen-bond acceptors and donors, respectively). The corresponding vector features are colored yellow, red, and green accordingly.

feature vectors is stored as an attribute of each graph node in the form of its appearance count. Once all subset and superset links are generated, the redundant paths are removed.

Visualization

For an easy comprehension of the information contained in the graph nodes and their links, the visualization plays an important role. From the feature vectors, information about the composition of the pharmacophore models and the hierarchical links between them is already present. Several visual parameters can be used to depict additional graph properties. For this publication the following properties have been chosen, unless otherwise indicated:

- The appearance count of the pharmacophore models is represented by the size of the nodes. Therefore, the higher the appearance count, the larger visual representation of the node.
- The "Observed" or "Artificial" nature of the node is represented by its color. Blue is used to depict "Observed" nodes, and orange for "Artificial" nodes.
- The specificity of the pharmacophore models is represented by organizing the nodes in the x axis based on the number of unique pharmacophore features of which they are composed. Considering a node, each other node in the same column has the same number of pharmacophore features, each node on its left side is composed of fewer features and every node on its right is composed of more features.
- The representation of pharmacophore model similarity is achieved by dimension reduction using a Multidimensional Scaling Method (MDS) (Mead, 1992; Borg and Groenen, 2003). This method places all the elements of a distance matrix in a single dimension, preserving the distance between nodes as much as possible. The distance matrix is obtained by calculating the Manhattan distance between the feature vectors of the nodes. The similarity between the pharmacophore models is then represented by the relative distance between the nodes projected on the vertical axis of the graph. The reliability of this process is visualized by displaying the proportion of variance of the scaled data.

Figure 3 depicts an example of the graph representation.

RESULTS AND DISCUSSION

Case Study: Glucokinase

The hexokinase IV, or glucokinase (GK) is an isoenzyme responsible for glucose phosphorylation (Beck and Miller, 2013). The concentration of glucose in the plasma determines the conformational switch of GK between its active and inactive states. The glucose level impact on GK activity makes this enzyme act as a sensor responsible for the glucose homeostasis in the human body (Bell and Polonsky, 2001). Therefore, GK has been a primary target for the development of antidiabetic drugs (Kamata et al., 2004; Osbak et al., 2009; Petit et al., 2011).

Two crystallographic structures of the active conformation of GK with bound activators have been selected for this study (PDB codes: 1v4s and 4no7). The structures of the ligands, their position in the binding pocket and the pharmacophore features derived from the independent x-ray experiments are depicted in **Figure 4**. The binding poses of the two ligands show similarities with respect to pharmacophoric hydrogen bond donor and acceptor features capable of forming interactions with the backbone of Arginine 63. As it has been observed in previous studies (Petit et al., 2011), the loop comprising the residues

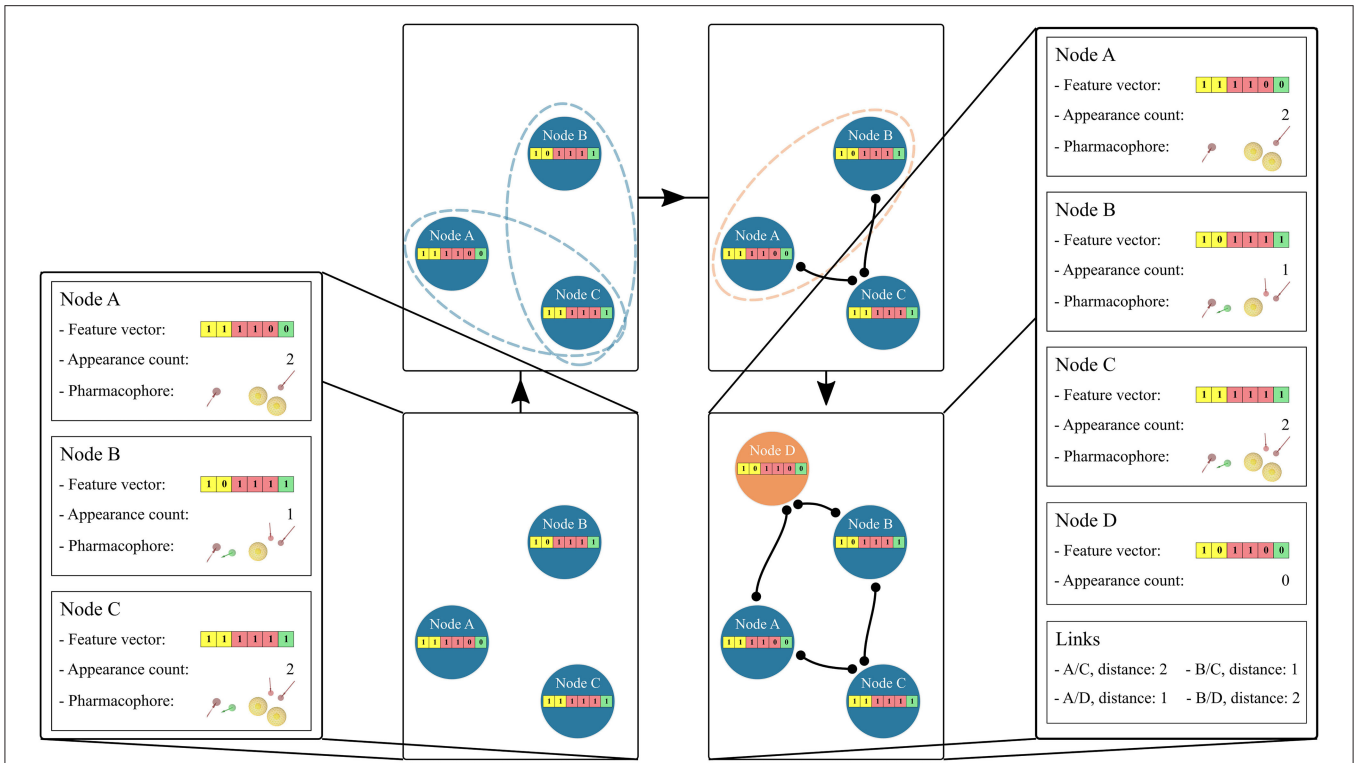


FIGURE 2 | Hierarchical linkage of the graph nodes. Starting from 3 "Observed" nodes in blue, the stored feature vectors are tested for subset/superset relations, represented as blue dotted ellipses. Edges are then created if the relation is found. In the case that two nodes do not depict this relation, represented as an orange ellipse, a new node is created and linked to them. The appearance count for this "Artificial" node is set to 0 and its color is changed to orange. The pharmacophore model features are: yellow spheres (hydrophobic), red and green arrows (hydrogen-bond acceptors and donors, respectively). The corresponding vector features are colored yellow, red, and green accordingly.

92–102 is poorly ordered, which results in the opening of an allosteric sub-pocket that can accommodate the chloro-phenyl-methane-sulfonate of the ligand in the 4no7 protein data bank (PDB) structure

Hierarchical Graphs of Pharmacophore Models

Three MD simulation runs each of 100 ns were performed for both protein-ligand complexes. From the MD simulation trajectories obtained, 10,000 frames were extracted and subsequently used for the generation of pharmacophore models as described in the Methods part. The hierarchical graphs were then generated from the frame-based pharmacophore models including also the crystallographic structure-based pharmacophore models. The pharmacophore models were filtered according to their appearance count before subjecting them to the hierarchical graph generation procedure. The graphs were generated for each individual run as well as for the reunification of all runs for each crystallographic structure. **Table 1** summarizes the composition of the graphs. In a previous publication (Wieder et al., 2017), we used a filtering criteria on the unique pharmacophore model appearance count in order to reduce the noise by removing pharmacophore models which appeared only once during the simulation. In this study, we investigated the impact of several values for the filtering criteria, discarding models appearing <2 frames up to <1% of the number of frames, in order to both reduce the noise and improve the readability of the graph. Source code for the generation and visualization of the graphs is available online (Source code for the HGPM Implementation, 2020). Source code for the generation and processing of LigandScout pharmacophore models has been excluded due to intellectual property reasons. Output data for an interactive demonstration of the Hierarchical Graph representation of Pharmacophore Models generated from the first MD run of 4no7 can also be accessed online (Demonstration of HGPM, 2020). A listing of all unique features for this system can be seen in the online demonstration when hovering the mouse over the feature vector.

Analysis of the Unique Feature Vectors

To visualize the composition of the pharmacophore models that were obtained for each performed MD simulation of the systems, the individual partitioning of the unique observed features is represented as Venn diagrams shown in **Figure 5**. For the 1v4s system, 37 out of 81 (46%) of all unique observed features are present in every of the 3 MDs runs if the appearance count filtering criteria is set to 2 frames or more. This ratio stays similar when the appearance count criteria is set to 10 frames or more,

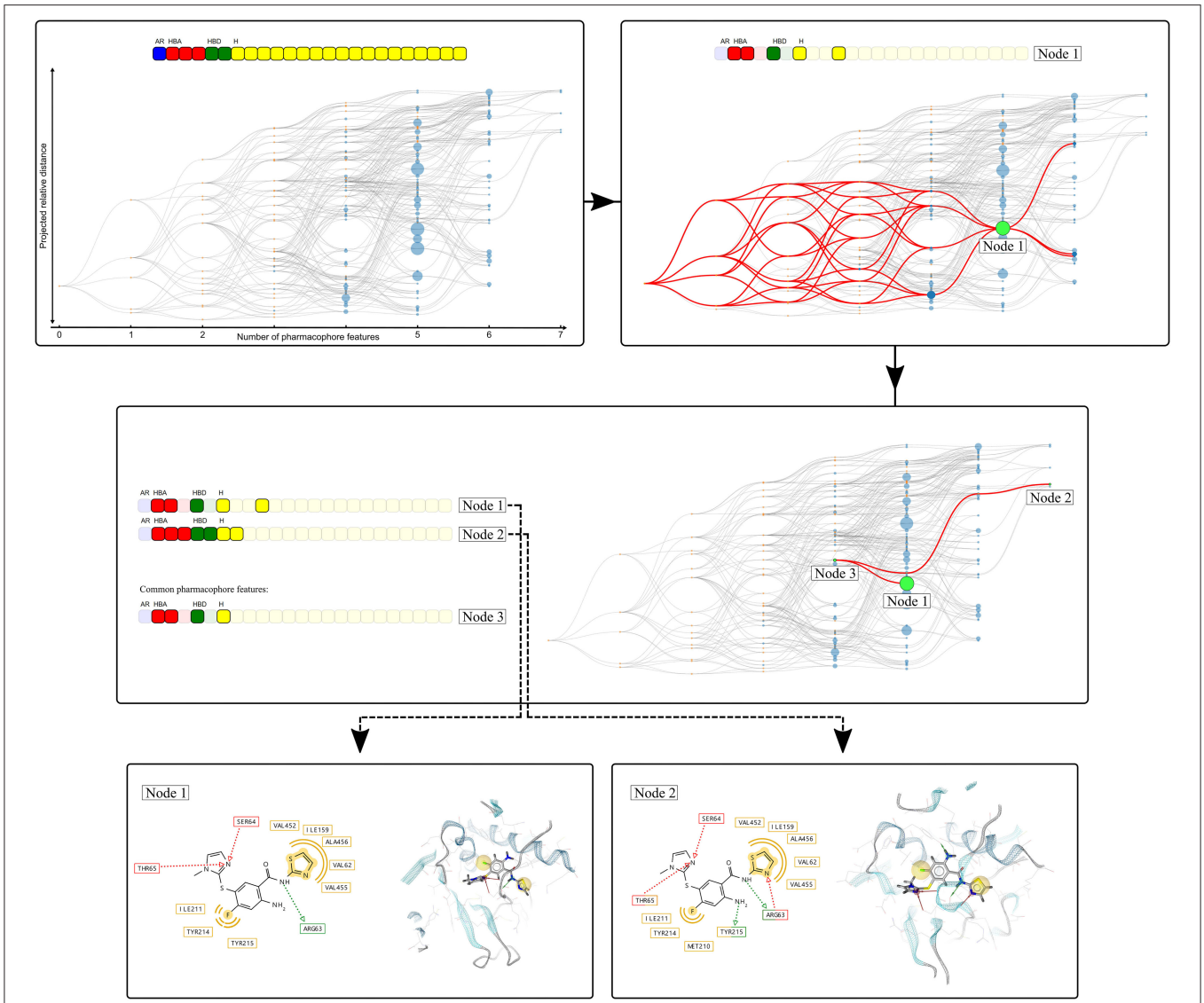


FIGURE 3 | Visualization of the Hierarchical graph representation of pharmacophores models derived from a molecular dynamics simulation of human glucokinase in complex with an activator (PDB code: 1v4s). The feature vector is represented on the top of the graph and each box represents a unique pharmacophore feature. The color of the boxes indicates the type of the corresponding feature: yellow (hydrophobic), red and green (hydrogen-bond acceptors and donors, respectively), and blue (aromatic). The hierarchical graph below the feature vector represents all pharmacophore models observed during the simulation. Nodes are linked by hierarchical relations and their color denotes their origin: blue ("Observed" pharmacophore models), or orange ("Artificial" models only composed of a subset of features from the "Observed" pharmacophores). The graph is interactive, nodes can be selected to depict all related pharmacophore models, as shown with the selection of node 1. When two nodes are selected, the node that depicts the pharmacophore feature intersection set is also highlighted (Node 3), as depicted with the selection of Nodes 1 and 2. For each node, the associated pharmacophore model can be easily retrieved.

resulting in 18 out of 39 (46%) unique observed features being in common. Although approximately half of the unique features are always observed, each individual run led to the observation of 5–17 unique features which are only present in this specific run (filtering criteria of 2 frames or more). When the filtering criteria is set up to 10, the number of unique features solely observed in a specific run decreases, but the ratio stays similar.

Considering the 4no7 system, the trend slightly differs with an increase in the ratio of unique features in common to the 3 runs. We observed 29 out of 50 (58%) unique common features with the appearance count filtering criteria set to 2 frames or more and 21 out of 29 (72%) with the filtering criteria set to 10 or more. Consequently, the number of unique features solely observed in single runs decreased between 5 and 7 with the filtering criteria set to 2 and 0–3 with the criteria set to 10. While several unique features were always present regardless of the run, each individual MD simulation provided exclusive information. Every further analysis presented in this work was made considering every pharmacophore model obtained from any of the 3 runs performed for each system.

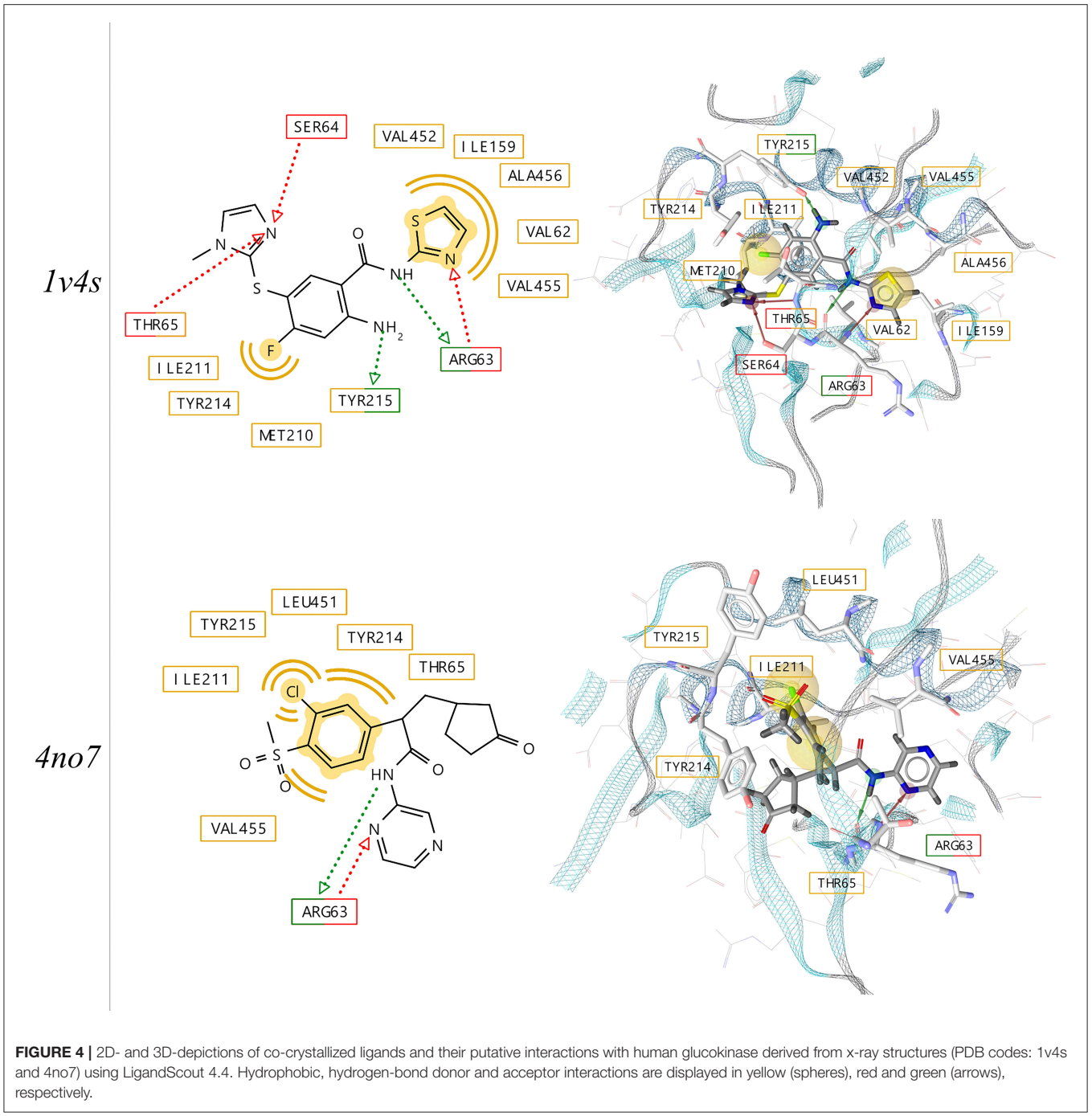


FIGURE 4 | 2D- and 3D-depictions of co-crystallized ligands and their putative interactions with human glucokinase derived from x-ray structures (PDB codes: 1v4s and 4no7) using LigandScout 4.4. Hydrophobic, hydrogen-bond donor and acceptor interactions are displayed in yellow (spheres), red and green (arrows), respectively.

The filtering criteria applied on the appearance count have been set to decrease the complexity of the hierarchical graphs in terms of number of nodes and links, as can be seen in **Table 1**. To investigate the impact of the filtering criterion on the feature vector composition, **Figure 6** depicts details of the nature of the unique pharmacophore features. Among all unique pharmacophore features observed in the MD simulations for 1v4s and 4no7 62 out of 81 and 31 out of 50 are hydrophobic interactions. This high number of hydrophobic features is partly caused by the nature of the feature serial generation algorithm. For example, a hydrophobic feature that involves the fluorine atom of the ligand in 1v4s and amino acids ILE211, TYR214, TYR215 will be considered different from another hydrophobic feature involving the same fluorine atom on the ligand, but different amino acids like THR65, MET210, ILE211, TYR214, TYR215, even though every amino acid of the first feature is also

TABLE 1 | Results from the HGPMs obtained for each MD simulation providing the graphs node composition, pharmacophore filtering criterion, variance of the MDS projection, and the generation time.

System	Run(s)	Number of pharmacophore models	Minimum appearance count of the pharmacophore model	Number of unique features	Number of "observed" nodes	Number of "artificial" nodes	Total number of nodes	Variance of the projection (%)	Time to generate (s)
1v4s	1	10,001	2	63	515	478	993	11.5	30
			10	24	135	100	235	15.1	19
	2		2	56	457	464	921	12.5	26
			10	27	145	119	264	15.5	19
	3		2	52	549	766	1,315	13.4	50
			10	28	160	243	403	14.5	20
	1,2,3	30,001	2	81	1,163	1,207	2,370	11.7	258
			10	41	346	337	683	13.3	54
			30	28	174	186	360	14.9	51
4no7	1	10,001	2	37	814	680	1,494	12.7	63
			10	26	177	139	316	17.1	18
	2		2	36	813	645	1,458	10.7	51
			10	23	172	133	305	14.0	17
	3		2	39	852	728	1,580	10.2	60
			10	25	191	150	341	13.4	21
	1,2,3	30,001	2	50	1,807	1,394	3,201	10.3	511
			10	29	461	285	746	13.5	51
			30	26	172	130	302	16.2	50

involved in the second one. This is inherent in the LigandScout definition of hydrophobic features and has been kept as is.

However, setting the appearance count filtering criterion to 10 frames or more tends to discard more hydrophobic features than any other type of pharmacophore interaction, as it can be observed for both systems shown in **Figure 6**. **Table 1** additionally provides information about hierarchical graphs with an appearance count set to 1% of the number of frames. For graph readability reasons, we chose to use a filtering criterion of 10 for the rest of this study.

Uses and Analysis of the Hierarchical Graphs

The hierarchical graphs of both investigated systems are shown in **Figure 7**. A detailed description of the information which can be visually retrieved from these graphs is given in detail in section Visualization.

For system 1v4s, the graph spans nine columns and thus includes models with up to 8 pharmacophore features. Several nodes are converging close to the pharmacophore model obtained from the crystallographic structure, labeled as PDB. However, no other pharmacophore models were observed that comprise a superset of the features represented by the PDB node. The node with the highest appearance count (HF node) among the 3 runs is vertically displaced from the PDB node and while they share 3 features, they are not directly related in terms of feature hierarchy as shown in more detail in **Supplementary Figure 3**. The node associated with the hydrogen bond donor and acceptor features that interact with Arginine 63 (CF node) is located on the third column of the graph, since it possesses a low number of features and is therefore not specialized. The PDB and HF nodes were used to perform VS runs against the active and decoy database as described in the Methods section. In addition, a selection of all pharmacophore models which represent a superset of the HF and CF nodes were used for a consensus screening run. The CHA (Wieder et al., 2017), that is a consensus approach considering all observed pharmacophores was also carried out. A summary of the VS results can be found in **Table 2**.

We observed that the PDB pharmacophore for 1v4s did not retrieve hits during the VS, which might be due to the number of defined and specific directional hydrogen bond vectors defining its specific binding mode, with 5 of the 7 pharmacophore features containing defined directions for hydrogen bonding. The 4 other pharmacophore selections for this system performed well, with area under the curve (AUC) values above 0.96 at 1% of the number of database molecules. The HF pharmacophore model performed as well as the other models regarding the AUC at 1% but slightly below for the other and retrieved 65 hits out of 20,756 molecules. The selection of all superset models of the CF node provided the most stable and best results regarding the AUC, even outperforming the CHA. Those good results can be linked to the high amount of pharmacophore models which were used for screening, 236 for the CF+ selection and 346 for the CHA. Lastly, the set of all pharmacophore models which are supersets of the

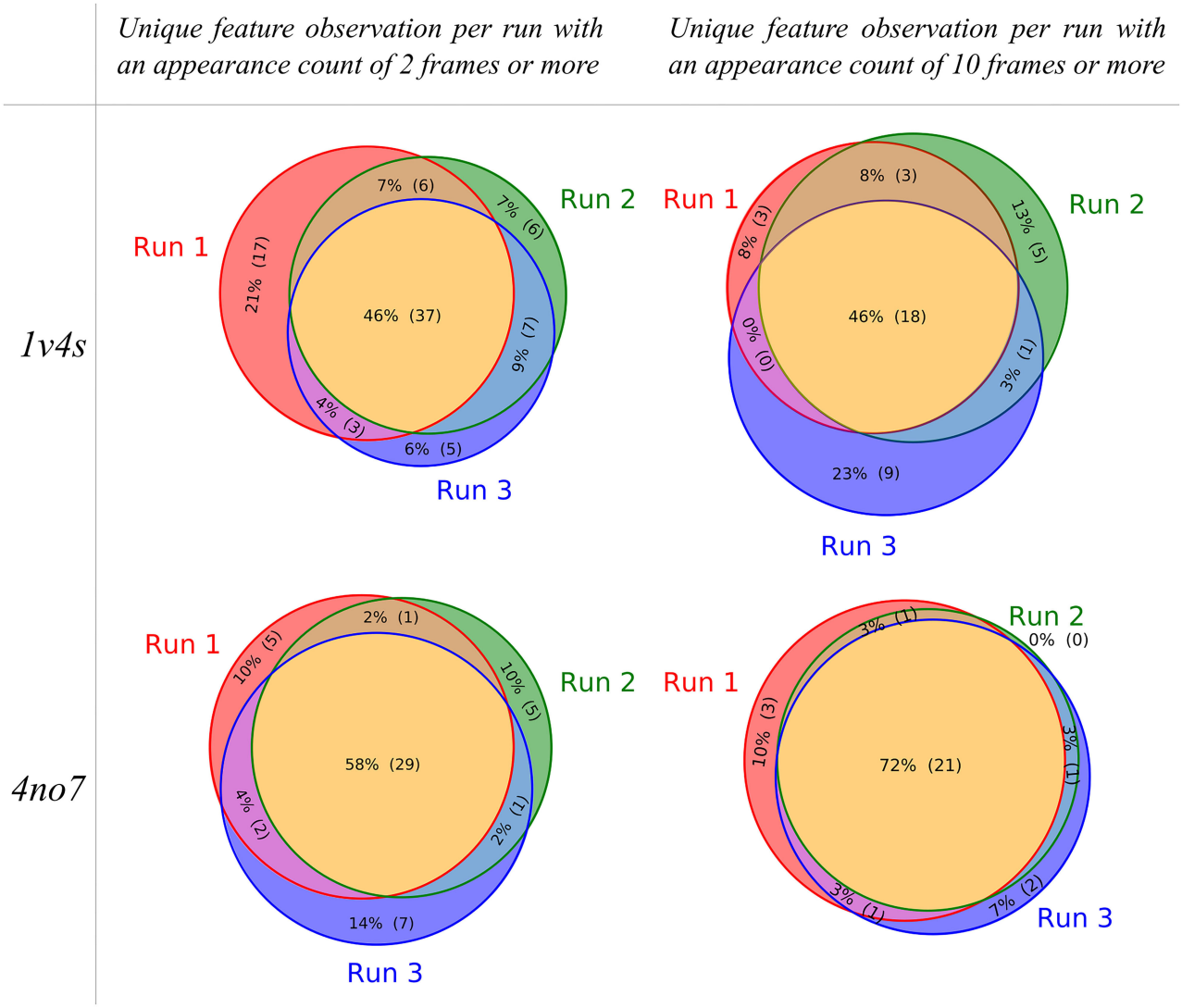


FIGURE 5 | Venn diagrams showing the unique pharmacophore feature similarities between the MD simulation runs performed for the 1v4s and 4no7 systems, respectively. The pharmacophore models considered for the sampling of the unique features were filtered based on an appearance count criteria, set to 2 for the diagrams on the left and 10 on the right.

HF nodes performed close to the HF node alone but retrieved 87 molecules.

The overall hierarchical graph of pharmacophore models of the 4no7 system differs from the one of the 1v4s system mainly by the two branches involving the nodes with the highest number of features on the right side of the graph. Additionally, the PDB, HF, and CF nodes are vertically quite close and are hierarchically linked as it can be seen in **Supplementary Figure 4**. This leads to the conclusion that these three nodes are part of only one of the two branches among the most specialized pharmacophore models. Therefore, an additional node has been selected to involve an arbitrary model of the second branch, which has been labeled as "Selection" in **Figure 7**. **Table 2** shows the VS results obtained with the PDB and HF nodes. The CHA was applied and additional consensus scoring runs were made considering the pharmacophore supersets of the PDB, HF, CF and custom selected nodes. For 4no7, the PDB pharmacophore obtained the best AUC results at 1% of the database and retrieved 401 molecules. The CHA performed as good as the PDB node regarding the AUC at 1% and showed its stability by delivering comparable AUC values at higher percentages than the other approaches while retrieving 13,868 molecules. The HF node model did not perform as well as other models with an AUC value of 0.81 at 1% that fell to 0.41 at 5%. The selection of all superset models of the HF node, however, achieved significantly better VS results than the HF node alone. Lastly, the selection based on the "Arbitrary" node delivered high AUC values with the best observed results at all thresholds superior to 1%. It is interesting

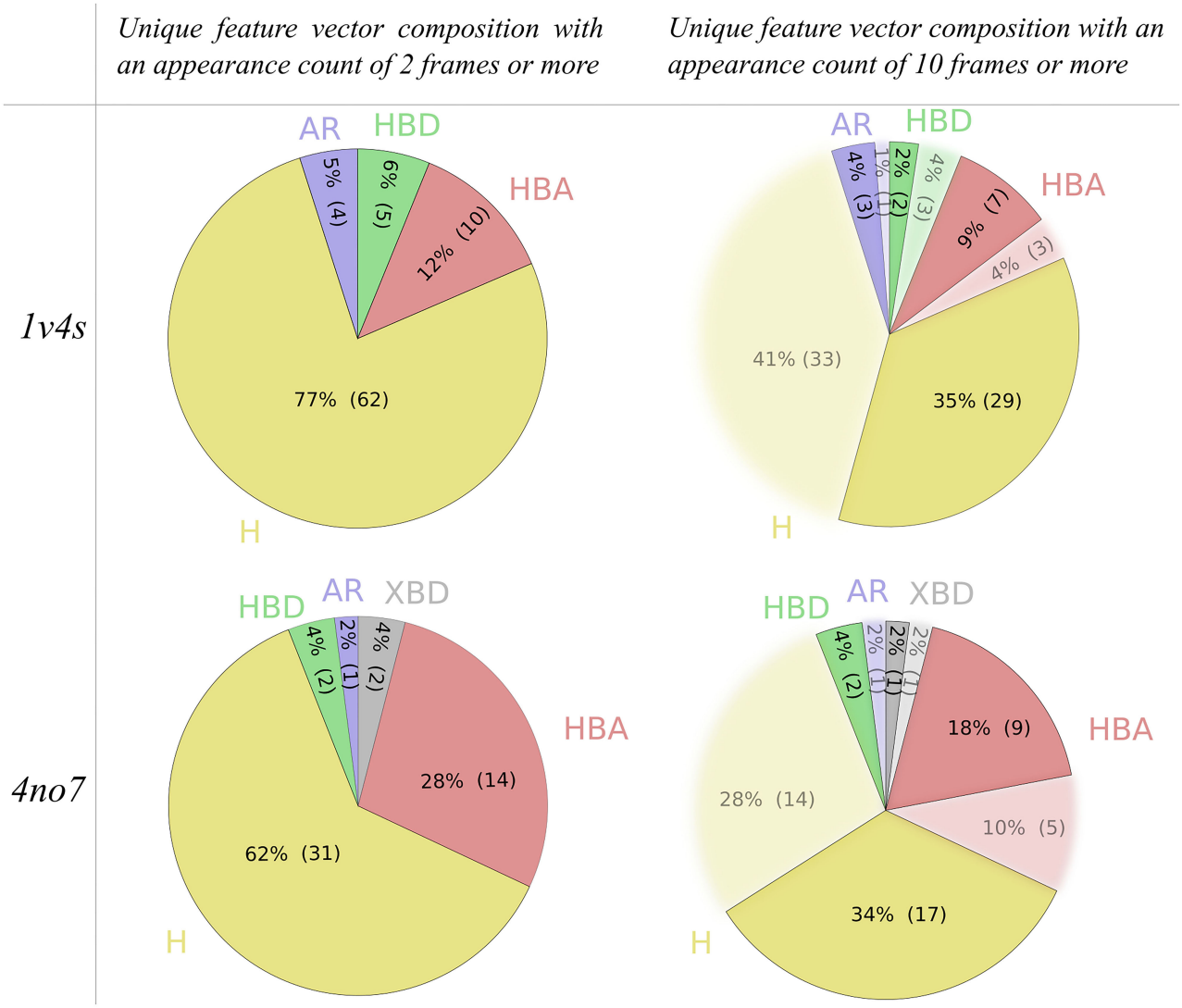


FIGURE 6 | Pie charts of the unique pharmacophore feature composition of the feature vectors generated from all 3 MD simulations, based on the applied appearance count filtering criterion for both the 1v4s and 4no7 systems. Hydrophobic interactions are colored yellow, aromatic interactions blue, hydrogen bond donors green, hydrogen bond acceptors red and the halogen bond donors gray.

to point out that the "Arbitrary" selected node and the HF node are of disparate feature composition, but still deliver high AUC values. To scrutinize this observation, we looked at the structural clusters of the active molecules retrieved by these two approaches, as detailed in **Supplementary Figure 5**. We observed that for a hitlist truncated at 2,076 molecules (10% of the database), the HF node retrieves a consequent number of molecules from all 5 clusters, when the arbitrary selected node selection mostly retrieves molecules from the cluster numbers 2, 3, and 5. This difference in the virtual screening hitlist composition suggests the presence of two different binding modes, involving structurally distinct active molecules. Therefore, investigating why these two different branches are nonetheless able to distinguish between active and inactive molecules can be of high value for elucidating the binding modes of highly affinity GK ligands. It can also be highlighted that without prior knowledge of the systems, the selections of multiple pharmacophore models proved to be especially stable and reliable in terms of AUC values in comparison to the single pharmacophore models.

Analysis of Hierarchical Graphs of Pharmacophore Models Colored by Virtual Screening Results

The hierarchical graph representations provided intuitive support for the selection and evaluation of pharmacophore models, as detailed in the previous part. Nonetheless, additional information can be depicted to emphasize special characteristics of the pharmacophore models. To better understand the

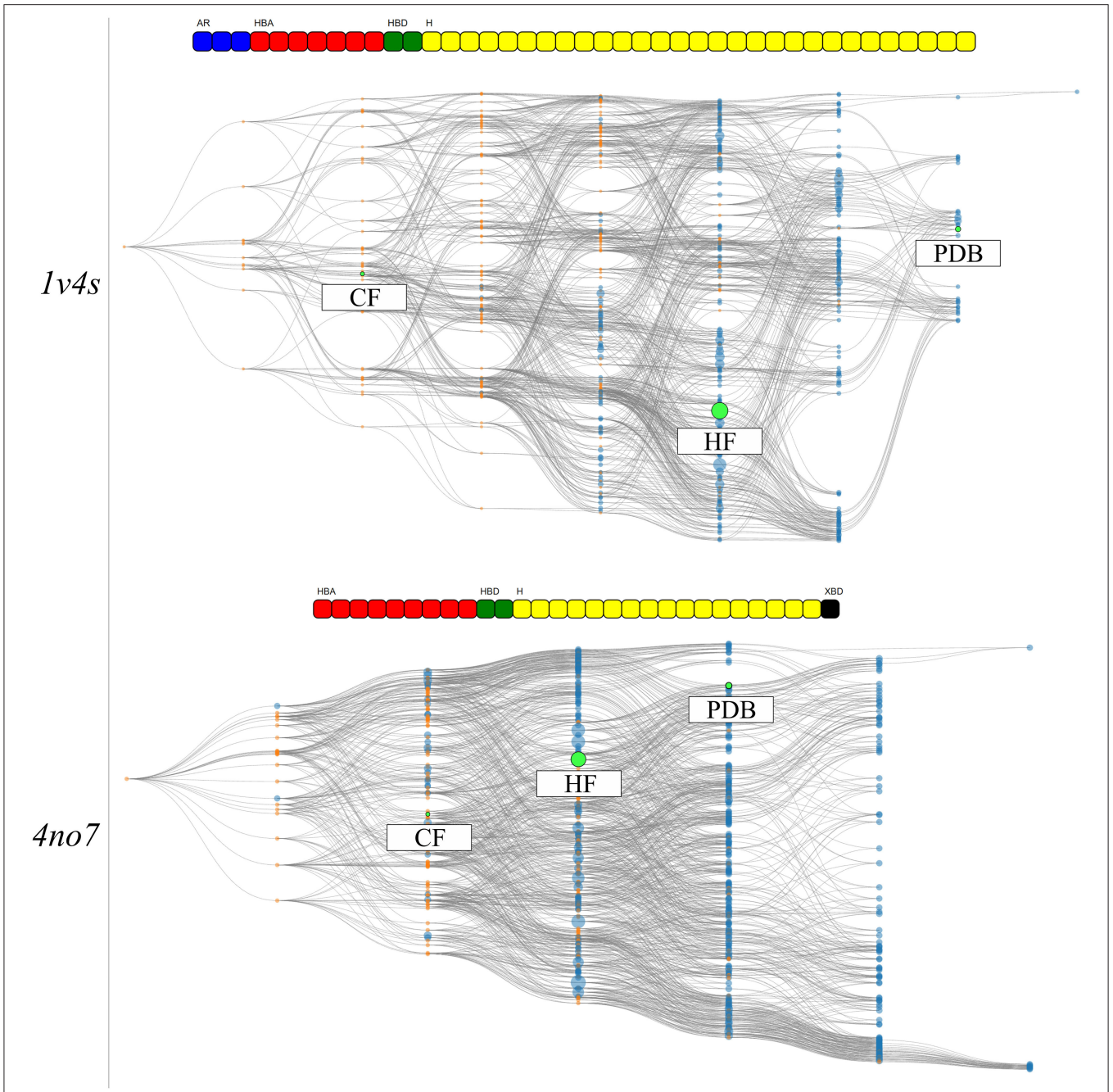


FIGURE 7 | The HGPMs based on the 3 MD simulation runs of the 1v4s system is shown on the top, and for the 4no7 system on the bottom. The nodes matching the pharmacophore models obtained from the PDB structure, the models with the highest frequency of appearance and with the highest count of common features between the crystallographic structures of the systems are highlighted in green and labeled HF and CF, respectively. The feature vectors of the highlighted models are shown above of the corresponding hierarchical graphs.

features involved in correctly distinguishing active and inactive molecules, all pharmacophore models associated with the feature vectors of the hierarchical graphs were used for VS. Then, each node was colored according to its AUC value at 10% of the database. For both systems, the corresponding results are shown in **Figure 8**. In the shown hierarchical graphs, the greener the node the closer its AUC value is to 1, and the redder its color becomes, the closer its AUC value is to 0. The "Artificial" nodes were not used for VS runs and are colored gray.

The graph of the 1v4s system shows a clear separation of colors. All nodes in the area of the HF node, as well as all its superset nodes are depicted in green. Therefore, prioritizing

TABLE 2 | Virtual screening results for both systems using different selections of pharmacophore models.

System	Selection	Number of common features	Number of pharmacophore model(s)	Number of hits	Auc at 1%	Auc at 5%	Auc at 10%	Auc at 50%	Auc at 100%
1v4s	pdb	7	1	0	0.00	0.00	0.00	0.00	0.00
	CHA	0	346	3,676	0.96	0.95	0.93	0.86	0.74
	HF	5	1	65	0.98	0.90	0.81	0.59	0.53
	HF+	5	7	87	0.98	0.91	0.82	0.59	0.53
	CF+	2	236	1,337	0.98	0.96	0.95	0.82	0.68
4no7	pdb	4	1	401	0.98	0.77	0.68	0.54	0.51
	pdb+	4	6	655	0.96	0.82	0.74	0.56	0.52
	CHA	0	461	13,868	0.98	0.79	0.71	0.64	0.58
	HF	3	1	2,761	0.81	0.41	0.34	0.51	0.50
	HF+	3	22	3,545	0.95	0.82	0.72	0.56	0.52
	CF+	2	49	2,546	0.93	0.81	0.74	0.59	0.54
	Selection +	3	30	3,421	0.96	0.88	0.81	0.58	0.53

Selections labeled with a + represent a subset of models used for consensus scoring. Model subsets comprise every pharmacophore with at least the same features as the initial node.

a selection of pharmacophore models from this area might be indicated in order to find the best balance between consensus virtual screening performance and number of considered models. On the other hand, the top half of the graph mainly depicts bad performing models, including the PDB node. The region around the CF node is depicted in gray as the VS was not performed for the pharmacophore model of the "Artificial" nodes. However, the consensus VS runs using all pharmacophore model supersets of the CF nodes performs well, although the sets contain both individual good and poorly performing models.

The graph for the 4no7 system is homogeneous and most of the nodes deliver good VS results (AUC $>$ 0.5). The HF node is colored brown, leading to below average results. However, the more specialized nodes located on the top branch of the graph performed well, including the PDB node. The region around the custom selected node, present on the other branch, also performed well, which can be interpreted as a different ligand binding mode. All models from the "Observed" nodes in the same columns as the CF node are displayed in red since models with less than three features do not lead to an unambiguous model alignment in 3D space and therefore return no hits.

Analysis of the Hierarchical Graph of Pharmacophore Models Projected to the GK Protein

The information gained for each individual system is present in the form of pharmacophore selection and in the comparison of the feature composition of the nodes. This is due to the pharmacophore feature label generation algorithm that considers which part of the ligand is interacting and allows a better accuracy in distinguishing unique features. Thus, the comparison between two systems with different ligands, as we did it for the 1v4s and 4no7 crystallographic structures is not possible. However, by disregarding the ligand identifier in the pharmacophore feature serial generation procedure we only keep the type of interaction of the pharmacophore feature and the protein identifier. In this way, we are trading the accuracy of where the interaction occurs on the ligand side, for the ability to 'project' the interaction on the protein side. This therefore unlocks the possibility to consider all pharmacophore models generated from a specific protein, regardless of which ligand was involved in the pharmacophore interactions. To demonstrate this alternative way of generating the pharmacophore feature serial, a HGPM based on all runs from both 1v4s and 4no7 systems has been created and is shown in **Figure 9**. Detailed information about the graph is provided in **Supplementary Table 3**.

The number of graph nodes and the number of unique pharmacophore features present in this hierarchical graph are both greater than those of the 2 previous graphs, which increases the complexity of the depiction. Among the 55 unique observed features, 7 where shared between 1v4s and 4no7: 3 hydrogen bond acceptor interactions with Arginine 63, Serine 64 and Threonine 65; a hydrogen bond donor interaction with Arginine 63; 3 hydrophobic features engaged in interactions with Tyrosine 24, the amino-acids Isoleucine 211, Threonine 65 and Tyrosine 214 and the amino-acids Threonine 65 and Tyrosine 214. Despite these common interactions, no "Observed" pharmacophore models were present in both the 1v4s and 4no7 MD simulation runs. We can nonetheless clearly observe a distinction between the two colored clusters, depicting the initial node affiliation. The top half of the graph with cyan colored nodes represent all pharmacophore models from the 1v4s system, where we can recognize a single branch leading to the specialized pharmacophore model of the PDB structure. For the bottom half of the graph with the orange nodes from the 4no7 system, we observe a more homogeneous spacing. The two branches previously observed are not easily distinguishable. It can be pointed out that the 4no7 PDB node is located near the nodes originating from 1v4s, and the custom selection node tends to fork more to the bottom part of the graph as more detailed

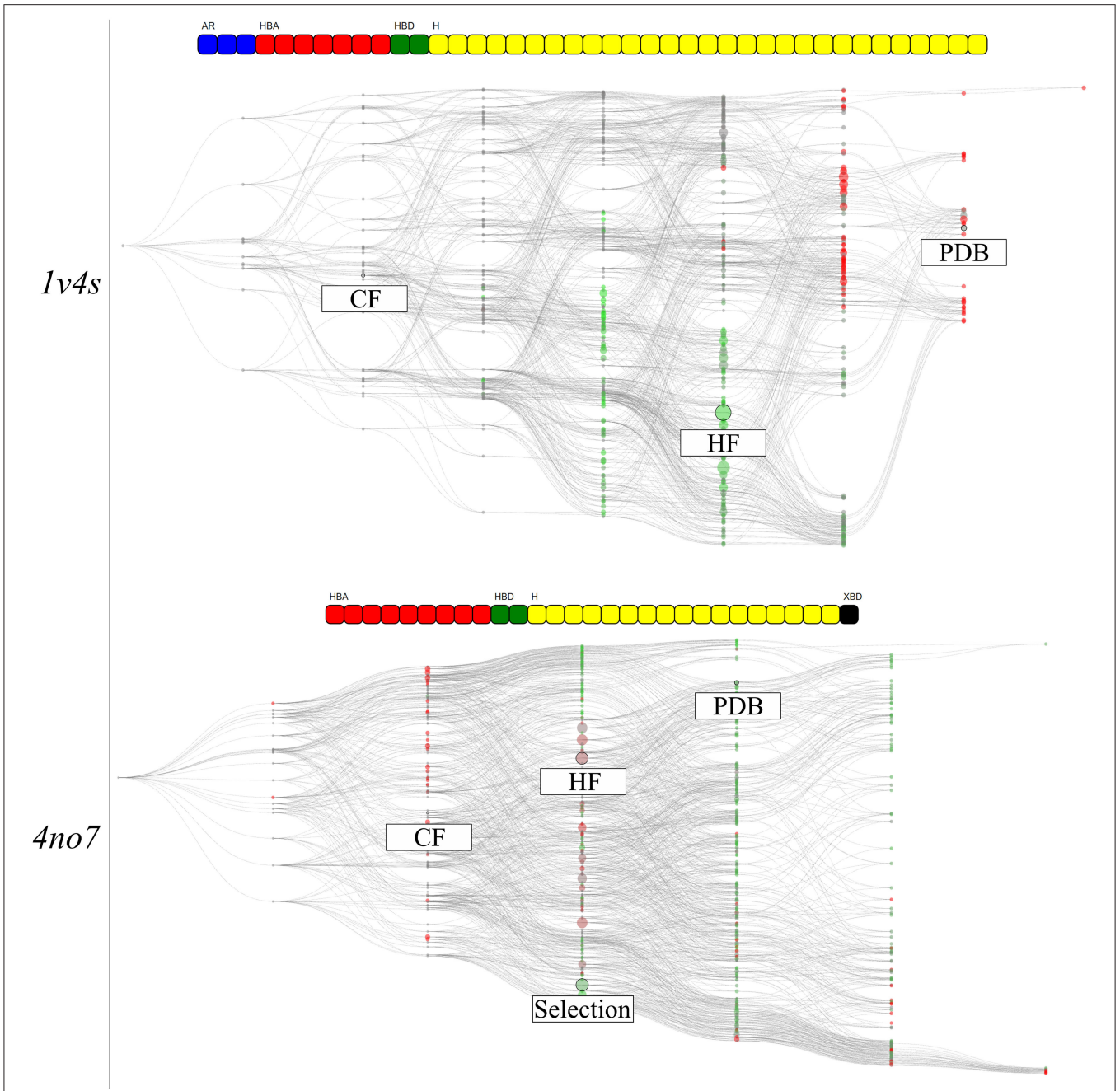


FIGURE 8 | The hierarchical graph of pharmacophore models derived from human glucokinase PDB codes of the 1v4s system is shown on the top, and of the 4no7 system on the bottom. The nodes are colored according to their AUC values at 10% of the database molecules. The greener the node, the closer is its AUC value to 1, and the redder the closer to 0. Nodes with "Artificial" pharmacophore models are colored gray. The nodes corresponding to the pharmacophore models obtained from the PDB structure, from the highest frequency of appearance and with the two common features between the systems are labeled accordingly. Associated feature vectors are shown above the corresponding graphs.

on **Supplementary Figure 6**. As the feature vector differs from the individual graphs, the pharmacophore selection for the CHA and for the subsets of pharmacophore models with the same feature as the CF nodes were different than previously and their virtual screening results details for this graph are presented on **Supplementary Table 4**.

Based on the heterogeneity of the graph, two pharmacophore models were selected. The hierarchical graph as well as two selected nodes, respectively binding mode 1 (BM1) and 2 (BM2) are shown in **Figure 10**. Both models are among the most specialized observed pharmacophores with 7 and 6 features, respectively. The single shared feature is a hydrogen bond

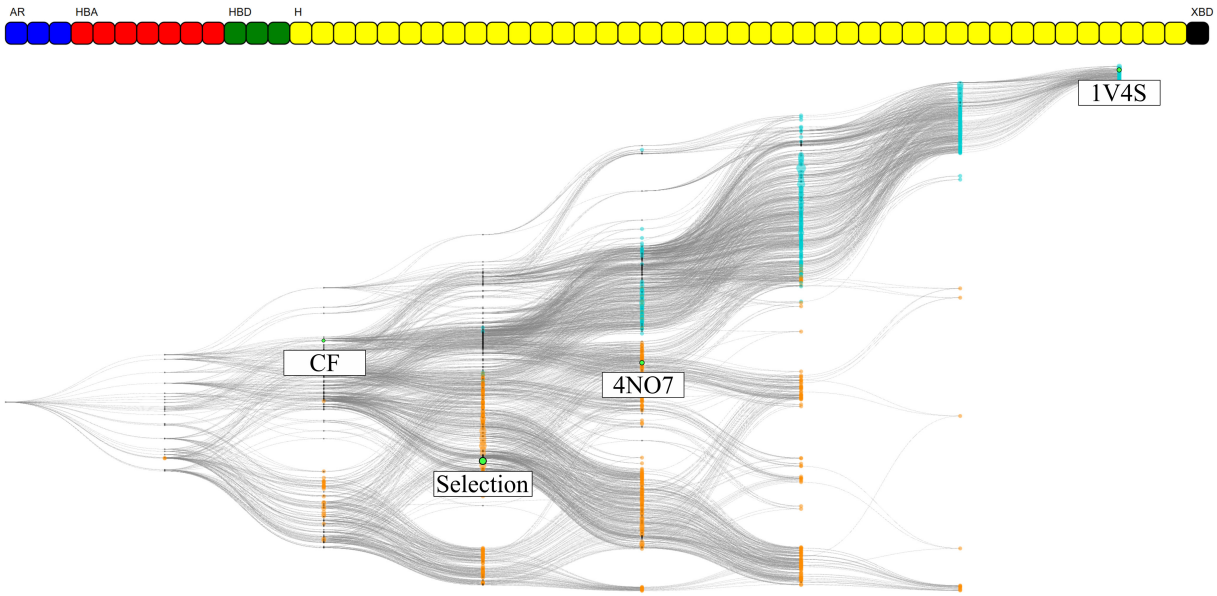


FIGURE 9 | HGPM obtained from the MD simulation runs of both the 1v4s and 4no7 system. The nodes are colored based on the system they were derived from: cyan for 1v4s and orange for 4no7. The nodes matching the pharmacophore models obtained from the PDB structures, the custom selection and the node with the two observed common features of the systems are labeled and highlighted in green. The feature vector is displayed on the top of the graph.

acceptor interaction with the Arginine 63, whose corresponding node is labeled CoreF. In the 3D representations of the pharmacophore models in **Figure 10**, we observed that the hydrophobic and halogen bonding interactions are present in different areas of the GK pocket as they involve different amino acids. Due to the important difference in the pharmacophore models in terms of feature composition and 3D alignment, the observed results can be linked to distinct binding modes due to the presence of the allosteric sub-pocket. This information can be used to provide a better depiction of the specific binding modes than the initial comparison between the two PDB pharmacophore models. Additionally, it allows the selection of smaller pharmacophore sets for consensus virtual screening in comparison with the CHA or the CF+ selection of models.

CONCLUSION

Motivated by the recent interest in consensus-based virtual screening methods involving pharmacophore models (Wieder et al., 2017; Polishchuk et al., 2019; Madzhidov et al., 2020), we developed an intuitive hierarchical graph representation of pharmacophore models. A user-friendly interactive visualization of the pharmacophore-based graph provides valuable information for computational chemists toward the understanding of protein-ligand interaction patterns and can aid in the selection of pharmacophore models for VS experiments. The graph can be created from sets of pharmacophore models generated for multiple crystallographic structures with identical macromolecular targets or for the output of MD simulation runs to provide insight into the dynamic aspects of the investigated systems. The graph generation has proven to be computationally inexpensive as it takes seconds to be created even for bigger ensembles with more than 10,000 models (see **Table 1**).

We selected two crystallographic structures of the human glucokinase to evaluate the HGPM generation algorithm in its ability to identify different binding modes and to select small representative pharmacophore model sets for consensus VS experiments. MD simulations and graph generations were performed individually for the two systems. Different selections of pharmacophore models were used to distinguish between active and inactive molecules for the two investigated systems. The selection of all models which possess a superset of the features contained in the pharmacophore model with the highest appearance count performed similarly to the CHA method in terms of AUC value and stability, while at the same time reducing the number of considered models for the 1v4s system by 20-fold and by more than 45-fold in the case of the 4no7 system (see **Table 2**), and thus helped to significantly reduce the required screening time. The hierarchical graph in **Figure 8** also helped to identify the best performing ensembles of pharmacophore features by depicting the virtual screening results for every pharmacophore model. Although the presence of two specialized branches in the hierarchical graph of the 4no7 system (in **Figure 7**) has already been noticed, the two different binding modes of GK were clearly identified by the hierarchical graph generated from the models extracted from both the 1v4s and 4no7 MD simulations (**Figure 10**), which is in perfect agreement with literature (Petit et al., 2011).

Depending on the goal, the graph representation can be adapted flexibly by either changing the pharmacophore

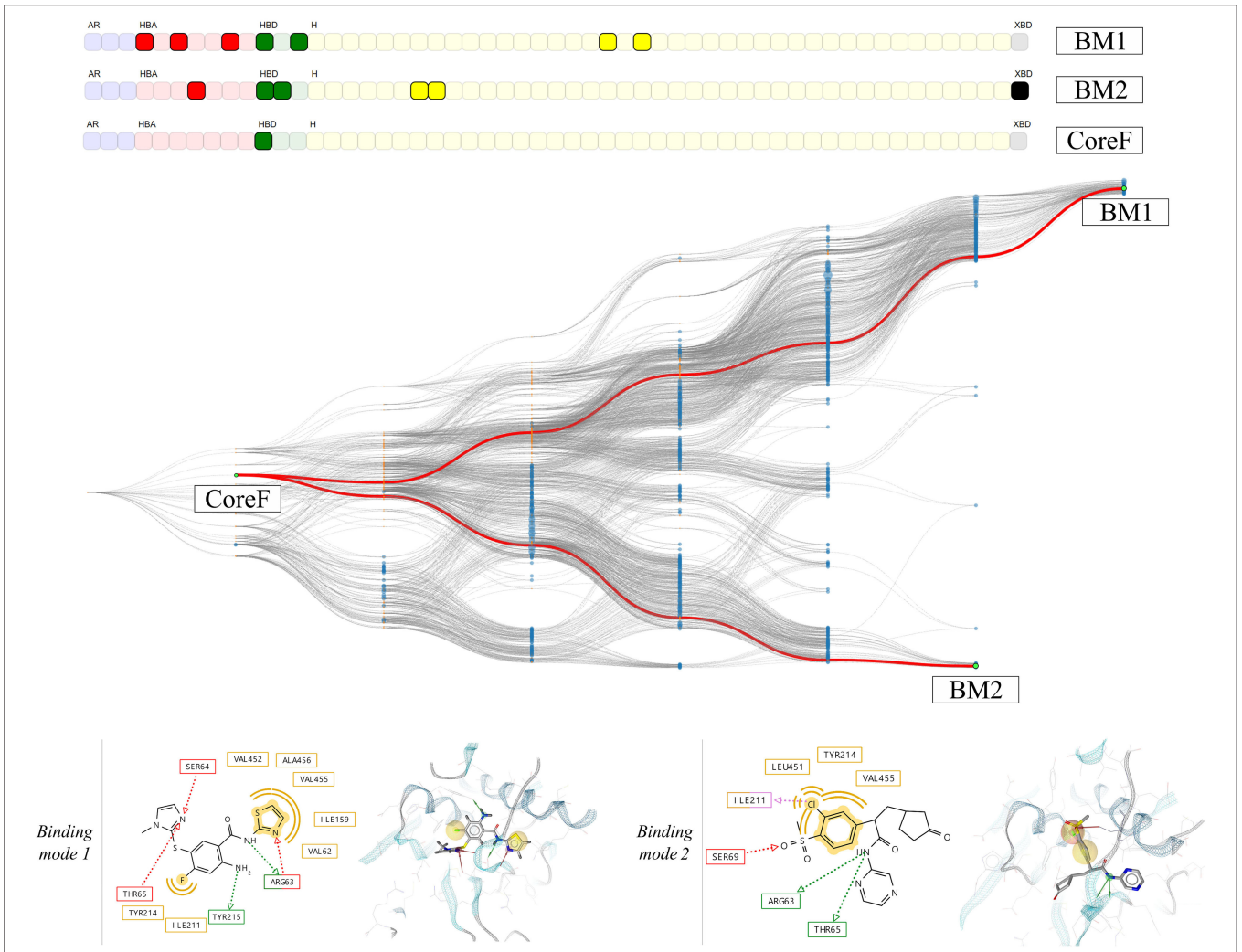


FIGURE 10 | HGPM obtained from the MD simulation runs human glucokinase starting from both the 1v4s and 4no7 systems. The nodes BM1 and BM2 were selected to show examples of different observed binding modes. 2D and 3D depictions of the corresponding pharmacophore models are shown below the graph.

feature serial generation algorithm or by showing additional properties. Highlighting the hierarchical relationship between pharmacophore models, this graph allows the user to analyze a target system by comparing the composition of several pharmacophore models in a single graphical representation, thus promoting the understanding of the binding process and the selection of pharmacophore models for consensus virtual screening runs. A typical workflow using the information provided by the graph representation is e.g., to first select a single pharmacophore model with the highest observed frequency, then perform a virtual screening run and finally add or remove individual features identified by following hierarchical links to build a refined model with the best ratio between accuracy and specificity. Presenting the hierarchical graph of pharmacophore models, we want to introduce an intuitive representation of multiple pharmacophore models and provide the computational and medicinal chemists with a new tool to enable an advanced understanding of the protein-ligand binding process, allowing for better decision support in the process of optimizing bio-active molecules.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: http://www.wwpdb.org/, 1v4s http://www.wwpdb.org/, 4no7.

AUTHOR CONTRIBUTIONS

GA, BS, TI, DP, and LT: project conception. GA, ST, TI, and LT: methodology and project administration. GA, WO,

BK, and IG: method implementation and validation. GA: writing and original draft preparation. WO, ST, BS, TI, DP, and LT: reviewing and editing. LT: supervision and funding resources. All authors contributed to the article and approved the submitted version.

FUNDING

GA gratefully acknowledges a PhD grant provided by IdRS.

ACKNOWLEDGMENTS

The authors gratefully acknowledge scientific discussions with scientists from IdRS and University of Vienna.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2020.599059/full#supplementary-material

REFERENCES

Beck, T., and Miller, B. G. (2013). Structural basis for regulation of human glucokinase by glucokinase regulatory protein. *Biochemistry* 52, 6232–6239. doi: 10.1021/bi400838t

Bell, G. I., and Polonsky, K. S. (2001). Diabetes mellitus and genetically programmed defects in β-cell function. *Nature* 414, 788–791. doi: 10.1038/414788a

Berman, H. M. (2000). The protein data bank. *Nucl. Acids Res.* 28, 235–242. doi: 10.1093/nar/28.1.235

Berthold, M. R., Cebron, N., Dill, F., Fatta, G. D., Gabriel, T. R., Georg, F., et al. (2009). Knime: The Konstanz Information Miner. 4.

Boehr, D. D., Nussinov, R., and Wright, P. E. (2009). The role of dynamic conformational ensembles in biomolecular recognition. *Nat. Chem. Biol.* 5, 789–796. doi: 10.1038/nchembio.232

Borg, I., and Groenen, P. (2003). Modern multidimensional scaling: theory and applications. *J. Educ. Measur.* 40, 277–280. doi: 10.1111/j.1745-3984.2003.tb01108.x

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

Choudhury, C., Priyakumar, U. D., and Sastry, G. N. (2015). Dynamics based pharmacophore models for screening potential inhibitors of mycobacterial cyclopropane synthase. *J. Chem. Inf. Model.* 55, 848–860. doi: 10.1021/ci500737b

Cozzini, P., Kellogg, G. E., Spyrakis, F., Abraham, D. J., Costantino, G., Emerson, A., et al. (2008). Target flexibility: an emerging consideration in drug discovery and design. *J. Med. Chem.* 51, 6237–6255. doi: 10.1021/jm800562d

De Vivo, M., Masetti, M., Bottegoni, G., and Cavalli, A. (2016). Role of molecular dynamics and related methods in drug discovery. *J. Med. Chem.* 59, 4035–4061. doi: 10.1021/acs.jmedchem.5b01684

Demonstration of HGPM (2020). *Individual Files Used by the Demonstration*. Available online at: https://www.inteligand.com/hgpharm/demo/ (accessed November 22, 2020).

Durrant, J. D., and McCammon, J. A. (2011). Molecular dynamics simulations and drug discovery. *BMC Biol.* 9:71. doi: 10.1186/1741-7007-9-71

Fenwick, R. B., Esteban-Martín, S., and Salvatella, X. (2011). Understanding biomolecular motion, recognition, and allostery by use of conformational ensembles. *Eur. Biophys. J.* 40, 1339–1355. doi: 10.1007/s00249-011-0754-8

Gaulton, A., Hersey, A., Nowotka, M., Bento, A. P., Chambers, J., Mendez, D., et al. (2017). The ChEMBL database in 2017. *Nucl. Acids Res.* 45, D945–D954. doi: 10.1093/nar/gkw1074

Jo, S., Kim, T., Iyer, V. G., 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

Kamata, K., Mitsuya, M., Nishimura, T., Eiki, J., and Nagata, Y. (2004). Structural basis for allosteric regulation of the monomeric allosteric enzyme human glucokinase. *Structure* 12, 429–438. doi: 10.1016/j.str.2004.02.005

Langer, T. (2010). Pharmacophores in drug research. *Mol. Inf.* 29, 470–475. doi: 10.1002/minf.201000022

Leach, A. R., Gillet, V. J., Lewis, R. A., and Taylor, R. (2010). Three-dimensional pharmacophore methods in drug discovery. *J. Med. Chem.* 53, 539–558. doi: 10.1021/jm900817u

Liu, X., Shi, D., Zhou, S., Liu, H., Liu, H., and Yao, X. (2018). Molecular dynamics simulations and novel drug discovery. *Expert Opin. Drug Discov.* 13, 23–37. doi: 10.1080/17460441.2018.1403419

Madzhidov, T. I., Rakhimbekova, A., Kutlushuna, A., and Polishchuk, P. (2020). Probabilistic approach for virtual screening based on multiple pharmacophores. *Molecules* 25:385. doi: 10.3390/molecules25020385

Maggiora, G. M., and Bajorath, J. (2014). Chemical space networks: a powerful new paradigm for the description of chemical space. *J. Comput. Aided Mol. Des.* 28, 795–802. doi: 10.1007/s10822-014-9760-0

Mead, A. (1992). Review of the development of multidimensional scaling methods. *Statistician* 41, 27–39. doi: 10.2307/2348634

Métivier, J.-P., Cuissart, B., Bureau, R., and Lepailleur, A. (2018). The pharmacophore network: a computational method for exploring structure–activity relationships from a large chemical data set. *J. Med. Chem.* 61, 3551–3564. doi: 10.1021/acs.jmedchem.7b01890

Osbak, K. K., Colclough, K., Saint-Martin, C., Beer, N. L., Bellanné-Chantelot, C., Ellard, S., et al. (2009). Update on mutations in glucokinase (GCK), which cause maturity-onset diabetes of the young, permanent neonatal diabetes, and hyperinsulinemic hypoglycemia. *Hum. Mutat.* 30, 1512–1526. doi: 10.1002/humu.21110

Petit, P., Antoine, M., Ferry, G., Boutin, J. A., Lagarde, A., Gluais, L., et al. (2011). The active conformation of human glucokinase is not altered by allosteric activators. *Acta Crystallogr. D Biol. Crystallogr.* 67, 929–935. doi: 10.1107/S0907444911036729

Poli, G., Seidel, T., and Langer, T. (2018). Conformational sampling of small molecules with icon: performance assessment in comparison with OMEGA. *Front. Chem.* 6:229. doi: 10.3389/fchem.2018.00229

Polishchuk, P., Kutlushina, A., Bashirova, D., Mokshyna, O., and Madzhidov, T. (2019). Virtual screening using pharmacophore models retrieved from molecular dynamic simulations. *IJMS* 20:5834. doi: 10.3390/ijms20235834

Prlić, A., Bliven, S., Rose, P. W., Bluhm, W. F., Bizon, C., Godzik, A., et al. (2010). Pre-calculated protein structure alignments at the RCSB PDB website: Fig. 1. *Bioinformatics* 26, 2983–2985. doi: 10.1093/bioinformatics/btq572

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, 327–341. doi: 10.1016/0021-9991(77)90098-5

Schrodinger (2010). *Maestro Version 9.1*. New York, NY: Schrodinger, LLC.

Schuster, D. (2010). 3D pharmacophores as tools for activity profiling. *Drug Discov. Today Technol.* 7, e203–e270. doi: 10.1016/j.ddtec.2010.11.006

Sohn, Y., Park, C., Lee, Y., Kim, S., Thangapandian, S., Kim, Y., et al. (2013). Multi-conformation dynamic pharmacophore modeling of the peroxisome proliferator-activated receptor γ for the discovery of novel agonists. *J. Mol. Graph. Modell.* 46, 1–9. doi: 10.1016/j.jmgm.2013.08.012

Source code for the HGPM and Implementation (2020). *Graph Generation. Visualization*. Available online at: https://www.inteligand.com/hgpharm/download/hgpharm_src.zip; https://www.inteligand.com/hgpharm/download/hgpharm_view.zip (accessed November 22, 2020).

Spyrakis, F., Benedetti, P., Decherchi, S., Rocchia, W., Cavalli, A., Alcaro, S., et al. (2015). A pipeline to enhance ligand virtual screening: integrating molecular dynamics and fingerprints for ligand and proteins. *J. Chem. Inf. Model.* 55, 2256–2274. doi: 10.1021/acs.jcim.5b00169

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

Wermuth, C. G., Ganellin, C. R., Lindberg, P., and Mitscher, L. A. (1998). Glossary of terms used in medicinal chemistry (IUPAC Recommendations 1998). *Pure Appl. Chem.* 70, 1129–1143. doi: 10.1351/pac199870051129

Wieder, M., Garon, A., Perricone, U., Boresch, S., Seidel, T., Almerico, A. M., et al. (2017). Common hits approach: combining pharmacophore modeling and molecular dynamics simulations. *J. Chem. Inf. Model.* 57, 365–385. doi: 10.1021/acs.jcim.6b00674

Wolber, G., and Langer, T. (2005). LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. *J. Chem. Inf. Model.* 45, 160–169. doi: 10.1021/ci049885e

Ye, Y., and Godzik, A. (2003). Flexible structure alignment by chaining aligned fragment pairs allowing twists. *Bioinformatics* 19, ii246–ii255. doi: 10.1093/bioinformatics/btg1086

Conflict of Interest: BS and IG are employees of the company Inte:Ligand Software-Entwicklungs und Consulting GmbH, Vienna, Austria. TI and DP are employees of Institut de Recherches Servier (IdRS), Croissy-sur-Seine, France. The herein presented work is the result of a joint research project funded by the aforementioned companies.

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: 17 December 2020
doi: 10.3389/fmolb.2020.601065

RASPD+: Fast Protein-Ligand Binding Free Energy Prediction Using Simplified Physicochemical Features

Stefan Holderbach[1,2], *Lukas Adam*[1,2], *B. Jayaram*[3], *Rebecca C. Wade*[1,4,5*] *and Goutam Mukherjee*[1,4*]

[1] Molecular and Cellular Modelling Group, Heidelberg Institute of Theoretical Studies, Heidelberg, Germany, [2] Institute of Pharmacy and Molecular Biotechnology (IPMB), Heidelberg University, Heidelberg, Germany, [3] Supercomputing Facility for Bioinformatics & Computational Biology, Department of Chemistry, Kusuma School of Biological Sciences, Indian Institute of Technology Delhi, New Delhi, India, [4] Center for Molecular Biology (ZMBH), DKFZ-ZMBH Alliance, Heidelberg University, Heidelberg, Germany, [5] Interdisciplinary Center for Scientific Computing (IWR), Heidelberg University, Heidelberg, Germany

The virtual screening of large numbers of compounds against target protein binding sites has become an integral component of drug discovery workflows. This screening is often done by computationally docking ligands into a protein binding site of interest, but this has the drawback of a large number of poses that must be evaluated to obtain accurate estimates of protein-ligand binding affinity. We here introduce a fast pre-filtering method for ligand prioritization that is based on a set of machine learning models and uses simple pose-invariant physicochemical descriptors of the ligands and the protein binding pocket. Our method, Rapid Screening with Physicochemical Descriptors + machine learning (RASPD+), is trained on PDBbind data and achieves a regression performance that is better than that of the original RASPD method and traditional scoring functions on a range of different test sets without the need for generating ligand poses. Additionally, we use RASPD+ to identify molecular features important for binding affinity and assess the ability of RASPD+ to enrich active molecules from decoys.

Keywords: structure based drug design, virtual screening, physicochemical molecular descriptors, machine learning, protein-ligand complex, binding free energy

OPEN ACCESS

Edited by:
Sergio Decherchi,
Italian Institute of Technology (IIT), Italy

Reviewed by:
Sophie Sacquin-Mora,
UPR9080 Laboratoire de Biochimie Théorique (LBT), France
Matteo Salvalaglio,
University College London, United Kingdom

****Correspondence:***
Rebecca C. Wade
rebecca.wade@h-its.org
Goutam Mukherjee
goutam.mukherjee@h-its.org

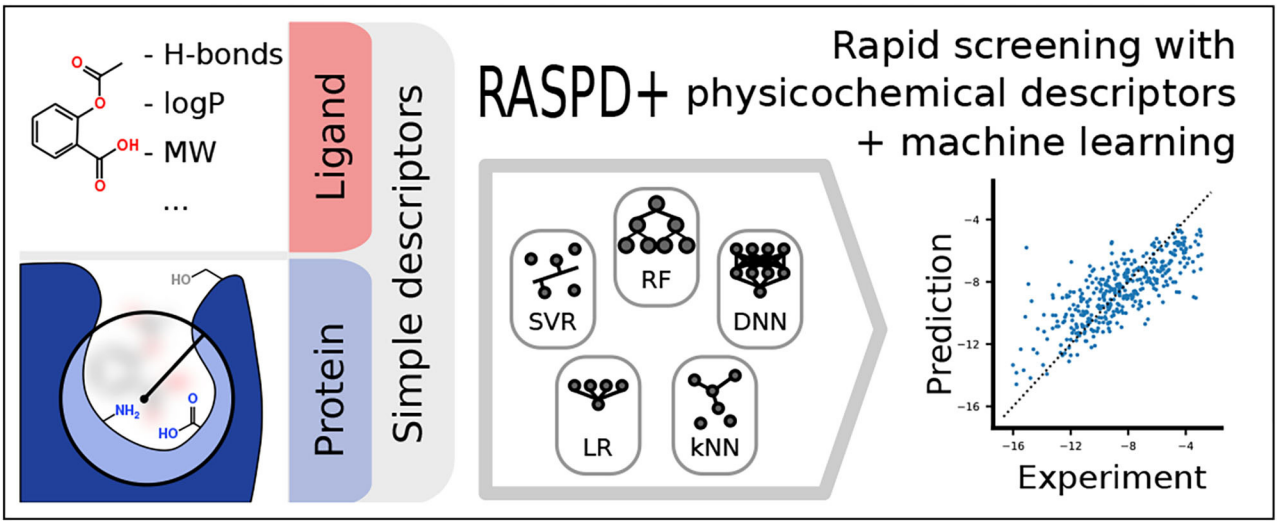
Specialty section:
This article was submitted to Biological Modeling and Simulation, a section of the journal Frontiers in Molecular Biosciences

Received: *31 August 2020*
Accepted: *13 November 2020*
Published: *17 December 2020*

Citation:
Holderbach S, Adam L, Jayaram B, Wade RC and Mukherjee G (2020) RASPD+: Fast Protein-Ligand Binding Free Energy Prediction Using Simplified Physicochemical Features. Front. Mol. Biosci. 7:601065. doi: 10.3389/fmolb.2020.601065

1. INTRODUCTION

Virtual screening to assess *in silico* the binding of candidate ligands to a target protein is a key component of structure-based drug design procedures (Torres et al., 2019; Wang et al., 2020). Typically, screening is done by docking the ligands at many different positions or poses in the three-dimensional structure of the target protein. At every position, a scoring function is evaluated to approximate the binding-free energy, and this is used to rank the binding poses and different candidate ligands for their ability to bind to the target protein. While correct docking poses are frequently generated, scoring functions often lack the accuracy necessary to correctly rank poses or ligands (Li et al., 2019). Docking procedures are therefore frequently supplemented by methods employing molecular dynamics simulations with the aim of computing more accurate binding affinities. However, both docking and molecular dynamics simulations often fail to provide predictions of binding free energy at the level of accuracy desired. Furthermore, they are demanding in terms of computational effort and expertise (Willems et al., 2020). There is therefore a need



Graphical Abstract | Overview of the overall workflow of RASPD+ (Rapid Screening with Physicochemical Descriptors + machine learning).

for quick approaches with robust predictive scoring functions to facilitate the screening and prioritization of large libraries of compounds prior to applying docking and simulation methods.

While the assessment of ligand properties, e.g., for drug-likeness (Lipinski et al., 2001), to filter ligand libraries is well established, we here address the need to filter and prioritize ligands based not only on ligand properties but also on the properties of the target protein. For this purpose, we previously developed a simple hybrid regression approach called RASPD (Rapid Screening with Physicochemical Descriptors) (Mukherjee and Jayaram, 2013). In this linear regression model, the binding-free energy ΔG was predicted using a minimal set of physicochemical descriptors for typical interactions. Hydrogen bonding was accounted for by counting potential donor and acceptor atoms. Van der Waals forces were approximated by the Wiener topology index (Wiener, 1947) and the molar refractivity, which describes the polarizability of a molecule (Ghose and Crippen, 1987). Additionally, the partition coefficient logP allowed for the estimation of the hydrophobic effect. While the descriptor values for the ligand are straightforward to compute, simplifying assumptions were made to obtain the physicochemical descriptors for the target protein. A sphere was centered on a known or assumed binding pocket position with a radius encompassing the maximum size of the ligand. This sphere was then used to select the amino acid residues for which descriptors were computed (Mukherjee and Jayaram, 2013) (**Figure 1A**).

However, the linear regression model used (Mukherjee and Jayaram, 2013) has limited abilities to capture complex feature interactions compared to non-linear models. Since RASPD was first developed, more high-quality data sets on protein-ligand complexes with associated binding-free energies have been made available (Liu et al., 2015; Gathiaka et al., 2016), and a large number of machine learning methods have been developed (Yang et al., 2019). Moreover, machine learning approaches have successfully been used to either replace (Gomes et al., 2017; Feinberg et al., 2018; Jiménez et al., 2018) or enhance (Pei et al., 2019; Boyles et al., 2020) the predictions of traditional scoring functions for protein-ligand binding.

We have thus developed RASPD+, which is a new tool that improves on the conceptual framework of the original RASPD method by using the following: (i) a set of diverse machine learning methods to derive an ensemble prediction, (ii) additional and more fine-grained descriptors for the target proteins, and (iii) larger training sets of newer protein-ligand binding data. We here describe the training, testing, and application of RASPD+. We demonstrate the capabilities of RASPD+ for binding free energy regression and compare its performance to established scoring functions. We also analyze the features contributing to the predictions to gain insights into the important features for binding affinity. Finally, we show that RASPD+ can enrich active molecules in tests with the Directory of Useful Decoys-Enhanced data set (DUD-E) (Mysinger et al., 2012).

2. METHODS

The computational workflow and the training and validation procedure used for RASPD+ are illustrated in **Figure 1**.

2.1. Datasets

The PDBbind refined data set (release 2018) (Wang et al., 2004; Liu et al., 2015), containing 4,463 protein-ligand crystal structures and experimentally measured binding affinities, served as the initial data set. Although the PDBbind refined data set contains data on binding from different measurements reported in the literature and no experimental method was specified as a requirement for the data to be incorporated in PDBbind,

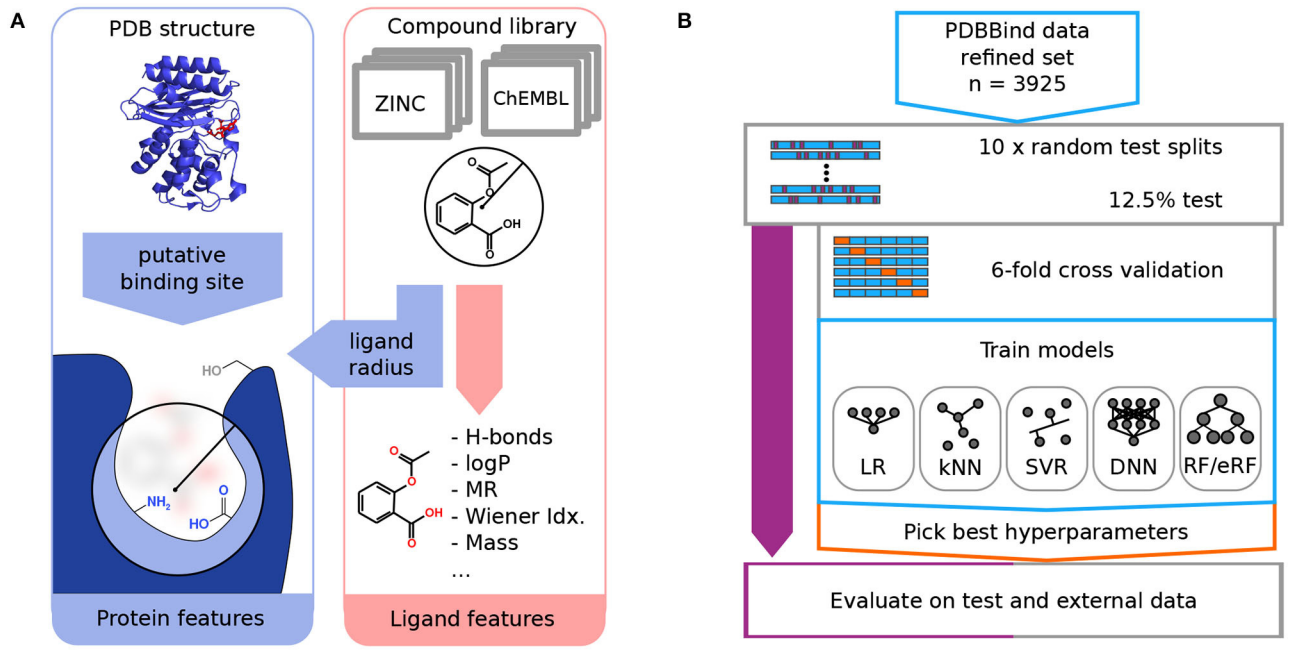


FIGURE 1 | The computational workflow of RASPD+ comprises featurization **(A)** and the training and evaluation of machine learning models **(B)**. **(A)** For each ligand molecule, simple physicochemical descriptors are computed based on atomic contributions. Information about the target protein is gathered within a sphere around a putative binding position whose size is determined by the radius of the ligand. For residues within this sphere, similar descriptors are computed. **(B)** Data from the PDBbind refined set featurized in this way served as the training data in a nested cross-validation strategy. To compare linear regression (LR), k-nearest neighbors (kNN), support vector regression (SVR), neural network (DNN), random forest (RF), and extremely random forest (eRF) models, test sets were split off in 10 replicates in an outer loop. In the inner loop, six-fold cross-validation was used to select the best hyperparameters for the given model.

it contains high-quality structures of non-covalent protein-ligand interactions with a resolution better than 2.5 Å and no steric clashes. The PDBbind refined data set is therefore extensively used as a benchmark set for protein-ligand binding affinity prediction (Liu et al., 2015). We thus obtained structural information about each protein in the data set, the position and structure of the ligand binding to it, and the corresponding binding constant. As we considered modeling the coordination of metal ions to be beyond the scope of our approach, the structures were filtered to exclude cases with metal ions within 2.1 Å of the ligand. Dissociation and inhibition constants and IC_{50} values were converted to binding-free energies using the following equation:

$$\begin{aligned} \Delta G &= -RT \ln K \text{ where } K \in \{K_d, K_i, IC_{50}\} \text{ assuming} \\ T &= 298.15\, K \end{aligned} \tag{1}$$

This processing resulted in a set of 3,925 protein-ligand complexes for training, validation, and testing.

For further testing, the following previously published benchmark sets served as external test sets: The Community Structure-Activity Resource (CSAR) NRC-HiQ 2010 selection (Dunbar et al., 2011; Smith et al., 2011), data sets from the CSAR 2012 (Dunbar et al., 2013), and CSAR 2014 (Carlson et al., 2016) challenges, and a data set described by Wang et al. (2015).

The CSAR-NRC 2010 HiQ release (Dunbar et al., 2011; Smith et al., 2011) contains two sets of protein-ligand complexes, with 55 and 49 docked complexes, respectively, as well as information about experimental binding affinities.

Another set of binding-free energies and corresponding structures was assembled from the CSAR 2012 (Dunbar et al., 2013) and CSAR 2014 (Carlson et al., 2016) data sets that are now curated by the Drug Design Data Resource (D3R) (drugdesigndata.org) (Gathiaka et al., 2016). For this set, which we refer to as the D3R data set, we downloaded the data for the proteins urokinase, cyclin-dependent kinase 2 (CDK2), checkpoint kinase 1 (CHK1), MAP kinase 1 (ERK2), LpxC deacetylase (LpxC), spleen tyrosine kinase (SYK), tRNA (m1G37) methyltransferase (tRMD), heat shock protein 90 (HSP90), and a CDK2-Cyclin A complex. The SMILES strings of 1,271 active inhibitors of these proteins in the D3R data set were converted to 3D structures in PDB format using Open Babel (O'Boyle et al., 2011). For HSP90, we excluded 46 compounds that were all assigned the same ΔG of -5.860 kcal/mol as this value, likely represented a threshold value for the experimental measurements rather than the actual binding affinity of the ligands.

Wang et al. (2015) aggregated previous experimental results and PDB structures for 283 complexes of seven different proteins: beta-secretase (BACE), CDK2, induced myeloid leukemia cell differentiation protein (Mcl-1), p38 MAP kinase, protein-tyrosine phosphatase 1B (PTP1B), thrombin, and tyrosine kinase 2 (TYK2). For this set, protein structures were retrieved from the RCSB protein data bank (http://www.rcsb.org) and hydrogen

atoms were added to the protein structures with the `tleap` module of AMBER 14 (Case et al., 2005). The structural data for inhibitors and experimental binding-free energies were obtained from the literature (Wilson et al., 2007; Baum et al., 2009; Goldstein et al., 2011; Cumming et al., 2012; Friberg et al., 2013; Liang et al., 2013a,b; Wang et al., 2013, 2015). This included additional ligands for Mcl-1 (Friberg et al., 2013) and TYK2 (Liang et al., 2013a,b) that were not used by Wang et al. (2015). The structures of the 283 inhibitors were redrawn and verified in the MOE software (Chemical Computing Group, Montreal, QC).

Further details on the source of structures and experimental binding affinities are given in **Supplementary Table 1**.

2.2. Generation of Molecular Descriptors

To model the non-covalent interactions, physicochemical molecular descriptors were computed using an improved pipeline based on that for the original RASPD procedure described in Mukherjee and Jayaram (2013) (**Figure 1A**). For each ligand, the molecular weight (here abbreviated as MASS), the number of hydrogen bond donors (D) and acceptors (A), an approximate octanol-water partition coefficient log P (logP) (Wildman and Crippen, 1999), the molar refractivity (MR) (Wildman and Crippen, 1999), and the Wiener topology index (W) (Wiener, 1947) were computed as described previously (Mukherjee and Jayaram, 2013). Based on the ligand position in the protein structure, the most likely interacting amino acid residues were selected using a sphere whose radius was derived from the maximum distance (*maxD*) between ligand atoms and the center of mass (**Figure 1A**). For the computation of the logP and MR descriptors, this sphere was extended by 0.9 Å over *maxD*, and residues were selected based on their center of mass. To count hydrogen bond donors and acceptors, a sphere extending 3 Å beyond *maxD* was used to select atoms. Details regarding the protein pocket selection procedure and the choice of the cut-off radii are given in Mukherjee and Jayaram (2013). To make the protein descriptors more fine grained than in the previous RASPD procedure, we computed molar refractivity and log P for aromatic and non-aromatic residues separately [PMR(Arom), PMR(Non-Arom), PlogP(Arom), PlogP(Non-Arom)]. Hydrogen bond donors were counted separately for the backbone amide group [PD(Amide-NH)] as well as for the following amino acid sets: Positively charged PD (K+R+HIP), neutral amino groups PD(K+N+Q), heteroaromatic donors PD (W+H), and hydroxyl-containing groups PD (T+S+Y+D+E). The number of hydrogen bond acceptors was determined for the backbone amide [PA(Amide-O)] and the following sets: negatively charged PA (D+E), neutral non-aromatic PA (N+Q+T+S+D-H+E-H), and aromatic acceptors PA (Y+H). The individual protein residue-derived descriptors were scaled by the ligand *maxD*. Additionally, the volume of the protein pocket (PVol) was computed using tools from the TRAPP software suite (Kokh et al., 2013; Yuan et al., 2020). In total, therefore, six ligand and 14 protein descriptors were computed per ligand-protein complex.

2.3. General Strategy for Training and Testing

To obtain a robust estimate of performance on the PDBbind data set as well as the test sets, a nested cross-validation strategy was used (**Figure 1B**). For 10 replicates, the PDBbind refined set was split into a test set covering 12.5% of the data and a set for cross-validation training. For each of these replicates, six-fold cross-validation training was performed to select the best hyperparameters for each replicate based on the Pearson correlation coefficient. For each replicate, therefore, 2,860 complexes were used for training, 572 for cross-validation, and 493 for testing.

The input features were robustly centered and scaled by the median and interquartile range (IQR) of the training set for each train-test split. All models obtained by the hyperparameter search were evaluated on the corresponding PDBbind test set as well as on the external test sets. We report the mean and standard deviation of the performance metrics.

2.4. Evaluation Metrics

To assess model performance, the root-mean-squared error (*RMSE*), Pearson (r), and Spearman (ρ) correlation coefficients, and the coefficient of determination, R^2, were computed using the `sklearn.metrics` and `scipy.stats` Python modules. Additionally, we report the Q^2_{F3} metric (Equation 2) (Consonni et al., 2009), as it is considered to be better suited for QSAR-like tasks than R^2 (Todeschini et al., 2016).

$$Q^2_{F3} = 1 - \frac{\sum_i^{n_{test}} (\hat{y} - y_{test})^2}{\sum_i^{n_{test}} (\hat{y} - \overline{y_{train}})^2} \tag{2}$$

2.5. Models and Hyperparameters

As part of this work, we evaluated different machine learning models. We considered linear regression (LR), as it was also used in the previous RASPD approach (Mukherjee and Jayaram, 2013), support vector regression (Drucker et al., 1997) (SVR), k-Nearest Neighbors (kNN), simple deep neural networks (DNN), random forests (Breiman, 2001) (RF), and a variant of the former, extremely random forests (Geurts et al., 2006) (eRF). The associated hyperparameters for each method were optimized by a grid search covering a typical space. Further details on each method and their associated hyperparameters are given in the **Supplementary Materials**. A comprehensive list of tested hyperparameters is given in **Supplementary Table 2**. All models except the neural networks were built using the `scikit-learn` Python package (version 0.20.2) (Pedregosa et al., 2011). For the neural networks, the Keras API (version 2.2.4) (Chollet et al., 2015) for TensorFlow (version 1.12) was used in conjunction with the `talos` package (version 0.4.6) (Kotila, 2018) for hyperparameter optimization.

2.6. Estimation of Feature Importance

To estimate the importance of individual input features, a simple permutation-based approach was used (Breiman, 2001). After prediction on a real-world test set, the column of each feature in the data set was shuffled in five replicates, and the mean

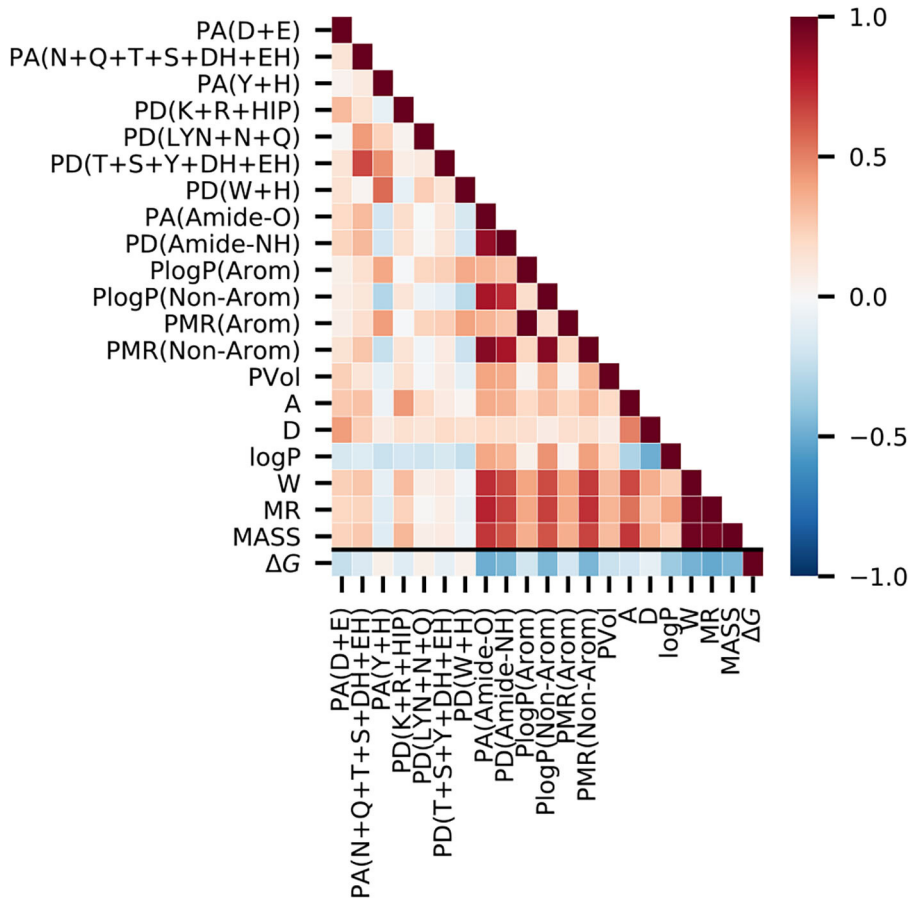


FIGURE 2 | Correlation analysis on the PDBbind data set reveals that the experimental binding-free energy has the strongest negative correlation with the ligand molar refractivity (MR, Spearman $\rho = -0.51$), and with the number of peptide bond oxygen atoms (hydrogen bond acceptors) present in the putative protein binding pocket [PA(Amide-O), Spearman $\rho = -0.49$]. The value of the Spearman's correlation coefficient is indicated by color.

change in Pearson correlation coefficient was computed. The model then has to make a prediction based on a random sample from a distribution with the same mean and variance. A drop in predictive performance indicates that the prediction is dependent on this feature.

2.7. Enrichment Analysis With Decoy Compounds From the DUD-E Dataset

To evaluate the performance of RASPD+ for capturing active molecules from a pool of computationally generated decoys, 3D coordinates of active and decoy molecules were retrieved from the DUD-E data set (Mysinger et al., 2012). This set contains 102 targets with on average ≈ 200 distinct and validated binding ligands and corresponding ≈ 14,000 selected decoys for each system. Information about the proteins, as well as the number of active and decoy molecules for each system, is given in **Supplementary Table 9**. Enrichment was performed by selecting a given percentage of molecules that scored highest in the given method. For scoring, the predictions across the six cross-validation folds of a replicate were averaged. The enrichment factor was defined as the ratio of the fraction of active molecules in the enriched set divided by the fraction of the active molecules in the total set. For failure case analysis, we additionally determined which systems contained another cofactor in the binding pocket by checking for non-protein atoms within the pocket structure. Surface-only binding sites were identified by filtering interactions with few amino-acids and manually validating surface binding. More detailed subsets of DUD-E were adopted from Vieira and Sousa (2019) instead of the more coarse-grained classification from Mysinger et al. (2012).

3. RESULTS

3.1. Analysis of the Descriptors and Data Sets

To confirm the usefulness of the chosen molecular descriptors, we performed correlation analysis on the PDBbind refined set (**Figure 2**, **Supplementary Table 3**). The Spearman correlations with the binding free energy, ΔG, were negative for most descriptors, as stronger binding is indicated by negative values

TABLE 1 | Comparison of the performance of the models derived with seven different machine learning methods for predicting the protein-ligand binding free energy for the PDBbind test set.

Model	RMSE	r	ρ	R^2	Q^2_{F3}
Null model	2.76 ± 0.05	0.0 ± 0.0	NA	-0.00 ± 0.00	-0.03 ± 0.05
LR	2.19 ± 0.05	0.61 ± 0.02	0.60 ± 0.02	0.37 ± 0.02	0.35 ± 0.03
kNN	2.03 ± 0.04	0.68 ± 0.02	0.67 ± 0.02	0.46 ± 0.03	0.44 ± 0.03
lSVR	2.20 ± 0.05	0.61 ± 0.02	0.60 ± 0.02	0.37 ± 0.02	0.35 ± 0.03
SVR	2.04 ± 0.05	0.68 ± 0.02	0.67 ± 0.02	0.45 ± 0.03	0.44 ± 0.03
DNN	2.05 ± 0.05	0.67 ± 0.02	0.66 ± 0.02	0.45 ± 0.02	0.43 ± 0.03
RF	1.88 ± 0.04	**0.74 ± 0.02**	0.73 ± 0.02	0.53 ± 0.02	0.52 ± 0.02
eRF	**1.86 ± 0.05**	**0.74 ± 0.02**	**0.74 ± 0.01**	**0.55 ± 0.02**	**0.54 ± 0.03**

The five metrics of performance are given as mean and standard deviation values computed by averaging from the 10 different random test set splits and six cross-validation folds. The RMSE is given in kcal/mol. NA, not applicable.

of ΔG. The strongest negative correlations were observed for the molar refractivity of the ligand molecule (MR, −0.51) and the abundance of peptide bond oxygen atoms (hydrogen bond acceptors) inside the protein binding pocket [PA(Amide-O), −0.49]. The correlations with ΔG for the features for specific amino acids were lower than 0.25 in magnitude, which is less than the corresponding correlation (> 0.4) obtained for the backbone [PA(Amide-O), PD(Amide-NH)] and non-aromatic amino acid [PlogP(Non-Arom), PMR(Non-Arom)] descriptors.

We next analyzed the correlations among the descriptors in particular to check for possible biases for certain interactions in the protein-ligand complexes of the PDBbind data. Amongst the ligand descriptors, the strongest correlations were observed between molecular weight, molar refractivity, and Wiener index (MASS, MR, and W). For the protein features, the strongest correlation was between the two descriptors for the aromatic amino acids, PlogP(Arom) and PMR(Arom). In addition, the backbone-based features, [PA(Amide-O) and PD(Amide-NH)], had a high correlation with the log P and molar refractivity values of the non-aromatic residues. Among the hydrogen bond contributions of the amino acids, we observed the strongest correlation with $\rho = 0.66$ between PD(T+S+Y+DH+EH) and PA(N+Q+T+S+DH+EH). This correlation is expected because they share the highest number of amino acids.

A higher correlation between the ligand and the protein features was observed between ligand features that directly scale with the size of the molecule (MASS, W, and MR) and the more general protein features, such as the backbone features and the log P and MR values of the non-aromatic residues. These protein features are expected to be related to the ligand size and, therefore, do not indicate any data set-specific bias of the PDBbind data set.

Comparing the distributions of binding-free energies ΔG between the PDBbind data set used for training and validation and the CSAR 2012 and 2014 (Dunbar et al., 2013; Carlson et al., 2016) and Wang et al. (2015) external data sets used for testing revealed that the PDBbind data set covers a wider range of binding free energies (**Supplementary Figure 1**). In contrast, the 101 protein-ligand complexes from the CSAR NRC-HiQ release cover a wider ΔG range than PDBbind.

From the distribution of the individual descriptors, it is clear that the PDBbind data set encompasses the full range of descriptor values covered by the other data sets (**Supplementary Figure 2**), even though there are differences in the mean values of the descriptors. For example, the average ligand molecular weight was lowest for the CSAR-NRC HiQ data and highest for the D3R data from CSAR 2012 and CSAR 2014.

3.2. Trained Models Random Forests Outperform Neural Networks

Initial tests revealed high variability in the performance metrics that depended on a random training and validation data split. We thus chose a nested cross-validation strategy to find the machine learning models best suited for the chosen descriptors (**Figure 1B**). Therefore, performance metrics are reported as the mean of sixty models resulting from 10 random data set draws and six-fold cross-validation. The corresponding standard deviation enables the quantification of the uncertainty of the performance metrics. Apart from the baseline correlation values between the individual descriptors and the target variable ΔG, we included a null model, which simply predicted the mean ΔG of the training data, to verify predictive power. The root-mean-squared error, *RMSE*, of 2.76 ± 0.05 measured for this null model is identical to the population standard deviation for the respective training folds (**Table 1**, **Figure 3**). The linear regression model derived by ordinary least squares fitting, similar to the original RASPD approach (Mukherjee and Jayaram, 2013), achieved a *RMSE* of 2.19 ± 0.05 kcal/mol on the test set. We tested six other methods and assessed whether they improved on this value.

SVR with a Gaussian radial basis function (RBF) kernel and a neural network with two hidden layers performed with *RMSE* values of 2.04 ± 0.05 kcal/mol and 2.05 ± 0.05 kcal/mol, respectively, these were similar to k-nearest neighbors with an *RMSE* of 2.03 ± 0.04 kcal/mol. Superior performance in terms of both deviation, quantified by *RMSE*, and ranking, as measured by the Spearman correlation ρ, was achieved with the two random forest-based models. The eRF model had a *RMSE* of

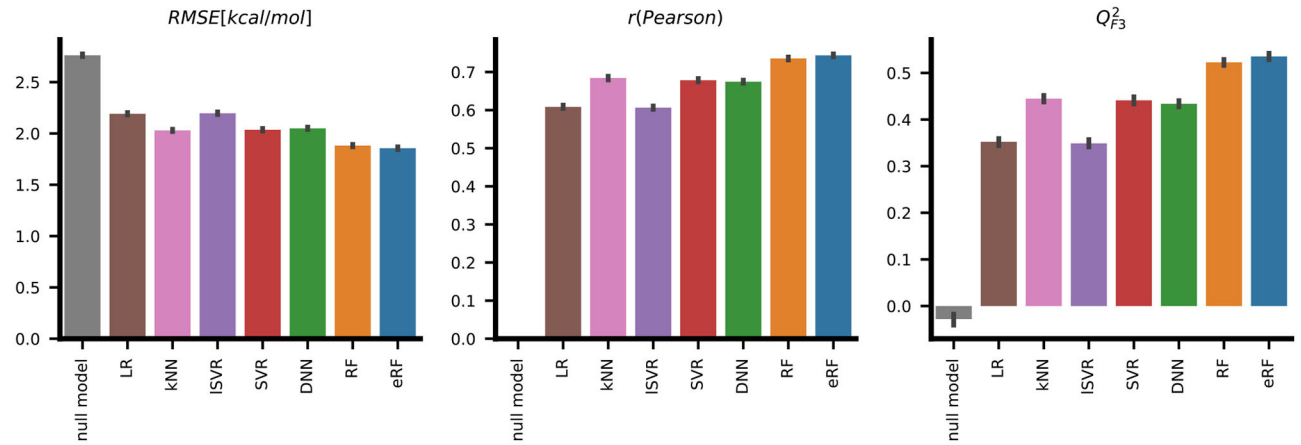


FIGURE 3 | Systematic evaluation of the predictive performance of the seven different machine learning methods on the PDBbind test set shows that, according to three metrics, the extremely random forests (eRF) model performs better than models derived by the other machine learning methods in predicting protein-ligand binding free energy. The error bars indicate the standard deviation for 10 replicates with six-fold cross-validation.

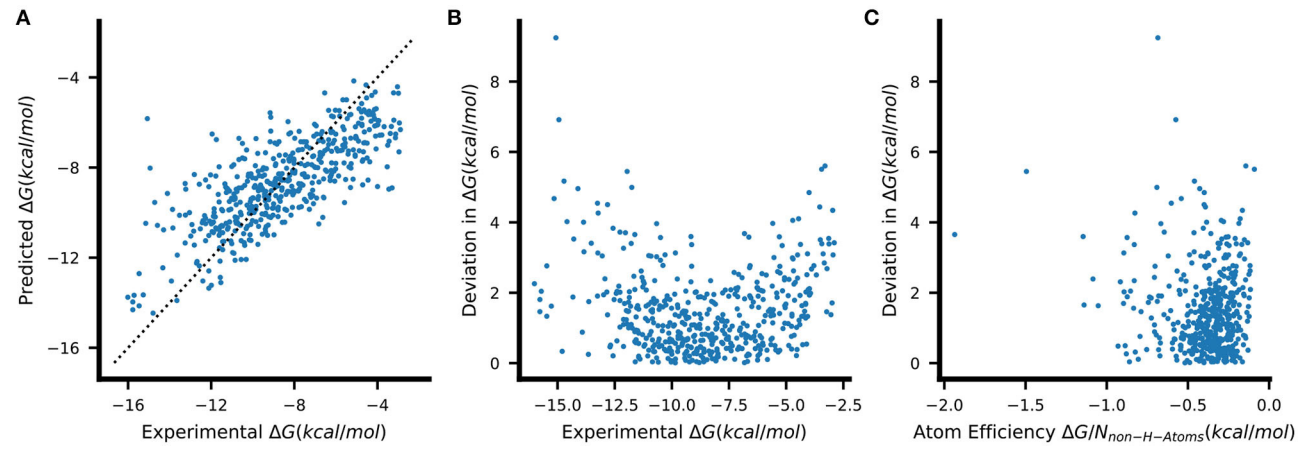


FIGURE 4 | (A) Binding-free energies predicted by a single eRF model on unseen PDBbind test data. The dashed line indicates the ideal prediction. The absolute errors of the predictions shown in **(A)** are plotted against **(B)** the respective true ΔG and **(C)** the atom efficiency, which describes the ΔG contributed on average by each non-hydrogen atom (atom efficiency $= \Delta G/N_{non-H-atoms}$).

1.86 ± 0.05 kcal/mol and a Pearson correlation r of 0.74 ± 0.02, and the RF model performed similarly (**Table 1**, **Figure 3**).

We therefore selected the resulting eRF models for further analysis. We note that these eRF regressors, which use 200 trees and have no limits on the number of samples per leaf, overfit the training set despite showing better validation set performance compared to more strongly regularized variants (see **Supplementary Tables 4, 5**). Nevertheless, an examination of the predictions of the eRF models on the PDBbind test data shows that the general trends in the data are captured although the lowest ΔG values are overestimated, and the highest ΔG values are underestimated (**Figure 4A**). The greatest deviations from the experimental values are thus observed for those complexes with extremely low or high binding free energies (**Figure 4B**). There is, however, no clear relation between having a higher error value and the atom efficiency (**Figure 4C**). The same trends were also observed with all the other machine learning methods.

3.3. Results on External Test Sets

To compare our RASPD+ approach using eRF models as well as LR with existing methods, we performed an evaluation on several external data sets from the literature (Dunbar et al., 2011, 2013; Wang et al., 2015; Carlson et al., 2016) that have different characteristics, as previously done by Jiménez et al. (2018) To compare to other methods for predicting protein-ligand binding free energy, we considered the previous RASPD approach (Mukherjee and Jayaram, 2013) as a method that does not rely on full docking, K_{DEEP} (Jiménez et al., 2018) as a representative deep learning-based method, RF-Score (Ballester and Mitchell, 2010) as a method using random forests, and cyScore (Cao and Li, 2014) and X-Score (Wang et al., 2002) as

TABLE 2 | Comparison of the performance of RASPD+ using eRF and LR models with five other methods to compute protein-ligand binding-free energy.

	RASPD+						
Data set	**eRF**	**LR**	**RASPD**	**KDeep***	**RF-Score***	**CyScore***	**X-Score***
RMSE							
CSAR HiQ 1	3.02 ± 0.04	3.07 ± 0.02	3.43	2.84	**2.71**	3.18	3.15
CSAR HiQ 2	**2.23 ± 0.04**	2.44 ± 0.02	2.79	2.60	2.26	3.00	2.51
CSAR12	1.50 ± 0.02	1.68 ± 0.02	1.93	2.17	1.36	2.84	**1.27**
CSAR14	1.36 ± 0.03	1.64 ± 0.02	2.05	2.39	**1.19**	2.03	1.36
Wang et al.	1.39 ± 0.03	1.39 ± 0.02	2.00	1.47	**1.19**	5.74	1.49
Pearson *r*							
CSAR HiQ 1	0.62 ± 0.02	0.58 ± 0.01	0.54	0.72	**0.77**	0.65	0.60
CSAR HiQ 2	**0.78 ± 0.01**	0.68 ± 0.01	0.67	0.65	0.75	0.64	0.65
CSAR12	0.40 ± 0.03	0.25 ± 0.01	0.29	0.37	0.46	0.26	**0.48**
CSAR14	0.55 ± 0.03	0.23 ± 0.02	0.32	0.61	0.80	0.67	**0.82**
Wang et al.	**0.70 ± 0.02**	0.68 ± 0.01	0.55	0.29	0.24	0.27	0.25

The RMSE [kcal/mol] and Pearson correlation coefficients for predictions on five external test sets are given. RASPD used simpler descriptors and the LR parameters from Mukherjee and Jayaram (2013). The values for the other methods are taken from Jiménez et al. (2018). The values for the best performing models are shown in bold.
**pK values reported by Jiménez et al. (2018) were converted to ΔG for comparison of RMSE values.*

traditional docking scoring functions. Previously reported *RMSE* values (Jiménez et al., 2018) were transformed from errors in *pK* values to errors in ΔG for the comparisons. The RASPD+ eRF model consistently achieved lower error and higher correlation compared to the linear regression using the RASPD+ descriptors and this difference was more pronounced when comparing to the original RASPD linear regression model. With respect to the absolute deviation, measured by *RMSE*, the established scoring functions, RF-Score and X-Score performed best (**Table 2**). Only on set 2 of the challenging CSAR-NRC HiQ release (Dunbar et al., 2011) did RASPD+ with the eRF model have a lower *RMSE*, with a value of 2.23 ± 0.04 kcal/mol, than the existing docking-based methods. When considering the Pearson correlation as a proxy for the ranking performance, RASPD+ with eRF models not only achieved the best result on the CSAR-NRC HiQ set 2 ($r = 0.78 \pm 0.01$) but also achieved $r = 0.70 \pm 0.02$ on the data set curated by Wang et al. (2015) (**Table 2**).

The good performance of the RASPD+ eRF on the Wang et al. data set is also borne out in the distribution of predictions (**Supplementary Figure 3**), which, compared to the results on CSAR-NRC HiQ (**Supplementary Figure 6**), not only ranks but also faithfully captures the range of energies. On both the CSAR 2012 and CSAR 2014 data sets, clear failures of the RASPD+ eRF and most other methods can be observed. For some cases, the RASPD+ model predicts energies in a very narrow range around −10.5 kcal/mol (**Supplementary Figures 4, 5**), but, interestingly, this value does not correspond to the mean ΔG value for the training data.

As the CSAR 2012 (Dunbar et al., 2013) and CSAR 2014 (Carlson et al., 2016) releases and the data set from Wang et al. (2015) provided data for several ligands for each individual protein target, we analyzed the failure cases at the level of the individual proteins (**Supplementary Tables 6–8**). The Pearson and Spearman correlations are below 0.3 for the BACE and CDK2 systems from the Wang et al. set (Wang et al., 2015) and CHK1 and SYK in the CSAR sets. In contrast, the CDK2 complexes in the CSAR 2012 set (Dunbar et al., 2013) achieved a Pearson correlation of $r = 0.50 \pm 0.05$. The highest correlations were observed for the PTP1B, Mcl-1, TYK2 systems in the Wang et al. (2015) data (**Supplementary Table 6**) and for the CDK2-Cyclin A complex (**Supplementary Table 7**) and TrmD on the CSAR data (**Supplementary Table 8**). Strikingly, only for PTP1B, TYK2, and TrmD was $R^2 > 0.3$ observed while all Q^2_{F3} values were above 0.5.

3.4. Feature Importance Analysis

To assess which features contribute to accurate predictions, two strategies were chosen. By permutation feature importance, the contribution to the prediction was quantified by the change in the Pearson correlation coefficient after shuffling the values in the individual feature columns randomly. Three different model types – namely, linear regression, support vector machine, and extremely random forests – showed different relative contributions of the individual features (**Figure 5A**). While LR assigned high contributions to a few features, the reduction in predictive performance for each shuffled feature was lower for eRF and the contribution signal was more evenly distributed among the different features. Molar refractivity (MR), which was the feature most strongly correlated with the target variable ΔG , showed the strongest effect in the LR, lSVR, and eRF models. For LR, randomizing MR almost completely removed the predictive power ($r < 0.2$). Among the protein features, the LR, SVR, and RF methods showed high contributions for the general descriptor PlogP(Non-Arom) (**Figure 5A**). While both SVR and eRF assigned high contributions to the PMR(Non-Arom), LR, and lSVR placed

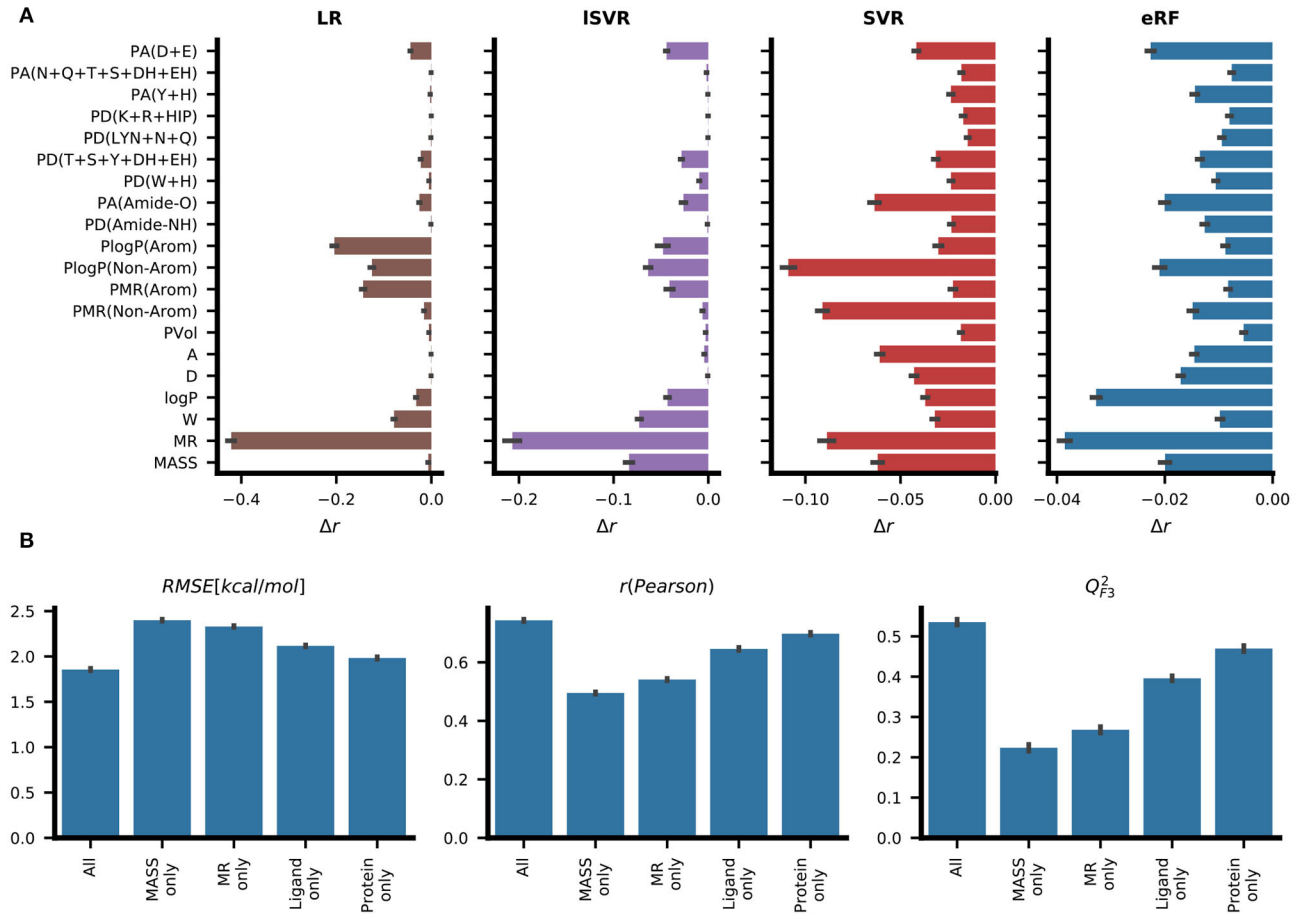


FIGURE 5 | Analysis of feature importance for the predictive performance of RASPD+. **(A)** The average change in predictive performance as measured by Pearson correlation when the corresponding feature column was shuffled. Results are reported for linear regression (LR), linear support vector regression (lSVR), support vector regression (SVR), and extremely random forests (eRF). **(B)** Ablation based analysis entailing training eRF models with different feature sets. The models trained on just the values for molar refractivity (MR only) and molecular weight (MASS only) serve as lower bounds for measuring performance. The "ligand only" and "protein only" models trained on either only ligand or only protein features, respectively, perform better than the "MR only" and "MASS only" models but not as well as the models derived from all features. The protein features contain implicit information on the ligand size and this may indicate why the performance of the "protein only" model is better than that of the "ligand only" model (see **Figure 1** for details).

higher contributions on PlogP(Arom) and PMR(Arom) among the general protein features. PA(Amide-O), which had the second-highest correlation with ΔG , showed a pronounced signal for the editedSVR and eRF models. The hydrogen bond acceptor count at the negatively charged amino acid residues [PA(D+E)] was informative for all these machine learning methods. In the eRF model, it had an importance value similar to the general protein features, such as the residue log P values. This is especially surprising as no information on the ligand charge was provided, and the count of positively charged amino acid hydrogen bond donors [PD(K+R+HIP)] did not contribute strongly to the predictions.

Additionally, we trained eRF models on subsets of the features and compared their performance to the full model (**Figure 5B**). Among the models trained on a single feature, the model trained on molar refractivity (MR) achieved better performance than that trained on molecular weight (MASS). Models trained on just the features of the protein pocket performed better than models using only ligand descriptors. In this case, the protein features still contained information about the ligand implicitly, as each protein descriptor is dependent on the size of the sphere surrounding the ligand. These reduced feature set models were also subjected to permutation feature importance analysis. For models with only ligand features, a very similar ranking of ligand features compared to the full training set was observed, illustrating the general preference for using those features for prediction (**Supplementary Figure 7**).

When examining the feature importance for protein-only models, the backbone hydrogen bond acceptor [PA(Amide-O)] stands out compared to the feature importance on the full feature set (**Supplementary Figure 8**). This could be partially explained by the fact that this feature showed a strong correlation

TABLE 3 | Average enrichment factors with corresponding standard deviations for the top 1, 5, and 10% of the data selected from the DUD-E systems.

	Full DUD-E $n_{systems} = 102$			w/o cofactor/surface binding $n_{systems} = 55$		
Method	**EF** 1%	**EF** 5%	**EF** 10%	**EF** 1%	**EF** 5%	**EF** 10%
eRF	1.8 ± 2.5	1.5 ± 1.4	1.3 ± 0.9	2.0 ± 3.0	1.7 ± 1.5	1.0 ± 1.0
RF	1.9 ± 2.4	1.5 ± 1.5	1.3 ± 1.0	2.3 ± 2.8	1.7 ± 1.7	1.4 ± 1.1
DNN	2.0 ± 2.0	1.4 ± 1.2	1.3 ± 0.9	1.8 ± 2.3	1.6 ± 1.3	1.4 ± 0.9
kNN	1.5 ± 1.9	1.3 ± 1.3	1.2 ± 0.9	1.8 ± 2.4	1.6 ± 1.6	1.0 ± 1.0
lSVR	2.6 ± 3.1	1.9 ± 1.6	1.6 ± 1.1	2.6 ± 3.3	2.0 ± 1.7	1.7 ± 1.2
SVR	2.3 ± 3.5	1.6 ± 1.5	1.4 ± 1.0	3.1 ± 4.3	1.9 ± 1.8	1.5 ± 1.1
LR	2.7 ± 3.4	2.0 ± 1.7	1.7 ± 1.2	2.9 ± 3.7	2.1 ± 1.8	1.8 ± 1.3
RASPD	4.1 ± 4.0	2.2 ± 1.7	1.7 ± 1.1	4.2 ± 4.2	2.3 ± 1.7	1.7 ± 1.2
Mean ensemble	2.0 ± 3.0	1.8 ± 1.6	1.6 ± 1.1	2.6 ± 3.4	2.0 ± 1.8	1.7 ± 1.1
Union	2.4 ± 2.3	1.8 ± 1.4	1.6 ± 1.0	2.8 ± 2.6	2.1 ± 1.6	1.8 ± 1.1
Union w/o kNN	2.5 ± 2.7	1.8 ± 1.5	2.0 ± 1.0	3.0 ± 3.0	2.1 ± 1.7	1.7 ± 1.1
Union Top 3	2.8 ± 3.6	1.9 ± 1.7	1.6 ± 1.2	3.3 ± 4.2	2.0 ± 2.0	1.8 ± 1.3

Union is the enrichment achieved by selecting the non-redundant set of candidate compounds obtained by combining the selections of each method. Performance when excluding cofactor and surface binding sites is also reported. Values are also given for the original RASPD method.

with general ligand features (**Figure 2**) and thereby provides information related to general ligand size.

3.5. Enrichment of Active Molecules From the DUD-E Data Set

To assess the usefulness of our RASPD+ method, we simulated a drug discovery setting using the benchmark DUD-E data set, which contains several computationally generated decoys per active compound (Mysinger et al., 2012). For each of the seven machine learning models, we calculated enrichment factors (EF) to quantify how effective ranking by predicted binding free energies was at enriching active molecules from the whole data set (**Table 3**). We also compared the RASPD+ results with those of RASPD (Mukherjee and Jayaram, 2013) and found that the linear regression models of both RASPD+ and RASPD were the most effective when filtering to 1, 5, and 10% of the samples, with EFs of 2.7 ± 3.4 and 4.1 ± 4.0, respectively, when filtering down to 1% of the samples. The high standard deviation in the mean EF resulted from high variability in the performance of different methods on individual proteins (**Supplementary Table 9**). As methods that ranked on average less favorably provided the only acceptable enrichment on some of the systems, we chose a conservative approach to interpreting the results by combining the predictions of all the methods. We thus also considered the union of the sets of top candidate molecules from all seven machine learning models. This combination achieved an enrichment at 1% of 2.4 ± 2.3, similar to the linear methods. By excluding the predictions of the worst-performing method kNN (Union w/o kNN), this set improved to 2.5 ± 2.7. When only combining the predictions of the three methods performing best on the DUD-E set LR, lSVR, and SVR (Union Top 3), this further increased to 2.8 ± 3.6. For comparison, the performance of scoring functions functions based on the docked structures of ligand-protein complexes on the DUD-E set has been assessed by Chen et al. (2019). The highest early (1%) and late (10%) enrichment factors were 6.67 and 2.55, respectively, and were obtained using the knowledge-based DLIGAND2 scoring method, whereas the corresponding values obtained with the widely used AutoDock Vina scoring function were 5.12 and 2.60. The late EF obtained with the RASPD+ union w/o kNN approach is similar to that of these docking-based methods.

One of the reasons for some of the poor predictions with DUD-E is that, in contrast to the training with PDBbind, the query ligands may bind at a different position to the co-crystallized ligand in the target whose center of mass is used to define the binding site for which protein properties are computed. If the query ligand binds in a somewhat different position, the computed protein features may not be so relevant. From the feature importance analysis (**Figure 5**), we see that for the eRF model, all the features contribute in a similar way to the final prediction. In contrast, for the LR and lSVR models, the dominant contributions to the prediction were from ligand molar refractivity and just three out of the 15 protein features. Since, in the LR and lSVR models, only a few protein features contribute to the final score, erroneous protein features may have less impact on the final predicted value compared to the random forest-based models. The average EF values for the LR method, and for the original RASPD LR model, for all the DUD-E sets are therefore higher than for the other methods. Another reason for low EF values for some targets is the presence of cofactors or structural water molecules in the binding site in some proteins as well as highly solvent exposed binding sites. To assess how much the performance is impacted by situations not properly modeled by RASPD+, we also considered whether a cofactor in the binding site or a mostly solvent-exposed surface binding site affects performance. For most methods, the exclusion of those challenging pockets, which by design could not be fully

TABLE 4 | Mean enrichment factors for different subsets of the DUD-E set.

		Union			Union top 3		
Target subset	**Count**	**EF** 1%	**EF** 5%	**EF** 10%	**EF** 1%	**EF** 5%	**EF** 10%
Cytochrome P450	2	1.86	1.25	1.28	1.11	1.65	1.42
GPCR	5	4.30	2.94	2.18	7.37	3.61	2.61
Ion channel	2	5.54	2.15	1.61	7.67	2.16	1.56
Kinase	23	3.04	2.55	2.15	4.43	3.03	2.49
Metal containing enzyme	18	2.47	1.66	1.42	2.89	1.84	1.55
Miscellaneous	30	1.83	1.49	1.41	1.57	1.34	1.12
Nuclear receptor	11	1.73	1.18	1.11	1.14	0.93	1.01
Protease	11	1.56	1.62	1.41	1.91	1.75	1.60

Performance reported both the union over the predictions of all methods and the union of the predictions from the three best ranking methods (LR, lSVR, and SVR).

modeled with RASPD and RASPD+ descriptors, improved the mean performance (**Table 3**).

Additionally, we analyzed the performance of the different protein subgroups in the DUD-E set (**Table 4**). Here we observed the lowest average performance for the protease subgroup when considering the union over all methods and when considering only the top three union, cytochrome P450, and nuclear receptor targets were the groups with the lowest enrichment. The poor performance for cytochrome P450s may be due to the heme cofactor in their binding site whereas it is notable that the eight proteases (out of the 11 in the DUD-E data set) with low EF factors have ligands that are solvent exposed in the crystal structure or there are structural water molecules bridging between polar atoms of the ligand and the protein. The highest enrichment was observed for ion channels, G-protein coupled receptors (GPCR), and kinases in both settings (**Table 4**).

4. DISCUSSION

As the global health crisis surrounding the SARS-CoV-2 pandemic (Wu et al., 2020) has demonstrated, there is a need for fast computational tools to accelerate drug design and development processes. The method we present here, RASPD+, is able to perform virtual screening of large libraries of compounds (Irwin and Shoichet, 2005; Wishart et al., 2006) at a fraction of the time typically required for protein-ligand docking methods. This enables quick prioritization of candidates for a follow up with more accurate yet computationally more demanding methods, such as docking. We achieved the speed up by training machine learning models on simple pose-invariant ligand and protein descriptors. With this simplified approach, we achieved results comparable to existing scoring functions (Wang et al., 2002; Ballester and Mitchell, 2010; Cao and Li, 2014; Jiménez et al., 2018) when predicting the binding free energy, ΔG , on several data sets. By splitting the PDBbind training, testing, and validation data in a nested cross-validation setup, we were able to assess reliably that random forest models, particularly the extremely random forest model, performed best on this type of data. While this splitting strategy increases confidence in the comparison of learning methods and feature importance analysis within the study, other data set splitting strategies, which explicitly control how similar proteins or ligands are between training and test sets (Feinberg et al., 2018; Sieg et al., 2019; Su et al., 2020), may be more appropriate to assess performance on completely different ligands or proteins directly.

We accounted for this deficiency by not only testing the regression performance on different external test sets but also by assessing the ability of the RASPD+ models to enrich active molecules from a set of inactive decoys. Although the achievable enrichment factors were not as high as state-of-the-art docking or free energy prediction methods (Li et al., 2014), RASPD+ still displayed appreciable enrichment of active molecules on the DUD-E data set (Mysinger et al., 2012). RASPD+ was able-without sampling docking poses-to achieve similar performance to an older scoring function in a docking method comparison (Li et al., 2013; Chen et al., 2019). This is remarkable for two reasons: First, the training set only includes molecules displaying binding to their specific target protein. Secondly, four of the six physicochemical descriptors (molecular weight, hydrogen bond donor and acceptor count, and logP value), used to describe the ligand molecule, were initially used to select decoys similar to the active molecules for the DUD-E data set (Huang et al., 2006; Mysinger et al., 2012). This makes the task of distinguishing active and inactive molecules particularly difficult for our models that employ only basic ligand descriptors (Lagarde et al., 2015). Notably, however, molar refractivity (MR), which was not used for the creation of the DUD-E decoys, was not only a powerful predictor on its own ($r > 0.5$) but was also consistently assigned the highest feature importance among the ligand features. The high importance of MR agrees with results from a recent study that used ligand descriptors to enhance the performance of a common docking scoring function (Boyles et al., 2020).

Not considering pockets containing metal ions or other cofactors, which are not taken into account by the simple RASPD+ descriptors, yielded slightly higher average enrichment than on the full DUD-E set. Random forest methods, which were best suited for the ΔG regression on known binders, were for most proteins outperformed by the simpler linear regression methods. This observation might support the recent finding that random forest methods, in particular, benefit from highly similar training molecules (Su et al., 2020). Considering

the strengths and weaknesses of the different machine learning methods, we therefore recommend that for applications of RASPD+, the results of the seven different machine learning methods are combined by picking top candidates from the rankings produced by each method. For this, we demonstrated different combinations using the union of selection sets from the different methods.

If this approach is applied to pick the top 10 % of RASPD+ candidates, this can provide a 10-fold reduction in the time spent for docking. Notably, we achieved computation times for RASPD+ that were over 100 times faster than Glide SP docking (Friesner et al., 2004) (Schrödinger Release 2019-4: Glide, Schrödinger, LLC, New York, NY) on a laptop grade CPU (data not shown), meaning that computation times for RASPD+ screening are negligible compared to times for docking and molecular dynamics simulation.

Thus, the use of RASPD+ is clearly beneficial in time-critical applications of virtual screening of large compound libraries against individual protein targets. Moreover, higher structure-based screening throughput could also enable more effective inverse virtual screening of protein databases to assess the specificity and potential side-effects of candidate molecules.

DATA AVAILABILITY STATEMENT

All data generated in this work necessary to reproduce the findings are available at https://doi.org/10.5281/zenodo.3937425 and the RASPD+ software is available on GitHub at https://github.com/HITS-MCM/RASPDplus.

AUTHOR CONTRIBUTIONS

GM developed the descriptor pipeline and performed enrichment analysis. LA and SH trained and evaluated machine learning models. SH performed feature importance analysis. SH, LA, GM, and RW analyzed the data. BJ provided advice for the development of the descriptor pipeline. RW provided guidance and supervised the work with GM. SH wrote the manuscript with input from all authors.

FUNDING

This work was supported by the Klaus Tschira Foundation.

ACKNOWLEDGMENTS

We thank Dr. Daria Kokh and Jui-Hung Yuan for providing code from the TRAPP software (Kokh et al., 2013; Yuan et al., 2020) for the calculation of binding pocket volume and Dr. Stefan Richter for technical assistance with the software and setting up the computing environment for the calculations. We thank Dr. Abhilash Jayaraj for help with program compilation. We thank Dr. Satyanarayana Rao and Vidhu S. Pandey for providing helpful utility code. We thank all those who made data on protein-ligand binding available in the databases that we have used. This work has been released as a pre-print at ChemRxiv (Holderbach et al., 2020).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2020.601065/full#supplementary-material

The code used for creating the models and figures in this paper, and scripts to run the RASPD+ pipeline are available on GitHub: https://github.com/HITS-MCM/RASPDplus. Precomputed descriptor values and the model weights can be downloaded from https://doi.org/10.5281/zenodo.3937425.

REFERENCES

Ballester, P. J., and Mitchell, J. B. O. (2010). A machine learning approach to predicting protein-ligand binding affinity with applications to molecular docking. *Bioinformatics* 26, 1169–1175. doi: 10.1093/bioinformatics/btq112

Baum, B., Mohamed, M., Zayed, M., Gerlach, C., Heine, A., Hangauer, D., et al. (2009). More than a simple lipophilic contact: a detailed thermodynamic analysis of nonbasic residues in the S1 pocket of thrombin. *J. Mol. Biol.* 390, 56–69. doi: 10.1016/j.jmb.2009.04.051

Boyles, F., Deane, C. M., and Morris, G. M. (2020). Learning from the ligand: using ligand- based features to improve binding affinity prediction. *Bioinformatics* 36, 758–764. doi: 10.26434/chemrxiv.8174525.v1

Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32. doi: 10.1023/A:1010933404324

Cao, Y., and Li, L. (2014). Improved protein-ligand binding affinity prediction by using a curvature-dependent surface-area model *Bioinformatics* 30, 1674–1680. doi: 10.1093/bioinformatics/btu104

Carlson, H. A., Smith, R. D., Damm-Ganamet, K. L., Stuckey, J. A., Ahmed, A., Convery, M. A., et al. (2016). CSAR 2014: a benchmark exercise using unpublished data from pharma. *J. Chem. Inf. Model.* 56, 1063–1077. doi: 10.1021/acs.jcim.5b00523

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, 1668–1688. doi: 10.1002/jcc.20290

Chen, P., Ke, Y., Lu, Y., Du, Y., Li, J., Yan, H., et al. (2019). DLIGAND2: an improved knowledge-based energy function for protein-ligand interactions using the distance-scaled, finite, ideal-gas reference state. *J. Cheminform.* 11:52. doi: 10.1186/s13321-019-0373-4

Chollet, F., et al. (2015). *Keras.* Available online at: https://keras.io

Consonni, V., Ballabio, D., and Todeschini, R. (2009). Comments on the definition of the Q2 parameter for QSAR validation. *J. Chem. Inf. Model.* 49, 1669–1678. doi: 10.1021/ci900115y

Cumming, J. N., et al. (2012). Structure based design of iminohydantoin BACE1 inhibitors: identification of an orally available, centrally active BACE1 inhibitor. *Bioorgan. Med. Chem. Lett.* 22, 2444–2449. doi: 10.1016/j.bmcl.2012.02.013

Drucker, H., Burges, C. J. C., Kaufman, L., Smola, A., and Vapnik, V. (1997). "Support vector regression machines," in *Advances in Neural Information Processing Systems 9: Proceedings of the 1996 Conference.* Available online at: http://papers.nips.cc/paper/1238-support-vector-regression-machines.pdf; http://books.google.com/bookl=en&lr=&id=QpD7n95ozWUC&pgis=1

Dunbar, J. B., Smith, R. D., Damm-Ganamet, K. L., Ahmed, A., Esposito, E. X., Delproposto, J., et al. (2013). CSAR data set release 2012: ligands, affinities, complexes, and docking decoys. *J. Chem. Inf. Model.* 53, 1842–1852. doi: 10.1021/ci4000486

Dunbar, J. B., Smith, R. D., Yang, C.-Y., Ung, P. M.-U., Lexa, K. W., Khazanov, N. A., et al. (2011). CSAR benchmark exercise of 2010: selection of the protein-ligand complexes. *J. Chem. Inf. Model.* 51, 2036–2046. doi: 10.1021/ci200082t

Feinberg, E. N., Sur, D., Wu, Z., Husic, B. E., Mai, H., Li, Y., et al. (2018). PotentialNet for molecular property prediction. *ACS Central Sci.* 4, 1520–1530. doi: 10.1021/acscentsci.8b00507

Friberg, A., Vigil, D., Zhao, B., Daniels, R. N., Burke, J. P., Garcia-Barrantes, P. M., et al. (2013). Discovery of potent myeloid cell Leukemia 1 (Mcl-1) inhibitors using fragment-based methods and structure-based design. *J. Med. Chem.* 56, 15–30. doi: 10.1021/jm301448p

Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., et al. (2004). Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* 47, 1739–1749. doi: 10.1021/jm0306430

Gathiaka, S., Liu, S., Chiu, M., Yang, H., Stuckey, J. A., Kang, Y. N., et al. (2016). Evaluation of protein-ligand pose and affinity predictions. *J. Comput. Aided Mol. Des.* 30, 651–668. doi: 10.1007/s10822-016-9946-8

Geurts, P., Ernst, D., and Wehenkel, L. (2006). Extremely randomized trees. *Mach. Learn.* 63, 3–42. doi: 10.1007/s10994-006-6226-1

Ghose, A. K., and Crippen, G. M. (1987). Atomic physicochemical parameters for threedimensional- structure-directed quantitative structure-activity relationships. 2. Modeling dispersive and hydrophobic interactions. *J. Chem. Inf. Model.* 27, 21–35.

Goldstein, D. M. et al. (2011). Discovery of 6-(2,4-di uorophenoxy)-2-[3-hydroxy-1-(2-hydroxyethyl) propylamino]-8-methyl-8 H -pyrido[2,3- d]pyrimidin-7-one (pamapimod) and 6-(2,4-diuorophenoxy)-8-methyl-2-(tetrahydro-2 H-pyran-4-ylamino)pyrido[2,3-d]pyrimidin-7(8 H)-one (R1487). *J. Med. Chem.* 54, 2255–2265. doi: 10.1021/jm101423y

Gomes, J., Ramsundar, B., Feinberg, E. N., and Pande, V. S. (2017). Atomic convolutional networks for predicting protein-ligand binding affinity. *[arXiv preprint]*. arXiv:1703.10603.

Holderbach, S., Adam, L., Jayaram, B., Wade, R., and Mukherjee, G. (2020). RASPD+: fast protein-ligand binding free energy prediction using simplified physicochemical features. *ChemRxiv Preprint*. doi: 10.26434/chemrxiv.12636704.v1

Huang, N., Shoichet, B. K., and Irwin, J. J. (2006). Benchmarking sets for molecular docking. *J. Med. Chem.* 49, 6789–6801. doi: 10.1021/jm0608356

Irwin, J. J., and Shoichet, B. K. (2005) ZINC:a free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.* 45, 177–182. doi: 10.1021/ci049714+

Jiménez, J., Škalič, M., Martínez-Rosell, G., and De Fabritiis, G. (2018). DEEP : protein- ligand absolute binding affinity prediction via 3D-convolutional neural networks. *J. Chem. Inf. Model.* 58, 287–296. doi: 10.1021/acs.jcim.7b00650

Kokh, D. B., Richter, S., Henrich, S., Czodrowski, P., Rippmann, F., and Wade, R. C. (2013). TRAPP: a tool for analysis of transient binding pockets in proteins. *J. Chem. Inf. Model.* 53, 1235–1252. doi: 10.1021/ci4000294

Kotila, M. (2018). *Talos.* Available online at: https://github.com/autonomio/talos

Lagarde, N., Zagury, J.-F., and Montes, M. (2015). Benchmarking data sets for the evaluation of virtual ligand screening methods: review and perspectives. *J. Chem. Inf. Model.* 55, 1297–1307. doi: 10.1021/acs.jcim.5b00090

Li, G.-B., Yang, L.-L., Wang, W.-J., Li, L.-L., and Yang, S.-Y. (2013). ID-score: a new empirical scoring function based on a comprehensive set of descriptors related to protein-ligand interactions. *J. Chem. Inf. Model.* 53, 592–600. doi: 10.1021/ci300493w

Li, J., Fu, A., and Zhang, L. (2019). An overview of scoring functions used for protein ligand interactions in molecular docking. *Interdisc. Sci.* 11, 320–328. doi: 10.1007/s12539-019-00327-w

Li, Y., Han, L., Liu, Z., and Wang, R. (2014). Comparative assessment of scoring functions on an updated benchmark: 2. evaluation methods and general results. *J. Chem. Inf. Model.* 54, 1717–1736. doi: 10.1021/ci500081m

Liang, J., Tsui, V., Van Abbema, A., Bao, L., Barrett, K., Beresini, M., et al. (2013a). Lead identification of novel and selective TYK2 inhibitors. *Eur. J. Med. Chem.* 67, 175–187. doi: 10.1016/j.ejmech.2013.03.070

Liang, J., van Abbema, A., Balazs, M., Barrett, K., Berezhkovsky, L., Blair, W., et al. (2013b). Lead optimization of a 4-aminopyridine benzamide scaffold to identify potent, selective, and orally bioavailable TYK2 inhibitors. *J. Med. Chem.* 56, 4521–4536. doi: 10.1021/jm400266t

Lipinski, C. A., Lombardo, F., Dominy, B. W., and Feeney, P. J. (2001). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* 46, 3–26. doi: 10.1016/S0169-409X(00)00129-0

Liu, Z., Li, Y., Han, L., Li, J., Liu, J., Zhao, Z., (2015). PDB-wide collection of binding data: current status of the PDBbind database. *Bioinformatics* 31, 405–412. doi: 10.1093/bioinformatics/btu626

Mukherjee, G., and Jayaram, B. (2013). A rapid identification of hit molecules for target proteins via physico-chemical descriptors. *Phys. Chem. Chem. Phys.* 15:9107. doi: 10.1039/c3cp44697b

Mysinger, M. M., Carchia, M., Irwin, J. J., and Shoichet, B. K. (2012). Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking. *J. Med. Chem.* 55, 6582–6594. doi: 10.1021/jm300687e

O'Boyle, N. M., Banck, M., James, C. A., Morley, C., Vandermeersch, T., Hutchison, G. R. (2011). Open Babel: an open chemical toolbox *J. Cheminform.* 3:33. doi: 10.1186/1758-2946-3-33

Pedregosa, F., Varoquaux, G., Michel, V., Thirion, B., Grisel, O., Blondel, M., et al. (2011). Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830.

Pei, J., Zheng, Z., Kim, H., Song, L. F., Walworth, S., Merz, M. R., et al. (2019). Random forest refinement of pairwise potentials for protein–ligand decoy detection. *J. Chem. Inf. Model.* 59, 3305–3315. doi: 10.1021/acs.jcim.9b00356

Sieg, J., Flachsenberg, F., and Rarey, M. (2019). In need of bias control: evaluating chemical data for machine learning in structure-based virtual screening. *J. Chem. Inf. Model.* 59, 947–961. doi: 10.1021/acs.jcim.8b00712

Smith, R. D., Dunbar, J. B., Ung, P. M.-U., Esposito, E. X., Yang, C.-Y., Wang, S., et al. (2011). CSAR benchmark exercise of 2010: combined evaluation across all submitted scoring functions. *J. Chem. Inf. Model.* 51, 2115–2131. doi: 10.1021/ci200269q

Su, M., Feng, G., Liu, Z., Li, Y., and Wang, R. (2020). Tapping on the black box: how is the scoring power of a machine-learning scoring function depended on the training set? *J. Chem. Inf. Model.* 60, 1122–1136. doi: 10.1021/acs.jcim.9b00714

Todeschini, R., Ballabio, D., and Grisoni, F. (2016). Beware of unreliable Q2 ! A comparative study of regression metrics for predictivity assessment of QSAR models. *J. Chem. Inf. Model.* 56, 1905–1913. doi: 10.1021/acs.jcim.6b00277

Torres, P. H. M., Sodero, A. C. R., Jofily, P., and Silva, F. P. Jr. (2019). Key topics in molecular docking for drug design. *Int. J. Mol. Sci.* 20:4574. doi: 10.3390/ijms20184574

Vieira, T. F., and Sousa, S. F. (2019). Comparing autoDock and vina in ligand/decoy discrimination for virtual screening. *Appl. Sci.* 9:4538. doi: 10.3390/app9214538

Wang, L., Deng, Y., Knight, J. L., Wu, Y., Kim, B., Sherman, W., et al. (2013). Modeling local structural rearrangements using FEP/REST: application to relative binding affinity predictions of CDK2 inhibitors *J. Chem. Theory Comput.* 9, 1282–1293. doi: 10.1021/ct300911a

Wang, L., Wu, Y., Deng, Y., Kim, B., Pierce, L., Krilov, G., et al. (2015). Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field. *J. Am. Chem. Soc.* 137, 2695–2703. doi: 10.1021/ja512751q

Wang, R., Fang, X., Lu, Y., and Wang, S. (2004). The PDBbind database: collection of binding affinities for protein-ligand complexes with known three-dimensional structures. *J. Med. Chem.* 47, 2977–2980. doi: 10.1021/jm030580l

Wang, R., Lai, L., and Wang, S. (2002). Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *J. Comput. Aided Mol. Des.* 16, 11–26. doi: 10.1023/A:1016357811882

Wang, Z., Sun, H., Shen, C., Hu, X., Gao, J., Li, D., et al. (2020). Combined strategies in structure-based virtual screening. *Phys. Chem. Chem. Phys.* 22, 3149–3159. doi: 10.1039/C9CP06303J

Wiener, H. (1947). Structural determination of paraffin boiling points. *J. Am. Chem. Soc.* 69, 17–20. doi: 10.1021/ja01193a005

Wildman, S. A., and Crippen, G. M. (1999). Prediction of physicochemical parameters by atomic contributions. *J. Chem. Inf. Comput. Sci.* 39, 868–873. doi: 10.1021/ci990307l

Willems, H., De Cesco, S., and Svensson, F. (2020). Computational Chemistry on a Budget: supporting drug discovery with limited resources. *J. Med. Chem.* 63, 10158–10169. doi: 10.1021/acs.jmedchem.9b02126

Wilson, D. P., et al. (2007). Structure-based optimization of protein tyrosine phosphatase 1B inhibitors: from the active site to the second phosphotyrosine binding site. *J. Med. Chem.* 50, 4681–4698. doi: 10.1021/jm0702478

Wishart, D. S., Knox, C., Guo, A. C., Shrivastava, S., Hassanali, M., Stothard, P., et al. (2006). DrugBank: a comprehensive resource for in silico drug

discovery and exploration. *Nucleic Acids Res.* 34, D668–D672. doi: 10.1093/nar/gkj067

Wu, F., Zhao, S., Yu, B., Chen, Y. M., Wang, W., Song, Z. G., et al. (2020). A new coronavirus associated with human respiratory disease in China. *Nature* 579, 265–269. doi: 10.1038/s41586-020-2008-3

Yang, X., Wang, Y., Byrne, R., Schneider, G., and Yang, S. (2019). Concepts of artificial intelligence for computer-assisted drug discovery. *Chem. Rev.* 119, 10520–10594. doi: 10.1021/acs.chemrev.8b00728

Yuan, J.-H., Han, S. B., Richter, S., Wade, R. C., and Kokh, D. B. (2020). Druggability assessment in TRAPP using machine learning approaches. *J. Chem. Inf. Model.* doi: 10.1021/acs.jcim.9b01185

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.

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