INSIGHTS IN STRUCTURAL BIOLOGY: 2021

EDITED BY: Annalisa Pastore, Caterina Alfano and Piero Andrea Temussi
PUBLISHED IN: Frontiers in Molecular Biosciences

frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-83250-571-7
DOI 10.3389/978-2-83250-571-7

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

INSIGHTS IN STRUCTURAL BIOLOGY: 2021

Topic Editors:
Annalisa Pastore, King's College London, United Kingdom
Caterina Alfano, Ri.MED Foundation, Italy
Piero Andrea Temussi, University of Naples Federico II, Italy

Citation: Pastore, A., Alfano, C., Temussi, P. A., eds. (2022). Insights in Structural Biology: 2021. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-83250-571-7

Table of Contents

frontiers | Frontiers in Molecular Biosciences

TYPE Editorial
PUBLISHED 07 October 2022
DOI 10.3389/fmolb.2022.1020473

OPEN ACCESS

EDITED AND REVIEWED BY
Anastassios C. Papageorgiou,
University of Alberta, Canada

*CORRESPONDENCE
Caterina Alfano,
calfano@fondazionerimed.com

SPECIALTY SECTION
This article was submitted to Structural Biology,
a section of the journal
Frontiers in Molecular Biosciences

RECEIVED 16 August 2022
ACCEPTED 20 September 2022
PUBLISHED 07 October 2022

CITATION
Alfano C, Pastore A and Temussi PA (2022), Editorial: Insights in Structural Biology: 2021.
Front. Mol. Biosci. 9:1020473.
doi: 10.3389/fmolb.2022.1020473

Editorial: Insights in Structural Biology: 2021

Caterina Alfano[1]*, Annalisa Pastore[2] and Piero Andrea Temussi[3]

[1]Structural Biology and Biophysics Unit, Fondazione Ri.MED, Palermo, Italy, [2]European Synchrotron Radiation Facility, Grenoble, France, [3]University of Naples "Federico II", Napoli, Italy

KEYWORDS

structural biology, molecular dynamics, NMR, mass spectrometry, *in vivo*

Editorial on the Research Topic
Insights in structural biology: 2021

The common theme of this Research Topic (RT) was to discuss examples of the most recent achievements made by scientists in the fast-growing field of Structural Biology. In particular, the goal of this special edition Research Topic was to shed light on the progress made in the past decade in Structural Biology and to provide an overview of the state-of-the-art of the field. We think that the articles received will inspire, inform and provide new directions and guidance to researchers.

One of the fields in which Structural Biology has become mature in within the last few years covers new approaches in computational methods.

Two of the papers accepted in the present research topic are those by the group of Patrick Senet. Both papers use a new software that was developed in this laboratory to extract conformational properties of proteins on the basis of the coordinates of α-carbons. The program is called CUTABI (CUrvature and Torsion based Alpha-helix and Beta-sheet Identification) and identifies residues in α-helices and β-sheets using solely α-carbon coordinates. This algorithm is ideal for the difficult cases in which the content of secondary structure is not directly determined.

The first of these papers (RT_1) by Guzzo et al. (2021) deals with α-synuclein and three single point mutations (A39P, A53T and E46K) of this protein related to familial forms of Parkinson disease (PD). The authors addressed the problem of identifying differences in conformational tendencies, possibly related to different tendencies to aggregate in β-sheet structures, that is the hallmark of PD. They calculated two-dimensional probability density maps and evaluated their differences by molecular dynamics trajectories. In the observed conformational states, the authors found some with a two-phase characteristics and a homogeneous (B, only β-sheets) and a heterogenous phase (HB, a mixture of α-helices and β-sheets). The B state is populated by 40% in the wild-type and in two mutants (A35T and A30P) whereas it is present only as a 25% population in the A53T mutant. The A53T mutant has also a rather larger propensity to forming helices than the wild-type and other mutants. The authors concluded that the equilibrium between the different conformations of the α-synuclein monomer is modified by the missense mutations in a subtle way. It is highly

significant that these authors could reach this conclusion in the case of an intrinsically unfolded protein because other parameters, such as the average gyration radius, could not be used to discriminate among conformational ensembles. The introduction of the algorithm CUTABI is promising for studies of intrinsically unfolded proteins.

The second paper by the same group (RT_9) focused on the properties of dimers of α-synuclein and the three single point mutations discussed in RT_1, i.e., A30P, A53T and E46K (Guzzo et al.). The authors found that there are two main segments in the sequence, with higher tendency to aggregate in the early stages of dimerization. The main result is that dimers of α-synuclein and of the three single point mutations do not adopt the same fold motif in fibrils but form disordered aggregates and a minority of prefibrillar dimers with intra- and inter-molecular β-sheets. This second contribution reinforces the importance of the development of CUTABI, an algorithm that could be useful in the investigation of unfolded proteins.

Most of the other papers of this RT are based on traditional spectroscopic methods. The paper presented by the Boeckmann's group (RT_2, Fogeron et al.) used mainly solution nuclear magnetic resonance (NMR) but also circular dichroism spectroscopy and mass spectrometry to investigate the importance of phosphorylation sites in the large envelope protein of human hepatitis virus (HBV). The large envelope protein occupies a central role for interactions with the HBV cellular receptor but also with the HBV capsid, the Hsc70 chaperone and cellular membranes during fusion. Using mass spectrometry and NMR, the authors established the phosphorylation patterns of human HBV L protein. Contrary to what was known for the analogous avian virus, the locations identified for the human virus might play a functional role because they were found at strategic sites previously predicted to be crucial for interactions of the large protein. Altogether, the paper by Fogeron et al. shows the impact of spectroscopic methods in the study of even large proteins.

The paper presented by Ami et al. (RT_3) is a review of the use of infrared (IR) spectroscopy to the study of aggregation phenomena in complex biological systems, such as intact cells and tissues (Ami et al.). IR spectroscopy has been regarded for many years as a physicochemical technique mainly useful for small molecules, but recently the use of this spectroscopy has been extended to *in situ* characterization of the conformational properties of protein aggregates and to the investigation of other biomolecules surrounding aggregates, e.g., those surrounding amyloids. The characterization of protein aggregates in their natural environment may help to narrow the gap between the knowledge of amyloid aggregation mechanisms *in vitro* and *in vivo*.

NMR spectroscopy is also the main technique used in the paper (RT_7) by Spyroulias and coworkers (Birkou et al.). These authors studied the impact of single nucleotide polymorphisms on Arkadia (RNF111), an E3 ubiquitin ligase that enhances transforming growth factor-beta (TGF-β) signalling by targeting negative regulators for degradation. In particular, a single nucleotide polymorphism generated on the enzymatic ring of Arkadia the substitution of Arginine 957 to cysteine. Detailed NMR investigations showed that the R957C mutation changes the electrostatic properties of the RING domain of Arkadia, without significant effects on the structure of the region. However, the R957C Arkadia mutant exhibits increased enzymatic activity, in agreement with increased aggressive and metastatic behaviour of Arkadia within tumor cells.

Several spectroscopic and computational methods are quoted in the review (RT_4) by Fedeles et al. that describes the relevance of tautomeric dynamics in nucleic acids and in antiviral nucleoside analogs. The chemical versatility of nucleic acids is due in part to the phenomenon of nucleobase tautomerism, because the bases can adopt multiple isomeric forms, known as tautomers. Tautomers of nucleobases refer to structural isomers that differ from one another by the position of protons. Thus, by altering the position of protons, tautomerism has profound effects on the biochemical processes involving nucleic acids. For instance, the transient formation of minor tautomers during replication could generate spontaneous mutations. In the review, the authors discuss the consequences of tautomerism on the fidelity of DNA replication but also on RNA systems such as riboswitches and self-cleaving ribozymes.

NMR and other physico-chemical techniques are at the basis of a paper (RT_8) by the Polshakov's group (Mariasina et al.). These authors tackled a complex problem related to a genetic disorder (Williams-Beuren syndrome) associated with the hemizygous deletion of several genes in chromosome 7. Malfunction of 26 proteins inducing multisystemic failure is well established for most of them, but remains elusive for methyltransferase WBSCR27. Considering the complexity of the problem, the authors tried several approaches. They first tried to find a substrate of methylation catalyzed by WBSCR27 by constructing mouse cell lines with a WBSCR27 gene knockout and studied these cells by several molecular biology and mass spectrometry techniques. In all cases, neither a direct substrate was identified nor the protein partner was detected. To reveal the nature of the putative methylation substrate, the authors determined the solution structure of WBSCR27 in the apo form and in a complex with Sadenosyl-L-homocysteine. The protein core adopts a canonical Rossman fold with a disordered N-terminus. Binding to S-adenosyl-L-homocysteine induces a transition to a well-formed binding state. The structure of the binding site suggests potential substrates of WBSCR27 methylation to be probed in further studies.

The mini review (RT_6) by Garnett and Atherton, while still quoting the importance of recent improvements in solid state NMR, is centred on the spectacular advances in cryogenic electron microscopy (cryo-EM) (Garnett and Atherton). The review deals with the notoriously difficult study of proteins that form highly polymeric and filamentous assemblies using high resolution structural techniques. The study of eukaryotic microtubules and

bacterial pili are good examples, and the review gives an overview of the advances that have been made for these two systems.

Last but not least, the review (RT_5) by Bellotti's collaborators tackles a general problem of structural studies on aggregation (Faravelli et al.). Their review focuses on the need to narrow the gap between *in vitro* and *in vivo* mechanisms when studying amyloid formation by globular proteins. Although all systemic amyloidoses are characterized by amyloid deposition, they are clinically heterogeneous because they affect different organs. It is thus essential to elucidate the structural events determining the protein metamorphosis from their globular to fibrillar state. Many studies have shown how to transform proteins into fibrillar polymers *in vitro* but the events occurring *in vivo* are more complex. Reviewing the major scientific attempts to understand the amyloidogenic metamorphosis of globular proteins in systems of increasing complexity may help to bridge the gap between the experimental models and the actual pathogenic events.

Overall, the nine papers of this exciting RT provide an up-to-date view of the most recent achievements made by scientists in the evergrowing field of Structural Biology.

Author contributions

PT wrote the first version which was then edited by CA and AP.

Conflict of interest

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

Publisher's note

frontiers in Molecular Biosciences

ORIGINAL RESEARCH
published: 29 November 2021
doi: 10.3389/fmolb.2021.786123

Missense Mutations Modify the Conformational Ensemble of the α-Synuclein Monomer Which Exhibits a Two-Phase Characteristic

*Adrien Guzzo[1], Patrice Delarue[1], Ana Rojas[2], Adrien Nicolaï[1], Gia G. Maisuradze[3] and Patrick Senet[1,3]**

[1]Laboratoire Interdisciplinaire Carnot de Bourgogne, UMR 6303 CNRS-Université de Bourgogne Franche-Comté, Dijon, France, [2]Schrödinger, Inc., New York, NY, United States, [3]Baker Laboratory of Chemistry and Chemical Biology, Cornell University, Ithaca, NY, United States

α-Synuclein is an intrinsically disordered protein occurring in different conformations and prone to aggregate in β-sheet structures, which are the hallmark of the Parkinson disease. Missense mutations are associated with familial forms of this neuropathy. How these single amino-acid substitutions modify the conformations of wild-type α-synuclein is unclear. Here, using coarse-grained molecular dynamics simulations, we sampled the conformational space of the wild type and mutants (A30P, A53P, and E46K) of α-synuclein monomers for an effective time scale of 29.7 ms. To characterize the structures, we developed an algorithm, CUTABI (CUrvature and Torsion based of Alpha-helix and Beta-sheet Identification), to identify residues in the α-helix and β-sheet from C^{α}-coordinates. CUTABI was built from the results of the analysis of 14,652 selected protein structures using the Dictionary of Secondary Structure of Proteins (DSSP) algorithm. DSSP results are reproduced with 93% of success for 10 times lower computational cost. A two-dimensional probability density map of α-synuclein as a function of the number of residues in the α-helix and β-sheet is computed for wild-type and mutated proteins from molecular dynamics trajectories. The density of conformational states reveals a two-phase characteristic with a homogeneous phase (state B, β-sheets) and a heterogeneous phase (state HB, mixture of α-helices and β-sheets). The B state represents 40% of the conformations for the wild-type, A30P, and E46K and only 25% for A53T. The density of conformational states of the B state for A53T and A30P mutants differs from the wild-type one. In addition, the mutant A53T has a larger propensity to form helices than the others. These findings indicate that the equilibrium between the different conformations of the α-synuclein monomer is modified by the missense mutations in a subtle way. The α-helix and β-sheet contents are promising order parameters for intrinsically disordered proteins, whereas other structural properties such as average gyration radius, R_g, or probability distribution of R_g cannot discriminate significantly the conformational ensembles of the wild type and mutants. When separated in states B and HB, the distributions of R_g are more significantly different, indicating that global structural parameters alone are insufficient to characterize the conformational ensembles of the α-synuclein monomer.

Keywords: α-synuclein, amyloid, Parkinson disease, molecular dynamics, dictionary of secondary structure of proteins, CUrvature and Torsion based of Alpha-helix and Beta-sheet Identification, PSEA

OPEN ACCESS

Edited by:
Piero Andrea Temussi, University of Naples Federico II, Italy

Reviewed by:
Leonid Breydo, West Pharmaceutical Services, United States
Fabio Sterpone, UPR9080 Laboratoire de Biochimie Théorique (LBT), France

****Correspondence:***
Patrick Senet
psenet@u-bourgogne.fr

Specialty section:
This article was submitted to Structural Biology, a section of the journal Frontiers in Molecular Biosciences

Received: *29 September 2021*
Accepted: *25 October 2021*
Published: *29 November 2021*

Citation:
Guzzo A, Delarue P, Rojas A, Nicolaï A, Maisuradze GG and Senet P (2021) Missense Mutations Modify the Conformational Ensemble of the α-Synuclein Monomer Which Exhibits a Two-Phase Characteristic. Front. Mol. Biosci. 8:786123. doi: 10.3389/fmolb.2021.786123

1 INTRODUCTION

α-helices and β-sheets are the two main secondary structures stabilized by backbone hydrogen bonds in proteins. The propensity of each residue to form an α-helix or a β-sheet depends on the amino-acid sequence and on the fold (Chou and Fasman, 1974; Smith et al., 1994; Pace and Scholtz, 1998; Bhattacharjee and Biswas, 2010). The formation of long β-sheets differs significantly from α-helices as the former necessarily involves long-distance contacts in the sequence and depends on the mean curvature of the fold (Koh et al., 2006; Bhattacharjee and Biswas, 2010). The content of these two secondary structure elements is a criterion for classification of protein native structures and characterization of protein folding kinetics and pathways (Konagurthu et al., 2020). In intrinsically disordered proteins (IDPs) (Wright and Dyson, 1999; Uversky et al., 2000; Uversky, 2019; Deiana et al., 2019), α-helices and β-sheets are metastable secondary structures. As for folded proteins, a natural extension of protein classification of IDP can be based on the content of these two secondary structural elements in a statistical sense. As an IDP has no native state, the relative content of α-helices (α) and of β-sheets (β) possibly defines a fingerprint of each conformational state in a two-dimensional (α, β) propensity map (Ullman et al., 2011). Here, we aim to build such an (α, β) effective free-energy map for α-synuclein (α-syn), wild-type (WT), and A30P, A53T, and E46K mutants from first principles by using coarse-grained molecular dynamics (MD) simulations with the UNRES (UNited RESidue) force field (Liwo et al., 2001; Maisuradze et al., 2010).

α-syn is a 140 amino-acid protein abundant in the brain (Jakes et al., 1994; Mollenhauer et al., 2008). It plays a central role in the onset of the Parkinson disease (PD) and other neurological disorders, named synucleopathies (Soto, 2003; Chiti and Dobson, 2006; Stefanis, 2012; Chiti and Dobson, 2017; Tanudjojo et al., 2021). In these diseases, α-syn is found in high concentration, as filaments, in spherical inclusions (Lewy bodies) located in the neuron cytoplasm (Spillantini et al., 1997; Breydo et al., 2012; Lashuel, 2020; Trinkaus et al., 2021). Familial cases of PD are induced either by an overexpression of WT α-syn due to α-syn gene triplication or by pathogenic mutations in α-syn gene corresponding to single amino-acid substitution, namely, H50Q, G51D, A53E, A30P, A53T, and E46K (Polymeropoulos et al., 1997; Krüger et al., 1998; Zarranz et al., 2004; Fuchs et al., 2008; Appel-Cresswell et al., 2013; Pasanen et al., 2014; Petrucci et al., 2016). These inherited forms of PD have phenotypes that are similar to the sporadic PD, but patients with the A53T mutant exhibit an early onset of PD (Petrucci et al., 2016). The role of α-syn in PD development and the mechanisms of its aggregation and of the aggregate propagation are still debatable (Henrich et al., 2020; Sang et al., 2021) and involve α-syn-lipid interactions (Galvagnion et al., 2015; Suzuki et al., 2018; Antonschmidt et al., 2021), a liquid-liquid phase transition (Ray et al., 2020), and a secondary nucleation and fibril fragmentation (Cremades et al., 2012; Cascella et al., 2021; Kumari et al., 2021). Moreover, α-syn aggregation might disregulate the mitochondrial function, and increasing the level of free radicals and alterations in this pathway may be involved in the pathogenesis of PD (Hsu et al., 2000; Devi et al., 2008).

From a structural point of view, α-syn is a chameleon protein (Ullman et al., 2011; Mor et al., 2016): it is disordered (IDP) in solution under physiological conditions, in equilibrium with a minor α-helical tetrameric form in the cytoplasm (Bartels et al., 2011; Wang et al., 2011) and α-helical when bounded to a cell membrane (Fusco et al., 2018). When incubated under physiological conditions *in vitro*, α-syn aggregates into fibrils with polymorphic cross-β-sheet conformations, in which a core of β-strands is aligned perpendicular to the fibril axis forming extended regular β-sheets with different arrangements (Tuttle et al., 2016; Guerrero-Ferreira et al., 2019; Guerrero-Ferreira et al., 2020). In addition to cylindrical fibrils, ribbon aggregates have been also observed (Bousset et al., 2013). The different polymorphs characterized *in vitro* are believed to mimic the α-syn filament structures in synucleopathies.

Three main regions were identified in the primary sequence of α-syn regarding their role in α-syn conformational dynamics and aggregation. The N-terminal region (residues 1–60) contains a number of imperfect repeats, with the consensus motif KTKEGV, strongly similar to that found in the amphipathic helices and responsible for membrane binding (Perrin et al., 2000). High-resolution NMR structures revealed a broken helix featuring two curved α-helices of α-syn (residues 2—37 and 45—92) bound to micelles (Ulmer et al., 2005), whereas electron paramagnetic resonance characterizations feature an extended helix for α-syn bound to a lipid membrane (first 97 residues) (Cheng et al., 2013). In solution, exposure of this N-terminal region to solvents is correlated to aggregation propensity (Stephens et al., 2020). The central hydrophobic region of α-syn (residues 61—95), called NAC (non-amyloid component), contains a hydrophobic stretch of 12 residues 71VTGVTAVAQKTV82 necessary for the aggregation (Giasson et al., 2001). Fibrils of the NAC region are also found in plaques of Alzheimer's disease (Uéda et al., 1993). The flexible acidic C-terminal region regulates fibril formation *in vitro* seeding experiments (Murray et al., 2003) and plays a role in the secondary nucleation process of amyloids via electrostatic interactions with the lysine-rich N-tail (Kumari et al., 2021). Transient, long-range interactions between the negatively charged C-terminus (residues 120—140) with the positively charged N-terminus and NAC (residues 30—100) were observed by paramagnetic relaxation enhancement (PRE) for WT α-syn (Dedmon et al., 2005). It is hypothesized that these transient interactions are responsible for the larger compactness of α-syn compared to a disordered chain (Dedmon et al., 2005). Structures with long-range contacts between the N-terminal and C-terminal represent about 14% of the conformational ensemble, and in a significant fraction of these structures, residues 68—78

(NAC) are exposed to a solvent, in contrast to the hypothesis that such long-range transient interactions prevent aggregation (Ullman et al., 2011).

Mutations affect the arrangement and growth of the fibrils *in vitro* (Guerrero-Ferreira et al., 2020). Compared to WT, A53T and E46K mutations aggregate faster, whereas A30P aggregates more slowly than WT (Tosatto et al., 2015; Stephens et al., 2020). In addition, WT, A53T, E46K, and A30P differ in the formation of the different early oligomeric moieties (Tosatto et al., 2015). The mutants A30P and A53T show a greater propensity to form non-fibrillar aggregates than WT (Li et al., 2002), and A53T promotes seeded aggregation in human neurons (Tanudjojo et al., 2021). The rate of lipid-induced aggregation and secondary nucleation have been found to differ by multiple orders of magnitude depending on which missense mutation is involved (Flagmeier et al., 2016). NMR studies have shown an increase flexibility of α-syn in nanosecond–microsecond time scales and a reduction of contacts between C- and N-terminals in mutants (Bertoncini et al., 2005). All single amino-acid substitutions have thus both a kinetic and structural effect on the formation of oligomeric structures. The purpose of the present theoretical study is the characterization of the differences between the conformational ensembles of monomeric WT and mutants, which may contribute to our understanding of the early steps of the aggregation process in solution.

The majority of monomeric WT α-syn conformations have no secondary structures (Ullman et al., 2011). The average helical and β-strand contents of WT α-syn deduced from restrained MD of α-syn fragments and NMR data are about 3 and 11%, respectively (Ullman et al., 2011). The maximum fraction of residues with helical and β-sheet secondary structures was found to be 20% (28 residues) and 28% (39 residues), respectively (Ullman et al., 2011). Little is known on how the missense mutations modify the propensity of α-syn to form secondary structures. As the α-helical/β-sheet equilibrium is central to the folding polymorphism of α-syn, we decided to characterize the propensities of these two secondary structures in WT, A30P, A53T, and E46K monomers using MD.

Because the huge conformational space of α-syn is out of range for all-atom MD simulations in the explicit solvent, we applied coarse-grained UNRES MD (Maisuradze et al., 2010) to sample the structures of WT and the most studied A30P, A53T, and E46K monomers. The effective time scale of UNRES is 3 orders of magnitude larger than the all-atom time scale (Khalili et al., 2005). To ensure the convergence of the conformational sampling of the monomeric states, we applied replica exchange MD (see the Material and Methods section). The total effective time scale of the present simulations is 29.7 ms (72 replicas of 412 μs each) for each protein studied. One of the gold standards to quantify the secondary structure elements of a protein from its structure is the Dictionary of Secondary Structure of Proteins (DSSP) (Kabsch and Sander, 1983; Touw et al., 2015) algorithm based on a simplified model of hydrogen bonds. Application of DSSP to coarse-grained structures simulated by UNRES requires to build a compatible all-atom structure from the C^{α} coordinates of the UNRES model using reconstruction programs (Feig et al., 2004; Rotkiewicz and Skolnick, 2008). To avoid the high computational cost of all-atom reconstructions from coarse-grained coordinates, we developed here an algorithm which assigns an α-helix or a β-sheet secondary structure to each residue based on the C^{α}-C^{α} distances and on the coarse-grained angles formed by C^{α}-C^{α} pseudobonds, which correspond to the local curvature and torsion of the protein main chain (Grassein et al., 2020). The accuracy of the present algorithm, named CUTABI (CUrvature and Torsion based of Alpha helix and Beta-sheet Identification), to quantify the β-sheet content of proteins is improved compared to an existing algorithm based on C^{α} coordinates [P-SEA (Labesse et al., 1997)] and is comparable to the accuracy of DSSP (see the Material and Methods section). For each structure of the conformational ensemble of WT and mutants monomers, the number of residues in α-helix or in β-sheet was computed with CUTABI. The probability density of this two-dimensional descriptor was computed by using the conformations at 300 and 310 K. Analysis of these maps and of the conformational ensembles of the WT protein and mutants revealed subtle effects of the single amino-acid substitutions which are possibly related to the differences observed in oligomerization between WT and mutants (see the Results and Discussion section).

2 MATERIALS AND METHODS

2.1 Coarse-Grained Molecular Dynamics Simulations

Detailed descriptions of the UNRES force field and its parameterization are available in the reference (Liwo et al., 2019) and at http://www.unres.pl. Therefore, it will be only briefly outlined here. In the UNRES force field, a polypeptide chain is represented as a sequence of C^{α} atoms with united peptide groups located halfway of the virtual C^{α}-C^{α} bonds and united side chains (SCs) attached to the C^{α} atoms. The force field has been derived as the potential of mean force (PMF) of a system of polypeptide chain(s) in the solvent, where all degrees of freedom except the coordinates of the C^{α} atoms and SC centers have been averaged out. The effective energy function contains local and site–site interactions as well as multibody terms, which have been obtained by decomposing the PMF into factors corresponding to clusters of interactions within and between coarse-grained sites (Liwo et al., 2001). The SC-SC interaction potentials implicitly include the contribution from solvation (Liwo et al., 2001; Maisuradze et al., 2010). The force field was calibrated to reproduce the structure and thermodynamics of small model proteins and applied with success to simulate protein folding (Maisuradze et al., 2010; Zhou et al., 2014; Sieradzan et al., 2021) and large-scale conformational dynamics (Gołaś et al., 2012).

All structures of α-syn (WT and mutants) were extracted from replica exchange MD trajectories generated with the UNRES force field. A total of 72 trajectories were computed for each protein: 32 trajectories at 300 K and 8 trajectories at each of the following temperatures, 310 K, 323 K, 337 K, 353 K, and 370 K. Each trajectory was started with two fully unfolded monomers separated by a distance of 25 Å. The Cartesian coordinates of C^{α} and SC beads were saved every 1,000 integration steps. The

integration time step in UNRES is 4.9 fs, corresponding to an effective actual time step of about 4.9 ps. Due to the implicit integration of fast motions, the time scale of UNRES compared to the experimental time scale is indeed accelerated by a factor of 1,000 (Khalili et al., 2005). For each trajectory, the first 4 million steps were discarded. After this relaxation period, 8 consecutive runs of 10 million steps were used for the analysis of each trajectory. Convergence of the conformational ensemble simulated for each protein was monitored by the probability densities of intra-chain and inter-chain contacts and of the radius of gyration. For each trajectory, the convergence was achieved for the last 3 runs of 10 million steps, i.e., for a statistics of 30 million steps (an effective time scale of 147 microseconds) out of 84 million steps (an effective time scale of 412 microseconds). Only structures at 300 and 310 K are reported here as they are close to the physiological temperature. The structures were saved every 1,000 integration steps, leading to about 1 million of converged conformations analyzed for each protein combining the results at 300 and 310 K.

Since the simulations are performed on two monomers, both isolated non-interacting monomer conformations and aggregated monomers were observed in the converged MD trajectories. The analysis and description of the dimeric conformations are out of the scope of the present paper, which is dedicated to isolated monomers, and they will be described elsewhere. The monomeric structures studied in the present work were extracted from the converged replica exchange MD trajectories if they obey the following condition: no residues at a distance smaller than 20 Å from the other monomer. The fraction of monomers out of all the conformers simulated at 300 and 310 K is 55% for WT, 65% for A30P, 50% for A53T, and 48% for E46K. The monomeric state is finally described here by about 1 million of structures of the converged production part of the MD trajectories at 300 and 310 K, representing a sampling on an effective time scale of 4.9 ms.

2.2 Curvature and Torsion Based of α-Helix and β-Sheet Identification: An Algorithm for Secondary Structure Determination Based on C^α Coordinates

Secondary structure elements (SSEs) are important descriptors of the native state of proteins (Konagurthu et al., 2020) and of the conformational ensemble of IDP, as shown in the next section for α-syn. The main local structures stabilized by backbone hydrogen bonds are helices (α-, 3_{10}-, and π-helices) and β-sheets. Deviations from the canonical definitions of these secondary structures based on Ramanchandran angles are common in the Protein Data Bank. Therefore, a practical and widely used method to assign an SSE to a residue is to apply the DSSP algorithm which is based on the calculation of a simplified energy function describing backbone hydrogen bonds (Kabsch and Sander, 1983). Application of DSSP to a protein structure requires the knowledge of the Cartesian coordinates of all backbone atoms. For structures simulated by coarse-grained force fields, like UNRES, only a subset of these coordinates is known, those of the C^α atoms. Assignment of the SSE based on C^α coordinates can be performed using DSSP but at the extra computational cost of the reconstruction of an all-atom representation of the coarse-grained structure using software like MMTSB (Feig et al., 2004) or PULCHRA (Rotkiewicz and Skolnick, 2008). Alternatively, the SSE can be defined from the C^α coordinates only [P-SEA algorithm (Labesse et al., 1997)]. Here, we have chosen the second faster option and have developed the algorithm CUTABI by analyzing 14,652 experimental structures with less than 40% of sequence identity extracted from the ASTRAL database (Fox et al., 2014; Chandonia et al., 2019). The parameters of CUTABI were adjusted to reproduce the DSSP assignment of helices and β-sheets for the ASTRAL database. The agreement between CUTABI and DSSP is 93% for the structures in the database with 10 times (DSSP + PULCHRA) to 30 times (DSSP + MMTSB) less computational cost.

2.2.1 Parameters of Curvature and Torsion Based of α-Helix and β-Sheet Identification

Assuming a constant virtual bond length between C^α atoms of successive residues, a chain of N amino acids is fully characterized by $N-3$ torsion angles γ_n, built from the positions of C^α_{n-1}, C^α_n, C^α_{n+1}, and C^α_{n+2} with $n = 2$ to $N-2$; and $N-2$ bond angles θ_n, built from C^α_{n-1}, C^α_n, and C^α_{n+1}, with $n = 2$ to $N-1$. These angles have clear geometrical meanings: they are respectively the discrete version of the local curvature (θ_n) and the local torsion (γ_n) of the chain formed by the successive $C^\alpha - C^\alpha$ virtual bonds (Grassein et al., 2020). From a mathematical point of view, the local curvature and torsion fully describe the structure of a string and form a complete set of local order parameters for protein folding (Grassein et al., 2020). For proteins, the curvature has a limited range with θ varying between 80° and 160°, whereas γ can take nearly any value between -180° and +180°. The SSEs correspond statistically to specific areas in the coarse-grained (γ, θ) maps. **Supplementary Figure S1** shows the distribution of (γ, θ) angles computed by the DSSP algorithm for the 14,652 experimental structures extracted from the ASTRAL database for helix (H + G + I) and β-sheet (E) (in brackets the one-letter SSE codes in DSSP). The areas corresponding to helix and β-sheet identified by DSSP in the (γ, θ) map are nicely reproduced by the CUTABI algorithm, as shown in **Supplementary Figure S1**, using an identification of the SSE based on the coordinates of the C^α atoms as follows.

In CUTABI, the minimal size of a helix is set to 4 residues. Helices with less than 3 residues, such as short 3_{10} helices, are thus not counted. Because helices involve short-range interactions along the amino-acid sequence, the local curvature θ and torsion γ parameters are sufficient to characterize this SSE. To decide if a set of four residues k, $k+1$, $k+2$, and $k+3$ pertains to a helix, a combination of the coarse-grained angles formed by residues (C^α atoms) $k-1$, k, $k+1$, $k+2$, $k+3$, and $k+4$ is considered, as shown in **Figure 1**. Residues k, $k+1$, $k+2$, and $k+3$ pertain to a helix if the angles θ_k, θ_{k+1}, θ_{k+2}, and θ_{k+3} belong to an interval between 80° and 105° and the angles γ_{k+1} and γ_{k+2} belong to an interval between 30° and 80°. A window of 4 residues is slid along the sequence to evaluate the residues pertaining to a helix in each α-syn conformation.

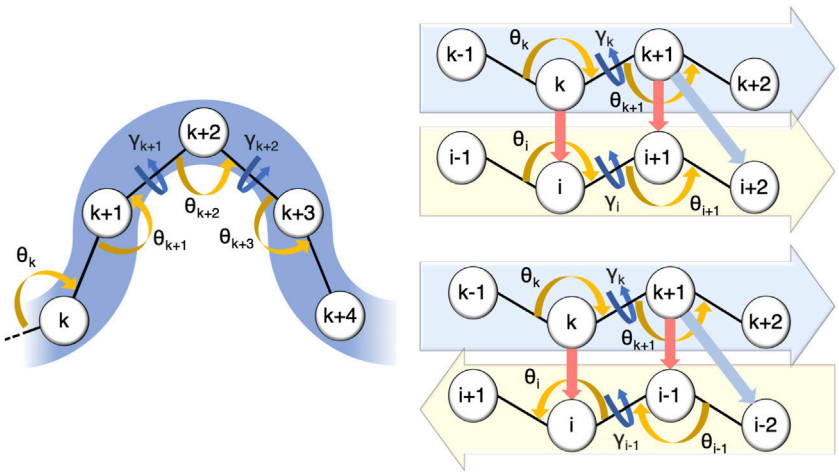

FIGURE 1 | Parameters defining residues pertaining to a helix and to a parallel or to an anti-parallel β-sheet in CUTABI (see text). The spheres represent the positions of C^{α} atoms of the residues. Orange and dark-blue arrows show the θ and γ angles involved, respectively. For β-sheets, red and light-blue arrows point out the distances between residues involved in their definitions.

In CUTABI, the minimal size of a β-strand is set to 2 residues, i.e., a β-sheet cannot be smaller than 4 residues. The β-strands of 1 residue forming β-bridges (B code in DSSP) are thus not considered. To define parallel and anti-parallel β-sheets (**Figure 1**), both local curvature and torsion parameters as well as distances between the two β-strands forming the sheet are necessary. If the four residues k, $k + 1$, i, and $i + 1$ pertain to a parallel β-sheet (**Figure 1**), then the angles θ_k, θ_{k+1}, θ_i, and θ_{i+1} must be between 100° and 155° and the angles γ_k and γ_i must be smaller than −80° or larger than 80°. In addition, the distances between k and i and $k + 1$ and $i + 1$ (**Figure 1**, red arrows) must be smaller than 5.5 Å, and the distance between $k + 1$ and $i + 2$ (**Figure 1**, light blue arrows) must be smaller than 6.8 Å. Similar conditions must be met for the anti-parallel β-sheet, as shown in **Figure 1**.

2.2.2 Performance of Curvature and Torsion Based of α-Helix and β-Sheet Identification Compared to Dictionary of Secondary Structure of Proteins and P-SEA

The SSEs computed with the parameters defined in **Figure 1** were compared to the SSEs calculated with DSSP (based on all-atom coordinates) (Kabsch and Sander, 1983) and P-SEA (based on C^{α} coordinates only) (Labesse et al., 1997). **Figure 2** shows the percentage of difference, i.e., the number of residues having an SSE different in the coarse-grained algorithms P-SEA and CUTABI compared to DSSP, divided by the sequence length. For helices, as shown in **Figure 2** (left panel), 96% of structures (5,550 + 8,585 = 14,235 structures) do not have more than 15% of residues with an SSE different in CUTABI and in DSSP. The results of P-SEA are similar. Considering DSSP as a gold standard, the CUTABI accuracy is improved compared with P-SEA on the evaluation of residues in the β-sheet. As shown in **Figure 2** (right panel), 5,739 structures examined (40% of the database) have less than 5% of residues pertaining to an SSE different in CUTABI compared to DSSP. This number is only 3,084 (20% of the database) in P-SEA. In total, assignments of SSE by CUTABI and DSSP agree for 84% of the structures studied (5,739 + 6,628 = 12,367 structures) with a difference compared to DSSP of maximum 15% of the sequence length. The average of the percentage of agreement over all the structures of the database (14,652) for all SSEs between CUTABI and DSSP is 93% for a computational time at least 10 times smaller.

As shown in **Figure 2**, one observes a large difference of secondary structure assignments between CUTABI and DSSP for a very small number of proteins. Most structures in this category were measured by X-ray diffraction but with a low resolution (>3Å). In addition, as hydrogen atom positions are not detected in the X-ray (except for ultra-high resolution), the application of DSSP to these structures may be less accurate as DSSP is based on the calculation of hydrogen bond energy. To illustrate the precise origin of this finding, we examined the structures and selected two representative structures with 35% of difference in SSE: one for the helix (**Figure 3**, left blue panel) and another for the β-sheet (**Figure 3**, right red panel).

Figure 3 (left panel) represents the fragment (from residue 19 to 98) of a calcium-gated potasium channel (PDB ID: 1lnq and ASTRAL ID: d1lnqa2). In this example, CUTABI detects a much larger number of residues in the helix than DSSP. However, the 3D representation shows that although these extra residues missed by DSSP are not in a canonical helix, the overall shape of the main chain is indeed helical.

Figure 3 (right panel) represents the fragment (from residue 1 to 90) of a glycosyltrehalose trehalohydrolase (PDB ID: 1eh9 and ASTRAL ID: d1eh9a1). The number of residues in the β-sheet is much larger in CUTABI than in DSSP. As in the case of the helix, the 3D representation indicates that the additional residues in the β-sheet in CUTABI are part of a main chain segment with the overall shape of the β-sheet, although not canonical, due probably to the low experimental resolution.

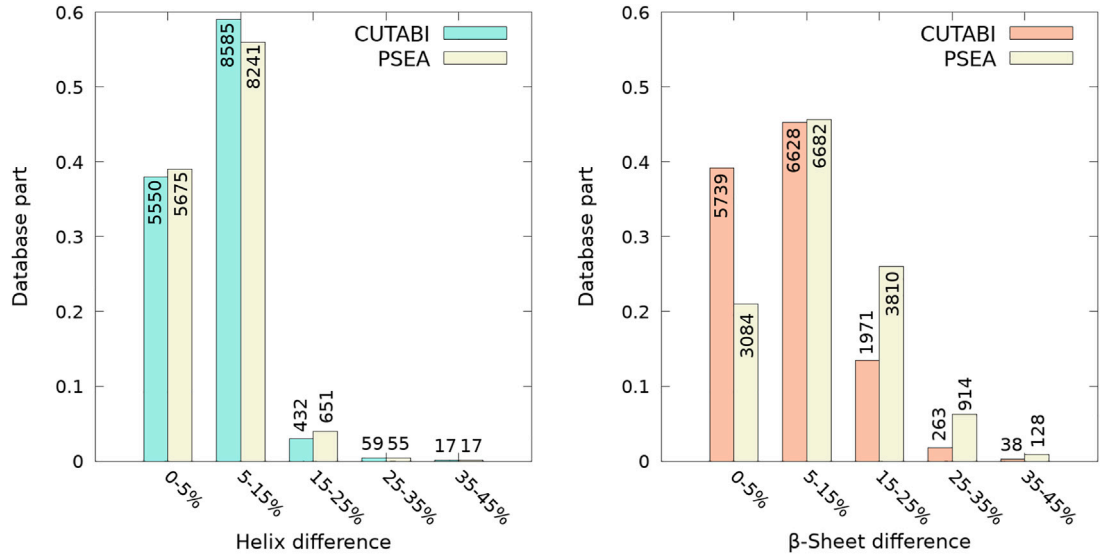

FIGURE 2 | Comparison between the identification of residues pertaining to the helix and to β-sheet between the all-atom DSSP algorithm and coarse-grained CUTABI and P-SEA algorithms. Each bar of the histograms represents the number of structures of the ASTRAL database as a function of a range of percentages of the difference between CUTABI (green and orange) and DSSP and between P-SEA (yellow) and DSSP. For each protein, the percentage is calculated as the number of residues having a different secondary structure in the coarse-grained algorithms and DSSP, divided by the protein sequence length.

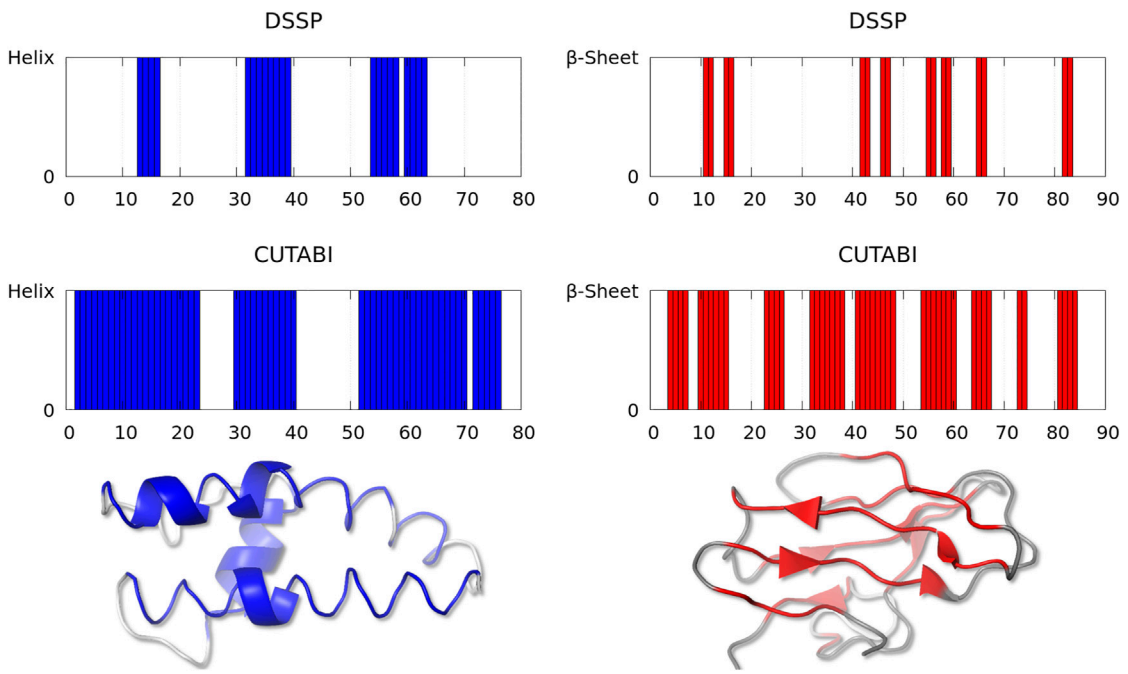

FIGURE 3 | Representative structures with a large discrepancy between the SSE assignment by DSSP and CUTABI for the helix (left, ASTRAL ID: d1lnqa2) and for the β-sheet (right, ASTRAL ID: d1eh9a1). In the left (right) panel, each blue (red) bar represents a residue pertaining to the helix (to the β-sheet). The 3D main chain is represented with a blue (red) tube for residues pertaining to the helix (to the β-sheet) for CUTABI and with a blue (red) cartoon for DSSP. Residues not in the helix or in the β-sheet are in gray. The 3D representations were made with the PyMOL software (Schrödinger, 2015).

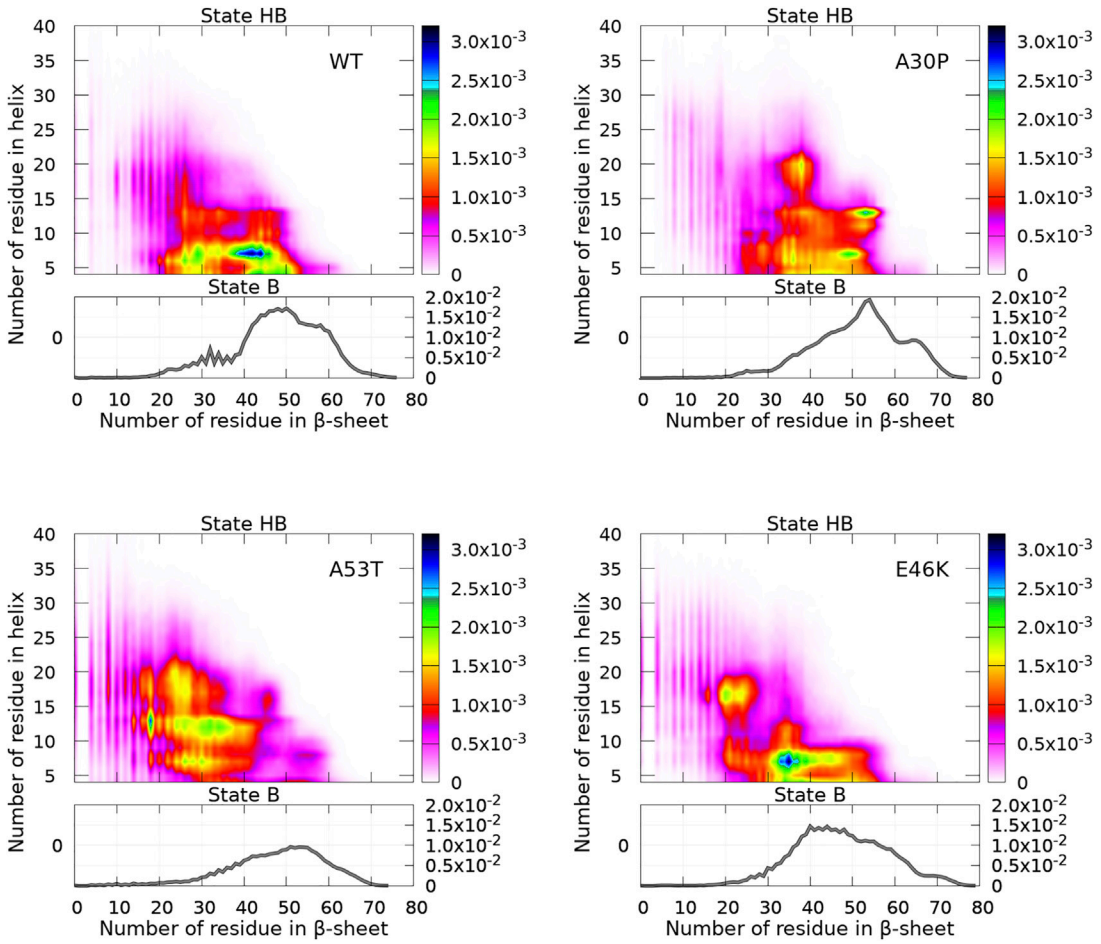

FIGURE 4 | Probability density of the number of residues in the α-helix and β-sheet for WT and mutants of α-syn. The probability density of state B (no helix) is represented by a function (gray) (right vertical axis), and the probability density of state HB is represented by a two-dimensional map (right color bar).

3 RESULTS

3.1 α-Helix and β-Sheet Propensities of α-syn Divide the Conformations Into Two Distinct States

The algorithm CUTABI was applied to the α-syn conformations to compute the number of residues in the α-helix (α) and β-sheet (β) of each structure for WT and mutants. Each conformation has (α, β) coordinates. The resulting probability densities in the (α, β) space are represented in **Figure 4**. In these maps, only the residues from the N-terminal and NAC regions were considered for the calculations [as can be seen in the next section (**Figure 8**), the C-terminal region does not contribute to SSE differences between WT and mutants].

A major observation is that the conformations are divided into two distinct states for the N-terminal + NAC region: an ensemble of conformations with no residue in the helix (state B) and the rest of conformations (state HB). The highest probability of observing a conformation in state B is an order of magnitude larger than that of state HB, as can be seen by comparing (the scale of) two-dimensional and one-dimensional plots in **Figure 4**. In addition, **Figure 4** reveals clear differences between WT and mutants.

First, we discuss the global differences between states B of WT and mutants (one-dimensional functions in **Figure 4**). The maximum of the distributions is at 50, 54, 53, and 40 for WT, A30P, A53T, and E46K, respectively. The distribution is sharper for A30P, which has the largest number of conformations with the largest number of residues (between 60 and 70) in the β-sheet. Clearly, A53T has the lowest number of conformations in state B. This is even better seen in **Figure 5**, showing the fraction of conformations within a free-energy difference cutoff from the global minimum of state B for each protein. With $P_{\max}$ being the maximum of probability at $(0, \beta)$ (in the B state) and P being the probability at (α, β) $(\alpha \geq 0$, in the B or HB states), the free-energy cutoff is computed as $-\ln(\frac{P_{\max}}{P})$ in kT units, where k is the Boltzmann constant and T is the temperature. Within 1 kT, there is 32% of the conformations found for WT, E46K, and A30P and only 20% for A53T, as shown in **Figure 5**.

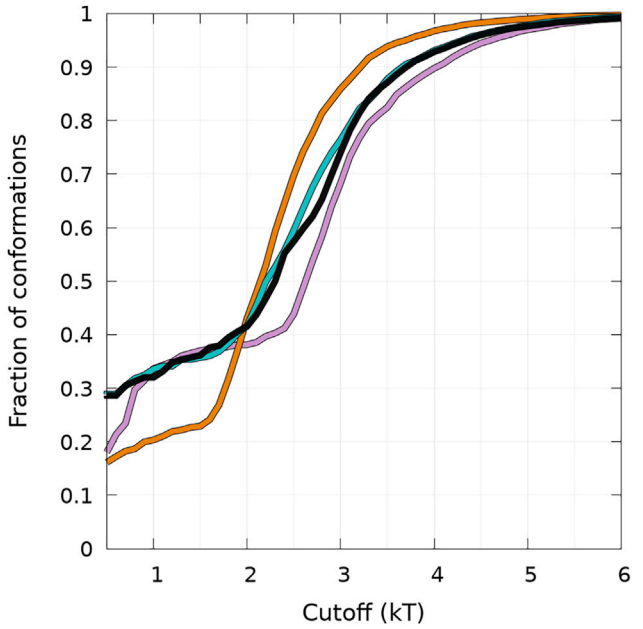

FIGURE 5 | Cumulative fraction of the total number of conformations simulated as a function of a free-energy cutoff (in kT units) from the global minimum of the B state (see text) represented in **Figure 4**. Curves are for WT (black), A30P (purple), E46K (turquoise), and A53T (orange).

The results illustrated in **Figure 5** emphasize the two-state behavior. For A30P, there is even a third state below 1 kT. By definition, the derivative of the curves represented in **Figure 5** represents the Density Of conformations or micro-States (DOS). The nearly linear behavior of the curves for WT and E46K up to 2 kT means a rather constant DOS corresponding to state B (one-dimensional probability densities in **Figure 4**). The change of slope at 2 kT for these proteins points up the separation between the states, i.e., the onset of state HB, i.e., a state with a mixture of α-helices and β-sheets. State B represents thus 40% of the conformations for WT and E46K. Between 2 kT and about 3 kT, the curves for WT and E46K in **Figure 5** are linear with a slope larger than that in the B state, which corresponds to a larger DOS in state HB. Beyond 3 kT, the DOS decreases and reaches a plateau for all proteins. For A53T, state B contains less conformations and the onset of the state HB occurs at about 1.5 kT. State B represents only about 25% of the conformations for A53T. The DOS of the HB state for A53T is higher than that in its B state, and it is also higher than the DOS of the HB state for WT and E46K. The case of A30P is special. State B has the highest DOS up to 1 kT, and it becomes similar to the DOS for WT and E46K up to 2.5 kT. The third state below 1 kT is clearly visible as a shoulder with a large proportion of conformations with 60—70 residues in the β-sheet in the one-dimensional probability distribution, shown in **Figure 4**. Finally, state B in A30P also concerns 40% of the structures. The DOS in the HB state for A30P is similar to the one of A53T.

Figure 5 shows that state HB encompasses about 60% of the conformations for WT, A30P, and E46K (all conformations beyond 2 kT for WT and E46K and beyond 2.5 kT for A30P) and 75% of the conformations for A53T (all conformations beyond 1.5 kT). Although the DOS is relatively constant, local maxima occur in the HB two-dimensional maps (**Figure 4**). The most probable (α, β) conformations occur at (7,44), (13,53), (13,18), and (7,35) for WT, A30P, A53T, and E46K. There is a significant difference between A53T and WT, A30P, and E46K. For each protein, selected structures for the maximum of probability of the B state and for the local maxima of the probability of the HB state are represented in **Figure 6**. They illustrate the expected diversity of conformations of an IDP. It is, however, important to emphasize that each (α, β) pair represents an ensemble of conformations. It is impossible to represent the entire diversity of these sub-ensembles. For example, for the WT protein, the maximum of the B state at (0,50) corresponds to 19,901 structures, and the maximum of the HB state at (7,44) corresponds to 3,569 structures. In **Figure 6**, the structures with a maximum of residues in the helix or in the β-sheet are shown. The maximum number of residues in the helix is 50 for A53T compared to 46 for the other proteins and compared to the 82 (Ulmer et al., 2005) and 97 residues (Cheng et al., 2013) in helices when α-syn is bound to membranes.

The average radius of gyration R_g is a common global structural parameter in polymer science, and it is interesting to relate this property to states B and HB. For a random coil represented by a self-avoiding walk in a good solvent (i.e., for which interactions between monomers and solvent molecules are energetically favorable), $R_g = 0.367bN^{\nu}$, where b is the length of the so-called statistical segment and ν is a fractal exponent. In three dimensions, we take b = 7.6 Å (twice the distance between two C^{α}), R_g = 54 Å for the size of α-syn (N = 140) (Victor et al., 1994). The average radius of gyration R_g computed from the UNRES trajectories is 24.7, 25.1, 26.0, and 25.2 Å for WT, A30P, A53T, and E46K, respectively. Only A53T has a significantly larger R_g than WT. Other authors reported an average radius of gyration of $R_g \simeq$ 23 Å for WT and a distribution of R_g narrower than the distribution of a random coil of a similar sequence length (Dedmon et al., 2005; Allison et al., 2009).

The distributions of R_g of the conformational ensemble simulated with UNRES for WT and mutants are very similar to each other, as shown in **Figure 7**. The peak of the probability density is lower for A30P. When the global distribution is divided into states B and HB, one reaches the same conclusion for state HB (**Figure 7**, middle panel), but one observes more significant differences between the proteins for state B (**Figure 7**, bottom panel). For state B, the peak of A53T is the highest and a sub-population appears clearly on the left side of the distribution for A30P and E46K. This sub-population is hardly visible in the global probability distribution (**Figure 7**, top panel) as a shoulder. Examination of the structures corresponding to this sub-state of the B state reveals that the structures of A30P and E46K have a large proportion of contacts between regions 1—20 and 96—140. The average number of contacts of the structures with 17.9Å < R_g < 18.1 Å is 5.2, 14.9, 0.1, and 9.0 for WT, A30P, A53T, and E46K, respectively. The large number of contacts between the two extremities of the protein for A30P and E46K mutants compared to WT and A53T explains the peak at 18 Å.

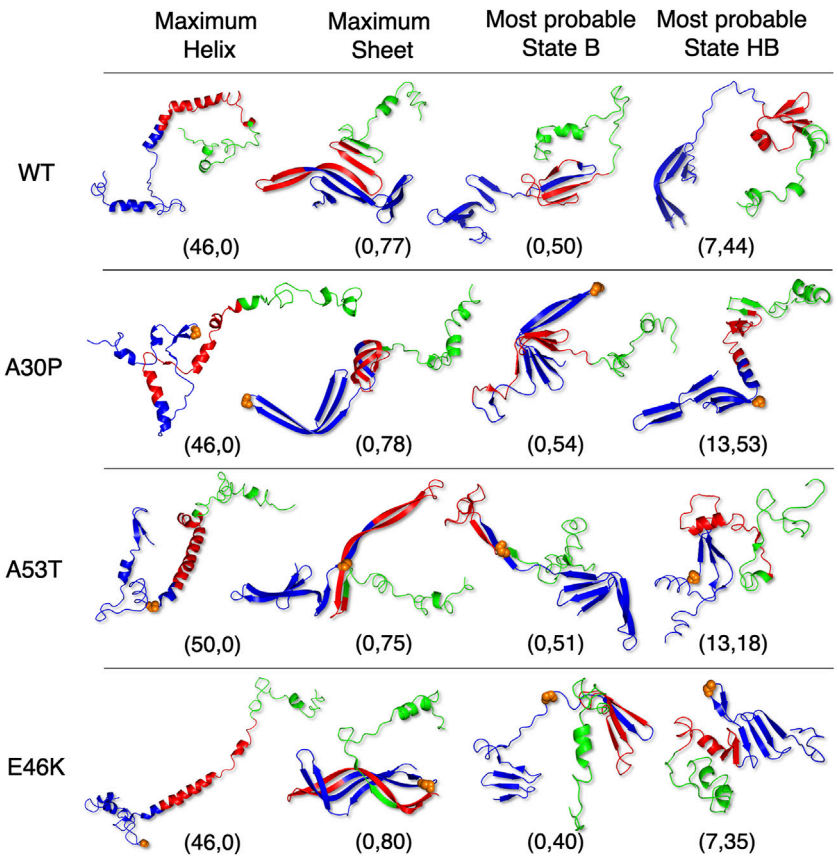

FIGURE 6 | Selected representative structures for WT and mutants extracted from the conformational sub-ensembles corresponding to the maximum fraction in the helix (first column), the maximum fraction in the β-sheet (second column), the most probable sub-state of state B (third column), and the most probable sub-state of state HB (last column). The coordinates correspond to the position in (α, β) maps of **Figure 4**. Orange spheres represent backbone atoms at the mutation location. The 3D representations were made with the PyMOL software (Schrödinger, 2015).

3.2 Secondary Structure Element Propensities as a Function of the Position in the Amino-Acid Sequence: Differences and Similarities Between the Wild Type and Mutants

As shown in **Figure 8**, helices are found in four main regions: two in the C-terminal region (residues 119—125 and 127—130), one in the NAC region (residues 75—82), and one overlapping the N-terminal and NAC regions (residues 53—65) for both WT and mutants. As claimed in the previous section, there are no significant differences between the propensities of SSE for WT and mutants in the C-terminal region. The major differences between WT and mutants occur in the region 53—65, which has a peak for the helix propensity at residue K58. For all mutants, the probability to form an α-helix in this region is larger than that for WT. For A53T, the probability of residue K58 to pertain to a helix is more than twice higher than one for the WT protein (P_{WT} = 0.25 compared to P_{A53T} = 0.56). Mutation 53 occurs in the N-terminal part of a helical region of WT, but its effect is not trivial as the mutation could be naively expected to decrease the helicity. Indeed, the propensity to be part of a helix is 1.45 for alanine and 0.82 for threonine according to the empirical helix propensity scale (Chou and Fasman, 1974). The increase of helicity observed in the 53—65 region of the sequence upon single mutation is clearly not a local effect. Similarly, single amino-acid substitutions, A30P and E46K, also increase significantly the helicity in the region of 53—65.

Compared to other proteins, another significant difference is observed for A53T in the helical region located at residues 18—22, where the probability to form an α-helix is significantly larger for this mutant. For example, the peak at E20 corresponds to the following probabilities: P_{WT} = 0.07, P_{A30P} = 0.04, P_{A53T} = 0.14, and P_{E46K} = 0.05. The significant increase of helicity in the 18—22 and 53—65 regions explains why state B of A53T is less populated, as shown in **Figures 4**, **5** and discussed in the previous section. The presence of an α-helix in the 18—22 region might explain why mutation A18T induces significant modification of the α-syn polymerization (Kumar et al., 2018). Simulations of this mutant are scheduled in the future. For A30P, the probability to observe a helix in the NAC region is slightly lower than for other proteins: the probabilities at A78 are P_{WT} = 0.37, P_{A30P} = 0.32, P_{A53T} = 0.40, and P_{E46K} = 0.38.

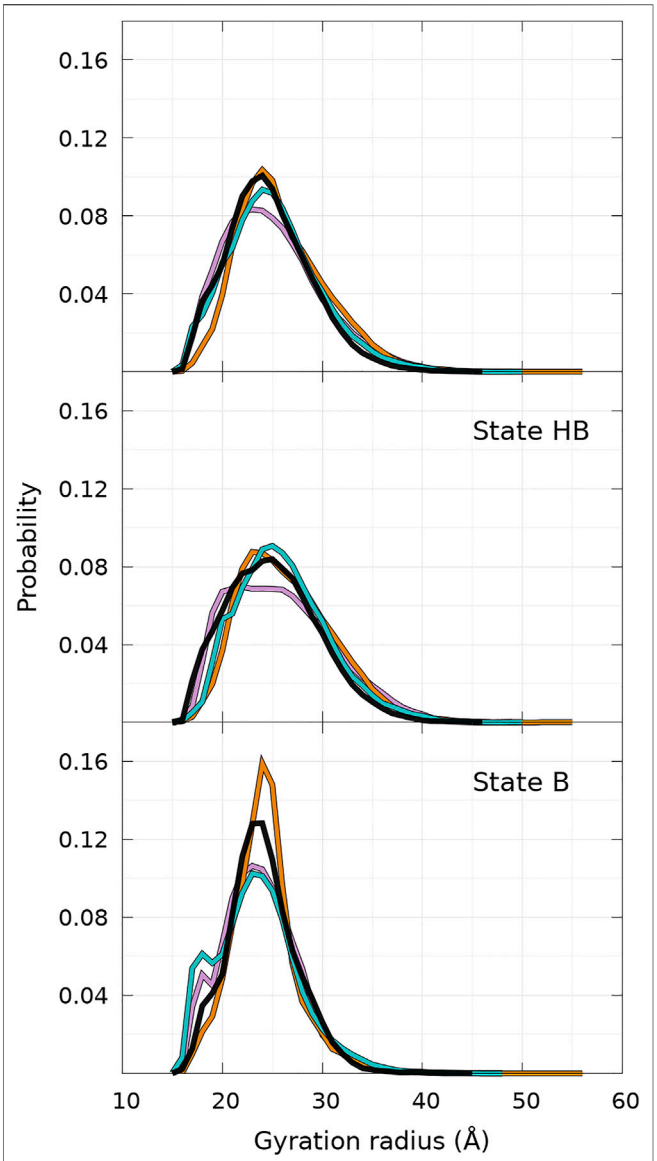

FIGURE 7 | Probability density of gyration radius computed for WT and mutants of α-syn for the complete conformational ensemble **(top panel),** the ensemble of state HB **(middle panel),** and the ensemble of state B **(bottom panel)**. The color code is WT (black), A30P (purple), E46K (turquoise), and A53T (orange). State B represents 40% of conformations for WT, A30P, and E46K and 25% for A53T (see text).

As shown in **Figure 8**, the probabilities to find residues in the β-sheet are significantly high in the N-terminal and NAC regions up to about residue 100 at the same locations for WT and mutants. For WT, the maximum of the peaks observed in **Figure 8** are P_{F4} = 0.57, P_{K10} = 0.61, P_{A17} = 0.87, P_{T22} = 0.76, P_{E28} = 0.65, P_{K34} = 0.61, P_{Y39} = 0.69, P_{K43} = 0.50, P_{V49} = 0.4, P_{V55} = 0.69, P_{V63} = 0.64, P_{V70} = 0.32, P_{K80} = 0.41, P_{A90} = 0.58, and P_{V95} = 0.35.

As shown in **Figure 8**, the β-sheet probabilities are very similar for WT and E46K along the sequence and differ significantly for A30P in the region 26—80 and for A53T in the region 35—65. The mutation A30P has a huge local impact on the probability of residue 30 to pertain to a β-sheet, $P_{WT,A30}$ = 0.49 and P_{P30} = 0.03. This induces an unexpected increase of the probability to occur in the β-sheet for the neighboring residues: P = 0.85 for residues 26—29 and 32—35. Long-range effects of A30P mutation on the propensities of other residues are observed by an increase of the peaks at P_{Y39} = 0.76, P_{K43} = 0.57, P_{V70} = 0.42, P_{T75} = 0.46, and P_{V95} = 0.46 and by a decrease of the peaks at P_{V55} = 0.53 and P_{V63} = 0.54, compared to WT. The decrease in β-sheet propensity in this region compared to WT agrees with their larger helical propensity in this region for the A30P mutant (**Figure 8**). For A53T, the most drastic effect of the single amino-acid substitution occurs in the region 53—65, where the probability to form β-sheets is significantly reduced compared to WT, P_{V55} = 0.32 and P_{V63} = 0.40, in agreement with their high probabilities to be in a helix (**Figure 8**). Other significant long-range effects of the amino-acid substitution are observed at peaks P_{Y39} = 0.55, P_{T44} = 0.37, and P_{V49} = 0.18, where the mutation A53T decreases the probability to form a β-sheet compared to WT. Overall, two crucial regions of the amino-acid sequence are affected by the mutations: the region 26—35, where the A30P mutation increases mainly the β-sheet formation, and the region 53—65, where the A53T mutation mainly decreases the β-sheet formation.

Finally, as shown in **Figure 9** for all proteins, the statistics of β-sheet propensity of state B differs from the global statistics (B + HB) in the region 50—100, which encompasses the two major helical regions centered at K58 and A78 (**Figure 9**, top panel). The most significant differences are observed for A53T, for which the probability to form a β-sheet is significantly lower in the HB state in the region 50—70, where the probability to form an helix is very high (**Figure 9**).

4 DISCUSSION

The present analysis is based on a first-principle (no bias or constraints applied) very large conformational sampling of the WT α-syn and mutants. As in any MD simulations, the sampling is never complete and each force-field has some bias. Next, we compare the predictions of UNRES for WT and mutants to available experimental data and previous theoretical studies.

To illustrate the difficulties of producing a conformational ensemble of α-syn, we compared (see **Figure 10**) the helix and β-sheet probabilities computed in four different works using MD simulations for WT α-syn (Ullman et al., 2011; Yu et al., 2015; Ramis et al., 2019; Coskuner and Wise-Scira, 2013). Compared to the UNRES simulations and Ref. Ullman et al., 2011, the calculations of Ref. Yu et al., 2015 largely overestimate the helical properties of α-syn, whereas the simulations of Ref. Ramis et al., 2019 largely underestimate them. The simulations of Ref. Coskuner and Wise-Scira, 2013 also predict a larger helical propensity than the one found in the present work (**Figure 8**) and in Ref. Ullman et al., 2011 (**Figure 10**). It is worth noting that the helical region nearby residue 60 is found in the UNRES simulations (**Figure 8**) and in Refs. Ullman et al., 2011; Coskuner and Wise-Scira, 2013 (**Figure 10**).

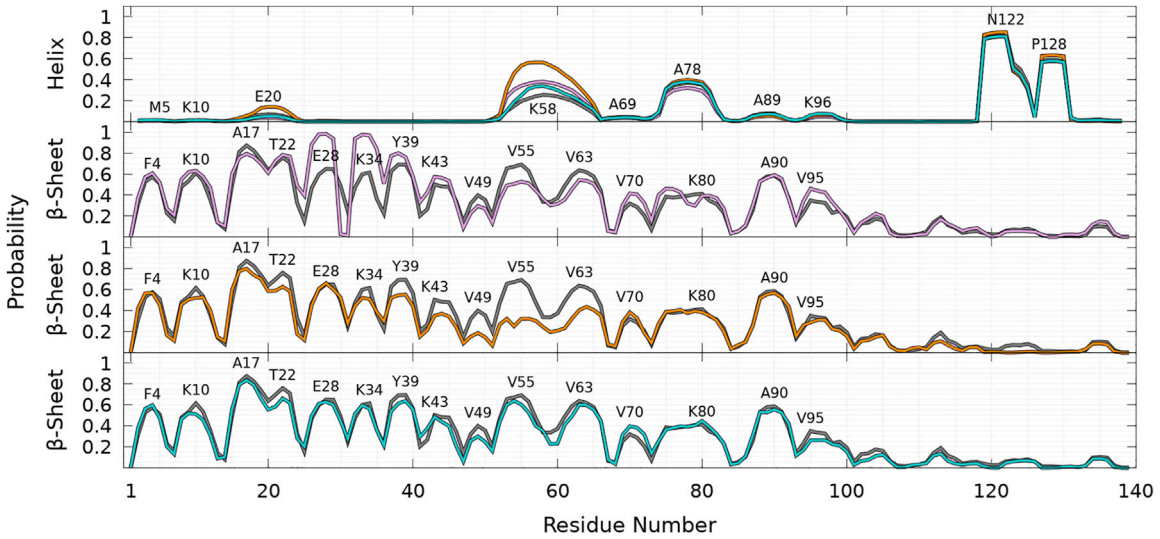

FIGURE 8 | Probability of each residue to pertain to an α-helix **(top panel)** and a β-sheet (other panels) for WT (gray), A30P (purple), A53T (orange), and E46K (turquoise) as a function of the residue position in the sequence. For the β-sheet, each mutant is compared to WT (gray) with labels pointing out residues at local maxima of the WT probability densities.

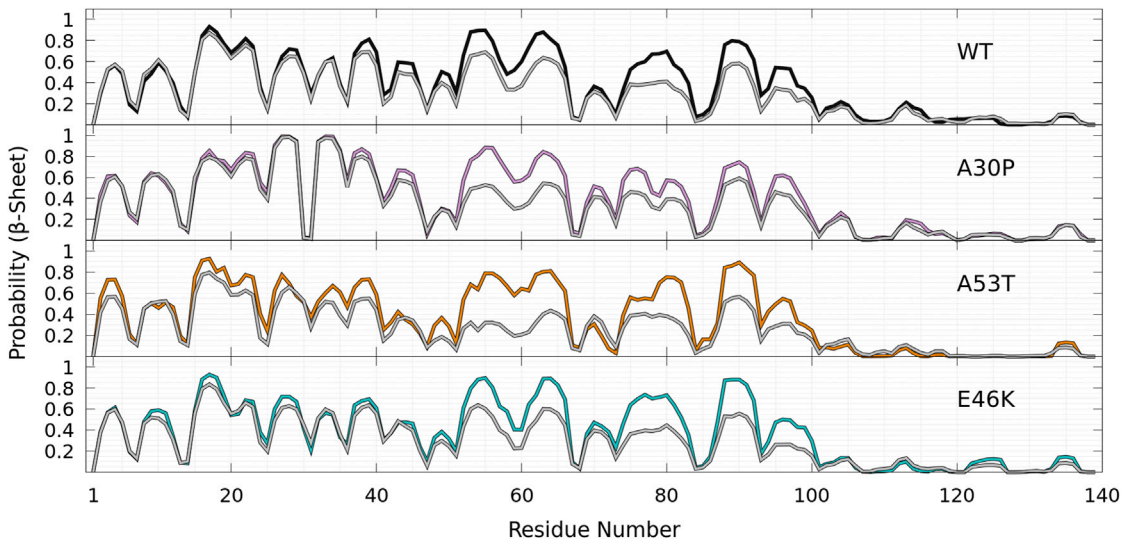

FIGURE 9 | Probability of each residue to pertain to a β-sheet for WT (black), A30P (purple), E46K (turquoise), and A53T (orange) in the B state compared to the probability in the corresponding entire (B + HB) ensemble of conformations for each protein (gray) as a function of the residue position in the sequence.

Comparison with our predictions shown in **Figure 8** indicates better agreement with Ref. Ullman et al., 2011 for the helical part, except that UNRES predicts a larger probability to form an α-helix at residue A78. In the present work, MD simulations are performed for the entire protein. In contrast, the method of Ref. Ullman et al., 2011 consists of a combination of MD simulations of peptides of α-syn with sequences which overlap each other and are extracted from the α-syn sequence. The conformation of the protein is rebuilt by a combination of these fragments with weights using NMR data. The formation of transient helical segments was observed experimentally in the WT α-syn tetramer in Ref. Wang et al., 2011 between residues 4—43 and 50—103. Larger probabilities of helical tendency were found experimentally in regions 20—23 and 48—90 (Wang et al., 2011). These findings are rather in good agreement with the transient existence of helices in the α-syn monomer, as shown in **Figure 8** (top panel), including the peak at A78. The stability of the tetrameric form is supported by previous all-atom MD simulations showing that the stability of this helical oligomer is due to a hydrophobic core formed by non-polar residues pertaining to the second α-helix of each chain, along with salt bridges formed by positively charged lysine residues and

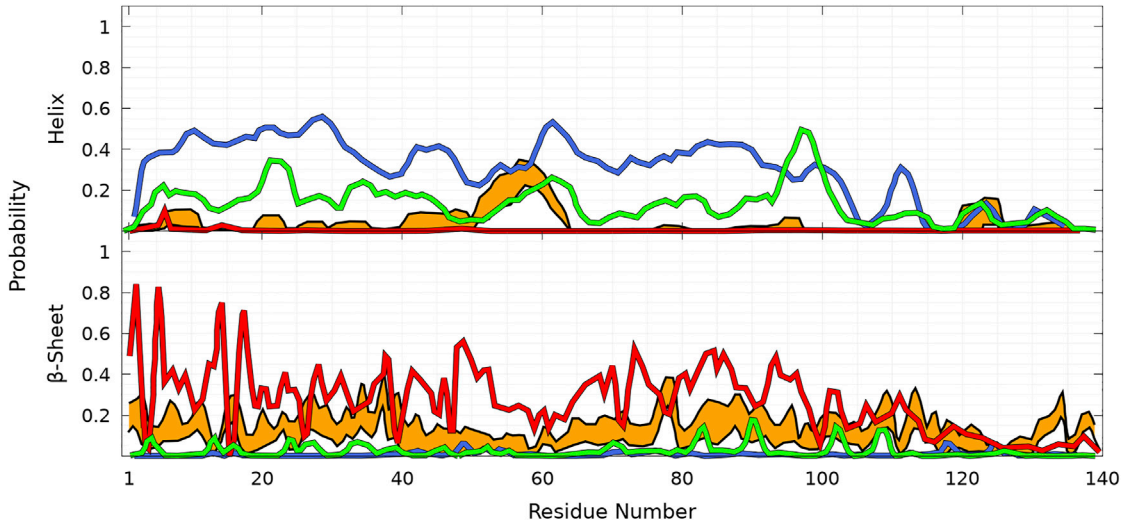

FIGURE 10 | Probability of each residue of WT to pertain to an α-helix **(top panel)** and a β-sheet **(bottom panel)** as a function of its position in the sequence from MD simulations according to four different works: Ullman et al., 2011 (orange), Yu et al., 2015 (blue), Ramis et al., 2019 (red), and Coskuner and Wise-Scira, 2013 (green). The thickness of the orange lines corresponds to the 95% confidence interval. Excerpted of the published figures of the works cited are reproduced with permissions. Further permissions related to the material excerpted should be directed to the corresponding publishers.

negatively charged aspartate and glutamate residues (Cote et al., 2018).

A detailed quantitative comparison between the UNRES β-sheet propensities (**Figure 8**) and those predicted in Refs. Ullman et al., 2011; Coskuner and Wise-Scira, 2013; Yu et al., 2015; Ramis et al., 2019 is difficult. However, we observe that there is a significant tendency to form β-sheet conformations over the entire sequence in Refs. Ullman et al., 2011; Ramis et al., 2019. On the contrary, simulations of Refs. Coskuner and Wise-Scira, 2013; Yu et al., 2015 predicted much localized and lower β-sheet probabilities.

Although the probabilities represented in **Figure 8** for intra-molecular β-sheets of an isolated monomer cannot be simply compared to the formation of intra-molecular contacts in fibrils, the peaks in UNRES predictions for the β-sheet propensities (**Figure 8**) agree with the pairs of residues forming intra-molecular contacts in protofilaments, i.e., 47–79, 48—78, 92—71, 93—70, 94—69, 95—68 (PDB ID: 2n0a), and 47—79, 48—78, 92—69, 93—68 (PDB ID 6h6b). In addition, the maxima in **Figure 8** for β-sheet propensity are located at or close to valine residues [P_{V3} = 0.52, P_{V26} = 0.46, P_{V40} = 0.57, P_{V49} = 0.4, P_{V55} = 0.69, P_{V63} = 0.64, P_{V70} = 0.32, P_{V82} = 0.31, P_{V95} = 0.35], which is the most frequently found amino acid in β-sheets (Chou and Fasman, 1974). The peak at A90 is in the hydrophobic stretch 88IAAA91. Present simulations show a significant propensity to form a β-sheet in the region 71—83 necessary for aggregation (Giasson et al., 2001). Except for G73, the probability is indeed about 40% in this region (**Figure 8**, bottom panel). However, as mentioned above, in the present simulations, the probability of helical tendency at A78 is also about 40%. The results shown in **Figure 8** are thus compatible with the probability to form both the β-sheet and helix in the region 71—83. The helix can be stabilized in the tetrameric oligomer (Bartels et al., 2011; Wang et al., 2011; Cote et al., 2018).

Although the propensities along the sequence are difficult to compare to experimental data, more global metrics can be used. Indeed, an interesting experimental parameter is the average content in helix and β-sheet conformations of WT α-syn in solution. Circular dichroism (CD) data reported an average of 2 ± 3% and 11 ± 7% for helix and β-sheet contents, respectively (Rekas et al., 2010). It is worth noting that the algorithms to extract β-sheet conformations from CD spectra are not as accurate as those for helices (Micsonai et al., 2015) and, as shown for α-syn, the CD spectrum is also dependent on the buffer and concentration (Araki et al., 2016). Other experimental values extracted from CD for the SSE content of WT monomers were reported: < 2% for the helix and 30% for β-sheet contents in Ref. Weinreb et al., 1996 and 3 ± 1% for the helix and 23 ± 8% for β-sheet fractions in Ref. Davidson et al., 1998. In their construction of the WT conformational ensemble based on MD simulations of α-syn fragments using the CHARMM force field constrained by NMR data, Ullman et al. reported the values 2% for the helix and 11% for β-sheet fractions (Ullman et al., 2011) (orange curve in **Figure 10**). In the present work, the global proportions of residues in the helix and β-sheet computed from the complete converged conformational WT ensemble are 10 and 31%, respectively. There are about 3 times more residues in the β-sheet than in the helix, a ratio which is in agreement with the one estimated from CD in Ref. Rekas et al., 2010. Interestingly, the theoretical values reported by coarse-grained simulations of WT α-syn with a different force field (red curve in **Figure 10**) are 20 ± 4% and 26.8 ± 6.8% for helix and β-sheet fractions, respectively (Ramis et al., 2019). In the present simulations, one notes that half of the helix fraction arises from the short helices located in the C-terminal region. The helix fraction without the C-terminal region is only 5% here. The fractions of the helix and β-sheet are on the same order of

magnitude for the mutants and WT: A30P (helix = 10%, β-sheet = 34%), A53T (helix = 13%, β-sheet = 26%), and E46K (helix = 10%, β-sheet = 29%). Again, A53T shows a larger deviation compared to WT in agreement with a less populated B state (**Figures 4** and **5**). Most likely, the present force field overestimates the formation of SSE but predicts the correct equilibrium between the two main SSEs.

Another global structural parameter is R_g. The average radius of gyration R_g measured by small-angle X-ray scattering for WT α-syn in solution depends on the protein concentration, the presence of dimers or trimers, buffer type, pH, acetylation, and the source of proteins (Araki et al., 2016). For recombinant α-syn in 10 mM ammonium acetate pH 7.4, R_g = 27.2 ± 0.44 Å extrapolated at infinite dilution, which is comparable to the average computed value (R_g = 24.7 Å, **Figure 7**). Addition of HCl and/or NaCl increases R_g significantly to 33—40 Å (Li et al., 2002; Allison et al., 2009; Araki et al., 2016), but these effects cannot be tested with the present model. It is interesting to note that R_g for dimers in the present simulations are about 10 Å larger than that of monomers, as will be discussed elsewhere. Therefore, any mixture of monomers and dimers increases the effective radius of gyration of the solution.

The effects of the missense mutations on the structural properties can be summarized as follows: for the SSE tendency (**Figure 8**), the most significant effects of the single mutations are for A30P and A53T. Residue P30 has a huge effect on neighboring residues for the formation of the β-sheet. Residue T53 has a great influence on the helical region centered at K58. For the B state, structures with low R_g (around 18 Å) are more compact for A30P and E46K due to the large average number of contacts between the residues 1—20 and 96—140. For the entire ensemble of conformations, the average number of contacts between these two regions is rather similar for WT (1.84), A30P (2.12), and E46K (1.61) but significantly less for A53T (0.75). The mutant A53T is thus expected to be more flexible and less compact as shown by its larger R_g. The analysis is slightly different if the average number of contacts is computed between the entire N-terminal (1—60) and C-terminal (96—140) regions for which the average number of contacts is 4.54 (WT), 4.22 (A30P), 2.71 (E46K), and 2.61 (A53T). One finds that both E46K and A53T are much more flexible than WT. NMR studies of mutants A30P and A53T showed a reduction of contacts between C- and N-terminals for both mutants compared to WT (Bertoncini et al., 2005). The present MD results fully agree with these data for A53T.

A major finding in the present work is that an isolated α-synuclein in solution occurs in two phases, which are clearly visible in **Figures 4** and **5**. Unfortunately, it is difficult to construct experimentally a two-dimensional map of SSE propensities such as in **Figure 4**. A possible road toward such experimental analysis could be the use of Raman single-molecule spectroscopy (Leray et al., 2016; Dai et al., 2021) as α-helices and β-sheets have been well described by Raman fingerprints (Dai et al., 2021). About 40% of conformations are in the B state for WT, A30P, and E46K and 25% for A53T, and such a significant fraction of the conformational ensemble might be detectable by spectroscopy.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

PS and GGM designed research. AG and PD performed research. AG, GGM, and PS analyzed the data. AR provided technical assistance with the UNRES package. AG did the figures and graphic representations, and PS wrote the paper. All authors discussed the results and reviewed the manuscript.

FUNDING

The calculations were performed using HPC resources from DSI-CCuB (Centre de Calcul de l'Université de Bourgogne). This work is part of the project NANO-NEURO-MED (2019–2022) supported by the EIPHI Graduate School (contract ANR-17-EUR-0002), the Conseil Régional de Bourgogne Franche-Comté, and the European Union through the PO FEDER-FSE Bourgogne 2014/2020 programs. This work was supported by a grant from the National Institutes of Health grant no. R01GM14312.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at https://www.frontiersin.org/articles/10.3389/fmolb.2021.786123/full#supplementary-material

REFERENCES

Allison, J. R., Varnai, P., Dobson, C. M., and Vendruscolo, M. (2009). Determination of the Free Energy Landscape of α-Synuclein Using Spin Label Nuclear Magnetic Resonance Measurements. *J. Am. Chem. Soc.* 131, 18314–18326. doi:10.1021/ja904716h

Antonschmidt, L., Dervişoğlu, R., Sant, V., Tekwani Movellan, K., Mey, I., Riedel, D., et al. (2021). Insights into the Molecular Mechanism of Amyloid Filament Formation: Segmental Folding of α-Synuclein on Lipid Membranes. *Sci. Adv.* 7, eabg2174. doi:10.1126/sciadv.abg2174

Appel-Cresswell, S., Vilarino-Guell, C., Encarnacion, M., Sherman, H., Yu, I., Shah, B., et al. (2013). Alpha-synuclein p.H50Q, a Novel Pathogenic Mutation for Parkinson's Disease. *Mov Disord.* 28, 811–813. doi:10.1002/mds.25421

Araki, K., Yagi, N., Nakatani, R., Sekiguchi, H., So, M., Yagi, H., et al. (2016). A Small-Angle X-ray Scattering Study of Alpha-Synuclein from Human Red Blood Cells. *Sci. Rep.* 6, 30473. doi:10.1038/srep30473

Bartels, T., Choi, J. G., and Selkoe, D. J. (2011). α-Synuclein Occurs Physiologically as a Helically Folded Tetramer that Resists Aggregation. *Nature* 477, 107–110. doi:10.1038/nature10324

Bertoncini, C. W., Jung, Y.-S., Fernandez, C. O., Hoyer, W., Griesinger, C., Jovin, T. M., et al. (2005). Release of Long-Range Tertiary Interactions Potentiates

Aggregation of Natively Unstructured -Synuclein. *Proc. Natl. Acad. Sci.* 102, 1430–1435. doi:10.1073/pnas.0407146102

Bhattacharjee, N., and Biswas, P. (2010). Position-Specific Propensities of Amino Acids in the Beta-Strand. *BMC Struct. Biol.* 10, 29. doi:10.1186/1472-6807-10-29

Bousset, L., Pieri, L., Ruiz-Arlandis, G., Gath, J., Jensen, P. H., Habenstein, B., et al. (2013). Structural and Functional Characterization of Two Alpha-Synuclein Strains. *Nat. Commun.* 4, 2575. doi:10.1038/ncomms3575

Breydo, L., Wu, J. W., and Uversky, V. N. (2012). α-Synuclein Misfolding and Parkinson's Disease. *Biochim. Biophys. Acta (Bba) - Mol. Basis Dis.* 1822, 261–285. doi:10.1016/j.bbadis.2011.10.002

Cascella, R., Chen, S. W., Bigi, A., Camino, J. D., Xu, C. K., Dobson, C. M., et al. (2021). The Release of Toxic Oligomers from α-Synuclein Fibrils Induces Dysfunction in Neuronal Cells. *Nat. Commun.* 12, 1814. doi:10.1038/s41467-021-21937-3

Chandonia, J.-M., Fox, N. K., and Brenner, S. E. (2019). SCOPe: Classification of Large Macromolecular Structures in the Structural Classification of Proteins-Extended Database. *Nucleic Acids Res.* 47, D475–D481. doi:10.1093/nar/gky1134

Cheng, C.-Y., Varkey, J., Ambroso, M. R., Langen, R., and Han, S. (2013). Hydration Dynamics as an Intrinsic Ruler for Refining Protein Structure at Lipid Membrane Interfaces. *Proc. Natl. Acad. Sci.* 110, 16838–16843. doi:10.1073/pnas.1307678110

Chiti, F., and Dobson, C. M. (2006). Protein Misfolding, Functional Amyloid, and Human Disease. *Annu. Rev. Biochem.* 75, 333–366. doi:10.1146/annurev.biochem.75.101304.123901

Chiti, F., and Dobson, C. M. (2017). Protein Misfolding, Amyloid Formation, and Human Disease: A Summary of Progress Over the Last Decade. *Annu. Rev. Biochem.* 86, 27–68. doi:10.1146/annurev-biochem-061516-045115

Chou, P. Y., and Fasman, G. D. (1974). Conformational Parameters for Amino Acids in Helical, β-Sheet, and Random Coil Regions Calculated from Proteins. *Biochemistry* 13, 211–222. doi:10.1021/bi00699a001

Coskuner, O., and Wise-Scira, O. (2013). Structures and Free Energy Landscapes of the A53T Mutant-type α-Synuclein Protein and Impact of A53T Mutation on the Structures of the Wild-Type α-Synuclein Protein with Dynamics. *ACS Chem. Neurosci.* 4, 1101–1113. doi:10.1021/cn400041j

Cote, Y., Delarue, P., Scheraga, H. A., Senet, P., and Maisuradze, G. G. (2018). From a Highly Disordered to a Metastable State: Uncovering Insights of α-Synuclein. *ACS Chem. Neurosci.* 9, 1051–1065. doi:10.1021/acschemneuro.7b00446

Cremades, N., Cohen, S. I. A., Deas, E., Abramov, A. Y., Chen, A. Y., Orte, A., et al. (2012). Direct Observation of the Interconversion of Normal and Toxic Forms of α-Synuclein. *Cell* 149, 1048–1059. doi:10.1016/j.cell.2012.03.037

Dai, X., Fu, W., Chi, H., Mesias, V. S. D., Zhu, H., Leung, C. W., et al. (2021). Optical Tweezers-Controlled Hotspot for Sensitive and Reproducible Surface-Enhanced Raman Spectroscopy Characterization of Native Protein Structures. *Nat. Commun.* 12, 1292. doi:10.1038/s41467-021-21543-3

Davidson, W. S., Jonas, A., Clayton, D. F., and George, J. M. (1998). Stabilization of α-Synuclein Secondary Structure upon Binding to Synthetic Membranes. *J. Biol. Chem.* 273, 9443–9449. doi:10.1074/jbc.273.16.9443

Dedmon, M. M., Lindorff-Larsen, K., Christodoulou, J., Vendruscolo, M., and Dobson, C. M. (2005). Mapping Long-Range Interactions in α-Synuclein Using Spin-Label NMR and Ensemble Molecular Dynamics Simulations. *J. Am. Chem. Soc.* 127, 476–477. doi:10.1021/ja044834j

Deiana, A., Forcelloni, S., Porrello, A., and Giansanti, A. (2019). Intrinsically Disordered Proteins and Structured Proteins with Intrinsically Disordered Regions Have Different Functional Roles in the Cell. *Plos One* 14, e0217889. doi:10.1371/journal.pone.0217889

Devi, L., Raghavendran, V., Prabhu, B. M., Avadhani, N. G., and Anandatheerthavarada, H. K. (2008). Mitochondrial Import and Accumulation of α-Synuclein Impair Complex I in Human Dopaminergic Neuronal Cultures and Parkinson Disease Brain. *J. Biol. Chem.* 283, 9089–9100. doi:10.1074/jbc.M710012200

Feig, M., Karanicolas, J., and Brooks, C. L. (2004). MMTSB Tool Set: Enhanced Sampling and Multiscale Modeling Methods for Applications in Structural Biology. *J. Mol. Graphics Model.* 22, 377–395. doi:10.1016/j.jmgm.2003.12.005

Flagmeier, P., Meisl, G., Vendruscolo, M., Knowles, T. P. J., Dobson, C. M., Buell, A. K., et al. (2016). Mutations Associated with Familial Parkinson's Disease Alter the Initiation and Amplification Steps of α-synuclein Aggregation. *Proc. Natl. Acad. Sci. USA* 113, 10328–10333. doi:10.1073/pnas.1604645113

Fox, N. K., Brenner, S. E., and Chandonia, J.-M. (2014). SCOPe: Structural Classification of Proteins-Extended, Integrating SCOP and ASTRAL Data and Classification of New Structures. *Nucl. Acids Res.* 42, D304–D309. doi:10.1093/nar/gkt1240

Fuchs, J., Tichopad, A., Golub, Y., Munz, M., Schweitzer, K. J., Wolf, B., et al. (2008). Genetic Variability in the SNCA Gene Influences α-Synuclein Levels in the Blood and Brain. *FASEB j.* 22, 1327–1334. doi:10.1096/fj.07-9348com

Fusco, G., Sanz-Hernandez, M., and De Simone, A. (2018). Order and Disorder in the Physiological Membrane Binding of α-Synuclein. *Curr. Opin. Struct. Biol.* 48, 49–57. doi:10.1016/j.sbi.2017.09.004

Galvagnion, C., Buell, A. K., Meisl, G., Michaels, T. C. T., Vendruscolo, M., Knowles, T. P. J., et al. (2015). Lipid Vesicles Trigger α-Synuclein Aggregation by Stimulating Primary Nucleation. *Nat. Chem. Biol.* 11, 229–234. doi:10.1038/nchembio.1750

Giasson, B. I., Murray, I. V. J., Trojanowski, J. Q., and Lee, V. M.-Y. (2001). A Hydrophobic Stretch of 12 Amino Acid Residues in the Middle of α-Synuclein Is Essential for Filament Assembly. *J. Biol. Chem.* 276, 2380–2386. doi:10.1074/jbc.M008919200

Gołaś, E., Maisuradze, G. G., Senet, P., Ołdziej, S., Czaplewski, C., Scheraga, H. A., et al. (2012). Simulation of the Opening and Closing of Hsp70 Chaperones by Coarse-Grained Molecular Dynamics. *J. Chem. Theor. Comput.* 8, 1750–1764. doi:10.1021/ct200680g

Grassein, P., Delarue, P., Nicolaï, A., Neiers, F., Scheraga, H. A., Maisuradze, G. G., et al. (2020). Curvature and Torsion of Protein Main Chain as Local Order Parameters of Protein Unfolding. *J. Phys. Chem. B* 124, 4391–4398. doi:10.1021/acs.jpcb.0c01230

Guerrero-Ferreira, R., Taylor, N. M., Arteni, A. A., Kumari, P., Mona, D., Ringler, P., et al. (2019). Two New Polymorphic Structures of Human Full-Length Alpha-Synuclein Fibrils Solved by Cryo-Electron Microscopy. *eLife* 8, e48907. doi:10.7554/eLife.48907

Guerrero-Ferreira, R., Kovacik, L., Ni, D., and Stahlberg, H. (2020). New Insights on the Structure of Alpha-Synuclein Fibrils Using Cryo-Electron Microscopy. *Curr. Opin. Neurobiol.* 61, 89–95. doi:10.1016/j.conb.2020.01.014

Henrich, M. T., Geibl, F. F., Lakshminarasimhan, H., Stegmann, A., Giasson, B. I., Mao, X., et al. (2020). Determinants of Seeding and Spreading of α-synuclein Pathology in the Brain. *Sci. Adv.* 6, eabc2487. doi:10.1126/sciadv.abc2487

Hsu, L. J., Sagara, Y., Arroyo, A., Rockenstein, E., Sisk, A., Mallory, M., et al. (2000). α-Synuclein Promotes Mitochondrial Deficit and Oxidative Stress. *Am. J. Pathol.* 157, 401–410. doi:10.1016/s0002-9440(10)64553-1

Jakes, R., Spillantini, M. G., and Goedert, M. (1994). Identification of Two Distinct Synucleins from Human Brain. *FEBS Lett.* 345, 27–32. doi:10.1016/0014-5793(94)00395-5

Kabsch, W., and Sander, C. (1983). Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-Bonded and Geometrical Features. *Biopolymers* 22, 2577–2637. doi:10.1002/bip.360221211

Khalili, M., Liwo, A., Jagielska, A., and Scheraga, H. A. (2005). Molecular Dynamics with the United-Residue Model of Polypeptide Chains. II. Langevin and Berendsen-Bath Dynamics and Tests on Model α-Helical Systems. *J. Phys. Chem. B.* 109, 13798–13810. doi:10.1021/jp058007w

Koh, E., Kim, T., and Cho, H.-S. (2006). Mean Curvature as a Major Determinant of Beta-Sheet Propensity. *Bioinformatics* 22, 297–302. doi:10.1093/bioinformatics/bti775

Konagurthu, A. S., Subramanian, R., Allison, L., Abramson, D., Stuckey, P. J., Garcia de la Banda, M., et al. (2020). Universal Architectural Concepts Underlying Protein Folding Patterns. *Front. Mol. Biosci.* 7, 612920. doi:10.3389/fmolb.2020.612920

Krüger, R., Kuhn, W., Müller, T., Woitalla, D., Graeber, M., Kösel, S., et al. (1998). AlaSOPro Mutation in the Gene Encoding α-synuclein in Parkinson's Disease. *Nat. Genet.* 18, 106–108. doi:10.1038/ng0298-106

Kumar, S., Jangir, D. K., Kumar, R., Kumari, M., Bhavesh, N. S., and Maiti, T. K. (2018). Role of Sporadic Parkinson Disease Associated Mutations A18T and A29S in Enhanced α-Synuclein Fibrillation and Cytotoxicity. *ACS Chem. Neurosci.* 9, 230–240. doi:10.1021/acschemneuro.6b00430

Kumari, P., Ghosh, D., Vanas, A., Fleischmann, Y., Wiegand, T., Jeschke, G., et al. (2021). Structural Insights into α-synuclein Monomer-Fibril Interactions. *Proc. Natl. Acad. Sci. USA* 118, e2012171118. doi:10.1073/pnas.2012171118

Labesse, G., Colloc'h, N., Pothier, J., and Mornon, J.-P. (1997). P-SEA: A New Efficient Assignment of Secondary Structure from C^{α} Trace of Proteins. *Bioinformatics* 13, 291–295. doi:10.1093/bioinformatics/13.3.291

Lashuel, H. A. (2020). Do Lewy Bodies Contain Alpha-Synuclein Fibrils? and Does it Matter? A Brief History and Critical Analysis of Recent Reports. *Neurobiol. Dis.* 141, 104876. doi:10.1016/j.nbd.2020.104876

Leray, A., Brulé, T., Buret, M., Colas des Francs, G., Bouhelier, A., Dereux, A., et al. (2016). Sorting of Single Biomolecules Based on Fourier Polar Representation of Surface Enhanced Raman Spectra. *Sci. Rep.* 6, 20383. doi:10.1038/srep20383

Li, J., Uversky, V. N., and Fink, A. L. (2002). Conformational Behavior of Human α-Synuclein Is Modulated by Familial Parkinson's Disease Point Mutations A30P and A53T. *Neuro.Toxicol.* 23, 553–567. doi:10.1016/S0161-813X(02)00066-9

Liwo, A., Czaplewski, C., Pillardy, J., and Scheraga, H. A. (2001). Cumulant-based Expressions for the Multibody Terms for the Correlation between Local and Electrostatic Interactions in the United-Residue Force Field. *J. Chem. Phys.* 115, 2323–2347. doi:10.1063/1.1383989

Liwo, A., Sieradzan, A. K., Lipska, A. G., Czaplewski, C., Joung, I., Żmudzińska, W., et al. (2019). A General Method for the Derivation of the Functional Forms of the Effective Energy Terms in Coarse-Grained Energy Functions of Polymers. III. Determination of Scale-Consistent Backbone-Local and Correlation Potentials in the UNRES Force Field and Force-Field Calibration and Validation. *J. Chem. Phys.* 150, 155104. doi:10.1063/1.5093015

Maisuradze, G. G., Senet, P., Czaplewski, C., Liwo, A., and Scheraga, H. A. (2010). Investigation of Protein Folding by Coarse-Grained Molecular Dynamics with the UNRES Force Field. *J. Phys. Chem. A.* 114, 4471–4485. doi:10.1021/jp9117776

Micsonai, A., Wien, F., Kernya, L., Lee, Y.-H., Goto, Y., Réfrégiers, M., et al. (2015). Accurate Secondary Structure Prediction and Fold Recognition for Circular Dichroism Spectroscopy. *Proc. Natl. Acad. Sci. USA* 112, E3095–E3103. doi:10.1073/pnas.1500851112

Mollenhauer, B., Cullen, V., Kahn, I., Krastins, B., Outeiro, T. F., Pepivani, I., et al. (2008). Direct Quantification of CSF α-Synuclein by ELISA and First Cross-Sectional Study in Patients with Neurodegeneration. *Exp. Neurol.* 213, 315–325. doi:10.1016/j.expneurol.2008.06.004

Mor, D. E., Ugras, S. E., Daniels, M. J., and Ischiropoulos, H. (2016). Dynamic Structural Flexibility of α-Synuclein. *Neurobiol. Dis.* 88, 66–74. doi:10.1016/j.nbd.2015.12.018

Murray, I. V. J., Giasson, B. I., Quinn, S. M., Koppaka, V., Axelsen, P. H., Ischiropoulos, H., et al. (2003). Role of α-Synuclein Carboxy-Terminus on Fibril Formation *In Vitro*. *Biochemistry* 42, 8530–8540. doi:10.1021/bi027363r

Pace, C. N., and Scholtz, J. M. (1998). A helix Propensity Scale Based on Experimental Studies of Peptides and Proteins. *Biophys. J.* 75, 422–427. doi:10.1016/s0006-3495(98)77529-0

Pasanen, P., Myllykangas, L., Siitonen, M., Raunio, A., Kaakkola, S., Lyytinen, J., et al. (2014). A Novel α-Synuclein Mutation A53E Associated with Atypical Multiple System Atrophy and Parkinson's Disease-type Pathology. *Neurobiol. Aging* 35, e1–2180. doi:10.1016/j.neurobiolaging.2014.03.024

Perrin, R. J., Woods, W. S., Clayton, D. F., and George, J. M. (2000). Interaction of Human α-Synuclein and Parkinson's Disease Variants with Phospholipids. *J. Biol. Chem.* 275, 34393–34398. doi:10.1074/jbc.M004851200

Petrucci, S., Ginevrino, M., and Valente, E. M. (2016). Phenotypic Spectrum of Alpha-Synuclein Mutations: New Insights from Patients and Cellular Models. *Parkinsonism Relat. Disord.* 22, S16–S20. doi:10.1016/j.parkreldis.2015.08.015

Polymeropoulos, M. H., Lavedan, C., Leroy, E., Ide, S. E., Dehejia, A., Dutra, A., et al. (1997). Mutation in the α-Synuclein Gene Identified in Families with Parkinson's Disease. *Science* 276, 2045–2047. doi:10.1126/science.276.5321.2045

Ramis, R., Ortega-Castro, J., Casasnovas, R., Mariño, L., Vilanova, B., Adrover, M., et al. (2019). A Coarse-Grained Molecular Dynamics Approach to the Study of the Intrinsically Disordered Protein α-Synuclein. *J. Chem. Inf. Model.* 59, 1458–1471. doi:10.1021/acs.jcim.8b00921

Ray, S., Singh, N., Kumar, R., Patel, K., Pandey, S., Datta, D., et al. (2020). α-Synuclein Aggregation Nucleates through Liquid-Liquid Phase Separation. *Nat. Chem.* 12, 705–716. doi:10.1038/s41557-020-0465-9

Rekas, A., Knott, R. B., Sokolova, A., Barnham, K. J., Perez, K. A., Masters, C. L., et al. (2010). The Structure of Dopamine Induced α-Synuclein Oligomers. *Eur. Biophys. J.* 39, 1407–1419. doi:10.1007/s00249-010-0595-x

Rotkiewicz, P., and Skolnick, J. (2008). Fast Procedure for Reconstruction of Full-Atom Protein Models from Reduced Representations. *J. Comput. Chem.* 29, 1460–1465. doi:10.1002/jcc.20906

Sang, J. C., Hidari, E., Meisl, G., Ranasinghe, R. T., Spillantini, M. G., and Klenerman, D. (2021). Super-Resolution Imaging Reveals α-Synuclein Seeded Aggregation in SH-SY5Y Cells. *Commun. Biol.* 4, 1–11. doi:10.1038/s42003-021-02126-w

Schrödinger, L. L. C. (2015). *The PyMOL Molecular Graphics System*. version 1.8. Available at https://pymol.org/2/support.html?.

Sieradzan, A. K., Korneev, A., Begun, A., Kachlishvili, K., Scheraga, H. A., Molochkov, A., et al. (2021). Investigation of Phosphorylation-Induced Folding of an Intrinsically Disordered Protein by Coarse-Grained Molecular Dynamics. *J. Chem. Theor. Comput.* 17, 3203–3220. doi:10.1021/acs.jctc.1c00155

Smith, C. K., Withka, J. M., and Regan, L. (1994). A Thermodynamic Scale for the Beta-Sheet Forming Tendencies of the Amino Acids. *Biochemistry* 33, 5510–5517. doi:10.1021/bi00184a020

Soto, C. (2003). Unfolding the Role of Protein Misfolding in Neurodegenerative Diseases. *Nat. Rev. Neurosci.* 4, 49–60. doi:10.1038/nrn1007

Spillantini, M. G., Schmidt, M. L., Lee, V. M.-Y., Trojanowski, J. Q., Jakes, R., and Goedert, M. (1997). α-Synuclein in Lewy Bodies. *Nature* 388, 839–840. doi:10.1038/42166

Stefanis, L. (2012). α-Synuclein in Parkinson's Disease. *Cold Spring Harbor Perspect. Med.* 2, a009399. doi:10.1101/cshperspect.a009399

Stephens, A. D., Zacharopoulou, M., Moons, R., Fusco, G., Seetaloo, N., Chiki, A., et al. (2020). Extent of N-Terminus Exposure of Monomeric Alpha-Synuclein Determines its Aggregation Propensity. *Nat. Commun.* 11, 2820. doi:10.1038/s41467-020-16564-3

Suzuki, M., Sango, K., Wada, K., and Nagai, Y. (2018). Pathological Role of Lipid Interaction with α-Synuclein in Parkinson's Disease. *Neurochem. Int.* 119, 97–106. doi:10.1016/j.neuint.2017.12.014

Tanudjojo, B., Shaikh, S. S., Fenyi, A., Bousset, L., Agarwal, D., Marsh, J., et al. (2021). Phenotypic Manifestation of α-synuclein Strains Derived from Parkinson's Disease and Multiple System Atrophy in Human Dopaminergic Neurons. *Nat. Commun.* 12, 3817. doi:10.1038/s41467-021-23682-z

Tosatto, L., Horrocks, M. H., Dear, A. J., Knowles, T. P. J., Dalla Serra, M., Cremades, N., et al. (2015). Single-molecule FRET Studies on Alpha-Synuclein Oligomerization of Parkinson's Disease Genetically Related Mutants. *Sci. Rep.* 5, 16696. doi:10.1038/srep16696

Touw, W. G., Baakman, C., Black, J., te Beek, T. A. H., Krieger, E., Joosten, R. P., et al. (2015). A Series of PDB-Related Databanks for Everyday Needs. *Nucleic Acids Res.* 43, D364–D368. doi:10.1093/nar/gku1028

Trinkaus, V. A., Riera-Tur, I., Martínez-Sánchez, A., Bäuerlein, F. J. B., Guo, Q., Arzberger, T., et al. (2021). *In Situ* Architecture of Neuronal α-Synuclein Inclusions. *Nat. Commun.* 12, 2110. doi:10.1038/s41467-021-22108-0

Tuttle, M. D., Comellas, G., Nieuwkoop, A. J., Covell, D. J., Berthold, D. A., Kloepper, K. D., et al. (2016). Solid-State NMR Structure of a Pathogenic Fibril of Full-Length Human α-synuclein. *Nat. Struct. Mol. Biol.* 23, 409–415. doi:10.1038/nsmb.3194

Uéda, K., Fukushima, H., Masliah, E., Xia, Y., Iwai, A., Yoshimoto, M., et al. (1993). Molecular Cloning of cDNA Encoding an Unrecognized Component of Amyloid in Alzheimer Disease. *Proc. Natl. Acad. Sci.* 90, 11282–11286. doi:10.1073/pnas.90.23.11282

Ullman, O., Fisher, C. K., and Stultz, C. M. (2011). Explaining the Structural Plasticity of α-Synuclein. *J. Am. Chem. Soc.* 133, 19536–19546. doi:10.1021/ja208657z

Ulmer, T. S., Bax, A., Cole, N. B., and Nussbaum, R. L. (2005). Structure and Dynamics of Micelle-Bound Human α-Synuclein. *J. Biol. Chem.* 280, 9595–9603. doi:10.1074/jbc.M411805200

Uversky, V. N., Gillespie, J. R., and Fink, A. L. (2000). Why Are "natively Unfolded" Proteins Unstructured under Physiologic Conditions? *Proteins* 41, 415–427. doi:10.1002/1097-0134(20001115)41:3<415:aid-prot130>3.0.co;2-7

Uversky, V. N. (2019). Intrinsically Disordered Proteins and Their "Mysterious" (Meta)Physics. *Front. Phys.* 7. doi:10.3389/fphy.2019.00010

Victor, J. M., Imbert, J. B., and Lhuillier, D. (1994). The Number of Contacts in a Self-Avoiding Walk of Variable Radius of Gyration in Two and Three Dimensions. *J. Chem. Phys.* 100, 5372–5377. doi:10.1063/1.467151

Wang, W., Perovic, I., Chittuluru, J., Kaganovich, A., Nguyen, L. T. T., Liao, J., et al. (2011). A Soluble α-synuclein Construct Forms a Dynamic Tetramer. *Proc. Natl. Acad. Sci.* 108, 17797–17802. doi:10.1073/pnas.1113260108

Weinreb, P. H., Zhen, W., Poon, A. W., Conway, K. A., and Lansbury, P. T. (1996). NACP, A Protein Implicated in Alzheimer's Disease and Learning, Is Natively Unfolded. *Biochemistry* 35, 13709–13715. doi:10.1021/bi961799n

Wright, P. E., and Dyson, H. J. (1999). Intrinsically Unstructured Proteins: Re-assessing the Protein Structure-Function Paradigm. *J. Mol. Biol.* 293, 321–331. doi:10.1006/jmbi.1999.3110

Yu, H., Han, W., Ma, W., and Schulten, K. (2015). Transient β-hairpin Formation in α-synuclein Monomer Revealed by Coarse-Grained Molecular Dynamics Simulation. *J. Chem. Phys.* 143, 243142. doi:10.1063/1.4936910

Zarranz, J. J., Alegre, J., Gómez-Esteban, J. C., Lezcano, E., Ros, R., Ampuero, I., et al. (2004). The New Mutation, E46K, of α-synuclein Causes Parkinson and Lewy Body Dementia. *Ann. Neurol.* 55, 164–173. doi:10.1002/ana.10795

Zhou, R., Maisuradze, G. G., Suñol, D., Todorovski, T., Macias, M. J., Xiao, Y., et al. (2014). Folding Kinetics of WW Domains with the United Residue Force Field for Bridging Microscopic Motions and Experimental Measurements. *Proc. Natl. Acad. Sci. USA.* 111, 18243–18248. doi:10.1073/pnas.1420914111

Conflict of Interest: Author AR is a employee of Schrödinger, Inc. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

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

frontiers in Molecular Biosciences

MINI REVIEW
published: 14 January 2022
doi: 10.3389/fmolb.2021.830304

Structure Determination of Microtubules and Pili: Past, Present, and Future Directions

James A. Garnett[1]* *and Joseph Atherton*[2]*

[1]Centre for Host-Microbiome Interactions, Faculty of Dental, Oral and Craniofacial Sciences, King's College London, London, United Kingdom, [2]Randall Centre for Cell and Molecular Biophysics, King's College London, London, United Kingdom

Historically proteins that form highly polymeric and filamentous assemblies have been notoriously difficult to study using high resolution structural techniques. This has been due to several factors that include structural heterogeneity, their large molecular mass, and available yields. However, over the past decade we are now seeing a major shift towards atomic resolution insight and the study of more complex heterogenous samples and *in situ/ex vivo* examination of multi-subunit complexes. Although supported by developments in solid state nuclear magnetic resonance spectroscopy (ssNMR) and computational approaches, this has primarily been due to advances in cryogenic electron microscopy (cryo-EM). The study of eukaryotic microtubules and bacterial pili are good examples, and in this review, we will give an overview of the technical innovations that have enabled this transition and highlight the advancements that have been made for these two systems. Looking to the future we will also describe systems that remain difficult to study and where further technical breakthroughs are required.

Keywords: cryo-EM, ssNMR, filament, fibre, microtubule, pilus

OPEN ACCESS

Edited by:
Annalisa Pastore,
King's College London,
United Kingdom

Reviewed by:
Louise Charlotte Serpell,
University of Sussex, United Kingdom

***Correspondence:**
James A. Garnett
james.garnett@kcl.ac.uk
Joseph Atherton
joseph.atherton@kcl.ac.uk

Specialty section:
This article was submitted to Structural Biology, a section of the journal Frontiers in Molecular Biosciences

Received: *07 December 2021*
Accepted: *28 December 2021*
Published: *14 January 2022*

Citation:
Garnett JA and Atherton J (2022) Structure Determination of Microtubules and Pili: Past, Present, and Future Directions. Front. Mol. Biosci. 8:830304. doi: 10.3389/fmolb.2021.830304

INTRODUCTION

In the 1930s-1950s, X-ray fibre diffraction studies of filamentous proteins by pioneers such as William Astbury, Francis Crick and Linus Pauling, laid the foundation for modern structural biology (Astbury and Street, 1932; Pauling and Corey, 1951a; Pauling and Corey, 1951b; Crick, 1952). These early studies provided new insights into the structural properties of fibrous substances such as keratin and collagen, however, they could only offer global information. With the advent of single crystal X-ray diffraction (Kendrew et al., 1958), from the 1960s structural studies of soluble proteins became the principal focus due to the delivery of atomic/subatomic resolutions. In the 1980s the first soluble protein was determined by solution state nuclear magnetic resonance (NMR) spectroscopy (Williamson et al., 1985) and with the development of recombinant protein expression systems (Itakura et al., 1977; Smith et al., 1983; Cregg et al., 1993), by the mid-1990s there became a clear exponential rise in soluble/globular macromolecular structures being deposited in the protein data bank (Berman et al., 2000). Likewise, new methods for isolating and reconstituting membrane proteins has led to significant numbers of these structures (Vinothkumar and Henderson, 2010), primarily elucidated by crystallographic methods, being deposited in the PDB since the 2000s.

Using solution NMR and X-ray crystallography we have gained significant understanding of filamentous systems through studying their lower-order subunits, however, due to technical limitations of these techniques, there has been a lack in our understanding of how these components interact and how this relates to their function. For example, there is a requirement

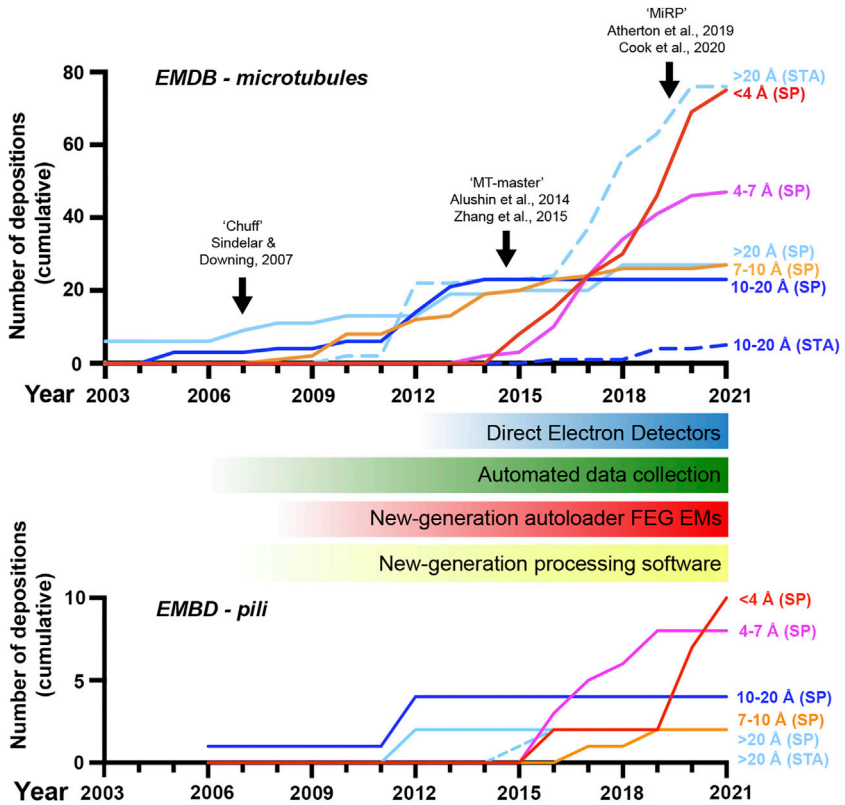

FIGURE 1 | A revolution in cryo-electron microscopy of MTs and pili. Graph showing the cumulative number of electron microscopy database (EMDB) depositions over time at different resolution levels as indicated; SP, single-particle cryo-EM, STA, sub-tomogram averaging. The emergence of new general hardware and software behind the revolution in cryo-EM is shown below the MT timeline, along with black arrows indicating the introduction of several image-processing pipelines for pseudo-helical single MTs in the top graph. Included data was based on the query title:MT AND status:REL or title:pilus AND status:REL at www.emdatasource.org.

for sample homogeneity (purity and molecular mass), high amounts of material, and for crystallography the sample must form highly ordered and relatively large crystals (generally >20–50 μm^3). Early advancements in the determination of high-resolution filamentous protein structures include the crystal structure of a synthetic peptide based collagen-like fragment with a defined length, published in 1994 (Bella et al., 1994). In the late 1990s the first proteinaceous structures were determined using solid state NMR (ssNMR) spectroscopy: the helical antibiotic peptide gramicidin, integrated within a lipid bilayer (Ketchem et al., 1996), and a peptide-based pentameric transmembrane helical bundle of the acetyl choline receptor (Opella et al., 1999). This ushered in a new direction for studying highly polymeric solids and by 2003 ssNMR had been successfully implemented to determine the structure of the fd filamentous bacteriophage particle coat protein (Zeri et al., 2003). Since then, the development of smaller magic angle spinning (MAS) sample rotors that spin at higher frequencies and require less sample, advances in isotopic labelling and partial deuteration of samples, and general enhancements of signal sensitivity (Ashbrook and Hodgkinson, 2018), has led to ~140 ssNMR-derived models being deposited in the PDB (Berman et al., 2000). Of these, ~20% are of filamentous proteins. Furthermore, in the last 10 years, hybrid approaches incorporating ssNMR, Rosetta-based *in silico* modelling and/or electron microscopy has provided atomic structures for the bacterial type III secretion system needle (Loquet et al., 2012; Demers et al., 2014), the M13 bacteriophage capsid (Morag et al., 2015) and Aβ amyloid fibrils (Sgourakis et al., 2015). However, over the past decade, cryo-electron microscopy (cryo-EM) has transformed all aspects of structural biology and has become the primary driving force in providing major advancements towards routine structure determination of filamentous protein assemblies (**Figure 1**).

The Cryo-EM Revolution

Pre-revolution, cryo-EM produced near-atomic resolutions only in ideal sample cases; usually large macromolecules with high symmetry such as particular viral capsids. In additional preferable cases, sub-nanometer resolutions allowing secondary structure visualisation were possible, but for most targets, particularly those of low symmetry and/or small size (<200 kDa), cryo-EM was most-often limited to nm resolutions. Now, cryo-EM can routinely produce near-atomic resolution structures of even asymmetrical macromolecules of small size (currently down to around ~50 kDa in ideal cases without a scaffold (Fan et al., 2019; Herzik et al., 2019)), with the first true atomic resolutions being reached in the last few years (Nakane et al., 2020; Yip et al., 2020).

Furthermore, a significant amount of sample heterogeneity can also now be tolerated and even utilised, with snapshots of different conformational states and multimeric arrangements revealing the dynamics of macromolecular machines. Whilst remaining more challenging than single-particle cryo-EM, in the last 5 years or so cryo-electron tomography (cryo-ET) with sub-tomogram averaging has become more capable of achieving sub-nanometer resolutions *in situ* and near-atomic resolutions with purified macromolecular preparations (Schur, 2019; Turk and Baumeister, 2020; Pyle and Zanetti, 2021).

Core to cryo-EM's transformation has been the introduction of direct electron detectors to replace charge-coupled devices and film, improving how faithfully transmitted electrons are recorded and allowing correction of global sample drift and local beam-induced motion (Li et al., 2013; Kuhlbrandt, 2014). Improvements to electron microscope hardware (including electron sources, stability, and energy filtration) and software (particularly increased automation in data collection) but also data processing hardware (increased computing power and improved storage) and software (in particular new Bayesian and artificial intelligence-based approaches) have also made significant contributions. The bottleneck to high-resolution cryo-EM is most often now the sample itself or the way it behaves during vitrification, however, innovation in sample preparation and data collection techniques are addressing these challenges.

In this review, we provide several examples of filamentous systems and show how our understanding has developed over the past decade due to these advances in structural biology techniques. We first discuss advancements in our understanding of microtubules (MTs) and MT complexes and then describe bacterial pili, primarily involved in adhesion. Looking to the future, we also highlight aspects that remain difficult to study and suggest where further advancements may be made.

Microtubules and Associated Proteins

MTs are tubular polymers of around 25 nm diameter built from longitudinally and laterally associated αβ-tubulin heterodimers and are a key component of the eukaryotic cytoskeleton. MTs display dynamic instability, in that they switch between polymerisation and depolymerisation phases modulated by β-tubulin's GTPase activity, post-translational modifications and the interaction of microtubule-binding proteins. MT dynamics generate forces vital to cell division, serve as intracellular signalling platforms and provide the tracks for intracellular transport and force generation by kinesin and dynein family motor proteins (Vale, 2003; Goodson and Jonasson, 2018). While MTs can form into a number of different architectures built from 10–16 protofilaments, 3-start 13-protofilament pseudo-helices are the most commonly observed in nature (Tilney et al., 1973; Pierson et al., 1978; Wade et al., 1990). Pseudo-helical MTs are those with a discontinuity in the helical lattice known as the seam that has heterotypic (α to β-tubulin) rather than homotypic (α to α/β to β) lateral interactions between tubulin dimers.

High-Resolution Studies of MTs Using Cryo-EM

MTs were amongst the first targets to be studied by cryo-EM (Mandelkow and Mandelkow, 1985) following its development (Lepault et al., 1983; McDowall et al., 1983). Cryo-EM at this early stage already provided key advantages over EM with heavy metal stains, enabled by rapid sample vitrification in near-native conditions, yet was chiefly limited to theoretical extrapolation of 3D MT lattice architectures *via* analysis of real and reciprocal space patterns from 2D projections (Chretien and Wade, 1991). In the late 90s and early 2000s, 3D reconstructions of MTs with or without associated motor proteins *via* helical and pseudo-tomographic back-projection methods were limited to nanometre resolutions (e.g. (Arnal et al., 1996; Hoenger et al., 1998; Kikkawa et al., 2000; Kikkawa et al., 2001; Metoz et al., 1997; Nogales et al., 1999; Sosa et al., 1997a; Sosa et al., 1997b)) (**Figure 1**). This work gave information on the MT polymer not available from emerging crystallographic structures of αβ-tubulin or protofilament subunits (Nogales et al., 1995; Nogales et al., 1998; Lowe et al., 2001). An important and lasting shift was treating MT segments as single-particles in reference-matching approaches (Li et al., 2002). This, often combined with refinement of helical symmetrisation parameters during iterative rounds of reference-based alignment (Egelman, 2000), helped move some MT reconstructions into the sub-nanometre range allowing secondary structure identifications (Bodey et al., 2009; Sui and Downing, 2010; Alushin et al., 2012) (**Figure 1**).

A major challenge is differentiating between highly similar α and β-tubulin monomers during MT image processing, resulting in significant blurring of α and β-tubulin and a failure to resolve the seam in more physiologically relevant pseudo-symmetric MTs (**Figure 2A**). This issue was particularly prominent in studies of MTs alone, while the presence of MT-bound proteins demarcating tubulin dimers would act as fiducials during processing, alleviating the severity of the artefact, particularly for pseudo-symmetric MT architectures. Combining statistical methods for seam-finding with pseudo-symmetrical averaging approaches efficiently identified α and β-tubulin register and seam-location with MT-binding proteins acting as fiducials (Sindelar and Downing, 2007). Nevertheless, CCD and film-derived reconstructions of pseudo-helical MTs and associated proteins could not break the ~4.5 Å resolution barrier required for visualisation of the peptide backbone and side chains (Alushin et al., 2014; Atherton et al., 2014; Fourniol et al., 2010; Maurer et al., 2012; Redwine et al., 2012; Sindelar and Downing, 2010). The introduction of direct electron detectors was central to MT reconstructions achieving near-atomic resolutions (<4.5 Å) (**Figure 1**). Ground-breaking work from the Nogales group, combining direct electron detector data with refined pseudo-helical processing methods achieved resolutions around 3.4 Å, allowing the authors to propose that small local nucleotide-dependent conformational changes leading to global changes in lattice compaction and twist govern dynamic instability (Zhang et al., 2015; Zhang and Nogales, 2015). This and later work to near-atomic resolutions continued to use MT binding proteins as fiducials for

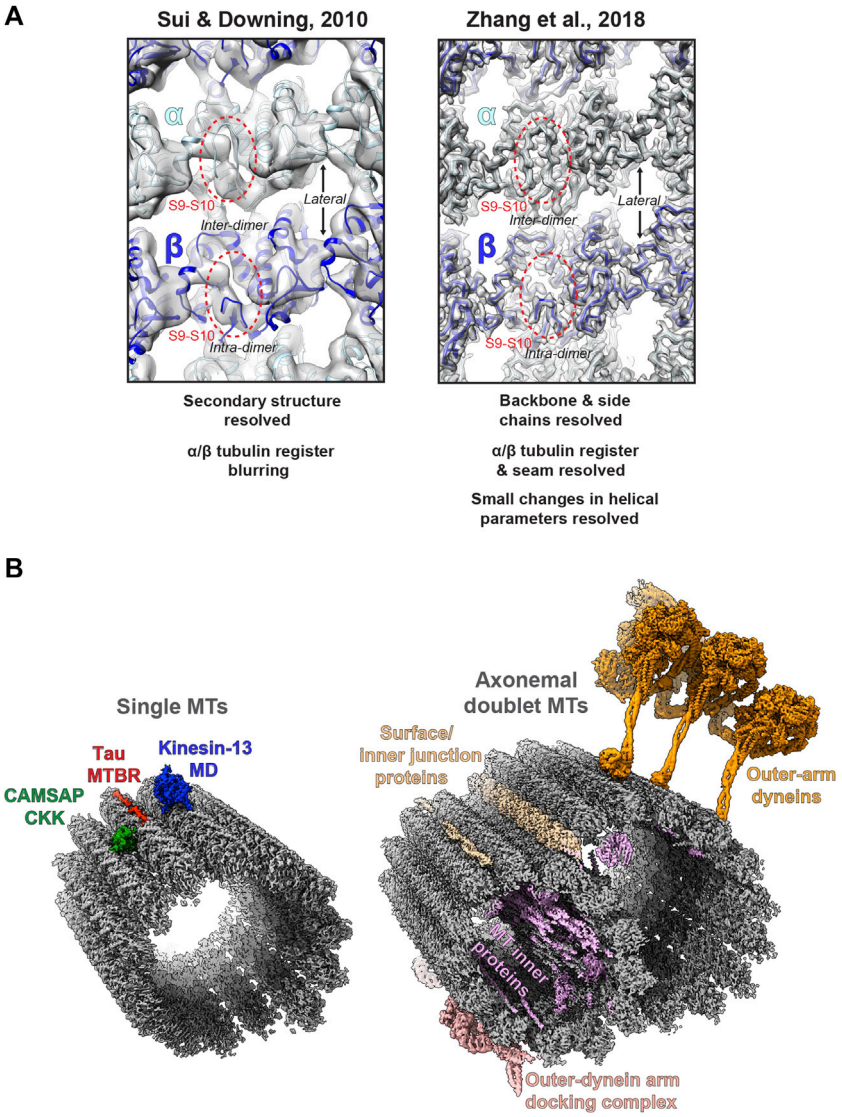

FIGURE 2 | Cryo-electron microscopy of MTs **(A)** Images of the lumenal face of undecorated MTs, centred on an inter-dimer interface, but also showing intra-dimer interfaces and lateral interfaces between protofilaments. Grey density is shown for reconstructions of the MT alone; top, EMDB:5193 at ~8 Å resolution (Sui and Downing, 2010); and bottom; EMDB:7973 at 3.1 Å resolution (Zhang et al., 2018), with the atomic model for the undecorated GMPCPP MT PDB:6dpu (Zhang et al., 2018) fitted into each reconstruction (α-tubulin light blue, β-tubulin dark blue). Failure to resolve differences between α and β-tubulin is a symptom of the earlier study (top), but not the more recent study (bottom), as illustrated by poor (top) or good (top) density differentiation between α and β-tubulin's S9-S10 loop (within red dashed oval) **(B)** Left; exemplar cryo-EM structures from single MTs. Coloured cryo-EM densities for MT-binding proteins are shown as indicated on the single MT alone cryo-EM density map EMDB:7973 (Zhang et al., 2018) coloured grey; CAMSAP1 CKK domain (Atherton et al., 2019), MT-binding repeat (MTBR) of Tau (Kellogg et al., 2018) and the motor domain (MD) of kinesin-13 (Benoit et al., 2018). Right; exemplar cryo-EM structures from axonemal doublet MTs. Coloured cryo-EM density for MT-binding proteins are shown as indicated on the bovine tracheal cilia doublet-MT cryo-EM density map EMDB:24664 (Gui et al., 2021); bovine tracheal cilia MT inner proteins (MIPs), surface and innter junction proteins and outer-dynein arm docking complex EMDB:24664 (Gui et al., 2021) and outer-arm dynein from *Tetrahymena thermophila* EMDB:22677 (Rao et al., 2021).

differentiation of α and β-tubulin in processing when working on tubulin isoforms or nucleotide or drug-induced conformational changes in the MT lattice; (Howes et al., 2018; Kellogg et al., 2017; Manka and Moores, 2018; Vemu et al., 2017; Vemu et al., 2016). However, this need for fiducials was overcome with additional image processing techniques, when near-atomic structural determination of pseudo-symmetric MTs without binding proteins in different nucleotide states was demonstrated (Zhang et al., 2018) (**Figure 2A**). Nevertheless, in many cases the MT-binding protein itself was the subject of structural interest, with notable examples solved near-atomic resolutions including the MT-binding regions of disease-related proteins tau (Kellogg et al., 2018) and doublecortin (Manka and Moores, 2020), minus-end binding CAMSAP (Atherton et al., 2019), MT-nucleator TPX2 (Zhang et al., 2017) and the MT-depolymerising kinesin-13 (Benoit et al., 2018) (**Figure 2A**).

In recent years, cryo-EM of MTs and binding partners has continued to develop (**Figure 1**). Near-atomic resolutions are being more readily achieved with accelerated automated and semi-automated data collection approaches (Tan et al., 2016) being adopted, allowing large amounts of data to be collected. Pipelines and methods for image processing of pseudo-symmetric MTs with or without MT-binding protein fiducials in the popular GUI-based program RELION have been introduced (Adib et al., 2019; Lacey et al., 2019; Cook et al., 2020). One challenge has been MT surface-binding proteins commonly being resolved at lower resolutions than the MT scaffold because heterogeneity in MT shape is generally amplified away from the centre of reconstructions, but also because of sample-dependent sub-stoichiometric occupancies and flexibility in the binder. In response to these challenges, new techniques have been developed to essentially subdivide MTs and bound proteins into sub-regions during processing, for example by using symmetry expansion and/or focused classification and refinement of individual MT subunits or binder sites to improve their quality (Liu et al., 2017; Debs et al., 2020; Cook et al., 2021). A recent stand-out study used focused classification methods to resolve MT inner proteins (MIPs) in the lumen of MTs extracted from *Toxoplasma gondii* to near-atomic resolution (Wang et al., 2021).

Although generally lagging-behind single-particle cryo-EM in resolution (**Figure 1**), cryo-ET can reveal 3D information on structurally heterogenous MT regions, such as lattice breaks and MT ends (Guesdon et al., 2016; Atherton et al., 2017; Gudimchuk et al., 2020). Furthermore, cryo-ET can be used to study MT architecture and organisation either *in situ* or in isolated *ex-vivo* preparations, where crowded overlapping environments render single-particle cryo-EM unsuitable (Atherton et al., 2018; McIntosh et al., 2018; Chakraborty et al., 2020). In particular cases where there are suitable repeating MT-associated sub-structures, sub-tomogram averaging (STA) can be employed to yield isotropic 3D reconstructions and improve resolution. For example, recently cryo-ET has resolved *ex-vivo* cytoplasmic dynein-dynactin transport teams on MTs (Chowdhury et al., 2015; Grotjahn et al., 2018), Parkinson's disease related LRRK2-decorated MTs in cells (Watanabe et al., 2020), EB-decorated singlet MTs inside primary cilia (Kiesel et al., 2020) and revealed MT intra-lumenal F-actin in kinesore-induced cell projections (Paul et al., 2020).

Finally, alongside work on single MTs and their binding proteins, there has been a recent flurry of exciting cryo-ET/STA (Jordan et al., 2018; Zabeo et al., 2018; Owa et al., 2019; Greenan et al., 2020) and single-particle cryo-EM (Ma et al., 2019; Khalifa et al., 2020; Gui et al., 2021; Rao et al., 2021; Walton et al., 2021) work on the axonemal MT doublet structures of primary and motile cilia (**Figure 2B**). These studies have utilised both intact *in situ* and reduced membranated or de-membranated *ex vivo* preparations and have revealed a wealth of information on gross cilia architecture and the organisation, identities, and structure of axonemal MIPs, dynein complexes and intraflagellar transport trains (IFTs).

Studying Dynamic Interactions of Microtubules and Associated Proteins

Prior to cryo-EM's revolution, ssNMR had provided a method of obtaining high-resolution information on drug-binding to MTs (Kumar et al., 2010). Nowadays, cryo-EM has become the tool of choice for studying the rigid interactions between MTs and their binding partners including small-molecules. However, cryo-EM struggles to resolve significant dynamics and flexible interactions due to the requirement for particle averaging and therefore ssNMR provides an ideal high-resolution method for studying the nature of these more dynamic interactions. Recent studies with labelled MT-binding domains of CAMSAPs, dynactin CAP-Gly, tau and plant companion of cellulose synthase 1 (CC1) have revealed atomic-level information on the dynamicity of their MT interfaces (Yan et al., 2015; Kadavath et al., 2018; Atherton et al., 2019; Kesten et al., 2019).

An exciting new development has been the purification of suitable amounts of isotopically labelled tubulin to produce MTs suitable for NMR studies, allowing labelled MTs and MT-binding proteins to be studied in parallel (Luo et al., 2021). In particular, this has enabled the study of the flexible and isoform-variable C-terminal tails of α and β tubulin involved in a plethora of interactions with MT-binding proteins within intact MTs (Janke, 2014; Roll-Mecak, 2020). This technical advance has now revealed the dynamic involvement of these C-terminal tails in CAMSAP-CKK domain and MAP7 MT binding at both slow and fast timescales (Luo et al., 2020) and has opened the door to further studies with a range of binding proteins.

Adhesive Bacterial Pili

The extracellular surfaces of bacteria are decorated with hair-like projections called pili or fimbriae, that are composed of smaller pilin subunits (Thanassi et al., 2012; Hospenthal et al., 2017a). Different types of pili range in their length and thickness and often have dedicated export and assembly systems that allow them to form on the bacterial surface. These include type IV-like and type V, chaperone-usher, amyloid-based, conjugative, type IV secretion, and sortase-mediated pili (Lukaszczyk et al., 2019). These pili have diverse functions including interacting with host cells during colonisation, promoting bacterial aggregation in biofilm formation, motility, conjugation, and secretion of proteins (Garnett and Matthews, 2012; Thanassi et al., 2012; Arutyunov and Frost, 2013; Berry and Pelicic, 2015). As such, these filamentous structures are often major virulence factors that drive the establishment of bacterial infection and the progression of disease. A decade ago, our primary understanding of pilus architectures was through crystallographic and solution NMR studies of monomeric pilin domains and the modelling of intact pili using low resolution (>10 Å) negative-stain and cryo-EM data (Craig et al., 2006; Salih et al., 2008; Garnett et al., 2012; Galkin et al., 2013). However, in the past 5 years ~15 intact pilus structures have now been deposited in the PDB derived using cryo-EM data at near near-atomic resolutions (<4.5 Å) (Costa et al., 2016; Hospenthal et al., 2016; Hospenthal et al., 2017b; Spaulding et al., 2018; Filman et al., 2019; Wang et al., 2019; Zheng et al., 2019; Neuhaus et al., 2020; Shibata et al., 2020; Zheng

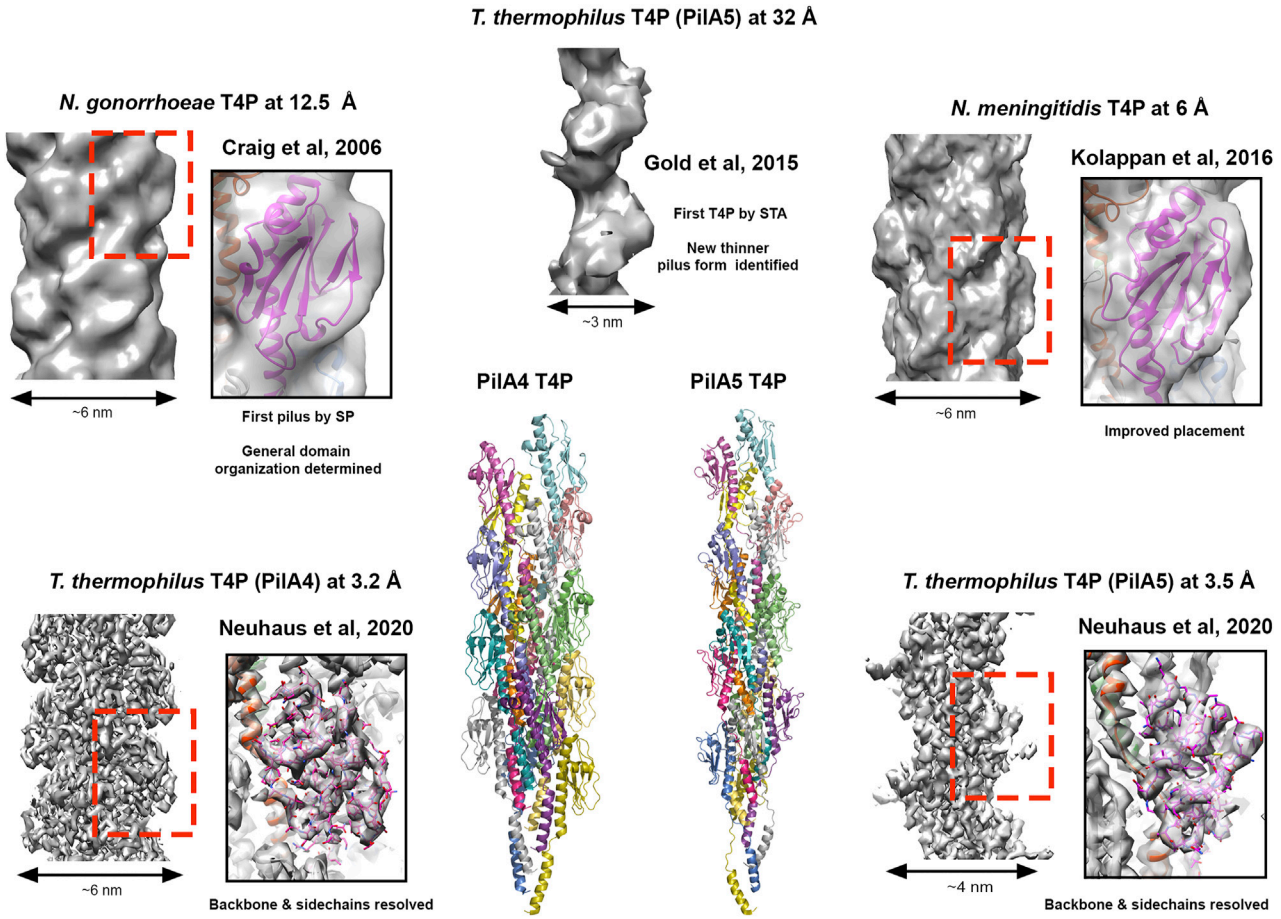

FIGURE 3 | Cryo-electron microscopy of type IV pili. Images of T4P models derived from different resolutions of SP data STA data. Structures providing resolution breakthroughs are shown as example. SP Cryo-EM density and models of "thick" pili (*N. meningitidis* T4P at 12.5 Å (EMDB 1236; PDBID 2hil) (Craig et al., 2006), *T. gonorrhoeae* T4P at 6.0 Å (EMDB 8287; PDBID 5kua) (Kolappan et al., 2016) and the *T. thermophilus* PilA4 T4P at 3.2 Å (EMDB 10647; PDBID 6xxd) (Neuhaus et al., 2020)) and a thin "pilus" (*T. thermophilus* PilA5 T4P at 3.5 Å (EMDB 10648; PDBID 6xxe) (Neuhaus et al., 2020)) are shown. Cryo-EM density for the thinner *T. thermophilus* PilA5 T4P derived by STA at 32 Å is also presented (EMDB 3024) (Gold et al., 2015). Cartoon representation of the *T. thermophilus* PilA4 and PilA5 are also given as the first Cryo-EM backbone/sidechain resolved T4P structures (Neuhaus et al., 2020).

et al., 2020; Gu et al., 2021; Pradhan et al., 2021). There has been a clear advancement in our ability to resolve high-resolution features of intact bacterial pili and this is well aligned with the introduction of direct electron detectors and processing techniques developed for filamentous helices (**Figure 1**). Here we will now discuss progress made in our understanding of two major classes of adhesive pili.

Type IV-like Pili

The first pilin structure was published in 1995 by John Tainer's group; the crystal structure of an intact type IV pilus (T4P) major subunit, PilE, from *Neisseria gonorrhoeae* (Parge et al., 1995). In *N. gonorrhea* this pilus is the only known virulence factor required for infection (Kellogg et al., 1963; Swanson et al., 1987) and is important for binding host cells, other bacteria and delivering antigenic variation (Quillin and Seifert, 2018). Furthermore, the ability of the T4P to both extend and retract, coupled with its adhesive properties, provide bacteria with twitching motility and the ability to conjugate genomic material (Pelicic, 2008). The structure of PilE indicated that a long N-terminal helix could mediate pilus assembly, and the globular C-terminal region may decorate the pilus and provide specific function. In 2006 this was realised using 12.5 Å cryo-EM data to model the intact pilus (Craig et al., 2006) which displayed a tight helical packing of N-terminal helices and with a width of ~6 nm (**Figure 3**). Then in 2015, cryo-ET was used to resolve the overall features of the PilA5 T4P isolated from Thermus thermophilus at 32 Å by STA (**Figure 3**). Strikingly, this pilus displayed a pronounced groove running along the fibre length and was much thinner at ~3 nm (Gold et al., 2015). Over the past 5 years we have seen significant improvements in resolution first with the *Neisseria meningitidis* T4P using cryo-EM maps reconstructed at 6 Å (Kolappan et al., 2016) and then several other T4P models published guided by sub-nanometre resolution data from *Escherichia coli*, *N. gonorrhoeae* and *Pseudomonas aeruginosa* PAK (Bardiaux et al., 2019; Wang et al., 2017)

(**Figure 3**). All these pili formed structures with widths of ~5–6 nm. However, in the last year Vicki Gold's group has now broken the ~4.5 Å barrier, with cryo-EM maps for the Thermus thermophilus PilA4 and PilA5 pili at 3.2 Å and 3.5 Å, respectively, allowing backbone and sidechain details to be resolved as well as sites of glycosylation (Neuhaus et al., 2020) (**Figure 3**). Similar to the *Neisseria*-like pili, the PilA4 pilus has a width of ~6 nm, while PilA5 is in line with the previous cryo-ET study with a width of ~4 nm. This study has provided a clear atomic rational for the differences in overall appearance of these two classes of pili, which likely reflects their specific functions in natural transformation and twitching motility, respectively, and is based on unique inter-subunit interactions, helical parameters, and surface charge.

Related to the T4P, pseudo-pili have similar structures but different functions. For example, components that form the T4P system and are essential for pilus assembly are related to the type II secretion system (T2SS) which transport substrates and effectors, many of which are virulence factors, from the Gram-negative bacterial periplasm into the extracellular space (Gu et al., 2017). While the T4P system utilise ATPases that can both drive the formation and retraction of the pilus, the T2SS only contains expresses the extension ATPase, which it uses to push cargo across the bacterial outer membrane *via* polymerisation of the pilus *via* a syringe-like mechanism. In 2017 the structure of the *Klebsiella oxytoca* T2SS pseudo-pilus, PulG, was resolved using solution NMR and cryo-EM maps at 5.0 Å resolution, and this study has revealed that this pilus is stabilised by calcium ions and disassembles in their absence (Lopez-Castilla et al., 2017; Naskar et al., 2021). *Geobacter sulfurreducens* is a Gram-negative bacterium that uses surface nanowires for extracellular electron transfer. A recent cryo-EM structure of the copolymerised PilA-N/C pseudo-pilus at 3.8 Å now indicates that like the T2SS, this structure is used to push nanowires out of the bacterium (Gu et al., 2021). Furthermore, cryo-EM structures of the OmcS nanowire at 3.7 Å and 3.4 Å has uncovered a new pilus type formed through the polymerisation of OmcS hexaheme cytochromes, with hemes packed within 3.5–6.0 Å of each other to allow electron transport (Filman et al., 2019; Wang et al., 2019).

Donor-Strand Exchanged Pili

The chaperone-usher (CU) pilus assembly pathway is another well characterised system in Gram-negative bacteria with the first structure, the uropathogenic *E. coli* (UPEC) type 1 pilus minor FimC-FimH chaperone-adhesin complex, published at the end of the 1990s by Stefan Knight's group. In CU systems, pilin domains consist of an incomplete Ig-like fold which lack the C-terminal strand, forming an acceptor groove, but have an additional unstructured extension at their N-terminus (Hospenthal et al., 2017a). Polymerisation proceeds through this N-terminal extension packing along the acceptor groove of an adjacent pilin subunit, which then stabilises and completes the Ig-like fold; a process called donor strand exchange. CU pili are highly variable but many consist of a major pilin subunit that makes up the majority of the fibre and then a minor pilin subunit (one or a few) at the tip, which is often an adhesin that binds carbohydrates or other receptors on the surface of host cells (e.g. *E. coli* type 1, P and common pili) (Garnett et al., 2012; Hospenthal et al., 2017a). However, in other CU systems, the major pilin domains that form the fibre shaft instead act as adhesive elements while the minor tip domain functions as an invasin that mediates invasion of host cells (e.g. *E. coli* AAF pili) (Berry et al., 2014); other arrangements also exist. Additionally, some CU pili are relatively thin (~2–3 nm) and exist with an extended 'beads on a string' architecture that are relatively dynamic, with others being much thicker (~10 nm) and forming rigid and more compacted helical arrangements.

A decade ago our understanding of CU pilus structures was primarily driven by crystallographic and NMR studies of generally monomeric or small engineered tandem subunits, however, again due to advances in cryo-EM we are now able to appreciate the functions of pilus packing through near-atomic resolution insights (Hospenthal et al., 2016; Hospenthal et al., 2017b; Spaulding et al., 2018; Zheng et al., 2019). In 2013 Lisa Craig's group determined the global features of the enterotoxigenic *E. coli* (ETEC) CS1 pilus at 20 Å resolution by cryo-EM and the major CS1 pilin subunit CooA by crystallography at 1.6 Å resolution (Galkin et al., 2013). Modelling the CooA subunits into different cryo-EM maps revealed that pilins could adopt multiple orientations and structural states, resulting in different pilin packing and a dynamic pilus structure. They proposed that CS1 and other thicker class of pili may stretch in response to shear forces that they experience during colonisation. In 2015 Adam Lange's group combined solution state NMR, ssNMR and scanning transmission electron microscopy (STEM) and determined the first atomic model of an intact UPEC type 1 pilus (Habenstein et al., 2015) while the following year, Gabriel Waksman's group published the first near-atomic resolution structure of an intact UPEC P-pilus by cryo-EM at 3.8 Å (Hospenthal et al., 2016). This structure along with a subsequent structure of the UPEC type 1 pilus by cryo-EM at 4.2 Å was able to explain precisely how changing inter-subunit interactions within these pili can mediate spring-like properties (Hospenthal et al., 2017b; Spaulding et al., 2018). Another recent study of the ETEC CFA/I pilus by cryo-EM at 4 Å has again shown that the helical quaternary structure of the pilus is influenced by shear forces and this is likely a common function of wound CU pili (Miller et al., 2006; Zheng et al., 2019).

Within the past 5 years a new type of adhesive fibre has also been discovered in Bacteroidales, named the type V pilus (T5P) (Xu et al., 2016). These are composed of an anchor, a stalk, an adapter and a tip pilin (Hospenthal et al., 2017a). Last year the first structure of an intact T5P was determined by cryo-EM at 3.6 Å, which consisted of the polymerised FimA pilin stalk subunit from the bacterium *Porphromonas gingivalis* (Shibata et al., 2020). This has revealed not just a new pilus architecture but also a new mode of pilus formation. Unlike the CU pathway, T5P pilin domains contain a C-terminal extended region and when they are exported to the bacterial surface, the N-terminal strand is cleaved by the protease RgpB and released. This forms an acceptor groove that can then accept a C-terminal extension *via* strand exchange with an adjacent pilin subunit.

CONCLUSIONS AND PERSPECTIVES

As has been discussed, structural biology of MTs and their binding proteins has advanced dramatically over the last decade, yet a number of samples and technical goals remain challenging, representing frontiers in the field. In single-particle cryo-EM of single MTs, an intriguing question is whether atomic resolutions are achievable. MTs include some intrinsic flexibility and heterogeneity, including bending and in some cases lattice breaks and even multiple seams (Cross, 2019; Debs et al., 2020). Future sample preparation, data collection and processing methods will have to reduce these variances, whilst essentially eliminating blurring of α and β-tubulin register (particularly challenging in the absence of fiducial binding proteins) in order to reach atomic resolutions. Resolving MT binding proteins with partial occupancy on the MT, or those adopting a mix of conformers remains difficult, although processing strategies including symmetry expansion, focused classifications and refinements and signal subtraction are proving fruitful (Liu et al., 2017; Debs et al., 2020; Cook et al., 2021). In some cases, rather than using the standard heterogenous tubulin purified from the brains of livestock, more homogenous sources such as purified single-isoform tubulins or tubulin from particular cell types can help improve resolve regions of tubulin isoform heterogeneity, such as the C-terminal tails and their interactions (Atherton et al., 2019; Li et al., 2020). Additionally, ssNMR is becoming a useful tool for particular cases where flexible interactions between MT-binding proteins and the MT are to be investigated.

As discussed, exciting recent data has also emerged from studies of cilia MTs that demonstrates the ability of cryo-EM to identify macromolecules *de novo* within *in situ* or *ex vivo* preparations (Gui et al., 2021; Kiesel et al., 2020; Wang et al., 2021). The ability to faithfully localise and identify macromolecules on MTs *in situ* is likely to be a prominent target in the future direction of the field. With the advent of a revolution in macromolecular structure prediction (Jumper et al., 2021) combined with expanding dataset sizes and increasing sub-tomogram averaging resolutions due to steady improvements in cryo-ET sample preparation (FIB), data-collection and image processing (**Figure 1**) (Turk and Baumeister, 2020; Hylton and Swulius, 2021; Pyle and Zanetti, 2021), this goal is looking more and more plausible.

On the back of cryo-EM method developments, primarily through the study of other filamentous systems such as MTs, there has also been a substantial increase in our understanding of bacterial pilus structures in recent years. This is highlighted by work shown here on type IV-like pili, CU pili and the newly characterised T5P, however, we still lack high-resolution insight of other types of pili, such as those formed from amyloids (Van Gerven et al., 2018), and lack an understanding of how minor pilin components interact within intact pilus assemblies. In many type IV pili, incorporation of minor pilins is required for pilus assembly and/or specific functions (Jacobsen et al., 2020) but it is unclear how these are distributed across the fibre or what effect these minor components have on the local pilus structure. Crystallographic and more recently cryo-EM studies have provided great insight into the atomic mechanisms of pilus biogenesis (Hospenthal et al., 2017a), however, pili often remain tethered to their secretion machinery, and there is a shortage of high-resolution observations in these states. The UPEC type 1 and P pili have become model systems to study CU pilus biogenesis and several groups have now provided snapshots of the initiation of pilus assembly at the outer membrane usher pore and the initial stages of adhesin exit into the extracellular space (Phan et al., 2011; Geibel et al., 2013; Du et al., 2018; Omattage et al., 2018; Du et al., 2021). With cryo-EM now leading the way here, the next leap will likely come once a defined number of major pilin subunits can be incorporated into these systems and then the maturation of the pilus can then be visualized in the context of secretion.

Although cryo-ET has not been used to study the CU pathway, it has been successfully implemented to elucidate the low-resolution features of T4P and T2SS pseudo-pilus assembly devices (Gold et al., 2015; Chang et al., 2016; Ghosal et al., 2019). Work on the T4P systems has revealed pilus structures emanating into the extracellular space (Gold et al., 2015; Chang et al., 2016) and with the expected future increases in cryo-ET resolution, as discussed above, in the future cryo-ET may be able to provide greater detail of how these pili form at the membrane and the role of pilus-specific assembly factors in their biogenesis (Hu et al., 2020). Cryo-EM has had a major impact on the of study filamentous systems over the past decade and it will be interesting to see how EM and its incorporation with other approaches (e.g. *in silico* prediction, solution/ssNMR, crystallography) will be able tackle more dynamic and heterogenous filamentous systems in the future.

AUTHOR CONTRIBUTIONS

JG and JA wrote the manuscript. Both the authors contributed to the article and approved the submitted version.

REFERENCES

Adib, R., Montgomery, J. M., Atherton, J., O'Regan, L., Richards, M. W., Straatman, K. R., et al. (2019). Mitotic Phosphorylation by NEK6 and NEK7 Reduces the Microtubule Affinity of EML4 to Promote Chromosome Congression. *Sci. Signal.* 12. doi:10.1126/scisignal.aaw2939

A. Garnett, J., and Matthews, S. (2012). Interactions in Bacterial Biofilm Development: a Structural Perspective. *Curr. Protein Pept. Sci.* 13, 739–755. doi:10.2174/138920312804871166

Alushin, G. M., Lander, G. C., Kellogg, E. H., Zhang, R., Baker, D., and Nogales, E. (2014). High-Resolution Microtubule Structures Reveal the Structural Transitions in αβ-Tubulin upon GTP Hydrolysis. *Cell* 157, 1117–1129. doi:10.1016/j.cell.2014.03.053

Alushin, G. M., Musinipally, V., Matson, D., Tooley, J., Stukenberg, P. T., and Nogales, E. (2012). Multimodal Microtubule Binding by the Ndc80 Kinetochore Complex. *Nat. Struct. Mol. Biol.* 19, 1161–1167. doi:10.1038/nsmb.2411

Arnal, I., Metoz, F., DeBonis, S., and Wade, R. H. (1996). Three-dimensional Structure of Functional Motor Proteins on Microtubules. *Curr. Biol.* 6, 1265–1270. doi:10.1016/s0960-9822(02)70712-4

Arutyunov, D., and Frost, L. S. (2013). F Conjugation: Back to the Beginning. *Plasmid* 70, 18–32. doi:10.1016/j.plasmid.2013.03.010

Ashbrook, S. E., and Hodgkinson, P. (2018). Perspective: Current Advances in Solid-State NMR Spectroscopy. *J. Chem. Phys.* 149, 040901. doi:10.1063/1.5038547

Astbury, W. T., and Street, A. (1932). X-Ray Studies of the Structure of Hair, Wool, and Related Fibres. I. General, *Series A, Containing Papers of a Mathematical or Physical Character*, 230. Philosophical Transactions of the Royal Society of London, 75–101. doi:10.1098/rsta.1932.0003

Atherton, J., Farabella, I., Yu, I. M., Rosenfeld, S. S., Houdusse, A., Topf, M., et al. (2014). Conserved Mechanisms of Microtubule-Stimulated ADP Release, ATP Binding, and Force Generation in Transport Kinesins. *Elife* 3, e03680. doi:10.7554/eLife.03680

Atherton, J., Jiang, K., Stangier, M. M., Luo, Y., Hua, S., Houben, K., et al. (2017). A Structural Model for Microtubule Minus-End Recognition and protection by CAMSAP Proteins. *Nat. Struct. Mol. Biol.* 24, 931–943. doi:10.1038/nsmb.3483

Atherton, J., Luo, Y., Xiang, S., Yang, C., Rai, A., Jiang, K., et al. (2019). Structural Determinants of Microtubule Minus End Preference in CAMSAP CKK Domains. *Nat. Commun.* 10, 5236. doi:10.1038/s41467-019-13247-6

Atherton, J., Stouffer, M., Francis, F., and Moores, C. A. (2018). Microtubule Architecture *In Vitro* and in Cells Revealed by Cryo-Electron Tomography. *Acta Cryst. Sect D Struct. Biol.* 74, 572–584. doi:10.1107/s2059798318001948

Bardiaux, B., de Amorim, G. C., Luna Rico, A., Zheng, W., Guilvout, I., Jollivet, C., et al. (2019). Structure and Assembly of the Enterohemorrhagic *Escherichia coli* Type 4 Pilus. *Structure* 27, 1082–1093. doi:10.1016/j.str.2019.03.021

Bella, J., Eaton, M., Brodsky, B., and Berman, H. M. (1994). Crystal and Molecular Structure of a Collagen-like Peptide at 1.9 Å Resolution. *Science* 266, 75–81. doi:10.1126/science.7695699

Benoit, M. P. M. H., Asenjo, A. B., and Sosa, H. (2018). Cryo-EM Reveals the Structural Basis of Microtubule Depolymerization by Kinesin-13s. *Nat. Commun.* 9, 1662. doi:10.1038/s41467-018-04044-8

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

Berry, A. A., Yang, Y., Pakharukova, N., Garnett, J. A., Lee, W.-c., Cota, E., et al. (2014). Structural Insight into Host Recognition by Aggregative Adherence Fimbriae of Enteroaggregative *Escherichia coli*. *Plos Pathog.* 10, e1004404. doi:10.1371/journal.ppat.1004404

Berry, J.-L., and Pelicic, V. (2015). Exceptionally Widespread Nanomachines Composed of Type IV Pilins: the Prokaryotic Swiss Army Knives. *FEMS Microbiol. Rev.* 39, 134–154. doi:10.1093/femsre/fuu001

Bodey, A. J., Kikkawa, M., and Moores, C. A. (2009). 9-Ångström Structure of a Microtubule-Bound Mitotic Motor. *J. Mol. Biol.* 388, 218–224. doi:10.1016/j.jmb.2009.03.008

Chakraborty, S., Mahamid, J., and Baumeister, W. (2020). Cryoelectron Tomography Reveals Nanoscale Organization of the Cytoskeleton and its Relation to Microtubule Curvature inside Cells. *Structure* 28, 991–1003. doi:10.1016/j.str.2020.05.013

Chang, Y. W., Rettberg, L. A., Treuner-Lange, A., Iwasa, J., Søgaard-Andersen, L., and Jensen, G. J. (2016). Architecture of the Type IVa Pilus Machine. *Science* 351, aad2001. doi:10.1126/science.aad2001

Chowdhury, S., Ketcham, S. A., Schroer, T. A., and Lander, G. C. (2015). Structural Organization of the Dynein-Dynactin Complex Bound to Microtubules. *Nat. Struct. Mol. Biol.* 22, 345–347. doi:10.1038/nsmb.2996

Chrétien, D., and Wade, R. H. (1991). New Data on the Microtubule Surface Lattice. *Biol. Cel* 71, 161–174. doi:10.1016/0248-4900(91)90062-r

Cook, A. D., Manka, S. W., Wang, S., Moores, C. A., and Atherton, J. (2020). A Microtubule RELION-Based Pipeline for Cryo-EM Image Processing. *J. Struct. Biol.* 209, 107402. doi:10.1016/j.jsb.2019.10.004

Cook, A. D., Roberts, A. J., Atherton, J., Tewari, R., Topf, M., and Moores, C. A. (2021). Cryo-EM Structure of a Microtubule-Bound Parasite Kinesin Motor and Implications for its Mechanism and Inhibition. *J. Biol. Chem.* 297, 101063. doi:10.1016/j.jbc.2021.101063

Costa, T. R. D., Ilangovan, A., Ukleja, M., Redzej, A., Santini, J. M., Smith, T. K., et al. (2016). Structure of the Bacterial Sex F Pilus Reveals an Assembly of a Stoichiometric Protein-Phospholipid Complex. *Cell* 166, 1436–1444. doi:10.1016/j.cell.2016.08.025

Craig, L., Volkmann, N., Arvai, A. S., Pique, M. E., Yeager, M., Egelman, E. H., et al. (2006). Type IV Pilus Structure by Cryo-Electron Microscopy and Crystallography: Implications for Pilus Assembly and Functions. *Mol. Cel* 23, 651–662. doi:10.1016/j.molcel.2006.07.004

Cregg, J. M., Vedvick, T. S., and Raschke, W. C. (1993). Recent Advances in the Expression of Foreign Genes in Pichia pastoris. *Nat. Biotechnol.* 11, 905–910. doi:10.1038/nbt0893-905

Crick, F. H. C. (1952). Is α-Keratin a Coiled Coil? *Nature* 170, 882–883. doi:10.1038/170882b0

Cross, R. A. (2019). Microtubule Lattice Plasticity. *Curr. Opin. Cel Biol.* 56, 88–93. doi:10.1016/j.ceb.2018.10.004

Debs, G. E., Cha, M., Liu, X., Huehn, A. R., and Sindelar, C. V. (2020). Dynamic and Asymmetric Fluctuations in the Microtubule wall Captured by High-Resolution Cryoelectron Microscopy. *Proc. Natl. Acad. Sci. USA* 117, 16976–16984. doi:10.1073/pnas.2001546117

Demers, J.-P., Habenstein, B., Loquet, A., Kumar Vasa, S., Giller, K., Becker, S., et al. (2014). High-resolution Structure of the Shigella Type-III Secretion Needle by Solid-State NMR and Cryo-Electron Microscopy. *Nat. Commun.* 5, 4976. doi:10.1038/ncomms5976

Du, M., Yuan, Z., Werneburg, G. T., Henderson, N. S., Chauhan, H., Kovach, A., et al. (2021). Processive Dynamics of the Usher Assembly Platform during Uropathogenic *Escherichia coli* P Pilus Biogenesis. *Nat. Commun.* 12, 5207. doi:10.1038/s41467-021-25522-6

Du, M., Yuan, Z., Yu, H., Henderson, N., Sarowar, S., Zhao, G., et al. (2018). Handover Mechanism of the Growing Pilus by the Bacterial Outer-Membrane Usher FimD. *Nature* 562, 444–447. doi:10.1038/s41586-018-0587-z

Egelman, E. H. (2000). A Robust Algorithm for the Reconstruction of Helical Filaments Using Single-Particle Methods. *Ultramicroscopy* 85, 225–234. doi:10.1016/s0304-3991(00)00062-0

Fan, X., Wang, J., Zhang, X., Yang, Z., Zhang, J.-C., Zhao, L., et al. (2019). Single Particle Cryo-EM Reconstruction of 52 kDa Streptavidin at 3.2 Angstrom Resolution. *Nat. Commun.* 10, 2386. doi:10.1038/s41467-019-10368-w

Filman, D. J., Marino, S. F., Ward, J. E., Yang, L., Mester, Z., Bullitt, E., et al. (2019). Cryo-EM Reveals the Structural Basis of Long-Range Electron Transport in a Cytochrome-Based Bacterial Nanowire. *Commun. Biol.* 2, 219. doi:10.1038/s42003-019-0448-9

Fourniol, F. J., Sindelar, C. V., Amigues, B., Clare, D. K., Thomas, G., Perderiset, M., et al. (2010). Template-free 13-protofilament Microtubule-MAP Assembly Visualized at 8 Å Resolution. *J. Cel Biol* 191, 463–470. doi:10.1083/jcb.201007081

Galkin, V. E., Kolappan, S., Ng, D., Zong, Z., Li, J., Yu, X., et al. (2013). The Structure of the CS1 Pilus of Enterotoxigenic *Escherichia coli* Reveals Structural Polymorphism. *J. Bacteriol.* 195, 1360–1370. doi:10.1128/jb.01989-12

Garnett, J. A., Martinez-Santos, V. I., Saldana, Z., Pape, T., Hawthorne, W., Chan, J., et al. (2012). Structural Insights into the Biogenesis and Biofilm Formation by the *Escherichia coli* Common Pilus. *Proc. Natl. Acad. Sci.* 109, 3950–3955. doi:10.1073/pnas.1106733109

Geibel, S., Procko, E., Hultgren, S. J., Baker, D., and Waksman, G. (2013). Structural and Energetic Basis of Folded-Protein Transport by the FimD Usher. *Nature* 496, 243–246. doi:10.1038/nature12007

Ghosal, D., Kim, K. W., Zheng, H., Kaplan, M., Truchan, H. K., Lopez, A. E., et al. (2019). *In Vivo* structure of the Legionella Type II Secretion System by Electron Cryotomography. *Nat. Microbiol.* 4, 2101–2108. doi:10.1038/s41564-019-0603-6

Gold, V. A., Salzer, R., Averhoff, B., and Kühlbrandt, W. (2015). Structure of a Type IV Pilus Machinery in the Open and Closed State. *Elife* 4, e07380. doi:10.7554/eLife.07380

Goodson, H. V., and Jonasson, E. M. (2018). Microtubules and Microtubule-Associated Proteins. *Cold Spring Harb Perspect. Biol.* 10, a022608. doi:10.1101/cshperspect.a022608

Greenan, G. A., Vale, R. D., and Agard, D. A. (2020). Electron Cryotomography of Intact Motile Cilia Defines the Basal Body to Axoneme Transition. *J. Cel Biol* 219, e201907060. doi:10.1083/jcb.201907060

Grotjahn, D. A., Chowdhury, S., Xu, Y., McKenney, R. J., Schroer, T. A., and Lander, G. C. (2018). Cryo-electron Tomography Reveals that Dynactin Recruits a Team of Dyneins for Processive Motility. *Nat. Struct. Mol. Biol.* 25, 203–207. doi:10.1038/s41594-018-0027-7

Gu, S., Shevchik, V. E., Shaw, R., Pickersgill, R. W., and Garnett, J. A. (2017). The Role of Intrinsic Disorder and Dynamics in the Assembly and Function of the Type II Secretion System. *Biochim. Biophys. Acta (Bba) - Proteins Proteomics* 1865, 1255–1266. doi:10.1016/j.bbapap.2017.07.006

Gu, Y., Srikanth, V., Salazar-Morales, A. I., Jain, R., O'Brien, J. P., Yi, S. M., et al. (2021). Structure of Geobacter Pili Reveals Secretory rather Than Nanowire Behaviour. *Nature* 597, 430–434. doi:10.1038/s41586-021-03857-w

Gudimchuk, N. B., Ulyanov, E. V., O'Toole, E., Page, C. L., Vinogradov, D. S., Morgan, G., et al. (2020). Mechanisms of Microtubule Dynamics and Force Generation Examined with Computational Modeling and Electron Cryotomography. *Nat. Commun.* 11, 3765. doi:10.1038/s41467-020-17553-2

Guesdon, A., Bazile, F., Buey, R. M., Mohan, R., Monier, S., García, R. R., et al. (2016). EB1 Interacts with Outwardly Curved and Straight Regions of the Microtubule Lattice. *Nat. Cel Biol* 18, 1102–1108. doi:10.1038/ncb3412

Gui, M., Farley, H., Anujan, P., Anderson, J. R., Maxwell, D. W., Whitchurch, J. B., et al. (2021). De Novo identification of Mammalian Ciliary Motility Proteins Using Cryo-EM. *Cell* 184, 5791–5806. doi:10.1016/j.cell.2021.10.007

Habenstein, B., Loquet, A., Hwang, S., Giller, K., Vasa, S. K., Becker, S., et al. (2015). Hybrid Structure of the Type 1 Pilus of Uropathogenic *Escherichia coli*. *Angew. Chem. Int. Ed.* 54, 11691–11695. doi:10.1002/anie.201505065

Herzik, M. A., Jr., Wu, M., and Lander, G. C. (2019). High-resolution Structure Determination of Sub-100 kDa Complexes Using Conventional Cryo-EM. *Nat. Commun.* 10, 1032. doi:10.1038/s41467-019-08991-8

Hoenger, A., Sack, S., Thormählen, M., Marx, A., Müller, J., Gross, H., et al. (1998). Image Reconstructions of Microtubules Decorated with Monomeric and Dimeric Kinesins: Comparison with X-ray Structure and Implications for Motility. *J. Cel Biol* 141, 419–430. doi:10.1083/jcb.141.2.419

Hospenthal, M. K., Costa, T. R. D., and Waksman, G. (2017a). A Comprehensive Guide to Pilus Biogenesis in Gram-Negative Bacteria. *Nat. Rev. Microbiol.* 15, 365–379. doi:10.1038/nrmicro.2017.40

Hospenthal, M. K., Redzej, A., Dodson, K., Ukleja, M., Frenz, B., Rodrigues, C., et al. (2016). Structure of a Chaperone-Usher Pilus Reveals the Molecular Basis of Rod Uncoiling. *Cell* 164, 269–278. doi:10.1016/j.cell.2015.11.049

Hospenthal, M. K., Zyla, D., Costa, T. R. D., Redzej, A., Giese, C., Lillington, J., et al. (2017b). The Cryoelectron Microscopy Structure of the Type 1 Chaperone-Usher Pilus Rod. *Structure* 25, 1829–1838. doi:10.1016/j.str.2017.10.004

Howes, S. C., Geyer, E. A., LaFrance, B., Zhang, R., Kellogg, E. H., Westermann, S., et al. (2018). Structural and Functional Differences between Porcine Brain and Budding Yeast Microtubules. *Cell Cycle* 17, 278–287. doi:10.1080/15384101.2017.1415680

Hu, L. I., Yin, S., Ozer, E. A., Sewell, L., Rehman, S., Garnett, J. A., et al. (2020). Discovery of a New Neisseria Gonorrhoeae Type IV Pilus Assembly Factor, TfpC. *mBio* 11, e02528–20. doi:10.1128/mBio.02528-20

Hylton, R. K., and Swulius, M. T. (2021). Challenges and Triumphs in Cryo-Electron Tomography. *iScience* 24, 102959. doi:10.1016/j.isci.2021.102959

Itakura, K., Hirose, T., Crea, R., Riggs, A. D., Heyneker, H. L., Bolivar, F., et al. (1977). Expression in *Escherichia coli* of a Chemically Synthesized Gene for the Hormone Somatostatin. *Science* 198, 1056–1063. doi:10.1126/science.412251

Jacobsen, T., Bardiaux, B., Francetic, O., Izadi-Pruneyre, N., and Nilges, M. (2020). Structure and Function of Minor Pilins of Type IV Pili. *Med. Microbiol. Immunol.* 209, 301–308. doi:10.1007/s00430-019-00642-5

Janke, C. (2014). The Tubulin Code: Molecular Components, Readout Mechanisms, and Functions. *J. Cel Biol* 206, 461–472. doi:10.1083/jcb.201406055

Jordan, M. A., Diener, D. R., Stepanek, L., and Pigino, G. (2018). The Cryo-EM Structure of Intraflagellar Transport Trains Reveals How Dynein Is Inactivated to Ensure Unidirectional Anterograde Movement in Cilia. *Nat. Cel Biol* 20, 1250–1255. doi:10.1038/s41556-018-0213-1

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* 596, 583–589. doi:10.1038/s41586-021-03819-2

Kadavath, H., Cabrales Fontela, Y., Jaremko, M., Jaremko, Ł., Overkamp, K., Biernat, J., et al. (2018). The Binding Mode of a Tau Peptide with Tubulin. *Angew. Chem. Int. Ed.* 57, 3246–3250. doi:10.1002/anie.201712089

Kellogg, D. S., Jr., Peacock, W. L., Jr., Deacon, W. E., Brown, L., and Pirkle, C. I. (1963). Neisseria Gonorrhoeae I. *J. Bacteriol.* 85, 1274–1279. doi:10.1128/jb.85.6.1274-1279.1963

Kellogg, E. H., Hejab, N. M. A., Howes, S., Northcote, P., Miller, J. H., Díaz, J. F., et al. (2017). Insights into the Distinct Mechanisms of Action of Taxane and Non-taxane Microtubule Stabilizers from Cryo-EM Structures. *J. Mol. Biol.* 429, 633–646. doi:10.1016/j.jmb.2017.01.001

Kellogg, E. H., Hejab, N. M. A., Poepsel, S., Downing, K. H., DiMaio, F., and Nogales, E. (2018). Near-atomic Model of Microtubule-Tau Interactions. *Science* 360, 1242–1246. doi:10.1126/science.aat1780

Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R. G., Wyckoff, H., and Phillips, D. C. (1958). A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-ray Analysis. *Nature* 181, 662–666. doi:10.1038/181662a0

Kesten, C., Wallmann, A., Schneider, R., McFarlane, H. E., Diehl, A., Khan, G. A., et al. (2019). The Companion of Cellulose Synthase 1 Confers Salt Tolerance through a Tau-like Mechanism in Plants. *Nat. Commun.* 10, 857. doi:10.1038/s41467-019-08780-3

Ketchem, R. R., Lee, K. C., Huo, S., and Cross, T. A. (1996). Macromolecular Structural Elucidation with Solid-State NMR-Derived Orientational Constraints. *J. Biomol. NMR* 8, 1–14. doi:10.1007/BF00198135

Khalifa, A. A. Z., Ichikawa, M., Dai, D., Kubo, S., Black, C. S., Peri, K., et al. (2020). The Inner junction Complex of the Cilia Is an Interaction Hub that Involves Tubulin post-translational Modifications. *Elife* 9, e52760. doi:10.7554/eLife.52760

Kiesel, P., Alvarez Viar, G., Tsoy, N., Maraspini, R., Gorilak, P., Varga, V., et al. (2020). The Molecular Structure of Mammalian Primary Cilia Revealed by Cryo-Electron Tomography. *Nat. Struct. Mol. Biol.* 27, 1115–1124. doi:10.1038/s41594-020-0507-4

Kikkawa, M., Okada, Y., and Hirokawa, N. (2000). 15 Å Resolution Model of the Monomeric Kinesin Motor, KIF1A. *Cell* 100, 241–252. doi:10.1016/s0092-8674(00)81562-7

Kikkawa, M., Sablin, E. P., Okada, Y., Yajima, H., Fletterick, R. J., and Hirokawa, N. (2001). Switch-based Mechanism of Kinesin Motors. *Nature* 411, 439–445. doi:10.1038/35078000

Kolappan, S., Coureuil, M., Yu, X., Nassif, X., Egelman, E. H., and Craig, L. (2016). Structure of the Neisseria Meningitidis Type IV Pilus. *Nat. Commun.* 7, 13015. doi:10.1038/ncomms13015

Kühlbrandt, W. (2014). The Resolution Revolution. *Science* 343, 1443–1444. doi:10.1126/science.1251652

Kumar, A., Heise, H., Blommers, M. J. J., Krastel, P., Schmitt, E., Petersen, F., et al. (2010). Interaction of Epothilone B (Patupilone) with Microtubules as Detected by Two-Dimensional Solid-State NMR Spectroscopy. *Angew. Chem. Int. Edition* 49, 7504–7507. doi:10.1002/anie.201001946

Lacey, S. E., He, S., Scheres, S. H., and Carter, A. P. (2019). Cryo-EM of Dynein Microtubule-Binding Domains Shows How an Axonemal Dynein Distorts the Microtubule. *Elife* 8, e47145. doi:10.7554/eLife.47145

Lepault, J., Booy, F. P., and Dubochet, J. (1983). Electron Microscopy of Frozen Biological Suspensions. *J. Microsc.* 129, 89–102. doi:10.1111/j.1365-2818.1983.tb04163.x

Li, F., Li, Y., Ye, X., Gao, H., Shi, Z., Luo, X., et al. (2020). Cryo-EM Structure of VASH1-SVBP Bound to Microtubules. *Elife* 9, e58157. doi:10.7554/eLife.58157

Li, H., DeRosier, D. J., Nicholson, W. V., Nogales, E., and Downing, K. H. (2002). Microtubule Structure at 8 Å Resolution. *Structure* 10, 1317–1328. doi:10.1016/s0969-2126(02)00827-4

Li, X., Mooney, P., Zheng, S., Booth, C. R., Braunfeld, M. B., Gubbens, S., et al. (2013). Electron Counting and Beam-Induced Motion Correction Enable Near-Atomic-Resolution Single-Particle Cryo-EM. *Nat. Methods* 10, 584–590. doi:10.1038/nmeth.2472

Liu, D., Liu, X., Shang, Z., and Sindelar, C. V. (2017). Structural Basis of Cooperativity in Kinesin Revealed by 3D Reconstruction of a Two-Head-Bound State on Microtubules. *Elife* 6, e24490. doi:10.7554/eLife.24490

López-Castilla, A., Thomassin, J.-L., Bardiaux, B., Zheng, W., Nivaskumar, M., Yu, X., et al. (2017). Structure of the Calcium-dependent Type 2 Secretion Pseudopilus. *Nat. Microbiol.* 2, 1686–1695. doi:10.1038/s41564-017-0041-2

Loquet, A., Sgourakis, N. G., Gupta, R., Giller, K., Riedel, D., Goosmann, C., et al. (2012). Atomic Model of the Type III Secretion System Needle. *Nature* 486, 276–279. doi:10.1038/nature11079

Löwe, J., Li, H., Downing, K. H., and Nogales, E. (2001). Refined Structure of αβ-tubulin at 3.5 Å Resolution 1 1Edited by I. A. Wilson. *J. Mol. Biol.* 313, 1045–1057. doi:10.1006/jmbi.2001.5077

Lukaszczyk, M., Pradhan, B., and Remaut, H. (2019). "The Biosynthesis and Structures of Bacterial Pili," in *Bacterial Cell Walls and Membranes*. Editor A. Kuhn (Cham: Springer International Publishing), 369–413. doi:10.1007/978-3-030-18768-2_12

Luo, Y., Xiang, S., Hooikaas, P. J., van Bezouwen, L., Jijumon, A. S., Janke, C., et al. (2020). Direct Observation of Dynamic Protein Interactions Involving Human Microtubules Using Solid-State NMR Spectroscopy. *Nat. Commun.* 11, 18. doi:10.1038/s41467-019-13876-x

Luo, Y., Xiang, S., Paioni, A. L., Adler, A., Hooikaas, P. J., Jijumon, A. S., et al. (2021). Solid-State NMR Spectroscopy for Studying Microtubules and Microtubule-Associated Proteins. *Methods Mol. Biol.* 2305, 193–201. doi:10.1007/978-1-0716-1406-8_10

Ma, M., Stoyanova, M., Rademacher, G., Dutcher, S. K., Brown, A., and Zhang, R. (2019). Structure of the Decorated Ciliary Doublet Microtubule. *Cell* 179, 909–922. doi:10.1016/j.cell.2019.09.030

Mandelkow, E.-M., and Mandelkow, E. (1985). Unstained Microtubules Studied by Cryo-Electron Microscopy. *J. Mol. Biol.* 181, 123–135. doi:10.1016/0022-2836(85)90330-4

Manka, S. W., and Moores, C. A. (2020). Pseudo-repeats in Doublecortin Make Distinct Mechanistic Contributions to Microtubule Regulation. *EMBO Rep.* 21, e51534. doi:10.15252/embr.202051534

Manka, S. W., and Moores, C. A. (2018). The Role of Tubulin-Tubulin Lattice Contacts in the Mechanism of Microtubule Dynamic Instability. *Nat. Struct. Mol. Biol.* 25, 607–615. doi:10.1038/s41594-018-0087-8

Maurer, S. P., Fourniol, F. J., Bohner, G., Moores, C. A., and Surrey, T. (2012). EBs Recognize a Nucleotide-dependent Structural Cap at Growing Microtubule Ends. *Cell* 149, 371–382. doi:10.1016/j.cell.2012.02.049

McDowall, A. W., Chang, J.-J., Freeman, R., Lepault, J., Walter, C. A., and Dubochet, J. (1983). Electron Microscopy of Frozen Hydrated Sections of Vitreous Ice and Vitrified Biological Samples. *J. Microsc.* 131, 1–9. doi:10.1111/j.1365-2818.1983.tb04225.x

McIntosh, J. R., O'Toole, E., Morgan, G., Austin, J., Ulyanov, E., Ataullakhanov, F., et al. (2018). Microtubules Grow by the Addition of Bent Guanosine Triphosphate Tubulin to the Tips of Curved Protofilaments. *J. Cel Biol* 217, 2691–2708. doi:10.1083/jcb.201802138

Metoz, F., Arnal, I., and Wade, R. H. (1997). Tomography without Tilt: Three-Dimensional Imaging of Microtubule/motor Complexes. *J. Struct. Biol.* 118, 159–168. doi:10.1006/jsbi.1997.3850

Miller, E., Garcia, T., Hultgren, S., and Oberhauser, A. F. (2006). The Mechanical Properties of *E. coli* Type 1 Pili Measured by Atomic Force Microscopy Techniques. *Biophysical J.* 91, 3848–3856. doi:10.1529/biophysj.106.088989

Morag, O., Sgourakis, N. G., Baker, D., and Goldbourt, A. (2015). The NMR-Rosetta Capsid Model of M13 Bacteriophage Reveals a Quadrupled Hydrophobic Packing Epitope. *Proc. Natl. Acad. Sci. USA* 112, 971–976. doi:10.1073/pnas.1415393112

Nakane, T., Kotecha, A., Sente, A., McMullan, G., Masiulis, S., Brown, P. M. G. E., et al. (2020). Single-particle Cryo-EM at Atomic Resolution. *Nature* 587, 152–156. doi:10.1038/s41586-020-2829-0

Naskar, S., Hohl, M., Tassinari, M., and Low, H. H. (2021). The Structure and Mechanism of the Bacterial Type II Secretion System. *Mol. Microbiol.* 115, 412–424. doi:10.1111/mmi.14664

Neuhaus, A., Selvaraj, M., Salzer, R., Langer, J. D., Kruse, K., Kirchner, L., et al. (2020). Cryo-electron Microscopy Reveals Two Distinct Type IV Pili Assembled by the Same Bacterium. *Nat. Commun.* 11, 2231. doi:10.1038/s41467-020-15650-w

Nogales, E., Grayer Wolf, S., Khan, I. A., Ludueña, R. F., and Downing, K. H. (1995). Structure of Tubulin at 6.5 Å and Location of the Taxol-Binding Site. *Nature* 375, 424–427. doi:10.1038/375424a0

Nogales, E., Whittaker, M., Milligan, R. A., and Downing, K. H. (1999). High-resolution Model of the Microtubule. *Cell* 96, 79–88. doi:10.1016/s0092-8674(00)80961-7

Nogales, E., Wolf, S. G., and Downing, K. H. (1998). Structure of the αβ Tubulin Dimer by Electron Crystallography. *Nature* 391, 199–203. doi:10.1038/34465

Omattage, N. S., Deng, Z., Pinkner, J. S., Dodson, K. W., Almqvist, F., Yuan, P., et al. (2018). Structural Basis for Usher Activation and Intramolecular Subunit Transfer in P Pilus Biogenesis in *Escherichia coli*. *Nat. Microbiol.* 3, 1362–1368. doi:10.1038/s41564-018-0255-y

Opella, S. J., Marassi, F. M., Gesell, J. J., Valente, A. P., Kim, Y., Oblatt-Montal, M., et al. (1999). Structures of the M2 Channel-Lining Segments from Nicotinic Acetylcholine and NMDA Receptors by NMR Spectroscopy. *Nat. Struct. Biol.* 6, 374–379. doi:10.1038/7610

Owa, M., Uchihashi, T., Yanagisawa, H.-a., Yamano, T., Iguchi, H., Fukuzawa, H., et al. (2019). Inner Lumen Proteins Stabilize Doublet Microtubules in Cilia and Flagella. *Nat. Commun.* 10, 1143. doi:10.1038/s41467-019-09051-x

Parge, H. E., Forest, K. T., Hickey, M. J., Christensen, D. A., Getzoff, E. D., and Tainer, J. A. (1995). Structure of the Fibre-Forming Protein Pilin at 2.6 Å Resolution. *Nature* 378, 32–38. doi:10.1038/378032a0

Paul, D. M., Mantell, J., Borucu, U., Coombs, J., Surridge, K. J., Squire, J. M., et al. (2020). *In Situ* cryo-electron Tomography Reveals Filamentous Actin within the Microtubule Lumen. *J. Cel Biol* 219. doi:10.1083/jcb.201911154

Pauling, L., and Corey, R. B. (1951a). The Structure of Fibrous Proteins of the Collagen-Gelatin Group. *Proc. Natl. Acad. Sci.* 37, 272–281. doi:10.1073/pnas.37.5.272

Pauling, L., and Corey, R. B. (1951b). The Structure of Hair, Muscle, and Related Proteins. *Proc. Natl. Acad. Sci.* 37, 261–271. doi:10.1073/pnas.37.5.261

Pelicic, V. (2008). Type IV pili: e pluribus unum? *Mol. Microbiol.* 68, 827–837. doi:10.1111/j.1365-2958.2008.06197.x

Phan, G., Remaut, H., Wang, T., Allen, W. J., Pirker, K. F., Lebedev, A., et al. (2011). Crystal Structure of the FimD Usher Bound to its Cognate FimC-FimH Substrate. *Nature* 474, 49–53. doi:10.1038/nature10109

Pierson, G. B., Burton, P. R., and Himes, R. H. (1978). Alterations in Number of Protofilaments in Microtubules Assembled *In Vitro*. *J. Cel Biol* 76, 223–228. doi:10.1083/jcb.76.1.223

Pradhan, B., Liedtke, J., Sleutel, M., Lindbäck, T., Zegeye, E. D., Sullivan, O, et al. (2021). Endospore Appendages: a Novel Pilus Superfamily from the Endospores of Pathogenic Bacilli. *EMBO J.* 40, e106887. doi:10.15252/embj.2020106887

Pyle, E., and Zanetti, G. (2021). Current Data Processing Strategies for Cryo-Electron Tomography and Subtomogram Averaging. *Biochem. J.* 478, 1827–1845. doi:10.1042/bcj20200715

Quillin, S. J., and Seifert, H. S. (2018). Neisseria Gonorrhoeae Host Adaptation and Pathogenesis. *Nat. Rev. Microbiol.* 16, 226–240. doi:10.1038/nrmicro.2017.169

Rao, Q., Han, L., Wang, Y., Chai, P., Kuo, Y.-w., Yang, R., et al. (2021). Structures of Outer-Arm Dynein Array on Microtubule Doublet Reveal a Motor Coordination Mechanism. *Nat. Struct. Mol. Biol.* 28, 799–810. doi:10.1038/s41594-021-00656-9

Redwine, W. B., Hernández-López, R., Zou, S., Huang, J., Reck-Peterson, S. L., and Leschziner, A. E. (2012). Structural Basis for Microtubule Binding and Release by Dynein. *Science* 337, 1532–1536. doi:10.1126/science.1224151

Roll-Mecak, A. (2020). The Tubulin Code in Microtubule Dynamics and Information Encoding. *Dev. Cel* 54, 7–20. doi:10.1016/j.devcel.2020.06.008

Salih, O., Remaut, H., Waksman, G., and Orlova, E. V. (2008). Structural Analysis of the Saf Pilus by Electron Microscopy and Image Processing. *J. Mol. Biol.* 379, 174–187. doi:10.1016/j.jmb.2008.03.056

Schur, F. K. (2019). Toward High-Resolution *In Situ* Structural Biology with Cryo-Electron Tomography and Subtomogram Averaging. *Curr. Opin. Struct. Biol.* 58, 1–9. doi:10.1016/j.sbi.2019.03.018

Sgourakis, N. G., Yau, W.-M., and Qiang, W. (2015). Modeling an In-Register, Parallel "Iowa" Aβ Fibril Structure Using Solid-State NMR Data from Labeled Samples with Rosetta. *Structure* 23, 216–227. doi:10.1016/j.str.2014.10.022

Shibata, S., Shoji, M., Okada, K., Matsunami, H., Matthews, M. M., Imada, K., et al. (2020). Structure of Polymerized Type V Pilin Reveals Assembly Mechanism Involving Protease-Mediated Strand Exchange. *Nat. Microbiol.* 5, 830–837. doi:10.1038/s41564-020-0705-1

Sindelar, C. V., and Downing, K. H. (2010). An Atomic-Level Mechanism for Activation of the Kinesin Molecular Motors. *Proc. Natl. Acad. Sci.* 107, 4111–4116. doi:10.1073/pnas.0911208107

Sindelar, C. V., and Downing, K. H. (2007). The Beginning of Kinesin's Force-Generating Cycle Visualized at 9-A˚ Resolution. *J. Cel Biol* 177, 377–385. doi:10.1083/jcb.200612090

Smith, G. E., Summers, M. D., and Fraser, M. J. (1983). Production of Human Beta Interferon in Insect Cells Infected with a Baculovirus Expression Vector. *Mol. Cel Biol* 3, 2156–2165. doi:10.1128/mcb.3.12.2156-2165.1983

Sosa, H., Dias, D. P., Hoenger, A., Whittaker, M., Wilson-Kubalek, E., Sablin, E., et al. (1997a). A Model for the Microtubule-Ncd Motor Protein Complex Obtained by Cryo-Electron Microscopy and Image Analysis. *Cell* 90, 217–224. doi:10.1016/s0092-8674(00)80330-x

Sosa, H., Hoenger, A., and Milligan, R. A. (1997b). Three Different Approaches for Calculating the Three-Dimensional Structure of Microtubules Decorated with Kinesin Motor Domains. *J. Struct. Biol.* 118, 149–158. doi:10.1006/jsbi.1997.3851

Spaulding, C. N., Schreiber, H. L., Zheng, W., Dodson, K. W., Hazen, J. E., Conover, M. S., et al. (2018). Functional Role of the Type 1 Pilus Rod Structure in Mediating Host-Pathogen Interactions. *Elife* 7. doi:10.7554/eLife.31662

Sui, H., and Downing, K. H. (2010). Structural Basis of Interprotofilament Interaction and Lateral Deformation of Microtubules. *Structure* 18, 1022–1031. doi:10.1016/j.str.2010.05.010

Swanson, J., Robbins, K., Barrera, O., Corwin, D., Boslego, J., Ciak, J., et al. (1987). Gonococcal Pilin Variants in Experimental Gonorrhea. *J. Exp. Med.* 165, 1344–1357. doi:10.1084/jem.165.5.1344

Tan, Y. Z., Cheng, A., Potter, C. S., and Carragher, B. (2016). Automated Data Collection in Single Particle Electron Microscopy. *Microscopy (Tokyo)* 65, 43–56. doi:10.1093/jmicro/dfv369

Thanassi, D. G., Bliska, J. B., and Christie, P. J. (2012). Surface Organelles Assembled by Secretion Systems of Gram-Negative Bacteria: Diversity in Structure and Function. *FEMS Microbiol. Rev.* 36, 1046–1082. doi:10.1111/j.1574-6976.2012.00342.x

Tilney, L. G., Bryan, J., Bush, D. J., Fujiwara, K., Mooseker, M. S., Murphy, D. B., et al. (1973). Microtubules: Evidence for 13 Protofilaments. *J. Cel Biol* 59, 267–275. doi:10.1083/jcb.59.2.267

Turk, M., and Baumeister, W. (2020). The Promise and the Challenges of Cryo-electron Tomography. *FEBS Lett.* 594, 3243–3261. doi:10.1002/1873-3468.13948

Vale, R. D. (2003). The Molecular Motor Toolbox for Intracellular Transport. *Cell* 112, 467–480. doi:10.1016/s0092-8674(03)00111-9

Van Gerven, N., Van der Verren, S. E., Reiter, D. M., and Remaut, H. (2018). The Role of Functional Amyloids in Bacterial Virulence. *J. Mol. Biol.* 430, 3657–3684. doi:10.1016/j.jmb.2018.07.010

Vemu, A., Atherton, J., Spector, J. O., Moores, C. A., and Roll-Mecak, A. (2017). Tubulin Isoform Composition Tunes Microtubule Dynamics. *MBoC* 28, 3564–3572. doi:10.1091/mbc.e17-02-0124

Vemu, A., Atherton, J., Spector, J. O., Szyk, A., Moores, C. A., and Roll-Mecak, A. (2016). Structure and Dynamics of Single-Isoform Recombinant Neuronal Human Tubulin. *J. Biol. Chem.* 291, 12907–12915. doi:10.1074/jbc.c116.731133

Vinothkumar, K. R., and Henderson, R. (2010). Structures of Membrane Proteins. *Quart. Rev. Biophys.* 43, 65–158. doi:10.1017/s0033583510000041

Wade, R. H., Chrétien, D., and Job, D. (1990). Characterization of Microtubule Protofilament Numbers. *J. Mol. Biol.* 212, 775–786. doi:10.1016/0022-2836(90)90236-f

Walton, T., Wu, H., and Brown, A. (2021). Structure of a Microtubule-Bound Axonemal Dynein. *Nat. Commun.* 12, 477. doi:10.1038/s41467-020-20735-7

Wang, F., Coureuil, M., Osinski, T., Orlova, A., Altindal, T., Gesbert, G., et al. (2017). Cryoelectron Microscopy Reconstructions of the *Pseudomonas aeruginosa* and Neisseria Gonorrhoeae Type IV Pili at Sub-nanometer Resolution. *Structure* 25, 1423–1435. doi:10.1016/j.str.2017.07.016

Wang, F., Gu, Y., O'Brien, J. P., Yi, S. M., Yalcin, S. E., Srikanth, V., et al. (2019). Structure of Microbial Nanowires Reveals Stacked Hemes that Transport Electrons over Micrometers. *Cell* 177, 361–369. doi:10.1016/j.cell.2019.03.029

Wang, X., Fu, Y., Beatty, W. L., Ma, M., Brown, A., Sibley, L. D., et al. (2021). Cryo-EM Structure of Cortical Microtubules from Human Parasite Toxoplasma Gondii Identifies Their Microtubule Inner Proteins. *Nat. Commun.* 12, 3065. doi:10.1038/s41467-021-23351-1

Watanabe, R., Buschauer, R., Böhning, J., Audagnotto, M., Lasker, K., Lu, T.-W., et al. (2020). The *In Situ* Structure of Parkinson's Disease-Linked LRRK2. *Cell* 182, 1508–1518. doi:10.1016/j.cell.2020.08.004

Williamson, M. P., Havel, T. F., and Wüthrich, K. (1985). Solution Conformation of Proteinase Inhibitor IIA from Bull Seminal Plasma by 1H Nuclear Magnetic Resonance and Distance Geometry. *J. Mol. Biol.* 182, 295–315. doi:10.1016/0022-2836(85)90347-x

Xu, Q., Shoji, M., Shibata, S., Naito, M., Sato, K., Elsliger, M.-A., et al. (2016). A Distinct Type of Pilus from the Human Microbiome. *Cell* 165, 690–703. doi:10.1016/j.cell.2016.03.016

Yan, S., Guo, C., Hou, G., Zhang, H., Lu, X., Williams, J. C., et al. (2015). Atomic-resolution Structure of the CAP-Gly Domain of Dynactin on Polymeric Microtubules Determined by Magic Angle Spinning NMR Spectroscopy. *Proc. Natl. Acad. Sci. USA* 112, 14611–14616. doi:10.1073/pnas.1509852112

Yip, K. M., Fischer, N., Paknia, E., Chari, A., and Stark, H. (2020). Atomic-resolution Protein Structure Determination by Cryo-EM. *Nature* 587, 157–161. doi:10.1038/s41586-020-2833-4

Zabeo, D., Heumann, J. M., Schwartz, C. L., Suzuki-Shinjo, A., Morgan, G., Widlund, P. O., et al. (2018). A Lumenal Interrupted helix in Human Sperm Tail Microtubules. *Sci. Rep.* 8, 2727. doi:10.1038/s41598-018-21165-8

Zeri, A. C., Mesleh, M. F., Nevzorov, A. A., and Opella, S. J. (2003). Structure of the Coat Protein in Fd Filamentous Bacteriophage Particles Determined by Solid-State NMR Spectroscopy. *Proc. Natl. Acad. Sci.* 100, 6458–6463. doi:10.1073/pnas.1132059100

Zhang, R., Roostalu, J., Surrey, T., and Nogales, E. (2017). Structural Insight into TPX2-Stimulated Microtubule Assembly. *Elife* 6. doi:10.7554/eLife.30959

Zhang, R., Alushin, G. M., Brown, A., and Nogales, E. (2015). Mechanistic Origin of Microtubule Dynamic Instability and its Modulation by EB Proteins. *Cell* 162, 849–859. doi:10.1016/j.cell.2015.07.012

Zhang, R., LaFrance, B., and Nogales, E. (2018). Separating the Effects of Nucleotide and EB Binding on Microtubule Structure. *Proc. Natl. Acad. Sci. USA* 115, E6191–E6200. doi:10.1073/pnas.1802637115

Zhang, R., and Nogales, E. (2015). A New Protocol to Accurately Determine Microtubule Lattice Seam Location. *J. Struct. Biol.* 192, 245–254. doi:10.1016/j.jsb.2015.09.015

Zheng, W., Andersson, M., Mortezaei, N., Bullitt, E., and Egelman, E. (2019). Cryo-EM Structure of the CFA/I Pilus Rod. *Int. Union Crystallogr. J.* 6, 815–821. doi:10.1107/s2052252519007966

Zheng, W., Pena, A., Low, W. W., Wong, J. L. C., Frankel, G., and Egelman, E. H. (2020). Cryoelectron-Microscopic Structure of the pKpQIL Conjugative Pili from Carbapenem-Resistant *Klebsiella pneumoniae*. *Structure* 28, 1321–1328. doi:10.1016/j.str.2020.08.010

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.

The handling editor declared a shared affiliation with the authors at time of review.

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

frontiers in Molecular Biosciences

REVIEW
published: 25 January 2022
doi: 10.3389/fmolb.2021.823253

Structural Insights Into Tautomeric Dynamics in Nucleic Acids and in Antiviral Nucleoside Analogs

Bogdan I. Fedeles[1], *Deyu Li*[2] *and Vipender Singh*[3]*

[1]Departments of Chemistry and Biological Engineering and Center for Environmental Health Sciences, Massachusetts Institute of Technology, Cambridge, MA, United States, [2]Department of Pharmaceutical Sciences, College of Pharmacy, University of Rhode Island, Kingston, RI, United States, [3]Department of Biochemistry and Biophysics, Novartis Institute of Biomedical Research, Cambridge, MA, United States

DNA (2′-deoxyribonucleic acid) and RNA (ribonucleic acid) play diverse functional roles in biology and disease. Despite being comprised primarily of only four cognate nucleobases, nucleic acids can adopt complex three-dimensional structures, and RNA in particular, can catalyze biochemical reactions to regulate a wide variety of biological processes. Such chemical versatility is due in part to the phenomenon of nucleobase tautomerism, whereby the bases can adopt multiple, yet distinct isomeric forms, known as tautomers. For nucleobases, tautomers refer to structural isomers that differ from one another by the position of protons. By altering the position of protons on nucleobases, many of which play critical roles for hydrogen bonding and base pairing interactions, tautomerism has profound effects on the biochemical processes involving nucleic acids. For example, the transient formation of minor tautomers during replication could generate spontaneous mutations. These mutations could arise from the stabilization of mismatches, in the active site of polymerases, in conformations involving minor tautomers that are indistinguishable from canonical base pairs. In this review, we discuss the evidence for tautomerism in DNA, and its consequences to the fidelity of DNA replication. Also reviewed are RNA systems, such as the riboswitches and self-cleaving ribozymes, in which tautomerism plays a functional role in ligand recognition and catalysis, respectively. We also discuss tautomeric nucleoside analogs that are efficacious as antiviral drug candidates such as molnupiravir for coronaviruses and KP1212 for HIV. The antiviral efficacy of these analogs is due, in part, to their ability to exist in multiple tautomeric forms and induce mutations in the replicating viral genomes. From a technical standpoint, minor tautomers of nucleobases are challenging to identify directly because they are rare and interconvert on a fast, millisecond to nanosecond, time scale. Nevertheless, many approaches including biochemical, structural, computational and spectroscopic methods have been developed to study tautomeric dynamics in RNA and DNA systems, and in antiviral nucleoside analogs. An overview of these methods and their applications is included here.

Keywords: Tautomerism, ribozymes, riboswitches, mutagenesis, COVID-19, antivirals, nucleoside analogs therapy, spontaneous mutations

OPEN ACCESS

Edited by:
Annalisa Pastore,
King's College London,
United Kingdom

Reviewed by:
Pablo D. Dans,
Institut Pasteur de Montevideo,
Uruguay
Pierre Barraud,
UMR8261 Expression Génétique
Microbienne, France

***Correspondence:**
Vipender Singh
vipender02118@gmail.com

Specialty section:
This article was submitted to
Structural Biology,
a section of the journal
Frontiers in Molecular Biosciences

Received: *27 November 2021*
Accepted: *30 December 2021*
Published: *25 January 2022*

Citation:
Fedeles BI, Li D and Singh V (2022)
Structural Insights Into Tautomeric
Dynamics in Nucleic Acids and in
Antiviral Nucleoside Analogs.
Front. Mol. Biosci. 8:823253.
doi: 10.3389/fmolb.2021.823253

INTRODUCTION

Nucleic acid bases exhibit keto-enol and amino-imino prototropic tautomerism due to the presence of multiple solvent-exchangeable protons (**Figure 1**) (Watson and Crick, 1953; Topal and Fresco, 1976; Brown et al., 1989; Colominas et al., 1996; Mons et al., 2002). The formation of minor tautomers can increase the overall structural and chemical diversity of nucleic acids, which enables their diverse functions in biology (Topal and Fresco, 1976; Cochrane and Strobel, 2008a; Singh et al., 2015). For example, many self-cleaving ribozymes (RNA enzymes) and some riboswitches (RNA aptamers) are proposed to utilize tautomerism to perform their biological function (**Figure 2**, **Figure 3**) (Singh et al., 2015; Singh et al., 2014). Formation of minor tautomers in DNA, at low frequency, is proposed to contribute to the phenomenon of 'spontaneous mutagenesis', which denotes the background level of mutations that appear during the replication of undamaged DNA (Watson and Crick, 1953; Topal and Fresco, 1976; Wang et al., 2011; Rangadurai et al., 2020). These mutations are thought to arise due, in part, to the altered base pairing properties of minor tautomers (**Figure 4**) (Watson and Crick, 1953; Topal and Fresco, 1976; Wang et al., 2011).

Therapeutics based on tautomeric nucleoside analogs have proven effective as antiviral agents against a range of retro- and ribo-viruses, including influenza (Delang et al., 2018), hepatitis-C-virus (HCV) (Crotty et al., 2001), human immunodeficiency virus (HIV) (Li et al., 2014) and coronaviruses, including COVID-19 (**Figure 5**) (Shannon et al., 2020; Kabinger et al., 2021). The antiviral efficacy of these analogs stems from their ability to exist in multiple tautomeric or rotameric states, which help mutagenize the viral genomes to error catastrophe and even to viral population extinction. The formation of minor tautomers in nucleic acids is a rare event, and their dynamics of interconversion is fast, on a millisecond to nanosecond time scale (Peng et al., 2013; Rangadurai et al., 2019). Recent developments in methods have allowed direct identification and quantification of minor tautomers in nucleic acid bases and in nucleoside/nucleotide analogs (Peng et al., 2011; Li et al., 2014; Singh et al., 2014; Peng et al., 2015; Rangadurai et al., 2019).

Small self-cleaving ribozymes and many riboswitches are proposed to utilize minor tautomeric forms of catalytic residues and ligands to perform their biological functions (**Figure 2**, **Figure 3**) (Cochrane et al., 2007; Cochrane and Strobel, 2008a; Cochrane and Strobel, 2008b; Thore et al., 2008; Gilbert et al., 2009; Wilcox and Bevilacqua, 2013; Singh et al., 2014; Singh et al., 2015). Among riboswitches, crystallographic evidence shows that the purine and the thiamine pyrophosphate riboswitches recognize the minor tautomeric forms of their non-natural ligands xanthine and oxythiamine pyrophosphate (OxyTPP), respectively (**Figure 2**) (Thore et al., 2008; Gilbert et al., 2009; Singh et al., 2015; Singh et al., 2014). Self-cleaving ribozymes such as hammerhead, hairpin, glmS, varkud satellite (VS), hepatitis delta virus (HDV), and twister catalyze nucleolytic intramolecular self-scission reactions (**Figure 3**) (Rupert and Ferré-D'Amaré, 2001; Martick and Scott, 2006; Cochrane et al., 2007; Cochrane and Strobel, 2008a; Suslov et al., 2015). In these ribozymes, it is proposed that the tautomeric forms of their catalytic guanosines act as a general base to activate the 2′-hydroxyl nucleophile to initiate the cleavage reactions (Klein et al., 2007; Cochrane and Strobel, 2008a).

During DNA replication, formation of minor tautomeric forms of nucleobases could lead to mispairing that generates mutations in the absence of any DNA lesion or other form of DNA damage (Watson and Crick, 1953; Löwdin, 1963; Topal and Fresco, 1976; Wang et al., 2011; Bebenek et al., 2011; Rangadurai et al., 2019; Rangadurai et al., 2020). This phenomenon is thought to explain the appearance of spontaneous mutations during the replication of undamaged DNA. The transient generation of minor tautomers in DNA allows stabilization of mismatches, in the polymerase active site, in conformations that are indistinguishable from the structures of canonical base pairs (**Figure 4**) (Watson and Crick, 1953; Topal and Fresco, 1976; Wang et al., 2011; Rangadurai et al., 2020).

While assembling their double-helix model of DNA, James Watson and Francis Crick stumbled over the phenomenon of base tautomerism; in fact, their model required that the bases adopt specific tautomeric forms in order to base-pair, with the suggestion that alternate tautomers would lead to mispairing and spontaneous mutations. (Watson and Crick, 1953). These ideas were later formalized by Michael Topal and Jacques Fresco, who described in 1976 the specific pairings between the minor tautomeric forms of canonical nucleic acid bases during replication that are likely to lead to mutations (Topal and Fresco, 1976). Since then, structural studies have shown that the formation of minor tautomers in DNA could allow wobble mismatches, such as the A-C and G-T base pairs, to adopt a geometry similar to that of canonical base pairs in Watson and Crick (W-C) conformations (**Figure 4**) (Wang et al., 2011; Szymanski et al., 2017). These studies provided structural evidence in support of the rare tautomer hypothesis for spontaneous mutations (Wang et al., 2011). Furthermore, tautomeric dynamics occur on a time scale that is similar to that of DNA replication, suggesting that the rate of rare tautomer formation could determine the probability of base misincorporation during replication (Rangadurai et al., 2019; Rangadurai et al., 2020).

In this review, we will discuss mechanisms involving tautomerism in RNA biology and in the generation of spontaneous mutations during DNA replication. We will also discuss the role of tautomerism in the development of nucleoside/nucleotide analogs based antiviral therapeutics. Specifically, we will focus on tautomerism in the catalysis of small self-cleaving ribozymes and in ligands recognition by riboswitches. We will discuss evidence of tautomerism in DNA and the associated mechanisms that contribute to the generation of spontaneous mutations. We will also review the role of tautomerism in the mechanism by which antiviral drug candidates based on nucleoside analogs, such as molnupiravir (isopropyl ester prodrug of the ribonucleoside analog β-D-N^4-hydroxycytidine) and KP1212 (5-aza-5,6-dihydro-2′-deoxycytidine), induce lethal mutagenesis in the genomes of SARS-CoV-2 and HIV,

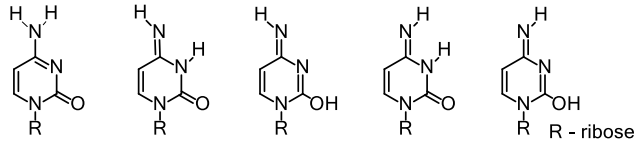

FIGURE 1 | Prototropic tautomerism in cytosine. Nucleic acid bases contain keto, amino, or both functional groups. These groups often participate in keto-enol and amino-imino tautomerism as shown here for the native base cytosine.

respectively. Minor tautomers are rare and their dynamics of interconversion is fast, which makes them challenging to identify in complex macromolecular systems, such as DNA and RNA. In the last section, we will summarize the current, state-of-the-art methods for studying tautomerism in nucleic acids and in antiviral nucleoside analogs, and discuss the future directions of the field.

Tautomerism in Nucleic Acids

Nucleic acids display keto-enol and amino-imino types of tautomerism (Singh et al., 2015), as shown in **Figure 1** using cytosine as an example. This type of tautomerism, called prototropic tautomerism (Laar, 1886; Antonov, 2013) involves the movement of protons in the form of deprotonation at one site and protonation at another site, which is accompanied by a rearrangement of the double bonds in between the two sites. The base pairing complementarity in nucleic acids depends on the position of protons on the Watson-Crick (W-C) interface (Watson and Crick, 1953; Wang et al., 2011). As the positions of protons change between different tautomers (**Figure 1**) (Singh et al., 2015), the distribution of hydrogen donor and acceptor sites is altered on the W-C face of each base. Therefore, certain minor tautomeric forms of canonical nucleic acid bases can engage in alternative base pairings (Watson and Crick, 1953; Topal and Fresco, 1976; Wang et al., 2011; Singh et al., 2015; Rangadurai et al., 2019; Rangadurai et al., 2020). Since base pairing interactions are predominantly responsible for the biological functions performed by nucleic acids, prototropic tautomerism is expected to influence the biology of both DNA and RNA (Singh et al., 2015). Below we discuss a few notable examples where nucleobase tautomerism is a key contributor or otherwise modulates the biological functions performed by nucleic acids.

Tautomerism in RNA Riboswitches

The ligand recognition by the purine and thiamine pyrophosphate riboswitches involves binding to minor tautomeric forms of their non-natural ligands (**Figure 2**) (Thore et al., 2008; Gilbert et al., 2009; Singh et al., 2014). The purine riboswitch regulates expression of genes involved in the metabolism of adenine and guanine. The regulation is mediated by the conformational change in response to binding to purine ligands (Gilbert et al., 2009; Gilbert et al., 2006). In addition to recognizing adenine and guanine, the purine riboswitch can also bind to metabolites of purines such as xanthine, hypoxanthine and 2,6-diaminopurine (Gilbert et al., 2009; Gilbert et al., 2006). Crystal structures of these ligands bound to the purine riboswitch show that the cytosine at the 74 position (C74) is critical for determining substrate specificity (**Figure 2A**) (Gilbert et al.,

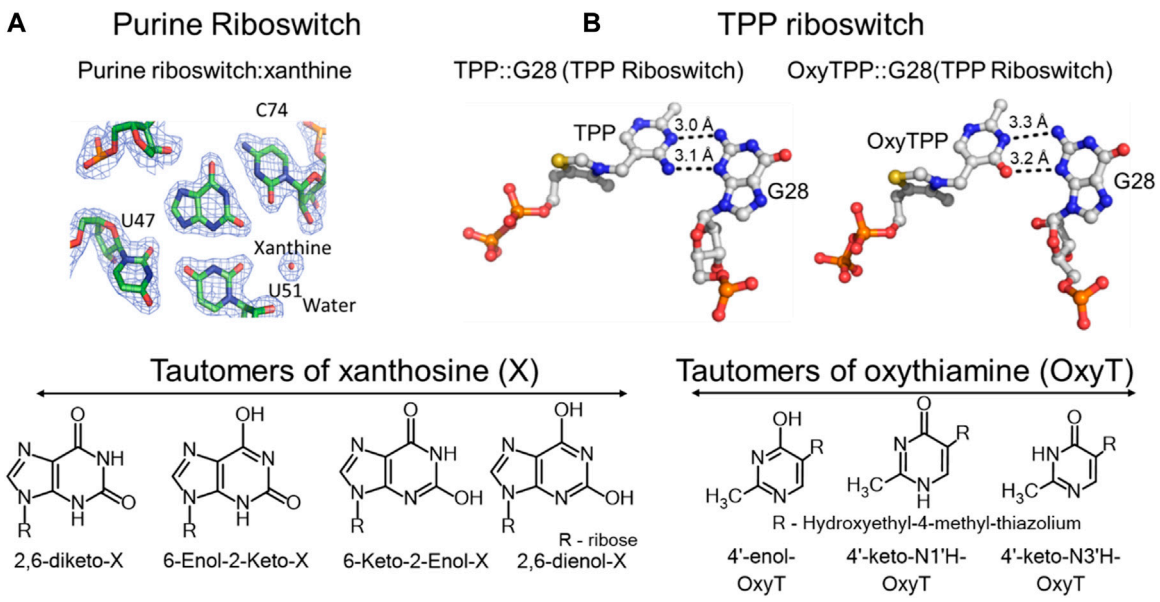

FIGURE 2 | Tautomerism in riboswitches. **(A)** Top left, crystal structure of the purine riboswitch with xanthine in its binding pocket interacting with carbonyl oxygens of C74, U47 and U51, and a water molecule. Bottom left, all possible tautomeric forms of xanthosine. **(B)** Crystal structures of the TPP riboswitch showing interactions of its binding site G28 (guanosine at the 28th position) with the thiamine pyrophosphate (TPP) and oxythiamine pyrophosphate (OxyTPP) ligands. Bottom, tautomeric forms of oxythiamine identified by NMR and vibrational spectroscopy (Singh et al., 2014). The figure is adapted from reference (Singh et al., 2014) (CC BY 4.0).

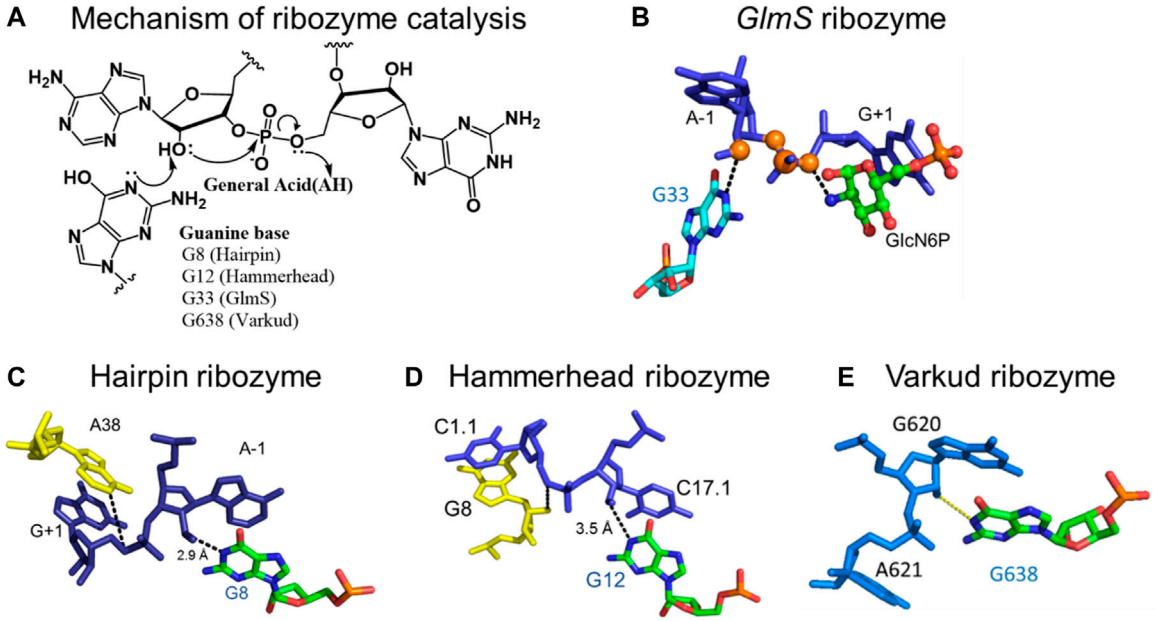

FIGURE 3 | Mechanism of the nucleolytic reaction catalyzed by small self-cleaving ribozymes. **(A)** Acid-base catalytic mechanism of small self-cleaving ribozymes and the positions of participating catalytic guanosines in the active sites of the respective ribozymes. **(B)** Catalytic guanosine (G33) in the active site of the *glmS* ribozyme, in close proximity to the 2′-hydroxyl of the A-1 nucleotide adjacent to the scissile phosphodiester bond. **(C)** The N1 of catalytic guanosine (G8) in the hairpin ribozyme in close proximity to the 2′-hydroxyl of the A-1 nucleotide. **(D)** The N1 of catalytic guanosine (G12) in the hammerhead ribozyme in close proximity to the 2′-hydroxyl of the A-1 nucleotide. **(E)** The N1 of the catalytic guanosine (G698) in the Varkud Satellite ribozyme in close proximity the 2′-hydroxyl of A-1 nucleotide. **Figure 3E** was shared by Joe Piccirilli's laboratory at the University of Chicago (Suslov et al., 2015), and parts of the figure are adapted from reference (Singh et al., 2014) (CC BY 4.0).

2009). The carbonyl oxygen (O2) of C74 and U51 form hydrogen bonds with the 2-amino functional group of the native ligand guanine (Gilbert et al., 2009), which explain the riboswitch preference for ligands that have a hydrogen bond donor at the 2-position of purines such as guanine and 2,6-diaminopurine (Gilbert et al., 2009; Gilbert et al., 2006). However, the riboswitch can also bind to xanthine, which has a carbonyl group at the 2-position, and hypoxanthine, which lacks any functional group at that positon, albeit with a weaker affinity (**Figure 2A**) (Gilbert et al., 2009). Xanthine binding has a dissociation constant (K_D) of 32 μM, three orders of magnitude higher than the nanomolar K_Ds for guanine and 2,6-diaminopurine (Gilbert et al., 2009). Hypoxanthine has 200-fold less affinity compared to guanine (Gilbert et al., 2009). However, the crystal structures of the riboswitch bound to xanthine or guanine ligands reveal no significant structural differences (Gilbert et al., 2009). To rationalize their identical structures, it was proposed that the 2-enol form of xanthine would alleviate the repulsive interactions that exist due to the presence of three negatively charged oxygens in close proximity (**Figure 2A**) (Singh et al., 2015). This proposal was also consistent with the biochemical observation that xanthine binds the riboswitch at a pH of 6 (K_D = 33 μM), but no binding is detected at a higher pH of 8.5 (Gilbert et al., 2009). The enol tautomers are expected to be more stable at lower pH (Gilbert et al., 2009). Despite the observed crystallographic and biochemical evidence in support of the minor 2-enol tautomer, the direct existence of 2-enol xanthine has not been established. This is due, in part, to the lack of sensitive methods for distinguishing keto and enol forms of xanthine in the background overlapping signals from polymeric RNA (Gilbert et al., 2009; Singh et al., 2014).

Tautomerism is also suggested to influence oxythiamine pyrophosphate (OxyTPP) binding to the thiamine pyrophosphate (TPP) riboswitch (Thore et al., 2006; Thore et al., 2008). The TPP riboswitch binds to its natural ligand, TPP, to negatively regulate the expression of genes involved in the biosynthesis and transport of thiamine (Thore et al., 2006). The X-ray structure of the riboswitch with the TPP ligand shows that the amino group at the 4′-position of TPP acts as a hydrogen bond donor to the N3 position of G28 (**Figure 2B**) (Thore et al., 2006; Thore et al., 2008). The OxyTPP has a carbonyl group at the 4′-position. Yet, under crystalline conditions, its hydrogen bonding interactions to G28 are almost identical to those of TPP (**Figure 2B**) (Thore et al., 2008). The 4′- position of OxyTPP can only act as a hydrogen bond donor in its enol form, leading to the hypothesis that OxyTPP binds to the riboswitch as an enol tautomer (Thore et al., 2008). The Oxythiamine (Oxy) portion of OxyTPP exists in three tautomeric forms including the proposed 4′-enol tautomer (Singh et al., 2014). Biochemical binding isotope effect (BIE) experiments combined with density functional theory (DFT) calculations performed using O-18 labelled OxyTPP could not unambiguously identify the tautomer of OxyTPP in the binding pocket of the TPP riboswitch (Singh et al., 2014). More sensitive methods, such as NMR relaxation dispersion,

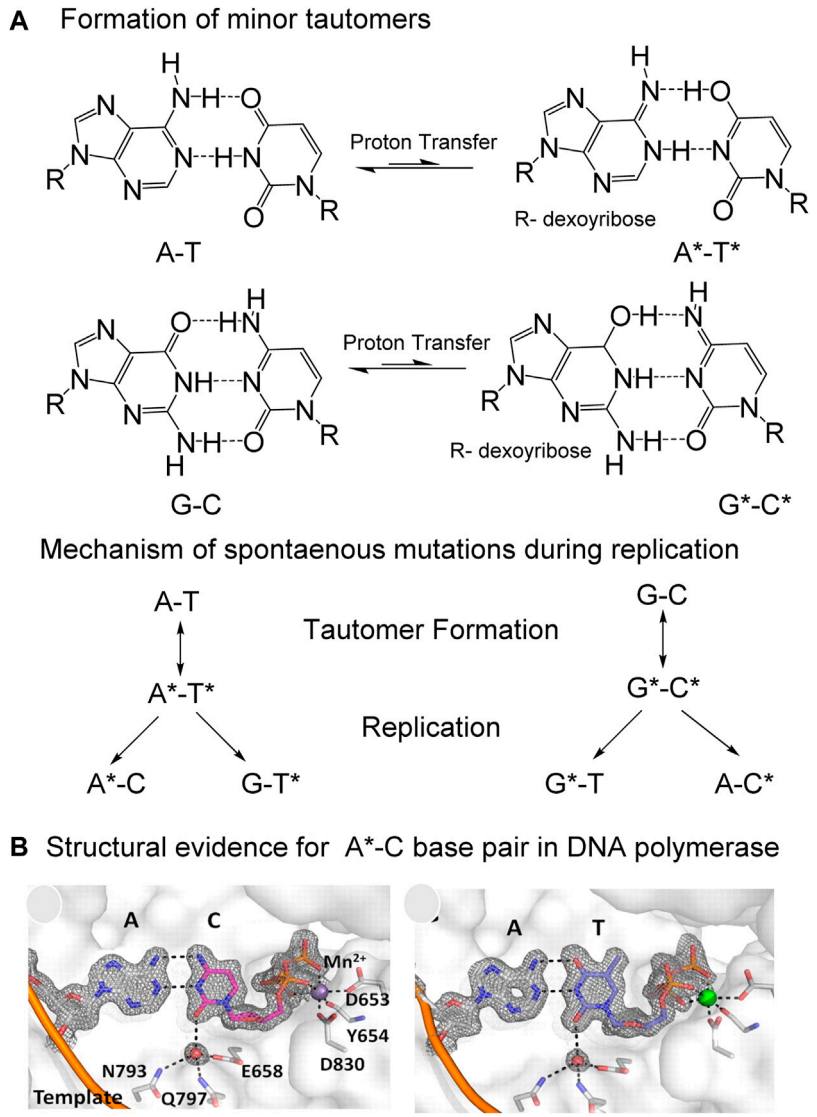

FIGURE 4 | Proposed mechanism by which "spontaneous" mutations are introduced by the inter-strand movement of protons in Watson-Crick (W–C) base pairs during replication. **(A)** Spontaneous transfer of two protons from one strand to another in A-T and G-C base pairs generate minor tautomers, which can form mismatches during replication to cause mutations. **(B)** Structural evidence for the stabilization of A-C base pair in W-C conformation, almost indistinguishable from the A-T base pair in active site of a high fidelity DNA polymerase. Figure is adapted from reference (Wang et al., 2011; Slocombe et al., 2021).

low temperature NMR, FTIR (Fourier Transform Infrared Spectroscopy) and 2D-IR (2-Dimentional Infrared Spectroscopy) (discussed in the Methods overview) are needed to unambiguously characterize the tautomeric form of OxyTPP or its binding partner G28 on the TPP riboswitch.

Tautomerism in Self-Cleaving Ribozymes

Small self-cleaving ribozymes are proposed to utilize the minor tautomeric form(s) of catalytic guanosines to execute the first step of the self-cleavage reactions (**Figure 3**) (Cochrane and Strobel, 2008a). As mentioned above, the hammerhead, hairpin, glmS, VS, HDV, and the twister ribozymes are examples of small self-cleaving ribozymes that catalyze the nucleolytic intramolecular self-scission reactions (**Figure 3**) (reviewed in (Cochrane and Strobel, 2008a)). Their reaction mechanism are very similar, and they typically involve base catalyzed activation of the site-specific 2′-hydroxyl that acts as a nucleophile to attack the adjacent scissile 3′-phosphate. This reaction yields two RNA products: one containing the 2′,3′-cyclic phosphate and the other the 5′-hydroxyl functional group (**Figure 3A**). The only exception to this mechanism is the *glmS* ribozyme, which requires an external cofactor in the self-cleavage reaction (Cochrane et al., 2007).

The self-scission reactions catalyzed by these ribozymes are proposed to utilize the minor tautomeric form of catalytic guanosines, in which the N1 is not protonated, to act as a general base in the 2′-hydroxyl activation step of the reaction

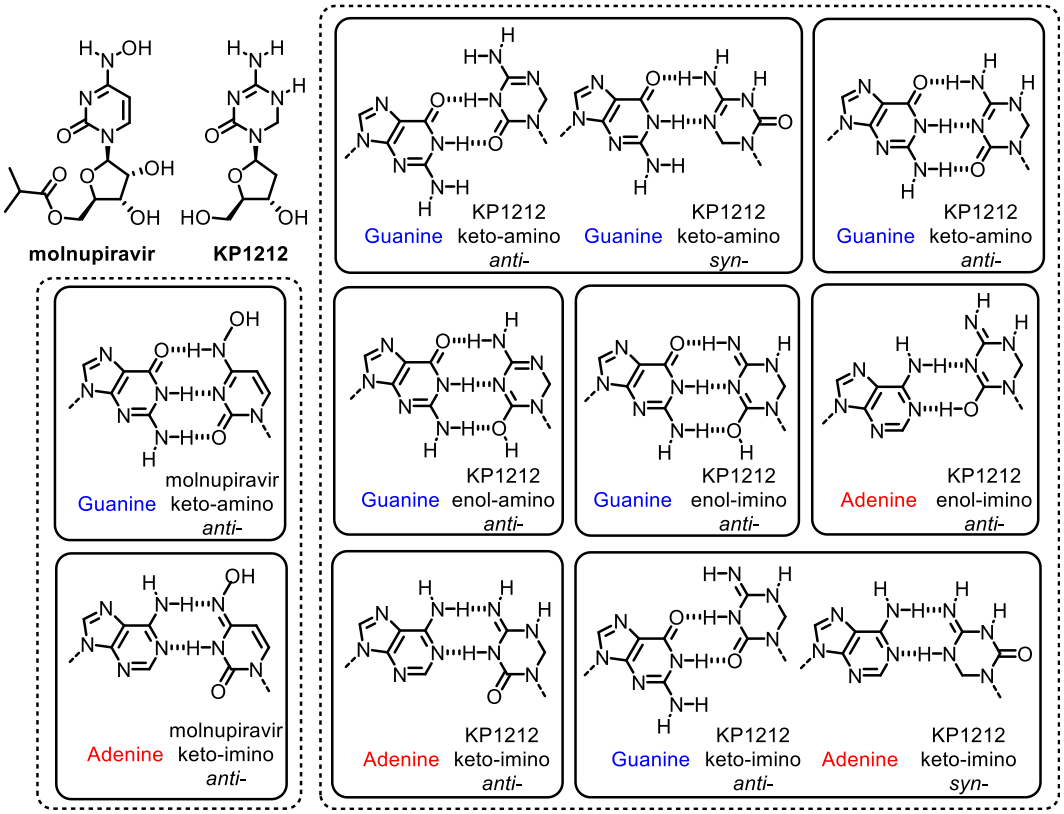

FIGURE 5 | Proposed tautomeric structures of antiviral drugs molnupiravir and KP1212, and base pairing properties during replication in SARS-CoV-2 and HIV viruses, respectively (Li et al., 2014; Kabinger et al., 2021).

(Singh et al., 2015) (**Figure 3A**). Structural studies of hammerhead (Martick and Scott, 2006), hairpin (Rupert and Ferré-D'Amaré, 2001), VS (Suslov et al., 2015) and the glmS (Cochrane et al., 2007; Klein and Ferre´-D'Amare´, 2006) ribozymes identified the N1 of catalytic guanosines in close proximity, within hydrogen bonding distance, to the 2′-hydroxyl nucleophile. These studies established that the N1 of G33 in *glmS* (**Figure 3B**) (Cochrane et al., 2007; Klein and Ferre´-D'Amare´, 2006), G8 in hairpin (**Figure 3C**) (Fedor, 2000; Pinard et al., 2001; Rupert and Ferré-D'Amaré, 2001; Kuzmin et al., 2004; Bevilacqua and Yajima, 2006), G12 in hammerhead (**Figure 3D**) (McKay, 1996; Han and Burke, 2005; Martick and Scott, 2006; Thomas and Perrin, 2008), and G638 in VS (**Figure 3E**) (Lafontaine et al., 2001; Hiley et al., 2002; Sood and Collins, 2002; Suslov et al., 2015) play the role of a general base in the self-cleavage reactions catalyzed by these ribozymes.

However, the N1 of guanosine has a pK_a of ~10 and is protonated at the physiological pH of ~7.4 (Singh et al., 2015). The protonated N1 of guanosine is a poor base to abstract a proton from the 2′-hydroxyl, which has a pK_a of ~13 (Velikyan et al., 2001). Therefore, tautomeric or ionic forms of the catalytic guanosines in which the N1 is not protonated are expected to form transiently during the catalysis. These minor tautomers are likely more nucleophilic at N1, and thus could extract the proton from the 2′-hydroxyl groups (Singh et al., 2015). Generation of N1 unprotonated guanosine would require perturbation of its pK_a towards neutrality. Significant perturbations in pK_a, by as much as four units, have been reported in RNA systems, including ribozymes and riboswitches (Legault and Pardi, 1997; Wilcox and Bevilacqua, 2013). Perturbation of pK_a towards neutrality would facilitate deprotonation at the N1 site of catalytic guanosines because prototropic tautomerism involves the deprotonation and the protonation steps, and is optimal in functional groups whose pK_a's are close to neutral (Singh et al., 2015).

Despite significant structural and biochemical evidence in support for the presence minor tautomeric forms of catalytic guanosines in small self-cleaving ribozymes, their direct identification has proven challenging, owing to the lack of sensitive methods (Singh et al., 2015).

Tautomerism in DNA and its Role in Replication Fidelity

The genetic integrity of genomic DNA relies on adenine, guanine, cytosine and thymine existing predominantly in their keto and amino tautomeric forms during replication and transcription (Watson and Crick, 1953; Topal and Fresco, 1976; Wang et al., 2011; Rangadurai et al., 2020).

Therefore, replication fidelity is expected to be influenced by the formation of minor tautomers (Watson and Crick, 1953; Topal and Fresco, 1976; Wang et al., 2011; Rangadurai et al., 2019; Rangadurai et al., 2020). In their work on the structure of DNA, Watson and Crick did appreciate that the formation of minor tautomeric forms would alter the base pairing properties of nucleic acid bases, potentially with mutagenic consequences (Watson and Crick, 1953). These minor tautomeric forms could arise from inter-helical transfer of protons in a DNA duplex (**Figure 4**) (Löwdin, 1963; Sevilla et al., 1995). Transient formation of minor tautomeric forms of DNA bases, and their stabilization in the active site of DNA polymerases during replication, could lead to incorporation of mismatched base pairs (Topal and Fresco, 1976; Watson and Crick, 1953; Wang et al., 2011). This phenomenon is plausible considering that the kinetics of minor tautomer formation and their lifetime in the active site is comparable with the kinetics of nucleotide incorporation by the polymerase. Therefore, the probability of base misincorporation during DNA replication may be correlated with the probability of rare tautomer formation (**Figure 4A**) (Topal and Fresco, 1976; Wang et al., 2011; Peng et al., 2015; Rangadurai et al., 2019).

Structural evidence for the rare tautomer hypothesis for spontaneous mutation came from the high-resolution crystal structure of a DNA polymerase that catalyzes replication in crystals (**Figure 4B**) (Wang et al., 2011). It was observed that a C•A mismatch mimics the shape of the cognate C•G base pair in the crystal (Wang et al., 2011). The movement of protons in the mismatched bases alter the hydrogen-bonding pattern such that the base pairs involving the minor tautomeric forms adopt an overall shape that is virtually indistinguishable from the canonical W-C base pair in DNA (Wang et al., 2011). This "shape mimicry" allows the mismatch to evade error detection mechanisms of human polymerases (**Figure 4B**). These observations provided structural support for the rare tautomer hypothesis of spontaneous mutagenesis.

NMR dispersion methods have allowed us to gain better understanding of tautomeric dynamics in DNA, and in RNA duplexes (Kimsey et al., 2018; Rangadurai et al., 2019; Rangadurai et al., 2020). Using W-C mismatches such as G-T or G-U, these studies established that G•T/U wobble mismatches exist in dynamic equilibrium between three distinct W-C mismatched base pairs within the DNA and RNA duplexes (Kimsey et al., 2018; Rangadurai et al., 2019; Rangadurai et al., 2020). The three distinct W-C mismatches include two tautomeric and one anionic species (Kimsey et al., 2018; Rangadurai et al., 2019; Rangadurai et al., 2020). The tautomeric forms were established using the chemical shifts of guanine N1 and thymidine/uridine N3. The chemical shifts, although consistent with G^{enol}•T/U base pair (minor enol tautomer of G paired with the dominant keto tautomeric form of T or U) were partially skewed toward G•T^{enol}/U^{enol} pairs (keto-amino tautomer of G paired with enol tautomer of T or U) (Kimsey et al., 2018). This skewness was interpreted as evidence for a rapid equilibrium between the major G^{enol}•T/U and the minor G•T^{enol}/U^{enol} tautomeric base pairs (Kimsey et al., 2018). The two rapidly exchanging tautomeric species (G^{enol}•T/U ⇌ G•T^{enol}/U^{enol}) were also quantitated and found to be around 0.4% of the total population at neutral pH. Increased understanding of base pairing mismatches like G•T, G•U, and A•C that nevertheless adopt W-C like geometry through either tautomerization or ionization allowed us to appreciate that these mismatches appear to be more common (Kimsey et al., 2018; Rangadurai et al., 2019; Rangadurai et al., 2020). Stabilization of W-C mismatches through tautomerization, either in the template strand or in the incoming nucleotide, could allow the incorporation of mismatches during replication (Topal and Fresco, 1976; Wang et al., 2011; Rangadurai et al., 2020), and remains a compelling mechanistic explanation for spontaneous mutagenesis.

Therapeutic Implications of Tautomeric Nucleosides

Nucleoside/nucleotide analogs enriched in minor tautomers are effective as antiviral agents against many viruses (Crotty et al., 2001; Baranovich et al., 2013; Li et al., 2014; Singh et al., 2015; Delang et al., 2018; Kabinger et al., 2021). The antiviral property of these analogs stem from their ability to induce mutations in viral genomes (Li et al., 2014; Delang et al., 2018; Gordon et al., 2021; Kabinger et al., 2021). Such nucleoside analogs are mutagenic, in part, because they form significant amounts of minor tautomeric forms, which enables them to engage in ambiguous base pairing, i.e., paring with more than one base (Li et al., 2014; Delang et al., 2018; Kabinger et al., 2021). The concept of specifically increasing the mutation rates of pathogens to drive their population collapse is particularly effective against retroviruses and riboviruses (Crotty et al., 2001; Anderson et al., 2004; Domingo et al., 2008; Manrubia et al., 2010; Domingo et al., 2012). These viruses have high mutation rates and their population lives close to the mutational threshold called the error catastrophe limit (ECL), a theoretical mutational rate above which producing viable viral progeny becomes impossible and leads to viral population collapse (Eigen, 2002). This phenomenon is called lethal mutagenesis. Additionally, unlike the human replicative polymerases that feature both high selectivity and high fidelity, the replication machinery of retroviruses is both promiscuous and error prone and thus, it can readily incorporate mutagenic nucleoside analogs (Anderson et al., 2004; Li et al., 2014; Kabinger et al., 2021). Therefore, these nucleoside analogs selectively mutagenize the viral genomes, while leaving the host (human) genome essentially untouched (Crotty et al., 2001; Anderson et al., 2004; Delang et al., 2018; Kabinger et al., 2021).

Nucleoside analogs that induce lethal mutagenesis are effective as antiviral drugs against many retroviruses and riboviruses, such as HIV, HCV, influenza virus and coronaviruses (Anderson et al., 2004; Baranovich et al., 2013; Li et al., 2014; Delang et al., 2018; Gordon et al., 2021 Kabinger et al., 2021). These viruses have high mutation rates and their replication machinery is error-prone (Watson and Crick, 1953; Crotty et al., 2001; Wang et al., 2011; Li et al., 2014; Rangadurai et al., 2020; Kabinger et al., 2021). The error prone nature of some viral polymerases allows incorporation of modified nucleosides (Crotty et al., 2001; Li

et al., 2014; Peng et al., 2015; Kabinger et al., 2021). The 5-aza-5,6-dihydro-2′-deoxycytidine (KP1212) is an experimental drug candidate that targets HIV by increasing the mutation rate of the virus (**Figure 5**) (Harris et al., 2005; Mullins et al., 2011; Li et al., 2014). The antiviral efficacy of KP1212 stems from its ability to exist in multiple tautomeric states (Li et al., 2014; Peng et al., 2015). These tautomers can engage in ambiguous base pairings to induce mutagenesis in HIV (Li et al., 2014; Peng et al., 2015). Other antiviral drugs such as ribavirin for HCV (Crotty et al., 2001; Li et al., 2014), and favipiravir for influenza (Baranovich et al., 2013; Delang et al., 2018) are also efficacious due, in part, to their ability to exist in multiple tautomeric or rotameric forms and induce mutagenesis in viral genomes. Lethal mutagenesis has also been implicated as the main mechanism of action for the recently developed small-molecule therapeutics, such as molnupiravir from Merck, for severe acute respiratory syndrome coronavirus (SARS-CoV-2, the virus that causes Covid-19) (Kabinger et al., 2021; Zhou, 2021).

The mechanism by which KP1212 induces lethal mutagenesis in HIV has been studied extensively (Li et al., 2014; Peng et al., 2015). KP1212 is mutagenic, inducing G to A and A to G transition mutations (**Figure 5**) (Harris et al., 2005; Li et al., 2014). The mutagenicity of KP1212 is due, in part, to its ability to exist in multiple tautomeric or rotameric forms (Peng et al., 2013; Li et al., 2014; Peng et al., 2015). While KP1212 is considered a dC (deoxycytidine) analog, because it features the functional groups and the W-C face of dC, it has a near neutral pKa of ~7, compared to pK_a of ~4 for dC. As a consequence, KP1212 in solution exists in multiple tautomeric states (Li et al., 2014; Peng et al., 2015). These include keto-amino, keto-imino and its rotameric form, enol-amino and enol-imino and its rotameric form (Li et al., 2014). The enol-amino and enol-imino are the dominant tautomeric forms of KP1212, in contrast to the dominant keto-amino form observed for dC (Li et al., 2014; Peng et al., 2015). The alternative tautomeric forms of KP1212 are expected to have perturbed base pairing properties (Li et al., 2014), and thus contribute to the mutagenic properties of the base. KP1212 was shown to induce G to A and A to G mutations in the HIV genomes in cellular models, in preclinical rodents studies and in HIV patients who participated in the clinical trials for KP1212 (Harris et al., 2005; Mullins et al., 2011; Li et al., 2014). However, these studies also revealed that KP1212 is not mutagenic to human cells. Lack of mutagenicity in humans is likely due to the higher fidelity and selectivity of human DNA replicative polymerases (Mullins et al., 2011). Unlike dC, KP1212 has a saturated carbon center at the 6-position, which causes the base ring to be puckered (Li et al., 2014). This geometric distortion is significant enough to be selected against by the human polymerases. KP1212, therefore, exemplifies the features of an ideal lethal mutagen because it specifically induces mutations in HIV, driving viral population collapse, while sparing the human genome.

Molnupiravir is a broad-spectrum nucleoside analog drug that is efficacious against many viruses including the proofreading-intact SARS-CoV-2 coronavirus with a high genetic barrier to resistance (Gordon et al., 2021; Agostini et al., 2019). Its efficacy stems from its ability to induce lethal mutagenesis in SARS-CoV-2, during the early stages of Covid-19 (**Figure 5**) (Agostini et al., 2019; Kabinger et al., 2021; Gordon et al., 2021; Menéndez-Arias, 2021). Molnupiravir is an isopropyl ester prodrug of the ribonucleoside analog β-D-N^4-hydroxycytidine (NHC) (Kabinger et al., 2021). It is currently in the phase-III clinical trial for Covid-19 (Kabinger et al., 2021). Biochemical studies using SARS-CoV-2 RNA-dependent-RNA polymerase (SARS-CoV-2 RdRp) have shown that the triphosphate form of NHC can be incorporated into RNA, albeit at a lower frequency compared to canonical nucleosides (Gordon et al., 2021; Menéndez-Arias, 2021). Once incorporated into the growing strand, it can be extended at its 3′-end. When present in the template strand it can form base pairs with G (NHC:G) or A (NHC: A) to induce G to A or A to G mutations (**Figure 5**) (Gordon et al., 2021). The NHC:A base pairs is more efficiently extended compared to NHC:G (Kabinger et al., 2021). Cellular studies have shown that NHC is 100-fold more active than ribavirin and favipiravir against SARS-CoV-2 (Zhou, 2021). The higher activity molnupiravir also induces a higher mutation frequency in the viral RNA (Menéndez-Arias, 2021; Zhou, 2021). Structural analysis of RdRp–RNA complexes containing mutagenesis products revealed that the NHC moiety can form stable base pairs with either G or A in the active site of RdRp (Kabinger et al., 2021). The formation of stable base mispairs with G or A, in W-C conformations, allow molnupiravir to escape the proofreading mechanism of the SARS-CoV-2 RdRp polymerase (Agostini et al., 2019; Gordon et al., 2021; Kabinger et al., 2021; Menéndez-Arias, 2021). The accumulation of mutations with each round of viral multiplication would lead to weakening of the virus (Kabinger et al., 2021; Menéndez-Arias, 2021). Similar to molnupiravir against SARS-CoV-2, ribavirin and favipiravir target HCV, and influenza, respectively, in part by the lethal mutagenesis mechanism (Crotty et al., 2001; Baranovich et al., 2013). Taken together, all the above studies underscore the importance of better understanding the mechanisms by which nucleoside analogs are incorporated and lead to mispaired bases in viral genomes.

Methods for Studying Tautomerism in Nucleic Acids and in Nucleoside Analogs

Minor tautomers of canonical nucleic acid bases and their analogs are rare, transient, and interconvert on a fast time scale (Topal and Fresco, 1976; Rangadurai et al., 2019; Peng et al., 2015). Recent developments of spectroscopic methods have allowed their direct identification in nucleic acids and in nucleoside analogs (**Figure 6**) (Wang et al., 2011; Peng and Tokmakoff, 2012; Peng et al., 2013; Singh et al., 2014; Li et al., 2014; Peng et al., 2015; Szymanski et al., 2017; Rangadurai et al., 2019; Rangadurai et al., 2020; Kabinger et al., 2021). Prototropic tautomers differ from one another based on the position of protons, and the position of double bonds. Therefore, methods that are capable of detecting and distinguishing protons in different chemical

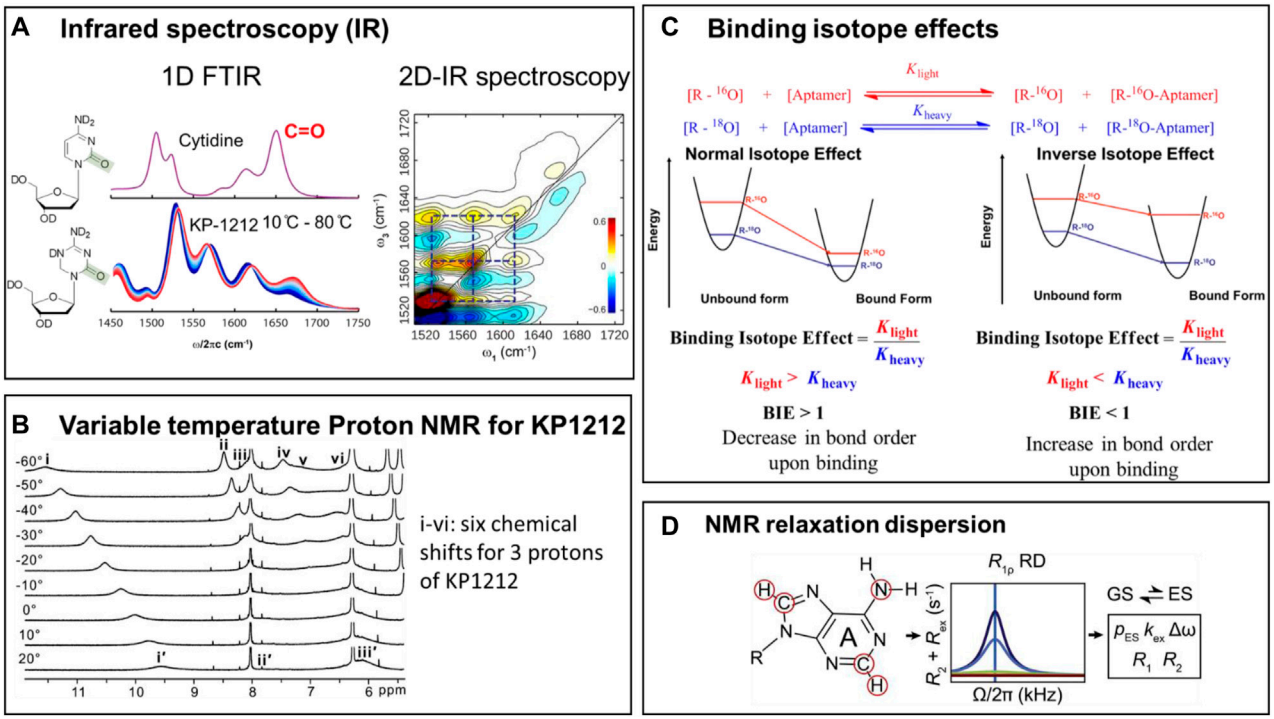

FIGURE 6 | Methods for studying tautomerism in nucleoside analogs and nucleic acids. **(A)** Left side: The Fourier-Transform Infrared (FTIR) spectra for 2′-deoxycytidine (dC) at room temperature, and the temperature dependence of vibrational frequencies in the. FTIR spectra for KP1212. Right side: The 2-Dimensional Infrared (2D-IR) spectroscopic data for KP1212, in aqueous solution at room temperature (Li et al., 2014). **(B)** Variable temperature 1H-NMR data for KP1212 (Li et al., 2014). **Figures 6A,B** are adapted from (Li et al., 2014). **(C)** Binding isotope effects combined with Density Functional theory (DFT) calculations used in the characterization of the tautomeric form of OxyTPP bound to the TPP riboswitch (Singh et al., 2015). Reprinted (adapted) with permission from (Singh et al., 2014) *Copyright © 2014, American Chemical Society.* **(D)** NMR relaxation dispersion methods used to identify and quantitate rare tautomers in a G●T mismatch in DNA/RNA duplexes in W-C conformations (Rangadurai et al., 2019). Figure courtesy of Hashim Al-Hashimi Duke University.

environment (for example, NMR) are useful for the direct identification of tautomeric forms (Singh et al., 2014; Li et al., 2014; Peng et al., 2013) in both nucleic acids and nucleoside analogs (**Figures 6B,C**) (Li et al., 2014; Peng et al., 2015; Rangadurai et al., 2019; Rangadurai et al., 2020). Similarly, methods that provide information on the bond order and strength of a specific chemical bond (such as IR and Raman spectroscopy) are useful for distinguishing between keto-enol or amino-imino tautomeric forms, because the functional groups in each tautomeric pair have very different and characteristic vibrational properties. In addition, biochemical binding isotope effects and their interpretation using computational DFT have been used for characterizing tautomeric forms of a ligand in the binding pocket of an RNA aptamer (**Figure 6C**) (Singh et al., 2014).

NMR measures the chemical shift of individual protons, a property that depends on the chemical environment of the proton. Since the protons in tautomeric pairs are attached to different heteroatoms, their NMR chemical shifts are very different and thus distinguishable. Often, because the tautomeric equilibria are faster than the NMR acquisition time scale, broader peaks are observed, that span areas of multiple tautomeric forms. However, by substantially lowering the temperature of the sample (a method called variable temperature NMR), the tautomeric equilibria can be slowed down sufficiently to resolve the individual tautomeric forms.

Infrared spectroscopy (IR) measures the vibrational properties of the chemical bonds. Depending on their chemical makeup and bond order, different bonds absorb infrared radiation at different wavelengths. IR spectra provide direct information on the types of bonds present in a molecule, and therefore can easily distinguish between keto-enol or amino-imino tautomeric forms. Two-dimensional IR (2D-IR), an advanced version of IR, uses the same principle but provides, in addition, information on bond connectivity. Bonds in the same molecule vibrate together, and thus give a characteristic cross peak in the spectrum, which allows the distinction between the individual tautomeric forms present at equilibrium.

In addition to the direct detection of rare tautomers in nucleic acids, which is very challenging, indirect approaches, based on binding isotope effects (BIEs), exist that allow inference of minor tautomer formation. Theoretically, BIEs report on changes in vibrational frequencies of a ligand upon binding to its target. Experimentally, they are estimated using the binding affinities of ligands carrying either light or heavy isotopes of atoms directly involved in the binding interaction. The experimental BIEs are then compared to BIEs calculated, using the Density Functional Theory (DFT), from frequencies of optimized tautomeric forms.

The exact characterization is completed by identifying a tautomeric structure whose calculated BIEs closely match the experimental BIEs. A good example of this approach is the identification of tautomeric form of OxyTPP recognized by the TPP riboswitch (**Figure 6C**), where spectroscopic observations indicated that each tautomer has a unique vibrational frequency pattern. This approach was used to study OxyTPP bound to the TPP riboswitch (**Figure 6C**), using ^{18}O and $^{32}P/^{33}P$ labeled OxyTPP (Singh et al., 2014; Singh et al., 2015). Although the calculated BIEs are different for the keto and enol forms of OxyTPP, the magnitude of the difference between the BIEs corresponding to various tautomers of OxyTPP was too small to yield a definite conclusion. More precise measurements of BIEs are nevertheless expected to provide clarity on the tautomeric form of OxyTPP bound to the TPP riboswitch (Sood and Collins, 2002; Singh et al., 2014). These studies also underscore the potential of indirect measurements like BIEs and kinetic isotope effects (KIEs) to increase our understanding of tautomerism in nucleic acids and base analogs.

The variable temperature NMR, FT-IR (Fourier Transform Infrared) and 2D-IR (2-dimensional Infrared) methods were used to identify the tautomers of KP1212 and oxythiamine (Singh et al., 2014; Li et al., 2014; Peng et al., 2015). Tautomeric dynamics of these nucleoside analogs are fast, typically on the nanosecond timescale (Peng et al., 2013; Peng et al., 2015). Therefore, attempts to distinguish between the multiple tautomers of KP1212 at room temperature using NMR were not successful. Lowering the temperature to −20°C allowed identification of individual tautomers of KP1212 and oxythiamine by 1-dimentional proton NMR (^{1}H-NMR) in the dimethylformamide (DMF) solvent (**Figure 6B**) (Singh et al., 2014; Li et al., 2014; Peng et al., 2015). While the characterization and distribution of tautomers of KP1212 and oxythiamine by ^{1}H-NMR utilized non-physiological conditions with low temperature and an aprotic solvent (DMF) (Li et al., 2014), these experiments clearly outlined the chemical plausibility of the tautomeric forms of these compounds. For detecting multiple tautomers under physiologically relevant aqueous conditions at room temperature, IR-based methods have proven to be more effective (Singh et al., 2014; Li et al., 2014; Peng et al., 2015). The temperature dependence of change in amplitude of vibrational frequencies in the 1D-FTIR spectrum confirmed the presence of multiple tautomers for KP1212 and oxythiamine (**Figure 6C**) (Li et al., 2014). To directly identify minor tautomers based on altered vibrational frequency of minor tautomers, 2D-IR data was combined with DFT calculations (**Figure 6A**) (Peng et al., 2015). The 2D IR spectrum has a greater spatial and temporal resolution than FTIR, and thus allows direct identification of multiple fast interconverting tautomers (Peng et al., 2013; Li et al., 2014; Peng et al., 2015). These studies established that KP1212 exists in seven different tautomeric or rotameric forms and oxythiamine exists in three different tautomeric forms (Li et al., 2014; Singh et al., 2014). Furthermore, the biochemical observations that KP1212 is 10% mutagenic when replicated *in vitro* and in living cells, and induces G to A mutations, can be rationalized by the tautomeric distribution identified from these spectroscopic studies (Li et al., 2014; Peng et al., 2015).

By contrast with the nucleoside analogs described above, the minor tautomers of the canonical DNA bases are significantly less abundant, shorter lived and more challenging to detect (Peng et al., 2013; Rangadurai et al., 2019; Peng et al., 2015). NMR relaxation dispersion (RD) methods allow identification of low-abundance short-lived conformational states in biomolecules (**Figure 6D**) (Mulder et al., 2001; Rangadurai et al., 2019; Rangadurai et al., 2020). NMR relaxation dispersion (RD) has been used to characterize rare tautomers in Hoogsteen base pairs versus Watson-Crick base pairs in different types of DNA/RNA, such as A or B type of DNA/RNA (Nikolova et al., 2011; Alvey et al., 2014; Zhou et al., 2016; Rangadurai et al., 2018). These results offer understandings into differences between A-RNA and B-DNA duplexes and provide possible explanations for how they respond to damage and modifications.

Using NMR RD, it has been shown that the G•T/U mismatches exist in dynamic equilibrium between tautomeric and anionic W-C conformations within the DNA and RNA duplexes. These studies also provided insights into the kinetics of rare tautomer formation in nucleic acids, and the probability of base misincorporation due to the formation of minor tautomers (Rangadurai et al., 2019; Rangadurai et al., 2020). These methods continue to enhance our understanding of the mechanisms by which the formation of minor tautomers contribute to the generation of spontaneous mutations.

Structural methods based on X-ray crystallography and Cryo-EM (Cryogenic electron microscopy) have proven effective in studying mismatched base pairs that potentially involve minor tautomers. While protons are very difficult to visualize directly by X-ray crystallography and Cryo-EM, the overall geometry of base pairs and the inter-heteroatom distances can be accurately measured, and thus allow to infer the presence of protons and hydrogen bonding interactions. Specifically, crystallographic studies have shown that mismatches in high fidelity DNA polymerases can adopt structures that closely mimic the W-C base pairing geometry (Wang et al., 2011; Kabinger et al., 2021). The existence of mismatches in conformations that are identical to W-C shapes may allow them to evade the polymerase proof reading mechanisms and lead to mutagenesis (**Figure 4B**, **Figure 5A**) (Bebenek et al., 2011; Wang et al., 2011; Fedeles et al., 2015; Kabinger et al., 2021). In sum, these studies have provided convincing structural evidence in support of rare tautomer hypothesis for the generation of spontaneous mutations that arise during nucleic acid replication.

FUTURE DIRECTIONS

The last decade has seen significant progress in our understanding of tautomerism in DNA and RNA. Despite the early realization that tautomerism could influence nucleic acid biology, directly identifying rare tautomers of nucleic acid bases proved very challenging. The minor tautomers of DNA bases are present at equilibrium in minute amounts (< 0.1%) and their dynamics of interconversion is generally very fast. Until recently,

little progress was made in directly visualizing minor tautomers of DNA bases. The NMR RD allowed direct detection of rare tautomers and characterized the dynamics of their formation in short DNA sequences. Structurally establishing that mismatches can be stabilized in W-C conformation by high fidelity DNA polymerases, provided crystallographic support for the rare tautomer hypothesis of spontaneous mutations. Advancement in methods based on BIEs, high-resolution crystallography, NMR and IR significantly enhanced our mechanistic understanding of mutagenesis from nucleoside analogs based antiviral drugs. While all these advancements are notable, they only reflect special cases of tautomerism in a limited set of experimental conditions. Much work remains to develop more broadly applicable methods that enable the direct visualization and identification of minor tautomers of canonical bases in complex biological systems, such as genomic DNA, and in the binding pocket of polymerases, under physiological conditions. Sensitivity improvements in all the methods summarized above are all expected to improve our understanding of tautomerism. In the case of BIEs, more precise measurements are expected to help identify minor tautomers in the context of RNA systems and may provide further insights into the role of tautomerism in RNA biology, including catalytic processes and ribozymes. More work is also needed to bridge the knowledge gap between the in-solution tautomeric equilibria and dynamics of nucleobases and analogs, and their biological consequences. In the case of antiviral nucleoside analogs, such as KP1212 and molnupiravir, there is great interest to understand how tautomeric equilibria translate into mutagenic potential inside the viral polymerases. These nucleoside analogs have already proven very effective as antiviral drugs. However, our understanding of the physico-chemical properties that enable these compounds to adopt multiple tautomeric forms is still limited. Further insight into tautomeric equilibria and dynamics should allow us to develop more potent and safer antiviral therapeutics.

AUTHOR CONTRIBUTIONS

All the authors made substantial, direct, and intellectual contribution to the manuscript and approved it for publication.

FUNDING

BF is supported by NIH grants R01-CA080024, P30-ES002109, P42-ES027707 and T32-ES007020. DL is supported by R01-ES028865. VS is supported by Novartis Institute of Biomedical Research.

ACKNOWLEDGMENTS

The authors thank Michael Romanowski of Novartis Institute of Biomedical Research (NIBR); Hashim Al-Hashimi of Duke University for the NMR dispersion figure (**Figure 6D**), and Joe Piccirilli of University of Chicago for providing the Varkud ribozyme figure (**Figure 3E**).

REFERENCES

Agostini, M. L., Pruijssers, A. J., Chappell, J. D., Gribble, J., Lu, X., Andres, E. L., et al. (2019). Small-Molecule Antiviral β-d-N 4-Hydroxycytidine Inhibits a Proofreading-Intact Coronavirus with a High Genetic Barrier to Resistance. *J. Virol.* 93 (24), e01348–19. doi:10.1128/JVI.01348-19

Alvey, H. S., Gottardo, F. L., Nikolova, E. N., and Al-Hashimi, H. M. (2014). Widespread Transient Hoogsteen Base Pairs in Canonical Duplex DNA with Variable Energetics. *Nat. Commun.* 5 (1), 4786. doi:10.1038/ncomms5786

Anderson, J. P., Daifuku, R., and Loeb, L. A. (2004). Viral Error Catastrophe by Mutagenic Nucleosides. *Annu. Rev. Microbiol.* 58, 183–205. doi:10.1146/annurev.micro.58.030603.123649

Antonov, L. (2013). *Tautomerism: Methods and Theories*. Sofia, Bulgaria: John Wiley & Sons.

Baranovich, T., Wong, S.-S., Armstrong, J., Marjuki, H., Webby, R. J., Webster, R. G., et al. (2013). T-705 (Favipiravir) Induces Lethal Mutagenesis in Influenza A H1N1 Viruses *In Vitro*. *J. Virol.* 87 (7), 3741–3751. doi:10.1128/jvi.02346-12

Bebenek, K., Pedersen, L. C., and Kunkel, T. A. (2011). Replication Infidelity via a Mismatch with Watson-Crick Geometry. *Proc. Natl. Acad. Sci. USA* 108 (5), 1862–1867. doi:10.1073/pnas.1012825108

Bevilacqua, P., and Yajima, R. (2006). Nucleobase Catalysis in Ribozyme Mechanism. *Curr. Opin. Chem. Biol.* 10 (5), 455–464. doi:10.1016/j.cbpa.2006.08.014

Brown, R. D., Godfrey, P. D., McNaughton, D., and Pierlot, A. P. (1989). Tautomers of Cytosine by Microwave Spectroscopy. *J. Am. Chem. Soc.* 111 (6), 2308–2310. doi:10.1021/ja00188a058

Cochrane, J. C., Lipchock, S. V., and Strobel, S. A. (2007). Structural Investigation of the GlmS Ribozyme Bound to its Catalytic Cofactor. *Chem. Biol.* 14 (1), 97–105. doi:10.1016/j.chembiol.2006.12.005

Cochrane, J. C., and Strobel, S. A. (2008). Catalytic Strategies of Self-Cleaving Ribozymes. *Acc. Chem. Res.* 41 (8), 1027–1035. doi:10.1021/ar800050c

Cochrane, J. C., and Strobel, S. A. (2008). Riboswitch Effectors as Protein Enzyme Cofactors. *Rna* 14 (6), 993–1002. doi:10.1261/rna.908408

Colominas, C., Luque, F. J., and Orozco, M. (1996). Tautomerism and Protonation of Guanine and Cytosine. Implications in the Formation of Hydrogen-Bonded Complexes. *J. Am. Chem. Soc.* 118 (29), 6811–6821. doi:10.1021/ja954293l

Crotty, S., Cameron, C. E., and Andino, R. (2001). RNA Virus Error Catastrophe: Direct Molecular Test by Using Ribavirin. *Proc. Natl. Acad. Sci.* 98 (12), 6895–6900. doi:10.1073/pnas.111085598

Delang, L., Abdelnabi, R., and Neyts, J. (2018). Favipiravir as a Potential Countermeasure against Neglected and Emerging RNA Viruses. *Antivir. Res.* 153, 85–94. doi:10.1016/j.antiviral.2018.03.003

Domingo, E., Grande-Pérez, A., and Martín, V. (2008). Future Prospects for the Treatment of Rapidly Evolving Viral Pathogens: Insights from Evolutionary Biology. *Expert Opin. Biol. Ther.* 8 (10), 1455–1460. doi:10.1517/14712598.8.10.1455

Domingo, E., Sheldon, J., and Perales, C. (2012). Viral Quasispecies Evolution. *Microbiol. Mol. Biol. Rev.* 76 (2), 159–216. doi:10.1128/mmbr.05023-11

Eigen, M. (2002). Error Catastrophe and Antiviral Strategy. *Proc. Natl. Acad. Sci.* 99 (21), 13374–13376. doi:10.1073/pnas.212514799

Fedeles, B. I., Freudenthal, B. D., Yau, E., Singh, V., Chang, S.-c., Li, D., et al. (2015). Intrinsic Mutagenic Properties of 5-chlorocytosine: A Mechanistic Connection between Chronic Inflammation and Cancer. *Proc. Natl. Acad. Sci. USA* 112 (33), E4571–E4580. doi:10.1073/pnas.1507709112

Fedor, M. J. (2000). Structure and Function of the Hairpin Ribozyme. *J. Mol. Biol.* 297 (2), 269–291. doi:10.1006/jmbi.2000.3560

Gilbert, S. D., Mediatore, S. J., and Batey, R. T. (2006). Modified Pyrimidines Specifically Bind the Purine Riboswitch. *J. Am. Chem. Soc.* 128 (44), 14214–14215. doi:10.1021/ja063645t

Gilbert, S. D., Reyes, F. E., Edwards, A. L., and Batey, R. T. (2009). Adaptive Ligand Binding by the Purine Riboswitch in the Recognition of Guanine and Adenine Analogs. *Structure* 17 (6), 857–868. doi:10.1016/j.str.2009.04.009

Gordon, C. J., Tchesnokov, E. P., Schinazi, R. F., and Götte, M. (2021). Molnupiravir Promotes SARS-CoV-2 Mutagenesis via the RNA Template. *J. Biol. Chem.* 297, 100770. doi:10.1016/j.jbc.2021.100770

Han, J., and Burke, J. M. (2005). Model for General Acid–Base Catalysis by the Hammerhead Ribozyme: pH−Activity Relationships of G8 and G12 Variants at the Putative Active Site. *Biochemistry* 44 (21), 7864–7870. doi:10.1021/bi047941z

Harris, K., Brabant, W., Styrchak, S., Gall, A., and Daifuku, R. (2005). KP-1212/1461, a Nucleoside Designed for the Treatment of HIV by Viral Mutagenesis. *Antivir. Res.* 67 (1), 1–9. doi:10.1016/j.antiviral.2005.03.004

Hiley, S. L., Sood, V. D., and Fan, J. (2002). 4-thio-U Cross-Linking Identifies the Active Site of the VS Ribozyme. *Embo J* 21 (17), 4691–4698. doi:10.1093/emboj/cdf462

Kabinger, F., Stiller, C., Schmitzová, J., Dienemann, C., Kokic, G., Hillen, H. S., et al. (2021). Mechanism of Molnupiravir-Induced SARS-CoV-2 Mutagenesis. *Nat. Struct. Mol. Biol.* 28 (9), 740–746. doi:10.1038/s41594-021-00651-0

Kimsey, I. J., Szymanski, E. S., Zahurancik, W. J., Shakya, A., Xue, Y., Chu, C.-C., et al. (2018). Dynamic Basis for dGdT Misincorporation via Tautomerization and Ionization. *Nature* 554 (7691), 195–201. doi:10.1038/nature25487

Klein, D. J., Been, M. D., and Ferré-D'Amaré, A. R. (2007). Essential Role of an Active-Site Guanine in glmS Ribozyme Catalysis. *J. Am. Chem. Soc.* 129 (48), 14858–14859. doi:10.1021/ja0768441

Klein, D. J., and Ferre´-D'Amare´, A. R. (2006). Structural Basis of glmS Ribozyme Activation by Glucosamine-6-Phosphate. *Science* 313 (5794), 1752–1756. doi:10.1126/science.1129666

Kuzmin, Y. I., Da Costa, C. P., and Fedor, M. J. (2004). Role of an Active Site Guanine in Hairpin Ribozyme Catalysis Probed by Exogenous Nucleobase rescue. *J. Mol. Biol.* 340 (2), 233–251. doi:10.1016/j.jmb.2004.04.067

Laar, C. (1886). Ueber die Hypothese der wechselnden Bindung. *Ber. Dtsch. Chem. Ges.* 19 (1), 730–741. doi:10.1002/cber.188601901165

Lafontaine, D. A., Wilson, T. J., Norman, D. G., and Lilley, D. M. J. (2001). The A730 Loop Is an Important Component of the Active Site of the VS Ribozyme 1 1Edited by J. Karn. *J. Mol. Biol.* 312 (4), 663–674. doi:10.1006/jmbi.2001.4996

Legault, P., and Pardi, A. (1997). Unusual Dynamics and pKa Shift at the Active Site of a Lead-Dependent Ribozyme. *J. Am. Chem. Soc.* 119 (28), 6621–6628. doi:10.1021/ja9640051

Li, D., Fedeles, B. I., Singh, V., Peng, C. S., Silvestre, K. J., Simi, A. K., et al. (2014). Tautomerism Provides a Molecular Explanation for the Mutagenic Properties of the Anti-HIV Nucleoside 5-Aza-5,6-Dihydro-2'-Deoxycytidine. *Proc. Natl. Acad. Sci.* 111 (32), E3252–E3259. doi:10.1073/pnas.1405635111

Löwdin, P.-O. (1963). Proton Tunneling in DNA and its Biological Implications. *Rev. Mod. Phys.* 35 (3), 724–732. doi:10.1103/revmodphys.35.724

Manrubia, S. C., Domingo, E., and Lázaro, E. (2010). Pathways to Extinction: beyond the Error Threshold. *Phil. Trans. R. Soc. B* 365 (1548), 1943–1952. doi:10.1098/rstb.2010.0076

Martick, M., and Scott, W. G. (2006). Tertiary Contacts Distant from the Active Site Prime a Ribozyme for Catalysis. *Cell* 126 (2), 309–320. doi:10.1016/j.cell.2006.06.036

McKay, D. B. (1996). Structure and Function of the Hammerhead Ribozyme: an Unfinished story. *Rna* 2 (5), 395–403.

Menéndez-Arias, L. (2021). Decoding Molnupiravir-Induced Mutagenesis in SARS-CoV-2. *J. Biol. Chem.* 297 (1), 100867. doi:10.1016/j.jbc.2021.100867

Mons, M., Dimicoli, I., Piuzzi, F., Tardivel, B., and Elhanine, M. (2002). Tautomerism of the DNA Base Guanine and its Methylated Derivatives as Studied by Gas-phase Infrared and Ultraviolet Spectroscopy. *J. Phys. Chem. A.* 106 (20), 5088–5094. doi:10.1021/jp0139742

Mulder, F. A. A., Mittermaier, A., Hon, B., Dahlquist, F. W., and Kay, L. E. (2001). Studying Excited States of Proteins by NMR Spectroscopy. *Nat. Struct. Biol.* 8 (11), 932–935. doi:10.1038/nsb1101-932

Mullins, J. I., Heath, L., Hughes, J. P., Kicha, J., Styrchak, S., Wong, K. G., et al. (2011). Mutation of HIV-1 Genomes in a Clinical Population Treated with the Mutagenic Nucleoside KP1461. *PloS one* 6 (1), e15135. doi:10.1371/journal.pone.0015135

Nikolova, E. N., Kim, E., Wise, A. A., O'Brien, P. J., Andricioaei, I., and Al-Hashimi, H. M. (2011). Transient Hoogsteen Base Pairs in Canonical Duplex DNA. *Nature* 470 (7335), 498–502. doi:10.1038/nature09775

Peng, C. S., Baiz, C. R., and Tokmakoff, A. (2013). Direct Observation of Ground-State Lactam-Lactim Tautomerization Using Temperature-Jump Transient 2D IR Spectroscopy. *Proc. Natl. Acad. Sci. USA* 110 (23), 9243–9248. doi:10.1073/pnas.1303235110

Peng, C. S., Fedeles, B. I., Singh, V., Li, D., Amariuta, T., Essigmann, J. M., et al. (2015). Two-dimensional IR Spectroscopy of the Anti-HIV Agent KP1212 Reveals Protonated and Neutral Tautomers that Influence pH-dependent Mutagenicity. *Proc. Natl. Acad. Sci. USA* 112 (11), 3229–3234. doi:10.1073/pnas.1415974112

Peng, C. S., Jones, K. C., and Tokmakoff, A. (2011). Anharmonic Vibrational Modes of Nucleic Acid Bases Revealed by 2D IR Spectroscopy. *J. Am. Chem. Soc.* 133 (39), 15650–15660. doi:10.1021/ja205636h

Peng, C. S., and Tokmakoff, A. (2012). Identification of Lactam-Lactim Tautomers of Aromatic Heterocycles in Aqueous Solution Using 2D IR Spectroscopy. *J. Phys. Chem. Lett.* 3 (22), 3302–3306. doi:10.1021/jz301706a

Pinard, R., Hampel, K. J., and Heckman, J. E. (2001). Functional Involvement of G8 in the Hairpin Ribozyme Cleavage Mechanism. *Embo j* 20 (22), 6434–6442. doi:10.1093/emboj/20.22.6434

Rangadurai, A., Szymanski, E. S., Kimsey, I., Shi, H., and Al-Hashimi, H. M. (2020). Probing Conformational Transitions towards Mutagenic Watson-crick-like G·T Mismatches Using Off-Resonance Sugar Carbon R1ρ Relaxation Dispersion. *J. Biomol. NMR* 74 (8-9), 457–471. doi:10.1007/s10858-020-00337-7

Rangadurai, A., Zhou, H., Merriman, D. K., Meiser, N., Liu, B., Shi, H., et al. (2018). Why Are Hoogsteen Base Pairs Energetically Disfavored in A-RNA Compared to B-DNA. *Nucleic Acids Res.* 46 (20), 11099–11114. doi:10.1093/nar/gky885

Rangadurai, A., Szymaski, E. S., Kimsey, I. J., Shi, H., and Al-Hashimi, H. M. (2019). Characterizing Micro-to-millisecond Chemical Exchange in Nucleic Acids Using Off-Resonance R1ρ Relaxation Dispersion. *Prog. Nucl. Magn. Reson. Spectrosc.* 112-113, 55–102. doi:10.1016/j.pnmrs.2019.05.002

Rupert, P. B., and Ferré-D'Amaré, A. R. (2001). Crystal Structure of a Hairpin Ribozyme-Inhibitor Complex with Implications for Catalysis. *Nature* 410 (6830), 780–786. doi:10.1038/35071009

Sevilla, M. D., Besler, B., and Colson, A.-O. (1995). Ab Initio molecular Orbital Calculations of DNA Radical Ions. 5. Scaling of Calculated Electron Affinities and Ionization Potentials to Experimental Values. *J. Phys. Chem.* 99 (3), 1060–1063. doi:10.1021/j100003a032

Shannon, A., Selisko, B., Le, N.-T. -T., Huchting, J., Touret, F., Piorkowski, G., et al. (2020). Rapid Incorporation of Favipiravir by the Fast and Permissive Viral RNA Polymerase Complex Results in SARS-CoV-2 Lethal Mutagenesis. *Nat. Commun.* 11 (1), 4682. doi:10.1038/s41467-020-18463-z

Singh, V., Fedeles, B. I., and Essigmann, J. M. (2015). Role of Tautomerism in RNA Biochemistry. *Rna* 21 (1), 1–13. doi:10.1261/rna.048371.114

Singh, V., Peng, C. S., Li, D., Mitra, K., Silvestre, K. J., Tokmakoff, A., et al. (2014). Direct Observation of Multiple Tautomers of Oxythiamine and Their Recognition by the Thiamine Pyrophosphate Riboswitch. *ACS Chem. Biol.* 9 (1), 227–236. doi:10.1021/cb400581f

Slocombe, L., Al-Khalili, J. S., and Sacchi, M. (2021). Quantum And Classical Effects In DNA Point Mutations: Watson-Crick Tautomerism In AT And GC Base Pairs. *Phy. Chemi.* 23 (7), 4141–4150. doi:10.1039/D0CP05781A

Sood, V. D., and Collins, R. A. (2002). Identification of the Catalytic Subdomain of the VS Ribozyme and Evidence for Remarkable Sequence Tolerance in the Active Site Loop. *J. Mol. Biol.* 320 (3), 443–454. doi:10.1016/s0022-2836(02)00521-1

Suslov, N. B., DasGupta, S., Huang, H., Fuller, J. R., Lilley, D. M. J., Rice, P. A., et al. (2015). Crystal Structure of the Varkud Satellite Ribozyme. *Nat. Chem. Biol.* 11 (11), 840–846. doi:10.1038/nchembio.1929

Szymanski, E. S., Kimsey, I. J., and Al-Hashimi, H. M. (2017). Direct NMR Evidence that Transient Tautomeric and Anionic States in dG·dT Form

Watson-crick-like Base Pairs. *J. Am. Chem. Soc.* 139 (12), 4326–4329. doi:10.1021/jacs.7b01156

Thomas, J. M., and Perrin, D. M. (2008). Probing General Base Catalysis in the Hammerhead Ribozyme. *J. Am. Chem. Soc.* 130 (46), 15467–15475. doi:10.1021/ja804496z

Thore, S., Frick, C., and Ban, N. (2008). Structural Basis of Thiamine Pyrophosphate Analogues Binding to the Eukaryotic Riboswitch. *J. Am. Chem. Soc.* 130 (26), 8116–8117. doi:10.1021/ja801708e

Thore, S., Leibundgut, M., and Ban, N. (2006). Structure of the Eukaryotic Thiamine Pyrophosphate Riboswitch with its Regulatory Ligand. *Science* 312 (5777), 1208–1211. doi:10.1126/science.1128451

Topal, M. D., and Fresco, J. R. (1976). Complementary Base Pairing and the Origin of Substitution Mutations. *Nature* 263 (5575), 285–289. doi:10.1038/263285a0

Velikyan, I., Acharya, S., Trifonova, A., Földesi, A., and Chattopadhyaya, J. (2001). The pKa's of 2'-Hydroxyl Group in Nucleosides and Nucleotides. *J. Am. Chem. Soc.* 123 (12), 2893–2894. doi:10.1021/ja0036312

Wang, W., Hellinga, H. W., and Beese, L. S. (2011). Structural Evidence for the Rare Tautomer Hypothesis of Spontaneous Mutagenesis. *Proc. Natl. Acad. Sci.* 108 (43), 17644–17648. doi:10.1073/pnas.1114496108

Watson, J. D., and Crick, F. H. C. (1953). Genetical Implications of the Structure of Deoxyribonucleic Acid. *Nature* 171 (4361), 964–967. doi:10.1038/171964b0

Wilcox, J. L., and Bevilacqua, P. C. (2013). A Simple Fluorescence Method for pKa Determination in RNA and DNA Reveals Highly Shifted pKa's. *J. Am. Chem. Soc.* 135 (20), 7390–7393. doi:10.1021/ja3125299

Zhou, H., Kimsey, I. J., Nikolova, E. N., Sathyamoorthy, B., Grazioli, G., McSally, J., et al. (2016). m1A and m1G Disrupt A-RNA Structure through the Intrinsic Instability of Hoogsteen Base Pairs. *Nat. Struct. Mol. Biol.* 23 (9), 803–810. doi:10.1038/nsmb.3270

Zhou, S., Hill, C. S., Sarkar, S., Tse, L. V., Woodburn, B. M. D., Schinazi, R. F., et al. (2021). β-DN 4-hydroxycytidine (NHC) Inhibits SARS-CoV-2 through Lethal Mutagenesis but Is Also Mutagenic to Mammalian Cells. *J. Infect. Dis.* 224 (3), 415–419. doi:10.1093/infdis/jiab247

Conflict of Interest: Declaration of interest DL is an author on the patent US9283242B2 (assigned to Massachusetts Institute of Technology), BF and VS are authors on the patent US9714265B2 (assigned to Massachusetts Institute of Technology). These patents describe two classes of mutagenic nucleoside analogs that can adopt multiple tautomeric forms and their potential uses as anti-viral therapeutics.

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

frontiers in Molecular Biosciences

REVIEW
published: 14 February 2022
doi: 10.3389/fmolb.2022.830006

Amyloid Formation by Globular Proteins: The Need to Narrow the Gap Between *in Vitro* and *in Vivo* Mechanisms

*Giulia Faravelli[1], Valentina Mondani[1], P. Patrizia Mangione[1,2], Sara Raimondi[1], Loredana Marchese[1], Francesca Lavatelli[1], Monica Stoppini[1], Alessandra Corazza[3,4], Diana Canetti[2], Guglielmo Verona[2], Laura Obici[5], Graham W. Taylor[2], Julian D. Gillmore[6], Sofia Giorgetti[1,4] and Vittorio Bellotti[1,2,4,7]**

[1]Unit of Biochemistry, Department of Molecular Medicine, University of Pavia, Pavia, Italy, [2]Wolfson Drug Discovery Unit, Division of Medicine, Centre for Amyloidosis and Acute Phase Proteins, University College London, London, United Kingdom, [3]Department of Medicine (DAME), University of Udine, Udine, Italy, [4]Istituto Nazionale Biostrutture e Biosistemi, Rome, Italy, [5]Amyloidosis Research and Treatment Centre, Fondazione IRCCS Policlinico San Matteo, Pavia, Italy, [6]National Amyloidosis Centre, University College London and Royal Free Hospital, London, United Kingdom, [7]Scientific Direction, Fondazione IRCCS Policlinico San Matteo, Pavia, Italy

OPEN ACCESS

Edited by:
Piero Andrea Temussi, University of Naples Federico II, Italy

Reviewed by:
Joost Schymkowitz, VIB & KU Leuven Center for Brain & Disease Research, Belgium
Carlo Camilloni, University of Milan, Italy

***Correspondence:**
Vittorio Bellotti
v.bellotti@ucl.ac.uk
vbellot@unipv.it

Specialty section:
This article was submitted to Structural Biology, a section of the journal Frontiers in Molecular Biosciences

Received: *06 December 2021*
Accepted: *20 January 2022*
Published: *14 February 2022*

Citation:
Faravelli G, Mondani V, Mangione PP, Raimondi S, Marchese L, Lavatelli F, Stoppini M, Corazza A, Canetti D, Verona G, Obici L, Taylor GW, Gillmore JD, Giorgetti S and Bellotti V (2022) Amyloid Formation by Globular Proteins: The Need to Narrow the Gap Between in Vitro *and* in Vivo *Mechanisms. Front. Mol. Biosci. 9:830006. doi: 10.3389/fmolb.2022.830006*

The globular to fibrillar transition of proteins represents a key pathogenic event in the development of amyloid diseases. Although systemic amyloidoses share the common characteristic of amyloid deposition in the extracellular matrix, they are clinically heterogeneous as the affected organs may vary. The observation that precursors of amyloid fibrils derived from circulating globular plasma proteins led to huge efforts in trying to elucidate the structural events determining the protein metamorphosis from their globular to fibrillar state. Whereas the process of metamorphosis has inspired poets and writers from Ovid to Kafka, protein metamorphism is a more recent concept. It is an ideal metaphor in biochemistry for studying the protein folding paradigm and investigating determinants of folding dynamics. Although we have learned how to transform both normal and pathogenic globular proteins into fibrillar polymers *in vitro*, the events occurring *in vivo*, are far more complex and yet to be explained. A major gap still exists between *in vivo* and *in vitro* models of fibrillogenesis as the biological complexity of the disease in living organisms cannot be reproduced at the same extent in the test tube. Reviewing the major scientific attempts to monitor the amyloidogenic metamorphosis of globular proteins in systems of increasing complexity, from cell culture to human tissues, may help to bridge the gap between the experimental models and the actual pathological events in patients.

Keywords: amyloidosis, transthyretin, β2-microglobulin, metamorphosis, amyloid

INTRODUCTION

Metamorphosis is a process of transformation of an entity into a different form, acquiring a different identity. Although the two entities are different, their basic substance is identical: the caterpillar and butterfly have the same genome, and the humanity of "Gregor Samsa" is unchanged whether caged in human form or in the semblance of a beetle. In the same way, proteins with the same primary sequence (substance) can assume two completely different shapes. Metamorphosis is an intrinsic

property of living creatures. It represents a fascinating phenomenon that has extraordinary evolutionary implications which increase the adaptation of living organisms to different environments and reduce competition for resources (Laudet, 2011). Entire generations of biologists have investigated this phenomenon which has, at the same time, stimulated imagination, dreams and artistic creativity. Metamorphosis has always inspired human imagination raising ambivalent sentiments from fear to amazement. Humankind have viewed the metamorphosis as an awe-inspiring biologic phenomenon (from the caterpillar to butterfly) (**Figure 1A**) or as a terrible nightmare (in the dream of being transformed into a horrible insect) (**Figure 1B**). More recently, the terms metamorphosis or metamorphism have been associated to the properties of particular proteins that may acquire multiple functions, such as moonlighting proteins (Dishman and Volkman, 2018) (http://www.moonlightingproteins.org), or exhibit different physiological and pathologic roles (**Figure 1C**). The prion protein represents an example of protein metamorphism in which the same protein sequence can adopt both a native globular fold (PrP^{C}) and a self-propagating infectious one (PrP^{Sc}) which is responsible for transmissible spongiform encephalopathies. The first structure of PrP^{C} was obtained by NMR (Riek et al., 1996) and shows a predominant alpha helical content (**Figure 2**). A model based on molecular dynamics simulations suggests that a high proportion of beta structures may characterize the conformation of the infectious prion. The mechanism by which the contact between the two isoforms triggers the conformational shift PrP^{C} to PrP^{SC} is not fully clarified yet, however specific contact sites suitable for triggering such a transition have been extensively studied (Brunori, 2021). A recent cryo-EM structure of a patient-derived amyloid fibril of PrP has provided highly resolved constraints that may shed light on a future structure-based understanding of prion biology (Kraus et al., 2021).

AMYLOIDOSIS IS AN AGING ASSOCIATED DISEASE

Whilst systemic amyloidosis has traditionally been associated with the presence of destabilized protein variants, such as lysozyme or transthyretin (TTR), or markedly increased plasma protein concentrations (SAA) (Merlini and Bellotti, 2003), epidemiological studies have shown that there is also an association of amyloid with aging (Blumenthal, 2004; Trigo et al., 2019; Tasaki et al., 2021) (**Table 1**). This is particularly true for amyloid diseases caused by normal circulating levels of wild type proteins such as TTR (Nativi-Nicolau et al., 2021). In the last decade, the comparative analysis of hereditary versus non-hereditary amyloidosis has represented an extraordinary source of information for the intrinsic amyloidogenic properties of these proteins. The age of onset of the genetic forms is, most often, much earlier compared with those caused by wild type proteins thus revealing that single residue changes, even if hidden within the protein structure, can strongly enhance the amyloidogenic propensity of the protein (**Table 1**). The metamorphic property of some globular proteins is particularly intriguing in cases where the whole protein can be converted into the amyloid fibrils, presenting two totally different shapes despite a completely identical sequence. Among globular proteins, investigated for their metamorphic and pathogenic behavior, we have extensively studied human β2-microglobulin (β2-m) and TTR. We believe that the wealth of information acquired on structure and folding dynamics of both proteins is extremely informative on the generic molecular basis of protein amyloidogenesis. Both proteins are assembled in the early phase of their life into their quaternary structure: β2-m as a light chain of the major antigen of histocompatibility type I (MHCI) and TTR as a homotetramer which forms a complex with the retinol binding protein (RBP). Both proteins can cause amyloid in either their wild type form or in their genetic variants associated with the familial forms of the disease.

GENETIC VERSUS ACQUIRED B2-M RELATED AMYLOIDOSIS

In 1985 the Japanese group of Prof Gejyo identified β2-m as the protein responsible for a peculiar form of amyloidosis (Gejyo et al., 1985) observed in the musculoskeletal system of patients under long term haemodialysis (Assenat et al., 1980). In patients under chronic haemodialysis, β2-m circulates at a concentration 5–20 fold higher that in normal subjects due to the limited efficacy of dialytic membranes to efficiently remove β2-m which, upon dissociation of MHCI, is continuously released from the surface of cells into circulation. The persistently high concentration of plasma β2-m is an essential requirement for the formation of amyloid *in vivo*, but other factors can modulate individual susceptibility to form amyloid deposits. Improved biocompatibility of dialytic membranes can reduce the inflammatory response and delay the onset of amyloid deposition. The physiological state of patients, and particularly their age, can affect their individual susceptibility to amyloid complication in hemodialysis. Van Ipersele de Strihou and his group has clearly shown that the risk of developing amyloid complications exponentially increases with the age of the patient at the onset of the dialysis treatment (**Figure 3**) (de Strihou et al., 1991). This finding is important because Dialysis Related Amyloidosis (DRA) is probably the only amyloid disease in which we can exactly establish the time (the first day of haemodialysis) in which a pro-amyloidogenic phase is triggered by simply increasing the concentration of plasma β2-m. In the last few decades, the intermediate steps along the pathway of β2-m amyloid formation *in vitro* have been identified (Michaels et al., 2018). It is plausible that primary nucleation, fragmentation, surface-catalyzed secondary nucleation, elongation and dissociation of fibrils can determine the rate of amyloid growth *in vivo*, but it is not known how the particular biological environment, within a living organism, could affect these intermediate steps. Even more challenging is to elucidate how aging can influence the rate of the intermediate steps of fibrillogenesis *in vivo*. There are several working hypotheses on the role of aging in amyloidogenesis:

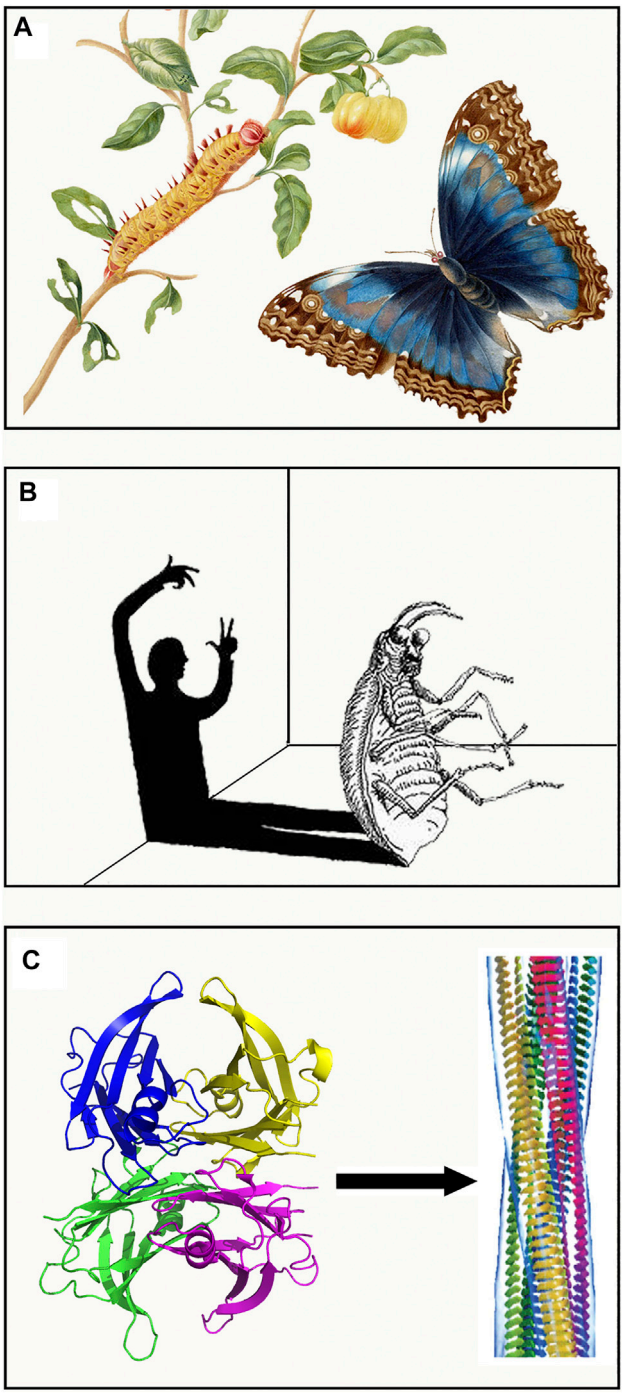

FIGURE 1 | Methamorphosis in nature and literature. **(A)** The caterpillar and butterfly captured by the scientific illustrator and naturalist Maria Sibylla Merian (1647–1717). **(B)** Representation of the nightmare in which Gregory Samsa is being transformed from human to a beetle. **(C)** The conversion of native tetrameric TTR into amyloid fibrillar structures marks the protein transition from a physiological to a pathologic role. Both artworks in **(A,B)** are in the public domain. Structure of amyloid fibrils from (Chiti and Dobson, 2017) are reprinted with permission from Annual Reviews, Copyright (2017).

malfunction of proteostasis (Trigo et al., 2019), modification of the extracellular matrix (Morawski et al., 2014) and, impairment in the bio-energetic efficiency (Swerdlow, 2018) are among the most accredited theories. The molecular transition of β2-m from its native to fibrillar conformation suggests that partial unfolding is required to prime aggregation. Early studies were carried out *in vitro* at a low pH revealing that fibrillogenesis of β2-m requires the unfolding of a large portion of the protein. Conditions used in the early studies to generate *in vitro* amyloid fibrils, both in the case of β2-m and, in general, for almost all the globular proteins including lysozyme, the archetypal protein of these studies (Booth et al., 1997), were not compatible with the physicochemical characteristics of the biologic environment even those of an aged organism. However, there was evidence that natural fibrils contained not only the full length β2-m, but also a significant proportion of a truncated species lacking 6 residues at N-terminal end (Bellotti et al., 1998). Further studies confirmed that the N-terminal truncation reduced the β2-m folding stability and generated fibrils even under a more physiological environment (Esposito et al., 2000). The metamorphism of the protein was somehow triggered by a selective proteolytic cleavage. This suggests that proteolysis-mediated remodeling may either be a physiological mechanism for priming β2-m degradation, or a purely a pathologic processing event enhanced in hemodialysis and, in particular, in elderly patients. The question remains unanswered due to the complexity of the system. Indeed, to prevent the adverse effects of dialysis related amyloidosis, more efforts have been devoted to ameliorate hemodialytic procedures rather than to discover drugs targeting the key steps of fibril formation. Our experience suggests that curiosity-driven programs in the biomedical field, more and more often, compete with translational projects that aim to solve medical needs diminishing the resources and efforts to elucidate the biological mechanisms underlying the disease and discover new therapeutic targets. Nevertheless, seminal observations have enabled advances in the field.

The observation that minimal truncation of β2-m is sufficient to prime β2-m fibrillogenesis in bio-compatible conditions was further strengthened by the demonstration that, in a physiological environment, truncated β2-m can play a prion like effect on full length β2 (Karamanos et al., 2014). The mechanism requires the formation of fibrillar seeds by the truncated species thus generating a suitable fibrillar template which triggers a chain polymerization mix of non-covalently linked oligomers of full length and truncated species, both acquiring an "amyloid conformation" (Natalello et al., 2016). Discovery of the first genetic variant of β2-microglobulin (Valleix et al., 2012) confirmed that destabilization of the tertiary structure is a generic prerequisite for amyloid transformation and that fibrillar seeds have a potent, potential, role in the propagation of amyloid formation. We say "potential role" because, although D76N variant was clearly shown to be a potent driver for the conversion of wild type β2-m into an amyloid conformation (Mangione et al., 2013), we were unable to demonstrate the co-deposition of the wild type within the natural D76N fibrils

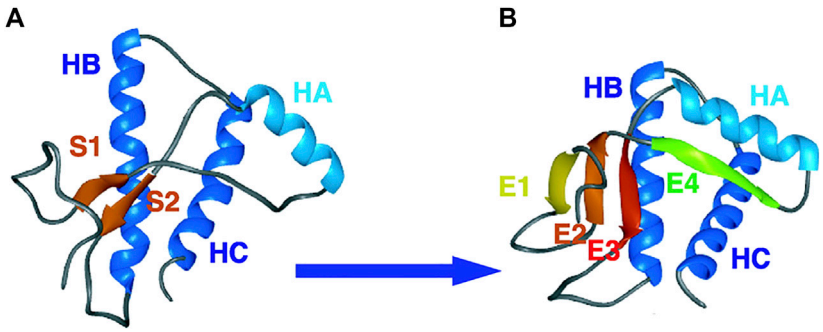

FIGURE 2 | Simulated conversion of Syrian hamster D147N PrP^{C} to PrP^{Sc} at low pH levels. The wild-type NMR structure is shown **(A)** with the helices and strands labeled. A representative PrP^{Sc}-like structure is shown **(B)**. Adapted from (DeMarco and Daggett, 2004), Copyright (2004) National Academy of Sciences, U.S.A.

TABLE 1 | Proteins that can cause amyloid deposits in humans either in their wild type form or in presence of mutations.

References	Fibril protein	Precursor protein	Age at onset	Acquired (A) or hereditary (H)	Target organs (clinically manifested)
(Merlini et al., 2018; Muchtar et al., 2019; Benson et al., 2020; Muchtar et al., 2021)	AL	Immunoglobulin light chain, monoclonal	Median 63 years	A	All organs, except CNS
Benson et al. (2015)		Immunoglobulin κ light chain (C_L domain)	n.a. (median age of end-stage renal disease 60 years)	H	Kidney
(Grogan et al., 2016; Muchtar et al., 2021; Nativi-Nicolau et al., 2021)	ATTR	TTR, wild type	Median age 70 years	A	Heart mainly in males, lung, ligaments, tenosynovium
Barroso et al. (2015)		TTR, variants	Mutation-dependent (median 39 years)	H	PNS, ANS, heart, eye, leptomeninges
(Gejyo et al., 1985; Stoppini and Bellotti, 2015)	β2-m	β2-m, wild type	n.a. (depending on dialysis onset)	A	Musculoskeletal system
(Valleix et al., 2012; Mizuno et al., 2021)		β2-m, variant(s)	Median age 50 years	H	GI, ANS, tongue, salivary glands
Koedam et al., (2010)	Aβ	Aβ protein precursor, wild type	>65 years	A	CNS
Latimer et al. (2021)		Aβ protein precursor, variant	<65 years (30 years–40 years)	H	CNS
Puoti et al. (2012)	APrP	Prion protein, wild type	55–75 years	A	CJD, fatal insomnia
Schmitz et al. (2017)		Prion protein variants	mutation-dependent (30–70 years)	H	CJD, GSS syndrome, fatal insomnia
(Mucchiano et al., 2001; Tasaki et al., 2021)	AApoAI	ApoAI, wild type[a]	Aging-associated	A	Aortic intima, carotid artery, meniscus with knee osteoarthritis, lumbar ligamentum flavum
Muchtar et al. (2021)		ApoAI, variants	Mutation-dependent (20–70 years)	H	Heart, liver, kidney, PNS, testis, larynx (C terminal variants), skin (C terminal variants)

C_L, light chain's constant domain; ApoAI, Apolipoprotein AI; GI, Gastrointestinal Tract; CNS, Central Nervous System; PNS, Peripheral Nervous System; ANS, Autonomic Nervous System; CJD, Creutzfeldt Jakob Disease; GSS, Gerstmann-Sträussler-Scheinker syndrome; n.a. not available.

[a]*ApoAI WT is not listed as an amyloidogenic protein in the Amyloid nomenclature 2020 (Benson et al., 2020).*

extracted from the spleen of one patient. A second genetically transmitted amyloidogenic variant of β2-m (V27M) has been recently described (Mizuno et al., 2021). In this case, the *in vitro* studies suggest a striking similarity with the biophysical properties of the D76N variant, both showing a similar level of destabilization in comparison to the wild type, and both forming amyloid fibrils in physiological buffer and with similar morphology when analyzed by electron microscopy. However, the natural fibrils extracted from the hip joint synovium contained a combination of 25% wild type and 75% V27M variant. This suggests that the "prion like" effect, in which the variant acts as primer to convert the wild type protein into fibrils, might be different, depending on the specific isoform and type of tissue that may offer more or less protection to co-propagate heterogeneous isoforms. Characterization of natural fibrils derived from amyloidogenic β2-m variants with similar

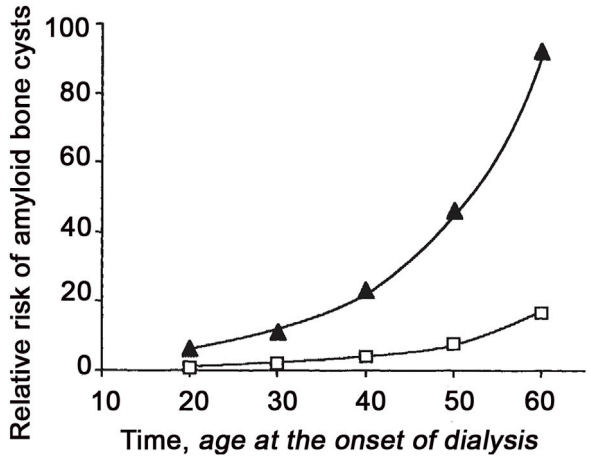

FIGURE 3 | Relative risk of bone lesions by amyloid deposition as a function of age at the initiation of dialysis either on AN69 or a cellulosic membrane. The risk of developing the lesion with AN69 for patients starting at age 20 is assumed to be one. Data are based on the best fitted model provided by Van Ypersele de Strihou work (de Strihou et al., 1991). Reprinted from (de Strihou et al., 1991) with permission from Elsevier, Copyright (1991).

biophysical properties, isolated in subjects with different ethnicity and, affected by two different types of syndromes, dramatically reveal the complexity of the phenomenon in which multiple factors play a pathogenic role.

ANIMAL MODELS OF B2-M

Several experimental models have been developed to investigate the molecular mechanisms of amyloidogenesis by human proteins *in vivo*, ranging from simple organisms such as *C. elegans*, to mammalians, such as mice and, even primates in rare cases. Our capacity to reproduce the clinical effects of β2-m in animal models, remains still a challenging task.

Particularly informative is the failure to create a mouse model of β2-m amyloidosis which was reported by Higuchi's group (Zhang et al., 2010) in 2010. Despite an overexpression of human β2-m in a murine β2-m knockout background, no β2-m related amyloid was formed in those mice even at their later stage. Finally, the amyloid observed was caused by Apolipoprotein A2 (ApoA2) which spontaneously forms amyloid in the specific strains of mice used in that laboratory. Even the attempt to induce amyloid deposition by injecting seeds of natural human fibrils failed. Failure to induce β2-m amyloidosis in transgenic mice expressing the D76N β2-m variant (Halabelian et al., 2014) was even more surprising and, very informative at the same time, as this protein is highly amyloidogenic both *in vivo* and *in vitro* (Valleix et al., 2012). The use of strains knocked out for murine β2-m excludes a protective role of the endogenous β2-m that otherwise is known to interfere with the aggregation of the human counterpart (Eichner and Radford, 2011; Karamanos et al., 2014; Achour et al., 2020). The specific protection of rodents against β2-m related amyloidosis is due to unidentified factors. Although mice can develop some specific forms of murine amyloidosis (such as AAPOA2), either spontaneously or upon stimulation, they are extremely resistant to other types of systemic amyloidosis, and in particular, to the type caused by immunoglobulin light chains. The only mouse models in which amyloid deposition is easily and consistently achieved, under experimental conditions, are those related to HDL related lipoproteins. Persistently high concentration of serum Amyloid A (SAA), in the presence (McAdam and Sipe, 1976) or even in the absence of chronic inflammation (Simons et al., 2013), cause AA amyloidosis. On the other hand, ApoA2 causes a senescence-accelerated and spontaneous form of amyloidosis in which the rate of deposition depends on specific polymorphisms of the protein (Higuchi et al., 1991). Why mice are susceptible to some types of systemic amyloidosis and totally resistant to others is not understood, and highlights the complexity of the mechanism of systemic deposition of amyloid in a living organism and, in particular, in vertebrates. Regarding systems of amyloidosis in invertebrates, *C. elegans* is an extremely interesting and versatile model for the study of natural aging in a living organism (Mack et al., 2018). Pathogenic effects of human soluble and oligomeric species of β2-m have extensively reported for wild type species as well as for the clinically pathogenic variants. Although worm models cannot recapitulate the clinical and pathologic complexity of the human disease, they can certainly highlight and single out specific steps within the process. *C. elegans* was specifically chosen to represent the pathophysiology of β2-m related amyloidosis because, as with all the invertebrates, it lacks the MHCI complex and therefore β2-m is expressed and secreted as free monomer. In the first models of *C. elegans* expressing human β2-m, effects of different isoforms of the protein on survival and performance of different transgenic strains were described (Diomede et al., 2012). At a time in which the pathogenic variants of β2-m were not known yet, worms expressing the natural truncated form of wild type β2-m (ΔN6β2-m), were generated. It was shown that ΔN6β2-m was not only highly amyloidogenic *in vitro* (Esposito et al., 2000), but it had also a strong tendency to form oligomers in the worms affecting both their survival and movement performance.

It was not possible to demonstrate the formation of genuine amyloid fibrils, but the evidence of a pathologic phenotype associated to β2-m aggregation and toxicity, made the *C. elegans* model the only available *in vivo* system to study the putative pathogenic property of β2-m. The discovery of the first natural amyloidogenic variant of β2-m in members of a French family (Valleix et al., 2012) and, the demonstration of its higher propensity to make fibrils *in vitro* compared to the wild type, inspired the generation of a *C. elegans* strain expressing the D76N β2-m variant. Interestingly, the protein appeared to be highly toxic for the worms which died before reaching their first larval stage. The only way for them to reach adulthood was achieved by using a temperature-inducible system to switch-on the D76N β2-m expression only at their larval phase. The reason of such a high toxicity, incompatible with the worm survival in the larval phase is unknown; it represents a paradox in the context of a disease which is known to be strictly associated to aging, rather than to a very early phase of development.

A nematode model, established in another laboratory has fully validated previous data and further demonstrated that the

expression of the highly amyloidogenic D76N β2-m and ΔN6 β2-m variants can cause global aggregation of endogenous *C. elegans* proteins at an advanced stage (Good et al., 2021). Disrupted ER homeostasis, combined with increased endogenous protein aggregation, were identified as drivers of β2-m associated toxicity *in vivo*. All the *C. elegans* β2-m models so far established display a reduced lifespan, impaired motility and developmental delays. Studies on the nematode models have also highlighted their importance as drug search tools. Use of a generic inhibitor of aggregation of amyloidogenic proteins, doxycycline, in the *C. elegans* strain expressing D76N β2m showed that the drug was able to rescue the adverse phenotype most likely by interfering with the formation of oligomeric conformers (Faravelli et al., 2019).

GENETIC VERSUS ACQUIRED TTR AMYLOIDOSIS

In 1978 the Portuguese group led by Pedro Costa demonstrated that Familial Amyloid Polyneuropathy (Andrade, 1952) was caused by the protein transthyretin (Costa et al., 1978) through a classical mechanism of autosomal dominant transmission (Saraiva, 2015). Histologically, TTR amyloid has been detected in all organs with the exception of brain parenchyma. After the discovery of the first amyloid associated mutation (V30M), more than one hundred amyloidogenic mutations have been identified (http://amyloidosismutations.com). Clinical pathologic features are well described with regard to the prominent involvement of one or multiple organs. Often the disease begins apparently with a single organ involvement and, in most cases, a systemic deposition of amyloid can further be observed. For example, following initial infiltration in the peripheral nerves with consequent polyneuropathy, amyloid deposition is frequently observed in the heart. There is a correlation between the type of single amino acid variant and the clinical phenotype. Some mutations cause a prominently or even an exclusively cardiac amyloidosis, whilst other mutations are primarily responsible for a polyneuropathic syndrome (**Figure 4**). Tissue specificity in systemic TTR amyloidosis sounds like a contradiction since it would be more appropriate to say that amyloid accumulates preferentially in some organs. In 1980 Sletten and Swedish colleagues (Sletten et al., 1980) showed that wild type transthyretin was also amyloidogenic and responsible for an already known non-familial form of cardiac amyloidosis mostly affecting the elderly population. Development of modern imaging techniques (Jurcuț et al., 2020) has enabled accurate and early identification of cardiac TTR amyloid leading to a highly significant increase in the diagnosis of this disease. It is estimated that 25% of people over 75 are likely to have cardiac deposition of TTR amyloid with a consequent high probability of developing heart failure with preserved ejection fraction (HFpEF) (Grogan et al., 2016). The number of elderly patients affected by wild type TTR amyloidosis (**Figure 5**) suggests that the category of TTR amyloidosis is moving from a rare to a common disease. The urgent need to understand the pathogenic mechanism and establish effective therapies can now benefit from the extensive investigation carried out in the last decades on the rare amyloidogenic variants associated to the familial forms of the disease. In the early nineties, the Kelly's group made the seminal observation that tetramer stability of TTR was impaired by the amyloidogenic mutations (Hammarström et al., 2002) and that the tetramer disassembly was a perquisite for the aggregation of TTR into an amyloid-like aggregates (Colon and Kelly, 1992). These observations were consistent with those made by the Wetzel's group on amyloidogenic immunoglobulin light chain (Hurle et al., 1994), and the Pepys's team on amyloidogenic lysozyme (Booth et al., 1997). The latter study highlighted the possibility of identifying a conformational intermediate along the folding pathway which was particularly susceptible to polymerization into a well-ordered cross-beta structure (**Figure 6**). After many decades of studies on the structural intermediates of the amyloidogenic pathway of TTR, we are still debating on the structure of misfolded monomers which lead to TTR fibrillogenesis. Nature offered an additional extraordinary tool for proving the causative role of folding destabilization in protein amyloidogenesis. Clinicians in Portugal observed that a small proportion of carriers of the amyloidogenic V30M mutation were protected from the disease and that the protection was provided by the concomitant T119M mutation in the TTR gene so that the resulting tetrameric protein was a mix of protomers presenting both the amyloidogenic V30M variant as well as the protective T119M. Demonstration that the presence of protomers with T119M has a stabilizing effect on the tetramer is a strong evidence in favor of the hypothesis that misfolding is at the basis of the disease. This widely accepted concept strongly encouraged the therapeutic strategy to stabilize TTR by small molecules. Colin Blake first showed that halogenated aromatic hydrocarbon can bind with high affinity to the thyroxine (T4)

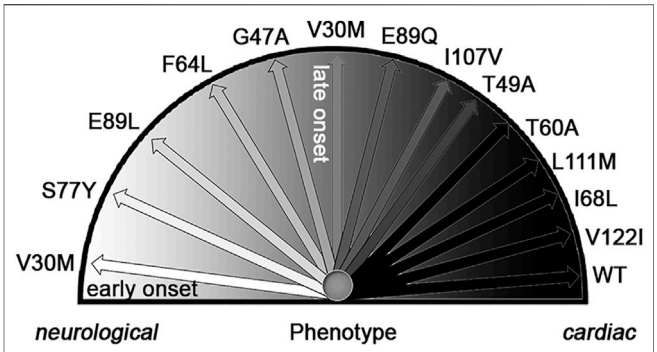

FIGURE 4 | Genotype-phenotype correlation in ATTR amyloidosis. Adapted from **Figure 2** in a previous publication by Rapezzi and collaborators (Rapezzi et al., 2013). Phenotypic expression of transthyretin-related amyloidosis varies widely from an almost exclusively neurological involvement (V30M mutation with early-onset disease) to a predominant or exclusively cardiac involvement (T60A, L111M, I68L, and V22I mutations and wild type). In between, several transthyretin-related amyloidosis mutations are associated with variable degrees of neurological and cardiac involvement, including V30M with late-onset disease. Adapted with permission by Oxford University Press, Copyright (2012).

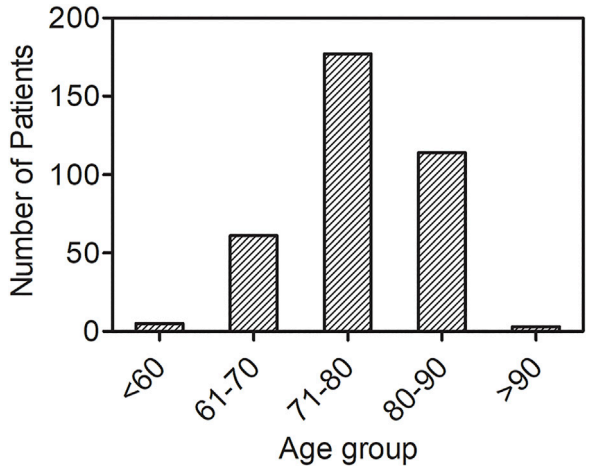

FIGURE 5 | Number of patients diagnosed, *ante mortem,* for wild type ATTR. Data shown are based on a retrospective review of patients diagnosed with wild type ATTR at the Mayo Clinic through 2013 with the purpose of studying the natural history of the disease (Grogan et al., 2016).

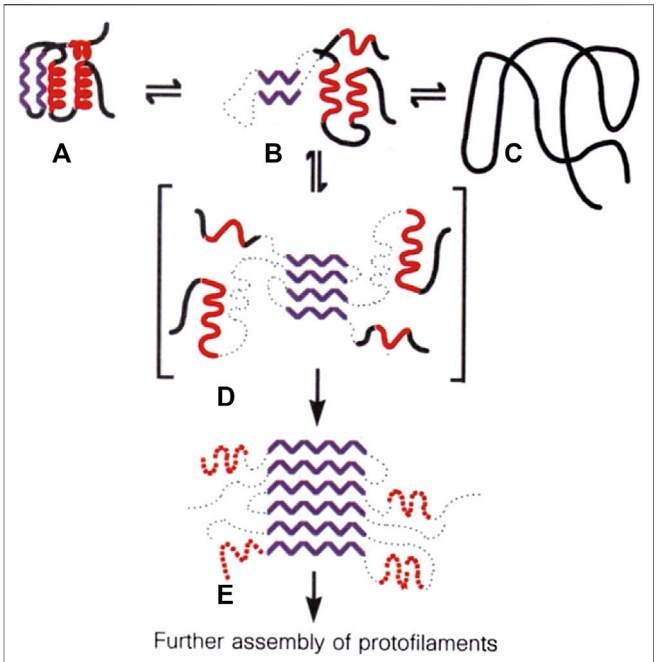

FIGURE 6 | Proposed mechanism for lysozyme amyloid fibril formation. A molten globule form of the protein **(B)**, distinct from the native **(A)** and denatured **(C)** states, self-associates through the β-domain **(D)** to initiate fibril formation and form the core structure of the fibril **(E)**. The scheme was included in a previous manuscript published in Nature, 1997 (Booth et al., 1997).

binding pocket of TTR thus designing the first study of the structurally-based design of TTR ligands (Rickenbacher et al., 1986). T4 and analogues became instrumental to the discovery of druggable stabilizers of the native tetrameric structure (Connelly et al., 2010) thus mimicking the enhanced stability observed in the presence of the T119M mutation. After almost 30 years, the approach of stabilizing the tetramer by a ligand has become a clinical reality. Tafamidis is now in the market for the treatment of TTR related amyloidoses and other compounds are in clinical trials such as acoramidis (Judge et al., 2019) or in the pipeline for being tested in clinical studies (Corazza et al., 2019). In the field of amyloidoses, it represents the most advanced experience of tackling disease by directly interfering with the misfolding mechanism. The discovery of drugs improving the stability of TTR highlights the importance of understanding more about the forces that can determine the TTR tetramer disassembly *in vivo*. Although it is well-known that destabilization is a pre-requisite for priming the amyloid pathway, little is known about the TTR misfolding energy landscape (Hartl and Hayer-Hartl, 2009) (**Figure 7**). Our efforts in looking for this source of energy were rewarded when we found out that biomechanical forces can provide sufficient energy to unfold the amyloidogenic proteins. Interestingly, these results represent a typical case of cross-fertilization between studies carried out in parallel on β2-m and TTR. The first observation on the capacity of biomechanical forces to misfold an amyloidogenic protein was reported for the characterization of the first pathogenic variant of β2-m. Later, this concept was applied to TTR amyloidogenesis, and, in this case, we discovered that local destabilization makes the protein more susceptible to cleavage by specific proteases. Proteolytic remodeling of protein exposed to biomechanical forces is a well-established process of great functional impact. The best example of mechano-enzymatic mechanism was described for the Von Willebrand factor (Zhang et al., 2009), where altering the protein structure *in vivo* can lead to both a positive and negative effect. Indeed, a specific genetic mutation either of the von Willebrand gene or of its specific protease, ADAMTS13 (Crawley et al., 2011), can respectively reduce or enhance the proteolytic activity. The system shifts from a sophisticated physiological control of hemostasis into a thrombotic or hemorrhagic syndrome. The mechano-enzymatic mechanism of TTR cleavage that our group has characterized in an *in vitro* and in an *in vivo* model of amyloidogenesis may be crucial in the pathogenesis of TTR related amyloidosis and, in particular, for the deposition of amyloid in organs where the biomechanical forces are particularly intense such as the heart and carpal tunnel. However, it is not known if it may have any physiologic role in the normal TTR catabolism.

ANIMAL MODELS OF TTR

More than 10 years after the discovery that the V30M variant TTR was the causative protein of familial amyloid polyneuropathy, the first mouse model of TTR amyloidosis was published by the Araki's group (Yi et al., 1991). Despite the presence of amyloid in the very old mice, the attempt to reproduce the clinical and pathologic features of amyloid polyneuropathy similar to the human counterpart was unsuccessful. This first model was followed by many other

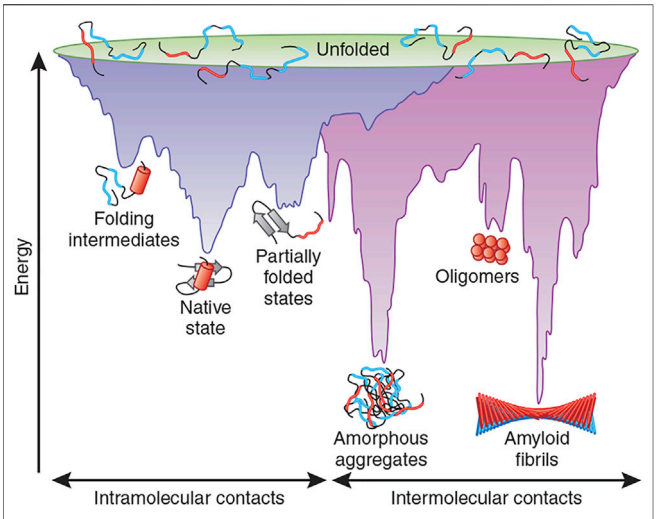

FIGURE 7 | Energy landscape scheme of protein folding and aggregation. The scheme shows multiple conformations "funneling" to the native state *via* intramolecular contacts (purple surface) as well as conformations moving toward amorphous aggregates or amyloid fibrils *via* intermolecular contacts (pink area). Reprinted from (Hartl and Hayer-Hartl, 2009) with permission from Springer Nature Customer Service Centre GmbH, Springer Nature, Copyright (2009).

transgenic strains extensively reviewed by Buxbaum in 2009 (Buxbaum, 2009). Mice expressing other human TTR variants were explored and, in particular, the highly unstable and clinically very aggressive L55P, and the I84S in which the mutation prevents the formation of the complex TTR-RBP. It is worth noting that only one third of mice expressing the L55P variant in a null-TTR background presented amyloid deposition (Sousa et al., 2002). When mice expressing V30M were crossed with those lacking the main heat shock transcription factor (Hsf1), Santos and collaborators observed amyloid deposits in mice at younger age with TTR deposits in sciatic nerve and dorsal root ganglia thus showing a striking similarity with the human disease (Santos et al., 2010). The mechanism by which Hsf1 can protect from amyloid deposition is still uncertain because this transcription factor has multiple beneficial effect on proteostasis, either directly inhibiting the TTR aggregation through specific chaperones or protecting the target tissue from amyloid toxicity. This mouse model, however, represents the first successful example of the *in vivo* multifactorial modulation of TTR amyloidosis. The complexity of the process of *in vivo* amyloidogenesis is highlighted by other two mouse models more recently described. Li and colleagues (Li et al., 2018) have successfully produced humanized mouse strains at both the *Ttr* and *Rbp4* loci, allowing human TTR (hTTR) to associate with human retinol binding protein 4 (hRPP4). Remarkably, these animals develop age-related amyloid deposits mainly in the peripheral nerves and gastro-intestinal tract. This model narrows the gap between experimental and clinical neuropathology since the deposition of amyloid affects the perineurium of sciatic nerve that was only previously achieved in TTR transgenic mice which did not express Hsf1 (Santos et al., 2010). The role of the V30M pathogenic mutation on amyloid formation was not so remarkable. In fact, apparently the onset of amyloid deposition was more anticipated in the wild type (V30) than in the variant (M30) strain, however, it was more abundant in elderly mice carrying the TTR variant (M30). All those mouse models developed a very limited amount of amyloid in the heart in which non-fibrillar amorphous aggregates of TTR were mostly observed. The deposition in the heart has been very recently obtained in a new transgenic mouse strain expressing the human S52P TTR variant in which the formation of amyloid requires pre-inoculation with seeds of natural amyloid fibrils (Slamova

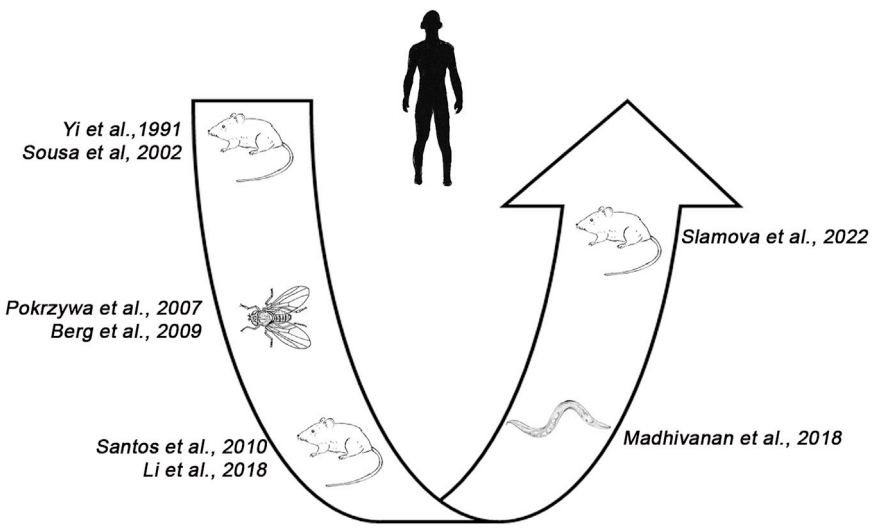

FIGURE 8 | Development of animal models for ATTR in humans. From the first mouse model (Yi et al., 1991), several attempts were made to reproduce the disease in invertebrates (i.e., *C. elegans*, *Drosophila melanogaster*) before developing a novel transgenic mouse model reproducing amyloid in the heart and tongue (Slamova et al., 2021).

et al., 2021). Here, the time frame of the onset of amyloid deposition and the amount of deposits are strictly dependent on the plasma concentration of TTR and, the amyloid mainly affects the heart and tongue. Interestingly, inhibition of the expression of α2-antiplasmin significantly anticipates the amyloid formation and aggravates the amyloid load, thus highlighting the possible pathogenic role of plasmin (Mangione et al., 2018).

In parallel to the development of disease models in mice, other interesting systems of TTR expression have been recently generated using *C. elegans* and *Drosophila melanogaster*. Nematode strains expressing wild type TTR, V30M TTR, D18G TTR and the non-aggregation-prone T119M TTR in the body-wall muscle have been generated (Madhivanan et al., 2018); among these, the ATTR V30M and D18G mutants displayed the accumulation of insoluble forms of TTR. Proteotoxicity, caused by each of the pathogenic TTR mutations, was monitored in the very short time frame of the worm's life (3 weeks). In particular, the expression of TTR variants resulted in defects in neuronal morphology, nociceptive response and locomotion of the nematodes. Tsuda et al. (2018) have confirmed that *C. elegans* is an excellent tool to investigate toxicity of TTR conformers with particular regard to truncated species which are abundant in natural fibrils (Bergstrom et al., 2005; Ihse et al., 2013) and highly fibrillogenic *in vitro* (Mangione et al., 2014). The transgenic strains expressing the 49–127 and 81–127 TTR fragments showed a reduced motility and a significantly shortened lifespan compared with those expressing the full-length protein.

Transgenic flies have been used to study familial amyloidotic polyneuropathy. *Drosophila* expressing the amyloidogenic L55P TTR and the engineered variant TTR-A (V14N/V16E TTR) have been generated and compared with flies expressing wild type TTR. Both variants exhibited time-dependent aggregation of misfolded TTR, neurodegeneration, shortened lifespan, and compromised flying ability (Pokrzywa et al., 2007). In contrast, flies expressing wild type TTR did not show any effect on either longevity or behavior. Expression of the V30M TTR variant in flies led to similar outcomes associated to neurological impairment with evidence of Congo red positive deposits in the brain of the aged transgenic flies (Berg et al., 2009). Studies conducted on transgenic strains of *drosophila* led also to the identification of peptides as potential inhibitors of TTR aggregation as they were able to delay ATTR progression in the flies whilst decreasing TTR deposition (Saelices et al., 2018).

Looking at the development of animal models for TTR related amyloidosis over the years, (**Figure 8**) it appears that we moved from complicated living organisms (i.e., mice) to quite simpler invertebrate models, such as *C. elegans* and *Drosophila* and, finally back to mice.

Nevertheless, we must also mention the extraordinarily informative primate counterpart of human disease, that cannot be considered a model. It is actually a disease that affects other species very close to humans in which aging and individual susceptibility play similar roles (Nakamura et al., 2008; Chambers et al., 2010; Ueda et al., 2012).

CONCLUSION

Despite extraordinary progress made in the last 3 decades on the theoretical and experimental elucidation of the molecular basis of metamorphogenesis of amyloidogenic proteins, we still do not understand which are the major molecular events that *in vivo* dictate the transformation of globular proteins into fibrils. The progress in the field has resulted in a challenge to Anfinsen's paradigm that, under physiological conditions, a protein's final structure is only determined by its primary sequence.

β2-m and TTR (**Figure 9**) exemplify the need for new paradigms. Along their folding pathway, both these proteins may generate globular or fibrillar structures. Despite their substantial differences in both their primary structure and native state conformation, they reach a similar fibrillar structure at the end of their pathologic transformation. We do not know when the different conformers start moving towards a common β-sheet pathological organization. Elucidation of structure and dynamics of the intermediate states populating the globular-fibrillar transition is of great value. Work on highly amyloidogenic truncated isoforms of β2-m and TTR is very promising and potentially supported by new systems of 3D prediction. However, the attempt to model the conformation of amyloidogenic fragments using AlphaFold2 leads to structures that differ from those based on molecular dynamics or experimental NMR restraints. In the predicted models, a higher proportion of the full-length parent protein is retained possibly because of the use of co-evolutionary methods in AlphaFold2 that may represent a major limitation in predicting the structure of protein aggregation intermediates (Pinheiro et al., 2021). Many proteins can be transformed into amyloid fibrils *in vitro*, but only a few make deposits *in vivo* causing amyloid disease (Chiti and Dobson, 2017). We must recognize that there is still a large gap between what we know about protein amyloidogenesis in experimental models and what happens in patients. Clinical, pathologic studies mirrored by the study of experimental models, which here we have summarized for β2-m and TTR-related amyloidosis, suggest that the disease is multifactorial and occurs only when the amyloidogenic protein acquires a conformation prone to form fibrils under conditions yet to clarify. However, the development of the final quaternary structure of amyloid fibrils is affected, apparently, not only by the protein primary structure but also by the tissue microenvironment that may influence which pathway the protein takes during its fibrillar conversion. Indeed, the recent work using cryo-EM has highlighted differences in the types of filament assembly on TTR fibrils isolated from the heart and vitreous body of the eye in carriers of the same amyloidogenic V30M TTR (Schmidt et al., 2019; Iakovleva et al., 2021).

A better understanding of the effects of the environmental determinants on the conformational dynamics of the protein may provide clues on the structural basis of the intrinsic amyloidogenic property of every specific protein. The complexity of this pathologic process arises through the effect of two dynamic entities. On one hand, we have proteins that can acquire different conformations and, on the other, we have a biological environment under dynamic transformation. In the field of amyloid disease, we are dealing with the very challenging interplay of two entities: protein metamorphosis linked to the organism's physiology which is

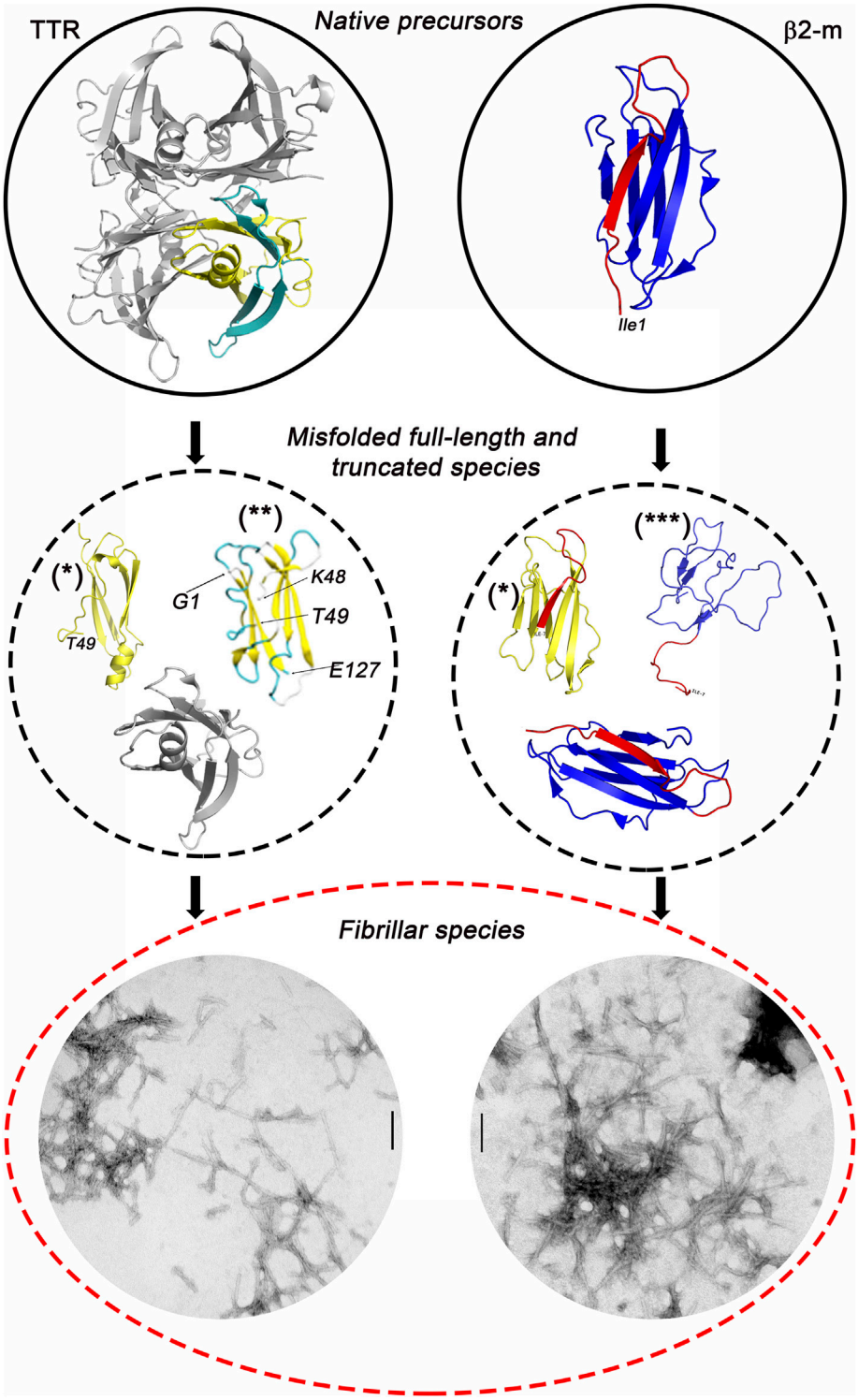

FIGURE 9 | TTR and β2m amyloid formation pathway. Both native proteins may undergo selective proteolytic cleavage generating misfolded amyloidogenic intermediates whose structures still remain elusive even applying advanced structure prediction software such as AlphaFold2. However, despite their different native structure, when misfolded, both TTR and β2-m convert into fibrillar structures that are morphologically similar as shown by negatively stained transmission electron microscopy (scale bar = 100 nm). (*) = structures predicted using the AlphaFold2 software; (**) = conformations predicted using molecular dynamics simulation (Marcoux et al., 2015); (***) = structures derived from NMR spectroscopy measurements (Esposito et al., 2000).

influenced by pathological features of environment, disease and age. An amyloidogenic protein and its environment are both in dynamic transformation and we cannot imagine to understand better the one without elucidate the role of the other in the time frame in which we develop the disease.

AUTHOR CONTRIBUTIONS

Topic selection, conceptualization, and writing first draft (VB); contribution to review literature data (GF, VM, SR, LM, DC, FL, GV, and SG). Contribution to writing (GF, AC, and PM). Editing and revision (GT, LO, JG, MS, SG, SR, and GV); editing and/or preparation of figures (GV, AC, and PM). All authors revised the review for intellectual content, contributed to discussion and approved the submitted version of the manuscript.

FUNDING

This work was supported by the investment from the University College London Technology Fund and grants from the U.K. Medical Research Council (MR/K000187/1 and MR/R016984/1), the Rosetrees Trust/Royal Free Charity PhD programme (M427), the Italian Ministry of Health (Ricerca Finalizzata RF 2013 02355259), the Italian Ministry of Research and University Dipartimenti di Eccellenza 2018–2022 grant to the Molecular Medicine Department (University of Pavia), and the Istituto Nazionale di Biostrutture e Biosistemi. Core support for the Wolfson Drug Discovery Unit is provided by the UK National Institute for Health Research Biomedical Research Centre and Unit Funding Scheme *via* the UCLH/UCL Biomedical Research Centre and by the UCL Amyloidosis Research Fund.

ACKNOWLEDGMENTS

This review is dedicated to the memory of Professor Robert Kisilevsky, a pioneer of modern research on amyloid diseases, always inspiring the need to merge clinical and experimental observations.

REFERENCES

Achour, A., Broggini, L., Han, X., Sun, R., Santambrogio, C., Buratto, J., et al. (2020). Biochemical and Biophysical Comparison of Human and Mouse Beta-2 Microglobulin Reveals the Molecular Determinants of Low Amyloid Propensity. *Febs J.* 287 (3), 546–560. doi:10.1111/febs.15046

Andrade, C. (1952). A Peculiar Form of Peripheral Neuropathy. *Brain* 75, 408–427. doi:10.1093/brain/75.3.408

Assenat, H., Calemard, E., Charra, B., Laurent, G., Terrat, J. C., and Vanel, T. (1980). Hemodialysis: Carpal Tunnel Syndrome and Amyloid Substance. *Nouv. Presse Med.* 9 (24), 1715.

Barroso, F., Waddington-Cruz, M., Mundayat, R., and Ong, M. L. (2015). THAOS - the Transthyretin Amyloidosis Outcomes Survey - Report on the Patient Demographic and Baseline Characteristics after 7years of Initiation. *J. Neurol. Sci.* 357, e67. doi:10.1016/j.jns.2015.08.249

Bellotti, V., Stoppini, M., Mangione, P., Sunde, M., Robinson, C., Asti, L., et al. (1998). Beta 2-microglobulin Can Be Refolded into a Native State from *Ex Vivo* Amyloid Fibrils. *Eur. J. Biochem.* 258 (1), 61–67. doi:10.1046/j.1432-1327.1998.2580061.x

Benson, M. D., Buxbaum, J. N., Eisenberg, D. S., Merlini, G., Saraiva, M. J. M., Sekijima, Y., et al. (2020). Amyloid Nomenclature 2020: Update and Recommendations by the International Society of Amyloidosis (ISA) Nomenclature Committee. *Amyloid* 27 (4), 217–222. doi:10.1080/13506129.2020.1835263

Benson, M. D., Liepnieks, J. J., and Kluve-Beckerman, B. (2015). Hereditary Systemic Immunoglobulin Light-Chain Amyloidosis. *Blood* 125 (21), 3281–3286. doi:10.1182/blood-2014-12-618108

Berg, I., Thor, S., and Hammarström, P. (2009). Modeling Familial Amyloidotic Polyneuropathy (Transthyretin V30M) in *Drosophila melanogaster*. *Neurodegenerative Dis.* 6 (3), 127–138. doi:10.1159/000213761

Bergström, J., Gustavsson, Å., Hellman, U., Sletten, K., Murphy, C. L., Weiss, D. T., et al. (2005). Amyloid Deposits in Transthyretin-Derived Amyloidosis: Cleaved Transthyretin Is Associated with Distinct Amyloid Morphology. *J. Pathol.* 206 (2), 224–232. doi:10.1002/path.1759

Blumenthal, H. T. (2004). Amyloidosis: A Universal Disease of Aging?The Journals of Gerontology. *J. Gerontol. A. Biol. Sci. Med. Sci.Series A.* 59 (4), M361–M369. doi:10.1093/gerona/59.4.m361

Booth, D. R., Sunde, M., Bellotti, V., Robinson, C. V., Hutchinson, W. L., Fraser, P. E., et al. (1997). Instability, Unfolding and Aggregation of Human Lysozyme Variants Underlying Amyloid Fibrillogenesis. *Nature* 385 (6619), 787–793. doi:10.1038/385787a0

Brunori, M. (2021). From Kuru to Alzheimer: A Personal Outlook. *Protein Sci.* 30 (9), 1776–1792. doi:10.1002/pro.4145

Buxbaum, J. N. (2009). Animal Models of Human Amyloidoses: Are Transgenic Mice worth the Time and Trouble. *FEBS Lett.* 583 (16), 2663–2673. doi:10.1016/j.febslet.2009.07.031

Chambers, J. K., Kanda, T., Shirai, A., Higuchi, K., Ikeda, S.-i., and Une, Y. (2010). Senile Systemic Amyloidosis in an Aged savannah Monkey (*Cercopithecus aethiops*) with Tenosynovial Degeneration. *J. Vet. Med. Sci.* 72 (5), 657–659. doi:10.1292/jvms.09-0394

Chiti, F., and Dobson, C. M. (2017). Protein Misfolding, Amyloid Formation, and Human Disease: A Summary of Progress over the Last Decade. *Annu. Rev. Biochem.* 86, 27–68. doi:10.1146/annurev-biochem-061516-045115

Colon, W., and Kelly, J. W. (1992). Partial Denaturation of Transthyretin Is Sufficient for Amyloid Fibril Formation *In Vitro*. *Biochemistry* 31 (36), 8654–8660. doi:10.1021/bi00151a036

Connelly, S., Choi, S., Johnson, S. M., Kelly, J. W., and Wilson, I. A. (2010). Structure-based Design of Kinetic Stabilizers that Ameliorate the Transthyretin Amyloidoses. *Curr. Opin. Struct. Biol.* 20 (1), 54–62. doi:10.1016/j.sbi.2009.12.009

Corazza, A., Verona, G., Waudby, C. A., Mangione, P. P., Bingham, R., Uings, I., et al. (2019). Binding of Monovalent and Bivalent Ligands by Transthyretin Causes Different Short- and Long-Distance Conformational Changes. *J. Med. Chem.* 62 (17), 8274–8283. doi:10.1021/acs.jmedchem.9b01037

Costa, P. P., Figueira, A. S., and Bravo, F. R. (1978). Amyloid Fibril Protein Related to Prealbumin in Familial Amyloidotic Polyneuropathy. *Proc. Natl. Acad. Sci.* 75 (9), 4499–4503. doi:10.1073/pnas.75.9.4499

Crawley, J. T. B., de Groot, R., Xiang, Y., Luken, B. M., and Lane, D. A. (2011). Unraveling the scissile bond: how ADAMTS13 recognizes and cleaves von Willebrand factor. *Blood* 118 (12), 3212–3221. doi:10.1182/blood-2011-02-306597

de Strihou, C. v. Y., Jadoul, M., Malghem, J., Maldague, B., and Jamart, J. (1991). Effect of Dialysis Membrane and Patient's Age on Signs of Dialysis-Related Amyloidosis. *Kidney Int.* 39 (5), 1012–1019. doi:10.1038/ki.1991.128

DeMarco, M. L., and Daggett, V. (2004). From Conversion to Aggregation: Protofibril Formation of the Prion Protein. *Proc. Natl. Acad. Sci.* 101 (8), 2293–2298. doi:10.1073/pnas.0307178101

Diomede, L., Soria, C., Romeo, M., Giorgetti, S., Marchese, L., Mangione, P. P., et al. (2012). *C. elegans* Expressing Human β2-Microglobulin: A Novel Model for

Studying the Relationship between the Molecular Assembly and the Toxic Phenotype. *PLoS One* 7 (12), e52314. doi:10.1371/journal.pone.0052314

Dishman, A. F., and Volkman, B. F. (2018). Unfolding the Mysteries of Protein Metamorphosis. *ACS Chem. Biol.* 13 (6), 1438–1446. doi:10.1021/acschembio.8b00276

Eichner, T., and Radford, S. E. (2011). A Diversity of Assembly Mechanisms of a Generic Amyloid Fold. *Mol. Cel* 43 (1), 8–18. doi:10.1016/j.molcel.2011.05.012

Esposito, G., Michelutti, R., Verdone, G., Viglino, P., ÁNdez, H. H., Robinson, C. V., et al. (2000). Removal of the N-Terminal Hexapeptide from Human β2-microglobulin Facilitates Protein Aggregation and Fibril Formation. *Protein Sci.* 9 (5), 831–845. doi:10.1110/ps.9.5.831

Faravelli, G., Raimondi, S., Marchese, L., Partridge, F. A., Soria, C., Mangione, P. P., et al. (2019). *C. elegans* Expressing D76N β2-microglobulin: a Model for *In Vivo* Screening of Drug Candidates Targeting Amyloidosis. *Sci. Rep.* 9 (1), 19960. doi:10.1038/s41598-019-56498-5

Gejyo, F., Yamada, T., Odani, S., Nakagawa, Y., Arakawa, M., Kunitomo, T., et al. (1985). A New Form of Amyloid Protein Associated with Chronic Hemodialysis Was Identified as β2-microglobulin. *Biochem. Biophysical Res. Commun.* 129 (3), 701–706. doi:10.1016/0006-291x(85)91948-5

Good, S. C., Dewison, K. M., Radford, S. E., and van Oosten-Hawle, P. (2021). Global Proteotoxicity Caused by Human β2 Microglobulin Variants Impairs the Unfolded Protein Response in *C. elegans*. *Ijms* 22 (19), 10752. doi:10.3390/ijms221910752

Grogan, M., Scott, C. G., Kyle, R. A., Zeldenrust, S. R., Gertz, M. A., Lin, G., et al. (2016). Natural History of Wild-type Transthyretin Cardiac Amyloidosis and Risk Stratification Using a Novel Staging System. *J. Am. Coll. Cardiol.* 68 (10), 1014–1020. doi:10.1016/j.jacc.2016.06.033

Halabelian, L., Ricagno, S., Giorgetti, S., Santambrogio, C., Barbiroli, A., Pellegrino, S., et al. (2014). Class I Major Histocompatibility Complex, the Trojan Horse for Secretion of Amyloidogenic β2-Microglobulin. *J. Biol. Chem.* 289 (6), 3318–3327. doi:10.1074/jbc.M113.524157

Hammarström, P., Jiang, X., Hurshman, A. R., Powers, E. T., and Kelly, J. W. (2002). Sequence-dependent Denaturation Energetics: A Major Determinant in Amyloid Disease Diversity. *Proc. Natl. Acad. Sci.* 99 (Suppl. 4), 16427–16432. doi:10.1073/pnas.202495199

Hartl, F. U., and Hayer-Hartl, M. (2009). Converging Concepts of Protein Folding *In Vitro* and *In Vivo*. *Nat. Struct. Mol. Biol.* 16 (6), 574–581. doi:10.1038/nsmb.1591

Higuchi, K., Kitagawa, K., Naiki, H., Hanada, K., Hosokawa, M., and Takeda, T. (1991). Polymorphism of Apolipoprotein A-II (apoA-II) Among Inbred Strains of Mice. Relationship between the Molecular Type of apoA-II and Mouse Senile Amyloidosis. *Biochem. J.* 279 (Pt 2), 427–433. doi:10.1042/bj2790427

Hurle, M. R., Helms, L. R., Li, L., Chan, W., and Wetzel, R. (1994). A Role for Destabilizing Amino Acid Replacements in Light-Chain Amyloidosis. *Proc. Natl. Acad. Sci.* 91 (12), 5446–5450. doi:10.1073/pnas.91.12.5446

Iakovleva, I., Hall, M., Oelker, M., Sandblad, L., Anan, I., and Sauer-Eriksson, A. E. (2021). Structural Basis for Transthyretin Amyloid Formation in Vitreous Body of the Eye. *Nat. Commun.* 12 (1), 7141. doi:10.1038/s41467-021-27481-4

Ihse, E., Rapezzi, C., Merlini, G., Benson, M. D., Ando, Y., Suhr, O. B., et al. (2013). Amyloid Fibrils Containing Fragmented ATTR May Be the Standard Fibril Composition in ATTR Amyloidosis. *Amyloid* 20 (3), 142–150. doi:10.3109/13506129.2013.797890

Judge, D. P., Heitner, S. B., Falk, R. H., Maurer, M. S., Shah, S. J., Witteles, R. M., et al. (2019). Transthyretin Stabilization by AG10 in Symptomatic Transthyretin Amyloid Cardiomyopathy. *J. Am. Coll. Cardiol.* 74 (3), 285–295. doi:10.1016/j.jacc.2019.03.012

Jurcuț, R., Onciul, S., Adam, R., Stan, C., Coriu, D., Rapezzi, C., et al. (2020). Multimodality Imaging in Cardiac Amyloidosis: a Primer for Cardiologists. *Eur. Heart J. Cardiovasc. Imaging* 21 (8), 833–844. doi:10.1093/ehjci/jeaa063

Karamanos, T. K., Kalverda, A. P., Thompson, G. S., and Radford, S. E. (2014). Visualization of Transient Protein-Protein Interactions that Promote or Inhibit Amyloid Assembly. *Mol. Cel* 55 (2), 214–226. doi:10.1016/j.molcel.2014.05.026

Koedam, E. L. G. E., Lauffer, V., van der Vlies, A. E., van der Flier, W. M., Scheltens, P., and Pijnenburg, Y. A. L. (2010). Early-versus Late-Onset Alzheimer's Disease: More Than Age Alone. *Jad* 19 (4), 1401–1408. doi:10.3233/jad-2010-1337

Kraus, A., Hoyt, F., Schwartz, C. L., Hansen, B., Artikis, E., Hughson, A. G., et al. (2021). High-resolution Structure and Strain Comparison of Infectious Mammalian Prions. *Mol. Cel* 81 (21), 4540–4551. e4546. doi:10.1016/j.molcel.2021.08.011

Latimer, C. S., Lucot, K. L., Keene, C. D., Cholerton, B., and Montine, T. J. (2021). Genetic Insights into Alzheimer's Disease. *Annu. Rev. Pathol. Mech. Dis.* 16, 351–376. doi:10.1146/annurev-pathmechdis-012419-032551

Laudet, V. (2011). The Origins and Evolution of Vertebrate Metamorphosis. *Curr. Biol.* 21 (18), R726–R737. doi:10.1016/j.cub.2011.07.030

Li, X., Lyu, Y., Shen, J., Mu, Y., Qiang, L., Liu, L., et al. (2018). Amyloid Deposition in a Mouse Model Humanized at the Transthyretin and Retinol-Binding Protein 4 Loci. *Lab. Invest.* 98 (4), 512–524. doi:10.1038/s41374-017-0019-y

Mack, H. I. D., Heimbucher, T., and Murphy, C. T. (2018). The Nematode *Caenorhabditis elegans* as a Model for Aging Research. *Drug Discov. Today Dis. Models* 27, 3–13. doi:10.1016/j.ddmod.2018.11.001

Madhivanan, K., Greiner, E. R., Alves-Ferreira, M., Soriano-Castell, D., Rouzbeh, N., Aguirre, C. A., et al. (2018). Cellular Clearance of Circulating Transthyretin Decreases Cell-Nonautonomous Proteotoxicity in *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. U.S.A.* 115 (33), E7710–e7719. doi:10.1073/pnas.1801117115

Mangione, P. P., Esposito, G., Relini, A., Raimondi, S., Porcari, R., Giorgetti, S., et al. (2013). Structure, Folding Dynamics, and Amyloidogenesis of D76N β2-Microglobulin. *J. Biol. Chem.* 288 (43), 30917–30930. doi:10.1074/jbc.M113.498857

Mangione, P. P., Porcari, R., Gillmore, J. D., Pucci, P., Monti, M., Porcari, M., et al. (2014). Proteolytic Cleavage of Ser52Pro Variant Transthyretin Triggers its Amyloid Fibrillogenesis. *Proc. Natl. Acad. Sci.* 111 (4), 1539–1544. doi:10.1073/pnas.1317488111

Mangione, P. P., Verona, G., Corazza, A., Marcoux, J., Canetti, D., Giorgetti, S., et al. (2018). Plasminogen Activation Triggers Transthyretin Amyloidogenesis *In Vitro*. *J. Biol. Chem.* 293 (37), 14192–14199. doi:10.1074/jbc.RA118.003990

Marcoux, J., Mangione, P. P., Porcari, R., Degiacomi, M. T., Verona, G., Taylor, G. W., et al. (2015). A Novel Mechano-enzymatic Cleavage Mechanism Underlies Transthyretin Amyloidogenesis. *EMBO Mol. Med.* 7 (10), 1337–1349. doi:10.15252/emmm.201505357

McAdam, K. P., and Sipe, J. D. (1976). Murine Model for Human Secondary Amyloidosis: Genetic Variability of the Acute-phase Serum Protein SAA Response to Endotoxins and Casein. *J. Exp. Med.* 144 (4), 1121–1127. doi:10.1084/jem.144.4.1121

Merlini, G., and Bellotti, V. (2003). Molecular Mechanisms of Amyloidosis. *N. Engl. J. Med.* 349 (6), 583–596. doi:10.1056/NEJMra023144

Merlini, G., Dispenzieri, A., Sanchorawala, V., Schönland, S. O., Palladini, G., Hawkins, P. N., et al. (2018). Systemic Immunoglobulin Light Chain Amyloidosis. *Nat. Rev. Dis. Primers* 4 (1), 38. doi:10.1038/s41572-018-0034-3

Michaels, T. C. T., Šarić, A., Habchi, J., Chia, S., Meisl, G., Vendruscolo, M., et al. (2018). Chemical Kinetics for Bridging Molecular Mechanisms and Macroscopic Measurements of Amyloid Fibril Formation. *Annu. Rev. Phys. Chem.* 69, 273–298. doi:10.1146/annurev-physchem-050317-021322

Mizuno, H., Hoshino, J., So, M., Kogure, Y., Fujii, T., Ubara, Y., et al. (2021). Dialysis-related Amyloidosis Associated with a Novel β2-microglobulin Variant. *Amyloid* 28 (1), 42–49. doi:10.1080/13506129.2020.1813097

Morawski, M., Filippov, M., Tzinia, A., Tsilibary, E., and Vargova, L. (2014). ECM in Brain Aging and Dementia. *Prog. Brain Res.* 214, 207–227. doi:10.1016/b978-0-444-63486-3.00010-4

Mucchiano, G. I., Häggqvist, B., Sletten, K., and Westermark, P. (2001). Apolipoprotein A-1-Derived Amyloid in Atherosclerotic Plaques of the Human Aorta. *J. Pathol.* 193 (2), 270–275. doi:10.1002/1096-9896(2000)9999:9999<:Aid-path753>3.0.Co;2-s

Muchtar, E., Dispenzieri, A., Magen, H., Grogan, M., Mauermann, M., McPhail, E. D., et al. (2021). Systemic Amyloidosis from A (AA) to T (ATTR): a Review. *J. Intern. Med.* 289 (3), 268–292. doi:10.1111/joim.13169

Muchtar, E., Gertz, M. A., Kyle, R. A., Lacy, M. Q., Dingli, D., Leung, N., et al. (2019). A Modern Primer on Light Chain Amyloidosis in 592 Patients with Mass Spectrometry-Verified Typing. *Mayo Clinic Proc.* 94 (3), 472–483. doi:10.1016/j.mayocp.2018.08.006

Nakamura, S., Okabayashi, S., Ageyama, N., Koie, H., Sankai, T., Ono, F., et al. (2008). Transthyretin Amyloidosis and Two Other Aging-Related Amyloidoses in an Aged Vervet Monkey. *Vet. Pathol.* 45 (1), 67–72. doi:10.1354/vp.45-1-67

Natalello, A., Mangione, P. P., Giorgetti, S., Porcari, R., Marchese, L., Zorzoli, I., et al. (2016). Co-fibrillogenesis of Wild-type and D76N β2-Microglobulin. *J. Biol. Chem.* 291 (18), 9678–9689. doi:10.1074/jbc.M116.720573

Nativi-Nicolau, J., Siu, A., Dispenzieri, A., Maurer, M. S., Rapezzi, C., Kristen, A. V., et al. (2021). Temporal Trends of Wild-type Transthyretin Amyloid Cardiomyopathy in the Transthyretin Amyloidosis Outcomes Survey. *JACC: CardioOncology* 3 (4), 537–546. doi:10.1016/j.jaccao.2021.08.009

Pinheiro, F., Santos, J., and Ventura, S. (2021). AlphaFold and the Amyloid Landscape. *J. Mol. Biol.* 433 (20), 167059. doi:10.1016/j.jmb.2021.167059

Pokrzywa, M., Dacklin, I., Hultmark, D., and Lundgren, E. (2007). Misfolded Transthyretin Causes Behavioral Changes in a Drosophila Model for Transthyretin-Associated Amyloidosis. *Eur. J. Neurosci.* 26 (4), 913–924. doi:10.1111/j.1460-9568.2007.05728.x

Puoti, G., Bizzi, A., Forloni, G., Safar, J. G., Tagliavini, F., and Gambetti, P. (2012). Sporadic Human Prion Diseases: Molecular Insights and Diagnosis. *Lancet Neurol.* 11 (7), 618–628. doi:10.1016/s1474-4422(12)70063-7

Rapezzi, C., Quarta, C. C., Obici, L., Perfetto, F., Longhi, S., Salvi, F., et al. (2013). Disease Profile and Differential Diagnosis of Hereditary Transthyretin-Related Amyloidosis with Exclusively Cardiac Phenotype: an Italian Perspective. *Eur. Heart J.* 34 (7), 520–528. doi:10.1093/eurheartj/ehs123

Rickenbacher, U., McKinney, J. D., Oatley, S. J., and Blake, C. C. F. (1986). Structurally Specific Binding of Halogenated Biphenyls to Thyroxine Transport Protein. *J. Med. Chem.* 29 (5), 641–648. doi:10.1021/jm00155a010

Riek, R., Hornemann, S., Wider, G., Billeter, M., Glockshuber, R., and Wüthrich, K. (1996). NMR Structure of the Mouse Prion Protein Domain PrP(121-231). *Nature* 382 (6587), 180–182. doi:10.1038/382180a0

Saelices, L., Pokrzywa, M., Pawelek, K., and Eisenberg, D. S. (2018). Assessment of the Effects of Transthyretin Peptide Inhibitors in Drosophila Models of Neuropathic ATTR. *Neurobiol. Dis.* 120, 118–125. doi:10.1016/j.nbd.2018.09.007

Santos, S. D., Fernandes, R., and Saraiva, M. J. (2010). The Heat Shock Response Modulates Transthyretin Deposition in the Peripheral and Autonomic Nervous Systems. *Neurobiol. Aging* 31 (2), 280–289. doi:10.1016/j.neurobiolaging.2008.04.001

Saraiva, M. J. (2015). TTR Amyloidosis: a Scientific Journey since Andrade. *Orphanet J. Rare Dis.* 10 (1), I19. doi:10.1186/1750-1172-10-S1-I19

Schmidt, M., Wiese, S., Adak, V., Engler, J., Agarwal, S., Fritz, G., et al. (2019). Cryo-EM Structure of a Transthyretin-Derived Amyloid Fibril from a Patient with Hereditary ATTR Amyloidosis. *Nat. Commun.* 10 (1), 5008. doi:10.1038/s41467-019-13038-z

Schmitz, M., Dittmar, K., Llorens, F., Gelpi, E., Ferrer, I., Schulz-Schaeffer, W. J., et al. (2017). Hereditary Human Prion Diseases: an Update. *Mol. Neurobiol.* 54 (6), 4138–4149. doi:10.1007/s12035-016-9918-y

Simons, J. P., Al-Shawi, R., Ellmerich, S., Speck, I., Aslam, S., Hutchinson, W. L., et al. (2013). Pathogenetic Mechanisms of Amyloid A Amyloidosis. *Proc. Natl. Acad. Sci.* 110 (40), 16115–16120. doi:10.1073/pnas.1306621110

Slamova, I., Adib, R., Ellmerich, S., Golos, M., Gilbertson, J. A., Botcher, N., et al. (2021). Plasmin Activity Promotes Amyloid Deposition in a Transgenic Model of Human Transthyretin Amyloidosis. *Nat. Commun* 12 (1), 7112. doi:10.1038/s41467-021-27416-z

Sletten, K., Westermark, P., and Natvig, J. B. (1980). Senile Cardiac Amyloid Is Related to Prealbumin. *Scand. J. Immunol.* 12 (6), 503–506. doi:10.1111/j.1365-3083.1980.tb00098.x

Sousa, M. M., Fernandes, R., Palha, J. A., Taboada, A., Vieira, P., and Saraiva, M. J. (2002). Evidence for Early Cytotoxic Aggregates in Transgenic Mice for Human Transthyretin Leu55Pro. *Am. J. Pathol.* 161 (5), 1935–1948. doi:10.1016/S0002-9440(10)64469-0

Stoppini, M., and Bellotti, V. (2015). Systemic Amyloidosis: Lessons from β2-Microglobulin. *J. Biol. Chem.* 290 (16), 9951–9958. doi:10.1074/jbc.R115.639799

Swerdlow, R. H. (2018). Mitochondria and Mitochondrial Cascades in Alzheimer's Disease. *Jad* 62 (3), 1403–1416. doi:10.3233/jad-170585

Tasaki, M., Lavatelli, F., Obici, L., Obayashi, K., Miyamoto, T., Merlini, G., et al. (2021). Age-related Amyloidosis outside the Brain: A State-Of-The-Art Review. *Ageing Res. Rev.* 70, 101388. doi:10.1016/j.arr.2021.101388

Trigo, D., Nadais, A., and da Cruz e Silva, O. A. B. (2019). Unravelling Protein Aggregation as an Ageing Related Process or a Neuropathological Response. *Ageing Res. Rev.* 51, 67–77. doi:10.1016/j.arr.2019.02.001

Tsuda, Y., Yamanaka, K., Toyoshima, R., Ueda, M., Masuda, T., Misumi, Y., et al. (2018). Development of Transgenic *Caenorhabditis elegans* Expressing Human Transthyretin as a Model for Drug Screening. *Sci. Rep.* 8 (1), 17884. doi:10.1038/s41598-018-36357-5

Ueda, M., Ageyama, N., Nakamura, S., Nakamura, M., Chambers, J. K., Misumi, Y., et al. (2012). Aged Vervet Monkeys Developing Transthyretin Amyloidosis with the Human Disease-Causing Ile122 Allele: a Valid Pathological Model of the Human Disease. *Lab. Invest.* 92 (3), 474–484. doi:10.1038/labinvest.2011.195

Valleix, S., Gillmore, J. D., Bridoux, F., Mangione, P. P., Dogan, A., Nedelec, B., et al. (2012). Hereditary Systemic Amyloidosis Due to Asp76Asn Variant β2-Microglobulin. *N. Engl. J. Med.* 366 (24), 2276–2283. doi:10.1056/NEJMoa1201356

Yi, S., Takahashi, K., Naito, M., Tashiro, F., Wakasugi, S., Maeda, S., et al. (1991). Systemic Amyloidosis in Transgenic Mice Carrying the Human Mutant Transthyretin (Met30) Gene. Pathologic Similarity to Human Familial Amyloidotic Polyneuropathy, Type I. *Am. J. Pathol.* 138 (2), 403–412.

Zhang, P., Fu, X., Sawashita, J., Yao, J., Zhang, B., Qian, J., et al. (2010). Mouse Model to Study Human A β2M Amyloidosis: Generation of a Transgenic Mouse with Excessive Expression of Human β2-microglobulin. *Amyloid* 17 (2), 50–62. doi:10.3109/13506129.2010.483116

Zhang, X., Halvorsen, K., Zhang, C.-Z., Wong, W. P., and Springer, T. A. (2009). Mechanoenzymatic cleavage of the ultralarge vascular protein von Willebrand factor. *Science* 324 (5932), 1330–1334. doi:10.1126/science.1170905

Conflict of Interest: For sake of clarity, we would like to disclose the following: research grant (2021–2023) from Pfizer (DC and VB); speaker and consulting honoraria from Alnylam, Pfizer, Akcea and SOBI (LO); participation to advisory boards for Pfizer, Alnylam Pharmaceuticals Inc. (VB and JG) and Eidos, Ionis, Intellia (JG). VB, GT, and PM are authors of a patent for ATTR amyloidosis agents.

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

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

frontiers in Molecular Biosciences

ORIGINAL RESEARCH
published: 23 February 2022
doi: 10.3389/fmolb.2022.844129

Impact of a Single Nucleotide Polymorphism on the 3D Protein Structure and Ubiquitination Activity of E3 Ubiquitin Ligase Arkadia

Maria Birkou[1], Vasilios Raptis[1†], Konstantinos D. Marousis[1†], Athanasios Tsevis[2], Kyriakos Bourikas[2], Detlef Bentrop[3], Vasso Episkopou[4] and Georgios A. Spyroulias[1]**

[1]Department of Pharmacy, University of Patras, Patras, Greece, [2]School of Science and Technology, Hellenic Open University, Patras, Greece, [3]Institute of Physiology II, Faculty of Medicine, University of Freiburg, Freiburg, Germany, [4]Faculty of Medicine, Imperial College London, Hammersmith Hospital Campus, Burlington Danes, London, United Kingdom

OPEN ACCESS

Edited by:
Piero Andrea Temussi, University of Naples Federico II, Italy

Reviewed by:
Christopher Berndsen, James Madison University, United States
Chi-Fon Chang, Academia Sinica, Taiwan

***Correspondence:**
Georgios A. Spyroulias
G.A.Spyroulias@upatras.gr
Vasso Episkopou
vasso.episkopou@imperial.ac.uk

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Structural Biology, a section of the journal Frontiers in Molecular Biosciences

Received: *27 December 2021*
Accepted: *31 January 2022*
Published: *23 February 2022*

Citation:
Birkou M, Raptis V, Marousis KD, Tsevis A, Bourikas K, Bentrop D, Episkopou V and Spyroulias GA (2022) Impact of a Single Nucleotide Polymorphism on the 3D Protein Structure and Ubiquitination Activity of E3 Ubiquitin Ligase Arkadia. Front. Mol. Biosci. 9:844129. doi: 10.3389/fmolb.2022.844129

Single nucleotide polymorphisms (SNPs) are genetic variations which can play a vital role in the study of human health. SNP studies are often used to identify point mutations that are associated with diseases. Arkadia (RNF111) is an E3 ubiquitin ligase that enhances transforming growth factor-beta (TGF-β) signaling by targeting negative regulators for degradation. Dysregulation of the TGF-β pathway is implicated in cancer because it exhibits tumor suppressive activity in normal cells while in tumor cells it promotes invasiveness and metastasis. The SNP CGT > TGT generated an amino-acid (aa) substitution of Arginine 957 to Cysteine on the enzymatic RING domain of Arkadia. This was more prevalent in a tumor than in a normal tissue sample of a patient with colorectal cancer. This prompted us to investigate the effect of this mutation in the structure and activity of Arkadia RING. We used nuclear magnetic resonance (NMR) to analyze at an atomic-level the structural and dynamic properties of the R957C Arkadia RING domain, while ubiquitination and luciferase assays provided information about its enzymatic functionality. Our study showed that the R957C mutation changed the electrostatic properties of the RING domain however, without significant effects on the structure of its core region. However, the functional studies revealed that the R957C Arkadia exhibits significantly increased enzymatic activity supporting literature data that Arkadia within tumor cells promotes aggressive and metastatic behavior.

Keywords: Arkadia, snps, NMR, RING, E3 ligase

INTRODUCTION

Ubiquitination is a three-step enzymatic reaction catalyzed by the action of three key-enzymes. Initially, ubiquitin is activated in an ATP-dependent manner by the E1 activating enzyme. Activated ubiquitin afterwards is transferred to E2 conjugating enzyme. In the last step of the enzymatic reaction ubiquitin is covalently attached to the target protein, through an isopeptide bond between its G76 and the ε-amino group of a lysine residue of the target protein. The attachment of ubiquitin to substrate is catalyzed by the E3 ubiquitin ligase. The E3 ubiquitin ligases are considered crucial partners of the ubiquitination machinery because they act as specific substrate recognition elements. The ubiquitin ligases are characterized by the presence of either a HECT (Homologous to E6-AP

Carboxyl Terminus) or a RING (Really Interesting New Gene) domain (Hershkos et al., 1983; Pickart, 2001). RING-type E3 ligases constitutes the largest class of E3 ligases and they are characterized by the following amino acids sequence: Cys-X_2-Cys-X_{9-39}-Cys-X_{1-3}-His-X_2-Cys/His-X_2-Cys-X_{4-48}-Cys-X_2-Cys, wherein X represents any amino acid. The RING domain motif is constructed by Cys2His2, Cys3His or Cys4 metal binding sites which coordinate two zinc ions in a "cross-brace" arrangement (Freemont, 1993; Saurin et al., 1996).

Ubiquitination is a signaling pathway that regulates a broad spectrum of biochemical processes, e.g., it enhances the tumor suppression activity of some proteins or the oncogenic and metastatic properties of others and has a significant role in the life cycle of cells. Thus, many studies have shown that mutations in E3 ligases are implicated in cancer (Shangary and Wang, 2008; Nelson and Holt, 2010; Sharma et al., 2011; Chan et al., 2013; Wang et al, 2017) and other diseases (Kumar et al., 2017). Two of the thoroughly studied and well characterized RING E3 ligases are BRCA1 and HDM2. Tumor suppressor BRCA1 is a RING E3 ligase whose mutations lead to a high predisposition of breast and ovarian cancers (Brzovic et al., 2001; Rosen, 2013). Mutations observed in the RING sequence are now considered as significant markers for these types of cancer. Similarly, Hdm2 is a RING E3 ligase whose overexpression associates with many human tumor types (breast, esophageal, lung carcinomas etc.,), whereas its binding and degradation of p53 tumor suppressor is inactivated in more than 50% of human cancers (Moll and Petrenko, 2003; Sun, 2006). Targeting specific E3 ligases, that play a significant role in cancer, may contribute to the development of therapeutic strategies (Lakshmanan et al., 2008).

Arkadia (RNF111) is a RING ubiquitin ligase that positively regulates the TGF-β pathway by targeting its negative regulator SKI and its close homologue SKIL (SNON), as well as the inhibitory SMAD7 for ubiquitin-dependent proteasomal degradation. Arkadia functions as E3 ligase through its C-terminal RING-H2 domain (amino acids 942–983) (Levy et al., 2007; Mavrakis et al., 2007; Nagano et al., 2007). Arkadia was shown to enhance and support TGF-β tumor suppressing function in colorectal cancer (CRC) (Sharma et al., 2011). Furthermore, Le Scolan, et al., 2008 and Briones-Orta et al., 2013, using tumor cell lines driven by the TGF-β pathway, demonstrated that Arkadia has a potent tumor-promoting activity. These data indicate that Arkadia supports both properties of the TGF-β pathway, i.e., tumor suppression in normal cells and metastasis in tumor cells. Moreover, deep sequencing studies of the mRNA from tumors of CRC patients (Bravou et al., 2009; Sharma et al., 2011), led to the identification of somatic mutations that diminish Arkadia's function (Sharma et al., 2011). They were listed in the COSMIC database of somatic mutations in cancer (Bravou et al., 2009; Sharma et al., 2011).

In the present study the single nucleotide polymorphism (SNP) of arginine (R) 957 (rs780099637, CGT > TGT, allele frequency = 0.000004 in gnomAD) enriched in CRC was studied. SNPs are genetic variations that are associated with individual susceptibility to diseases (Shastry, 2007). The R957C SNP was identified both in CRC and adjacent normal tissue samples (**Supplementary Table S1**). Existence of mutations in normal tissues adjacent to tumors is frequently observed and their association with cancer is under investigation (Risques and Kennedy 2018; Fiala and Diamandis, 2020; Oh and Sung, 2020). SNP studies are crucial to identify amino acids substitutions; in the protein coding regions that potentially alter the function or structure of a protein (Bhattacharya et al., 2017). Herein, we present a thorough overview about the effects of the R957C point mutation. Replacement of the positively charged arginine with cysteine affects the electrostatic properties of the protein, which may lead to alterations of structure, stability, and function of the enzymatic RING domain of the E3 ubiquitin ligase Arkadia. To predict the effect of the R957C mutation on Arkadia protein, the SNPs and GO and I-Mutant pathogenicity prediction servers were used (Capriotti et al., 2005; Capriotti et al., 2013). In order to investigate the impact of this mutation on the conformational dynamics and the function of the Arkadia RING domain, we prepared the recombinant polypeptide bearing a cysteine residue instead of arginine in position 957 of the human sequence and we conducted a structural and functional study to obtain an atomic-level insight into its conformational dynamics and activity in comparison with the native Arkadia RING domain.

MATERIALS AND METHODS

Prediction of Single Nucleotide Polymorphism Impact

The effect of the R957C substitution was analyzed using SNPs and GO and I-Mutant tools. SNPs and GO is a tool which predicts disease associated amino acid substitution at a single position in a specific protein including functional classifications with >82% prediction accuracy. I-Mutant predicts the effects of single point mutation with 80% accuracy. A probability score higher than 0.5 reveals a disease related effect of the mutation on protein function. The input given to the SNPs and GO and I-Mutant tools was the UniProt accession number (Q6ZNA4) of Arkadia isoform-1 protein, the sequence position of the wild type amino acid and the mutated amino acid.

Protein Expression and Purification

A human RNF111 gene encoding amino acids 927–994 of the full length Arkadia was sub-cloned in a pGEX-4T-1 vector and transformed in *Escherichia coli* (*E. cloni®*) EXPRESS BL21 (DE3) cells (Lucigen). Cells were grown at 37°C in minimal medium (M9) supplemented with 1 g/L ^{15}N ammonium chloride, 4 g/L glucose or ^{13}C-glucose and 1 ml/L ^{15}N or ^{15}N/^{13}C Bioxpress™ (CIL) for single or double labeled samples, respectively. Cell cultures were induced at an optical absorption of 0.6–0.9 with 1 mM IPTG (final concentration). After 4 h incubation at 37°C cells were harvested and the resulting pellet was lysed by sonication in PBS (Phosphate Buffered Saline) pH 7.4 ± 0.2 containing a protease inhibitor cocktail (Sigma) and DNAase I. The cell lysate was cleared by centrifugation at 20.000rpm (rounds per minute) for 30 min. For protein

purification the supernatant was loaded onto a GST-trap 5 ml column (GE Healthcare). The GST-tag and Arkadia$^{927-994}$ were separated after overnight incubation with the protease thrombin (Merck Millipore) at 4°C. Arkadia$^{927-994}$ was eluted in PBS pH 7.4 ± 0.2 and concentrated using Amicon® Ultra centrifugal filter units (3 kDa cutoff) (Merck Millipore). The protein was further purified by size exclusion chromatography on a Superdex75 column (GE Healthcare).

The E2 enzyme UBCH5B (UBE2D2) was expressed and purified as described elsewhere (Birkou et al., 2017) and ubiquitin was expressed in *E. coli* (*E. cloni*®) EXPRESS BL21 (DE3) cells (Lucigen). Cells were grown at 37°C in minimal medium (M9) supplemented with 4 g/L glucose and 1 g/L NH_4Cl and induced with 1 mM (final concentration) IPTG. After 4 h of incubation at 37°C cells were harvested and lysed by sonication in PBS pH 7.4. The cell lysate was cleared by centrifugation at 20.000 rpm for 30 min. The supernatant was heated at 85°C for 15 min and centrifuged at 10.000 rpm for 20 min. For further purification ubiquitin was passed through a Superdex75 column (GE Healthcare).

Nuclear Magnetic Resonance Data and Structure Calculation

NMR spectra were recorded on a Bruker Avance 600 MHz spectrometer equipped with a TXI cryoprobe and Bruker Avance III 700 MHz spectrometer equipped with a four-channel 5 mm cryogenically cooled TCI gradient probe. Protein samples were prepared in a mixed solvent of 90% H_20 (50 mM K_2HPO_4, 50 mM KH_2PO_4 pH 7), 10% D_20, 1 mM NaN_3 and 0.25 mM DSS (4,4-dimethyl-4-silapentane-1-sulfonic acid) as internal chemical shift standard. All NMR experiments are included in the Bruker pulse program library. Data were processed with Topspin 3.5 pl5 software and analyzed with CARA (Keller, 2004) and XEASY (Bartels et al., 1995). The sequence specific assignment of R957C Arkadia$^{927–994}$ domain was obtained using conventional backbone and side chain assignment methods (Wüthrich, 1986; Ferentz and Wagner, 2000) and were deposited in the BioMagResBank (BMRB; accession no: 50,985). The NMR solution structure of the R957C mutant was determined with DYANA (Güntert et al., 1997) and the ensemble of 30 models with the lowest RMSD and target function values was deposited in the ProteinDataBank after energy minimization with AMBER (Pearlman et al., 1995) (PDB, accession no: 7P2K; statistical analysis is reported in **Supplementary Table S2**). The assignment of the backbone ^{1}H-^{15}N resonances of all non-proline residues of E2 enzyme UBCH5B was obtained from the BMRB database (accession no: 6,277).

^{15}N Relaxation Data

The backbone mobility of the R957C Arkadia$^{927-994}$ mutant on the ps-ns time scale was investigated through ^{15}N relaxation measurements (^{15}N T_1 and T_2, $\{^1H^N\}$-^{15}N NOE at 298 K) on Bruker Avance 600 and Avance III 700 spectrometers (experiments are reported in **Supplementary Table S3**). Standard pulse programs of the Bruker library were used. The overall delays between scans were 1.8 s for the T_1 and T_2 measurements and 4 s for the interleaved heteronuclear NOE experiment, respectively. The delays used for the T_1 experiments were 7, 18, 40, 85, 150, 230, 350, 500, 680, and 900 ms, whereas delays of 17, 34, 51, 68, 85, 102, 136, 187, and 221 ms were used for the T_2 experiments. ^{15}N relaxation data were analyzed according to the model-free approach as implemented in the Tensor 2 program (Dosset et al., 2000) (R_1, R_2 and (^{1}H) ^{15}N NOE diagrams are shown in **Supplementary Figure S2**).

Atomic Absorption Spectroscopy

To identify whether two zinc ions are bound to the R957C mutant, atomic absorption spectroscopy was performed (Perkin Elmer AAnalyst 700). For the zinc concentration determination proteins were purified as described above. Zinc ions standards of 0.1–1 ppm (Zinc Pure Standard, 1000 μg/ml, 2% HNO_3, Perkin Elmer) were used and an appropriate calibration curve was constructed (**Supplementary Figure S1**). The concentration of zinc in the protein samples was determined dividing the intercept by the slope of calibration curve.

Metal-Chelation Experiments

To evaluate the stability of the R957C Arkadia$^{927–994}$ domain metal-chelation experiments were performed with the ethylenediaminetetraacetic acid (EDTA). ^{1}H-^{15}N HSQC spectrums were recorded after each EDTA addition at a ratio 1:0.5, 1:1, 1:1.5 and 1:2 (Zn: EDTA).

Titration Experiments Monitored by Nuclear Magnetic Resonance

In order to identify the interaction interface between E3 R957C Arkadia$^{927–994}$ and E2 UBCH5B, titration experiments were monitored by ^{1}H–^{15}N HSQC spectra of labeled ^{15}N R957C Arkadia$^{927–994}$ or ^{15}N UBCH5B after each addition of the unlabeled protein partner. The unlabeled protein was added in eight steps in order to reach the following ratios and saturation of labeled/unlabeled protein: 1:0.25, 1:0.5, 1:0.75, 1:1, 1:1.25, 1:1.5, 1:1.75, 1:2. Combined chemical shift perturbations (CSPs) after binding were calculated using the equation: $\Delta\delta_{ppm} = \sqrt{(\Delta\delta_{HN})^2 + (\frac{\Delta\delta_N}{5})^2}$ (Garrett et al., 1997; Baker et al., 2006; Marousis et al., 2018).

Ubiquitination Assay

For the ubiquitination assays the Arkadia$^{876–994}$ polypeptide bearing the R957C mutation was used. Briefly, ubiquitination assay was performed by incubating 0.5 μM hUbe1 enzyme (Boston Biochem), 5 μM UBCH5B, 100–150 μM ubiquitin and 15 μM wt/R957C Arkadia$^{876-994}$ in 20 mM Tris-HCl pH 7.5, 50 mM NaCl, 5 mM ATP, 2 mM $MgCl_2$ and 2 mM DTT at 37°C for 0–60 min. Time points were collected after the addition of ATP and the reactions were stopped by addition of SDS loading buffer. Samples were resolved by 15% SDS-PAGE and visualized by western blotting for ubiquitin using the anti-Ub antibody P4D1 (Santa Cruz Biotechnologies, SCBT) and the goat anti-mouse m-IgGk BP-HRP sc-516102 (SCBT).

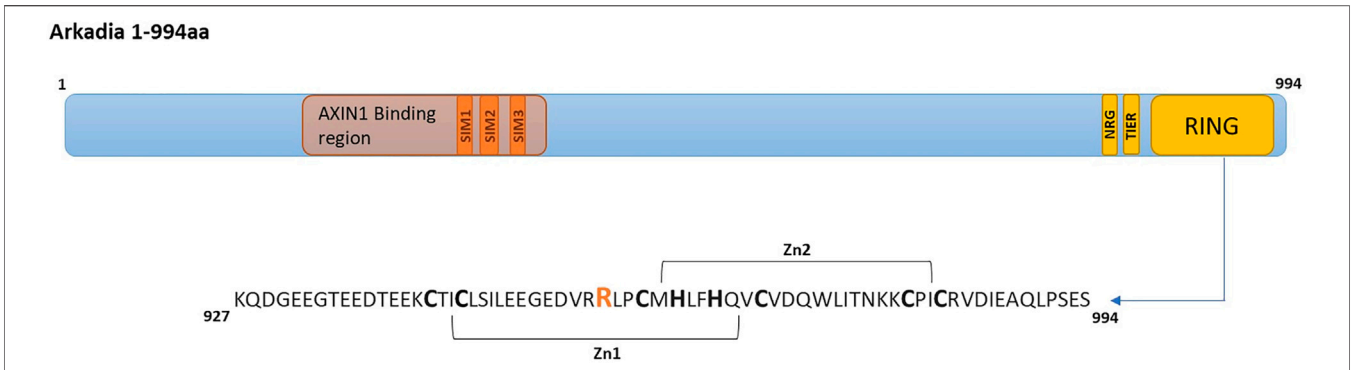

FIGURE 1 | Arkadia E3 ubiquitin ligase domain organization and RING mutated sequence. SIMs (SUMO interacting motifs) (Poulsen et al., 2013) colored in orange, NRG, TIER segments and RING domain are colored in yellow.

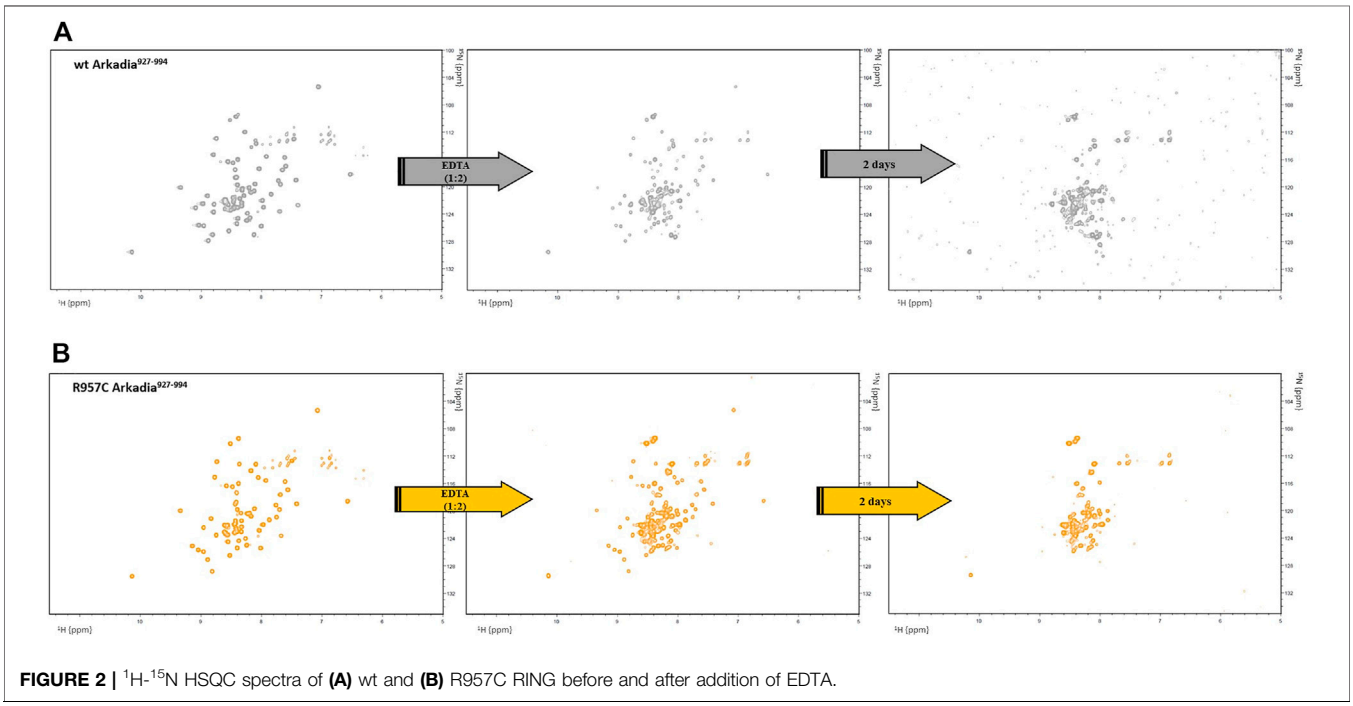

FIGURE 2 | ^{1}H-^{15}N HSQC spectra of **(A)** wt and **(B)** R957C RING before and after addition of EDTA.

Luciferase Assays

Luciferase assays were conducted in HEK293T cells (Cancer Research UK). using the Dual Luciferase reporter system (Promega), as previously described by Levy at al. 2007. Briefly, cells were transfected with the appropriate combination of promoter-reporter constructs and expression plasmids using FuGENE transfection reagent (Promega) and were cultured for 24 h after transfection. Luciferase activities in the cell lysates were measured following the manufacturer's protocol. The experiments were repeated at least three times.

RESULTS

The SNP and GO and I-Mutant pathogenicity prediction tools indicated, with high probability, that the R957C mutation in the Arkadia RING domain (**Figure 1**) is a disease-causing mutation (**Supplementary Table S1**). To determine the effect of the R957C mutation on Arkadia's RING domain structure, stability and interaction properties NMR studies were performed. In addition, auto-ubiquitination and luciferase assays were carried out to investigate the biochemical activity of the R957C mutant.

Determination of R957C Arkadia$^{927–994}$ Zinc Ions Content and Stability Upon Addition of Ethylenediaminetetraacetic Acid.

Atomic absorption spectroscopy was performed on R957C mutant to measure the Zn (II) content. A protein sample of 0.3 mM gave zinc concentration of 0.71 mM. These concentrations correspond to two zinc ions per protein molecule.

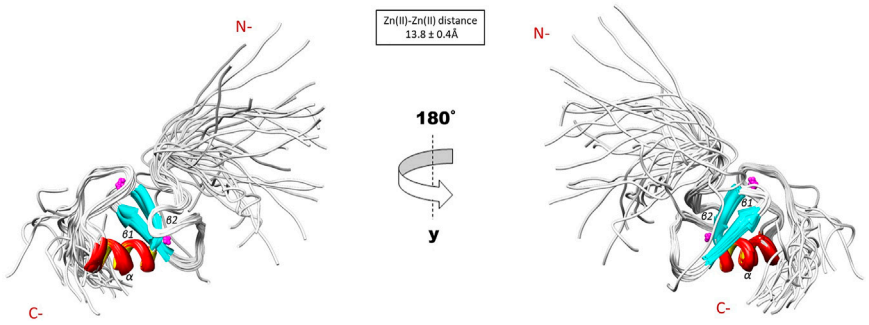

FIGURE 3 | Solution structure of the R957C RING domain as represented by the final family of 30 energy minimized models.

According to the structural role of the Zn (II) in RING domains, addition of the metal-chelating agent EDTA to a final ratio 1:2 (Zn: EDTA) led to the complete unfolding of the protein after 48 h as indicated by the loss of chemical shift dispersion in the ^{1}H-^{15}N HSQC spectra (**Figure 2A**). Similarly, addition of EDTA to the wt RING domain caused complete unfolding of the protein after 48 h of incubation with the chelating agent (**Figure 2B**). These results suggest that the addition of one more Cys residue in the RING domain of Arkadia has not affected the zinc ions binding and the stability of the mutated protein. According to the C_{α} (56 ppm) and C_{β} (27 ppm) chemical shifts Cys957 in reduced state and does not participate in zinc ions binding or in disulfide bond formations (Sharma and Rajarathnam, 2000).

Solution Structure and ^{15}N Relaxation Studies of the R957C Arkadia$^{927–994}$ RING Domain

The NMR solution structure of the R957C mutant was determined based on a total of 906 NOEs distance constraints. The final family energy minimized NMR models contains two β-strands, namely β1 (Val955-Leu958) and β2 (His962-His965) forming an antiparallel β-sheet, two zinc binding loops and a 3-turn α-helix encompassing the residues Gln966-Thr975 (**Figure 3**). The two β-strands of the R957C mutant are one amino acid longer compared to wt RING Val955-Arg957 and Leu963-His965 β-strands. The *a*-helix observed in the R957C mutant is identical to the wt Arkadia RING helix in length and sequence position (Chasapis et al., 2012). The secondary structure elements of the R957C RING domain exhibit a $\beta\beta\alpha$ topology as observed in the wt Arkadia, RNF24 (PDB: 2EP4) and RNF168 RING domains (Zhang et al., 2013). The average distance between the two metal centers in the final 30 NMR models is 13.8 ± 0.4 Å (**Figure 3**). A detailed view of the coordination of the two Zn (II) ions is shown in **Figure 4**. The overall backbone and heavy atom RMSDs of the 30 energy-minimized models of R957C are 1.22 ± 0.48 Å and 2.08 ± 0.45 Å, respectively, for the core region between Glu939 and Ile986. Quality assessment of the final models reveals that 100% of the non-glycine/non-proline residues fall into favorable or allowed regions of the φ/ψ dihedral angle space in the Ramachandran plot (**Supplementary Table S2**).

^{15}N relaxation studies of R957C Arkadia$^{927–994}$ revealed that mutant is a monomer in solution since the correlation time for isotropic tumbling measured based on the R_2/R_1 ratio is 4.69 ns corresponding to a MW of ~7.81 kDa (theoretical MW of R957C RING is 7767 Da). In addition, model-free analysis of ^{15}N relaxation data as implemented in the Tensor2 program showed that the core exhibited a rather rigid structure (**Figure 5**). The order parameters of the region Lys941-Ile986 (average $S^2 = 0.77$) are higher than those of the N- and C-terminal residues (Lys927-Glu940 and Glu987-Ser994, respectively) closely resembling the ^{15}N-relaxation properties of the wt Arkadia RING (Chasapis et al., 2012; Birkou et al., 2017) (**Figure 5**). The above data clearly support that there are no significant differences in the structure and dynamic properties between wt and R957C Arkadia RING domain.

Interaction Studies of R957C Arkadia$^{927–994}$ Really Interesting New Gene Domain and UBCH5B

To determine whether the R957C mutation affects the interaction of the RING with E2 we used NMR driven interaction studies of ^{15}N R957C with the E2 enzyme UBCH5B, which functions as a partner of Arkadia *in vitro* (Mavrakis et al., 2007). The analysis showed that the resonances exhibit either fast or intermediate exchange behavior, suggesting a moderate affinity for the two proteins. The largest chemical shift changes were observed for the sequential stretches Cys942–Thr943, Cys945-Ile947, Val955-Leu958, Phe964-His965, Val967-Asp970 and Trp972-Asn976 and for residue Ile981 (**Figure 6A**). The first three sequential stretches comprise the first and the third Zn(II)-binding motif, while the fourth and fifth are located at the C-end of the α-helix Gln966-Thr975. The above regions are essentially identical with those exhibiting the largest chemical shift changes upon interaction with UBCH5B in the native RING (Chasapis et al., 2012; Birkou et al., 2017).

The addition of unlabeled R957C RING to ^{15}N-labeled UBCH5B resulted in the loss of four UBCH5B amide resonances (Ser94, Ile99, Ser100 and Leu103) in the ^{1}H–^{15}N HSQC spectra during the titration. The largest CSPs in UBCH5B were observed for residues Ala2-Leu3, Ile6-Glu9, Asn11-Asp12, Arg15-Asp16 (N-terminal helix α1) Ala19,

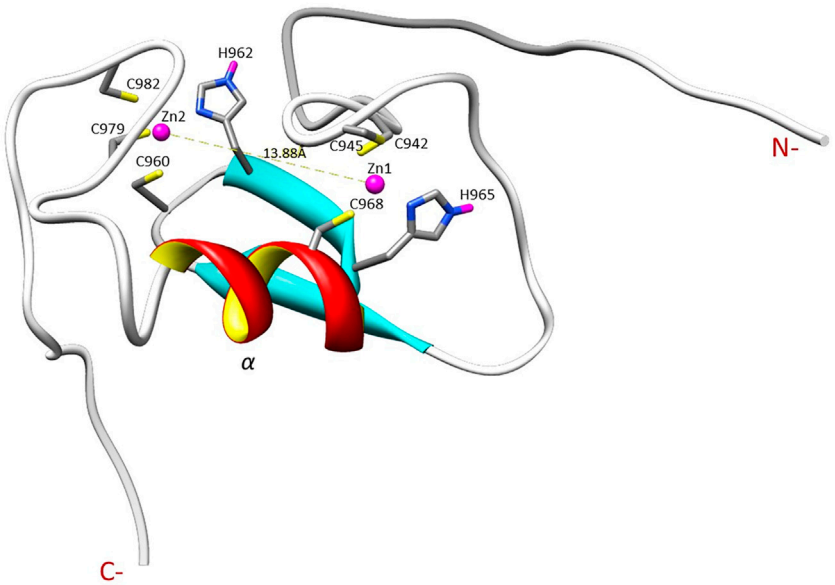

FIGURE 4 | Illustration of the zinc ions binding sites and the Zn ions distance in the R957C RING domain as appeared in the first model of the family.

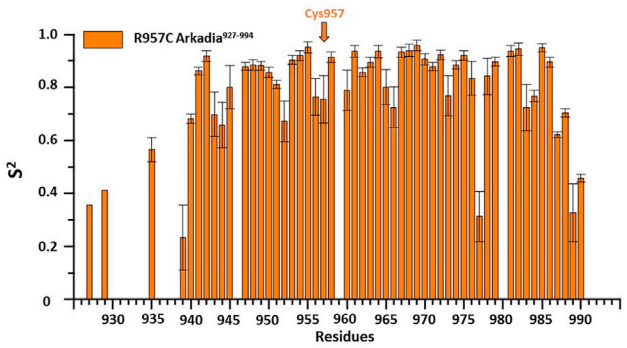

FIGURE 5 | Order parameters (S^2) derived from ^{15}N-relaxation measurements [R_1, R_2 and heteronuclear ($^1H^N$)– ^{15}N NOE] for R957C Arkadia$^{927–994}$ domain at 600 MHz and 298 K.

Ser22 (β1 strand), Ile37 (β2 strand), Phe62-Lys63 (loop L1 between the β3 and β4 strands), Ser91-Glu92, Ala96-Thr98 (loop L2b), Lys101, and Leu104 (the first half of helix α2) (**Figure 6B**). All these residues were also found to participate in wt Arkadia RING–UBCH5B interactions (Chasapis et al., 2012; Birkou et al., 2017), strongly suggesting that UBCH5B utilizes the same interface for its interaction with the Arkadia RING mutant R957C.

The above data show that R957C Arkadia mutant and UBCH5B E2 enzyme interact via the surfaces that are expected to be involved in a canonical E2-E3 RING interaction (Chasapis et al., 2012; Gundogdu and Walden, 2019). However, it seems that the CSPs observed for the mutant are somewhat larger compared with those observed for the wt-E2 interaction [0.09/0.06 ppm for the R957C and wt RING, respectively and 0.042/0.038 ppm for the interacting UBCH5B with R957C and wt RING, respectively (**Figure 5** and Birkou et al., 2017)].

Effect of the R957C Mutation on Arkadia's Function

To examine the actual activity of the mutant protein we developed an *in vitro* ubiquitination assay. For the auto-ubiquitination assays, wt or mutated polypeptides containing NRG and TIER segments along with the RING domain (Arkadia$^{876–994}$) were used, because the RING domain alone does not auto-ubiquitinate. We performed the assay in the presence of E2 UBCH5B and ubiquitin. Immunoblot analysis of these reactions, with anti-ubiquitin antibodies, showed that both wt and R957C Arkadia $^{876–994}$ were functional in this *in vitro* auto-ubiquitination assay, as indicated from the ladder-like bands that appeared after addition of ATP to the reaction mixture (**Figure 7**) due to the formation of a polyubiquitin chain on the substrates wt or R957C Arkadia $^{876–994}$. We found that the R957C Arkadia$^{876–994}$ exhibit significantly rapid and enhanced autoubiquitination (**Figure 7**), indicating a gain of function and not loss of function.

Moreover, R957C Arkadia's functionality was tested *in vivo* by luciferase assays. Luciferase experiments were carried out in HEK293T cells using the TGF-β SMAD-dependent reporter $CAGA_{12}$-Luc, empty plasmid, the full length wt, C937A and R957C Arkadia proteins as described in Birkou et al., 2017. Overexpression of the R957C Arkadia protein enhanced reporter expression a slightly higher than that of the wt Arkadia, however, this was not statistical significant in this assay (**Figure 8**). These *in vitro* and *in vivo* results show that the R957C mutation on the RING domain does not reduce the activity of Arkadia as E3 ubiquitin ligase.

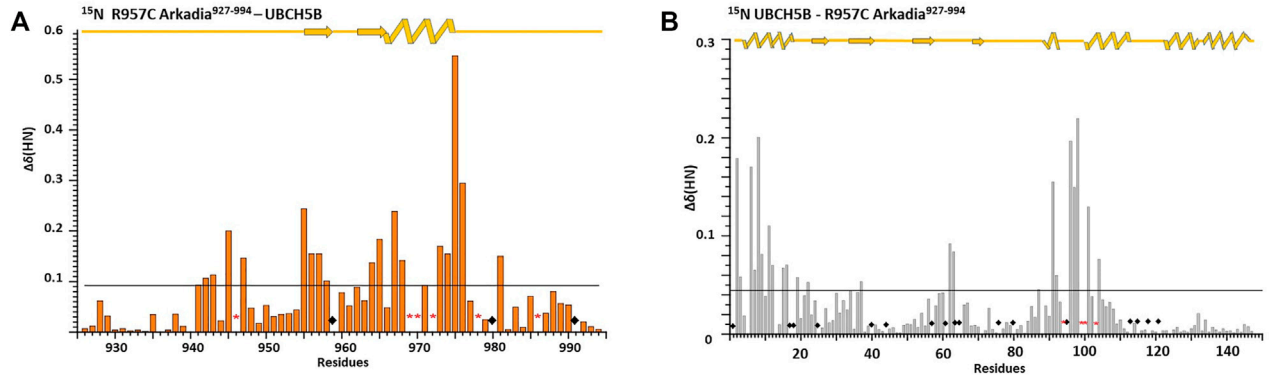

FIGURE 6 | Interaction of E3 Ub ligase Arkadia RING with the E2 UBCH5B enzyme monitored by ^{1}H-^{15}N HSQC spectra. Diagram **(A)** illustrates the total CSPs measured at 1:2 M ratio for ^{15}N R957C Arkadia/^{14}N UBCH5B (left), while diagram **(B)** illustrates the CSPs for ^{15}N UBCH5B/^{14}N R957C Arkadia RING. Straight lines indicate the applied thresholds of CSPs representing the respective average CSP (Kornhaber et al., 2006). ◆represents proline residues and residues with no information.

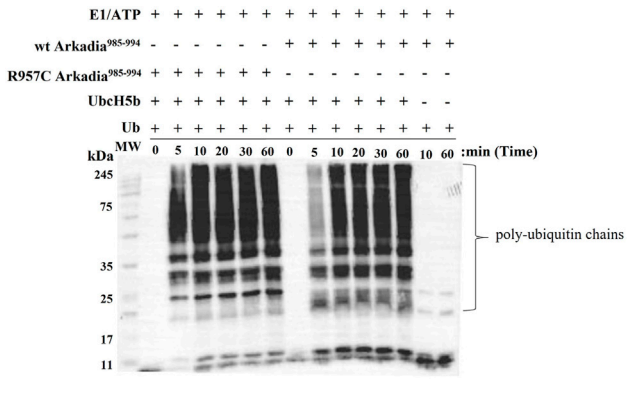

FIGURE 7 | Western blot of *in vitro* ubiquitination assays of R957C Arkadia$^{985–994}$ (left) and wt Arkadia$^{985–994}$ (right).

DISCUSSION

Mutations within the RING domain of E3 ligases such as BRCA1 are correlated with high risk of cancer (Ruffner et al., 2001). Usually, most of the SNPs cause missense mutations that are neutral or deleterious and unclassified variants. Deleterious SNPs/mutations cause phenotypic differences leading to various types of cancer (Ramensky et al., 2002). Here, we study the structural and functional consequences of the R957C SNP in the RING domain of Arkadia that was identified in normal and tumor tissues of a colorectal cancer patient.

Does the R957C mutation affect the solution structure and dynamics of Arkadia?

The NMR structure of the R957C mutant of Arkadia closely agrees with the NMR structure of the native RING domain (PDB: 2KIZ). Overlay of the first model of R957C RING domain with the first model of wt RING domain shows that there is no significant difference in the overall topology (**Figure 9**). Both contain a 3-turn α-helix, although in the final family of the R957C mutant 10 out of 30 models show a shightly distorted third turn of the α-helix due to the lack of helix-diagnostic NOEs. Replacement of Arg957 with Cys results in a slight bending of the $\beta 1$-strand which is slightly bended, affecting the conformation of a loop that comprises the acidic residues Glu936 and Asp937. In wt, the positively charged NH_1 and NH_2 atoms of Arg957 interact electrostatically with the carboxyl groups of Glu936 and Asp937, respectively, leading to a compact and less flexible N-terminus (**Figure 10**). This electrostatic interaction is abolished in the R957C mutant resulting in an increased flexibility of the N-terminus as indicated by the backbone mobility data (**Figure 5**) and by the distances between the carboxyl groups of Glu936 and Asp937 and the sulfur atom of Cys957. As far as the mobility of the RING core is concerned, the ^{15}N relaxation measurements show that the replacement of Arg957 with Cys has essentially no effects on the ps–ns time scale (**Supplementary Figure S2**).

Comparison of the NMR structure of the R957C mutant with the RING domains of RNF24 (PDB: 2EP4) and RNF168 (PDB: 4GB0) that exhibit a $\beta\beta\alpha$ topology reveals the similarity of the Arkadia mutant with the RING domain of RNF24 (**Figure 11**). Interestingly, both domains have β-strands, that are four residues long and adopt the same orientation (**Figures 11A,B**). Moreover, their β-strands do not exhibit sequence homology but rather physicochemical similarity in their amino acids composition (**Figure 12**).

How does the R957C mutation affects Arkadia's interaction properties and function?

The R957C Arkadia mutant is capable to interact with the E2 enzyme UBCH5B as shown by titration experiments monitored through ^{1}H-^{15}N HSQC spectra and auto-ubiquitination assays (**Figure 6** and **Figure 7**). Interestingly, the observed CSPs suggest a native-like interaction, involving the two metal-binding loops and parts of the α-helix of the mutated RING (**Figure 6**). A similar surface, formed by the same RING structural elements participates in the E2-E3 interaction of BRCA1 (Brzovik et al., 2003), c-Cbl (Zheng et al., 2000) and RNF38 (Buetow et al., 2015) RING domains. Additionally, the UBCH5B regions that

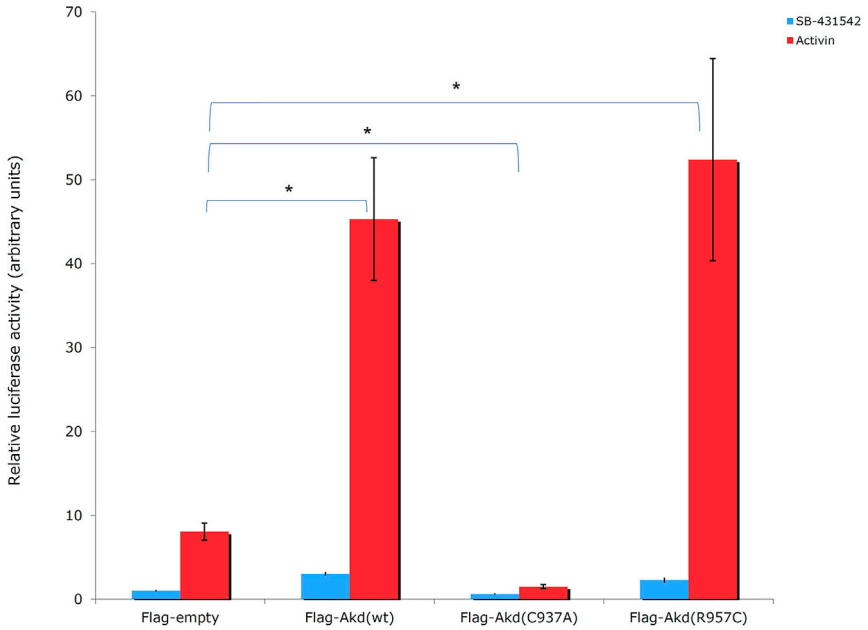

FIGURE 8 | Luciferase $CAGA_{12}$ reporter assay values. Activin: activator of the TGF-β pathway. SB-431542: inhibitor of TGF-β superfamily type I activin receptor-like kinase receptors. C937A is a ligase defective mutant, which forms mixed dimers (Erker et al., 2013) with the endogenous wt protein leading on suppression of wt function in HEK293T cells (Sharma et al., 2011).

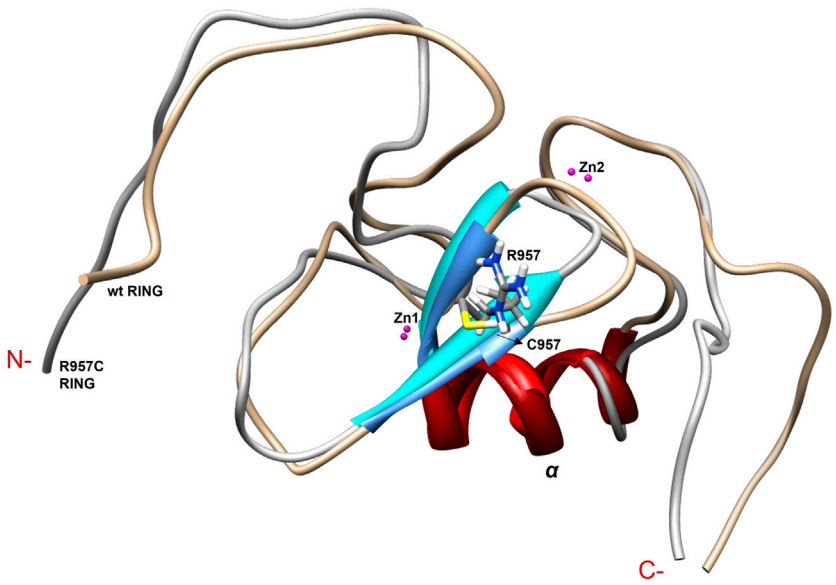

FIGURE 9 | Overlay of the first model of the NMR family models of wt and R957C mutant of Arkadia RING.

participate in the E2-E3 interaction are the same as those observed in the vast majority of E2-E3 pairs (Gundogdu and Walden, 2019).

Thus, the R957 mutation which is located on the $\beta 1$-strand, does not seem to have an impact on the RING domain interaction properties. Since R957 is not considered as a residue that is crucial for E2-E3 interaction (Chasapis et al., 2012) its mutation to cysteine is not expected to disrupt the interaction with the E2 enzyme UBCH5B. Generally, it is well established that mutations of conserved residues in RING domains disrupt E2 interaction and result in the loss of E3 ligase enzymatic activity (Deshaies and Joazeiro et al., 2009). For example, mutation of the conserved tryptophan (Trp) in the α-helix of most RING domains disrupts the recruitment of E2 and this is sufficient to abolish the ligase

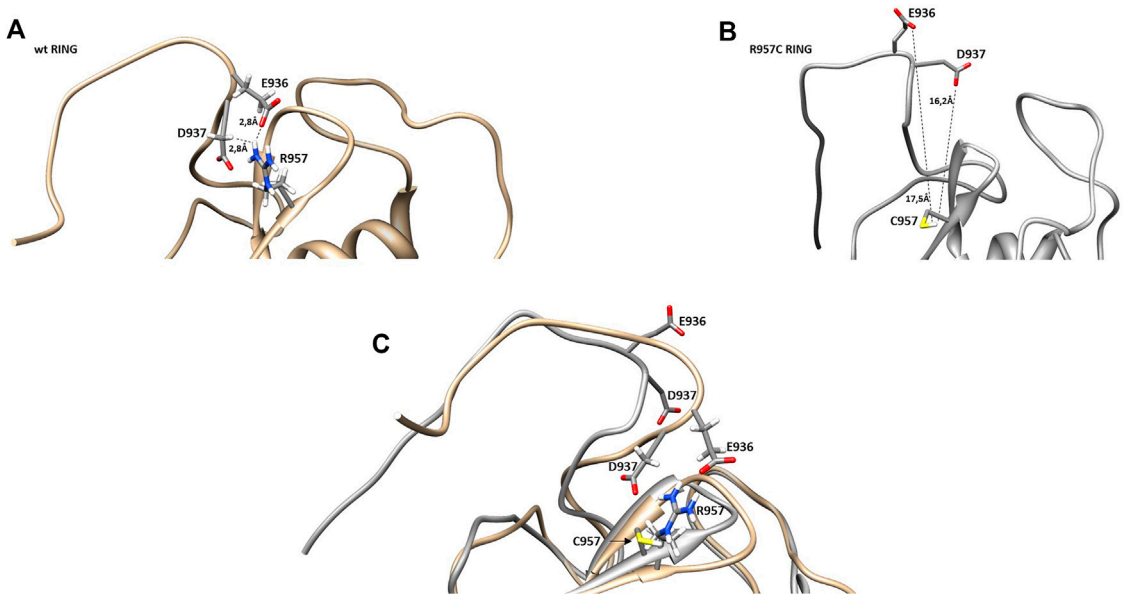

FIGURE 10 | Representation of β-sheets region in **(A)** wt and **(B)** R957C RING domain. **(C)** Overlay of the β-sheet region of wt (gold) and R957C mutant (white).

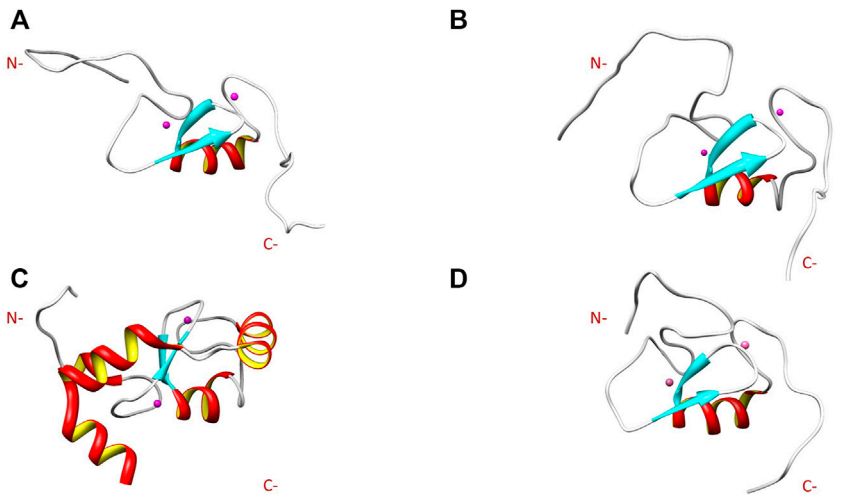

FIGURE 11 | NMR solution structure of **(A)** RNF24 (PDB: 2EP4), **(B)** R957C Arkadia, **(C)** RNF168 (PDB: 4GB0) and **(D)** wt Arkadia (PDB: 2KIZ) RING domains.

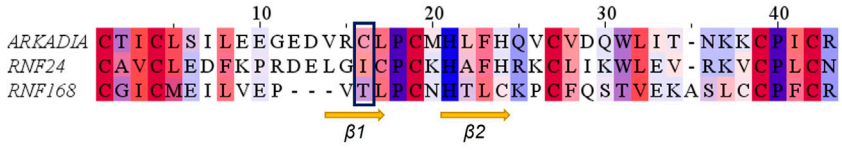

FIGURE 12 | Alignment of the RING domains of R957C Arkadia, RNF24 and RNF168 E3 ubiquitin ligases. Coloring by residue hydrophobicity and sequence conservation.

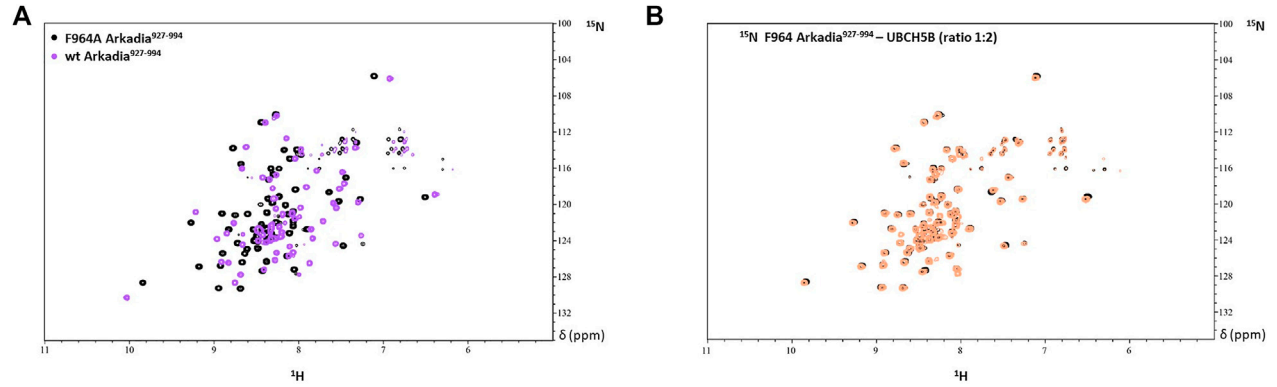

FIGURE 13 | (A) Overlay of the ^{1}H-^{15}N HSQC spectrum of F964A and wt Arkadia RING domain (black and purple, respectively). **(B)** Overlay of the ^{1}H-^{15}N HSQC spectrum of F964A before (black) and after (orange) the addition of UBCH5B to a 1:2 ratio.

activity. A recent study on Arkadia's RING domain revealed that mutation of Trp972 to arginine affected Arkadia's interaction interface and abolished the E3 ligase function (Birkou et al., 2017). Another example is the mutation of Phe964 in the Arkadia RING domain. This phenylalanine is conserved in RING-H2 ligases and precedes the fifth zinc binding residue. Mutation of F964 to alanine (Ala, A) changed the folding of the RING domain (**Figure 13A**), which resulted in the disruption of E2-E3 interaction (**Figure 13B**).

The auto-ubiquitination assay revealed R957C mutation is a gain of function mutation that enhances the enzymatic activity of the RING (**Figure 7**), and consistent to this, the luciferase assays shows that the R957C mutation exhibits slightly increased activity of Arkadia although it is not statistically significant here (**Figure 8**). The TGF-β pathway has tumor suppressive properties in normal tissues and promotes metastasis within tumors. Similarly, Arkadia enhances TGF-β signaling and exhibit the same bidiretional properties (Sharma et al., 2011; Briones-Orta et al., 2013). Indeed, Sharma et al., 2011 has shown that reduction of Arkadia in normal cells increases significantly the susceptibility to cancer. Furthermore, Le Scolan et al., 2008 and Briones-Orta et al., 2013 performing loss of function experiments in tumor cell lines have shown that Arkadia supports metastatic phenotypes via a TGF-β dependent manner. Our data presented here support the pro-tumorigenic functions of Arkadia in a context-cell type-dependent manner because it was found enriched in tumor compared to the adjacent normal tissue. However, it is possible that the presence of the R957C mutation in the adjacent normal cells represent infiltration of a metastatic tumor cells and not a polymorphism. More experiments are necessary to verify the role of this hyperactive mutant Arkadia within tumor cells.

CONCLUSION

Analysis of the consequences of SNPs on the 3D protein structure is beneficial in understanding both their function and their role in diseases. SNPs at the level of proteins usually affect their activity, binding/association, assembly, and rearrangement and promote their aggregation (Bhattacharya et al., 2017). Most studies analyzing the effects of SNPs on the 3D protein structure are restricted to specific diseases. The goal of this study is to provide experimental data about an SNP that was found in CRC and adjacent normal tissue samples of a patient and its effects on structure, stability, binding properties, and function of the Arkadia RING domain. Our NMR structural data show that the R957C mutation, in the *β1* strand of this domain is located far away from the two Zn (II) binding motifs. It does not interfere with the metal binding and does not alter the geometry of the cross-brace metal coordination; moreover, it does not change the conformation of the E2 docking surface of the RING domain. The R957C mutation does not lead to conformational change in the *β1–β2* surface, so no significant change of the overall conformational features of the RING domain is observed, and the same holds for the backbone dynamics of the RING core. Finally, the functional assays (**Figures 7**, **8**) also showed that Arkadia does not lose its E3 ligase activity, despite the R957C mutation. It is worth mentioning that auto-ubiquitination assays indicated a gain of function effect of the R957C mutation of Arkadia, while the luciferase functional assay showed milder hyperactivity. The finding that this SNP was present in normal tissue and enriched within the tumor, suggests that it segrates with the tumor and possibly with infiltration of adjacent tissue by metastatic tumor cells.

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 in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

Conceptualization: VE, GS, Methodology: MB, KM, VR, and AT, Validation: MB, KM, KB, DB, VE, and GS, Writing-original draft

preparation: MB, Writing-review and editing: KM, DB, VE, and GS, Resources: GS. All authors reviewed the results and approved the final version of the paper.

FUNDING

This research is co-financed by Greece and the European Union (European Social Fund—ESF) through the Operational Programme «Human Resources Development, Education and Lifelong Learning» in the context of the project "Reinforcement of Postdoctoral Researchers—2nd Cycle" (MIS5033021), implemented by the State Scholarships Foundation (IKY). The work was also supported by the INSPIRED (MIS 5002550) which is implemented under the Action "Reinforcement of the Research and Innovation Infrastructure," funded by the Operational Program "Competitiveness, Entrepreneurship and Innovation" (NSRF 2014–2020) and co-financed by Greece and the European Union (European Regional Development Fund).

ACKNOWLEDGMENTS

EU FP7 REGPOT CT-2011-285950—"SEE-DRUG" project is acknowledged for the purchase of UPAT's 700 MHz NMR equipment. State Scholarships Foundation-MIS 5033021 and INSPIRED-MIS 5002550 are acknowledged for the financial support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2022.844129/full#supplementary-material

REFERENCES

Baker, K. A., Hilty, C., Peti, W., Prince, A., Pfaffinger, P. J., Wider, G., et al. (2006). NMR-derived Dynamic Aspects of N-type Inactivation of a Kv Channel Suggest a Transient Interaction with the T1 Domain. *Biochemistry* 45, 1663–1672. doi:10.1021/bi0516430

Bartels, C., Xia, T.-h., Billeter, M., Güntert, P., and Wüthrich, K. (1995). The Program XEASY for Computer-Supported NMR Spectral Analysis of Biological Macromolecules. *J. Biomol. NMR* 6, 1–10. doi:10.1007/bf00417486

Bhattacharya, R., Rose, P. W., Burley, S. K., and Prlić, A. (2017). Impact of Genetic Variation on Three Dimensional Structure and Function of Proteins. *PLoS One* 12 (3), e0171355. doi:10.1371/journal.pone.0171355

Birkou, M., Chasapis, C. T., Marousis, K. D., Loutsidou, A. K., Bentrop, D., Lelli, M., et al. (2017). A Residue Specific Insight into the Arkadia E3 Ubiquitin Ligase Activity and Conformational Plasticity. *J. Mol. Biol.* 429, 2373–2386. doi:10.1016/j.jmb.2017.06.012

Bravou, V., Antonacopoulou, A., Papadaki, H., Floratou, K., Stavropoulos, M., Episkopou, V., et al. (2009). TGF-β Repressors SnoN and Ski Are Implicated in Human Colorectal Carcinogenesis. *Anal. Cell Pathol.* 31, 41–51. doi:10.3233/clo-2009-046010.1155/2009/975602

Briones-Orta, M. A., Levy, L., Madsen, C. D., Das, D., Erker, Y., Sahai, E., et al. (2013). Arkadia Regulates Tumor Metastasis by Modulation of the TGF-β Pathway. *Cancer Res.* 73, 1800–1810. doi:10.1158/0008-5472.CAN-12-1916

Brzovic, P. S., Keeffe, J. R., Nishikawa, H., Miyamoto, K., Fox, D., Fukuda, M., et al. (2003). Binding and Recognition in the Assembly of an Active BRCA1/BARD1 Ubiquitin-Ligase Complex. *Proc. Natl. Acad. Sci.* 100 (10), 5646–5651. doi:10.1073/pnas.0836054100

Brzovic, P. S., Rajagopal, P., Hoyt, D. W., King, M.-C., and Klevit, R. E. (2001). Structure of a BRCA1-BARD1 Heterodimeric RING-RING Complex. *Nat. Struct. Biol.* 8, 833–837. doi:10.1038/nsb1001-833

Buetow, L., Gabrielsen, M., Anthony, N. G., Dou, H., Patel, A., Aitkenhead, H., et al. (2015). Activation of a Primed RING E3-E2-Ubiquitin Complex by Non-covalent Ubiquitin. *Mol. Cel* 58 (2), 297–310. doi:10.1016/j.molcel.2015.02.017

Capriotti, E., Calabrese, R., Fariselli, P., Martelli, P., Altman, R. B., and Casadio, R. (2013). WS-SNPs&GO: a Web Server for Predicting the Deleterious Effect of Human Protein Variants Using Functional Annotation. *BMC Genomics* 14, S6–S7. doi:10.1186/1471-2164-14-S3-S6

Capriotti, E., Fariselli, P., and Casadio, R. (2005). I-Mutant2.0: Predicting Stability Changes upon Mutation from the Protein Sequence or Structure. *Nucleic Acids Res.* 33, W306–W310. doi:10.1093/nar/gki375

Chan, C.-H., Morrow, J. K., Li, C.-F., Gao, Y., Jin, G., Moten, A., et al. (2013). Pharmacological Inactivation of Skp2 SCF Ubiquitin Ligase Restricts Cancer Stem Cell Traits and Cancer Progression. *Cell* 154, 556–568. doi:10.1016/j.cell.2013.06.048

Chasapis, C. T., Kandias, N. G., Episkopou, V., Bentrop, D., and Spyroulias, G. A. (2012). NMR-based Insights into the Conformational and Interaction Properties of Arkadia RING-H2 E3 Ub Ligase. *Proteins* 80, 1484–1489. doi:10.1002/prot.24048

Deshaies, R. J., and Joazeiro, C. A. P. (2009). RING Domain E3 Ubiquitin Ligases. *Annu. Rev. Biochem.* 78, 399–434. doi:10.1146/annurev.biochem.78.101807.093809

Dosset, P., Hus, J.-C., Blackledge, M., and Marion, D. (2000). Efficient Analysis of Macromolecular Rotational Diffusion from Heteronuclear Relaxation Data. *J. Biomol. NMR* 16, 23–28. doi:10.1023/a:1008305808620

Erker, Y., Neyret-Kahn, H., Seeler, J. S., Dejean, A., Atfi, A., and Levy, L. (2013). Arkadia, a Novel SUMO-Targeted Ubiquitin Ligase Involved in PML Degradation. *Mol. Cel Biol* 33 (11), 2163–2177. doi:10.1128/MCB.01019-12

Ferentz, A. E., and Wagner, G. (2000). NMR Spectroscopy: A Multifaceted Approach to Macromolecular Structure. *Quart. Rev. Biophys.* 33, 29–65. doi:10.1017/s0033583500003589

Fiala, C., and Diamandis, E. P. (2020). Mutations in normal Tissues-Some Diagnostic and Clinical Implications. *BMC Med.* 18, 1–9. doi:10.1186/s12916-020-01763-y

Freemont, P. S. (1993). The RING Finger. *Ann. NY Acad. Sci.* 684, 174–192. doi:10.1111/j.1749-6632.1993.tb32280.x

Garrett, D. S., Seok, Y.-J., Peterkofsky, A., Clore, G. M., and Gronenborn, A. M. (1997). Identification by NMR of the Binding Surface for the Histidine-Containing Phosphocarrier Protein HPr on the N-Terminal Domain of Enzyme I of the *Escherichia coli* Phosphotransferase System. *Biochemistry* 36, 4393–4398. doi:10.1021/bi970221q

Gundogdu, M., and Walden, H. (2019). Structural Basis of Generic versus Specific E2-RING E3 Interactions in Protein Ubiquitination. *Protein Sci.* 28, 1758–1770. doi:10.1002/pro.3690

Güntert, P., Mumenthaler, C., and Wüthrich, K. (1997). Torsion Angle Dynamics for NMR Structure Calculation with the New Program DYANA. *J. Mol. Biol.* 273 (1), 283–298. doi:10.1006/jmbi.1997.1284

Hershko, A., Heller, H., Elias, S., and Ciechanover, A. (1983). Components of Ubiquitin-Protein Ligase System. Resolution, Affinity Purification, and Role in Protein Breakdown. *J. Biol. Chem.* 258 (13), 8206–8214. doi:10.1016/s0021-9258(20)82050-x

Keller, R. (2004). *The Computer Aided Resonance Assignment Tutorial*. 1st Edition. Goldau, Switzerland: CANTINA Verlag.

Kornhaber, G. J., Snyder, D., Moseley, H. N. B., and Montelione, G. T. (2006). Identification of Zinc-Ligated Cysteine Residues Based on 13Cα and 13Cβ Chemical Shift Data. *J. Biomol. NMR* 34, 259–269. doi:10.1007/s10858-006-0027-5

Kumar, A., Chaugule, V. K., Condos, T. E. C., Barber, K. R., Johnson, C., Toth, R., et al. (2017). Parkin-phosphoubiquitin Complex Reveals Cryptic Ubiquitin-Binding Site Required for RBR Ligase Activity. *Nat. Struct. Mol. Biol.* 24, 475–483. doi:10.1038/nsmb.3400

Lakshmanan, M., Bughani, U., Duraisamy, S., Diwan, M., Dastidar, S., and Ray, A. (2008). Molecular Targeting of E3 Ligases - a Therapeutic Approach for Cancer. *Expert Opin. Ther. Targets* 12 (7), 855–870. doi:10.1517/14728222.12.7.855

Le Scolan, E., Zhu, Q., Wang, L., Bandyopadhyay, A., Javelaud, D., and Mauviel, A. (2008). Transforming Growth Factor-Beta Suppresses the Ability of Ski to Inhibit Tumor Metastasis by Inducing its Degradation. *Cancer Res.* 68 (9), 3277–85. doi:10.1158/0008-5472.CAN-07-6793

Levy, L., Howell, M., Das, D., Harkin, S., Episkopou, V., and Hill, C. S. (2007). Arkadia Activates Smad3/Smad4-dependent Transcription by Triggering Signal-Induced SnoN Degradation. *Mol. Cel. Biol.* 27 (17), 6068–6083. doi:10.1128/MCB.00664-07

Marousis, K. D., Tsika, A. C., Birkou, M., Matsoukas, M.-T., and Spyroulias, G. A. (2018). Lead Identification through the Synergistic Action of Biomolecular NMR and In Silico Methodologies. *Methods Mol. Biol.* 1824, 299–316. doi:10.1007/978-1-4939-8630-9_18

Mavrakis, K. J., Andrew, R. L., Lee, K. L., Petropoulou, C., Dixon, J. E., Navaratnam, N., et al. (2007). Arkadia Enhances Nodal/TGF-β Signaling by Coupling Phospho-Smad2/3 Activity and Turnover. *Plos Biol.* 5, e67. doi:10.1371/journal.pbio.0050067

Moll, U. M., and Petrenko, O. (2003). The MDM2-P53 Interaction. *Mol. Cancer Res.* 1, 1001–1008.

Nagano, Y., Mavrakis, K. J., Lee, K. L., Fujii, T., Koinuma, D., Sase, H., et al. (2007). Arkadia Induces Degradation of SnoN and C-Ski to Enhance Transforming Growth Factor-β Signaling. *J. Biol. Chem.* 282, 20492–20501. doi:10.1074/jbc.M701294200

Nelson, A. C., and Holt, J. T. (2010). Impact of RING and BRCT Domain Mutations on BRCA1 Protein Stability, Localization and Recruitment to DNA Damage. *Radiat. Res.* 174, 1–13. doi:10.1667/RR1290.1

Oh, J.-H., and Sung, C. O. (2020). Comprehensive Characteristics of Somatic Mutations in the normal Tissues of Patients with Cancer and Existence of Somatic Mutant Clones Linked to Cancer Development. *J. Med. Genet.* 58, 433–441. doi:10.1136/jmedgenet-2020-106905

Pearlman, D. A., Case, D. A., Caldwell, J. W., Ross, W. S., Cheatham, T. E., DeBolt, S., et al. (1995). AMBER, a Package of Computer Programs for Applying Molecular Mechanics, normal Mode Analysis, Molecular Dynamics and Free Energy Calculations to Simulate the Structural and Energetic Properties of Molecules. *Comp. Phys. Commun.* 91, 1–41. doi:10.1016/0010-4655(95)00041-d

Pickart, C. M. (2001). Mechanisms Underlying Ubiquitination. *Annu. Rev. Biochem.* 70, 503–533. doi:10.1146/annurev.biochem.70.1.503

Poulsen, S. L., Hansen, R. K., Wagner, S. A., van Cuijk, L., van Belle, G. J., Streicher, W., et al. (2013). RNF111/Arkadia Is a SUMO-Targeted Ubiquitin Ligase that Facilitates the DNA Damage Response. *J. Cel Biol.* 201 (6), 797–807. doi:10.1083/jcb.201212075

Ramensky, V., Bork, P., and Sunyaev, S. (2002). Human Non-synonymous SNPs: Server and Survey. *Nucleic Acids Res.* 30 (17), 3894–3900. doi:10.1093/nar/gkf493

Risques, R. A., and Kennedy, S. R. (2018). Aging and the Rise of Somatic Cancer-Associated Mutations in normal Tissues. *Plos Genet.* 14, e1007108. doi:10.1371/journal.pgen.1007108

Rosen, E. M. (2013). BRCA1 in the DNA Damage Response and at Telomeres. *Front. Genet.* 4, 1–14. doi:10.3389/fgene.2013.00085

Ruffner, H., Joazeiro, C. A. P., Hemmati, D., Hunter, T., and Verma, I. M. (2001). Cancer-predisposing Mutations within the RING Domain of BRCA1: Loss of Ubiquitin Protein Ligase Activity and protection from Radiation Hypersensitivity. *Proc. Natl. Acad. Sci.* 98 (9), 5134–5139. doi:10.1073/pnas.081068398

Saurin, A. J., Borden, K. L. B., Boddy, M. N., and Freemont, P. S. (1996). Does This Have a Familiar RING? *Trends Biochem. Sci.* 21 (6), 208–214. doi:10.1016/s0968-0004(96)80017-x

Shangary, S., and Wang, S. (2008). Targeting the MDM2-P53 Interaction for Cancer Therapy. *Clin. Cancer Res.* 14, 5318–5324. doi:10.1158/1078-0432.CCR-07-5136

Sharma, D., and Rajarathnam, K. (2000). ^{13}C Nmr Chemical Shifts Can Predict Disulfide Bond Formation. *J. Biomol. NMR* 18, 165–171. doi:10.1023/a:1008398416292

Sharma, V., Antonacopoulou, A. G., Tanaka, S., Panoutsopoulos, A. A., Bravou, V., Kalofonos, H. P., et al. (2011). Enhancement of TGF-β Signaling Responses by the E3 Ubiquitin Ligase Arkadia Provides Tumor Suppression in Colorectal Cancer. *Cancer Res.* 71, 6438–6449. doi:10.1158/0008-5472.CAN-11-1645

Shastry, B. S. (2007). SNPs in Disease Gene Mapping, Medicinal Drug Development and Evolution. *J. Hum. Genet.* 52, 871–880. doi:10.1007/s10038-007-0200-z

Sun, Y. (2006). E3 Ubiquitin Ligases as Cancer Targets and Biomarkers. *Neoplasia* 8, 645–654. doi:10.1593/neo.06376

Wang, D., Ma, L., Wang, B., Liu, J., and Wei, W. (2017). E3 Ubiquitin Ligases in Cancer and Implications for Therapies. *Cancer Metastasis Rev.* 36, 683–702. doi:10.1007/s10555-017-9703-z

Wüthrich, K. (1986). *NMR of Proteins and Nucleic Acids*. New York: Wiley.

Zhang, X., Chen, J., Wu, M., Wu, H., Arokiaraj, A. W., Wang, C., et al. (2013). Structural Basis for Role of Ring finger Protein RNF168 RING Domain. *Cell Cycle* 12, 312–321. doi:10.416/Fcc.23104

Zheng, N., Wang, P., Jeffrey, P. D., and Pavletich, N. P. (2000). Structure of a C-Cbl-UbcH7 Complex. *Cell* 102 (4), 533–539. doi:10.1016/s0092-8674(00)00057-x

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

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

frontiers in Molecular Biosciences

ORIGINAL RESEARCH
published: 23 February 2022
doi: 10.3389/fmolb.2021.821755

Phosphorylation of the Hepatitis B Virus Large Envelope Protein

*Marie-Laure Fogeron[1†], Lauriane Lecoq[1†], Laura Cole[1], Roland Montserret[1], Guillaume David[1], Adeline Page[2], Frédéric Delolme[2], Michael Nassal[3] and Anja Böckmann[1]**

[1]*Molecular Microbiology and Structural Biochemistry, Labex Ecofect, UMR 5086 CNRS/Université de Lyon, Lyon, France,* [2]*Protein Science Facility, SFR BioSciences CNRS UAR 3444, Inserm US8, UCBL, ENS de Lyon, Lyon, France,* [3]*Department of Medicine II / Molecular Biology, Medical Center, University Hospital Freiburg, University of Freiburg, Freiburg im Breisgau, Germany*

We here establish the phosphorylation sites in the human hepatitis B virus (HBV) large envelope protein (L). L is involved in several functionally important interactions in the viral life cycle, including with the HBV cellular receptor, HBV capsid, Hsc70 chaperone, and cellular membranes during fusion. We have recently shown that cell-free synthesis of the homologous L protein of duck HBV in wheat germ extract results in very similar phosphorylation events to those previously observed in animal cells. Here, we used mass spectrometry and NMR to establish the phosphorylation patterns of human HBV L protein produced by both *in vitro* cell-free synthesis and in *E. coli* with the co-expression of the human MAPK14 kinase. While in the avian virus the phosphorylation of L has been shown to be dispensable for infectivity, the identified locations in the human virus protein, both in the PreS1 and PreS2 domains, raise the intriguing possibility that they might play a functional role, since they are found at strategic sites predicted to be involved in L interactions. This would warrant the further investigation of a possible function in virion formation or cell entry.

Keywords: hepatitis B, L HBsAg, phosphorylation, NMR, cell-free (CF) protein synthesis, mass spectrometry

OPEN ACCESS

Edited by:
Piero Andrea Temussi, University of Naples Federico II, Italy

Reviewed by:
Giovanna Musco, San Raffaele Hospital (IRCCS), Italy
Richard J. Bingham, University of Huddersfield, United Kingdom

****Correspondence:***
Anja Böckmann
a.bockmann@ibcp.fr

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Structural Biology, a section of the journal Frontiers in Molecular Biosciences

Received: *24 November 2021*
Accepted: *22 December 2021*
Published: *23 February 2022*

Citation:
Fogeron M-L Lecoq L, Cole L, Montserret R, David G, Page A, Delolme F, Nassal M and Böckmann A (2022) Phosphorylation of the Hepatitis B Virus Large Envelope Protein. Front. Mol. Biosci. 8:821755. doi: 10.3389/fmolb.2021.821755

INTRODUCTION

The three hepatitis B virus (HBV) envelope proteins L, M, and S [large, middle and small hepatitis B surface antigen (HBsAg)] form the viral envelope. S is an integral membrane protein for which structural models predict four transmembrane-spanning helices, interrupted by an N-proximal cytosolic loop and, after the second helix, the antigenic loop, which presents a complex yet poorly defined structure stabilized by multiple disulfide bridges. M and L share their C-termini with S but carry in addition the PreS1 plus PreS2 (for L) and PreS2 (for M) domains. PreS1 and PreS2 together, collectively termed PreS, are suspected to represent an intrinsically disordered protein domain. This was experimentally supported for the avian (duck) homolog, DHBV PreS (Urban et al., 2000), while the corresponding data for PreS from human HBV lacked.

The PreS part of L plays a central role in a variety of interactions with viral and cellular proteins. The very N-terminal residues of PreS represent a first important site, with the first 48 amino acids involved in binding to the HBV receptor (Glebe et al., 2005; Gripon et al., 2005), the sodium taurocholate co-transporting polypeptide (NTCP) (Yan et al., 2012). N-terminal myristoylation is central in this context (Persing et al., 1987). Further down the PreS sequence, several hydrophobic stretches can be identified that are concentrated between residues 50–70. Most models favor that different residues distributed along preS are responsible for membrane interaction, hinting at a

possible fusion mechanism (Núñez et al., 2009; Delgado et al., 2015), but a recent proposal locates the fusion peptide exactly to this hydrophobic region (Pérez-Vargas et al., 2021). A third region of interest is localized at the PreS1/PreS2 border, comprising residues between approximately amino acids 90–120, which is believed to be involved in interactions with the viral capsid during particle formation (Gudima et al., 2007; Xi et al., 2021). Finally, another site in PreS1 is involved in the interaction with the Hsc70 chaperone, reported to be a determinant in the i-PreS orientation observed in immature viral particles (Prange et al., 1999; Prange, 2012). In sum, several specific sites in preS have been identified to be involved in the important functional interactions of L.

Post-translational modifications have been shown to play a central role in the HBV life cycle (Heger-Stevic et al., 2018; Yang, 2018). The phosphorylation of PreS has been thoroughly investigated in the DHBV L variant (Grgacic and Anderson, 1994; Grgacic et al., 1998). Metabolic labeling with ^{32}P and digestion with residue-specific phosphatases (Grgacic and Anderson, 1994) have identified several phosphorylation sites, whose functional importance has, in part, been assessed by the mutational analysis of Ser and Thr residues (Grgacic et al., 1998). Mutations mimicking absent or constitutive PreS phosphorylation had no detectable effect on DHBV replication (Grgacic et al., 1998), although DHBV L phosphorylation was found important in host cell–virus cross talk (Rothmann et al., 1998).

We have recently found, during the wheat-germ cell-free protein synthesis of DHBs proteins for structural studies, that the DHBV L protein is phosphorylated in this system (David et al., 2018; David et al., 2019). While it was known that the cell-free extract indeed contains kinases, allowing post-translational phosphorylation (Badillo et al., 2017), our analysis of the DHBV L protein confirmed that the phosphorylation sites are the same as in animal cells (Grgacic et al., 1998) and identified two additional sites (David et al., 2019).

As an important step in the ongoing studies in our group on HBV preS–capsid interactions, we here investigate preS phosphorylation, in the protein, from the human virus, as obtained in recombinant systems, and apply a similar analysis to the HBV L protein and PreS fragments thereof. Based on the results, we designed a bacterial kinase co-expression system able to phosphorylate PreS, which we show to reproduce the results obtained in the acellular system. In addition, we used NMR to structurally analyze the PreS protein and to confirm major phosphorylation sites by NMR chemical-shifts. Our approach confirmed the disordered nature of the protein, and identified four consensus sites as well as several additional potential PreS phosphorylation sites.

MATERIALS AND METHODS

Plasmids

For cell-free expression, PreS1, PreS, and full-length HBV L sequences were amplified by PCR from the HBV isolate H2815 (genotype D5, GenBank Accession Number KP322603.1), and cloned into a pEU-E01-MCS vector (CellFree Sciences, Matsuyama, Japan). A *Strep*-tag II, shortly named "tag" in the following, was fused either to the N- or C-terminal end for purification (Schmidt and Skerra, 2007), resulting in the four constructs PreS1$_{tag}$, $_{tag}$PreS, PreS$_{tag}$, and $_{tag}$L. The plasmids were amplified in *Escherichia coli* TOP10 cells (Life Technologies, Carlsbad, CA, USA). DNA was isolated using a NucleoBond Xtra Maxi kit (Macherey-Nagel). Plasmids were further purified by phenol/chloroform extraction, according to CellFree Sciences recommendations.

For the bacterial expression of PreS ($^{E.coli}$PreS$_{tag}$), cDNA-encoding PreS was cloned into a pRSF-T7 vector and into an analogous pRSF-T7 vector carrying a MAPK14 gene under a Tet promotor in order to promote PreS phosphorylation. A solubility-enhancing fusion protein, GB1, with an N-terminal His-Tag, was fused at the N-terminus of PreS and a *Strep*-tag II to its C-terminus. A TEV protease cleavage site was inserted between GB1 and PreS, resulting in the following constructs: pRSF_T7-H6-GB1-ENLYFQG-preS-*Strep*-tag-II and pRSF_Tet-H6MAPK14_T7-H6-GB1-ENLYFQG-preS-*Strep*-tag-II.
Plasmids were amplified in *E. coli* TOP10 cells (Life Technologies). DNA was isolated using QIAprep Spin Miniprep Kit (Qiagen, Hilden, Germany).

Wheat Germ Cell-free Expression and Purification of PreS and L

Homemade wheat germ extract was prepared using non-treated durum wheat seeds (Semences du Sud, Vic-Fezensac, France), as described in Fogeron et al. (2017). Protein synthesis was performed with uncoupled transcription and translation. Transcription was performed using 100 µg/ml plasmid, 2.5 mM NTP mix (Promega, Charbonnières-les-Bains, France), 1 U/µl RNase inhibitor (CellFree Sciences, Matsuyama, Japan), and 1 U/µl SP6 RNA polymerase (CellFree Sciences, Matsuyama, Japan) in transcription buffer (CellFree Sciences, Matsuyama, Japan) composed of 80 mM HEPES-KOH pH 7.6, 16 mM magnesium acetate, 10 mM DTT, and 2 mM spermidine in nuclease-free water. The solution was incubated for 6 h at 37°C; the produced mRNA solution was then used directly for translation. Translation was performed using the bilayer method (Takai et al., 2010; Fogeron et al., 2015), either at a small scale with the one well of a 6-well plate (6 ml total reaction volume), or at a large scale with two 6-well plates (2 × 36 ml total reaction volume) in order to obtain a sufficient amount of protein to perform NMR experiments. The feeding buffer composition was 30 mM HEPES-KOH pH 7.6, 100 mM potassium acetate, 2.7 mM magnesium acetate, 16 mM creatine phosphate, 0.4 mM spermidine, 1.2 mM ATP, 0.25 mM GTP, and 4 mM DTT supplemented with 6 mM amino acid mix (0.3 mM average concentration per amino acid). The translation mix was prepared with the mRNA solution, wheat germ extract (250 µl for each well of a 6-well plate), 6 mM amino acid mix, and 40 µg/ml creatine kinase. On the bottom of the each well, the translation mix (518 µl for each well) was then deposited under the feeding buffer (5.5 ml for each well), allowing for the formation of a bilayer. The plate was incubated overnight at 22°C without

shaking. For NMR sample preparation, large-scale production was performed in presence of a mixture of (^{15}N) or (^{2}H-^{13}C-^{15}N)-labeled amino acids (Cambridge Isotope Laboratories) added to the reaction solution and the feeding buffer. A summary of the different sample preparations is given in **Supplementary Table S1**.

For *Strep*-Tactin affinity chromatography, the total cell-free reaction was incubated with homemade benzonase 250 U/µl (50 µl per well) on a rolling wheel for 30 min at room temperature. This solution was then centrifuged at 20,000 *g*, 4°C for 30 min. The supernatant obtained was loaded either on a 200-µl (small-scale production) or on two 1-ml (large-scale production) *Strep*-Tactin Superflow® gravity flow columns (IBA Lifesciences, Göttingen, Germany). Purification was performed as described previously (Fogeron et al., 2015; Fogeron et al., 2016). The protein of interest was eluted in 100 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM EDTA, and 2.5 mM D-desthiobiotin (IBA Lifesciences, Göttingen, Germany).

All experiments were assessed using 15% polyacrylamide SDS-PAGE gels. Samples were resuspended in a loading buffer containing 62.5 mM Tris-HCl pH 6.8, 10% glycerol (v/v), 2% SDS (w/v), 5% β-mercaptoethanol (v/v), and 0.01% bromophenol blue (w/v).

Bacterial Expression and Purification of $^{E.coli}$PreS$_{tag}$

E. coli BL21 (DE3) or *E. coli* BL21*CP (DE3) cells were transformed with the plasmids for PreS expression or PreS plus MAPK14 co-expression, respectively, and grown at 37°C, either in LB (Lysogeny Broth) for purification setup and mass spectrometry or in M9 minimal medium containing 2 g/L of ^{13}C-labeled glucose and ^{15}N-labeled ammonium chloride for NMR spectroscopy. T7 promoter-controlled protein expression was induced at an OD_{600nm} of 1.2 using 1 mM of isopropyl-ß-D-1-thiogalactopyranoside (IPTG) for 17 h at 30°C and cells were harvested (6,000 *g*, 20 min, 4°C). Purification procedures were the same for labeled and unlabeled samples. Cells were resuspended in 4 ml/g of cell pellet lysis buffer (20 mM $NaPO_4$, pH 7.5, 500 mM NaCl, 40 mM imidazole) supplemented with EDTA-free protease inhibitor (Roche). Cell lysis was performed by incubating with 1 mg/ml of lysozyme (Sigma) for 50 min at 4°C under rotation, and nucleic acids were digested with Benzonase nuclease 250 U/µl (6 µl/L of culture) supplemented with 2 mM $MgCl_2$, for 30 min at room temperature. Cellular membranes were broken by passing three times through a microfluidizer (Microfluidics M-110P) at 15,000 psi. Soluble proteins were isolated by centrifugation (25,000 *g*, 30 min, 4°C) and filtrated using a 0.45 µm filter, before being loaded into a 5 ml HisTrap (GE Healthcare) affinity column, connected to a Biorad NGC chromatography system. The HisTrap column was washed with lysis buffer until A_{280nm} returned to baseline, and His-tagged proteins were eluted through a one-step elution with elution buffer (20 mM $NaPO_4$ pH 7.5, 500 mM NaCl, 1 M imidazole). The eluted proteins were dialyzed using a 3.5 kDa cut-off membrane (Spectrum labs) in TEV protease reaction buffer (50 mM Tris-HCl pH 7.5, 1 mM EDTA, 5 mM DTT), at 4°C under slow stirring. GB1 fusion proteins were cleaved using homemade TEV protease 10 U/µl (0.2 ml or 2,000 U per milligram of PreS protein) by incubating overnight at 4°C under rotation. Cleaved PreS protein was recovered using *Strep*-Tactin resin (IBA Lifesciences), in batch mode, following the manufacturer's recommendations. EDTA-free protease inhibitor (Roche) 1X was added to the eluted protein.

PreS was dialyzed using a 3.5 kDa cut-off membrane (spectrum labs) overnight in final NMR buffer (20 mM $NaPO_4$ pH 6, 50 mM NaCl). Protein concentration was determined using a Nanodrop instrument (Thermo Fisher) and the absorbance at 280 nm. Subsequently, the protein solution was concentrated by immerging the dialysis bag containing the protein and protease inhibitor into Sephadex G-25 powder at 4°C. Concentration was followed by weighing the dialysis bag before adding the powder and every 12–20 h thereafter. This approach was used to avoid PreS sticking to cellulose membranes used in concentrators as Amicon (Merck) or Vivaspin (Sartorius). The protein concentration was measured by Nanodrop before storage at −80°C. Protein solutions in all purification steps were analyzed using 15% polyacrylamide SDS-PAGE gels. The typical yields of PreS and PreS-MAPK14 co-expressed protein were around 30 mg in LB and 10 mg per liter of M9 medium culture.

Solution-State NMR Spectroscopy

Isotopically labeled samples were dialyzed in 20 mM HEPES-KOH pH 7.5 containing 50 mM NaCl for ^{15}N- and ^{2}H-^{13}C-^{15}N-PreS1$_{tag}$, in 20–50 mM phosphate buffer pH 6.0 containing 50 mM NaCl for ^{2}H-^{13}C-^{15}N-PreS$_{tag}$ and $_{tag}$PreS produced in cell-free, and ^{13}C-^{15}N-$^{E.coli}$PreS$_{tag}$ proteins ± MAPK14 produced in the bacterial expression system. A pH of 6.0 could not be used for PreS1$_{tag}$ due to its close theoretical isoelectric point of 6.37, which could result in protein aggregation. D_2O was added to a final volume of 7%, and protein concentration was quantified by NanoDrop. Concentrations were estimated for cell-free samples at 150 µM for PreS1$_{tag}$, 60 µM for PreS$_{tag}$ and 120 µM for $_{tag}$PreS, and 50 µM for both bacterial samples MAPK14-$^{E.coli}$PreS$_{tag}$ and $^{E.\ coli}$PreS$_{tag}$. About 1 µl of 2,2-dimethyl-2-silapentane-5-sulfonate (DSS) was added to each sample for chemical-shift referencing. NMR experiments were recorded at 298 K (25°C) on Bruker Avance II spectrometers operating at 600 MHz ($_{tag}$PreS and $^{E.\ coli}$PreS$_{tag}$ samples) and 950 MHz (PreS1$_{tag}$ and PreS$_{tag}$). Backbone resonances were assigned using two-dimensional (2D) BEST-TROSY spectra and three-dimensional (3D) BEST-TROSY versions of HNCA, HNcaCO, as well as HNCACB_2H, HNcoCACB_2H, and HncaCO_2H optimized for deuterated proteins (cell-free samples) (Solyom et al., 2013) and HNCACB, HNcoCACB, and HncaCO for non-deuterated proteins ($^{E.coli}$PreS$_{tag}$ samples). Pulse sequences were installed on the 600 MHz using the NMRlib tool (Favier and Brutscher, 2019). For details on NMR samples, NMR experiments, and assignment, see **Supplementary Table S2**. All PreS samples started to show degradation products in the NMR spectra after 1 week, even in the presence of a protease inhibitor. NMR data were processed using TopSpin 4.0.7 (Bruker) and analyzed with CcpNmr Analysis 2.4.2 (Vranken et al., 2005; Stevens et al., 2011).

For the secondary chemical shifts, the Cα and Cβ chemical shifts of each residue for PreS1 and both PreS constructs were compared to their random coil shift taken from Wang and Jardetzky (2002). The difference between ΔCα and ΔCβ were calculated, where positive values indicate the tendency to form an α-helix and negative values indicate the tendency to form a β-strand.

MALDI-TOF Mass Spectrometry

Mass spectra were acquired with a Voyager-DE PRO (Sciex, Framingham, MA, USA) equipped with a nitrogen laser emitting at 337 nm. Ions were accelerated to a final potential of 20 kV, and the mass spectrum was the sum of 300 laser shots. An external mass calibration was used (a mixture of peptides from the Sequazyme™ standards kit, AB Sciex). The analysis was performed in linear mode (instrumental mass accuracy is 0.05%). Samples were prepared by diluting 10-fold the protein solution (0.2 mg/ml) in the matrix sinapinic acid (Sigma-Aldrich, St. Louis, MI, USA), used without further purification and dissolved in 0.1%TFA/acetonitrile (70/30 v/v). About 1 µl of the mixture was deposited onto the MALDI sample plate and let dried to complete co-crystallization.

Nano LC-MS/MS Analysis

The solutions of all purified PreS1 and PreS samples were digested overnight at 37°C with 1/100 the amount (w/w) of trypsin (trypsin porcine; Promega, Charbonnières-les-Bains, France). $_{tag}$L protein solution was reduced with 5 mM TCEP for 45 min at 57°C, then alkylated with 10 mM iodoacetamide for 30 min in the dark at room temperature and under agitation (850 rpm) and digested overnight at 25°C with chymotrypsin 1/100 ratio (bovine pancreas chymotrypsin; Promega, Charbonnières-les-Bains, France). Peptide digests were desalted using C18 spin columns (Thermo Scientific). Peptides were dried in a speed-vac and suspended in 0.1% HCOOH. Samples were analyzed using an Ultimate 3,000 nano-RSLC (Thermo Scientific) coupled on line with a Q-Exactive HF mass spectrometer *via* a nano-electrospray ionization source in positive ionization mode (Thermo Scientific). Peptide mixtures were loaded on a C18 Acclaim PepMap100 trap-column and then separated on a C18 Acclaim Pepmap100 nano-column 50 cm × 75 µm i. d, 2 µm, 100 Å (Thermo Scientific) with a 60 min linear gradient from 3.2 to 40% buffer 0.1% formic acid in © at a flow rate of 300 nl/min. Samples were analyzed using TOP20 HCD; mass data were acquired in a data-dependent strategy, selecting the fragmentation events based on the 20 most abundant precursor ions in the survey scan (375–1,600 Th). The resolution of the survey scan was 60,000 at m/z 200 Th. The ion target values for the survey scans in the Orbitrap and the MS^2 mode were set to 3E6 and 1E5, respectively, and the maximum injection time was set to 60 ms for both scan modes. The parameters for acquiring HCD MS/MS spectra were set to a collision energy of 27 and an isolation width of 2 m/z. The precursors with unknown charge state or a charge state of 1 were excluded. The peptides selected for MS/MS acquisition were then placed on an exclusion list for 20 s using the dynamic exclusion mode to limit duplicate spectra. Peptides were identified by database searching using Sequest HT and MS Amanda with Proteome Discoverer 2.2 software (Thermo Scientific) against the HBV sequence. Precursor mass tolerance was set at 10 ppm, and fragment mass tolerance was set at 0.02 Da, and up to two missed cleavages were allowed. Oxidation (M), acetylation (protein N-terminus, K) and phosphorylation (S, T, Y) were set as variable modification and carbamidomethylation (C) as fixed modification. Peptides were filtered with a fixed-value PSM validator and rank 1. Phosphorylation sites were then manually validated.

Circular Dichroism

Far UV circular dichroism (CD) spectra were recorded on a Chirascan spectrometer (Applied Photophysics, Leatherhead, United Kingdom) calibrated with 1S-(+)-10-camphorsulfonic acid. Measurements were carried out at room temperature in a 0.1 cm path length quartz cuvette (Hellma). Spectra were measured in a 180–260 nm wavelength range with an increment of 0.2 nm, band pass of 0.5 nm, and integration time of 1 s. Spectra were processed, baseline-corrected, and smoothed with the Chirascan software. $_{tag}$PreS was in 50 mM phosphate buffer pH 6.5 at a concentration of 9.5 µM.

Data Availability

The ^{13}C and ^{15}N backbone chemical shifts of PreS have been deposited in the BioMagResBank (http://www.bmrb.wisc.edu/) under accession code 51186.

RESULTS

Cell-free Synthesis and Purification of HBV L and PreS Peptides

We synthesized full-length HBV L, as well as PreS1 and PreS fragments thereof, using wheat-germ cell-free protein synthesis (WG-CFPS) (Sawasaki et al., 2002; Takai et al., 2010; Fogeron et al., 2017; Fogeron et al., 2021). All cell-free synthesized proteins carried a Strep-tag II for affinity purification, for L at the N-terminus and for PreS1$_{tag}$ at the C-terminus; we produced both versions for PreS, named $_{tag}$PreS and PreS$_{tag}$. **Figure 1A** shows the results of the WG-CFPS of PreS1$_{tag}$, followed by affinity purification, as analyzed by SDS-PAGE, followed by Coomassie blue staining. The protein was fully soluble even in the absence of detergent, as no PreS1 was detected in the pellet fraction. The band corresponding to PreS1$_{tag}$ is clearly visible on the gel (yields are given in **Supplementary Table S1**), and interestingly, two separate bands with different intensities are observed for all elution fractions. We recently reported a similar observation for the duck HBV L protein, where it resulted from alternative translation initiation in addition to phosphorylation (David et al., 2019). Both PreS$_{tag}$ and $_{tag}$PreS have also been successfully purified by affinity chromatography (**Figure 1B** and **Figure 1C**, respectively). Final yields (**Supplementary Table S1**) might be slightly overestimated as some degradation is visible on the gels, indicating that these constructs might not be stable on the long term unless protease inhibitors are present. A CD spectrum was recorded on the purified protein

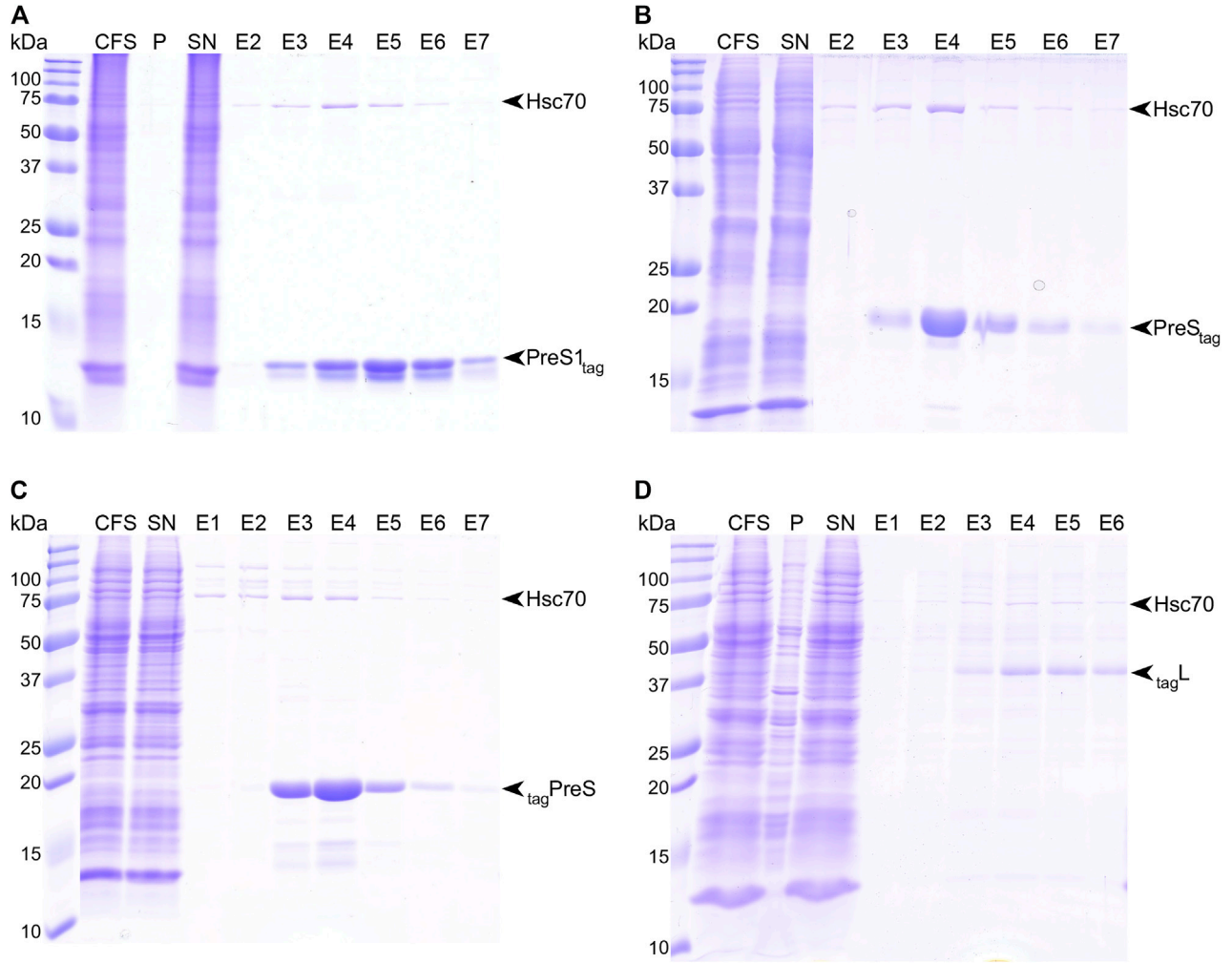

FIGURE 1 | Strep-Tactin© affinity purification of PreS1tag **(A)**, $PreS_{tag}$ **(B)**, $_{tag}PreS$ **(C)**, and $_{tag}L$ **(D)**. Purification steps have been analyzed by SDS-PAGE, followed by Coomassie blue staining; CFS, total cell-free sample; P, pellet; SN, supernatant; E#, elution fractions.

(**Supplementary Figure S1**), which already indicates that PreS is likely unstructured.

As a membrane protein, the full-length construct $_{tag}L$ required the presence of detergent to be synthesized in a solubilized form for further purification. We have identified Brij-58 as a suitable detergent for the soluble production of this construct, with a concentration of 0.05% being optimal. Affinity purification was successfully performed in the presence of 0.1% n-dodecyl-β-D-maltoside (DDM) (**Figure 1D**). While L_{tag} could be expressed in a soluble form as well, the purification of this construct was unsuccessful, so it was not included in our analyses.

One can note that, for all constructs, in addition to the two bands corresponding to $PreS1_{tag}$, a third band eluted from the column with an apparent molecular weight below 75 kDa, which corresponds to Hsc70, as identified by the mass spectrometry of the band cut from the gel. Hsc70 is known to be an interactant of PreS in cells, and our work thus identifies at least one binding site to be localized in PreS1, in line with previous studies that assigned it to residues 70–107 of PreS (Prange et al., 1999).

In summary, all PreS-containing constructs expressed well using WG-CFPS, with yields between 0.3 and 1.7 mg protein per milliliter of wheat-germ extract (**Supplementary Table S1**), and in a soluble form, which was obtained for L through addition of detergent to the reaction. All proteins could be purified to high homogeneity, using in some cases detergent, resulting in higher purity.

Identification of PreS Phosphorylation Sites Using Mass Spectrometry

As multiple bands observed by SDS-PAGE (typically for PreS, **Figure 1A**) can be an indication of a post-translational modification such as phosphorylation (Ubersax et al., 2003) we set out to assess the modifications using mass spectrometry, in line with previous work on the DHBV L protein (David et al., 2019) and also HBV core (Heger-Stevic et al., 2018). First, the amount of post-translational modifications of $PreS1_{tag}$ were evaluated while analyzing the total mass of the

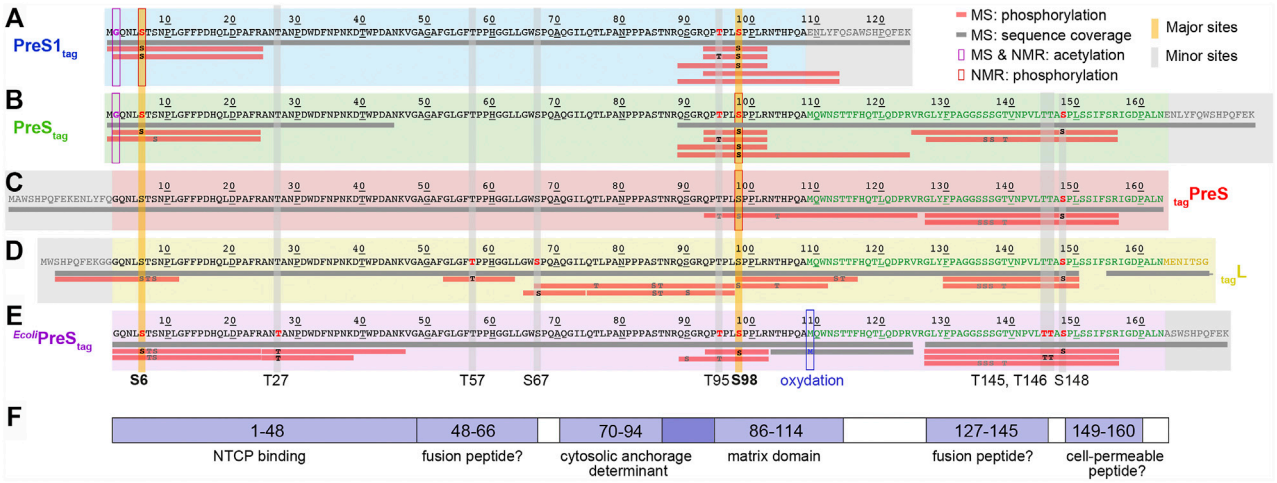

FIGURE 2 | Identification of acetylation and phosphorylation sites in PreS1, PreS, and HBV L using mass spectrometry and NMR. Mass spectrometry analysis has been performed on fractions from purified protein samples shown on **Figure 1** for cell-free synthesized $PreS1_{tag}$ **(A)**, $PreS_{tag}$ **(B)**, $_{tag}$PreS **(C)**, $_{tag}$L **(D)**; and MAPK14-$^{E.\ coli}PreS_{tag}$ **(E)**. Horizontal gray bars below the sequences represent the sequence coverage by LC-MS/MS. Red bars represent peptides where one or more phosphorylation sites have been confirmed; the residues in black correspond to the phosphorylation sites formally identified, and the ones in gray correspond to possible but not confirmed sites. The PreS1 part of the amino-acid sequence is typed in black, the PreS2 part in green, the beginning of the S part in yellow (the full $_{tag}$L protein was analyzed, but the sequence coverage of S was poor and no phosphorylation sites were detected), and the tag sequences in gray. Additional information extracted from NMR chemical shifts (see below) is shown as purple squares for acetylation and red squares for phosphorylation. Major phosphorylation sites identified by mass spectrometry and NMR are highlighted by vertical yellow bars; minor sites (only mass spectrometry) by vertical gray bars. **(F)** Functional regions of the HBV PreS: binding to NTCP (Yan et al., 2012); possible fusion peptides (Pérez-Vargas et al., 2021); the MD (Bruss and Thomssen, 1994); and a possible cell-permeable peptide (Oess and Hildt, 2000).

protein by MALDI-TOF mass spectrometry. The region of interest in **Supplementary Figure S2** showed four peaks, all corresponding to $PreS1_{tag}$. The first peak from the left corresponds to the protein from which the N-terminal methionine (13,364.0 Da) was stripped. Peaks representing $PreS1_{tag}$ with one additional acetyl group (+42 Da, 13,405.8 Da), with one phosphoryl group (+80 Da, 13,444.2 Da), and with both acetyl and phosphoryl groups (+122 Da, 13,486.3 Da) were observed as well. These signals indicate a single phosphorylation per chain, though not necessarily on the same residue in all proteins in the sample. Major and minor sites can overlap and yield a global sum of one site.

In order to localize the modifications on the protein sequence, $PreS1_{tag}$ was further analyzed by LC-MS/MS mass spectrometry (**Supplementary Figure S3**), and one can see that identified peptides completely covered the sequence of $PreS1_{tag}$ (**Figure 2A**). Phosphorylation in $PreS1_{tag}$ was unambiguously identified for three amino acids, namely, S6, T95, and S98, which are highlighted in red bold font on the sequence. The analysis also confirmed the removal of M1, as well as acetylation on G2 (**Supplementary Figure S4**).

For the $PreS_{tag}$ construct, even if the entire sequence was not covered, the analysis by LC-MS/MS mass spectrometry (**Figure 2B**) revealed phosphorylation on four amino acids: S6, T95, S98, and S148 (in bold red type, **Supplementary Figure S5**). Further unconfirmed sites pointed to S8, and S136/S137/T139. As for $PreS1_{tag}$, N-terminal methionine processing and G2 acetylation could be identified (data not shown). We also analyzed the complementary construct $_{tag}$PreS, for which LC-MS/MS mass spectrometry shows full sequence coverage (**Figure 2C**). However, phosphorylation was unambiguously identified only for one single amino acid, namely, S148 (**Supplementary Figure S6**). Further ambiguous sites are proposed for T95/S98/T104, and also S135/S136/S137/T139. This indicates that the N-terminal tag interferes with phosphorylation in the N-terminal portion.

The LC-MS/MS mass spectra of HBV L (**Figure 2D**) show that the region between amino acids 50–70 is best covered in L. Three phosphorylation sites could be unambiguously identified in L by LC-MS/MS: T57, S67, and S148 (**Supplementary Figure S7**). Several ambiguous sites, including S6/T7/S8 and T76/S85/T86, were revealed as well. As the phosphorylation of S6 was clearly identified for the isolated $PreS1_{tag}$ and $PreS_{tag}$ forms as described above, the phosphorylation of HBV L thus most probably also occurs on this residue, and not on T7 or S8.

When combining the results obtained for L and its fragments as produced by WG-CFPS, LC-MS/MS mass spectrometry thus identified S6, T57, S67, T95, S98, and S148 as phosphorylated, and highlights other possible but unconfirmed sites. At the same time, the MALDI-TOF analyses of PreS1 suggested that only one major site exists (S6, T95, S98) and that other sites are minor. These findings are summarized in **Figures 2A–D**, where horizontal bars represent the peptides that could be analyzed, with those for which phosphorylation was unambiguously confirmed in red.

Identification of Phosphorylation Sites by NMR Spectroscopy.

We produced $PreS1_{tag}$, $PreS_{tag}$, and also $_{tag}$PreS on a large scale and uniformly ^{2}H-^{13}C-^{15}N labeled for NMR studies, with yields

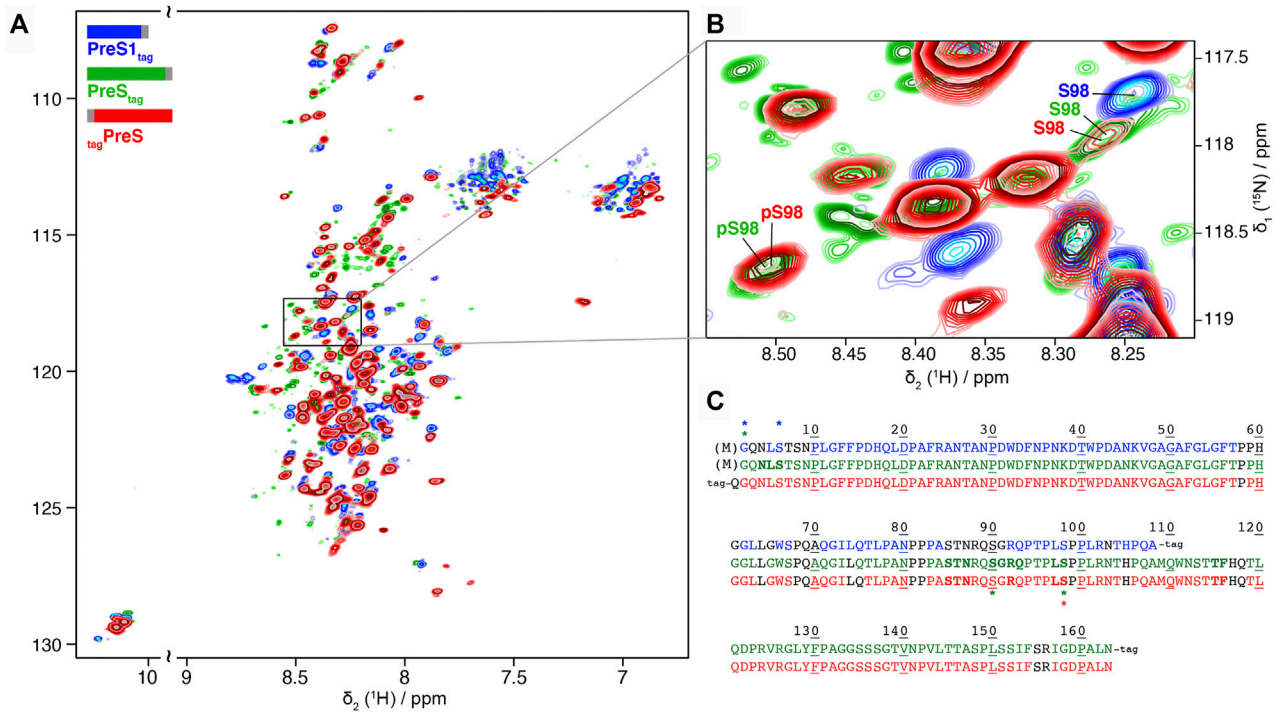

FIGURE 3 | Partial phosphorylation is detected by NMR for PreS and PreS1 produced using cell-free synthesis: **(A)** Solution NMR ^{1}H-^{15}N BEST-TROSY spectra of ^{2}H-^{13}C-^{15}N PreS1$_{tag}$ at pH 7.5 (in blue), PreS$_{tag}$ at pH 6 (in green) and $_{tag}$PreS at pH 6 (in red). **(B)** Extract of 2D BEST-TROSY spectra showing a peak corresponding to unphosphorylated S98 for the 3 constructs on the top right. The peak is weak for both PreS$_{tag}$ and $_{tag}$PreS, with another peak corresponding to phosphorylated S98 detected at the bottom left. This means that S98 is mainly, but not fully phosphorylated in PreS. The Cβ chemical shift was used to confirm the phosphorylation state of this serine (**Supplementary Figure S10**). **(C)** Sequence of the three constructs showing assigned residues for PreS1$_{tag}$ (top, blue), PreS$_{tag}$ (middle, green) and $_{tag}$PreS (bottom, red). Assigned residues are colored, while unassigned residues are in black type. Proline residues are colored when their Cα, Cβ and C′ are assigned. Residues which display two forms on the NMR spectra are shown in bold, and residues with chemical-shifts typical of phosphorylation and acetylation are indicated by a star with the corresponding color-code. Processed residue M1 is shown in brackets.

between 0.3 and 0.6 mg protein per ml WGE (**Supplementary Table S1**). Solution NMR experiments were recorded on PreS1$_{tag}$, PreS$_{tag}$ and $_{tag}$PreS, and 2D BEST-TROSY spectra are shown in **Figure 3A**, and an extract in **Figure 3B**. The peak pattern reveals a narrow chemical shift dispersion, with $^{1}H_{N}$ resonances observed between 7.5 and 8.5 ppm, revealing with atomic detail the intrinsically disordered nature of all three PreS fragments. Backbone resonances were assigned using a combination of 3D NMR spectra (Solyom et al., 2013). Assigned residues are shown on the sequences of the three constructs in **Figure 3C**, and assignment statistics are given in **Supplementary Table S2**. A 2D-HN assigned spectrum is shown in **Supplementary Figure S8**. Secondary chemical shifts derived from the sequential assignments reveal that the proteins do not display any partial secondary structures (**Supplementary Figure S9**). HBV PreS is thus, as DHBV PreS, an intrinsically disordered protein.

NMR chemical shifts are sensitive to phosphorylation events, as addition of this group typically strongly affects the electronic environment of the neighboring spins. We thus listed random-coil chemical shifts of phosphorylated Ser and Thr residues taken from Hendus-Altenburger et al. (2019) (**Supplementary Table S3** and **Supplementary Table S4**), and compared them to all assigned chemical shifts of Ser and Thr residues in the three samples. This allowed to clearly identify the phosphorylation of S6 in PreS1$_{tag}$, and partial phosphorylation of S98 in both PreS$_{tag}$ and $_{tag}$PreS constructs. S6 and S98 thus correspond to major phosphorylation sites in the WG-CFPS system. S98 is, however, only partially phosphorylated, as can be seen from the presence of a second, weaker signal for this residue at a chemical shift indicative for the non-modified residue, as shown in **Figure 3B**. The ratios of the peaks allowed to estimate that around 50% of S98 is phosphorylated in PreS$_{tag}$ and 65% in $_{tag}$PreS. The observed peak splitting could possibly also be due to cis–trans isomerization of neighboring proline residues (Hull and Kricheldorf, 1980); still, the chemical shift of the Cβ is typical for a phosphorylated residue (**Supplementary Figure S10**), and thus clearly points to phosphorylation.

The NMR spectra reveal phosphorylation only for a subset of the residues identified by mass spectrometry as being phosphorylated. This means that several residues, in the different constructs, are only phosphorylated on a subset of proteins in the sample, which is below the detection limit of NMR (about 10%). They thus correspond to minor sites, in agreement with the PreS1$_{tag}$ MALDI-TOF analysis that pointed to a single major phosphorylation site in this construct (**Supplementary Figure S1**), which thus must be S6 in PreS1$_{tag}$, according to the NMR analysis.

Interestingly, the analysis of the NMR spectra revealed signals corresponding to phosphorylation of an unconfirmed site, S90 in $PreS_{tag}$ (**Supplementary Table S3**). This site however has not been observed in the two other PreS constructs, and its identity thus remains to be confirmed.

In addition, NMR corroborates removal of the N-terminal methionine and the acetylation of G2 for both $PreS1_{tag}$ and $PreS_{tag}$, as shown in the 3D extracts of the HNcoCAB in **Supplementary Figure S11**. Indeed, a peak at 24.6 ppm displays a ^{13}C chemical shift typical of an acetyl glycine, while this is not observed in N-terminally tagged $_{tag}$PreS and $_{tag}$L, where no M1 processing and acetylation can take place due to the tag.

Taken together, the NMR data show that among the phosphorylation sites identified by mass spectrometry, S6 and S98 represent major phosphorylation sites, while T27, T57, S67, T95, T145, T146, and S148 must be minor sites. This is summarized on **Figure 2** by the yellow vertical bars, annotated in bold type for major sites, and the gray bars for minor sites.

Design of an *E. coli* Expression System for Phosphorylated PreS

With major phosphorylation sites being identified according to **Figure 2**, we predicted, combining information from several web programs (Li et al., 2018; Wang et al., 2020) (http://gps.biocuckoo.org/links.php), the kind of kinase that could generate such a pattern, in order to design a recombinant system to produce phosphorylated PreS in larger amounts for structural and interaction studies. cdk5 and MAPK14 were the best hits, and since cdk5 is reputed to be difficult to produce in bacteria, we included MAPK14 in the *E. coli* co-expression system. The design of the plasmid was based on a previous one developed for the HBV core protein (Heger-Stevic et al., 2018), and included a cleavable GB1 fusion protein in N-terminus. Cleavage results in PreS with a C-terminal *Strep*-tag II as shown in **Supplementary Figure S12**. We expressed PreS in bacteria with and without MAPK14 co-expression and analyzed the obtained protein with mass spectrometry for phosphorylation. The total mass of the $^{E.\ coli}PreS_{tag}$ co-expressed with MAPK14 was analyzed by MALDI-TOF mass spectrometry, revealing the presence of heterogeneous phosphorylation, with up to five cumulative phosphorylated sites (**Supplementary Figure S13**). The identified phosphorylation sites by LC-MS/MS mass spectrometry are shown in **Figure 2E**, with confirmed sites at S6, T27, T95, S98, T145, T146, and S148 (**Supplementary Figure S14**), with T95 possibly also assigned to S90. Also, similarly as for L_{tag}, T7 and S8 have been proposed as alternatives to S6; yet with S6 clearly confirmed for $PreS1_{tag}$ and $PreS_{tag}$, T7 and S8 are unlikely options. Four sites overlap with the previously identified phosphorylation sites using WG-CFPS, namely, S6, T95, S98, and S148. Surprisingly, while mass spectrometry allowed to evidence phosphorylation in $^{E.\ coli}PreS_{tag}$, it could not be detected by NMR as shown in **Supplementary Figure S15**. Most likely, phosphorylation in the current recombinant system is not quantitative as seen in the mass spectrum from **Supplementary Figure S13** where the major species remains the unphosphorylated protein, which probably places the phosphorylated residues below the detection level of 2D and 3D NMR spectra. This highlights the interesting ability of the cell-free system to efficiently induce phosphorylation by endogenous kinases.

DISCUSSION

We have shown that the HBV L protein is phosphorylated at several sites *in vitro* when synthesized both in a wheat-germ cell-free system or by bacterial co-expression with MAPK14. Proteomics studies (Mak et al., 2006) on wheat germs have highlighted the presence of several kinases, including also serine/threonine kinases; a study of kinases present in different organs of the wheat plant has revealed that germs contain a variety of kinases as well (Wei and Li, 2019). The phosphorylation of DHBV L in WG-CFPS has been observed to mainly occur at sites, followed by a proline residue, however, with exceptions (David et al., 2019). This is also the case for HBV L, where five sites are followed by Pro. Interestingly, as also for DHBV L, the N-terminal site (S6 in HBV L and S8 in DHBV L, respectively) is not preceding a Pro. Most residues for which phosphorylation could be confirmed are highly conserved in the sequence, and also most subsequent Pro residues (with the exception of 149), as shown in **Supplementary Figure S15**. An exception is S27, which is in about 20% of cases a Thr. We identified a total of nine phosphorylation sites, with several observed in different constructs by mass spectrometry and two confirmed by NMR as being major.

All identified sites are located in or just next to regions closely linked to different PreS functions [reviewed recently in Sun et al. (2021)] as shown in **Figure 2F**. Indeed, S6 is located in the interaction sequence of L with the cellular HBV receptor NTCP (Gripon et al., 2005; Yan et al., 2012). Since S6 is close to the central myristoylation site, its phosphorylation clearly has the potential to impact NTCP binding. T27 is located inside the identified NTCP-binding peptide as well. T57, S67, and T145 are located in or just next to the putative fusion peptides that have been proposed recently using a combined computational and experimental approach (Pérez-Vargas et al., 2021). Residues A70–P94 have been described to contain the amino-acid stretch that determines the cytosolic anchorage of PreS, presumably through interaction with the cognate heat shock protein Hsc70 (Prange et al., 1999). T95 and S98 are located in the so-called matrix domain (MD), which comprises the stretch of amino acids T86–T114 in the large envelope protein L. MD is the presumed interaction site of PreS with the core protein of the HBV capsid, central in the process of envelopment (Poisson et al., 1997; Le Pogam and Shih, 2002). This region is crucial for virion formation, and is believed to establish contact to the nucleocapsid, since truncations up to G92 were compatible with envelopment (Bruss and Thomssen, 1994), and several point mutations within the MD-blocked virion formation (Bruss, 1997). It has been shown also that a peptide comprising this domain interacts with the core particle (Poisson et al., 1997). Phosphorylation on T95 and S98 is centrally located in this

domain and has the potential to change the required interaction interfaces, resulting in productive envelopment. S148 is located just downstream from the proposed amphipathic PreS2 translocation motif (Oess and Hildt, 2000; Sun et al., 2021). It is intriguing that in, or just next to, each proposed functional sites, phosphorylation is observed in the present experiments, pointing to a possible role thereof in the regulation of PreS function. No role for phosphorylation has yet been identified for the phosphorylation sites in DHBV PreS (Grgacic and Anderson, 1994; Grgacic et al., 1998). However, DHBV differs from HBV in fundamental aspects, including by a much-large core protein (Makbul et al., 2020) and the lack of an HBx-like transactivator that is crucial in HBV infection (Slagle and Bouchard, 2018). Hence, a functional relevance of PreS phosphorylation in the human virus would not be surprising.

CONCLUSION

We here reported phosphorylation on nine different sites of the PreS domain of the human HBV L envelope protein, enabled by endogenous kinases in the wheat germ extract used for cell-free protein synthesis, or in *E. coli* by co-expression with the MAPK14 kinase. We identified two major phosphorylation sites, on S6 and S98, and seven minor ones, using a combination of mass spectrometry and NMR. We found phosphorylation to occur in all major functional regions of PreS, which raises the possibility that phosphorylation is involved in the regulation of these functions in the human virus protein, although this has not been identified in the avian virus homolog. We however found also that full phosphorylation is provided in neither recombinant system, which positions phosphorylation mimics by S/T to E mutations as the best strategy to explore the impact of phosphorylation on preS interactions in structural studies.

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 below: https://bmrb.io, 51186.

AUTHOR CONTRIBUTIONS

AB, M-LF, and LL conceived and planned the experiments. M-LF, RM, and GD carried out cell-free syntheses, LC carried out bacterial expression, and MN designed bacterial expression vectors. LL carried out NMR analyses. AP and FD carried out mass spectrometry analyses. M-LF, LL, MN, and AB interpreted results and wrote the manuscript, with input from all authors.

ACKNOWLEDGMENTS

Financial support from the ANRS (ECTZ71388), the CNRS (CNRS-Momentum 2018) and the IR-RMN-THC Fr3050 CNRS for conducting the research is gratefully acknowledged. The authors acknowledge the financial support from ITMO Cancer AVIESAN (Alliance Nationale pour les Sciences de la Vie et de la Santé, National Alliance for Life Sciences and Health) within the framework of the cancer plan for Orbitrap mass spectrometer funding.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2021.821755/full#supplementary-material

REFERENCES

Badillo, A., Receveur-Brechot, V., Sarrazin, S., Cantrelle, F.-X., Delolme, F., Fogeron, M.-L., et al. (2017). Overall Structural Model of NS5A Protein from Hepatitis C Virus and Modulation by Mutations Confering Resistance of Virus Replication to Cyclosporin a. *Biochemistry* 56, 3029–3048. doi:10.1021/acs.biochem.7b00212

Bruss, V. (1997). A Short Linear Sequence in the Pre-S Domain of the Large Hepatitis B Virus Envelope Protein Required for Virion Formation. *J. Virol.* 71, 9350–9357. doi:10.1128/jvi.71.12.9350-9357.1997

Bruss, V., and Thomssen, R. (1994). Mapping a Region of the Large Envelope Protein Required for Hepatitis B Virion Maturation. *J. Virol.* 68, 1643–1650. doi:10.1128/jvi.68.3.1643-1650.1994

David, G., Fogeron, M.-L., Montserret, R., Lecoq, L., Page, A., Delolme, F., et al. (2019). Phosphorylation and Alternative Translation on Wheat Germ Cell-free Protein Synthesis of the DHBV Large Envelope Protein. *Front. Mol. Biosci.* 6, 138. doi:10.3389/fmolb.2019.00138

David, G., Fogeron, M. L., Schledorn, M., Montserret, R., Haselmann, U., Penzel, S., et al. (2018). Structural Studies of Self-Assembled Subviral Particles: Combining Cell-Free Expression with 110 kHz MAS NMR Spectroscopy. *Angew. Chem. Int. Ed.* 57, 4787–4791. doi:10.1002/anie.201712091

Delgado, C. L., Núñez, E., Yélamos, B., Gómez-Gutiérrez, J., Peterson, D. L., and Gavilanes, F. (2015). Study of the Putative Fusion Regions of the preS Domain of Hepatitis B Virus. *Biochim. Biophys. Acta (Bba) - Biomembranes* 1848, 895–906. doi:10.1016/j.bbamem.2014.12.020

Favier, A., and Brutscher, B. (2019). NMRlib: User-Friendly Pulse Sequence Tools for Bruker NMR Spectrometers. *J. Biomol. NMR* 73, 199–211. doi:10.1007/s10858-019-00249-1

Fogeron, M.-L., Badillo, A., Penin, F., and Böckmann, A. (2017). Wheat Germ Cell-free Overexpression for the Production of Membrane Proteins. *Methods Mol. Biol.* 1635, 91–108. doi:10.1007/978-1-4939-7151-0_5

Fogeron, M.-L., Jirasko, V., Penzel, S., Paul, D., Montserret, R., Danis, C., et al. (2016). Cell-free Expression, Purification, and Membrane Reconstitution for NMR Studies of the Nonstructural Protein 4B from Hepatitis C Virus. *J. Biomol. NMR* 65, 87–98. doi:10.1007/s10858-016-0040-2

Fogeron, M.-L., Lecoq, L., Cole, L., Harbers, M., and Böckmann, A. (2021). Easy Synthesis of Complex Biomolecular Assemblies: Wheat Germ Cell-free Protein Expression in Structural Biology. *Front. Mol. Biosci.* 8, 639587. doi:10.3389/fmolb.2021.639587

Fogeron, M.-L., Paul, D., Jirasko, V., Montserret, R., Lacabanne, D., Molle, J., et al. (2015). Functional Expression, Purification, Characterization, and Membrane Reconstitution of Non-structural Protein 2 from Hepatitis C Virus. *Protein Expr. Purif.* 116, 1–6. doi:10.1016/j.pep.2015.08.027

Glebe, D., Urban, S., Knoop, E. V., Çağ, N., Krass, P., Grün, S., et al. (2005). Mapping of the Hepatitis B Virus Attachment Site by Use of Infection-Inhibiting preS1 Lipopeptides and tupaia Hepatocytes. *Gastroenterology* 129, 234–245. doi:10.1053/j.gastro.2005.03.090

Grgacic, E. V., and Anderson, D. A. (1994). The Large Surface Protein of Duck Hepatitis B Virus Is Phosphorylated in the Pre-S Domain. *J. Virol.* 68, 7344–7350. doi:10.1128/jvi.68.11.7344-7350.1994

Grgacic, E. V., Gazina, E. V., Lin, B., Anderson, D. A., and Snooks, M. J. (1998). Normal Phosphorylation of Duck Hepatitis B Virus L Protein Is Dispensable for Infectivity. *J. Gen. Virol.* 79 (Pt 11), 2743–2751. doi:10.1099/0022-1317-79-11-2743

Gripon, P., Cannie, I., and Urban, S. (2005). Efficient Inhibition of Hepatitis B Virus Infection by Acylated Peptides Derived from the Large Viral Surface Protein. *J. Virol.* 79, 1613–1622. doi:10.1128/JVI.79.3.1613-1622.2005

Gudima, S., Meier, A., Dunbrack, R., Taylor, J., and Bruss, V. (2007). Two Potentially Important Elements of the Hepatitis B Virus Large Envelope Protein Are Dispensable for the Infectivity of Hepatitis Delta Virus. *J. Virol.* 81, 4343–4347. doi:10.1128/JVI.02478-06

Heger-Stevic, J., Zimmermann, P., Lecoq, L., Böttcher, B., and Nassal, M. (2018). Hepatitis B Virus Core Protein Phosphorylation: Identification of the SRPK1 Target Sites and Impact of Their Occupancy on RNA Binding and Capsid Structure. *Plos Pathog.* 14, e1007488. doi:10.1371/journal.ppat.1007488

Hendus-Altenburger, R., Fernandes, C. B., Bugge, K., Kunze, M. B. A., Boomsma, W., and Kragelund, B. B. (2019). Random Coil Chemical Shifts for Serine, Threonine and Tyrosine Phosphorylation over a Broad pH Range. *J. Biomol. NMR* 73, 713–725. doi:10.1007/s10858-019-00283-z

Hull, W. E., and Kricheldorf, H. R. (1980). 15N-nmr Spectroscopy. 20.Cis/trans Isomerism and Neighboring Residue Effects of Proline-Containing peptidesCis/trans Isomerism and Neighboring Residue Effects of Proline-Containing Peptides. *Biopolymers* 19, 1103–1122. doi:10.1002/bip.1980.360190602

Le Pogam, S., and Shih, C. (2002). Influence of a Putative Intermolecular Interaction between Core and the Pre-S1 Domain of the Large Envelope Protein on Hepatitis B Virus Secretion. *J. Virol.* 76, 6510–6517. doi:10.1128/JVI.76.13.6510-6517.2002

Li, F., Li, C., Marquez-Lago, T. T., Leier, A., Akutsu, T., Purcell, A. W., et al. (2018). Quokka: a Comprehensive Tool for Rapid and Accurate Prediction of Kinase Family-specific Phosphorylation Sites in the Human Proteome. *Bioinformatics* 34, 4223–4231. doi:10.1093/bioinformatics/bty522

Mak, Y., Skylas, D. J., Willows, R., Connolly, A., Cordwell, S. J., Wrigley, C. W., et al. (2006). A Proteomic Approach to the Identification and Characterisation of Protein Composition in Wheat Germ. *Funct. Integr. Genomics* 6, 322–337. doi:10.1007/s10142-005-0018-8

Makbul, C., Nassal, M., and Böttcher, B. (2020). Slowly Folding Surface Extension in the Prototypic Avian Hepatitis B Virus Capsid Governs Stability. *eLife* 9, e57277. doi:10.7554/eLife.57277

Núñez, E., Yélamos, B., Delgado, C., Gómez-Gutiérrez, J., Peterson, D. L., and Gavilanes, F. (2009). Interaction of preS Domains of Hepatitis B Virus with Phospholipid Vesicles. *Biochim. Biophys. Acta (Bba) - Biomembranes* 1788, 417–424. doi:10.1016/j.bbamem.2008.10.014

Oess, S., and Hildt, E. (2000). Novel Cell Permeable Motif Derived from the PreS2-Domain of Hepatitis-B Virus Surface Antigens. *Gene Ther.* 7, 750–758. doi:10.1038/sj.gt.3301154

Pérez-Vargas, J., Teppa, E., Amirache, F., Boson, B., Pereira de Oliveira, R., Combet, C., et al. (2021). A Fusion Peptide in preS1 and the Human Protein Disulfide Isomerase ERp57 Are Involved in Hepatitis B Virus Membrane Fusion Process. *eLife* 10, e64507. doi:10.7554/eLife.64507

Persing, D. H., Varmus, H. E., and Ganem, D. (1987). The preS1 Protein of Hepatitis B Virus Is Acylated at its Amino Terminus with Myristic Acid. *J. Virol.* 61, 1672–1677. doi:10.1128/JVI.61.5.1672-1677.1987

Poisson, F., Severac, A., Hourioux, C., Goudeau, A., and Roingeard, P. (1997). Both pre-S1 and S Domains of Hepatitis B Virus Envelope Proteins Interact with the Core Particle. *Virology* 228, 115–120. doi:10.1006/viro.1996.8367

Prange, R. (2012). Host Factors Involved in Hepatitis B Virus Maturation, Assembly, and Egress. *Med. Microbiol. Immunol.* 201, 449–461. doi:10.1007/s00430-012-0267-9

Prange, R., Werr, M., and Löffler-Mary, H. (1999). Chaperones Involved in Hepatitis B Virus Morphogenesis. *Biol. Chem.* 380, 305–314. doi:10.1515/BC.1999.042

Rothmann, K., Schnölzer, M., Radziwill, G., Hildt, E., Moelling, K., and Schaller, H. (1998). Host Cell-Virus Cross Talk: Phosphorylation of a Hepatitis B Virus Envelope Protein Mediates Intracellular Signaling. *J. Virol.* 72, 10138–10147. doi:10.1128/jvi.72.12.10138-10147.1998

Sawasaki, T., Ogasawara, T., Morishita, R., and Endo, Y. (2002). A Cell-free Protein Synthesis System for High-Throughput Proteomics. *Proc. Natl. Acad. Sci.* 99, 14652–14657. doi:10.1073/pnas.232580399

Schmidt, T. G., and Skerra, A. (2007). The Strep-Tag System for One-step Purification and High-Affinity Detection or Capturing of Proteins. *Nat. Protoc.* 2, 1528–1535. doi:10.1038/nprot.2007.209

Slagle, B. L., and Bouchard, M. J. (2018). Role of HBx in Hepatitis B Virus Persistence and its Therapeutic Implications. *Curr. Opin. Virol.* 30, 32–38. doi:10.1016/j.coviro.2018.01.007

Solyom, Z., Schwarten, M., Geist, L., Konrat, R., Willbold, D., and Brutscher, B. (2013). BEST-TROSY Experiments for Time-Efficient Sequential Resonance Assignment of Large Disordered Proteins. *J. Biomol. NMR* 55, 311–321. doi:10.1007/s10858-013-9715-0

Stevens, T. J., Fogh, R. H., Boucher, W., Higman, V. A., Eisenmenger, F., Bardiaux, B., et al. (2011). A Software Framework for Analysing Solid-State MAS NMR Data. *J. Biomol. NMR* 51, 437–447. doi:10.1007/s10858-011-9569-2

Sun, H., Chang, L., Yan, Y., and Wang, L. (2021). Hepatitis B Virus pre-S Region: Clinical Implications and Applications. *Rev. Med. Virol.* 31, e2201. doi:10.1002/rmv.2201

Takai, K., Sawasaki, T., and Endo, Y. (2010). Practical Cell-free Protein Synthesis System Using Purified Wheat Embryos. *Nat. Protoc.* 5, 227–238. doi:10.1038/nprot.2009.207

Ubersax, J. A., Woodbury, E. L., Quang, P. N., Paraz, M., Blethrow, J. D., Shah, K., et al. (2003). Targets of the Cyclin-dependent Kinase Cdk1. *Nature* 425, 859–864. doi:10.1038/nature02062

Urban, S., Schwarz, C., Marx, U. C., Zentgraf, H., Schaller, H., and Multhaup, G. (2000). Receptor Recognition by a Hepatitis B Virus Reveals a Novel Mode of High Affinity Virus-Receptor Interaction. *EMBO J.* 19, 1217–1227. doi:10.1093/emboj/19.6.1217

Vranken, W. F., Boucher, W., Stevens, T. J., Fogh, R. H., Pajon, A., Llinas, M., et al. (2005). The CCPN Data Model for NMR Spectroscopy: Development of a Software Pipeline. *Proteins* 59, 687–696. doi:10.1002/prot.20449

Wang, C., Xu, H., Lin, S., Deng, W., Zhou, J., Zhang, Y., et al. (2020). GPS 5.0: An Update on the Prediction of Kinase-specific Phosphorylation Sites in Proteins. *Genomics, Proteomics & Bioinformatics* 18, 72–80. doi:10.1016/j.gpb.2020.01.001

Wang, Y., and Jardetzky, O. (2002). Probability-based Protein Secondary Structure Identification Using Combined NMR Chemical-Shift Data. *Protein Sci.* 11, 852–861. doi:10.1110/ps.3180102

Wei, K., and Li, Y. (2019). Functional Genomics of the Protein Kinase Superfamily from Wheat. *Mol. Breed.* 39, 1–23. doi:10.1007/s11032-019-1045-9

Xi, J., Liu, H., and Hu, J. (2021). Regulation of Hepatitis B Virus Virion Release and Envelopment Timing by Nucleocapsid and Envelope Interactions. *J. Virol.*, JVI0130521. doi:10.1128/JVI.01305-21

Yan, H., Zhong, G., Xu, G., He, W., Jing, Z., Gao, Z., et al. (2012). Sodium Taurocholate Cotransporting Polypeptide Is a Functional Receptor for Human Hepatitis B and D Virus. *eLife* 1, e00049. doi:10.7554/eLife.00049

Yang, F. (2018). Post-translational Modification Control of HBV Biological Processes. *Front. Microbiol.* 9, 2661. doi:10.3389/fmicb.2018.02661

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

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

frontiers | Frontiers in Molecular Biosciences

MINI REVIEW
published: 08 April 2022
doi: 10.3389/fmolb.2022.822852

Contribution of Infrared Spectroscopy to the Understanding of Amyloid Protein Aggregation in Complex Systems

Diletta Ami[1]*, *Paolo Mereghetti*[2] *and Antonino Natalello*[1]*

[1]Department of Biotechnology and Biosciences, University of Milano-Bicocca, Milano, Italy, [2]Bioinformatics Consultant, Arquata Scrivia, Italy

Infrared (IR) spectroscopy is a label-free and non-invasive technique that probes the vibrational modes of molecules, thus providing a structure-specific spectrum. The development of infrared spectroscopic approaches that enable the collection of the IR spectrum from a selected sample area, from micro- to nano-scale lateral resolutions, allowed to extend their application to more complex biological systems, such as intact cells and tissues, thus exerting an enormous attraction in biology and medicine. Here, we will present recent works that illustrate in particular the applications of IR spectroscopy to the *in situ* characterization of the conformational properties of protein aggregates and to the investigation of the other biomolecules surrounding the amyloids. Moreover, we will discuss the potential of IR spectroscopy to the monitoring of cell perturbations induced by protein aggregates. The essential support of multivariate analyses to objectively pull out the significant and non-redundant information from the spectra of highly complex systems will be also outlined.

Keywords: amyloid, sub-cellular, microspectroscopy, proteins, infrared spectroscopy

OPEN ACCESS

Edited by:
Piero Andrea Temussi,
University of Naples Federico II, Italy

Reviewed by:
Vittorio Bellotti,
University College London,
United Kingdom

***Correspondence:**
Diletta Ami
diletta.ami@unimib.it
Antonino Natalello
antonino.natalello@unimib.it

Specialty section:
This article was submitted to Structural Biology, a section of the journal Frontiers in Molecular Biosciences

Received: *26 November 2021*
Accepted: *14 March 2022*
Published: *08 April 2022*

Citation:
Ami D, Mereghetti P and Natalello A (2022) Contribution of Infrared Spectroscopy to the Understanding of Amyloid Protein Aggregation in Complex Systems. Front. Mol. Biosci. 9:822852. doi: 10.3389/fmolb.2022.822852

INTRODUCTION

The average life expectancy has been increasing at a rapid rate in industrialized countries, but unfortunately with a longer life expectancy comes a higher chance of developing a neurodegenerative disease, such as Alzheimer's, Parkinson's, Huntington's, and others. A common key molecular pathway implicated in different neurodegenerative and non-neurodegenerative amyloid diseases is the misfolding and aggregation of proteins that cause cellular toxicity and contribute to cell proteostasis collapse, a key hallmark of cell aging (Chiti and Dobson, 2006; Chiti and Dobson, 2017; Moreno et al., 2019; Santra et al., 2019). In this regard, despite significant efforts aimed at understanding and counteracting the pathogenic cascades induced by misfolded or aggregated proteins, the comprehension of the molecular bases of the mechanisms that lead to their formation and to the induced cell toxicity is still lacking. To this aim, the development of approaches that enable the characterization of protein aggregates in their natural environment and that provide a global insight in the physiological alterations induced by their formation is of huge importance. Noteworthy, in the 90's one of the first direct evidence of the β-sheet structures in the β-amyloid plaque was provided *in situ* by synchrotron radiation (SR) Fourier transform infrared (IR) microspectroscopy (SR-μFTIR), analyzing the brain from an Alzheimer's disease patient (Choo et al., 1996). This important result paved the way to other *in situ* studies that employed vibrational

spectroscopies to shed light on the molecular bases of protein misfolding diseases, emerging as promising tools in the field of neurodegeneration.

IR spectroscopy holds a great deal of promise for studying *in situ* the amyloid structural properties and aggregation mechanisms. In particular, this vibrational tool enables the sample characterization in a label-free and non-destructive way, a characteristic that allows complementary analyses on the same sample, which does not require pre-analytical chemical modification. However, since the standard protocols for medical histopathology diagnosis include sample fixing, several works on infrared spectroscopy of fixed specimens are reported (Meuse and Barker, 2009). Indeed, fixation facilitates sample storage and could be useful to correlate routine histomorphologic microscopic examinations with the chemico-physical information obtained by IR spectroscopy. Nonetheless, several drawbacks of fixed sample characterization exist, because it has become evident that how the sample is handled prior to IR analysis is crucial to obtain reliable structural information, and particularly the interference due to external substances, such as dyes or fixatives or even culture media, has to be taken into consideration (Zohdi et al., 2015). In this perspective, to apply IR spectroscopy for the investigation of the structural properties of biological samples, it is imperative to evaluate how different methods of sample preparation could affect the IR spectra, and to analyze unfixed samples when possible, thus avoiding misleading interpretation of the spectral data.

Interestingly, numerous works are reported in literature, illustrating the ability of this vibrational approach to detect *in situ* changes in protein structural features and aggregate properties associated with different disorders. The possibility to highlight biochemical and structural changes directly in cells, in tissues or biofluids provides on one hand a tool for an early diagnosis of the disease, on the other the potential to get new insights into the mechanisms of amyloid formation and toxicity. Since the IR Amide I band is sensitive to fine protein structural details (Barth, 2007; Zandomeneghi et al., 2004; Seshadri et al., 1999; Sarroukh et al., 2013), representing a fingerprint of its conformational properties, IR spectroscopy has been used to characterize aggregation intermediates and amyloid fibrils. For instance, conventional FTIR spectroscopy (Surmacz-Chwedoruk et al., 2012; Bhak et al., 2014) and super-resolution IR imaging methods (Klementieva et al., 2020) have been employed to study amyloid polymorphism. Interestingly, isotope-edited IR has been applied to study conformational changes and aggregation or co-aggregation of two proteins copresent in the same sample (Natalello et al., 2016). Then, an interesting example of the application of IR spectroscopy at nanometer lateral resolution is related to the aggregation mechanism of the Josephin domain of ataxin-3 protein. The combination of atomic force microscopy and infrared spectroscopy allowed the identification of oligomeric intermediates sharing the structural properties of the native protein or being more similar to fibrils. These results indicate an aggregation process involving an initial formation of native-like assemblies, therefore suggesting a model of "first-aggregation-then-misfolding" for this protein (Ruggeri et al., 2015).

Notably, the possibility to characterize biofluids—including blood derivatives, saliva, urine and tears—could make it possible to develop a spectroscopic method for the early detection of neurodegenerative biomarkers in patient' samples collected in a non-invasive way (Paraskevaidi et al., 2017). For instance, FTIR spectroscopy of blood plasma samples showed a decreasing trend in Alzheimer's patients compared with healthy controls (HCs) for the ratios of lipid-to-protein, phosphate-to-carbohydrate, and RNA-to-DNA (Paraskevaidi et al., 2017). Moreover, El Khoury and others reported that the FTIR characterization of serum samples from patients affected by multiple sclerosis and amyotrophic lateral sclerosis (ALS) led to a satisfactory differentiation of the two pathologies (El Khoury et al., 2019). In particular, the Authors found that the most important spectral differences were ascribable to variations in the structure of DNA and RNA and to modifications of glycolipids, glycoproteins, and collagen (El Khoury et al., 2019). Recently, we applied vibrational spectroscopy supported by multivariate analysis for the characterization of tears—a very promising biofluid for diagnostic purposes—from patients affected by ALS and HCs. Particularly, an excellent discrimination of the two sample classes was obtained, and it was possible to disclose ALS spectroscopic markers related to protein and lipid alterations compared with HCs (Ami et al., 2021a).

It should be emphasized that these objectives are also achievable thanks to important advances in both instrumentation—that allows extending applications to spatially resolved analysis on nanometric scale—and in multivariate analysis methods that enable the investigation of big spectral data-set in a large variety of samples.

Infrared Spectroscopy: From Micro- to Nano-Scale Lateral Resolutions

Infrared spectroscopy has a long history in biophysics, being extensively applied to the study of the secondary structure and aggregation of proteins *in vitro* (Barth, 2007; Natalello and Doglia, 2015; Seshadri et al., 1999; Sarroukh et al., 2013), as well as of the structural properties of lipids, carbohydrates, and nucleic acids (Casal and MantschKuhn, 1950; Sinclair at al., 1952a; Sinclair at al., 1952b; Zhizhina and Oleinik, 1972, 1984; Taillandier and Liquier, 1992; Kačuráková and Wilson, 2001). Particularly, IR spectroscopy allows to characterize the protein secondary structures mainly through the analysis of the Amide I band (1,700–1,600 cm^{-1}), due to the CO stretching vibrations of the peptide bond, where β-sheets in native proteins and in protein aggregates display different components (Seshadri et al., 1999; Zandomeneghi et al., 2004; Barth, 2007; Sarroukh et al., 2013; Natalello and Doglia, 2015).

In the last few decades, supported by technological advancements that led to the improvement of infrared sources, optics, and detectors, the combination of microscopy with FTIR spectroscopy made it possible to extend the IR investigations to more complex biological systems, including intact cells and tissues, thus bringing about its renaissance in structural biology (Miller and Dumas, 2010). In particular, the possibility to study selected areas of the sample under investigation makes

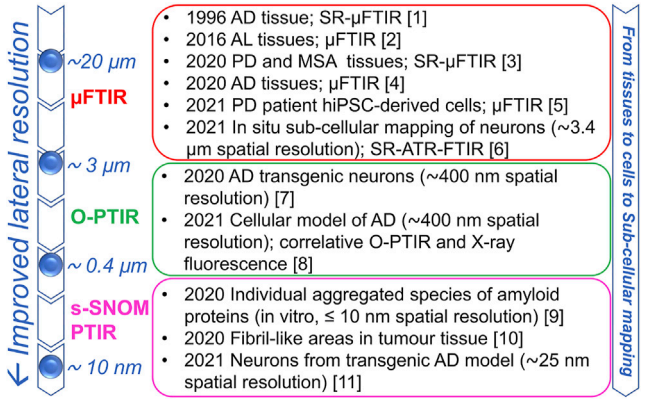

FIGURE 1 | Infrared spectroscopy approaches: from the micro- to the nano-scale characterization of biological samples. The lateral resolutions and selected examples for the *in situ* study of protein aggregation are reported for different IR approaches. The increased resolution enables characterizing protein aggregates in complex systems from tissues, to intact cells and to sub-cellular areas. References: (1) Choo et al., 1996; (2) Ami et al., 2016; (3) Araki et al., 2020; (4) Röhr et al., 2020; (5) Gustavsson et al., 2021a; (6) Hartnell et al., 2021; (7) Klementieva et al., 2020; (8) Gustavsson et al., 2021b; (9) Waeytens et al., 2020 (10) Paluszkiewicz et al., 2020 (11) Freitas et al., 2021.

the μFTIR approach particularly useful for the analysis of systems characterized by an intrinsic heterogeneity, such as precisely biological systems. Moreover, depending on the instrument characteristics, diffraction-limited spatial resolution can be achieved, thus extending the application of this technique to cells and subcellular compartments (**Figure 1**). In particular, when the microscope is equipped with conventional IR source and detector, the spatial resolution is of about a few tens of microns per side. Then, with a focal plane array detector that enables not only to collect the IR absorption spectrum of the sample but also an IR chemical imaging, a sampling area of a few microns, near to the diffraction limit, can be measured with a good signal to noise ratio (Centrone, 2015). Furthermore, the use of a synchrotron IR light source, with a brightness of at least two orders of magnitude higher than that of a conventional thermal source, even allows achieving diffraction-limited spatial resolution with enhanced signal to noise ratio, making it possible to explore the IR spectra at the subcellular level. However, even under the best conditions, the employed long wavelength and the lack of high numerical aperture objectives limit the spatial resolution at about 3–10 μm in the mid-IR range and it is a function of the wavenumber employed in the μFTIR imaging (Centrone, 2015).

A submicrometer spatial resolution can, instead, be achieved by the new technique called optical photothermal infrared (O-PTIR) super-resolution imaging. In this technique, a pulsed IR tunable laser is focused onto the sample where the IR absorption causes a temperature increase that changes the refractive index. The IR absorption is locally detected as a change in sample reflectivity using a visible laser which is collinear with the IR beam. In this set-up it is the visible laser at a fixed wavelength that determines the spatial resolution. Under the optimal conditions, spatial resolution of about 1 μm–400 nm has been reported (Klementieva et al., 2020; Zhang et al., 2016).

Interestingly, to achieve a spectral resolution beyond the diffraction limit two main methods based on the coupling of IR spectroscopy with atomic force microscopy (AFM) have been developed: the scattering scanning near-field optical microscopy (s-SNOM) and the photothermal-induced resonance (PTIR) (Centrone, 2015). In the s-SNOM method, the measurement of the amplitude and phase of light scattered from the tip close to the sample makes it possible to obtain the IR absorption coefficient of a thin portion of the sample in proximity to the tip. Although this approach allows a nanometer (≈20 nm) resolution, it should be noted that the IR s-SNOM spectra are not directly comparable to the standard far-field IR spectra (Centrone, 2015).

In PTIR, the absorption of light from a pulsed, wavelength-tunable laser induces the sample thermal expansion that can be recorded by an AFM cantilever. In this case, the lateral resolution (which is in the order of 20–100 nm) depends on the tip size and on the sample thermomechanical properties (Centrone, 2015). Therefore, PTIR and O-PTIR provide an indirect measurement of the vibrational absorption using a different set-up, which results in different lateral resolution, penetration capacity and requirements of the sample conditions. Of note, both PTIR and O-PTIR spectra are expected to be directly comparable with those measured by traditional FTIR spectrometers (Centrone, 2015; Zhang et al., 2016; Klementieva et al., 2020).

Multivariate Analysis of Complex Biological System Spectroscopic Data

IR (micro)-spectroscopy monitors the global biochemical composition of the sample under investigation; therefore, the resulting spectrum is a superposition of the vibrational signatures of the different sample biomolecules. Particularly when applied to investigate the response of living systems under diverse physiological and pathological conditions, spectral evaluation may become significantly complicated, also because in some cases minor differences may contain critical information (Ami et al., 2013; Morais et al., 2020). For this reason, the interpretation of the spectral results requires the support of appropriate multivariate analysis approaches able to tackle the study of high-dimensional data and to explore the whole spectral information simultaneously. Multivariate analysis creates a statistical model which describes the relationship between the spectra and the cellular response. In doing this, multivariate analysis also verifies the statistical significance of the results and pulls out the spectral components carrying the higher spectral variance and which are important in describing the underlying cellular conditions.

In the early decades of the 20th century, the pioneering work of R. A. Fisher contributed to the development of multivariate statistical methodology for applications in various fields, marking thus the beginning of research in multivariate analysis as a statistical technique for drawing inferences from multivariate data and opening up the doors to new knowledge through appropriate analyses of data (Radhakrishna Rao, 1992;

Anderson, 1996; Rencher, 2002). Inspired by his work, multivariate analysis soon became popular, and especially applied to those systems which can be described using a large number of variables (i.e., biological systems), whose connections or associations need to be investigated. Then, in the last two decades, the increasing processing power of computers and the improvements in statistical algorithms allowed an impressive development of multivariate analysis methodologies (Baştanlar and Ozuysal, 2014) including discriminant analysis and class-modeling techniques where multiple spectral variables are analyzed simultaneously to discriminate and assign unknown samples to pre-defined classes (Morais et al., 2020; Zeng et al., 2021). Indeed, the rapid successful development of IR spectroscopy in biomedical research further highlighted the crucial role provided by advanced machine learning techniques, e.g., deep-learning neural networks of complex datasets to extract important information and visualize it in a readily interpretable form (Morais et al., 2020; Krafft et al., 2009; Marini et al., 2008).

Selected Recent Examples: Exploiting the Potentialities of Infrared Spectroscopy for Amyloid Studies in Complex Systems

In the following, we will discuss a few selected works where the Authors applied IR spectroscopic methods to investigate protein aggregates, both *in vitro* and *in situ* within complex systems, with particular attention to amyloid aggregation.

Having the unique advantage of being a label-free method, spatially resolved FTIR microspectroscopy can provide information on the content and structure of different biomolecules simultaneously. This peculiarity enabled us to characterize not only amyloid deposits within unfixed tissues from patients affected by light chain (AL) amyloidosis, but also the biomolecules within or surrounding the protein aggregates. This possibility contributed to demonstrate a connection between lipid distribution and amyloidogenesis also in AL amyloidosis, further highlighting the presence of common features among different protein misfolding diseases (Ami et al., 2016). The possibility of investigating these properties *in situ* and at the individual level is priceless. Indeed, it makes it possible to relate the biophysical features of the fibrils to their specific pathogenic profile and to the histological properties of the amyloid plaques, thus contributing to fill the gap in the knowledge about the bases of amyloidosis at the individual level.

The fundamental role of the brain environment in determining the structural and pathological features of protein aggregates has been investigated in several diseases such as in Parkinson's disease (PD) and multiple system atrophy (MSA). Of note, in both cases α-synuclein is the most abundant protein in the respective brain inclusions: Lewy bodies (LBs) in PD and glial cytoplasmic inclusions (GCIs) in MSA. Unfixed post-mortem brain sections from individuals with PD or MSA diagnosis have been analyzed by synchrotron microFTIR, which disclosed significantly higher β-sheet structures in LBs compared to GCIs. This result not only confirmed that α-synuclein deposits formed in the complex biological environment are enriched in β-sheet elements, as found for amyloid-like structures induced in the test tube, but also disclosed important differences in the main conformation of the proteins accumulated in the brain of these different synucleinopathies (Araki et al., 2020).

Notably, the possibility offered by FTIR microspectroscopy to analyze unlabeled brain tissue sections from AD cases, with a spatial resolution able to investigate the amyloid plaque substructures, has recently provided evidence of the conformational events associated with plaque developments. In particular, the Amide I band analysis of different plaque types indicated an increased content of β-sheet structures from diffuse plaques to compact and classic cored plaques. Moreover, an increased weight of the low-wavenumber β-sheet band (at 1,628 cm^{-1}) in proportion to the high-wavenumber band (at 1,693 cm^{-1}) was observed in mature plaques (Röhr et al., 2020). These results suggest the formation of A-beta fibrils with a parallel β-sheet conformation during plaque developments in the brain of AD patients, as previously described *in vitro*.

One of the drawbacks of the analysis of post-mortem brain samples is that extensive protein aggregation already occurred, making it difficult to characterize the early stage of the process and, of course, to translate IR spectroscopy in a diagnostic setting in the case of neurodegenerative diseases. These limitations have been recently addressed using human induced pluripotent stem cells (hiPSC) from familial and idiopathic PD patients (Gustavsson et al., 2021a). FTIR microspectroscopy analyses allowed the simultaneous study of the protein secondary structures, the lipid oxidation and the lipid chain length, enabling to correlate differences in the structure and composition of the main biomolecules with the peculiar intracellular environment induced by each patient genome (Gustavsson et al., 2021a). The validation of such results in large cohorts of patients could represent the starting point for the development of new tools for PD diagnosis and patient stratification towards personalized medicine.

A further advancement to improve the *in situ* investigation of protein aggregates has been recently described by Hartnell and others (Hartnell et al., 2021). Here, the potential of the synchrotron attenuated total reflection (ATR) FTIR technique to study macromolecular inclusions, such as protein aggregates, at sub-cellular level, has been reported. In particular, it has been shown that SR-ATR-FTIR mapping technique makes it possible to achieve a diffraction limited spatial resolution of a few microns, allowing the *in situ* characterization of sub-cellular protein aggregates in degenerating neurons of 20–30 μm in size. This important result opens the possibility to get new insights into the mechanism of protein aggregate formation in brain tissue (Hartnell et al., 2021). Overall, the merit of this technological advancement in the improvement of spatial resolution is that it offers the possibility to visualize at the sub-cellular level markers of cell physiology and disease pathology, an ongoing challenge in biomedicine.

Among the methodologies recently developed for improving spatial resolution, Klementieva and others (Klementieva et al., 2020) illustrated the application of O-PTIR spectroscopy to detect protein aggregates directly in neurons, at a sub-cellular level.

O-PTIR, indeed, allows to tackle a very challenging task, that is to overcome the diffraction limit of IR light, providing a spatial resolution that is determined by the focusing visible laser achieving a few hundred nanometer resolution (Klementieva et al., 2020; Zhang et al., 2016).

By O-PTIR analysis on intact AD transgenic neurons, the Authors detected, for the first time, polymorphic protein aggregates at sub-cellular level, suggesting multiple Aβ aggregation mechanisms leading to assemblies with different structures and likely different neurotoxic effects (Klementieva et al., 2020).

Thanks to the recent technical advances of the O-PTIR instrumentations, this approach is also successfully applied to living cell imaging at sub-micrometer resolution. Interestingly, with this methodology Raman and IR spectra can be also collected simultaneously from the same sample area (Spadea et al., 2021).

Of note, very recently Gustavsson and others illustrated the application of O-PTIR spectroscopy combined to synchrotron-based X-ray fluorescence (S-XRF) to study amyloid aggregation processes in a cellular model of AD, at a sub-cellular level. In this way, it was possible to demonstrate directly in neurons the co-localization of iron clusters with elevated amyloid β-sheet structures and oxidized lipids (Gustavsson et al., 2021b).

Recent works that deserve mention illustrate the combination of two widely employed techniques for the characterization both *in vitro* (Waeytens et al., 2020) and *in situ* (Paluszkiewicz et al., 2020) of protein assemblies, namely atomic force microscopy and infrared spectroscopy. This method allows collecting images and infrared spectra at the nanometer scale, resulting, therefore, promising for the analysis of amyloid fibrils and other protein assemblies. Indeed, AFM-IR enables the correlation of the information contained in AFM images with the secondary structure and other information provided by the infrared spectra. As an *in situ* case study, in the AFM-IR analysis of the pleomorphic adenoma and the marginal tissue, fibril-like areas were observed in the AFM morphological characterization of the tumour tissue. Noteworthy, the AFM-IR approach enabled the collection of the IR spectra of the observed fibrils and of the surrounding area (Paluszkiewicz et al., 2020).

Moreover, the s-SNOM approach has been employed to measure individual cultured neurons from wild-type and transgenic AD model, enabling the identification of protein aggregates with a lateral resolution of ~25 nm. The obtained results provided a strong mark that β-sheet rich structures associated with cell membranes can be sensed at a nanometer lateral resolution by this label-free method (Freitas et al., 2021).

Interestingly, the possibility to provide the global biochemical composition of the probed sample supported the extension of IR spectroscopy to the investigation of the molecular bases of amyloid toxicity. As an example, we exploited FTIR microspectroscopy, combined with multivariate analysis, to analyze *in situ* the spectral changes occurring in cultured intact HL-1 cardiomyocytes exposed to transthyretin (TTR)—wild type (WT) or the highly amyloidogenic variant L55P—in native or amyloid-like conformations (Ami et al., 2018). In particular, this study allowed us to detect the cell perturbations—affecting lipid bilayer fluidity/compactness and the cell metabolic/phosphorylation status—induced by the proteins in their different conformational states (Ami et al., 2018).

Furthermore, with the purpose of investigating possible perturbations induced on intact cells by intracellular amyloidogenic proteins, we applied FTIR microspectroscopy, coupled to multivariate analysis, to study *Escherichia coli* cells—taken as model system—expressing the human ataxin-3 (ATX3), both a physiological (ATX3-Q24) and a pathological ATX3 variant (ATX3-Q55) (Ami et al., 2021b). Notably, by the FTIR analysis it was possible not only to monitor the protein aggregation, but also the induced cell perturbations. In particular, we found that the toxic oligomeric species associated with the expression of the pathological variant ATX3-Q55 were responsible for the main spectral changes, mostly ascribable to the cell envelope modifications, a result that was also supported by electron microscopy analysis (Ami et al., 2021b).

CONCLUSION AND FUTURE PERSPECTIVES

The growing interest in the application of IR spectroscopy in different fields of biological and biomedical research is accompanied by an ongoing technological advancement that makes it possible to extend the IR investigations to more complex biological systems. In particular, the development of more and more sophisticated methodologies, allowing to achieve or even to overcome diffraction-limited spatial resolution, enables the exploration of the IR response at sub-micrometer scale. This is a crucial result to obtain more detailed structural information on protein aggregates in intact cells and tissues, making IR spectroscopy a powerful tool for basic research as well as for clinical purposes. The possibility to characterize the conformational features of protein aggregates in their natural microenvironment has the potential to help us narrow the gap between our knowledge of amyloid aggregation mechanisms *in vitro* and *in vivo* (Faravelli et al., 2022).

AUTHOR CONTRIBUTIONS

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

FUNDING

This work was supported by the University of Milano-Bicocca (grant "Fondo di Ateneo per la Ricerca" to AN).

ACKNOWLEDGMENTS

DA acknowledges a postdoctoral fellowship of the University of Milano-Bicocca.

REFERENCES

Ami, D., Duse, A., Mereghetti, P., Cozza, F., Ambrosio, F., Ponzini, E., et al. (2021a). Tear-Based Vibrational Spectroscopy Applied to Amyotrophic Lateral Sclerosis. *Anal. Chem.* 93, 16995–17002. doi:10.1021/acs.analchem.1c02546

Ami, D., Lavatelli, F., Rognoni, P., Palladini, G., Raimondi, S., Giorgetti, S., et al. (2016). *In Situ* characterization of Protein Aggregates in Human Tissues Affected by Light Chain Amyloidosis: a FTIR Microspectroscopy Study. *Sci. Rep.* 6, 29096. doi:10.1038/srep29096

Ami, D., Mereghetti, P., Leri, M., Giorgetti, S., Natalello, A., Doglia, S. M., et al. (2018). A FTIR Microspectroscopy Study of the Structural and Biochemical Perturbations Induced by Natively Folded and Aggregated Transthyretin in HL-1 Cardiomyocytes. *Sci. Rep.* 8, 12508. doi:10.1038/s41598-018-30995-5

Ami, D., Mereghetti, P., and Doglia, S. M. (2013). "Multivariate Analysis for Fourier Transform Infrared Spectra of Complex Biological Systems and Processes," in *Multivariate Analysis*. Editors L. V. de Freitas and A. P. B. R. de Freitas (Norderstedt, Germany: Books on Demand). doi:10.5772/53850

Ami, D., Sciandrone, B., Mereghetti, P., Falvo, J., Catelani, T., Visentin, C., et al. (2021b). Pathological ATX3 Expression Induces Cell Perturbations in *E. coli* as Revealed by Biochemical and Biophysical Investigations. *Ijms* 22, 943. doi:10.3390/ijms22020943

Anderson, T. W. (1996). R. A. Fisher and Multivariate Analysis. *Statist. Sci.* 11, 20–34. doi:10.1214/ss/1032209662

Araki, K., Yagi, N., Ikemoto, Y., Hayakawa, H., Fujimura, H., Moriwaki, T., et al. (2020). The Secondary Structural Difference between Lewy Body and Glial Cytoplasmic Inclusion in Autopsy Brain with Synchrotron FTIR Micro-spectroscopy. *Sci. Rep.* 10, 19423. doi:10.1038/s41598-020-76565-6

Barth, A. (2007). Infrared Spectroscopy of Proteins. *Biochim. Biophys. Acta (Bba) - Bioenerg.* 1767, 1073–1101. doi:10.1016/j.bbabio.2007.06.004

Baştanlar, Y., and Özuysal, M. (2014). Introduction to Machine Learning. *Methods Mol. Biol.* 1107, 105–128. doi:10.1007/978-1-62703-748-8_7

Bhak, G., Lee, J., Kim, T.-H., Lee, S., Lee, D., and Paik, S. R. (2014). Molecular Inscription of Environmental Information into Protein Suprastructures: Temperature Effects on Unit Assembly of α-synuclein Oligomers into Polymorphic Amyloid Fibrils. *Biochem. J.* 464, 259–269. doi:10.1042/BJ20140723

Casal, H. L., and Mantsch, H. H. (1984). Polymorphic Phase Behaviour of Phospholipid Membranes Studied by Infrared Spectroscopy. *Biochim. Biophys. Acta (Bba) - Rev. Biomembranes* 779, 381–401. doi:10.1016/0304-4157(84)90017-0

Centrone, A. (2015). Infrared Imaging and Spectroscopy beyond the Diffraction Limit. *Annu. Rev. Anal. Chem.* 8, 101–126. doi:10.1146/annurev-anchem-071114-040435

Chiti, F., and Dobson, C. M. (2017). Protein Misfolding, Amyloid Formation, and Human Disease: A Summary of Progress over the Last Decade. *Annu. Rev. Biochem.* 86, 27–68. doi:10.1146/annurev-biochem-061516-045115

Chiti, F., and Dobson, C. M. (2006). Protein Misfolding, Functional Amyloid, and Human Disease. *Annu. Rev. Biochem.* 75, 333–366. doi:10.1146/annurev.biochem.75.101304.123901

Choo, L. P., Wetzel, D. L., Halliday, W. C., Jackson, M., LeVine, S. M., and Mantsch, H. H. (1996). *In Situ* characterization of Beta-Amyloid in Alzheimer's Diseased Tissue by Synchrotron Fourier Transform Infrared Microspectroscopy. *Biophysical J.* 71, 1672–1679. doi:10.1016/S0006-3495(96)79411-0

El Khoury, Y., Collongues, N., De Sèze, J., Gulsari, V., Patte-Mensah, C., Marcou, G., et al. (2019). Serum-based Differentiation between Multiple Sclerosis and Amyotrophic Lateral Sclerosis by Random Forest Classification of FTIR Spectra. *Analyst* 144, 4647–4652. doi:10.1039/c9an00754g

Faravelli, G., Mondani, V., Mangione, P. P., Raimondi, S., Marchese, L., Lavatelli, F., et al. (2022). Amyloid Formation by Globular Proteins: The Need to Narrow the Gap between *In Vitro* and *In Vivo* Mechanisms. *Front. Mol. Biosci.* 9, 830006. doi:10.3389/fmolb.2022.830006

Freitas, R. O., Cernescu, A., Engdahl, A., Paulus, A., Levandoski, J. E., Martinsson, I., et al. (2021). Nano-Infrared Imaging of Primary Neurons. *Cells* 10, 2559. doi:10.3390/cells10102559

Gustavsson, N., Paulus, A., Martinsson, I., Engdahl, A., Medjoubi, K., Klementiev, K., et al. (2021b). Correlative Optical Photothermal Infrared and X-ray Fluorescence for Chemical Imaging of Trace Elements and Relevant Molecular Structures Directly in Neurons. *Light Sci. Appl.* 10, 151. doi:10.1038/s41377-021-00590-x

Gustavsson, N., Savchenko, E., Klementieva, O., and Roybon, L. (2021a). The Intracellular Milieu of Parkinson's Disease Patient Brain Cells Modulates Alpha-Synuclein Protein Aggregation. *Acta Neuropathol. Commun.* 9, 153. doi:10.1186/s40478-021-01256-w

Hartnell, D., Hollings, A., Ranieri, A. M., Lamichhane, H. B., Becker, T., Sylvain, N. J., et al. (2021). Mapping Sub-cellular Protein Aggregates and Lipid Inclusions Using Synchrotron ATR-FTIR Microspectroscopy. *Analyst* 146, 3516–3525. doi:10.1039/d1an00136a

Kačuráková, M., and Wilson, R. H. (2001). Developments in Mid-infrared FT-IR Spectroscopy of Selected Carbohydrates. *Carbohydr. Polym.* 44, 291–303. doi:10.1016/S0144-8617(00)00245-9

Klementieva, O., Sandt, C., Martinsson, I., Kansiz, M., Gouras, G. K., and Borondics, F. (2020). Super-Resolution Infrared Imaging of Polymorphic Amyloid Aggregates Directly in Neurons. *Adv. Sci.* 7, 1903004. doi:10.1002/advs.201903004

Krafft, C., Steiner, G., Beleites, C., and Salzer, R. (2009). Disease Recognition by Infrared and Raman Spectroscopy. *J. Biophoton.* 2, 13–28. doi:10.1002/jbio.200810024

Kuhn, L. P. (1950). Infrared Spectra of Carbohydrates. *Anal. Chem.* 22, 276–283. doi:10.1021/ac60038a015

Marini, F., Bucci, R., Magrì, A. L., and Magrì, A. D. (2008). Artificial Neural Networks in Chemometrics: History, Examples and Perspectives. *Microchemical J.* 88, 178–185. doi:10.1016/j.microc.2007.11.008

Meuse, C. W., and Barker, P. E. (2009). Quantitative Infrared Spectroscopy of Formalin-Fixed, Paraffin-Embedded Tissue Specimens. *Appl. Immunohistochem. Mol. Morphol.* 17, 547–552. doi:10.1097/PAI.0b013e3181a9300e

Miller, L. M., and Dumas, P. (2010). From Structure to Cellular Mechanism with Infrared Microspectroscopy. *Curr. Opin. Struct. Biol.* 20, 649–656. doi:10.1016/j.sbi.2010.07.007

Morais, C. L. M., Lima, K. M. G., Singh, M., and Martin, F. L. (2020). Tutorial: Multivariate Classification for Vibrational Spectroscopy in Biological Samples. *Nat. Protoc.* 15, 2143–2162. doi:10.1038/s41596-020-0322-8

Moreno, D. F., Jenkins, K., Morlot, S., Charvin, G., Csikasz-Nagy, A., and Aldea, M. (2019). Proteostasis Collapse, a Hallmark of Aging, Hinders the Chaperone-Start Network and Arrests Cells in G1. *Elife* 8, e48240. doi:10.7554/eLife.48240

Natalello, A., and Doglia, S. M. (2015). Insoluble Protein Assemblies Characterized by Fourier Transform Infrared Spectroscopy. *Methods Mol. Biol.* 1258, 347–369. doi:10.1007/978-1-4939-2205-5_20

Natalello, A., Mangione, P. P., Giorgetti, S., Porcari, R., Marchese, L., Zorzoli, I., et al. (2016). Co-fibrillogenesis of Wild-type and D76N β2-Microglobulin: the Crucial Role of Fibrillar Seeds. *J. Biol. Chem.* 291, 9678–9689. doi:10.1074/jbc.M116.720573

Paluszkiewicz, C., Piergies, N., Guidi, M. C., Pięta, E., Ścierski, W., Misiołek, M., et al. (2020). Nanoscale Infrared Probing of Amyloid Formation within the Pleomorphic Adenoma Tissue. *Biochim. Biophys. Acta (Bba) - Gen. Subjects* 1864, 129677. doi:10.1016/j.bbagen.2020.129677

Paraskevaidi, M., Morais, C. L. M., Lima, K. M. G., Snowden, J. S., Saxon, J. A., Richardson, A. M. T., et al. (2017). Differential Diagnosis of Alzheimer's Disease Using Spectrochemical Analysis of Blood. *Proc. Natl. Acad. Sci. U.S.A.* 114, E7929–E7938. doi:10.1073/pnas.1701517114

Rao, C. R. (1992). R. A. Fisher: The Founder of Modern Statistics. *Statist. Sci.* 7, 34–48. doi:10.1214/ss/1177011442

Rencher, A. C. (2002). *Methods of Multivariate Analysis*. Hoboken: Wiley.

Röhr, D., Boon, B. D. C., Schuler, M., Kremer, K., Hoozemans, J. J. M., Bouwman, F. H., et al. (2020). Label-free Vibrational Imaging of Different Aβ Plaque Types in Alzheimer's Disease Reveals Sequential Events in Plaque Development. *Acta Neuropathol. Commun.* 8, 222. doi:10.1186/s40478-020-01091-5

Ruggeri, F. S., Longo, G., Faggiano, S., Lipiec, E., Pastore, A., and Dietler, G. (2015). Infrared Nanospectroscopy Characterization of Oligomeric and Fibrillar Aggregates during Amyloid Formation. *Nat. Commun.* 6, 7831. doi:10.1038/ncomms8831

Santra, M., Dill, K. A., and de Graff, A. M. R. (2019). Proteostasis Collapse Is a Driver of Cell Aging and Death. *Proc. Natl. Acad. Sci. U.S.A.* 116, 22173–22178. doi:10.1073/pnas.1906592116

Sarroukh, R., Goormaghtigh, E., Ruysschaert, J.-M., and Raussens, V. (2013). ATR-FTIR: a "rejuvenated" Tool to Investigate Amyloid Proteins. *Biochim. Biophys. Acta (Bba) - Biomembranes* 1828, 2328–2338. doi:10.1016/j.bbamem.2013.04.012

Seshadri, S., Khurana, R., and Fink, A. L. (1999). [36] Fourier Transform Infrared Spectroscopy in Analysis of Protein Deposits. *Methods Enzymol.* 309, 559–576. doi:10.1016/s0076-6879(99)09038-2

Sinclair, R. G., McKay, A. F., and Jones, R. N. (1952a). The Infrared Absorption Spectra of Saturated Fatty Acids and Esters1. *J. Am. Chem. Soc.* 74, 2570–2575. doi:10.1021/ja01130a033

Sinclair, R. G., McKay, A. F., Myers, G. S., and Jones, R. N. (1952b). The Infrared Absorption Spectra of Unsaturated Fatty Acids and Esters1. *J. Am. Chem. Soc.* 74, 2578–2585. doi:10.1021/ja01130a035

Spadea, A., Denbigh, J., Lawrence, M. J., Kansiz, M., and Gardner, P. (2021). Analysis of Fixed and Live Single Cells Using Optical Photothermal Infrared with Concomitant Raman Spectroscopy. *Anal. Chem.* 93, 3938–3950. doi:10.1021/acs.analchem.0c04846

Surmacz-Chwedoruk, W., Nieznańska, H., Wójcik, S., and Dzwolak, W. (2012). Cross-seeding of Fibrils from Two Types of Insulin Induces New Amyloid Strains. *Biochemistry* 51, 9460–9469. doi:10.1021/bi301144d

Taillandier, E., and Liquier, J. (1992). [16] Infrared Spectroscopy of DNA. *Methods Enzymol.* 211, 307–335. doi:10.1016/0076-6879(92)11018-e

Waeytens, J., Van Hemelryck, V., Deniset-Besseau, A., Ruysschaert, J.-M., Dazzi, A., and Raussens, V. (2020). Characterization by Nano-Infrared Spectroscopy of Individual Aggregated Species of Amyloid Proteins. *Molecules* 25, 2899. doi:10.3390/molecules25122899

Zandomeneghi, G., Krebs, M. R. H., McCammon, M. G., and Fändrich, M. (2004). FTIR Reveals Structural Differences between Native β-sheet Proteins and Amyloid Fibrils. *Protein Sci.* 13, 3314–3321. doi:10.1110/ps.041024904

Zeng, J., Guo, Y., Han, Y., Li, Z., Yang, Z., Chai, Q., et al. (2021). A Review of the Discriminant Analysis Methods for Food Quality Based on Near-Infrared Spectroscopy and Pattern Recognition. *Molecules* 26, 749. doi:10.3390/molecules26030749

Zhang, D., Li, C., Zhang, C., Slipchenko, M. N., Eakins, G., and Cheng, J.-X. (2016). Depth-resolved Mid-infrared Photothermal Imaging of Living Cells and Organisms with Submicrometer Spatial Resolution. *Sci. Adv.* 2, e1600521. doi:10.1126/sciadv.1600521

Zhizhina, G. P., and Oleinik, E. F. (1972). Infrared Spectroscopy of Nucleic Acids. *Russ. Chem. Rev.* 41, 258–280. doi:10.1070/RC1972v041n03ABEH002043

Zohdi, V., Whelan, D. R., Wood, B. R., Pearson, J. T., Bambery, K. R., and Black, M. J. (2015). Importance of Tissue Preparation Methods in FTIR Micro-spectroscopical Analysis of Biological Tissues: 'Traps for New Users'. *PLoS One* 10 (2), e0116491. doi:10.1371/journal.pone.0116491

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

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

frontiers | Frontiers in Molecular Biosciences

ORIGINAL RESEARCH
published: 15 June 2022
doi: 10.3389/fmolb.2022.865743

Williams-Beuren Syndrome Related Methyltransferase WBSCR27: From Structure to Possible Function

*Sofia S. Mariasina[1,2], Chi-Fon Chang[3], Tsimafei L. Navalayeu[4], Anastasia A. Chugunova[4], Sergey V. Efimov[5], Viktor G. Zgoda[6], Vasily A. Ivlev[7], Olga A. Dontsova[4,8], Petr V. Sergiev[2,4,8] and Vladimir I. Polshakov[1]**

[1]Faculty of Fundamental Medicine, M.V. Lomonosov Moscow State University, Moscow, Russia, [2]Institute of Functional Genomics, M.V. Lomonosov Moscow State University, Moscow, Russia, [3]Genomics Research Center, Academia Sinica, Taipei, Taiwan, [4]Chemical Department, M.V. Lomonosov Moscow State University, Moscow, Russia, [5]NMR Laboratory, Institute of Physics, Kazan Federal University, Kazan, Russia, [6]Institute of Biomedical Chemistry, Moscow, Russia, [7]Pharmacy Resource Center, RUDN University, Moscow, Russia, [8]Skolkovo Institute of Science and Technology, Moscow, Russia

Williams-Beuren syndrome (WBS) is a genetic disorder associated with the hemizygous deletion of several genes in chromosome 7, encoding 26 proteins. Malfunction of these proteins induce multisystemic failure in an organism. While biological functions of most proteins are more or less established, the one of methyltransferase WBSCR27 remains elusive. To find the substrate of methylation catalyzed by WBSCR27 we constructed mouse cell lines with a *Wbscr27* gene knockout and studied the obtained cells using several molecular biology and mass spectrometry techniques. We attempted to pinpoint the methylation target among the RNAs and proteins, but in all cases neither a direct substrate has been identified nor the protein partners have been detected. To reveal the nature of the putative methylation substrate we determined the solution structure and studied the conformational dynamic properties of WBSCR27 in apo state and in complex with S-adenosyl-L-homocysteine (SAH). The protein core was found to form a canonical Rossman fold common for Class I methyltransferases. N-terminus of the protein and the β6–β7 loop were disordered in apo-form, but binding of SAH induced the transition of these fragments to a well-formed substrate binding site. Analyzing the structure of this binding site allows us to suggest potential substrates of WBSCR27 methylation to be probed in further research.

Keywords: Williams-Beuren syndrome (WBS), methyltransferase (MTase), NMR, protein structure in solution, protein dynamics, S-adenosyl-L-homocysteine (SAH)

OPEN ACCESS

Edited by:
Annalisa Pastore, King's College London, United Kingdom

Reviewed by:
Anna Maria D'Ursi, University of Salerno, Italy
Sunny Sharma, Rutgers, The State University of New Jersey, United States

****Correspondence:***
Vladimir I. Polshakov
vpolsha@fbm.msu.ru

Specialty section:
This article was submitted to Structural Biology, a section of the journal Frontiers in Molecular Biosciences

Received: *30 January 2022*
Accepted: *16 May 2022*
Published: *15 June 2022*

Citation:
Mariasina SS, Chang C-F, Navalayeu TL, Chugunova AA, Efimov SV, Zgoda VG, Ivlev VA, Dontsova OA, Sergiev PV and Polshakov VI (2022) Williams-Beuren Syndrome Related Methyltransferase WBSCR27: From Structure to Possible Function. Front. Mol. Biosci. 9:865743. doi: 10.3389/fmolb.2022.865743

INTRODUCTION

Williams-Beuren syndrome (WBS) is a complex developmental disorder, induced by haploinsufficiency of 24–26 genes in the chromosome region 7q11 (Schubert, 2009). Multisystem disorders associated with this disease include aortic stenosis, hypercalcemia, impaired glucose metabolism, thyroid dysfunction, growth retardation, characteristic facial appearance, mental deficiency, and "friendly" personality which is usually considered as hyper-friendliness (Jones et al., 2000; Pober, 2010; Masserini et al., 2013).

For some phenotypic features, the impact of a specific gene deletion is already well-established. For instance, the ELN gene, being a part of WBS deletion, encodes the protein elastin, a component of

vascular walls, and its insufficiency leads to aortic stenosis (Ewart et al., 1993). Deletion of gene LIMK1, encoding LIM-kinase 1, brings about impaired visuospatial constructive cognition (Frangiskakis et al., 1996). There is also solid evidence on the insufficiency of BAZ1B contributing to hypercalcemia through interaction with the vitamin D receptor (Kitagawa et al., 2003). However, the impact of other genes lost in case of WBS manifestation remains unclear.

Uncovering the physiological consequences of gene loss on the behavior characteristic phenotype of WBS patients is among the most complicated directions in WBS studies. There is no direct evidence, but the decreased expression level of some gene products from the WBS chromosome region is likely to be involved. Interestingly, domestic dogs exhibit some of the behavioral traits typical of humans with WBS (vonHoldt et al., 2017): as compared to their ancestor, the gray wolf, domestic dogs have heightened propensity to initiate social contacts showing "hyper-sociability." Comparing the dog genome to the Yellowstone gray wolf one revealed mobile element insertions affecting transcriptional regulation in the genes responsible for WBS (vonHoldt et al., 2018). Transcriptome sequencing confirmed that the expression levels of six genes placed in WBS chromosome region (WBSCR17, LIMK1, GTF2I, WBSCR27, BAZ1B, and BCL7B) differ between these animals accounting for different behavior patterns.

One of the plausible candidate proteins associated with the behavioral aspects of WBS is WBSCR27. Human and chimpanzee genome sequences were compared and nine human-specific frameshift mutations were identified (Hahn and Lee, 2005). One of these mutations is placed within the WBSCR27 gene coding sequence: there is an 11 bp insertion in human WBSCR27. The insertion occurred specifically in the human lineage and probably could somehow affect the functioning of the protein; and thereby directly or indirectly alter human social behavior in comparison with chimpanzees. However, there is no direct experimental proof of this hypothesis yet.

There is a dearth of information about the biological function of WBSCR27. To outline some functional role of this protein only differential gene expression was measured. The expression level of WBSCR27 was reported to change in response to different external conditions. The overexpression of WBSCR27 was found in three tumor types: esophageal carcinoma, stomach adenocarcinoma, and kidney renal papillary cell carcinoma (Campeanu et al., 2021) by bioinformatic analysis of data available in TCGA (The Cancer Genome Atlas). WBSCR27 is overexpressed also in colon cancer and can be used as a prognostic marker of this desease (Wang et al., 2022).

Salvianolic acid B treatment was recently studied as a potential therapeutic approach for obesity (An et al., 2019). To examine the differential gene expression in mouse white adipose tissue caused by treating with Salvianolic acid B RNA-Seq was performed demonstrating that 234 lncRNAs, 19 circRNAs, and 132 mRNAs were differentially expressed. Among the mRNAs, the upregulated expression of WBSCR27 was the highest, with a fold change of 2.053. These results were confirmed by the qPCR. The other upregulated protein-coding genes were involved in the insulin resistance pathway, while the downregulated genes mainly participated in the IL-17 signaling pathway.

Nevertheless, these data do not shed any light on the possible role of WBSCR27 in WBS, as well as on its biological function in general. Metzger and colleagues (Metzger et al., 2019) previously identified C21orf127 MTase (later renamed to KMT9) to be responsible for histone lysine methylation. Surprisingly, this protein turned out to be a seven-β-stranded methyltransferase (MTase), while all histone lysine methylating proteins known before belonged to SET-domain family (Husmann and Gozani, 2019). To find out other histone lysine MTases within the seven-β-stranded MTase family, cluster analysis on multiple amino acid sequence alignments of putative seven-β-stranded MTase domains was performed. WBSCR27 was found among the seven closest homologs of C21orf127 and was tested for histone methylation in the *in vitro* assay, but did not show any methylation activity.

In our previous work (Mariasina et al., 2020) we demonstrated that WBSCR27 effectively interacts with the cofactor S-(5′-adenosyl)-L-methionine (SAM) and has a canonical Rossman fold, typical of Class I MTases. This information supports the bioinformatic assignment of WBSCR27 to MTases; however, the substrate of methylation catalyzed by this enzyme is still unknown. Here we determined the solution structure of WBSCR27 in apo-state and in complex with the cofactor. This work may shed light on the possible biological role of this protein making the complete mapping of gene deletions in WBS and physiological consequences one step closer.

MATERIALS AND METHODS

WBSCR27 Expression and Purification

The uniformly ^{15}N and $^{15}N/^{13}C$ enriched protein was expressed in *E. coli* cells grown on ^{15}N or $^{15}N/^{13}C$ M9 minimal media using the glucose-^{13}C (2 g/L) and/or ammonium sulphate-^{15}N (1 g/L) as a source of stable isotopes. The protein selectively ^{13}C-labelled in the methyl groups of Thr and Met residues was expressed in ^{15}N M9 media containing 100% D_2O and ISOGRO®-D supplemented by Met-ε-$^{13}CH_3$ and Thr-γ-$^{13}CH_3$ as well as fully deuterated 2-ketobutyrate and Gly-d_2 to prevent cross-labelling (Kerfah et al., 2015).

Protein samples were purified as described in Mariasina et al. (2018). Refolding was an important step in the purification procedure of the apo-form of the protein. In this case after affinity chromatography on the Ni-NTA column and subsequent His-tag cleavage, the protein samples were denatured in 6 M urea, washed from endogenous SAH by dialysis, and refolded back to the native form (Mariasina et al., 2018). Samples of the WBSCR27-SAH complex with identical content of the ^{13}C and ^{15}N isotopes in both protein and ligand were purified without refolding. In all other cases samples for NMR structural studies were prepared by adding the corresponding ligand to the apo-form of WBSCR27 followed by the dialysis against the buffer containing 50 mM NaCl, 50 mM sodium phosphate (pH 7.0), 10 mM DTT and 0.02% NaN_3.

Synthesis of [Methyl ^{13}C]-SAM and Preparation of ^{13}C,^{15}N Uniformly Labelled SAH

[Methyl ^{13}C]-labelled SAM was synthesized from SAH (Sigma) and $^{13}CH_3I$ (Cambridge Isotope Laboratories) according to the described method (Huber et al., 2016). 15 mg of SAH were dissolved in 500 µl of deuterated formic acid. 300 µl of $^{13}CH_3I$ were added to the resulting solution. The mixture was vortexed for 2 h and then stirred at room temperature in the dark. The completeness of the reaction was monitored by ^{1}H NMR. After 5 days, 1 ml of water was added and the unreacted ^{13}C methyl iodide was extracted with 2 ml of diethyl ether twice. The pH of the aqueous phase was adjusted to 7.15, after which the sample was applied to an ion exchange chromatographic column (732-0003 BioRad cartridge) preliminarily equilibrated with 0.01 M sodium phosphate buffer solution (pH 7.15). The column was washed with 55 ml of the same buffer solution (at first, the uncharged unreacted SAH and subproduct MTA were washed off, then the positively charged SAM was washed off). Next, the column was washed first with 20 ml of 0.1 M acetic acid, then with 20 ml of 4 M acetic acid. The main part of the product came off the column in the interval between 13 and 26 ml.

The SAM-containing fractions were evaporated in SpeedVac and dried in a freeze-dryer. A mixture of (S,S)- and (R,S)-SAM diastereomers was obtained, which is in agreement with the earlier described results (Bennett et al., 2017). The purity of the obtained product was controlled by NMR spectroscopy, the concentration was determined by UV spectrophotometry (ε_{260} = 16,000 M^{-1} cm^{-1}). The overall conversion rate can be estimated at 50% of the initial SAH. The yield of pure substance after purification was 10%.

^{13}C, ^{15}N uniformly labelled SAH was obtained from *E. coli* cell line overexpressing the WBSCR27 protein and grown on ^{13}C,^{15}N M9 medium. The method was based on the propensity of WBSCR27 to be co-purified with the cofactor and isolated in the form of a complex with SAH. The details of this protocol will be published elsewhere. The quality of the obtained ^{13}C,^{15}N-SAH was confirmed by 1D and 2D NMR spectra (**Supplementary Figure S1**). The sample contained DTT as an impurity, which does not interfere with the following procedure and was not removed from the product. The concentration of the obtained SAH was measured using UV absorbance at 260 nm (ε_{260} = 16,000 M^{-1} cm^{-1}). In total, we obtained 3.2 µmol of ^{13}C,^{15}N-SAH from 2 L of ^{13}C,^{15}N M9 medium.

Cell Lines

Mouse embryonic fibroblast cells NIH3T3 were cultured in DMEM/F12 medium (Gibco), supplemented with 10% FBS (Gibco), 1% Penicillin/Streptomycin (Gibco), and 1% Glutamax (Gibco) at 37°C, 5% CO_2 and used for all genetic manipulations.

The cell lines created for this study are schematically shown in **Supplementary Figure S2**. The cell lines A and B were prepared for studying intercellular localization of WBSCR27. The cell line A ectopically expressed fusion of WBSCR27 with far-red fluorescent protein mKate2 on N-terminal (Shemiakina et al., 2012). Similarly, the cell line B ectopically expressed hemagglutinin epitope YPYDVPDYA known as an HA tag (Field et al., 1988). The cell line B together with an endogenous C-terminal WBSCR27-HA fusion (cell line C) were used in co-immunoprecipitation experiments aimed at finding the possible macromolecular partners of WBSCR27. The knockout line (D) containing point mutations in the 2nd exon was used to study the phenotypic consequences of WBSCR27 depletion. To confirm WBSCR27 depletion on the protein level an HA-tag was added to the C-terminus of WBSCR27 in the knockout line (E). Two cell lines F and G were created from the knockout line for proximity labelling in the BioID experiment (Roux et al., 2018). These cell lines ectopically expressed prokaryotic biotin ligase BirA mutant (R118G) from *E.coli* (designated as BirA*), fused to HA and WBSCR27 (HA-BirA*-WBSCR27) or only to HA as the control (HA-BirA*). The details of cloning and constructing these cell lines are provided in **Supplementary Data**.

WBSCR27 Localization in the Cell

The intercellular localization of WBSCR27 was verified using cells with ectopic expression of HA or mKate2 (Shemiakina et al., 2012) fusions with WBSCR27. mKate2-WBSCR27 and HA-WBSCR27 expressing cells were seeded on coverslips and kept in the incubator overnight. The next day, the cells were washed 3 times for 5 min with PBST (PBS + 0.1% Triton X-100) and fixed with freshly-prepared 4% paraformaldehyde (in PBS) for 10 min at room temperature. The coverslips were rinsed with PBST (3 times for 5 min), followed by permeabilization with 1% Triton X-100 (in PBS) for 15 min at room temperature and subsequent PBST wash (3 times for 5 min).

The permeabilized coverslips with mKate2-WBSCR27 cells were incubated with 100 mM DAPI in PBS for 5 min at room temperature, followed by washing with PBST (2 times for 7 min). The coverslips were mounted with Mowiol (Sigma) and dried overnight.

The permeabilized coverslips with HA-WBSCR27 cells were blocked with 3% BSA (in PBST) for 1 h at room temperature, followed by incubation with primary anti-HA antibodies (Sigma, 3F10) and then with secondary goat anti-rat Alexa555 conjugated antibodies (Thermo Fisher Scientific) overnight at 4°C in PBST. After washing with PBST (3 times for 5 min), the coverslips were subjected to DAPI staining and mounting as described for mKate2-WBSCR27. Imaging was done with the Nikon Ti-E fluorescence microscope.

Co-Immunoprecipitation

Cells were cultured in five 15 cm plates to 95% confluency in a DMEM-F12 medium supplemented with 10% FBS, Glutamax, penicillin and streptomycin, and doxycycline hyclate (Sigma) at a concentration of 1 µg/ml at 37°C, 5% CO_2. The cells were harvested by trypsin, washed twice with 5 ml of PBS and kept frozen at −80°C prior to the immunoprecipitation (IP) experiment.

The frozen cells were resuspended in a lysis buffer [100 mM Hepes-KOH pH 7.5, 150 mM NaCl, 0.05% Triton X-100, 1 mM DTT and complete protease inhibitor cocktail (Roche)]. After

centrifugation at 13,000 g for 30 min, the supernatant was transferred to a new tube. For IP, 100 µl of anti HA-beads (Sigma Aldrich) were added to the lysate obtained from 1 g of wet cell mass and incubated overnight at 4°C. After five washes with the lysis buffer, proteins were eluted with a PAGE loading buffer for 5 min at 95°C. The protein eluates were analyzed by PAGE followed by silver staining and Western-blotting. Several bands present exclusively in the samples corresponding to the HA-tagged WBSCR27 were analyzed using MALDI according to the standard protocol (Chugunova et al., 2019).

For cross-linking experiments, the cells were resuspended in 1% formaldehyde (in PBS) and mixed for 7 min at room temperature. The cells were pelleted (500 g, 3 min) and washed twice with 1.25 M glycine (in PBS) to quench the remaining formaldehyde. Next, the cells were lysed as described in the paragraph above. The control sample was subjected to the same protocol, but omitting the formaldehyde addition step.

BioID Pull-Down

The BioID experiment was performed in accordance with the published protocol (Roux et al., 2018). Three cell lines based on NIH3T3 ΔWBSCR27 were used in the experiment: expressing HA-BirA*-WBSCR27, HA-BirA*, and parental ΔWBSCR27.

Ten 15 cm dishes for each cell line were seeded. Biotin labelling was performed when the cells reached approximately 80% confluency, the medium was changed to a fresh complete medium containing 50 µM biotin and incubated for 16–18 h. In the next stage the medium was completely removed by aspiration, the cells were rinsed twice with 5 ml/dish of PBS, and treated by 600 µl of lysis buffer/dish. The cells were harvested by gentle scraping. After centrifugation we obtained 1–2 g of the cells. The affinity purification of biotinylated proteins was performed using Dynabeads M-280 Streptavidin (Thermo Fisher Scientific).

The eluates from streptavidin beads were treated by trypsin and subsequently analyzed by shotgun proteomics (a technique for identifying proteins in complex mixtures such as cell lysates using a combination of high performance liquid chromatography and tandem mass spectrometry). Mass spectroscopy analysis was performed in triplicates with a Q Exactive HF-X mass spectrometer (Q Exactive HF-X Hybrid Quadrupole-Orbitrap™ Mass spectrometer, Thermo Fisher Scientific, Rockwell, IL, United States). The experimental details were published earlier (Laptev et al., 2020). The obtained raw data were processed using SearchGui (Barsnes and Vaudel, 2018) and PeptideShaker (Vaudel et al., 2015) programs with built-in search engines X! Tandem, MS Amanda, OMSSA, and Comet. Protein sequences of the complete mouse proteome provided by Uniprot (August 2019) were used for protein identification. N-terminal acetylation as well as the oxidation of methionine residues were set as variable modifications for the peptide search. Up to two missed cleavages were allowed for trypsin digestion. The false discovery rates for peptide and protein identifications were set to 1%.

The Primer Extension Assay

The experiment was performed as described earlier (Lesnyak et al., 2006). Total RNA was purified from NIH3T3 cell lines (WT and ΔWBSCR27) using Trizol reagent (ThermoFisher). The reverse transcription was performed with Maxima Reverse Transcriptase (ThermoFisher) using ^{32}P-labelled oligonucleotide complementary to the 28 S rRNA fragment 4,537–4,551. The products of the reverse transcription were separated by electrophoresis in the 10% (w/v) denaturing polyacrylamide gel and visualized by phosphorimagery.

NMR Spectroscopy

NMR samples at a concentration of 0.2–0.6 mM for ^{13}C and/or ^{15}N-labelled WBSCR27 and its complex with SAH were prepared in 95% H_2O/5% D_2O, 50 mM NaCl, 50 mM sodium phosphate buffer (pH 7.0), 10 mM DTT, and 0.02% NaN_3. NMR spectra were recorded at 308 K on Bruker AVANCE 600, 700, 800, and 850 MHz spectrometers equipped with a triple resonance (^{1}H, ^{13}C, ^{15}N) room temperature probe (600 MHz), Prodigy probe (700 MHz), and CryoProbe (800 and 850 MHz), or a quadruple resonance (^{1}H, ^{13}C, ^{15}N, ^{31}P) CryoProbe (700 MHz). 1D NMR spectra were processed and analyzed using Mnova software (Mestrelab Research, Spain). 2D and 3D spectra were processed by NMRPipe (Delaglio et al., 1995) and analyzed using NMRFAM-Sparky (Lee et al., 2015).

NMR Structure Determination

Earlier we reported the backbone and side chain signal assignments for the complex SAH-WBSCR27 (BMRB-27417, Mariasina et al., 2018) and for the apo form of the protein (BMRB-27578, Mariasina et al., 2020) and deposited these data in BioMagResBank (https://bmrb.io). This information was used to determine NMR restraints. Backbone φ and ψ dihedral angle restraints were determined from the chemical shift values of the backbone atoms ^{13}Cα, ^{13}Cβ, ^{13}CO, ^{1}Hα, ^{1}HN, and ^{15}N using TALOS+ software (Shen et al., 2009). Two independent sets of residual dipolar coupling (RDC) constants were measured in the nematic phase of a colloidal suspension of filamentous Pf1 phages (Hansen et al., 1998) and in a dilute liquid crystalline medium, consisting of DMPC/DHPC bicelles (Ottiger and Bax, 1999). The RDC values were calculated as a difference of the ^{15}N-^{1}H splitting values measured in the IPAP-HSQC spectrum (Ottiger et al., 1998) acquired in anisotropic and isotropic conditions. Hydrogen bond restraints were assigned to the amide groups having slow H/D exchange rates and located near carbonyl groups, as identified in the initial set of structures. NOE distance restraints were determined from the ^{1}H,^{13}C HSQC-NOESY and ^{1}H,^{15}N HSQC-NOESY spectra measured with a 100 ms mixing time. The initial set of NOE restraints (mainly intra-residue and sequential correlations) was selected manually. The rest of the cross-peaks were assigned using the automatic iterative procedure of spectra assignment/structure calculation implemented in ARIA 2.3 software (Bardiaux et al., 2012). The assignments were further manually verified by multiple steps of structure refinement using the simulated annealing protocol of the CNS 1.21 software package (Brunger et al., 1998). Database values of conformational torsion angle pseudopotentials (Kuszewski

et al., 1997) were used during the final cycles of the structure refinement to improve the quality of protein backbone conformation. The structure quality and restraint violations have been analyzed using the CNS tools, Procheck-NMR (Laskowski et al., 1996) and an in-house written NMRest program (Ivanova et al., 2007). The final families of 20 NMR structures of the SAH-WBSCR27 complex and apo-WBSCR27 were selected from 200 calculated conformers in accordance with the lowest-energy criterion and the absence of the residues in the disallowed regions of the Ramachandran map. The restraints used in structure calculations and statistics for the obtained NMR structures are presented in **Supplementary Tables S1, S2**. Additional details of structure calculations are provided in **Supplementary Material**. Structure visualization and analysis were carried out using PyMOL (Schrodinger LLC) and Discovery Studio Visualizer (Dassault Systemes Biovia Corp.).

Relaxation Measurements and Data Analysis

R_1 and R_2 relaxation rates and ^{1}H-^{15}N heteronuclear NOEs for ^{15}N uniformly labelled WBSCR27-SAH complex and the apo-form of WBSCR27 were measured at 308 K on a Bruker AVANCE III HD 700 MHz spectrometer. The measured experimental values were analyzed with a model-free formalism using the program RelaxFit written in-house (Polshakov et al., 1999). All the details of the NMR relaxation data collection and analysis are provided in **Supplementary Material**.

H/D Exchange

Amide H/D exchange rates in both ^{15}N labelled apo-WBSCR27 and complex WBSCR27-SAH were measured using heteronuclear ^{15}N–^{1}H NMR spectroscopy (at 35°C and pH 7.0) on a Bruker AVANCE Neo 700 MHz spectrometer. The details of the H/D exchange rate measurements and the calculation of the protection factors of the amide H_N atoms are given in the **Supplementary Data**.

NMR Studies of WBSCR27-Ligand Interactions

NMR experiments were carried out to test the binding of amino acids, nucleosides, and short DNA fragments to WBSCR27. The WBSCR27-SAM complex was prepared using ^{15}N-WBSCR27 (0.3 mM) and SAM (1.2 mM) in 320 µl of 90% H_2O/10% D_2O.

To test possible binding of amino acids to WBSCR27, a mixture of seven amino acids (Thr, Ser, Arg, Tyr, Cys, Glu, and Lys) in equimolar ratio was prepared. This mixture was added to WBSCR27-SAM samples to obtain molar ratios WBSCR27:SAM:mixture of 1:4:5 and 1:4:10. To test possible interactions of WBSCR27 with the fragments of nucleic acids, we mixed each of nucleosides (guanosine, uridine, cytidine, thymidine) with WBSCR27-SAM separately to obtain molar ratios WBSCR27:SAM: nucleoside = 1:4:10. We additionally prepared a mixture of desoxyoligonucleotide AAACCTCGCATTACGAACGGC TCC with the WBSCR27-SAM sample with a ratio of WBSCR27:SAM:DNA = 1:4:1. The purpose of testing an interaction of this arbitrary oligonucleotide with WBSCR27 was to check the possibility of oligonucleotide chain binding by protein. For each sample the ^{15}N,^{1}H HSQC spectrum was measured at 308 K and 600 MHz.

Interaction of SAM Epimers With WBSCR27

NMR spectroscopy was used to determine the binding ability of SAM epimers towards WBSCR27. The [methyl ^{13}C]-SAM obtained *via* a chemical synthesis from SAH in a concentration of 0.1 mM was used for this purpose. Since the synthetic product is an equimolar mixture of (S,S)- and (R,S)-SAM diastereomers, the concentration of individual components was 0.05 mM. ^{13}C,^{1}H HSQC spectra were recorded at 308 K and 600 MHz ^{1}H frequency for four samples: 1) a free SAM [a mixture of 0.05 mM (S,S)-SAM and 0.05 mM (R,S)-SAM], 2) a free apo-WBSCR27 (protein concentration 0.05 mM), 3) a mixture of 0.05 mM (S,S)-SAM, 0.05 mM (R,S)-SAM, and 0.05 mM WBSCR27 (1:1:1 ratio), and 4) a mixture of 0.05 mM (S,S)-SAM, 0.05 mM (R,S)-SAM, and 0.1 mM WBSCR27 (1:1:2 ratio).

RESULTS

WBSCR27 Is Localized in Both Cytoplasm and Nucleus

To determine the intracellular localization of WBSCR27 we inserted HA-WBSCR27 and mKate2-WBSCR27 fusion protein genes under a doxycycline inducible promoter into the NIH3T3 cell line *via* Sleeping Beauty transposase (Mátés et al., 2009). Visualizing both fusion proteins by anti-HA epitope immunocytochemical staining of fixed and permeabilized cells for HA-WBSCR27, and fluorescent microscopy of cells expressing mKate2-WBSCR27 allowed us to reveal both the cytoplasmic and nucleus distribution of WBSCR27 (**Figure 1**).

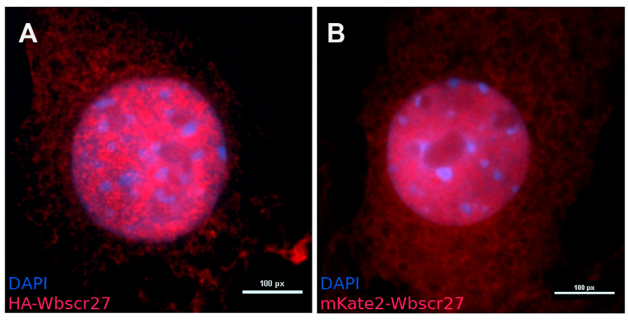

FIGURE 1 | The intracellular localization of WBSCR27 in NIH3T3 cells. **(A)** HA-WBSCR27; **(B)** mKate2-WBSCR27. The signal is present in both the cytoplasm and nucleus. The nucleus was visualized with DAPI.

WBSCR27 Apparently Does Not Establish Stable Interactions With Proteins and RNA

To outline the putative partners WBSCR27 protein interacts with, we applied a NIH3T3 cell line with ectopic expression of HA-WBSCR27. After inducing the fusion gene expression by doxycycline, HA-WBSCR27 was immunoprecipitated by an immobilized anti-HA antibody (**Supplementary Figure S3A**). While a protein of 27 kDa mass identified as HA-WBSCR27 by immunoblotting (**Supplementary Figure S3B**) was clearly present in the immunoprecipitate, no bands of its putative protein partners were observed. To exclude the possibility that the lack of identifiable partner proteins results from HA-WBSCR27 overexpression (which should lead to a significant decrease in the portion of the complex of WBSCR27 with its potential partner against the background of free protein), or that an N-terminally located HA tag prevents interaction with partner proteins, we used CRISPR/Cas9 directed cleavage and subsequent homologous recombination to create an NIH3T3 cell line with the natural *Wbscr27* gene C-terminally appended with an HA coding part. Immunoprecipitation of WBSCR27-HA from the extracts of the latter cell line (**Supplementary Figure S3C**) did not lead to the identification of potential WBSCR27 partner proteins.

The homolog of WBSCR27, rRNA MTase WBSCR22, and a number of other MTases form a complex with TRMT112. To specifically address the possibility that TRMT112 might coprecipitate with WBSCR27, we analyzed the eluate after HA-WBSCR27 immunoprecipitation by TRMT112 specific antibodies (data not shown) and found no evidence favoring the interaction between WBSCR27 and TRMT112.

The WBSCR27 MTase might establish only transient contacts with its possible substrates, being disengaged in the process of immunopurification. To this end, we immunopurified ectopically expressed HA-WBSCR27 following the formaldehyde treatment of the cells. Application of the formaldehyde cross-linking to isolate WBSCR27 partner proteins also did not help to identify the interacting proteins due to the absence or negligible yield of covalently crosslinked products (**Supplementary Figure S4**).

An alternative approach to address the short-lived protein-protein interactions is BioID, a proximity-dependent protein biotinylation *in vivo* by mutant promiscuously active biotin ligase BirA* followed by biotin-affinity capture (Roux et al., 2018). To apply BioID to search for WBSCR27 protein partners we created a cell line expressing HA-BirA*-WBSCR27 fusion and a control cell line expressing HA-BirA*. After labelling was induced by supplementing the cell cultures with biotin, the biotinylated proteins were purified from the cell extracts *via* streptavidine affinity capture and analyzed by shotgun proteomics. While both HA-BirA*-WBSCR27 and HA-BirA* fusion proteins were successfully expressed (**Supplementary Figure S5A**) and biotinylated endogenous proteins (**Supplementary Figure S5B**), panoramic proteome analysis of the biotinylated proteins (**Supplementary Table S3**) did not identify proteins modified specifically by the WBSCR27 protein fusion.

To identify potential RNA partners of WBSCR27 protein we applied PAR-CLIP protocol as described by Gopanenko and co-authors (Gopanenko et al., 2017). The cells of the NIH3T3 line expressing HA-WBSCR27 protein grown in the presence of 4-thiouridine were subsequently subjected to mild UV irradiation to induce RNA-protein cross-linking. After RNA fragmentation and HA-WBSCR27 immunopurification, cross-linked RNA was labelled with γ-[^{32}P]ATP and analyzed by gel electrophoresis and autoradiography (**Supplementary Figure S6**). Despite our efforts, the PAR-CLIP method did not allow identifying an RNA partner/substrate of the WBSCR27 enzyme.

WBSCR27 Gene Inactivation

To search for potential WBSCR27 substrates we used CRISPR/Cas9 guided WBSCR27 gene inactivation. Biallelic mutations disrupting the WBSCR27 reading frame were introduced to the second exon of the gene. Due to the absence of sufficient specificity of anti-WBSCR27 antibodies (data not shown), we verified the lack of WBSCR27 protein in the knockout cell line by biallelic extension of the *Wbscr27* reading frame with the HA coding region. Disrupting the reading frame of the *Wbscr27*-HA gene resulted in the disappearance of the band stained by anti-HA antibodies (**Supplementary Figure S7**).

Probing rRNA Methylating Activity of WBSCR27

For most methylated rRNA nucleotides, the enzymes responsible for their modification are known (Sergiev et al., 2018). The only methylated nucleotide of mammalian ribosomal RNA for which the enzyme responsible for the modification has not yet been identified is the m^{3}U4530 of the 28S rRNA (human rRNA numbering). This nucleotide is located in the peptidyl transferase center of the large ribosomal subunit (Sergiev et al., 2018). To validate whether this modification is due to the enzymatic activity of WBSCR27, we carried out a reverse transcription experiment similar to that used to identify a bacterial MTase modifying G2445 of the 23 S rRNA (Lesnyak et al., 2006). The method is based on m^{3}U inducing reverse transcription arrest. The experiment showed that the reverse transcription arrest is observed in rRNA from both the WT and WBSCR27 knockout cells (**Supplementary Figure S8**). These data clearly indicate that WBSCR27 is not responsible for modifying U4530 and, therefore, rRNA cannot be a methylation substrate for this enzyme.

WBSCR27 Does Not Recognize Fragments of Potential Substrates

The fragment-based lead discovery approach was used to probe the interactions of the possible WBSCR27 substrate fragments with protein. NMR techniques are usually able to detect highly specific interactions of small fragments of a larger ligand with a protein, even in the case of weak binding (Polshakov et al., 2019). We investigated the interactions of small compounds mimicking the fragments of

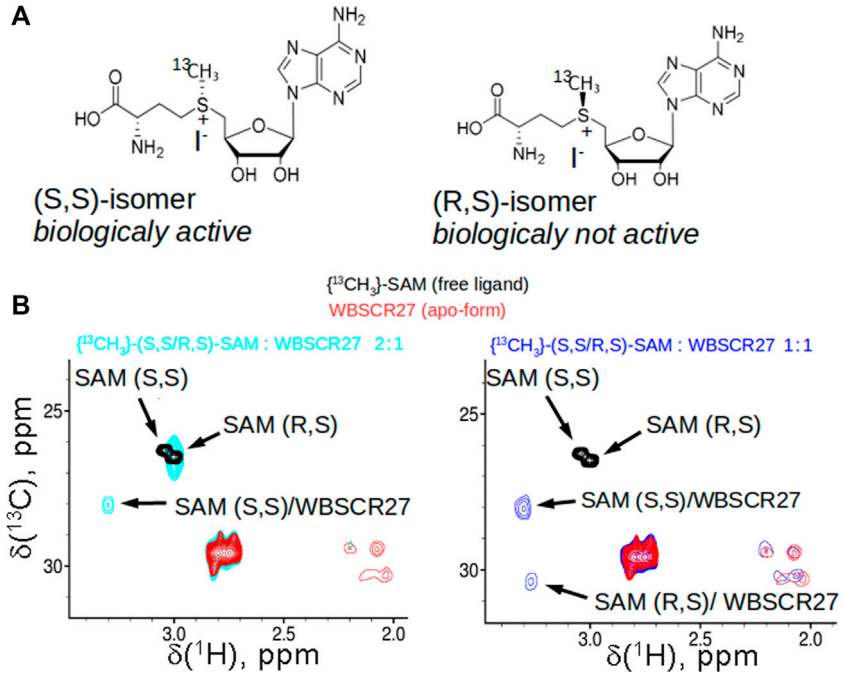

FIGURE 2 | SAM stereoisomers and their binding to WBSCR27. **(A)** Two stable epimers of SAM: biologically active (S,S)-SAM, synthesized enzymatically from ATP and methionine in cells, and (R,S)-SAM, which is obtained as a by-product synthetically. **(B)** Overlay of the fragments of ^{13}C,^{1}H HSQC spectra of racemic $^{13}CH_3$-labelled (R,S/S,S)-SAM (black), apo-WBSCR27 (red) and the mixture of SAM and protein in the ratio 2:1 (cyan, shown in the left panel) or 0.9:1 (blue, shown in the right panel).

macromolecules, namely: amino acids (Thr, Ser, Arg, Tyr, Cys, Glu, and Lys), nucleosides (guanosine, uridine, cytidine, timidine), and desoxyoligonucleotides with ^{15}N-labelled WBSCR27-SAM complex using the methods of heteronuclear NMR spectroscopy. In none of the studied fragments was specific binding observed, leading to a change of the ^{1}H and/or ^{15}N chemical shifts of the amide groups of certain WBSCR27 amino acids upon adding ligands to the protein.

WBSCR27 Preferably Binds the Biologically Active (S,S)-SAM Epimer

The sulfonium atom of SAM represents a chiral center, and both (R) and (S)-epimers are stable (**Figure 2A**). In living cells, the natural SAM (S,S-epimer) is biosynthesized from L-methionine and ATP by methionine adenosyltransferase (Zhang and Zheng, 2016). Chemically synthesized SAM contains both stereoisomers in equal amounts. We demonstrated by 2D NMR using synthesized [methyl ^{13}C]-labelled racemic (S,S/R,S)-SAM that WBSCR27 preferably binds the (S,S) stereoisomer. With an excess of WBSCR27, only the (S,S)-SAM isomer is bound, while the (R,S)-epimer remains in the free form (**Figure 2B**, left panel). With an increase in WBSCR27 content, the signals of the bound form (R,S)-SAM appear, but their intensity is significantly lower than that of the signals of the (S,S)-SAM-WBSCR27 complex (**Figure 2B**, right panel). This indicates a significantly lower affinity of the (R,S)-epimer compared with the (S,S)-epimer.

Solution Structure of Apo-Form of WBSCR27 and Its Complex With S-Adenosyl-L-Homocysteine

Earlier, we found that both SAM and SAH strongly bind to WBSCR27 (Mariasina et al., 2020). Relatively small changes in the chemical shifts of the signals of the residues in the binding site of these two ligands indicate that the structure of WBSCR27-SAM and WBSCR27-SAH complexes is similar (Mariasina et al., 2020). At the same time, bound SAM rapidly (several hours) decomposes to SAH; as a result, only the WBSCR27-SAH complex remains sufficiently stable to measure a series of heteronuclear NMR spectra. To determine the structures of the complex WBSCR27-SAH and the protein in apo-form traditional heteronuclear NMR and restrained molecular dynamics techniques were used. To assign the signals of bound SAH and protein-ligand NOEs, the NMR spectra for the complexes of ^{15}N-WBSCR27 with ^{13}C-labelled and unlabelled SAH were measured and compared. Several protein-ligand NOEs were also identified in NOESY spectra of the complex of ^{13}C,^{15}N-WBSCR27 with unlabelled SAH. In total, 21 protein-ligand NOEs were used in structure calculation (**Supplementary Table S1**).

The solution structure of the WBSCR27 apo-form (**Figure 3A**) shows that the first 51 amino acid residues of the protein as well as the loop between residues 204 and 228 (loop67 in **Figure 3C**) are unstructured. SAH binding to the protein puts in order the N-terminal protein fragment by forming three well-structured α-helices (α1–α3) and a short helical part in the first nine residues (**Figure 3B**). The structure of WBSCR27 has a canonical Rossman

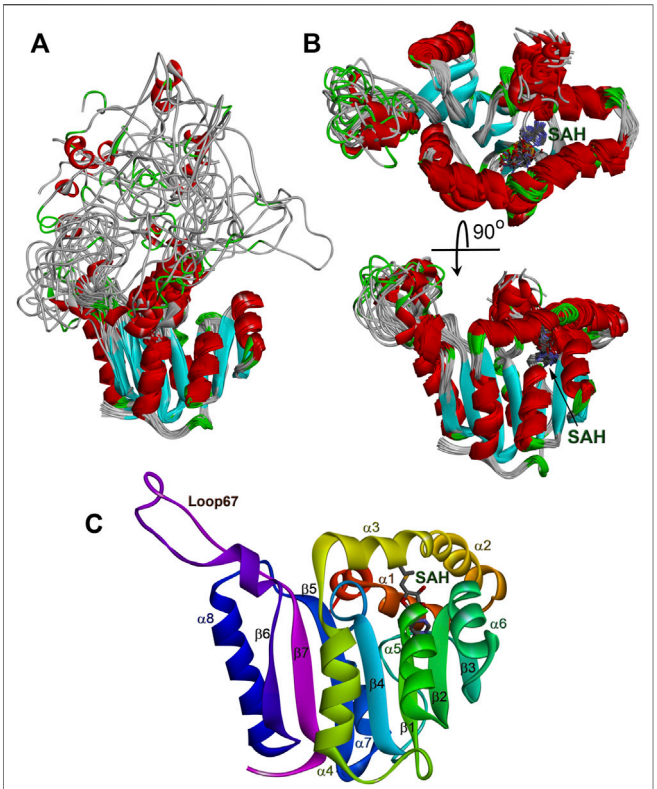

FIGURE 3 | Structure of WBSCR27 in apo-form **(A)** and in complex with SAH **(B,C)**. Families of 20 NMR conformers **(A,B)** and rainbow colored structure of WBSCR27-SAH complex with labelled elements of the protein structure are shown.

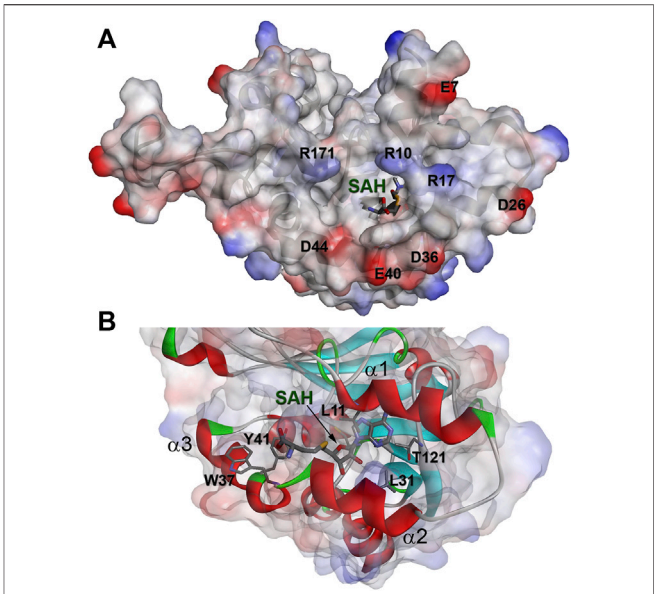

FIGURE 4 | The binding site of SAH in complex with WBSCR27. **(A)** Electrostatic surface potential of the protein. **(B)** Amino acids of WBSCR27 participating in the interaction with SAH.

fold, typical of most of the Class I MTases (**Figure 3C**). The protein core consists of seven β-strands (β1–β7) surrounded by five α-helices (α4–α8, **Figure 3C**). SAH binds to the residues on the tips of three β-strands (β1–β3) and strengthens these interactions by several hydrophobic and electrostatic contacts with the amino acid residues in the helices α1, α2, and α3 (**Figure 4**). The purine fragment of SAH binds predominantly to the amino acid residues of the protein β-core, while the methionine fragment interacts exclusively with the residues of helices α1–α3. The position of SAH methionine fragment is determined less precisely. The pairwise RMSD of the coordinates of the heavy atoms of the methionine fragment of SAH in the final family of structures is 2.3 ± 0.5 Å. For the adenosine fragment of SAH this value is 2.0 ± 0.6 Å. This may be due to higher mobility of the first three α-helices relative to the protein core.

WBSCR27 Backbone Dynamics

The protein backbone dynamics of WBSCR27 in both the apo-form and the complex with SAH was investigated by analyzing ^{15}N relaxation experiments and hydrogen-to-deuterium (H/D) exchange rates of the amide protons. The values of the overall rotational correlation time τ_c, calculated from the ^{15}N T_1 and T_2 data measured at 308 K (Kay et al., 1989), are 12.5 ± 0.2 and 10.4 ± 0.3 ns for the apo-form and the WBSCR27-SAH complex, respectively. The differences in the τ_c values for the apo-form and the complex appear to reflect the distinctions in the shape of the protein molecule. The protein globule is apparently more compact in the case of the WBSCR27-SAH complex, while the long unstructured N-terminal tail in the apo-form slows down protein tumbling. The experimentally measured ^{15}N relaxation parameters were interpreted using the model-free formalism (Lipari and Szabo, 1982) with extension to include chemical exchange contributions R_{ex} to the transverse relaxation rates (Clore et al., 1990) (**Supplementary Material** for details). **Figure 5** (for the apo-form of WBSCR27) and **Figure 6** (for WBSCR27-SAH) show the measured ^{15}N relaxation parameters R_1, R_2, and NOE with the calculated order parameters S^2 and R_{ex} values, plotted against the corresponding residue numbers. The mobility of the WBSCR27 backbone in the apo-form is significantly higher than that in the complex with SAH, which agrees well with the observed results of the structural studies.

Figure 7 shows the distribution of the protection factors PF for the amino acid residues of the apo-form of WBSCR27 and the WBSCR27-SAH complex, as analyzed from the measured proton-to-deuterium exchange rates. Binding of SAH leads to a significant increase in the PF of most amino acid residues of WBSCR27, which reflects the slowing down of high-amplitude protein backbone motions upon the ligand binding and the strengthening of the hydrogen bond network within the protein molecule.

DISCUSSION

Dissecting the WBSCR27 structure and delineating its functions may pave the way to understanding the molecular mechanisms underlying the clinical manifestations of WBS. In its turn, this

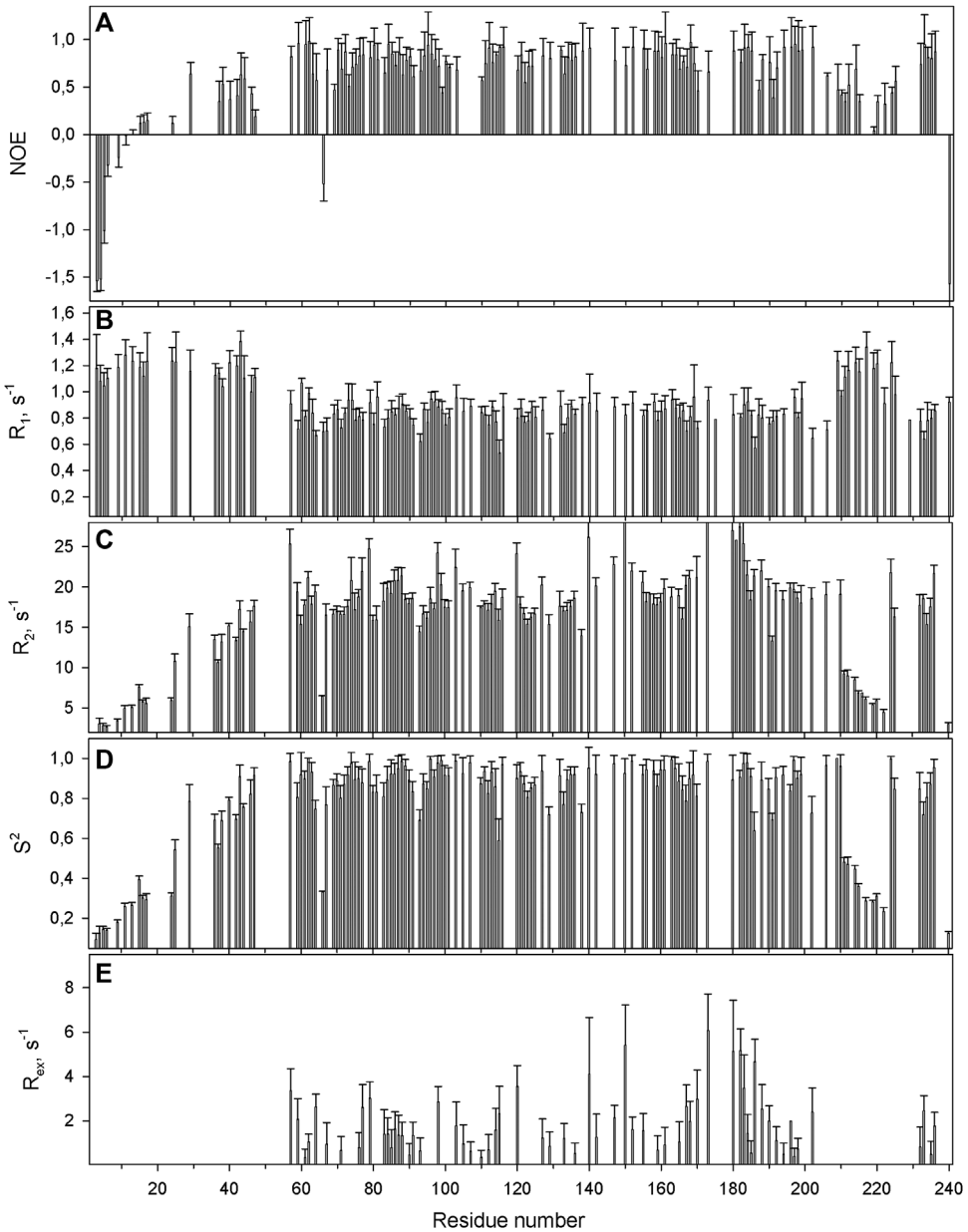

FIGURE 5 | The relaxation parameters of the amide ^{15}N nuclei of each residue of the apo-form of WBSCR27, measured at 16.3 T (700 MHz proton resonance frequency) and 308 K. **(A)** The heteronuclear ^{15}N,^{1}H-steady-state NOE values. **(B)** The longitudinal relaxation rate R_1 (s^{-1}). **(C)** The transverse relaxation rate R_2 (s^{-1}). **(D)** The order parameter S^2 determined by model-free analysis. **(E)** Chemical exchange R_{ex} contributions to the transverse relaxation rates (s^{-1}).

may contribute to developing clinical interventions aiming to compensate for the symptoms of this genetic disease.

Binding of SAH Causes Structuration of the N-Terminal Tail of WBSCR27 and General Tightening of Protein Structure in Solution

Comparing structures of the WBSCR27 apo-form and its complex with SAH indicates the formation of three additional α-helices at the N-terminal tail of the protein upon the cofactor binding (**Figure 3**). In the apo-form the first 50 residues forming these helices turn out to be disordered. These structural observations are clearly confirmed by the results of the protein backbone dynamics studies (**Figures 5–8**). For many residues of the apo-WBSCR27 the order parameters of amide NH bonds, determined *via* the ^{15}N relaxation measurements, are much lower than the corresponding values for the WBSCR27-SAH complex (**Figures 5**, **6**). These results indicate a high amplitude backbone motion of the apo-form of the protein in a time scale from ps to ns. There are also significant differences in the rates of backbone

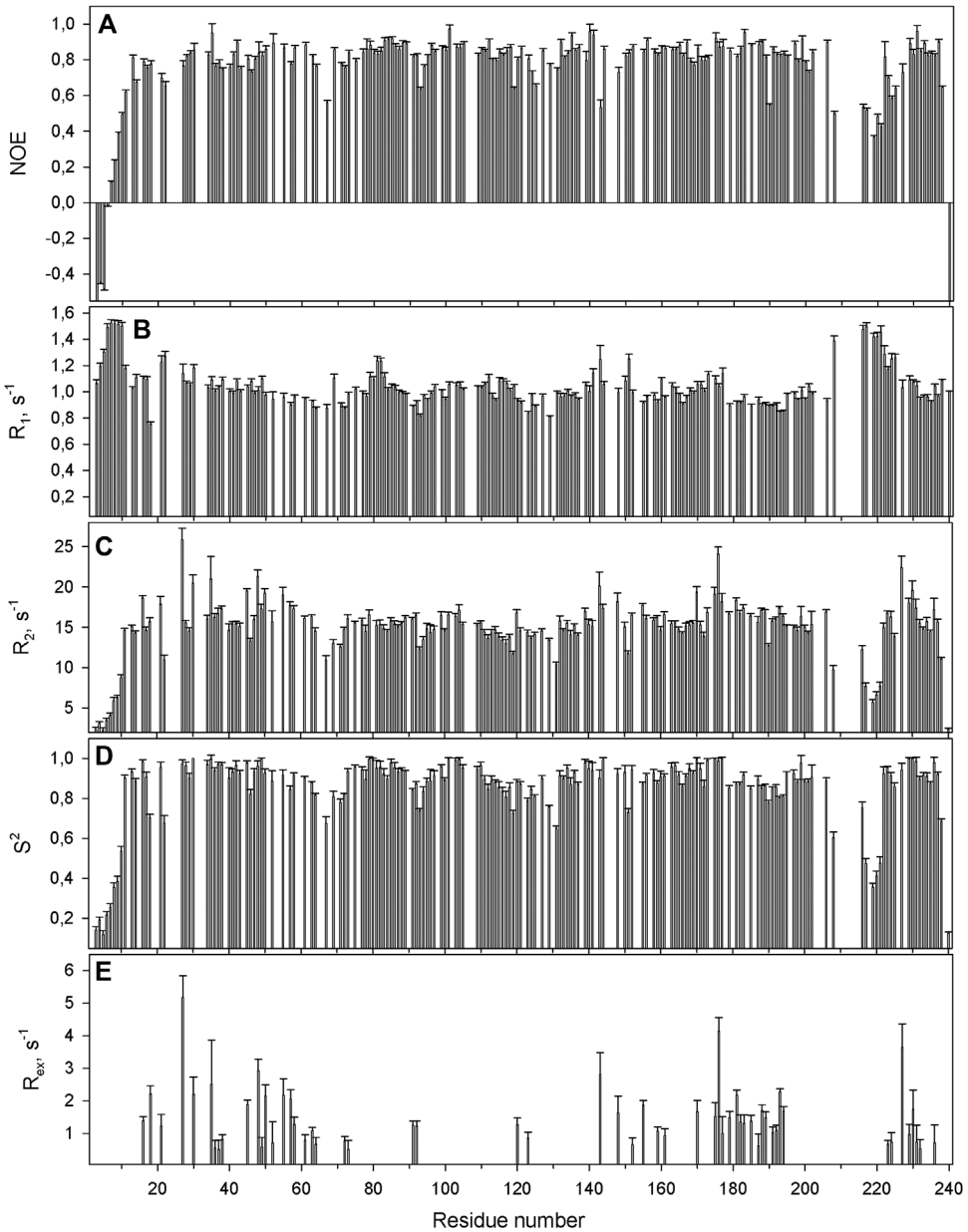

FIGURE 6 | The relaxation parameters of the amide ^{15}N nuclei of each residue of the SAH-WBSCR27 complex, measured at 16.3 T (700 MHz proton resonance frequency) and 308 K. **(A)** The heteronuclear ^{15}N,^{1}H-steady-state NOE values. **(B)** The longitudinal relaxation rate R_1 (s^{-1}). **(C)** The transverse relaxation rate R_2 (s^{-1}). **(D)** The order parameter S^2 determined by model-free analysis. **(E)** Chemical exchange R_{ex} contributions to the transverse relaxation rates (s^{-1}).

motions occurring in the ms time scale and characterized by the conformational exchange. The protein fragments which contain residues participating in conformational exchange are colored in orange in **Figure 8**.

The amplitude of the fast protein backbone motions, as well as the conformational transitions occurring in the ms timescale, are much greater for the case of the apo-form than for the WBSCR27-SAH complex. This difference in protein dynamics is observed not only for the first 50 amino acid residues unstructured in the apo-form, but also for the well-structured protein core. For example, the fragments of β4, β5, and α8 are highly mobile in the apo-form. Loop 67 remains highly mobile, both in the apo-form and in the complex. This loop is likely to be involved in recognizing the substrate molecule, and following its binding with helices α1–α3, loop 67 fixes this interaction, after which its mobility should disappear. Notably, the mobility of the G66 residue is high in both forms: the apo-form and the WBSCR27-SAH complex, the magnitude of the ^{15}N-^{1}H NOE of the amide group of this residue is negative. However, the mobility of this residue located at the apex of the loop between α1 and β1 is unlikely to play any functional role.

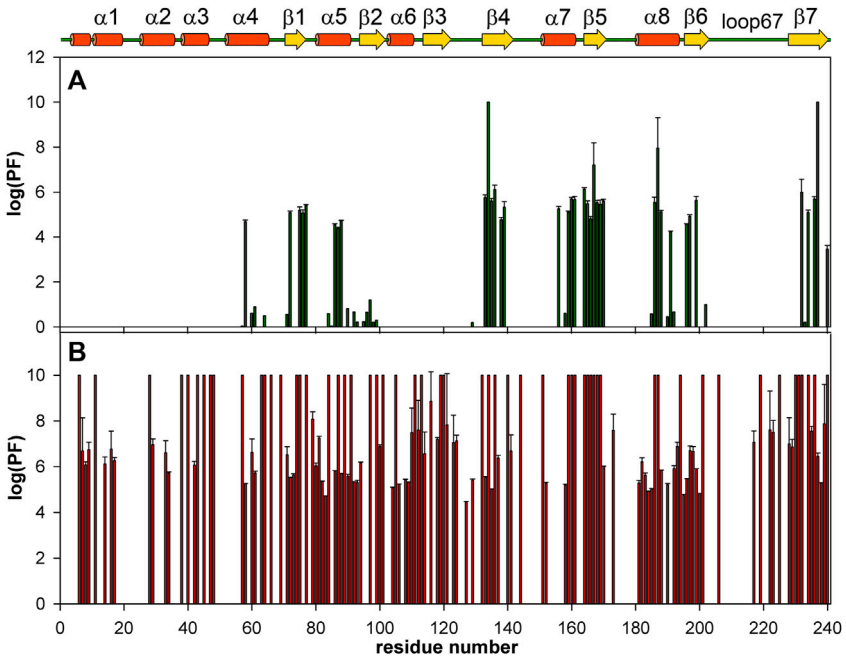

FIGURE 7 | Bar charts showing the amide HN protection factors of the apo-WBSCR27 **(A)** and complex WBSCR27-SAH **(B)**. The elements of protein secondary structure identified for WBSCR27-SAH solution structure are shown on the top.

The differences in the amplitude of the slow protein backbone motions, occurring on a time scale from minutes to hours and determined from analyzing the H/D exchange, are even more obvious when we compare the apo-form and the protein-ligand complex. Relatively high values of amide NH protection factors for the apo-form of WBSCR27 are observed only for the very central part of the protein core (**Figures 7**, **8**). Interestingly, all the amide groups of the residues from outer β-strand β3 and helices α4 and α6 are unprotected in the apo-form of the protein indicating a high mobility of these elements of the secondary structure. After SAH binding, almost the whole protein molecule, except for loop regions, turns out to be well protected from exchanging amide protons with water. This follows from large values of the protection factors of the corresponding NH groups (**Figure 8**).

Cofactor Binding Site

Helices α1–α3 in the WBSCR27-SAH complex surround the SAH molecule and partially form its binding site. These three helices also form a binding site for the potential substrate of the methylation reaction, catalyzed by WBSCR27. The adenosine fragment of SAH binds to the tips of the three strands β1, β2, and β3, while the methionine chain is positioned between the helices α2 and α3. The backbone carbonyl groups and amide hydrogens of the residues 100, 101, and 121 form several hydrogen bonds with the adenosine fragment. The side chains of the residues L11, L31, A77, and T121 form hydrophobic interactions with the adenine moiety of the SAH. In the region of the methionine fragment, there are also two aromatic residues, Y41 and W37. If the role of Y41

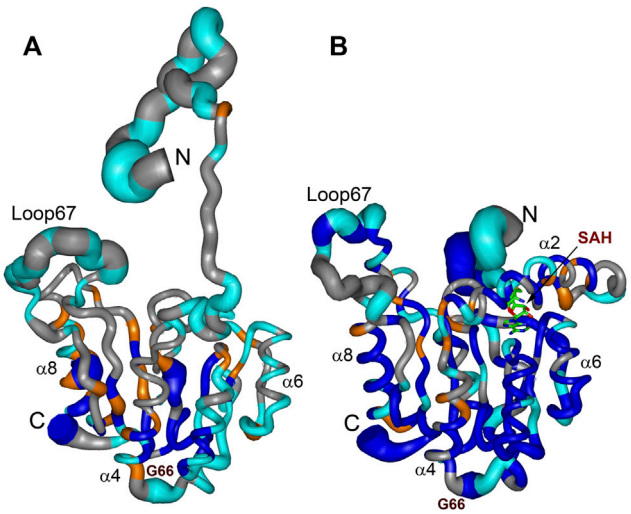

FIGURE 8 | Summary of dynamic and conformational behaviour of the WBSCR27 in apo-form **(A)** and in complex with SAH **(B)**. Representative NMR models of apo-WBSCR27 and complex SAH-WBSCR27 are shown according to protein mobility in a broad time scale. The thickness of the chain is proportional to the value ($1 - S^2$) representing the extent of the local amplitude of protein backbone motion in ps-ns time scale. Fragments of the protein backbone containing amino acids undergoing conformational exchange in ms time scale (with values R_{ex} exceeding 2 s^{-1}) are colored orange. Protein backbone fragments with residues having high values of protection factors determined from the analysis of hydrogen-to-deuterium HN exchange rates are colored blue. Residues for which ^{15}N relaxation data could not be obtained (proline residues and those with overlapped HN signals) are shown in gray. Representative secondary structure elements and SAH molecule are labelled.

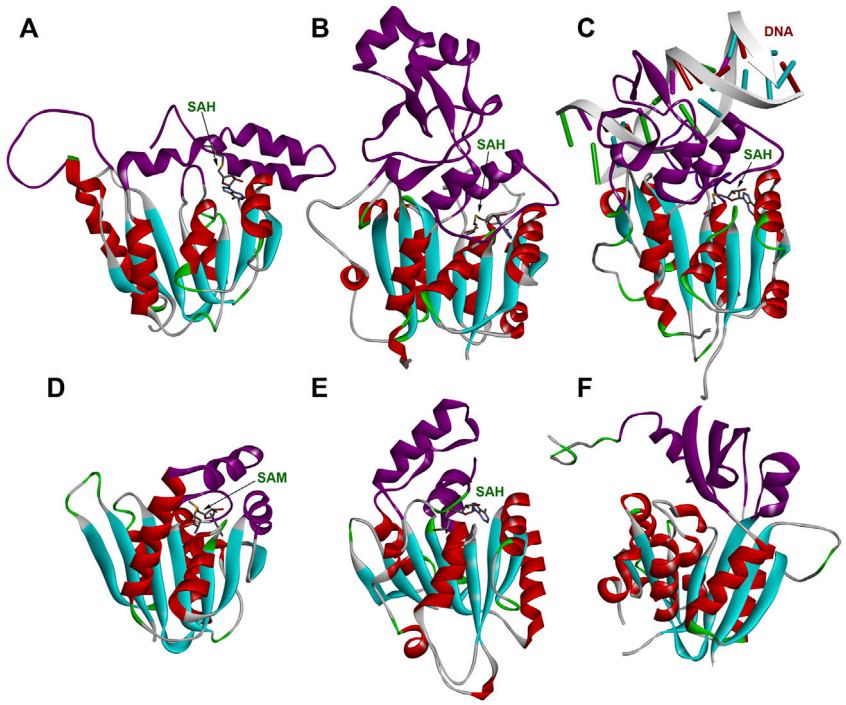

FIGURE 9 | Structures of several MTases. **(A)** WBSCR27 in complex with SAH (PDB id 7QCB); **(B)** *S. pombe* $tRNA^{Asp}$ MTase DNMT2 in complex with SAH (PDB id 6FDF); **(C)** human DNA MTase DNMT3A in complex with SAH and DNA fragment (PDB id 6BRR); **(D)** *S. cerevisiae* rRNA MTase Bud23 (PDB id 4QTU); **(E)** human $tRNA^{His}$ MTase BCDIN3D in complex with SAH (PDB id 6L8U); **(F)** human glycine MTase (PDB id 2AZT). Rossmann-fold of the enzymes' core is colored in cyan (β-strands) and red (α-helices). Substrate-binding domains are colored purple.

is most likely related to interacting with the carboxyl or amino group of the SAH methionine fragment, then W37 may participate in the interaction with a potential substrate fragment. It may, for instance, hold the aromatic base of RNA or DNA by stacking interaction in the position favorable for methylating this nucleotide or the neighboring one. This possibility is evidenced by the outward orientation of the side chain of W37 from the protein core, and its proximity to the sulfur atom of the SAH methionine residue.

Comparing WBSCR27 Structure With Other Class I MTases

The structure of WBSCR27 represents a classical Rossmann fold (**Figure 3C**), typical of all Class I SAM-dependent MTases. More than 120 various members of this enzyme family were classified (Martin and McMillan, 2002). They have different methylation substrates and very little sequence identity, but a highly conserved structural fold (**Figure 9**) and a β-sheet core, formed by seven β-strands (**Figure 9**). There are some variations in the number, length and orientations of α-helices surrounding this β-sheet, but they are still rather conservative in their structure. The greatest differences, as expected, are observed in the structure of the substrate-binding regions (colored purple in **Figure 9**), although the substrate molecule can also interact with the residues in the core region of the protein. The substrate-binding domain should ensure the selectivity of the substrate molecule binding and its correct positioning relative to the methyl group of the co-factor. However, the variability in the structure of the substrate-binding domain is great even for one type of substrate. **Figures 9B,D,E** show the structures of RNA MTases with significantly different substrate-binding fragments. Notably, there are two or three α-helices in the proximity to the SAM binding site in RNA MTases. A similar topology is observed in the case of DNA MTases (**Figure 9C**). The topology of the substrate-binding domain for small molecule methylation MTase (the structure of glycine MTase is shown in **Figure 9F** as an example, Luka et al., 2007) is markedly different. Based on these structural considerations, a nucleic acid would be the most likely substrate for WBSCR27, but other options cannot be ruled out.

One of the closest sequence homologs of WBSCR27 is a human protein WBSCR22 and its yeast ortholog Bud23 (Mariasina et al., 2018). WBSCR22 is a 18S rRNA MTase involved in pre-rRNA processing and ribosome 40 S subunit biogenesis (Haag et al., 2015). WBSCR22 has an interaction partner—the protein TRMT112 which is vital for the functional activity of this MTase in mammalian cells (Õunap et al., 2015). The known 3D structure of the complex of Bud23 with TRMT112 (Létoquart et al., 2014) allows comparing the TRMT112-binding interface on the surface of Bud23 with the similar area of the molecular surface of WBSCR27 (**Supplementary Figure S9**). The patterns of electrostatic potentials on the surface of Bud23 and WBSCR27 are quite different, and it is unlikely that TRMT112 can be the functional partner of WBSCR27. However, we tested this

hypothesis using co-immunoprecipitation and antibody staining for TRMT112. The experimental data obtained confirm the conclusion that these proteins do not interact with each other.

The protein folding topology and three-dimensional structure of WBSCR27 are similar to those of RNA MTase GidB (Romanowski et al., 2002). Notably, for GidB, as well as for WBSCR27, the enzyme function and methylation substrate were initially unknown. At the same time, establishing the three-dimensional structure of this MTase accelerated identifying the methylation substrate. This enzyme (alias RsmG) was recently shown to be responsible for N7 methylation in G527 of 16 S bacterial rRNA (Abedeera et al., 2020).

Possible Substrates of WBSCR27 and Its Potential Function

While our work has not yielded WBSCR27 substrates and partners, the negative result of this kind is also informative potentially narrowing down the range of possibilities for future research. First and foremost, WBSCR27 seems not to form any stable interaction with a substrate, unlike MTases functioning as the molecular switches in ribosome assembly, i.e., bacterial KsgA (Connolly et al., 2008) or mammalian METTL15 (Laptev et al., 2020). A number of MTases responsible for modifying the translation apparatus components and other substrates form a stable complex with the TRMT112 protein (Zorbas et al., 2015; Metzger et al., 2019; van Tran et al., 2019; Yang et al., 2021), while mRNA specific MTase METTL3 forms a stable complex with METTL14 (Liu et al., 2014; Wang et al., 2014) and WTAP (Ping et al., 2014). The results of this study disfavor the scenario that WBSCR27 forms a stable functional complex with other proteins. WBSCR27 is likely to be a standalone MTase only transiently interacting with its substrate. Moreover, WBSCR27 is unlikely to catalyze protein methylation, otherwise there must be a target protein that would be found in one of the experiments described above.

Our results do not exclude that WBSCR27 possibly participates in the methylation of a small molecule, whose chemical properties do not allow it to be identified in cell lysates by NMR methods. An overarching theory unpacking WBSCR27 functions is still ahead, and we hope that determining the three-dimensional structure of this enzyme in the apo-form and in the form of a complex with SAH will help achieve this goal.

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 in the article/ **Supplementary Material**.

AUTHOR CONTRIBUTIONS

SM, OD, PS, and VP designed the study. SM, C-FC, AC, TN, VZ, SE, and VI performed experiments and acquired data. SM and VP carried out structure calculations. SM, PS, and VP analyzed and interpreted data and drafted the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

The Russian Foundation for Basic Research (grant 20-04-00318) supported the NMR and structural studies. Gene editing experiments were supported by the Russian Science Foundation (grant 21-64-00006). Search for potential substrates was supported by the Russian Foundation for Basic Research (grant 20-04-00736). This work was done also with support of the Interdisciplinary Scientific and Educational School of Moscow University «Molecular Technologies of the Living Systems and Synthetic Biology». The 800 and 850 MHz NMR experiments were carried out using the equipment of the High-Field Nuclear Magnetic Resonance Center (HFNMRC) supported by Academia Sinica (AS-CFII-108-112), Taipei, Taiwan.

ACKNOWLEDGMENTS

The authors are grateful to Moscow State University (Russia) for the opportunity to use the NMR facilities and the supercomputer SKIF Lomonosov. The authors are grateful to Advanced Isotope Labelling for NMR School in Grenoble (France, 2019) for the opportunity to prepare specifically labelled WBSCR27 samples.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2022.865743/full#supplementary-material

REFERENCES

Abedeera, S. M., Hawkins, C. M., and Abeysirigunawardena, S. C. (2020). RsmG Forms Stable Complexes with Premature Small Subunit rRNA during Bacterial Ribosome Biogenesis. *RSC Adv.* 10 (38), 22361–22369. doi:10.1039/d0ra02732d

An, T., Zhang, J., Lv, B., Liu, Y., Huang, J., Lian, J., et al. (2019). Salvianolic Acid B Plays an Anti-obesity Role in High Fat Diet-Induced Obese Mice by Regulating the Expression of mRNA, circRNA, and lncRNA. *PeerJ.* 7, e6506. doi:10.7717/peerj.6506

Bardiaux, B., Malliavin, T., and Nilges, M. (2012). "ARIA for Solution and Solid-State NMR," in *Protein NMR Techniques*. Editors A. Shekhtman and D.S. Burz (Totowa, NJ: Humana Press), 453–483. doi:10.1007/978-1-61779-480-3_23

Barsnes, H., and Vaudel, M. (2018). SearchGUI: A Highly Adaptable Common Interface for Proteomics Search and De Novo Engines. *J. Proteome Res.* 17 (7), 2552–2555. doi:10.1021/acs.jproteome.8b00175

Bennett, M. R., Shepherd, S. A., Cronin, V. A., and Micklefield, J. (2017). Recent Advances in Methyltransferase Biocatalysis. *Curr. Opin. Chem. Biol.* 37, 97–106. doi:10.1016/j.cbpa.2017.01.020

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., et al. (1998). Crystallography & NMR System: A New Software Suite for Macromolecular Structure Determination. *Acta Cryst. D.* 54 (5), 905–921. doi:10.1107/S0907444998003254

Campeanu, I. J., Jiang, Y., Liu, L., Pilecki, M., Najor, A., Cobani, E., et al. (2021). Multi-omics Integration of Methyltransferase-like Protein Family Reveals Clinical Outcomes and Functional Signatures in Human Cancer. *Sci. Rep.* 11 (1), 14784. doi:10.1038/s41598-021-94019-5

Chugunova, A., Loseva, E., Mazin, P., Mitina, A., Navalayeu, T., Bilan, D., et al. (2019). *LINC00116* Codes for a Mitochondrial Peptide Linking Respiration and Lipid Metabolism. *Proc. Natl. Acad. Sci. U.S.A.* 116 (11), 4940–4945. doi:10.1073/pnas.1809105116

Clore, G. M., Driscoll, P. C., Wingfield, P. T., and Gronenborn, A. M. (1990). Analysis of the Backbone Dynamics of interleukin-1.Beta. Using Two-Dimensional Inverse Detected Heteronuclear Nitrogen-15-Proton NMR Spectroscopy. *Biochemistry* 29 (32), 7387–7401. doi:10.1021/bi00484a006

Connolly, K., Rife, J. P., and Culver, G. (2008). Mechanistic Insight into the Ribosome Biogenesis Functions of the Ancient Protein KsgA. *Mol. Microbiol.* 70 (5), 1062–1075. doi:10.1111/j.1365-2958.2008.06485.x

Delaglio, F., Grzesiek, S., Vuister, G., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: a Multidimensional Spectral Processing System Based on UNIX Pipes. *J. Biomol. NMR* 6 (3), 277–293. doi:10.1007/BF00197809

Ewart, A. K., Morris, C. A., Atkinson, D., Jin, W., Sternes, K., Spallone, P., et al. (1993). Hemizygosity at the Elastin Locus in a Developmental Disorder, Williams Syndrome. *Nat. Genet.* 5, 11–16. doi:10.1038/ng0993-11

Field, J., Nikawa, J., Broek, D., MacDonald, B., Rodgers, L., Wilson, I. A., et al. (1988). Purification of a RAS-Responsive Adenylyl Cyclase Complex from *Saccharomyces cerevisiae* by Use of an Epitope Addition Method. *Mol. Cell Biol.* 8 (5), 2159–2165. doi:10.1128/mcb.8.5.2159-2165.1988

Frangiskakis, J. M., Ewart, A. K., Morris, C. A., Mervis, C. B., Bertrand, J., Robinson, B. F., et al. (1996). LIM-kinase1 Hemizygosity Implicated in Impaired Visuospatial Constructive Cognition. *Cell* 86 (1), 59–69. doi:10.1016/S0092-8674(00)80077-X

Gopanenko, A. V., Malygin, A. A., Tupikin, A. E., Laktionov, P. P., Kabilov, M. R., and Karpova, G. G. (2017). Human Ribosomal Protein eS1 Is Engaged in Cellular Events Related to Processing and Functioning of U11 snRNA. *Nucleic Acids Res.* 45 (15), 9121–9137. doi:10.1093/nar/gkx559

Haag, S., Kretschmer, J., and Bohnsack, M. T. (2015). WBSCR22/Merm1 Is Required for Late Nuclear Pre-ribosomal RNA Processing and Mediates N7-Methylation of G1639 in Human 18S rRNA. *RNA* 21 (2), 180–187. doi:10.1261/rna.047910.114

Hahn, Y., and Lee, B. (2005). Identification of Nine Human-specific Frameshift Mutations by Comparative Analysis of the Human and the Chimpanzee Genome Sequences. *Bioinformatics* 21 (Suppl. l_1), i186–i194. doi:10.1093/bioinformatics/bti1000

Hansen, M. R., Mueller, L., and Pardi, A. (1998). Tunable Alignment of Macromolecules by Filamentous Phage Yields Dipolar Coupling Interactions. *Nat. Struct. Mol. Biol.* 5 (12), 1065–1074. doi:10.1038/4176

Huber, T. D., Johnson, B. R., Zhang, J., and Thorson, J. S. (2016). AdoMet Analog Synthesis and Utilization: Current State of the Art. *Curr. Opin. Biotechnol.* 42, 189–197. doi:10.1016/j.copbio.2016.07.005

Husmann, D., and Gozani, O. (2019). Histone Lysine Methyltransferases in Biology and Disease. *Nat. Struct. Mol. Biol.* 26 (10), 880–889. doi:10.1038/s41594-019-0298-7

Ivanova, E. V., Kolosov, P. M., Birdsall, B., Kelly, G., Pastore, A., Kisselev, L. L., et al. (2007). Eukaryotic Class 1 Translation Termination Factor eRF1 – the NMR Structure and Dynamics of the Middle Domain Involved in Triggering Ribosome-dependent Peptidyl-tRNA Hydrolysis. *FEBS J.* 274 (16), 4223–4237. doi:10.1111/j.1742-4658.2007.05949.x

Jones, W., Bellugi, U., Lai, Z., Chiles, M., Reilly, J., Lincoln, A., et al. (2000). II. Hypersociability in Williams Syndrome. *J. Cogn. Neurosci.* 12 (Suppl. 1), 30–46. doi:10.1162/089892900561968

Kay, L. E., Torchia, D. A., and Bax, A. (1989). Backbone Dynamics of Proteins as Studied by ^{15}N Inverse Detected Heteronuclear NMR Spectroscopy: Application to Staphylococcal Nuclease. *Biochemistry* 28 (23), 8972–8979. doi:10.1021/bi00449a003

Kerfah, R., Plevin, M. J., Sounier, R., Gans, P., and Boisbouvier, J. (2015). Methyl-specific Isotopic Labeling: a Molecular Tool Box for Solution NMR Studies of Large Proteins. *Curr. Opin. Struct. Biol.* 32, 113–122. doi:10.1016/j.sbi.2015.03.009

Kitagawa, H., Fujiki, R., Yoshimura, K., Mezaki, Y., Uematsu, Y., Matsui, D., et al. (2003). RETRACTED: The Chromatin-Remodeling Complex WINAC Targets a Nuclear Receptor to Promoters and Is Impaired in Williams Syndrome. *Cell* 113 (7), 905–917. doi:10.1016/S0092-8674(03)00436-7

Kuszewski, J., Gronenborn, A. M., and Clore, G. M. (1997). Improvements and Extensions in the Conformational Database Potential for the Refinement of NMR and X-Ray Structures of Proteins and Nucleic Acids. *J. Magnetic Reson.* 125 (1), 171–177. doi:10.1006/jmre.1997.1116

Laptev, I., Shvetsova, E., Levitskii, S., Serebryakova, M., Rubtsova, M., Zgoda, V., et al. (2020). METTL15 Interacts with the Assembly Intermediate of Murine Mitochondrial Small Ribosomal Subunit to Form m4C840 12S rRNA Residue. *Nucl. Acids Res.* 48 (14), 8022–8034. doi:10.1093/nar/gkaa522

Laskowski, R., Rullmann, J. A., MacArthur, M., Kaptein, R., and Thornton, J. (1996). AQUA and PROCHECK-NMR: Programs for Checking the Quality of Protein Structures Solved by NMR. *J. Biomol. NMR* 8 (4), 477–486. doi:10.1007/BF00228148

Lee, W., Tonelli, M., and Markley, J. L. (2015). NMRFAM-SPARKY: Enhanced Software for Biomolecular NMR Spectroscopy. *Bioinformatics* 31 (8), 1325–1327. doi:10.1093/bioinformatics/btu830

Lesnyak, D. V., Sergiev, P. V., Bogdanov, A. A., and Dontsova, O. A. (2006). Identification of *Escherichia coli* m2G Methyltransferases: I. The ycbY Gene Encodes a Methyltransferase Specific for G2445 of the 23 S rRNA. *J. Mol. Biol.* 364 (1), 20–25. doi:10.1016/j.jmb.2006.09.009

Létoquart, J., Huvelle, E., Wacheul, L., Bourgeois, G., Zorbas, C., Graille, M., et al. (2014). Structural and Functional Studies of Bud23-Trm112 Reveal 18S rRNA N 7 -G1575 Methylation Occurs on Late 40S Precursor Ribosomes. *Proc. Natl. Acad. Sci. U.S.A.* 111 (51), E5518–E5526. doi:10.1073/pnas.1413089111

Lipari, G., and Szabo, A. (1982). Model-Free Approach to the Interpretation of Nuclear Magnetic-Resonance Relaxation in Macromolecules .1. Theory and Range of Validity. *J. Am. Chem. Soc.* 104 (17), 4546–4559. doi:10.1021/ja00381a009

Liu, J., Yue, Y., Han, D., Wang, X., Fu, Y., Zhang, L., et al. (2014). A METTL3-METTL14 Complex Mediates Mammalian Nuclear RNA N6-Adenosine Methylation. *Nat. Chem. Biol.* 10 (2), 93–95. doi:10.1038/nchembio.1432

Luka, Z., Pakhomova, S., Luka, Y., Newcomer, M. E., and Wagner, C. (2007). Destabilization of Human glycine N-Methyltransferase by H176N Mutation. *Protein Sci.* 16 (9), 1957–1964. doi:10.1110/ps.072921507

Mariasina, S. S., Chang, C. F., Petrova, O. A., Efimov, S. V., Klochkov, V. V., Kechko, O. I., et al. (2020). Williams–Beuren Syndrome-Related Methyltransferase WBSCR27: Cofactor Binding and Cleavage. *Febs J.* 287 (24), 5375–5393. doi:10.1111/febs.15320

Mariasina, S. S., Petrova, O. A., Osterman, I. A., Sergeeva, O. V., Efimov, S. V., Klochkov, V. V., et al. (2018). NMR Assignments of the WBSCR27 Protein Related to Williams-Beuren Syndrome. *Biomol. NMR Assign.* 12 (2), 303–308. doi:10.1007/s12104-018-9827-2

Martin, J., and McMillan, F. M. (2002). SAM (Dependent) I AM: the S-adenosylmethionine-dependent Methyltransferase Fold. *Curr. Opin. Struct. Biol.* 12 (6), 783–793. doi:10.1016/S0959-440X(02)00391-3

Masserini, B., Bedeschi, M. F., Bianchi, V., Scuvera, G., Beck-Peccoz, P., Lalatta, F., et al. (2013). Prevalence of Diabetes and Pre-diabetes in a Cohort of Italian Young Adults with Williams Syndrome. *Am. J. Med. Genet.* 161 (4), 817–821. doi:10.1002/ajmg.a.35655

Mátés, L., Chuah, M. K. L., Belay, E., Jerchow, B., Manoj, N., Acosta-Sanchez, A., et al. (2009). Molecular Evolution of a Novel Hyperactive Sleeping Beauty Transposase Enables Robust Stable Gene Transfer in Vertebrates. *Nat. Genet.* 41 (6), 753–761. doi:10.1038/ng.343

Metzger, E., Wang, S., Urban, S., Willmann, D., Schmidt, A., Offermann, A., et al. (2019). KMT9 Monomethylates Histone H4 Lysine 12 and Controls Proliferation of Prostate Cancer Cells. *Nat. Struct. Mol. Biol.* 26 (5), 361–371. doi:10.1038/s41594-019-0219-9

Ottiger, M., and Bax, A. (1999). Bicelle-based Liquid Crystals for NMR-Measurement of Dipolar Couplings at Acidic and Basic pH Values. *J. Biomol. NMR* 13 (2), 187–191. doi:10.1023/a:1008395916985

Ottiger, M., Delaglio, F., and Bax, A. (1998). Measurement ofJand Dipolar Couplings from Simplified Two-Dimensional NMR Spectra. *J. Magnetic Reson.* 131 (2), 373–378. doi:10.1006/jmre.1998.1361

Õunap, K., Leetsi, L., Matsoo, M., and Kurg, R. (2015). The Stability of Ribosome Biogenesis Factor WBSCR22 Is Regulated by Interaction with TRMT112 via Ubiquitin-Proteasome Pathway. *PLOS One* 10 (7), e0133841. doi:10.1371/journal.pone.0133841

Ping, X.-L., Sun, B.-F., Wang, L., Xiao, W., Yang, X., Wang, W.-J., et al. (2014). Mammalian WTAP Is a Regulatory Subunit of the RNA N6-Methyladenosine Methyltransferase. *Cell Res.* 24 (2), 177–189. doi:10.1038/cr.2014.3

Pober, B. R. (2010). Williams-beuren Syndrome. *N. Engl. J. Med.* 362 (3), 239–252. doi:10.1056/NEJMra0903074

Polshakov, V. I., Batuev, E. A., and Mantsyzov, A. B. (2019). NMR Screening and Studies of Target - Ligand Interactions. *Russ. Chem. Rev.* 88 (1), 59–98. doi:10.1070/rcr4836

Polshakov, V. I., Birdsall, B., Frenkiel, T. A., Gargaro, A. R., and Feeney, J. (1999). Structure and Dynamics in Solution of the Complex of *Lactobacillus Casei* Dihydrofolate Reductase with the New Lipophilic Antifolate Drug Trimetrexate. *Protein Sci.* 8 (3), 467–481. doi:10.1110/ps.8.3.467

Romanowski, M. J., Bonanno, J. B., and Burley, S. K. (2002). Crystal Structure of the *Escherichia coli* Glucose-Inhibited Division Protein B (GidB) Reveals a Methyltransferase Fold. *Proteins* 47 (4), 563–567. doi:10.1002/prot.10121

Schubert, C. (2009). The Genomic Basis of the Williams–Beuren Syndrome. *Cell. Mol. Life Sci.* 66 (7), 1178–1197. doi:10.1007/s00018-008-8401-y

Roux, K. J., Kim, D. I., Burke, B., and May, D. G. (2018). BioID: A Screen for Protein-Protein Interactions. *Curr. Protoc. Protein Sci.* 91 (1), 19.23.1–19.23.15. doi:10.1002/cpps.51

Sergiev, P. V., Aleksashin, N. A., Chugunova, A. A., Polikanov, Y. S., and Dontsova, O. A. (2018). Structural and Evolutionary Insights into Ribosomal RNA Methylation. *Nat. Chem. Biol.* 14, 226–235. doi:10.1038/nchembio.2569

Shemiakina, I. I., Ermakova, G. V., Cranfill, P. J., Baird, M. A., Evans, R. A., Souslova, E. A., et al. (2012). A Monomeric Red Fluorescent Protein with Low Cytotoxicity. *Nat. Commun.* 3 (1), 1204. doi:10.1038/ncomms2208

Shen, Y., Delaglio, F., Cornilescu, G., and Bax, A. (2009). TALOS Plus : a Hybrid Method for Predicting Protein Backbone Torsion Angles from NMR Chemical Shifts. *J. Biomol. NMR* 44 (4), 213–223. doi:10.1007/s10858-009-9333-z

van Tran, N., Ernst, F. G. M., Hawley, B. R., Zorbas, C., Ulryck, N., Hackert, P., et al. (2019). The Human 18S rRNA m6A Methyltransferase METTL5 Is Stabilized by TRMT112. *Nucl. Acids Res.* 47 (15), 7719–7733. doi:10.1093/nar/gkz619

Vaudel, M., Burkhart, J. M., Zahedi, R. P., Oveland, E., Berven, F. S., Sickmann, A., et al. (2015). PeptideShaker Enables Reanalysis of MS-derived Proteomics Data Sets. *Nat. Biotechnol.* 33 (1), 22–24. doi:10.1038/nbt.3109

vonHoldt, B. M., Ji, S. S., Aardema, M. L., Stahler, D. R., Udell, M. A. R., and Sinsheimer, J. S. (2018). Activity of Genes with Functions in Human Williams-Beuren Syndrome Is Impacted by Mobile Element Insertions in the Gray Wolf Genome. *Genome Biol. Evol.* 10 (6), 1546–1553. doi:10.1093/gbe/evy112

vonHoldt, B. M., Shuldiner, E., Koch, I. J., Kartzinel, R. Y., Hogan, A., Brubaker, L., et al. (2017). Structural Variants in Genes Associated with Human Williams-Beuren Syndrome Underlie Stereotypical Hypersociability in Domestic Dogs. *Sci. Adv.* 3 (7), e1700398. doi:10.1126/sciadv.1700398

Wang, K., Zhang, J., Deng, M., Ju, Y., and Ouyang, M. (2022). METTL27 is a Prognostic Biomarker of Colon Cancer and Associated With Immune Invasion. *J. South. Med. Univ.* 42 (4), 486–497. doi:10.12122/j.issn.1673-4254.2022.04.04

Wang, Y., Li, Y., Toth, J. I., Petroski, M. D., Zhang, Z., and Zhao, J. C. (2014). N6-methyladenosine Modification Destabilizes Developmental Regulators in Embryonic Stem Cells. *Nat. Cell Biol.* 16 (2), 191–198. doi:10.1038/ncb2902

Yang, W.-Q., Xiong, Q.-P., Ge, J.-Y., Li, H., Zhu, W.-Y., Nie, Y., et al. (2021). THUMPD3-TRMT112 Is a m2G Methyltransferase Working on a Broad Range of tRNA Substrates. *Nucl. Acids Res.* 49 (20), 11900–11919. doi:10.1093/nar/gkab927

Zhang, J., and Zheng, Y. G. (2016). SAM/SAH Analogs as Versatile Tools for SAM-dependent Methyltransferases. *ACS Chem. Biol.* 11 (3), 583–597. doi:10.1021/acschembio.5b00812

Zorbas, C., Nicolas, E., Wacheul, L., Huvelle, E., Heurgué-Hamard, V., and Lafontaine, D. L. J. (2015). The Human 18S rRNA Base Methyltransferases DIMT1L and WBSCR22-Trmt112 but Not rRNA Modification Are Required for Ribosome Biogenesis. *MBoC* 26 (11), 2080–2095. doi:10.1091/mbc.E15-02-0073

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

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

frontiers | Frontiers in Molecular Biosciences

ORIGINAL RESEARCH
published: 28 June 2022
doi: 10.3389/fmolb.2022.910104

Wild-Type α-Synuclein and Variants Occur in Different Disordered Dimers and Pre-Fibrillar Conformations in Early Stage of Aggregation

*Adrien Guzzo [1], Patrice Delarue [1], Ana Rojas[2], Adrien Nicolaï [1], Gia G. Maisuradze [3] and Patrick Senet [1,3]**

[1]Laboratoire Interdisciplinaire Carnot de Bourgogne, UMR 6303 CNRS-Université de Bourgogne Franche-Comté, Dijon, France, [2]Schrödinger, Inc., New York, NY, United States, [3]Baker Laboratory of Chemistry and Chemical Biology, Cornell University, Ithaca, NY, United States

α-Synuclein is a 140 amino-acid intrinsically disordered protein mainly found in the brain. Toxic α-synuclein aggregates are the molecular hallmarks of Parkinson's disease. *In vitro* studies showed that α-synuclein aggregates in oligomeric structures of several 10th of monomers and into cylindrical structures (fibrils), comprising hundred to thousands of proteins, with polymorphic cross-β-sheet conformations. Oligomeric species, formed at the early stage of aggregation remain, however, poorly understood and are hypothesized to be the most toxic aggregates. Here, we studied the formation of wild-type (WT) and mutant (A30P, A53T, and E46K) dimers of α-synuclein using coarse-grained molecular dynamics. We identified two principal segments of the sequence with a higher propensity to aggregate in the early stage of dimerization: residues 36–55 and residues 66–95. The transient α-helices (residues 53–65 and 73–82) of α-synuclein monomers are destabilized by A53T and E46K mutations, which favors the formation of fibril native contacts in the N-terminal region, whereas the helix 53–65 prevents the propagation of fibril native contacts along the sequence for the WT in the early stages of dimerization. The present results indicate that dimers do not adopt the Greek key motif of the monomer fold in fibrils but form a majority of disordered aggregates and a minority (9–15%) of pre-fibrillar dimers both with intra-molecular and intermolecular β-sheets. The percentage of residues in parallel β-sheets is by increasing order monomer < disordered dimers < pre-fibrillar dimers. Native fibril contacts between the two monomers are present in the NAC domain for WT, A30P, and A53T and in the N-domain for A53T and E46K. Structural properties of pre-fibrillar dimers agree with rupture-force atomic force microscopy and single-molecule Förster resonance energy transfer available data. This suggests that the pre-fibrillar dimers might correspond to the smallest type B toxic oligomers. The probability density of the dimer gyration radius is multi-peaks with an average radius that is 10 Å larger than the one of the monomers for all proteins. The present results indicate that even the elementary α-synuclein aggregation step, the dimerization, is a complicated phenomenon that does not only involve the NAC region.

Keywords: α-synuclein, amyloid, Parkinson's disease, molecular dynamics, dimers, CUTABI

OPEN ACCESS

Edited by:
Piero Andrea Temussi, University of Naples Federico II, Italy

Reviewed by:
Leonid Breydo, West Pharmaceutical Services, United States
Carsten Kötting, Ruhr University Bochum, Germany

****Correspondence:***
Patrick Senet
psenet@u-bourgogne.fr

Specialty section:
This article was submitted to Structural Biology, a section of the journal Frontiers in Molecular Biosciences

Received: *31 March 2022*
Accepted: *16 May 2022*
Published: *28 June 2022*

Citation:
Guzzo A, Delarue P, Rojas A, Nicolaï A, Maisuradze GG and Senet P (2022) Wild-Type α-Synuclein and Variants Occur in Different Disordered Dimers and Pre-Fibrillar Conformations in Early Stage of Aggregation. Front. Mol. Biosci. 9:910104. doi: 10.3389/fmolb.2022.910104

1 INTRODUCTION

α-Synuclein (α-syn) is a soluble 140 amino-acid intrinsically disordered protein (IDP) (Wright and Dyson, 1999) abundant in the brain (Jakes et al., 1994; Mollenhauer et al., 2008), the function of which remains unclear. Abnormal aggregation of α-syn is central to the onset of diseases, and its inhibition is an intensive research area (Teil et al., 2020). In synucleopathies (Soto, 2003; Chiti and Dobson, 2006; Stefanis, 2012; Chiti and Dobson, 2017; Tanudjojo et al., 2021), such as the Parkinson disease (PD), α-syn is found in high concentration, as filamentous aggregates, in intraneuronal inclusions (Lewy bodies) and in extracellular deposits (Lewy neurites) (Spillantini et al., 1997; Breydo et al., 2012; Lashuel, 2020; Trinkaus et al., 2021). In addition to sporadic PD, rare familial cases of PD with similar phenotypes are induced either by an overexpression of wild-type (WT) α-syn due to α-syn gene triplication or by pathogenic mutations in the α-syn gene corresponding to single amino-acid substitution (Polymeropoulos et al., 1997; Krüger et al., 1998; Zarranz et al., 2004; Fuchs et al., 2008; Appel-Cresswell et al., 2013; Pasanen et al., 2014; Petrucci et al., 2016).

The origin of aggregation of α-syn is unknown. The formation of α-syn amyloid shares many common properties with the toxic aggregation of other IDPs, as, for example, the aggregation of Aβ in Alzheimer's disease (Cremades et al., 2017; Iadanza et al., 2018; Nguyen et al., 2021). When incubated in physiological conditions *in vitro*, α-syn aggregates spontaneously into fibrils with polymorphic cross-β-sheet conformations, in which a core of β-strands is aligned perpendicularly to the fibril axis, forming extended regular β-sheets with different arrangements (Tuttle et al., 2016; Guerrero-Ferreira et al., 2019; Guerrero-Ferreira et al., 2020). These different *in vitro* regular β-sheet structures are believed to mimic the α-syn aggregates in synucleopathies. Post-mortem X-ray diffraction of brains from patients with PD reveals patterns of β-sheet arrangements (Araki et al., 2019). The first *ex vivo* structures of α-syn fibrils from patients with multiple system atrophy have topology similar to the recombinant *in vitro* fibrils but with structural differences and occur with non-proteinaceous moieties (Schweighauser et al., 2020). In all known structures of fibrils, the monomer chain folds according to a three-layered L-shaped motif bearing resemblance to a Greek key motif (Guerrero-Ferreira et al., 2020). New toxic polymorphs were recently detected for which no structures are yet available (De Giorgi et al., 2020).

A recent study using solid-state nuclear magnetic resonance (NMR) of WT α-syn aggregation on phospholipid small unilamellar vesicles identified pre-fibrillar and early fibrillar species (Antonschmidt et al., 2021). NMR data suggest a pre-fibrillar segmental folding of the α-syn monomer in a three-layered L-shaped motif resembling one of the mature fibrils. The folding occurs with about the first half of the residues of the N-terminal region of the protein bounded to the phospholipid in an α-helical conformation (Antonschmidt et al., 2021). Another study of α-syn aggregation in solution identifies helix-rich intermediates in the transition of unstructured conformation α-syn into β-sheet-rich fibril formation during the growth of fibrils (elongation phase) (Ghosh et al., 2015). Interestingly, α-syn may also form ribbon aggregates (Bousset et al., 2013).

α-Syn fibrils are composed of thousands of α-syn monomers (Iadanza et al., 2018). Kinetic studies of α-syn aggregation showed that the onset of fibril formation is preceded by a long lag phase, indicating that the initial rate of aggregation is controlled by the nucleation of fibrils. It is hypothesized that pre-fibrillar aggregates of α-syn, α-syn protofibrils, i.e., oligomers composed of several dozen monomers occurring in the lag phase, might be the most toxic aggregates (Conway et al., 2000; Winner et al., 2011; Lázaro et al., 2014). These α-syn oligomers are transient intermediate species with heterogeneous structures that are consumed as formed fibrils (Cremades et al., 2017). Structural properties of α-syn oligomers were observed by different techniques, including single-molecule Förster resonance energy transfer (FRET), atomic force microscopy (AFM), small angle x-ray scattering (SAXS), small angle neutron scattering (SANS), transmission electron microscopy (TEM), and cryo-electron microscopy (cryo-EM) (Conway et al., 2000; Volles et al., 2001; Fredenburg et al., 2007; Giehm et al., 2011; Cremades et al., 2012; Lorenzen et al., 2014; Tosatto et al., 2015; Cremades et al., 2017; Bhak et al., 2018; Li et al., 2019; Zhou and Kurouski, 2020).

Single-molecule FRET provides information on the distance probability distribution between two fluorophores covalently linked to the α-synuclein molecule. The FRET study of pre-fibrillar oligomerization of α-syn (oligomers composed of 2–150 monomers) revealed two types of oligomers: type A with a low FRET efficiency (large internal distances) and type B with a high FRET efficiency (low internal distances) (Cremades et al., 2012; Horrocks et al., 2015; Tosatto et al., 2015). Type A occurred first in the kinetic of aggregation. Primary nucleation of α-syn may result in the creation of oligomers of type A from monomeric protein molecules that can grow through monomer addition, but they can also convert into type B oligomers (Cremades et al., 2012). The type B oligomers are toxic, more compact than the initially formed type A oligomers, and more proteinase-K-resistant than the type A and the monomer (Cremades et al., 2012). Unlike type B, type A oligomers dissociate in distilled water (Horrocks et al., 2015). Single-molecule FRET showed that missense mutations (A30P, A53T, and E46K) observed in familial forms of PD influence the kinetics of formation of oligomers and induce structural differences compared with the WT (Tosatto et al., 2015). The concentration of oligomeric species of variants (A30P, A53T, and E46K) in the lag phase of fibril formation is similar to the one of the WT, which indicates that the effect of the single amino acid substitutions on the oligomer structure might be more relevant for their neurotoxicity (Tosatto et al., 2015). Both type A and B oligomers are observed for mutants with a conversion rate from A to B higher for A53T and lower for A30P compared to the WT (Tosatto et al., 2015). Interestingly, for E46K mutants, it is not possible to separate clearly the two types of oligomeric species. Electron microscopy analysis of α-syn aggregation showed that the A30P mutant promotes the formation of pore-like protofibrils, whereas A53T promotes the formation of annular and tubular protofibrillar structures (Lashuel et al., 2002). The E46K variant generates annular

oligomers similar to those observed for the A53T and A30P mutants (Fredenburg et al., 2007).

Spherical metastable oligomers, with no significant secondary structures, with a diameter of 100 Å, were identified by SANS (Bhak et al., 2018). Oligomers composed of about 30 monomers forming a compact core with a flexible shell were characterized by SAXS (Lorenzen et al., 2014). These ellipsoidal oligomers, of the order of 100 Å size, inhibit both the primary nucleation and the subsequent elongation steps of α-syn fibrils (Lorenzen et al., 2014). The structure of two main subgroups of small oligomers, composed of 18 and 29 monomers on average, with a pore-like shape was identified by cryo-EM (Chen et al., 2015). The β-sheet structure in these oligomeric species is predominantly antiparallel (in opposite to the parallel β-sheet structure of fibrils) and amounts to 35% of the residues. Interestingly, spherical oligomers observed by AFM-IR are also mainly composed of the anti-parallel β-sheet (Zhou and Kurouski, 2020). AFM studies of α-syn pre-fibrillar formation on different solid surfaces reported the observation of heterogenous oligomers with spherical, elongated, and annular shapes (Conway et al., 2000; Zhou and Kurouski, 2020). Chains of spherical oligomers were also observed (Conway et al., 2000). Oligomers are formed in the primary nucleation phase of fibril formation but can also be generated upon fibril disaggregation (Bousset et al., 2013; Cremades et al., 2017).

The observation of annular oligomers for the WT and mutants suggests that α-syn may form a pore-like β-barrel structure in membranes that disrupt the membrane integrity (Volles et al., 2001; Fredenburg et al., 2007; Laganowsky et al., 2012). Monomeric α-syn and fibrils did not show membrane permeabilization activity (Volles et al., 2001). Exposure of hydrophobic side chains of oligomers to solvent may facilitate membrane permeabilization and be a source of cellular dysfunction (Cremades et al., 2012). Excessive production of reactive oxygen species, which may lead to cell apoptosis, is induced by α-syn oligomers (Cremades et al., 2012). The selective binding between transmembrane protein LAG3 (lymphocyte-activation gene 3) and α-syn preformed fibril species is also suggested as a mechanism of neurotoxicity in which LAG3 facilitates the endocytosis of toxic oligomeric species and the neuron-to-neuron transmission of pathological α-synuclein aggregates (Mao et al., 2016). In addition, the observation of α-syn helical tetramers (Bartels et al., 2011; Wang et al., 2011; Dettmer et al., 2015; Cote et al., 2018; Lucas and Fernández, 2020) stabilized by hydrophobic interactions and salt bridges between the monomers (Cote et al., 2018) suggests another mechanism of toxicity related to the equilibrium between the tetrameric form and other (β-sheet) oligomeric species (Bartels et al., 2011). Replica exchange molecular dynamics (MD) biased by NMR data showed that helix-rich and β-strand-rich trimers and tetramers are stable and may represent a minor population of α-syn in solution (Gurry et al., 2013).

Although oligomers composed of dozens of α-syn monomers are identified as toxic (type B oligomers), the role of dimers in the neurotoxicity of α-syn cannot be discarded. It has been hypothesized that the critical rate-limiting step in the primary nucleation is the oxidative formation and accumulation of a dityrosine cross-linked dimer (Krishnan et al., 2003). The peak of accumulation of dimers coincides with the rapid onset of fibrillation for the WT and mutant (A53T and A30P) proteins. Dimer formation is accelerated for the A30P and A53T variants (Krishnan et al., 2003). It is worth noting that dimers have a detectable membrane permeabilization activity (Fredenburg et al., 2007). As dimerization is the most initial step in self-assembling of monomers, dimers may play an important role in the different pathways of aggregation leading to type B oligomers. α-Syn dimers are systematically observed in the lag phase of α-syn aggregation and were characterized by FRET (Cremades et al., 2012; Horrocks et al., 2015; Tosatto et al., 2015), fluorescence (Krishnan et al., 2003; Fredenburg et al., 2007; Lv et al., 2015), AFM (Yu et al., 2008; Zhang et al., 2018), and circular dichroism (Roostaee et al., 2013).

Dimerization between fluorophore-free WT and mutant (A30P, A53T, and E46K) α-syn monomers on a substrate and fluorophore-labeled monomers in solution was analyzed by internal reflection fluorescence microscopy (Lv et al., 2015). Two types of dimers were identified which differ by an order of magnitude in their lifetimes. For the WT, E46K, and A53T, the less stable (type 1) and most stable (type 2) dimers have a lifetime of the order of about 200–300 ms and about 3s, respectively (Lv et al., 2015). The variant A30P is significantly different with lifetimes of the order of 700 ms (type 1) and 5 s (type 2), suggesting a more stable dimeric structure (Lv et al., 2015). Type 2 dimers were also observed in single-molecule AFM spectroscopy findings of WT α-syn (Yu et al., 2008). The lifetimes (types 1 and 2) of the mutants are systematically longer than the one of the WT, which indicates that the missense mutations seem to increase the stability of the dimer aggregate. Most dimeric structures are of type 1, but A30P and A53T single amino-acid substitutions increase significantly the type 2 population. The type 2 dimer might be related to the B-type toxic oligomer (Tosatto et al., 2015). The α-syn dimers have a low FRET efficiency but might be a mixture of types A and B, which cannot be resolved experimentally (Horrocks et al., 2015). The fraction of dimers formed in the lag phase of aggregation is by increasing order WT < A53T < A30P as measured by FRET. The A30P variant generates the largest population of B oligomers at the end of the lag phase of fibrillation (Tosatto et al., 2015). These data (Lv et al., 2015; Tosatto et al., 2015) taken together indicate a significant structural difference between A30P dimers compared to WT and A53T dimers.

Differences between WT and mutant dimers were also observed by single-molecule AFM force spectroscopy at low pH (promoting aggregation) (Krasnoslobodtsev et al., 2013). Mutiple segments' interactions between the monomers are favored in A53T and E46K compared to the WT as reflected by a larger number of multiple rupture force events (Krasnoslobodtsev et al., 2013), whereas A30P favored single segment dimerization compared to the WT (Krasnoslobodtsev et al., 2013). α-Syn monomer and dimer structures and dynamics were also measured at neutral pH by high-speed AFM for proteins adsorbed on a solid surface immersed in an aqueous solution (Zhang et al., 2018). The monomer adopts mainly a

compact and stable spherical structure, but one-tail and two-tail transient structures were also identified (Zhang et al., 2018). Conformational transitions between different conformations occurred on a second-to-minute time scale. Dimers formed by the association of two globular monomers (major population) or of one globular monomer and a one-tail monomer (minor population) were observed (Zhang et al., 2018). However, the α-syn structure and dynamics on a surface might be different from those of α-syn in solution, as the confinement of a polymer on a surface is known to modify the dynamics and stability of the polymer conformations. For example, an α-syn monomer adopts an α-helical structure on the negatively charged (phospholipid) membrane (Fusco et al., 2018).

No experimental technique has so far provided the atomistic description of dimer heterogeneity. Detailed structural information on the influence of the missense mutations on the fundamental dimerization step of α-synuclein is missing. The structural properties and the role of α-syn dimers in different parallel pathways of the formation of larger oligomers are still unclear. MD simulation is a complementary means to single-molecule experimental techniques that may give insights on the dimerization process. Previous simulations of α-synuclein dimers were limited to the ultra-simplified model (discrete MD (Zhang et al., 2018) or small fragments of α-synuclein (Yamauchi and Okumura, 2021), which does not take into account properly the dynamics of the polymer or biased the initial conditions of the simulations by docking of the monomer native structure adopted by α-synuclein on a phospholipid membrane (Sahu et al., 2015). Here, we are going a step further by using unbiased replica exchange MD simulations of two α-synuclein molecules in an implicit solvent by using a physics-based coarse-grained UNited-RESidue (UNRES) force field (Maisuradze et al., 2010; Liwo et al., 2019) on a time scale of 29.7 milliseconds (72 replicas of 412 μs each for each variant studied), which is three orders of magnitude larger than typical all-atom MD simulations (Khalili et al., 2005). The force field was calibrated to reproduce the structure and thermodynamics of small model proteins and applied with success to simulate protein folding (Maisuradze et al., 2010; Zhou et al., 2014; Sieradzan et al., 2021), large-scale conformational dynamics (Gołaś et al., 2012), Aβ-amyloids (Rojas et al., 2017), and the effect of Aβ-fibrils on the aggregation of tau protein (Rojas et al., 2018). In the present MD simulations, most of the α-syn molecules do not aggregate and remain thus in a monomeric conformation. These monomers for the WT, A30P, A53T, and E46K were recently described elsewhere (Guzzo et al., 2021). Here, we describe the structure and populations of the different dimeric states found in the MD simulations for the WT and the same mutants and compare those to data extracted from experiments.

2 MATERIALS AND METHODS

All structures of α-syn (WT and A30P, A53T, and E46K variants) were extracted from replica exchange MD trajectories generated with the coarse-grained UNRES force field (Maisuradze et al., 2010; Liwo et al., 2019). In the UNRES force field, a polypeptide chain is represented as a sequence of C^α atoms with united peptide groups located halfway of the virtual C^α–C^α bonds and united side chains (SC) attached to the C^α atoms. The SC–SC interaction potentials implicitly include the contribution from solvation (Liwo et al., 2001; Maisuradze et al., 2010). Descriptions of the UNRES force field and its parameterization are available in the reference (Liwo et al., 2019) and at http://www.unres.pl.

A total of 72 trajectories were computed for *each* protein: 32 trajectories at 300 K and 8 trajectories at each of the following temperatures: 310 K, 323 K, 337 K, 353 K, and 370 K, using replica exchange MD as described in Guzzo et al. (2021). Each trajectory was started with 2 fully unfolded monomers separated by a distance of 25 Å. The integration time step in UNRES is 4.9 fs corresponding to an effective actual time step of about 4.9 ps (Khalili et al., 2005). The convergence of each trajectory was monitored by computing the probability density of contacts of each residue, resulting in 30 million steps (effective time scale of 147 microseconds) of consolidated data out of 84 million steps (effective time scale of 412 microseconds) of simulation for *each* trajectory. The Cartesian coordinates of C^α and SC beads were saved for every 1,000 integration steps. Only structures at 300 and 310 K (40 trajectories) are reported here as they are close to the physiological temperature.

Since the simulations are performed on two monomers, both isolated non-interacting monomer conformations and aggregated monomers were observed in the converged MD trajectories. A dimeric conformation was defined as two monomers with more than 10 pairs of residues with at least one intermolecular distance between their C^α atoms smaller than 5 Å. This cut-off value was chosen because the average distance between two residues in an intermolecular β-sheet of an α-syn fibril is 4.8 Å (PDB ID: 2n0a). With this dimer definition, the fraction of dimers out of all the conformers simulated at 300 and 310 K is 31% for the WT, 23% for A30P, 29% for A53T, and 35% for E46K, respectively. The formation of dimers is increased for the E46K variant and is reduced for A30P (significantly) and A53T (weakly) mutants. The dimeric state of each protein is finally described here by about 300,000 structures extracted every 1,000 integration steps from the converged part of the replica exchange MD trajectories at 300 and 310 K, representing a sampling on an effective time scale of 1.47 milliseconds.

Analysis of the secondary structures of α-synuclein conformations was performed directly from the C^α coordinates with the CUTABI (CUrvature and Torsion based of Alpha-helix and Beta-sheet Identification) algorithm recently developed in our group (Guzzo et al., 2021). CUTABI is 10–30 times faster than the commonly used DSSP algorithm (Dictionary of Secondary Structure of Proteins) (Kabsch and Sander, 1983; Touw et al., 2015) because it avoids the construction of an all-atom description of the protein backbone from the coarse-grained UNRES structure, as required to apply DSSP. In CUTABI, the minimal size of a helix is set to 4 residues. Helices with less than 3 residues, such as short 3_{10} helices, are thus not counted. The minimal size of a β-strand is set to 2 residues, i.e., a β-sheet cannot be smaller than 4 residues. The β-strands of 1 residue forming β-bridges are thus not considered. A detailed description of the MD trajectories and of the CUTABI algorithm can be found in our previous work on monomeric state (Guzzo et al., 2021).

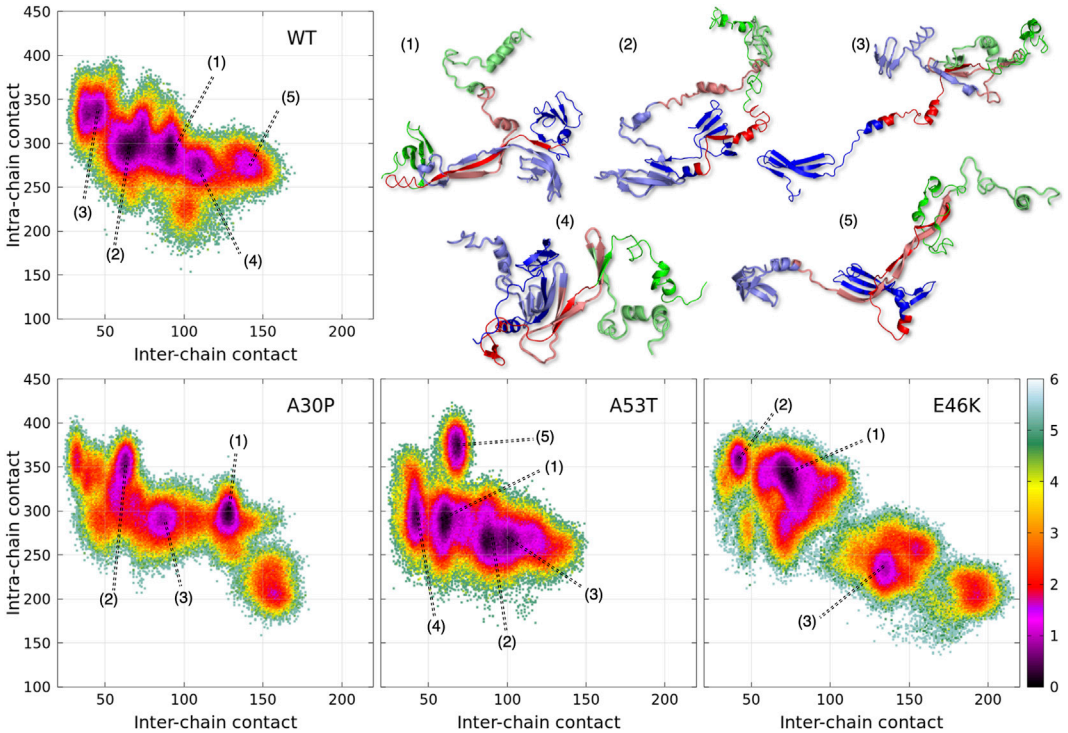

FIGURE 1 | Color maps of $-\ln[\frac{P(n_{inter},n_{intra})}{P_{max}}]$ computed from the 2D probability density function P of the number of inter-chain (n_{inter}) and intra-chain (n_{intra}) contacts between the residues of α-syn for the complete ensemble of dimers found in molecular dynamics trajectories at 300 and 310 K of the WT and variants. In each map, P_{max} is the maximum value of the probability of the map. The local minima within one unit from the global minimum of each map are numbered. Examples of 3D structures associated with each local minimum of the WT are represented with the N-terminus in blue, NAC in red, and the C-term in green. The light and dark colors differentiate the structure of the two monomers within the dimers.

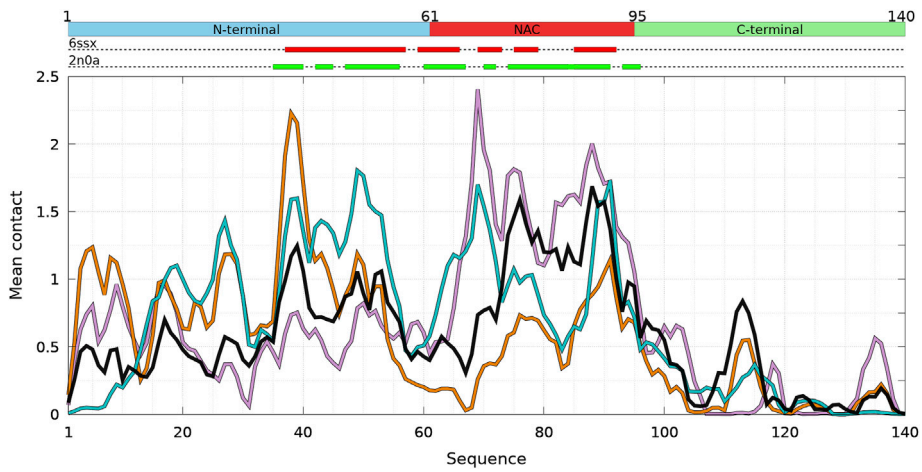

FIGURE 2 | Mean intermolecular contact of a residue of α-syn WT and mutants for the complete ensemble of dimers found in molecular dynamics trajectories at 300 and 310 K. The color code is black (WT), violet (A30P), orange (A53T), and turquoise (E46K). The upper bar represents the N-terminal (blue), NAC (red), and C-terminal (green) regions of the sequence. The red and green rectangles on the upper dotted bars describe the locations of the intermolecular β-sheets identified by CUTABI in the experimental structures of α-syn fibrils with the PDB IDs 6ssx and 2n0a, respectively.

The statistics of contacts between amino acids (**Figure 1**, **Figure 2**, **Figure 3**) were computed by defining a contact as a pair of 2 C^{α} atoms belonging to different residues at a distance smaller than 6 Å. By definition, the C^{α} atom of a residue may form several contacts. The mean contact of a residue (**Figure 2**) is computed as the average of all *intermolecular* contacts made

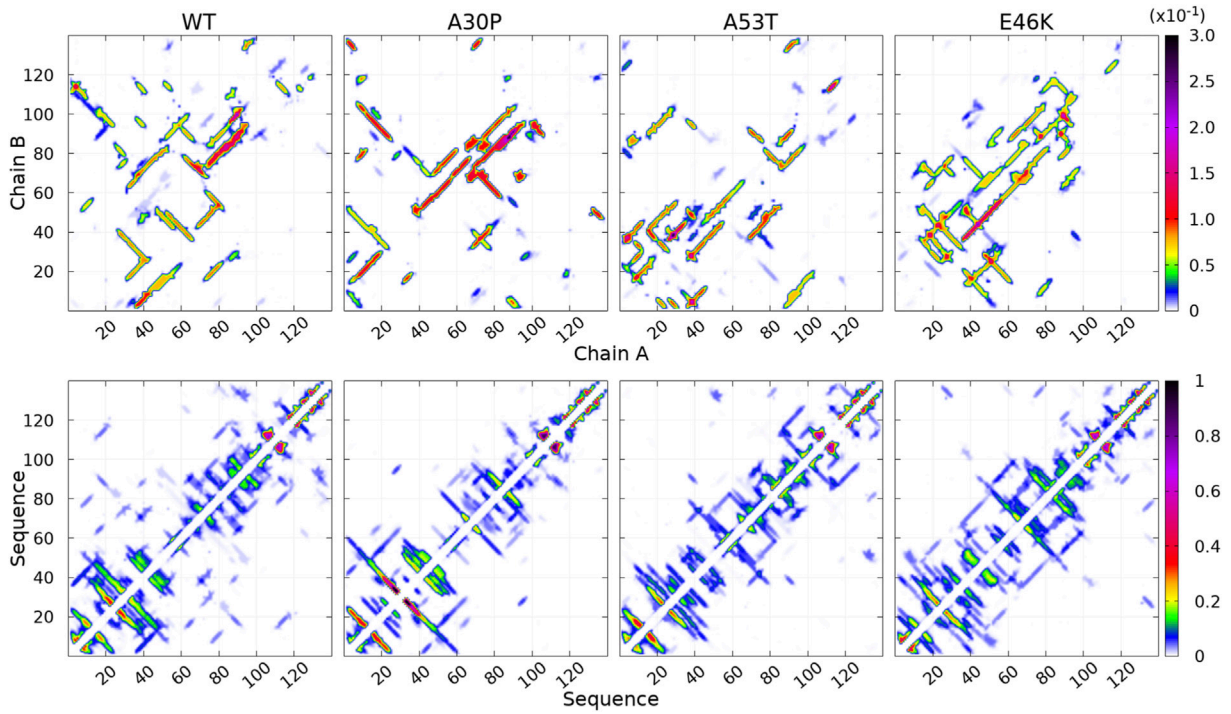

FIGURE 3 | Contact maps computed for the complete ensemble of α-syn dimers found in molecular dynamics simulations for the WT and the mutants. The upper panels show intermolecular contacts between chains A and B of the dimers, and the lower panels show intra-molecular contacts within monomers A + B. The color bars indicate the probability. Note the different scales for intra- and intermolecular contact maps.

by this residue in all snapshots of all trajectories at 300 and 310 K.

3 RESULTS

3.1 Propensity of α-Syn Dimerization

A dimensionless (effective) free-energy landscape of the α-syn dimers, $-\ln[\frac{P(n_{inter},n_{intra})}{P_{max}}]$, was computed from the two-dimensional probability density P of the number of inter-chain (n_{inter}) and intra-chain (n_{intra}) contacts between the residues for the WT and the mutants (**Figure 1**). $P_{\max}$ is the maximum value of the probability for each protein. For all proteins, n_{intra} varies between about 200 and 400. The minimum value of n_{inter} is 11 by definition of a dimer (see Material and Methods) and the maximum about 150 for the WT, A30P, and A53T and 200 for E46K. As for the monomer, the α-syn dimer has no unique native conformation but is represented by a large variety of dimer conformations. The structural diversity of the dimer conformations is shown in **Figure 1** where one structure was randomly selected from each subpopulation of the minima of the WT effective free energy.

The numbered local minima correspond to different subpopulations of dimers with similar free energies (**Table 1**). These minima are separated by low barriers for the WT and A30P (violet to red regions between the minima in **Figure 1**). A larger barrier exists between the minima 1 and 5 of A53T and the minimum 3 and the minima 1 and 2 for E46K. It is worth noting that only the WT has a map without large barriers, whereas the maps of A30P, A53T, and E46K show small islands separated from the rest by large barriers. The inter-conversions between different subpopulations of dimer structures seem easier in the WT compared to the variants.

TABLE 1 | Effective (dimensionless) free-energy difference ($-ln[\frac{P_i}{P_1}]$), where P_1 and P_i are the probabilities of the minimum 1 and the of *i*th minima shown in **Figure 1** for the WT and the variants.

Protein	Min 2	Min 3	Min 4	Min 5
WT	0.08	0.39	0.40	0.84
A30P	0.42	0.54	-	-
A53T	0.04	0.12	0.14	0.24
E46K	0.32	0.72	-	-

Figure 1 is muted on how the numerous intermolecular contacts are spread along the sequence. To answer this question, we first calculate the mean intermolecular contacts of residues along the sequence, as shown in **Figure 2**. For the WT, we observe seven representative peaks with maximum at A17, A27, Y39, and V49-A53 in the N-terminus; at A76 and I88-A90 in the NAC; and at L113 in the C-terminus. All these positions correspond to hydrophobic residues. In particular, the two largest peaks correspond to segments of three, A76-V77-A78, and four, I88-A89-A90-A91, hydrophobic residues, respectively. As mentioned previously, alanine plays a particular role in the

α-syn dynamics (Cote et al., 2018). It is worth noting that the mean contact variable counts both intermolecular random coil contacts and intermolecular β-sheet contacts. However, in *in vitro* fibrils, as in dimers computed from MD, the intermolecular contacts are mainly from residues in β-sheets (see the next subsection). Therefore, we compare the mean contact curves with the location of intermolecular β-sheets computed by CUTABI in two different experimental *in vitro* structures of WT α-syn fibrils (PDB IDs: 6ssx and 2n0a) in **Figure 2**. The peaks at Y39 and V49-A53 in MD are within a region of β-sheets in both experimental WT fibril structures and the peaks at A76 and I88-A90. A more precise comparison between intermolecular β-sheets found in MD and these structures will be examined in **Figures 6**, **8**.

As shown in **Figure 2**, it is remarkable to observe that a single mutation has huge effects on the relative weights of the different local regions having high aggregation propensity in the WT. The effects of a single amino-acid substitution are not limited to residues close to the mutation: the single amino-acid substitution has long-range effects on the mean contact. Compared to the WT, the main differences are as follows: 1) a larger propensity to aggregate in the N-terminal region for A53T and E46K and smaller for A30P, 2) a larger propensity to aggregate in the NAC region for A30P and lower for A53T, and 3) E46K has a propensity to aggregate larger than the WT all along the sequence excepted in region A76-I88 of the NAC. The largest peak of A30P is located at G69-V70-V71 corresponding to a succession of three hydrophobic residues. The highest propensity to aggregate is maximum at L38 for A53T and at V49 for E46K. The contribution of the C-term is modest and only significant close to L113 (except for A30P) and to Y136 (mainly for A30P).

How the different regions of the protein interact with each other is summarized by the intra-molecular and intermolecular contact maps in **Figure 3** for the WT and the mutants. In such a map, the colored lines parallel to the diagonal indicate contacts between residues in parallel segments, whereas the colored lines perpendicular to the diagonal show contacts between residues in anti-parallel segments. The parallel lines are the most probable for the intermolecular contacts for all proteins. In the opposite, the perpendicular lines are the most probable and numerous for the intra-molecular contacts for the WT and mutants.

For the WT, the most probable intermolecular contacts are on the diagonal of the map in region 75–93 (NAC), which means that this region in one chain interacts with the corresponding region in the other chain. Such contacts are similar to those in fibrils. Using (A) and (B) for chains A and B, we find more precisely contacts between residues 75–85 (A) and 74–84 (B) and between 83 and 93 (A) and 83–93 (B). A similar region in the NAC has contact along a parallel to the diagonal, between residues 75–91 (A) and 84–100 (B). Other long regions of contacts shifted by several residues occur between residues 36–51 (A) and 2–17 (B), 32–52 (A) and 63–83 (B), 52–63 (A) and 88–99 (B) (parallel segments), and 52 and 66 (A) and 37–50 (B) (anti-parallel segments). There are also a large number of small contacts between parallel segments in different regions. An important characteristic of the WT map is the absence of intermolecular contacts along the diagonal in the region 50–65, which is favorable to the helical structure in the monomer (Guzzo et al., 2021). As mentioned, the contact map of intra-molecular contacts is characterized by a large number of anti-parallel segments in contacts for the WT. Using the numbers in the horizontal axis of the maps, the most probables are 2–23, 12–39, 14–47, and 33/36–50/51 (N-terminus) and 67–77 and 87–101 (NAC). The WT map shows a lot of weakly probable contacts in distant regions.

For A30P, the most probable intermolecular contact is along the diagonal or parallel to the diagonal between residues 38 and 95 (end of the N-terminus and NAC). Contacts strictly similar to those in fibrils are found between residues 80 and 95. Contacts between parallel segments close to the diagonal (shifted by several residues) are between residues 69–79 (A) and 73–83 (B), 38–67 (A) and 50–78 (B), and 17–25 (A). Probable anti-parallel intermolecular contacts are observed mainly between residues 3–23 (A) and 32–51 (B), and 8–25 (A) and 87–105 (B). Compared to the WT, the number of regions interacting with each other is less numerous. The intra-molecular contacts are mainly between anti-parallel segments in the N-terminus, as in the WT, with an extended region of high probability for segments 8–53 and 77–93.

For A53T, the most probable intermolecular contacts are in the N-terminal region [residues 1–60 (A)] with many contacts between shifted segments as, for example, 15–29 (A) and 41–55 (B), 26–35 (A) and 36–45 (B), and 37–55 (A) and 27–45 (B). Contrary to A30P, the WT, and E46K (discussed next), there are only short regions of intermolecular contacts parallel to the diagonal, namely, in the NAC region (85–92) and in the C-terminus (111–116). The intra-molecular contacts in the N-terminus are similar to those of the WT, whereas contacts in the NAC are similar to those of A30P.

The E46K mutant is characterized by a large region of high probability of aligned intermolecular contacts (on the diagonal) from the N-terminus to the NAC (regions 36–81 and 87–90) and a large region of parallel intermolecular contacts between shifted segments: between regions 15–95 (A) and 35–109 (B). Intra-molecular contacts between anti-parallel segments are found in regions 7–20, 12–38, and 72–94.

3.2 Heterogeneity of Secondary Structure Elements in α-Syn Dimers

For each protein, the algorithm CUTABI (Guzzo et al., 2021) was applied to the ensemble of α-syn dimers to compute the sum of residues in the α-helix (α) and in the β-sheet (β). Each dimer conformation thus has (α, β) coordinates. The resulting probability densities in the (α, β) space are represented in **Figure 4**. In these maps, only residues from the N-terminal and NAC regions were considered for the calculations as the C-terminal region does not contribute to secondary structure differences between the WT and mutants (Guzzo et al., 2021).

For the isolated monomers, a major observation was that the conformations of the monomers were divided into two distinct states for the N-terminal + NAC region: an ensemble of conformations with no residue in the helix (state B) and the rest of conformations (state HB). The highest probability of observing a conformation in state B was an order of

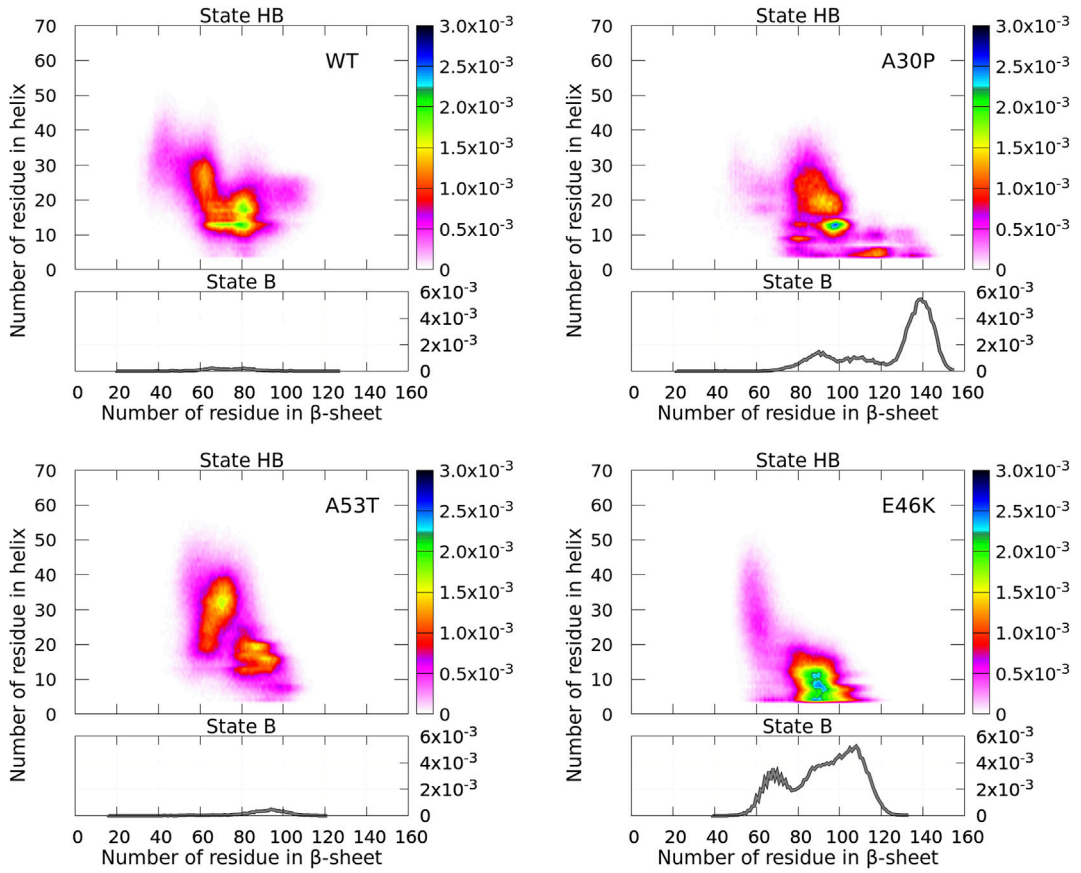

FIGURE 4 | Probability density of the number of residues in α-helices and β-sheets for the WT and mutants of α-syn. The probability density of state B (no helix) is represented by a function (gray) (right vertical axis), and the probability density of state HB is represented by a two-dimensional map (right color bar).

magnitude larger than that of state HB (Guzzo et al., 2021). In **Figure 4**, states B are clearly visible for the monomers forming A30P and E46K dimers with a probability only twice larger than state HB. On the opposite, states B disappeared in the WT and A53T dimers where the probability to find a dimer with no helix is completely negligible. For A30P, the maximum number of residues in β-sheets is 155 and the highest probable number is 139, i.e., half of the residues of the dimer. A sub-state B is found at about 90 for A30P. For E46K, the maximum number of residues in state B is also large (133) with a peak at 108 and a second peak at 66.

States B and HB can be distinguished from the function presented in **Figure 5** showing the fraction of conformations within an effective free-energy difference cut-off from the global minimum of state B for each protein (Guzzo et al., 2021). With $P_{\max}$, the maximum of probability at $(0, \beta)$ (in the B state), and P, the probability at (α, β) ($\alpha \geq 0$, in the B or HB states), the effective free-energy cut-off is computed as $-\ln(\frac{P}{P_{max}})$ in kT units, where k is the Boltzmann constant and T is the temperature. By definition, the *derivative of the curves* represented in **Figure 5** represents the Density Of conformations or micro-States (DOS). For A30P and E46K, a change in slope (DOS) is observed at about 1.2 kT, corresponding to about 10 and 45% of the conformations, respectively. The change in slope at 1.2 kT points up the separation between main states, i.e., the onset of state HB, i.e., a state with a mixture of α-helices and β-sheets. The case of E46K is special. We observe in fact two changes in slope at about 0.8 and 1.2 kT corresponding to the major peak at 110 residues and the minor peak at 70 residues in the one-dimensional probability density of state B in **Figure 4**. The HB state of E46K (**Figure 4**) is also very different from the HB states of the other proteins: they are rare dimers with helical regions larger than about 15 residues. Conformations with a maximum number of residues in helix are a few on the time scale of the present simulations with a maximum number of residues in the helix of 59, 50, 63, and 61 for the WT, A30P, A53T, and E46K, respectively. As for the monomers, A53T is the structure with the largest number of residues in helix (Guzzo et al., 2021). It is worth noting that the free-energy map of the HB state of A53T is also the most diffuse. The global minima of the HB maps are (13,80), (13,98), (31,70), and (8,91) for the WT, A30P, A53T, and E46K, respectively. The positions of these minima reflect also the largest propensities to form helical segments and β-sheets for A53T and E46K, respectively.

The probability to find the different secondary structures along the amino-acid sequence of the WT and mutants of

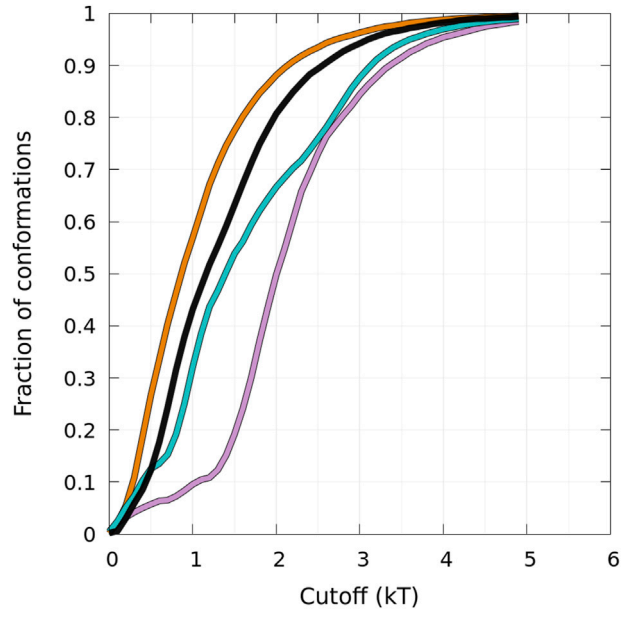

FIGURE 5 | Cumulative fraction of the total number of conformations simulated as a function of an effective free-energy cut-off from the global minimum of the B state (see text) represented in **Figure 4**. The curves are for the WT (black), A30P (purple), E46K (turquoise), and A53T (orange).

α-syn dimers was analyzed with CUTABI. As shown in **Figure 6**, helices are found in four main regions: two in the C-terminal region (residues 119–125 and 127–130), one in the NAC region (residues 75–82), and one overlapping the N-terminal and NAC regions (residues 53–65) for both the WT and mutants. There is no significant difference between the propensities of helices for the WT and mutants in the C-terminal region. These results for the helical propensity are similar to those found for the isolated monomers (Guzzo et al., 2021). The main differences are that the probability to form the helix 53–65 is twice larger in the dimer than in an isolated monomer for the WT and that the probability to form the 75–82 helix is about twice smaller in the dimer than in an isolated monomer for A30P and E46K.

The intra-molecular and intermolecular β-sheets were analyzed separately. The probability for a residue to pertain to an intra-molecular β-sheet has peaks at the same positions than for an isolated monomer. For the WT, the maximum of the peaks observed in **Figure 6** is as follows (in brackets for an isolated monomer (Guzzo et al., 2021)): $P_{F4} = 0.44$ (0.57), $P_{K10} = 0.53$ (0.61), $P_{A17} = 0.71$ (0.87), $P_{T22} = 0.74$ (0.76), $P_{E28} = 0.60$ (0.65), $P_{K34} = 0.64$ (0.61), $P_{38} = 0.65$ (0.69), and $P_{Y39} = 0.63$ (0.69); $P_{K43} = 0.36$ (0.50), $P_{V49} = 0.26$ (0.4), $P_{53} = 0.19$, and $P_{V55} = 0.13$ (0.69); $P_{V63} = 0.14$ (0.64), $P_{V70} = 0.15$ (0.32), $P_{81} = 0.18$ (0.35), and $P_{K80} = 0.17$ (0.41); $P_{88} = 0.43$ and $P_{A90} = 0.42$ (0.58); and $P_{V95} = 0.20$ (0.35) and $P_{97} = 0.27$. These peaks are at the same locations for the mutants as they were in isolated monomers. These maxima are located at or close to valine residues, which is the most frequently found amino acid in β-sheets (Chou and Fasman, 1974). The empirical prediction of the dimensionless propensity of the formation of the β-sheet computed from the scale of Chou and Fasman (CF) (Chou and Fasman, 1974) is shown in **Figure 6** for the WT. The CF simple algorithm, using only the primary sequence, predicts some peaks at similar locations than predicted by MD, namely, at 39, 52, 71, and 94. Of course, there are quantitative differences in the shape and relative heights predicted by CF compared to MD (they are not exactly the same quantities), but both methods identify qualitatively similar local regions of sequences which are important for β-sheets. It is worth noting that the CF algorithm predicts only a relatively small local change of the β-sheet propensity for the mutants, as the influence of a residue is limited to a few residues in the vicinity of its position.

Compared to the WT, the increase in probability of intra-molecular β-sheets in the region 26–35 for A30P and its decrease in the region 53–65 for A53T were also found in the isolated monomers (Guzzo et al., 2021). Compared to isolated monomers, the probability for a residue to pertain to an intra-molecular β-sheet decreases significantly from residues 43 to 100 for the WT, A30P, and A53T. Therefore, the intra-molecular β-sheets form mainly in the N-terminal region for these three proteins. For E46K, they are found equally in the N-terminal and NAC regions. Note that the CF algorithm underestimates the large propensity of formation of β-sheets in the N-terminal region.

The maxima of probability to form intermolecular β-sheets are located at the same positions than those of the probability to form intra-molecular β-sheets. The amplitude of these peaks is however different between the proteins. They are four major peaks for the WT and A30P: $P_{Y39,WT} = 0.33$, $P_{T75,WT} = 0.31$, $P_{T81,WT} = 0.29$, and $P_{I88,WT} = 0.36$, and $P_{V70,A30p} = 0.40$, $P_{T75,A30p} = 0.46$, $P_{I88,A30p} = 0.40$, and $P_{A90,A30p} = 0.52$, respectively. Except for the peak at Y39, the larger propensities to form intermolecular β-sheets are located in the NAC for the WT and A30P. For A53T, all the major peaks are in the N-terminal region: $P_{F4,A53T} = 0.36$, $P_{E28,A53T} = 0.37$, $P_{L38,A53T} = 0.60$, and $P_{K43,WT} = 0.34$. For the variant E46K, the propensity to form intermolecular β-sheets is mainly in the N-terminal region with peaks at $P_{A27,E46K} = 0.35$, $P_{Y39,E46K} = 0.45$, $P_{T44,E46K} = 0.37$, $P_{V49,E46K} = 0.40$, and $P_{V53,E46K} = 0.35$ but also in the NAC region with peaks at $P_{V71,E46K} = 0.31$ and $P_{A90,E46K} = 0.40$. A single amino-acid substitution modifies significantly the most probable regions of the formation of intermolecular β-sheets: highest in the NAC for the WT and A30P, in the N-terminal region for A53T, and in both the N-terminal region and in the NAC for E46K. These results compare well with the mean intermolecular contact profiles (**Figure 2**). As shown in **Figure 6** for the WT, the mean contact curve follows quite well the probability of intermolecular contact (except for the 80–85 region). Finally, the probability to form intermolecular β-sheets between the same residues in both monomers (diagonal in the contact maps in **Figure 3**), named native fibril-like contacts (Nfcs), is also shown for comparison in **Figure 6**. For all proteins, the probability of Nfcs is low. The sub-ensemble of dimers showing these contacts is analyzed and discussed next.

3.3 Dimers With Fibril Native Contacts (Dfncs)

Contact maps of the WT and mutants show contacts along their diagonal as in the experimental structures of fibrils (PDB IDs:

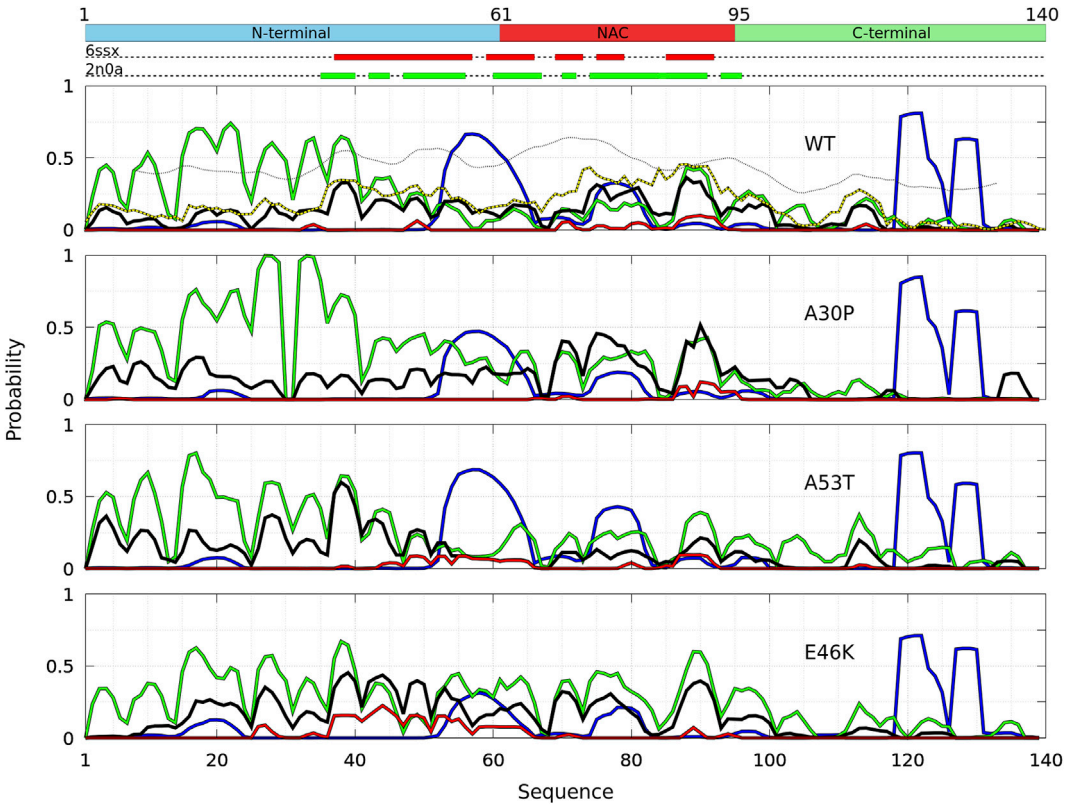

FIGURE 6 | Analysis of the secondary structures for the complete ensemble of α-syn dimers found in molecular dynamics simulations for the WT and the mutants using the CUTABI algorithm (Guzzo et al., 2021). Probabilities for each residue to pertain to an α-helix (blue), to an intra-molecular β-sheet (green), to an intermolecular β-sheet (black), and to a native intermolecular β-sheet fibril type (red). The probability for each residue to form an intermolecular contact (**Figure 2**) is shown for the WT (black/yellow broken line) for comparison. The propensity to form a β-sheet calculated from the empirical scale of Chou and Fasman with the Protscale online tool (https://web.expasy.org/protscale/) using a window of 15 residues and a linear interpolation with a weight of 0.10 for the window edges is also represented (dotted line) for the WT. The red and green rectangles on the upper dotted bars describe the locations of the intermolecular β-sheets identified by CUTABI in the experimental structures of α-syn fibrils with the PDB IDs 6ssx and 2n0a, respectively.

6ssx and 2n0a). To identify conformations belonging to the diagonal of the contact maps, we extracted from the dimer statistics a sub-ensemble of dimers with Nfcs. We define a dimer with fibril-like native contacts (Dfncs) as a dimer conformation that has at least 5 consecutive intermolecular contacts made between residues with the same indices at a contact distance less than 5Å. This value was chosen as the distances between native contacts in fibrils are 4.9Å and 4.7Å in the measured structures with PDB IDs 6ssx and 2n0a. We may consider the Dfncs ensemble as conformations that are the most probable pre-formed fibrils in the present work.

The dimensionless (effective) free-energy landscape of the Dfncs was computed from the two-dimensional probability density P of the number of inter-chain (n_{inter}) and intra-chain (n_{intra}) contacts between the residues for the WT and the mutants (**Figure 7**). For the WT and A53T, the Dfncs is located in one basin, whereas for A30P and E46K, two distinct clusters are visible. Some of the fibril clusters are located at the local minima of the global free-energy landscape (**Figure 1**), except for A53T. For A30P and E46K, one of the two clusters of Dfncs occurs at the global minimum of the global free-energy landscape (**Figure 1**). This indicates that Dfncs are more probable to be formed for these two mutants. In fact, the sub-ensemble of Dfncs represents 14.04% and 15.73% for A30P and E46K, respectively. For the WT and A53T, the population of Dfncs is only 8.33 and 10.65%, respectively. It is worth noting that these numbers indicate that the majority of dimers are disordered aggregates at least on the effective time scale (millisecond) of the present simulations.

The calculation of the probability to find the different secondary structure elements along the amino-acid sequence of the WT and mutants of α-syn dimers was repeated for the sub-ensemble of Dfncs using CUTABI (**Figure 8**). Compared to the ensemble of dimers, the helical regions of Dfncs show significant differences for the WT and A53T: the region 53–65 can only form a helix in the WT, preventing the dimerization of this region, whereas this helical region disappears in A53T and the helical region 75–82 disappears in the WT.

The peaks of probability for a residue to pertain in an intra-molecular β-sheet occur at the same positions in the Dfncs sub-ensemble and in the ensemble of the dimers for all proteins (see **Figure 6**). For the WT, the propensity of intra-molecular β-sheets is strongly increased in the N-terminal region of Dfncs with the

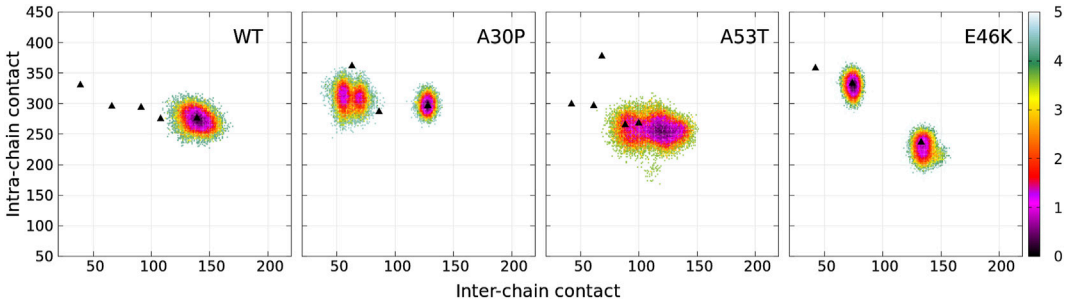

FIGURE 7 | Color maps of $-\ln[\frac{P(n_{Inter},n_{Intra})}{P_{max}}]$ computed from the 2D probability density function P of the number of inter-chain (n_{Inter}) and intra-chain (n_{Intra}) contacts between the residues of α-syn for the sub-ensemble of dimers with fibril native contacts found in molecular dynamics trajectories at 300 and 310 K of the WT and variants. In each map, P_{max} is the maximum value of the probability of the map. The local minima numbered in **Figure 1** are shown by black triangles.

peaks with a probability near 1 meaning residues pertaining to this type of secondary structure in all Dfncs conformations. On the contrary, all residues above position 50 have a totally negligible probability to form an intra-molecular β-sheet in Dfncs. Unlike the ensemble of WT dimers for which intermolecular β-sheets can be found nearly everywhere along the sequence (**Figure 6**), in the Dfncs sub-ensemble, they are found only for residues 43–51 (N-terminal region), 69–95 (NAC region), and 99–100 and 134–135 (C-terminal region). We observe a propensity to form Nfcs, as in fibrils, only for specific regions in the NAC: residues 75–76, 80–82, 85, and 87–93. For the segment 87–93, the superposition of the black and red curves in **Figure 8** means that only Nfcs are formed there. Considering the limited time scale of the present simulations, these residues are probably nucleation centers of pre-fibril-like dimers in the WT.

For A30P, the probability to form intra-molecular β-sheets is similar in Dfncs and in the full ensemble of dimers (**Figure 6**) but with an increase in the region 36–58. Contrary to the ensemble of A30P dimers, for which intermolecular β-sheets can be found nearly everywhere between residues 3–104 along the sequence (**Figure 6**), in Dfncs, they are found in the NAC for residues 70–83 and 87–96 as well as with smaller probabilities in the C-terminal region at 103–105 and in the N-terminal region at 18–20 and 26–28. The propensity to form Nfcs occurs only in the NAC at 87–88 and 90–95, which are probably the nucleation centers for this type of structure for this variant. Clearly, the WT and A30P share the same region of formation of Nfcs in the NAC.

For A53T, the propensity for a residue to pertain to an intra-molecular β-sheet increases in the NAC and C-terminal regions in Dfncs compared to the complete ensemble of dimers (**Figure 6**). The formation of intermolecular β-sheets occurs all along the sequence: residues 3–11, 22–30, and 35–60 (N-terminal region); 61–65, 71, and 87–92 (NAC region), and 112–116 (C-terminal region). A53T differs thus from the WT and A30P by a large number of regions of intermolecular β-sheets, including the N-terminal region. In addition, the probability of segments 44–65 and 87–92 in the NAC to form Nfcs is identical to the probability of intermolecular β-sheets, meaning that these residues form only Nfcs in the Dfncs sub-ensemble and are probably nucleation centers for A53T pre-fibril-like dimers.

For E46K, the probability of intra-molecular β-sheets is similar between Dfncs and the ensemble of dimers. The exception is the peak at residues 38 and 39 for which the probability is 1 as well as the probability for these residues to form Nfcs, meaning that these residues are involved in both Nfc and intra-molecular β-sheets. As for the other proteins, the presence of intermolecular β-sheets is limited to specific regions in contrast to the ensemble of dimers where they were found along all the sequence (**Figure 6**). In addition, nearly all intermolecular β-sheets are Nfcs (the black and red curves in **Figure 8** overlap), namely, at 26–28 and 37–60 (N-terminal region) and 61–65, 75–78, and 88–92 (NAC).

Figure 8 shows that regions of high propensity of Nfcs in A53T and E46K are significantly different from those of the WT and A30P. This can be more clearly seen for the Nfcs in **Figure 9**. **Figure 9** shows the different lengths of the Nfc regions and their probabilities along the sequence. A30P has the shortest consecutive segments of Nfcs, with the longest between residues 86 and 97 (NAC and C-terminal region), with the largest probability of Nfcs at 87–88 (NAC). The maximum number of consecutive Nfcs for the WT is 18 between residues 78 and 95 (NAC). As already mentioned, the formation of Nfcs in A53T and E46K occurs differently compared to the WT and A30P: they are mainly in the N-terminal region. These two variants have the largest segments of consecutive Nfcs on the time scale of the simulation: 25 between residues 43 and 66 for A53T and 37 between residues 34 and 70 for E46K. The structures of the proteins with the largest number of Nfcs are represented in the insets of **Figure 9**, and they can be considered the most probable nucleation structures of Dfncs on the time scale of the present simulations (millisecond). As mentioned earlier, the number of Dfncs structures is the largest for A30P and E46K. One may deduce that the nucleation is easier for these two structures. On the other hand, if we consider the maximum number of Nfcs possibly formed on the time scale of the simulations as a criterion in the difficulty of the growth of Nfcs from a nucleation center, we found by increasing order of difficulty: A30P, WT, A53T, and E46K.

As shown in **Figure 9**, some Dfncs of A53T and E46K have segments that do not start before residue 39 and do not expand above residue 58. It is worth noting that residue 58 is the location

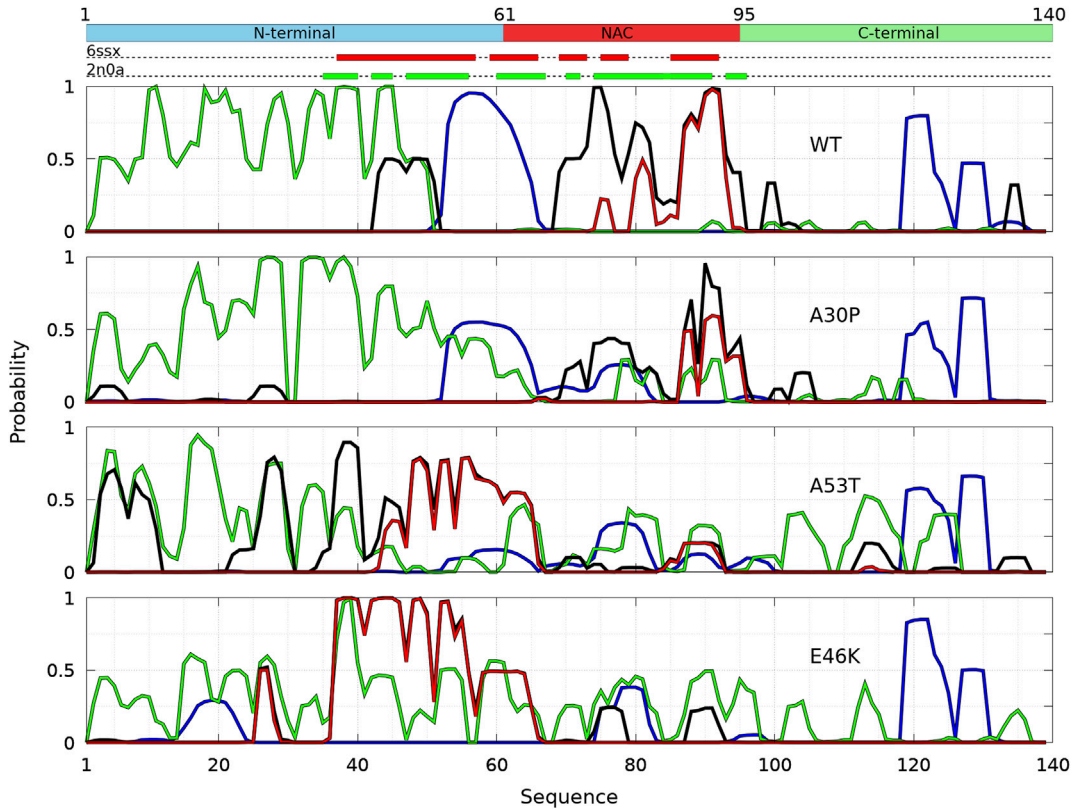

FIGURE 8 | Analysis of the secondary structures in the sub-ensemble of α-syn dimers with native fibril-type contacts found in molecular dynamics simulations for the WT and the mutants at 300 and 310 K. Probabilities for each residue to pertain to an α-helix (blue), to an intra-molecular β-sheet (green), to an intermolecular β-sheet (black), and to a native intermolecular β-sheet fibril type (red).

of the maximum of probability to form a helix and residue 39 has a high propensity to form intra-molecular β-sheets in both disordered dimers (**Figure 6**) and the monomers (Guzzo et al., 2021).

3.4 Selected Local and Global Properties Related to Experimental Techniques

In dimers, the residues may form both intra-molecular and intermolecular parallel and anti-parallel β-sheets that can be distinguished in infrared and Raman spectroscopies using amide bands. **Table 2** shows a significant variation in the amount of parallel β-sheets between the monomers and the dimers. The formation of disordered dimers and Dfncs is clearly characterized by an increase in the proportion of residues in parallel β-sheets compared to monomers: by increasing percentage, one finds monomer, disordered dimers, and Dfncs. The amount of parallel β-sheets in Dfncs is the largest for the WT despite the fact that the length of segments of Nfcs is quite short (**Figure 9**), which implies that most of the parallel β-sheets are in segments of WT monomers that are not aligned as in fibrils. For the proportion of anti-parallel β-sheets, we do not observe a clear systematic variation between the monomers and the dimers. For A53T and E46K, the percentage of anti-parallel β-sheets is increased from the monomers to disordered dimers and Dfncs. No significant change is found for the WT. The case of A30P is special: the proportion of anti-parallel β-sheets is decreased in Dfncs and increased in disordered dimers compared to the monomers. It is worth noting that on the time scale of the simulations, the Dfncs represent about 9–15% of the structures, with the other dimers being disordered.

A global parameter that can be measured by SAXS is the gyration radius. The probability density functions of the gyration radius of the ensemble of dimers are presented in **Figure 10**. Each function can be represented by the sum of several subpopulations described by Gaussian functions. The Gaussian parameters, given in **Table 3**, were computed with the Gaussian mixture model (GMM) algorithm (Reynolds et al., 2009). However, the GMM clustering is misleading as the subpopulations cover large areas of the free-energy landscape maps of the contacts (n_{inter}, n_{intra}) (**Figure 1**) and of the secondary structures (α, β) (**Figure 4**), as shown in **Supplementary Figure S1** and **Supplementary Figure S2**, respectively. Finally, the average gyration radius (**Table 3**) is the smallest for E46K and the largest for A30P.

Single-molecule FRET allows us to extract information on local properties. In Cremades et al. (2012), Horrocks et al. (2015), and Tosatto et al. (2015), the fluorophores (Alexa Fluor 488 and Alexa Fluor 647) were covalently linked at position 90 of α-syn mutants A90C. The FRET efficiencies depend on the distance

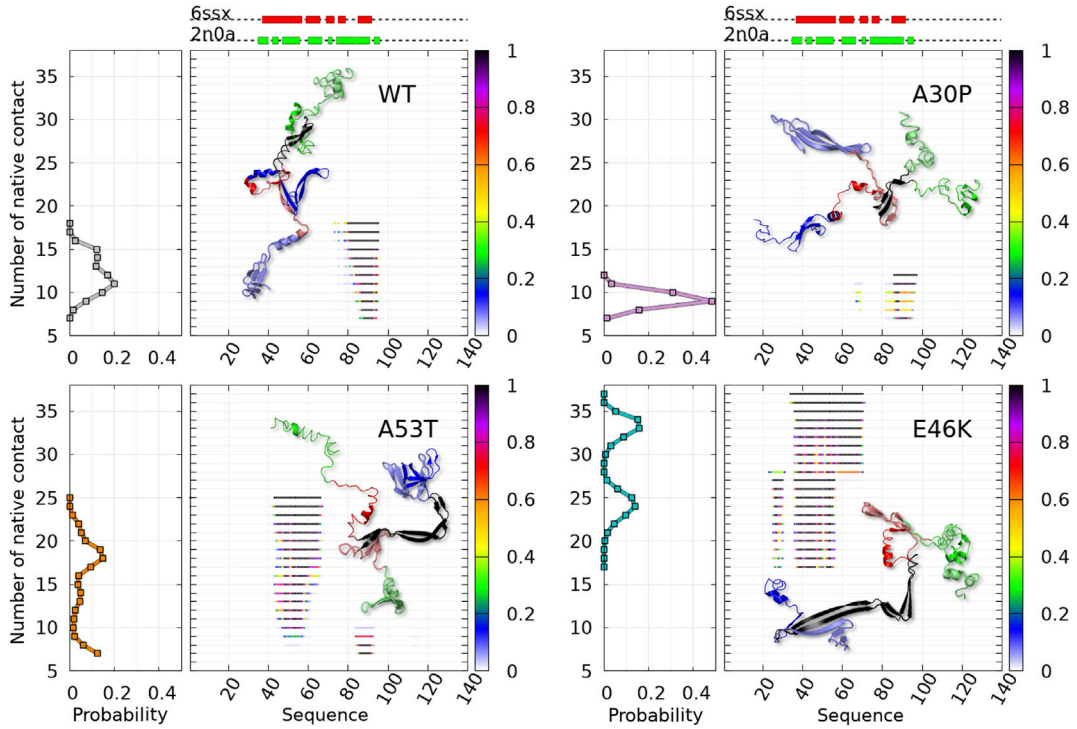

FIGURE 9 | Analysis of the mean contact probability along the amino-acid sequence for the sub-ensemble of Dfncs found in MD trajectories for the WT and variants. For each protein, the left panel represents the probability density of the number of contacts and the right panel represents its distribution along the sequence for each point of the distribution, with a color code giving the probability at each residue according to the color bar on the right. 3D structure representatives of the maximum number of native fibril-type contacts are shown with the following color: N-terminal (blue), NAC (red), C-terminal (green), and native contact region (black).

TABLE 2 | Percentages of residues in parallel and anti-parallel intra- and intermolecular β-sheets computed with CUTABI for the ensemble of dimers, the sub-ensemble of Dfncs, and isolated monomers.

	Dimer	Dimer	Dfncs	Dfncs	Monomer	Monomer
Protein	**Parallel**	**Anti-parallel**	**Parallel**	**Anti-parallel**	**Parallel**	**Anti-parallel**
WT	16.5	27.9	29.1	28.7	10.6	28.4
A30P	18.2	35.1	21.0	27.0	10.9	31.1
A53T	17.1	29.4	22.8	30.8	11.0	22.4
E46K	18.9	31.8	23.1	30.0	10.3	25.8

between the fluorophores. Therefore, the FRET experimental procedure observed the aggregation of α-syn using a specific local property that is directly correlated with the distance between residues 90 of interacting monomers. It is, therefore, interesting to compute the probability density functions of the distances between the C^{α} of residues A90 in disordered dimers and in Dfncs. These functions are presented in **Figure 11** (panels A and B) for the WT and mutants. In disordered dimers, we observe a heterogeneity of these distances with a group of peaks below 15 Å and other peaks between 20 Å and 80 Å (panel A). In the subpopulation of Dfncs, there is a drastic change in the distance probability distribution (panel B). All proteins, except E46K, have a peak around 3.8 Å. This can be explained by **Figures 8**, **9**, which show that residue 90 is involved in Nfcs for all proteins, except E46K. All other peaks present in disordered dimers have disappeared in Dfncs, except for E46K, for which we observe a peak at around 15 Å and a large background. Regarding this specific A90–A90 distance, the difference between the probability densities of disordered dimers and Dfncs is spectacular. In addition, this clearly shows that E46K has a different behavior regarding this local parameter.

4 DISCUSSION

The first important finding of the present simulations is that the ensemble of α-syn conformations (WT and mutants) is strongly heterogeneous, as shown by the computed probability

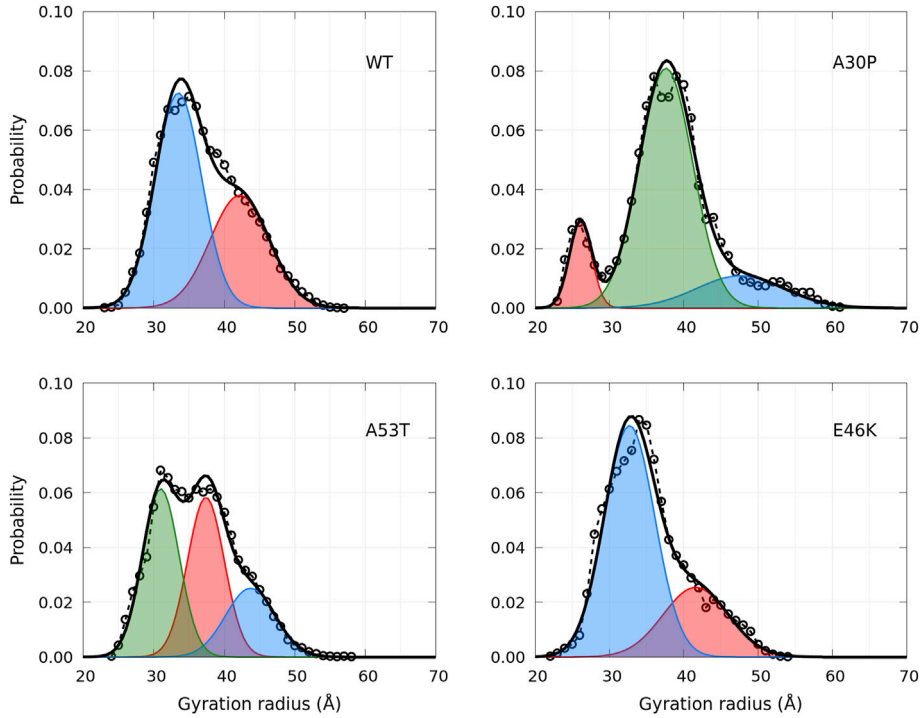

FIGURE 10 | Probability density functions of the gyration radius for the WT and mutants computed from MD simulations (circle symbols and dashed lines). Each distribution is represented by a set of Gaussians (colored areas), the number and the parameters of which were found by applying the GMM algorithm. The sum of Gaussians is represented by a full black line.

TABLE 3 | Clustering of the gyration radius probability density using the GMM algorithm. The values in brackets are the corresponding % of the ensemble of the conformations.

Protein	Cluster 1	Cluster 2	Cluster 3	Average value
WT	33.5 Å (60%)	42.2 Å (40%)	-	37.0 Å (100%)
A30P	37.6 Å (73%)	47.6 Å (16%)	26.1 Å (11%)	37.9 Å (100%)
A53T	31.1 Å (41%)	37.6 Å (38%)	43.8 Å (21%)	36.2 Å (100%)
E46K	32.7 Å (72%)	41.7 Å (28%)	-	35.2 Å (100%)

density functions of the gyration radius (**Figure 10**), the effective free-energy landscape of the contacts (**Figure 1**), the effective free-energy landscape of the secondary structures (**Figure 4**), and the probability density function of the A90–A90 distance (**Figure 11**). The second significant finding is that the ensemble of heterogeneous dimer conformations is divided in a majority of disordered dimers and a complementary minority of Dfncs (**Figure 9**). The third important finding is that a single amino-acid substitution has a huge effect on the contact probability between (hydrophobic) residues of α-syn, which is not limited to the vicinity of the mutated residue as shown in **Figure 6** for disordered dimers and in **Figures 8**, **9** for Dfncs. Keeping in mind that the effective time scale of the present simulations (millisecond) represents the early lag phase of fibril growth, as it is three to four orders of magnitude smaller than the lag phase of fibrils observed experimentally, we will, however, attempt next to make qualitative comparisons between these three significant results and different experimental findings.

Unfortunately, there are no SAXS or SANS measurements of α-syn dimers to compare directly with **Figure 10** to the best of our knowledge. It is worth noting that the distribution of the gyration radius of α-syn monomers is single peak, except for the sub-ensemble of A30P and E46K monomers without an alpha-helix (B monomer state), for which a small peak at R_g = 18 Å was found (Guzzo et al., 2021). For recombinant monomeric α-syn, the average gyration radius, extrapolated at infinite dilution, is 27.2 ± 0.44 Å (Araki et al., 2016), comparable to the value calculated previously in our MD simulations (24.7Å) (Guzzo et al., 2021). The experimental value of the gyration radius of a monomer varies among the experiments: for example, in Tris-buffer, it is increased to 42.7Å (Araki et al., 2016). We hypothesize that the presence of dimers in a monomeric solution may give an apparent larger gyration radius of the monomers because the dimer conformational ensembles have an average gyration radius of about 10Å larger than one of the monomer ensembles (**Table 2**).

The heterogeneity of the size distribution (**Figure 10**) and of the contacts (**Figure 1**) of α-syn dimers can be related to the rupture-force AFM experiments on WT, A30P, A53T, and E46K α-syn dimers in solution (Krasnoslobodtsev et al., 2013). These experiments were performed at low pH to increase the aggregation propensity of α-syn and with the additional mutation A40C needed to hang one of the monomers on the AFM tip and the other on the surface. Contour lengths of the

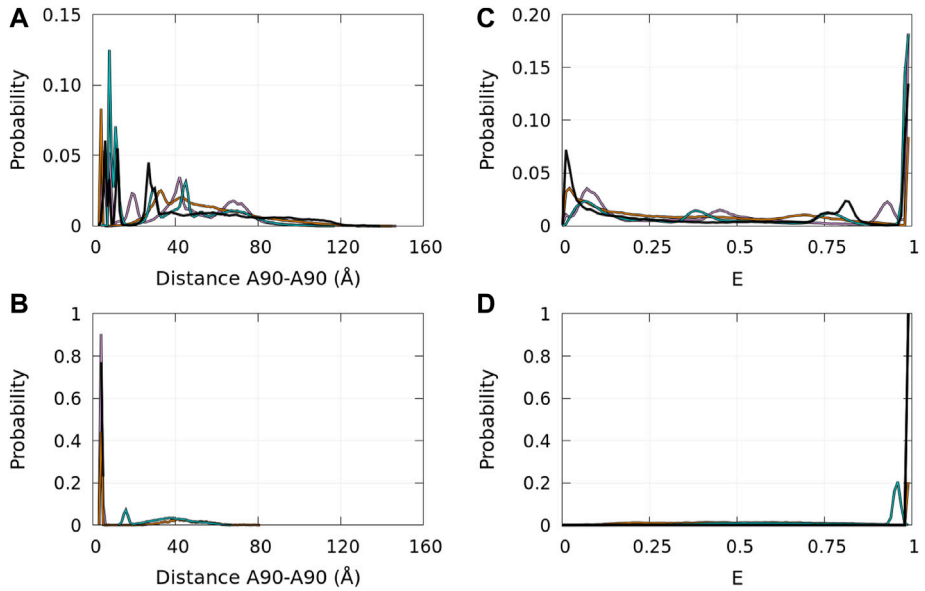

FIGURE 11 | Probability density functions of the distance between the C^{α} atoms of residues A90 in the ensembles of disordered dimers [panel **(A)**] and of Dfncs [panel **(B)**] and the corresponding probability distribution functions of the approximated FRET efficiency E (see text) for the ensembles of disordered dimers [panel **(C)**] and of Dfncs [panel **(D)**]. Color code: WT—black, A30P—purple, E46K—turquoise, and A53T—orange.

dimers were extracted from force–distance curves. They were interpreted as total lengths of stretchable parts of α-syn molecules that are not involved in dimeric interaction and, thus, as the sum of the total length from the C-terminal anchor point to the first residue of a dimerization region in each monomer. The data show a multi-peak distribution of contour lengths (Krasnoslobodtsev et al., 2013), which might reflect the multiple minima observed in **Figure 1**, as the contour length of a dimer conformation is related to the number of intra-molecular and intermolecular contacts. Moreover, the distribution of contour lengths is different for each α-syn variant in the AFM experiments, as in **Figure 1** for the contacts. We hypothesize that the number of single and multiple rupture force events might be correlated with the propensity of the dimers to form Nfcs. A larger number of multiple rupture force events for A53T and E46K compared to the WT and A30P were interpreted as multiple interaction segments for the former (Krasnoslobodtsev et al., 2013). This might be supported by **Figure 8** (black and red curves) as the WT and A30P form Nfcs and intermolecular β-sheets in shorter localized segments than A53T and E46K. Moreover, contour lengths can be estimated for residues not involved in Nfcs shown in **Figure 8** by assuming a C^{α} – C^{α} virtual bond distance of 3.8 Å and evaluating the contour length as twice (for the two monomers) the distance starting from the C-terminal to the first residue forming a Nfc (Krasnoslobodtsev et al., 2013). For WT Dfncs, we found a contour length of 357 Å (residues 140 to 94). For A30P Dfncs, the distance is similar. For A53T Dfncs, two contour lengths are estimated: 380 Å (residues 140 to 91) and 585 Å (residues 140 to 64). This last value is the same for E46K Dfncs. These values are on the same order of magnitude than the most probable contour lengths in the experiment (340 Å and 440 Å) (Krasnoslobodtsev et al., 2013). The comparison, although qualitative, indicates that the heterogeneity of α-syn dimers and the influence of a single amino-acid substitution on their structural properties, found in MD, agree qualitatively with the AFM force–distance data.

The heterogeneity of the (α, β) map could be tested in the early lag phase of fibril formation by using single-molecule Raman spectroscopy (Dai et al., 2021), as α-helices and β-sheets have been well described by Raman fingerprints. Such data are not available so far. In the present simulations, the ensemble of heterogeneous dimer conformations is divided in a majority of disordered dimers and a complementary minority of Dfncs for both the WT and mutants. Dfncs could be identified by spectroscopy. Indeed, the percentage of parallel intermolecular β-sheets is characteristic of Dfncs as it is significantly larger than the one in monomers and in disordered dimers (**Table 2**). This result agrees with FTIR spectroscopy measurements indicating that WT α-syn fibrils have a majority of parallel β-sheets, contrary to oligomers which show a majority of anti-parallel β-sheets (Chen et al., 2015). The total amount of β-sheets estimated from FTIR (Chen et al., 2015) is 35 ± 5% for oligomers of 10–40 molecules and 65 ± 10% for fibrils. In spite of the smaller size of dimers and the shorter time-scale of the present simulations, these values compare qualitatively with the sum of parallel and anti-parallel β-sheets in WT disordered dimers (44.4%) and in Dfncs (57,8%), respectively (**Table 2**).

Probably, the best technique to identify Dfncs from disordered dimers would be single-molecule FRET. We tentatively compared the results of **Figure 11** with FRET studies of α-syn oligomerization (Cremades et al., 2017; Cremades et al., 2012; Tosatto et al., 2015; Horrocks et al., 2015) in the early lag phase of fibril formation. A quantitative prediction of the FRET efficiencies from α-syn conformations would require to

TABLE 4 | Main residues or segments identified by MD as important for the dimerization of α-syn from the maxima of propensity for the mean contact, intermolecular β-sheets, and Nfcs.

Protein	Mean contact	Intermolecular β-sheet (all dimers)	Nfcs in Dfncs
WT	I88	I88	T75, A76, K80, T81,V82, A85, S87-G93
A30P	V70	A90	S87, I88, A90-V95
A53T	L38	L38	T44-N65, S87-T92
E46K	V49	Y39	V26-E28, V37-N65

simulate the protein labeled with the fluorescent molecules and to model the FRET mechanism (Hoefling et al., 2011). Such an approach is difficult to apply for a large and complicated protein like α-syn. In addition, to predict protein structure modifications introduced by the cysteine mutations and the fluorescent organic molecules, it is necessary to build and test a new coarse-grained force field for the fluorescent molecules covalently linked to α-syn. This task is extremely demanding in computer time and is well beyond the scope of the present work. For these reasons, we chose to compute an approximate efficiency E from the distance distributions between the C^{α} atoms shown in **Figure 11** (panels A and B). From these distances, we extracted an approximated FRET efficiency E as $E = 1/[1 + (d/R_0)^6]$, where d is the distance between the C^{α} atoms increased by 20Å and R_0 is the Förster distance with $R_0 = 60$Å for the fluorophores used (dos Remedios and Moens, 1995). Note that a single value of R_0 is a severe approximation that assumes an averaging of the fluorophore orientation. The shift of 20Å takes into account twice the estimated distance between the dye center of a fluorophore and the C^{α} atom to which it is linked (Nicolaï et al., 2013).

The probability distribution of E is shown in **Figure 11** (panels C and D). In disordered dimers (panel C), we observed two regions with peaks of low (< 0.15) and of high efficiency (> 0.75). In Dfncs (panel D), peaks are observed only at high efficiency (> 0.9) with a small background at all efficiencies. It is tempting to associate the Dfncs with type B dimers. The disordered dimers are both A and B (panel C). It was proposed that the α-syn dimers might be a mixture of types A and B, which cannot be resolved experimentally (Horrocks et al., 2015). The time scale simulated here is of course extremely short compared to that in FRET experiments, but it is very interesting to observe that the formation of consecutive Nfcs leads to an increase in the FRET efficiency. It is worth noting that these results are not very sensitive to the exact value of the estimated distance between the dye center of a fluorophore and the C^{α} atom. Using a distance of 7.5Å (**Supplementary Figure S3**) and 15Å (**Supplementary Figure S4**) instead of 10Å (**Figure 11**) leads to the same conclusions. Comparison with **Figure 9** shows that A90 belongs to all segments of consecutive Nfcs, except for E46K. However, **Figure 8** shows that the Dfncs of the E46K variant form other contacts in the A90 region. Thus, E46K behaves differently from the WT, A30P, and A53T in the A90 region, which may be a hint to explain the difficulty in reproducing the probability distribution of the FRET signal in the experiments (Horrocks et al., 2015). If the population of dimers formed in the early lag phase measured by FRET for A30P is taken as 1 for reference, the population of WT and A53T dimers was 0.4 and 0.64, respectively (Horrocks et al., 2015). In the present simulations, the populations of Dfncs using the same reference are 0.59 and 0.75 for the WT and A53T. However, the total populations of dimers out of all the conformations (monomers + dimers) found in the simulations are different. Taking A30P as a reference (=1), we found 1.34 and 1.26 for the WT and A53T, respectively.

The effect of a single amino-acid substitution on the propensity to form contacts (**Figure 1**) and on the formation of Nfcs (**Figures 6**, **8**, **9**) is spectacular. How could these results be compared to various differences observed experimentally between the variants and the WT? The aggregation of A30P in fibrils is slower than the WT (Conway et al., 2000), whereas E46K and A53T aggregate faster than the WT (Conway et al., 1998; Greenbaum et al., 2005). However, the A30P monomer was consumed at a comparable rate or slightly more rapidly than the WT monomer, whereas A53T was consumed even more rapidly (Conway et al., 2000). This might be explained by the early nucleation steps of the fibrils, which is presented by **Figure 8**. Indeed, E46K and A53T form *larger* regions of Nfcs than the A30P and WT, which might be interpreted as they are "faster" to form pre-fibrils. For A30P, the region of Nfcs is extremely short. However, the sub-ensemble of Dfncs in the dimer ensemble is larger for A30P (14.04%) and E46K (15.73%) in comparison with the WT (8.33%) and A53T (10.65%). On the opposite, the number of dimers found in the MD simulations is the lowest for A30P (23%), followed by an increasing order by A53T (29%), the WT (31%), and E46K (35%). To make these percentages more concrete, it is better to use an example. For 123 monomers in solution, A30P forms 20 disordered dimers, and 3 Dfncs and 77 monomers remain free. For 121 monomers in solution, the WT forms 28 disordered dimers, and 3 Dfncs and 69 monomers remain free. For 129 monomers, A53T forms 26 disordered dimers, and 3 Dfncs and 71 monomers remain free. Finally, for 135 monomers, E46K forms 30 disordered dimers, and 5 to 6 Dfncs and 65 monomers remain free. Thus, all proteins will form approximately the same number of Dfncs, except E46K, but the size of Dfncs on the same time scale is much larger for A53T and E46K (**Figure 9**).

In order to compare the present results to WT mutagenesis experiments, we list the key residues that play a role in α-syn dimerization in the present simulations in **Table 4** for the WT and mutants. These residues were selected from the maxima of probabilities of mean contacts (**Figure 1**) and of the formation of intermolecular β-sheets and Nfcs (**Figures 6**, **8**).

For the WT, residue I88 is of paramount importance. The segment S87-G93 forms only Nfcs in the sub-ensemble of Dfncs (the black and red curves are superposed in **Figure 8**). This agrees with the observation that the removal of the segment A85-E94 reduces the α-syn polymerization (Waxman et al., 2009). Truncation of V71-V82 prevents the polymerization of fibrils

(Waxman et al., 2009). This agrees with the fact that K80-V82 forms Nfcs. We found that the pair T75-A76 has non-negligible probabilities to form Nfcs and other intermolecular β-sheets (**Figure 8**). However, removal of residue A76 or V77 alone has no effect on the polymerization, but the missing pair A76-V77 prevents the polymerization in fibrils (Waxman et al., 2009). The role of A76 cannot be neglected. It is worth noting that **Table 4** suggests that mutations or cutting segments in the N-terminus of A53T and E46K could provide information on the key residues promoting the aggregation of these variants.

The role of the N-terminal region in α-syn in the dimerization is not to be underestimated as shown by the formation of Nfcs for A53T and E46K and by the high probability to form intra-molecular β-sheets for the WT and A30P in this region (**Figures 8**, **9**). In a recent work, two segments in the N-terminal region that regulate the α-syn polymerization have been identified: G36-S42 (named P1) and K45-E57 (named P2) (Doherty et al., 2020; Tripathi, 2020). The removal of P1 at pH 7.5 prevents α-syn aggregation but not at pH 4.5. The removal of P1 and P2 prevents the aggregation at both pH. The present simulations are calibrated at pH 7 for which the force field was developed. As a single amino-acid substitution has a huge effect on the propensity of aggregation, it is difficult to compare the results of P1 truncated protein with the present results. However, dimerization of A53T and E46K shows that a mutation in this region drastically change the Nfc propensity. Moreover, for the WT, the region G36-S42 is a region with high propensity of intra-molecular β-sheets in Dfncs which precedes a region of intermolecular β-sheets (43–51) (black curve in **Figure 8**). It is possible that the formation of intra-molecular β-sheets may be coupled to the formation of fibrils. Finally, region P2 overlaps the region of the main helical segment of WT monomers (Guzzo et al., 2021) and dimers (**Figures 6**, **8**). Clearly, the high probability of this helical segment must play a role in the polymerization.

In conclusion, the present MD simulations show that the dimer conformations are largely heterogeneous with both disordered and pre-fibrillar-like dimers that differ between the WT and the variants A30P, A53T, and E46K. Despite the limitations inherent to any MD simulations (accuracy of the force field and the limited time scale), the present findings agree quite well with available experimental data and suggest possible further spectroscopic and mutagenesis experiments.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**; further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

PS and GGM designed the research. AG performed the research. AG, PS, and GGM analyzed the data. AR provided technical assistance with the UNRES package. AN performed the Gaussian clustering of gyration radii. PS wrote the manuscript, and AG did the figures. All authors discussed the results and reviewed the manuscript.

FUNDING

The calculations were performed using HPC resources from DSI-CCuB (Centre de Calcul de l'Université de Bourgogne) and 588-processor Beowulf cluster at the Baker Laboratory of Chemistry and Chemical Biology, Cornell University. This work is part of the project NANO-NEURO-MED (2019–2022) supported by the EIPHI Graduate School (contract ANR-17-EUR-0002), the Conseil Régional de Bourgogne Franche-Comté, and the European Union through the PO FEDER-FSE Bourgogne 2014/2020 programs. This work is supported by a grant from the National Institutes of Health (Grant no. R01GM14312).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2022.910104/full#supplementary-material

REFERENCES

Antonschmidt, L., Dervişoğlu, R., Sant, V., Tekwani Movellan, K., Mey, I., Riedel, D., et al. (2021). Insights into the Molecular Mechanism of Amyloid Filament Formation: Segmental Folding of α-synuclein on Lipid Membranes. *Sci. Adv.* 7, eabg2174. doi:10.1126/sciadv.abg2174

Appel-Cresswell, S., Vilarino-Guell, C., Encarnacion, M., Sherman, H., Yu, I., Shah, B., et al. (2013). Alpha-synuclein p.H50Q, a Novel Pathogenic Mutation for Parkinson's Disease. *Mov. Disord.* 28, 811–813. doi:10.1002/mds.25421

Araki, K., Yagi, N., Aoyama, K., Choong, C.-J., Hayakawa, H., Fujimura, H., et al. (2019). Parkinson's Disease Is a Type of Amyloidosis Featuring Accumulation of Amyloid Fibrils of α-synuclein. *Proc. Natl. Acad. Sci. U.S.A.* 116, 17963–17969. ISBN: 9781906124113 Publisher: National Academy of Sciences Section: Biological Sciences. doi:10.1073/pnas.1906124116

Araki, K., Yagi, N., Nakatani, R., Sekiguchi, H., So, M., Yagi, H., et al. (2016). A Small-Angle X-Ray Scattering Study of Alpha-Synuclein from Human Red Blood Cells. *Sci. Rep.* 6, 30473. doi:10.1038/srep30473

Bartels, T., Choi, J. G., and Selkoe, D. J. (2011). α-Synuclein Occurs Physiologically as a Helically Folded Tetramer that Resists Aggregation. *Nature* 477, 107–110. doi:10.1038/nature10324

Bhak, G., Lee, S., Kim, T. H., Lee, J. H., Yang, J. E., Joo, K., et al. (2018). Morphological Evaluation of Meta-Stable Oligomers of α-Synuclein with Small-Angle Neutron Scattering. *Sci. Rep.* 8, 14295. doi:10.1038/s41598-018-32655-0

Bousset, L., Pieri, L., Ruiz-Arlandis, G., Gath, J., Jensen, P. H., Habenstein, B., et al. (2013). Structural and Functional Characterization of Two Alpha-Synuclein Strains. *Nat. Commun.* 4, 2575. doi:10.1038/ncomms3575

Breydo, L., Wu, J. W., and Uversky, V. N. (2012). α-Synuclein Misfolding and Parkinson's Disease. *Biochimica Biophysica Acta (BBA) - Mol. Basis Dis.* 1822, 261–285. doi:10.1016/j.bbadis.2011.10.002

Chen, S. W., Drakulic, S., Deas, E., Ouberai, M., Aprile, F. A., Arranz, R., et al. (2015). Structural Characterization of Toxic Oligomers that Are Kinetically Trapped during α-synuclein Fibril Formation. *Proc. Natl. Acad. Sci. U.S.A.* 112, E1994–E2003. Publisher: National Academy of Sciences Section: PNAS Plus. doi:10.1073/pnas.1421204112

Chiti, F., and Dobson, C. M. (2017). Protein Misfolding, Amyloid Formation, and Human Disease: A Summary of Progress over the Last Decade. *Annu. Rev. Biochem.* 86, 27–68. doi:10.1146/annurev-biochem-061516-045115

Chiti, F., and Dobson, C. M. (2006). Protein Misfolding, Functional Amyloid, and Human Disease. *Annu. Rev. Biochem.* 75, 333–366. doi:10.1146/annurev.biochem.75.101304.123901

Chou, P. Y., and Fasman, G. D. (1974). Conformational Parameters for Amino Acids in Helical, β-sheet, and Random Coil Regions Calculated from Proteins. *Biochemistry* 13, 211–222. doi:10.1021/bi00699a001

Conway, K. A., Harper, J. D., and Lansbury, P. T. (1998). Accelerated *In Vitro* Fibril Formation by a Mutant α-synuclein Linked to Early-Onset Parkinson Disease. *Nat. Med.* 4, 1318–1320. doi:10.1038/3311

Conway, K. A., Lee, S.-J., Rochet, J.-C., Ding, T. T., Williamson, R. E., and Lansbury, P. T. (2000). Acceleration of Oligomerization, Not Fibrillization, Is a Shared Property of Both α-synuclein Mutations Linked to Early-Onset Parkinson's Disease: Implications for Pathogenesis and Therapy. *Proc. Natl. Acad. Sci. U.S.A.* 97, 571–576. Publisher: National Academy of Sciences Section: Biological Sciences. doi:10.1073/pnas.97.2.571

Cote, Y., Delarue, P., Scheraga, H. A., Senet, P., and Maisuradze, G. G. (2018). From a Highly Disordered to a Metastable State: Uncovering Insights of α-Synuclein. *ACS Chem. Neurosci.* 9, 1051–1065. doi:10.1021/acschemneuro.7b00446

Cremades, N., Chen, S. W., and Dobson, C. M. (2017). "Structural Characteristics of α-Synuclein Oligomers," in *International Review of Cell and Molecular Biology. Early Stage Protein Misfolding and Amyloid Aggregation*. Editor M Sandal (Cambridge, MA, San Diego, CA, Oxford, London: Academic Press), 329, 79–143. doi:10.1016/bs.ircmb.2016.08.010

Cremades, N., Cohen, S. I. A., Deas, E., Abramov, A. Y., Chen, A. Y., Orte, A., et al. (2012). Direct Observation of the Interconversion of Normal and Toxic Forms of α-Synuclein. *Cell.* 149, 1048–1059. doi:10.1016/j.cell.2012.03.037

Dai, X., Fu, W., Chi, H., Mesias, V. S. D., Zhu, H., Leung, C. W., et al. (2021). Optical Tweezers-Controlled Hotspot for Sensitive and Reproducible Surface-Enhanced Raman Spectroscopy Characterization of Native Protein Structures. *Nat. Commun.* 12, 1292. Bandiera_abtest: a Cc_license_type: cc_by Cg_type: Nature Research Journals Number: 1 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Bioanalytical chemistry;Intrinsically disordered proteins;Nanoparticles;Optical manipulation and tweezers;Raman spectroscopy Subject_term_id: bioanalytical-chemistry;intrinsically-disordered-proteins; nanoparticles;optical-manipulation-and-tweezers;raman-spectroscopy. doi:10.1038/s41467-021-21543-3

De Giorgi, F., Laferrière, F., Zinghirino, F., Faggiani, E., Lends, A., Bertoni, M., et al. (2020). Novel Self-Replicating α-synuclein Polymorphs that Escape ThT Monitoring Can Spontaneously Emerge and Acutely Spread in Neurons. *Sci. Adv.* 6, eabc4364. doi:10.1126/sciadv.abc4364

Dettmer, U., Newman, A. J., von Saucken, V. E., Bartels, T., and Selkoe, D. (2015). KTKEGV Repeat Motifs Are Key Mediators of Normal α-synuclein Tetramerization: Their Mutation Causes Excess Monomers and Neurotoxicity. *Proc. Natl. Acad. Sci. U.S.A.* 112, 9596–9601. doi:10.1073/pnas.1505953112

Doherty, C. P. A., Ulamec, S. M., Maya-Martinez, R., Good, S. C., Makepeace, J., Khan, G. N., et al. (2020). A Short Motif in the N-Terminal Region of α-synuclein Is Critical for Both Aggregation and Function. *Nat. Struct. Mol. Biol.* 27, 249–259. doi:10.1038/s41594-020-0384-x

dos Remedios, C. G., and Moens, P. D. J. (1995). Fluorescence Resonance Energy Transfer Spectroscopy Is a Reliable "Ruler" for Measuring Structural Changes in Proteins. *J. Struct. Biol.* 115, 175–185. doi:10.1006/jsbi.1995.1042

Fredenburg, R. A., Rospigliosi, C., Meray, R. K., Kessler, J. C., Lashuel, H. A., Eliezer, D., et al. (2007). The Impact of the E46K Mutation on the Properties of α-Synuclein in its Monomeric and Oligomeric States. *Biochemistry* 46, 7107–7118. Publisher: American Chemical Society. doi:10.1021/bi7000246

Fuchs, J., Tichopad, A., Golub, Y., Munz, M., Schweitzer, K. J., Wolf, B., et al. (2008). Genetic Variability in the SNCA Gene Influences α-synuclein Levels in the Blood and Brain. *FASEB J.* 22, 1327–1334. doi:10.1096/fj.07-9348com

Fusco, G., Sanz-Hernandez, M., and De Simone, A. (2018). Order and Disorder in the Physiological Membrane Binding of α-synuclein. *Curr. Opin. Struct. Biol.* 48, 49–57. doi:10.1016/j.sbi.2017.09.004

Ghosh, D., Singh, P. K., Sahay, S., Jha, N. N., Jacob, R. S., Sen, S., et al. (2015). Structure Based Aggregation Studies Reveal the Presence of Helix-Rich Intermediate during α-Synuclein Aggregation. *Sci. Rep.* 5, 9228. doi:10.1038/srep09228

Giehm, L., Svergun, D. I., Otzen, D. E., and Vestergaard, B. (2011). Low-resolution Structure of a Vesicle Disrupting α-synuclein Oligomer that Accumulates during Fibrillation. *Proc. Natl. Acad. Sci. U.S.A.* 108, 3246–3251. Publisher: National Academy of Sciences Section: Biological Sciences. doi:10.1073/pnas.1013225108

Gołaś, E., Maisuradze, G. G., Senet, P., Ołdziej, S., Czaplewski, C., Scheraga, H. A., et al. (2012). Simulation of the Opening and Closing of Hsp70 Chaperones by Coarse-Grained Molecular Dynamics. *J. Chem. Theory Comput.* 8, 1750–1764. doi:10.1021/ct200680g

Greenbaum, E. A., Graves, C. L., Mishizen-Eberz, A. J., Lupoli, M. A., Lynch, D. R., Englander, S. W., et al. (2005). The E46K Mutation in Alpha-Synuclein Increases Amyloid Fibril Formation. *J. Biol. Chem.* 280, 7800–7807. Publisher: Elsevier. doi:10.1074/jbc.M411638200

Guerrero-Ferreira, R., Kovacik, L., Ni, D., and Stahlberg, H. (2020). New Insights on the Structure of Alpha-Synuclein Fibrils Using Cryo-Electron Microscopy. *Curr. Opin. Neurobiol.* 61, 89–95. doi:10.1016/j.conb.2020.01.014

Guerrero-Ferreira, R., Taylor, N. M., Arteni, A.-A., Kumari, P., Mona, D., Ringler, P., et al. (2019). Two New Polymorphic Structures of Human Full-Length Alpha-Synuclein Fibrils Solved by Cryo-Electron Microscopy. *eLife* 8, e48907. doi:10.7554/eLife.48907

Gurry, T., Ullman, O., Fisher, C. K., Perovic, I., Pochapsky, T., and Stultz, C. M. (2013). The Dynamic Structure of α-Synuclein Multimers. *J. Am. Chem. Soc.* 135, 3865–3872. Publisher: American Chemical Society. doi:10.1021/ja310518p

Guzzo, A., Delarue, P., Rojas, A., Nicolaï, A., Maisuradze, G. G., and Senet, P. (2021). Missense Mutations Modify the Conformational Ensemble of the α-Synuclein Monomer Which Exhibits a Two-phase Characteristic. *Front. Mol. Biosci.* 8, 1104. doi:10.3389/fmolb.2021.786123

Hoefling, M., Lima, N., Haenni, D., Seidel, C. A. M., Schuler, B., and Grubmüller, H. (2011). Structural Heterogeneity and Quantitative FRET Efficiency Distributions of Polyprolines through a Hybrid Atomistic Simulation and Monte Carlo Approach. *PLOS ONE* 6, e19791. Publisher: Public Library of Science. doi:10.1371/journal.pone.0019791

Horrocks, M. H., Tosatto, L., Dear, A. J., Garcia, G. A., Iljina, M., Cremades, N., et al. (2015). Fast Flow Microfluidics and Single-Molecule Fluorescence for the Rapid Characterization of α-Synuclein Oligomers. *Anal. Chem.* 87, 8818–8826. Publisher: American Chemical Society. doi:10.1021/acs.analchem.5b01811

Iadanza, M. G., Jackson, M. P., Hewitt, E. W., Ranson, N. A., and Radford, S. E. (2018). A New Era for Understanding Amyloid Structures and Disease. *Nat. Rev. Mol. Cell. Biol.* 19, 755–773. Bandiera_abtest: a Cg_type: Nature Research Journals Number: 12 Primary_atype: Reviews Publisher: Nature Publishing Group Subject_term: Alzheimer's disease;Neurodegenerative diseases;Prions; Protein aggregation;Protein folding;Structural biology Subject_term_id: alzheimers-disease;neurodegenerative-diseases;prions;protein-aggregation; protein-folding;structural-biology. doi:10.1038/s41580-018-0060-8

Jakes, R., Spillantini, M. G., and Goedert, M. (1994). Identification of Two Distinct Synucleins from Human Brain. *FEBS Lett.* 345, 27–32. doi:10.1016/0014-5793(94)00395-5

Kabsch, W., and Sander, C. (1983). Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-Bonded and Geometrical Features. *Biopolymers* 22, 2577–2637. doi:10.1002/bip.360221211

Khalili, M., Liwo, A., Jagielska, A., and Scheraga, H. A. (2005). Molecular Dynamics with the United-Residue Model of Polypeptide Chains. II. Langevin and Berendsen-Bath Dynamics and Tests on Model α-Helical Systems. *J. Phys. Chem. B* 109, 13798–13810. doi:10.1021/jp058007w

Krasnoslobodtsev, A. V., Volkov, I. L., Asiago, J. M., Hindupur, J., Rochet, J.-C., and Lyubchenko, Y. L. (2013). α-Synuclein Misfolding Assessed with Single Molecule AFM Force Spectroscopy: Effect of Pathogenic Mutations. *Biochemistry* 52, 7377–7386. Publisher: American Chemical Society. doi:10.1021/bi401037z

Krishnan, S., Chi, E. Y., Wood, S. J., Kendrick, B. S., Li, C., Garzon-Rodriguez, W., et al. (2003). Oxidative Dimer Formation Is the Critical Rate-Limiting Step for

Parkinson's Disease α-Synuclein Fibrillogenesis. *Biochemistry* 42, 829–837. Publisher: American Chemical Society. doi:10.1021/bi026528t

Krüger, R., Kuhn, W., Müller, T., Woitalla, D., Graeber, M., Kösel, S., et al. (1998). AlaSOPro Mutation in the Gene Encoding α-synuclein in Parkinson's Disease. *Nat. Genet.* 18, 106–108. doi:10.1038/ng0298-106

Laganowsky, A., Liu, C., Sawaya, M. R., Whitelegge, J. P., Park, J., Zhao, M., et al. (2012). Atomic View of a Toxic Amyloid Small Oligomer. *Science* 335, 1228–1231. Publisher: American Association for the Advancement of Science. doi:10.1126/science.1213151

Lashuel, H. A., Petre, B. M., Wall, J., Simon, M., Nowak, R. J., Walz, T., et al. (2002). Alpha-synuclein, Especially the Parkinson's Disease-Associated Mutants, Forms Pore-like Annular and Tubular Protofibrils. *J. Mol. Biol.* 322, 1089–1102. doi:10.1016/s0022-2836(02)00735-0

Lashuel, H. A. (2020). Do Lewy Bodies Contain Alpha-Synuclein Fibrils? and Does it Matter? A Brief History and Critical Analysis of Recent Reports. *Neurobiol. Dis.* 141, 104876. doi:10.1016/j.nbd.2020.104876

Lázaro, D. F., Rodrigues, E. F., Langohr, R., Shahpasandzadeh, H., Ribeiro, T., Guerreiro, P., et al. (2014). Systematic Comparison of the Effects of Alpha-Synuclein Mutations on its Oligomerization and Aggregation. *PLoS Genet.* 10, e1004741. Publisher: Public Library of Science. doi:10.1371/journal.pgen.1004741

Li, X., Dong, C., Hoffmann, M., Garen, C. R., Cortez, L. M., Petersen, N. O., et al. (2019). Early Stages of Aggregation of Engineered α-synuclein Monomers and Oligomers in Solution. *Sci. Rep.* 9, 1734. Bandiera_abtest: a Cc_license_type: cc_by Cg_type: Nature Research Journals Number: 1 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Intrinsically disordered proteins;Protein aggregation Subject_term_id: intrinsically-disordered-proteins;protein-aggregation. doi:10.1038/s41598-018-37584-6

Liwo, A., Czaplewski, C., Pillardy, J., and Scheraga, H. A. (2001). Cumulant-based Expressions for the Multibody Terms for the Correlation between Local and Electrostatic Interactions in the United-Residue Force Field. *J. Chem. Phys.* 115, 2323–2347. doi:10.1063/1.1383989

Liwo, A., Sieradzan, A. K., Lipska, A. G., Czaplewski, C., Joung, I., Żmudzińska, W., et al. (2019). A General Method for the Derivation of the Functional Forms of the Effective Energy Terms in Coarse-Grained Energy Functions of Polymers. III. Determination of Scale-Consistent Backbone-Local and Correlation Potentials in the UNRES Force Field and Force-Field Calibration and Validation. *J. Chem. Phys.* 150, 155104. doi:10.1063/1.5093015

Lorenzen, N., Nielsen, S. B., Buell, A. K., Kaspersen, J. D., Arosio, P., Vad, B. S., et al. (2014). The Role of Stable α-Synuclein Oligomers in the Molecular Events Underlying Amyloid Formation. *J. Am. Chem. Soc.* 136, 3859–3868. Publisher: American Chemical Society. doi:10.1021/ja411577t

Lucas, H., and Fernández, R. (2020). Navigating the Dynamic Landscape of Alpha-Synuclein Morphology: a Review of the Physiologically Relevant Tetrameric Conformation. *Neural Regen. Res.* 15, 407. doi:10.4103/1673-5374.265792

Lv, Z., Krasnoslobodtsev, A. V., Zhang, Y., Ysselstein, D., Rochet, J.-C., Blanchard, S. C., et al. (2015). Direct Detection of α-Synuclein Dimerization Dynamics: Single-Molecule Fluorescence Analysis. *Biophysical J.* 108, 2038–2047. Publisher: Elsevier. doi:10.1016/j.bpj.2015.03.010

Maisuradze, G. G., Senet, P., Czaplewski, C., Liwo, A., and Scheraga, H. A. (2010). Investigation of Protein Folding by Coarse-Grained Molecular Dynamics with the UNRES Force Field. *J. Phys. Chem. A* 114, 4471–4485. doi:10.1021/jp9117776

Mao, X., Ou, M. T., Karuppagounder, S. S., Kam, T.-I., Yin, X., Xiong, Y., et al. (2016). Pathological α-synuclein Transmission Initiated by Binding Lymphocyte-Activation Gene 3. *Science* 353, aah3374. Publisher: American Association for the Advancement of Science. doi:10.1126/science.aah3374

Mollenhauer, B., Cullen, V., Kahn, I., Krastins, B., Outeiro, T. F., Pepivani, I., et al. (2008). Direct Quantification of CSF α-synuclein by ELISA and First Cross-Sectional Study in Patients with Neurodegeneration. *Exp. Neurol.* 213, 315–325. doi:10.1016/j.expneurol.2008.06.004

Nguyen, P. H., Ramamoorthy, A., Sahoo, B. R., Zheng, J., Faller, P., Straub, J. E., et al. (2021). Amyloid Oligomers: A Joint Experimental/Computational Perspective on Alzheimer's Disease, Parkinson's Disease, Type II Diabetes, and Amyotrophic Lateral Sclerosis. *Chem. Rev.* 121, 2545–2647. Publisher: American Chemical Society. doi:10.1021/acs.chemrev.0c01122

Nicolaï, A., Delarue, P., and Senet, P. (2013). Conformational Dynamics of Full-Length Inducible Human Hsp70 Derived from Microsecond Molecular Dynamics Simulations in Explicit Solvent. *J. Biomol. Struct. Dyn.* 31, 1111–1126. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/07391102.2012.726190. doi:10.1080/07391102.2012.726190

Pasanen, P., Myllykangas, L., Siitonen, M., Raunio, A., Kaakkola, S., Lyytinen, J., et al. (2014). A Novel α-synuclein Mutation A53E Associated with Atypical Multiple System Atrophy and Parkinson's Disease-type Pathology. *Neurobiol. Aging* 35, e1–2180. doi:10.1016/j.neurobiolaging.2014.03.024

Petrucci, S., Ginevrino, M., and Valente, E. M. (2016). Phenotypic Spectrum of Alpha-Synuclein Mutations: New Insights from Patients and Cellular Models. *Park. Relat. Disord.* 22, S16–S20. doi:10.1016/j.parkreldis.2015.08.015

Polymeropoulos, M. H., Lavedan, C., Leroy, E., Ide, S. E., Dehejia, A., Dutra, A., et al. (1997). Mutation in the α-Synuclein Gene Identified in Families with Parkinson's Disease. *Science* 276, 2045–2047. doi:10.1126/science.276.5321.2045

Reynolds, D. (2009). "Gaussian Mixture Models," in *Encyclopedia of Biometrics*. Editors SZ Li and A Jain (Boston, MA: Springer US), 659–663. doi:10.1007/978-0-387-73003-5_196

Rojas, A., Maisuradze, N., Kachlishvili, K., Scheraga, H. A., and Maisuradze, G. G. (2017). Elucidating Important Sites and the Mechanism for Amyloid Fibril Formation by Coarse-Grained Molecular Dynamics. *ACS Chem. Neurosci.* 8, 201–209. Publisher: American Chemical Society. doi:10.1021/acschemneuro.6b00331

Rojas, A. V., Maisuradze, G. G., and Scheraga, H. A. (2018). Dependence of the Formation of Tau and Aβ Peptide Mixed Aggregates on the Secondary Structure of the N-Terminal Region of Aβ. *J. Phys. Chem. B* 122, 7049–7056. Publisher: American Chemical Society. doi:10.1021/acs.jpcb.8b04647

Roostaee, A., Beaudoin, S., Staskevicius, A., and Roucou, X. (2013). Aggregation and Neurotoxicity of Recombinant α-synuclein Aggregates Initiated by Dimerization. *Mol. Neurodegener.* 8, 5. doi:10.1186/1750-1326-8-5

Sahu, K. K., Woodside, M. T., and Tuszynski, J. A. (2015). α-Synuclein Dimer Structures Found from Computational Simulations. *Biochimie* 116, 133–140. doi:10.1016/j.biochi.2015.07.011

Schweighauser, M., Shi, Y., Tarutani, A., Kametani, F., Murzin, A. G., Ghetti, B., et al. (2020). Structures of α-synuclein Filaments from Multiple System Atrophy. *Nature* 585, 464–469. Bandiera_abtest: a Cg_type: Nature Research Journals Number: 7825 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Cryoelectron microscopy;Neurodegeneration Subject_term_id: cryoelectron-microscopy;neurodegeneration. doi:10.1038/s41586-020-2317-6

Sieradzan, A. K., Korneev, A., Begun, A., Kachlishvili, K., Scheraga, H. A., Molochkov, A., et al. (2021). Investigation of Phosphorylation-Induced Folding of an Intrinsically Disordered Protein by Coarse-Grained Molecular Dynamics. *J. Chem. Theory Comput.* 17, 3203–3220. doi:10.1021/acs.jctc.1c00155

Soto, C. (2003). Unfolding the Role of Protein Misfolding in Neurodegenerative Diseases. *Nat. Rev. Neurosci.* 4, 49–60. doi:10.1038/nrn1007

Spillantini, M. G., Schmidt, M. L., Lee, V. M.-Y., Trojanowski, J. Q., Jakes, R., and Goedert, M. (1997). α-Synuclein in Lewy Bodies. *Nature* 388, 839–840. doi:10.1038/42166

Stefanis, L. (2012). -Synuclein in Parkinson's Disease. *Cold Spring Harb. Perspect. Med.* 2, a009399. doi:10.1101/cshperspect.a009399

Tanudjojo, B., Shaikh, S. S., Fenyi, A., Bousset, L., Agarwal, D., Marsh, J., et al. (2021). Phenotypic Manifestation of α-synuclein Strains Derived from Parkinson's Disease and Multiple System Atrophy in Human Dopaminergic Neurons. *Nat. Commun.* 12, 3817. doi:10.1038/s41467-021-23682-z

Teil, M., Arotcarena, M. L., Faggiani, E., Laferriere, F., Bezard, E., and Dehay, B. (2020). Targeting α-synuclein for PD Therapeutics: A Pursuit on All Fronts. *Biomolecules* 10, 391. Number: 3 Publisher: Multidisciplinary Digital Publishing Institute. doi:10.3390/biom10030391

Tosatto, L., Horrocks, M. H., Dear, A. J., Knowles, T. P. J., Dalla Serra, M., Cremades, N., et al. (2015). Single-molecule FRET Studies on Alpha-Synuclein Oligomerization of Parkinson's Disease Genetically Related Mutants. *Sci. Rep.* 5, 16696. doi:10.1038/srep16696

Touw, W. G., Baakman, C., Black, J., te Beek, T. A. H., Krieger, E., Joosten, R. P., et al. (2015). A Series of PDB-Related Databanks for Everyday Needs. *Nucleic Acids Res.* 43, D364–D368. doi:10.1093/nar/gku1028

Trinkaus, V. A., Riera-Tur, I., Martínez-Sánchez, A., Bäuerlein, F. J. B., Guo, Q., Arzberger, T., et al. (2021). *In Situ* architecture of Neuronal α-Synuclein Inclusions. *Nat. Commun.* 12, 2110. doi:10.1038/s41467-021-22108-0

Tripathi, T. (2020). A Master Regulator of α-Synuclein Aggregation. *ACS Chem. Neurosci.* 11, 1376–1378. doi:10.1021/acschemneuro.0c00216

Tuttle, M. D., Comellas, G., Nieuwkoop, A. J., Covell, D. J., Berthold, D. A., Kloepper, K. D., et al. (2016). Solid-state NMR Structure of a Pathogenic Fibril of Full-Length Human α-synuclein. *Nat. Struct. Mol. Biol.* 23, 409–415. doi:10.1038/nsmb.3194

Volles, M. J., Lee, S.-J., Rochet, J.-C., Shtilerman, M. D., Ding, T. T., Kessler, J. C., et al. (2001). Vesicle Permeabilization by Protofibrillar α-Synuclein: Implications for the Pathogenesis and Treatment of Parkinson's Disease. *Biochemistry* 40, 7812–7819. Publisher: American Chemical Society. doi:10.1021/bi0102398

Wang, W., Perovic, I., Chittuluru, J., Kaganovich, A., Nguyen, L. T. T., Liao, J., et al. (2011). A Soluble α-synuclein Construct Forms a Dynamic Tetramer. *Proc. Natl. Acad. Sci. U.S.A.* 108, 17797–17802. doi:10.1073/pnas.1113260108

Waxman, E. A., Mazzulli, J. R., and Giasson, B. I. (2009). Characterization of Hydrophobic Residue Requirements for α-Synuclein Fibrillization. *Biochemistry* 48, 9427–9436. doi:10.1021/bi900539p

Winner, B., Jappelli, R., Maji, S. K., Desplats, P. A., Boyer, L., Aigner, S., et al. (2011). *In Vivo* demonstration that α-synuclein Oligomers Are Toxic. *Proc. Natl. Acad. Sci. U.S.A.* 108, 4194–4199. ISBN: 9781100976105 Publisher: National Academy of Sciences Section: Biological Sciences. doi:10.1073/pnas.1100976108

Wright, P. E., and Dyson, H. J. (1999). Intrinsically Unstructured Proteins: Re-assessing the Protein Structure-Function Paradigm. *J. Mol. Biol.* 293, 321–331. doi:10.1006/jmbi.1999.3110

Yamauchi, M., and Okumura, H. (2021). Dimerization of α-Synuclein Fragments Studied by Isothermal-Isobaric Replica-Permutation Molecular Dynamics Simulation. *J. Chem. Inf. Model.* 61, 1307–1321. doi:10.1021/acs.jcim.0c01056

Yu, J., Malkova, S., and Lyubchenko, Y. L. (2008). α-Synuclein Misfolding: Single Molecule AFM Force Spectroscopy Study. *J. Mol. Biol.* 384, 992–1001. doi:10.1016/j.jmb.2008.10.006

Zarranz, J. J., Alegre, J., Gómez-Esteban, J. C., Lezcano, E., Ros, R., Ampuero, I., et al. (2004). The New Mutation, E46K, of α-synuclein Causes Parkinson and Lewy Body Dementia. *Ann. Neurol.* 55, 164–173. doi:10.1002/ana.10795

Zhang, Y., Hashemi, M., Lv, Z., Williams, B., Popov, K. I., Dokholyan, N. V., et al. (2018). High-speed Atomic Force Microscopy Reveals Structural Dynamics of α-synuclein Monomers and Dimers. *J. Chem. Phys.* 148, 123322. Publisher: American Institute of Physics. doi:10.1063/1.5008874

Zhou, L., and Kurouski, D. (2020). Structural Characterization of Individual α-Synuclein Oligomers Formed at Different Stages of Protein Aggregation by Atomic Force Microscopy-Infrared Spectroscopy. *Anal. Chem.* 92, 6806–6810. doi:10.1021/acs.analchem.0c00593

Zhou, R., Maisuradze, G. G., Suñol, D., Todorovski, T., Macias, M. J., Xiao, Y., et al. (2014). Folding Kinetics of WW Domains with the United Residue Force Field for Bridging Microscopic Motions and Experimental Measurements. *Proc. Natl. Acad. Sci. U.S.A.* 111, 18243–18248. doi:10.1073/pnas.1420914111

Conflict of Interest: AR was employed by Schrödinger, Inc.

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

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

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