In memory of Professor Marek Cieplak

REVIEW

Recent Advances in Mapping Protein Self-Assembly and Aggregation for Common Proteinopathies

S. Bhattacharya and D. Thompson*

Department of Physics, Bernal Institute, University of Limerick, V94 T9PX, Ireland

Doi: 10.12693/APhysPolA.145.S37 *e-mail: damien.thompson@ul.ie

The accumulation of abnormal conformation by brain peptides and proteins followed by their aberrant self-assembly into insoluble aggregates is the hallmark of "proteinopathies", common across many neurodegenerative disorders. Experiments suggest that soluble low-molecular-weight oligomers formed in the early stages of assembly are neurotoxic, and hence, drug targets. However, the inherent polymorphic nature of these short-lived oligomers restricts their experimental characterisation in pathological protein self-assembly pathways. Here, we shed light on the latest contributions from atomic-level modelling techniques, such as computer-based molecular dynamics simulations in bulk solution and on surfaces, which are guiding experimental efforts to map early stages of protein self-assembly in common proteinopathies, including Alzheimer's and Parkinson's diseases, which could potentially aid in molecular-level understanding of disease pathologies. Predictive computational modelling of amyloid-β and tau protein assemblies in Alzheimer's disease and α-synuclein protein assemblies in Parkinson's disease highlights the potential for identification and characterisation of new therapeutic targets for currently incurable neurodegeneration.

topics: proteinopathies, self-assembly, computational modelling, molecular dynamics simulations

1. Introduction

Deposition of protein fibrillar aggregates is a characteristic shared by > 50 human diseases [1]. Pathological protein self-assembly with the formation of inclusion bodies, such as fibrils, is the hallmark of many neurodegenerative disorders (ND) [2], or broadly, "proteinopathies". NDs are a heterogeneous group of lethal brain disorders that may be characterised by symptomatic gradual decline of the structure and function of central and peripheral nervous systems [3]. They share a significant Global Burden of Disease (GBD) [4], with World Health Organization (WHO) projections that dementia will account for $>$ 1% of total deaths by 2030 [5]. NDs, including Alzheimer's disease (AD), Parkinson's disease (PD), Huntington's disease (HD) [6], prion diseases, amyotrophic lateral sclerosis (ALS), and other systemic amyloidosis diseases [3], exhibit distinct aetiologies but share common pathologies. These disorders could be characterised by amyloidosis or the production of amyloids, where abnormal protein conformations form through spontaneous misfolding and self-assembly starting from their intrinsically disordered native state (intrinsically disordered proteins, IDPs) [7]. AD and PD are the most common proteinopathies [8]. Currently, only five Food and Drug Administration (FDA)-approved drugs are available to treat cognitive symptoms of AD or slow its progression by removal of brain amyloid [9–11], and a handful of FDA-approved treatment options address the motor symptoms associated with PD [12]. Yet, to date, there exists no clinically effective disease-modifying strategy for AD and PD multifactorial diseases, creating a massive burden on the management of symptoms and patient care [13, 14]. To translate disease-modifying strategies into effective clinical targets, urgent re-evaluation of current therapeutic and molecular targets is required.

The development of effective treatment for AD and PD is hampered by an insufficient understanding of the events that trigger the self-assembly of the monomeric IDPs into higher-order assemblies and, eventually, fibrils [15]. Knowledge to date is summarised in Fig. 1 (see also [16]), showing the potential molecular processes from misfolded monomeric proteins to self-assembled aggregates. In addition, other mechanisms of amyloid toxicity are also proposed from a misfolded monomer [6, 17, 18], including investigations of elastic and thermodynamic properties of amyloid-β and α-synuclein fibrils from molecular simulations to understand experimental nanomechanical characterisation techniques [19]. Mechanical properties of fibrillar assemblies can also serve as a diagnostic fingerprint for potential applications or pathology [20–22], supported by co-development

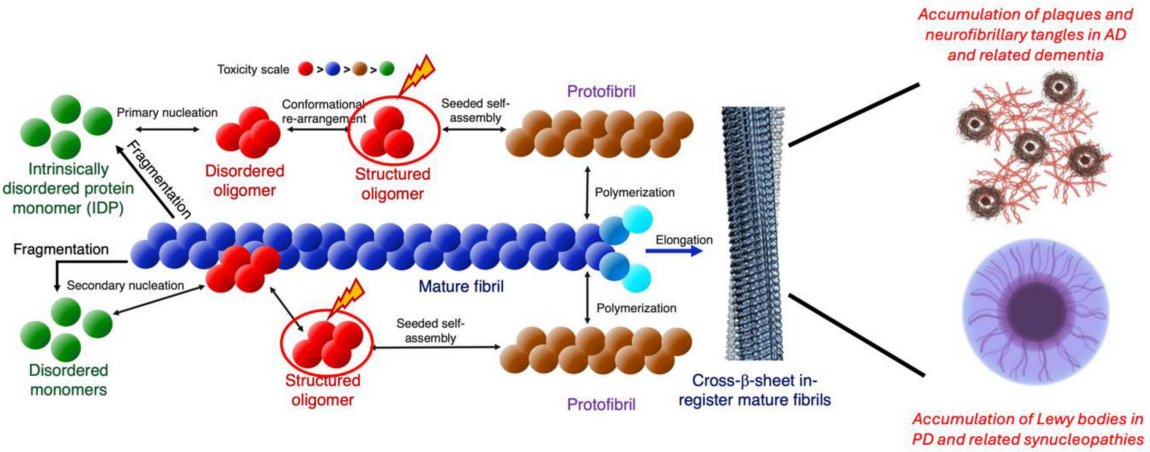


Fig. 1. Known molecular mechanisms underlying amyloid pathogenesis to self-assembled insoluble aggregates of mature fibrils further elongated to cross-β sheet fibrils and eventually plaques and neurofibrillary tangles in AD and related dementia, and Lewy bodies in PD and related synucleopathies. IDPs of Aβ, tau, and αS may undergo primary or predominantly secondary nucleation [16] to form oligomers, which are the critical nuclei for the growth of seeds into protofibrils and fibrils, which, through further polymerisation, lead to mature fibrils. Oligomers and mature fibrils may further undergo fragmentation or dissociate to form new seeds.

of reliable measurement techniques and predictive modelling [23–25]. Amyloid-β (Aβ) [26] and tau [27] form plaques and neurofibrillary tangles in AD and related dementia, and α-synuclein (αS) [28] form Lewy bodies in PD and other synucleopathies. The neurotoxicity in AD and PD is now generally attributed to low-molecular-weight aggregates or small oligomers rather than large amyloid fibrils and plaques [29, 30]. However, detection and isolation of these soluble oligomers has proven difficult due to their short lifetimes, low concentration, and structural heterogeneity [29] in solution, unless they are artificially engineered to be structurally stable [31].

The relationship between aggregation rates of amyloidogenic peptides and their morphological changes over time is important not only to comprehensively map the pathological protein self-assembly, but also to characterise the shapes and sizes of the small population of cytotoxic, short-lived, and structurally diverse misfolded oligomers to design new potential drug targets [32]. To detect and measure the metastable oligomers in physiological solution, several *in vitro* and *in vivo* techniques have been employed in the past, generating mainly only qualitative or semi-quantitative data [32], and therefore computational molecular modelling and computer simulations have played a major role in guiding experiments on protein self-assembly. The presence of biological and non-biological surfaces or surface–water interfaces is well known to mediate pathogenic protein self-assembly [33]. The interaction between biological lipid membranes and pathogenic peptides in the context of fibril formation has been extensively studied [34–36]. The interaction between non-biological surfaces and amyloid aggregation depends on the nature of the surfaces, which can play a crucial role in either catalysing or inhibiting the aggregation of amyloid proteins and is an interesting area that is only beginning to be explored [37, 38]. In this mini-review, we provide a perspective on the latest findings and recent advances made in mapping pathological protein self-assembly. We highlight the importance of computer-based molecular modelling and simulations supported by experimental investigations in bulk solution and on biological and non-biological surfaces to reveal molecular-level details of the assembly mechanisms, identifying potential early stages of self-assembly, the role of secondary structures, and routes to resist toxic aggregation with potential therapeutic intervening targets focussing on the proteins responsible for AD and PD pathologies, Aβ, tau, and αS.

2. Demystifying stages of pathogenic protein self-assembly from molecular simulations

As discussed above, it is experimentally very difficult or impossible to characterise thermally accessible states of Aβ, tau, and αS protein assemblies in AD and PD pathophysiology, so molecular modelling with appropriate benchmarking and experimental validation can help to identify the thermodynamic driving force behind specific protein morphologies formed in the self-assembly pathway, and also to estimate their kinetics (i.e., how fast these morphologies form) of formation and dissociation [39]. Computational modelling, in particular molecular dynamics (MD) simulations, can predict dynamic local and long-range interactions driving heterogeneous assemblies [40, 41]. Predictive models from MD may also provide mechanistic insights

into previously unknown self-assembly features at different stages of pathological protein aggregation, including those that could be validated, directly or indirectly, by experiments [42].

One of the earliest instances of very short MD simulations provided mechanistic insights into amyloidogenic misfolding that is involved in the multimerisation of PrP_{Sc} (pathogenic prion) [43]. The only information available from experiments was that the conformational conversion of PrP_C (cellular prion) to PrP_{Sc} occurred at low pH. Within ten nanoseconds (ns) of molecular dynamics, the simulations mapped conformational shifts in the N-terminal region that were mainly due to the breaking of charge-stabilised hydrogen bonded interactions at low pH, which was later confirmed by amide-proton exchange nuclear magnetic resonance (NMR) experiments [44]. The computational predictive power of molecular simulations is ever-increasing due to improvements in hardware and increased accessibility of high-performance computing platforms coupled with software developments in particular advanced sampling methods [45, 46] and the latest improved force fields [47, 48] and water models [7] for MD, allowing to reach extended timescales to map physically realistic, and biologically relevant, protein aggregation pathways.

3. Atomic models of pathological protein self-assembly

3.1. Modelling self-assembly of Aβ protein in AD

The mechanisms of initial misfolding and aggregation of Aβ in AD have been studied extensively by MD simulations supported by single-molecule experimental techniques [49–51]. As Aβ dimers were identified as the smallest toxic oligomers that could potentially assemble into neurotoxic protofibrils [52], recent long, multi-microsecond MD studies have investigated their detailed assembly to reveal differences in dimer morphologies from the U-shaped and S-shaped fibrillar morphologies indicating significant conformational re-arrangements in aggregation from small toxic oligomers to fibrils [53, 54]. These atomic scale insights are not readily available from high-resolution experimental techniques due to the broad distribution of short-lived dimeric shapes and their rapid self-assembly into higher-order structures. Similarly, recent transition path theory (TPT) network and Markov state models (MSM) based on MD-generated ensembles of dimers and higher-order oligomers have shown the predominance of Aβ oligomer shapes in directing self-assembly propensities; the compact metastable dimer matching to the oligomer distribution has been observed experimentally and may be more toxic than the extended dimers that self-assemble into larger fibrils [54, 55].

A comprehensive study involving microseconds scale MD simulations of ten different protein force fields and cross-correlation network analysis benchmarked by experimental NMR revealed the very nascent aggregation-favouring and aggregation-impeding propensities of fully folded, unfolded, and partially folded helical states of both $A\beta_{1-42}$ and αS_{1-140} (see Fig. 2a). The fully folded helical states optimise the direct intra-protein hydrophobic contacts between the termini and the central hydrophobic domain (CHD) of both proteins which resist aggregation (see Fig. 2b) [56], while the partially folded helical states may promote initial self-assembly by long-range allosteric coupling between the terminal residues and the CHD (Fig. 2c) [57].

In another recent work [58], extensive MD simulations predicted the molecular signatures of the difference in aggregation profiles visualised by atomic force microscopy (AFM) experiments on preformed oligomers in LS-shaped fibril fold (profibrillar 12-mers) between peptides $A\beta_{40}$ and $A\beta_{42}$ that may account for the higher pathogenicity of $A\beta_{42}$ in AD (Fig. 3). Modelling the orientation of both peptide assemblies on a single layer of graphene as the interface between graphene and water is an ideal platform to study peptide assemblies with AFM. From oligomer–graphene binding energies, the models predicted that amyloid beta undergoes chain elongation along the graphene sheet (orientation III, Fig. 3a, b). Predictions of oligomer model height profiles on top of graphene and hydrogen bond (H-bond) occupancies in three dimers of hexamer (dimer at one end of the oligomer, denoted as E1, dimer in the centre of oligomer, denoted as C, and dimer at the other end of oligomer, denoted as E2) forming two layers of the 12-mer and validated from AFM maps (Fig. 3c) revealed unidirectional growth profile for $A\beta_{40}$ and bidirectional growth for $A\beta_{42}$ at the graphene–water interface (Fig. 3d) that may explain the highly aggregation-prone nature and toxicity of $A\beta_{42}$.

3.2. Modelling self-assembly of tau protein in AD

The microtubule-associated protein tau (MAPT or simply τ) [59] is implicated in the pathogenesis of AD. Tau is an IDP responsible for the polymerisation and stabilisation of microtubules and has two major domains: (i) the projection domain, which includes the N-terminal and points away from the microtubule surface, and (ii) the C-terminal domain, which binds to microtubules [60]. The polymorphic nature of hTau40 [61] has precluded attempts to resolve its full atomic structure experimentally, and recent cryo-EM structures of tau filaments capture the structural polymorphism at fibrillar level with paired helical filaments (PHFs) [62–65], straight filaments (SF) [62, 63], narrow Pick's filament (NPF) [66] in frontotemporal

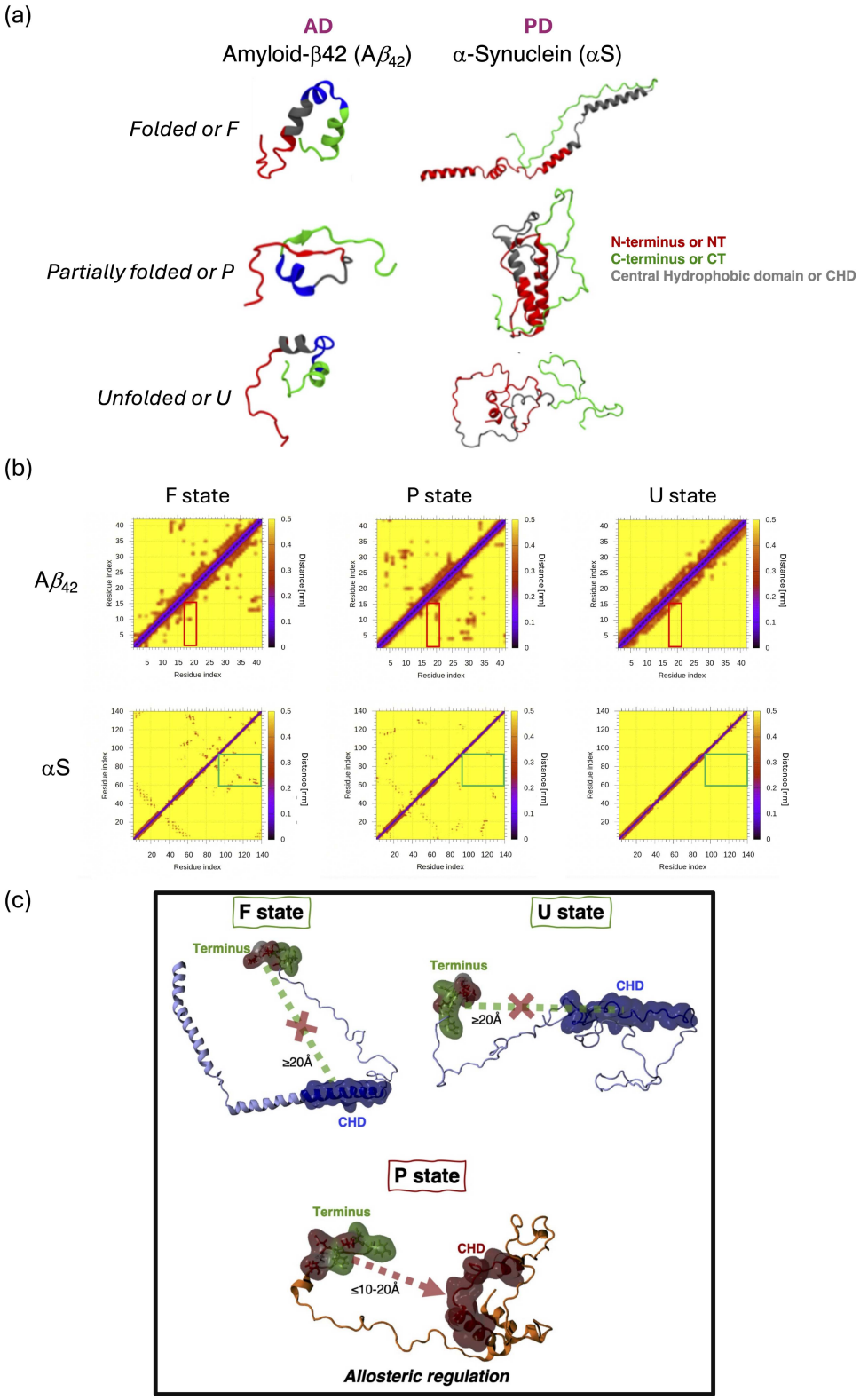
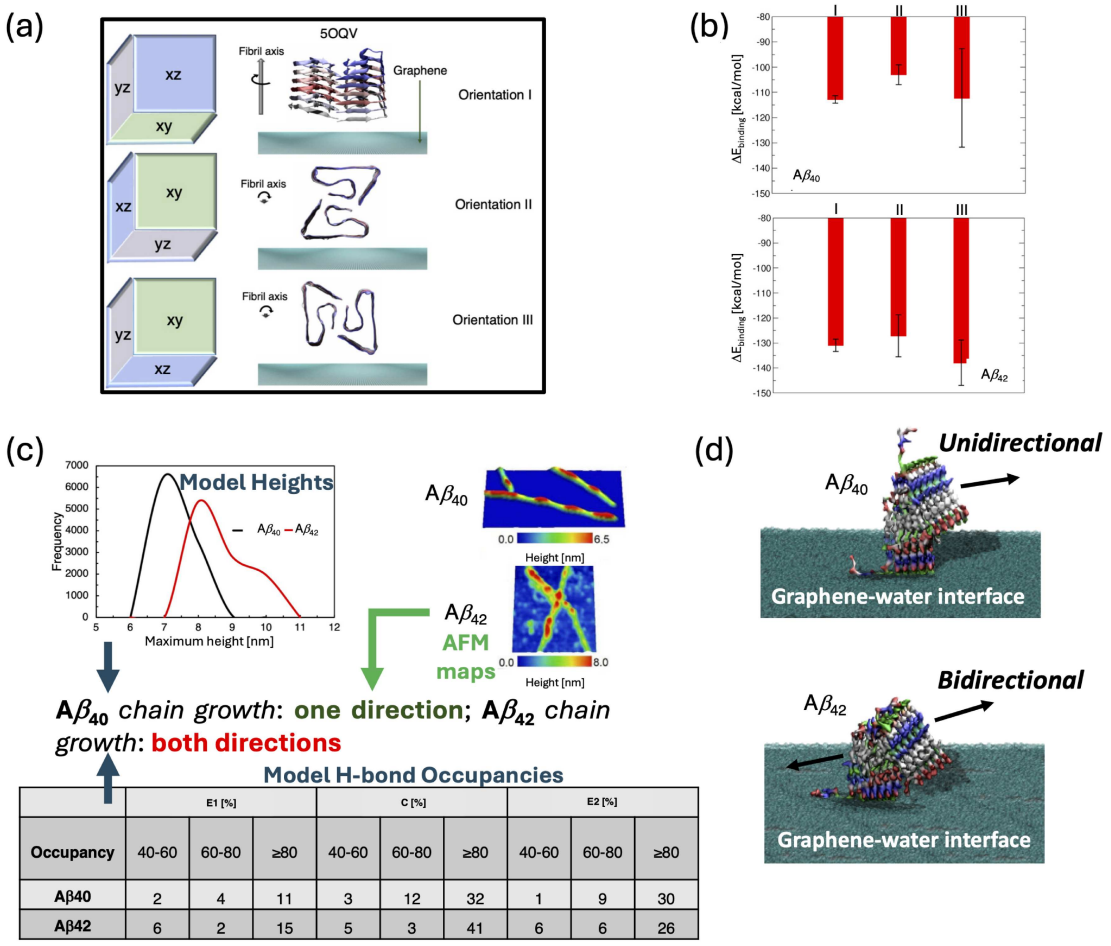


Fig. 2. (a) Representative folded (F), partially folded (P), and unfolded (U) helical state structures of Aβ_{42} in AD and αS in PD sampled from $\sim$ 36 microseconds MD simulations of helical structures employing ten alternative force fields and water models combinations. (b) Residue–residue contact maps showing the helical F state stabilising the interactions between the N-terminus and CHD of Aβ_{42}, and the C-terminus and CHD of αS that inhibit aggregation and missing in the P and the U states. (c) Long-range ($\leq$ 20 Å) allosteric regulation of the CHD by the termini of helical peptides Aβ_{42} and αS in their P state, which makes them aggregation-prone.



	E1 [%]			C [%]			E2 [%]		
Occupancy	40-60	60-80	≥80	40-60	60-80	≥80	40-60	60-80	≥80
Aβ40	2	4	11	3	12	32	1	9	30
Aβ42	6	2	15	5	3	41	6	6	26

Fig. 3. (a) Different starting orientations of Aβ preformed oligomers in fibril fold on top of a graphene sheet and at the graphene–water interface for running MD simulations. (b) Oligomer–graphene binding energies, showing that orientation III with a fibril axis along the graphene sheet is most favourable. (c) Model prediction of maximum oligomer heights on top of graphene and comparison of H-bond occupancies [%] between two ends of oligomers and the centre of oligomer supports visualisation of AFM maps to reveal Aβ_{40} chain elongation in one direction and Aβ_{42} chain elongation in both directions. (d) Depiction of unidirectional and bidirectional chain growth of Aβ_{40} and Aβ_{42}, respectively, at the graphene–water interface.

dementia, and very recently non-helical filaments [67] in AD, leaving out details of the morphological features of oligomer assemblies that could not be characterised experimentally. Post-translational modifications, including hyper-phosphorylation of tau, trigger its self-assembly by decreasing its microtubule-stabilising ability [60] and may act as an important target for disease-modifying therapies [68]. In addition to the formation of PHFs in AD, hyper-phosphorylated tau aggregates may form neurofibrillary or gliofibrillary tangles commonly known as "tauopathies" [69].

In this regard, MD simulations have helped reveal the dynamics of monomer misfolding and dimerisation of the four microtubule-binding (MTB) repeat domains (R1–R4) constituting the core of tau fibrils with R3 monomers forming β-sheets while both R2 and R3 repeats aggregate into metastable β-sheet-rich dimers, especially residues composed of the PHF6 hexapeptide of R3 [70]. A more recent MD simulation study predicted the aggregation propensity of the repeat domains of tau peptide where the R3–R4 (residues 306–378) monomer may form transient β-hairpins within the R3 repeat and between the R3 and R4 repeats in bulk solution, but spontaneous β-sheets insertion was not observed in modelling on the membrane surface [71]. MD simulations have also been coupled with MSM and TPT models to uncover the tau misfolding kinetics and structural features of the key R3 repeat domain at the atomic scale, where a critical intermediate state was noted for the formation of two target β-sheet structures [72].

A number of recent simulations have focused on investigating the morphologies of tau oligomers and fibrils. For instance, it was predicted through atomistic MD simulations that the C-shaped conformation of the fibril core is retained only by the R3–R4 repeat domains, while the R1–R2, first and second repeat domains, tend to have a linear

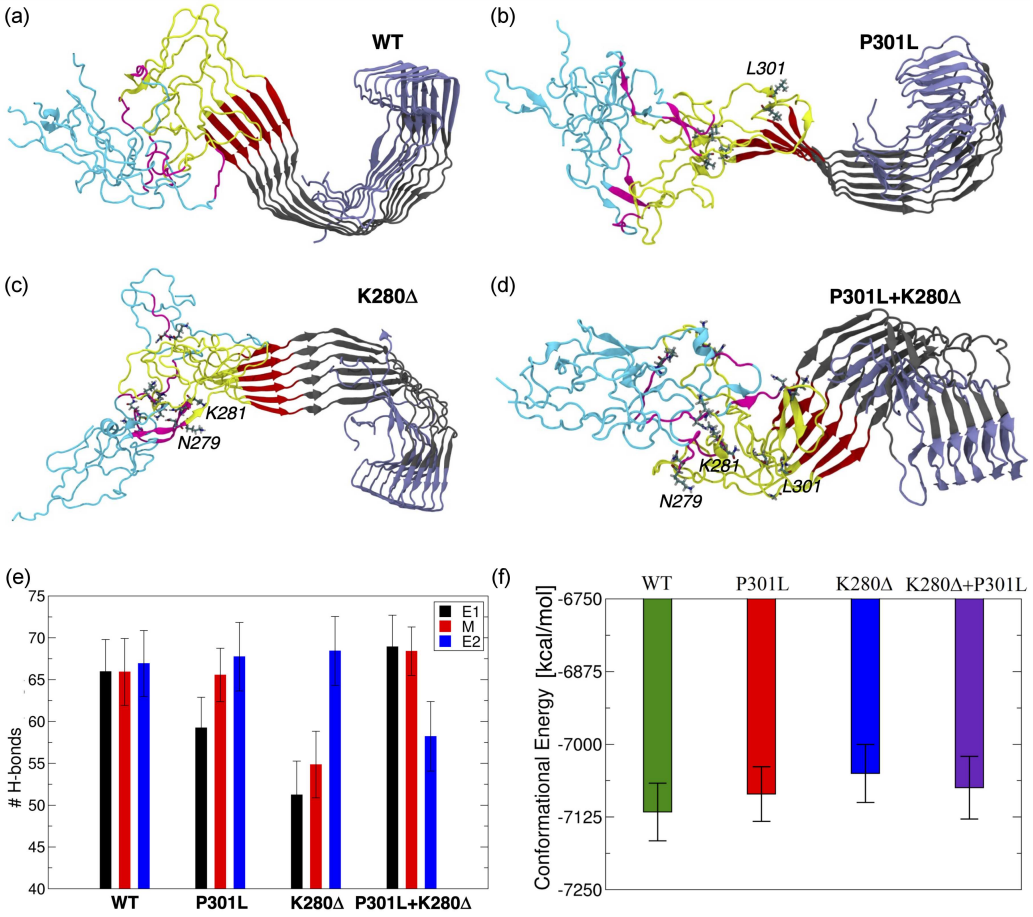


Fig. 4. (a–d) Final conformation of tau microtubule-binding domain (MBD) pro-fibrillar structure formed after 1 μs each of free molecular dynamics (MD) in water under physiological conditions for (a) WT, (b) P301L, (c) K280Δ, (d) and double mutant P301L + K280Δ. (e) Comparison of the average number of H-bonds (computed over the last 250 ns dynamics) between two monomer chains at one end (E1), central or middle dimer (M), and the other end dimer (E2) of each hexamer tau MBD model. (f) Comparison of conformational energies for the crystalline domain of the tau MBD hexamers.

shape [73]. This model finding of a C-shaped fibrillar core formed by R3 and R4 domains was later confirmed by experimental cryo-EM structures of tau protofibrillar straight filaments (SF) [62]. Multiscale MD simulations (both atomistic and coarse-grained MD) to explore the conformational features of hyper-phosphorylation on tau repeat domains (R1–R4) showed that hyper-phosphorylation exposes the repeats to bulk solution, which could further promote tau filament self-assembly [74]. The latest study also proposed MD models of the hyper-phosphorylated NPF fibril repeat domains at three experimentally observed phosphorylated Ser sites (S262, S324, and S356) in the MTB domain [75]. Mutations E264G and D358G were engineered on the wild-type (WT) narrow Pick's filament to understand the function of E264 and D358 residues on the local conformations and compare them with the fibrillar architecture of hyper-phosphorylated NPF from microseconds scale atomistic MD simulations. The models revealed that the mutant and hyper-phosphorylated NPF showed a major morphological departure from the WT narrow Pick's filament and that the repeat-specific sequence of the C-terminal hexapeptide strongly guides and influences the conformational properties of the PGGG motif that flanks the hexapeptide in tau [75].

Through four-microsecond atomistic MD simulations, we recently modelled the driving forces behind the assembly of pro-fibrillar hexameric oligomers of two familial mutations within the MTB repeat R1–R4 domains of tau, namely the P301L substitution and the deletion mutation K280Δ (both known to cause frontotemporal dementia). The models identify their pro-aggregation capability due to their innate core packing by R2 and R3 and the overall stability of hexamers in their fibrillar fold that facilitates pro-fibrillar elongation compared to the WT and a control double mutant, P301L + K280Δ of the MTB R1–R4 domain (Fig. 4a–d) [76]. Based on H-bond networks, the models predicted H-bond strengths with regard to three dimers in the hexamer: the dimer at one end of the fibril (denoted as E1), the dimer in the middle of

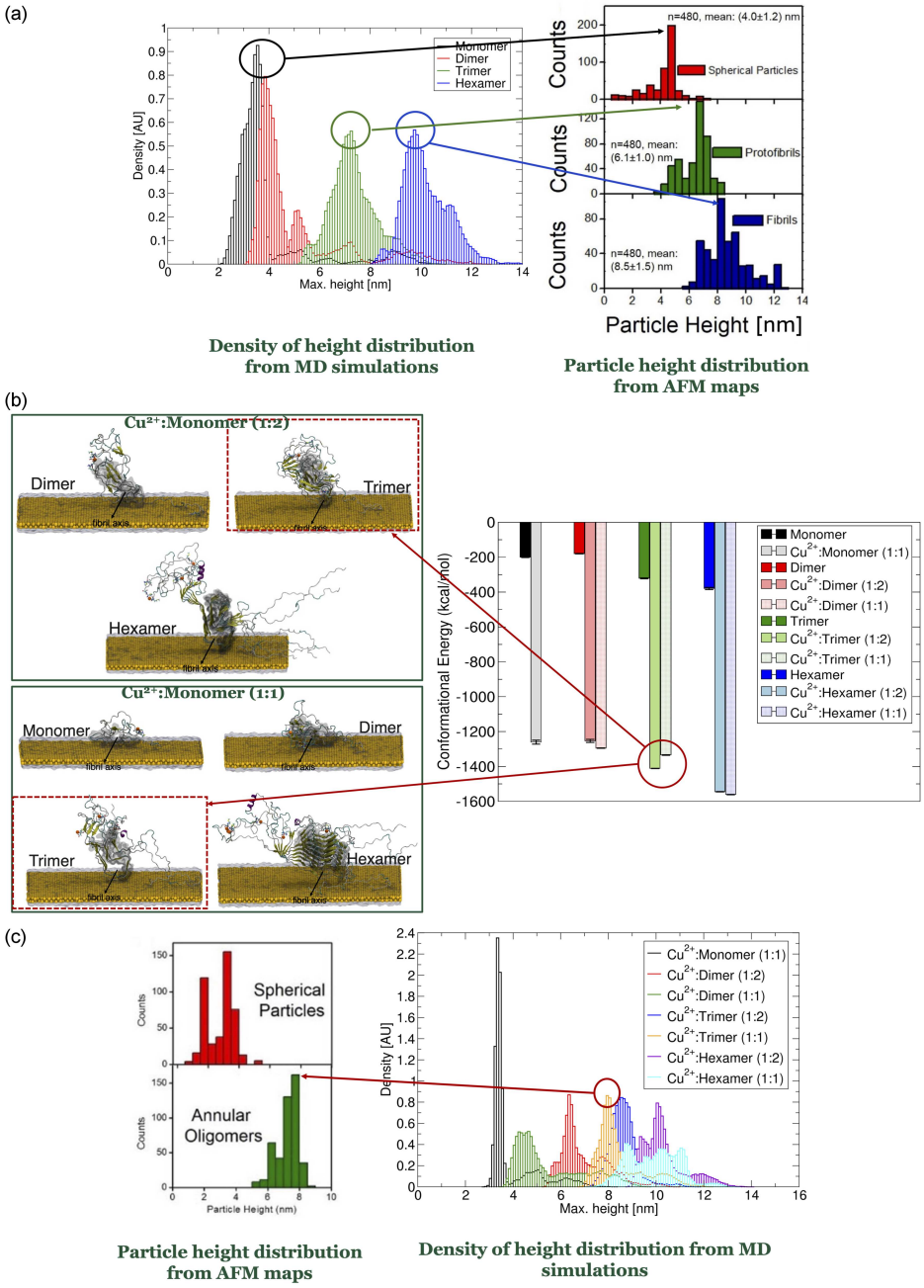


Fig. 5. (a) Density of height distribution profile of αS assemblies without Cu(II) at the gold–water interface from MD simulations matched up with particle height distributions from AFM maps on ultra-flat gold predicts αS trimer as the critical nucleus for fibril growth. (b) Representative snapshots of Cu (II)-bound monomer and oligomer constructs at different Cu(II) concentrations (low concentration — Cu^{2+}: monomer = 1 : 2 and high concentration — Cu^{2+} : monomer = 1 : 1). Increase in Cu(II) concentration do not sample more stable trimers, as confirmed by their conformational energies. (c) Model predicted trimer height distributions at higher Cu(II) concentrations on the gold–water interface confirmed to be annular oligomers from AFM particle height profile on ultra-flat gold.

the fibril (denoted as M), and the dimer at the other end of the fibril (denoted as E2). We noted almost no difference in H-bond strengths of E1 and E2 for WT tau, indicating that the WT hexamers may not undergo pro-fibrillar growth in one direction. The maximum difference in H-bond strength of E1 and E2 is observed for K280Δ followed by P301L + K280Δ, with P301L showing the least difference (Fig. 4e), which supported previous experimental observations that K280Δ mutation leads to enhanced overall aggregation kinetics with increased nucleation and elongation rates, while the P301L mutation leads to a stunted fibril growth rate compared to K280Δ. The models predicted that the

thermodynamic stabilities of hexamers in the fibrillar fold and core packing of the tau crystalline domain are significantly altered by the missense mutations, P301L and K280Δ, indicating they may be more oligomer-like (Fig. 4f).

3.3. Modelling self-assembly of αS protein in PD

Several lines of evidence have shown that β-sheet-rich αS oligomers may trigger neurotoxicity mainly by disrupting membrane integrity, including impairment of protein degradation and function of mitochondria and endoplasmic reticulum [77]. In a previous work [78], we used extensive MD simulations to model the location of several hotspots in the hydrophobic segment (residues 71–82) of non-amyloid β component (NAC region) fibrils that initiates αS self-assembly, which in both termini of NAC could change the populations of different fold morphologies adopted by NAC. The models predicted that at a lower temperature, both WT and mutant αS are sensitive to the solution environment, including the physiological salt concentration, which decreases the stability of WT NAC fibrils and may shift the relative stability of different NAC mutants. The models provide new insights into the polymorphic conformational states of αS fibrils to help predict the binding sites of new and existing protective inhibitors.

A recent MD study of full-length αS monomer misfolding and dimerisation revealed that both monomers and dimers mainly adopt disordered conformations with partial helices around the N-terminus, which is known to bind lipids and form α-helices [79]. β-sheets were mainly formed in the N-terminal tail and the NAC region, with the C-terminus remaining mostly unstructured. Further, dimerisation enhanced the β-sheet content with a subsequent decrease in disorder, with the interaction of the C-terminus with the N-terminal tail and NAC regions indicating the prevention of αS self-assembly [80]. Multi-scale MD was also recently used in conjunction with NMR and cross-linking mass spectrometry (XLMS) to probe the interactions of αS with anionic lipid cellular membrane [81]. The computational and experimental models reveal a break in the helical structure of the NAC region of αS that possibly promotes oligomer formation. Specifically, liposome-bound αS showed β-strand formation in the NAC region, and MSM models indicated a membrane-interacting αS mechanism *via* the dynamic helix break in the NAC region for pathogenesis in PD. To identify the cellular lipid membrane-mediated polymorphic folds of αS fibrils, six structures of αS fibril–lipid complexes were identified with cryo-EM, and the lipid–fibril interactions were revealed using MD simulations along with solid-state NMR (ss-NMR) spectroscopy [82]. The models revealed that phospholipids promote an unusual arrangement of protofilaments, which fill the fibril central cavities, identifying a potential mechanism for the neurotoxicity in PD by fibril-induced lipid extraction.

In recent work [83], we modelled different oligomeric assemblies (monomers, dimers, trimers, and hexamers) of the αS protein at low and high copper (Cu(II)) concentrations, because copper is one of the metals found in high concentration in the post-mortem PD patient brains. MD simulations were performed at the gold–water interface, as αS aggregates were visualised and quantified on ultra-flat gold in water through AFM (Fig. 5) [83]. Our model distribution of density of the heights of assemblies in a Cu(II)-free environment predicted proximity of αS monomer and dimer with spherical particles measured in AFM, trimer with the protofibrillar fold in AFM and hexamer having a fibrillar fold (see Fig. 5a). So, we propose that trimers are the minimal critical nucleus for elongation of αS protofibrils at the gold–water interface. Our simulations show that there are significantly tighter assemblies with increasing concentrations of Cu(II), as seen from the assembly height distributions at the gold–water interface and their conformational energy profiles, except for the trimers for which we noted that increased copper concentration does not make the trimers thermodynamically more stable (Fig. 5b). At low Cu(II) concentration, the trimer retains their assembly fold, but at higher Cu(II) concentration, the trimer shifts to a different conformation indicative of an atypical fold, which was confirmed from AFM particle heights to be annular-shaped oligomers corresponding to our model heights of trimers at high copper concentration (Fig. 5c). We propose such highly toxic annular oligomers as potential drug targets for treating PD.

3.4. Modelling the aggregation-resistant αS helical tetramer protein in PD

The latest experimental findings by a number of groups have proposed that αS may exist as α-helically folded tetramer that resists further aggregation under normal physiological conditions [84, 85]. There are many contradictory viewpoints on whether αS is a predominantly disordered cytosolic monomer that is aggregation-prone [86], as long understood [87], or a cytosolic α-helically folded tetramer that is aggregation-impeding, as recently discovered [88–90]. It is now believed that the unfolded monomeric and helically folded tetrameric states may be in dynamic equilibrium with each other [91, 92], as evident from the familial PD causing missense mutations which shifted the tetramers to pro-aggregating monomers precipitating neurotoxicity by decreasing αS solubility [93]. It was also shown that homologous E→K mutations destabilise αS multimers (including helical tetramers) and induce monomer aggregation at membranes to form vesicle-rich inclusions [89, 94].

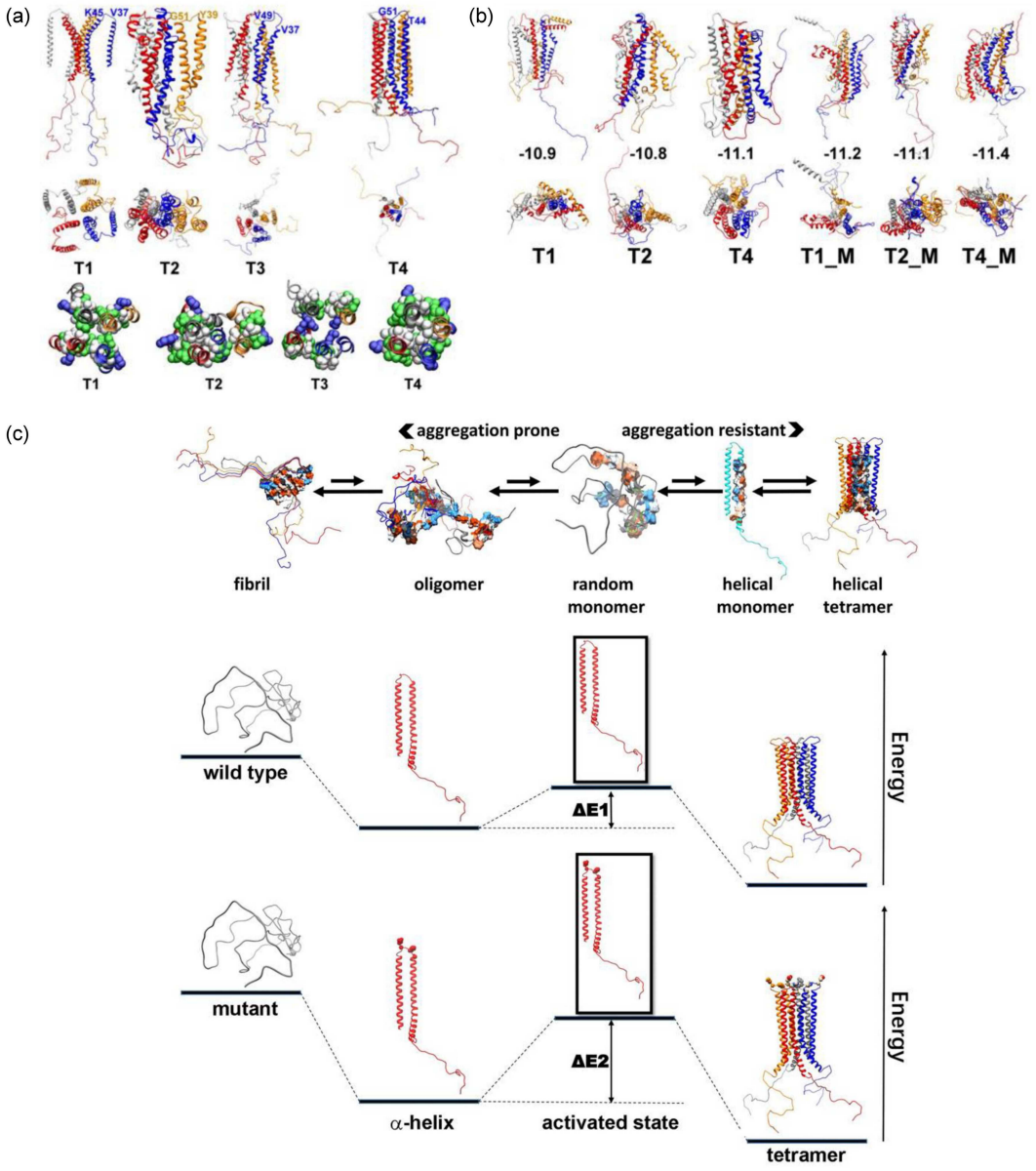


Fig. 6. (a) Initial conformations of four full-length α-helical αS tetramers (T1–T4) in both side and top views. The first and last residues of the loop that connects α1 and α2 segments are labelled. The heavy atoms (top view) of residues 71–82 are represented as van der Waals spheres with hydrophobic, acidic, basic, and polar residues coloured white, red, blue, and green, respectively. (b) Final structures of different helical αS tetramer conformations (T1, T2, and T4 for wild type, T1_M, T2_M, and T4_M for mutant) computed following 200 ns simulations, shown in side and top views. The tetramer T3 is not stable and is not included here. The corresponding conformational energy (in 10^3 kcal/mol) averaged over the final 20 ns of dynamics is shown for each tetramer. (c) Top: Role of helical tetramer in the pathological aggregation of αS. The hydrophobic NAC is also represented as the surface. Bottom: The proposed molecular pathways to decreased tetramer:monomer ratios in the mutant. Energy levels are estimated according to the calculated conformational energy of disordered αS monomers, helical αS monomers and tetramers.

However, there are very limited computationally verified models or MD simulation studies to date on the helical αS tetramers. One MD study that used a fragment-based approach to construct energetically favourable full-length αS suggested that the sampled structures with amphipathic helices can self-assemble via hydrophobic contacts to form tetramers [95]. In another study, a combination of replica exchange MD (REMD) and variational Bayesian weighting (VBW) methods was used to generate monomers, α-helical- and β-strand-rich αS trimers and tetramers in an attempt to resolve the controversy regarding experimentally observed αS native structure [96]. The authors noted that the ensemble is dominated by disordered monomers, with very few helical trimers and tetramers, although

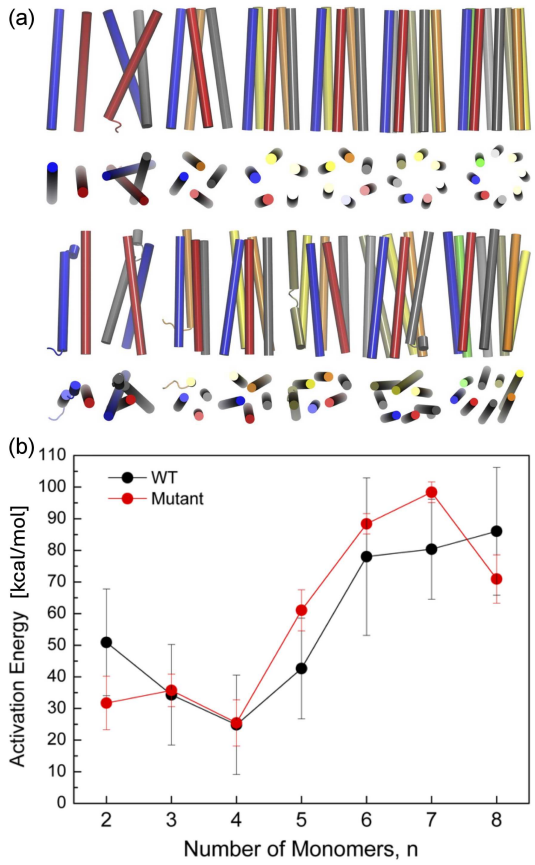


Fig. 7. (a) Designed αS NAC oligomers from dimer to octamer in side and top views at the start of the MD simulations (top panel, after minimization and equilibration) and after 200 ns of unconstrained dynamics (bottom panel) in water. (b) Activation energy of formation of αS multimers from monomer to octamer. Horizontal axis number n from 1 to 8 indicates growth from monomer to octamer.

the tetrameric states had significant helical content. Another simulation study observed that αS tetramer from a completely disordered state exhibited appreciably reduced stable β-sheets in comparison to dimers and a more stable helical content than either monomer or dimer [97]. Finally, by employing a steric parameter, correlations were obtained between the main and side chains of αS monomers and tetramers, revealing residues consisting mostly of parts of KTKEGV repeats that could potentially mediate the formation of helical tetramers [98].

Given that hydrophobic packing plays an important role in the folding and stabilisation of globular proteins [99], for IDPs like αS that display a broad distribution of conformational substates under physiological environment [7], hydrophobic interactions have been suggested to be a major driving force governing their self-assembly into oligomers, as "hydrophobes" also pack along the fibril growth axis [100]. We note that most homomers have high structural symmetry [101], and cyclic symmetries are thought to be the basic building blocks for the *de novo* design of self-assembling proteins such as water-soluble α-helical barrels [102] and helical bundles with high thermodynamic stability [103]. We used this preliminary knowledge to model and rationally design αS tetramers, which are nevertheless homomeric assemblies. We probe the free energy landscape of αS tetramerisation from four alternatively designed WT broken α-helical constructs (T1–T4) and their familial mutants T1_M, T2_M, and T4_M (Fig. 6a,b) and identify the active state corresponding to the conformation attained by a monomer when bound in a stable tetramer [104]. In the process, we designed the most thermodynamically stable (tetramer T4) *de novo* broken α-helical tetramer with a reconstructed loop motif using available experimental data [105]. Our results highlight that optimisation of inter-monomeric hydrophobic packing in NAC regions facilitates assembly to a stable broken water-soluble α-helical tetrameric construct, with secondary roles of the termini in regulating stability [104]. Moreover, we show that PD-causing familial mutations may create a much higher energy barrier for association of α-helical monomers into the aggregation-resistant α-helical tetramer, shifting tetramer–monomer equilibrium back towards aggregation-prone disordered monomers (Fig. 6c).

Following our designed *de novo* broken α-helical αS tetramer assembly in [104] with residues Val3–Val44 and Lys51–Thr92 forming two α-helices, we further designed a more stable αS broken α-helical tetramer construct using the same αS helical monomer as the building unit. Additionally, oligomers/multimers from dimers to octamers were modelled using the same designed broken α-helical monomer structure (Fig. 7a). The initial helical multimeric structures contained NAC regions with Cn symmetry. We characterize the thermodynamic and kinetic properties from MD simulations of both WT and quadruple mutated (E46K + H50Q + G51D + A53T) α-helical tetramers from the re-designed, more stable *de novo* α-helical tetramer assembly in order to elucidate the proposed hypothesis that tetramers may be ubiquitous in nature compared to α-helical αS oligomers from dimer to octamer [106]. Our models revealed that although the conformational stability of αS oligomers increases linearly with the number of monomers, the assembly of αS multimers proceeds *via* multiple energy barriers. The tetramer shows the lowest activation energy (Fig. 7b), which may explain its ubiquity.

3.5. Modelling co-assembly of pathological proteins in AD and PD

It is becoming increasingly clear that co-aggregation or cross-seeding assembly of amyloid proteins may be more neurotoxic than their self-assembly in AD and PD pathogenesis [107], and

there are several clinical overlaps in symptoms and pathologies between AD and PD. A previous MD study investigated the plausible early assembly pathways in PD and AD through cross-dimerisation of αS_{1-95} and $A\beta_{1-42}$ to reveal that the imperfect KTKEGV repeats in the N-terminus of αS may be responsible for forming inter-protein salt bridges with Aβ and NAC in αS may closely interact with Aβ hydrophobic core to form these hetero-assembled pathological protein complexes [108]. In a more recent study, MD simulations investigated the impact of αS–Aβ hetero dimerisation, showing that αS directly interacted with Aβ monomers and dimers, aggregating to potentially toxic β-barrel intermediates [109]. The αS–Aβ binding was mediated by the N-terminal end and NAC region in αS and the central hydrophobic cluster (CHC) C-terminus in Aβ.

Recent REMD simulations identified the conformational ensembles formed by the co-aggregation of CHC of Aβ ($A\beta_{16-22}$) and each of two core segments of tau (PHF6* and PHF6) [110]. The hetero-oligomers formed were found to be rich β-sheet, with PHF6 and $A\beta_{16-22}$ aggregate forming closed β-barrels, while PHF6* and $A\beta_{16-22}$ aggregate form open β-barrels. Hydrophobic and π–π stacking interactions were found to be crucial for the formation of toxic closed β-barrel between PHF6 and $A\beta_{16-22}$.

4. Conclusions

To identify new therapeutic targets for common proteinopathies such as Alzheimer's (AD) and Parkinson's diseases (PD), a comprehensive map of the molecular-level detailed pathway of early stages of self-assembly of pathological proteins is required, especially to structurally define the rare, polymorphic and short-lived toxic oligomeric intermediates. A number of *in vitro* and *in vivo* experimental techniques in the past have attempted to uncover the morphologies of pathogenic oligomers, but with little to no success, mainly due to a lack of a reliable quantification method. Computer-based molecular modelling and molecular dynamics (MD) simulations then have provided significant insights in guiding experiments on protein self-assembly. In this mini-review, we have focussed on the recent advances and latest findings from modelling and MD simulations of pathological protein self-assembly in bulk solution and on surfaces and interfaces that reveal some key structural and morphological details, thermodynamic driving forces, and kinetics of formation of several assembly constructs, including oligomers of proteins amyloid-β (Aβ) and tau in AD and α-synuclein in PD. We remain hopeful that fundamental simulation-guided research will uncover new therapeutic targets for these common proteinopathies and foster efforts to re-engineer functionality in pathogenic amyloids, like designer nanostructured materials [111–113].

Acknowledgments

This review article is dedicated with love to the memory of Professor Marek Cieplak, a dear mentor and friend of DT. To paraphrase a colleague's description of the generous and knowledgeable Marek, even when Google did not work, Marek worked. His passion and enthusiasm for modelling disordered linker regions in proteins informed much of our current work in IDPs as drug targets.

References

[1] T.P.J. Knowles, M. Vendruscolo, C.M. Dobson, *Nat. Rev. Mol. Cell Biol.* **15**, 384 (2014).

[2] C.A. Ross, M.A. Poirier, *Nat. Rev. Mol. Cell Biol.* **6**, 891 (2005).

[3] M.T. Heemels, *Nature* **539**, 179 (2016).

[4] V.L. Feigin, T. Vos, E. Nichols et al., *Lancet Neurol.* **19**, 255 (2020).

[5] World Health Organization (WHO), *Neurological Disorders. Public Health Challenges*, World Health Organization, Geneva 2006.

[6] À. Gómez-Sicilia, M. Sikora, M. Cieplak, M. Carrión-Vázquez, *PLoS Comput. Biol.* **11**, e1004541 (2015).

[7] S. Bhattacharya, L. Xu, D. Thompson, *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **8**, e1359 (2018).

[8] R.N.L. Lamptey, B. Chaulagain, R. Trivedi, A. Gothwal, B. Layek, J. Singh, *Int. J. Mol. Sci.* **23**, 1851 (2022).

[9] N. Bittner, C.S.M. Funk, A. Schmidt, F. Bermpohl, E.J. Brandl, E.E.A. Algharably, R. Kreutz, T.G. Riemer, *Drugs Aging* **40**, 953 (2023).

[10] S. Reardon, *Nature* **613**, 227 (2023).

[11] R. McShane, M.J. Westbya, E. Roberts, N. Minakaran, L. Schneider, L.E. Farrimond, N. Maayan, J. Ware, J. Debarros, *Cochrane Database Syst. Rev.* **2019**, CD003154 (2019).

[12] P. Chopade, N. Chopade, Z. Zhao, S. Mitragotri, R. Liao, V.C. Suja, *Bioeng. Transl. Med.* **8**, e10367 (2023).

[13] S. Srivastava, R. Ahmad, S.K. Khare, *Eur. J. Med. Chem.* **216**, 113320 (2021).

[14] J.A. Szász, V.A. Constantin, K. Orbán-Kis et al., *Brain Sci.* **11**, 826 (2021).

[15] T. Sinnige, *Chem. Sci.* **13**, 7080 (2022).

[16] J. Habchi, S. Chia, R. Limbocker et al., *Proc. Natl. Acad. Sci. USA* **114**, E200 (2017).

[17] Ł. Mioduszewski, M. Cieplak, *Phys. Chem. Chem. Phys.* **20**, 19057 (2018).

[18] M. Wojciechowski, À. Gómez-Sicilia, M. Carrión-Vázquez, M. Cieplak, *Mol. Biosyst.* **12**, 2700 (2016).

[19] A.B. Poma, H.V. Guzman, M.S. Li, P.E. Theodorakis, *Beilstein J. Nanotechnol.* **10**, 500 (2019).

[20] A. Kamada, A. Levin, Z. Toprakcioglu, Y. Shen, V. Lutz-Bueno, K.N. Baumann, P. Mohammadi, M.B. Linder, R. Mezzenga, T.P.J. Knowles, *Small* **16**, 1904190 (2020).

[21] L.R. Volpatti, T.P.J. Knowles, *J. Polym. Sci. B Polym. Phys.* **52**, 281 (2014).

[22] T.P.J. Knowles, M.J. Buehler, *Nat. Nanotechnol.* **6**, 469 (2011).

[23] S. Guerin, S.A.M. Tofail, D. Thompson, *Cryst. Growth Des.* **18**, 4844 (2018).

[24] K. Tao, J. O' Donnell, H. Yuan et al., *Energy Environ. Sci.* **13**, 96 (2020).

[25] R.B. Svensson, H. Mulder, V. Kovanen, S.P. Magnusson, *Biophys. J.* **104**, 2476 (2013).

[26] M.G. Iadanza, M.P. Jackson, E.W. Hewitt, N.A. Ranson, S.E. Radford, *Nat. Rev. Mol. Cell Biol.* **19**, 755 (2018).

[27] M. Goedert, D.S. Eisenberg, R.A. Crowther, *Annu. Rev. Neurosci.* **40**, 189 (2017).

[28] A.L. Mahul-Mellier, J. Burtscher, N. Maharjan, L. Weerens, M. Croisier, F. Kuttler, M. Leleu, G.W. Knott, H.A. Lashuel, *Proc. Natl. Acad. Sci. USA* **117**, 4971 (2020).

[29] A.J. Dear, T.C.T. Michaels, G. Meisl, D. Klenerman, S. Wu, S. Perrett, S. Linse, C.M. Dobson, T.P.J. Knowles, *Proc Natl Acad Sci USA* **117**, 12087 (2020).

[30] P.H. Nguyen, A. Ramamoorthy, B.R. Sahoo et al., *Chem. Rev.* **121**, 2545 (2021).

[31] R. Limbocker, N. Cremades, R. Cascella, P.M. Tessier, M. Vendruscolo, F. Chiti, *Acc. Chem. Res.* **56**, 1395 (2023).

[32] K. Kulenkampff, A.M. Wolf Perez, P. Sormanni, J. Habchi, M. Vendruscolo, *Nat. Rev. Chem.* **5**, 277 (2021).

[33] F. Grigolato, P. Arosio, *Biophys. Chem.* **270**, 106533 (2021).

[34] J.M. Kenyaga, Q. Cheng, W. Qiang, *J. Biol. Chem.* **298**, 102491 (2022).

[35] N. El Mammeri, O. Gampp, P. Duan, M. Hong, *Commun. Biol.* **6**, 467 (2023).

[36] A.S. Kurochka, D.A. Yushchenko, P. Bouř, V.V. Shvadchak, *ACS Chem. Neurosci.* **12**, 825 (2021).

[37] T. John, A. Gladytz, C. Kubeil, L.L. Martin, H.J. Risselada, B. Abel, *Nanoscale* **10**, 20894 (2018).

[38] A. Morriss-Andrews, J.E. Shea, *J. Chem. Phys.* **136**, 065103 (2012).

[39] P. Ricchiuto, A.V. Brukhno, S. Auer, *J. Phys. Chem. B* **116**, 5384 (2012).

[40] L. Xu, S. Bhattacharya, D. Thompson, *Methods Mol. Biol.* **2340**, 379 (2022).

[41] S. Bhattacharya, L. Xu, D. Thompson, *Methods Mol. Biol.* **2340**, 401 (2022).

[42] D. Matthes, V. Gapsys, J.T. Brennecke, B.L. De Groot, *Sci. Rep.* **6**, 33156 (2016).

[43] D.O.V. Alonso, S.J. DeArmond, F.E. Cohen, V. Daggett, *Proc. Natl. Acad. Sci. USA* **98**, 2985 (2001).

[44] L. Calzolai, R. Zahn, *J. Biol. Chem.* **278**, 35592 (2003).

[45] S. Samantray, W. Schumann, A.-M. Illig, M. Carballo-Pacheco, A. Paul, B. Barz, B. Strodel, *Computer Simulations of Aggregation of Proteins and Peptides*, Vol. 2340, 2022, p. 235.

[46] B. Strodel, *Curr. Opin. Struct. Biol.* **67**, 145 (2021).

[47] J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B.L. de Groot, H. Grubmüller, A.D. MacKerell Jr., *Nat. Methods* **14**, 71 (2016).

[48] P. Robustelli, S. Piana, D.E. Shaw, *Proc. Natl. Acad. Sci. USA* **115**, E4758 (2018).

[49] Y. Zhang, Y.L. Lyubchenko, *Biophys. J.* **107**, 2903 (2014).

[50] W. Zheng, M.Y. Tsai, M. Chen, P.G. Wolynes, *Proc. Natl. Acad. Sci. USA* **113**, 11835 (2016).

[51] M. Hashemi, Y. Zhang, Z. Lv, Y.L. Lyubchenko, *Nanoscale Adv.* **1**, 3892 (2019).

[52] G.M. Shankar, S. Li, T.H. Mehta et al., *Nat. Med.* **14**, 837 (2008).

[53] V.H. Man, P.H. Nguyen, P. Derreumaux, *J. Phys. Chem. B* **121**, 2434 (2017).

[54] V.H. Man, P.H. Nguyen, P. Derreumaux, *J. Phys. Chem. B* **121**, 5977 (2017).

[55] A.M. Illig, B. Strodel, *J. Chem. Theory Comput.* **16**, 7825 (2020).

[56] S. Bhattacharya, L. Xu, D. Thompson, *ACS Chem. Neurosci.* **10**, 2830 (2019).

[57] S. Bhattacharya, L. Xu, D. Thompson, *Sci. Rep.* **10**, 7597 (2020).

[58] P.N. Nirmalraj J. List, S. Battacharya, G. Howe, L. Xu, D. Thompson, M. Mayer, *Sci. Adv.* **6**, eaaz6014 (2020).

[59] M. Morris, S. Maeda, K. Vossel, L. Mucke, *Neuron* **70**, 410 (2011).

[60] Y. Wang, E. Mandelkow, *Nat. Rev. Neurosci.* **17**, 22 (2016).

[61] M.D. Mukrasch, S. Bibow, J. Korukottu, S. Jeganathan, J. Biernat, C. Griesinger, E. Mandelkow, M. Zweckstetter, *PLoS Biol.* **7**, e1000034 (2009).

[62] A.W.P. Fitzpatrick, B. Falcon, S. He et al., *Nature* **547**, 185 (2017).

[63] Y. Shi, A.G. Murzin, B. Falcon et al., *Acta Neuropathol.* **141**, 697 (2021).

[64] G.I. Hallinan, M.R. Hoq, M. Ghosh, F.S. Vago, A. Fernandez, H.J. Garringer, R. Vidal, W. Jing, B. Ghetti, *Acta Neuropathol.* **142**, 227 (2021).

[65] B. Falcon, W. Zhang, M. Schweighauser, A.G. Murzin, R. Vidal, H.J. Garringer, B. Ghetti, S.H.W. Scheres, M. Goedert, *Acta Neuropathol.* **136**, 699 (2018).

[66] B. Falcon, W. Zhang, A.G. Murzin, G. Murshudov, H.J. Garringer, R. Vidal, R.A. Crowther, B. Ghetti, S.H.W. Scheres, M. Goedert, *Nature* **561**, 137 (2018).

[67] P. Duan, A.J. Dregni, N. El Mammeri, M. Hong, *Proc. Natl. Acad. Sci. USA* **120**, e2310067120 (2023).

[68] N. Basheer, T. Smolek, I. Hassan, F. Liu, K. Iqbal, N. Zilka, P Novak, *Mol. Psychiatry* **28**, 2197 (2023).

[69] Y. Zhang, K.M. Wu, L. Yang, Q. Dong, J.T. Yu, *Mol. Neurodegener.* **17**, 28 (2022).

[70] H. He, Y. Liu, Y. Sun, F. Ding, *J. Chem. Inf. Model.* **61**, 2916 (2021).

[71] P.H. Nguyen, P. Derreumaux, *J. Phys. Chem. B* **126**, 3431 (2022).

[72] H. Liu, H. Zhong, Z. Xu, Q. Zhang, S.J.A. Shah, H. Liu, X. Yao, *Phys. Chem. Chem. Phys.* **22**, 10968 (2020).

[73] X. Li, X. Dong, G. Wei, M. Margittai, R. Nussinov, B. Ma, *Chem. Commun.* **54**, 5700 (2018).

[74] L. Xu, J. Zheng, M. Margittai, R. Nussinov, B. Ma, *ACS Chem. Neurosci.* **7**, 565 (2016).

[75] A.E. Sahayaraj, R. Viswanathan, F. Pinhero, A. Abdul Vahid, V. Vijayan, *ACS Chem. Neurosci.* **14**, 136 (2023).

[76] O. Maraba, S. Bhattacharya, M. Conda-Sheridan, D. Thompson, *Nano Express* **3**, 044004 (2022).

[77] G. Fusco, S.W. Chen, P.T.F. Williamson et al., *Science* **358**, 1440 (2017).

[78] L. Xu, S. Bhattacharya, D. Thompson, *Phys. Chem. Chem. Phys.* **20**, 4502 (2018).

[79] H.A. Lashuel, C.R. Overk, A. Oueslati, E. Masliah, *Nat. Rev. Neurosci.* **14**, 38 (2012).

[80] Y. Zhang, Y. Wang, Y. Liu, G. Wei, F. Ding, Y. Sun, *ACS Chem. Neurosci.* **13**, 3126 (2022).

[81] S.B.T.A. Amos, T.C. Schwarz, J. Shi, B.P. Cossins, T.S. Baker, R.J. Taylor, R. Konrat, M.S.P. Sansom, *J. Phys. Chem. B* **125**, 2929 (2021).

[82] B. Frieg, L. Antonschmidt, C. Dienemann et al., *Nat. Commun.* **13**, 6810 (2022).

[83] O. Synhaivska, S. Bhattacharya, S. Campioni, D. Thompson, P.N. Nirmalraj, *ACS Chem. Neurosci.* **13**, 1410 (2022).

[84] W. Wang, I. Perovic, J. Chittuluru et al., *Proc. Natl. Acad. Sci. USA* **108**, 17797 (2011).

[85] T. Bartels, J.G. Choi, D.J. Selkoe, *Nature* **477**, 107 (2011).

[86] F.X. Theillet, A. Binolfi, B. Bekei et al., *Nature* **530**, 45 (2016).

[87] P.H. Weinreb, W. Zhen, A.W. Poon, K.A. Conway, P.T. Lansbury, *Biochemistry* **35**, 13709 (1996).

[88] B. Fauvet, M.K. Mbefo, M.-B. Fares et al., *J. Biol. Chem.* **287**, 15345 (2012).

[89] U. Dettmer, N. Ramalingam, V.E. von Saucken et al., *Hum. Mol. Genet.* **26**, 3466 (2017).

[90] E.S. Luth, T. Bartels, U. Dettmer, N.C. Kim, D.J. Selkoe, *Biochemistry* **54**, 279 (2015).

[91] D. Selkoe, U. Dettmer, E. Luth, N. Kim, A. Newman, T. Bartels, *Neurodegener. Dis.* **13**, 114 (2014).

[92] U. Dettmer, D. Selkoe, T. Bartels, *Curr. Opin. Neurobiol.* **36**, 15 (2016).

[93] U. Dettmer, A.J. Newman, F. Soldner et al., *Nat. Commun.* **6**, 7314 (2015).

[94] S. Nuber, M. Rajsombath, G. Minakaki, J. Winkler, C.P. Müller, M. Ericsson, B. Caldarone, U. Dettmer, D.J. Selkoe, *Neuron* **100**, 75 (2018).

[95] O. Ullman, C.K. Fisher, C.M. Stultz, *J. Am. Chem. Soc.* **133**, 19536 (2011).

[96] T. Gurry, O. Ullman, C.K. Fisher, I. Perovic, T. Pochapsky, C.M. Stultz, *J. Am. Chem. Soc.* **135**, 3865 (2013).

[97] J.Y. Mane, M. Stepanova, *FEBS Open Bio* **6**, 666 (2016).

[98] Y. Cote, P. Delarue, H.A. Scheraga, P. Senet, G.G. Maisuradze, *ACS Chem. Neurosci.* **9**, 1051 (2018).

[99] G.D. Rose, R. Wolfenden, *Annu. Rev. Biophys. Biomol. Struct.* **22**, 381 (1993).

[100] B. Li, P. Ge, K.A. Murray et al., *Nat. Commun.* **9**, 3609 (2018).

[101] C.H. Norn, I. André, *Curr. Opin. Struct. Biol.* **39**, 39 (2016).

[102] A.R. Thomson, C.W. Wood, A.J. Burton, G.J. Bartlett, R.B. Sessions, R.L. Brady, D.N. Woolfson, *Science* **346**, 485 (2014).

[103] P.S. Huang, G. Oberdorfer, C. Xu et al., *Science* **346**, 481 (2014).

[104] L. Xu, S. Bhattacharya, D. Thompson, *Chem. Commun.* **54**, 8080 (2018).

[105] E. Kara, P.A. Lewis, H. Ling, C. Proukakis, H. Houlden, J. Hardy, *Neurosci. Lett.* **546**, 67 (2013).

[106] L. Xu, S. Bhattacharya, D. Thompson, *Phys. Chem. Chem. Phys.* **21**, 12036 (2019).

[107] B. Ren, Y. Zhang, M. Zhang et al., *J. Mater. Chem. B* **7**, 7267 (2019).

[108] J.C. Jose, P. Chatterjee, N. Sengupta, *PLoS One* **9**, e106883 (2014).

[109] F. Huang, Y. Liu, Y. Wang, J. Xu, J. Lian, Y. Zou, C. Wang, F. Ding, Y. Sun, *Phys. Chem. Chem. Phys.* **25**, 31604 (2023).

[110] X. Li, Y. Chen, Z. Yang, S. Zhang, G. Wei, L. Zhang, *Int. J. Biol. Macromol.* **254**, 127841 (2024).

[111] M. Gunnoo, P.-A. Cazade, A. Orlowski, M. Chwastyk, H. Liu, D.T. Ta, M. Cieplak, M. Nashde, D. Thompson, *Phys. Chem. Chem. Phys.* **20**, 22674 (2018).

[112] M. Gunnoo, P.-A. Cazade, A. Galera-Prat et al., *Adv. Mater.* **28**, 5619 (2016).

[113] G. Nawrocki, P.A. Cazade, D. Thompson, M. Cieplak, *J. Phys. Chem. C* **119**, 24404 (2015).