



PhD-FSTC-2019-57

The Faculty of Sciences, Technology and Communication

DISSERTATION

Defence held on 02/09/2019 in Luxembourg
to obtain the degree of

DOCTEUR DE L'UNIVERSITÉ DU LUXEMBOURG
EN PHYSIQUE

by

Ali Asghar HAKAMI ZANJANI

Born on 28 March 1974 in Zanjan (Iran)

THERMODYNAMICS AND AGGREGATION KINETICS OF LYSOZYME-DERIVED PEPTIDES

Dissertation defence committee

Dr. Tanja Schilling, dissertation supervisor

Professor, Universität Freiburg

Dr. Josh Berryman

Université du Luxembourg

Dr. Alexandre Tkatchenko, Chairman

Professor, Université du Luxembourg

Dr. Rhoda Hawkins

University of Sheffield

Dr. Thomas Waigh, Vice-Chairman

University of Manchester

Acknowledgements

I would like to thank my supervisors Dr. Joshua T. Berryman and Prof. Tanja Schilling for giving me the opportunity to work on this project and for their advice and support during the last four years.

I would like to express my gratitude to all our collaborators especially Prof. Raffaele Mezzenga (*Department of Health Science and Technology, ETH Zurich, Switzerland*) and Dr. Nicholas P. Reynolds (*Swinburne University of Technology, Australia*) for their valuable contributions.

I wish to thank Prof. Alexander Tkatchenko and all my friends and colleagues in *Theoretical Chemical Physics* (TCP) group in the University of Luxembourg and *Statistical Physics of Soft Matter and Complex Systems* (SO Φ T) group in the University of Freiburg who have given me their support and suggestions during these years.

Resources from the Luxembourgish *Fonds Nationale de la Recherche* (C14/MS/8329720) and the University of Luxembourg High-Performance Computing facility are gratefully acknowledged.

This text was created with \LaTeX using the Texmaker under Linux and Python was used for plotting mathematical functions. Some other free and open-source programs have been used in this project which are mentioned in the text. I would like to appreciate the generosity of all who have made these programs.

I would also like to particularly thank my wife, Bitu, for her patience, encouragement and continuous support.

Abstract

When multiple similar protein or peptide chains form non-covalent aggregates, this is termed ‘amyloid’. Many serious progressive diseases such as Alzheimer’s and Parkinson’s are related to undesirable amyloid aggregation. From a positive perspective, functional amyloids have applications as robust and versatile biomaterials in nature, nanotechnology, and biomedicine. To probe the properties of the amyloid aggregation process in terms of the structure of molecules and the microscopic interactions between them, molecular simulation methods such as molecular dynamics (MD) and Monte Carlo (MC) can be used. These tools are especially valuable to illustrate short length and time scales not easily accessible for systems in solution via current experimental techniques.

In this work the thermodynamics and aggregation kinetics of the ILQINS hexapeptide are studied. ILQINS is a biological material derived from hen’s egg-white lysozyme. Two ILQINS homologues, IFQINS and TFQINS are compared to ILQINS and some of the complex physics which leads to the increased amyloidogenicity of these species, which is not expected from first-order consideration of amino acid properties, is discussed.

The IFQINS hexapeptide is of particular interest as the human homologue of ILQINS. Solution X-ray and X-ray crystallography are compared to simulation, verifying that at least two metastable polymorphic structures exist for this system which are substantially different at the atomistic scale, and illustrating the physics driving kinetic competition between polymorphs.

Contents

1	Introduction	1
1.1	Peptide Structure	1
1.2	Peptide aggregation	5
1.2.1	Amyloids	5
1.2.2	Noncovalent Interactions	6
1.2.3	Polymorphism	8
1.3	ILQINS, a Lysozyme-Derived Hexapeptide	9
1.4	Methods	11
1.4.1	Classical Molecular Dynamics (MD)	11
1.4.1.1	Solvent Models	13
1.4.2	Replica Exchange Molecular Dynamics (REMD)	15
1.4.2.1	Hamiltonian Replica Exchange Molecular Dynamics (HREMD)	17
1.4.3	Kinetic Monte Carlo (KMC)	17
1.4.4	Molecular Graphics	20
1.5	Overview	20
2	Antiparallel Strand Alignment as the Default for Amyloid Oligomers at Early Times	21
2.1	Introduction	21
2.2	Results	24
2.2.1	3D Aggregates of ILQINS in Different Symmetry Classes and Corresponding WAXS Curves	24
2.2.2	Energy Decomposition	27
2.2.3	Hamiltonian Replica Exchange Molecular Dynamics (HREMD)	28
2.2.4	Classical MD of P and AP Single β -Sheets	30
2.3	Discussion	32
2.4	Materials and Methods	33

2.4.1	MD for 3D Aggregates and Scattering Calculations	33
2.4.2	Energy Decomposition	33
2.4.3	Hamiltonian Replica Exchange MD	33
2.4.4	Synthesis and Light Scattering	35
3	Kinetic Control of Parallel versus Antiparallel Amyloid Aggregation via Mesostructural Polymorphism	37
3.1	Introduction	37
3.2	Results	39
3.2.1	Atomistic Simulations Compared to WAXS	39
3.2.2	Peptide-Level Assembly Thermodynamics	40
3.2.3	Complex Kinetic Competition	43
3.3	Methods	46
3.3.1	Molecular Simulation	46
3.3.2	Desolvation Energy Calculation	46
3.3.3	Kinetic Rate Equation Network	47
3.3.4	Experimental Methods	50
3.4	Discussion	50
4	Discussion	53
4.1	Antiparallel Structure Dominates Initially	53
4.2	Kinetic Competition Between P and AP Aggregation	54
4.3	Context in Aggregation Mechanisms and Kinetics	54
4.4	Outlook	57
	Bibliography	58

List of Figures

1.1	Two amino acids make a dipeptide and a water molecule is released.	1
1.2	The backbone of polypeptide chains as a series of rigid planes with consecutive planes sharing a common point of rotation at C_α . a Parallel β -sheet conformation of polypeptide chains. b Antiparallel β -sheet conformation in which the amino-terminal to carboxyl-terminal orientation of adjacent chains is reversed.	3
1.3	Ramachandran plot of the torsional angles in a peptide for various allowed secondary structures. The torsional angles determine the conformation of the residues and the peptide. Blue areas show the fully allowed conformations that involve no steric overlap; green areas indicate conformations allowed subject to slightly unfavourable atomic contacts; yellow area reflects conformations with significant strain [1].	4
1.4	Eight different fibril symmetry classes [2] illustrated with left hands. R_a , R_b and R_c show rotations of 180° about a , b and c axis. Classes 1–4 have parallel alignment of β -strands within each β -sheet and classes 5–8 have antiparallel arrangement of β -strands within their β -sheets. Each peptide sequence may have the ability to form one or more of these morphologies.	8
1.5	Three homologous hexapeptides: a ILQINS (from d hen’s egg-white lysozyme [3WUN]), b IFQINS (from e human lysozyme [1REX]) and c TFQINS (from f mutant human lysozyme [1IOC]). The hexapeptide sequence is exposed to solvent even in the folded protein, as it is in the active-site cleft.	10
1.6	An overview of the MD procedure, using the AMBER force field and the Velocity-Verlet integration algorithm.	12
1.7	In this schematic figure colored arrows represent MD trajectories of replicas and the black arrows represent attempted swaps between replicas in state space. Question marks represent Monte Carlo exchange attempts (check marks for successful, cross for failed).	16

1.8	Collision of two species along a axis (a) Two species have matching-sized planes in a and they join together after collision. (b) Two species have mismatched planes in a , but they have the same length in b and the collision requires a split in the c axis. (c) Two species have mismatching-sized planes with different lengths in b and c ; collision requires a split first in b , then another split in c or a split first in c , then another split in b . Depending on the energy cost of the two reactions of panel (c), only the least expensive is implemented in order to control the complexity of the simulation.	19
2.1	Amyloid-like lattices based on the eight different symmetry classes, illustrated with images of left hands. R_a , R_b and R_c show rotations of 180° about a , b and c axis. Classes 1–4 have parallel alignment of β -strands within each β -sheet and classes 5–8 have antiparallel arrangement of β -strands within their β -sheets. A given peptide sequence may have metastable aggregates with one or more of these symmetries. A given symmetry defines only the sheet and steric zipper relationships: multiple space groups not shown, and also intra-peptide degrees of freedom, may be compatible with a given class.	23
2.2	Eight initial structures before MD simulations corresponding to the symmetry classes in Figure 2.1. View is an orthoscopic projection along the c (hydrogen-bonding) crystal axis, such that the P β structures 1-4 appear to be a 2D wallpaper.	25
2.3	WAXS of ILQINS peptide solution recorded after 24 h of self-assembly (black), and calculated scattering for the MD snapshot of the 1296-peptide nanocrystal ($6 \times 12 \times 18$) taken at 10 ns (AP class 5: blue, P class 1: red). The AP lattice parameters are $a = 20.6\text{\AA}$, $b = 19.1\text{\AA}$ and $\gamma = 82^\circ$, peaks relevant to lattice shape are annotated with the corresponding Bragg spacing. <i>ILQINS WAXS data supplied by Nicholas P. Reynolds, Swinburne University of Technology, Australia.</i>	26
2.4	Class 5 AP structure. a Two sequential planes perpendicular to the c axis: termini are in close contact on the a axis, but not across the steric zipper. b plane perpendicular to the b axis, showing backbones of peptides only. Termini are also in contact along the c axis and hydrogen bonding is strong. c A molecular dynamics snapshot with semi-ordered dynamic water molecules shown as orange spheres. The AP lattice parameters are $a = 20.6\text{\AA}$, $b = 19.1\text{\AA}$ and $\gamma = 82^\circ$	27

2.5	Number of hydrogen bonds in parallel (red) and antiparallel (blue) bridges calculated by DSSP [3] analysis for $\lambda = 0$ during the simulation time confirms that the majority of aggregates are forming in antiparallel structures. Thick lines indicate a moving average over 50 ps	29
2.6	(a-c) Extended reference peptides used for HREMD simulations and (d-f) representative AP single β -sheets formed in these simulations. Sidechain orientation in the AP sheets initially formed is not regular: the structures (d-f) are mixed between the intrasheet orderings of classes 5/6 and 7/8.	30
2.7	Classical MD over 100 ns in explicit water shows that antiparallel single β -sheets have greater configurational stability than parallel. (a-f): P and AP β structures consistent with classes 1 or 2 (P) and 5 or 6 (AP), at 100 ns for the three sequences studied. (g-i): backbone hydrogen bonding declines more rapidly for the P structure in each case. Thick lines indicate a moving average over a 1 ns window.	31
2.8	Replica exchange convergence for different sequences; $\lambda_i = 0.032(i/63)^2$ was verified empirically to provide good mixing between replicas (Replica index i runs over [0..63]). Replica exchanges were attempted at 5 ps intervals. The black trace in each system shows an arbitrary reference replica.	34
3.1	(a) Calculated scattering curve based on computationally derived AP structure, compared to solution scattering; lattice parameters are $a = 20.0 \text{ \AA}$, $b = 19.1 \text{ \AA}$ and $\gamma = 83^\circ$. (b) AP structure, with semi-ordered dynamic water molecules shown as orange spheres. (c) Calculated scattering based on solid-phase crystal structure after evolution in aqueous environment for 15 ns. (d) Solid-phase P crystal structure 4R0P of IFQINS (four unit cells) reported by Sievers; lattice parameters are $a = 43.2 \text{ \AA}$, $b = 19.6 \text{ \AA}$ and $\gamma = 90^\circ$. A full translational unit cell is shown as a rectangle in the center. <i>IFQINS WAXS data supplied by Nicholas P. Reynolds, Swinburne University of Technology, Australia.</i>	40
3.2	Cartoon showing binding free energies for interfaces in the a (backbone axis), b (sidechain axis) and c (hydrogen bond axis). (a-d): AP structure has generic in-register steric zipper 3D assembly. (e-h): Sievers' structure has herringbone assembly characterised by two energy terms ϵ and ϵ' . Binding free energies are (f): $\Delta G = (2n_b - 1)\epsilon/2$ in the a direction and (g): $\Delta G = (n_a - 1)\epsilon + n_a\epsilon'$ in the b direction, where n_a and n_b are the number of peptides in the a and b directions respectively.	42

3.3	Evolution of oligomer formation colored by parallel or antiparallel. Curves i-viii show concentrations increasing from 5 nM in multiples of 10. Spreads show minimum and maximum values reached in 10 replicates. Traces are averages over the replicates. ‘Frac. 1D+’ (a) shows peptides incorporated into any aggregate which is identifiably P or AP, therefore initially describes mostly single-sheet aggregates, for which AP is the most stable geometry. (b) shows formation of 2D aggregates, where complex kinetics driven by the availability of 1D aggregates and the stability of 2D+ aggregates begin to take effect. (c) shows the quite sharp kinetically-driven phase transition between AP dominance and AP-P coexistence in the final sample.	44
3.4	Evolution of cross-sectional aspect ratio colored by parallel or antiparallel. Monomers are counted as both P and AP for purposes of averaging. Spreads show minima and maxima reached in 10 replicates. Traces are averages over the replicates. Both P and AP show complex behaviour crossing from μM to mM concentration.	45
4.1	Schematic overview of peptide aggregation pathways. <i>Vesicle, Micelle and Bilayer images CC licenced by Mariana Ruiz Villarreal. β-barrel structure is 4RLC.</i>	55

Chapter 1

Introduction

In this introductory chapter the basic concepts of peptide sequence, or ‘primary structure’ are introduced. Peptide local configuration, or ‘secondary structure’ is also introduced with emphasis on the parallel (P) and antiparallel (AP) β -sheet secondary structures which are a common signature of amyloid formation. The varying mesoscale morphologies of peptide aggregates which may have similar primary and secondary structures is motivated. Finally, the specific lysozyme-derived hexapeptides studied in this thesis and the methods used are introduced.

1.1 Peptide Structure

Polypeptides are chains of amino acids held together by peptide bonds. A peptide bond is formed when the carboxyl group of one amino acid binds the amino group of another amino acid in a condensation reaction and a molecule of water is released [4] (see Figure 1.1).

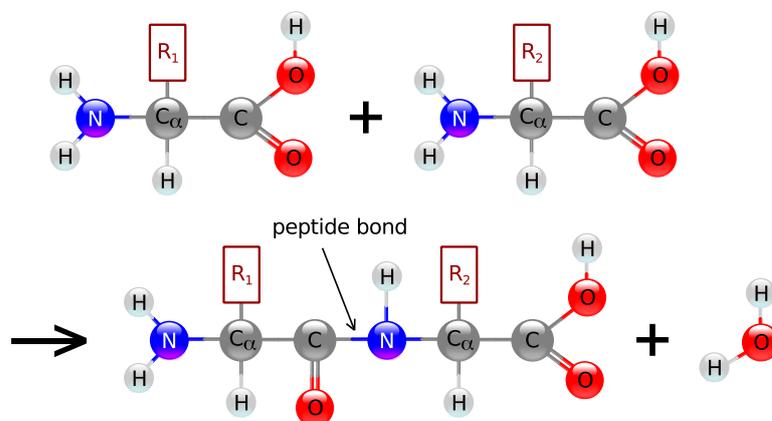


Figure 1.1: Two amino acids make a dipeptide and a water molecule is released.

The difference between a peptide and a protein is defined such that all proteins are also peptides but not all peptides are proteins. Typically we define a peptide as a protein if it has more than 50 residues (50 amino acids) and is composed of biologically available amino acids. If a small or non-natural peptide nonetheless has a stable globular fold we may still call it a protein.

Peptide bonds are rigid, shorter than the C–N bond in a simple amine, and unable to rotate freely: the atoms associated with the peptide bond are co-planar. The major rotational degrees of freedom are about the N–C $_{\alpha}$ and C $_{\alpha}$ –C bonds, so we can consider the backbone of a polypeptide chain as a series of rigid planes with consecutive planes sharing a common point of rotation at C $_{\alpha}$ as illustrated in Figure 1.2. Rotations about N–C $_{\alpha}$ and C $_{\alpha}$ –C bonds are labeled ϕ and ψ respectively and are known as torsion or dihedral angles [5].

The values of torsion angles define the backbone conformation of the peptide chain [1,6,7], however the presence of sidechains (for all amino-acids apart from Glycine) constrains the allowed backbone torsions. Flexible sidechains will have further degrees of freedom, and sidechain-sidechain interactions are decisive in determining preferred conformations. Steric, electrostatic and dispersion interactions between sidechains make determination of the structure or structural ensemble for a chain or set of chains a complex many-body optimisation problem. In nature, the sequence of sidechains is selected by evolution to determine the structure and/or assembly of the protein such that it serves its function (or often many functions) in the organism.

Although torsion angles can formally have any value between -180° and $+180^{\circ}$ many values are not possible because of steric clashes between atoms in the polypeptide backbone and side chains. Allowed values for torsion angles vary considerably by sequence, however a general pattern exists which is graphically shown in the Ramachandran plot [8] of Figure 1.3.

One of the most prominent secondary structures of polypeptides which occurs widely in folded globular proteins is β conformation where polypeptide chains form sheets [9,10]. The backbone of the chain is extended and the chains are arranged to form (possibly curved) planar sheets with hydrogen bonds between adjacent strands. Neighbouring β -strands in a β -sheet can be either parallel (having the same amino-to-carboxyl orientation) or antiparallel (having opposite amino-to-carboxyl orientation) (see Figure 1.2). Three-dimensional protein structure data from the Protein Data Bank (PDB) show that the majority of β -sheets in folded proteins are antiparallel sheets [11], however this arises mostly from the topological constraint that sheets are formed from a continuous peptide chain. Although hydrogen bonding geometry appears different the free energy of formation for assemblies of inde-

pendent chains in P or AP β -sheet is roughly the same [12], with the more stable of the two being determined by the sidechains.

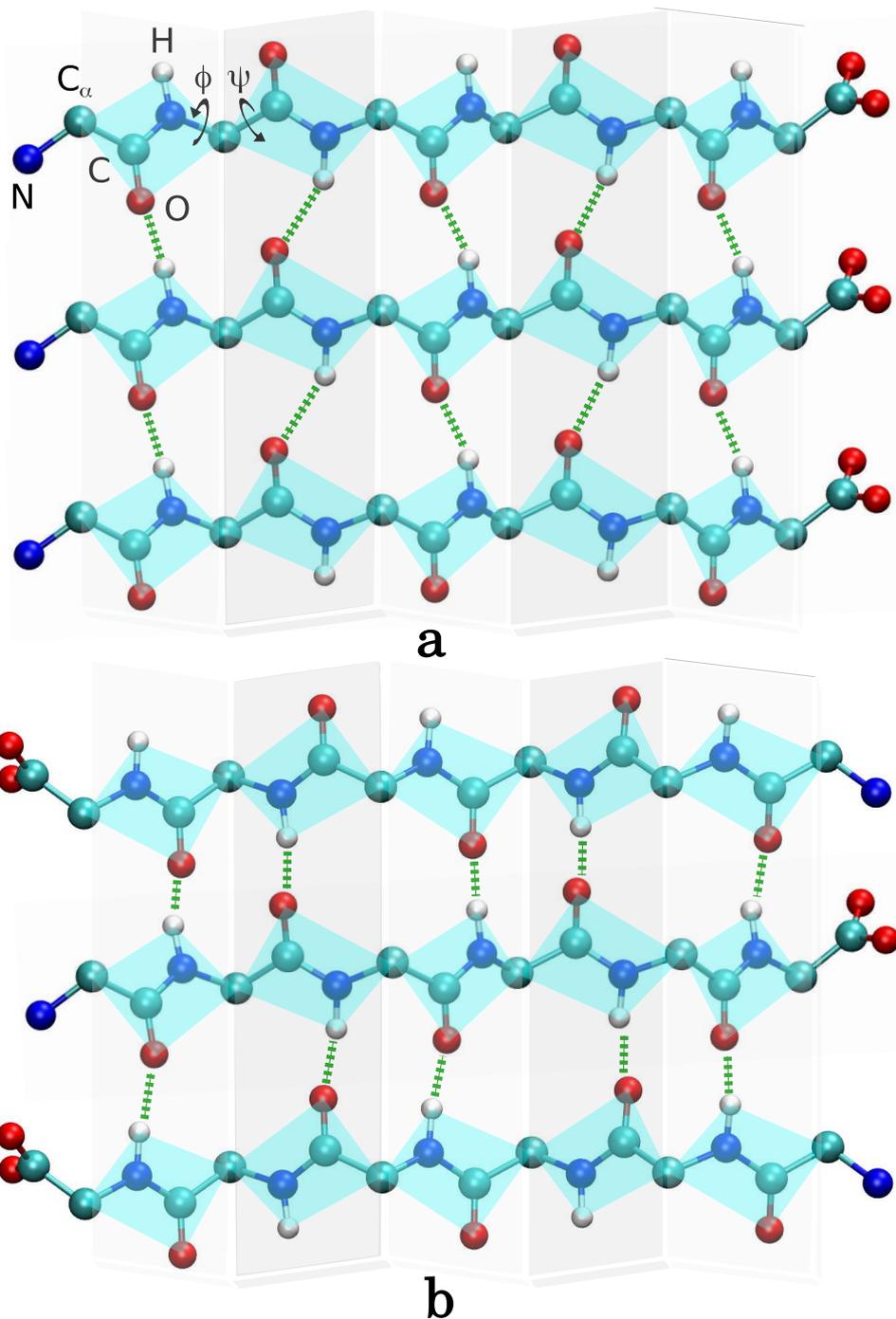


Figure 1.2: The backbone of polypeptide chains as a series of rigid planes with consecutive planes sharing a common point of rotation at C_α . **a** Parallel β -sheet conformation of polypeptide chains. **b** Antiparallel β -sheet conformation in which the amino-terminal to carboxyl-terminal orientation of adjacent chains is reversed.

A currently debated possible structure of particular interest in relation to amyloid formation is α -sheet, which is similar to β -sheet, but with different orientation of the carbonyl and amino groups in the peptide bond units; in each strand, all the carbonyl groups are oriented in the same direction on one side, and all the amino groups are oriented in the same direction on the opposite side [13, 14]. On the Ramachandran plane (Figure 1.3) this equates to alternating between the α -L and α -R regions to give near-zero net twist. Like β -sheet, α -sheet may be formed between separate chains, and α -sheet to β -sheet interconversion via peptide plane flipping may be a stage in the assembly of some multichain aggregates [15]. Further interesting conformations in relation to multichain assembly include β -sheet curved to form a cylindrical tube structure, called β -barrel, in which the first strand is bonded to the last strand via hydrogen bonds. β -barrel membrane proteins carry out various functions in different organisms [16, 17], and β -barrel oligomers have also been advanced as a neurotoxic species [18, 19].

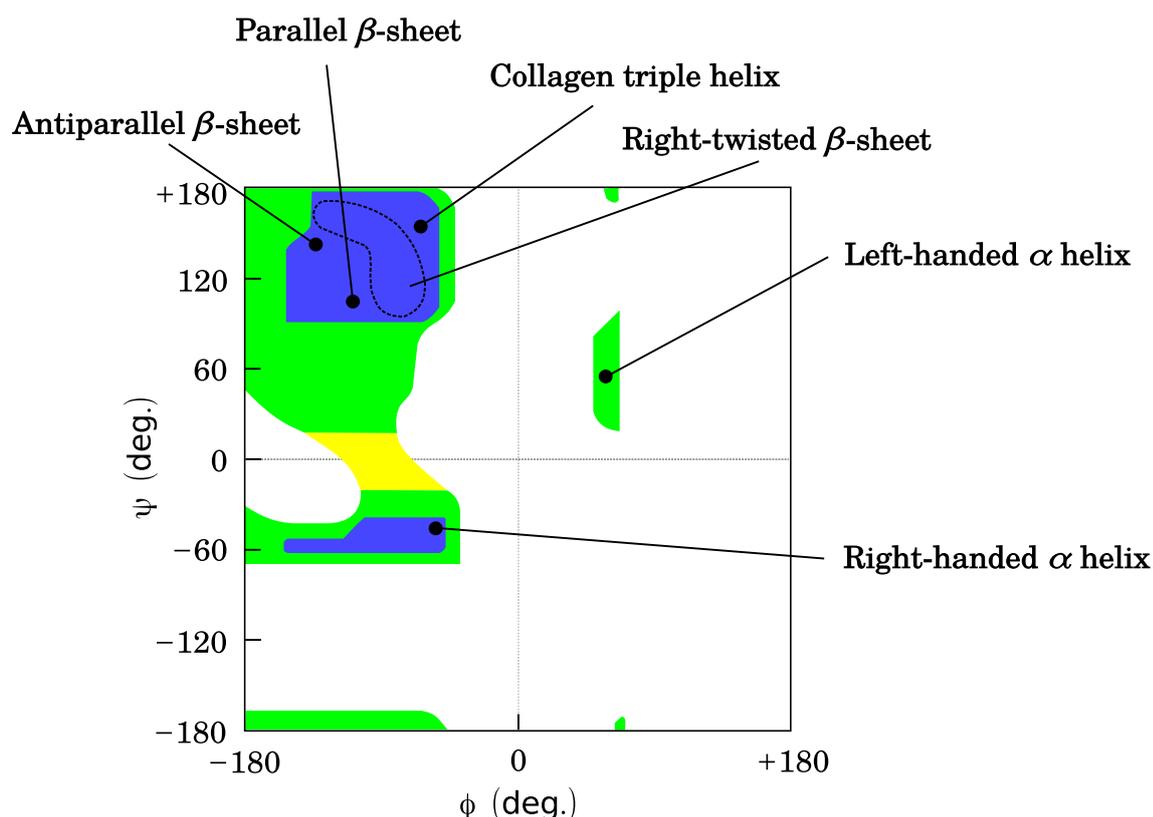


Figure 1.3: Ramachandran plot of the torsional angles in a peptide for various allowed secondary structures. The torsional angles determine the conformation of the residues and the peptide. Blue areas show the fully allowed conformations that involve no steric overlap; green areas indicate conformations allowed subject to slightly unfavourable atomic contacts; yellow area reflects conformations with significant strain [1].

1.2 Peptide aggregation

1.2.1 Amyloids

One of the most important problems in protein chemistry and molecular medicine is to understand and control the misfolding and aggregation of so-called ‘amyloidogenic’ proteins which contain sequences that interact strongly with other proteins of the same sequence, forming intramolecular assemblies which are typically β structured [20–23]. Amyloid oligomers play a key role in amyloid deposition diseases; many degenerative diseases such as Alzheimer’s, Parkinson’s, and Huntington’s, are related to undesirable amyloid aggregation [24–26].

Proteins can aggregate in regular assemblies by changing their structure from native forms to cross-beta-sheets, forming amyloids [27–29]. Although the mechanisms of protein self-assembly are often complex and still not well understood, experimental studies suggest that nucleation and growth processes are usually an important part of the aggregation kinetics [30–34]. Amyloid formation in physiological conditions is typically a process of nucleated polymerization, which begins slowly with aggregation of monomers to form oligomeric nuclei, then is accelerated with addition of monomers, formation of fibrils, and their growth [35–38]. This may include homogenous nucleation in that no foreign surface is involved, heterogeneous nucleation on the surface of some other entity, or secondary nucleation from a parent seed aggregate of the same type [39]. Homogenous primary nucleation occurs directly from a supersaturated solution and is typically dominant only during the start up phase of a non-seeded crystallization process, whereas the secondary nucleation uses crystals present in the solution as a site and source for the formation of nuclei and is observed at lower supersaturation compared to primary nucleation [40,41].

Amyloid β ($A\beta$) is a small self-assembling peptide that is crucially involved in such diseases. It can aggregate to form flexible soluble oligomers that are toxic to nerve cells [26,42–44]. Small oligomers and protofibrils also have important and strong neurological activities and targeting these oligomers might play a key role in prevention of amyloid diseases [45]. Nevertheless, $A\beta$ has a normal biological function in host organisms and may be non-toxic in its form as mature fibrils [46]. Further natural and designed amyloid-forming systems which appear to be benevolent or at least non-toxic have applications as unique biomaterials in nanotechnology and biomedicine [47–51]. The explanation for some amyloid species being non-toxic seems in many cases to be that they rapidly pass through the small-oligomer stage, producing mature fibrils which are immobile, insoluble and inert [52]. This opens the fascinating possibility that therapies for amyloid disease

might in future include *acceleration* of the amyloid aggregation process, or use of small-molecule agents to nudge the assembly into more stable polymorphs which can absorb more monomers and oligomers from solution.

1.2.2 Noncovalent Interactions

Noncovalent interactions are essential in maintaining the three-dimensional structure of proteins and nucleic acids. They are also involved in many biological processes in which large molecules bind specifically but transiently to one another [53]. These interactions also play a key role in amyloid fibril formation and their stability [54–56]. Noncovalent interactions range from ~ 60 kcal/mol for ion-ion interactions in vacuum to ~ 0.5 kcal/mol for dispersion interactions [57]. They can be classified into different categories [58, 59], such as electrostatic, hydrogen bonding, hydrophobic and van der Waals interactions.

Electrostatic Interaction

Interaction between charges is a well-known type of noncovalent interaction which is usually employed to induce structural specificity for charged peptides in formation of nanostructures [60, 61]. Electrostatic bonds based on Coulombic attractions between opposite charges lead to the formation of salt bridges with the strongest interaction energies in the range of $\sim 50 - 70$ kcal/mol in vacuum [57]. The solvent strongly influences electrostatic interactions, for example in water the strength of electrostatic interactions are weakened by about 80-fold [62] (dielectric constant of water is about 80). Minimization of the distance between two oppositely charged ions will be important in the structure of the supramolecular aggregate, even though there is no particular directionality in the ion-ion interaction. Interactions between ions and dipoles are weaker ($\sim 10 - 50$ kcal/mol) but in this case the orientation of the dipole with respect to the charge is important [58].

Hydrogen Bonding

Hydrogen bonding plays an important role in molecular self-assembly processes such as micelle formation, biological membrane structure, fibril formation and the determination of protein conformation [59, 63]. Hydrogen bonds are weaker than ionic interactions and occur between a proton donor group, which is the strongly polar group in a molecule such as F-H, O-H, N-H, S-H, and an slightly electronegative proton acceptor atom such as fluorine, oxygen, nitrogen and sulfur. Hydrogen bonds have energies in the range $\sim 15 - 30$ kcal/mol (heteroatom-heteroatom distances between $2.2 - 2.5$ Å) for strong hydrogen bonds, $\sim 3.5 - 15$ kcal/mol ($2.2 - 3.2$ Å) for moderate hydrogen bonds and below

3.5 kcal/mol (long donor-acceptor distances of up to 4Å) for weak hydrogen bonds. Also, the range of possible hydrogen bond angles for strong hydrogen bonds is 175°-180°, while it is more flexible for moderate (130°-180°) and weak (90°-150°) hydrogen bonds [58]. Among different noncovalent interactions, hydrogen bonding is probably the most important in peptide self-assembly. Conversion of peptides into diverse one dimensional (1D), 2D, and 3D nanostructures can be induced by the selectivity and high directionality of hydrogen bonds [64].

Hydrophobic Interactions

Hydrophobic interaction arises from the minimization of surface between polar/protic and unpolar/aprotic molecules which is energetically more favorable [58]. In addition to charge interactions between amino acids, their ability to form hydrogen bonds with surrounding water molecules is an important driving force for the conformation of proteins. In compact conformation of globular proteins hydrophobic groups are buried in the center to avoid contact with the surrounding water. Also, hydrophobic parts of amphipathic molecules try to aggregate to minimize their surface area in contact with water, leaving the hydrophilic parts exposed to water [64].

Hydrophobic effects are also important in guest binding by cyclodextrins. By replacing the water molecules residing inside the unpolar cavity with an unpolar guest, their interaction with other water molecules outside the cavity will be much stronger, and favorable for the enthalpy. This is also favorable for entropy because by replacing several water molecules with one guest the total number of translationally free molecules increases [58].

Van der Waals Interactions

Van der Waals forces are on the weak end of noncovalent interactions with energies about 1 kcal/mol, but the forces act between all types of atom and molecule. These forces arise from the interaction of an electron cloud polarized by adjacent nuclei. In other word they are the result of fluctuations of the electron distribution of two closely spaced molecules and can be treated as a type of instantaneous electrostatic interaction [64].

The effective range of van der Waals forces can reach hundreds of angstroms. They are attractive dispersion interactions, which decrease with the distance r in a r^{-6} dependence, and are counteracted at close range by strong electron-electron repulsion forces. Similar to the interaction between aliphatic tails in peptide amphiphiles, van der Waals forces provide an important contribution to various noncovalent interactions and they are ubiquitous in assembly systems. However, because van der Waals forces do not have the same specificity as electrostatic forces or complex interactions such as hydrogen bonding, only

a few examples engage van der Waals interactions as a dominant force for the control of self-assembly [65,66].

1.2.3 Polymorphism

In order to understand the mechanism of oligomer formation detailed structures of amyloid oligomers are needed. Amyloids are generally polymorphic at the molecular level, and a given peptide or protein can form different morphologies [12,67–70] including filaments [71], nanotubes [72], helical ribbons [73–75], twisted ribbons [74,75] and crystals [75,76]. Amyloid fibrils and crystals formed from the same polypeptide can have different arrangements of the β -strands and make different symmetry classes [2,77].

Figure 1.4 shows the eight allowed symmetries to form a two- β -sheet ‘steric zipper’ [2]. The different fibril symmetry classes are illustrated with left hands. The arrangement of β -strands within each β -sheet is either parallel (classes 1–4) or antiparallel (classes 5–8).

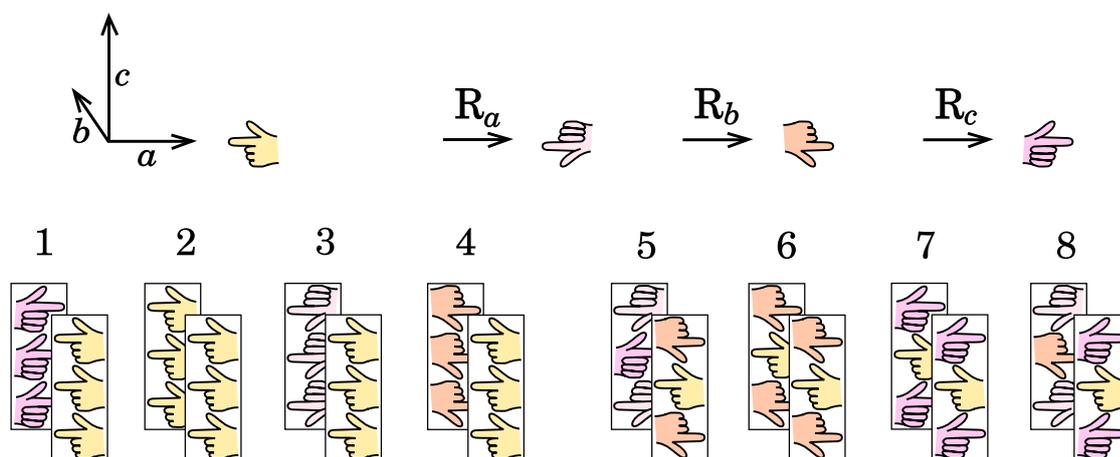


Figure 1.4: Eight different fibril symmetry classes [2] illustrated with left hands. R_a , R_b and R_c show rotations of 180° about a , b and c axis. Classes 1–4 have parallel alignment of β -strands within each β -sheet and classes 5–8 have antiparallel arrangement of β -strands within their β -sheets. Each peptide sequence may have the ability to form one or more of these morphologies.

The selection of symmetry class represents the ‘middle’ level of amyloid polymorphism. At the smaller level, amyloid assemblies may differ in sidechain conformation, or in a shift of register between chains while having the same symmetry class, and at the meso level aggregates with the same symmetry class may have different curvature or aspect ratio. Polymorphism at the level of symmetry class is of particular interest because it can be

understood physically in terms of the terminus-terminus and other interactions associated to the peptide chains, which have differing preferences for different symmetries [2].

The mesoscopic shape polymorphism of amyloid aggregates, easily inspected through a microscope, is very significant in limiting the progress of aggregation towards thermodynamic equilibrium, creating arrested states with a large proportion of smaller soluble and neurotoxic aggregates. Twisted and curved aggregates in particular have reduced rates to assemble hierarchically in comparison to rectangular aggregates, as by definition they must pay a thermodynamic cost to un-twist in order to create assembly-competent flat planes. Factors determining twist are complex: from the Ramachandran diagram (Figure 1.3) it is apparent that AP β strands have slightly less twist than P β , however the relaxed twist of a β -sheet also depends on the strand length (sheet width): the wider the sheet at a given pitch, the greater the shear at the sheet edge, therefore wider sheets will tend to have a longer pitch. Sheet-sheet assembly may take place in multiple geometries, altering the total curvature of the aggregate in ways which are unpredictable. Parallel in-register β -sheets have the special property for sheet-sheet assembly that the stacks of identical side-chains create smooth grooves on the buried surface, allowing sheets to slide past each other and alter the assembled twist in a labile manner [78]. For the (I/T)(L/F)QINS systems studied here, assembly at the sheet-sheet interface adds twist with the opposite sign to assembly in the terminus-terminus direction [75], such that cross-section aspect ratios close to one imply rectangular microcrystal-like aggregates, while very small or very large cross-section aspect ratios imply bent or twisted aggregates. It has not been shown how general this effect may be.

Although the hypothesis is extant in the literature that formation of large aggregates is protective, as these are less toxic than the equivalent number of peptides organised into a large number of small oligomers [52], it is quite feasible that diversion of oligomers into bent or twisted nano-fibrils might also be a beneficial strategy in cases that these oligomeric aggregates are themselves only weakly metastable with limited further assembly, remaining in balance with a large population of harmless (or even functional) non-amyloid individual peptides.

1.3 ILQINS, a Lysozyme-Derived Hexapeptide

The ILQINS hexapeptide is a subsequence of Hen's Egg White Lysozyme, shown empirically to be a significant driver of amyloid formation for the full-length protein [79] and also for aggregation of digestion fragments [73]. Self-assembly of ILQINS has been investigated leading to the discovery that pH, concentration, and specific biorelevant mutations

can select the mesostructure, with the mesostructure then controlling the total amount of aggregation [73, 75]. IFQINS is the human wild-type homologue of ILQINS (with enhanced *in vitro* aggregation), and TFQINS is a disease-associated mutation leading to further enhanced aggregation *in vitro* and also to *in vivo* aggregation associated with disease [80, 81] (see Figure 1.5).

Angstrom-resolution structures have been discovered for IFQINS and TFQINS based on solid-phase crystallography, with parallel β -strands within β -sheets [82], but solution assemblies and some precipitated nano- and micro-crystals of these peptides show strong signals of antiparallel β -sheet. The similar peptides GNNQQNY and NNQQNY (subsequences of the Sup35 yeast prion [83]) have been crystallized in parallel β -strands [84] and also been the subject of simulation or other studies from different groups. The results of simulations and solid-state NMR show the formation of antiparallel β -strands structures [12, 85–88] as well as parallel.

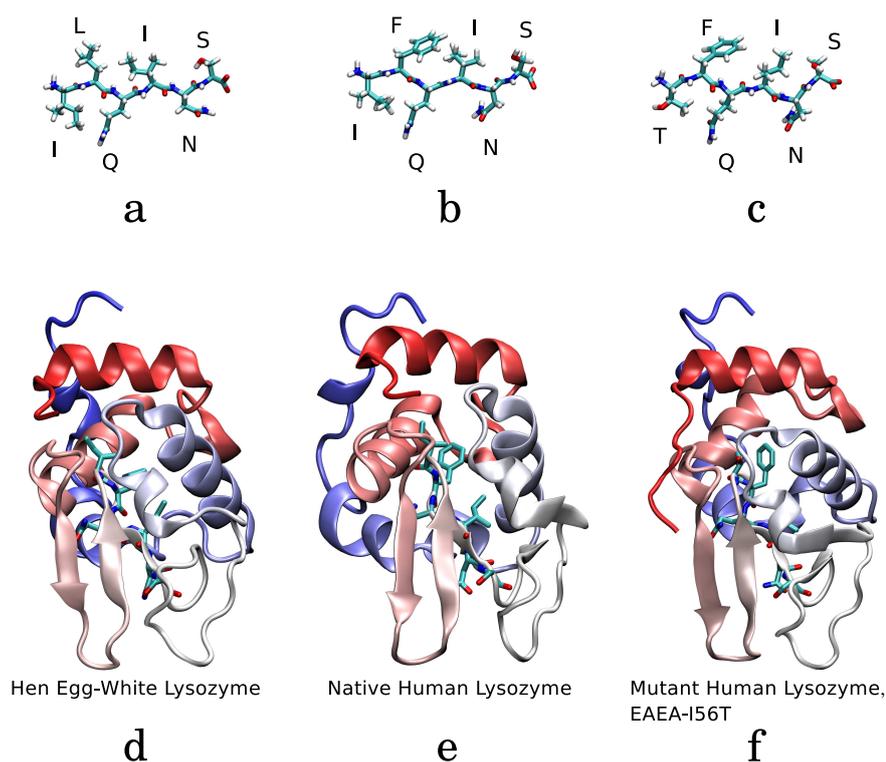


Figure 1.5: Three homologous hexapeptides: **a** ILQINS (from **d** hen’s egg-white lysozyme [3WUN]), **b** IFQINS (from **e** human lysozyme [1REX]) and **c** TFQINS (from **f** mutant human lysozyme [1IOC]). The hexapeptide sequence is exposed to solvent even in the folded protein, as it is in the active-site cleft.

The tau protein, related to Parkinson’s disease [89], contains ILQINS homologue se-

quences VQIVYK and VQIINK. Aggregation inhibitor design targeting these sequences was successful by taking into account the polymorphic steric zippers for VQIVYK and VQIINK [90], including structural information from soluble nanocrystal or fibril structures as well as from microcrystals amenable to solid-phase crystallography. Potential aggregation inhibitor design for lysozyme-related amyloidosis should therefore also take account of the polymorphic amyloid structures available. Further disease-related ILQINS homologues include the recently discovered pathogenic amyloid-former TDP-43 [91], which contains GNNQNQ and ASQQNQS subsequences.

In this project the thermodynamics and aggregation kinetics of ILQINS and related peptides are studied using molecular simulation methods. The simulation results are analysed in the context of novel experimental results relating to the bio-relevant solution structures, as well as with reference to the extant X-ray crystallography results collected from large solid-phase samples.

1.4 Methods

1.4.1 Classical Molecular Dynamics (MD)

As a tool for microscopic analysis and a complement to conventional experiments, molecular dynamics (MD) and Monte Carlo (MC) methods can be used in the hope of understanding the properties of the amyloid aggregation process in terms of the structure of molecules and the microscopic interactions between them [92–95].

The motion of a particle in classical mechanics is specified by the position and velocity of that particle. If all forces acting on a particle are known, we can specify the state of that particle at any time by integration of Newton’s equations of motion provided that we know the initial state of the particle. For many body systems like proteins and peptides, forces acting on particles can be derived from the negative gradient of the potential energy with respect to the displacement in a specified direction. We state an analytic expression for the energy of a system, known as the force field. Force field parameters usually are obtained from experimental and quantum mechanical studies of small molecules, assuming that such parameters may be transferred to larger molecules on an additive basis [96]. The force field function is composed of intermolecular and intramolecular energetic contributions, including bond stretching, angle bending, bond rotations and non-bonded terms [97]. A wide variety of force fields for biological molecules are available. Throughout this work AMBER (Assisted Model Building with Energy Requirement) force field is used.

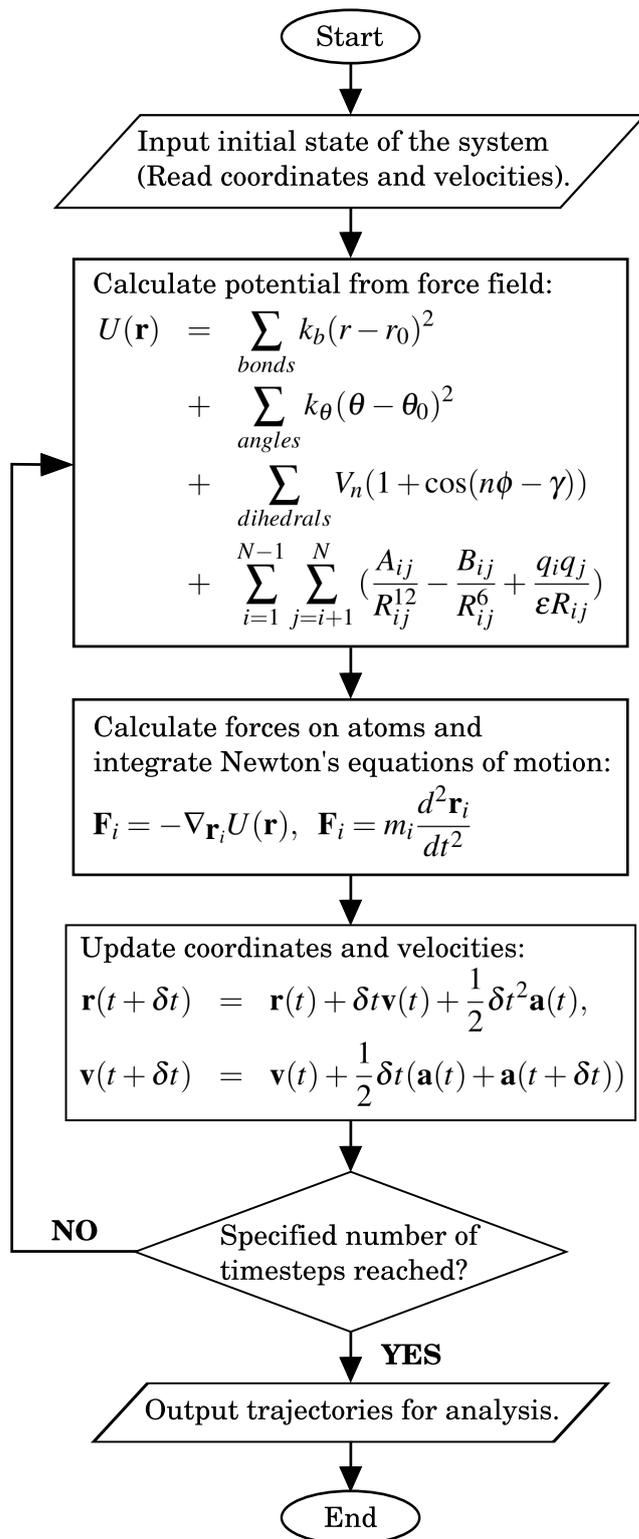


Figure 1.6: An overview of the MD procedure, using the AMBER force field and the Velocity-Verlet integration algorithm.

The simplest form of the Amber force field uses the following equation [98]:

$$\begin{aligned}
 U(\mathbf{r}) &= \sum_{\text{bonds}} k_b(r - r_0)^2 \\
 &+ \sum_{\text{angles}} k_\theta(\theta - \theta_0)^2 \\
 &+ \sum_{\text{dihedrals}} V_n(1 + \cos(n\phi - \gamma)) \\
 &+ \sum_{i=1}^{N-1} \sum_{j=i+1}^N \left(\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + \frac{q_i q_j}{\epsilon R_{ij}} \right). \tag{1.1}
 \end{aligned}$$

In Equation 1.1 the individual terms for bond lengths and angles are based on simple harmonic potentials on the basis of an energetic penalty associated with a deviation from the equilibrium value. To model the barriers associated with the rotation of bonds a torsional angle potential function is used. Coulomb and Lennard-Jones (LJ) potentials are also used to express electrostatic and van der Waals forces, respectively for non-bonding interactions.

In large systems, the solution to Newton’s equation of motion can be gained numerically by finite difference methods; the continuous time evolution of the system is broken down in small time steps such that the total force acting on each particle in the system changes little in a given timestep. An overview of the integration procedure, using the AMBER force field and the Velocity-Verlet integration algorithm is given in Figure 1.6.

1.4.1.1 Solvent Models

Atomistic MD simulation can be done in explicit or implicit [99] solvent. In explicit-solvent methods, positions and interactions of both solute and solvent atoms are explicitly computed. The particle mesh Ewald (PME) is the most commonly used explicit-solvent method for biomolecular simulations. In this method periodic boundary conditions are imposed such that no water ‘surface’ is generated. Short-range electrostatics (on lengthscales typically 8-12Å, much less than the box size) are treated directly on a pairwise basis, and long-range interactions are treated by binning the charge distribution onto a grid, such that it can be subjected to a Fourier transform [100] and electrostatic forces can be solved efficiently in the frequency domain. Forces arising from the direct-pairwise and Fourier-mesh parts of the calculation are mixed using an interpolation function.

In implicit methods, solvent is approximated as a dielectric continuum, so the number of particles to explicitly keep track of in the system is reduced drastically. An additional benefit is that the random collisions of solvent molecules to solute can be modelled using the Langevin equation, either aiming for physical realism or for enhanced sampling by increasing or decreasing the solvent friction/collision rate [101, 102].

The generalized Born (GB) solvation model is an implicit solvent model for atomistic MD simulation. The total solvation free energy of a molecule, ΔG_{solv} is decomposed into the non-electrostatic and electrostatic parts:

$$\Delta G_{solv} = \Delta G_{noel} + \Delta G_{el} \quad (1.2)$$

where ΔG_{noel} is the free energy of solvating a molecule when the partial charges of every atom are set to zero and comes from the combined effect of the favorable van der Waals attraction between the solute and solvent molecules, and the unfavorable cost of breaking the structure of the solvent around the solute. This is usually taken to be proportional to the total solvent accessible surface area (SA) of the molecule, with a proportionality constant derived from experimental solvation energies of small non-polar molecules.

ΔG_{el} is the free energy of first removing all charges in the vacuum, and then adding them back in the presence of a continuum solvent environment, which is of course more favourable for the presence of charges: this term dominates the solvation energy in most cases, certainly for ionic solutes in a high-dielectric medium such as water. The electrostatic energy depends not only on the charge distribution but on the shape of the cavity created by the exclusion of solvent: a macromolecule such as a protein may have a large volume at its core which is more akin to vacuum than to solvent in electrostatic terms. The energy to place an arbitrary charge distribution in an arbitrary solvent cavity is not analytically available and is traditionally calculated on a grid by numerical solution of the Poisson-Boltzmann equation. In MD, solution of the Poisson-Boltzmann equation at every step is however prohibitively costly, so the GB method is presented as a numerically cheaper alternative [98, 103].

In this model each atom in a molecule is considered as a sphere of radius R_i with a charge q_i at its center and the interior of the atom is assumed to be filled uniformly with a material of dielectric constant 1. The molecule is surrounded by a solvent of a high dielectric ϵ (80 for water at 300 K) and ΔG_{el} is approximated by an analytical formula:

$$\Delta G_{el} \approx -\frac{1}{2} \sum_{ij} \frac{q_i q_j}{f_{GB}(r_{ij}, R_i, R_j)} \left(1 - \frac{\exp[-\kappa f_{GB}]}{\epsilon} \right) \quad (1.3)$$

where r_{ij} is the distance between atoms i and j , the R_i are the so-called ‘‘effective Born radii’’ and correspond to the self-energy contributions of individual atoms. The atom-pair contributions are estimated by an analytical certain function f_{GB} that depends upon the effective Born radii and interatomic distance of the atom pairs [104]. The electrostatic screening effects of low salt concentration is considered via the Debye-Hückel screening parameter κ [105].

f_{GB} is chosen for a specific GB implementation, an example form is [103]:

$$f_{GB} = [r_{ij}^2 + R_i R_j \exp(-r_{ij}^2/4R_i R_j)]^{1/2}. \quad (1.4)$$

For an atom i in a molecule, R_i reflects the degree of its burial inside the molecule: for an isolated ion, it is equal to its van der Waals radius ρ_i and assuming $\kappa = 0$ for pure water we obtain the simple form of the solvation energy of a single ion:

$$\Delta G_{el} = -\frac{q_i^2}{2\rho_i} \left(1 - \frac{1}{\epsilon}\right). \quad (1.5)$$

The function f_{GB} interpolates between the limit $r_{ij} \rightarrow 0$, when atomic spheres merge into one, and $r_{ij} \rightarrow \infty$, when the ions can be treated as point charges obeying the Coulomb's law [98].

1.4.2 Replica Exchange Molecular Dynamics (REMD)

The timescale of protein aggregation is often inaccessible using straightforward MD because the stable states, defined by the minima in free energy, are usually separated by large free energy barriers, and straightforward MD simulations frequently become stuck in local, metastable minima, and therefore we must use advanced sampling techniques. One of the successful methods to enhance conformational sampling in molecular simulations is parallel tempering or REMD [106–110].

In REMD, several replicas of the system are simulated independently and simultaneously using classical MD or MC methods at different simulation temperatures [111] or Hamiltonians [112, 113]. At certain intervals, neighboring pairs are exchanged with a specified transition probability. Usually the temperature is the parameter that varies among the replicas (T-REMD) and the random walk in temperature allows the trapped conformations at a low simulation temperature to escape from locally stable states by exchanging with replicas at higher simulation temperature. This exchange is essentially an MC move in temperature space, and is accepted with the probability:

$$P_{acc}(i \longleftrightarrow j) = \min(1, \exp(-\Delta)) \quad (1.6)$$

$$\Delta = (U_i \beta_j + U_j \beta_i) - (U_i \beta_i + U_j \beta_j) \quad (1.7)$$

where the inverse temperature β is defined by $\beta = 1/k_B T$ (k_B is the Boltzmann constant) and U represents the potential energy of the system for a given configuration. The above expression enforces that lower-energy states move to lower-temperature replicas, subject

to some random mixing. Sampling is enhanced by the communication between high-temperature replicas, in which free energy barriers are reduced, and low-temperature replicas in which the dynamics are close to gradient descent. It has been recognized that (for equilibrium classical systems) increased temperature for a given replica is equivalent to inversely scaling the energy associated with its associated configurations, and that in fact further more sophisticated modifications to the energy landscape can be made in order to enhance sampling.

Usually, the replicas are constructed so that one end of the replica set is the ensemble from which sampling is wanted, and the other end is one where barriers can be crossed more easily [114]. The resulting sampling will be statistically correct and closer to ergodicity if the attempts to exchange replicas produce valid ensembles, and the probability of exchange attempts succeeding is high enough, and the resulting flow of replicas over the ensembles produces good mixing. The probability of a replica containing a given configuration depends on the potential energy and the temperature ($P(x) \propto e^{-\beta U(x)}$). If two ensembles are chosen so that their distributions of states have significant overlap, then states have a good chance to be exchanged between them.

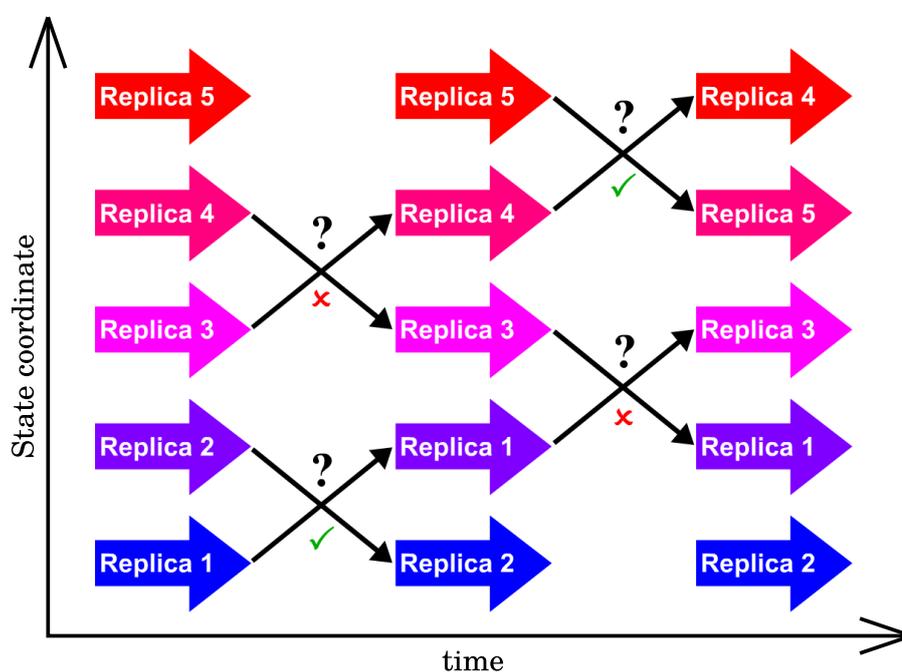


Figure 1.7: In this schematic figure colored arrows represent MD trajectories of replicas and the black arrows represent attempted swaps between replicas in state space. Question marks represent Monte Carlo exchange attempts (check marks for successful, cross for failed).

The simplest workflow for replica exchange simulations is illustrated in Figure 1.7. Each replica runs a pre-specified number of MD steps before stopping to attempt exchanges with one of its nearest neighbours. In general restricting exchange attempts to pairs is not required and exchanges can involve more than two replicas [98].

1.4.2.1 Hamiltonian Replica Exchange Molecular Dynamics (HREMD)

To investigate the early stage nucleation processes of peptides, biasing potential REMD (BP-REMD), which is a Hamiltonian REMD (HREMD) method is used. The Hamiltonian is modified by the addition of a biasing potential for each configuration, imposing a harmonic force restraining each peptide to an extended reference conformation. The reference conformation is aligned to each peptide in each MD step so that whole-molecule diffusion is not altered.

The restraint potential is defined as $U(\vec{x}, \lambda) = \frac{1}{2}\lambda|\vec{x}|^2$, where \vec{x} is the vector of $3N$ displacements from the atomic coordinates to the corresponding reference coordinates. We apply different biasing potential levels in each replica (one replica runs without any biasing potential as reference replica) and replica exchanges between biasing levels are attempted at preset intervals, and accepted or rejected according to a Metropolis criterion:

$$\begin{aligned} w(R_i \longleftrightarrow R_j) &= 1 & : \Delta U_{ij} \leq 0, \\ w(R_i \longleftrightarrow R_j) &= \exp(-\beta\Delta U_{ij}) & : \Delta U_{ij} > 0. \end{aligned} \quad (1.8)$$

The energy change for a candidate exchange event is:

$$\Delta U_{ij} = [U(\lambda_i, \vec{x}_j) + U(\lambda_j, \vec{x}_i)] - [U(\lambda_j, \vec{x}_j) + U(\lambda_i, \vec{x}_i)].$$

At one end of the replica set, one replica runs without any biasing potential as the reference replica which we want to sample from and at the other end a replica runs with nearly-rigid extended peptides which have no intra-molecular hydrogen bonds and are thus highly aggregation prone. The details of specific applications of the method will be discussed in coming chapters.

1.4.3 Kinetic Monte Carlo (KMC)

Kinetic Monte Carlo (KMC) is a stochastic algorithm used to simulate the time evolution of dynamic systems from state to state [115]. In this work the event-driven Doob-Gillespie [116] algorithm is used to investigate the evolution of oligomer formation. This algorithm is based on the collision of anisotropic bodies within a reaction vessel. It is assumed that all reactions within the Gillespie framework are initiated by at most two bodies

and reactions involving more than two bodies can be treated as a sequence of binary reactions. The algorithm begins by initializing the number of bodies in the system, reaction rate constants, and random number generators. The second step is generating random numbers to determine the next reaction event, and also the wait time to that event. The next step is updating the time and the counts for each species present, based on the reaction that occurred. We iterate the procedure by going back to the second step unless the simulation time has been exceeded.

In order to define reaction rate constants for the KMC algorithm, it is necessary to have diffusion constants and surface cross-sections for each species (to find collision rates) and also to define energy barriers for joining following a collision, and also to define energy barriers for spontaneous fission.

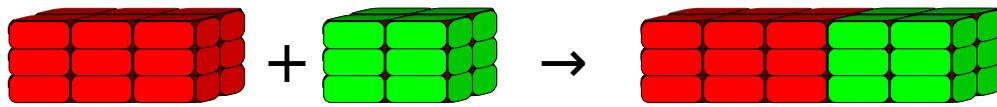
Given that assembly in the hydrogen-bonding axis is considerably stronger than in lateral axes, leading to approximately rod-like aggregates, we employ the formula for diffusion of rod-like bodies given by Ortega and de la Torre [117]. To reduce the number of distinct species which need to be tracked, we consider only processes for assembly or splitting of rigid cuboidal assemblies of β -structured peptides, neglecting conformational changes. The rate for association of two oligomers i and j , drawn respectively from populations of size N_i and N_j , having matching-sized planes in a can be defined via the diffusive collision rate for anisotropic bodies with surface $bL_b cL_c$ in a volume V , subject to an exponential barrier of $3k_B T$ representing the entropic cost to remove translational and rotational degrees of freedom by joining two rigid bodies (a , b , and c are terminus-terminus, sidechain and hydrogen bonding axes respectively and in this formula they are lattice parameters or the dimensions of a cuboidal unit cell. L_a , L_b and L_c are the number of unit cells along a , b , and c axes respectively).

From this, we are able to define a manageable system of Arrhenius-type rate equations, suitable for Monte Carlo sampling via the Doob-Gillespie method. Here rate equations are presented in the form directly useful for simulation, such that the units are s^{-1} :

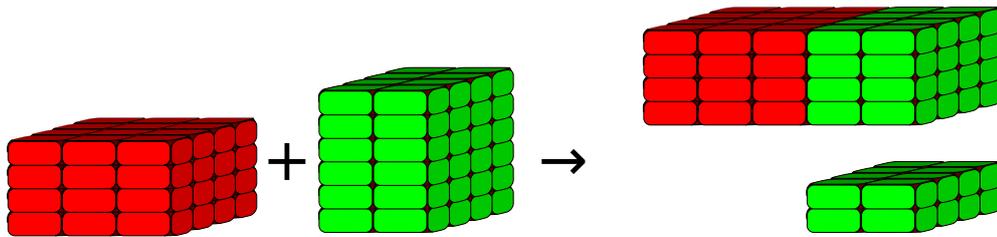
$$r_{aa} = \frac{2N_i N_j}{V} e^{-3} (D_i + D_j) \sqrt{2bL_b cL_c}, \quad (1.9)$$

where D_i and D_j are the diffusion constants [117].

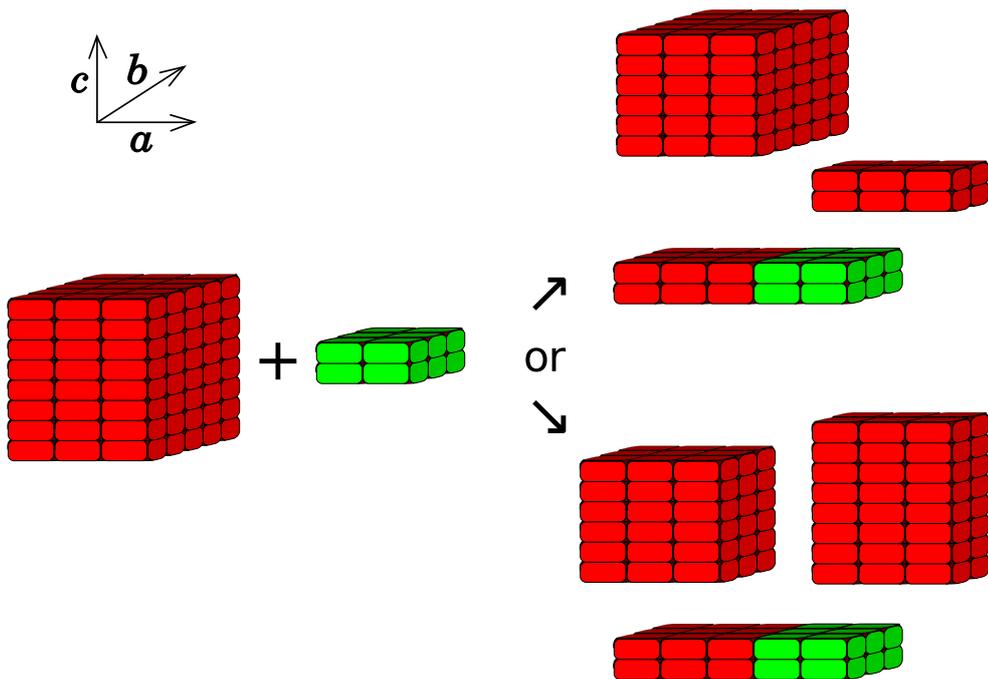
To calculate a rate for the splitting processes, an energetic scale is necessary. Binding free energies for interfaces in the a , b and c axes of the simulated microcrystal structures can be found by calculating the free energy of blocks of peptides separately and joined together, eg: $\Delta G_c^\circ = [G_{118}^\circ - (G_{114}^\circ + G_{114}^\circ)]$, where subscript triplets are the dimensions of a peptide block or sub-block.



(a) 2 to 1: collision



(b) 2 to 2: collision-and-split



(c) 2 to 3: collision and two splits

Figure 1.8: Collision of two species along a axis (a) Two species have matching-sized planes in a and they join together after collision. (b) Two species have mismatched planes in a , but they have the same length in b and the collision requires a split in the c axis. (c) Two species have mismatching-sized planes with different lengths in b and c ; collision requires a split first in b , then another split in c or a split first in c , then another split in b . Depending on the energy cost of the two reactions of panel (c), only the least expensive is implemented in order to control the complexity of the simulation.

The rate for a cuboidal oligomer of size $L_a \times L_b \times L_c$ unit cells to split along some plane perpendicular to the a crystal axis is defined as:

$$r_a = \frac{1}{\tau_0} N(L_a - 1) e^{\frac{L_b L_c \Delta G_a^\circ}{k_B T}}. \quad (1.10)$$

Here τ_0 is a constant which sets the timescale (taken as the single-peptide diffusion time [117]) and N is the number present of the given species. The term in the exponential rapidly becomes large and negative as the area of the interface to be cleaved grows. Rates for splitting in b or c are available by permuting the indices.

Although simulation of non-cuboidal aggregates was avoided due to the large number of different species which would then need to be tracked, an approximation to this was made by allowing non-matching faces to join providing that the non-cuboidal aggregate formed then splits to form cuboidal products. Figure 1.8 illustrate the three different possibilities for collision along the a axis.

1.4.4 Molecular Graphics

In this work molecular graphics were prepared using PyMOL [118] and VMD [119]. Secondary structure types for structure images were assigned using DSSP [3]. We have also used Inkscape as a free and open-source vector graphics editor to create or edit vector graphics.

1.5 Overview

In the following chapters the aggregation behaviour of the ILQINS, IFQINS and TFQINS peptides is quantified and explored. In chapter 2 early-stage competition and cooperation between antiparallel and parallel β -sheet aggregation is identified and examined using MD and REMD simulations, verified by solution X-ray diffraction (wide-angle X-ray scattering, WAXS). In chapter 3 the presence of mature (late-stage) parallel and antiparallel structures in solution is shown for the IFQINS peptide by comparing simulation to WAXS, and mechanisms for concentration-driven competition between large AP and P aggregates are investigated using MD and KMC, for systems up to millions of coarse-grained peptides in size. The thesis closes with a discussion chapter 4 in which a synthesis of the results is presented and implications of the findings are discussed.

Chapter 2

Antiparallel Strand Alignment as the Default for Amyloid Oligomers at Early Times

Amyloid aggregates of a given polypeptide may differ at the atomic or the meso scale. The most basic difference is between parallel and antiparallel alignment of peptide strands. Antiparallel contact between short peptides should be more favourable, due to the proximity of like charges in parallel alignment, however parallel β -sheet is most often observed via crystallography of mature aggregates. We combine solution X-ray with molecular dynamics for a set of hexapeptides and find that antiparallel structure dominates for the small oligomers. Initial formation of sheets is more favourable with antiparallel structure, while the assembly of two or more existing sheets is more favourable for parallel oligomers. The present peptide system, similar instances from the literature, and the basic physics of peptide interaction all argue that antiparallel order should be the first formation of amyloid in most or all amyloid aggregation processes, regardless of the structure of the final mesoscale aggregate.

2.1 Introduction

Small assemblies of peptides, known as amyloid oligomers, are now verified as toxic species in multiple neurodegenerative diseases [120]. Oligomers are hard to characterise. They may be on or off-pathway to the highly stable mature aggregates, and they may have a variety of structures including parallel or antiparallel peptide strands, in alpha or beta conformations. Here a computationally generated antiparallel β -sheet conformation is presented and linked to observed solution X-ray diffraction for the given peptide system, showing that this conformation is dominant in the early phase of aggregation. Thermodynamic

calculations are used to explain that this early dominance of antiparallel structure should be common.

Toxic amyloid oligomers play a key role in Alzheimer's, Parkinson's and other degenerative diseases [26, 42–44], while functional amyloids can be applied as unique and versatile biomaterials in nanotechnology and biomedicine [48–50]. Detailed knowledge of morphology and structure of amyloids is necessary to fully understand the process of oligomerization and fibrillization (see Chapter 1). It has been proposed that α -sheet, differing from β -sheet in that hydrogen bond donors/acceptors are aligned on opposite sides of the chain, is an important contributor to amyloid- β protein toxicity and early-stage aggregation [121]. High-temperature simulations have also indicated α -sheet in early lysozyme aggregation [122], with an antiparallel alignment. No α -sheet is observed in the present work, however the discussion of the importance of antiparallel alignment is to some extent independent of the α/β question.

Allowed arrangements for pairs of β -sheets (forming 'steric zippers' or 'Q-N zippers' for the present peptides, where the QxN sequence motif strongly joins pairs of sheets) have been catalogued as a set of eight 'symmetry classes' [2, 77, 123]. In the introduction chapter, these symmetries were presented diagrammatically (Figure 1.4). Here, Figure 2.1 shows a set of full (space-filling) crystal lattices generated based on these symmetry classes. A given symmetry class is compatible with multiple expanded lattices, however solution scattering suggests unit cell dimensions of aggregates in solution are comparable in the observed systems to a single steric zipper level (one pair of left hands) [75], supporting the choice of the simplest, translation-only, lattice.

The I₅₆LQINS₆₁ hexapeptide subsequence of hen's egg white lysozyme has been shown to be a significant driver of aggregation in digested or full-length lysozyme [73] and to have controllable mesostructure-scale polymorphism, with mutations, pH, and initial peptide concentration being used to select *in vitro* between twisted fibrils (low total aggregation) and rectangular rod-like fibrils or microcrystals (with higher total aggregation) [75], even preserving roughly the same atomistic structure and contact topology. Parallel β -sheet atomistic structures for the ILQINS homologues IFQINS and TFQINS have been documented using solid-phase X-ray crystallography [82], but solution-phase X-ray scattering from these peptides (and also ILQINS) shows signals which consistent with antiparallel structure in solution, demonstrating atomistic-level polymorphism alongside the mesostructural polymorphism which has already been discussed.

The argument of this chapter is that antiparallel structure has a significant thermodynamic advantage relative to parallel structure at the early stages of aggregation. Simulation studies confirm the formation of antiparallel β structures in similar systems such as

GNNQQNY which have been crystallized in parallel β -sheet [12, 86, 87]. The amyloid beta [124–126] and the alpha synuclein [127] peptides have been shown to form antiparallel (or mixed parallel/antiparallel) oligomers but parallel fibrils, with important implications for our understanding of neurodegenerative diseases and our strategies for molecular therapy. Only slightly different fibril morphologies may be associated with substantially different diagnoses of amyloid disease [128].

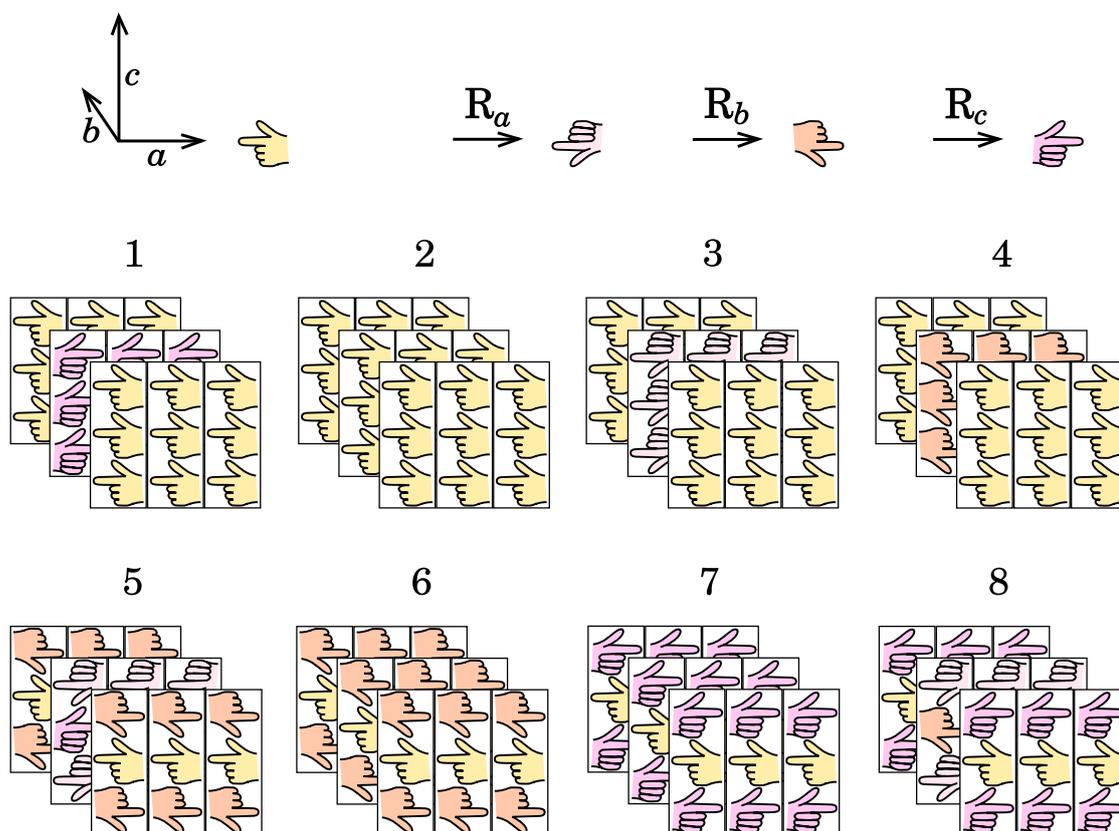


Figure 2.1: Amyloid-like lattices based on the eight different symmetry classes, illustrated with images of left hands. R_a , R_b and R_c show rotations of 180° about a , b and c axis. Classes 1–4 have parallel alignment of β -strands within each β -sheet and classes 5–8 have antiparallel arrangement of β -strands within their β -sheets. A given peptide sequence may have metastable aggregates with one or more of these symmetries. A given symmetry defines only the sheet and steric zipper relationships: multiple space groups not shown, and also intra-peptide degrees of freedom, may be compatible with a given class.

Here MD simulations of the different symmetry classes of ILQINS aggregates are used to find the atomistic structure most consistent with the results from solution wide angle X-ray scattering (WAXS) experiments. The most consistent structure (based on calculated scattering profiles) has antiparallel β -sheet conformation, contrary to the crystallography

[82] and to the earlier proposed oligomer/nanocrystal structure [73].

The early stage self-assembly mechanism of ILQINS and its homologues is investigated via Hamiltonian replica exchange molecular dynamics (HREMD) simulations. The majority of the initial single-sheet aggregates for these systems at room temperature are observed to be antiparallel. Higher stability of the AP sheets is confirmed relative to single P β -sheets via un-accelerated MD simulations.

2.2 Results

2.2.1 3D Aggregates of ILQINS in Different Symmetry Classes and Corresponding WAXS Curves

In order to make a search for the best 3D ILQINS structure in relation to the WAXS data, one ILQINS model nanocrystallite for each of the eight symmetry classes in Figure 2.1 was built. All structures were based on the class 1 structure already refined to yield a qualitative match to scattering data [73]. The new structures in classes 2–8 were subjected to rotations so as to fulfill the appropriate symmetries and also to minor alterations applied by hand so as to reduce clashes and generate favourable contacts (see Figure 2.2 for the eight initial structures).

The calculated scattering changed substantially in the course of the simulation for each system, even in class 1 which did not show major structural rearrangement to the naked eye. Among the calculated scattering curves, the structure with symmetry class 5 ended the simulation in a state overall most consistent with the WAXS experiment (Figure 2.3), although we should note that the experimental scattering showed a peak at 4.8Å which is a signature of strands in P β alignment, but is not seen in the AP β structure due the 9.7Å distance on the c axis between translated equivalent peptides in this geometry. Absence of the 4.8Å reflection in antiparallel amyloid has been confirmed elsewhere via WAXS on AP α -synuclein fibrils [129].

The inset of Figure 2.3 shows that the Bragg reflections between 20Å and 10Å which are relevant for unit cell geometry agreed better with the AP structure than with the P structure. The large peaks at $\approx 20\text{\AA}$ and $\approx 10\text{\AA}$ are hypothesised to be related to unit multiples of individual axes (i.e. 100, 200, 010, 020), and are fit by a variety of available peptide structures as the a and b lattice parameters are roughly constant across structures (and roughly equal to each other). The two smaller peaks at $\approx 17.5\text{\AA}$ and $\approx 13.4\text{\AA}$ are more difficult to assign but have observed to be sensitive to γ indicating that they involve mixing of two or more non-zero Miller indices, the AP structure class 5 was the only one to stably capture these two peaks.

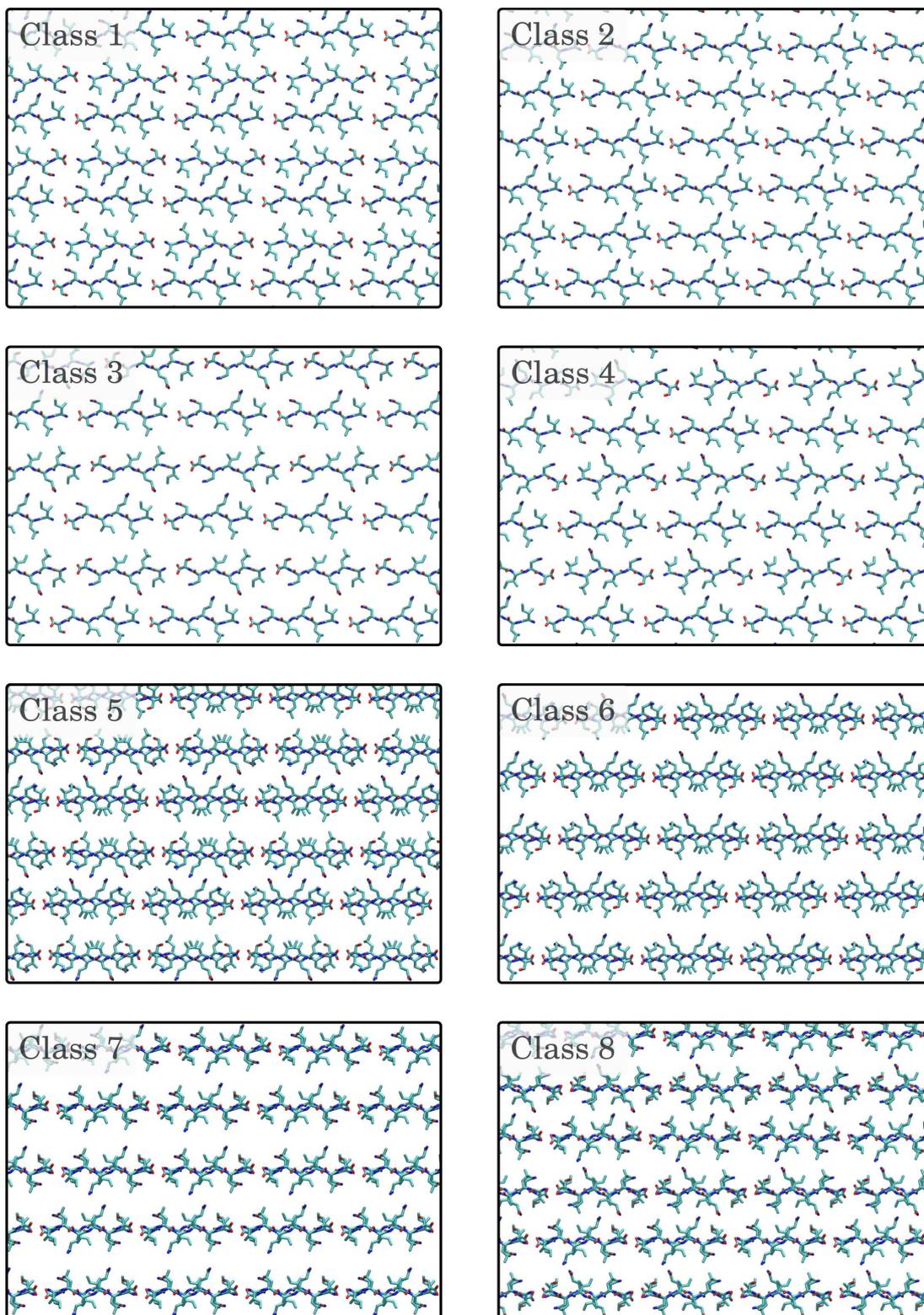


Figure 2.2: Eight initial structures before MD simulations corresponding to the symmetry classes in Figure 2.1. View is an orthoscopic projection along the c (hydrogen-bonding) crystal axis, such that the $P\beta$ structures 1-4 appear to be a 2D wallpaper.

As both of the Class 1 and Class 5 structures were missing observed peaks, until further refinement can generate a single structure which reproduces the entire scattering it is assumed that the real solution contains some proportion of both structures. As seen in Figure 2.1, symmetry class 5 has an antiparallel alignment of β -strands within β -sheets. Figure 2.4 shows this structure in detail. The structure arrived at through MD and WAXS does not have the status of 1Å-resolution crystallography outputs: the limited information from solution scattering compared to crystallography means that while a match of calculated to observed scattering is information in favour of a given structure being correct, further computational analysis is needed to provide thermodynamic or kinetic explanations of why it should be formed.

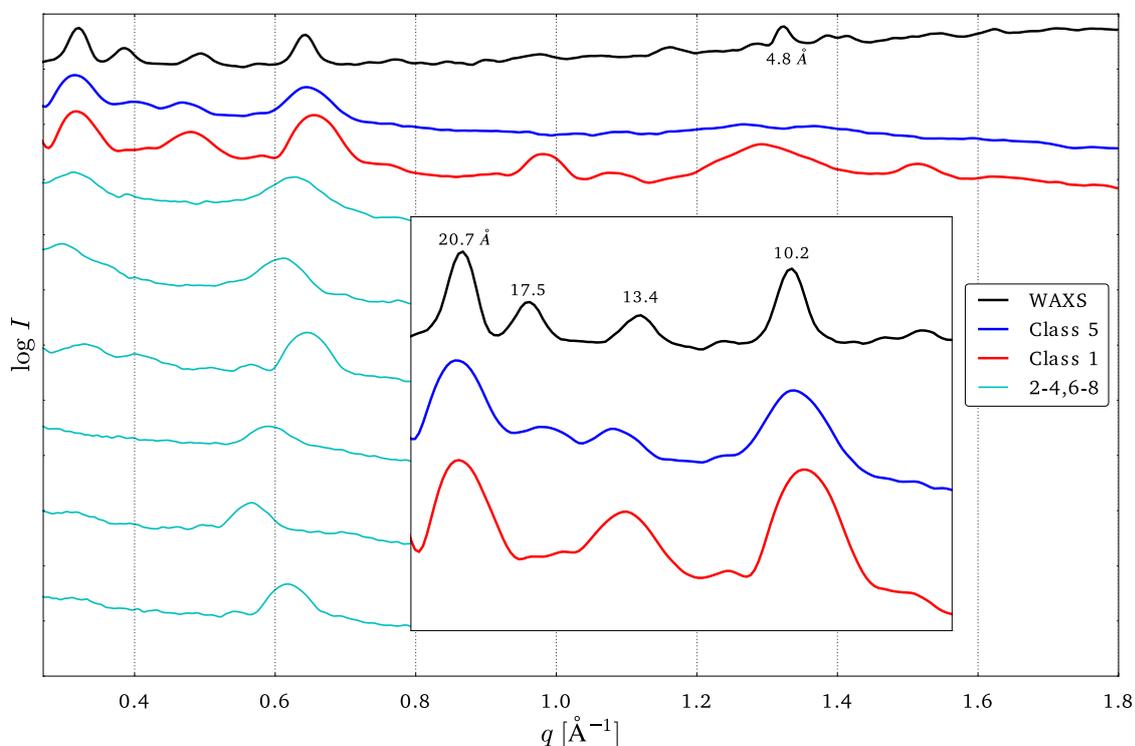


Figure 2.3: WAXS of ILQINS peptide solution recorded after 24 h of self-assembly (black), and calculated scattering for the MD snapshot of the 1296-peptide nanocrystal ($6 \times 12 \times 18$) taken at 10 ns (AP class 5: blue, P class 1: red). The AP lattice parameters are $a = 20.6\text{\AA}$, $b = 19.1\text{\AA}$ and $\gamma = 82^\circ$, peaks relevant to lattice shape are annotated with the corresponding Bragg spacing. *ILQINS WAXS data supplied by Nicholas P. Reynolds, Swinburne University of Technology, Australia.*

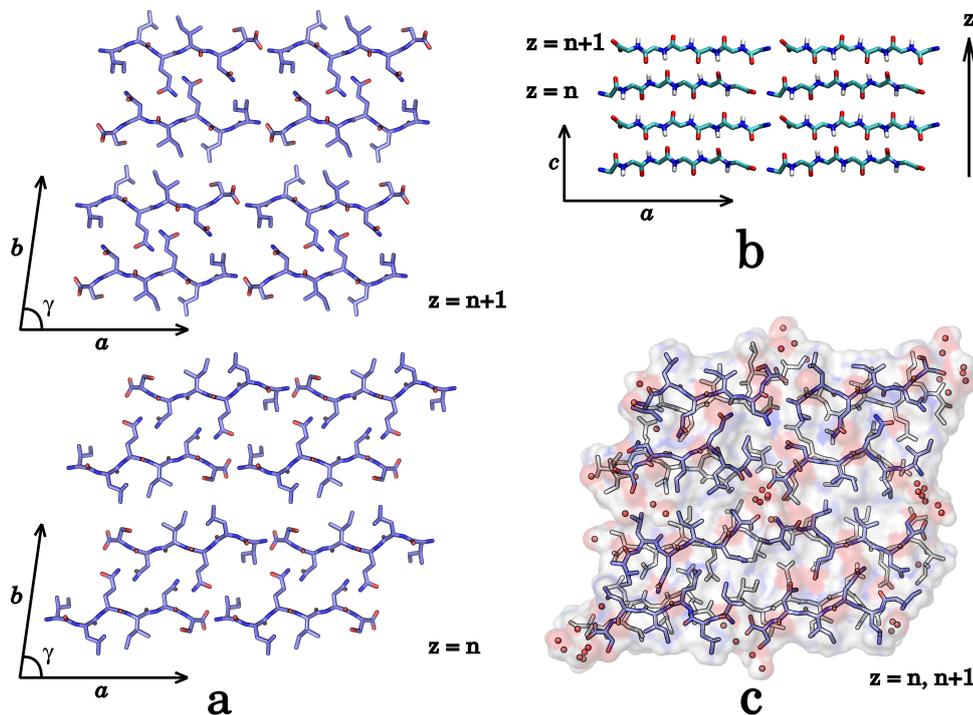


Figure 2.4: Class 5 AP structure. **a** Two sequential planes perpendicular to the c axis: termini are in close contact on the a axis, but not across the steric zipper. **b** plane perpendicular to the b axis, showing backbones of peptides only. Termini are also in contact along the c axis and hydrogen bonding is strong. **c** A molecular dynamics snapshot with semi-ordered dynamic water molecules shown as orange spheres. The AP lattice parameters are $a = 20.6\text{\AA}$, $b = 19.1\text{\AA}$ and $\gamma = 82^\circ$.

2.2.2 Energy Decomposition

As an initial test of the idea that AP structures should lead aggregation even though P structures are more common as insoluble fibrils or crystals, we evaluate binding free energy gains to join AP and P β -sheets in the c axis and at the steric zipper interface. This is done by calculating the difference between free energies of joined sheets of peptides (for example two four-peptide sheets making an eight-peptide sheet) and separated sheets (for example, the same two four-peptide sheets no longer in contact) e.g.:

$$\Delta G_c^\circ = [G_{118} - (G_{114} + G_{114})]$$

where integer subscript triplets are the number of peptides in each dimension of a rectangular nanocrystal. Reference block free energies G_{ijk} are calculated as averages over not less than 50 blocks sampled from the converged part (the tenth nanosecond) of the MD simulation. After a block is ‘cut’ from the simulation system, its energy is minimised in a

continuum solvent [130], so that the final free energy accounts for the electrostatics of solvent exposure, and also contains at least some of the appropriate physical entropy change from creating an interface, particularly that related to ordering of the solvent.

Table 1 shows the interface formation energy compared for class 1 and class 5 structures. The larger free energy gain in c to stabilise AP structures is consistent with the tendency to see single-sheet AP oligomers more than P, while the stronger steric zipper formation energy (ΔG_{zip}°) for P structures is consistent with eventual dominance of P β crystals or of P multi-sheet fibrils. Interestingly, the pattern becomes more pronounced in line with increasing amyloidogenicity [75] of the sequence.

			ΔG_c°	ΔG_{zip}°
ILQINS	P	class 1	-28.1(2)	-18.6(1)
	AP	class 5	-28.9(2)	-16.3(2)
IFQINS	P	class 1	-27.1(1)	-20.5(1)
	AP	class 5	-29.3(2)	-16.4(3)
TFQINS	P	class 1	-26.1(1)	-21.7(2)
	AP	class 5	-29.5(2)	-13.1(2)

Table 2.1: Standard binding free energy gain to construct a buried interface in the direction of the beta-sheet hydrogen bonding axis (ΔG_c°), and the sidechain steric zipper interface (ΔG_{zip}°). Units are kcal/mol/peptide buried by the interface. AP is stronger in terms of sheet elongation but is weaker in zipper stability.

2.2.3 Hamiltonian Replica Exchange Molecular Dynamics (HREMD)

To further investigate the early aggregation process of each system of peptides (ILQINS, IFQINS and TFQINS), making sure to place no initial assumption on the structure formed, a Hamiltonian Replica Exchange Molecular Dynamics (HREMD) method was implemented (as described in the introduction 1.4.2 and in the detailed methods below 1.4.2.1). Each system, each of 64 replicas, contained 64 peptides confined to a spherical volume of 50Å radius, giving an effective concentration of 0.2 M. With only 64 peptides, the full aggregation process cannot be investigated however accelerated conformational search via this method allows the initial steps to be probed. The majority of aggregates formed in the HREMD calculations had antiparallel alignment of peptide strands, which is consistent with the ΔG calculations and with the terminus-terminus interactions: each uncapped peptide has an unsatisfied charge at the end, favouring antiparallel alignment.

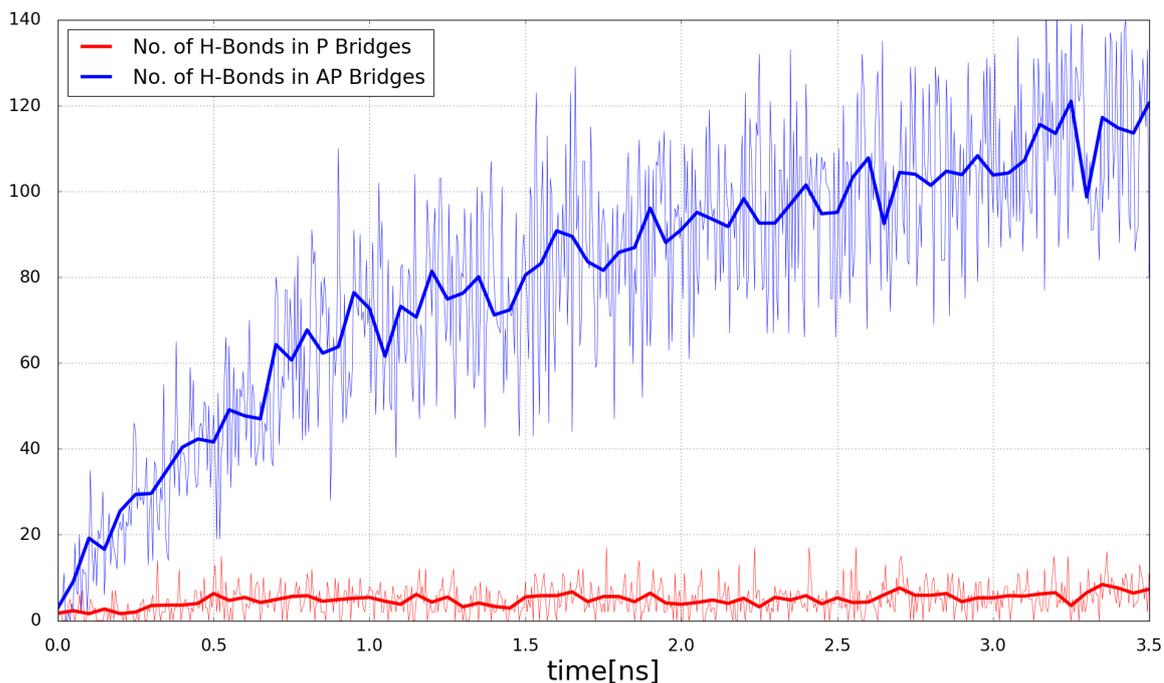


Figure 2.5: Number of hydrogen bonds in parallel (red) and antiparallel (blue) bridges calculated by DSSP [3] analysis for $\lambda = 0$ during the simulation time confirms that the majority of aggregates are forming in antiparallel structures. Thick lines indicate a moving average over 50 ps

Dominance of the antiparallel aggregates during the simulation time is confirmed by DSSP [3] analysis for calculating the number of hydrogen bonds in parallel and antiparallel bridges. Time evolution of the number of H-bonds in P and AP bridges for $\lambda = 0$ shows that the majority of aggregates are in antiparallel structures. Figure 2.5 shows the result of the analysis for ILQINS system.

Figure 2.6 (a-c) shows the extended reference conformations and Figure 2.6 (d-f) shows representative single β -sheets for each system taken from the simulations.

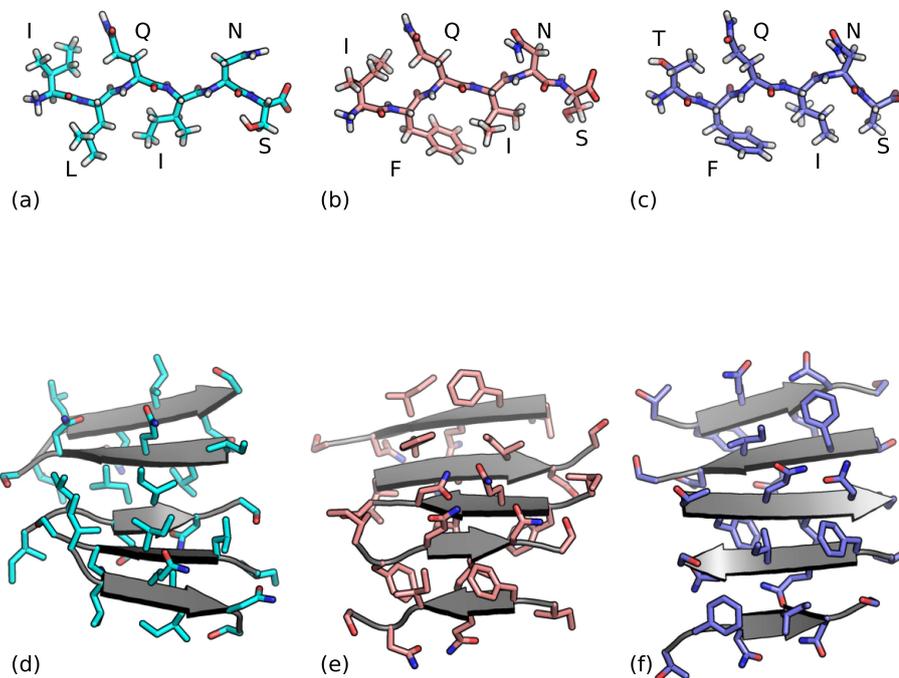


Figure 2.6: (a-c) Extended reference peptides used for HREMD simulations and (d-f) representative AP single β -sheets formed in these simulations. Sidechain orientation in the AP sheets initially formed is not regular: the structures (d-f) are mixed between the intrasheet orderings of classes 5/6 and 7/8.

2.2.4 Classical MD of P and AP Single β -Sheets

We confirm the stability of the antiparallel aggregates by implementing classical molecular dynamics simulations up to 100 ns in explicit water in 300 K and 1 atm, finding greater configurational stability for the antiparallel rather than parallel single β -sheets. Final structures from these simulations are shown in Figure 2.7.

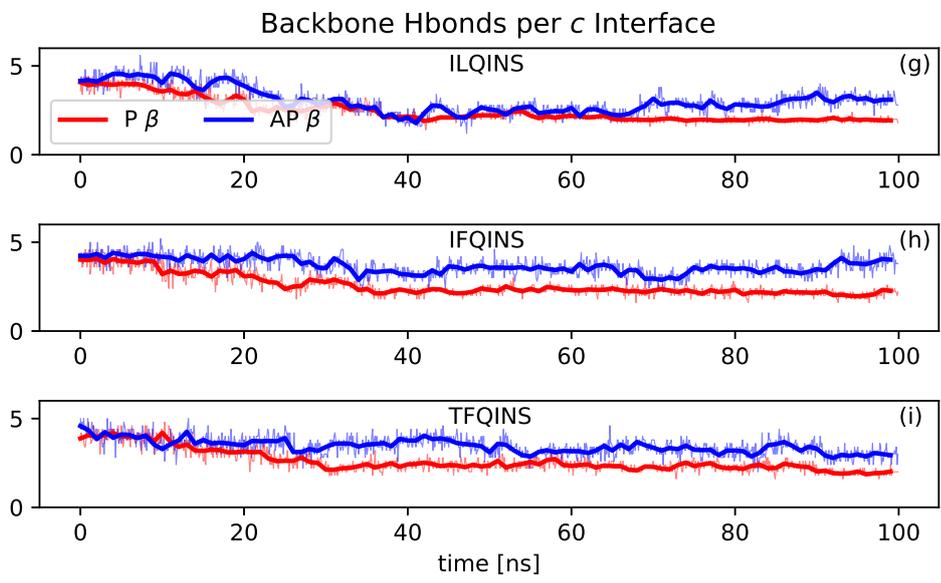
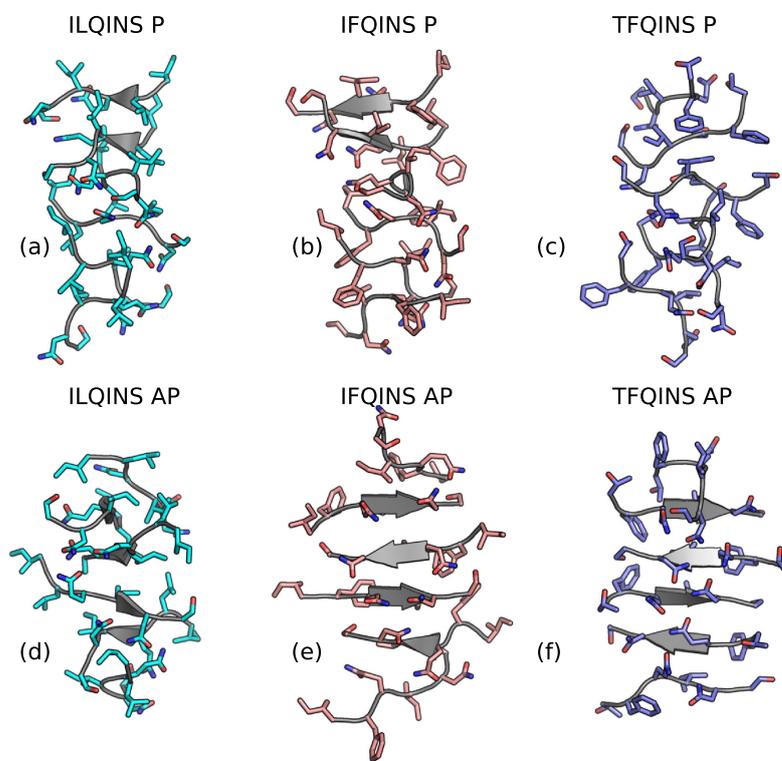


Figure 2.7: Classical MD over 100 ns in explicit water shows that antiparallel single β -sheets have greater configurational stability than parallel. (a-f): P and AP β structures consistent with classes 1 or 2 (P) and 5 or 6 (AP), at 100 ns for the three sequences studied. (g-i): backbone hydrogen bonding declines more rapidly for the P structure in each case. Thick lines indicate a moving average over a 1 ns window.

2.3 Discussion

We have analysed the aggregation of a set of amyloidogenic lysozyme-derived peptides. Calculated scattering profiles show that the symmetry with best agreement to solution WAXS experiments has antiparallel arrangement of β -strands within the β -sheets, although indications of parallel β -sheet are also detected, suggesting a mixture of parallel and antiparallel sheet at early stages of aggregation. Accelerated molecular dynamics was used to search the conformational space for small assemblies (not more than 64 peptides) and the peptides ILQINS, IFQINS and TFQINS all formed single-sheet antiparallel structures which were stable on longer timescales (order 100ns) than the equivalent parallel structures. This observation is consistent with growing evidence from the wider literature that antiparallel oligomers precede parallel aggregates for many amyloid systems [12,86,87,124–127]. This observation is physically motivated by the stronger electrostatic interaction of two extended peptide strands in an antiparallel rather than parallel arrangement, competing with the countervailing effect (observed for the present systems, and probably very common) that parallel β -sheet, with its smoother side surface, should have more favourable sidechain-sidechain stacking in the lateral phase of assembly. The formation of a stable lateral steric zipper is a crucial stage in amyloid self-assembly, as the doubled thickness therefore roughly doubles the energetic cost to break the growing aggregate perpendicular to the c axis, therefore squaring the timescale for which it can be expected to endure in solution.

This P-follows-AP effect should be quite general to amyloid assembly, particularly for small peptides where the relative importance of the termini is larger, but evidence also supports this effect for the much larger $A\beta$ [124–126] and α -synuclein peptides [127]. For longer peptides, whatever charges or whatever electric dipoles parallel to the strand axis are present, they cannot prefer a parallel in-register alignment because such a structure necessarily stacks like charges with like.

Counterexamples to the discussed phenomenon do exist, for example similar accelerated MD simulations to those documented here have found that for a pair of human Islet Amyloid Polypeptide (hIAPP) chains in solution, parallel- β dimerisation is the initial step [131]: this exception manifests what is a quite common pattern in amyloid formation, of each chain making a strand-turn-strand ‘horseshoe’ motif, such that the interface labelled *zip* in the present work (strong for $P\beta$) is formed at the same time as, or before, the interface labelled *c*. In the case that axial and zipper ordering are cooperative, or that zipper ordering precedes axial ordering, the arguments presented here do not apply.

2.4 Materials and Methods

2.4.1 MD for 3D Aggregates and Scattering Calculations

Each model crystal had 1296 ($6 \times 12 \times 18$) peptides along the a (terminus-terminus) $\times b$ (sidechain) $\times c$ (hydrogen bonding) axes. Each structure was immersed in a periodic box of TIP3P atomistic water [132], then relaxed for 10 ns by MD simulation at 300 K and 1 atm using the ff14SB atomistic forcefield [133] and the pmemd software [134]. Scattering was calculated by an orientationally averaged Fourier transform of the electron density, using CRY SOL [135].

2.4.2 Energy Decomposition

The formula applied to estimate a ‘docking’ free energy assumes linearity with the number of peptides buried by the interface and independence with respect to the block size in the axis perpendicular to the cut plane. This assumption should hold approximately true for aggregate sizes above the Bjerrum length. The Bjerrum length is $\approx 7\text{\AA}$ when measured through water, or $\approx 40\text{\AA}$ measured through the weaker dielectric of a peptide assembly (assuming $\epsilon_r = 15$ for a small amyloid oligomer). Beyond the use of non-polarisable classical forcefields, the effective length at which linearity sets in has been found using quantum methods as $\approx 15\text{\AA}$ (3 cells) along the c -axis [136], implying an even shorter effective Bjerrum length than assumed. Cooperativity (non-linearity) arises mainly in the hydrogen bonding direction, and should be stronger for AP than for P fibrils due to the less pleated alignment of dipoles, so should magnify rather than diminish the size of the effect in ΔG_c discussed here.

The a and b axes were calculated by splitting blocks of size $2 \times 4 \times 3$ peptides down the middle. The zipper axis was calculated by splitting a geometry $1 \times 2 \times 14$, the longer protofibril being chosen to minimise end effects due to an overhang of 2.4\AA at the zipper interface in some candidate structures examined. The c axis was calculated by splitting a geometry $1 \times 1 \times 14$, under the assumption that aggregation in c precedes assembly in other directions, thus the most relevant regime of ΔG_c is for single-sheet assembly.

2.4.3 Hamiltonian Replica Exchange MD

The HREMD search method [112, 113] was implemented using the NAB (Nucleic Acid Builder) molecular manipulation language [137]. A generalized Born model [130] was used to represent the solvent as a continuum. Sixty-four replicas of each system were run simultaneously using the developed NAB program and the Amber ff14SB forcefield [133].

The modification to the Hamiltonian for replicas $i > 0$ was to add a harmonic restraint driving each peptide towards an extended reference configuration, with the position and orientation of the reference superimposed on the peptide at each step but the intra-chain degrees of freedom for the reference held constant. The restraint potential was defined as $U(\vec{x}, \lambda) = \frac{1}{2}\lambda|\vec{x}|^2$, where \vec{x} is the vector of $3N$ displacements from the atomic coordinates to the corresponding reference coordinates, $\lambda_i = 0.032(i/63)^2$, and the replica index i runs over $[0..63]$. The given functional form for λ was verified empirically to provide good mixing between replicas (see Figure 2.8). Replica exchanges were attempted at 5 ps intervals.

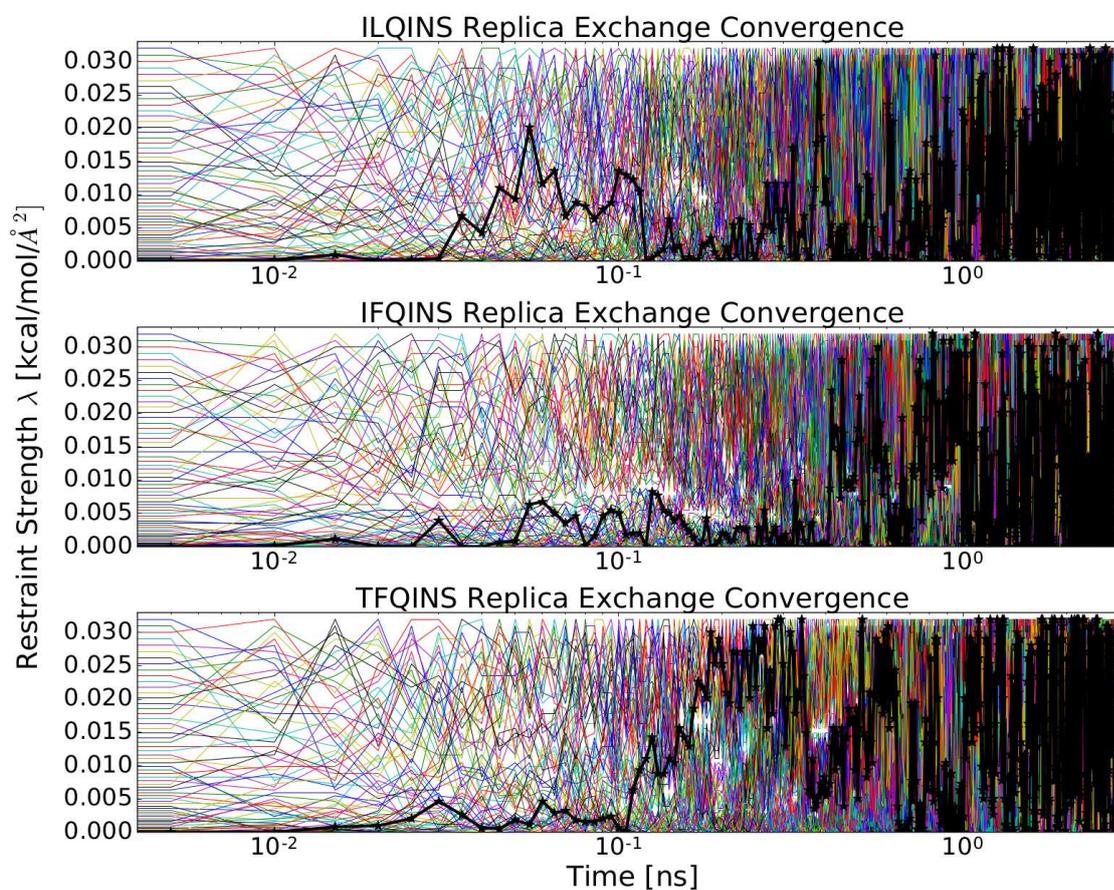


Figure 2.8: Replica exchange convergence for different sequences; $\lambda_i = 0.032(i/63)^2$ was verified empirically to provide good mixing between replicas (Replica index i runs over $[0..63]$). Replica exchanges were attempted at 5 ps intervals. The black trace in each system shows an arbitrary reference replica.

2.4.4 Synthesis and Light Scattering

Synthesis was carried out by Wen Li, Kun Liu and Afang Zhang of the University of Shanghai, China. Light scattering was carried out by Nicholas P Reynolds, Swinburne University of Technology, Australia.

ILQINS was made by solid phase peptide synthesis from Wangresin with O-(benzotriazole-1-yl)-1,1,3,3-tetramethylcarbamide tetrafluoroborate (TBTU) as the coupling reagent and N,N'-diisopropylethylenamine (DiPEA) as the base. 1-Hydroxybenzotriazole (HOBt) was used to avoid intramolecular cyclization to form dike-topiperazine. A typical synthetic procedure was: After swelling the resin in DMF overnight, Fmoc-protected amino acid (4 equiv), TBTU (4 equiv), DiPEA (4 equiv) and HOBt (4 equiv) in DMF were added and shaken. After coupling for 1 h, the resin was washed with DMF (4 × 1 min) and DCM (4 × 1 min). Removal of the Fmoc group was performed with piperidine (15 min). The peptide was cleaved from the resin with HF in the presence of 10% anisole for 1 h at 0°C. The crude peptide was precipitated with anhydrous tert-butylmethyl ether, dissolved in AcOH and lyophilized, which was further purified by RP-HPLC with gradients of water and acetonitrile.

In order to initiate self-assembly lyophilized ILQINS was mixed with MilliQ water at 1.5 mM then left to stand for 24 hours. Wide-Angle X-Ray Scattering (WAXS) was then performed at room temperature using the SAXS/WAXS beamline of the Australian synchrotron. Samples were loaded into a 96 well plate held on a robotically controlled x-y stage and transferred to the beamline via a quartz capillary connected to a syringe pump. A beam of wavelength $\lambda = 1.03320\text{\AA}$ (12.0 KeV), cross-section $300\ \mu\text{m} \times 200\ \mu\text{m}$ and a typical flux of 1.2×10^{13} photons per second was used. 2D diffraction images were recorded on a Pilatus 1M detector. q ranges were between $0.03\text{-}1.5\text{\AA}^{-1}$. Spectra were recorded under flow ($0.15\ \text{ml min}^{-1}$) in order to spread the beam damage. A set of 15 spectra was recorded (exposure time = 1 s) and the average spectrum is shown after background subtraction against MQ water in the same capillary.

Chapter 3

Kinetic Control of Parallel versus Antiparallel Amyloid Aggregation via Mesostructural Polymorphism

By combining atomistic and higher-level modelling with solution X-ray diffraction, self-assembly pathways are analysed for the IFQINS hexapeptide, a bio-relevant amyloid former derived from human lysozyme. It is verified that (at least) two metastable polymorphic structures exist for this system which are substantially different at the atomistic scale, and the conditions under which they are kinetically accessible are compared.

The higher-level polymorphism for these systems, which manifests as shape differences between structures instead of or as well as differences in the small-scale contact topology, is further examined at the nanometre to micrometre scales. Any future design of structure based inhibitors based on the IFQINS steric zipper should take account of this polymorphic assembly.

3.1 Introduction

The hydrogen-bonding, hydrophobic and electrostatic interactions which stabilise globular proteins can also drive the formation of tough multi-chain ‘amyloid’ aggregates which are often associated in biology with disease [21, 22]. Amyloid formation is implicated in various pathologies, particularly fatal neurodegenerative diseases such as Alzheimer’s, Parkinson’s and Huntington’s [26, 42–44]. Beyond the neurodegenerative diseases, certain inherited amyloidoses may be systemic or else localised in non-brain tissues: Lysozyme amyloidosis is an example of this class, in which a mutation in the IFQINS subsequence (to TFQINS) leads to accumulation of amyloid and eventual multiple organ failure [80, 81]. A further relevant sequence variation is ILQINS, the wild type subsequence in *Gallus gallus*,

which demonstrates reduced *in vitro* amyloid formation relative to the human wild type IFQINS.

In the study of amyloid aggregation it is common to use truncations or subsequences of longer bio-relevant proteins, in some cases because the protein is indeed truncated in the biological context but often also for simple convenience. The (I/T)(L/F)QINS peptide system has elements of both of these motivations: it was shown using mass spectrometry that in conditions of warm acid similar to the stomach, full-length lysozyme is hydrolysed into fragments, and fragments containing (I/T)(L/F)QINS such as Y₅₄GILQINSRWWCND₆₇ dominated the aggregation process [73]. As well as I56T, W64R and D67H are disease-associated mutations [138], nevertheless in order to develop full understanding and control of the aggregation process the focus of the present work is on the 6-residue fragment only. It has been shown that aggregation propensity increases following the sequence *ILQINS* → *IFQINS* → *TFQINS*, and *pH7* → *pH2* [75]. In the same paper, a computational prediction that amyloid formation should in some cases *decrease* with increasing concentration (fewer macro-scale crystals should form, and more monomeric peptide should remain in solution) was made, and validated experimentally. It should be noted that this previous paper used different candidate atomistic structures for the hexapeptide systems but that, as the mutation series only alters sidechains at the unit-cell surface, conclusions for self-assembly from that study remain qualitatively unchanged when repeated using the newer candidate atomistic structures.

Analysis of short peptide steric zippers has in the past led to successful design of inhibitors for aggregation of the full-length chain, including aggregation of the A β [139] and tau peptides [140]. Tau includes the VQIVYK and VQIINK homologue hexapeptides of IFQINS, and effectiveness of inhibitor design was improved by targeting the polymorphic steric zippers for VQIVYK and VQIINK [90], including structural information from soluble nanocrystal or fibril structures as well as from microcrystals amenable to solid-phase crystallography. Effective design of inhibitors for human Lysozyme aggregation should therefore also benefit from understanding of IFQINS polymorphic steric zippers. The kinetic process by which polymorphs compete or cross-seed can potentially be complex. If amyloid aggregates propagating as prions are a form of highly simplified quasi-lifeform, then this network of polymorph interactions is the quasi-ecology which determines dominance or extinction of a given fold.

Research into amyloid is not only driven by medical goals, but also seeks to develop peptide biomaterials [48–50]. One of the motivations to consider amyloid as a biomaterial is the potential for versatility in material properties driven by polymorphism at the atomistic or mesoscopic levels: it is common that a given sequence can stably take on a variety of

morphologies [141] including filaments [71], nanotubes [72], helical ribbons [73–75], twisted ribbons [74, 75] and crystals [75, 76] depending on the growth conditions.

Here, solution scattering taken during the aggregation process at high peptide concentration in water is examined showing an aggregated structure for IFQINS that is consistent with a solid-phase crystal structure previously published (by Sievers *et al.* [82], pdb code **4R0P** [47]), but which is different to the solution scattering previously observed. We show that the medium-concentration structures previously studied are composed of antiparallel (AP) β sheet, while the crystal and higher-concentration solution are composed of parallel (P) β sheet.

The experimental data contrasting these two polymorphic structures which differ in the symmetry of assembly permits the modelling of the hexapeptide aggregation process and examination of the physics of selection between polymorphs formed from P and AP β sheet to be extended beyond the early times that were considered for ILQINS, IFQINS and TFQINS in the previous chapter (2), up to millions of peptides and timescales in the range of microseconds to seconds.

3.2 Results

3.2.1 Atomistic Simulations Compared to WAXS

Atomistic models of the IFQINS **4R0P** P β crystal structure and a designed AP β structure were placed in a virtual aqueous environment and allowed to relax for 15 ns (simulation details in section 3.3), and calculated scattering was compared to WAXS spectra collected from real solutions with high (5 mM) and low (1.5 mM) concentrations of peptide. The WAXS for IFQINS at low concentration is very similar to the ILQINS WAXS discussed in the previous chapter (Figure 2.3) however the high-concentration WAXS is significantly different.

The lower-concentration experimental scattering agrees quite well with calculations based on the designed class 5 structure (Figure 3.1 (a,b)), while the X-ray data for a high-concentration (5 mM) solution of IFQINS after 24 h agrees with scattering calculated based on the **4R0P** deposited crystal structure (Figure 3.1 (c,d)). Despite the differing symmetry of **4R0P** to the AP structure, the overall scattering is not completely dissimilar, however the 180° rotation which accompanies translation along the *a*-axis in **4R0P** leads to fewer (but not much shifted) peaks in the angular window considered than were observed from IFQINS aggregated at lower concentrations. Peaks from the **4R0P** structure are much sharper than from the AP structure, both in experiment and simulation, indicating stronger ordering.

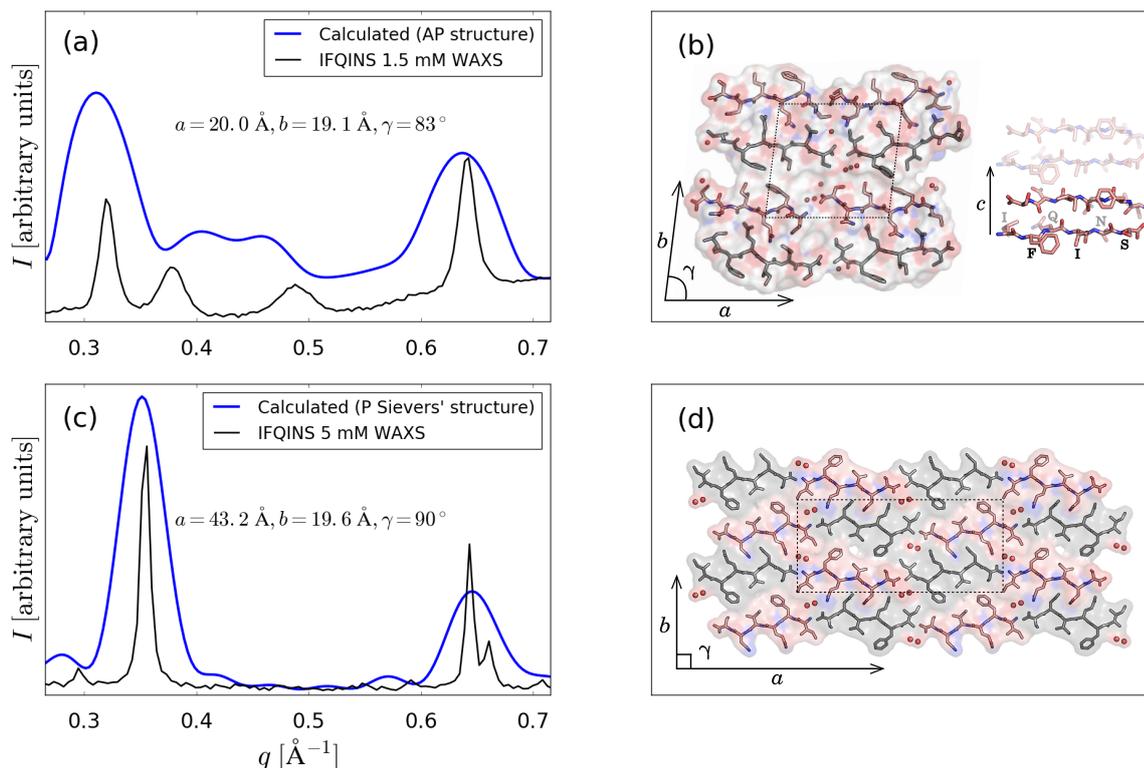


Figure 3.1: (a) Calculated scattering curve based on computationally derived AP structure, compared to solution scattering; lattice parameters are $a = 20.0 \text{ \AA}$, $b = 19.1 \text{ \AA}$ and $\gamma = 83^\circ$. (b) AP structure, with semi-ordered dynamic water molecules shown as orange spheres. (c) Calculated scattering based on solid-phase crystal structure after evolution in aqueous environment for 15 ns. (d) Solid-phase P crystal structure **4R0P** of IFQINS (four unit cells) reported by Sievers; lattice parameters are $a = 43.2 \text{ \AA}$, $b = 19.6 \text{ \AA}$ and $\gamma = 90^\circ$. A full translational unit cell is shown as a rectangle in the center. *IFQINS WAXS data supplied by Nicholas P. Reynolds, Swinburne University of Technology, Australia.*

3.2.2 Peptide-Level Assembly Thermodynamics

Having arrived at atomistic models for solution nanocrystals, standard free energies are evaluated to join together the different lattice planes of the nanocrystals under a linear approximation such that the total energy scales proportionally to the number of peptides buried by the interface. We find the free energy per peptide for interfaces perpendicular to the a (terminus-terminus axis), b (sidechain interaction axis) and c (hydrogen bonding axis) by calculating the difference between free energies of joined blocks of peptides and separated blocks, for example:

$$\Delta G_a^\circ = (G_{243} - 2G_{143}) / (4 \times 3),$$

where integer subscript triplets ijk are the number of peptides in each dimension of a rectangular peptide block or sub-block, and the denominator term is the number of peptides buried in the reference interface. The free energy to form a steric zipper, creating an interface which cuts through a unit cell of the crystal lattice, was also calculated. This interface, ΔG_{zip} , is in the plane perpendicular to the b axis, but is stronger than ΔG_b . Where a splitting event changes between a single block with even j and two blocks with odd j , ΔG_{zip} is relevant rather than ΔG_b .

Reference block free energies G_{ijk} are calculated as averages over 100 blocks sampled from the converged part of the MD simulation. After a block is ‘cut’ from the simulation system, its energy is minimised in a continuum solvent [130], so that the final free energy accounts for the electrostatics of solvent exposure, and also contains part of the appropriate physical entropy change from creating an interface, particularly that related to ordering of the solvent.

Because the **4R0P** structure has a herringbone symmetry (group $p2$ in the ab plane) rather than pure translational (group $p1$), the edges of an assembly are jagged with substantial overhang, and writing the free energy to join two blocks as a straightforward linear sum is less appropriate than for the class 5 structure (Figure 3.2). As well as these edge irregularities visible in projections onto the ab plane, adjacent sheets are also stepped by $\pm 0.5c$ in the vertical c axis, so in general the calculated interface energy based on the ab plane should be multiplied by $n_c - 1/2$ rather than by n_c as is the case for strictly rectangular blocks. Figure 3.2 (f,g) gives definitions for two components of the interface energy, labelled ϵ and ϵ' , that can be used to compose the binding free energies in the a direction as:

$$\Delta G/(n_c - 1/2) = (2n_b - 1)\epsilon/2 \quad (3.1)$$

and in the b direction as:

$$\Delta G/(n_c - 1/2) = (n_a - 1)\epsilon + n_a\epsilon' \quad (3.2)$$

where n_a , n_b and n_c are the numbers of peptides in the a , b and c directions respectively. That the P lattice free energy is initially gained more slowly than the AP lattice energy seems like an immediate qualitative explanation for the **4R0P** structure to have a higher nucleation barrier than the AP structure, and therefore for it to form later (or never) under conditions of lower concentration where nucleation or meta-nucleation is a more significant limit to the aggregation process, however other differences between the two systems exist and interact with this phenomenon.

Assembly Interface Definitions

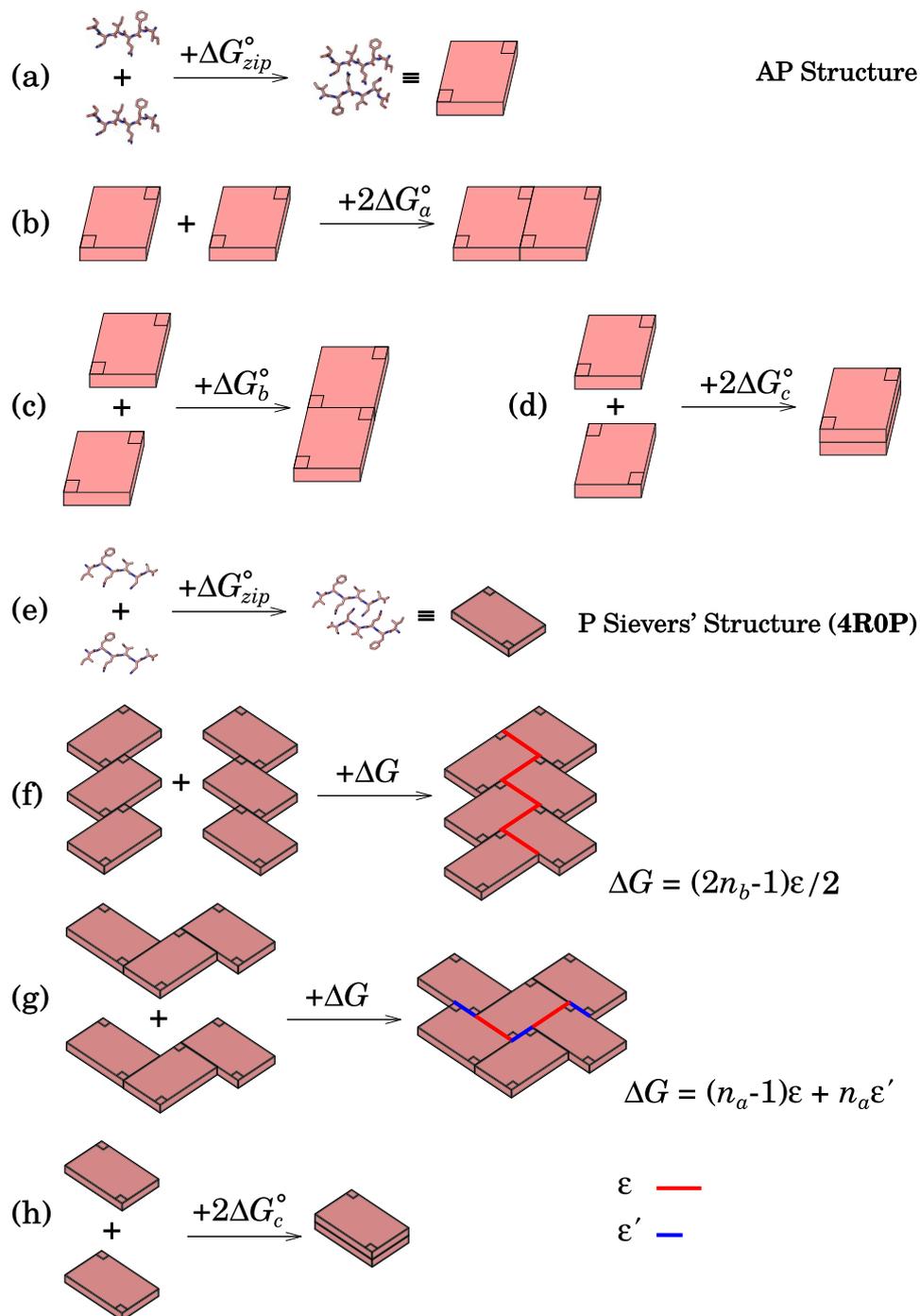


Figure 3.2: Cartoon showing binding free energies for interfaces in the a (backbone axis), b (sidechain axis) and c (hydrogen bond axis). (a-d): AP structure has generic in-register steric zipper 3D assembly. (e-h): Sievers' structure has herringbone assembly characterised by two energy terms ϵ and ϵ' . Binding free energies are (f): $\Delta G = (2n_b - 1)\epsilon/2$ in the a direction and (g): $\Delta G = (n_a - 1)\epsilon + n_a\epsilon'$ in the b direction, where n_a and n_b are the number of peptides in the a and b directions respectively.

The standard binding free energies to construct a buried interface associated with each cleavage plane ΔG_a° , ΔG_b° , ΔG_{zip}° and ΔG_c° are written in TABLE 3.1 in kcal/mol/peptides. The parameters $\varepsilon = -18.84$, and $\varepsilon' = -5.99$ were found in the same way as the others, by comparing blocks of peptides (see Section 3.3).

IFQINS Structure			ΔG_{zip}°	ΔG_c°
designed AP class 5	$\Delta G_a^\circ = -9.2(1)$	$\Delta G_b^\circ = -6.2(1)$	-16.4(3)	-29.3(2)
4R0P P structure	$\varepsilon/2 = -9.4(2)$	$\varepsilon' = -6.0(1)$	-25.3(2)	-25.3(2)

Table 3.1: Standard binding free energy gain to construct a buried interface between unit cells in the direction of backbone axis (ΔG_a°), sidechain axis (ΔG_b°), intra-cell zipper (ΔG_{zip}° , in the plane perpendicular to b), and hydrogen bond axis (ΔG_c°). Units are kcal/mol/peptide buried by the interface. **4R0P** structure has herringbone assembly characterised by two energy terms ε and ε' (see Figure 3.2 (f,g)). For the crystal structure, energies are initially not linear with n_a , n_b and n_c however the linear change per increase in dimension at the limit of large aggregates is shown.

3.2.3 Complex Kinetic Competition

Having identified selection *in vitro* between two dissimilar structures, the event-driven Gillespie algorithm (introduced in section 1.4.3) was used to make a kinetic simulation investigating the competition between the AP and P β -sheet structures over a range of concentrations. Figure 3.3 shows the evolving mass of aggregated peptides, broken down by elongation (Figure 3.3 (a)), then formation of 2D and 3D aggregates (Figure 3.3 (b,c)). A complex kinetic with two regimes is evident, at low and high concentrations.

At low concentrations (nM-mM) the AP structure elongates noticeably faster, as the weaker ΔG_c° for P β -sheet makes the formation of single-sheet aggregates reversible for these structures on timescales approximating that of collisions. The lead of AP in forming 1D aggregates translates into formation of 2D and 3D aggregates by hierarchical self-assembly and the AP system dominates assembly at lower concentrations. The small amount of P assembly which does take place in this regime shows a stochastic distribution of wait times, indicating that rare nucleation events are needed for highly stable 3D P aggregates to form. At high concentrations (mM and up), the gain of the AP system in forming 1D aggregates is overtaken by the P system in forming 2D aggregates: the stronger steric zipper and lateral assembly in the P geometry allows it to form 2D structures with long-term stability while the pool of free monomers is still not fully depleted. At these higher concentrations the proportion of 1D or higher AP aggregated peptide even takes a gradual downward trend, as peptides leave the 1D+ AP aggregates and are recruited into

2D and higher P aggregates. Within the simulation timescale, dominance of P over AP is never dramatic. The turnover at which the two are roughly equal is located at around 5 mM, the concentration at which a mixed population of fibrils was observed experimentally.

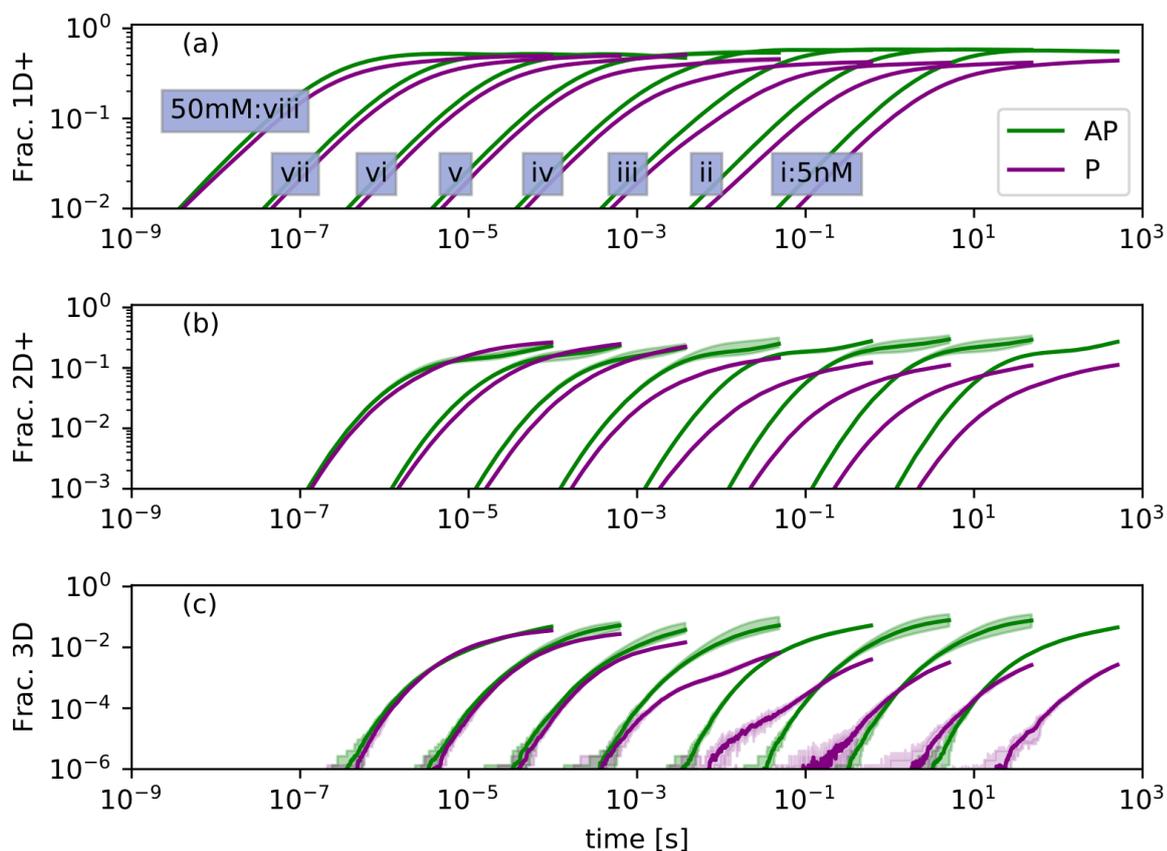


Figure 3.3: Evolution of oligomer formation colored by parallel or antiparallel. Curves i-viii show concentrations increasing from 5 nM in multiples of 10. Spreads show minimum and maximum values reached in 10 replicates. Traces are averages over the replicates. ‘Frac. 1D+’ (a) shows peptides incorporated into any aggregate which is identifiably P or AP, therefore initially describes mostly single-sheet aggregates, for which AP is the most stable geometry. (b) shows formation of 2D aggregates, where complex kinetics driven by the availability of 1D aggregates and the stability of 2D+ aggregates begin to take effect. (c) shows the quite sharp kinetically-driven phase transition between AP dominance and AP-P coexistence in the final sample.

The mesoscopic shape of the growing aggregates (twist, bend and aspect ratio) has an important but complex connection to the kinetics, in that bent and twisted aggregates have reduced possibilities for hierarchical self-assembly without paying an energetic penalty to un-twist or un-bend (introduced in Section 1.2.3). In the physical system twist and bend are coupled to cross-section area and aspect ratio, with smaller area implying less cost to

twist, and an aspect ratio further from one (large N_a/N_b or large N_b/N_a) implying less cost to bend [142], although the relationship may be complex. Elastic deformation was not treated directly in the models presented here, however it is apparent that the turnover in experimentally observed aggregation at the 1.5 mM to 5 mM range of concentrations corresponds to an inflection in the aspect ratio behaviour for both P and AP aggregates (Figure 3.4).

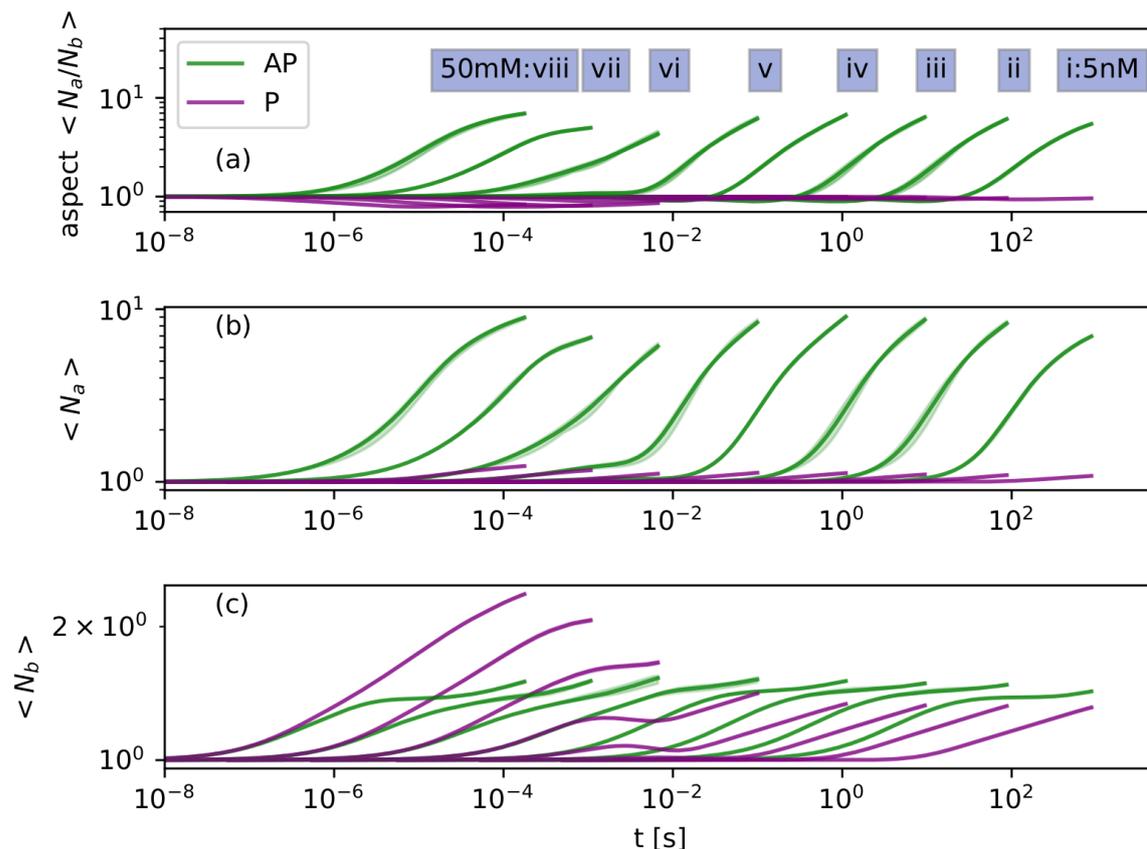


Figure 3.4: Evolution of cross-sectional aspect ratio colored by parallel or antiparallel. Monomers are counted as both P and AP for purposes of averaging. Spreads show minima and maxima reached in 10 replicates. Traces are averages over the replicates. Both P and AP show complex behaviour crossing from μM to mM concentration.

At μM concentrations and below, the AP aspect ratio N_a/N_b initially drops very slightly, driven by steric zipper formation in the plane perpendicular to the b axis ($N_b = 2$ implies a steric zipper). The non-zipper hydrophobic b interface is however less stable than the terminus-terminus a interface (see Table 3.1), so the more stable a interface then takes over and leads growth, driving formation of large quasi-2D sheet pairs which are likely to become curled ribbons or twisted fibrils in the physical system. Above mM concentrations, a

and b interfaces are both stable on the timescales of oligomer collision, and the aspect ratios do not run away to the same extent. Based on aspect ratios, the region of concentrations from $500\mu\text{M}$ (Figure 3.4 trace *vi*) to 5 mM (trace *vii*) emerges as optimal for formation of more rectangular, less twisted nanocrystals or thick fibrils in the AP geometry.

The P aspect ratio trace shows that at low concentrations the P system is reliant on formation of a 3D aggregate for stability, so has slow but roughly isotropic growth following the initial formation of the steric zipper such that $N_b = 2$. At high concentrations (where P aggregates are observed empirically), growth in b is enough to stabilise the P system without much growth in a , so this axis leads the lateral growth but not so much as to cause dramatic anisotropy.

The crystallisation experiment leading to P β structures used a buffer solution not modelled in simulation, however it may not be a coincidence that it took place via hanging drop method from an initial concentration of 5 mM [82], the region where P formation is also strong in the simulation.

3.3 Methods

3.3.1 Molecular Simulation

To relate the crystallography to the nanocrystallites studied via solution X-ray, an atomistic model nanocrystal of 1296 peptides (roughly cubic, dimension $\approx 10\text{ nm}$ or $6 \times 6 \times 18$ two-peptide steric zipper) was built following the **4R0P** crystal structure, and immersed in a periodic box of TIP3P atomistic water [132]. The structure was thermalised and allowed to relax for 15 ns in a simulated aqueous environment using the AMBER molecular dynamics system [134] and ff14SB forcefield [133] without showing any major reordering. Scattering was calculated by an orientationally averaged Fourier transform using CRY SOL [135]. Figure 3.1 (c) shows that the calculated scattering curve is compatible with the solution WAXS spectrum. Four unit cells of the crystal structure are shown in Figure 3.1 (d). The translational unit cell parameters are $a = 43.2\text{ \AA}$, $b = 19.6\text{ \AA}$ and $\gamma = 90^\circ$, however if lattice transforms including rotations are permitted, the a cell length becomes 21.6 \AA .

3.3.2 Desolvation Energy Calculation

To understand the anisotropic growth kinetics of the competing P and AP structures, free energies per buried peptide to desolvate a lattice plane ΔG_a° , ΔG_b° , ΔG_c° were calculated, and also ΔG_{zip}° to bury a steric zipper. Within a linear approximation, the free energy cost

to break an interface for example in a should be writeable in the form:

$$\Delta G = n_b n_c \Delta G_a^\circ. \quad (3.3)$$

This linear approximation should become increasingly valid for larger aggregates, as edge effects, cooperativity and finite-size thermodynamics become progressively less significant. The measured values for $\Delta G_{a,b,c,zip}^\circ$ therefore depend to some extent on the size of the blocks which are broken or joined in order to evaluate them. In order to have more accurate free energies close to the decisive region of small initial aggregates, capturing cooperativity at approximately the right lengthscales 20-40Å, block size combinations for each interface were chosen as follows, where $G_{i,j,k}$ is the calculated reference free energy for a block of size $i \times j \times k$:

$$\Delta G_c^\circ = (G_{1,1,10} - 10G_{1,1,1})/9 \quad (3.4)$$

$$\Delta G_{zip-P}^\circ = (G_{1,2,10} - 2G_{1,1,10})/9.5 \quad (3.5)$$

$$\Delta G_{zip-AP}^\circ = (G_{1,2,10} - 2G_{1,1,10})/10 \quad (3.6)$$

$$\Delta G_b^\circ = (G_{2,4,3} - 2G_{2,2,3})/6 \quad (3.7)$$

$$\Delta G_a^\circ = (G_{2,4,3} - 2G_{1,4,3})/12. \quad (3.8)$$

Each $G_{i,j,k}$ is found as an average over 50 blocks cut from the large nanocrystal MD simulations of P or AP IFQINS structures already used to calculate scattering. Individual blocks are minimised in a Generalised Born solvent model [130] and the converged energies averaged. The P zipper energies (ΔG_{zip-P}°) were found by joining two sheets of ten peptides and then dividing by only 9.5 because of the $0.5c$ overhang in the c axis between adjacent sheets defined by the deposited crystal structure. As discussed under the assembly thermodynamics (Equations (3.1) and (3.2)), the buried area for interfaces is then scaled up again by $n_c - 0.5$ for the P structure rather than by n_c as for the AP structure, thus recovering the original measured ΔG for the measured interface sizes.

3.3.3 Kinetic Rate Equation Network

The model for self-assembly of the peptides was defined as a system of N rigid bodies each with six faces labelled $a+, a-, b, zip, c+, c-$. Collision rates were calculated for the peptides (and assemblies of them) based on the equations for diffusion coefficients D of rod-like particles due to Ortega and de la Torre [117]. Any coupling between orientation and direction of movement was ignored such that collisions were resolved based on the surface area of the assembly-competent planes, determined from the crystal lattice. In this

form an example collision rate constant for two rectangular crystallites of shape i, j, k and u, v, w matching $a+$ and $a-$ planes such that $v = j$ and $w = k$ is:

$$k_{a+a-}(u, v, w, i, j, k) = 2e^{-3} (D(u, v, w) + D(i, j, k)) \sqrt{2bc \cdot jk}. \quad (3.9)$$

The barrier term e^{-3} is assigned based on the loss of translational and rotational degrees of freedom on joining two peptide blocks, calorimetric studies have found that the true barrier for small peptides to assemble is indeed of the order $2-5k_B T$ at 300K [143]. For assembly of larger peptides and proteins an energetic cost to unfold should be added to the barrier term for monomer collisions, for example a barrier of $10.1k_B T$ (25 kJ/mol) is quoted for full-length human lysozyme.

Number density of given species in the simulation volume enters the rate equation directly, with no account made of any spatial correlations in the solution:

$$r_{a+a-} = k_{a+a-} N_{ijk} N_{uvw} / V. \quad (3.10)$$

In the case of homodimerisation, such that $ijk = uvw$, the symmetry factor $N_{ijk} N_{uvw}$ is replaced by $N_{ijk}(N_{ijk} - 1)/2$. In the case that two monomers collide, selection of P or AP geometry was made with a 50% chance for each. Once a monomer in an assembly was committed to P or AP, it could return to an uncommitted state only by leaving the assembly. This model does not allow for heterogenous nucleation of P fibrils from AP, or for formation of mixed P/AP fibrils: P and AP structures interact only indirectly, by competing for monomers. Heterogenous nucleation may be added in future iterations of the research.

To define Arrhenius-like rates for a given aggregate to split, it is necessary to set a dynamical timescale. For example the rate for aggregates of a given geometry u, v, w to split on some c plane is set as:

$$r_c(u, v, w) = k_c(u, v) N_{uvw} (w - 1) \quad (3.11)$$

$$k_c(u, v) = \frac{1}{\tau_0} \exp[uv\Delta G_c^\circ / k_B T] \quad (3.12)$$

where τ_0 is chosen as the time for a single peptide to diffuse its own length.

The above system allows a rate equation network for collision of rectangular objects which have at least one matching face to be constructed, however such a network quickly and unphysically leads to three populations of aggregates which are extended in each of the lattice axes, and which have zero rates to combine between populations. In order to control complexity of the calculation it was not feasible to track the full space of non-rectangular

aggregates, however these were treated ‘virtually’ by allowing complex collisions including a splitting process into the rate, such that objects with only one or zero matching dimensions could still collide, and the final state after the reaction would contain again only rectangular aggregates. Figure 1.8 illustrates the multistep reactions treated. Figure 1.8 (a) corresponds to the single-step collision of Equation (3.10), while for Figure 1.8 (b) we combine rate constants for joining (k_{a+a-}) and splitting (k_c):

$$r_{a+a-} = \tau_0 k_c(2,4) k_{a+a-}(3,4,4,2,4,6) \frac{N_{3,4,4} N_{2,4,6}}{V}. \quad (3.13)$$

The two-to-three body process of Figure 1.8 (c) allows a choice of two pairs of cleavage planes given a collision surface of two non-matching bodies. To manage the complexity of the calculation, rates for each choice of planes were calculated, and only the fastest one retained in the kinetic system. Any process with $k < 10^{-50}$ was also automatically discarded from the rate equation system.

The rate equations for single collisions and collision-plus-split are balanced by simple splitting for single collisions, and by the reverse two body process for the collision-plus-split, however no three-to-two process was constructed as a simple calculation of likely rates for this found extremely small values. The two-to-three process also had extremely low rates, but was retained in order to avoid pathological situations such that a $10 \times 12 \times 102$ aggregate could not ever assemble with a $9 \times 10 \times 101$ aggregate.

Given the set of rate equations described above (a ‘kinetic master equation’) it should be possible to make an analytical statement of the non-equilibrium kinetics and the final equilibrium state of the system [144]. Such analyses typically do not capture stochastic effects such as nucleation, which are often important for amyloid formation, so instead the decision was made to sample the rate equation set numerically using the event-driven Gillespie algorithm [116]. In this method, the rates for all possible forward or backward processes given the current state of the system are calculated, and a single process to carry out is then selected randomly with a weight proportional to the rate for that process. The system is then updated according to the reaction chosen, and the rates re-calculated with re-use of information from the previous iteration. Rate constants for given reactions are calculated only once, on the first occasion that given reactants are added to the system, and then cached so that future rate calculations for that reaction can be made cheaply. For each simulation system, 6 million peptides were used, and concentration was controlled by setting the volume V .

3.3.4 Experimental Methods

Synthesis was carried out by Wen Li, Kun Liu and Afang Zhang of the University of Shanghai, China. Light scattering was carried out by Nicholas P Reynolds, Swinburne University of Technology, Australia.

IFQINS was made by solid phase peptide synthesis from Wangresin with O-(benzotriazole-1-yl)-1,1,3,3-tetramethylcarbamide tetrafluoroborate (TBTU) as the coupling reagent and N,N'-diisopropylethylenamine (DiPEA) as the base. 1-Hydroxybenzotriazole (HOBt) was used to avoid intramolecular cyclization to form diketopiperazine. A typical synthetic procedure was: After swelling the resin in DMF overnight, Fmoc-protected amino acid (4 equiv), TBTU (4 equiv), DiPEA (4 equiv) and HOBt (4 equiv) in DMF were added and shaken. After coupling for 1 h, the resin was washed with DMF (4 × 1 min) and DCM (4 × 1 min). Removal of the Fmoc group was performed with piperidine (15 min). The peptide was cleaved from the resin with HF in the presence of 10% anisole for 1 h at 0°C. The crude peptide was precipitated with anhydrous tert-butylmethyl ether, dissolved in AcOH and lyophilized, which was further purified by RP-HPLC with gradients of water and acetonitrile. For IFQINS the molecular weight of 720.8 Da measured was found to be in good agreement with expected mass of 720.82 Da.

In order to initiate self-assembly lyophilized IFQINS was mixed with MilliQ water at either 1.5 mM or 5 mM then left to stand for 24 hours, at which point WAXS was used to characterise the spectra of the assemblies formed.

Experiments were performed at room temperature at the SAXS/WAXS beamline at the Australian synchrotron. Samples were loaded into a 96 well plate held on a robotically controlled x-y stage and transferred to the beamline via a quartz capillary connected to a syringe pump. The experiments used a beam of wavelength of $\lambda = 1.03320\text{\AA}$ (12.0 KeV) with dimensions $300\ \mu\text{m} \times 200\ \mu\text{m}$ and a typical flux of 1.2×10^{13} photons per second. 2D diffraction images were recorded on a Pilatus 1M detector. Experiments were performed at q ranges between $0.03\text{-}1.5\ \text{\AA}^{-1}$. Spectra were recorded under flow ($0.15\ \text{ml min}^{-1}$) in order to prevent X-ray damage from the beam. Multiples of 15 spectra were recorded for each time point (exposure time = 1 s) and the averaged spectra are shown after background subtraction against MQ water in the same capillary.

3.4 Discussion

Here a system with kinetic competition between parallel and antiparallel aggregation is examined, showing a somewhat counterintuitive pattern whereby the structure with a smaller

free energy of formation per unit volume (AP) is nonetheless favoured, particularly at low concentrations, due to having no single high barrier in its metanucleation pathway.

In general, P and AP sheets contrast in that AP systems have stronger axial interactions in the direction of the β -sheet, while P systems which are antiparallel across the steric zipper interface can compensate for this by having stronger lateral interactions. In this specific system the contrast between the P and AP structures is not limited to the β -sheet symmetry, the two also differ in the relative arrangement of unit cells with the P structure having a herringbone (or parquet) pattern which buries less surface per lattice plane in the early stages of lateral growth, even though this growth is ultimately more isotropic and stronger.

Quasi-2D aggregates, those with a cross-sectional aspect ratio far from 1, are known to readily form twisted or curled fibrils which are then geometrically hindered from hierarchical assembly, leading to slower kinetics, thereby slowing or limiting aggregation [75]. It is now necessary to add a counterexample where the anisotropic type of lateral growth which leads to fibrils rather than to microcrystals may overall slow the kinetics relative to 3D growth, but where it is still better for a given polymorph to be growing laterally than to stay longer at the stage of pure 1D aggregation. This case of early anisotropic lateral growth leading to eventual dominance is relevant in the context of competition for monomers against other polymorphs with a longer lag phase.

In this study it was attempted to understand kinetics by quantitatively following the route *structure* \rightarrow *energetics* \rightarrow *kinetics*, however it is feasible to build intuition such that the lattice parameters can directly suggest the conditions which will favour or disfavour a given aggregation scheme. The picture which now emerges is that strong lateral interactions are necessary in order to have a fast assembly kinetic, and that lateral interactions must be of roughly equal strength (suggested by roughly equal-sized lattice planes) in order to have isotropic aggregates which will ultimately dominate the aggregation process and progress to form a large amount of precipitate.

Amyloid kinetics are multifaceted. Before the formation of amyloid, oligomeric or disordered droplet assemblies may or may not form, depending on the sequence and solution, and these may compete with β -structured assemblies, or seed them, or mature into them [145]. Once β -sheet has been formed, even within a quasi-1D paradigm distinctions can be drawn between elongation following unconnected nucleation events, nucleation plus secondary nucleation, and self-seeding by fragmentation; and these distinctions have measurable consequences to the kinetic [146]. This quasi-1D approximation should allow meaningful investigation into the aggregation kinetics in particular at low concentrations, where all fibrils must nucleate to a finite thickness in order to be stable, but few

fibrils will grow to much greater thickness than that required. A 1D picture is also trivially valid in the case that the chain has steric, electrostatic, or other constraints which prevent assembly in higher dimensions.

With increasing concentration or increasing interaction strength, reduced free energy barriers allow lateral assembly either hierarchically (as modelled in the present work) or *via* secondary nucleation of new beta sheets at the surface of existing sheets [147]. The resulting intermediate-dimensionality extended structures, between pseudo-1D fibrils and pseudo-infinite 3D crystals, may be present in various competing polymorphic shapes. This study has presented and discussed an example in which two of these mesopolymorphic structures compete with each other, preventing or indefinitely delaying dominance of the more thermodynamically stable polymorph in a manner which is compatible with the Ostwald step rule.

Chapter 4

Discussion

Atomistic computer models give structural information at the most detailed level possible, subject to systematic error but with arbitrary spatial precision. Here, simulation methods were used to understand the formation of amyloid oligomers and mesostructures. Although this project was specifically focused on the ILQINS hexapeptide and its homologues IFQINS and TFQINS, the computational methods developed can be exploited to study and describe many similar biological systems.

The results of the project reveal that at early times in the aggregation process, antiparallel structure dominates among the small oligomers formed. In a specific system (IFQINS), simulation methods were used to confirm and elucidate kinetic competition *in vitro* between structures with parallel and antiparallel alignment of β -strands.

4.1 Antiparallel Structure Dominates Initially

Comparing the calculated scattering curves of 3D aggregates of ILQINS in different symmetry classes with the experimental data shows that the symmetry with best agreement to solution WAXS experiments has antiparallel arrangement of β -strands within the β -sheets, although indications of parallel β -sheet are also detected, suggesting a mixture of P and AP sheet at early stages of aggregation. The single-sheet aggregates resulting from HREMD simulations spontaneously formed antiparallel structures, which were verified to be stable on timescales longer than the equivalent parallel structures. These results and reasoning imply that antiparallel strand alignment is the default for amyloid oligomers at early stage self-assembly of peptides (Chapter 2).

4.2 Kinetic Competition Between P and AP Aggregation

Comparing the calculated scattering curves of two different 3D structures of IFQINS verifies that (at least) two metastable polymorphic structures exist for this system which are substantially different at the atomistic scale. One of the computational structures derived, with antiparallel arrangement of β -strands within the β -sheets, was consistent with scattering observed from multiple peptides at low and high concentrations whereas the crystallographic structure **4R0P**, with parallel arrangement of β -strands within the β -sheets, has only so far been verified to form starting from high initial peptide concentration. A kinetic simulation investigating the competition between the AP and P β -sheet structures over a range of concentrations was presented in Chapter 3.

4.3 Context in Aggregation Mechanisms and Kinetics

It has recently been remarked in reference to nucleation in a system of metal atoms that ‘*all nuclei that adopt an equilibrium shape are alike; every non-equilibrium-structured nucleus has its own shape*’ [148]. If the nucleation of metals is complex and various, amyloid formation is many times more so. Amyloid nucleation is certainly not a quasi-equilibrium process: rearrangement times for the internal degrees of freedom of nuclei are comparable to or longer than the timescales for addition or removal of monomers to the nuclei.

Amyloid is the endpoint of a complex kinetic network containing several reaction steps of nucleation and growth [149]. Kinetics and the mechanisms of aggregation are very important in determining the properties and functions of the final products and from an overview of the various biophysical studies, a broad overall mechanism for the formation of amyloids can be proposed [24, 150]. The variety of amyloid reactions can then be expressed as a choice of pathways through this complex network of states, with the probability for each choice determined by sequence, concentration and other experimental conditions (Figure 4.1).

As a default for *in-vitro* and *in-silico* systems, amyloid aggregation is activated initially by primary nucleation, which generates the first pre-critical nuclei from dissolved monomers. The nuclei can be parallel or antiparallel dimers or various other small soluble species collectively known as oligomers. As nuclei grow they typically become more ordered, either in a continuous way with collectivity strengthening intramolecular interactions, or through a sharply defined ‘ripening’ from liquid-like micelles or other droplets into solid-like sheets.

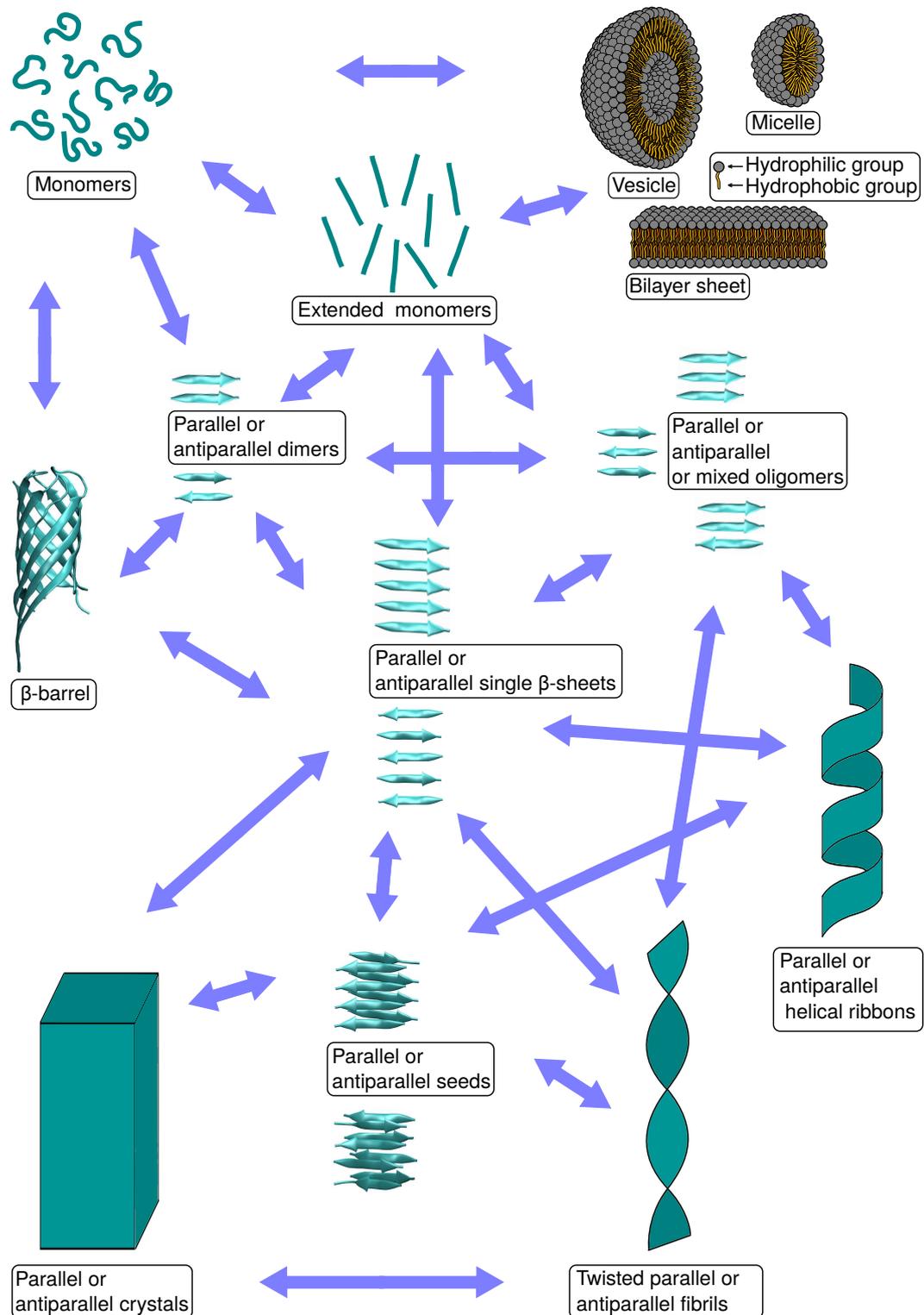


Figure 4.1: Schematic overview of peptide aggregation pathways. *Vesicle, Micelle and Bilayer images CC licenced by Mariana Ruiz Villarreal. β -barrel structure is 4RLC.*

Some oligomeric species represent kinetic dead-ends, but others can grow into larger species like ordered sheets with strong translational symmetry, from sheets into protofibrils, and eventually into mature aggregates. The ‘maturity’ of an aggregate is relative: twisted fibrils and helical ribbons may represent the end state or may eventually convert to crystals. It was observed that fibril-to-crystal conversion occurs within single aggregates via untwisting of twisted fibrils possessing saddle-like curvature and cross-sectional aspect ratios approaching unity [75]. Changing sequence, pH or concentration can shift the growth towards larger aspect ratio species assembling into stable helical ribbons possessing mean-curvature [75]. Although in most amyloid systems fibril growth is thermodynamically downhill after the initial barrier, elongation reactions can in principle be reversible.

Amyloid formation may include secondary nucleation processes which involve typically the fragmentation of fibrils induced by either thermal energy or mechanical forces [151, 152]. These breakage events multiply the number of fibrils and increase the concentration of reactive fibril ends that can recruit monomers and elongate. Secondary nucleation process may also involve the generation of new oligomers catalyzed by the presence of the surfaces of existing fibrils [41, 153].

Secondary nucleation of monomers on fibrils surface involves at least three molecular events: arrival of peptide at the fibrils surface, formation of product and release. Secondary nucleation is unsaturated at low monomer concentration, and the observed overall aggregation profiles are strongly dependent on monomer concentration. The process may saturate at high monomer (substrate) concentration and the observed overall aggregation profiles show little dependence on monomer concentration [41]. Such saturation is observed for $A\beta$ upon reduced electrostatic repulsion due to salt screening, pH variation, or mutation, implying that secondary nucleation is not a single-step reaction [154–157].

The overview of the peptide aggregation landscape in Figure 4.1 includes pathways for which novel physics which was introduced and quantified in this work: the route from monomers to parallel/antiparallel β -sheet was found to be more rapid for antiparallel β -sheet, while the routes from sheet to ribbons/fibrils/crystals were found to be faster from parallel β -sheet. It was found for the specific competing P/AP polymorphs of the IFQINS system that the AP polymorph had a (concentration-dependent) faster path to helical ribbons, with an inflection in the amount of expected ribbon at the mM- μ M concentration range.

That there is no route from helical ribbons to fibrils or crystals is an important assumption made in this work, which gives ribbons the status of kinetic traps, ultimately limiting the total aggregation in the system. Whether this limitation of the aggregation is harmful or beneficial in a medical context is beyond the present study to specify.

4.4 Outlook

Clinical trials have been underway for various aggregation inhibitors over the previous years (an example review is [158]), although as a rare genetic disease lysozyme amyloidosis is not much studied in a clinical context, better understanding of the amyloid structural biology and kinetics should inform further work on this topic. Close homologues of the sequences studied are important for various other more widespread (and probably also more complex) diseases; it is possible that the present work can inform research into these major public health problems.

The developed computational methods should be beneficially applicable to a wide range of peptide systems, but are still definitely susceptible to improvement. The choice of Hamiltonian modification for the replica exchange method was made *ad hoc* and other augmented Hamiltonians might be more effective. The kinetic simulation included only a subset of the entities in Figure 4.1, and considerably more sophistication could be added to this treatment without a major conceptual revision, such as some method of treating mixed parallel and antiparallel oligomers, or even the full range of symmetries. The kinetic assembly algorithm treats all oligomers as cuboidal and this is justified also by the lack of terracing in the experimentally observed mature structures (see Fig. 1 in [75]). The lack of treatment of terracing is related to a lack of treatment of secondary nucleation between heterogenous or similar fibrils and future work can potentially include such phenomena. For example, the current algorithm could be developed by considering the attachment of the building blocks (one peptide or two-peptide steric zipper) to the preformed crystal surface. In the next steps, blocks attaching to the surface will be more stable if they adsorb to the surface in sites adjacent to the preadsorbed blocks because in this case they will form more than one bound (one with surface and other(s) with the adjacent block(s)). So another layer will be growing on the crystal surface. Next layers can be formed and grow on the previously formed and growing layers as the same way.

A comprehensive multiscale modelling tool for intelligent control of peptide structure formation would be a wonderful thing, but would probably require major conceptual revision. The combination of molecular mechanics and the Gillespie algorithm has been surprisingly effective but the assumption of perfect mixing in the Gillespie method seems increasingly suspect as aggregates increase in size, and although incorporation of elastic degrees of freedom into the growing aggregates in the Gillespie algorithm can be conceived, in practice the calculations needed to parameterise such a treatment are expensive and complicated. With imagination, perhaps something which moves past these limitations can be assembled with only a few more years of work.

Bibliography

- [1] David L Nelson, Albert L Lehninger, and Michael M Cox. *Lehninger principles of biochemistry*. Macmillan, 2008.
- [2] Michael R Sawaya, Shilpa Sambashivan, Rebecca Nelson, Magdalena I Ivanova, Stuart A Sievers, Marcin I Apostol, Michael J Thompson, Melinda Balbirnie, Jed JW Wiltzius, Heather T McFarlane, et al. Atomic structures of amyloid cross- β spines reveal varied steric zippers. *Nature*, 447(7143):453, 2007.
- [3] Wolfgang Kabsch and Christian Sander. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, 22(12):2577–2637, 1983.
- [4] Lizabeth A Allison. *Fundamental molecular biology*. Blackwell Pub., 2007.
- [5] Engelbert Buxbaum. *Fundamentals of protein structure and function*, volume 31. Springer, 2007.
- [6] Peter H Raven, George B Johnson, Kenneth A Mason, Jonathan B Losos, and Susan R Singer. *Biology*. McGraw-Hill Education, 2017.
- [7] C.W. Pratt and K. Cornely. *Essential Biochemistry 3rd Ed*. John Wiley and Sons, Inc., 2014.
- [8] Gopalamudram Narayana Ramachandran. Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.*, 7:95–99, 1963.
- [9] Andrzej Bierzyński^{1/2}. Methods of peptide conformation studies. *Acta Biochimica Polonica*, 48(4):1091–1099, 2001.
- [10] James S Nowick. Chemical models of protein β -sheets. *Accounts of chemical research*, 32(4):287–296, 1999.

- [11] Motosuke Tsutsumi and Joji M Otaki. Parallel and antiparallel β -strands differ in amino acid composition and availability of short constituent sequences. *Journal of chemical information and modeling*, 51(6):1457–1464, 2011.
- [12] Joshua T Berryman, Sheena E Radford, and Sarah A Harris. Thermodynamic description of polymorphism in q-and n-rich peptide aggregates revealed by atomistic simulation. *Biophysical journal*, 97(1):1–11, 2009.
- [13] Linus Pauling and Robert B Corey. The pleated sheet, a new layer configuration of polypeptide chains. *Proceedings of the National Academy of Sciences of the United States of America*, 37(5):251, 1951.
- [14] Steven Hayward and E James Milner-White. The geometry of α -sheet: Implications for its possible function as amyloid precursor in proteins. *Proteins: Structure, Function, and Bioinformatics*, 71(1):415–425, 2008.
- [15] James E Milner-White, James D Watson, Guoying Qi, and Steven Hayward. Amyloid formation may involve α -to β sheet interconversion via peptide plane flipping. *Structure*, 14(9):1369–1376, 2006.
- [16] Georg E Schulz. The structure of bacterial outer membrane proteins. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, 1565(2):308–317, 2002.
- [17] William C Wimley. The versatile β -barrel membrane protein. *Current opinion in structural biology*, 13(4):404–411, 2003.
- [18] Yunxiang Sun, Xinwei Ge, and Feng Ding. Oligomerization and fibrillization dynamics of amyloid peptides and beta-barrel oligomer intermediates. *Biophysical Journal*, 114(3):227a–228a, 2018.
- [19] Nicklas Österlund, Rani Moons, Leopold L Ilag, Frank Sobott, and Astrid Gräslund. Native ion mobility-mass spectrometry reveals the formation of β -barrel shaped amyloid- β hexamers in a membrane-mimicking environment. *Journal of the American Chemical Society*, 2019.
- [20] Diletta Ami, Antonino Natalello, Marina Lotti, and Silvia Maria Doglia. Why and how protein aggregation has to be studied in vivo. *Microbial cell factories*, 12(1):1, 2013.

- [21] Massimo Stefani. Protein misfolding and aggregation: new examples in medicine and biology of the dark side of the protein world. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, 1739(1):5–25, 2004.
- [22] Jurriaan A. Luiken. *Self-assembly of functionalized colloids and short amyloidogenic peptides: Modelling, theory and simulations*. Van 't Hoff Institute for Molecular Sciences (HIMS), 2015.
- [23] LG Rizzi and S Auer. Amyloid fibril solubility. *The Journal of Physical Chemistry B*, 119(46):14631–14636, 2015.
- [24] Alan S Cohen. Amyloid proteins: The beta sheet conformation and disease: Volumes 1 and 2. *Amyloid*, 13(4):263, 2006.
- [25] Altaira D Dearborn, Joseph S Wall, Naiqian Cheng, J Bernard Heymann, Andrey V Kajava, Jobin Varkey, Ralf Langen, and Alasdair C Steven. α -synuclein amyloid fibrils with two entwined, asymmetrically associated protofibrils. *Journal of Biological Chemistry*, 291(5):2310–2318, 2016.
- [26] M Di Carlo, D Giacomazza, and PL San Biagio. Alzheimer's disease: biological aspects, therapeutic perspectives and diagnostic tools. *Journal of Physics: Condensed Matter*, 24(24):244102, 2012.
- [27] Martino Calamai, Fabrizio Chiti, and Christopher M Dobson. Amyloid fibril formation can proceed from different conformations of a partially unfolded protein. *Biophysical journal*, 89(6):4201–4210, 2005.
- [28] Karina MM Carneiro, Halei Zhai, Li Zhu, Jeremy A Horst, Melody Sitlin, Mychi Nguyen, Martin Wagner, Cheryl Simpliciano, Melissa Milder, Chun-Long Chen, et al. Amyloid-like ribbons of amelogenins in enamel mineralization. *Scientific reports*, 6:23105, 2016.
- [29] Martijn FBG Gebbink, Dennis Claessen, Barend Bouma, Lubbert Dijkhuizen, and Han AB Wösten. Amyloids—a functional coat for microorganisms. *Nature reviews microbiology*, 3(4):333–341, 2005.
- [30] JE Gillam and CE MacPhee. Modelling amyloid fibril formation kinetics: mechanisms of nucleation and growth. *Journal of Physics: Condensed Matter*, 25(37):373101, 2013.

- [31] Jason T Giurleo, Xianglan He, and David S Talaga. β -lactoglobulin assembles into amyloid through sequential aggregated intermediates. *Journal of molecular biology*, 381(5):1332–1348, 2008.
- [32] Jessica Nasica-Labouze and Normand Mousseau. Kinetics of amyloid aggregation: a study of the gnnqqny prion sequence. *PLoS Comput Biol*, 8(11):e1002782, 2012.
- [33] Phuong H Nguyen, Mai Suan Li, Gerhard Stock, John E Straub, and D Thirumalai. Monomer adds to preformed structured oligomers of $\alpha\beta$ -peptides by a two-stage dock–lock mechanism. *Proceedings of the National Academy of Sciences*, 104(1):111–116, 2007.
- [34] Marieke Schor, Jocelyne Vreede, and Peter G Bolhuis. Elucidating the locking mechanism of peptides onto growing amyloid fibrils through transition path sampling. *Biophysical journal*, 103(6):1296–1304, 2012.
- [35] Jae Sun Jeong, Annalisa Ansaloni, Raffaele Mezzenga, Hilal A Lashuel, and Giovanni Dietler. Novel mechanistic insight into the molecular basis of amyloid polymorphism and secondary nucleation during amyloid formation. *Journal of molecular biology*, 425(10):1765–1781, 2013.
- [36] Evan T Powers and David L Powers. The kinetics of nucleated polymerizations at high concentrations: amyloid fibril formation near and above the “supercritical concentration”. *Biophysical journal*, 91(1):122–132, 2006.
- [37] Jie Yang, Alexander J Dear, Thomas CT Michaels, Christopher M Dobson, Thomas PJ Knowles, Si Wu, and Sarah Perrett. Direct observation of oligomerization by single molecule fluorescence reveals a multistep aggregation mechanism for the yeast prion protein ure2. *Journal of the American Chemical Society*, 140(7):2493–2503, 2018.
- [38] Anthony WP Fitzpatrick, Giovanni M Vanacore, and Ahmed H Zewail. Nanomechanics and intermolecular forces of amyloid revealed by four-dimensional electron microscopy. *Proceedings of the National Academy of Sciences*, 112(11):3380–3385, 2015.
- [39] Jurriaan A Luiken and Peter G Bolhuis. Primary nucleation kinetics of short fibril-forming amyloidogenic peptides. *The Journal of Physical Chemistry B*, 119(39):12568–12579, 2015.

- [40] SG Agrawal and AHJ Paterson. Secondary nucleation: mechanisms and models. *Chemical Engineering Communications*, 202(5):698–706, 2015.
- [41] Mattias Törnquist, Thomas CT Michaels, Kalyani Sanagavarapu, Xiaoting Yang, Georg Meisl, Samuel IA Cohen, Tuomas PJ Knowles, and Sara Linse. Secondary nucleation in amyloid formation. *Chemical Communications*, 54(63):8667–8684, 2018.
- [42] Christian Haass and Dennis J Selkoe. Soluble protein oligomers in neurodegeneration: lessons from the alzheimer’s amyloid β -peptide. *Nature reviews Molecular cell biology*, 8(2):101–112, 2007.
- [43] Dominic M Walsh and Dennis J Selkoe. $A\beta$ oligomers—a decade of discovery. *Journal of neurochemistry*, 101(5):1172–1184, 2007.
- [44] Adelin Gustot, José Ignacio Gallea, Rabia Sarroukh, María Soledad Celej, Jean-Marie Ruyschaert, and Vincent Raussens. Amyloid fibrils are the molecular trigger of inflammation in parkinson’s disease. *Biochemical Journal*, 471(3):323–333, 2015.
- [45] William L Klein, Grant A Krafft, and Caleb E Finch. Targeting small $a\beta$ oligomers: the solution to an alzheimer’s disease conundrum? *Trends in neurosciences*, 24(4):219–224, 2001.
- [46] Hugh A Pearson and Chris Peers. Physiological roles for amyloid β peptides. *The Journal of physiology*, 575(1):5–10, 2006.
- [47] Dan Li, Eric M Jones, Michael R Sawaya, Hiroyasu Furukawa, Fang Luo, Magdalena Ivanova, Stuart A Sievers, Wenyuan Wang, Omar M Yaghi, Cong Liu, et al. Structure-based design of functional amyloid materials. *Journal of the American Chemical Society*, 136(52):18044–18051, 2014.
- [48] Raffaele Mezzenga, Peter Schurtenberger, Adam Burbidge, and Martin Michel. Understanding foods as soft materials. *Nature materials*, 4(10):729–740, 2005.
- [49] Samir K Maji, Marilyn H Perrin, Michael R Sawaya, Sebastian Jessberger, Krishna Vadodaria, Robert A Rissman, Praful S Singru, K Peter R Nilsson, Rozalyn Simon, David Schubert, et al. Functional amyloids as natural storage of peptide hormones in pituitary secretory granules. *Science*, 325(5938):328–332, 2009.

- [50] Shruti Mankar, A Anoop, Shamik Sen, and Samir Kumar Maji. Nanomaterials: amyloids reflect their brighter side. *Nano Reviews & Experiments*, 2(1):6032, 2011.
- [51] Annalisa Pastore and Pierandrea Temussi. Protein aggregation and misfolding: good or evil? *Journal of Physics: Condensed Matter*, 24(24):244101, 2012.
- [52] Roland G Roberts. Good amyloid, bad amyloid—what’s the difference? *PLoS biology*, 14(1):e1002362, 2016.
- [53] Ramnath Anjana, Marthandan Kirti Vaishnavi, Durairaj Sherlin, Surapaneni Pavan Kumar, Kora Naveen, Pasam Sandeep Kanth, and Kanagaraj Sekar. Aromatic-aromatic interactions in structures of proteins and protein-dna complexes: a study based on orientation and distance. *Bioinformatics*, 8(24):1220, 2012.
- [54] Karen E Marshall, Kyle L Morris, Deborah Charlton, Nicola O’Reilly, Laurence Lewis, Helen Walden, and Louise C Serpell. Hydrophobic, aromatic, and electrostatic interactions play a central role in amyloid fibril formation and stability. *Biochemistry*, 50(12):2061–2071, 2011.
- [55] Angel Mozo-Villarías and Enrique Querol. A protein self-assembly model guided by electrostatic and hydrophobic dipole moments. *PloS one*, 14(4):e0216253, 2019.
- [56] John C Stendahl, Mukti S Rao, Mustafa O Guler, and Samuel I Stupp. Intermolecular forces in the self-assembly of peptide amphiphile nanofibers. *Advanced Functional Materials*, 16(4):499–508, 2006.
- [57] Jonathan W Steed, David R Turner, and Karl Wallace. *Core concepts in supramolecular chemistry and nanochemistry*. John Wiley & Sons, 2007.
- [58] Christoph A Schalley. *Analytical methods in supramolecular chemistry*, volume 1. John Wiley & Sons, 2012.
- [59] Thomas Andrew Waigh. *Applied biophysics: a molecular approach for physical scientists*. John Wiley & Sons, 2007.
- [60] Charl FJ Faul and Markus Antonietti. Ionic self-assembly: Facile synthesis of supramolecular materials. *Advanced Materials*, 15(9):673–683, 2003.
- [61] Thomas H Rehm and Carsten Schmuck. Ion-pair induced self-assembly in aqueous solvents. *Chemical Society Reviews*, 39(10):3597–3611, 2010.

- [62] John Kuriyan, Boyana Konforti, and David Wemmer. *The molecules of life: Physical and chemical principles*. Garland Science, 2012.
- [63] Sergey E Paramonov, Ho-Wook Jun, and Jeffrey D Hartgerink. Self-assembly of peptide- amphiphile nanofibers: the roles of hydrogen bonding and amphiphilic packing. *Journal of the American Chemical Society*, 128(22):7291–7298, 2006.
- [64] Juan Wang, Kai Liu, Ruirui Xing, and Xuehai Yan. Peptide self-assembly: thermodynamics and kinetics. *Chemical Society Reviews*, 45(20):5589–5604, 2016.
- [65] Michael R Caplan, Peter N Moore, Shuguang Zhang, Roger D Kamm, and Douglas A Lauffenburger. Self-assembly of a β -sheet protein governed by relief of electrostatic repulsion relative to van der waals attraction. *Biomacromolecules*, 1(4):627–631, 2000.
- [66] Kazukuni Tahara, Shengbin Lei, Jinne Adisojoso, Steven De Feyter, and Yoshito Tobe. Supramolecular surface-confined architectures created by self-assembly of triangular phenylene–ethynylene macrocycles via van der waals interaction. *Chemical Communications*, 46(45):8507–8525, 2010.
- [67] Robert Tycko. Amyloid polymorphism: structural basis and neurobiological relevance. *Neuron*, 86(3):632–645, 2015.
- [68] Johnny D Pham, Borries Demeler, and James S Nowick. Polymorphism of oligomers of a peptide from β -amyloid. *Journal of the American Chemical Society*, 136(14):5432–5442, 2014.
- [69] Ankur P Chaudhary, Neha H Vispute, Vaibhav Kumar Shukla, and Basir Ahmad. A comparative study of fibrillation kinetics of two homologous proteins under identical solution condition. *Biochimie*, 132:75–84, 2017.
- [70] Marcus Fändrich, Jessica Meinhardt, and Nikolaus Grigorieff. Structural polymorphism of alzheimer $a\beta$ and other amyloid fibrils. *Prion*, 3(2):89–93, 2009.
- [71] Michael Schleege, Tanja Deckert-Gaudig, Volker Deckert, Krassimir P Velikov, Gijse Koenderink, Mischa Bonn, et al. Amyloids: from molecular structure to mechanical properties. *Polymer*, 54(10):2473–2488, 2013.
- [72] Céline Valéry, Franck Artzner, and Maité Paternostre. Peptide nanotubes: molecular organisations, self-assembly mechanisms and applications. *Soft Matter*, 7(20):9583–9594, 2011.

- [73] Cecile Lara, Nicholas Reynolds, Joshua Berryman, Afang Zhang, Anqiu Xu, and Raffaele Mezzenga. Ilqins hexapeptide, identified in lysozyme left-handed helical ribbons and nanotubes, forms right-handed helical ribbons and crystals. *Journal of the American Chemical Society*, 136(12):4732–4739, 2014.
- [74] Shuai Zhang, Maria Andreassen, Jakob T Nielsen, Lei Liu, Erik H Nielsen, Jie Song, Gang Ji, Fei Sun, Troels Skrydstrup, Flemming Besenbacher, et al. Coexistence of ribbon and helical fibrils originating from hiapp20–29 revealed by quantitative nanomechanical atomic force microscopy. *Proceedings of the National Academy of Sciences*, 110(8):2798–2803, 2013.
- [75] Nicholas P Reynolds, Jozef Adamcik, Joshua T Berryman, Stephan Handschin, Ali Asghar Hakami Zanjani, Wen Li, Kun Liu, Afang Zhang, and Raffaele Mezzenga. Competition between crystal and fibril formation in molecular mutations of amyloidogenic peptides. *Nature Communications*, 8(1):1338, 2017.
- [76] Karen E Marshall, Matthew R Hicks, Thomas L Williams, Søren Vrønning Hoffmann, Alison Rodger, Timothy R Dafforn, and Louise C Serpell. Characterizing the assembly of the sup35 yeast prion fragment, gnnqqny: structural changes accompany a fiber-to-crystal switch. *Biophysical journal*, 98(2):330–338, 2010.
- [77] Jed JW Wiltzius, Meytal Landau, Rebecca Nelson, Michael R Sawaya, Marcin I Apostol, Lukasz Goldschmidt, Angela B Soriaga, Duilio Cascio, Kanagalaghatta Rajashankar, and David Eisenberg. Molecular mechanisms for protein-encoded inheritance. *Nature structural & molecular biology*, 16(9):973, 2009.
- [78] Josh Berryman, S Radford, and S Harris. Prediction of twist of amyloid fibrils using molecular dynamics. *Proceedings of the NIC Workshop 2008*, 40:169–171, 2008.
- [79] Takeshi Chiba, Yoshihisa Hagihara, Takashi Higurashi, Kazuhiro Hasegawa, Hironobu Naiki, and Yuji Goto. Amyloid fibril formation in the context of full-length protein effects of proline mutations on the amyloid fibril formation of β 2-microglobulin. *Journal of Biological Chemistry*, 278(47):47016–47024, 2003.
- [80] MB Pepys, PN Hawkins, DR Booth, DM Vigushin, GA Tennent, AK Soutar, N Totty, O Nguyen, CCF Blake, CJ Terry, et al. Human lysozyme gene mutations cause hereditary systemic amyloidosis. *Nature*, 362(6420):553, 1993.
- [81] Christopher Pleyer, Jan Flesche, and Fahad Saeed. Lysozyme amyloidosis—a case report and review of the literature. *Clinical nephrology. Case studies*, 3:42, 2015.

- [82] Stuart Aaron Sievers. *Structural characterization of amyloid-like protein segments and the rational design of peptide inhibitors of fibrillation*. University of California, Los Angeles, 2008.
- [83] Vitaly V Kushnirov, Natalia V Kochneva-Pervukhova, Maria B Chechenova, Natalia S Frolova, and Michael D Ter-Avanesyan. Prion properties of the sup35 protein of yeast *pichia methanolica*. *The EMBO journal*, 19(3):324–331, 2000.
- [84] Rebecca Nelson, Michael R Sawaya, Melinda Balbirnie, Anders Ø Madsen, Christian Riek, Robert Grothe, and David Eisenberg. Structure of the cross- β spine of amyloid-like fibrils. *Nature*, 435(7043):773, 2005.
- [85] Patrick CA van der Wel, Józef R Lewandowski, and Robert G Griffin. Solid-state nmr study of amyloid nanocrystals and fibrils formed by the peptide gnnqny from yeast prion protein sup35p. *Journal of the American Chemical Society*, 129(16):5117–5130, 2007.
- [86] Massimiliano Meli, Giulia Morra, and Giorgio Colombo. Investigating the mechanism of peptide aggregation: insights from mixed monte carlo-molecular dynamics simulations. *Biophysical journal*, 94(11):4414–4426, 2008.
- [87] Luigi Vitagliano, Luciana Esposito, Carlo Pedone, and Alfonso De Simone. Stability of single sheet gnnqny aggregates analyzed by replica exchange molecular dynamics: antiparallel versus parallel association. *Biochemical and biophysical research communications*, 377(4):1036–1041, 2008.
- [88] Birgit Strodel, Chris S Whittleston, and David J Wales. Thermodynamics and kinetics of aggregation for the gnnqny peptide. *Journal of the American Chemical Society*, 129(51):16005–16014, 2007.
- [89] J Paul Taylor, John Hardy, and Kenneth H Fischbeck. Toxic proteins in neurodegenerative disease. *Science*, 296(5575):1991–1995, 2002.
- [90] PM Seidler, DR Boyer, JA Rodriguez, MR Sawaya, D Cascio, K Murray, T Gonen, and DS Eisenberg. Structure-based inhibitors of tau aggregation. *Nature chemistry*, 10(2):170, 2018.
- [91] Yu-Sheng Fang, Kuen-Jer Tsai, Yu-Jen Chang, Patricia Kao, Rima Woods, Pan-Hsien Kuo, Cheng-Chun Wu, Jhih-Ying Liao, Shih-Chieh Chou, Vinson Lin, et al. Full-length tdp-43 forms toxic amyloid oligomers that are present in frontotemporal lobar dementia-tdp patients. *Nature communications*, 5:4824, 2014.

- [92] Michael P Allen. Introduction to molecular dynamics simulation. *Computational soft matter: from synthetic polymers to proteins*, 23:1–28, 2004.
- [93] D. Frenkel and B. Smit. *Understanding Molecular Simulation: From Algorithms to Applications*. Computational science series. Elsevier Science, 2001.
- [94] M Cecchini, R Curcio, M Pappalardo, R Melki, and A Caffisch. A molecular dynamics approach to the structural characterization of amyloid aggregation. *Journal of molecular biology*, 357(4):1306–1321, 2006.
- [95] Alka Srivastava and Petety V Balaji. Molecular events during the early stages of aggregation of gnnqny: An all atom md simulation study of randomly dispersed peptides. *Journal of structural biology*, 192(3):376–391, 2015.
- [96] Sachin Patodia, Ashima Bagaria, and Deepak Chopra. Molecular dynamics simulation of proteins: A brief overview. *Journal of Physical Chemistry & Biophysics*, 4(6):1, 2014.
- [97] Carmen Domene. *Computational Biophysics of Membrane Proteins*. Royal Society of Chemistry, 2016.
- [98] D.A. Case, I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, T.E. Cheatham III, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, D. Ghoreishi, G. Giambasu, T. Giese, M.K. Gilson, H. Gohlke, A.W. Goetz, D. Greene, R Harris, N. Homeyer, Y. Huang, S. Izadi, A. Kovalenko, R. Krasny, T. Kurtzman, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, V. Man, D.J. Mermelstein, K.M. Merz, Y. Miao, G. Monard, C. Nguyen, H. Nguyen, A. Onufriev, F. Pan, R. Qi, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C.L. Simmerling, J. Smith, J. Swails, R.C. Walker, J. Wang, H. Wei, L. Wilson, R.M. Wolf, X. Wu, L. Xiao, Y. Xiong, D.M. York, and P.A. Kollman. *Amber 2019 (comprised of AmberTools19 and Amber18)*. University of California, 2019.
- [99] Alexey Onufriev. Continuum electrostatics solvent modeling with the generalized born model. *Modeling Solvent Environments*, 1, 2010.
- [100] Tom Darden, Darrin York, and Lee Pedersen. Particle mesh ewald: An nlog(n) method for ewald sums in large systems. *The Journal of chemical physics*, 98(12):10089–10092, 1993.

- [101] Laurent David, Ray Luo, and Michael K Gilson. Comparison of generalized born and poisson models: energetics and dynamics of hiv protease. *Journal of Computational Chemistry*, 21(4):295–309, 2000.
- [102] Michael Feig. Kinetics from implicit solvent simulations of biomolecules as a function of viscosity. *Journal of chemical theory and computation*, 3(5):1734–1748, 2007.
- [103] W Clark Still, Anna Tempczyk, Ronald C Hawley, and Thomas Hendrickson. Semi-analytical treatment of solvation for molecular mechanics and dynamics. *Journal of the American Chemical Society*, 112(16):6127–6129, 1990.
- [104] Alexey Onufriev, David A Case, and Donald Bashford. Effective born radii in the generalized born approximation: the importance of being perfect. *Journal of computational chemistry*, 23(14):1297–1304, 2002.
- [105] Jayashree Srinivasan, Megan W Trevathan, Paul Beroza, and David A Case. Application of a pairwise generalized born model to proteins and nucleic acids: inclusion of salt effects. *Theoretical Chemistry Accounts*, 101(6):426–434, 1999.
- [106] David J Earl and Michael W Deem. Parallel tempering: Theory, applications, and new perspectives. *Physical Chemistry Chemical Physics*, 7(23):3910–3916, 2005.
- [107] Ulrich HE Hansmann. Parallel tempering algorithm for conformational studies of biological molecules. *Chemical Physics Letters*, 281(1):140–150, 1997.
- [108] Yuji Sugita and Yuko Okamoto. Replica-exchange molecular dynamics method for protein folding. *Chemical physics letters*, 314(1):141–151, 1999.
- [109] Benjamin Perry Cooke Isard. *Theory and practice in replica-exchange molecular dynamics simulation*. ProQuest, 2008.
- [110] Jianfeng Lu and Eric Vanden-Eijnden. Infinite swapping replica exchange molecular dynamics leads to a simple simulation patch using mixture potentials. *The Journal of chemical physics*, 138(8):084105, 2013.
- [111] Christopher J. Woods, Muan Hong Ng, Steven Johnston, Stuart E. Murdock, Bing Wu, Kaihsu Tai, Hans Fangohr, Paul Jeffreys, Simon Cox, Jeremy G. Frey, Mark S. P. Sansom, and Jonathan W. Essex. Grid computing and biomolecular simulation. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 363(1833):2017–2035, 2005.

- [112] Srinivasaraghavan Kannan and Martin Zacharias. Folding simulations of trp-cage mini protein in explicit solvent using biasing potential replica-exchange molecular dynamics simulations. *Proteins: Structure, Function, and Bioinformatics*, 76(2):448–460, 2009.
- [113] Katja Ostermeir and Martin Zacharias. Hamiltonian replica-exchange simulations with adaptive biasing of peptide backbone and side chain dihedral angles. *Journal of computational chemistry*, 35(2):150–158, 2014.
- [114] Mark James Abraham, Teemu Murtola, Roland Schulz, Szilárd Páll, Jeremy C Smith, Berk Hess, and Erik Lindahl. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1:19–25, 2015.
- [115] Arthur F Voter. Introduction to the kinetic monte carlo method. In *Radiation effects in solids*, pages 1–23. Springer, 2007.
- [116] Daniel T Gillespie. Stochastic simulation of chemical kinetics. *Annu. Rev. Phys. Chem.*, 58:35–55, 2007.
- [117] A Ortega and J Garcia de la Torre. Hydrodynamic properties of rodlike and disklike particles in dilute solution. *The Journal of chemical physics*, 119(18):9914–9919, 2003.
- [118] Schrödinger, LLC. The PyMOL molecular graphics system, version 1.8. November 2015.
- [119] William Humphrey, Andrew Dalke, and Klaus Schulten. Vmd: visual molecular dynamics. *Journal of molecular graphics*, 14(1):33–38, 1996.
- [120] Katharine Gammon. Neurodegenerative disease: brain windfall. *Nature*, 515(7526):299–300, 2014.
- [121] Dylan Shea, Cheng-Chieh Hsu, Timothy M Bi, Natasha Paranjapye, Matthew Carter Childers, Joshua Cochran, Colson P Tomberlin, Libo Wang, Daniel Paris, Jeffrey Zonderman, Gabriele Varani, Christopher D Link, Mike Mullan, and Valerie Daggett. α -Sheet secondary structure in amyloid β -peptide drives aggregation and toxicity in Alzheimer’s disease. *Proceedings of the National Academy of Sciences of the United States of America*, 116:8895–8900, April 2019.

- [122] Roger S Armen, Mari L DeMarco, Darwin O V Alonso, and Valerie Daggett. Pauling and Corey's alpha-pleated sheet structure may define the prefibrillar amyloidogenic intermediate in amyloid disease. *Proceedings of the National Academy of Sciences of the United States of America*, 101:11622–11627, August 2004.
- [123] Joshua T Berryman, Sheena E Radford, and Sarah A Harris. Systematic examination of polymorphism in amyloid fibrils by molecular-dynamics simulation. *Biophysical journal*, 100(9):2234–2242, 2011.
- [124] Emilie Cerf, Rabia Sarroukh, Shiori Tamamizu-Kato, Leonid Breydo, Sylvie Derclaye, Yves F Dufrière, Vasanthi Narayanaswami, Erik Goormaghtigh, Jean-Marie Ruyschaert, and Vincent Raussens. Antiparallel beta-sheet: a signature structure of the oligomeric amyloid beta-peptide. *The Biochemical journal*, 421:415–423, July 2009.
- [125] Liping Yu, Rohinton Edalji, John E Harlan, Thomas F Holzman, Ana Pereda Lopez, Boris Labkovsky, Heinz Hillen, Stefan Barghorn, Ulrich Ebert, Paul L Richardson, Laura Miesbauer, Larry Solomon, Diane Bartley, Karl Walter, Robert W Johnson, Philip J Hajduk, and Edward T Olejniczak. Structural characterization of a soluble amyloid beta-peptide oligomer. *Biochemistry*, 48:1870–1877, March 2009.
- [126] Samuel J. Bunce, Yiming Wang, Katie L. Stewart, Alison E. Ashcroft, Sheena E. Radford, Carol K. Hall, and Andrew J. Wilson. Molecular insights into the surface-catalyzed secondary nucleation of amyloid- β 40 ($a\beta$ 40) by the peptide fragment $a\beta$ 16–22. *Science Advances*, 5(6), 2019.
- [127] Maria Soledad Celej, Rabia Sarroukh, Erik Goormaghtigh, Gerardo Fidelio, Jean-Marie Ruyschaert, and Vincent Raussens. Alpha-Synuclein Amyloid Oligomers Exhibit Beta-Sheet Antiparallel Structure as Revealed by FTIR Spectroscopy. *Biophysical Journal*, 102(3):440a–441a, June 2019.
- [128] Benjamin Falcon, Wenjuan Zhang, Alexey G Murzin, Garib Murshudov, Holly J Garringer, Ruben Vidal, R Anthony Crowther, Bernardino Ghetti, Sjors H W Scheres, and Michel Goedert. Structures of filaments from Pick's disease reveal a novel tau protein fold. *Nature*, 561:137–140, September 2018.
- [129] Steven J Roeters, Aditya Iyer, Galja Pletikapić, Vladimir Kogan, Vinod Subramaniam, and Sander Woutersen. Evidence for Intramolecular Antiparallel Beta-Sheet

- Structure in Alpha-Synuclein Fibrils from a Combination of Two-Dimensional Infrared Spectroscopy and Atomic Force Microscopy. *Scientific reports*, 7:41051, January 2017.
- [130] Alexey Onufriev, Donald Bashford, and David A Case. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins: Structure, Function, and Bioinformatics*, 55(2):383–394, 2004.
- [131] Ashley Z. Guo, Aaron M. Fluitt, and Juan J. de Pablo. Early-stage human islet amyloid polypeptide aggregation: Mechanisms behind dimer formation. *The Journal of Chemical Physics*, 149(2):025101, 2018.
- [132] William L Jorgensen, Jayaraman Chandrasekhar, Jeffrey D Madura, Roger W Impey, and Michael L Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics*, 79(2):926–935, 1983.
- [133] James A Maier, Carmenza Martinez, Koushik Kasavajhala, Lauren Wickstrom, Kevin E Hauser, and Carlos Simmerling. ff14sb: improving the accuracy of protein side chain and backbone parameters from ff99sb. *Journal of chemical theory and computation*, 11(8):3696–3713, 2015.
- [134] Tai-Sung Lee, David S Cerutti, Dan Mermelstein, Charles Lin, Scott LeGrand, Timothy J Giese, Adrian Roitberg, David A Case, Ross C Walker, and Darrin M York. GPU-Accelerated Molecular Dynamics and Free Energy Methods in Amber18: Performance Enhancements and New Features. *Journal of chemical information and modeling*, 58:2043–2050, October 2018.
- [135] DIBCKMHJ Svergun, Claudio Barberato, and Michel HJ Koch. Crysol—a program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates. *Journal of applied crystallography*, 28(6):768–773, 1995.
- [136] Kiril Tsemekhman, Lukasz Goldschmidt, David Eisenberg, and David Baker. Cooperative hydrogen bonding in amyloid formation. *Protein science : a publication of the Protein Society*, 16:761–764, April 2007.
- [137] Thomas J. Macke and David Case. Modeling unusual nucleic acid structures. *ACS Symp. Ser.*, 682, 11 1997.
- [138] PT Sattianayagam, SDJ Gibbs, D Rowczenio, JH Pinney, AD Wechalekar, JA Gilbertson, PN Hawkins, HJ Lachmann, and JD Gillmore. Hereditary lysozyme

- amyloidosis—phenotypic heterogeneity and the role of solid organ transplantation. *Journal of internal medicine*, 272(1):36–44, 2012.
- [139] Jinxia Lu, Qin Cao, Chuchu Wang, Jing Zheng, Feng Luo, Jingfei Xie, Yichen Li, Xiaojuan Ma, Lin He, David Eisenberg, James Nowick, Lin Jiang, and Dan Li. Structure-Based Peptide Inhibitor Design of Amyloid- β Aggregation. *Frontiers in molecular neuroscience*, 12:54, 2019.
- [140] Stuart A Sievers, John Karanicolas, Howard W Chang, Anni Zhao, Lin Jiang, Onofrio Zirafi, Jason T Stevens, Jan Münch, David Baker, and David Eisenberg. Structure-based design of non-natural amino-acid inhibitors of amyloid fibril formation. *Nature*, 475:96–100, June 2011.
- [141] Jozef Adamcik and Raffaele Mezzenga. Amyloid polymorphism in the protein folding and aggregation energy landscape. *Angewandte Chemie International Edition*, 57(28):8370–8382, 2018.
- [142] Amalia Aggeli, Irina A Nyrkova, Mark Bell, Richard Harding, L Carrick, Tom CB McLeish, Alexander N Semenov, and Neville Boden. Hierarchical self-assembly of chiral rod-like molecules as a model for peptide β -sheet tapes, ribbons, fibrils, and fibers. *Proceedings of the National Academy of Sciences*, 98(21):11857–11862, 2001.
- [143] Alexander K Buell, Anne Dhulesia, Duncan A White, Tuomas P J Knowles, Christopher M Dobson, and Mark E Welland. Detailed analysis of the energy barriers for amyloid fibril growth. *Angewandte Chemie (International ed. in English)*, 51:5247–5251, May 2012.
- [144] Artur Wachtel, Riccardo Rao, and Massimiliano Esposito. Thermodynamically consistent coarse graining of biocatalysts beyond michaelis–menten. *New Journal of Physics*, 20(4):042002, 2018.
- [145] Alexander J Dear, Anđela Šarić, Thomas CT Michaels, Christopher M Dobson, and Tuomas PJ Knowles. Statistical Mechanics of Globular Oligomer Formation by Protein Molecules. *Journal of Physical Chemistry B*, 49(122):11721–11730, 2018.
- [146] Georg Meisl, Julius B Kirkegaard, Paolo Arosio, Thomas C T Michaels, Michele Vendruscolo, Christopher M Dobson, Sara Linse, and Tuomas P J Knowles. Molecular mechanisms of protein aggregation from global fitting of kinetic models. *Nature protocols*, 11:252–272, February 2016.

- [147] Stefan Auer. Simple Model of the Effect of Solution Conditions on the Nucleation of Amyloid Fibrils. *Journal of Physical Chemistry B*, 121(38):8893–8901, 2017.
- [148] Peter G Vekilov. Crystallization tracked atom by atom. 2019.
- [149] Samuel IA Cohen, Michele Vendruscolo, Mark E Welland, Christopher M Dobson, Eugene M Terentjev, and Tuomas PJ Knowles. Nucleated polymerization with secondary pathways. i. time evolution of the principal moments. *The Journal of chemical physics*, 135(6):08B615, 2011.
- [150] Gang Wei, Zhiqiang Su, Nicholas P Reynolds, Paolo Arosio, Ian W Hamley, Ehud Gazit, and Raffaele Mezzenga. Self-assembling peptide and protein amyloids: from structure to tailored function in nanotechnology. *Chemical Society Reviews*, 46(15):4661–4708, 2017.
- [151] Tuomas PJ Knowles, Christopher A Waudby, Glyn L Devlin, Samuel IA Cohen, Adriano Aguzzi, Michele Vendruscolo, Eugene M Terentjev, Mark E Welland, and Christopher M Dobson. An analytical solution to the kinetics of breakable filament assembly. *Science*, 326(5959):1533–1537, 2009.
- [152] Motomasa Tanaka, Sean R Collins, Brandon H Toyama, and Jonathan S Weissman. The physical basis of how prion conformations determine strain phenotypes. *Nature*, 442(7102):585, 2006.
- [153] Anđela Šarić, Yasmine C Chebaro, Tuomas PJ Knowles, and Daan Frenkel. Crucial role of nonspecific interactions in amyloid nucleation. *Proceedings of the National Academy of Sciences*, 111(50):17869–17874, 2014.
- [154] Georg Meisl, Xiaoting Yang, Christopher M Dobson, Sara Linse, and Tuomas PJ Knowles. Modulation of electrostatic interactions to reveal a reaction network unifying the aggregation behaviour of the $\alpha\beta$ 42 peptide and its variants. *Chemical science*, 8(6):4352–4362, 2017.
- [155] Axel Abelein, Jüri Jarvet, Andreas Barth, Astrid Gräslund, and Jens Danielsson. Ionic strength modulation of the free energy landscape of $\alpha\beta$ 40 peptide fibril formation. *Journal of the American Chemical Society*, 138(21):6893–6902, 2016.
- [156] Georg Meisl, Xiaoting Yang, Birgitta Frohm, Tuomas PJ Knowles, and Sara Linse. Quantitative analysis of intrinsic and extrinsic factors in the aggregation mechanism of alzheimer-associated $\alpha\beta$ -peptide. *Scientific reports*, 6:18728, 2016.

- [157] Georg Meisl, Xiaoting Yang, Erik Hellstrand, Birgitta Frohm, Julius B Kirkegaard, Samuel IA Cohen, Christopher M Dobson, Sara Linse, and Tuomas PJ Knowles. Differences in nucleation behavior underlie the contrasting aggregation kinetics of the $a\beta$ 40 and $a\beta$ 42 peptides. *Proceedings of the National Academy of Sciences*, 111(26):9384–9389, 2014.
- [158] Sofia Giorgetti, Claudio Greco, Paolo Tortora, and Francesco Aprile. Targeting amyloid aggregation: an overview of strategies and mechanisms. *International journal of molecular sciences*, 19(9):2677, 2018.