This is a repository copy of *Thermodynamic phase diagram of amyloid-β (16-22) peptide*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/141149/

Version: Accepted Version

Classification: Biological Science; Biophysics and Computational Biology

Thermodynamic phase diagram of amyloid-β (16-22) peptide

Yiming Wang[1], Samuel J. Bunce[2,3], Sheena E. Radford[2,4], Andrew J. Wilson[2,3], Stefan Auer [3], Carol K. Hall[1*]

1. Department of Chemical and Biomolecular Engineering, North Carolina State University, Raleigh, NC 27695-7905, United States

2. Astbury Centre for Structural Molecular Biology, University of Leeds, Woodhouse Lane, Leeds LS2 9JT, United Kingdom.

3. School of Chemistry, University of Leeds, Woodhouse Lane, Leeds LS2 9JT, United Kingdom.

4. School of Molecular and Cellular Biology, Faculty of Biological Sciences, University of Leeds, Woodhouse Lane, Leeds LS2 9JT, United Kingdom.

Corresponding authors

E-mail: **hall@ncsu.edu (CKH)**

Keywords: solubility, phase diagram, amyloidogenic peptide.

Abbreviations: Aβ-Amyloid β protein; TEM-transmission electron microscopy

Abstract

The aggregation of monomeric Aβ peptide into oligomers and amyloid fibrils in the mammalian brain is associated with Alzheimer's disease. Insight into the thermodynamic stability of the Aβ peptide in different polymeric states is fundamental to defining and predicting the aggregation process. Experimental determination of Aβ thermodynamic behavior is challenging due to the transient nature of Aβ oligomers and the low peptide solubility. Furthermore, quantitative calculation of a thermodynamic phase diagram for a specific peptide requires extremely long computational times. Here, using a coarse-grained protein model, molecular dynamics simulations are performed to determine an equilibrium concentration and temperature phase diagram for the amyloidogenic peptide fragment, A$\beta_{16\text{-}22}$. Our results reveal that the only thermodynamically stable phases are the solution phase and the macroscopic fibrillar phase, and that there also exists a hierarchy of metastable phases. The boundary line between the solution phase and fibril phase is found by calculating the temperature-dependent solubility of a macroscopic A$\beta_{16\text{-}22}$ fibril consisting of an infinite number of β-sheet layers. To our knowledge, this is the first *in silico* determination of an equilibrium (solubility) thermodynamic phase diagram for a real amyloid-forming peptide. Furthermore, the *in silico* prediction of A$\beta_{16\text{-}22}$ solubilities over the temperature range of 277-330K agrees well with fibrillation experiments and transmission electron microscopy measurements of the fibril morphologies formed. This *in silico* approach of predicting peptide solubility is also potentially useful for optimizing biopharmaceutical production and manufacturing nanofiber scaffolds for tissue engineering.

Significance Statement

Phase diagrams of atomic systems are calculated routinely by computer simulations, but such calculations are absent for even the simplest peptides. Previous simulations are mainly non-equilibrium, and focus on the assembly of peptides from the monomeric to the aggregated state. To obtain accurate equilibrium solubilities, however, it is necessary to simulate many assembly and disassembly events of fibrillar aggregates, which is notoriously difficult, as it requires breaking many hydrogen bonds. We overcome these challenges and calculate the equilibrium phase diagram of $A\beta_{16\text{-}22}$, the archetypal amyloid former, in the first quantitative calculation of a peptide phase diagram using a realistic protein model. Importantly, our prediction of $A\beta_{16\text{-}22}$ solubility over temperatures from 277K to 330K agrees well with experimental measurements.

\body

Introduction

A$\beta_{16\text{-}22}$ is a 7-residue amyloidogenic peptide (Ac-K-L-V-F-F-A-E-NH$_2$) comprising the central, fibril-forming core of the full-length Aβ peptide, a major constituent of the extracellular plaques associated with Alzheimer's disease. (1, 2) A$\beta_{16\text{-}22}$ has been widely studied due to its relative ease of synthesis and its ability to form well-characterized antiparallel β-sheet fibril structures at concentrations above 55 μM.(3, 4) At neutral pH, monomeric A$\beta_{16\text{-}22}$ is predicted to adopt a random coil configuration that then oligomerizes, passing through an intermediate out-of-register state prior to the final in-register anti-parallel alignment. (5-8) Atomistic simulations of a small number of peptides on the nanosecond time-scale are possible and have been used to examine the structural stability of a variety of A$\beta_{16\text{-}22}$ oligomers, including β-barrels(9), β-sheet-rich dimers(10), trimers(11, 12) and hexamers.(13, 14) Coarse-grained molecular dynamics (MD) simulations have succeeded in predicting spontaneous formation of octamer(15), tri-, tetra- (6) and hepta-layer(16) twisted A$\beta_{16\text{-}22}$ nanofibrils involving up to 192 peptides on the microsecond time-scale.(17) A prerequisite to understanding which structures A$\beta_{16\text{-}22}$ peptides form, and under what conditions, is knowledge of the phase diagram, which characterizes the thermodynamic stability of the peptide phases as a function of relevant parameters, most importantly temperature and peptide concentration. However, despite being studied under a wide range of conditions, a quantitative thermodynamic phase diagram has never been generated for A$\beta_{16\text{-}22}$, or any other peptide, using a realistic protein model.

In general, experiments (18-20) determining the coexistence lines (solubilities) between oligomeric or fibrillar aggregates and the protein solution are difficult to perform because

4

of the metastable nature of the oligomers and the low solubilities of fibrils. Most experiments are performed by preparing a solution at a given concentration and temperature and left for up to weeks to determine whether oligomeric or fibrillar aggregates have formed. (21, 22) The so-observed concentrations and temperatures at which fibrils form can often be far from the true equilibrium concentrations (solubilities) as the observed structures can be kinetically trapped and thus not representative of the thermodynamically most-stable state (23). Experiments in which pre-formed fibrils are put in solution and the concentration at which they elongate or dissolve is measured can also provide estimates of solubility (24, 25).

The challenge in determining a peptide phase diagram for realistic protein models *in silico* is the long computational time needed to calculate the solubility. Molecular simulations aimed at constructing phase diagrams of peptides (26, 27) are all non-equilibrium and only simulate the self-assembly of peptides from the monomeric to the aggregate state. There is no guarantee, however, that structures so obtained will coincide with the thermodynamically stable phases. Furthermore, previous studies do not consider the existence of meta-stable phases. It has only recently become possible to perform Monte Carlo simulations that are capable of constructing a concentration and temperature phase diagram at thermodynamic equilibrium for simple generic peptide models. (28, 29) Given the complexity of biomolecules, the relevance of this finding to any real peptide system is not clear.

Here we construct a temperature-concentration phase diagram for the $A\beta_{16-22}$ peptide by determining the solubility of oligomeric aggregates and fibrils using coarse-grained molecular dynamics and classical nucleation theory. The phase diagram reveals that the

solution phase and fibril phase are the only two thermodynamic stable phases, and that there also exists a hierarchy of metastable aggregate phases. The thermodynamic boundary between the solution phase and the fibril phase is taken to be the solubility line of an infinite layer β-sheet fibril predicted using discontinuous molecular dynamics simulation and classical nucleation theory. The strength and soundness of this approach lies in the fact that our force field, PRIME20, is knowledge-based, sequence specific and accurately predicts peptide orientation, strand-strand and sheet-sheet distances of $A\beta_{16\text{-}22}$ fibril structures. (6) To validate the phase diagram predicted *in silico*, we conduct $A\beta_{16\text{-}22}$ fibrillation experiments at peptide concentrations varying from 10 μM to 200μM at 277K-330K. Judging from TEM images, the experimental conditions under which the $A\beta_{16\text{-}22}$ solution forms fibrils agrees well with the predicted solution-fibril phase boundary. To the authors' knowledge, this is the first *in silico* determination of the thermodynamic phase diagram for a real amyloid-forming peptide and the first to be validated by experimental measurements.

Results and Discussion

Protein model, Structure of $A\beta_{16\text{-}22}$ oligomer and fibril

The $A\beta_{16\text{-}22}$ aggregates are modelled using a coarse-grained protein model, called PRIME20 (30, 31), in which each amino acid residue is represented by three backbone spheres, one each for N-H, C-H, and C=O, and a single side-chain sphere R (Fig. 1A). The non-directional interaction between sidechain spheres and the directional hydrogen bond between N-H and C=O spheres are modelled as square-well potentials (Fig. 1B and 1C). While the strength of the directional backbone hydrogen bonding interaction is the same for all residues, the strength of the non-directional sidechain-sidechain interaction is sidechain specific. See Methods for details.

Here we focus on five types of aggregates formed by Aβ_{16-22}: non-hydrogen-bonded (non-HB) oligomers (Fig. 1D), hydrogen-bonded (HB) oligomers (Fig. 1E), and fibrils composed of 2, 3 and 4 β-sheet layers (Figs. 1F, G, H). The non-HB oligomer is defined to be an aggregate formed solely by inter-peptide hydrophobic sidechain-sidechain (HP) interactions with no inter-peptide hydrogen bonds; intra-peptide hydrogen bonds are allowed. The HB oligomer is defined to be an aggregate that has at most one inter-peptide hydrogen bond between any two neighboring peptides and any number of intra-peptide hydrogen bonds. We consider the non-HB and HB oligomers to be two distinct states because the peptides in these two states have different numbers of hydrogen bonds and hydrophobic sidechain contacts with neighboring peptides (see Fig. S1). The β-sheet aggregates are stabilized by having sidechain HP contacts between tightly-packed anti-parallel β-sheets and a maximum number of backbone HBs between fully-extended peptides within each β-sheet, consistent with X-ray diffraction and solid-state NMR measurements (2, 6). Our simulations show that the probability distributions for the number of HB and HP sidechain contacts for the five aggregates do not change significantly with temperature over the studied temperature range.

Measurement of solubility for Aβ_{16-22} oligomers and fibrils

In order to determine the solubility of aggregates (oligomers and fibrils) formed by the Aβ_{16-22} peptides, we perform molecular dynamics simulations in the canonical ensemble. The solubility is defined to be the peptide concentration above which the peptide will join with others to form an aggregate and below which the peptide will remain in solution. In our simulations, the solubilities for the different Aβ_{16-22} aggregates (oligomers and fibrils) at a given temperature are measured as the equilibrium monomer concentration at which

7

the aggregate neither grows nor shrinks. (See Methods for details). This approach was used by Bai and Li (32) to measure solid-liquid phase equilibrium for the Lennard-Jones (LJ) fluid. Their calculation of the solid-liquid interfacial energy (surface tension) using classical nucleation theory was in qualitative agreement with results calculated using other theoretical approaches, e.g. interface fluctuation method (33) and reversible work integration (34). Later, Auer and co-workers used this method along with Monte Carlo simulation to calculate a phase diagram for a simple homo-polypeptide model where each amino acid is represented by one Cα bead (28, 29). Their predicted values for the surface tension were in agreement with theoretical predictions for protein crystals using a simple lattice model (35).

We first measured the solubility for the non-HB and HB oligomers. Fig. 2A shows that the solubilities for the non-HB and HB oligomers increase with increasing temperature. This is expected as at higher temperatures the thermal energy makes it easier for peptides to detach from the oligomer so that higher concentrations are required to stabilize these aggregates. The latent heat of aggregation for monomeric peptides into oligomers can be obtained by fitting the van't Hoff equation,

$$C_e = C_r \exp(-\frac{L}{k_B T}) \tag{1}$$

to the simulation data (Fig. 2B), where C_e is the equilibrium fibril solubility, C_r is a temperature-independent reference concentration, k_B is the Boltzmann constant, T is the simulation temperature, and L is the latent heat of monomer peptide aggregation into the oligomer or fibril. From the simulation, we also calculated the average peptide binding energy (the sum of the HP and HB energies per peptide) within the non-HB and HB oligomers, including all the peptides on the surface and inside the aggregate, and found

that the latent heat is about half of the average peptide binding energy E_B, (see Fig. 2C) consistent with earlier work(29).

Although the solubility of the oligomers should depend on the oligomer size as described by the Ostwald formula, (28, 29, 36) (which predicts that solubility decreases with increasing oligomer size), we do not consider the effect of oligomer size here. This is because large oligomers do not form for this A$\beta_{16\text{-}22}$ system. The simulations show that the non-HB and HB A$\beta_{16\text{-}22}$ oligomers are relatively small (containing around ten peptides) and easily break apart into smaller aggregates when they contain more than fifteen peptides.

Next we determined the solubilities for 2, 3, 4 and ∞ β-sheet A$\beta_{16\text{-}22}$ fibrils. The simulation results (Fig. 2A) show that at a given temperature, the solubilities for the 2, 3 and 4 β-sheets decrease with increasing thickness (number of β-sheet layers) of the fibril. As the number of β-sheets in the fibril increases, the average number of HP contacts per peptide increases, as shown in Fig. S1C, leading to increased stability of the fibril. A quantitative description of the dependence of the solubility of a fibril with i layers (i=1, 2, 3, etc.) of a β-sheet, $C_{i\beta}$, on thickness at fixed temperature can be derived from classical nucleation theory and is given by (37)

$$C_{i\beta} = C_{\infty\beta}\exp\left(\frac{2\psi_h}{k_BT}\cdot\frac{1}{i}\right) \tag{2}$$

where $C_{\infty\beta}$ is the solubility of an infinite thick fibril, $\psi_h = a_h\sigma_h$ is the fibril surface energy parallel to the fibril thickening axis, σ_h is the specific surface energy of the face perpendicular to the fibril axis, and a_h is the lateral surface area occupied by each peptide within one β-sheet as defined in (37). A linear fit of our simulation data ($lnC_{i\beta}$ vs 1/i) to Eq. (2) (Fig. 2D) yields an estimate for the solubility $C_{\infty\beta}$ as a function of temperature for the infinitely thick fibril. As before, the latent heats of peptide aggregation into fibril phases

from solution can be obtained by a fit of our data to Eq. (1), see Fig. 2B, and the obtained values are shown in Fig. 2C. From Fig. 2C, we find that for 2, 3 and 4 β-sheet fibrils, the values of the latent heat are comparable to the average peptide binding energy. The relation between latent heat and protein-protein interaction energy for a protein crystal was first derived by Haas and Drenth using a simple lattice model (35). They found that the latent heat should be half of the average protein-protein interaction energy; this was derived by using the zero-temperature approximation, which does not consider the entropic effects. We speculate that it might be the neglect of the entropic effect that causes the relation between latent heat and peptide binding energy to hold for oligomers but not for 2, 3 and 4 β-sheet fibrils in this work. In addition, the fit of our simulation data to Eq. (2) also yields values for $\psi_h = a_h \sigma_h$, and knowledge of a_h enables us to estimate the average surface tension σ_h of the A$\beta_{16\text{-}22}$ fibril (Fig. S2). The lateral surface area per peptide, $a_h = 106.1\text{Å}$, is the product of the length of a peptide within a β-sheet (22.1Å) and the inter-peptide distance in one β-sheet (4.8Å). Hence, depending on temperature, the average surface tension σ_h for the A$\beta_{16\text{-}22}$ fibril ranges from 24 to 30 mJ/m^2 (see Fig. S2), which is in the range of 0.1-30 mJ/m^2 reported for protein crystals in aqueous solution (35, 38-40).

A$\beta_{16\text{-}22}$ phase diagram

The major finding of this study is our quantitative calculation of the thermodynamic phase diagram for the A$\beta_{16\text{-}22}$ peptide fragment (Fig. 3A). The peptide solution phase (cyan region) and the macroscopic fibrillar phase (yellow region) are the only two thermodynamically stable phases. The thermodynamic boundary between these phases is taken to be the solubility line for a fibril with an infinite number of β-sheets. (Error bars are not shown in Figs. 3 and 4A because the R^2 value for the linear fit of $C_{\infty\beta}$ vs T data to

obtain $C_{\infty\beta}(\text{T})$ is close to 1). Within the fibril phase region of the phase diagram, there exists a series of metastable aggregate phases separated by the metastable solubility lines for the non-HB oligomer, HB oligomer and 2, 3 and 4 β-sheet fibrils. A hierarchy of metastable phases with the ultimate stable boundary being between solution and infinitely thick fibril was first observed by Auer et al. (28).

To justify the infinite-thickness approximation in calculating fibril solubility, we calculated A$\beta_{16\text{-}22}$ fibril solubility versus fibril thickness (Fig. S3) and found that fibril solubility remains unchanged after the thickness exceeds 30 layers. We think that an actual A$\beta_{16\text{ -}22}$ fibril may contain more than 30 layers of β-sheets at neutral pH, because A$\beta_{1\text{-}42}$ (41), A$\beta_{10\text{-}35}$ (42) and A$\beta_{18\text{-}28}$ (43) form fibrils *in vitro* with 2, 6 and 24 layers of β-sheets, respectively at neutral pH. This is consistent with a hypothesis by Lu et al. (44) that shorter amyloidogenic peptides tend to adopt more planar β-sheets and maintain more layers in the fibril than longer peptides.

To better illustrate why the solubility line for the fibril is regarded as the coexistence line, and why the oligomer and various multi-sheet fibril phases are metastable with respect to the fibril phase, we put our calculations in the context of more familiar thermodynamic arguments in which the compositions of two phases in equilibrium are found by minimizing their free energies. Fig. S4 plots the Helmholtz free energies of the solution, oligomer and fibril phases as a function of the peptide mole fraction, x_p, (which is simply related to the peptide concentration for a dilute solution) in a peptide-water mixture at fixed temperature and volume. A double tangent (also called convex-envelop) construction on the free energies of the solution and fibril phase determines the peptide mole fractions, $x_p^{soln(1)}$ and x_p^{fib} at which the solution and fibril phases are in equilibrium. The construction locates the

peptide mole fractions at which the peptide chemical potentials (the $x_p = 1$ intercepts of lines tangent to the free energy curves) are equal, i.e. $\mu_p^{soln(1)}(x_p^{soln(1)}) = \mu_p^{fib}(x_p^{fib})$. The composition, $x_p^{soln(1)}$, is equivalent to the peptide concentration above which the monomer goes through a phase transition to another phase, the solubility line determined in our simulations. Similarly a double tangent construction on the solution and oligomer phases yields peptide mole fractions $x_p^{soln(2)}$ and x_p^{olig} with equal peptide chemical potentials, $\mu_p^{soln(2)}(x_p^{soln(2)}) = \mu_p^{olig}(x_p^{olig})$, signifying equilibrium between solution and oligomer phases in this concentration range. This is, however, a metastable equilibrium compared to the equilibrium between solution at $x_p^{soln(1)}$ and fibril at x_p^{fib} because the peptide chemical potential in solution is higher than that when the solution is in equilibrium with the fibril. Note also that having $\mu_p^{soln(2)}\left(x_p^{soln(2)}\right) > \mu_p^{soln(1)}(x_p^{soln(1)})$ means that $x_p^{soln(2)} > x_p^{soln(1)}$, consistent with what is measured in simulations. Notice also that in Fig. S4, there are no metastable phases when $x_p^{soln(1)} < x_p < x_p^{soln(2)}$; the oligomer phase only becomes metastable with respect to the fibril phase when $x_p > x_p^{soln(2)}$. This is also reflected in the solubility versus temperature phase diagram, Fig. 3A, which shows a region of concentrations where the HB oligomer is unstable (below the dashed purple line and above the black solubility line), and a region where the oligomer is metastable (above the dashed purple line). Similar arguments can be made for the other metastable phases, such as the 2, 3, and 4 β-sheet fibrils phases. We emphasize that we did not calculate free energies, as we were able to calculate the solubilities directly; the discussion here is mainly presented for pedagogical purposes.

Knowledge of the thermodynamically stable and metastable phases in the phase diagram

allow us to determine the conditions under which the oligomer or fibril can form. A previous discontinuous molecular dynamics (DMD) study of Aβ_{16-22} kinetic aggregation by Cheon *et al.* (6) found that at 20mM and T*=0.2 corresponding to 342K, a system of peptides in a random configuration first forms a fibril nucleus which then grows into fibrils. At the same concentration but at a lower temperature T*=0.17 corresponding to T=273K, the peptides first form HB oligomers that later merge and rearrange to form fibrils. Snapshots from new simulations at T=342K and 273K are shown in Fig. 3B, top and bottom panels. Our thermodynamic phase diagram helps explain why Aβ_{16-22} aggregates via a "one-step" pathway at high temperature and a "two-step" pathway at low temperature. From Fig. 3A, at point A (20mM, T=342K), the peptide concentration is greater than the solubility of fibril, C>$C_{e,fibril}$(T=342K) but smaller than the solubility of the HB oligomer, C<$C_{e,HB\ oligomer}$ (T=342K). As Point A satisfies the thermodynamic criteria for stable fibril formation but not for HB oligomer formation, it is not surprising that the aggregation kinetics is characterized by nucleus formation and subsequent fibril growth. In comparison, point B (20mM, T=273K) satisfies the thermodynamic condition for formation of a stable fibril and of a metastable HB oligomer C>$C_{e,fibril}$(T=273K) and C>$C_{e,HB\ oligomer}$ (T=273K); this is conducive to two-step aggregation kinetics in which several oligomers form and later merge and rearrange to form fibrils. These arguments about the connection between the fibril formation kinetics and aggregate stability and metastability are meant here only to be suggestive. The best way to predict Aβ_{16-22} aggregation pathways at different temperatures and concentrations would be to construct aggregation free energy landscapes along the appropriate reaction coordinates.

To validate the *in silico* prediction of Aβ_{16-22} solubility at biophysically relevant

temperatures, transmission electron microscopy was used to determine whether fibrils had formed after a predetermined time (2 weeks incubation). Fig. 4 is a summary of our experimental results near the predicted solubility boundary line at temperatures ranging from 277K to 330K. At T=330K, fibrils form at both 200μM and 300μM but no fibrils are observed when peptide concentration is at or below 100μM, which is consistent with the predicted solubility (177μM) at 330K. Furthermore, at T=277K, 296K and 310K, fibrils form when the concentration is equal to or above 20μM, which also agrees with our predicted solubility at T=277K, 296K and 310K of 0.2μM, 2.6μM and 16.5μM, respectively. It is important to note that even though fibrils were not observed in the TEM images at 10 μM at all temperatures tested, this does not rule out the presence of low populations of fibrils under these conditions which are below the detection limits of this method, i.e. low possible surface adsorption of the fibrils. Thus, we conclude that the boundaries predicted by the phase diagram quantitatively agree with the experimental findings, demonstrating the power of this approach to understand the thermodynamic stability of peptide assemblies.

The concentration and temperature conditions at which $A\beta_{16-22}$ forms fibrils reported from a number of other *in vitro* studies also agree with our simulation prediction (Figs. 4 and S5). (2, 3, 8, 45) It should be noted that the only studies included in this comparison use $A\beta_{16-22}$ with capped N- and C-termini, since uncapping the termini can have a substantial twisting effect on the fibril supramolecular structure.(46) Most of these studies were performed at high concentrations (>200μM), placing them in regions of the phase diagram that would predict fibril formation, even at the highest temperature studied (55 °C, 328K). (7) Only one study by Senguen *et al*. (3) demonstrated that fibrils are formed at a

14

concentration of 55 μM at 37 °C (310K), again in agreement with our phase diagram.

Knowledge of the peptide solution-fibril phase boundary could allow us to modulate the quality of nanofiber structures formed. This can be accomplished by carefully controlling the driving force for fibril formation, e.g. highly-ordered fibrils can be obtained at solution conditions close to the phase boundary while less-ordered entangled fibrils can be obtained at conditions away from the phase boundary. This suggests that our computation-based approach could inform future fabrication of amyloid nanofiber substrates / scaffolds with desired microstructure properties for various tissue engineering and biomedical applications.(47, 48) In addition, controlling protein aggregation is a key problem for the purification and storage of therapeutic proteins.(49, 50) Thus, the knowledge of solubility of specific peptides and proteins over a wide range of temperatures could be used to effectively prevent their unwanted agglomeration.

Our approach can be applied to investigate the effect of mutations on peptide solubility. For example, in preliminary simulations, we show that F19A mutant of A$\beta_{16\text{-}22}$ remains soluble at C=20mM and T=342K in contrast to the fibril-forming behavior of the wild type at the same conditions. In fact, our predicted solubility of the F19A mutant is about 600 mM at 307K. (see Fig. S6) This result is consistent with experimental finding by Senguen *et al.* that the F19A mutant of A$\beta_{16\text{-}22}$ is much more soluble (at least 350 μM) than its wild type at body temperature (3). One of our next steps will be to calculate the solubility of A$\beta_{16\text{-}22}$ mutants at positions 19 and 20 and compare our results with experimental measurements.

Conclusion

In this work we demonstrated that by using a relatively realistic coarse-grained protein

model it is possible to calculate *via* simulation an equilibrium concentration and temperature phase diagram for a short amyloidogenic peptide, A$\beta_{16\text{-}22}$. The predicted phase diagram provides insight into the thermodynamic stability of the different types of aggregates formed by A$\beta_{16\text{-}22}$ under biophysically relevant conditions. This helps us to understand the conditions under which they can form, and how to prevent them from forming. The solubilities predicted *in silico* are in the range expected from experiments and as such also provide a rigorous test of the validity of the model and methodology used. As is shown in Fig. S6, the solubility of Aβ(16-22) peptide calculated in this work is two to three orders of magnitudes higher than that of the generic polypeptide calculated by Auer et al. (28,29). To the authors' knowledge, our work is the first *in silico* determination of an equilibrium (solubility) thermodynamic phase diagram for a non-trivial peptide. Although peptide phase diagrams have been calculated via simulations (26, 27), they are usually based on the kinetics of phase formation as opposed to a direct solubility measurement. Furthermore, the *in silico* prediction of A$\beta_{16\text{-}22}$ solubilities over the temperature range of 4-57°C agrees well with fibrillation experiments and transmission electron microscopy measurements of the fibril morphologies formed. This *in silico* approach of predicting peptide solubility is also potentially useful for engineering nanofiber scaffolds for biomedical applications as well as manufacturing biopharmaceuticals.

Methods

In silico peptide model

The molecular model that we use is a four-sphere-per-residue model in which each amino acid residue is represented by three backbone spheres, one each for N-H, C-H, and C=O, and one for the side-chain sphere, R. (30, 31) For illustrative purposes, Fig. S7 shows

a simulation snapshot of the structure of a two-layer β-sheet fibril formed by Aβ_{16-22} peptides. The PRIME20 model has unique geometric and energetic parameters for each of the 20 amino acids. Specifically, each sidechain sphere of the 20 amino acids has a distinct hard sphere diameter (effective van der Waals radius) and sidechain-to-backbone distances (R-Cα, R-NH, and R-CO). The interactions between the atoms are described by discontinuous potentials including the hard sphere, square-well and square-shoulder potentials. The strength of the directional hydrogen bond is described by the well depth ε_{HB} and is the same for all residues, and the strength of the non-directional hydrophobic interaction between side chains i and j is described by a sidechain specific well depth (ε_{HP-ij}). The potential energy parameters between the 20 different amino acids include 210 independent square well widths and 19 independent square well depths derived by using a perceptron learning algorithm that optimizes the energy gap between 711 known native states from the PDB and decoy structures. (31) For example, the interaction strength, ε_{HP-AA}, between two alanine sidechains is 0.084ε_{HB}, where the hydrogen bonding strength ε_{HB} is chosen to be 12.47kJ/mol as was used in previous work(17). The details of the assignment of the values of the interaction parameters of the PRIME20 model are described in earlier work.(17, 30, 31) The reduced temperature is defined as $T^* = k_BT/\varepsilon_{HB}$. It can be related to a real temperature by T [K] = 2288.46T^*-115.79; this equation was obtained by matching the folding temperature of alanine-rich polypeptides in our previous discontinuous molecular dynamic (DMD) simulations (17) to the experimental values(51). The reduced time unit is $\Delta t_{reduced}$ = 0.96 ns, which has been obtained by matching the self-diffusion coefficient of Aβ_{16-22} obtained from DMD simulations to that calculated from atomistic MD simulation.(17)

DMD simulations with the PRIME20 model are ideally suited for simulating protein aggregation in three respects. (i) They allow simulation of the complete aggregation process from a system of random coil peptides to a fibrillar structure (up to 200 peptides) at time scales up to 100μs (16, 17). (ii) The PRIME20 model is realistic enough to ensure that the predicted fibril structures agree well with X-ray diffraction and solid-state NMR measurements (6). Its adequacy and efficiency has been demonstrated by its application to short peptide systems including $A\beta_{16-22}$ (6), the designed hexapeptide sequences of Lopez de la Paz *et al.*(52, 53), the tau fragment(54), and longer peptide systems including $A\beta_{17-36}$ (55) and $A\beta_{17-42}$(56). (iii) The sidechain spheres of the twenty possible amino acids each have distinct geometric and energetic parameters, ensuring that the peptide is modelled in a sequence specific manner.

Simulation method

The solubilities for the different $A\beta_{16-22}$ aggregates (oligomers and fibrils) were measured using an approach first proposed by Bai and Li (32) and later adopted by Auer.(29) The method is based on the definition of the solubility of an aggregate at a given temperature as the equilibrium monomer concentration at which the aggregate neither grows nor shrinks. The pre-formed aggregate is initially placed in the center of a cubic box surrounded by monomer peptides. Periodic boundary conditions are imposed. At constant temperature, we constrain the fibril to avoid new β-sheet creation while allowing it to elongate or shrink; the concentration of monomeric peptides is monitored until it reaches a plateau and remains constant thereafter for a long time. (Fig. S8) The solubility of the aggregate at any given temperature is defined to be the monomeric peptide concentration in that final equilibrium state. Although this method is simple in principle, it is very costly in terms of simulation

time. In fact, to observe whether a pre-formed aggregate grows or shrinks requires simulation of hundreds of attachment and detachment events to and from the aggregate. Detachment events of peptides from fibrils are particularly rare as they require breakage of several strong hydrogen bonds. Due to the infrequency of detachment events, the total number of peptides used in our simulations is relatively small (sixteen to eighty peptides).

Here we measured the solubilities of five types of $A\beta_{16-22}$ aggregates including a non-hydrogen-bonded oligomer (non-HB oligomer), a hydrogen-bonded oligomer (HB oligomer), and fibrils containing 2, 3, and 4 β-sheets. We choose to focus our attention on the non-HB and HB oligomers because they represent two limiting examples of the forces that can stabilize oligomers formed early in the fibril formation process. The solubility profiles of HB oligomers with more than one HB formed between any peptide pair are expected to be similar to the ones we study here and thus are omitted. The 2, 3 and 4 β-sheet fibrils are defined to be aggregates that contain two, three and four stacked β-sheets. The system size and simulation details are described in supporting information. To measure the solubility of potentially metastable aggregates, constraints are required to avoid structural conversion of one type of metastable aggregate into another (i.e. the growth of a non-HB oligomer into a HB oligomer or the growth of a 2β-sheet fibril into a 3β-sheet fibril). Methods used to accomplish this are described in supporting information.

$A\beta_{16-22}$ solid-phase peptide synthesis, aggregation conditions and Transmission Electron Microscopy (TEM)

The experimental details are described in the supporting information.

Acknowledgements

References

1. Knowles TPJ, Vendruscolo M, & Dobson CM (2014) The amyloid state and its association with protein misfolding diseases. *Nat. Rev. Mol. Cell Biol.* 15(6):384-396.

2. Balbach JJ*, et al.* (2000) Amyloid fibril formation by A beta(16-22), a seven-residue fragment of the Alzheimer's beta-amyloid peptide, and structural characterization by solid state NMR. *Biochemistry* 39(45):13748-13759.

3. Senguen FT*, et al.* (2011) Probing aromatic, hydrophobic, and steric effects on the self-assembly of an amyloid-beta fragment peptide. *Mol. Biosyst.* 7(2):486-496.

4. Preston GW, Radford SE, Ashcroft AE, & Wilson AJ (2012) Covalent Cross-Linking within Supramolecular Peptide Structures. *Anal. Chem.* 84(15):6790-6797.

5. Santini S, Wei GH, Mousseau N, & Derreumaux P (2004) Pathway complexity of Alzheimer's beta-amyloid A beta(16-22) peptide assembly. *Structure* 12(7):1245-1255.

6. Cheon M, Chang I, & Hall CK (2011) Spontaneous Formation of Twisted A beta(16-22) Fibrils in Large-Scale Molecular-Dynamics Simulations. *Biophys. J.* 101(10):2493-2501.

7. Petty SA & Decatur SM (2005) Experimental Evidence for the Reorganization of β-Strands within Aggregates of the Aβ(16−22) Peptide. *J. Am. Chem. Soc.* 127(39):13488-13489.

8. Hsieh M-C, Liang C, Mehta AK, Lynn DG, & Grover MA (2017) Multistep Conformation Selection in Amyloid Assembly. *J. Am. Chem. Soc.* 139(47):17007-17010.

9. Irbäck A & Mitternacht S (2008) Spontaneous β-barrel formation: An all-atom Monte Carlo study of Aβ16–22 oligomerization. *Proteins: Struct., Funct., Bioinf.* 71(1):207-214.

10. Gnanakaran S, Nussinov R, & García AE (2006) Atomic-Level Description of Amyloid β-Dimer Formation. *J. Am. Chem. Soc.* 128(7):2158-2159.

11. Klimov DK & Thirumalai D (2003) Dissecting the Assembly of Aβ16–22 Amyloid Peptides into Antiparallel β Sheets. *Structure* 11(3):295-307.

12. Nguyen PH, Li MS, & Derreumaux P (2011) Effects of all-atom force fields on amyloid oligomerization: replica exchange molecular dynamics simulations of the A[small beta]16-22 dimer and trimer. *Phys. Chem. Chem. Phys.* 13(20):9778-9788.

13. Favrin G, Irbäck A, & Mohanty S (2004) Oligomerization of Amyloid Aβ16–22 Peptides Using Hydrogen Bonds and Hydrophobicity Forces. *Biophys. J.* 87(6):3657-3664.

14. Rohrig UF, Laio A, Tantalo N, Parrinello M, & Petronzio R (2006) Stability and structure of oligomers of the Alzheimer peptide A beta(16-22): From the dimer to the 32-mer. *Biophys. J.* 91(9):3217-3229.

15. Lu Y, Derreumaux P, Guo Z, Mousseau N, & Wei G (2009) Thermodynamics and dynamics of amyloid peptide oligomerization are sequence dependent. *Proteins: Struct., Funct., Bioinf.s* 75(4):954-963.

16. Latshaw DC, Cheon M, & Hall CK (2014) Effects of Macromolecular Crowding on Amyloid Beta (16-22) Aggregation Using Coarse-Grained Simulations. *J. Phys. Chem. B* 118(47):13513-13526.

17. Wang Y, Shao Q, & Hall CK (2016) N-terminal Prion Protein Peptides (PrP(120–144)) Form Parallel In-register β-Sheets via Multiple Nucleation-dependent Pathways. *J. Biol. Chem.* 291(42):22093-22105; Correction: 292(50), 20655-20655.

18. Aggeli A*, et al.* (2001) Hierarchical self-assembly of chiral rod-like molecules as a model for peptide beta-sheet tapes, ribbons, fibrils, and fibers. *Proc. Natl. Acad. Sci. U. S. A.* 98(21):11857-11862.

19. Hamley IW*, et al.* (2010) Influence of the Solvent on the Self-Assembly of a Modified Amyloid Beta Peptide Fragment. II. NMR and Computer Simulation Investigation. *J. Phys. Chem. B* 114(2):940-951.

20. Annunziata O*, et al.* (2005) Oligomerization and Phase Transitions in Aqueous Solutions of Native and Truncated Human βB1-Crystallin. *Biochemistry* 44(4):1316-1328.

21. Hellstrand E, Boland B, Walsh DM, & Linse S (2010) Amyloid beta-Protein

Aggregation Produces Highly Reproducible Kinetic Data and Occurs by a Two-Phase Process. *ACS Chem. Neurosci.* 1(1):13-18.

22. Mason TO*, et al.* (2017) Thermodynamics of Polypeptide Supramolecular Assembly in the Short-Chain Limit. *J. Am. Chem. Soc.* 139(45):16134-16142.

23. Ricchiuto P, Brukhno AV, & Auer S (2012) Protein Aggregation: Kinetics versus Thermodynamics. *J. Phys. Chem. B* 116(18):5384-5390.

24. Chen S & Wetzel R (2001) Solubilization and disaggregation of polyglutamine peptides. *Protein Sci.* 10(4):887-891.

25. O'Nuallain B, Shivaprasad S, Kheterpal I, & Wetzel R (2005) Thermodynamics of Aβ(1−40) Amyloid Fibril Elongation. *Biochemistry* 44(38):12709-12718.

26. Dima RI & Thirumalai D (2002) Exploring protein aggregation and self-propagation using lattice models: Phase diagram and kinetics. *Protein Sci.* 11(5):1036-1049.

27. Nguyen HD & Hall CK (2004) Phase diagrams describing fibrillization by polyalanine peptides. *Biophys. J.* 87(6):4122-4134.

28. Auer S & Kashchiev D (2010) Phase Diagram of alpha-Helical and beta-Sheet Forming Peptides. *Phys. Rev. Lett.* 104(16):168105.

29. Auer S (2011) Phase diagram of polypeptide chains. *J. Chem. Phys.* 135(17):175103.

30. Nguyen HD & Hall CK (2004) Molecular dynamics simulations of spontaneous fibril formation by random-coil peptides. *Proc. Natl. Acad. Sci. U. S. A.* 101(46):16180-16185.

31. Cheon M, Chang I, & Hall CK (2010) Extending the PRIME model for protein

aggregation to all 20 amino acids. *Proteins: Struct. Funct. Bioinf.* 78(14):2950-2960.

32.   Bai XM & Li M (2006) Calculation of solid-liquid interfacial free energy: A classical nucleation theory based approach. *J. Chem. Phys.* 124(12):124707.

33.   Morris JR & Song XY (2003) The anisotropic free energy of the Lennard-Jones crystal-melt interface. *J. Chem. Phys.* 119(7):3920-3925.

34.   Broughton JQ & Gilmer GH (1986) Molecular dynamics investigation of the crystal–fluid interface. VI. Excess surface free energies of crystal–liquid systems. *J. Chem. Phys.* 84(10):5759-5768.

35.   Haas C & Drenth J (1995) The interaction energy between two protein molecules related to physical properties of their solution and their crystals and implications for crystal growth. *J. Cryst. Growth* 154(1):126-135.

36.   Kashchiev D (2000) *Nucleation: Basic Theory with Applications* (Butterworth-Heinemann,Oxford).

37.   Kashchiev D & Auer S (2010) Nucleation of amyloid fibrils. *J. Chem. Phys.* 132(21):215101.

38.   Garai K, Sahoo B, Sengupta P, & Maiti S (2008) Quasihomogeneous nucleation of amyloid beta yields numerical bounds for the critical radius, the surface tension, and the free energy barrier for nucleus formation. *J. Chem. Phys.* 128(4):045102.

39.   Gosavi RA, Bhamidi V, Varanasi S, & Schall CA (2009) Beneficial Effect of Solubility Enhancers on Protein Crystal Nucleation and Growth. *Langmuir* 25(8):4579-4587.

40. Bromberg L, Rashba-Step J, & Scott T (2005) Insulin Particle Formation in Supersaturated Aqueous Solutions of Poly(Ethylene Glycol). *Biophys. J.* 89(5):3424-3433.

41. Gremer L*, et al.* (2017) Fibril structure of amyloid-beta(1-42) by cryo-electron microscopy. *Science* 358(6359):116-119.

42. Burkoth TS*, et al.* (2000) Structure of the beta-amyloid((10-35)) fibril. *J. Am. Chem. Soc.* 122(33):7883-7889.

43. Kirschner DA*, et al.* (1987) Synthetic peptide homologous to beta protein from Alzheimer disease forms amyloid-like fibrils in vitro. *Proc. Natl. Acad. Sci. U. S. A.* 84(19):6953-6957.

44. Lu K, Jacob J, Thiyagarajan P, Conticello VP, & Lynn DG (2003) Exploiting amyloid fibril lamination for nanotube self-assembly. *J. Am. Chem. Soc.* 125(21):6391-6393.

45. Childers WS, Anthony NR, Mehta AK, Berland KM, & Lynn DG (2012) Phase Networks of Cross-beta Peptide Assemblies. *Langmuir* 28(15):6386-6395.

46. Tao K*, et al.* (2011) Self-Assembly of Short A beta(16-22) Peptides: Effect of Terminal Capping and the Role of Electrostatic Interaction. *Langmuir* 27(6):2723-2730.

47. Zhang SG (2003) Fabrication of novel biomaterials through molecular self-assembly. *Nat. Biotechnol.* 21(10):1171-1178.

48. Krieg E, Bastings MMC, Besenius P, & Rybtchinski B (2016) Supramolecular Polymers in Aqueous Media. *Chem. Rev.* 116(4):2414-2477.

49. Roberts CJ (2014) Therapeutic protein aggregation: mechanisms, design, and

control. *Trends Biotechnol.* 32(7):372-380.

50. Rosa M, Roberts CJ, & Rodrigues MA (2017) Connecting high-temperature and low-temperature protein stability and aggregation. *PLoS One* 12(5):e0176748.

51. Muñoz V & Serrano L (1995) Elucidating the Folding Problem of Helical Peptides using Empirical Parameters. III>Temperature and pH Dependence. *J. Mol. Biol.* 245(3):297-308.

52. de la Paz ML & Serrano L (2004) Sequence determinants of amyloid fibril formation. *Proc. Natl. Acad. Sci. U. S. A.* 101(1):87-92.

53. Wagoner VA, Cheon M, Chang I, & Hall CK (2012) Fibrillization Propensity for Short Designed Hexapeptides Predicted by Computer Simulation. *J. Mol. Biol.* 416(4):598-609.

54. Cheon M, Chang I, & Hall CK (2012) Influence of temperature on formation of perfect tau fragment fibrils using PRIME20/DMD simulations. *Protein Sci.* 21(10):1514-1527.

55. Wang Y, Latshaw DC, & Hall CK (2017) Aggregation of Aβ(17–36) in the Presence of Naturally Occurring Phenolic Inhibitors Using Coarse-Grained Simulations. *J. Mol. Biol.* 429(24):3893-3908.

56. Cheon M, Hall CK, & Chang I (2015) Structural Conversion of A beta(17-42) Peptides from Disordered Oligomers to U-Shape Protofilaments via Multiple Kinetic Pathways. *PLoS Comput. Biol.* 11(5):e1004258.
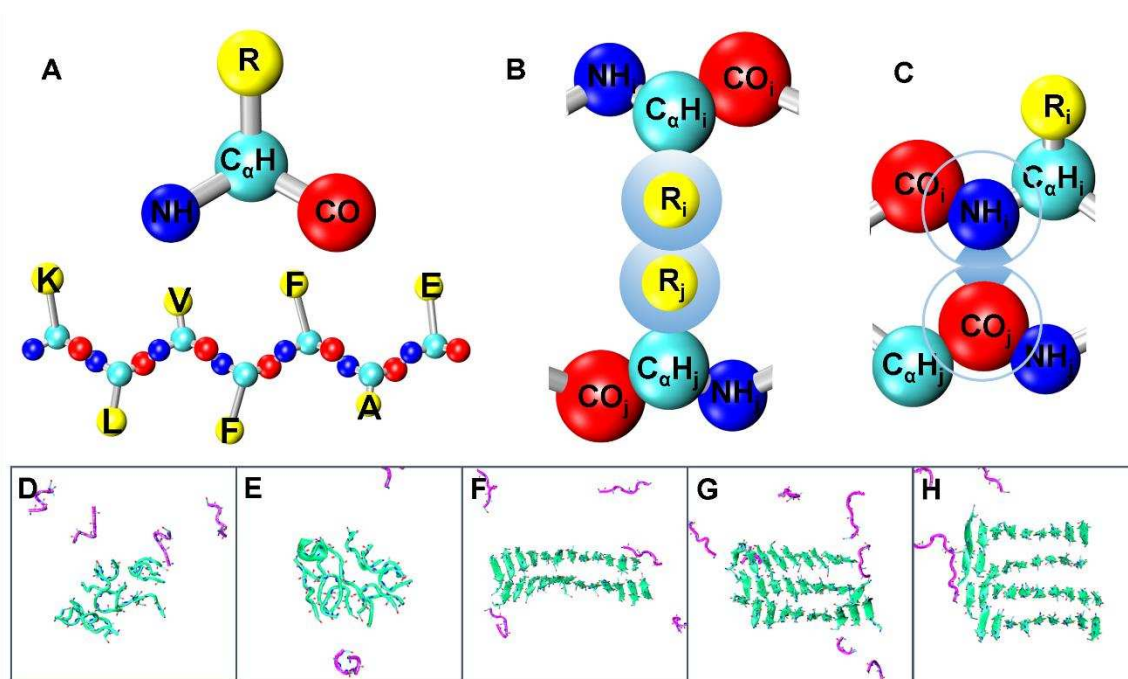
Figure Legends

Figure 1. (A) Representation of the A$\beta_{16\text{-}22}$ peptide in the PRIME20 model. (B) and (C) are schematic representations of the square-well interaction between two sidechains and the directional square-well interaction (hydrogen bonding) between backbone NH and C=O spheres. (The spheres are not drawn to scale for ease of viewing.) D-H are simulation snapshots of non-HB oligomer, and HB oligomer, 2, 3 and 4 β-sheet fibrils, respectively, generated by visual molecular dynamics software (VMD). The peptides in the five aggregates and in solution are shown in green and magenta, respectively.
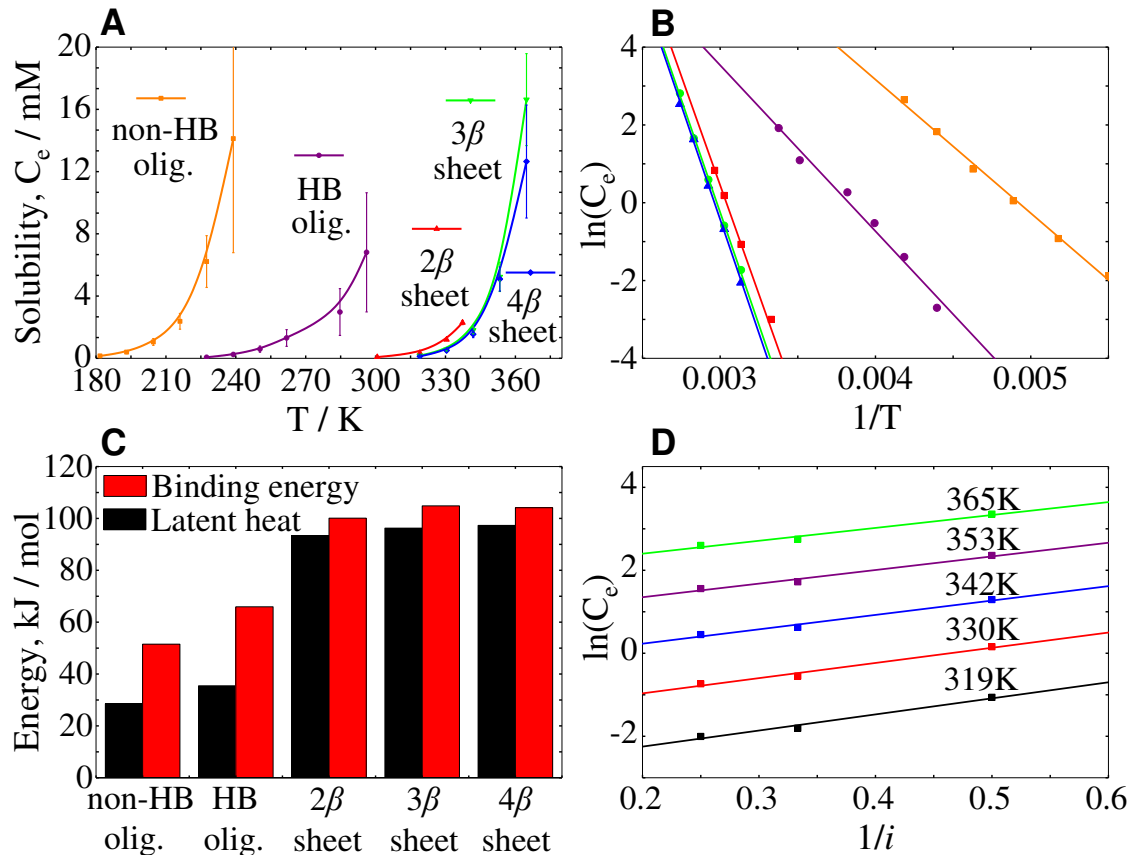
Figure 2. (A) A plot of the solubility C_e for non-HB oligomer, HB oligomer, 2, 3 and 4 β-sheet fibrils as a function of the temperature; (B) Solubility data for non-HB oligomer, HB oligomer, 2, 3 and 4 β-sheet fibrils at four to six different temperatures; lines are fitted to Eq. (1); the color scheme is the same as plot A; (C) Comparison of the average binding energy per peptide within the aggregate and the corresponding latent heat of peptide aggregation into oligomer and fibril.; (D) Dependence of the fibril solubility on thickness (i=2, 3 and 4) for 2, 3 and 4 β-sheet fibrils, respectively; lines are fitted to Eq. (2).
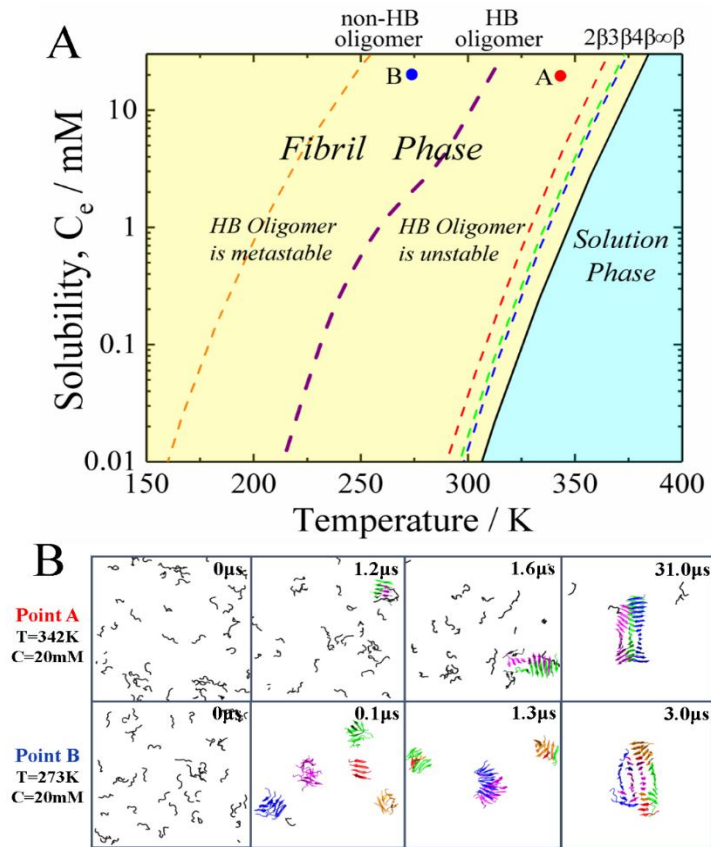
Figure 3. (A) Phase diagram for A$\beta_{16\text{-}22}$ peptide. Solubility (C$_e$)-vs-temperature (T) data are presented for non-HB oligomer (orange dashed line), HB oligomer (purple dashed line), 2β-sheet fibril (red dashed line), 3β-sheet fibril (green dashed line), 4β-sheet fibril (blue dashed line) and the infinite layer (∞) β-sheet fibril (black line). The fibril and solution phases are colored yellow and cyan, respectively. (B) DMD/PRIME20 simulation snapshots at the concentration and temperature of phase point A and B are taken at different time points, respectively.
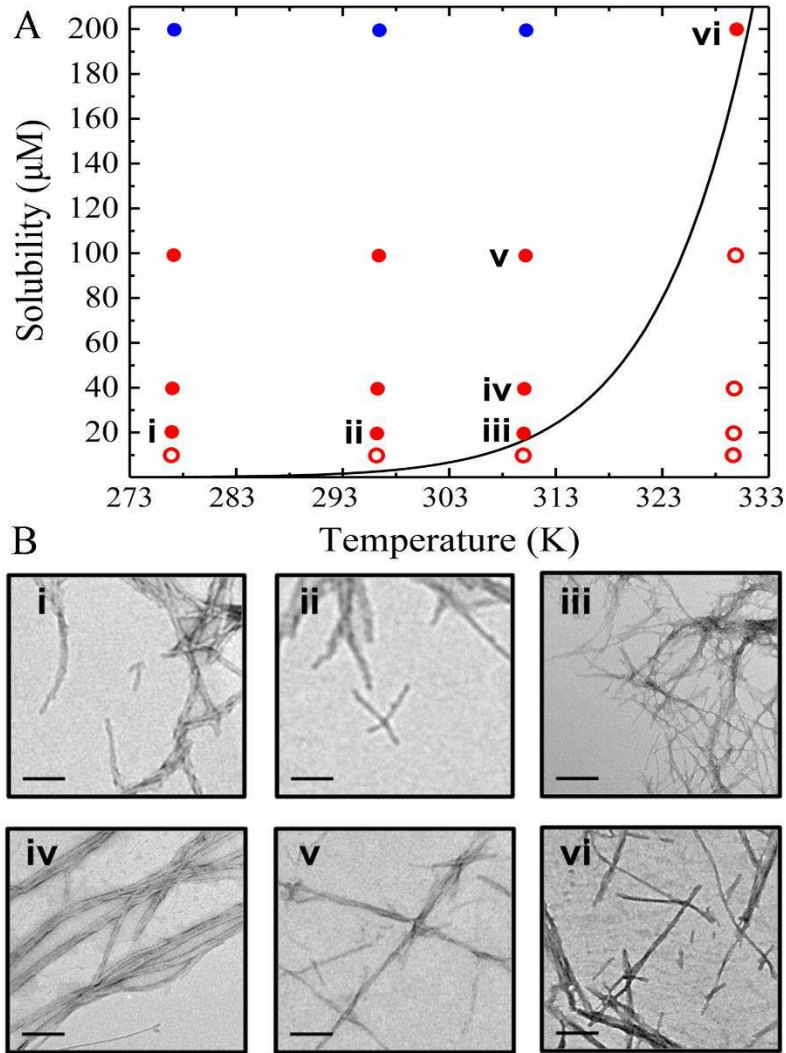
Figure 4. (A) Summary of both the simulation-predicted temperature-dependent solubility line C_e (T) for A$\beta_{16\text{-}22}$ peptide (black curve) and the fibrillation experiments performed under given conditions. Red dots and red circles indicate conditions at which fibrils have been found to form, and not to form, respectively, via TEM (at T= 277K, 296K, 310K and 330K). Blue dots indicate that fibrils have been reported in the literature to form under these conditions (see Fig. S5). (B) Six selected transmission electron micrographs (TEM) images (i-vi) showing that A$\beta_{16\text{-}22}$ form fibrils under the conditions that correspond to the six red dots labeled i-vi in (A). Buffer: 100 mM ammonium bicarbonate, pH=7, with a final concentration of 1% DMSO (v/v). Scale bar: 200 nm.