# Molecular determinants of A$\beta_{42}$ adsorption to amyloid fibril surfaces

Mathias M. J. Bellaiche[†,‡] and Robert B. Best[*,†]

†*Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892-0520*

‡*Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW*

E-mail: robertbe@nih.gov

## Simulation details

To construct our system, we placed an A$\beta_{1-42}$ fibril segment containing ten monomers in an orthorhombic cell of size $(L_x, L_y, L_z) = (14, 14, 4.87)$ nm whose $z$ dimension was chosen to match the length of ten layers of an infinite fibril. To this, we added a full-length monomer and solvated the box with TIP4P/2005 water,[1] 150 mM NaCl and excess neutralizing cations. The monomer was added in two different places to generate two initial conditions, and both systems were equilibrated in the isobaric-isothermal ensemble for 200 ns at 300 K using Gromacs 4.6.7.[2]

For the production runs, we used a Hamiltonian replica exchange scheme to facilitate several binding and unbinding events. The energy function of the $i^{\text{th}}$ window was:

$$U_i(r_p^{3N_p}, r_s^{3N_s}) = U_{\text{pp}}(r_p^{3N_p}) + U_{\text{ss}}(r_s^{3N_s}) + \lambda_i \varepsilon_{\text{ps}}(r_p^{3N_p}, r_s^{3N_s}) + U'_{\text{ps}}(r_p^{3N_p}, r_s^{3N_s}), \qquad (1)$$

where $r_p$ and $r_s$ refer respectively to protein and solvent atomic coordinates and $N_p$ and $N_s$ are the numbers of protein and solvent atoms, respectively. $U_{\text{pp}}$ and $U_{\text{ss}}$ are the protein-protein and solvent-solvent energy terms, $\varepsilon_{\text{ps}}$ is the Lennard-Jones interaction term between protein atoms and water oxygens, and $U'_{\text{ps}}$ contains all other non-bonded protein-solvent terms, i.e. Coulombic interactions. The parameter $\lambda$ is the scaling factor that sets the solvation energy strength, with $\lambda = 1.10$ the value corresponding to experimentally-consistent IDP behavior and the value used for all equilibration runs in the structure preparation stage.[3,4] The parameter space $\lambda > 1.10$ results in proteins being more solvated than realistic, whilst the region $\lambda < 1.10$ results in the converse.

We then ran long isobaric-isothermal (NPT) simulations using Gromacs version 4.6.7 with Plumed version 2.1, using 16 copies of the system with different $\lambda$ and Hamiltonian replica exchange swaps attempted every 1 ps between adjacent windows.[2,5] Two sets of such simu-

lations were run, starting from two different initial geometries (shown for the protein coordinates in S2). The first simulation set was 1 $\mu$s long, and the second was 0.9279 $\mu$s, for a total of 30.8 $\mu$s of chemical time simulated.

Forces were evaluated using the Amber03ws force field,[3] which has been shown to reproduce experimental observations of the A$\beta_{42}$ monomer.[4] A leap-frog integrator was used to propagate Langevin dynamics with force evaluations every 2 fs, constrained by a velocity-rescaling thermostat with a time constant of 0.1 ps. The pressures were kept constant using a semi-isotropic Parrinello-Rahman barostat at reference pressure 1 bar in both the $x/y$ and $z$ directions using a coupling time constant of 5 ps.[6,7] Nearest neighbor searches were of grid type, using a group cutoff scheme with an update frequency of 10 steps and a neighbor list cutoff of 1.2 nm. All bonds were constrained with the LINCS algorithm.[8] Non-bonded interactions were truncated at 1.2 nm for both van der Waals and Coulombic energies, with both energy and pressure dispersion corrections. Long-range electrostatics were treated with a fourth-order particle mesh Ewald sum of spacing 0.12 nm$^{-1}$ and relative tolerance of $1\times10^{-5}$.[9]

In all analyses, the first 100 ns were discarded as equilibration and frames in which the free peptide was within 4.5 Å of its periodic image were discarded to ignore non-physical configurations (corresponding to 6.8% or 2.8% of frames in each simulation). Analyses were performed with the help of the GrocasWrapper Python module (developed by Oliver Beckstein et al, doi: 10.5281/zenodo.17901).

## Calculation of free energies and thermodynamics

To estimate the free energy surfaces of adsorption, we built the potentials of mean force on $r_{\mathrm{min}}$ through a Boltzmann inversion of the probability density $p(r_{\mathrm{min}})$,

$$G(r_{\mathrm{min}}) = -RT \ln \left[ p(r_{\mathrm{min}}) \right], \tag{2}$$

for $R$ the universal gas constant and $T$ the temperature. Note that addition of an arbitrary constant to Eq. (2) does not affect the distribution function, so we set the lowest free energy to zero. Standard errors in the probability distribution were calculated using block averaging, with 10 blocks, and propagated to the free energy in quadrature.

To average the free energy surfaces from the two simulations, we used Boltzmann averaging:

$$\exp \left[ -\beta \langle G(r_{\mathrm{min}}) \rangle \right] = \sum_{i=1,2} \exp \left[ -\beta G_i(r_{\mathrm{min}}) \right], \tag{3}$$

where the index $i$ goes over both simulations. Errors from the individual free energies were propagated in quadrature.

To calculate the free energy change upon adsorption, we first calculated the dimensionless association constant $K_{\mathrm{bind}}$ from the probability density function $p(r_{\mathrm{min}})$, averaged over both simulations, as

$$K_{\mathrm{bind}} = \frac{1 - p_{\mathrm{bound}}}{p_{\mathrm{bound}}}, \tag{4}$$

for the total bound probability $p_{\mathrm{bound}} = \int_0^{r_{\mathrm{bound}}} dr_{\mathrm{min}} p(r_{\mathrm{min}})$ using a cutoff $r_{\mathrm{bound}} = 4.5$ Å. Then we could estimate the averaged standard binding free energies as:

$$\Delta_{\mathrm{bind}} G^{\ominus} = RT \ln \left[ K_{\mathrm{bind}} \frac{c_{\mathrm{sim}}}{c^{\ominus}} \right], \tag{5}$$

with the $c_{\mathrm{sim}}/c^{\ominus}$ factor correcting for the difference between the simulation monomer concentration (1.74 mM) and the experimental standard state (1 M). Errors were propagated in quadrature from the individual probability distribution functions.

To calculate the thermodynamics of adsorption, we considered the thermodynamic definition of the enthalpy change $\Delta H$ under isobaric and isothermal constraints:

$$\Delta H)_{p,T} = \Delta U)_{p,T} + p\Delta V)_{p,T}\,, \tag{6}$$

where the $\Delta$ refers to the difference between bound and unbound states, $U$ is the potential energy, $p$ is the pressure and $V$ is the volume. In these simulations, the volume fluctuations were negligible compared to the potential energy changes, so that the expansion work term in Eq. (6) could be ignored and $\Delta_{\mathrm{bind}}H \approx \Delta_{\mathrm{bind}}U$. Hence, by averaging the potential energies when the peptide was bound ($r_{\mathrm{min}} < 4.5$ Å) and unbound ($r_{\mathrm{min}} > 4.5$ Å) and subtracting, we could derive the enthalpy of binding. Standard errors were calculated using block averaging, with 10 blocks. The individual estimates of binding enthalpies were then averaged to yield an averaged adsorption enthalpy, with weights determined by estimate variance:

$$\Delta_{\mathrm{bind}}H = \frac{\sum_{i=1,2} w_i \Delta_{\mathrm{bind}}H_i}{\sum_{i=1,2} w_i}, \tag{7}$$

where $\Delta_{\mathrm{bind}}H_i$ and $\delta(\Delta_{\mathrm{bind}}H_i)$ are respectively the estimate of $\Delta_{\mathrm{bind}}H$ and its error from the $i$th simulation, and $w_i = \delta(\Delta_{\mathrm{bind}}H_i)^{-2}$. Individual errors were propagated to the averaged error via:

$$\delta\Delta_{\mathrm{bind}}H = \sqrt{\frac{1}{\sum_{i=1,2} w_i}}. \tag{8}$$

By knowing the total free energy change of adsorption and the enthalpic contribution, we

could calculate the entropy change upon binding as:

$$-T\Delta_{\text{bind}}S^{\ominus} = \Delta_{\text{bind}}G^{\ominus} - \Delta_{\text{bind}}H, \tag{9}$$

with errors propagated in quadrature.

To verify our estimate of the binding free energy, we estimated the range of binding free energies consistent with the coarse-grained simulations in[10] by calculating the binding constant $K_D$ as:

$$K_D^{-1} = 2\pi N_A L \int_{\sigma}^{1.5\sigma} dr\, r \exp(-\beta V(r)), \tag{10}$$

where $N_A$ is Avogadro's Constant, $L$ is a characteristic lengthscale of binding, $V(r)$ is the interaction potential and $\sigma$ is the particle width. In this work, the interaction potential was non-zero only over the range $r1.5\sigma$ and given by $V(r) = -\varepsilon(\sigma/r)^6$, for $\varepsilon$ the peptide-fibril contact energy. Using the value of $\sigma$ =2 nm and assuming a binding lengthscale equal to this, $L = \sigma$, then the published range of contact energies consistent with secondary nucleation $\varepsilon \in (6, 16)$ $\text{k}_B\text{T}$ yielded values of standard binding free energies $\Delta_{\text{bind}}G^{\ominus} = RT\ln(K_D/c^{\ominus})$ of $\Delta_{\text{bind}}G^{\ominus} \in (-37, -15)$ kJ.mol$^{-1}$, bracketing our estimate of $\Delta_{\text{bind}}G^{\ominus} = -19 \pm 2$ kJ.mol$^{-1}$. The maximum in catalytic efficiency occurred for a value of $\varepsilon = 8$ $\text{k}_B\text{T}$, corresponding to a standard binding free energy of $\Delta_{\text{bind}}G^{\ominus} = -19$ kJ.mol$^{-1}$.[10]

## Intermolecular contact analysis

To investigate the nature of peptide-fibril interactions, we generated distance matrices between all protein non-hydrogen atoms using the Gromacs `g_mdmat` tool, and converted to a system contact probability matrix $\mathcal{C}$, whose elements $\mathcal{C}_{i,j}$ measure the fraction of frames

in which residues $i$ and $j$ are within 4.5 Å of each other, provided that some fibril atom is within 4.5 Å of some peptide atom (i.e. the two species are in contact). This system-level contact map was transformed to a reduced $42 \times 42$ contact map $\mathcal{C}'$ whose elements $\mathcal{C}'_{i,j}$ measure the fraction of frames in which any fibril residue $i$ is within 4.5 Å of free peptide residue $j$, which was then integrated along rows and columns respectively to yield one-dimensional arrays that measure the fibril and peptide contact hotspots. This was done for each individual simulation, with the results averaged and standard errors calculated to generate the curves shown in Fig. 3(a) of the main text.

To obtain a two-dimensional view of the binding process by identifying specific residue pairs responsible for contact, we used a machine learning algorithm to extract the dominant binding modes observed in our simulations. We did this by performing a $k$-means clustering analysis on all the matrices $\mathcal{C}'(t)$ (from both simulations). We created instantaneous interaction matrices from the distance matrices $\mathcal{D}$ through the continuous transformation Eq. (11) to the distances, for $\beta = 50$ nm$^{-1}$ and $r_0 = 4.5$ Å:

$$\mathcal{C}_{ij}(t) = \frac{1}{1 + \exp(\beta(\mathcal{D}_{ij}(t) - r_0))}, \tag{11}$$

Clustering was done using $k = 4$ clusters, as this was the best choice as determined by our validation method described below. The clustering algorithm was the `vq.kmeans` function of the `scipy.cluster` Python module. The results of the clustering are shown in Fig. 3(c) of the main text, with the fourth minor cluster corresponding to a region of the fibril surface that would be sequestered in inter-fibril binding to generate the two-fold symmetric surface, and hence not biophysically relevant.

The validation of our choice of $k$ follows the method described in,[4] which balances the need to describe the data adequately against the risk of overfitting. This is done by generating

the metric $\Delta$, defined as:

$$\Delta = ||\langle \mathcal{C} \rangle_t, \langle \mathcal{C} \rangle_k||, \tag{12}$$

with $||\cdots||$ the $L_2$ norm, $\langle \mathcal{C} \rangle_t$ the contact map averaged over all simulation frames and $\langle \mathcal{C} \rangle_k$ the contact map averaged over just the frames closest to the cluster centers. This metric, then, measures the ability of the clusters to describe the ensemble data. We fit the data to several values of $k$, finding $\Delta$ at each iteration, and determined the point at which adding a cluster did not improve the fit (and hence entered the regime of overfitting). This was determined as the point at which the fractional decrease in $\Delta$ relative to using just two clusters was no longer above 10%, which as shown in S5 occurs for $k = 4$.

## Monomer configurational analysis

Intramolecular contact maps for the monomer were generated by building time-dependent distance matrices of the peptide residues against themselves and converting each of these to contact probability matrices by applying Eq. (11), with $\beta = 50$ nm$^{-1}$ and $r_0 = 4.5$ Å. These contact probability maps were then averaged within the bound ($r_{\min} < 4.5$ Å) and free sets ($r_{\min} > 4.5$ Å), and subtracted to generate the average free and difference maps shown in Fig. 4 of the main text. The individual simulation contact maps were averaged, weighted according to the fraction of time each simulation spent either as bound or free.

The $\beta$-propensity of the monomer was calculated as:

$$\langle \Delta p(\beta) \rangle = \frac{1}{42} \sum_{i=1}^{42} \Delta p(\beta)_i, \tag{13}$$

with $\Delta p(\beta)_i$ the difference in the fraction of frames in which the DSSP algorithm[11] classified residue $i$ as $\beta$-sheet when the peptide was bound to or free from the fibril. This analysis gave values of 2.0% residue$^{-1}$ and 3.2% residue$^{-1}$ for each individual simulation.

## Rational mutation design

To leverage the atomistic information we gleaned from our simulations, we designed several peptide constructs which we believe will exhibit perturbed rates of secondary nucleation. Before describing how we design mutations, it is helpful first to define two energies $\sigma$ and $\varepsilon$:

$$\sigma(s_{\text{fib}}) = \sum_{i=15}^{42} \mathcal{B}_{s_{\text{fib}}(i),s_{\text{fib}}(i)} \tag{14}$$

and

$$\varepsilon(s_{\text{fib}}, s_{\text{pep}}, \mathcal{C}') = \sum_{i=1}^{42}\sum_{j=1}^{42} \mathcal{V}_{s_{\text{fib}}(i),s_{\text{pep}}(j)}\mathcal{C}'_{ij}, \tag{15}$$

where $i$ and $j$ are residue indices, $s_{\text{species}}(i)$ is the residue type of the $i$th element of the primary structure of A$\beta_{1-42}$ that makes up either the fibril (fib) or peptide (pep) species, $\mathcal{C}'$ is the reduced fibril-peptide interaction contact map described previously, $\mathcal{B}$ is the Pasta potential matrix (a knowledge-based potential describing interaction strengths between residue types in amyloid structures)[12] and $\mathcal{V}$ is the MJ statistical potential matrix.[13] With such definitions, then, $\sigma$ captures the stability of a fibril made of the A$\beta_{1-42}$ assuming parallel, in-register structure where each residue mostly interacts with itself in the next amyloid chain counting only the fibril core from the NMR structure. In contrast, $\varepsilon$ is a measure of the interaction energy between the fibril and the peptide, as captured in the concatenated reduced contact map $\mathcal{C}'$.

To predict rational point mutations, we first identified each fibril residue that appeared in some contact in $\mathcal{C}'$ with probability greater than a threshold (we took $p_{\text{min}} = 0.05$). For each of these important fibril residues $j$, then, we systematically changed its residue type $s(j)$ to each of the other 19 amino acids, and calculated the effect of this mutation on both the fibril stability and the interaction energy (taking both the cases that the fibril is mutated but the free peptide is wild type, as well as both species mutated), assuming the same interaction

contact map $\mathcal{C}'$. This algorithm then generated a collection of point mutants of the wild-type peptide $s_0 \rightarrow \{s'_k\}$, with each mutated sequence $s'_k$ having the following associated stability and interaction energy changes:

$$\Delta\sigma_k = \sigma(s'_k) - \sigma(s_0) \tag{16}$$

$$\Delta\varepsilon_k^{\mathrm{M/WT}} = \varepsilon(s'_k, s_0, \mathcal{C}') - \varepsilon(s_0, s_0, \mathcal{C}') \tag{17}$$

$$\Delta\varepsilon_k^{\mathrm{M/M}} = \varepsilon(s'_k, s'_k, \mathcal{C}') - \varepsilon(s_0, s_0, \mathcal{C}'), \tag{18}$$

where the superscript in the $\varepsilon$ terms indicate whether a mutant fibril is interacting with a wild-type (WT) or mutant (M) monomer. Finally, for all generated mutants, we imposed different criteria on $\Delta\sigma$ and $\Delta\varepsilon$, described below, to create three lists of rational mutants which we predict will have experimentally observable effects on the fibril-peptide adsorption process and, hence, the secondary nucleation rate.

As described in the type key S1, we partitioned the identified mutants into three mutually exclusive classes: Type I (S3) are those that are predicted not to affect the fibril stability while affecting the adsorption energy; Type II (S4) are those that stabilize the fibril and affect the adsorption energy; and Type III (S5) are those that affect the adsorption energy (with the possibility of severely destabilizing the fibril). In the tables, $\Delta\sigma$ is the effect of the mutation on the fibril stability, $\Delta\varepsilon^{\mathrm{M/WT}}$ the effect on the interaction energy between a mutant fibril and a wild type peptide and $\Delta\varepsilon^{\mathrm{M/M}}$ the effect on the interaction energy between a mutant fibril and a mutant peptide, all in units of $k_BT$. S2 summarizes the meaning of the symbols in each of the mutant tables in terms of , the effect on fibril stability or adsorption energy. An 'x' entry denotes that the mutated residue lay in the disordered region of the fibril, and hence would not be expected to affect fibril stability.

## Reweighting to different ionic strengths

Since it would have been prohibitive to run additional simulations at other ionic strengths than the 150 mM used in our work, we reweighted the existing simulation data by using the difference between the fibril-peptide Debye-Hückel electrostatic energy at 150 mM salt and other ionic strengths from 20–300 mM. Specifically, the Debye-Hückel energy of coordinates $\mathbf{R}$ at ionic strength $I$ is given by:

$$E_{\mathrm{DH}}(\mathbf{R}, I) = \sum_{i \in \mathrm{fibril}} \sum_{j \in \mathrm{peptide}} \frac{q_i q_j}{4\pi\epsilon_0\epsilon_r r_{ij}} \exp[-r_{ij}/\lambda_D(I)] \tag{19}$$

where $q_i$ and $q_j$ are two atomic partial charges on the fibril and peptide respectively, $r_ij$ the distance between them, and the Debye length $\lambda_D$ is given by $\lambda_D(I) = ((\epsilon_0\epsilon_r k_{\mathrm{B}}T)/(2 \times 10^3 N_A e^2 I))^{1/2}$, and $k_{\mathrm{B}}$ is the Boltzmann constant, $T$ the temperature, $\epsilon_0$ the permittivity of free space, $\epsilon_r$ the dielectric constant, $N_A$ Avogadro's number, $e$ the elementary charge and $I$ the ionic strength in molar units. We used the intermolecular fibril-peptide energy rather than the total electrostatic energy to reduce noise from the much larger fibril electrostatic energy. The weight of each frame in computing averages was thus $w \propto \exp[-\beta(E_{\mathrm{DH}}(I) - E_{\mathrm{DH}}(150\mathrm{mM}))]$, where $\beta = (k_{\mathrm{B}}T)^{-1}$

# Supporting Figures

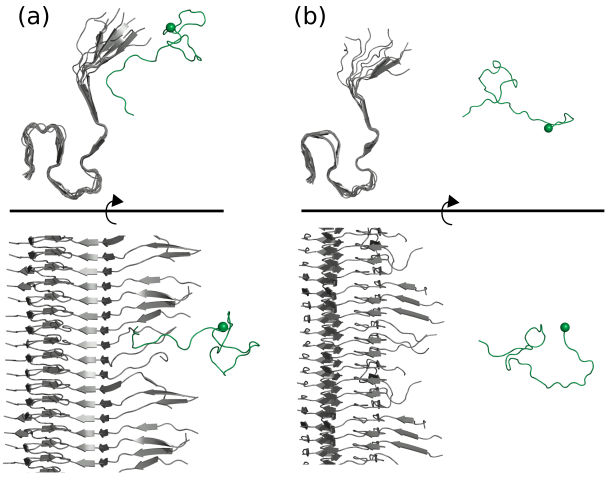


Figure S1: Starting configurations of protein atoms in both simulations (a) and (b), as (top) view down fibril axis and (bottom) side view. The free monomer is shown in green, with its N-terminus highlighted as a sphere, and the fibril in gray. Water molecules and ions not shown for clarity. Figures made with PyMOL.

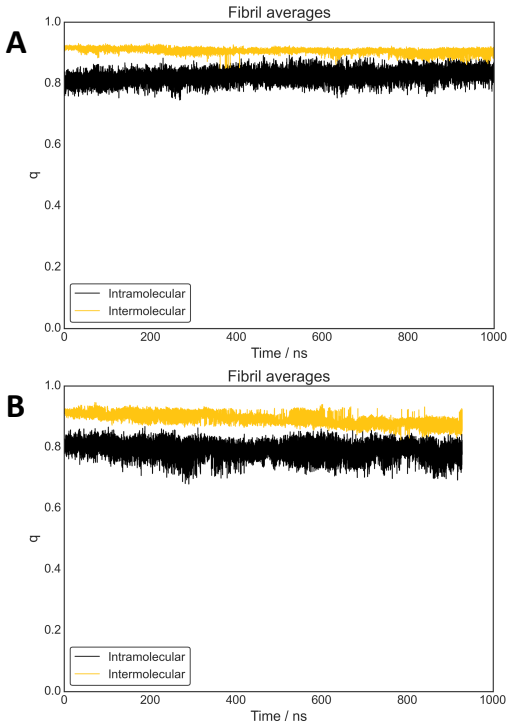


Figure S2: Average inter- and intramolecular contacts between ordered regions (residues 14-42) of peptides which are part of the fibril. Contacts are any pairs of heavy atoms within 0.45 nm in the experimental structure. Results are shown for simulations 1 and 2 in (A) and (B) respectively.

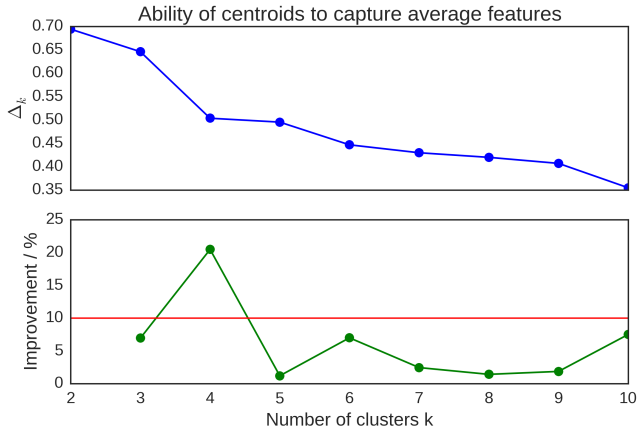


Figure S3: Ability of $k$-means clustering algorithm to describe binding data. (Top) Distance metric $\Delta$ and (bottom) relative improvement upon adding another cluster. Red line indicates the 10% threshold used to avoid overfitting.

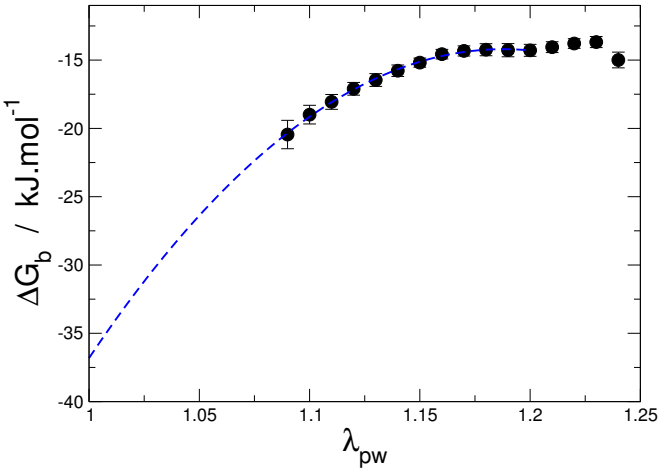


Figure S4: Extrapolation of binding free energy to $\lambda = 1.0$, corresponding to the Amber ff03w force field.[14,15] Blue curve shows quadratic fit to the data to guide the eye.

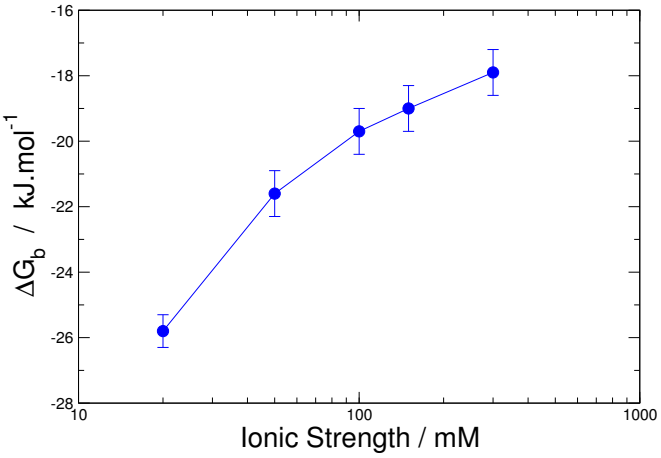


Figure S5: Esimate of the effect of ionic strength on binding free energy using a Debye-Hückel approximation to the difference in peptide-fibril electrostatic energy between the simulated (150 mM) and target ionic strengths.

# Supporting Tables

**Table S1: Mutation type key.**

| Mutation type | Effect on fibril stability | Effect on adsorption energy |
|---|---|---|
| Type I | $\lvert\Delta\sigma\rvert < 1\,k_BT$ | $\lvert\Delta\varepsilon\rvert > 1\,k_BT$ |
| Type II | $\Delta\sigma < 1\,k_BT$ | $\lvert\Delta\varepsilon\rvert > 1\,k_BT$ |
| Type III | – | $\lvert\Delta\varepsilon\rvert > 1\,k_BT$ |

**Table S2: Symbol key.**

| Symbol | $\Delta E$ range / $k_BT$ | Symbol | $\Delta E$ range / $k_BT$ |
|---|---|---|---|
| - | (-1,0) | + | (0,1) |
| -- | (-2,-1) | ++ | (1,2) |
| --- | (-3,-2) | +++ | (2,3) |
| ---- | $< -3$ | ++++ | >3 |

**Table S3: Type I mutations.**

| Mutant | $\Delta\sigma$ | $\Delta\varepsilon^{\mathrm{M/WT}}$ | $\Delta\varepsilon^{\mathrm{M/M}}$ |
|---|---|---|---|
| E11F | x | - | -- |
| E11L | x | - | -- |
| H13F | x | -- | --- |
| H13I | x | -- | -- |
| H13K | x | ++ | ++ |
| H13L | x | -- | --- |
| H13M | x | -- | -- |
| H13V | x | -- | -- |
| H13W | x | -- | -- |
| Q15A | - | -- | -- |
| Q15K | + | ++ | ++ |
| K16R | + | -- | -- |
| S26C | - | - | -- |
| S26M | - | -- | -- |
| S26W | - | -- | -- |
| S26Y | - | - | -- |

**Table S4: Type II mutations.**

| Mutant | $\Delta\sigma$ | $\Delta\varepsilon^{M/WT}$ | $\Delta\varepsilon^{M/M}$ | Mutant | $\Delta\sigma$ | $\Delta\varepsilon^{M/WT}$ | $\Delta\varepsilon^{M/M}$ |
|---|---|---|---|---|---|---|---|
| Q15C | -- | --- | ---- | D23F | ---- | - | -- |
| Q15F | --- | ---- | ---- | D23I | ---- | - | -- |
| Q15H | -- | -- | -- | D23L | --- | - | -- |
| Q15I | ---- | ---- | ---- | D23M | -- | - | -- |
| Q15L | -- | ---- | ---- | D23V | ---- | - | -- |
| Q15M | -- | ---- | ---- | D23W | --- | - | -- |
| Q15V | ---- | ---- | ---- | D23Y | -- | - | -- |
| Q15W | -- | ---- | ---- | S26F | -- | -- | --- |
| Q15Y | -- | ---- | ---- | S26I | --- | -- | -- |
| K16C | -- | --- | --- | S26L | -- | -- | --- |
| K16F | --- | ---- | ---- | S26V | --- | -- | -- |
| K16H | -- | -- | -- | K28F | --- | - | -- |
| K16I | ---- | ---- | ---- | K28I | ---- | - | -- |
| K16L | --- | ---- | ---- | K28L | --- | - | -- |
| K16M | -- | --- | ---- | K28M | -- | - | -- |
| K16V | ---- | --- | ---- | K28V | ---- | - | -- |
| K16W | -- | ---- | ---- | K28W | -- | - | -- |
| K16Y | -- | --- | ---- | K28Y | -- | - | -- |
| E22F | --- | - | -- | G37F | --- | -- | -- |
| E22I | ---- | - | -- | G37I | ---- | - | -- |
| E22L | --- | - | -- | G37L | --- | -- | -- |
| E22M | -- | - | -- | | | | |
| E22V | ---- | - | -- | | | | |
| E22W | --- | - | -- | | | | |
| E22Y | -- | - | -- | | | | |

# References


(1) Abascal, J. L.; Vega, C. A general purpose model for the condensed phases of water: TIP4P/2005. *J. Chem. Phys.* **2005**, *123*, 234505.

(2) Hess, B.; Kutzner, C.; Van Der Spoel, D.; Lindahl, E. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.

(3) Best, R. B.; Zheng, W.; Mittal, J. Balanced protein–water interactions improve prop-

**Table S5: Type III mutations.**

| **Mutant** | $\Delta\sigma$ | $\Delta\varepsilon^{\mathrm{M/WT}}$ | $\Delta\varepsilon^{\mathrm{M/M}}$ |
|---|---|---|---|
| V36A | +++ | + | ++ |
| V36D | ++++ | ++ | ++ |
| V36E | ++++ | ++ | ++ |
| V36G | ++++ | ++ | ++ |
| V36H | ++ | + | ++ |
| V36K | ++++ | ++ | +++ |
| V36N | ++ | ++ | ++ |
| V36P | ++++ | ++ | ++ |
| V36Q | ++++ | ++ | ++ |
| V36R | ++++ | ++ | ++ |
| V36S | +++ | ++ | ++ |
| V36T | +++ | ++ | ++ |

erties of disordered proteins and non-specific protein association. *J. Chem. Theory Comput.* **2014**, *10*, 5113–5124.

(4) Meng, F.; Bellaiche, M. M.; Kim, J.-Y.; Zerze, G. H.; Best, R. B.; Chung, H. S. Highly Disordered Amyloid-$\beta$ Monomer Probed by Single-Molecule FRET and MD Simulation. *Biophys. J.* **2018**, *4*, 870–884.

(5) Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R. A. et al. PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Comput. Phys. Commun.* **2009**, *180*, 1961–1972.

(6) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182–7190.

(7) Nosé, S.; Klein, M. Constant pressure molecular dynamics for molecular systems. *Mol. Phys.* **1983**, *50*, 1055–1076.

(8) others,, et al. LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997**, *18*, 1463–1472.

(9) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577–8593.

(10) Šarić, A.; Buell, A. K.; Meisl, G.; Michaels, T. C.; Dobson, C. M.; Linse, S.; Knowles, T. P.; Frenkel, D. Physical determinants of the self-replication of protein fibrils. *Nat. Phys.* **2016**, *12*, 874.

(11) Kabsch, W.; Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983**, *22*, 2577–2637.

(12) Walsh, I.; Seno, F.; Tosatto, S. C.; Trovato, A. PASTA 2.0: an improved server for protein aggregation prediction. *Nucleic Acids Res.* **2014**, *42*, W301–W307.

(13) Miyazawa, S.; Jernigan, R. L. Estimation of effective interresidue contact energies from protein crystal structures: quasi-chemical approximation. *Macromolecules* **1985**, *18*, 534–552.

(14) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T. et al. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum chemical calculations. *J. Comp. Chem.* **2003**, *24*, 1999–2012.

(15) Best, R. B.; Mittal, J. Protein simulations with an optimized water model: cooperative helix formation and temperature-induced unfolded state collapse. *J. Phys. Chem. B* **2010**, *114*, 14916–14923.