Computational Design of Amyloid Self-Assembling Peptides Bearing Aromatic Residues and the Cell Adhesive Motif Arg-Gly-Asp

SUPPLEMENTARY INFORMATION

Sai Vamshi R. Jonnalagadda[1,*], Eirini Ornithopoulou[2,3,*], Asuka A. Orr[1], Estelle Mossou[4,5], V. Trevor Forsyth[4,5], Edward P. Mitchell[5,6], Matthew W. Bowler[7,8] , Anna Mitraki[2,3,‡], and Phanourios Tamamis[1,‡]

[1] Artie McFerrin Department of Chemical Engineering, Texas A&M University, College Station, Texas 77843-3122, United States of America

[2] Department of Materials Science and Technology, University of Crete, Voutes Campus, GR-70013 Heraklion, Crete, Greece

[3] Institute for Electronic Structure and Laser, FORTH, N. Plastira 100, GR 70013, Heraklion, Greece

[4] Institut Laue Langevin, 71 Avenue des Martyrs, 38042 Grenoble Cedex 9, France

[5] Faculty of Natural Sciences / Institute for Science and Technology in Medicine, Keele University, Staffordshire ST5 5BG, United Kingdom

[6] European Synchrotron Radiation Facility, 6 rue Jules Horowitz, 38043 Grenoble Cedex 9, France

[7] European Molecular Biology Laboratory, Grenoble Outstation, 71 avenue des Martyrs, CS 90181 F-38042 Grenoble, France

[8] Unit for Virus Host Cell Interactions, Univ. Grenoble Alpes-EMBL-CNRS, 71 avenue des Martyrs, CS 90181 F-38042 Grenoble, France

Supplementary Information Methods

Computational Methods

Computational Design – Modification at Position 11.

We used the computational design model, presented in *Eq. S1*, to introduce mutations at residue position 11 of two sets of 4-stranded highly ordered and well-aligned β-sheet fibrils, hereinafter referred to as designable scaffolds of self-assembling peptide with sequence RGDSGAITIGC (previously investigated by us; ref. 47 in the main text). Of the two sets, one set comprised of highly ordered and well-aligned 4-stranded parallel β-sheet fibrils and the other set comprised of highly ordered and well-aligned 4-stranded antiparallel β-sheet fibrils (ref. 47 in the main text). For both sets, the specific highly ordered and well-aligned β-sheet fibrils which were used as designable scaffolds were extracted from the global free energy minima of free energy landscapes constructed using polar and nematic order parameters from the replica exchange molecular dynamics self-assembly simulations of peptide RGDSGAITIGC (ref. 47 in the main text). As the dominant configuration of amyloid forming peptide RGDSGAITIGC corresponds to an antiparallel β-sheet configuration, the parallel β-sheet configurations formed by RGDSGAITIGC were not presented in-detail in ref. 47 in the main text; yet, they were analyzed in-depth and were considered for the purpose of the present study. The extracted β-sheet states were selected as designable scaffolds because of their high degree of order and alignment which suggests that they most likely correspond to the naturally occurring β-sheet structures in the amyloid fibrils. We refined our sets of highly ordered and well-aligned β-sheet states by discarding certain β-sheet states at which an artificial bent was present in one of the two outer peptides which occurred owing the limited number of peptides used in the simulations (ref. 47 in the main text).

$$\min_{y_i^j, y_k^l, w(s)_{ik}^{jl}} \left(\frac{1}{f} \left(\sum_{a=1}^{p} \sum_{b=1}^{p} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \sum_{\substack{k=1 \\ k \ge i}}^{n} \sum_{l=1}^{m_k} \sum_{s=1}^{f} E_{iakb}^{jl} y_i^j y_k^l w(s)_{ik}^{jl} \right) + \frac{1}{f} \left(\sum_{a=1}^{p} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \sum_{s=1}^{f} \gamma(SASA(s)_{ia}^j) y_i^j \right) \right) \qquad \textit{Eq. S1}$$

Subject to
$$\sum_{j=1}^{m_i} y_i^j = 1 \ \forall \ i$$

$$w(s)_{ik}^{jl} \, d(s)_{iakb}^{j,l} \le cutoff$$

$$d(s)_{iakb}^{jl} \ge \left(1 - w(s)_{ik}^{jl}\right) cutoff$$

$$k \ne i \, \forall \, a = b$$

$$b > a \forall k = i$$

$$l = j \forall (k \in \{\alpha_1, \alpha_2, ..., \alpha_v\}, k = i, a \ne b)$$

$$i \in \{\alpha_1, \alpha_2, ..., \alpha_v\}$$

$$w(s)_{ik}^{jl} \in \{0,1\} \; \forall \; i, j, k, l$$

$$y_i^j, y_k^l \in \{0,1\} \; \forall \; i, j, k, l$$

The aforementioned model simultaneously minimizes the pairwise interaction energy, E_{iakb}^{jl}, accounting for protein-protein interactions, and the solvent accessible surface area, $SASA(s)_{ia}^j$, multiplied by the surface tension coefficient, γ, implicitly accounting for protein-water interactions. The pairwise interaction energy, E_{iakb}^{jl}, corresponds to the interaction energy between residue j at mutable position i in strand a and residue l at position k in strand b (taken from the SIPPER force field; ref. 72 in the main text). The positions i and k can belong to the same peptide strand (intra-molecular interactions) or different peptide strands (inter-molecular interactions). In this study, the total number of residue positions, n, is 11 per strand and the total number of peptide strands, p, is 4. $\{\alpha_1, \alpha_2, ..., \alpha_v\}$ is a set of residue positions that are amenable for modification in the peptide under the condition that $\alpha_1 < \alpha_2 < ... < \alpha_v$. In the present study, position 11 is mutable, i.e. $\{\alpha_1 = 11\}$. The pairwise interaction energy of each designable scaffold, s, was summed and averaged over the total number of designable scaffolds, f. In the present study, the number of designable scaffolds used, f, was equal to 50 for both parallel and antiparallel configurations, independently. The computational design model was solved independently for both antiparallel and parallel configurations of designable scaffolds; thus, there are two change in energy values from the objective function (*Eq. S1*) per amino acid substitution. The binary variable $y_i^{\,j}$ equals one if position i is occupied by amino acid j, and zero otherwise.

The binary variable y_k^l equals one if position k is occupied by amino acid l, and zero otherwise. The binary variable $w(s)_{ik}^{jl}$ equals one if, $d(s)_{iakb}^{jl}$, the distance between side chain geometric centers of a residue j at position i in strand a and residue l at position k in strand b is less than a specific distance, *cutoff*, and zero otherwise. The specific distance, *cutoff*, in our case is taken to be 6.5 Å, as this distance proved optimal to account interactions of the mutated residues with its neighboring counterpart residues. The specific distance (6.5 Å between side chain geometric centers) matches well the distance cutoff of 5.0 Å between any non-hydrogen atom of two interacting residues based on which the SIPPER force field was trained on. The constraint, $l = j \forall (k \in \{\alpha_1, \alpha_2, ..., \alpha_v\}; k = i; a \neq b)$, defined under *Eq. S1* is introduced in order to ensure that mutations occurring at a position in a peptide strand simultaneously occur at that position in each of the peptide strands of the self-assembled structures.

The second term of *Eq. S1*, approximately estimates implicitly the non-polar solvation free energy change upon a mutation in the elementary structural unit of the designed fibrils. The non-polar solvation free energy is required for (1) the formation of a cavity in the solvent to accommodate the solute and (2) establishing solute-solvent dispersion interactions. The formation of the cavity requires entropic and solvent-reorganization energy, and due to the strong self-attraction of water stemming from its ability to form hydrogen bond networks with itself, energy is also lost due to the introduction of hydrophobic molecules, which cannot form hydrogen bonds and disrupt the hydrogen bond network[1,2,3]. Thus, the second term may also be considered as a penalty for the introduction of larger hydrophobic residues into the designable scaffold, which would result in the formation of a larger cavity and the disruption of more solvent-solvent interactions. To account for solute-solvent interactions, we additionally used a coarse-grained approach to approximate the contribution of each residue to the solvation free energy by assuming a linear dependence between the solvation free energy and the solvent accessible surface area (SASA)[4,5,6,7]. Under this assumption, the cavity formed to accommodate a solute is proportional to the SASA of the introduced solute, and the solute-solvent dispersion interaction energy also correlates with SASA as solute atoms at the solute-solvent interface interact more strongly with the solvent compared to buried solute atoms; in the present study. The $SASA(s)_{ia}^j$ term is the estimated solvent accessible surface area of the introduced amino acid j

at position i in strand a. The surface tension coefficient, γ used in *Eq. S1* for this study is equal to 0.002 kcal/(mol·Å^2). The $SASA(s)_{ia}^j$ term is approximated through the following equation:

$$SASA(s)_{ia}^j = (theoretical\ SASA\ of\ the\ introduced\ residue) - \\ [(theoretical\ SASA\ of\ the\ native\ residue) - (SASA\ of\ the\ native\ residue\ in\ scaffold(s))]$$

Eq. S2

where the second term, [(theoretical SASA of the native residue) – (SASA of the native residue in the scaffold(s))], represents the solvent-excluded surface, or how "buried" the residue is. The theoretical SASA values are ASA values of the whole residue (X) in the tripeptide, Ala-X-Ala, taken from Table 2 of reference[8] and the structural SASA values are calculated using the GEPOL[9] algorithm in WORDOM (ref. 85 in the main text) for the whole residue in the designable scaffold. According to our results, the second term introduced in *Eq. S1* to approximately estimate implicitly the non-polar solvation free energy change upon a mutation in the elementary structural unit of the designed fibrils, primarily affects the order of magnitude of the total energy term under minimization and to a lesser extent the relative ranking of the designed-mutated peptides. Additional investigation of the term's role in the computational protein design in general, and more specifically in the computational design of amyloid biomaterials, as well as the appropriate value of the surface tension coefficient, γ, will be performed in future studies.

The model presented *Eq. S1* was solved exhaustively and the mutations associated with a favorable (< 0) energy (first term under minimization in *Eq. S1*) are presented rank ordered in Figure S1. As the model was solved independently for parallel and antiparallel β-sheets, two independent columns are provided per amino acid mutation at position 11.

Upon the solution of the model, we introduced an additional constraint (*Eq. S3*) during the solution of the computational design problem to investigate the feasibility of crosslinking between adjacent tyrosine residues upon fibril formation, in the parallel β-sheet designable scaffolds of peptide RGDSGAITIGY.

$$\left(\sum_{a=1}^{p} \sum_{b=1}^{p} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \sum_{k=1}^{n} \sum_{l=1}^{m_k} \sum_{s=1}^{f} \left(D(s)_{iakb}^{jl} \leq 10.65 \, \mathring{A} \right) \right) \forall \, (i = 11, k = 11, b = a + 1)$$

Eq. S3
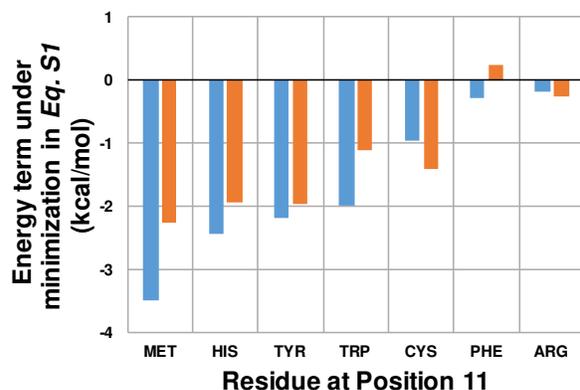
Figure S1. The energy term in the objective function under minimization in *Eq. S1*, in kcal/mol, (Y-axis) with respect to the mutated residue introduced at position 11 (X-axis) in either parallel or antiparallel β-sheets). Blue bars indicate the value of the energy term associated with mutations introduced to the parallel configuration of the designable scaffolds. Red bars indicate the value of the energy term associated with mutations introduced to the antiparallel configuration of the designable scaffolds. Mutations are rank ordered according to the energy value (first term under minimization in *Eq. S1*) for the parallel designable scaffolds. Only mutations associated with a favorable (< 0) energy (first term under minimization in *Eq. S1*) are presented.
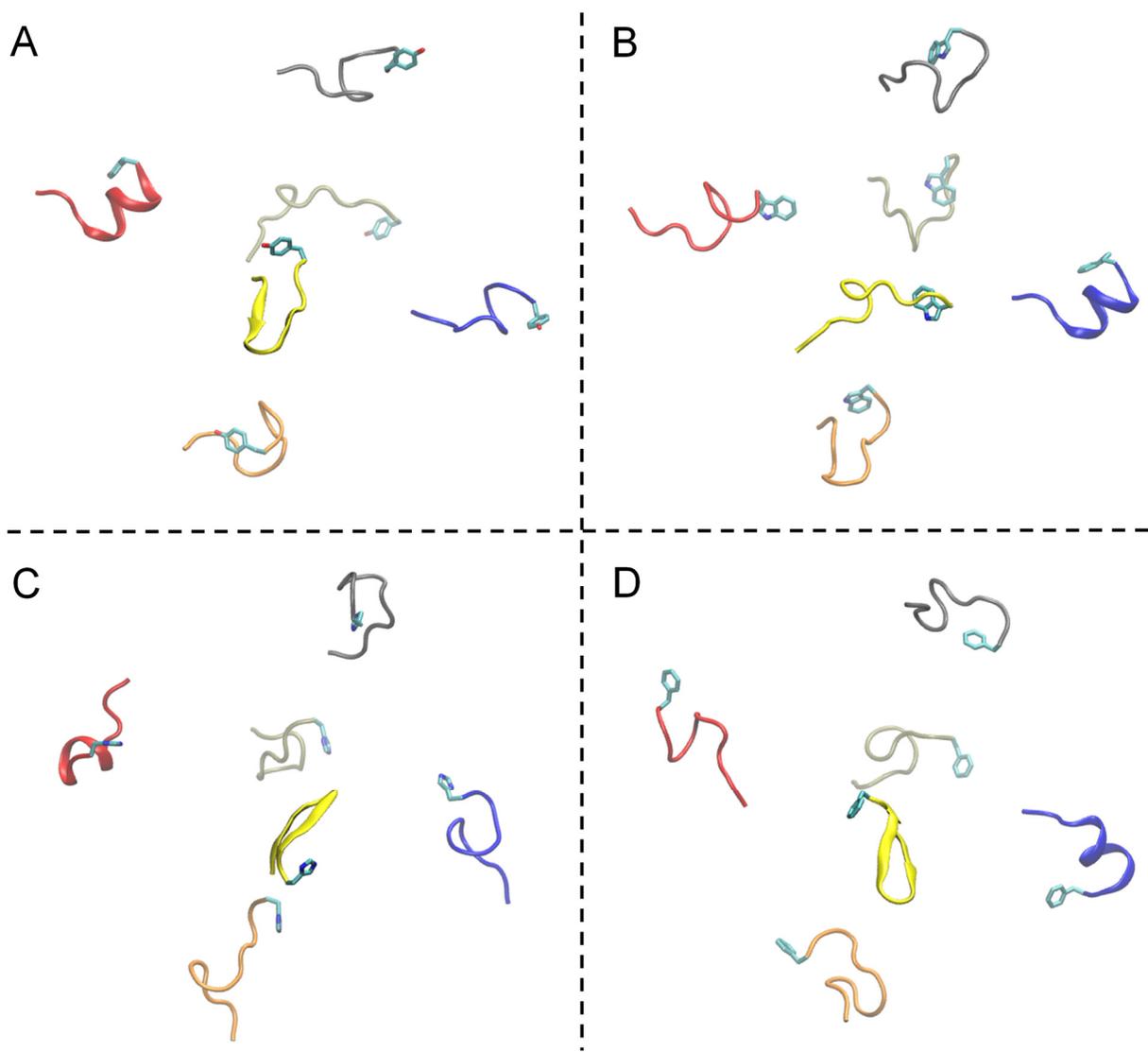
Figure S2. The initially modeled structures showing six copies of the designed peptides RGDSGAITIGY, RGDSGAITIGW, RGDSGAITIGH, RGDSGAITIGF, prior to the execution of the REMD simulations, is presented in panels A, B, C and D, respectively. The panels are divided using black dashed lines for clarity. In each panel, the six copies of peptides are shown in different color in cartoon representation.
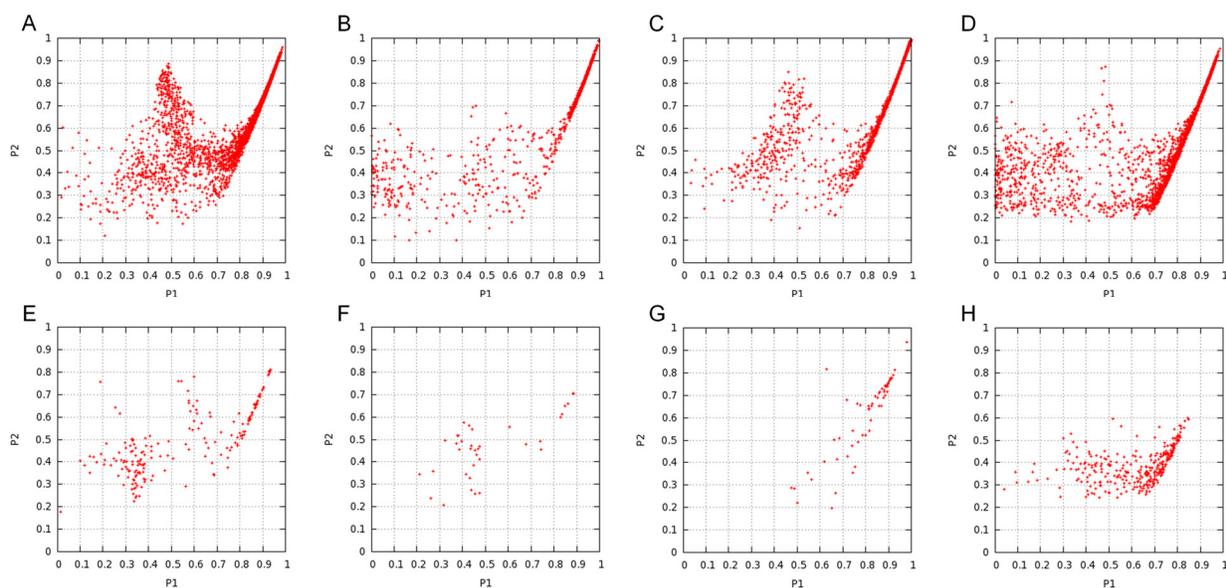
Figure S3. Plots of P_2 (Y-axis) as a function of P_1 (X-axis) for the 4- and 5- stranded parallel β-sheet conformations observed in the REMD simulations at 300 K. Figures A and E correspond to 4- and 5- stranded parallel configurations of peptide RGDSGAITIGY, respectively. Figures B and F correspond to 4- and 5- stranded parallel configurations of peptide RGDSGAITIGW, respectively. Figures C and G correspond to 4- and 5- stranded parallel configurations of peptide RGDSGAITIGH, respectively. Figures D and H correspond to 4- and 5- stranded parallel configurations of peptide RGDSGAITIGF, respectively.
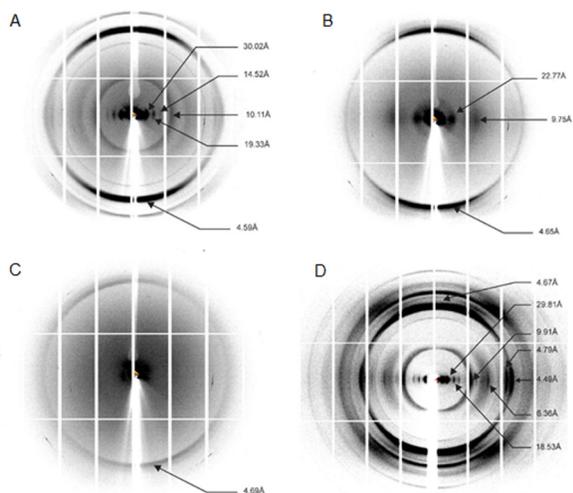
Figure S4. X-ray fiber diffraction patterns for (A) RGDSGAITIGY, (B) RGDSGAITIGW, (C) RGDSGAITIGH and (D) RGDSGAITIGF peptides recorded at the European Synchrotron Radiation facility. All peptides present the characteristic reflections of cross β-sheet structure.

Supplementary References

[1] D. Laage, G. Stirnemann and J. T. Hynes, *J. Phys. Chem. B.,* 2009, **113**, 2428-2435.

[2] Q. Du, E. Freysz and Y. R. Shen, *Science,* 1994, **264**, 826-828.

[3] F. Despa and R. S. Berry, *Biophys J.,* 2007, **92**, 373-378.

[4] C. Cyrus, *Nature,* 1974, **248**, 338-339.

[5] H. Eisenberg, *Clarendon Press*, 1976.

[6] R. S. Spolar, J. Hat and M. T. Record, *Proc. Natl. Acad. Sci. USA.,* 1989, **86**, 8382-8385.

[7] J. M. J. Swanson, R. H. Henchman and J. A. McCammon, *Biophysics. J.*, 2004, **86**, 67–74.

[8] U. Samanta, R. P. Bahadur and P. Chakrabarti, *Protein Eng.,* 2002, **15**, 659-667.

[9] J. L. Pascual-Ahuir, E. Silla and I. Tuñon, *J. Comput. Chem.,* 1994, **15**, 1127 – 1138.