Supplementary Information

for

Exploring the candidates for a new protein folding – cross-α amyloid – in available protein databases

Mitradip Das[*a,b] and B. L. Bhargava[*b]

[a]*Department of Chemical Sciences, Tata Institute of Fundamental Research, Colaba, Mumbai, India 400005.*

[b]*School of Chemical Sciences, National Institute of Science Education and Research – Bhubaneswar, HBNI, Jatni, Odisha, India 752050.*

*Corresponding authors:

Mitradip Das: mitradip.das@tifr.res.in

B. L. Bhargava: bhargava@niser.ac.in

Table of Contents

Simulation and Computational Details

The systems were generated by extracting the required α-helix from the protein, with a padding of one or two residues at both ends for flexibility. This was carried out using the Solution Builder module in CHARMM-GUI. The extracted α-helix was then taken, stripped of all water and ions, and the PSF and CRD files were used as an input of the Multicomponent Assembler module. Large enough box length was considered to ensure that the individual α-helices are free to not "see" each other, i.e. they are free to be away from each other beyond the Coulomb and van der Waals cutoff, and the system was solvated with water. While solvating, Na^+ and Cl^- ions were added at the concentration of 0.15 M, besides neutralizing ions, in order to mimic biophysical conditions. Details of each system can be found in Table S1.

The simulations contain two steps: solvent equilibration and production. In case of solvent equilibration, additional restraints were put on the peptides and the system was equilibrated at constant NVT conditions for 1 ns with temperature set at 310.15 K and a step size of 1 fs. For production run, the system was simulated using constant NPT conditions for 500 ns (for proto-fibrillation) and 1 μs (for fibril formation), by fixing the temperature at 310.15 K and pressure at 1.0 bar. The integration was carried out using leap frog algorithm, where the integration time step was selected to be 2 fs, with the neighbors updated every 20 steps (40 fs). The electrostatic calculations were carried out using the particle mesh ewald (PME) model with a coulomb cutoff at 1.2 nm. Same cutoff (1.2 nm) was also used for calculation of van der Waals interactions. Nosé-Hoover thermostat was used for maintaining the temperature, where the coupling time was selected to be 1 ps. For controlling the pressure, Parrinello-Rahman barostat was used, where the coupling time was set to 5 ps and compressibility to $4.5 \times 10^{-5}\ bar^{-1}$. The bonds involving hydrogen were constrained using linear constraint solver (LINCS) algorithm. Throughout the simulation, the frames were stored at an interval of 2 ps.

Table S1: Details of simulation

No.	PDB ID, Chain, Residue ID	No of α-helices	No of water molecules	No of Ions (Na^+,Cl^-)	Total No. of atoms	Box dimensions (nm × nm × nm)	Simulation time (ns)
1	1A8I, A, 614-630	3	22643	64, 64	69236	9.0 × 9.0 × 9.0	500
2	1EVH, A, 94-111	3	15864	50, 44	48601	8.0 × 8.0 × 8.0	500
3	1UW4, B, 824-835	3	10574	29, 29	32515	7.0 × 7.0 × 7.0	500
4	1X9I, A, 137-153	3	15857	44, 50	48661	8.0 × 8.0 × 8.0	500
5	2BW0, A, 168-185	3	15797	44, 50	48571	8.0 × 8.0 × 8.0	500
6	2DFD, D, 230-249	3	19034	53, 56	58264	8.5 × 8.5 × 8.5	500
7	3O3U, N, 186-201	3	12994	36, 39	39981	7.5 × 7.5 × 7.5	500
8	3P1A, A, 285-301	3	15829	50, 44	48493	8.0 × 8.0 × 8.0	500
9	1X9I, A, 137-153	8	54036	152, 160	164596	12.0 × 12.0 × 12.0	1000

Cross Helix Distance Map (CHDM) Analysis

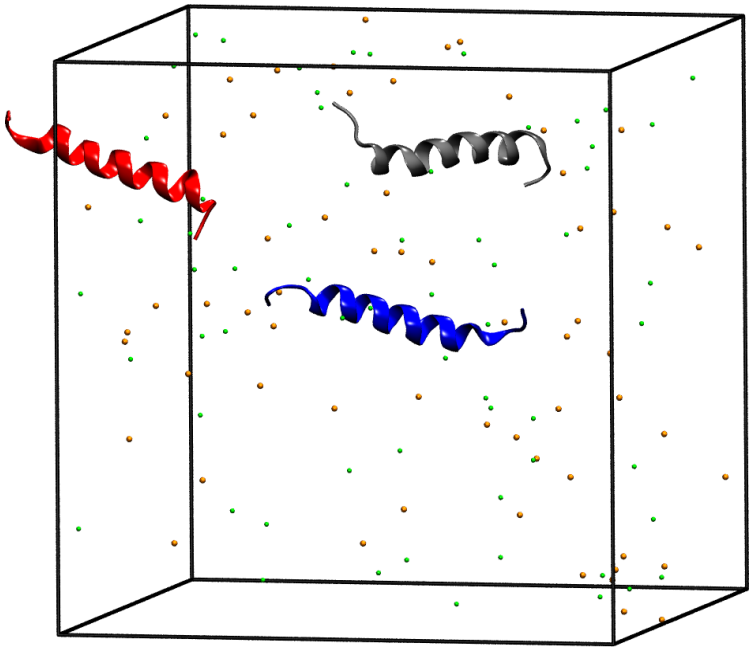
Cross helix distance map (CHDM) is a matrix which essentially calculates the distance between two residues in two different helices of the system, with averaging over time. The idea for CHDM is similar to that of contact map, except being quantitative rather than binary. In order to understand CHDM analysis, let us consider two α-helices, A and B. At first, a representative atom from each residues (in this case, C_α) is selected. Considering one residue in helix A, the minimum distance (considering periodic boundary condition) with each residue in helix B is calculated. The process is then repeated for each residue in helix A to obtain the whole matrix. The elements of the matrix are then averaged over multiple frames to obtain CHDM. A continuous low lying area in the CHDM plot shows the residues which are close to each other, and gives a general idea how the system is arranged.

A derived analysis from the CHDM is the minimum distance histogram. In this case, the minimum along each row (or column) of the CHDM is considered. A region of the minimum distance histogram with periodic behavior with a periodicity of ~3.6 residues (pitch of α-helix) indicates that the two helices are parallel (or antiparallel) to each other, implying the formation of cross-α amyloid structure.

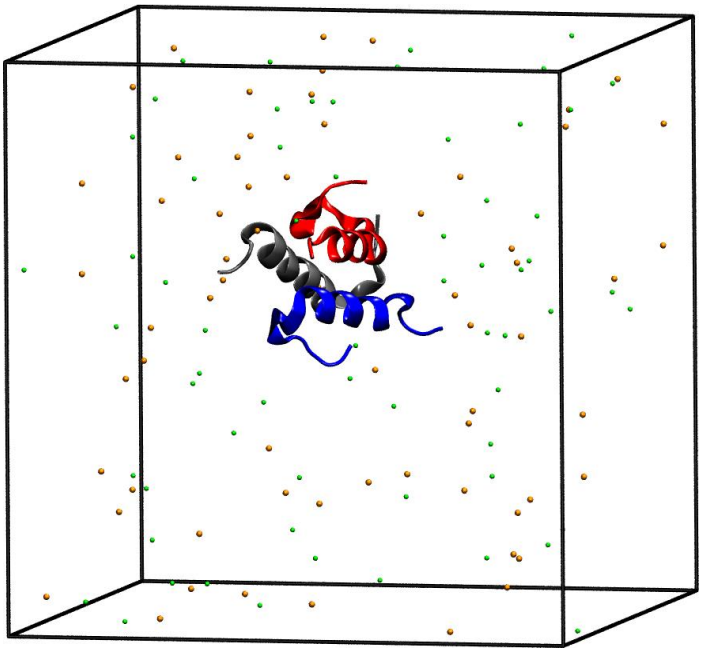
Dot product between helix vectors

In order to calculate the angle between the helices, a helix vector for each of the helices have been defined and the dot product between unit vectors in the same direction has been considered. To consider a helix vector owing to some helix unwinding in solution, we consider the residues which retains helical character (α-helix, π-helix or 3_{10}-helix) thorough at least half of the trajectory to define the start and end of the helices. Since an alpha helix has 3.6 residues per turn (7.2 residues for every two turn), we consider the center of mass of the C_α atoms of the residue marking the start of the helix to six residues after it (total seven residues) as one point in the helix vector. Similarly, the other point in the helix vector is considered as the center of mass of the C_α atoms of the residue marking the end of the helix from six residues before it.

Similarly, helix vectors for other helices are defined and in order to get an idea about the orientation, dot product between the unit vectors along the concerned helix vectors is considered. The angle can be obtained, if required, by calculation of the inverse cosine of the dot product.

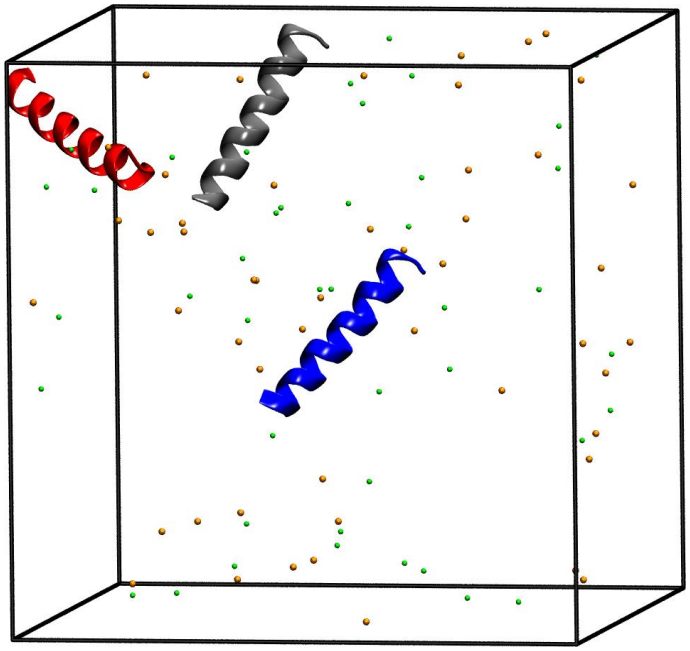


(a) Initial State

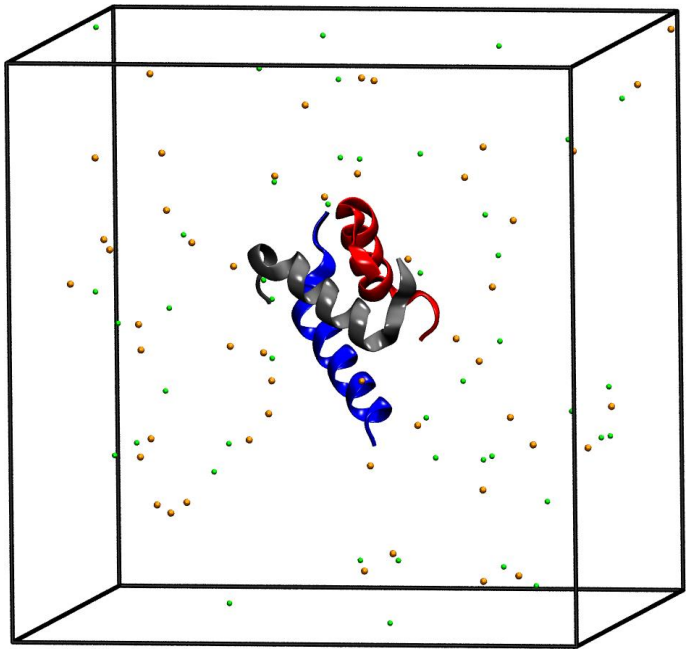


(b) Final State

Figure S1: Snapshots for simulation during proto-fibrillation process of PDB ID: 1A8I, Residue ID: 614-630, Chain: A. The red, blue and black colours represent the helices while the orange and green colour represents sodium and chloride ions respectively.

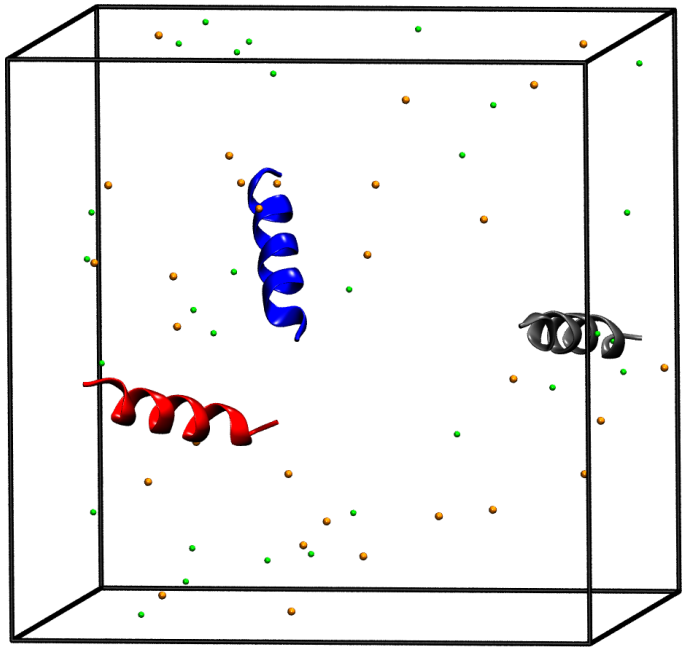


(a) Initial State

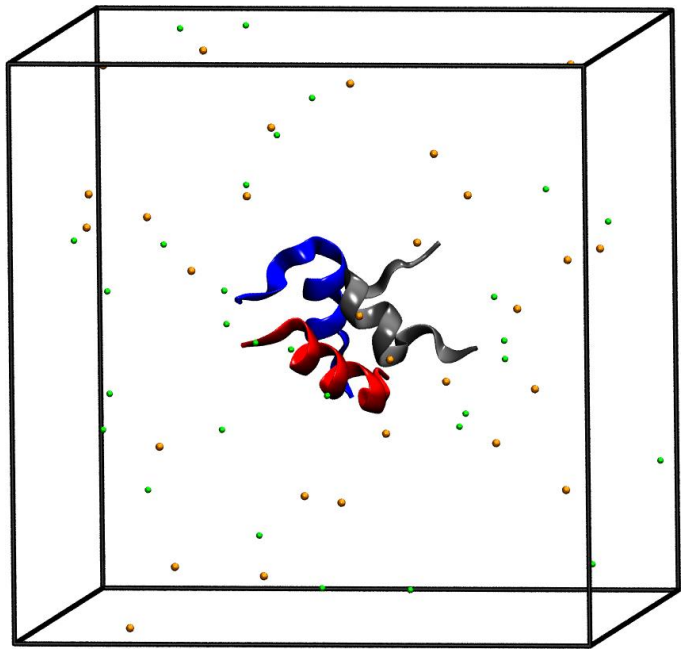


(b) Final State

Figure S2: Snapshots for simulation during proto-fibrillation process of PDB ID: 1EVH, Residue ID: 94-111, Chain: A. The red, blue and black colours represent the helices while the orange and green colour represents sodium and chloride ions respectively.

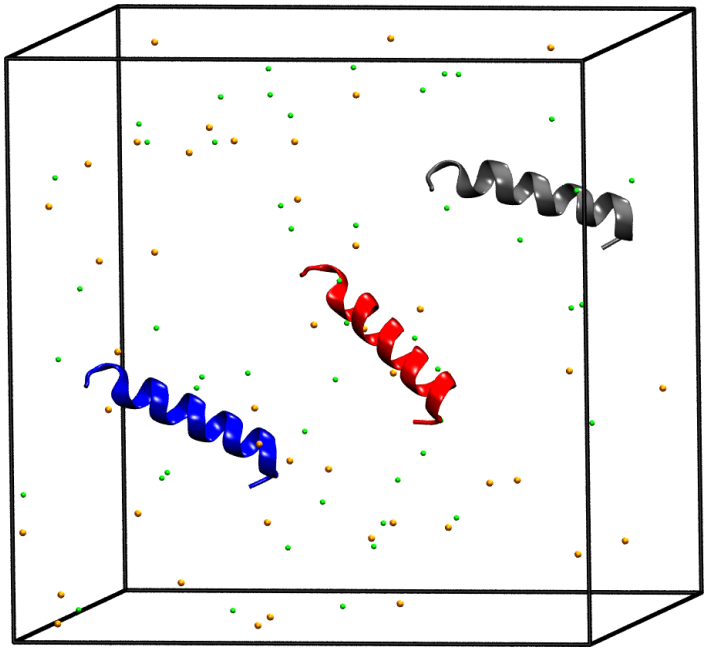


(a) Initial State

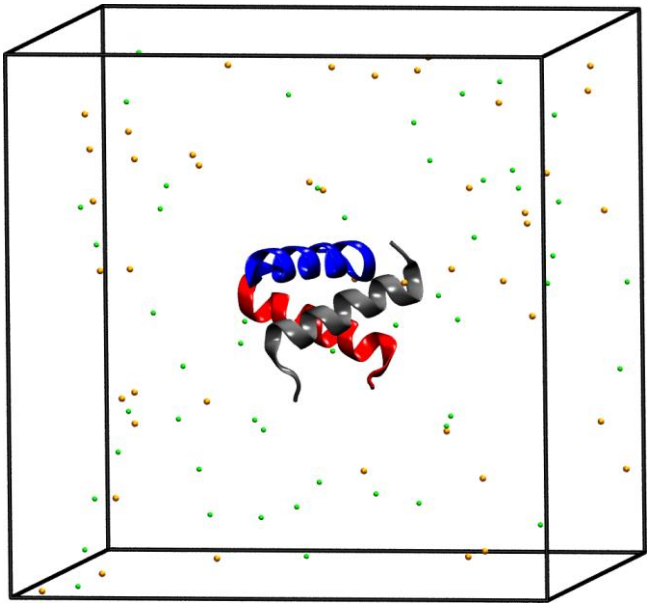


(b) Final state

Figure S3: Snapshots for simulation during proto-fibrillation process of PDB ID: 1UW4, Residue ID: 824-835, Chain: B. The red, blue and black colours represent the helices while the orange and green colour represents sodium and chloride ions respectively.

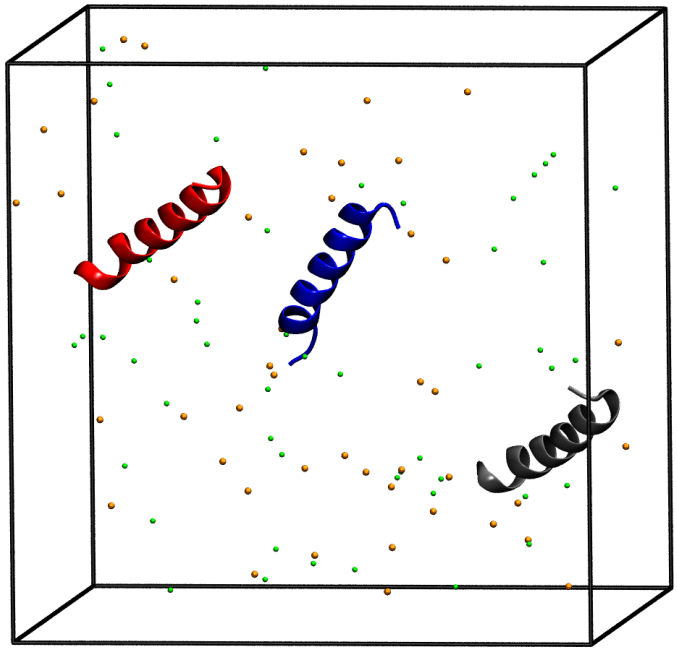


(a) Initial State

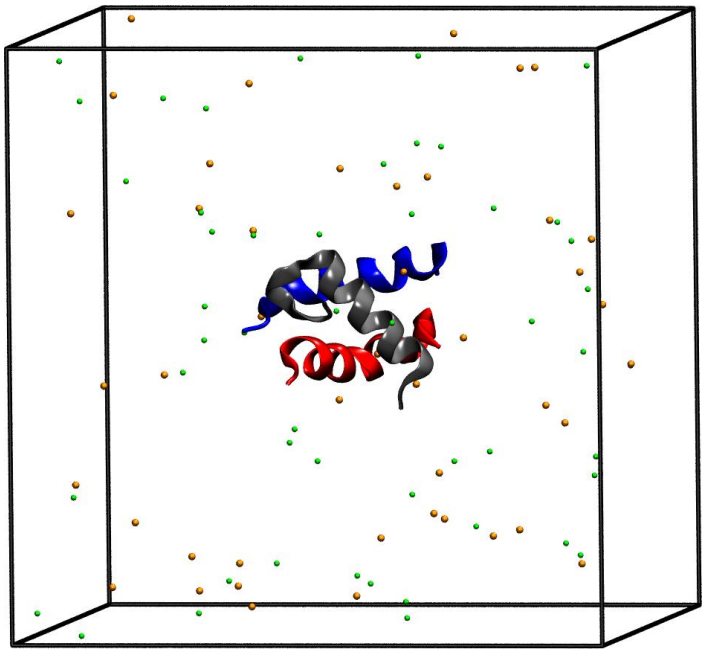


(b) Final state

Figure S4: Snapshots for simulation during proto-fibrillation process of PDB ID: 1X9I, Residue ID: 137-153, Chain: A. The red, blue and black colours represent the helices while the orange and green colour represents sodium and chloride ions respectively.

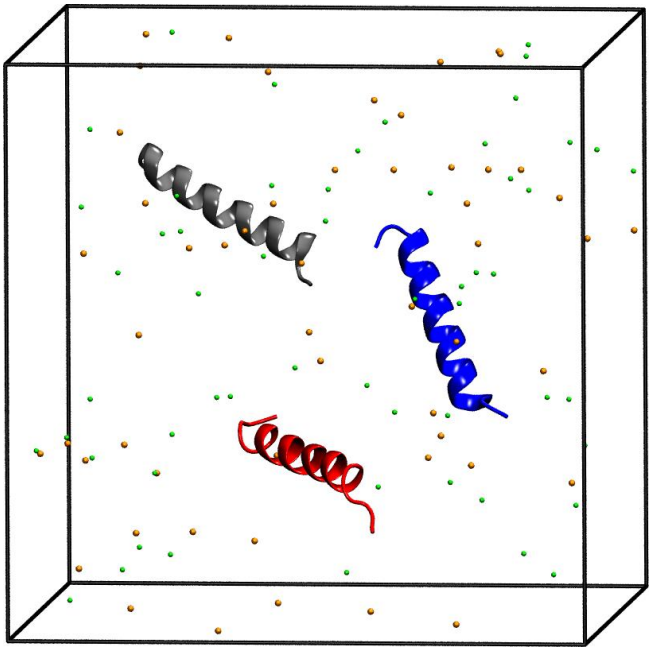


(a) Initial State

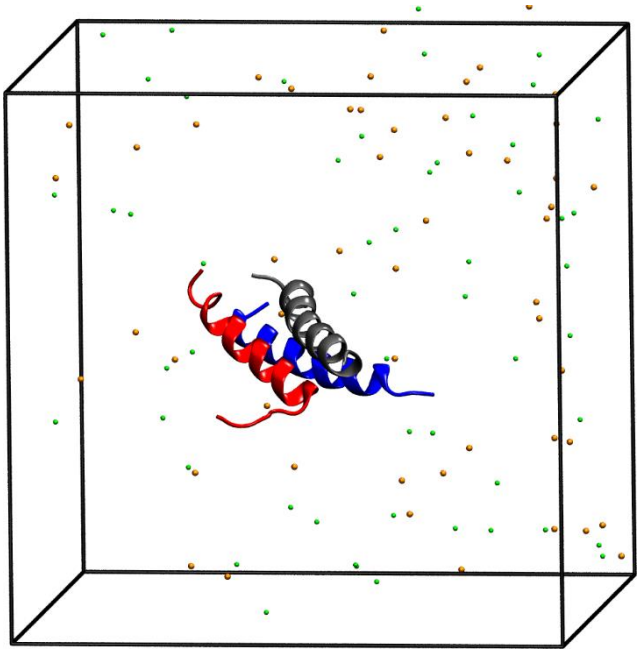


(b) Final state

Figure S5: Snapshots for simulation during proto-fibrillation process of PDB ID: 2BW0, Residue ID: 168-185, Chain: A. The red, blue and black colours represent the helices while the orange and green colour represents sodium and chloride ions respectively.

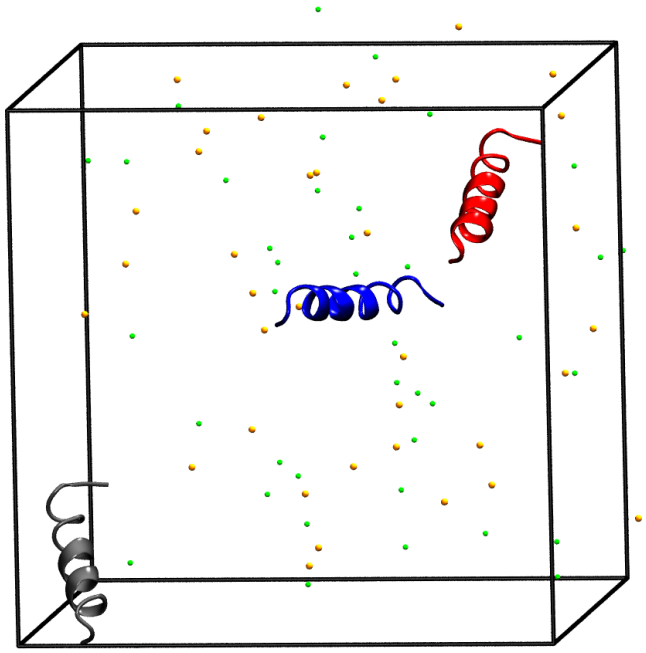


(a) Initial State

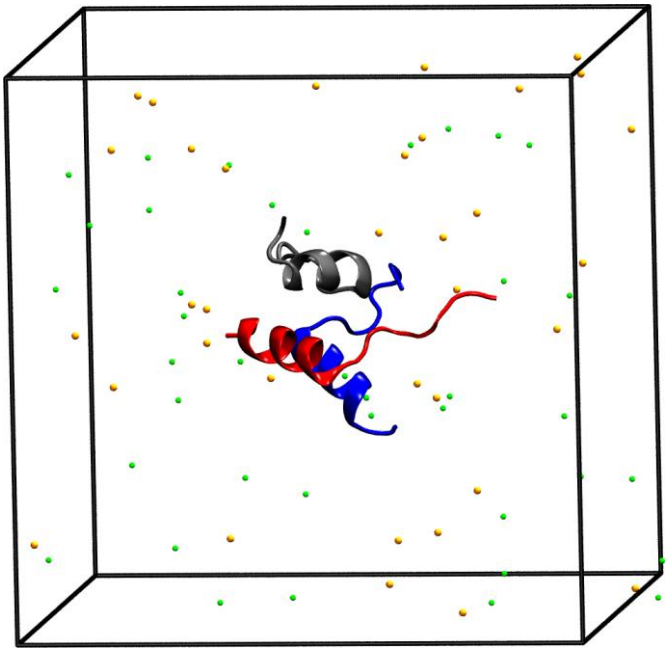


(b) Final state

Figure S6: Snapshots for simulation during proto-fibrillation process of PDB ID: 2DFD, Residue ID: 230-249, Chain: D. The red, blue and black colours represent the helices while the orange and green colour represents sodium and chloride ions respectively.

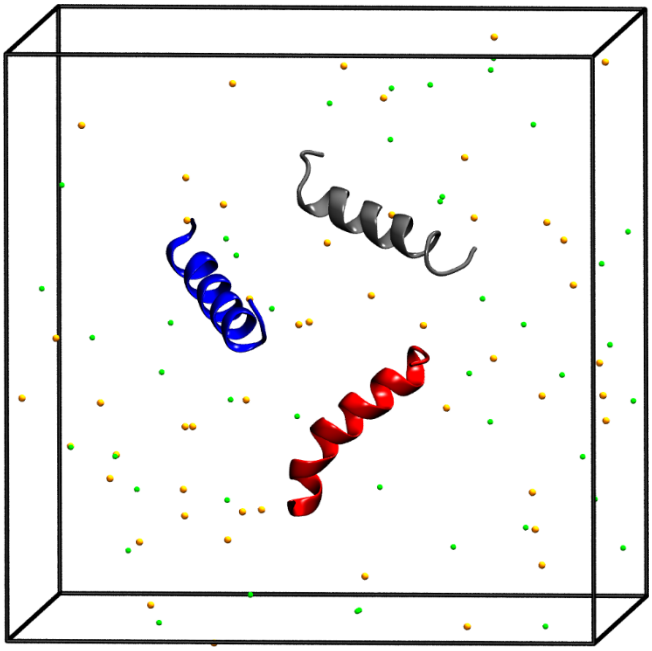


(a) Initial State

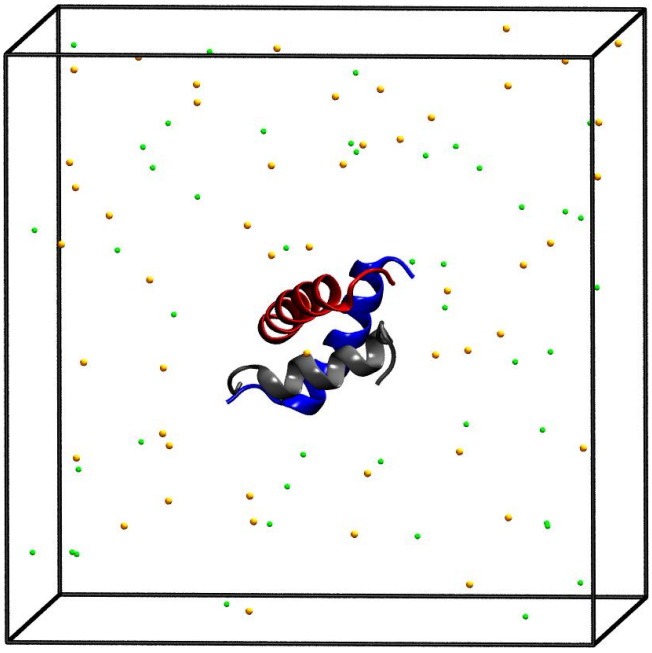


(b) Final state

Figure S7: Snapshots for simulation during proto-fibrillation process of PDB ID: 3O3U, Residue ID: 186-201, Chain: N. The red, blue and black colours represent the helices while the orange and green colour represents sodium and chloride ions respectively.

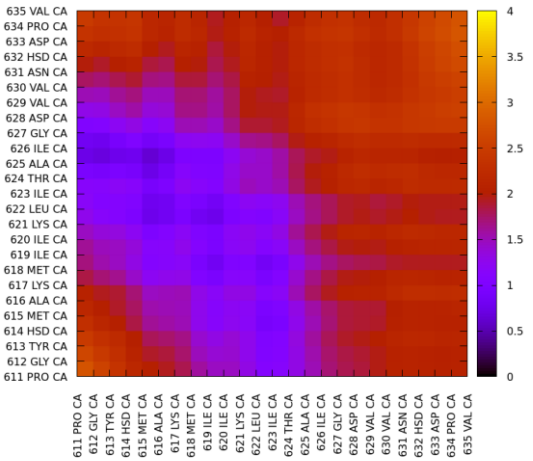


(a) Initial State

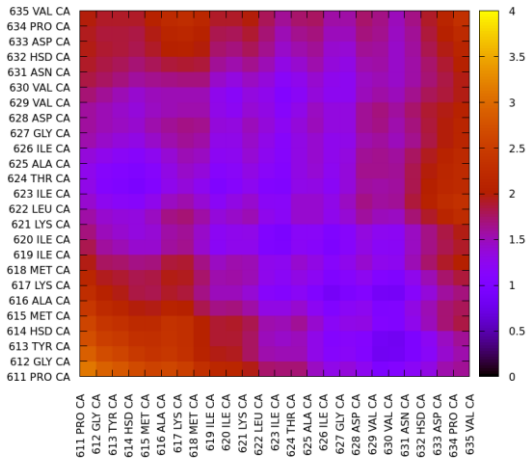


(b) Final state

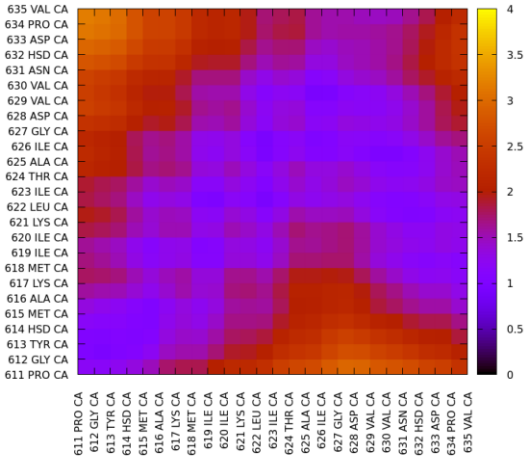
Figure S8: Snapshots for simulation during proto-fibrillation process of PDB ID: 3P1A, Residue ID: 285-301, Chain: A. The red, blue and black colours represent the helices while the orange and green colour represents sodium and chloride ions respectively.



(a) Helices 1(X axis) and 2 (Y axis)

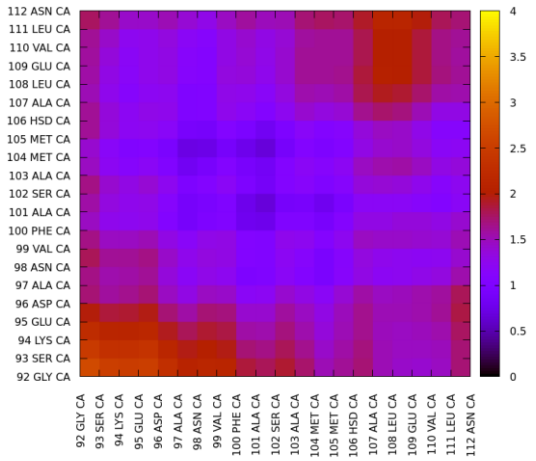


(b) Helices 2 (X axis) and 3 (Y axis)

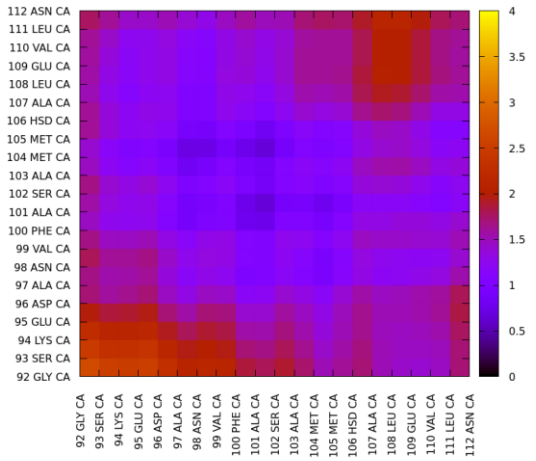


(c) Helices 3 (X axis) and 1 (Y axis)

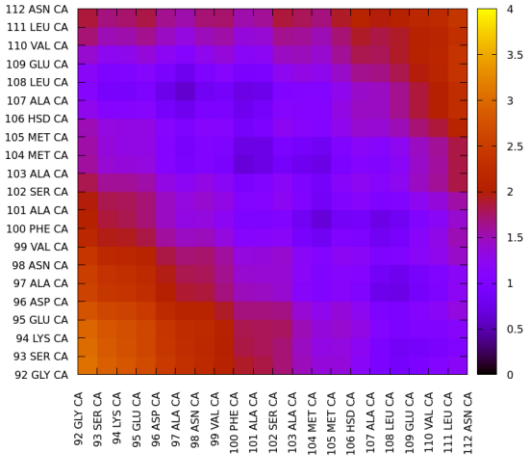
Figure S9: CHDM Maps for proto-fibrillation process of PDB ID: 1A8I, Residue ID: 614-630, Chain: A. The colour axis represents distance in nm.



(a) Helices 1(X axis) and 2 (Y axis)

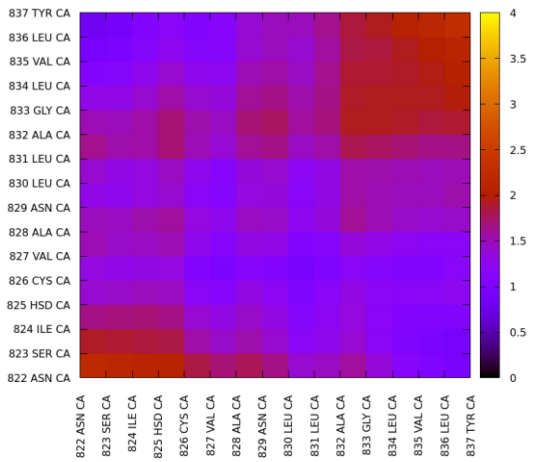


(b) Helices 2 (X axis) and 3 (Y axis)

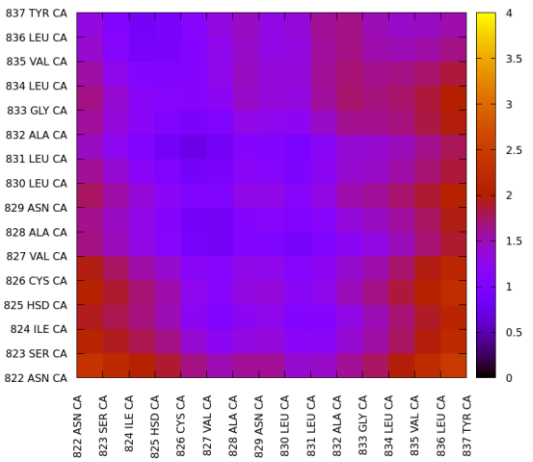


(c) Helices 3 (X axis) and 1 (Y axis)

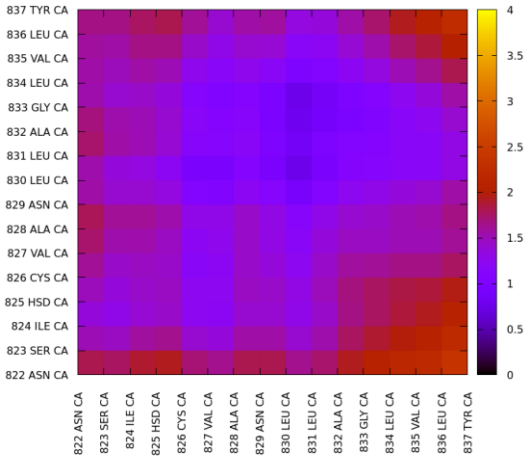
Figure S10: CHDM Maps for proto-fibrillation process of PDB ID: 1EVH, Residue ID: 94-111, Chain: A. The colour axis represents distance in nm.



(a) Helices 1(X axis) and 2 (Y axis)

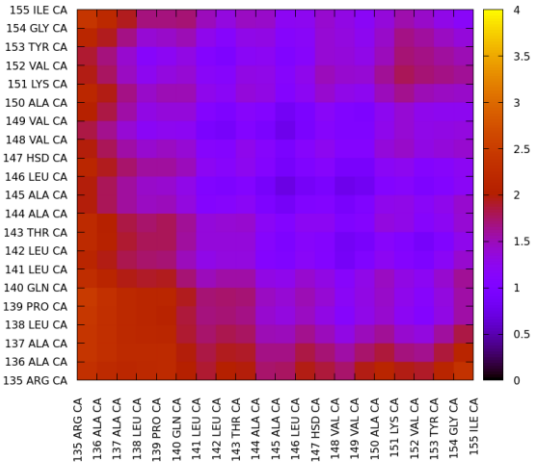


(b) Helices 2 (X axis) and 3 (Y axis)

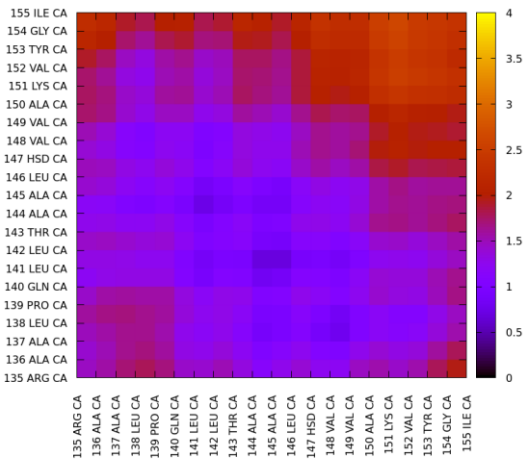


(c) Helices 3 (X axis) and 1 (Y axis)

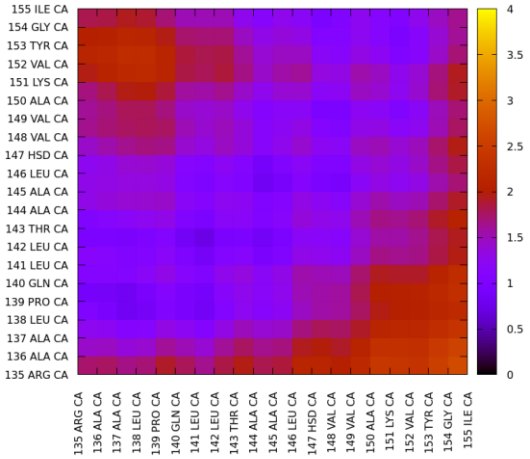
Figure S11: CHDM Maps for proto-fibrillation process of PDB ID: 1UW4, Residue ID: 824-835, Chain: B. The colour axis represents distance in nm.



(a) Helices 1(X axis) and 2 (Y axis)

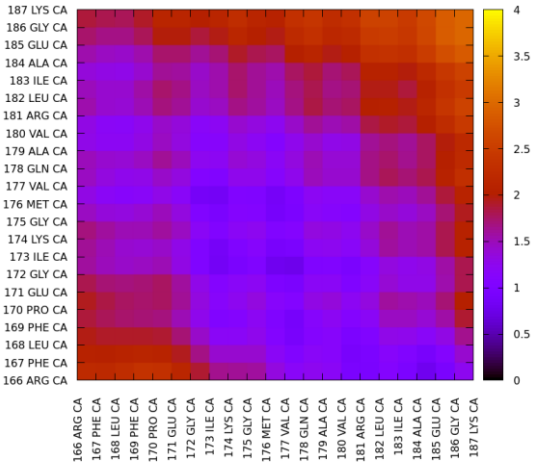


(b) Helices 2 (X axis) and 3 (Y axis)

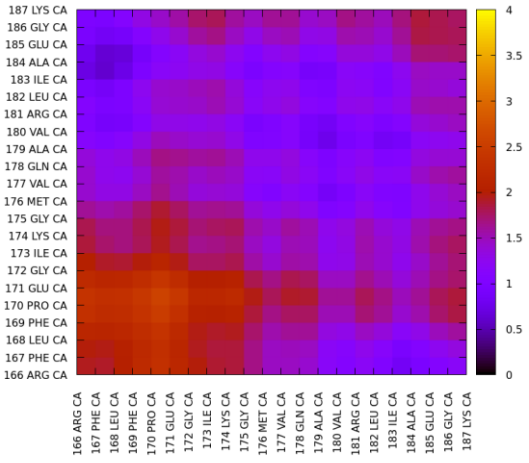


(c) Helices 3 (X axis) and 1 (Y axis)

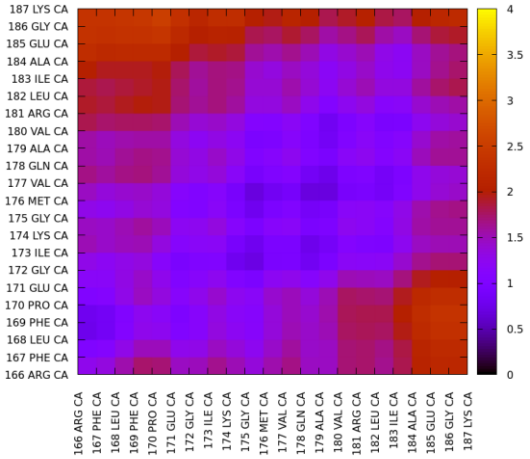
Figure S12: CHDM Maps for proto-fibrillation process of PDB ID: 1X9I, Residue ID: 137-153, Chain: A. The colour axis represents distance in nm.



(a) Helices 1(X axis) and 2 (Y axis)

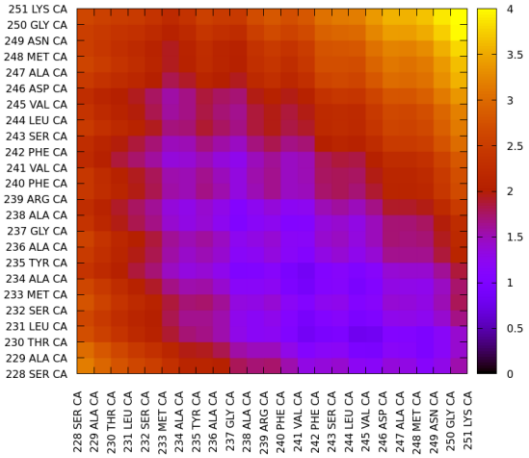


(b) Helices 2 (X axis) and 3 (Y axis)

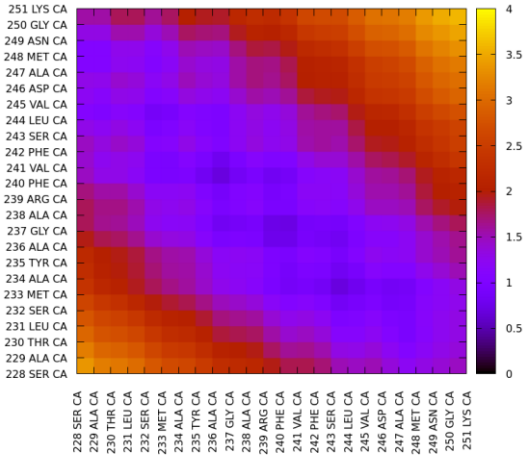


(c) Helices 3 (X axis) and 1 (Y axis)

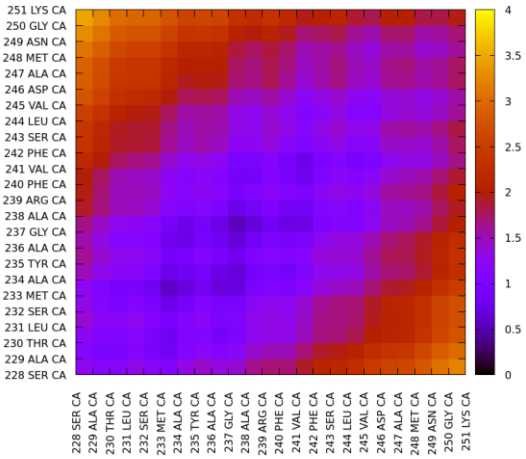
Figure S13: CHDM Maps for proto-fibrillation process of PDB ID: 2BW0, Residue ID: 168-185, Chain: A. The colour axis represents distance in nm.



(a) Helices 1(X axis) and 2 (Y axis)

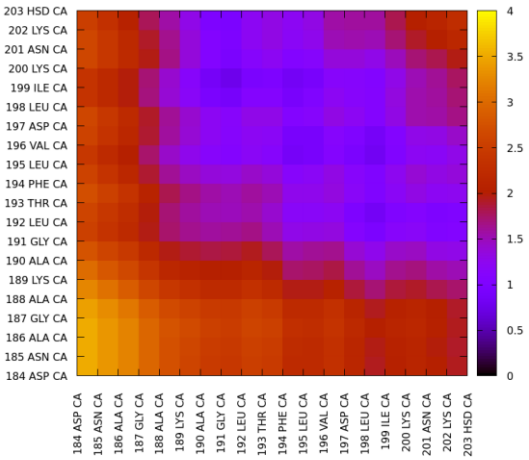


(b) Helices 2 (X axis) and 3 (Y axis)

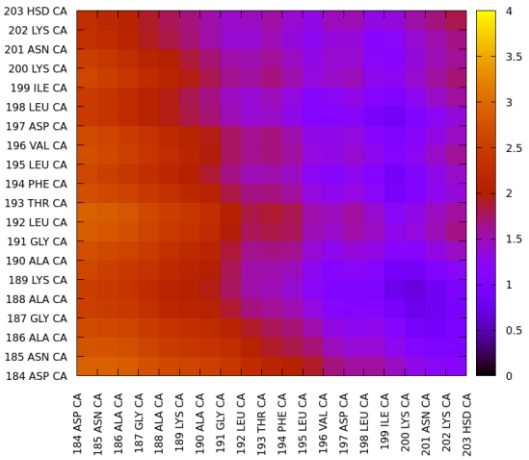


(c) Helices 3 (X axis) and 1 (Y axis)

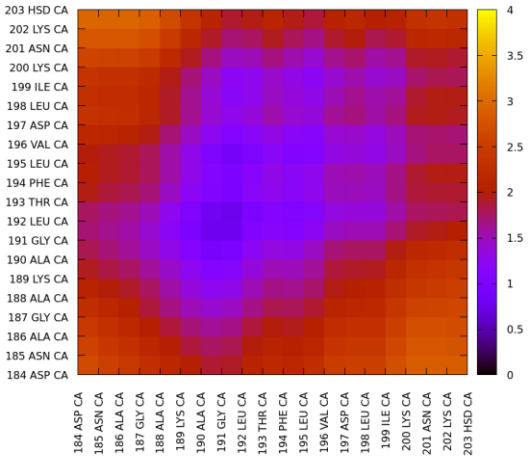
Figure S14: CHDM Maps for proto-fibrillation process of PDB ID: 2DFD, Residue ID: 230-249, Chain: D. The colour axis represents distance in nm.



(a) Helices 1(X axis) and 2 (Y axis)

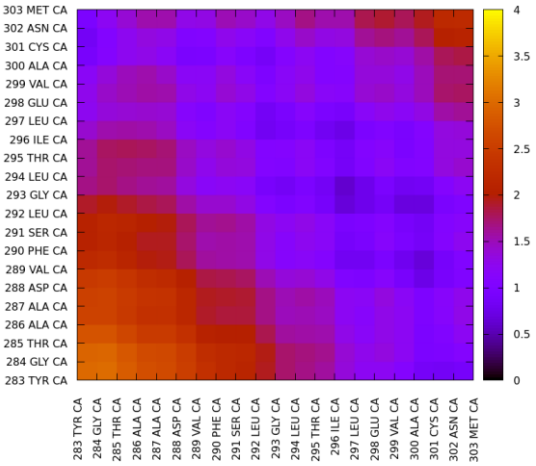


(b) Helices 2 (X axis) and 3 (Y axis)

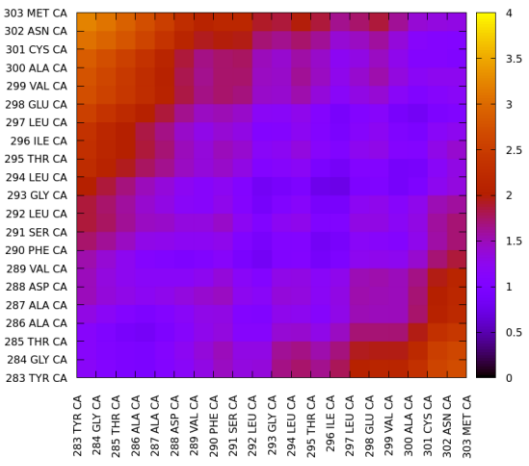


(c) Helices 3 (X axis) and 1 (Y axis)

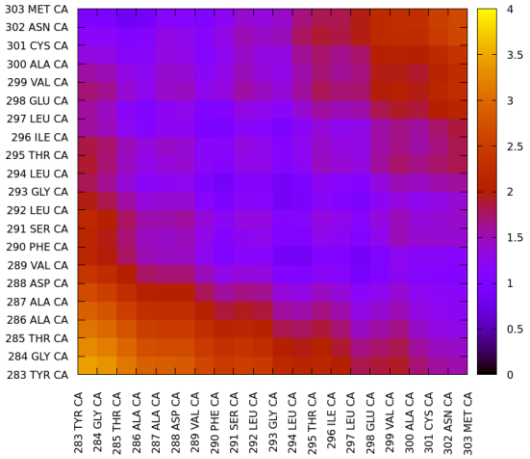
Figure S15: CHDM Maps for proto-fibrillation process of PDB ID: 3O3U, Residue ID: 186-201, Chain: N. The colour axis represents distance in nm.



(a) Helices 1(X axis) and 2 (Y axis)

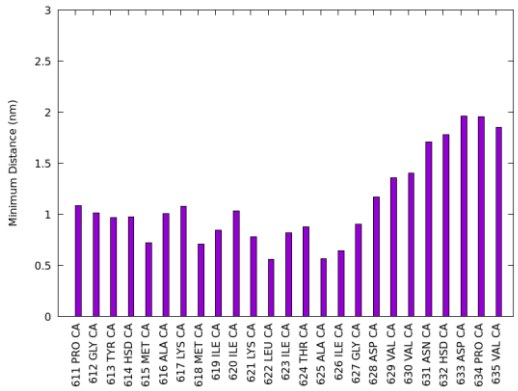


(b) Helices 2 (X axis) and 3 (Y axis)

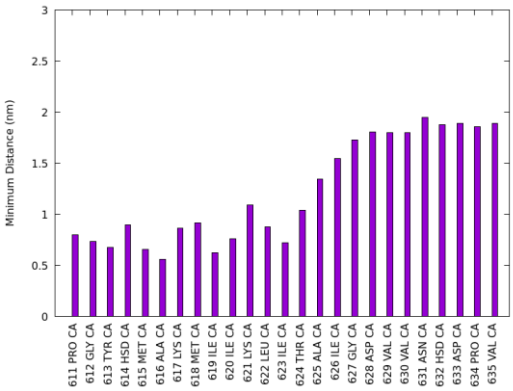


(c) Helices 3 (X axis) and 1 (Y axis)

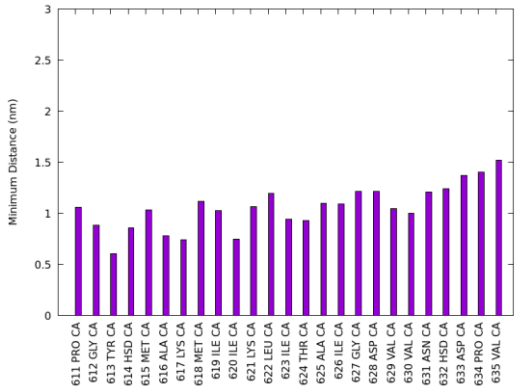
Figure S16: CHDM Maps for proto-fibrillation process of PDB ID: 3P1A, Residue ID: 285-301, Chain: A. The colour axis represents distance in nm.



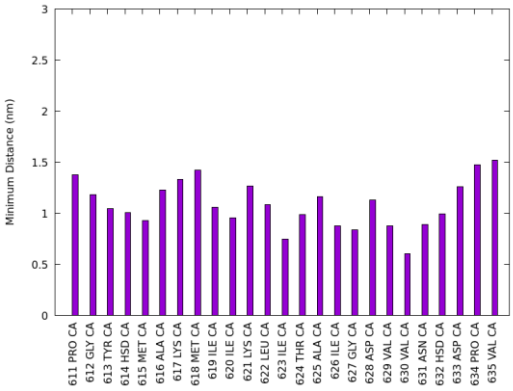
(a) Helix 1 from Helix 2



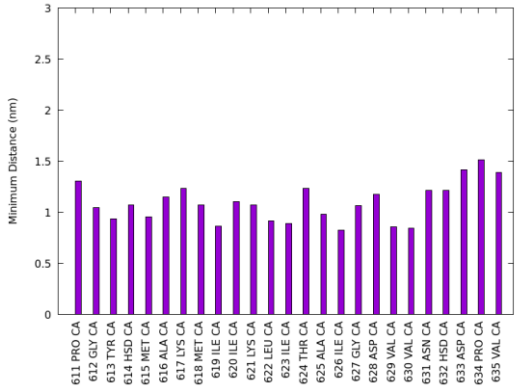
(b) Helix 2 from Helix 1



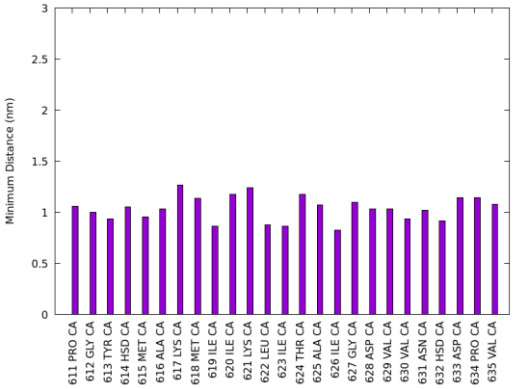
(c) Helix 2 from Helix 3



(d) Helix 3 from Helix 2

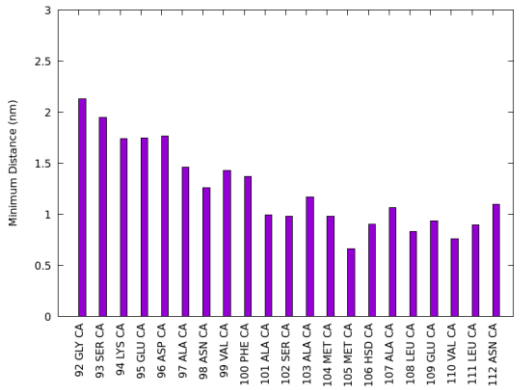


(e) Helix 3 from Helix 1

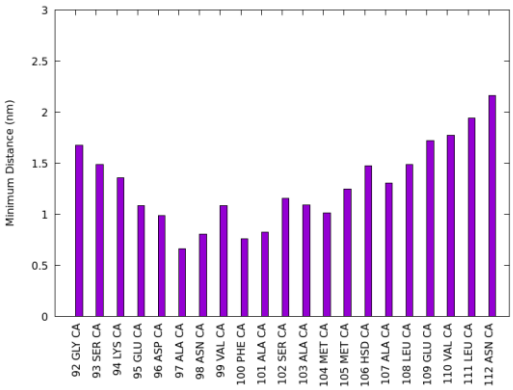


(f) Helix 1 from Helix 3

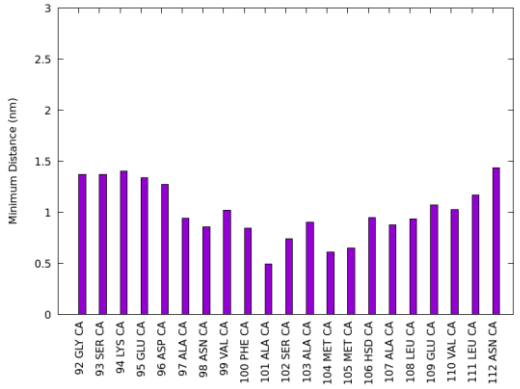
Figure S17: Minimum distance of residues of one helix from another helix, calculated using CHDM, during proto-fibrillation of PDB ID: 1A8I, Residue ID: 614-630, Chain: A.



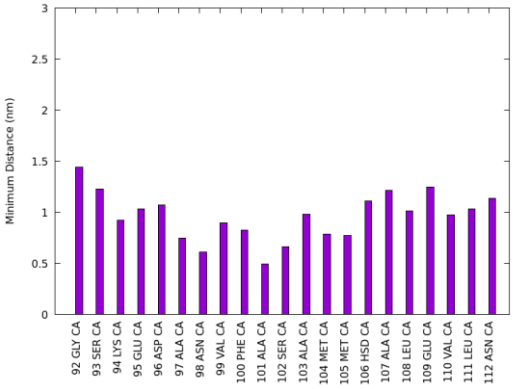
(a) Helix 1 from Helix 2



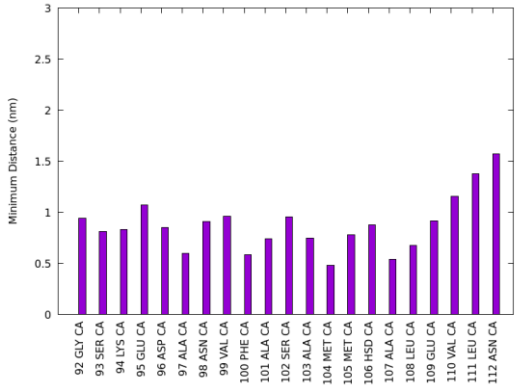
(b) Helix 2 from Helix 1



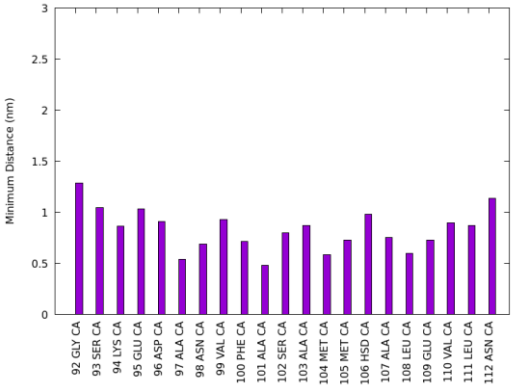
(c) Helix 2 from Helix 3



(d) Helix 3 from Helix 2

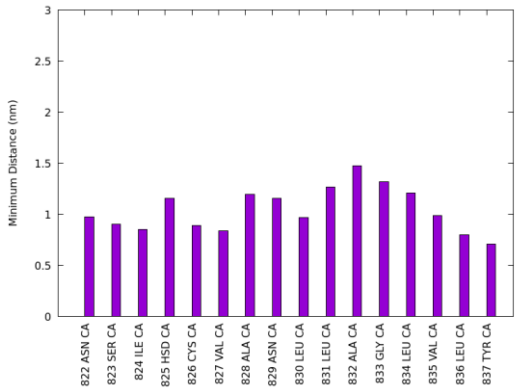


(e) Helix 3 from Helix 1

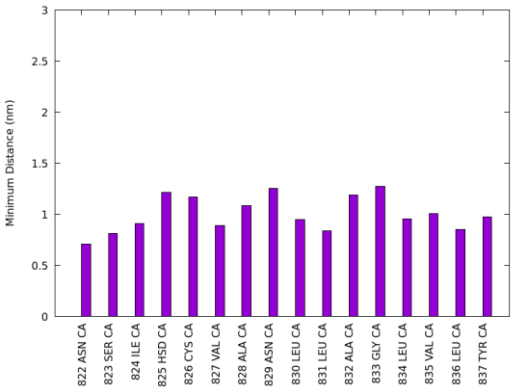


(f) Helix 1 from Helix 3

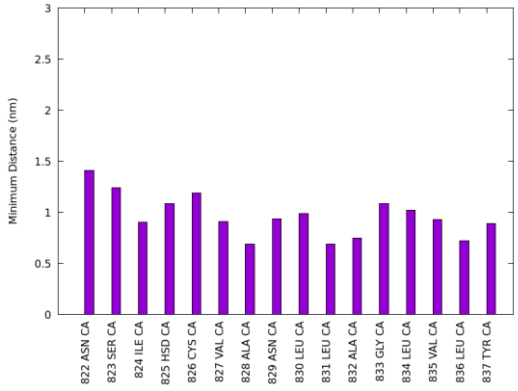
Figure S18: Minimum distance of residues of one helix from another helix, calculated using CHDM, during proto-fibrillation of PDB ID: 1EVH, Residue ID: 94-111, Chain: A.



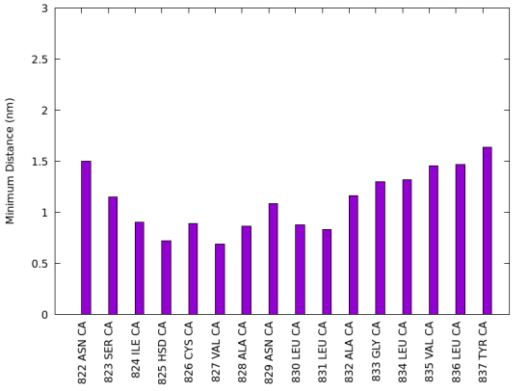
(a) Helix 1 from Helix 2



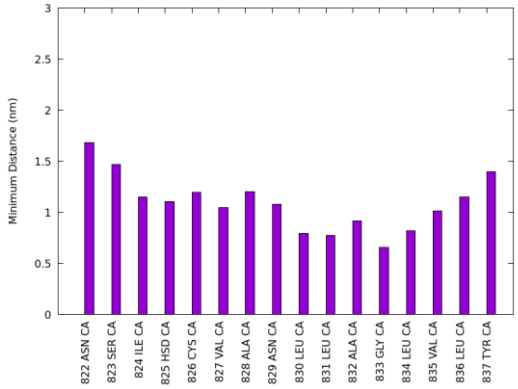
(b) Helix 2 from Helix 1



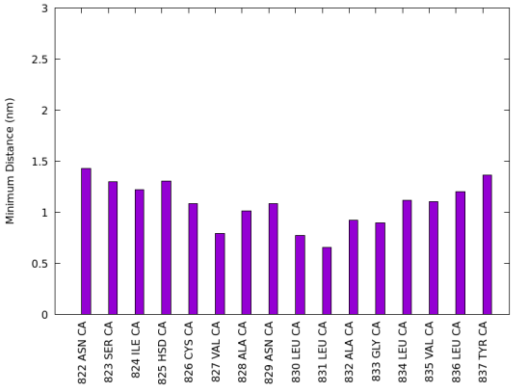
(c) Helix 2 from Helix 3



(d) Helix 3 from Helix 2

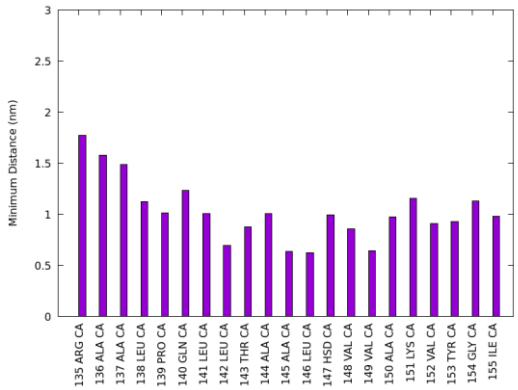


(e) Helix 3 from Helix 1

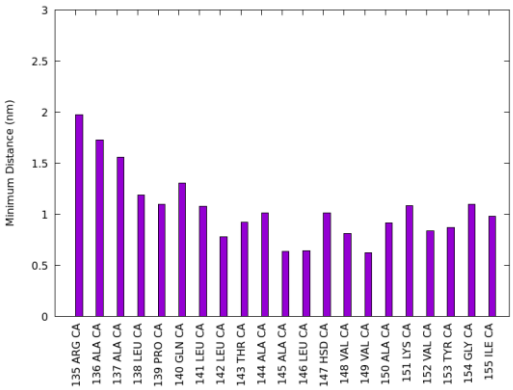


(f) Helix 1 from Helix 3

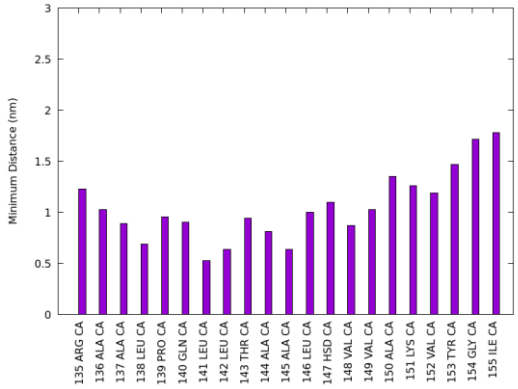
Figure S19: Minimum distance of residues of one helix from another helix, calculated using CHDM, during proto-fibrillation of PDB ID: 1UW4, Residue ID: 824-835, Chain: B.



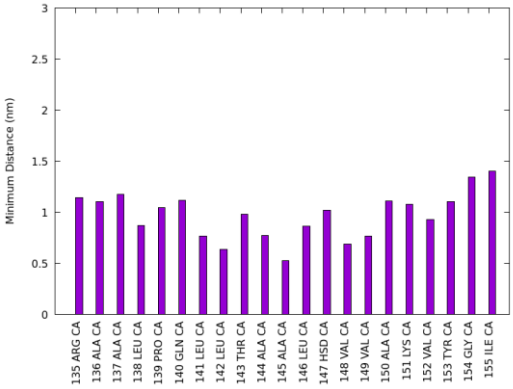
(a) Helix 1 from Helix 2



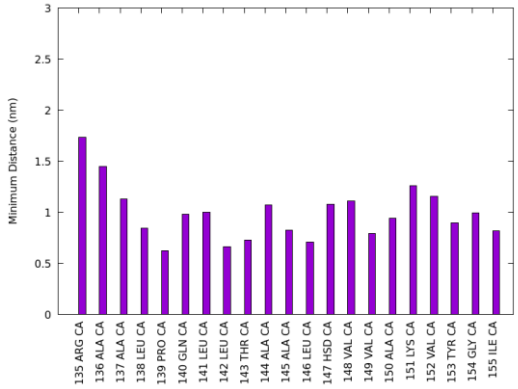
(b) Helix 2 from Helix 1



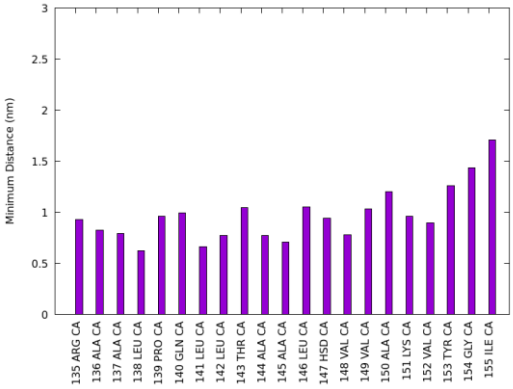
(c) Helix 2 from Helix 3



(d) Helix 3 from Helix 2

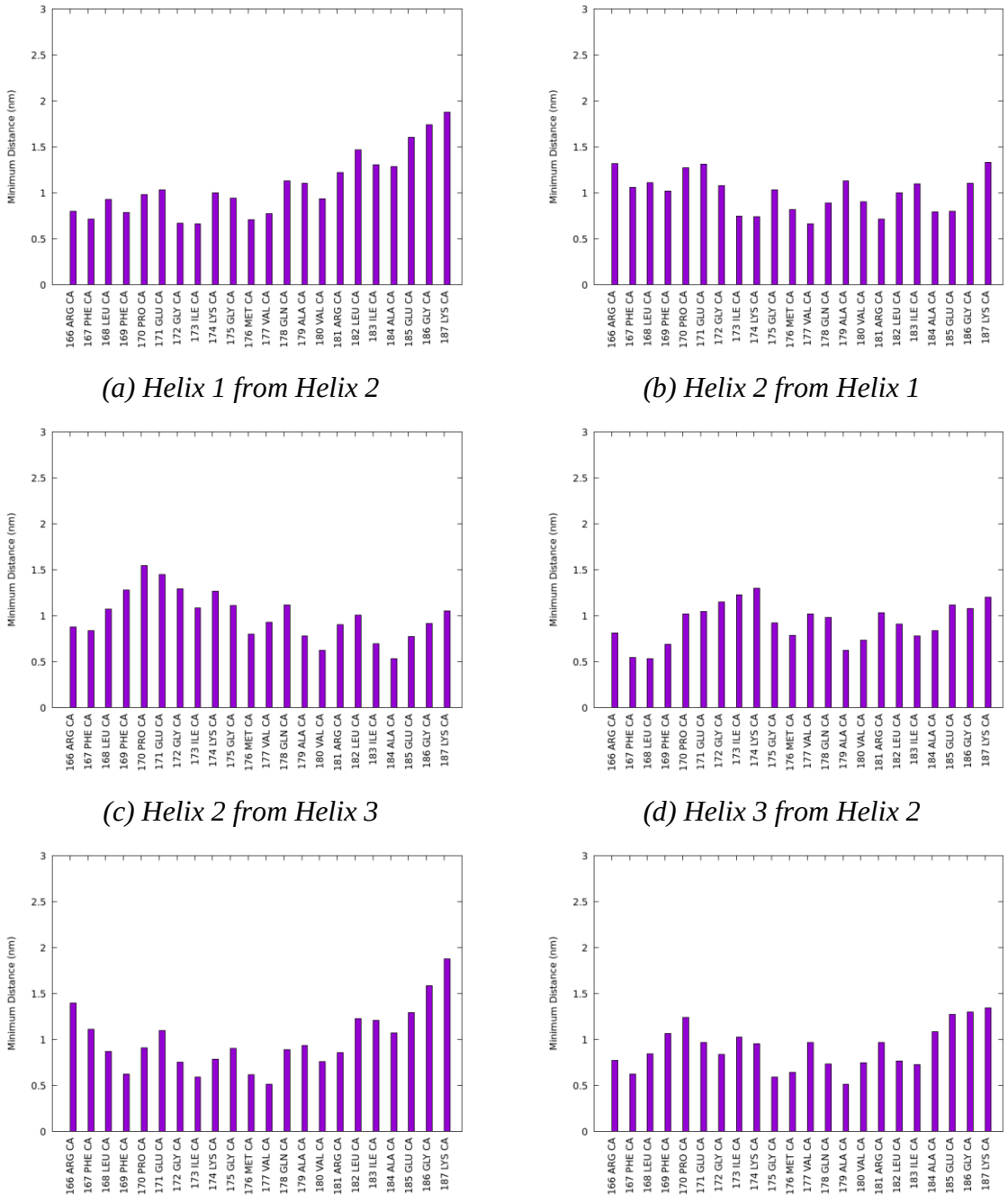


(e) Helix 3 from Helix 1



(f) Helix 1 from Helix 3

Figure S20: Minimum distance of residues of one helix from another helix, calculated using CHDM, during proto-fibrillation of PDB ID: 1X9I, Residue ID: 137-153, Chain: A.



(a) Helix 1 from Helix 2

(b) Helix 2 from Helix 1

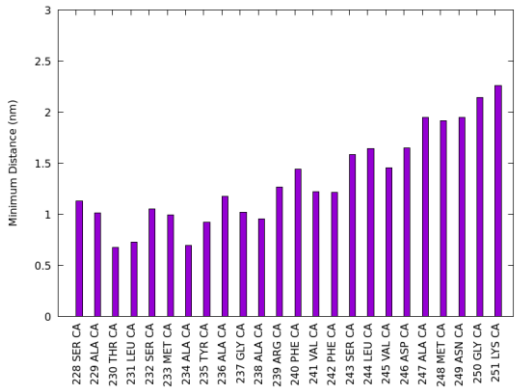
(c) Helix 2 from Helix 3

(d) Helix 3 from Helix 2

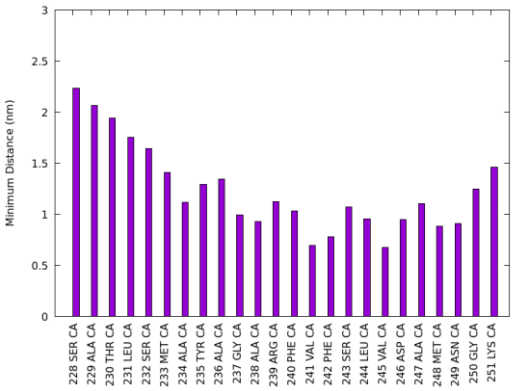
(e) Helix 3 from Helix 1

(f) Helix 1 from Helix 3

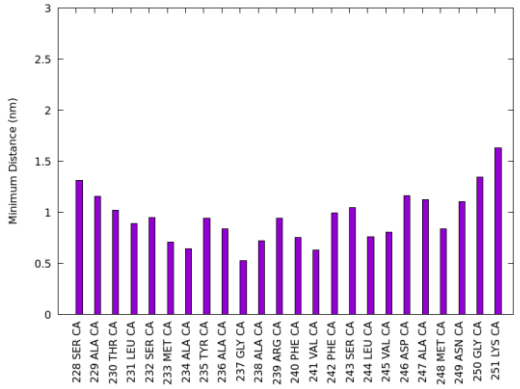
Figure S21: Minimum distance of residues of one helix from another helix, calculated using CHDM, during proto-fibrillation of PDB ID: 2BW0, Residue ID: 168-185, Chain: A.



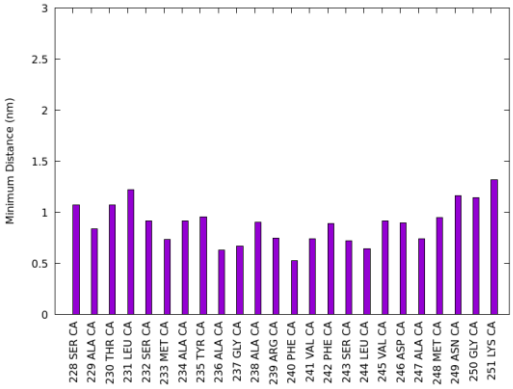
(a) Helix 1 from Helix 2



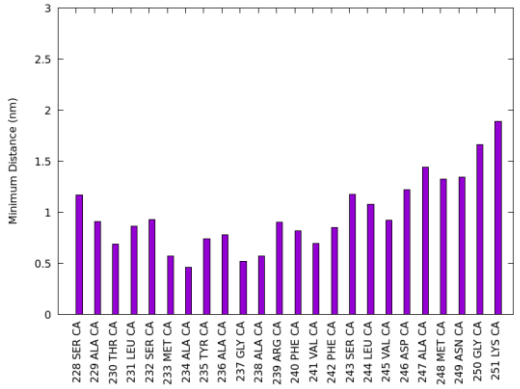
(b) Helix 2 from Helix 1



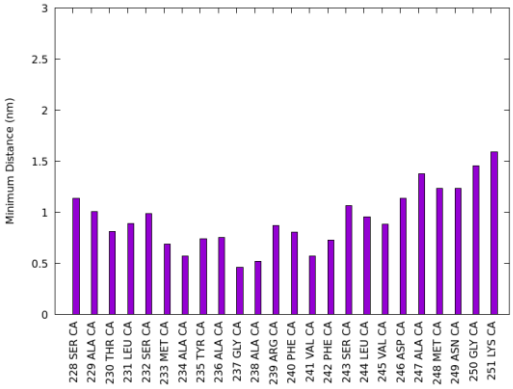
(c) Helix 2 from Helix 3



(d) Helix 3 from Helix 2

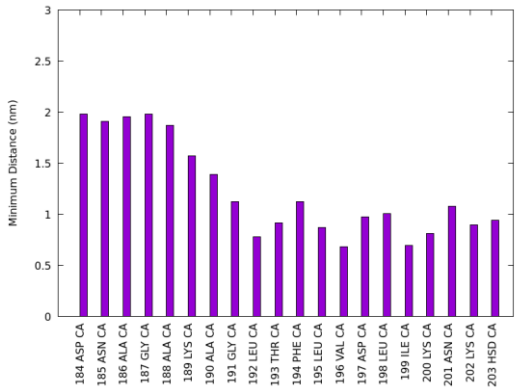


(e) Helix 3 from Helix 1

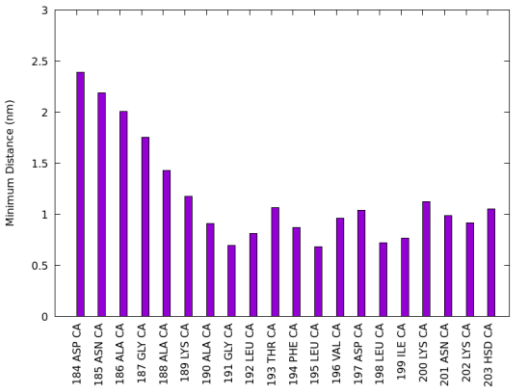


(f) Helix 1 from Helix 3

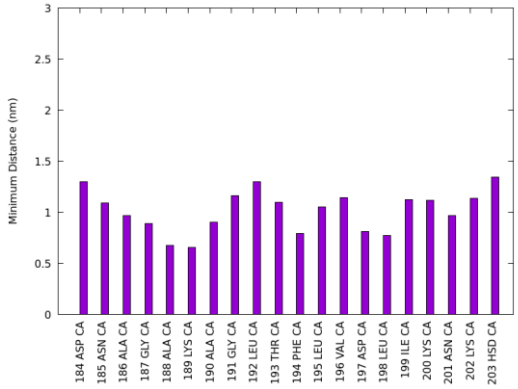
Figure S22: Minimum distance of residues of one helix from another helix, calculated using CHDM, during proto-fibrillation of PDB ID: 2DFD, Residue ID: 230-249, Chain: D.



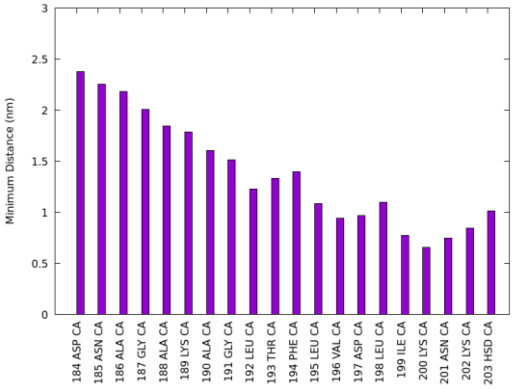
(a) Helix 1 from Helix 2



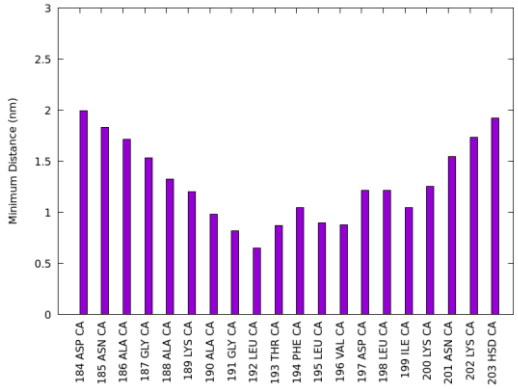
(b) Helix 2 from Helix 1



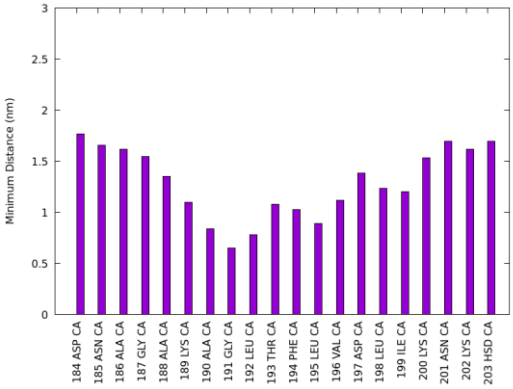
(c) Helix 2 from Helix 3



(d) Helix 3 from Helix 2

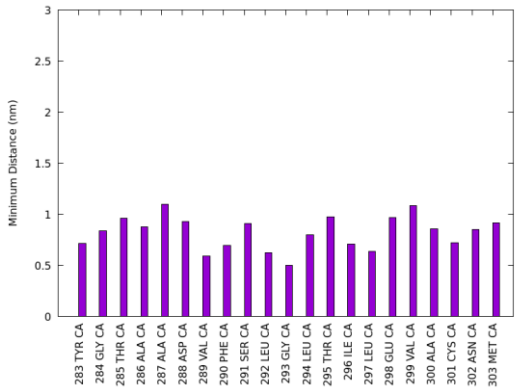


(e) Helix 3 from Helix 1

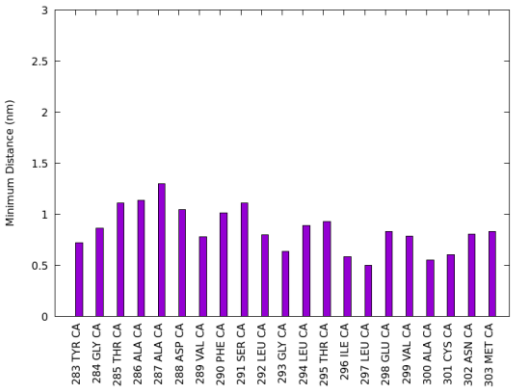


(f) Helix 1 from Helix 3

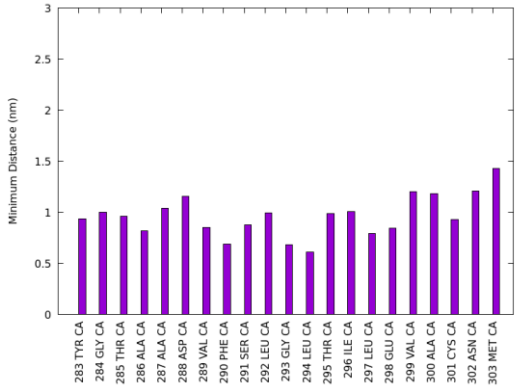
Figure S23: Minimum distance of residues of one helix from another helix, calculated using CHDM, during proto-fibrillation of PDB ID: 3O3U, Residue ID: 186-201, Chain: N.



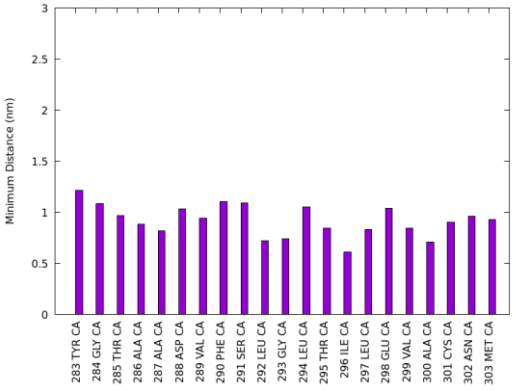
(a) Helix 1 from Helix 2



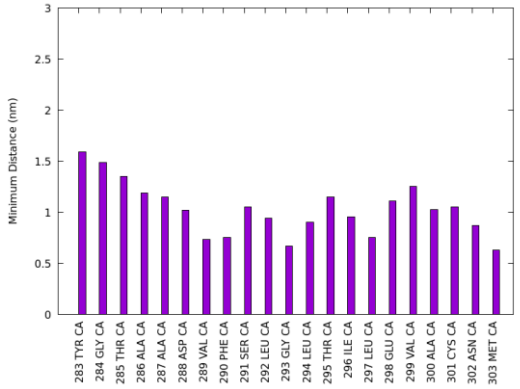
(b) Helix 2 from Helix 1



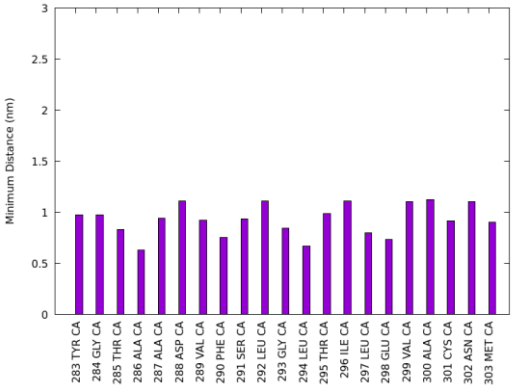
(c) Helix 2 from Helix 3



(d) Helix 3 from Helix 2

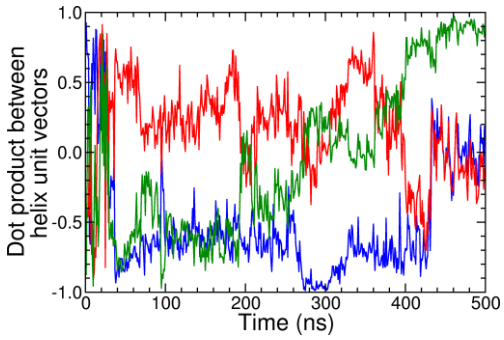


(e) Helix 3 from Helix 1

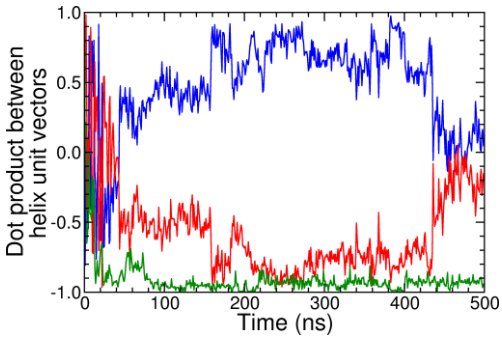


(f) Helix 1 from Helix 3

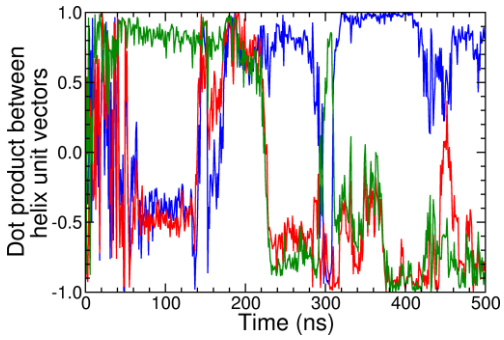
Figure S24: Minimum distance of residues of one helix from another helix, calculated using CHDM, during proto-fibrillation of PDB ID: 3P1A, Residue ID: 285-301, Chain: A.



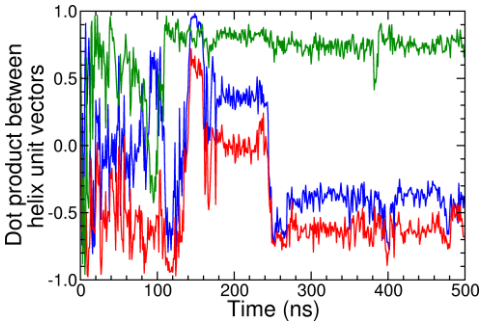
(a) 1A8I, 614-630, A



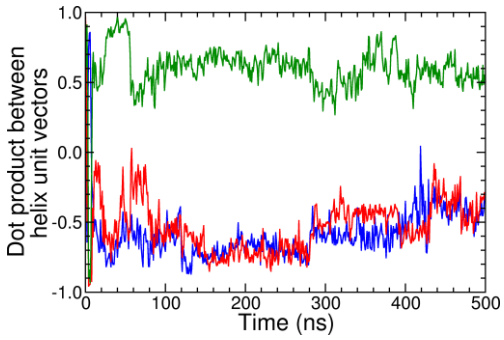
(b) 1EVH, 94-111, A



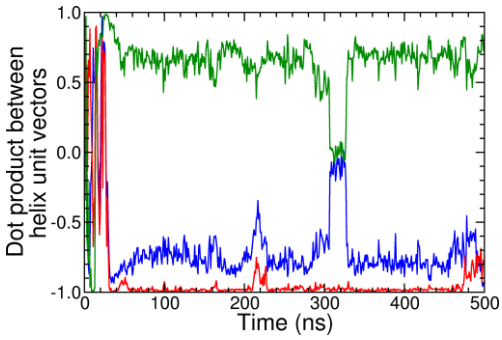
(c) 1UW4, 824-835, B



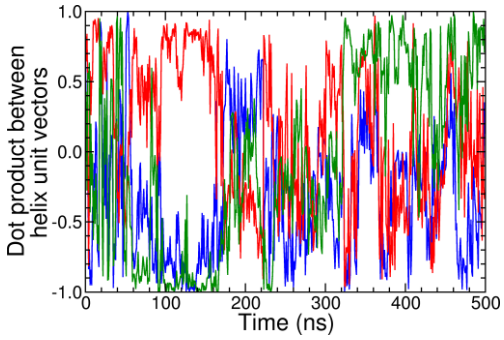
(d) 1X9I, 137-153, A



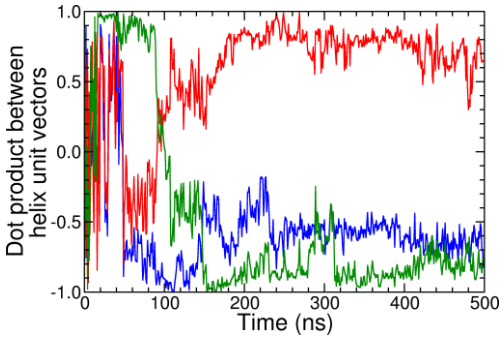
(e) 2BW0, 168-185, A



(f) 2DFD, 230-249, A

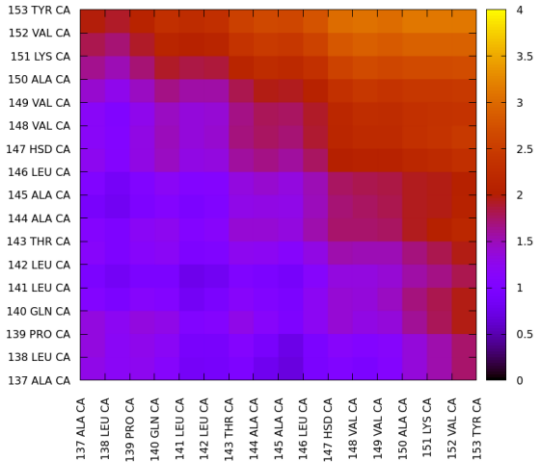


(g) 3O3U, 186-201, N

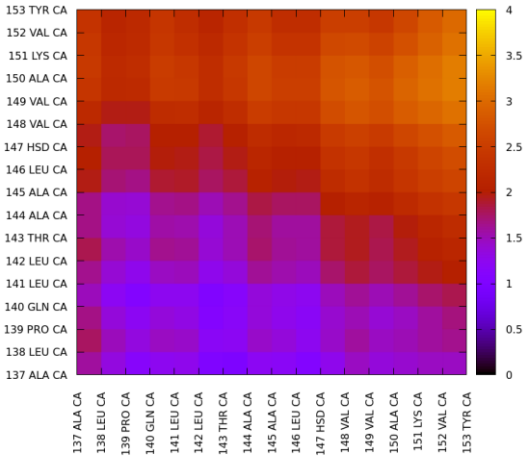


(h) 3P1A, 285-301, A

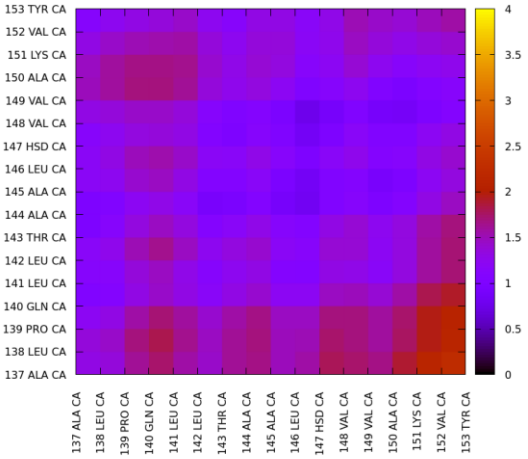
Figure S25: The dot product between the helix vectors with time during formation of proto-fibril in various cross-α amyloid candidates. The colours represent pairs of helices between which the dot product has been calculated and the subcaptions have been given in format PDB ID, Residue ID and chain ID.



(a) Helices A1 (X axis) and A2 (Y axis)

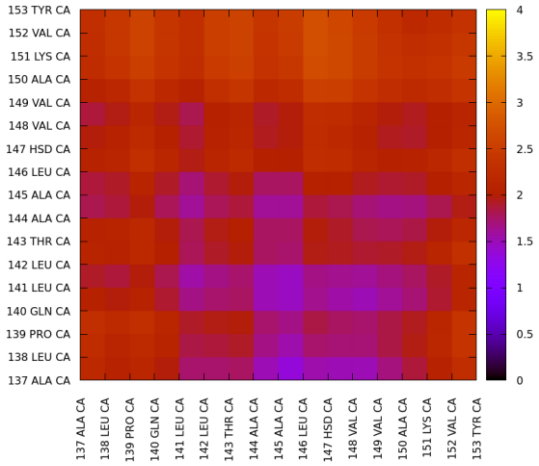


(b) Helices A2 (X axis) and A3 (Y axis)

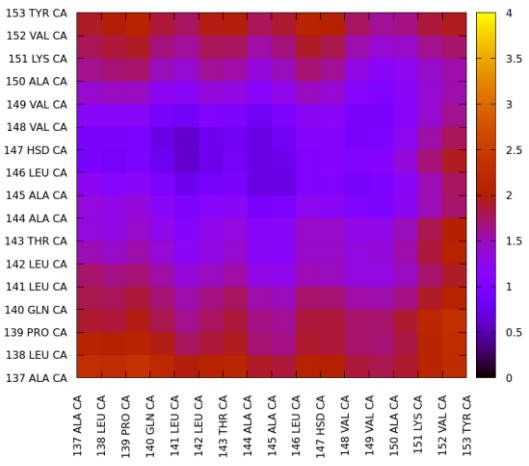


(c) Helices A3 (X axis) and A4 (Y axis)

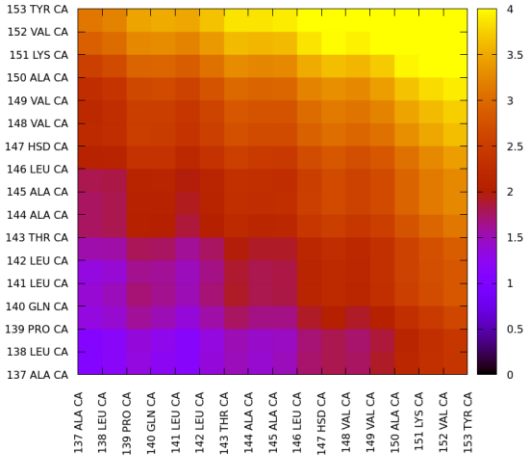
Figure S26: CHDM Maps for helices in layer A in fibrillation process of cross-α amyloid formed using eight copies of monomer of helix with residue numbers 137-153 (chain A) of peptide with PDB ID 1X9I. The colour axis represents distance in nm.



(a) Helices B1 (X axis) and B2 (Y axis)

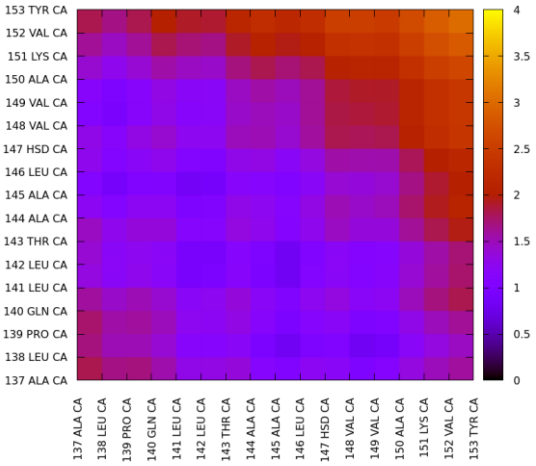


(b) Helices B2 (X axis) and B3 (Y axis)

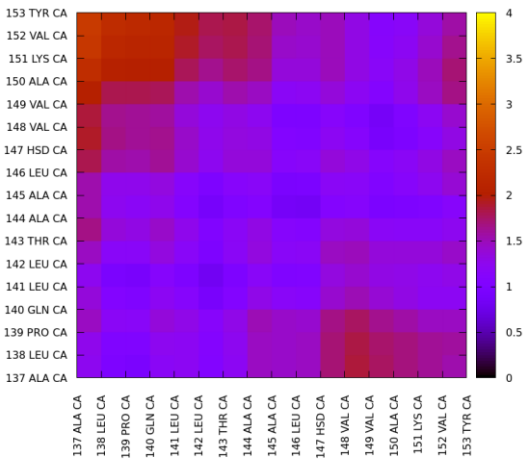


(c) Helices B3 (X axis) and B4 (Y axis)

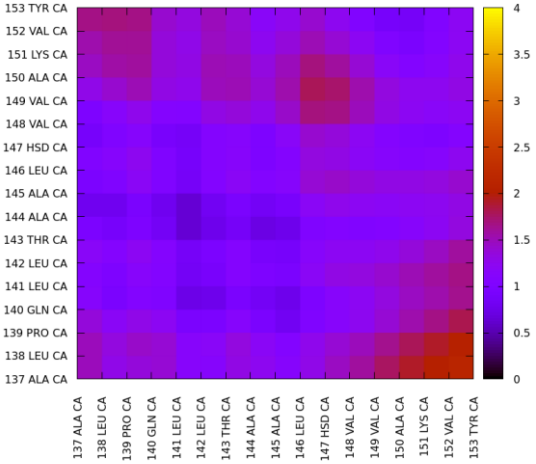
Figure S27: CHDM Maps for helices within layer B in fibrillation process of cross-α amyloid formed using eight copies of monomer of helix with residue numbers 137-153 (chain A) of peptide with PDB ID 1X9I. The colour axis represents distance in nm.



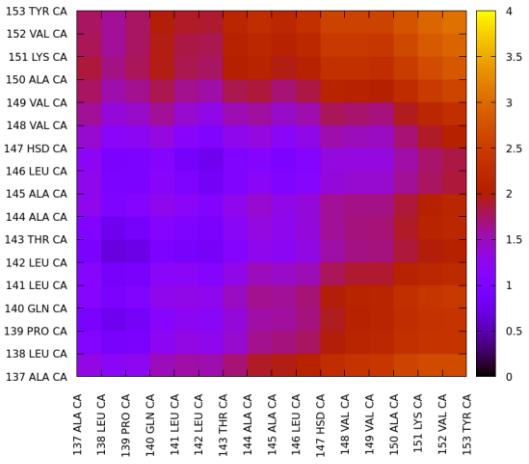
(a) Helices A1 (X axis) and B1 (Y axis)



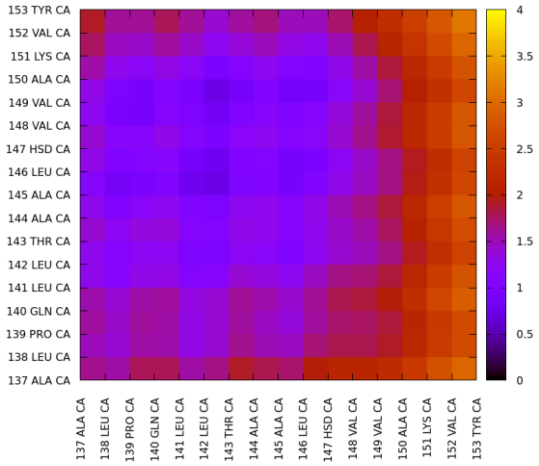
(b) Helices B1 (X axis) and A2 (Y axis)



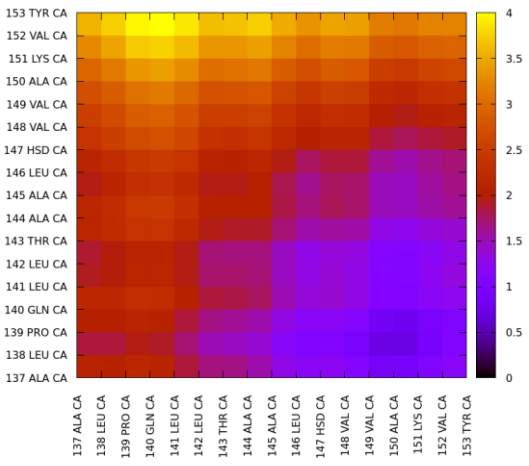
(c) Helices A2 (X axis) and B2 (Y axis)



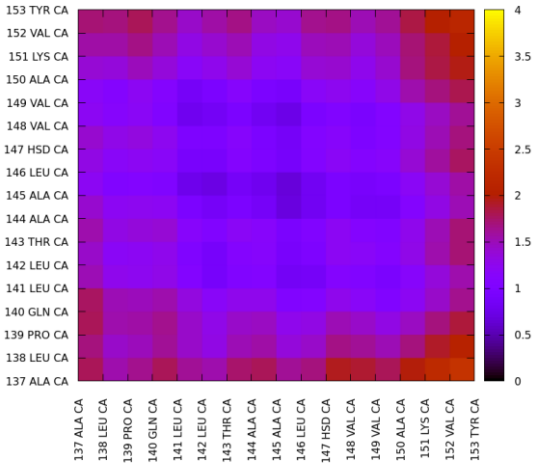
(d) Helices B2 (X axis) and A3 (Y axis)



(e) Helices A3 (X axis) and B3 (Y axis)

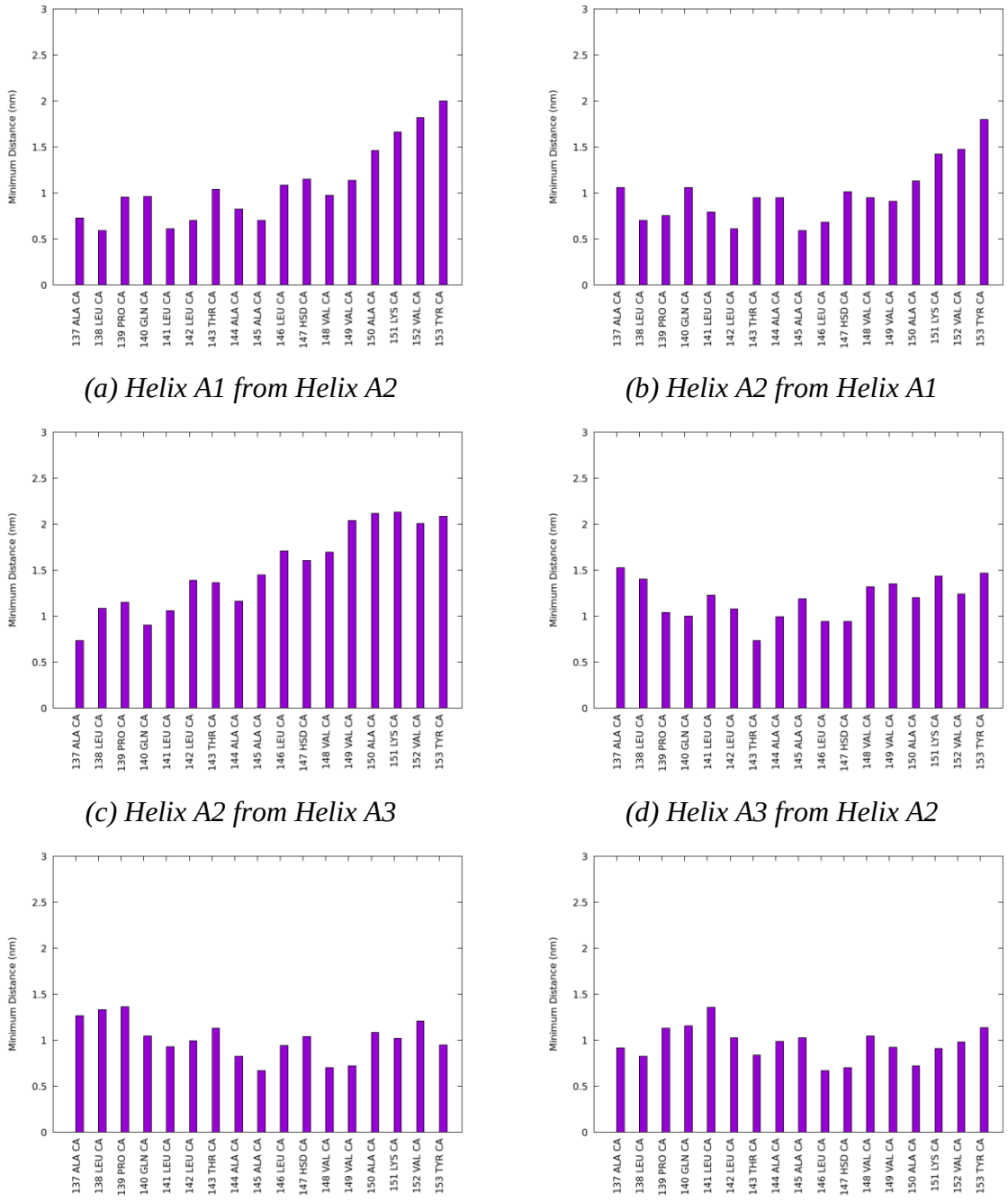


(f) Helices B3 (X axis) and A4 (Y axis)



(g) Helices A4 (X axis) and B4 (Y axis)

Figure S28: CHDM Maps for helices between the layers A and B in fibrillation process of cross-α amyloid formed using eight copies of monomer of helix with residue numbers 137-153 (chain A) of peptide with PDB ID 1X9I. The colour axis represents distance in nm.



(a) Helix A1 from Helix A2

(b) Helix A2 from Helix A1

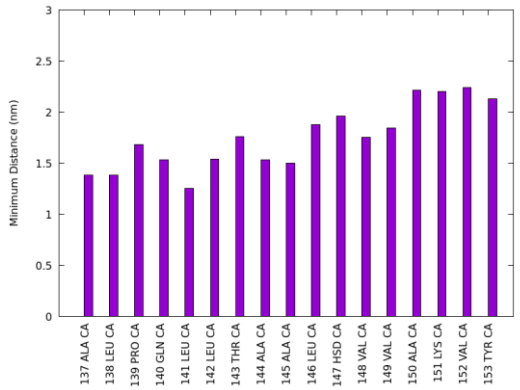
(c) Helix A2 from Helix A3

(d) Helix A3 from Helix A2

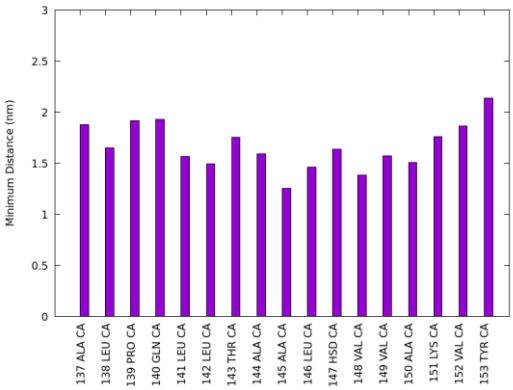
(e) Helix A3 from Helix A4

(f) Helix A4 from Helix A3

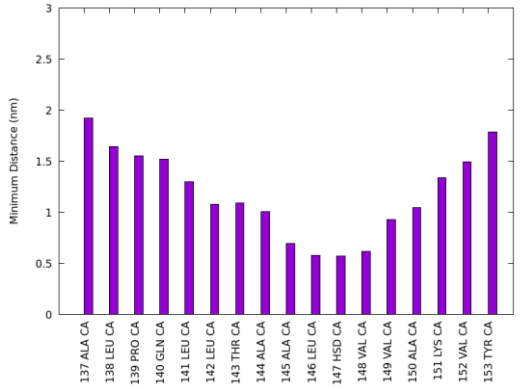
Figure S29: Minimum distance of residues of one helix from another helix, calculated using CHDM, for helices within layer A in fibrillation process of cross-α amyloid formed using eight copies of monomer of helix with residue numbers 137-153 (chain A) of peptide with PDB ID 1X9I. The colour axis represents distance in nm.



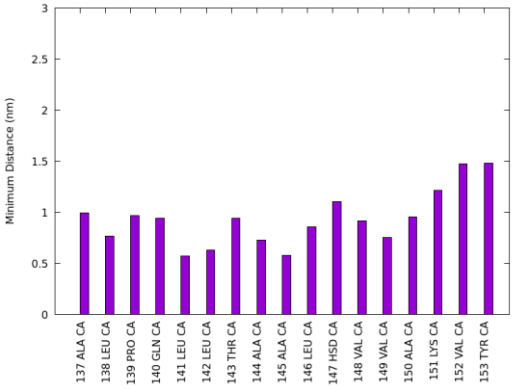
(a) Helix B1 from Helix B2



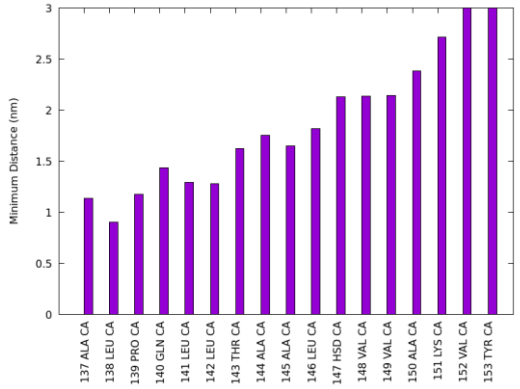
(b) Helix B2 from Helix B1



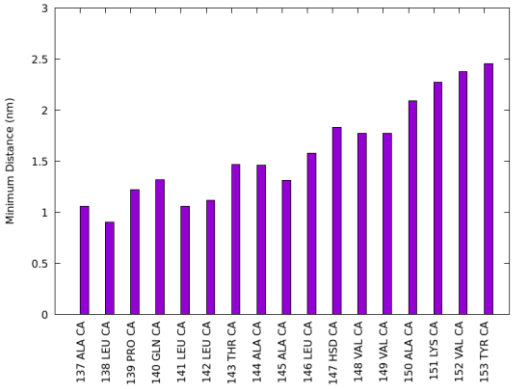
(c) Helix B2 from Helix B3



(d) Helix B3 from Helix B2

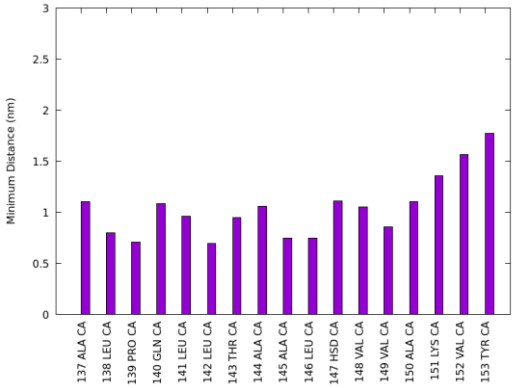


(e) Helix B3 from Helix B4

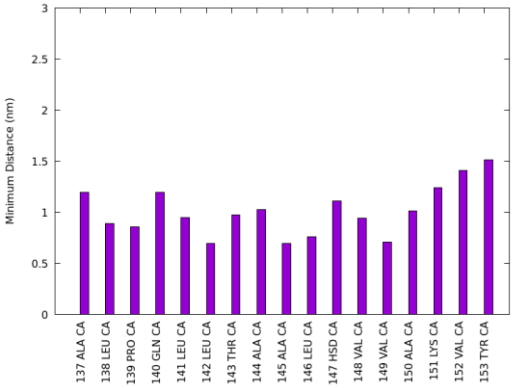


(f) Helix B4 from Helix B3

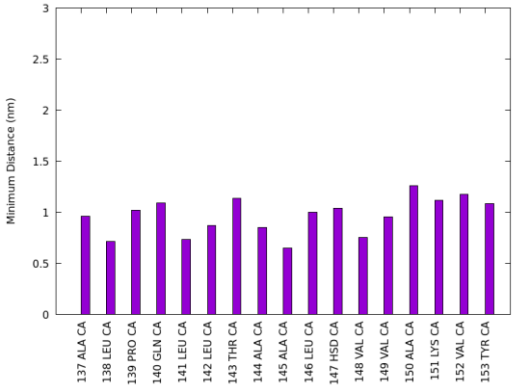
Figure S30: Minimum distance of residues of one helix from another helix, calculated using CHDM, for helices within layer B in fibrillation process of cross-α amyloid formed using eight copies of monomer of helix with residue numbers 137-153 (chain A) of peptide with PDB ID 1X9I. The colour axis represents distance in nm.



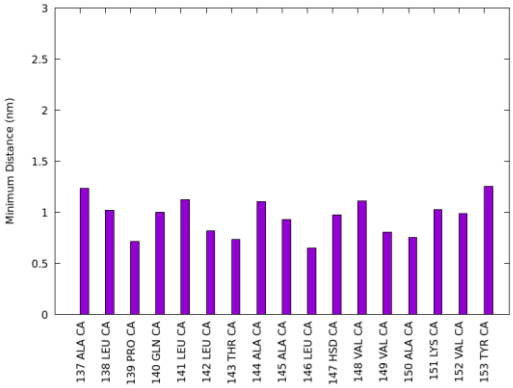
(a) Helix A1 from Helix B1



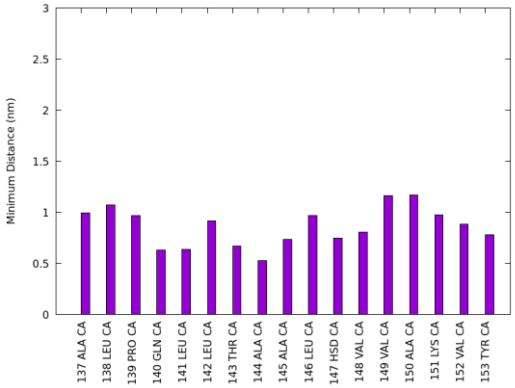
(b) Helix B1 from Helix A1



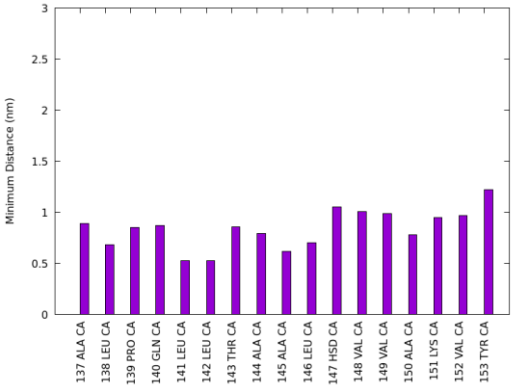
(c) Helix B1 from Helix A2



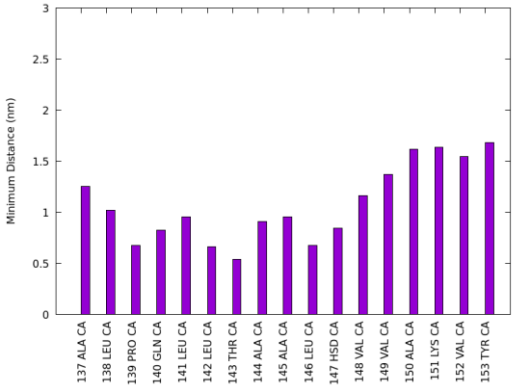
(d) Helix A2 from Helix B1



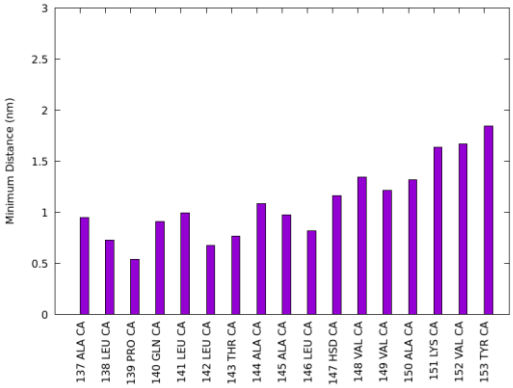
(e) Helix A2 from Helix B2



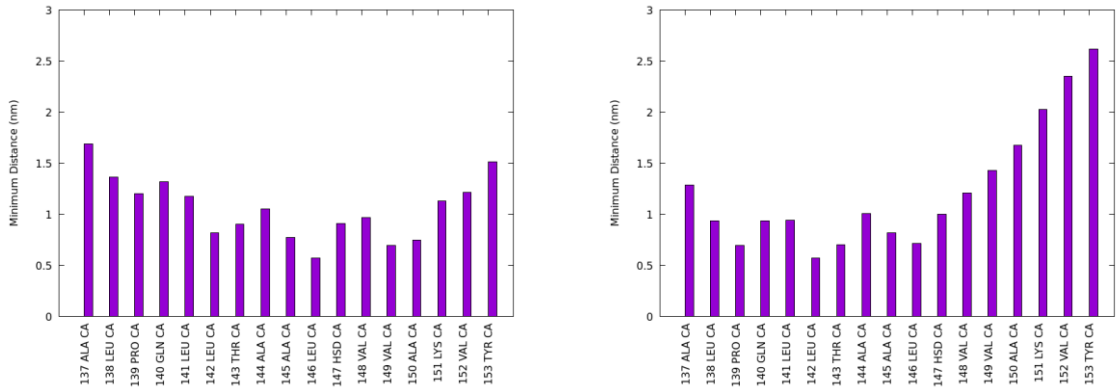
(f) Helix B2 from Helix A2



(g) Helix B2 from Helix A3

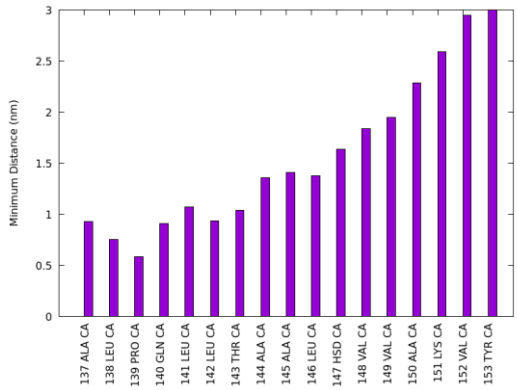


(h) Helix A3 from Helix B2

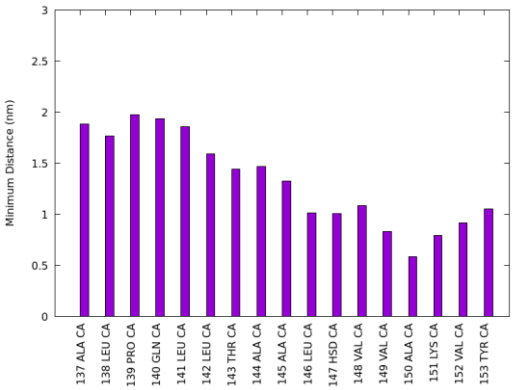


(i) Helix A3 from Helix B3

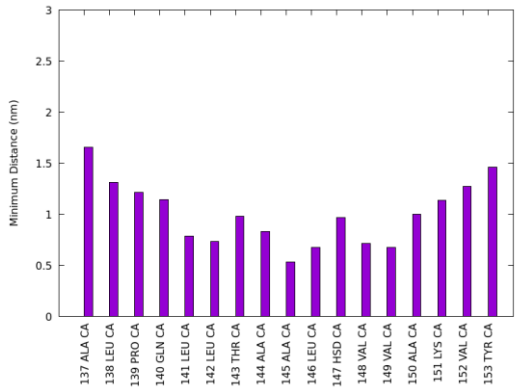
(j) Helix B3 from Helix A3



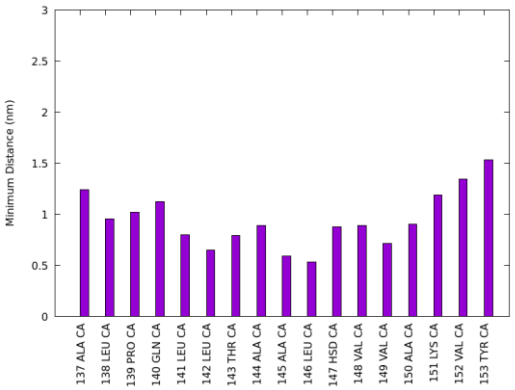
(k) Helix B3 from Helix A4



(l) Helix A4 from Helix B3

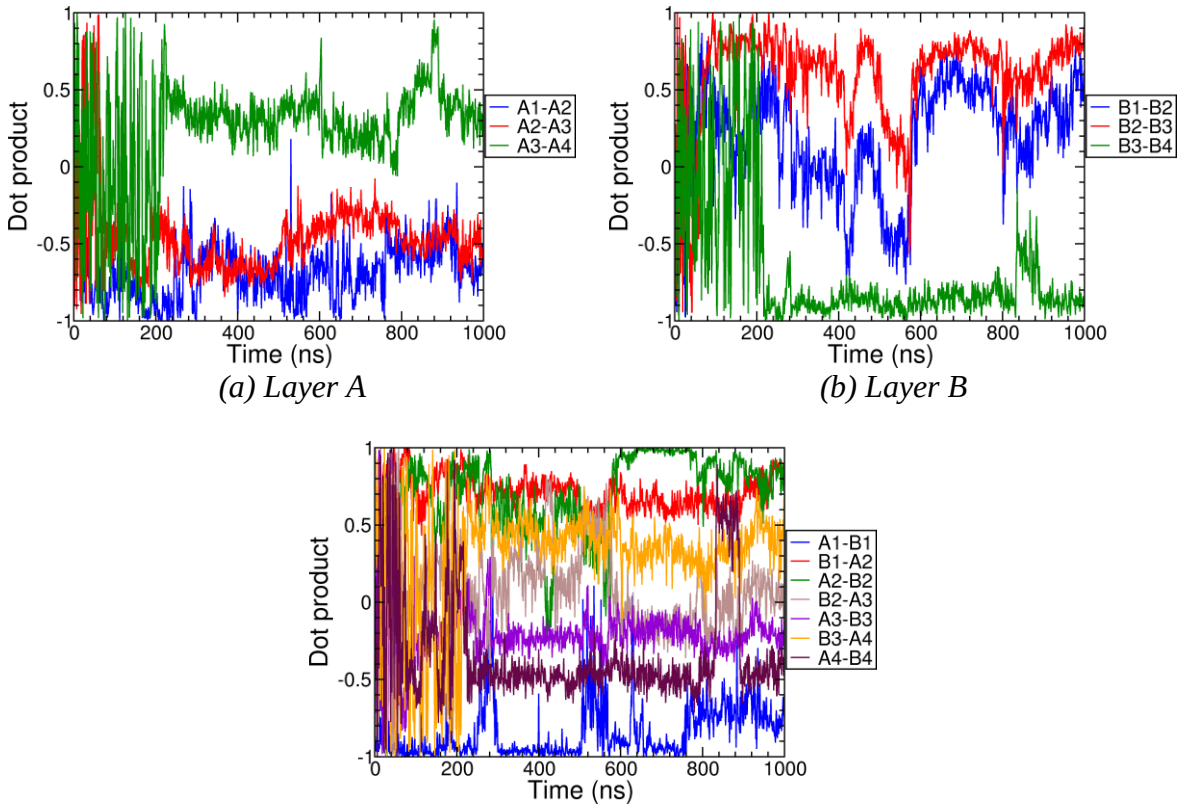


(m) Helix A4 from Helix B4



(n) Helix B4 from Helix A4

Figure S31: Minimum distance of residues of one helix from another helix, calculated using CHDM, for helices between the layers A and B in fibrillation process of cross-α amyloid formed using eight copies of monomer of helix with residue numbers 137-153 (chain A) of peptide with PDB ID 1X9I. The colour axis represents distance in nm.



(a) Layer A

(b) Layer B

(c) In between layers

Figure S32: Dot product between various α-helices in fibrillation process of cross-α amyloid formed using eight copies of monomer of helix with residue numbers 137-153 (chain A) of peptide with PDB ID 1X9I.

Table S2: Cross-alpha amyloid propensities of sequences from MetAmyl[1] satisfying the required helix length and minimum hydrophobic residues

No	Sequence	Length	Similarity	Propensity
1	RVQGVCFRMYTEDEAR	16	0.714	7.38
2	SKLEYSNFSIRY	12	0.600	5.24
3	MSSYAFFVQTCREEHK	16	0.500	7.07
4	ATQRLANFLVHSS	13	0.727	5.09
5	EVHHQKLVFFAEDVG	15	0.692	7.48
6	DWSFYLLYYTEFT	13	0.636	6.49
7	MLEGKVKWFNSEKGFGFIEVEG	22	0.700	11.72
8	GLSDGEWQQVLNVWGKVEADIAGHGQEVL	29	0.704	15.23
9	IKYLEFISDAIIHVLHSK	18	0.813	7.57
10	FSMLVAAIQSAGLTETLN	18	0.750	7.19
11	KTNMKHMAGAAAAGAVVGGLG	21	0.789	9.35
12	GYMLGSAMSRPIIHFGSDYED	21	0.737	13.37
13	DCVNITIKQHTVTTTT	16	0.571	9.28
14	DVKMMERVVEQMCITQY	17	0.733	7.57
15	GAARCQVTLRDLFDR	15	0.692	6.31
16	DRAVVLSHYIHNLSS	15	0.692	6.85
17	RYTHGRGFITKAINS	15	0.769	7.65
18	RSFFSFLGEAFD	12	1.000	5.67
19	AVVTGVTAVAQKTV	14	0.833	6.68
20	GKVQIINKKLDL	12	0.700	5.36

[1] M. Emily, A. Talvas and C. Delamarche, *PLoS One*, 2013, **8**, e79722.

Figure S33: Step by step screenshots for using Cross-Alpha-Det.