Seongwon Kim,[†] Wenling E. Chang,[†‡] Rashmi Kumar,[†] and Dmitri K. Klimov[†]

[†]Department of Bioinformatics and Computational Biology, George Mason University, Manassas, Virginia; and [‡]The MITRE Corporation, San Diego, California

Supplemental Materials

Implicit solvent model: The solvent accessible surface area (SASA) model has been developed by Caflisch and coworkers (see ref. [37] in the paper). In this model, solvent is treated implicitly and hydrophobic effect is accounted by solvation free energy, which scales linearly with accessible surface area of atoms. The combination of SASA model and CHARMM19 united atom force field has been used to fold α-helix and β-sheet polypeptides [1,2] and for studying aggregation of amyloidogenic peptides (ref. [23] in the paper). Earlier we have used CHARMM19+SASA model to probe the thermodynamics of Aβ fibril growth and have obtained the results consistent with experimental data, including the value of dissociation temperature and the deposition mechanism (ref. [26] in the paper). Because the number of degrees of freedom in CHARMM19+SASA model is reduced, the simulations are several orders of magnitude faster than those employing explicit solvent. Consequently, CHARMM19+SASA model allows one to collect exhaustive sampling of conformational space that cannot be usually achieved with explicit solvent models.

Tests of the naproxen force field have indicated that the conformational ensembles of naproxen computed in our simulations and determined by *ab initio* methods and NMR technique are consistent (ref. [34] in the paper). Two other findings support the use of CHARMM19+SASA model for naproxen binding. First, in our previous studies (ref. [34] in the paper) we compared the free energy ΔF_b of naproxen and ibuprofen binding to Aβ fibril. We found that ΔF_b is -7.6RT for naproxen and -5.2RT for ibuprofen at 360K. Stronger affinity of naproxen compared to ibuprofen for binding to Aβ fibril has also been reported experimentally (ref. [19] in the paper). Second, our previous implicit solvent simulations (ref. [34] in the paper) revealed the mechanism of naproxen and ibuprofen binding to amyloid fibril, which is very similar to that deduced from explicit water simulations of structurally similar thioflavin-T (ref. [52] in the paper). In both mechanisms, fibril surface geometry is a dominant binding factor and grooves are the primary binding locations. Incidentally, similar binding pattern has been observed for biomarker ^{18}FFDDNP, which also bears structural similarity to naproxen [3].

Selection of Aβ species: Our simulations were performed using the N-terminal truncated fragment of the full-length peptide, Aβ10-40 (Fig. 1b). Several observations suggest similarities in the aggregation propensities of Aβ1-40 and Aβ10-40. First, solid-state NMR studies have shown that both peptides form similar two-fold symmetry fibril structures ([4] and ref. [48] from the paper). In the Aβ1-40 fibril the first nine N-terminal residues are disordered [5,6]. In addition, Aβ1-40 fibrils were reported to seed the growth of Aβ10-40 [4]. Second, according to the experiments [7] and simulations [8] truncation of the first nine N-terminal amino acids leads to minor changes in the conformational ensembles of Aβ monomers and oligomers. These observations constitute a rationale for using Aβ10-40 as a model of the full-length Aβ1-40 peptide.

Computation of structural probes: To assign contacts formed between naproxen and Aβ, three structural groups G1, G2, G3 in the ligand are distinguished (Fig. 1a). If the

distance between the centers of mass of a side chain and one of these groups is less than 6.5 Å, a contact is formed. A ligand is bound, if it forms at least one contact with Aβ side chain. Ligand is assumed to occur at the dimer aggregation interface, if it forms contacts with two peptides simultaneously.

To analyze interpeptide interactions in Aβ dimer, we computed the radial probability distribution function $g_{pp}(r)$, where r is a minimum distance between amino acids from different peptides. Using similar definition we computed the radial probability distribution function $g_{lp}(r)$, which probed binding of naproxen ligands to Aβ side chains. The functions $g_{pp}(r)$ and $g_{lp}(r)$ were constructed using the bin size of 0.5 Å.

Convergence of REMD simulations: The convergence of REMD sampling of Aβ dimer was checked using the number N_s of unique states (E_{eff},C) sampled at least once in the course of simulations. Each state (E_{eff},C) is defined by the effective energy of the entire simulation system, E_{eff}, which includes the potential and solvation energies, and by the number of interpeptide side chain contacts in the dimer, C. Fig. S1 shows N_s as a function of the cumulative equilibrium simulation time τ_{sim}. At $\tau_{sim} > 20$ μs N_s starts to level off and becomes almost constant after 200 μs. Similar results were obtained, if the convergence of REMD is probed using the states (E_{eff},C_h), where C_h is the number of interpeptide hydrophobic side chain contacts, or the states (E_{eff},N_{ihb}), where N_{ihb} is the number of intrapeptide hydrogen bonds. The number $N_{s,i}$ of unique states (E_{eff},N_{ihb}) is shown in Fig. S1. The behavior of N_s and $N_{s,i}$ in Fig. S1 suggests that sampling of Aβ dimer conformational ensemble becomes exhaustive and approximate convergence of REMD simulations takes place.

The convergence of REMD sampling of Aβ-naproxen interactions was probed in a similar way using the number $N_{s,l}$ of the unique states (E_{eff} ,L), where L is the number of ligands bound Aβ dimer. The behavior of $N_{s,l}$ as a function of the cumulative equilibrium simulation time τ_{sim} is analogous to that of N_s or $N_{s,i}$ (Fig. S1).

To further test the reliability of REMD sampling the simulation data were divided into two equal subsets and analyzed independently at 360K. All thermodynamic quantities from the two subsets related to inter- or intrapeptide interactions have differed by no more than 4%. For example, the fractions of residues in β-strand and helix conformations differed by 1 and 3%, respectively. All quantities related to naproxen-Aβ interactions have the errors below 1%.

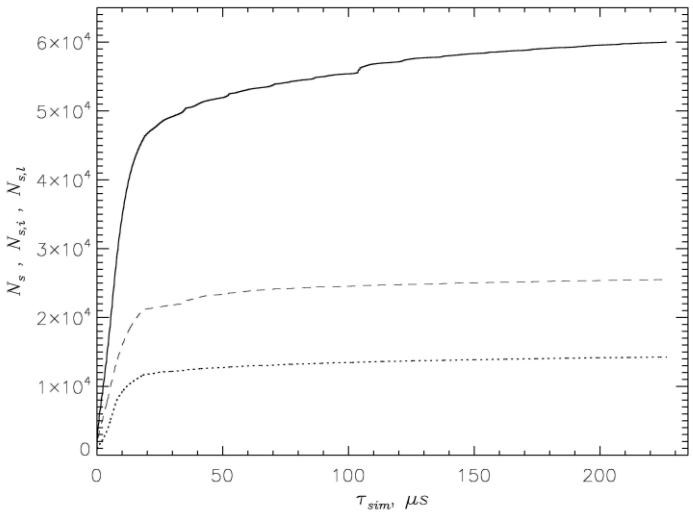

Fig. S1 The number of unique states N_s probing interpeptide interactions in Aβ dimer in the course of REMD simulations as a function of cumulative equilibrium simulation time τ_{sim}(solid line). The number of unique states $N_{s,i}$ probing intrapeptide interactions is shown by dashed line. Dotted line represents the number of unique states $N_{s,l}$ sampling interactions between Aβ dimer and naproxen ligands.

Testing naproxen binding mechanism with AutoDock: Analysis of naproxen binding to Aβ using a software package AutoDock4.2 (ref. [47] in the paper) provides an independent test of binding mechanism proposed on the basis of CHARMM19+SASA force field. To this end we designed the following procedure. A drawback of typical docking simulations is that protein backbone structure is assumed rigid. To address this issue we utilized the ensemble of Aβ monomer conformations in naproxen solution produced by REMD implicit solvent simulations in our previous studies (ref. [35] in the paper). We have randomly selected 640 Aβ structures equilibrated at 330K and passed them as initial conditions to AutoDock simulations. Although this conformational ensemble was initially produced using CHARMM19+SASA model, AutoDock applies its own force field to examine binding of ligands to Aβ. (AutoDock force field is parameterized using known structures and inhibition constants for protein-ligand complexes and is unrelated to CHARMM.) To compute the probabilities of naproxen binding to amino acids in Aβ sequence, we applied the Lamarckian Genetic Algorithm (LGA), which generates naproxen bound conformations. In these computations naproxen structure was treated as flexible. For each Aβ conformation out of 640 we obtained 10 lowest free energy states for bound naproxen resulting in the total of 6400 bound conformers. (We checked that increasing the number of Aβ structures in the ensemble does not change qualitatively binding results. To check the results of LGA, we used simulated annealing (SA) search in AutoDock. The ensemble of bound states generated with SA at 330K was very similar to that obtained with LGA.) In AutoDock simulations naproxen is assumed bound to amino acid, if the distance between any pair of atoms from the ligand and side chain is less than 6.5Å.

Using bound conformers we computed the probability $P_b(i)$ of naproxen binding to amino acid i in Aβ monomer. Fig. S2 compares $P_b(i)$ with the number of contacts $<C_l(i)>$ formed by the side chain of amino acid i in Aβ dimer with naproxen ligands. It follows from this figure that variations in both quantities along Aβ sequence are consistent as the correlation factor computed between $P_b(i)$ and $<C_l(i)>$ is 0.66. Both AutoDock and CHARMM19+SASA simulations reveal that the affinity of the Aβ N-terminal for binding naproxen is higher than of the C-terminal (see Results section in the paper). To further compare naproxen binding, we select 10 residues, which form the largest number of contacts with naproxen in each of the two force fields. According to CHARMM19+SASA model these are (in the order of descending $<C_l(i)>$) Glu11, Tyr10, His13, Phe20, Lys16, Phe19, His14, Val12, Asp23, and Lys28. The list computed based on AutoDock results includes (in the order of descending $P_b(i)$) Phe20, Lys28, Lys16, Ile32, Phe19, Glu11, Asn27, Gln15, Tyr10, and His14. Although there are differences in binding affinities of individual amino acids in the force fields, seven residues out of 10 are present in both lists. Taken together, it appears that binding propensities along Aβ sequence are similar in CHARMM19+SASA and AutoDock simulations, which utilize unrelated force fields. This analysis supports the notion that CHARMM19+SASA model captures the mechanism of naproxen binding.

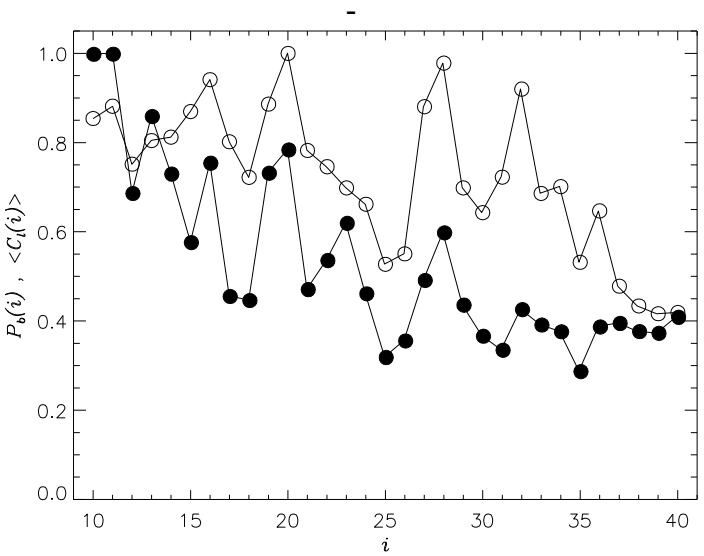

Fig. S2 Probabilities $P_b(i)$ of naproxen binding to residues i in Aβ10-40 monomer obtained using AutoDock (open circles). Numbers of contacts with naproxen, $<C_l(i)>$, formed by amino acids i in Aβ10–40 dimer are shown by filled circles. The distribution $<C_l(i)>$ is computed at 330K using REMD simulations employing CHARMM19+SASA force field. Both quantities, $P_b(i)$ and $<C_l(i)>$, are normalized.

Distribution of amino acids in the dimer: To probe the composition of dimer core we have computed the probability of occurrence of amino acid i in the core, $P_c(i)$ (Fig. S3, see paper text for details).

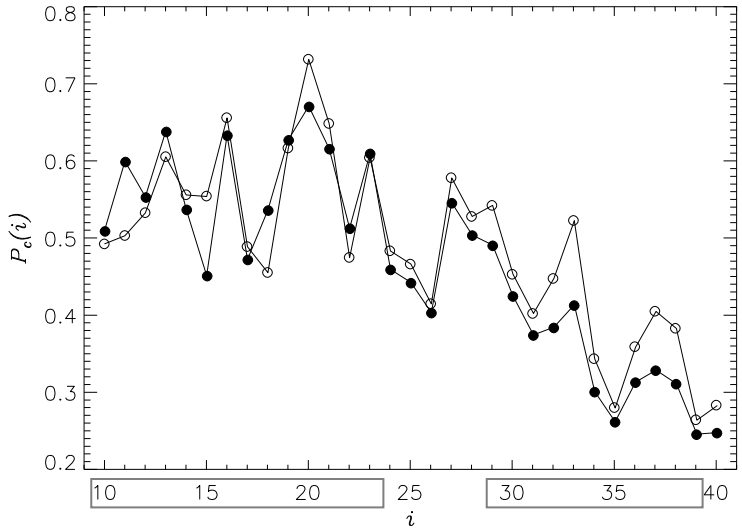

Fig. S3 Probabilities of finding amino acids i in the dimer core, $P_c(i)$, computed at 360K: naproxen solution (filled circles, the core radius r_c=12Å), water (open circles, data from ref. [49] in the paper). The plot shows that the N-terminal Nt tends to bury in the dimer core, while the C-terminal Ct is usually exposed to the solvent. The figure also demonstrates that naproxen does not change significantly the distribution of amino acids in the dimer volume. N- and C-terminals are boxed.

Fig. S3 suggests that Abeta dimer aggregation interface is largely formed by the Nt terminal. Two factors are likely to contribute to this observation. First, the Ct contains four Gly residues (Fig. 1b), whereas Nt has none. Gly residues enhance local fluctuations in backbone structure resulting in larger entropic penalty for confining the Ct to the dimer core than the Nt (refs. [38,49] in the paper). Second, interpeptide interactions in the Nt are promoted by aromatic Phe residues. In Fig. S3 Phe20 has the largest probability (>0.70, data for water) among all residues to occur in the dimer core. The importance of aromatic residues for amyloid formation was noted in the previous studies [9,10].

Distributions of ligands bound to Aβ dimer: To analyze the mechanism of naproxen binding to Aβ dimer the distribution of bound ligands $<L(S_c)>$ was computed (Fig. S4). At 330K almost all naproxen molecules ($<L_c>\approx$19.2 out of 20) participate in the clusters bound to Aβ dimer, of which 17.7 (or 92%) are included in large clusters (S_c>6).

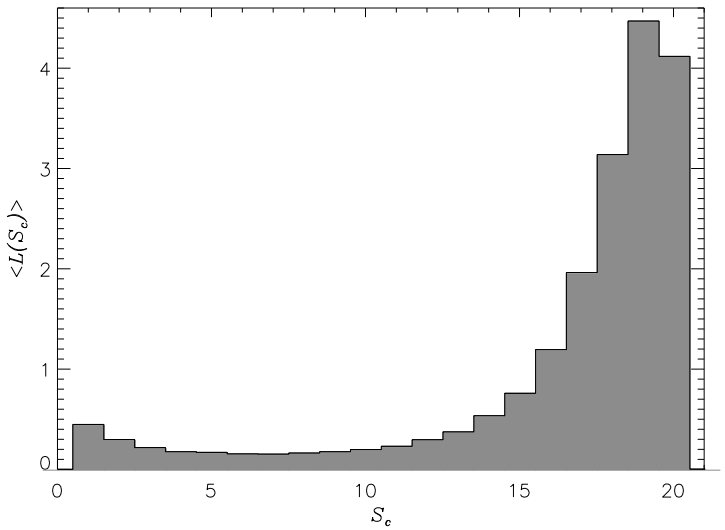

Fig. S4 Distribution of the numbers of naproxen ligands $<L(S_c)>$ with respect to bound cluster size S_c. The plot demonstrates that formation of large bound clusters is strongly preferred.

References

[1] Ferrara, P, Caflisch, A (2000) Folding simulations of a three-stranded antiparallel beta-sheet peptide. Proc. Natl. Acad. Sci. USA 97:10780-10785.

[2] Hiltpold, A, Ferrara, P, Gsponer, J, Caflisch, A (2000) Free energy surface of the helical Peptide Y(MEARA)6. J. Phys. Chem. B 104:10080-10086.

[3] Agdeppa, E, Kepe, V, Liu, J, Flores-Torres, S, Satyamurthy, N, Petric, A, Cole, GM, Small, GW, Huang, S-C, Barrio, JR (2001) Binding characteristics of radiofluorinated 6-dialkylamino-2-naphthylethylidene derivatives as positron emission tomography imaging probes for β-amyloid plaques in Alzheimer's disease. J Neurosc 21:1–5.

[4] Paravastu, AK, Petkova, AT, Tycko, R (2006) Polymorphic fibril formation by residues 10-40 of the Alzheimer's β-Amyloid Peptide. Biophys. J. 90:4618-4629.

[5] Paravastua, AK, Leapman, RD, Yau, W-M, Tycko, R. (2008) Molecular structural basis for polymorphism in Alzheimers β-amyloid fibrils. Proc. Natl. Acad. Sci. USA 105:18349–18354.

[6] O'Nuallain, B, Shivaprasad, S, Kheterpal, I, Wetzel, R. (2005) Hydrogen-deuterium (H/D) exchange mapping of Aβ1-40 amyloid fibril secondary structure using nuclear magnetic resonance spectroscopy. Biochemistry 44:4434-4441.

[7] Bitan, G, Vollers, SS, Teplow, DB (2003) Elucidation of primary structure elements controlling early amyloid beta-protein oligomerization. J. Biol. Chem. 278:34882-34889.

[8] Takeda, T, Klimov, DK (2009) Probing the effect of amino-terminal truncation for Abeta1-40 Peptides. J. Phys. Chem. B 113: 6692-6702.

[9] Reches, M., Gazit, E. (2004) Formation of closed-cage nanostructures by self-assembly of aromatic dipeptides. Nano Lett. 4: 581–585

[10] Wurth, C, Guimard, NK, Hecht, MH (2002) Mutations that reduce aggregation of the Alzheimer's Aβ42 Peptide: an unbiased search for the sequence determinants of Aβ amyloidogenesis. J. Mol. Biol. 319: 1279-1290.