Supporting Information

Chong and Ham 10.1073/pnas.1120646109

SI Text

Molecular Dynamics (MD) Simulation Method and Calculation of Collision Cross-Section. The dimerization simulations of 42-residue amyloid-β (Aβ42) proteins were performed for 100 ns at 300 K and 1 bar under neutral pH with SANDER module of AMBER9 program package (1) using the ff99 force filed (2). Each Aβ42 monomer has a sequence of $^{(1)}$DAEFR HDSGY EVHHQ KLVFF AEDVG SNKGA IIGLM VGGVV IA$^{(42)}$. Two monomers, each having the unfolded structure in water studied in ref. 3, were initially placed at 45 Å apart from each other with a random orientation, and no artificial attraction force was employed between them. Two monomers were explicitly solvated with 24,708 transferable intermolecular potential 3 point (TIP3P) water molecules (4) in the rectangular box with 20 Å buffer, and periodic boundary condition was applied. Six Na^+ counter ions were added to neutralize the system. The particle mesh Ewald method (5) was applied for treating long-range electrostatic interactions, whereas a 10 Å cutoff was used for the short-range nonbonded interactions. The hydrogen atoms were constrained to the equilibrium bond length using the SHAKE algorithm (6). To remove unfavorable van der Waals contacts, the system was initially subjected to 500 steps of steepest decent minimization followed by 500 steps of conjugate gradient minimization while each monomer was constrained by 500 kcal/(mol Å^2) harmonic potential. Then, the whole system was minimized using 1,000 steps of steepest decent minimization followed by 1,500 steps of conjugate gradient minimization without harmonic restraints. The system was subsequently subjected to 20 ps equilibration process in which the temperature was gradually raised from 0 to 300 K. After the equilibration step, the production run was carried out for 100 ns with 2 fs time step and with NPT ensemble, i.e., a constant number of particles (N), pressure (P), and temperature (T). Temperature and pressure were controlled by Berendsen's thermostat and barostat with coupling constants of 1.0 and 2.0 ps, respectively (7). Three independent dimerization simulations were performed with different random initial relative orientations and velocities.

The Aβ42 dimerization process was monitored via the center-of-mass distance between two monomers, the number of intermonomer heavy atom contacts, and the collision cross-section. The heavy atom contact is counted when the distance between two heavy atoms belonging to different monomers is less than 5.4 Å. The collision cross-sections to be compared with the ion-mobility mass spectrometry measurements were calculated following the procedure described in refs. 8 and 9. To better correlate with the solvent-free experiments on samples electrosprayed from solution phase, protein structure from simulations was instantaneously dehydrated through energy minimization in vacuum (500,000 steps). The collision cross-section of the dehydrated structure was then calculated using the trajectory method implemented in the MOBCAL software (10).

Structural Comparison with the Previous Aβ Dimer Studies. Up to now, related computational studies have been performed for the dimer formation of Aβ fragments (11–13) and for the full-length Aβ42 dimerization in a continuum solvent (14, 15). We found a common structural feature—a salt-bridge formation between Glu11 in one monomer and Lys28 in the other—observed in the dimer formation of Aβ(10–35) fragments (12). The presence of such a salt bridge was the characteristic feature of the Aβ(10–35) dimer generated by a docking protocol that emphasizes the intermonomer electrostatic interaction, which, however, was found to have a short lifetime and did not contribute to the stability of this dimer (12). The intermonomer Glu11-Lys28 salt-bridge formation observed in our simulation was also quite transient and did not contribute to stabilize the Aβ42 dimer conformation as can be inferred from Fig. 6*A* in the main text. On the other hand, we did not observe further noteworthy common structural aspects with the previous studies, in particular, on the full-length Aβ42 dimer based on the discrete molecular dynamics simulations with a coarse-grained model for protein (14) and on the Monte Carlo simulations with implicit water and an effective potential for protein (15). This is possibly because of the differences in the solvation model and in the force fields employed in the simulations and of the intrinsically disordered nature of Aβ42 protein. In fact, the applicability of the force fields to intrinsically disordered proteins such as Aβ42 protein is one of the recent topics (16) because those proteins have not been designed to fold in a cooperative fashion, and hence, their structures are considered to be very susceptible to small differences in the force fields. The intrinsically disordered nature of Aβ42 protein has also made it difficult to determine its atomic-resolution structure by traditional methods such as X-ray crystallography and solution NMR. To the best of our knowledge, the collision cross-sections measured by the ion-mobility mass spectrometry (17) are the only structural characteristics of Aβ42 oligomers including dimers currently available from experiments. The agreement of the Aβ42 dimer structures from our simulations with the ion-mobility mass spectrometry measurement as discussed in the main text suggests the relevance of our simulations to experiment.

Solvation Thermodynamics Based on the Integral-Equation Theory of Liquids. For each Aβ42 dimer conformation generated by the MD simulations, we applied the three-dimensional reference interaction site model (3D-RISM) theory (18, 19) to calculate thermodynamic functions of solvation. The 3D-RISM theory is an integral-equation theory based on statistical mechanics for obtaining the 3D distribution function $g_\gamma(\mathbf{r})$ of the site γ, oxygen or hydrogen, of water at position $\mathbf{r}$ around a molecular solute such as protein. For a solute–solvent system at infinite dilution, the 3D-RISM equation is given by

$$h_\gamma(\mathbf{r}) = \sum_{\gamma'} c_{\gamma'}(\mathbf{r}) * [w^{vv}_{\gamma'\gamma}(r) + \rho h^{vv}_{\gamma'\gamma}(r)]. \qquad \text{[S1]}$$

Here $h_\gamma(\mathbf{r})$ and $c_\gamma(\mathbf{r})$ refer to the 3D total and direct correlation functions of the water site γ, respectively; the asterisk denotes a convolution integral; $w^{vv}_{\gamma'\gamma}(r)$ and $h^{vv}_{\gamma'\gamma}(r)$ are the site–site intramolecular and total correlation functions of water; and ρ represents the average number density of water. This equation is to be supplemented by an approximate closure relation, and in the present study we adopted the one suggested by Kovalenko and Hirata (18)

$$h_\gamma(\mathbf{r}) = \begin{cases} \exp[d_\gamma(\mathbf{r})] - 1 & \text{for } d_\gamma(\mathbf{r}) \le 0, \\ d_\gamma(\mathbf{r}) & \text{for } d_\gamma(\mathbf{r}) > 0, \end{cases} \qquad \text{[S2]}$$

in which $d_\gamma(\mathbf{r}) = -u_\gamma(\mathbf{r})/(k_B T) + h_\gamma(\mathbf{r}) - c_\gamma(\mathbf{r})$ with k_B denoting Boltzmann's constant. $u_\gamma(\mathbf{r})$ refers to the interaction potential acting on the water site γ that is generated by atoms in protein and is represented by a sum of radially symmetric Lennard–Jones (LJ) and Coulomb electrostatic terms centered on the protein interaction site α of position $\mathbf{r}_\alpha$, $u_\gamma(\mathbf{r}) = \sum_\alpha [u^{(LJ)}_{\alpha\gamma}(|\mathbf{r} - \mathbf{r}_\alpha|) + u^{(elec)}_{\alpha\gamma}(|\mathbf{r} - \mathbf{r}_\alpha|)]$. Here $u^{(LJ)}_{\alpha\gamma}(r) = 4\epsilon_{\alpha\gamma}[(\sigma_{\alpha\gamma}/r)^{12} - (\sigma_{\alpha\gamma}/r)^6]$ and

$u_{\alpha\gamma}^{(\mathrm{elec})}(r) = q_\alpha q_\gamma / r$ with $\epsilon_{\alpha\gamma}$, $\sigma_{\alpha\gamma}$, q_α, and q_γ being the LJ parameters and atomic charges.

The 3D-RISM calculation for $g_\gamma(\mathbf{r})$ was performed as follows. For each Aβ42 dimer conformation generated by MD simulations, one can determine the interaction potential $u_\gamma(\mathbf{r})$. Based on the knowledge of $u_\gamma(\mathbf{r})$, the two unknown functions $h_\gamma(\mathbf{r})$ and $c_\gamma(\mathbf{r})$ can be determined by solving Eqs. **S1** and **S2** self-consistently, and the 3D water distribution function is obtained via $g_\gamma(\mathbf{r}) = h_\gamma(\mathbf{r}) + 1$. We used the dielectrically consistent RISM theory (20) for the site–site correlation functions $w_{\gamma'\gamma}^{vv}(r)$ and $h_{\gamma'\gamma}^{vv}(r)$ of water determined at $T = 300$ K and $\rho = 1$ g/cm^3 and with the dielectric constant of 78.4. Technical details concerning the 3D-RISM calculation can be found in ref. 18.

Thermodynamic functions of solvation can be obtained based on the water distribution function. For the solvation free energy $\Delta\mu$, the following analytical expression is available under the use of the Kovalenko–Hirata closure given in Eq. **S2** (18):

$$\Delta\mu = \rho k_\mathrm{B} T \sum_\gamma \int d\mathbf{r} \left[\frac{1}{2} h_\gamma(\mathbf{r})^2 \Theta(-h_\gamma(\mathbf{r})) - c_\gamma(\mathbf{r}) - \frac{1}{2} h_\gamma(\mathbf{r}) c_\gamma(\mathbf{r}) \right]. \qquad \textbf{[S3]}$$

Here $\Theta(x)$ is the Heaviside step function.

Although the solvation free energy does not depend on whether the solute insertion is done under the isochoric (constant volume) or isobaric (constant pressure) condition, its energetic and entropic components depend on such a condition (21). Under the isochoric condition, the solvation free energy $\Delta\mu$ comprises the solvation energy $\Delta\epsilon_V$ and isochoric solvation entropy Δs_V. The latter is given by the temperature derivative of $\Delta\mu$ at constant density,

$$\Delta s_V = -\left(\frac{\partial \Delta\mu}{\partial T} \right)_\rho. \qquad \textbf{[S4]}$$

In the present work, the temperature derivative was calculated numerically using the first order finite difference with $\Delta T = 2$ K. The solvation energy can then be obtained from

$$\Delta\epsilon_V = \Delta\mu + T\Delta s_V. \qquad \textbf{[S5]}$$

Under the isobaric condition, which is more relevant to the present study because the MD simulations were performed at constant pressure, the solvation free energy consists of the solvation enthalpy Δh and isobaric solvation entropy Δs. The relations between the quantities under isochoric and isobaric conditions are given by (21)

$$\Delta h = \Delta\epsilon_V + \frac{T\alpha_P}{\kappa_T} V_u, \qquad T\Delta s = T\Delta s_V + \frac{T\alpha_P}{\kappa_T} V_u \qquad \textbf{[S6]}$$

in terms of the isobaric thermal expansion coefficient α_P and the isothermal compressibility κ_T of the solvent, and the partial molar volume V_u of the solute.

Using the experimental values for α_P and κ_T for water at $T = 300$ K and $P = 1$ bar, the term $T(\alpha_P/\kappa_T)V_u$ in Eq. **S6** can be estimated as $0.041 \times V_u$ kcal/mol when V_u is measured in cm^3/mol. The partial molar volume can be obtained in terms of the 3D direct correlation function via $V_u = k_\mathrm{B} T \kappa_T [1 - \rho \sum_\gamma \int d\mathbf{r} c_\gamma(\mathbf{r})]$ (22). We confirmed that the contribution from the term $T(\alpha_P/\kappa_T)V_u$ is practically negligible as far as the changes in Δh vs. $\Delta\epsilon_V$ and $T\Delta s$ vs. $T\Delta s_V$ are concerned.

The main limitation of the 3D-RISM theory lies in the use of an approximate closure relation such as Eq. **S2**, which is inherent in all the integral-equation theories. In particular, the absolute value of the solvation free energy depends on the closure relation used (18, 19). However, it is known that relative values of the solvation free energies are reasonably accurate (18). We note in this connection that only relative values of thermodynamic functions matter in the present study. It is therefore expected that our results in the main text do not significantly suffer from the limitation of the integral-equation theory.

Decomposition Method of the Solvation Thermodynamic Functions. The nonelectrostatic and electrostatic contributions to the solvation free energy $\Delta\mu$, solvation enthalpy Δh, and solvation entropy Δs can be obtained as follows. First, we calculate $\Delta\mu$, Δh, and Δs with the full protein–water interaction. Next, we repeat this calculation with the electrostatic protein–water interaction turned off, $u_{\alpha\gamma}^{(\mathrm{elec})}(r) = 0$, which yields the nonelectrostatic contributions to be denoted as $\Delta\mu^{(\mathrm{LJ})}$, $\Delta h^{(\mathrm{LJ})}$, and $\Delta s^{(\mathrm{LJ})}$. The electrostatic contributions can then be obtained by subtraction: $\Delta\mu^{(\mathrm{elec})} = \Delta\mu - \Delta\mu^{(\mathrm{LJ})}$, $\Delta h^{(\mathrm{elec})} = \Delta h - \Delta h^{(\mathrm{LJ})}$, and $\Delta s^{(\mathrm{elec})} = \Delta s - \Delta s^{(\mathrm{LJ})}$.

Further decomposition of the solvation thermodynamic quantities into atomic contributions can be carried out using the exact partitioning method developed in ref. 23 that is based on the Kirkwood charging formula. The atomic decomposition of the solvation free energy $\Delta\mu$ into contribution $\Delta\mu_\alpha$ from atom α in protein is given by

$$\Delta\mu = \sum_\alpha \Delta\mu_\alpha \quad \text{with } \Delta\mu_\alpha = \Delta\mu_\alpha^{(\mathrm{LJ})} + \Delta\mu_\alpha^{(\mathrm{elec})}, \qquad \textbf{[S7]}$$

in which

$$\Delta\mu_\alpha^{(\mathrm{LJ})} = 4\pi\rho \sum_\gamma \int_0^1 d\lambda_1 \int r^2 dr \frac{\partial u_{\alpha\gamma}^{(\mathrm{LJ})}(r;\lambda_1)}{\partial\lambda_1} g_{\alpha\gamma}(r;\lambda_1,\lambda_2 = 0), \qquad \textbf{[S8]}$$

$$\Delta\mu_\alpha^{(\mathrm{elec})} = 4\pi\rho \sum_\gamma \int_0^1 d\lambda_2 \int r^2 dr \frac{\partial u_{\alpha\gamma}^{(\mathrm{elec})}(r;\lambda_2)}{\partial\lambda_2} g_{\alpha\gamma}(r;\lambda_1 = 1,\lambda_2). \qquad \textbf{[S9]}$$

Here, λ_1 and λ_2 are the parameters for scaling the LJ parameter ($\lambda_1\sigma_{\alpha\gamma}$) and the atomic charge ($\lambda_2 q_\alpha$) of the protein, respectively, and the resulting interaction potentials are denoted as $u_{\alpha\gamma}^{(\mathrm{LJ})}(r;\lambda_1)$ and $u_{\alpha\gamma}^{(\mathrm{elec})}(r;\lambda_2)$. $g_{\alpha\gamma}(r;\lambda_1,\lambda_2)$ refers to the radial distribution function, related to the 3D distribution function via $g_{\alpha\gamma}(r;\lambda_1,\lambda_2) = (1/4\pi)\int d\hat{\mathbf{r}} g_\gamma(\mathbf{r}_\alpha + \mathbf{r};\lambda_1,\lambda_2)$ with $\hat{\mathbf{r}} = \mathbf{r}/r$ and $r = |\mathbf{r}|$, when the protein–water interaction potential is given by $u_\gamma(\mathbf{r};\lambda_1,\lambda_2) = \sum_\alpha [u_{\alpha\gamma}^{(\mathrm{LJ})}(|\mathbf{r} - \mathbf{r}_\alpha|;\lambda_1) + u_{\alpha\gamma}^{(\mathrm{elec})}(|\mathbf{r} - \mathbf{r}_\alpha|;\lambda_2)]$. $g_{\alpha\gamma}(r;\lambda_1,\lambda_2)$ can also be obtained from the 3D-RISM theory with the aforementioned procedure. We used the Kovalenko–Hirata closure for this purpose as well, in which case both Eqs. **S7**–**S9** and the analytical expression (**S3**) yield the identical numerical value of the total solvation free energy $\Delta\mu$.

A corresponding partitioning of the isochoric solvation entropy Δs_V can be derived by applying the thermodynamic relation (**S4**) to Eq. **S7**. The decomposition of the solvation energy $\Delta\epsilon_V$ is then obtained using Eq. **S5** for each atomic component.

As mentioned above, isochoric ($\Delta\epsilon_V$ and $T\Delta s_V$) and isobaric (Δh and $T\Delta s$) quantities behave practically the same as far as their changes are concerned. Therefore, the changes in the atomic decomposition of Δh and $T\Delta s$ can be well approximated by those of $\Delta\epsilon_V$ and $T\Delta s_V$, and this approximation was used in the main text.

1. Case DA, et al. (2006) *AMBER9* (University of California, San Francisco).

2. Wang J, Cieplak P, Kollman PA (2000) How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J Comput Chem* 21:1049–1074.
3. Lee C, Ham S (2011) Characterizing amyloid-beta protein misfolding from molecular dynamics simulations with explicit water. *J Comput Chem* 32:349–355.
4. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 79:926–935.
5. Darden T,York D, Pedersen L (1993) Particle mesh Ewald: An *N*·log(*N*) method for Ewald sums in large systems. *J Chem Phys* 98:10089–10092.
6. Ryckaert J-P, Ciccotti G, Berendsen HJC (1977) Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of *n*-alkanes. *J Comput Phys* 23:327–341.
7. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR (1984) Molecular dynamics with coupling to an external bath. *J Chem Phys* 81:3684–3690.
8. Bernstein SL, et al. (2005) Amyloid β-protein: Monomer structure and early aggregation states of Aβ42 and its Pro^{19} alloform. *J Am Chem Soc* 127:2075–2084.
9. Baumketner A, et al. (2006) Amyloid β-protein monomer structure: A computational and experimental study. *Protein Sci*. 15:420–428.
10. Mesleh MF, Hunter JM, Shvartsburg AA, Schatz GC, Jarrold MF (1996) Structural information from ion mobility measurements: Effects of the long-range potential. *J Phys Chem* 100:16082–16086.
11. Krone MG, et al. (2008) Role of water in mediating the assembly of Alzheimer amyloid-β Aβ16–22 protofilaments. *J Am Chem Soc* 130:11066–11072.
12. Tarus B, Straub JE, Thirumalai D (2005) Probing the initial stage of aggregation of the $A\beta_{10-35}$-protein: Assessing the propensity for peptide dimerization. *J Mol Biol* 345:1141–1156.
13. Gnanakaran S, Nussinov R, Garcia AE (2006) Atomic-level description of amyloid β-dimer formation. *J Am Chem Soc* 128:2158–2159.
14. Urbanc B, et al. (2004) Molecular dynamics simulation of amyloid β dimer formation. *Biophys J* 87:2310–2321.
15. Mitternacht S, Staneva I, Härd T, Irbäck A (2011) Monte Carlo study of the formation and conformational properties of dimers of Aβ42 variants. *J Mol Biol* 410:357–367.
16. Best RB (2012) Atomistic molecular simulations of protein folding. *Curr Opin Struct Biol* 22:52–61.
17. Bernstein SL, et al. (2009) Amyloid β-protein oligomerization and the importance of tetramers and dodecamers in the aetiology of Alzheimer's disease. *Nat Chem* 1:326–331.
18. Kovalenko A (2003) *Molecular Theory of Solvation*, ed Hirata F (Kluwer Academic, Dordrecht, The Netherlands), pp 164–275.
19. Imai T, Harano Y, Kinoshita M, Kovalenko A, Hirata, F (2006) A theoretical analysis on hydration thermodynamics of proteins. *J Chem Phys* 125:024911.
20. Perkyns J, Pettitt BM (1992) A site-site theory for finite concentration saline solutions. *J Chem Phys* 97:7656–7666.
21. Cann NM, Patey GN (1997) An investigation of the influence of solute size and insertion conditions on solvation thermodynamics. *J Chem Phys* 106:8165–8195.
22. Harano Y, Imai T, Kovalenko A, Kinoshita M, Hirata F (2001) Theoretical study for partial molar volume of amino acids and polypeptides by the three-dimensional reference interaction site model. *J Chem Phys* 114:9506–9511.
23. Chong S-H, Ham S (2011) Atomic decomposition of the protein solvation free energy and its application to amyloid-beta protein in water. *J Chem Phys* 135:034506.

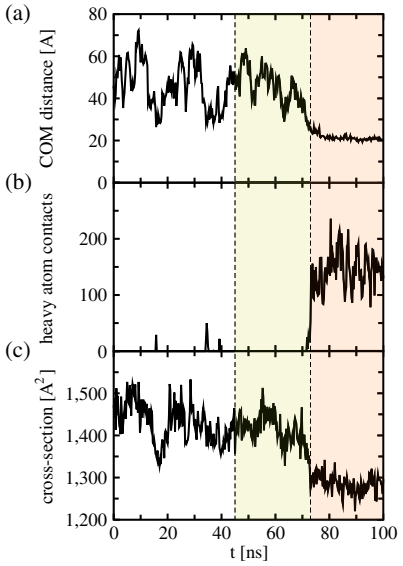

Fig. S1. Structural characteristics of the second Aβ42 dimerization simulation trajectory. (*A*) Center-of-mass (COM) distance between two monomers, (*B*) the number of intermonomer heavy atom contacts, and (*C*) the collision cross-section as a function of time. Vertical dashed lines refer to 45 and 73 ns separating the diffusive regime (0 to 45 ns), the approach stage (45 to 73 ns, colored by light yellow), and the structural adjustment stage (73 to 100 ns, colored by light orange).

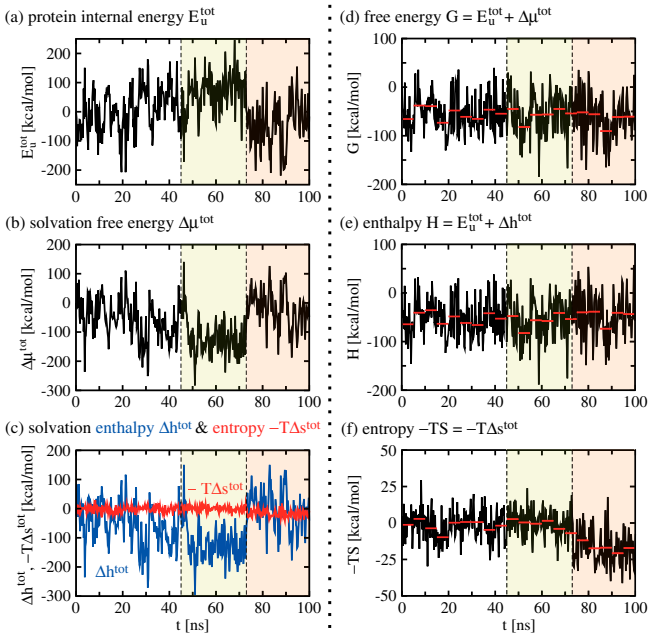

Fig. S2. Thermodynamics of the second Aβ42 dimerization simulation trajectory. (*A*) Total protein internal energy E_u^{tot}, (*B*) solvation free energy $\Delta\mu^{tot}$, (*C*) solvation enthalpy Δh^{tot} and solvation entropy $-T\Delta s^{tot}$, (*D*) free energy $\mathcal{G} = E_u^{tot} + \Delta\mu^{tot}$, and (*E*) its enthalpy component ($\mathcal{H} = E_u^{tot} + \Delta h^{tot}$) and (*F*) entropy component ($-T\mathcal{S} = -T\Delta s^{tot}$) for dimer conformation along the simulation trajectory. In these panels, the initial values are set to zero, and vertical dashed lines refer to 45 and 73 ns indicating the approach stage (45 to 73 ns, colored by light yellow) and the structural adjustment stage (73 to 100 ns, colored by light orange). Red horizontal bars in *D*–*F* represent averages over each 5-ns time interval.

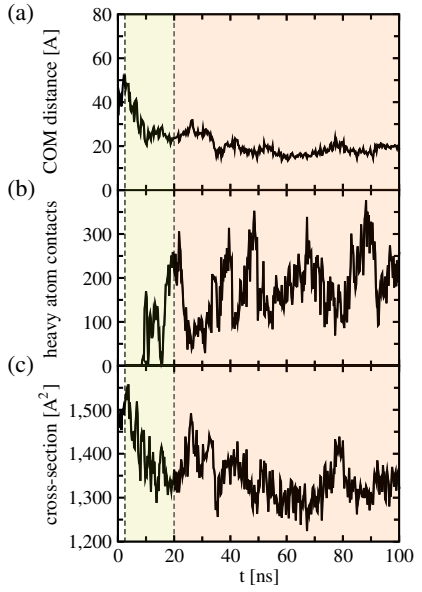

Fig. S3. Structural characteristics of the third Aβ42 dimerization simulation trajectory. (*A*) Center-of-mass (COM) distance between two monomers, (*B*) the number of intermonomer heavy atom contacts, and (*C*) the collision cross-section as a function of time. Vertical dashed lines refer to 2.5 and 20 ns separating the diffusive regime (0 to 2.5 ns), the approach stage (2.5 to 20 ns, colored by light yellow), and the structural adjustment stage (20 to 100 ns, colored by light orange).

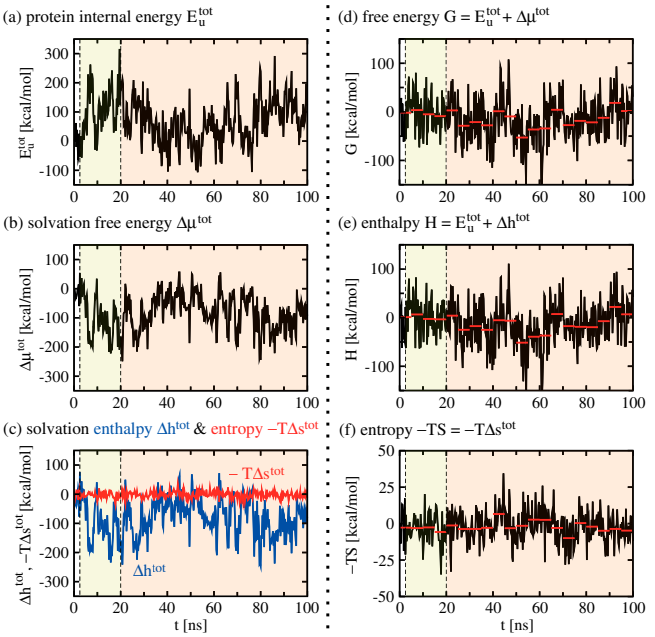

Fig. S4. Thermodynamics of the third Aβ42 dimerization simulation trajectory. (*A*) Total protein internal energy E_u^{tot}, (*B*) solvation free energy $\Delta\mu^{tot}$, (*C*) solvation enthalpy Δh^{tot} and solvation entropy $-T\Delta s^{tot}$, (*D*) free energy $\mathcal{G} = E_u^{tot} + \Delta\mu^{tot}$, and (*E*) its enthalpy component ($\mathcal{H} = E_u^{tot} + \Delta h^{tot}$) and (*F*) entropy component ($-T\mathcal{S} = -T\Delta s^{tot}$) for dimer conformation along the simulation trajectory. In these panels, the initial values are set to zero, and vertical dashed lines refer to 2.5 and 20 ns indicating the approach stage (2.5 to 20 ns, colored by light yellow) and the structural adjustment stage (20 to 100 ns, colored by light orange). Red horizontal bars in panels *D–F* represent averages over each 5-ns time interval.

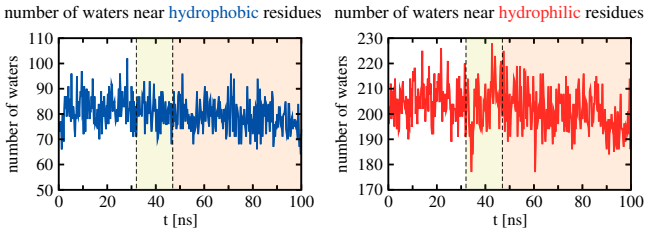

Fig. S5. Number of water molecules in the first solvation shell of hydrophobic residues (*Left*) and of hydrophilic residues (*Right*). The first solvation shell is defined with water molecules whose oxygen is within 3.4 Å from heavy atoms in protein. In both panels, dashed lines refer to 32 and 47 ns indicating the approach stage (32 to 47 ns, colored by light yellow) and structural adjustment stage (47 to 100 ns, colored by light orange).

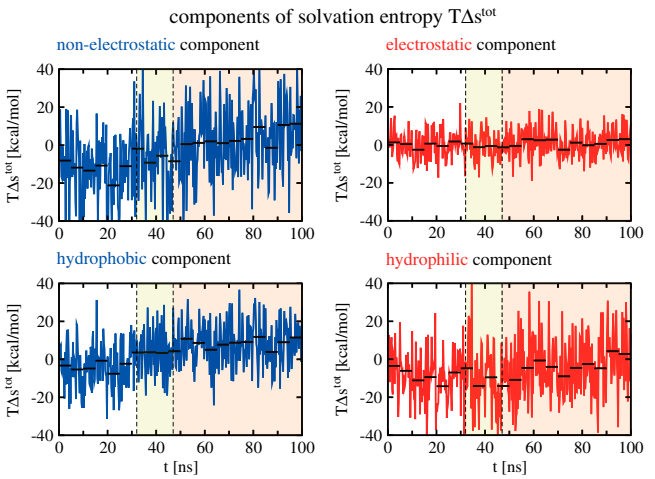

Fig. S6. Component analysis of the solvation entropy. The solvation entropy $T\Delta s^{tot}$ is partitioned into nonelectrostatic and electrostatic components (*Upper*) and into hydrophobic- and hydrophilic-residue components (*Lower*). In all the panels, the initial values are set to zero, dashed lines refer to 32 and 47 ns indicating the approach stage (32 to 47 ns, colored by light yellow) and structural adjustment stage (47 to 100 ns, colored by light orange), and horizontal bars represent averages over each 5-ns time interval.