Special Issue
"Progress of Theoretical and Computational Biophysics"

Vol. 16, pp. 185–195 (2019)
doi: 10.2142/biophysico.16.0_185

Biophysics and Physicobiology

https://www.jstage.jst.go.jp/browse/biophysico/

Review Article

Energetics of cosolvent effect on peptide aggregation

Nobuyuki Matubayasi and Keiichi Masutani

Division of Chemical Engineering, Graduate School of Engineering Science, Osaka University, Toyonaka, Osaka 560-8531, Japan

Received April 23, 2019; accepted May 16, 2019

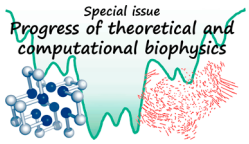
The cosolvent effect on the equilibrium of peptide aggregation is reviewed from the energetic perspective. It is shown that the excess chemical potential is stationary against the variation of the distribution function for the configuration of a flexible solute species and that the derivative of the excess chemical potential with respect to the cosolvent concentration is determined by the corresponding derivative of the solvation free energy averaged over the solute configurations. The effect of a cosolvent at low concentrations on a chemical equilibrium can then be addressed in terms of the difference in the solvation free energy between pure-water solvent and the mixed solvent with the cosolvent, and illustrative analyses with all-atom model are presented for the aggregation of an 11-residue peptide by employing the energy-representation method to compute the solvation free energy. The solvation becomes more favorable with addition of the urea or DMSO cosolvent, and the extent of stabilization is smaller for larger aggregate. This implies that these cosolvents inhibit the formation of an aggregate, and the roles of such interaction components as the electrostatic, van der Waals, and excluded-volume are discussed.

Key words: aggregation, solvation, free energy, cosolvent, MD simulation

Corresponding author: Nobuyuki Matubayasi, Division of Chemical Engineering, Graduate School of Engineering Science, Osaka University, 1-3 Machikaneyama-cho, Toyonaka, Osaka 560-8531, Japan.
e-mail: nobuyuki@cheng.es.osaka-u.ac.jp

Protein or peptide can be harmful upon formation of aggregate [1–4]. It is considered that such amyloidoses as Alzheimer's, Parkinson's, and prion diseases are caused by the amyloid aggregation, and the formation of inclusion body is an impeding factor for massive expression in the field of protein engineering. The extent of aggregate formation of a peptide is determined by the interactions among the peptides forming the aggregate and those between the peptides and the surrounding solvent. In fact, the potential function is fixed for the interactions within the peptide aggregate when the peptide species is identified, whereas it can be tuned for the interactions with the surrounding environment. A common scheme for treating a peptide aggregate is thus to control the solvent environment with cosolvent [5–10], and since the cosolvent interacts with the peptide in a different manner from water, the interactions among the peptide, cosolvent, and water needs to be elucidated at the molecular level toward rational design of a cosolvent.

The aggregation mechanism of peptide has been investigated both theoretically and experimentally [11–39]. Molecular dynamics (MD) simulation has proven to be useful to address atomic-level processes, in particular, and the peptide-peptide and peptide-solvent interactions were analyzed to reveal the driving force of aggregate formation [12,14,16,18–24,30]. In the context of statistical thermodynamics, the extent of peptide aggregation is described by the equilibrium constant of aggregation. It is determined by the balance of the



◀ ***Significance*** ▶

The cosolvent effect on peptide aggregation is addressed with all-atom molecular dynamics simulation and free-energy calculation. It is shown from the variational principle that the stability of a flexible solute species is modulated by a cosolvent through the solvation free energy averaged over the solute structures. The computational analyses are conducted for an 11-residue peptide, and the solvation free energies of the peptide and its aggregates are obtained with the energy-representation method of solutions. It is demonstrated quantitatively that urea and DMSO inhibit the aggregation since these cosolvents stabilize the monomer more strongly than the aggregates.

intermolecular interactions among the peptide and solvent molecules, and an all-atom treatment of (free-)energetics is desirable to understand and control the aggregation and dissolution of the peptide.

The purpose of the present review is to show the energetics of peptide aggregation by focusing on NACore [38]. It corresponds to the 68th to 78th residues of α-synuclein, which are considered as the key region to cause Parkinson's disease through aggregate formation [2,40–42]. All-atom MD simulation was recently conducted to examine the roles of the intra-aggregate and aggregate-solvent interactions in the formations of the 8-mer, 16-mer, and 24-mer of NACore, and the effect of urea or dimethyl sulfoxide (DMSO) as a cosolvent was elucidated through free-energy calculations [30]. This review introduces a statistical-thermodynamic treatment for the aggregation of the peptide, with emphasis on the free energy of solvation of a peptide or its aggregate as the solute.

When the solvent environment is varied by addition of a cosolvent, the extent of fluctuation will change for the structure of a flexible solute molecule. The "configurational entropy" (chain entropy) is a quantity to assess the effect of structural fluctuation, though its computation is still a subject of active development [43–50]. In the following section, in fact, we show that the derivative of the excess chemical potential of the solute species with respect to the cosolvent concentration is determined by the corresponding derivative of the solvation free energy. It is not necessary to handle the configurational entropy, and the effect of cosolvent can be addressed only in terms of the solvation free energy. The key variable is then the solvation free energy and its derivative with respect to the cosolvent concentration.

The solvation free energy requires much computational demand for a peptide or its aggregate in explicit solvent, however, when the free-energy perturbation or thermodynamic integration method is utilized in molecular simulation [51]. Indeed, these standard methods are carried out by introducing a number of intermediate states connecting the solution system of interest and the reference-solvent system without the solute of interest. The method of energy representation is a useful alternative for the computation of the solvation free energy [52–56]. It is a density-functional method for solutions and evaluates the solvation free energy with a functional of a set of distribution functions for the pair-interaction energy between the solute and solvent. A variety of approximate methods have been proposed for free-energy computation [57–75], and the energy-representation method is unique in compromising the accuracy, efficiency, and range of applicability [30,53,55,56,76–96]. In the method of energy representation, the molecular simulation is to be performed only at the endpoints of solute insertion (solution and reference-solvent systems of interest), leading to the reduction of the computational load. The solvation free energy can then be obtained feasibly for a protein or its complex consisting of a few hundred residues not only in pure-water solvent but also in more complex environments involving cosolvent or lipid membrane, that can be beyond the reach of exact free-energy methods or integral equation theories [30,81,86,90–93,97–102]. The method of energy representation is approximate in practical applications, while it has been observed that the error from the approximation in the free-energy functional is not larger than the error from the use of force field [80,84,85]. The solvation free energies presented in this review were obtained by the energy-representation method [30].

Theory

When the aggregation of NACore is treated, the solute refers to the peptide or its aggregate. It should be noted that the aggregate is treated as a single, solute particle as a whole, and the (relative) stability of the solute species is described by its excess chemical potential μ^{ex}. Let ψ denote the coordinate of the solute particle collectively and $P(\psi)$ be the probability distribution function of ψ in the solution system of interest. μ^{ex} is then expressed in the canonical (NVT) ensemble as [30]

$$\mu^{\mathrm{ex}} = \int d\psi P(\psi) E_S(\psi) + \int d\psi P(\psi) \Delta v(\psi) + k_B T \int d\psi P(\psi) \log (P(\psi) V) \tag{1}$$

where $E_S(\psi)$ is the one-body energy of the solute, k_B and T are the Boltzmann constant and temperature, respectively, V is the volume of the system, and $P(\psi)$ satisfies the normalization condition of

$$\int d\psi P(\psi) = 1. \tag{2}$$

$\Delta v(\psi)$ is the solvation free energy at fixed configuration ψ of the solute particle, and is introduced by

$$\Delta v(\psi) = -k_B T \log \left(\frac{\int d\mathbf{X} \exp (-\beta \{ U(\mathbf{X}) + u(\psi, \mathbf{X}) \})}{\int d\mathbf{X} \exp (-\beta U(\mathbf{X}))} \right), \tag{3}$$

where $\beta = 1/k_B T$, all the solvent particles are collectively written with the coordinate $\mathbf{X}$ and the energy $U(\mathbf{X})$, and $u(\psi, \mathbf{X})$ is the interaction between the solute and solvent. It is further supposed that $E_S(\psi)$ is a function only of the solute configuration ψ and does not depend on the solvent coordinate. The first term of Eq. 1 is the average of the one-body energy of the solute in the solution system of interest, and the second term is the averaged free energy of solvation. The third term corresponds to the "configurational entropy" (chain entropy) of the solute particle for which the configuration ψ distributes with $P(\psi)$. Although the solvation free energy refers to the "reference-solvent" system without solute (denominator within the logarithm of Eq. 3), the other variables in Eq. 1 can be obtained in principle from a simulation of the solution system of interest.

The equilibrium of n-mer formation is determined by the excess chemical potentials of the n-mer and monomer. When ρ_n and μ_n^{ex} are the concentration and excess chemical potential of the n-mer, respectively, and ρ_1 and μ_1^{ex} are the corresponding quantities for the monomer,

$$-k_B T \log\left(\frac{\rho_n}{\rho_1^n}\right) = n\left(\frac{\mu_n^{\mathrm{ex}}}{n} - \mu_1^{\mathrm{ex}}\right) \tag{4}$$

is satisfied at equilibrium. ρ_n/ρ_1^n is the equilibrium constant for the n-mer formation, and the difference from μ_1^{ex} appears in Eq. 4 for μ_n^{ex} per monomer basis. Actually, Eqs. 1 and 4 are valid even when the solute species is at finite concentration. In such a case, one of the solute particles is treated with ψ and the others are expressed as part of **X**. The activity coefficient is incorporated in the excess chemical potential μ^{ex}, furthermore.

When Eq. 1 is viewed as a functional of the distribution function $P(\psi)$, the variational property of

$$\left.\frac{\delta\mu^{\mathrm{ex}}}{\delta P(\psi)}\right|_{E_S(\psi),\Delta v(\psi),T} = \text{constant independent of } \psi \tag{5}$$

holds at the $P(\psi)$ that is realized in the solution system of interest and is given by

$$P(\psi) = \frac{\int d\mathbf{X} \exp\left(-\beta\{E_S(\psi) + U(\mathbf{X}) + u(\psi,\mathbf{X})\}\right)}{\int d\psi d\mathbf{X} \exp\left(-\beta\{E_S(\psi) + U(\mathbf{X}) + u(\psi,\mathbf{X})\}\right)}. \tag{6}$$

This is because $P(\psi)$ of Eq. 6 satisfies

$$-k_B T \log\left(P(\psi)\right) = E_S(\psi) + \Delta v(\psi) + \text{constant independent of } \psi \tag{7}$$

and the variation of the excess chemical potential μ^{ex} is expressed as

$$\delta\mu^{\mathrm{ex}} = \int d\psi \left(\delta P(\psi)\right)\left[E_S(\psi) + \Delta v(\psi) + k_B T \log\left(P(\psi)V\right) + k_B T\right] \tag{8}$$

at fixed $E_S(\psi)$, $\Delta v(\psi)$, V, and T. $\Delta v(\psi)$ and $P(\psi)$ may change when a cosolvent is added at constant temperature. Due to the variational property of Eq. 5, though, the change in $P(\psi)$ does not contribute to the derivative of the excess chemical potential μ^{ex} with respect to the cosolvent contribution c. The normalization condition of

$$\int d\psi \left(\delta P(\psi)\right) = 0 \tag{9}$$

and Eq. 5 lead to

$$\frac{\partial\mu^{\mathrm{ex}}}{\partial c} = \int d\psi P(\psi) \frac{\partial(\Delta v(\psi))}{\partial c}. \tag{10}$$

Equation 10 shows that the derivative of μ^{ex} can be computed by averaging the corresponding derivative of $\Delta v(\psi)$. Actually, ψ is high-dimensional unless the solute is simple and it is often prohibitive to obtain $P(\psi)$ (and $\log(P(\psi))$) explicitly. At $c=0$, Eq. 10 reduces to

$$\begin{aligned}\left.\frac{\partial\mu^{\mathrm{ex}}}{\partial c}\right|_{c=0} &= \int d\psi P_0(\psi) \left.\frac{\partial(\Delta v(\psi))}{\partial c}\right|_{c=0} \\ &\approx \int d\psi P_0(\psi) \frac{\Delta v(\psi;c) - \Delta v(\psi;0)}{c} \\ &= \frac{\langle \Delta v(\psi;c)\rangle - \langle \Delta v(\psi;0)\rangle}{c}\end{aligned} \tag{11}$$

where $P_0(\psi)$ is the distribution function of ψ in pure-water solvent ($c=0$, without cosolvent), $\Delta v(\psi;c)$ is the solvatin free energy for the solute configuration ψ at the cosolvent concentration of c, and $\langle ... \rangle$ is the average over the distribution $P_0(\psi)$. The second equation is the finite-difference approximation, and is valid when c is small enough. To compute the value of the rightmost side, both of $\Delta v(\psi;c)$ and $\Delta v(\psi;0)$ are to be averaged over a set of solute configurations ψ sampled in pure-water solvent ($c=0$).

$(\partial\Delta v(\psi)/\partial c)$ is related to the Kirkwood-Buff integral through [77,103–107]

$$-\frac{\partial(\Delta v(\psi))}{\partial c} = k_B T \int dr (4\pi r^2)\left(g_c(r;\psi) - g_w(r;\psi)\right), \tag{12}$$

where $g_c(r;\psi)$ and $g_w(r;\psi)$ are the radial distribution functions of the cosolvent and water, respectively, with the solute at fixed configuration ψ; the spatial integrals of $(g_c - 1)$ and $(g_w - 1)$ are called Kirkwood-Buff integrals and their values are independent of the sites that specify points within the molecules and define the radial distance r. Equations 10 and 12 then lead to

$$-\frac{\partial\mu^{\mathrm{ex}}}{\partial c} = k_B T \int dr (4\pi r^2)\left(g_c(r) - g_w(r)\right), \tag{13}$$

where $g_c(r)$ and $g_w(r)$ are the radial distribution functions obtained by averaging $g_c(r;\psi)$ and $g_w(r;\psi)$ over the distribution $P(\psi)$ of the solute configuration, respectively. Equation 13 shows that the accumulation of the cosolvent around the solute and/or the depletion of water stabilizes the solute with more negative μ^{ex}. The Kirkwood-Buff integral requires a careful treatment in its computation with molecular simulation, however [108–111], and Eq. 10 can be a more viable route to evaluating the cosolvent effect on the excess chemical potential through molecular simulation.

Let W be the peptide concentration where α% of the peptide molecules are in the aggregate form. W is a function of the cosolvent concentration c, and Eqs. 4 and 11 lead to

$$\left(\frac{\partial \log W}{\partial c}\right)_{c=0} = \frac{n}{(n-1)k_B T}\left[\frac{1}{n}\int d\psi_n P_0(\psi_n)\left(\frac{\partial(\Delta v(\psi_n))}{\partial c}\right)_{c=0} - \int d\psi_1 P_0(\psi_1)\left(\frac{\partial(\Delta v(\psi_1))}{\partial c}\right)_{c=0}\right]$$

$$\approx \frac{n}{(n-1)k_B T}\left[\frac{1}{n}\int d\psi_n P_0(\psi_n)\frac{\Delta v(\psi_n;c)-\Delta v(\psi_n;0)}{c} - \int d\psi_1 P_0(\psi_1)\frac{\Delta v(\psi_1;c)-\Delta v(\psi_1;0)}{c}\right] \quad (14)$$

where ψ_n and ψ_1 denote the configurations of the n-mer and monomer, respectively, $P_0(\psi_n)$ and $P_0(\psi_1)$ are the distribution functions of ψ_n and ψ_1 in pure-water solvent (c=0), respectively, $\Delta v(\psi_n;c)$ and $\Delta v(\psi_1;c)$ are the solvation free energies of the n-mer at the configuration ψ_n and of the monomer at ψ_1, respectively, when the cosolvent concentration is c, and $\Delta v(\psi_n;0)$ and $\Delta v(\psi_1;0)$ are the corresponding values in pure-water solvent. Actually, Eq. 14 holds at any percent α ($0<\alpha<100$) of aggregate formation.

Computational Methods

The present review shows the energetic analyses of NACore, its 8-mer, 16-mer, and 24-mer in pure-water solvent and 3.0 M mixtures of water with urea and with DMSO (M=mol/L). The NACore peptide consists of 11 residues, with the amino-acid sequence of Gly-Ala-Val-Val-Thr-Gly-Val-Thr-Ala-Val-Ala [38]. When the peptide was simulated, it was treated as a neutral species by capping the N- and C-termini with -CO-CH_3 and -NH-CH_3, respectively. The crystal structure (PDB code: 4RIL) was utilized to build the initial structures of the aggregates. The 8-mer, 16-mer, and 24-mer were prepared by extracting 4 parallel and 2 anti-parallel layers of β-sheet, 4 parallel and 4 anti-parallel layers, and 4 parallel and 6 anti-parallel layers, respectively. Illustrative structures are provided in Figure 1. The original TIP3P model was adopted for water, and the Amber99sb force field with the RESP procedure was employed for NACore, urea, and DMSO [30,84,112–114]. The configurations of the system were generated by carrying out the molecular dynamics (MD) simulation in the canonical (*NVT*) ensemble with a cubic unit cell. GROMACS 2016.4 was used at a temperature of 300 K and the cell size that corresponds to the pressure of 1 bar [115,116]. In all the MD simulations, the number of solvent molecules in the unit cell was 30000 in total, and when the urea-water and DMSO-water mixtures were simulated, 1770 urea molecules and 1960 DMSO were located, respectively, to make the cosolvent concentration 3.0 M at 300 K and 1 bar. The solute refers to NACore, its 8-mer, 16-mer, or 24-mer, and the number of solute particle in the unit cell was always unity. It was actually observed for the 8-mer that the β-sheet structure is lost

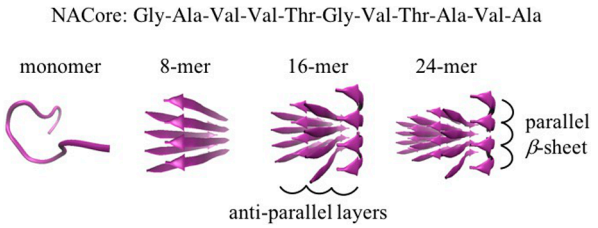


Figure 1 Illustrative structures for the NACore peptide, its 8-mer, the 16-mer, and the 24-mer.

with the time evolution in MD when no restraint was applied. A restraining potential in the flat-bottomed form of

$$\frac{1}{2}k(r-d)^2 H(r-d) \quad (15)$$

was thus employed for the 8-mer, 16-mer, and 24-mer, where k is the force constant of 2500 kcal/mol/Å^2, d is the threshold distance of 8 Å, H is the Heaviside step function, and r is the distance between the atoms in the terminal residues of neighboring pairs of parallel peptide chains. The anti-parallel pairs were not subject to any restraint, and r refers to the N atom within -NH-CO-CH_3 of the N-terminus or the carbonyl C atom within -CO-NH-CH_3 of the C-terminus. The detailed procedures of MD simulation are presented in Ref. [30].

Two types of MD simulation were conducted to obtain the computational results in the present review. One was performed to prepare a set of solute structures, and the other was done to analyze the solute-solvent energetics. In the former, MD was conducted in pure-water solvent by treating the solute as a flexible species. The simulation length was 150 ns including the first 50 ns for equilibration, and 50 samples of the solute structure were collected at an interval of 2 ns in the course of equilibrium fluctuation. Each (snapshot) structure of solute thus sampled was then subject to the energetic analysis in the latter type of MD. The free-energy calculation was carried out by freezing the solute structure, and the solvation free energy was evaluated for each rigid structure of the solute in all of pure-water solvent, the urea-water mixed solvent, and the DMSO-water mixed solvent. The method of energy representation was adopted to obtain the solvation free energy [53,55,56], with its detailed procedures described in Ref. [30].

The solute structures subject to the energetic analysis were taken from trajectories in pure-water solvent. In accordance with Eqs. 11 and 14, the solvation free energies in the urea-water and DMSO-water mixed solvents were also calculated over the solute structures sampled in pure water. The error estimate is then necessary for the average of a quantity A with the form of

$$\langle A \rangle = \int d\psi P(\psi) A(\psi), \quad (16)$$

where ψ is the solute configuration and $P(\psi)$ is the probability distribution function of ψ. $A(\psi)$ is evaluated at fixed

ψ by allowing only the solvent molecules to move, and $\langle A\rangle$ denotes the average over the solute degrees of freedom. The first and second terms of Eq. 1 are in the form of Eq. 16 by setting A to the intra-solute energy E_S and the solvation free energy Δv, respectively, and in Eqs. 11 and 14, A is involved as the cosolvent-induced change in the solvation free energy Δv.

When n configurations are prepared for ψ according to the distribution function $P(\psi)$, $\langle A\rangle$ is computed as

$$\langle A\rangle = \frac{1}{n}\sum_{i=1}^{n} A_i, \tag{17}$$

where A_i is the value of $A(\psi)$ for the ith sample. In fact, the number of sampled configurations for the solvent is also finite in actual simulation, and accordingly, a statistical error is involved in the calculation of A_i (i=1, ..., n). When A_i is calculated m_i times through finite sampling of the solvent, we let $(A_i+e_i^j)$ denote the numerical result from the jth run (j=1, ..., m_i) and A_i stand for the value at infinite sampling. In practical computation, only $(A_i+e_i^j)$ can be obtained. Equation 17 should then read as

$$\langle A\rangle = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{m_i}\sum_{j=1}^{m_i}(A_i+e_i^j) \tag{18}$$

and the standard error for $\langle A\rangle$ is expressed as

$$\frac{1}{\sqrt{n}}\sqrt{\mathrm{var}_u(A)+\frac{1}{n}\sum_{i=1}^{n}(\mathrm{ste}_v(e_i))^2} \tag{19}$$

when all of A_i and e_i^j are supposed to be uncorrelated to one another. In Eq. 19, $\mathrm{var}_u(A)$ refers to the variance of $A(\psi)$ and is evaluated over the distribution function $P(\psi)$ of the solute configuration ψ. The subscript u is attached to mean that var_u is the variance over the solute degrees of the freedom, and in numerical treatment with $n \gg 1$,

$$\mathrm{var}_u(A) \approx \frac{1}{n}\sum_{i=1}^{n}\left[\left(\frac{1}{m_i}\sum_{j=1}^{m_i}(A_i+e_i^j)\right)-\langle A\rangle\right]^2 \tag{20}$$

is employed with $\langle A\rangle$ of Eq. 18. $\mathrm{ste}_v(e_i)$ denotes the standard error for A_i due to the finite sampling of the solvent at fixed, solute configuration ψ, where the subscript v is for the solvent. $\mathrm{ste}_v(e_i)$ is calculated from

$$(\mathrm{ste}_v(e_i))^2 \approx \frac{1}{m_i}\left[\frac{1}{m_i}\sum_{j=1}^{m_i}\left((A_i+e_i^j)-\frac{1}{m_i}\sum_{k=1}^{m_i}(A_i+e_i^k)\right)^2\right] \tag{21}$$

by virtue of $m_i \gg 1$. It should be noted that $\mathrm{ste}_v(e_i)$ converges to zero for each i when the solvent degrees of freedom are sampled enough, while $\mathrm{var}_u(A)$ remains finite even when a large number of solute configurations ψ is explored. The average of $(\mathrm{ste}_v(e_i))^2$ over i=1, ..., n is involved within the square root of Eq. 19, and is reduced by a factor of $\sqrt{n}$ when the standard error of $\langle A\rangle$ is estimated.

To obtain the computational results presented in the next section, 50 configurations were sampled for the solute structure as noted in the 2nd paragraph of the present section. The dependence of the results on the number of sampled configurations was analyzed in Ref. [30], and it was observed there that 50 samples are enough for the NACore peptide and its aggregates.

Results and Discussion

Figure 2 shows the solvent-accessible surface area (SASA) per monomer; note that the first 50 ns is for equilibration and that the last 100 ns corresponds to the production period. The SASA fluctuates stably, and its average (per monomer) is 1.30, 0.59, 0.48, and 0.45 in units of 10^3 Å^2 for the monomer, 8-mer, 16-mer, and 24-mer, respectively. The peptide is thus exposed less to the solvent with aggregation. In the following, the aggregate means a set of configurations of the 8-mer, 16-mer, or 24-mer sampled during the course of MD in pure-water solvent. Each of the 8-mer, 16-mer, and 24-mer is further treated as a single particle as a whole without adopting any criterion to identify whether the particle is aggregated or not.

We next describe the β-sheet structure of NACore and its aggregates in pure-water solvent. At the monomer state, NACore does not keep the β-sheet conformation in the sense that the average number of residues with that conformation is zero. The β sheet forms for the aggregates, and the number of β-sheet residues per monomer is 7.3, 8.0, and 7.9 on average for the 8-mer, 16-mer, and 24-mer, respectively, with the standard deviation of 0.6, 0.2, and 0.2. The β-sheet is thus more "rigid" for the larger aggregates, given that the crystal structure has 9 residues with β-sheet conformation (per monomer) [38].

The energetics in pure-water solvent can be examined on the basis of Eq. 1. The first and second terms of Eq. 1 are denoted by $\langle E_S\rangle$ and $\langle \Delta v\rangle$, respectively, and in accordance with Eq. 4, we discuss the energetics of n-mer formation in

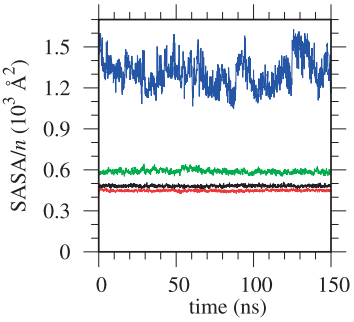


Figure 2 Solvent-accessible surface area (SASA) divided by the degree of aggregation n against the MD time for the monomer (blue), 8-mer (green), 16-mer (black), and 24-mer (red). The MD was performed in pure-water solvent, and the SASA was calculated using the solute surface in contact with the spherical probe which mimics the water molecule and has a radius of 1.4 Å.

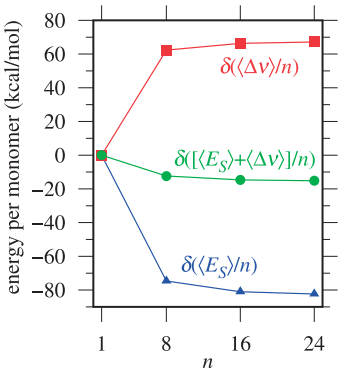


Figure 3 Difference $\delta(\langle E_S\rangle/n)$ of the averaged intra-aggregate energy from the value for the monomer (n=1), the difference $\delta(\langle\Delta v\rangle/n)$ of the averaged solvation free energy from the monomer value, and their sum $\delta([\langle E_S\rangle+\langle\Delta v\rangle]/n)$ plotted against the degree of aggregation n. $\langle E_S\rangle$ and $\langle\Delta v\rangle$ are the first and second terms of Eq. 1, respectively, where the averages were taken over the solute structures sampled in pure-water solvent. $\delta(\langle E_S\rangle/n)$, $\delta(\langle\Delta v\rangle/n)$, and $\delta([\langle E_S\rangle+\langle\Delta v\rangle]/n)$ are zero at n=1 by definition, and the lines connecting the data are drawn for eye guide.

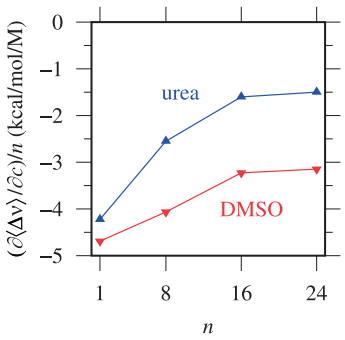


Figure 4 Cosolvent-induced change in the averaged solvation free energy $(\partial\langle\Delta v\rangle/\partial c)/n$ plotted per monomer against the degree of aggregation n, where c is the cosolvent concentration and the derivative with respect to c is taken at c=0 through Eq. 11. The lines connecting the data are drawn for eye guide.

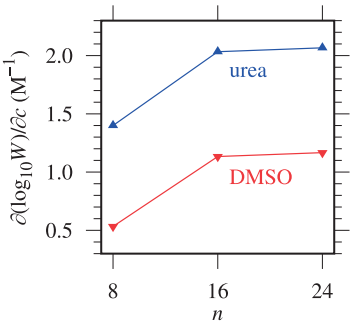


Figure 5 $(\partial\log_{10}W/\partial c)$ at c=0 against the degree of aggregation n, where c is the cosolvent concentration and W is the peptide concentration at which α% of the peptide molecules are in the aggregate form. Actually, $(\partial\log_{10}W/\partial c)$ is independent of the percentage value α when α is fixed in the variation of c. The numerical value is expressed as a common logarithm (with base 10) only in this figure, and the lines connecting the data are drawn for eye guide.

terms of $\delta(\langle E_S\rangle/n)$ and $\delta(\langle\Delta v\rangle/n)$ defined respectively as the differences of $\langle E_S\rangle/n$ and $\langle\Delta v\rangle/n$ from their values for the monomer (n=1). Figure 3 depicts $\delta(\langle E_S\rangle/n)$ and $\delta(\langle\Delta v\rangle/n)$ against the degree of aggregation n. As noted in the previous section, $\langle E_S\rangle$ and $\langle\Delta v\rangle$ were computed by averaging over the 50 structures sampled in pure-water solvent. Note that E_S consists both of the intramolecular energy of the peptide and the intermolecular interaction among the peptides when the solute of interest is an aggregate (n>1). It is seen in Figure 3 that $\delta(\langle E_S\rangle/n)$ is more favorable (more negative) at larger n. The interactions among the peptide molecules favor the formation of a larger aggregate, and this reflects the increase of the number of β-sheets with n as noted in the preceding paragraph. The solvent effect in the n-mer formation is expressed as $\delta(\langle\Delta v\rangle/n)$ and is less favorable (more positive) at larger degree of aggregation. The solvent water inhibits the aggregation, and in terms of $\langle E_S\rangle$ and $\langle\Delta v\rangle$, the former is the driving force of aggregate formation. The increase of $\delta(\langle\Delta v\rangle/n)$ is overwhelmed by the decrease of $\delta(\langle E_S\rangle/n)$, and the aggregate of NACore is energetically more stable than the monomer.

As noted with respect to Eq. 11, the effect of a cosolvent on the stability of a solute species is determined by the change in $\langle\Delta v\rangle$ due to addition of the cosolvent. Figure 4 shows $(\partial\langle\Delta v\rangle/\partial c)/n$ against the degree of aggregation n, where c is the cosolvent concentration and the derivative with respect to c is taken at c=0. It is seen from $(\partial\langle\Delta v\rangle/\partial c)/n$<0 that the solvation is more favorable in the mixed solvent of water with urea or DMSO than in pure-water solvent. The peptide and its aggregates are stabilized with these cosolvents, and the stabilization effect is stronger with DMSO. The cosolvent-induced change in $\langle\Delta v\rangle/n$ is smaller in magnitude at larger n, furthermore, implying that the urea or DMSO cosolvent stabilizes a larger aggregate to a lesser extent. In other words, the cosolvent favors the monomeric form of the peptide and acts to inhibit the aggregation. The n dependence of $(\partial\langle\Delta v\rangle/\partial c)/n$ is stronger with urea than with DMSO. The difference in the cosolvent effect of stabilization between the monomer and aggregates is thus larger with urea, which means in turn that urea is a better inhibitor of aggregation than DMSO.

Since the cosolvent stabilizes the monomer more strongly than the aggregate, its addition shifts the equilibrium of aggregate formation to the monomer side. Equation 14 connects the cosolvent effect of inhibiting the aggregate formation to $(\partial\langle\Delta v\rangle/\partial c)/n$ in Figure 4, and $(\partial\log_{10}W/\partial c)$ at c=0 is depicted in Figure 5 at each degree of aggregation n; W denotes the peptide concentration at which the aggregation occurs (see the description above Eq. 14 and the caption of Fig. 5). It should be noted that the common logarithm (with base 10) is adopted in Figure 5 to show the order of magnitude more clearly. According to Figure 5, the peptide aggregates at concentrations higher by orders of magnitude in the mixed solvent with urea or DMSO than in pure-water

solvent. NACore is prepared in practice at concentrations of ~10^{-3} M [38], and Figure 5 indicates that the aggregation is prevented by urea or DMSO when it is added on the order of M. $(\partial \log_{10} W/\partial c)$ is larger with urea than with DMSO. This means that DMSO is a weaker inhibitor of aggregation than urea, and corresponds to the computational result in Figure 4 that the n dependence of $(\partial\langle\Delta v\rangle/\partial c)/n$ is weaker with DMSO. Experimental reports have noted indeed that both of the urea and DMSO cosolvents can inhibit the aggregation at cosolvent concentrations of M scale, with the effect being stronger for the former [117–119].

The effect of solvation is determined by the cooperation and/or competition among a variety of intermolecular interactions, and it is of interest to examine the roles of interaction components such as electrostatic, van der Waals (dispersion), and excluded-volume in the equilibrium of aggregate formation. The electrostatic and van der Waals interactions are usually operative as attractive components of the solute-solvent interaction, and the former is more residue-specific. The excluded-volume effect refers to the free-energy penalty for displacing solvent molecules from the region to be occupied by the solute. It is part of the repulsive component of solute-solvent interaction, and can play a decisive role in structure formation of biomolecules [87,93,120–122]. In the energy-representation formalism, the solvation free energy Δv is expressed as

$$\Delta v = u_{\mathrm{elec}} + u_{\mathrm{vdW}} + \sum_i \int d\epsilon_i f(\epsilon_i) \tag{22}$$

where ϵ_i is the pair energy of the solute with solvent species i (i is water, urea, or DMSO) and it is supposed that the intermolecular interaction between the solute and solvent is a sum of the electrostatic and van der Waals terms. In Eq. 22, the first and second terms are the electrostatic and van der Waals components of the average sum of the solute-solvent interaction energy in the solution of interest, respectively. They are obtained exactly (within the statistical error) under the used set of force field through a simulation of the solution system with fully coupled solute-solvent interaction. $f(\epsilon_i)$ describes the effect of solvent reorganization due to the introduction of the solute, and the excluded-volume component Δv_{excl} can be introduced by restricting the domain of integration in the third term of Eq. 22 to $\epsilon_i > \epsilon^c$ as [86–88, 90–93]

$$\Delta v_{\mathrm{excl}} = \sum_i \int_{\epsilon^c}^{\infty} d\epsilon_i f(\epsilon_i). \tag{23}$$

ϵ^c is a threshold value for specifying the excluded-volume domain and $\epsilon_i > \epsilon^c$ is a high-energy region of pair interaction corresponding to the solute-solvent overlap. The value of ϵ^c is rather arbitrary with a requirement that a configuration with $\epsilon_i > \epsilon^c$ is practically inaccessible in the solution system and does not appear in the solution MD. We set ϵ^c=20 kcal/mol in the results shown below, while the following discussion is valid irrespective of the (reasonable) setting of the ϵ^c value.

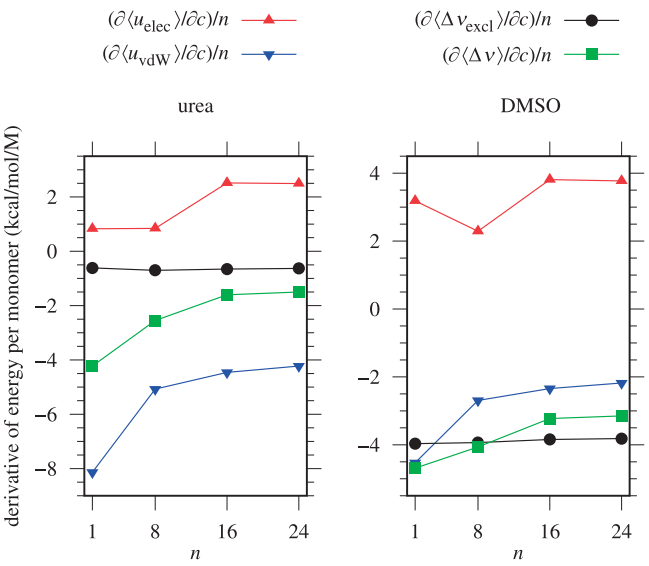


Figure 6 Dependence per monomer on the degree of aggregation n of the derivative of the electrostatic component of the solute-solvent energy $(\partial\langle u_{\mathrm{elec}}\rangle/\partial c)/n$, the derivative of the van der Waals component $(\partial\langle u_{\mathrm{vdW}}\rangle/\partial c)/n$, the derivative of the excluded-volume component $(\partial\langle \Delta v_{\mathrm{excl}}\rangle/\partial c)/n$, and the derivative of the solvation free energy $(\partial\langle \Delta v\rangle/\partial c)/n$, where c denotes the cosolvent concentration and the energetic quantities are differentiated at c=0. $(\partial\langle \Delta v\rangle/\partial c)/n$ is the same as that in Figure 4, and the lines connecting the data are drawn for eye guide.

In correspondence to Eq. 11, the cosolvent effects on the electrostatic, van der Waals, and excluded-volume components are described by $(\partial\langle u_{\mathrm{elec}}\rangle/\partial c)$, $(\partial\langle u_{\mathrm{vdW}}\rangle/\partial c)$, and $(\partial\langle \Delta v_{\mathrm{excl}}\rangle/\partial c)$, respectively, where c is the cosolvent concentration and the differentiation is done at c=0. These derivatives were obtained with the finite-difference approximation, and $\langle ... \rangle$ denotes the average over the 50 structures of the solute sampled in pure-water solvent. Figure 6 shows $(\partial\langle \Delta u_{\mathrm{elec}}\rangle/\partial c)/n$, $(\partial\langle u_{\mathrm{vdW}}\rangle/\partial c)/n$, $(\partial\langle \Delta v_{\mathrm{excl}}\rangle/\partial c)/n$, and $(\partial\langle \Delta v\rangle/\partial c)/n$ against the degree of aggregation n; these energetic quantities are plotted per monomer as in Figures 3 and 4 and the derivative of the free energy is provided for comparison. When the interaction components are compared among the electrostatic, van der Waals, and excluded-volume, the n dependence is the strongest for the van der Waals component. It was actually observed in Ref. [30] that the contributions from the cosolvent and water vary with n in compensating manner for the electrostatic and excluded-volume components and that the van der Waals component depends on n in parallel with its cosolvent contribution. According to Figure 6, furthermore, the n dependence of $(\partial\langle \Delta v\rangle/\partial c)/n$ is correlated most strongly to the van der Waals component among the three interaction components examined. The van der Waals interaction is most influential on the cosolvent-induced change in the solvation free energy, which determines the cosolvent effect on the aggregate formation in turn. The dominant role of the van der Waals component

was also pointed out when the effect of the urea cosolvent on the solvation energetics was analyzed for amino-acid analogs, cytochrome *c*, and T4-lysozyme [84,86,90,91]. Figure 6 shows the importance of the van der Waals component in the cosolvent effects of urea and DMSO on the equilibrium of aggregate formation.

Conclusion

The present review described a computational study of the cosolvent effect on the equilibrium of peptide aggregation. The analysis was based on the property that the modulation of the excess chemical potential of a flexible solute species due to addition of a cosolvent is determined by the corresponding change in the solvation free energy averaged over the solute structures. This is a general feature reflecting the variational principle, and when the cosolvent concentration is low enough, the averaging is to be done only over the solute configurations sampled in pure-water solvent. The cosolvent effects of urea and DMSO were then analyzed by focusing on an 11-residue peptide, NACore, and adopting the energy-representation method to compute the solvation free energy with all-atom model. It was observed that the solvation free energy becomes more favorable upon addition of the cosolvent at each degree of aggregation examined and varies to a lesser extent for larger aggregate. The urea and DMSO cosolvents thus stabilize the monomer more strongly than the aggregates, and act to inhibit the formation of the aggregate. Urea is a better inhibitor than DMSO, and the importance of the van der Waals component was pointed out in the solvation thermodynamics of aggregate formation.

According to Eq. 10, the cosolvent-induced change in the solvation free energy determines the cosolvent effect on the equilibrium for a conversion of chemical species. Equation 10 is an exact expression derived from the variational principle of Eq. 5, and an explicit treatment for the configurational entropy (chain entropy) is made unnecessary even for a highly flexible molecule such as protein. This is useful in theoretical-computational analyses of cosolvent effects since it is still not straightforward to compute the configurational entropy. The affinities of ligand-receptor binding and protein-protein docking are affected strongly by the solvent water, and may thus be modulated by changing the solvent environment. Therefore, Eq. 10 or 11 can be a basis for exploring the effect of the environment on biomolecular functions systematically.

Acknowledgments

The authors are grateful to Prof. Kang Kim of Osaka University, Dr. Yu Yamamori of National Institute of Advanced Industrial Science and Technology, and Prof. Seishi Shimizu of University of York. This work is supported by the Grants-in-Aid for Scientific Research (Nos. JP26240045 and JP19H04206) from the Japan Society for the Promotion of Science and by the Elements Strategy Initiative for Catalysts and Batteries and the Post-K Supercomputing Project from the Ministry of Education, Culture, Sports, Science, and Technology. Finally but not less importantly, we congratulate the 80th anniversary of Prof. Nobuhiro Gō. N. M. started his theoretical/computational studies of solutions and biomolecules under Prof. Gō's influence and is honored to make a contribution to this anniversary issue.

Conflict of Interest

The authors declare no conflict of interest.

Author Contribution

N. M. directed the entire work, analyzed the data, and wrote the manuscript. K. M. performed the calculations and analyzed the data.

References

[1] Jarvis, S. & Mostaert, A. (eds.) *The Functional Fold*. (Pan Stanford, Singapore, 2012).
[2] Otzen, D. E. (ed.) *Amyloid Fibrils and Prefibrillar Aggregates*. (Wiley-VCH, Weinheim, 2013).
[3] Kim, Y. E., Hipp, M. S., Bracher, A., Hayer-Hartl, M. & Ulrich Hartl, F. Molecular Chaperone Functions in Protein Folding and Proteostasis. *Annu. Rev. Biochem.* **82**, 323–355 (2013).
[4] Labbadia, J. & Morimoto, R. I. The biology of proteostasis in aging and disease. *Annu. Rev. Biochem.* **84**, 435–464 (2015).
[5] Silva, J. L. & Weber, G. Pressure Stability of Proteins. *Annu. Rev. Phys. Chem.* **44**, 89–113 (1993).
[6] Damas, A. M. & Saraiva, M. J. Review: TTR amyloidosis—Structural features leading to protein aggregation and their implications on therapeutic strategies. *J. Struct. Biol.* **130**, 290–299 (2000).
[7] Morris, A. M., Watzky, M. A. & Finke, R. G. Protein aggregation kinetics, mechanism, and curve-fitting: A review of the literature. *Biochim. Biophys. Acta* **1794**, 375–397 (2009).
[8] Wang, W., Nema, S. & Teagarden, D. Protein aggregation-Pathways and influencing factors. *Int. J. Pharm.* **390**, 89–99 (2010).
[9] Johnson, S. M., Connelly, S., Fearns, C., Powers, E. T. & Kelly, J. W. The transthyretin amyloidoses: From delineating the molecular mechanism of aggregation linked to pathology to a regulatory-agency-approved drug. *J. Mol. Biol.* **421**, 185–203 (2012).
[10] Perchiacca, J. M. & Tessier, P. M. Engineering Aggregation-Resistant Antibodies. *Annu. Rev. Chem. Biomol. Eng.* **3**, 263–286 (2012).
[11] Klimov, D. K. & Thirumalai, D. Dissecting the assembly of Abeta16-22 amyloid peptides into antiparallel beta sheets. *Structure* **11**, 295–307 (2003).
[12] Klimov, D. K., Straub, J. E. & Thirumalai, D. Aqueous urea solution destabilizes Abeta(16–22) oligomers. *Proc. Natl. Acad. Sci. USA* **101**, 14760–14765 (2004).
[13] Urbanc, B., Cruz, L., Ding, F., Sammond, D., Khare, S., Buldyrev, S. V., *et al.* Molecular dynamics simulation of amyloid b dimer formation. *Biophys. J.* **87**, 2310–2321 (2004).
[14] Zheng, J., Ma, B., Tsai, C.-J. & Nussinov, R. Structural stability and dynamics of an amyloid-forming peptide GNNQQNY from the yeast prion sup-35. *Biophys. J.* **91**, 824–833 (2006).

[15] Nguyen, P. H., Li, M. S., Stock, G., Straub, J. E. & Thirumalai, D. Monomer adds to preformed structured oligomers of Abeta-peptides by a two-stage dock-lock mechanism. *Proc. Natl. Acad. Sci. USA* **104**, 111–116 (2007).
[16] Reddy, G., Straub, J. E. & Thirumalai, D. Dynamics of locking of peptides onto growing amyloid fibrils. *Proc. Natl. Acad. Sci. USA* **106**, 11948–11953 (2009).
[17] Berhanu, W. M. & Masunov, A. E. Alternative packing modes leading to amyloid polymorphism in five fragments studied with molecular dynamics. *Biopolymers* **98**, 131–144 (2012).
[18] Chong, S.-H. & Ham, S. Atomic-level investigations on the amyloid-β dimerization process and its driving forces in water. *Phys. Chem. Chem. Phys.* **14**, 1573–1575 (2012).
[19] Itoh, S. G. & Okumura, H. Dimerization process of amyloid-β (29–42) studied by the hamiltonian replica-permutation molecular dynamics simulations. *J. Phys. Chem. B* **118**, 11428–11436 (2014).
[20] Chong, S.-H. & Ham, S. Distinct role of hydration water in protein misfolding and aggregation revealed by fluctuating thermodynamics analysis. *Acc. Chem. Res.* **48**, 956–965 (2015).
[21] Sahu, K. K., Woodside, M. T. & Tuszynski, J. A. α-Synuclein Dimer Structures Found From Computational Simulations. *Biochimie* **116**, 133–140 (2015).
[22] Tomar, D. S., Weber, V., Pettitt, B. M. & Asthagiri, D. Importance of Hydrophilic Hydration and Intramolecular Interactions in the Thermodynamics of Helix-Coil Transition and Helix-Helix Assembly in a Deca-Alanine Peptide. *J. Phys. Chem. B* **120**, 69–76 (2016).
[23] Carballo-Pacheco, M. & Strodel, B. Advances in the Simulation of Protein Aggregation at the Atomistic Scale. *J. Phys. Chem. B* **120**, 2991–2999 (2016).
[24] Itoh, S. G. & Okumura, H. Oligomer Formation of Amyloid-β (29–42) from Its Monomers Using the Hamiltonian Replica-Permutation Molecular Dynamics Simulation. *J. Phys. Chem. B* **120**, 6555–6561 (2016).
[25] Shirai, N. C. & Kikuchi, M. The interplay of intrinsic disorder and macromolecular crowding on α-synuclein fibril formation. *J. Chem. Phys.* **144**, 055101 (2016).
[26] Kuroda, Y., Suenaga, A., Sato, Y., Kosuda, S. & Taiji, M. All-atom molecular dynamics analysis of multi-peptide systems reproduces peptide solubility in line with experimental observations. *Sci. Rep.* **6**, 19479 (2016).
[27] Man, V. H., Nguyen, P. H. & Derreumaux, P. High-Resolution Structures of the Amyloid-β 1-42 Dimers from the Comparison of Four Atomistic Force Fields. *J. Phys. Chem. B* **121**, 5977–5987 (2017).
[28] Tesei, G., Vazdar, M., Jensen, M. R., Cragnell, C., Mason, P. E., Heyda, J., *et al.* Self-association of a highly charged arginine-rich cell-penetrating peptide. *Proc. Natl. Acad. Sci. USA* **114**, 11428–11433 (2017).
[29] van der Munnik, N. P., Sajib, M. S. J., Moss, M. A., Wei, T. & Uline, M. J. Determining the Potential of Mean Force for Amyloid-β Dimerization: Combining Self-Consistent Field Theory with Molecular Dynamics Simulation. *J. Chem. Theory Comput.* **14**, 2696–2704 (2018).
[30] Masutani, K., Yamamori, Y., Kim, K. & Matubayasi, N. Free-energy analysis of the hydration and cosolvent effects on the β-sheet aggregation through all-atom molecular dynamics simulation. *J. Chem. Phys.* **150**, 145101 (2019).
[31] Harper, J. D. & Lansbury, P. T. Models of Amyloid Seeding in Alzheimer's Disease and Scrapie: Mechanistic Truths and Physiological Consequences of the Time-Dependent Solubility of Amyloid Proteins. *Annu. Rev. Biochem.* **66**, 385–407 (1997).
[32] McLaurin, J., Yang, D. S., Yip, C. M. & Fraser, P. E. Review: Modulating factors in amyloid-β fibril formation. *J. Struct. Biol.* **130**, 259–270 (2000).
[33] Collinge, J. Prion Diseases of Humans and Animals: Their Causes and Molecular Basis. *Annu. Rev. Neurosci.* **24**, 519–550 (2001).
[34] Chatani, E. & Goto, Y. Structural stability of amyloid fibrils of β_2-microglobulin in comparison with its native fold. *Biochim. Biophys. Acta* **1753**, 64–75 (2005).
[35] Tycko, R. Molecular structure of amyloid fibrils: Insights from solid-state NMR. *Q. Rev. Biophys.* **39**, 1–55 (2006).
[36] Eisenberg, D. & Jucker, M. The amyloid state of proteins in human diseases. *Cell* **148**, 1188–1203 (2012).
[37] Bemporad, F. & Chiti, F. Protein misfolded oligomers: Experimental approaches, mechanism of formation, and structure-toxicity relationships. *Chem. Biol.* **19**, 315–327 (2012).
[38] Rodriguez, J. A., Ivanova, M. I., Sawaya, M. R., Cascio, D., Reyes, F. E., Shi, D., *et al.* Structure of the toxic core of α-synuclein from invisible crystals. *Nature* **525**, 486–490 (2015).
[39] Watanabe-Nakayama, T., Ono, K., Itami, M., Takahashi, R., Teplow, D. B. & Yamada, M. High-speed atomic force microscopy reveals structural dynamics of amyloid β_{1-42} aggregates. *Proc. Natl. Acad. Sci. USA* **113**, 5835–5840 (2016).
[40] Auluck, P. K., Caraveo, G. & Lindquist, S. α-Synuclein: Membrane Interactions and Toxicity in Parkinson's Disease. *Annu. Rev. Cell Dev. Biol.* **26**, 211–233 (2010).
[41] Recasens, A. & Dehay, B. Alpha-synuclein spreading in Parkinson's disease. *Front. Neuroanat.* **8**, 159 (2014).
[42] Xu, L. & Pu, J. Alpha-Synuclein in Parkinson's Disease: From Pathogenetic Dysfunction to Potential Clinical Application. *Parkinson's Disease* **2016**, 1720621 (2016).
[43] Karplus, M. & Kushick, J. N. Method for estimating the configurational entropy of macromolecules. *Macromolecules* **14**, 325–332 (1981).
[44] Nola, A. D., Berendsen, H. & Edholm, O. Free energy determination of polypeptide conformations generated by molecular dynamics. *Macromolecules* **17**, 2044–2050 (1984).
[45] Schlitter, J. Estimation of absolute and relative entropies of macromolecules using the covariance matrix. *Chem. Phys. Lett.* **215**, 617–621 (1993).
[46] Chang, C. E. & Gilson, M. K. Free Energy, Entropy, and Induced Fit in Host-Guest Recognition: Calculations with the Second-Generation Mining Minima Algorithm. *J. Am. Chem. Soc.* **126**, 13156–13164 (2004).
[47] King, B. M. & Tidor, B. MIST: Maximum Information Spanning Trees for dimension reduction of biological data sets. *Bioinformatics* **25**, 1165–1172 (2009).
[48] Hensen, U., Lange, O. F. & Grubmüller, H. Estimating Absolute Configurational Entropies of Macromolecules: The Minimally Coupled Subspace Approach. *PLoS ONE* **5**, e9179 (2010).
[49] Harpole, K. W. & Sharp, K. A. Calculation of Configurational Entropy with a Boltzmann-Quasiharmonic Model: The Origin of High-Affinity Protein-Ligand Binding. *J. Phys. Chem. B* **115**, 9461–9472 (2011).
[50] Hikiri, S., Yoshidome, T. & Ikeguchi, M. Computational Methods for Configurational Entropy Using Internal and Cartesian Coordinates. *J. Chem. Theory Comput.* **12**, 5990–6000 (2016).
[51] Frenkel, D. & Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications* (Academic Press, London, UK, 1996).
[52] Matubayasi, N. & Nakahara, M. Theory of Solutions in the Energetic Representation. I. Formulation. *J. Chem. Phys.* **113**, 6070–6081 (2000).
[53] Matubayasi, N. & Nakahara, M. Theory of Solutions in the Energy Representation. II. Functional for the Chemical Potential. *J. Chem. Phys.* **117**, 3605–3616 (2002).
[54] Matubayasi, N. & Nakahara, M. Erratum: "Theory of Solutions in the Energy Representation. II. Functional for the

Chemical Potential" [J. Chem. Phys. 117, 3605 (2002)]. *J. Chem. Phys.* **118**, 2446 (2003).
[55] Matubayasi, N. & Nakahara, M. Theory of Solutions in the Energy Representation. III. Treatment of the Molecular Flexibility. *J. Chem. Phys.* **119**, 9686–9702 (2003).
[56] Sakuraba, S. & Matubayasi, N. Ermod: Fast and Versatile Computation Software for Solvation Free Energy with Approximate Theory of Solutions. *J. Comput. Chem.* **35**, 1592–1608 (2014). Footnote 1 on page 1598 of this paper was incorrectly stated. The TIP3P model used for the calculations was the CHARMM-modified one with a Lennard-Jones term on the hydrogen atom.
[57] Levy, R. M., Belhadj, M. & Kitchen, D. B. Gaussian Fluctuation Formula for Electrostatic Free-Energy Changes in Solution. *J. Chem. Phys.* **95**, 3627–3633 (1991).
[58] Luzhkov, V. & Warshel, A. Microscopic Models for Quantum Mechanical Calculations of Chemical Processes in Solutions: LD/AMPAC and SCAAS/AMPAC Calculations of Solvation Energies. *J. Comput. Chem.* **13**, 199–213 (1992).
[59] Åqvist, J., Medina, C. & Samuelsson, J. E. A New Method for Predicting Binding Affinity in Computer-Aided Drug Design. *Protein Eng. Des. Sel.* **7**, 385–391 (1994).
[60] Carlson, H. A. & Jorgensen, W. L. An Extended Linear Response Method for Determining Free Energies of Hydration. *J. Phys. Chem.* **99**, 10667–10673 (1995).
[61] Kast, S. M. Combinations of Simulation and Integral Equation Theory. *Phys. Chem. Chem. Phys.* **3**, 5087–5092 (2001).
[62] Vener, M. V., Leontyev, I. V., Dyakov, Y. A., Basilevsky, M. V. & Newton, M. D. Application of the Linearized MD Approach for Computing Equilibrium Solvation Free Energies of Charged and Dipolar Solutes in Polar Solvents. *J. Phys. Chem. B* **106**, 13078–13088 (2002).
[63] Hirata, F. (ed.) *Molecular Theory of Solvation*. (Kluwer Academic Publishers Dordrecht, Netherlands, 2003).
[64] Galván, I. F., Sanchez, M. L., Martin, M. E., Olivares del Valle, F. J. & Aguilar, M. A. Geometry Optimization of Molecules in Solution: Joint Use of the Mean Field Approximation and the Free-Energy Gradient Method. *J. Chem. Phys.* **118**, 255–263 (2003).
[65] Freedman, H. & Truong, T. N. Coupled Reference Interaction Site Model/Simulation Approach for Thermochemistry of Solvation: Theory and Prospects. *J. Chem. Phys.* **121**, 2187–2198 (2004).
[66] Chuev, G. N., Fedorov, M. V. & Crain, J. Improved Estimates for Hydration Free Energy Obtained by the Reference Interaction Site Model. *Chem. Phys. Lett.* **448**, 198–202 (2007).
[67] Yamamoto, T. Variational and Perturbative Formulations of Quantum Mechanical/Molecular Mechanical Free Energy with Mean-Field Embedding and Its Analytical Gradients. *J. Chem. Phys.* **129**, 244104 (2008).
[68] Kokubo, H., Hu, C. Y. & Pettitt, B. M. Peptide Conformational Preferences in Osmolyte Solutions: Transfer Free Energies of Decaalanine. *J. Am. Chem. Soc.* **133**, 1849–1858 (2011).
[69] Frolov, A. I., Ratkova, E. L., Palmer, D. S. & Fedorov, M. V. Hydration Thermodynamics Using the Reference Interaction Site Model: Speed or Accuracy? *J. Phys. Chem. B* **115**, 6011–6022 (2011).
[70] Weber, V. & Asthagiri, D. Regularizing Binding Energy Distributions and the Hydration Free Energy of Protein Cytochrome C from All-Atom Simulations. *J. Chem. Theory Comput.* **8**, 3409–3415 (2012).
[71] Liu, Y., Zhao, S. & Wu, J. A Site Density Functional Theory for Water: Application to Solvation of Amino Acid Side Chains. *J. Chem. Theory Comput.* **9**, 1896–1908 (2013).
[72] Sergiievskyi, V., Jeanmairet, G., Levesque, M. & Borgis, D. Solvation free-energy pressure corrections in the three dimensional reference interaction site model. *J. Chem. Phys.* **143**, 184116 (2015).
[73] Ratkova, E. L., Palmer, D. S. & Fedorov, M. V. Solvation Thermodynamics of Organic Molecules by the Molecular Integral Equation Theory: Approaching Chemical Accuracy. *Chem. Rev.* **115**, 6312–6356 (2015).
[74] Ou, S.-C., Drake, J. A. & Pettitt, B. M. Nonpolar Solvation Free Energy from Proximal Distribution Functions. *J. Phys. Chem. B* **121**, 3555–3564 (2017).
[75] Remsing, R. C., Xi, E. & Patel, A. J. Protein Hydration Thermodynamics: The Influence of Flexibility and Salt on Hydrophobin II Hydration. *J. Phys. Chem. B* **122**, 3635–3646 (2018).
[76] Takahashi, H., Matubayasi, N., Nakahara, M. & Nitta, T. A Quantum Chemical Approach to the Free Energy Calculations in Condensed Systems: The QM/MM Method Combined with the Theory of Energy Representation. *J. Chem. Phys.* **121**, 3989–3999 (2004).
[77] Matubayasi, N., Liang, K. K. & Nakahara, M. Free-Energy Analysis of Solubilization in Micelle. *J. Chem. Phys.* **124**, 154908 (2006).
[78] Matubayasi, N., Shinoda, W. & Nakahara, M. Free-Energy Analysis of the Molecular Binding into Lipid Membrane with the Method of Energy Representation. *J. Chem. Phys.* **128**, 195107 (2008).
[79] Takahashi, H., Ohno, H., Kishi, R., Nakano, M. & Matubayasi, N. Computation of the Free Energy Change Associated with One-Electron Reduction of Coenzyme Immersed in Water: A Novel Approach within the Framework of the Quantum Mechanical/Molecular Mechanical Method Combined with the Theory of Energy Representation. *J. Chem. Phys.* **129**, 205103 (2008).
[80] Karino, Y., Fedorov, M. V. & Matubayasi, N. End-Point Calculation of Solvation Free Energy of Amino-Acid Analogs by Molecular Theories of Solution. *Chem. Phys. Lett.* **496**, 351–355 (2010).
[81] Karino, Y. & Matubayasi, N. Communication: Free-Energy Analysis of Hydration Effect on Protein with Explicit Solvent: Equilibrium Fluctuation of Cytochrome *c*. *J. Chem. Phys.* **134**, 041105 (2011).
[82] Kawakami, T., Shigemoto, I. & Matubayasi, N. Free-Energy Analysis of Water Affinity in Polymer Studied by Atomistic Molecular Simulation Combined with the Theory of Solutions in the Energy Representation. *J. Chem. Phys.* **137**, 234903 (2012).
[83] Kawakami, T., Shigemoto, I. & Matubayasi, N. Erratum: "Free-Energy Analysis of Water Affinity in Polymer Studied by Atomistic Molecular Simulation Combined with the Theory of Solutions in the Energy Representation" [J. Chem. Phys. 137, 234903 (2012)]. *J. Chem. Phys.* **140**, 169903 (2014).
[84] Karino, Y. & Matubayasi, N. Interaction-Component Analysis of the Urea Effect on Amino Acid Analogs. *Phys. Chem. Chem. Phys.* **15**, 4377–4391 (2013).
[85] Frolov, A. I. Accurate Calculation of Solvation Free Energies in Supercritical Fluids by Fully Atomistic Simulations: Probing the Theory of Solutions in Energy Representation. *J. Chem. Theory Comput.* **11**, 2245–2256 (2015).
[86] Yamamori, Y., Ishizuka, R., Karino, Y., Sakuraba, S. & Matubayasi, N. Interaction-Component Analysis of the Hydration and Urea Effects on Cytochrome *c*. *J. Chem. Phys.* **144**, 085102 (2016).
[87] Kamo, F., Ishizuka, R. & Matubayasi, N. Correlation analysis for heat denaturation of Trp-cage miniprotein with explicit solvent. *Protein Sci.* **25**, 56–66 (2016).
[88] Date, A., Ishizuka, R. & Matubayasi, N. Energetics of Nonpolar and Polar Compounds in Cationic, Anionic, and Nonionic Micelles Studied by All-Atom Molecular Dynamics Simulation Combined with a Theory of Solutions. *Phys. Chem. Chem. Phys.* **18**, 13223–13231 (2016).
[89] Harris, R. C., Deng, N., Levy, R. M., Ishizuka, R. & Matubayasi,

N. Computing Conformational Free Energy Differences in Explicit Solvent: An Efficient Thermodynamic Cycle Using an Auxiliary Potential and a Free Energy Functional Constructed from the End Points Solvation Free Energy with Approximate Theory of Solutions. *J. Comput. Chem.* **38**, 1198–1208 (2017).

[90] Yamamori, Y. & Matubayasi, N. Interaction-Component Analysis of the Effects of Urea and Its Alkylated Derivatives on the Structure of T4-Lysozyme. *J. Chem. Phys.* **146**, 225103 (2017).

[91] Matubayasi, N. Free-energy analysis of protein solvation with all-atom molecular dynamics simulation combined with a theory of solutions. *Curr. Opin. Struct. Biol.* **43**, 45–54 (2017).

[92] Mizuguchi, T. & Matubayasi, N. Free-Energy Analysis of Peptide Binding in Lipid Membrane Using All-Atom Molecular Dynamics Simulation Combined with Theory of Solutions. *J. Phys. Chem. B* **122**, 3219–3229 (2018).

[93] Tokunaga, Y., Yamamori, Y. & Matubayasi, N. Probabilistic analysis for identifying the driving force of protein folding. *J. Chem. Phys.* **148**, 125101 (2018).

[94] Kawakami, T., Shigemoto, I. & Matubayasi, N. Structure and permeability of ionomers studied by atomistic molecular simulation combined with the theory of solutions in the energy representation. *J. Chem. Phys.* **148**, 214903 (2018).

[95] Yamamoto, N., Nakakuki, I. & Matubayasi, N. Free-energy analysis of physisorption on solid-liquid interface with the solution theory in the energy representation. *J. Chem. Phys.* **149**, 014504 (2018).

[96] Ishii, Y., Yamamoto, N., Matubayasi, N., Zhang, B. W., Cui, D. & Levy, R. M. Spatially-Decomposed Free Energy of Solvation Based on the Endpoint Density-Functional Method. *J. Chem. Theory Comput.* **15**, 2896–2912 (2019).

[97] Saito, H., Matubayasi, N., Nishikawa, K. & Nagao, H. Hydration property of globular proteins: An analysis of solvation free energy by energy representation method. *Chem. Phys. Lett.* **497**, 218–222 (2010).

[98] Mizukami, T., Saito, H., Kawamoto, S., Miyakawa, T., Iwayama, M., Takasu, M., *et al.* Solvation Effect on the Structural Change of a Globular Protein: A Molecular Dynamics Study. *Int. J. Quant. Chem.* **112**, 344–350 (2012).

[99] Takemura, K., Guo, H., Sakuraba, S., Matubayasi, N. & Kitao, A. Evaluation of protein-protein docking model structures using all-atom molecular dynamics simulations combined with the solution theory in the energy representation. *J. Chem. Phys.* **137**, 215105 (2012).

[100] Saito, H., Iwayama, M., Mizukami, T., Kang, J., Tateno, M. & Nagao, H. Molecular dynamics study on binding free energy of AzurinCytochrome *c*551 complex. *Chem. Phys. Lett.* **556**, 297–302 (2013).

[101] Takemura, K., Matubayasi, N. & Kitao, A. Binding free energy analysis of protein-protein docking model structures by evERdock. *J. Chem. Phys.* **148**, 105101 (2018).

[102] Shinobu, A., Takemura, K., Matubayasi, N. & Kitao, A. Refining evERdock: Improved selection of good protein-protein complex models achieved by MD optimization and use of multiple conformations. *J. Chem. Phys.* **149**, 195101 (2018).

[103] Ben-Naim, A. *Solvation Thermodynamics*. (Plenum Press, New York, 1987).

[104] Shimizu, S. Estimating hydration changes upon biomolecular reactions from osmotic stress, high pressure, and preferential hydration experiments. *Proc. Natl. Acad. Sci. USA* **101**, 1195–1199 (2004).

[105] Shulgin, I. L. & Ruckenstein, E. A protein molecule in an aqueous mixed solvent: Fluctuation theory outlook. *J. Chem. Phys.* **123**, 054909 (2005).

[106] Jiao, Y. & Smith, P. E. Fluctuation theory of molecular association and conformational equilibria. *J. Chem. Phys.* **135**, 014502 (2011).

[107] Shimizu, S. & Matubayasi, N. Hydrotropy: Monomer-Micelle Equilibrium and Minimum Hydrotrope Concentration. *J. Phys. Chem. B* **118**, 10515–10524 (2014).

[108] Matubayasi, N. & Levy, R. M. Thermodynamics of the Hydration Shell. 2. Excess Volume and Compressibility of a Hydrophobic Solute. *J. Phys. Chem.* **100**, 2681–2688 (1996).

[109] Schnell, S. K., Liu, X., Simon, J.-M., Bardow, A., Bedeaux, D., Vlugt, T. J. H., *et al.* Calculating Thermodynamic Properties from Fluctuations at Small Scales. *J. Phys. Chem. B* **115**, 10911–10918 (2011).

[110] Ganguly, P. & van der Vegt, N. F. A. Convergence of Sampling Kirkwood-Buff Integrals of Aqueous Solutions with Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **9**, 1347–1355 (2013).

[111] Cortes-Huerto, R., Kremer, K. & Potestio, R. Communication: Kirkwood-Buff integrals in the thermodynamic limit from small-sized molecular dynamics simulations. *J. Chem. Phys.* **145**, 141103 (2016).

[112] Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983). Water is treated as rigid and the hydrogen atom does not have the Lennard-Jones term.

[113] Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A. & Simmerling, C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins: Struct. Funct. Bioinf.* **65**, 712–725 (2006).

[114] Bayly, C. I., Cieplak, P., Cornell, W. D. & Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.* **97**, 10269–10280 (1993).

[115] Pronk, S., Páll, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., *et al.* GROMACS 4.5: A High-Throughput and Highly Parallel Open Source Molecular Simulation Toolkit. *Bioinformatics* **29**, 845–854 (2013).

[116] Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., *et al.* GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **1-2**, 19–25 (2015).

[117] Hirota-Nakaoka, N., Hasegawa, K., Naiki, H. & Goto, Y. Dissolution of β_2-Microglobulin Amyloid Fibrils by Dimethylsulfoxide. *J. Biochem.* **134**, 159–164 (2003).

[118] Kim, J. R., Muresan, A., Lee, K. Y. C. & Murphy, R. M. Urea modulation of β-amyloid fibril growth: Experimental studies and kinetic models. *Protein Sci.* **13**, 2888–2898 (2004).

[119] Hayden, E. Y., Conovaloff, J. L., Mason, A., Bitan, G. & Teplow, D. B. Preparation of pure populations of covalently stabilized amyloid β-protein oligomers of specific sizes. *Anal. Biochem.* **518**, 78–85 (2017).

[120] Kinoshita, M. Roles of Translational Motion of Water Molecules in Sustaining Life. *Front. Biosci.* **14**, 3419–3454 (2009).

[121] Kinoshita, M. A New Theoretical Approach to Biological Self-Assembly. *Biophys. Rev.* **5**, 283–293 (2013).

[122] Oshima, H. & Kinoshita, M. Essential roles of protein-solvent many-body correlation in solvent-entropy effect on protein folding and denaturation: Comparison between hard-sphere solvent and water. *J. Chem. Phys.* **142**, 145103 (2015).