ACCOUNTS of chemical research

Article

pubs.acs.org/accounts

Distinct Role of Hydration Water in Protein Misfolding and Aggregation Revealed by Fluctuating Thermodynamics Analysis

Song-Ho Chong and Sihyun Ham*

Department of Chemistry, Sookmyung Women's University, Cheongpa-ro 47-gil 100, Yongsan-Ku, Seoul 140-742, Korea

CONSPECTUS: Protein aggregation in aqueous cellular environments is linked to diverse human diseases. Protein aggregation proceeds through a multistep process initiated by conformational transitions, called protein misfolding, of monomer species toward aggregation-prone structures. Various forms of aggregate species are generated through the association of misfolded monomers including soluble oligomers and amyloid fibrils. Elucidating the molecular mechanisms and driving forces involved in the misfolding and subsequent association has been a central issue for understanding and preventing protein aggregation diseases such as Alzheimer's, Parkinson's, and type II diabetes.

In this Account, we provide a thermodynamic perspective of the misfolding and aggregation of the amyloid-beta (Aβ) protein implicated in Alzheimer's disease through the application of fluctuating thermodynamics. This approach "dissects" the conventional thermodynamic characterization of the end states into the one of the fluctuating processes connecting them, and enables one to analyze variations in the thermodynamic functions that occur during the course of protein conformational changes. The central quantity in this approach is the solvent-averaged effective energy, $f = E_u + G_{solv}$, comprising the protein potential energy (E_u) and the solvation free energy (G_{solv}), whose time variation reflects the protein dynamics on the free energy landscape. Protein configurational entropy is quantified by the magnitude of fluctuations in f. We find that misfolding of the Aβ monomer when released from a membrane environment to an aqueous phase is driven by favorable changes in protein potential energy and configurational entropy, but it is also accompanied by an unfavorable increase in solvation free energy. The subsequent dimerization of the misfolded Aβ monomers occurs in two steps. The first step, where two widely separated monomers come into contact distance, is driven by water-mediated attraction, that is, by a decrease in solvation free energy, harnessing the monomer solvation free energy earned during the misfolding. The second step, where a compact dimer structure is formed, is driven by direct protein−protein interactions, but again it is accompanied by an increase in solvation free energy. The increased solvation free energy of the dimer will function as the driving force to recruit another Aβ protein in the approach stage of subsequent oligomerizations.

The fluctuating thermodynamics analysis of the misfolding and dimerization of the Aβ protein indicates that the interaction of the protein with surrounding water plays a critical role in protein aggregation. Such a water-centric perspective is further corroborated by demonstrating that, for a large number of Aβ mutants and mutants of other protein systems, the change in the experimental aggregation propensity upon mutation has a significant correlation with the protein solvation free energy change. We also find striking discrimination between the positively and negatively charged residues on the protein surface by surrounding water molecules, which is shown to play a crucial role in determining the protein aggregation propensity. We argue that the protein total charge dictates such striking behavior of the surrounding water molecules. Our results provide new insights for understanding and predicting the protein aggregation propensity, thereby offering novel design principles for producing aggregation-resistant proteins for biotherapeutics.

1. INTRODUCTION

Proteins after synthesis normally fold into a specific, functional three-dimensional structure that is soluble in aqueous environments. However, under certain intrinsic or external perturbations, proteins convert from their native conformations into non-native ones, termed as protein misfolding.[1] The misfolded proteins tend to aggregate by themselves to form dimers, oligomers, and eventually fibrillar deposits called amyloid fibrils.[2] Such protein aggregates are often toxic and are associated with a number of human diseases ranging from neurodegenerative disorders to systemic amyloidoses.[3] Elucidating the molecular mechanisms and driving forces involved in the misfolding and subsequent association has therefore been a central issue for understanding and preventing protein aggregation diseases such as Alzheimer's, Parkinson's, and type II diabetes.[4,5]

In this Account, we present a thermodynamic perspective of the misfolding and aggregation of the amyloid-beta (Aβ) protein implicated in Alzheimer's disease (AD).[6] The Aβ protein is derived through proteolytic cleavage from the transmembrane amyloid-beta precursor protein.[6] The Aβ protein produced in this manner is presumed to possess a helical structure, as inferred from the NMR structure

Received: January 23, 2015
Published: April 6, 2015

DOI: 10.1021/acs.accounts.5b00032
Acc. Chem. Res. 2015, 48, 956−965

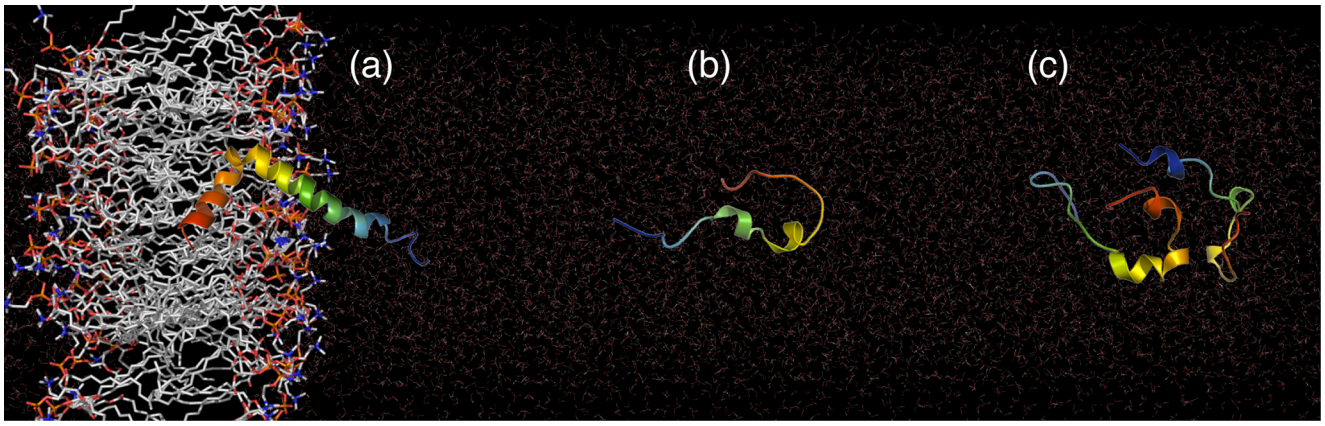

Figure 1. Misfolding of the Aβ protein in an aqueous phase when transferred from a membrane environment, and subsequent dimerization. Here, the initial helical structure (a) is taken from NMR determined in a membrane-mimicking organic solvent (PDB ID: 1IYT),[7] whereas the misfolded (b) and dimer structures (c) are taken from our simulations.

determined in an organic solvent that mimics the membrane environment[7] (Figure 1a). When transferred from a membrane to an aqueous environment, the Aβ protein undergoes conformational transitions (Figure 1b) and exhibits a strong propensity to aggregate.[8] Misfolded Aβ proteins in an aqueous phase self-assemble to form soluble oligomers and eventually insoluble amyloid fibrils. Although amyloid fibrils have long been suspected to be the disease agents in AD, a number of recent studies indicate that Aβ oligomers are the main pathogenic agents.[9] This lead to more focus on the early stage oligomers,[10] in particular the Aβ dimer (Figure 1c), which is the smallest oligomeric species impairing function and synapse structure in the AD brain.[11]

Thermodynamics provides both a conceptual and a practical framework to understand the occurrence of certain process in terms of the free energy. However, the conventional statistical thermodynamic approaches, for example, for protein folding[12] and protein−ligand binding,[13] are concerned only with the end states of a process (the folded and unfolded states in the case of protein folding). For example, Aβ dimerization is just characterized by the free energy difference between the dimer and two monomers. Yet, such a characterization does not answer the following types of questions relevant to intervention and the prevention of aggregation: (1) How and in what sense does the Aβ monomer acquire its aggregation-prone nature during misfolding? (2) In what way is the acquired aggregation-prone nature actually used to drive the subsequent aggregation? A *process-based* description is required to address these fundamental questions, and fluctuating thermodynamics—the main subject of this Account—offers a method for arriving at such a description. This approach "dissects" the conventional thermodynamic characterization of the two end states into the thermodynamic description of the *time-dependent, fluctuating* processes that connect them, and enables one to analyze how the thermodynamic functions vary during biomolecular processes of interest. By further applying the site-directed thermodynamics analysis detailed below, it is possible to identify the parts of the protein contributing the most to the thermodynamic properties (thermodynamic "hot spots"). Using this novel analysis method, we aim to provide a detailed thermodynamic picture on the protein misfolding and aggregation, which may be valuable in developing drugs to prevent protein aggregation and also impact strategies to design aggregation-resistant proteins as biotherapeutics.

2. FLUCTUATING THERMODYNAMICS

Free energy is the fundamental state function in statistical thermodynamics. The free energy of state X (such as the folded and unfolded states), denoted as F_X, is related to the configuration integral Z_X, that is, the potential energy part in the partition function, via $F_X = -k_B T \log Z_X$ with Boltzmann's constant k_B and the temperature T. For a protein dissolved in water, Z_X is given by the integration of the Boltzmann factor over protein and water configurations. When one is primarily interested in the protein configuration (collectively represented as $\boldsymbol{r}_u$), integration can be performed over water configurations concerning the parts in the Boltzmann factor associated with protein−water and water−water interactions. Up to an irrelevant constant and with $\beta^{-1} = k_B T$, this results in the following expression:[12,14]

$$Z_X = \int_X d\boldsymbol{r}_u \exp[-\beta f(\boldsymbol{r}_u)] \tag{1}$$

in terms of the solvent-averaged Boltzmann factor $e^{-\beta f(\boldsymbol{r}_u)}$. Here, the integration is restricted to the region of the protein conformational space associated with state X, and $f(\boldsymbol{r}_u) = E_u(\boldsymbol{r}_u) + G_{solv}(\boldsymbol{r}_u)$ is the solvent-averaged effective energy consisting of the protein potential energy $E_u(\boldsymbol{r}_u)$ and the solvation free energy $G_{solv}(\boldsymbol{r}_u)$.

The free energy difference $F_B - F_A$ between two states A and B is the reversible work for transforming the system from state A to state B. Similarly, the free energy difference can be introduced for a microscopic, or fluctuating, process in which the protein conformation is changed from $\boldsymbol{r}_u$ to $\boldsymbol{r}_u'$ in water, and it is given by $f(\boldsymbol{r}_u') - f(\boldsymbol{r}_u)$. This holds because $f(\boldsymbol{r}_u)$ is precisely the reversible work for a process in which protein constituent atoms are moved from infinite separation in vacuum to a particular conformation $\boldsymbol{r}_u$ in water: $E_u(\boldsymbol{r}_u)$ corresponds to the reversible work to form a protein of conformation $\boldsymbol{r}_u$ in vacuum, and $G_{solv}(\boldsymbol{r}_u)$ accounts for the solvation process. The quantity $f(\boldsymbol{r}_u)$ thus defines a hypersurface in the protein configuration space referred to as the free energy landscape.[12]

This notion of $f(\boldsymbol{r}_u)$ allows us to perform thermodynamic analysis on protein fluctuating processes. Two different types of approaches are conceivable for such fluctuating thermodynamics. The first one is based on the following Langevin-type equation (with $\boldsymbol{r}_{u,i}$ denoting the position of a constituent atom i and m_i its mass)

DOI: 10.1021/acs.accounts.5b00032
Acc. Chem. Res. 2015, 48, 956−965

$$m_i\ddot{\boldsymbol{r}}_{u,i} = -\frac{\partial f(\boldsymbol{r}_u)}{\partial \boldsymbol{r}_{u,i}} - \gamma^0\dot{\boldsymbol{r}}_{u,i} + \boldsymbol{R}_{u,i}(t) \tag{2}$$

in which the bare friction coefficient γ^0 and the random force $\boldsymbol{R}_{u,i}(t)$ are connected via the fluctuation−dissipation relation. By identifying $-\partial f(\boldsymbol{r}_u)/\partial \boldsymbol{r}_u \, \mathrm{d}\boldsymbol{r}_u(t)$ as the reversible work and $(-\gamma^0\dot{\boldsymbol{r}}_u + \boldsymbol{R}_u(t))\mathrm{d}\boldsymbol{r}_u(t)$ as the heat transferred from the environment during the infinitesimal time dt, eq 2 not only generates protein dynamics on the free energy landscape $f(\boldsymbol{r}_u)$, but also allows one to analyze thermodynamics along the trajectory $\boldsymbol{r}_u(t)$.[15] However, this approach is not realistic for proteins since a reliable estimation of $G_{\mathrm{solv}}(\boldsymbol{r}_u)$ to determine $f(\boldsymbol{r}_u)$ is a nontrivial task for such complex systems and repeating this calculation for each integration time step is computationally quite demanding.

Alternatively—and this is the analysis method we employ in this Account under the name of fluctuating thermodynamics—one can perform thermodynamic analysis for protein fluctuating processes by combining molecular dynamics (MD) simulations with the equilibrium distribution function theory for surrounding water molecules. In this approach, MD simulations are first carried out with explicit water to generate protein conformational fluctuations in an aqueous environment. From the simulation trajectories, one then extracts the protein conformation $\boldsymbol{r}_u$ only, replacing surrounding explicit water molecules with their equilibrium distribution. This replacement amounts to assuming time scale separation, which is a reasonable approximation since the time scales for sensible protein conformational changes (typically, nanosecond or longer time scales) are much longer than those for water dynamics (picosecond time scales). The protein potential energy $E_u(\boldsymbol{r}_u)$ can be calculated using the force field employed in the simulations. The equilibrium water distribution can be obtained based on the integral-equation theory of molecular liquids, from which solvation thermodynamic functions such as $G_{\mathrm{solv}}(\boldsymbol{r}_u)$ and its enthalpy and entropy components can be computed: in the applications presented below, the three-dimensional reference interaction site model (3D-RISM) theory[16,17] was employed for this purpose. In this way, one obtains the free energy $f(\boldsymbol{r}_u) = E_u(\boldsymbol{r}_u) + G_{\mathrm{solv}}(\boldsymbol{r}_u)$ along the protein conformational fluctuations. This allows us to address questions such as the role of the intrinsic properties of protein (E_u) and water (G_{solv}) and whether the process is enthalpically or entropically driven.

One can also "integrate" the fluctuating thermodynamics to recover the conventional thermodynamics description. Indeed, the free energy difference $F_{\mathrm{B}} - F_{\mathrm{A}}$ between states A and B can be computed from the statistical properties of $f(\boldsymbol{r}_u)$ in these states.[14,18] To see this, one first rewrites eq 1 as $Z_{\mathrm{X}} = \int \mathrm{d}f W(f)\, \mathrm{e}^{-\beta f}$ with the distribution of f, $W(f) \propto \int_{\mathrm{X}} \mathrm{d}\boldsymbol{r}_u \, \delta(f - f(\boldsymbol{r}_u))$, normalized such that $\int \mathrm{d}f W(f) = 1$. Here, the terms irrelevant to the free energy difference are omitted (see ref 18 for details). $W(f)$ has been found to be close to the Gaussian distribution for folded and unfolded proteins,[18] as well as for intrinsically disordered proteins.[19,20] This holds as a result of the central limit theorem since $f(\boldsymbol{r}_u) = E_u(\boldsymbol{r}_u) + G_{\mathrm{solv}}(\boldsymbol{r}_u)$ comprises the sum of a number of canceling energy terms. In this case, the configuration integral Z_{X}, and hence the free energy F_{X}, can be expressed in terms of the mean ($\bar{f}$) and fluctuations (σ_f^2) of f values sampled in state X[14,18]

$$F_{\mathrm{X}} = \bar{f} - TS_{\mathrm{conf}}; \qquad TS_{\mathrm{conf}} = (\beta/2)\sigma_f^2 \tag{3}$$

Here, the term associated with σ_f^2 is identified as the protein configurational entropy (S_{conf}) since it quantifies the extent to which a protein explores the free energy landscape. We previously demonstrated for the protein villin headpiece subdomain that the protein-folding free energy computed based on eq 3 is in accord with the experimental results.[18]

3. SITE-DIRECTED THERMODYNAMICS

Fluctuating thermodynamics thus allow us to carry out not only the conventional thermodynamic characterization of the end states, but also thermodynamic analysis on fluctuating processes through $f(\boldsymbol{r}_u)$. However, $f(\boldsymbol{r}_u)$ is just a single number for a particular protein conformation $\boldsymbol{r}_u$; it does not by itself tell us which parts of the protein contribute the most to the change in $f(\boldsymbol{r}_u)$. Therefore, a method identifying such thermodynamic hot spots is desirable. To this end, decomposing $f(\boldsymbol{r}_u)$ into various useful forms, for example, into contributions from individual energy components such as electrostatic and Lennard−Jones (LJ) terms and from atom groups such as amino-acid residues and functional groups, is indispensable.

Such a decomposition is straightforward for the protein potential energy $E_u(\boldsymbol{r}_u)$. In contrast, the decomposition of solvation thermodynamic functions such as $G_{\mathrm{solv}}(\boldsymbol{r}_u)$ is not obvious. To demonstrate their decomposability, we first analyze the average protein−water interaction energy,[21,22] $E_{uv}(\boldsymbol{r}_u) = \langle E_{uv}(\boldsymbol{r}_u,\boldsymbol{r}_v)\rangle$, where $E_{uv}(\boldsymbol{r}_u,\boldsymbol{r}_v)$ is the protein−water interaction potential and $\langle\cdots\rangle$ is the equilibrium average over water configurations $\boldsymbol{r}_v$. Since $E_{uv}(\boldsymbol{r}_u,\boldsymbol{r}_v)$ consists of interaction potentials (u_{ij}) between the constituent atoms of the protein and those of water, we obtain $E_{uv}(\boldsymbol{r}_u) = \Sigma_{i\in u}\Sigma_{j\in v}\langle u_{ij}(|\boldsymbol{r}_{u,i} - \boldsymbol{r}_{v,j}|)\rangle$. Clearly, $E_{uv}(\boldsymbol{r}_u)$ is decomposable into atomic contributions. Since u_{ij} is typically a sum of LJ and electrostatic terms, $E_{uv}(\boldsymbol{r}_u)$ can further be partitioned into these energy terms. Finally, we note that this expression for $E_{uv}(\boldsymbol{r}_u)$ can be written in terms of the equilibrium radial distribution function $g_{ij}(r;\boldsymbol{r}_u)$ as

$$E_{uv}(\boldsymbol{r}_u) = \sum_{i\in u}\sum_{j\in v}\int 4\pi r^2 \, \mathrm{d}r u_{ij}(r)\rho g_{ij}(r;\, \boldsymbol{r}_u) \tag{4}$$

where ρ is the average number density of water.

An exact expression similar to eq 4 can be derived for the solvation free energy $G_{\mathrm{solv}}(\boldsymbol{r}_u)$,[22]

$$G_{\mathrm{solv}}(\boldsymbol{r}_u) = \sum_{i\in u}\sum_{j\in v}\int_0^1 \mathrm{d}\lambda \int 4\pi r^2 \, \mathrm{d}r \frac{\partial u_{ij}(r;\,\lambda)}{\partial\lambda}\rho g_{ij}(r;\, \boldsymbol{r}_u;\, \lambda) \tag{5}$$

This expression explicitly demonstrates that $G_{\mathrm{solv}}(\boldsymbol{r}_u)$ is decomposable. The only essential difference from eq 4 is the "charging parameter" λ, which gradually turns on the protein−water interaction from no interaction ($\lambda = 0$) to full interaction ($\lambda = 1$) along some path. $g_{ij}(r;\boldsymbol{r}_u;\lambda)$ is the equilibrium radial distribution function when the charging parameter is λ, which can also be computed from the integral-equation theory such as the 3D-RISM theory. Various paths for λ may be considered, and so the decomposition represented by eq 5 is exact but not unique. (We notice that the integrated result for $G_{\mathrm{solv}}(\boldsymbol{r}_u)$ is independent of the path.) However, the most "natural" path exists, which is to first turn on the LJ interaction and then the electrostatic interaction.[22] The corresponding decomposition of the solvation enthalpy and solvation entropy also follows from the temperature derivative of eq 5.[23,24] Thus, $f(\boldsymbol{r}_u)$ and its enthalpy and entropy components are decomposable, and such site-directed analysis elucidates the molecular origin of the

DOI: 10.1021/acs.accounts.5b00032
Acc. Chem. Res. 2015, 48, 956−965

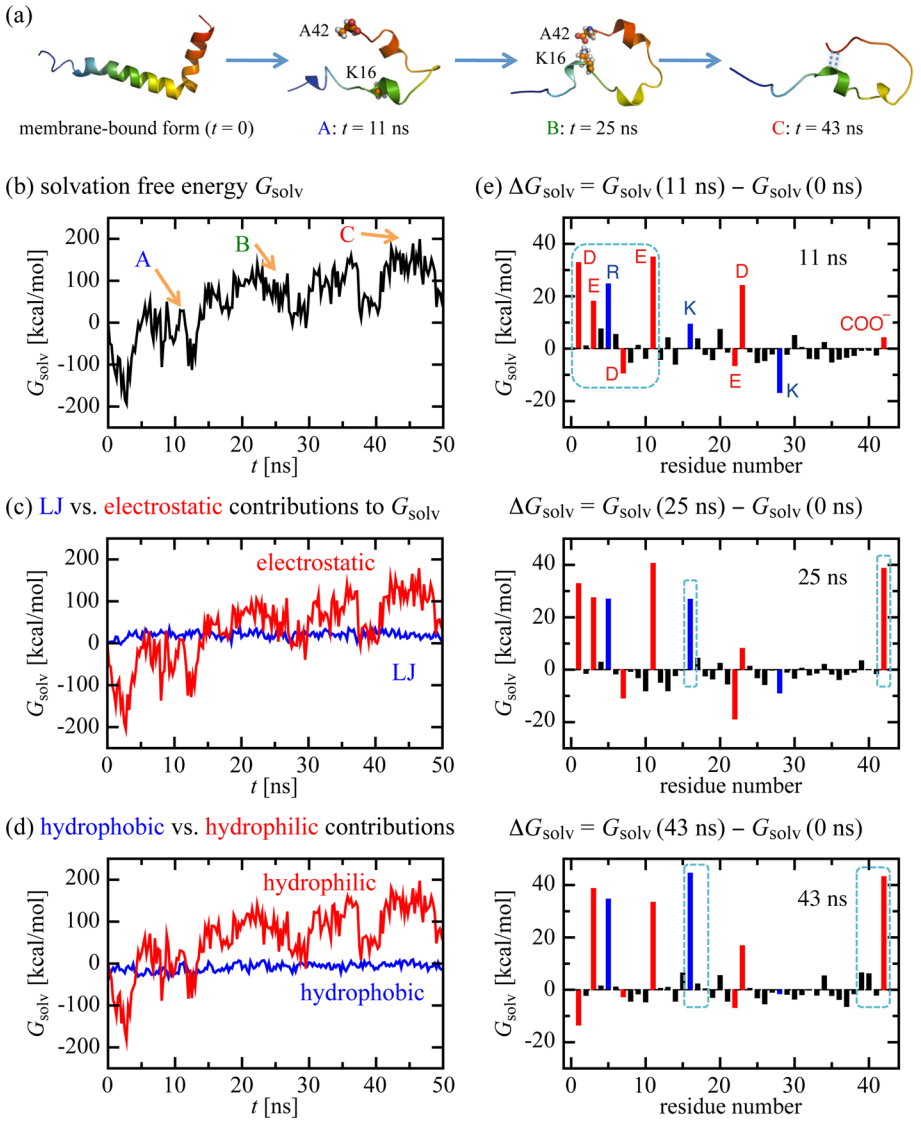

Figure 2. Fluctuating thermodynamics analysis of the Aβ42 misfolding. (a) A representative misfolding process. The positions of the residues K16 and A42 are indicated with sphere representation in the 11 and 25 ns structures. (b) Solvation free energy G_{solv} versus simulation time (the initial value is set to zero). (c) Decomposition of G_{solv} into the LJ (blue) and electrostatic (red) terms. (d) Decomposition of G_{solv} into hydrophobic residue (blue) and hydrophilic residue (red) contributions. (e) Site-resolved G_{solv} at 11 ns (top), 25 ns (middle), and 43 ns (bottom). Contributions from positively charged residues are colored with blue, negatively charged residues with red, and neutral residues with black. Dashed boxes indicate the residue regions referred to in the main text.

change in thermodynamic properties during fluctuating processes.

4. MISFOLDING OF Aβ42 PROTEIN

We applied the fluctuating thermodynamics analysis to the misfolding of the 42-residue form of the Aβ protein (Aβ42).[19] We first conducted MD simulations by placing in explicit water an Aβ42 monomer of helical form as determined in an organic solvent mimicking a membrane environment.[25] A representative misfolding process from our simulations is shown in Figure 2a. We computed $f(\boldsymbol{r}_u) = E_u(\boldsymbol{r}_u) + G_{solv}(\boldsymbol{r}_u)$ values along the simulation trajectories, and then constructed the distribution $W(f)$ of f for the initial helical state and the one for the misfolded state. Both distributions were found to be close to the Gaussian distribution,[19] allowing us to compute the change in free energy upon misfolding (see eq 3) and its components. The value of ΔF was found to be −39.6 kcal/mol, which indicates that the misfolding of the Aβ42 helical structure in an aqueous environment is highly spontaneous. This free energy difference consists of ΔE_u = −91.0 kcal/mol, ΔG_{solv} = +72.2 kcal/mol, and $T\Delta S_{conf}$ = +20.9 kcal/mol, that is, the Aβ42 misfolding is driven by a decrease in intraprotein energy and an increase in protein configurational entropy, but it is also accompanied by an increase in solvation free energy (Figure 2b). We note here that G_{solv} measures the affinity toward the solvent water, and thus protein conformations with larger G_{solv} values can be considered to be more hydrophobic and exhibit a greater tendency to cluster in water. Thus, the misfolded Aβ42 monomer is more prone to aggregation in aqueous environments.

The fluctuating thermodynamics analysis of the trajectory combined with the site-directed analysis further elucidates the molecular origin of the increase in the hydrophobicity of the Aβ42 monomer during its misfolding.[19] The increase in G_{solv} upon misfolding is predominantly determined by the electro-

DOI: 10.1021/acs.accounts.5b00032
Acc. Chem. Res. 2015, 48, 956−965

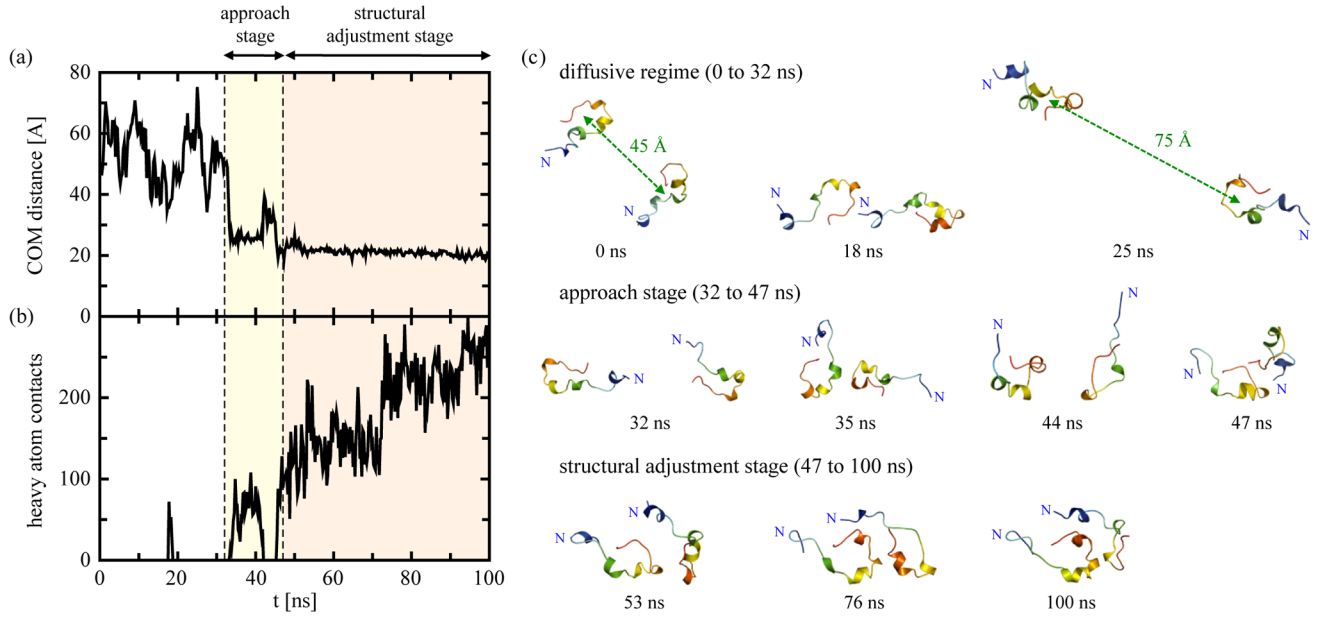

Figure 3. Structural characterization of the Aβ42 dimerization. (a) Intermonomer COM distance and (b) heavy atom contacts versus simulation time. In this and the following Figure 4, the diffusive regime (0−32 ns), the approach stage (32−47 ns, colored light yellow), and the structural adjustment stage (47−100 ns, colored light orange) are separated by vertical dashed lines. (c) Representative Aβ42 dimer conformations from the simulation.

static term (Figure 2c) and the contribution from the hydrophilic residues (Figure 2d). To further explore the structural origin of the increase in G_{solv}, the simulation trajectory was divided into three time regions. The structural characteristic in the first (up to 18 ns) time regime is the unfolding of the initial helix structure, in particular in the N-terminal (residues 1−11) region which is rich in charged residues. Upon the disruption of the helical structure, these charged residues start to form salt-bridges, and the concomitant dehydration of these residues increases G_{solv} in this time regime. This feature can be confirmed from the residue-decomposed result (Figure 2e), from which one observes that G_{solv} in the N-terminal portion already increased significantly at 11 ns. In the second (18−43 ns) time regime, a salt-bridge between K16 and A42 is formed, which further increases G_{solv}; this can also be verified from the residue-resolved G_{solv} at 25 ns. In the final (43−50 ns) time regime, nonlocal backbone hydrogen bonds develop between the central (residues 16−18) and C-terminal (residues 39−42) regions, and the dehydration of these regions additionally increases G_{solv}. In this way, one understands how the Aβ42 monomer acquires its aggregation-prone nature during misfolding.

5. DIMERIZATION OF Aβ42 PROTEINS

To study the dimerization of the misfolded Aβ42 monomers, unbiased dimerization simulations were carried out by placing the two misfolded monomers 45 Å apart from each other with a random orientation.[24,26] To characterize the dimerization, we monitored the intermonomer center-of-mass (COM) distance and heavy atom contacts (Figures 3a and b).

Two distinct stages, termed the approach and structural adjustment stages (Figure 3c), are discernible following the diffusive regime of the two monomers. Up to 32 ns, the two monomers make transient contacts, but readily dissociate into two diffusing monomers. The COM distance exhibits a large drop between 32 and 47 ns, and this time interval is referred to

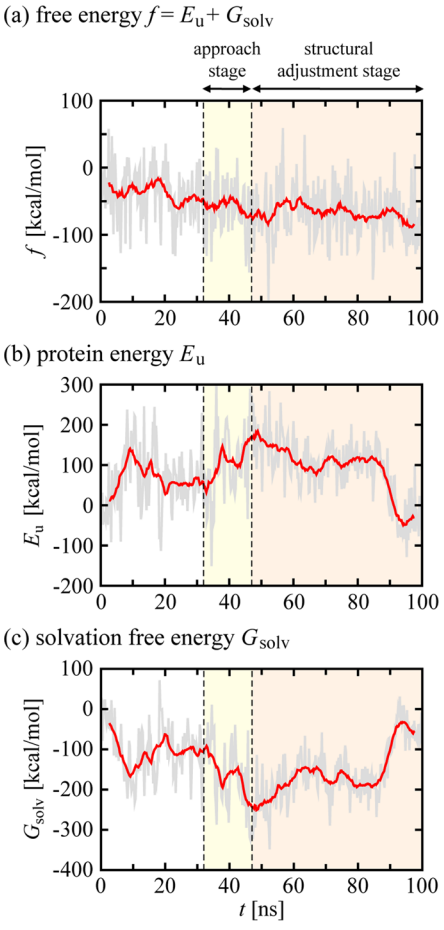

Figure 4. Thermodynamic characterization of the Aβ42 dimerization. (a) Free energy f, (b) protein energy E_u, and (c) solvation free energy G_{solv} versus simulation time. The initial values are set to zero, and running averages over 5 ns are shown in red.

DOI: 10.1021/acs.accounts.5b00032
Acc. Chem. Res. 2015, 48, 956−965

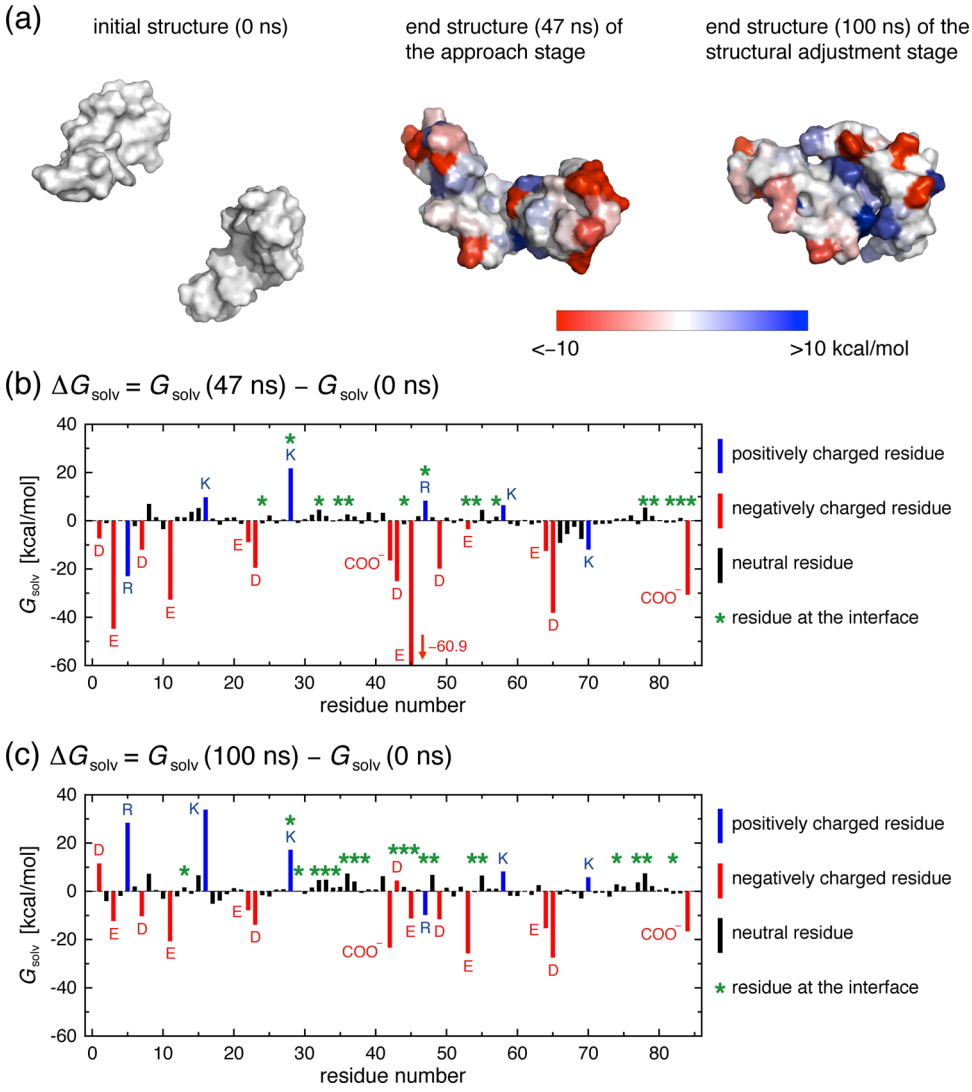

Figure 5. Site-directed thermodynamics analysis of the Aβ42 dimerization. (a) The initial structure (0 ns) and the end structures of the approach stage (47 ns) and the structural adjustment stage (100 ns) are shown with surface representation in which each residue is colored according to its site-resolved G_{solv} value relative to that of the initial structure. Numerical values are plotted in (b) for the 47 ns structure and (c) for the 100 ns structure, where the residue number for the first (second) monomer ranges from 1 (43) to 42 (84). Contributions from positively charged residues are colored with blue, negatively charged residues with red, and neutral residues with black. Residues located at the dimer interface are indicated with green star.

as the approach stage. The dynamics in this stage are not monotonic. For example, another transient dimer with a different relative orientation is formed at 35 ns, but this falls apart at 44 ns since a sufficient number of intermonomer contacts are not developed to stabilize this dimer conformation. The association of the two monomers is observed again at 47 ns with different conformations and a contact interface. After 47 ns, the number of intermonomer contacts is significantly increased, and the two monomers remain in contact until the end of the simulation time (100 ns). Since the conformational adjustment in the two monomers to make up favorable intermonomer interactions is the most distinctive feature after 47 ns, this time regime is termed as the structural adjustment stage. Such conformational changes result in a more compact dimer structure.

Fluctuating thermodynamics analysis was then applied to uncover the driving factors of the respective stages of dimerization.[24] The driving forces must originate from $f = E_u + G_{solv}$ since the protein configurational entropy decreases upon dimer formation. (We note that E_u here comprises both intra- and intermonomer contributions.) Indeed, we confirmed that the free energy f decreases as the dimerization proceeds (Figure 4a). Interestingly, the decrease in f in the approach and structural adjustment regimes has different origins. We find that the thermodynamic force driving the approach of two monomers is the decrease in G_{solv} (Figure 4c). Thus, the large hydrophobicity (G_{solv}) of the misfolded monomers acquired during their conformational changes in water as a monomer is harnessed to drive the approach process of the two monomers to a contact distance.

Our result on the thermodynamic driving force in the approach stage is understandable in light of the fact that the two negatively charged Aβ42 monomers (the total charge of an Aβ42 monomer at neutral pH is −3) would not approach each other if there were no water-mediated attraction that overcomes the electrostatic repulsion. To gain more insight into the underlying molecular mechanisms, we provide a pictorial representation of the end structure of the approach

DOI: 10.1021/acs.accounts.5b00032
Acc. Chem. Res. 2015, 48, 956–965

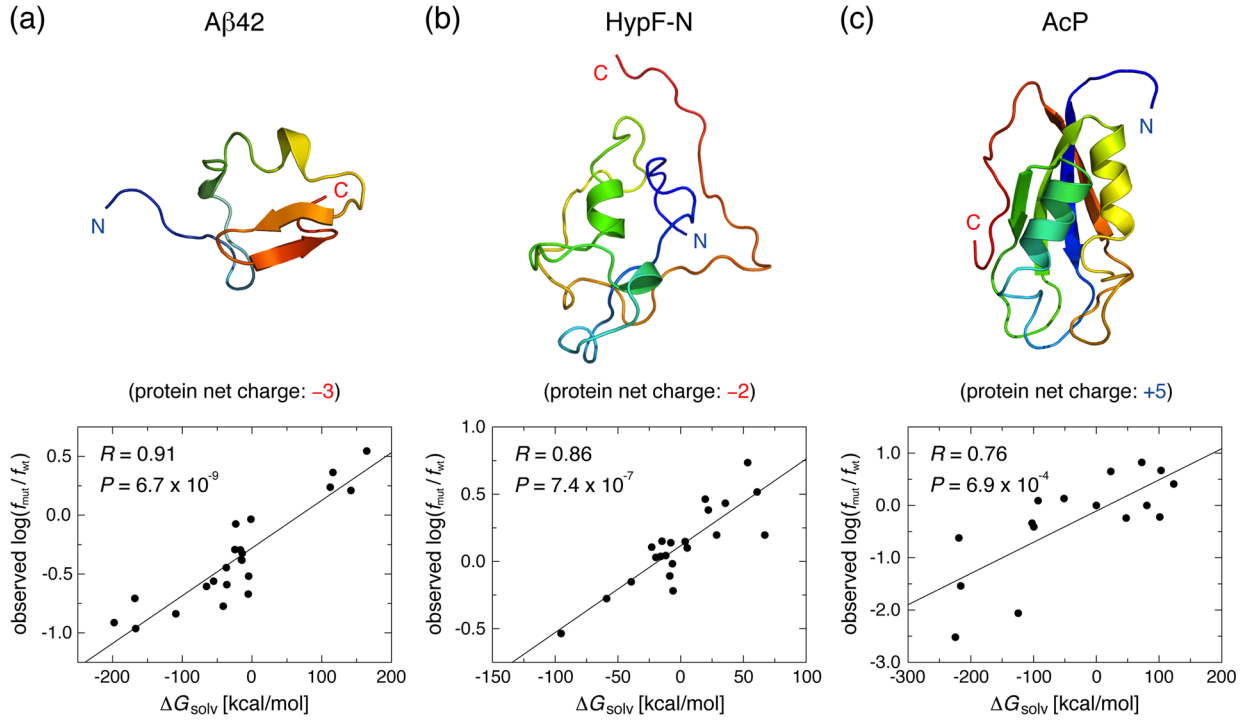

Figure 6. Aggregation propensity versus solvation free energy. Experimental aggregation propensity change $\log(f_{mut}/f_{wt})$ is plotted versus solvation free energy change, $\Delta G_{solv} = G_{solv}(mut) - G_{solv}(wt)$, of the mutant (mut) and wild-type (wt) proteins for (a) Aβ42, (b) HypF-N, and (c) AcP. Pearson correlation coefficient (R) and statistical significance (P value) are shown. Representative simulation structures of the wild-type proteins and their net charges are also displayed.

stage in which the protein residues are colored according to the site-resolved G_{solv} (Figure 5a and b). We find that the hydrophilic residues, in particular the negatively charged residues, get more hydrated in the approach stage of dimerization. Thus, it is the water-mediated force (solvation free energy) acting more strongly on the hydrophilic residues that plays a crucial role to bring two proteins from a long distance to a contact length.

In contrast, the thermodynamic force driving the structural adjustment stage is the protein potential energy E_u, which decreases in this stage (Figure 4b). Direct protein−protein interactions such as intermonomer van der Waals contacts and hydrogen bonds come into play after atomic contacts are developed between two monomers. The decrease in E_u as the structural adjustment proceeds toward a compact dimer structure reflects its energetic stabilization. Interestingly, such structural adjustment also accompanies an increase in the solvation free energy (Figure 4c), which originates from the dehydration of the protein surface and of the interfacial region. This feature can be confirmed from the end structure of the structural adjustment stage in which the protein residues are color-coded by the site-resolved G_{solv} (Figure 5a and c). The enhanced solvation free energy of the Aβ42 dimer will function as the driving force to recruit another Aβ42 protein in the approach stage of subsequent oligomerizations.

6. DISTINCT ROLE OF HYDRATION WATER IN DETERMINING PROTEIN AGGREGATION PROPENSITY

The fluctuating thermodynamics analysis of the Aβ42 misfolding and dimerization presented so far indicates that the protein−water interaction quantified by the solvation free energy plays a crucial role in initiating protein aggregation. In particular, it suggests that proteins with higher solvation free energies are more prone to aggregate. To further corroborate such a water-centric perspective, we examined how the effect of mutations on protein aggregation propensity can be accounted for by the solvation free energy. For this purpose, 22 mutants of the Aβ42 protein, 21 mutants of the N-terminal domain of the protein HypF (HypF-N), and 15 mutants of acylphosphatase (AcP) were studied, for which the experimental data for the aggregation propensity are available.[27] The equilibrium solution structures of these mutants and their wild types were sampled via explicit-water MD simulations, followed by computation of the average solvation free energy G_{solv}.[27] It is observed that the experimental change in aggregation propensity upon mutation has a significant correlation with the change in G_{solv} (Figure 6). This result clearly indicates the distinct role of hydration water in dictating protein aggregation propensity.

As shown in Figure 2c and d, the electrostatic interaction exerted on the hydrophilic residues is the major determinant of the change in G_{solv}. The dominance of the electrostatic term implies that the charged residues are more important than the neutral ones. Indeed, the charged residues more significantly affect the solvation free energy than the neutral ones (Figure 7). Interestingly, the positively and negatively charged residues exhibit contrasting behavior depending on the protein total charge.[27,28] When the total charge of a protein is negative as in Aβ42 (Figure 7a) and HypF-N (Figure 7b), the negatively charged residues display a much stronger affinity for water (i.e., much more negative solvation free energy) than the positively charged ones. AcP, which has a positive total charge, exhibits just the opposite trend: the solvation free energy of the positively charged residues is much more negative than that of the negatively charged ones (Figure 7c).

DOI: 10.1021/acs.accounts.5b00032
Acc. Chem. Res. 2015, 48, 956−965

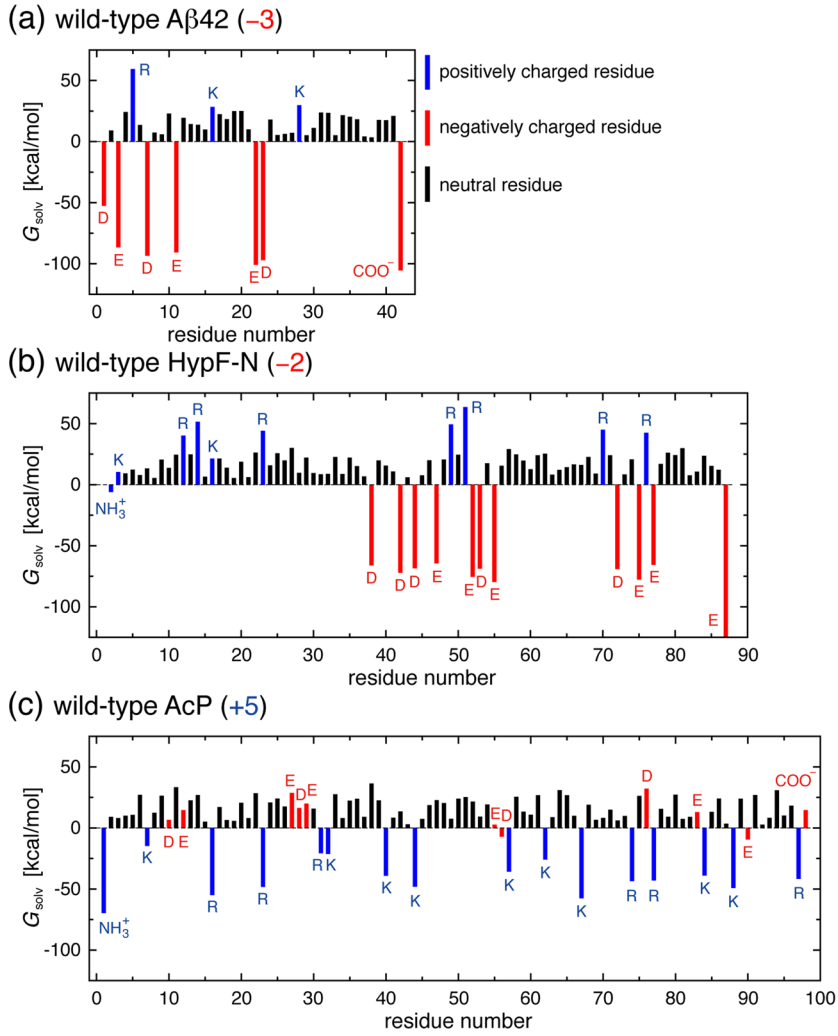

Figure 7. Site-directed analysis of solvation free energy. Site-resolved G_{solv} is shown for (a) the wild-type Aβ42, (b) HypF-N, and (c) AcP. Contributions from positively charged residues are colored with blue, negatively charged residues with red, and neutral residues with black.

The long-distance hydration structure is the key to understanding this striking behavior.[27] Because the electrostatic interaction is long-range, the long-distance hydration structure of a charged residue is also influenced by neighboring charged residues. For example, the negative net charge of Aβ42 and HypF-N generates such an equilibrium orientational distribution of the surrounding water in which the hydrogen of water points toward the protein. Such a water orientational distribution at long distances provides favorable electrostatic interaction with the negatively charged residues, but an unfavorable interaction with the positively charged residues, leading to more preferential solvation of the negatively charged residues. Applying the same argument, it is understood that more preferential solvation of the positively charged residues is brought about on the surface of a protein whose total charge is positive as in AcP. Together with the observation that G_{solv} strongly depends on the protein conformation (Figure 2b), these results demonstrate that the solvation free energy of amino acids depends on the context; consequently, the overall hydrophobicity of a protein cannot be understood solely from the sequence using, for example, conventional hydrophobicity scales determined for free amino acids.[29]

7. CONCLUSION AND PERSPECTIVE

Since the first protein crystal structure was resolved, "static" or "rigid" views such as the lock-and-key enzyme specificity and antibody–antigen structural complementarity have served as the guiding principles to elucidate the structure–function relationships. Although such views are useful and work for a number of proteins, recent discovery that many proteins or protein regions are inherently unstructured has directed our attention to the relevance of conformational disorders and their fluctuating processes.[30,31] Because biomolecular assembly processes are largely under thermodynamic control,[32] dynamic extension of thermodynamics is necessary to uncover the mechanisms and driving factors of fluctuating processes. The fluctuating thermodynamics method presented in this Account offers a practical means for the thermodynamic characterization of conformational dynamics in biomolecules. In particular, this approach does not call for the presence of a well-defined folded structure. Indeed, the Aβ42 protein mainly studied in this Account is disordered in aqueous environments, and the thermodynamic characterization of its conformational disorder in terms of the configurational entropy using widely employed methods such as the quasi-harmonic approximation[33] has been difficult. Together with site-directed analysis which allows one to indentify thermodynamic hot spots, the use of fluctuating

DOI: 10.1021/acs.accounts.5b00032
Acc. Chem. Res. 2015, 48, 956–965

thermodynamics has the potential to provide a comprehensive picture of fluctuating phenomena in diverse biological processes. It will also be valuable for establishing a dynamic paradigm on how biological activity is encoded in protein conformational disorder.

Through the application of fluctuating thermodynamics, we provide a thermodynamic perspective on the misfolding and aggregation of the Aβ42 protein which is implicated in Alzheimer's disease. We find that the Aβ42 monomer acquires a large solvation free energy, that is, becomes more "hydrophobic", through the misfolding that occurs when it is released into an extracellular aqueous environment from the membrane. The hydrophobic nature of the misfolded Aβ42 monomer contributes to its aggregation propensity since hydrophobic particles tend to cluster in aqueous environments.[34,35] Indeed, our observation that the decrease in solvation free energy, that is, water-induced attraction,[36] is the driving factor in the approach stage of forming a contact dimer confirms this view. Interestingly, after two monomers form a contact dimer, additional conformational changes occur, resulting in a more compact structure, and during this structural adjustment stage, the Aβ42 dimer acquires more hydrophobicity and aggregation-prone nature. The increased hydrophobicity of the Aβ42 dimer will then function as the driving force in subsequent oligomerizations. The relevance of the solvation free energy in determining the aggregation-prone nature was further corroborated by demonstrating that, for a large number of proteins, the experimental aggregation propensity has a significant correlation with the solvation free energy.

Site-directed thermodynamics analysis further elucidates the molecular origin of the change in the overall protein hydrophobicity (solvation free energy) during the misfolding and aggregation processes. It is observed that the electrostatic term and the contribution from hydrophilic residues dominate the change in solvation free energy in these processes. In particular, we find that, depending on the protein total charge, the positively and negatively charged residues each have a distinct role in determining the protein aggregation propensity. This finding has significant implications in controlling protein aggregation propensity through site-directed mutagenesis. When the protein total charge is negative, the solubility of the protein in principle increases (decreases) if a neutral amino acid is mutated to a negatively (positively) charged one. For proteins with a positive net charge, the role of negatively and positively charged residues is reversed. Our analyses thus provide new insights for understanding and predicting protein aggregation propensity, thereby offering novel design principles for producing aggregation-resistant proteins for biotherapeutics.

Fluctuating thermodynamics presented in this Account offers a general framework for obtaining site-specific and time-resolved thermodynamic quantities associated with fluctuating processes. On the other hand, there are several possibilities in implementing this general approach with specific computational methods. The application of enhanced sampling algorithms[37] and methods such as the Markov state model[38] may be valuable for more efficient exploration of the protein conformational state, the combination of which has recently been used to identify metastable conformations relevant to aggregation of an intrinsically disordered protein.[39] For an accurate description of the free energy landscape, it is essential to have a reliable method for computing the solvation free energy. The 3D-RISM theory adopted here is known as one of the most successful methods, but it has several drawbacks, in particular in handling hydrophobic solutes, because of which its applicability to systems exhibiting extended hydrophobic surfaces[40] is questionable. In this regard, we have recently developed a density-functional theory that exactly takes into account the intramolecular correlations of the solvent water.[41] Promising results on the solvation free energy have been reported based on such a theory for an extensive number of small neutral molecules.[42] Fluctuating thermodynamics analysis combined with these advanced computational methods will significantly enlarge the range of its applicability.

AUTHOR INFORMATION

Corresponding Author

*E-mail sihyun@sookmyung.ac.kr. Phone +82-2-710-9410. Fax +82-2-2077-7321.

Notes

The authors declare no competing financial interest.

Biographies

S.-H. Chong is a research associate in Department of Chemistry at Sookmyung Women's University. His research interests are in physical chemistry and biology.

S. Ham is Professor of Chemistry and the Director of the Center for Biomolecular Network at Sookmyung Women's University. Her current research interests focus on understanding biomolecular interactions at the atomic level.

ACKNOWLEDGMENTS

This work was supported by Samsung Science and Technology Foundation under Project Number SSTF-BA1401-13.

REFERENCES

(1) Dobson, C. M. Protein Folding and Misfolding. *Nature* **2003**, *426*, 884−890.

(2) Eisenberg, D.; Jucker, M. The Amyloid State of Proteins in Human Diseases. *Cell* **2012**, *148*, 1188−1203.

(3) Chiti, F.; Dobson, C. M. Protein Misfolding, Functional Amyloid, and Human Disease. *Annu. Rev. Biochem.* **2006**, *75*, 333−366.

(4) Chiti, F.; Stefani, M.; Taddei, N.; Ramponi, G.; Dobson, C. M. Rationalization of the Effects of Mutations on Peptide and Protein Aggregation Rates. *Nature* **2003**, *424*, 805−808.

(5) Belli, M.; Ramazzotti, M.; Chiti, F. Prediction of Amyloid Aggregation *In Vivo*. *EMBO Rep.* **2011**, *12*, 657−663.

(6) Goedert, M.; Spillantini, M. G. A Century of Alzheimer's Disease. *Science* **2006**, *314*, 777−781.

(7) Crescenzi, O.; Tomaselli, S.; Guerrini, R.; Salvadori, S.; D'Ursi, A. M.; Temussi, P. A.; Picone, D. Solution Structure of the Alzheimer Amyloid β-Peptide (1−42) in an Apolar Microenvironment. *Eur. J. Biochem.* **2002**, *269*, 5642−5648.

(8) Tomaselli, S.; Esposito, V.; Vangone, P.; van Nuland, N. A. J.; Bonvin, A. M. J. J.; Guerrini, R.; Tancredi, T.; Temussi, P. A.; Picone, D. The α-to-β Conformational Transition of Alzheimer's Aβ−(1−42) Peptide in Aqueous Media Is Reversible: A Step by Step Conformational Analysis Suggests the Location of β Conformation Seeding. *ChemBioChem* **2006**, *7*, 257−267.

(9) Walsh, D. M.; Selkoe, D. J. Aβ Oligomers—A Decade of Discovery. *J. Neurochem.* **2007**, *101*, 1172−1184.

(10) Bernstein, S. L.; Dupuis, N. F.; Lazo, N. D.; Wyttenbach, T.; Condron, M. M.; Bitan, G.; Teplow, D. B.; Shea, J.-E.; Ruotolo, B. T.; Robinson, C. V.; Bowers, M. T. Amyloid-β Protein Oligomerization and the Importance of Tetramers and Dodecamers in the Aetiology of Alzheimer's Disease. *Nat. Chem.* **2009**, *1*, 326−331.

(11) Shankar, G. M.; Li, S.; Mehta, T. H.; Garcia-Munoz, A.; Shepardson, N. E.; Smith, I.; Brett, F. M.; Farrell, M. A.; Rowan, M. J.;

DOI: 10.1021/acs.accounts.5b00032
Acc. Chem. Res. 2015, 48, 956−965

Lemere, C. A.; Regan, C. M.; Walsh, D. M.; Sabatini, B. L.; Selkoe, D. J. Amyloid-β Protein Dimers Isolated Directly from Alzheimer's Brains Impair Synaptic Plasticity and Memory. *Nat. Med.* **2008**, *14*, 837−842.
(12) Lazaridis, T.; Karplus, M. Thermodynamics of Protein Folding: A Microscopic View. *Biophys. Chem.* **2003**, *100*, 367−395.
(13) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. The Statistical-Thermodynamic Basis for Computation of Binding Affinities: A Critical Review. *Biophys. J.* **1997**, *72*, 1047−1069.
(14) Chong, S.-H.; Ham, S. Configurational Entropy of Protein: A Combined Approach Based on Molecular Simulation and Integral-Equation Theory of Liquids. *Chem. Phys. Lett.* **2011**, *504*, 225−229.
(15) Sekimoto, K. *Stochastic Energetics*; Springer: New York, 2010.
(16) Imai, T.; Harano, Y.; Kinoshita, M.; Kovalenko, A.; Hirata, F. A Theoretical Analysis on Hydration Thermodynamics of Proteins. *J. Chem. Phys.* **2006**, *125*, 024911.
(17) Kovalenko, A. In *Molecular Theory of Solvation*; Hirata, F., Ed.; Kluwer Academic: Dordrecht, The Netherlands, 2003; pp 169−275.
(18) Chong, S.-H.; Ham, S. Protein Folding Thermodynamics: A New Computational Approach. *J. Phys. Chem. B* **2014**, *118*, 5017−5025.
(19) Chong, S.-H.; Park, M.; Ham, S. Structural and Thermodynamic Characteristics That Seed Aggregation of Amyloid-β Protein in Water. *J. Chem. Theory Comput.* **2012**, *8*, 724−734.
(20) Chong, S.-H.; Ham, S. Conformational Entropy of Intrinsically Disordered Protein. *J. Phys. Chem. B* **2013**, *117*, 5503−5509.
(21) Chong, S.-H.; Lee, C.; Kang, G.; Park, M.; Ham, S. Structural and Thermodynamic Investigations on the Aggregation and Folding of Acylphosphatase by Molecular Dynamics Simulations and Solvation Free Energy Analysis. *J. Am. Chem. Soc.* **2011**, *133*, 7075−7083.
(22) Chong, S.-H.; Ham, S. Atomic Decomposition of the Protein Solvation Free Energy and its Application to Amyloid-Beta Protein in Water. *J. Chem. Phys.* **2011**, *135*, 034506.
(23) Chong, S.-H.; Ham, S. Component Analysis of the Protein Hydration Entropy. *Chem. Phys. Lett.* **2012**, *535*, 152−156.
(24) Chong, S.-H.; Ham, S. Impact of Chemical Heterogeneity on Protein Self-Assembly in Water. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 7636−7641.
(25) Lee, C.; Ham, S. Characterizing Amyloid-Beta Protein Misfolding from Molecular Dynamics Simulations with Explicit Water. *J. Comput. Chem.* **2011**, *32*, 349−355.
(26) Chong, S.-H.; Ham, S. Atomic-Level Investigations on the Amyloid-β Dimerization Process and its Driving Forces in Water. *Phys. Chem. Chem. Phys.* **2012**, *14*, 1573−1575.
(27) Chong, S.-H.; Ham, S. Interaction with the Surrounding Water Plays a Key Role in Determining the Aggregation Propensity of Proteins. *Angew. Chem., Int. Ed.* **2014**, *53*, 3961−3964.
(28) Chong, S.-H.; Ham, S. Site-Directed Analysis on Protein Hydrophobicity. *J. Comput. Chem.* **2014**, *35*, 1364−1370.
(29) Kyte, J.; Doolittle, R. F. A Simple Method for Displaying the Hydropathic Character of a Protein. *J. Mol. Biol.* **1982**, *157*, 105−132.
(30) Dyson, H. J.; Wright, P. E. Intrinsically Unstructured Proteins and Their Functions. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 197−208.
(31) Tompa, P. *Structure and Function of Intrinsically Disordered Proteins*; CRC Press: Boca Raton, FL, 2010.
(32) Tanford, C. The Hydrophobic Effect and the Organization of Living Matter. *Science* **1978**, *200*, 1012−1018.
(33) Schlitter, J. Estimation of Absolute and Relative Entropies of Macromolecules Using the Covariance Matrix. *Chem. Phys. Lett.* **1993**, *215*, 617−621.
(34) Kauzmann, W. Some Factors in the Interpretation of Protein Denaturation. *Adv. Protein Chem.* **1959**, *14*, 1−63.
(35) Chandler, D. Interfaces and the Driving Force of Hydrophobic Assembly. *Nature* **2005**, *437*, 640−647.
(36) Ben-Naim, A. *Hydrophobic Interactions*; Plenum: New York, 1980.
(37) Bernardi, R. C.; Melo, M. C. R.; Schulten, K. Enhanced Sampling Techniques in Molecular Dynamics Simulations of Biological Systems. *Biochim. Biophys. Acta* **2015**, *1850*, 872−877.
(38) Shukla, D.; Hernández, C. X.; Weber, J. K.; Pande, V. S. Markov State Models Provide Insights into Dynamic Modulation of Protein Function. *Acc. Chem. Res.* **2015**, *48*, 414−422.
(39) Qiao, Q.; Bowman, G. R.; Huang, X. Dynamics of an Intrinsically Disordered Protein Reveal Metastable Conformations That Potentially Seed Aggregation. *J. Am. Chem. Soc.* **2013**, *135*, 16092−16101.
(40) Liu, P.; Huang, X.; Zhou, R.; Berne, B. J. Observation of a Dewetting Transition in the Collapse of the Melittin Tetramer. *Nature* **2005**, *437*, 159−162.
(41) Chong, S.-H.; Ham, S. Aqueous Interaction Site Integral-Equation Theory That Exactly Takes into Account Intramolecular Correlations. *J. Chem. Phys.* **2012**, *137*, 154101.
(42) Liu, Y.; Fu, J.; Wu, J. High-Throughput Prediction of the Hydration Free Energies of Small Molecules from a Classical Density Functional Theory. *J. Phys. Chem. Lett.* **2013**, *4*, 3687−3691.

DOI: 10.1021/acs.accounts.5b00032
Acc. Chem. Res. 2015, 48, 956−965