Protein Aggregation | Hot Paper

DOI: 10.1002/anie.201309317

Interaction with the Surrounding Water Plays a Key Role in Determining the Aggregation Propensity of Proteins**

*Song-Ho Chong and Sihyun Ham**

Abstract: *Understanding the molecular determinants of the relative propensities of proteins to aggregate in a cellular environment is a central issue for treating protein-aggregation diseases and developing peptide-based therapeutics. Despite the expectation that protein aggregation can largely be attributed to direct protein–protein interactions, a crucial role the surrounding water in determining the aggregation propensity of proteins both in vitro and in vivo was identified. The overall protein hydrophobicity, defined solely by the hydration free energy of a protein in its monomeric state sampling its equilibrium structures, was shown to be the predominant determinant of protein aggregation propensity in aqueous solution. Striking discrimination of positively and negatively charged residues by the surrounding water was also found. This effect depends on the protein net charge and plays a crucial role in regulating the solubility of the protein. These results pave the way for the design of aggregation-resistant proteins as biotherapeutics.*

A number of human diseases are associated with the formation of toxic protein aggregates.[1] Protein aggregation also represents a major bottleneck for the biotechnological production of polypeptide-based drugs[2] and antibody-based reagents.[3] Several sequence-based algorithms have thus been developed to rationalize and predict protein aggregation propensity.[4–7] However, the role of water in protein aggregation has been largely unexplored owing to the perception that protein adhesion (protein–protein interaction) is the main actor and the surrounding water is just a spectator and can be disregarded when trying to understand protein aggregation phenomena. Indeed, almost all studies on the prediction of protein aggregation have mainly focused on sequence characteristics such as intermolecular β-structure-forming propensity.[8–10] Such a protein-centric view, however, ignores water as an active constituent of biological systems.[11–13] In fact, whether a protein remains soluble or becomes aggregation-prone in aqueous environments should intrinsically rely on its affinity toward water in its monomeric state. The main focus of the present work is to uncover the controlling role of water in discriminating aggregation-prone proteins from soluble ones and thereby to contribute to the understanding and prediction of protein aggregation propensity.

An interaction induced by water such as the hydrophobic effect is often invoked to explain biomolecular self-assembly in aqueous media.[14–16] According to statistical thermodynamics, the water-mediated force is quantified by the hydration free energy G_{hyd} (the change in free energy from gas to aqueous phases), which characterizes the affinity of a protein for the solvent water.[17] Molecules with larger G_{hyd} values are more hydrophobic and the hydrophobic interaction between them is thus more effective. A key question is whether this concept of hydrophobicity applies beyond the conventional levels of small molecules or individual amino acids and whether a more hydrophobic protein is more prone to aggregate in aqueous environments. While the overall protein hydrophobicity (protein hydration free energy G_{hyd}) is not directly experimentally accessible,[17] recent developments in computational tools make it possible to approach this quantity. First, owing to advances in computing power and refinements of force fields, molecular dynamics (MD) simulations have developed into a powerful and accurate tool for protein conformational studies.[18,19] Second, the molecular theory of solvation has been much advanced in recent years, thus allowing us to compute[20] and analyze[21–23] protein hydration free energy based on simulated protein conformations.

In this study, we quantified the overall protein hydrophobicity based on these advanced computational tools by applying them to intrinsically disordered proteins [22 mutants of amyloid-β (Aβ42) protein and 21 mutants of the N-terminal domain of the *Escherichia coli* protein HypF (HypF-N)] as well as to natively folded proteins [15 mutants of human muscle acylphosphatase (AcP)], the aggregation propensities of which have been experimentally characterized in vitro[24,25] and in vivo.[26–28] (see Figure 1b,c for the structures and sequences of the wild-type proteins; all of the mutants studied are specified in Tables S1–S3 in the Supporting Information.) For each of these proteins, we performed extensive all-atom, explicit-water MD simulations to sample their equilibrium solution structures. (See the Supporting Information.) The hydration free energy G_{hyd} for each protein was then computed by applying the integral-equation theory of liquids[20–23] to simulated protein conformations and the resulting average G_{hyd} is identified as the overall protein hydrophobicity: a larger (i.e., more positive) G_{hyd} value is associated with increased protein hydrophobicity.

[*] Dr. S.-H. Chong, Prof. S. Ham
Department of Chemistry, Sookmyung Women's University
Cheongpa-ro 47-gil 100, Yongsan-Ku, Seoul 140-742 (Korea)
E-mail: sihyun@sookmyung.ac.kr
Homepage: http://nbcc.sm.ac.kr

[**] This work was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (Nos. 2012-0003068 and 2012R1A2A01004687).

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201309317.

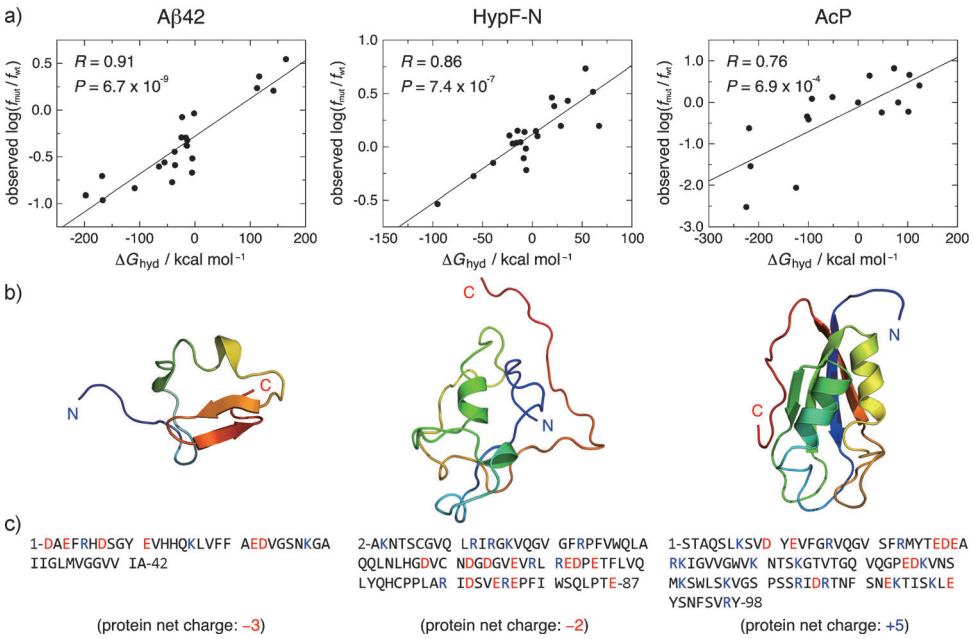

Figure 1. Correlation between experimental protein aggregation propensity and calculated protein hydrophobicity. a) Changes in the experimental aggregation propensity $\log(f_{mut}/f_{wt})$ upon mutation are plotted against the difference in the hydration free energy $[\Delta G_{hyd} = G_{hyd}(mut) - G_{hyd}(wt)]$ for the mutant and wild-type variants of Aβ42 (left-hand panels), HypF-N (middle panels), and AcP (right-hand pannels). The Pearson correlation coefficient (R) and statistical significance (P value) are also displayed. b) Snapshot structures generated by simulations for the intrinsically disordered Aβ42, the folding-incompetent HypF-N, and the folded form of AcP. Each structure is color-coded according to primary sequence, ranging from blue at the N terminus to red at the C terminus. c) Sequences for wild-type Aβ42, HypF-N, and AcP. Positively charged residues are indicated in blue, negatively charged residues in red, and neutral residues in black. The total net charge of each protein is also displayed.

The correlation between the experimental aggregation propensity and the computed protein hydrophobicity (Figure 1a) is striking considering the large variation in the aggregation propensities, the distinct nature of the underlying protein structures, and different types of mutations involved. These results show that it is the water surrounding a protein, rather than the protein itself, which is the predominant factor that controls the extent to which a protein is aggregation-prone in aqueous environments. This is corroborated by the observation that representative sequence-based models focusing on intermolecular β-structure-forming propensity do not consistently yield such high correlation with the experimental data (Figure S1 in the Supporting Information). Furthermore, we found that it is the electrostatic term in the protein–water interaction that dictates the protein hydrophobicity (Figure S2). This finding implies that the water-induced interaction that causes proteins to aggregate is long-range in nature. In fact, we have recently demonstrated that the long-range water-mediated force is primarily responsible for two aggregating proteins approaching each other from a large separation to within contact distance.[23] After proteins make atomic contacts, direct protein–protein interactions will then come into play. The intermolecular β-sheet-forming propensity, taken into account in sequence-based algorithms to identify aggregation-prone regions,[8–10] will play a major role in the subsequent structural reorganization step toward the formation of ordered aggregates called amyloid fibrils. Methods to identify interaction regions through estimation of the desolvation free energy[29] will also be useful since protein contact formation involves the removal of interfacial water.

The dominance of the electrostatic interaction in regulating protein hydrophobicity implies a greater relevance for charged residues compared to neutral ones. In fact, we find from the site-directed thermodynamic analysis that the charged residues more significantly influence the hydration free energy than the neutral ones (Figure 2). Quite unexpectedly, we observe a contrasting role of positively (Lys and Arg) and negatively (Glu and Asp) charged residues when embedded in the protein context (Figure 2). The negatively charged residues exhibit much more negative hydration free energy (i.e., a much greater affinity for water) than the positively charged ones in both Aβ42 (Figure 2b) and HypF-N (Figure 2c), the net charges of which are negative. Just the opposite trend is observed for AcP, the net charge of which is positive: the positively charged residues display significantly more negative hydration free energy than the negatively charged ones (Figure 2d). It is well recognized that chaotropes (Lys and Arg side chains) bind water more weakly than kosmotropes (Glu and Asp side chains).[30] In addition, the three or four methylenes in Arg and Lys side chains may play more significant hydrophobic roles compared to the one or two methylenes in Asp and Glu side chains. These features are in fact reflected in the hydration free energy for the free charged amino acids (Figure 2a). However, this argument based on the free charged amino acids does not rationalize the strikingly distinct behavior of positively and negatively charged residues when embedded in the protein.

The key to understanding this contrasting behavior lies in the long-distance hydration structure surrounding the protein surface (Figure 3). As a result of the long-range nature of the electrostatic term dominating the charged-residue–water interaction, the long-distance hydration structure of a charged residue is also affected by the neighboring charged residues: the equilibrium water distribution is determined primarily by the net charge produced by those residues. In the case of a protein with a negative net charge, such as Aβ42 and HypF-N, this yields a long-range water orientational distribution in which the net dipole moment of water is directed toward the protein (blue water molecules shown in Figure 3b,c). The

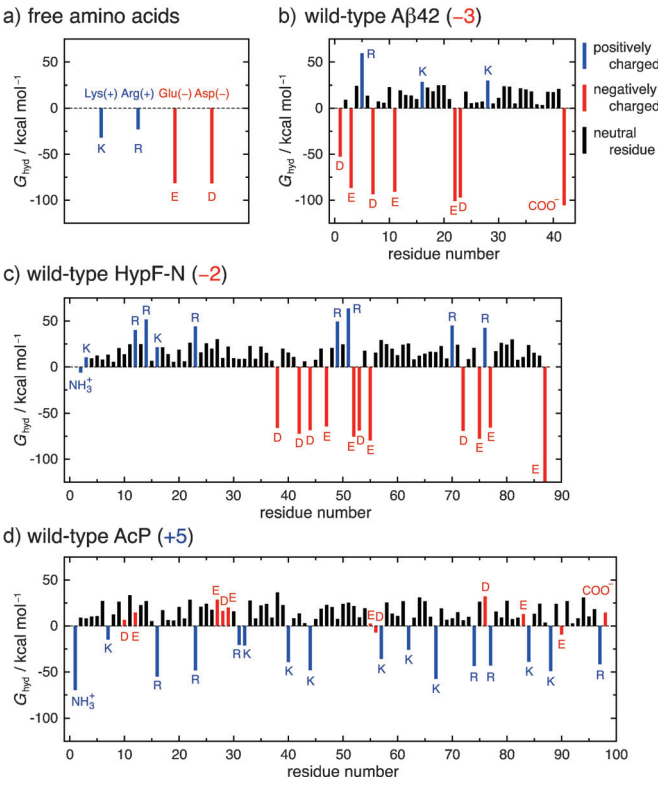

Figure 2. Site-directed analysis of protein hydration free energy. a) Hydration free energy of the free charged amino acids in the form of *N*-acetyl-X-*N*′-methylamide (CH_3-CO-NH-CHR-CO-NH-CH_3 with R denoting the side-chain group). b–d) The contribution of each of constituent amino acids to the hydration free energy of wild-type Aβ42 (b), HypF-N (c), and AcP (d). The contributions from positively charged residues are shown in blue, those from negatively charged residues in red, and those from neutral residues in black.

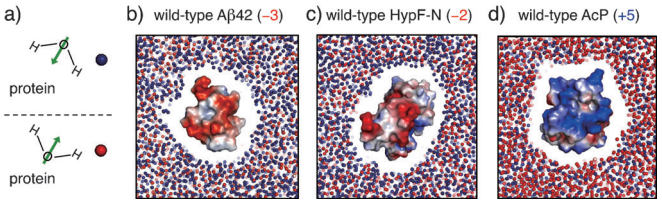

Figure 3. Long-range orientational distribution of the water surrounding the protein surface. The protein surfaces are colored according to the surface charge (blue for positively and red for negatively charged regions) and the water molecules, shown as spheres, are colored according to the direction of the dipole moment [blue indicates a dipole moment directed toward the protein and red indicates a dipole moment pointed away from the protein; see panel (a)]. Water molecules close (< 8 Å) to the protein surface, the orientation which depends on the local chemical details of the surface, are not shown. The orientational distribution of the water molecules at greater distances is determined primarily by the net charge of the protein. There are thus more "dipole-towards" (blue) water molecules around the negatively charged Aβ42 (b) and HypF-N (c), whereas there are more "dipole-away" (red) water molecules around the positively charged AcP (d). These long-range "polarized" water molecules in turn discriminate between charged residues on the protein surface.

resulting interaction between charges on the protein surface and the water dipole at long-range becomes unfavorable for the positively charged residues and favorable for the negatively charged residues. The long-range hydration structure surrounding a protein with a negative net charge thus results in a thermodynamic preference for negatively charged residues over positively charged ones on the protein surface. Applying the same argument, the opposite discrimination of positively and negatively charged residues results on the surface of a protein with a positive net charge, such as AcP, because the net dipole moment of solvent water at long distances will in this case be reversed, i.e., directed outward from the protein (red water molecules displayed in Figure 3 d). The contrasting behavior of positively and negatively charged residues is thus caused by "cross talk" between charged residues on the protein surface and the solvent water molecules in the long-range hydration layers.

The striking discrimination of positively and negatively charged residues depending on the protein net charge has significant implications for various fields including biotechnology and biomedical sciences. One can control the aggregation propensity of a protein through site-directed mutagenesis by using the following design principles. When the net charge of a protein is negative, mutating neutral amino acid to negatively (positively) charged ones will in principle increase (decrease) the solubility in water. Just the opposite will result when the protein net charge is positive: mutating neutral amino acid to negatively (positively) charged ones will diminish (enhance) the protein solubility in water. These design principles are consistent with the recently reported solubility measurements for ribonuclease Sa, which has a negative net charge:[31] 1) Glu and Asp contribute more favorably to protein solubility than do any of the other amino acids, 2) Lys and Arg contribute unfavorably to protein solubility, and 3) Lys provides a favorable contribution to the solubility when there is an excess of positive charge on the protein. The contrasting behavior of charged amino acids depending on the net protein charge also provides an explanation for the experimental observation that, in the entire *Escherichia coli* protein ensemble, proteins with a higher content of negatively charged residues tend to be more soluble than those with more positively charged residues[32] because the net charge of soluble proteins is mostly negative under physiological conditions.

The identification of the factors that determine whether a protein is aggregation prone or remains soluble in aqueous environments is of central importance in addressing protein-aggregation diseases. The major findings of the present study are that the overall protein hydrophobicity, as quantified by the hydration free energy, is the predominant determinant of protein aggregation propensity both in vitro and in vivo and that protein hydrophobicity is significantly influenced by the contrasting behavior of positively and negatively charged residues depending on the net charge of the protein. Our computational tools that deal with protein three-dimensional structure and hydration thermodynamics enable the ab initio prediction of the effects of mutations on protein aggregation propensity, and the derived design principles for developing

aggregation-resistant proteins will find a wide range of applications in biotechnology and biotherapeutics.

Received: October 25, 2013
Revised: December 13, 2013
Published online: February 24, 2014

Keywords: amyloid beta-peptides · hydrophobic effect · proteins · thermodynamics · water

[1] F. Chiti, C. M. Dobson, *Annu. Rev. Biochem.* **2006**, *75*, 333 – 366.
[2] S. Ventura, A. Villaverde, *Trends Biotechnol.* **2006**, *24*, 179 – 185.
[3] D. Lowe, K. Dudgeon, R. Rouet, P. Schofield, L. Jermutus, D. Christ, *Adv. Protein Chem.* **2011**, *84*, 41 – 61.
[4] F. Chiti, M. Stefani, N. Taddei, G. Ramponi, C. M. Dobson, *Nature* **2003**, *424*, 805 – 808.
[5] A.-M. Fernandez-Escamilla, F. Rousseau, J. Schymkowitz, L. Serrano, *Nat. Biotechnol.* **2004**, *22*, 1302 – 1306.
[6] G. G. Tartaglia, M. Vendruscolo, *Chem. Soc. Rev.* **2008**, *37*, 1395 – 1401.
[7] S. Maurer-Stroh, M. Debulpaep, N. Kuemmerer, M. L. de la Paz, I. C. Martins, J. Reumers, K. L. Morris, A. Copland, L. Serpell, L. Serrano, J. W. H. Schymkowitz, F. Rousseau, *Nat. Methods* **2010**, *7*, 237 – 242.
[8] M. Belli, M. Ramazzotti, F. Chiti, *EMBO Rep.* **2011**, *12*, 657 – 663.
[9] V. Castillo, R. Graña-Montes, R. Sabate, S. Ventura, *Biotechnol. J.* **2011**, *6*, 674 – 685.
[10] A. B. Ahmed, A. V. Kajava, *FEBS Lett.* **2013**, *587*, 1089 – 1095.
[11] M. Chaplin, *Nat. Rev. Mol. Cell Biol.* **2006**, *7*, 861 – 866.
[12] P. Ball, *Chem. Rev.* **2008**, *108*, 74 – 108.
[13] P. Ball, *Nature* **2011**, *478*, 467 – 477.
[14] W. Kauzmann, *Adv. Protein Chem.* **1959**, *12*, 1 – 63.
[15] C. Tanford, *Science* **1978**, *200*, 1012 – 1018.
[16] K. A. Dill, *Biochemistry* **1990**, *29*, 7133 – 7155.
[17] A. Ben-Naim, *Hydrophobic Interactions*, Plenum, New York, **1980**.
[18] P. L. Freddolino, C. B. Harrison, Y. Liu, K. Schulten, *Nat. Phys.* **2010**, *6*, 751 – 785.
[19] S. Piana, K. Lindorff-Larsen, D. E. Shaw, *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 17845 – 17850.
[20] T. Imai, Y. Harano, M. Kinoshita, A. Kovalenko, F. Hirata, *J. Chem. Phys.* **2006**, *125*, 024911.
[21] S.-H. Chong, S. Ham, *J. Chem. Phys.* **2011**, *135*, 034506.
[22] S.-H. Chong, C. Lee, G. Kang, M. Park, S. Ham, *J. Am. Chem. Soc.* **2011**, *133*, 7075 – 7083.
[23] S.-H. Chong, S. Ham, *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 7636 – 7641.
[24] K. Murakami, K. Irie, A. Morimoto, H. Ohigashi, M. Shindo, M. Nagao, T. Shimizu, T. Shirasawa, *J. Biol. Chem.* **2003**, *278*, 46179 – 46187.
[25] F. Chiti, M. Calamai, N. Taddei, M. Stefani, G. Ramponi, C. M. Dobson, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 16419 – 16426.
[26] C. Nilsberth, A. Westlind-Danielsson, C. B. Eckman, M. M. Condron, K. Axelman, C. Forsell, C. Stenh, J. Luthman, D. B. Teplow, S. G. Younkin, J. Naslund, L. Lannfelt, *Nat. Neurosci.* **2001**, *4*, 887 – 893.
[27] W. Kim, M. H. Hecht, *J. Biol. Chem.* **2005**, *280*, 35069 – 35076.
[28] J. Winkelmann, G. Calloni, S. Campioni, B. Mannini, N. Taddei, F. Chiti, *J. Mol. Biol.* **2010**, *398*, 600 – 613.
[29] S. Fiorucci, M. Zacharias, *Biophys. J.* **2010**, *98*, 1921 – 1930.
[30] K. D. Collins, *Biophys. J.* **1997**, *72*, 65 – 76.
[31] S. R. Trevino, J. M. Scholtz, C. N. Pace, *J. Mol. Biol.* **2007**, *366*, 449 – 460.
[32] T. Niwa, B.-W. Ying, K. Saito, W. Jin, S. Takada, T. Ueda, H. Taguchi, *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 4201 – 4206.

