Protein Context-Dependent Hydrophobicity of Amino Acids in Protein

Hanul Cho and Sihyun Ham*

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

Abstract: Hydrophobicity is the key concept to understand the water plays in protein folding, protein aggregation, and protein-protein interaction. Traditionally, the hydrophobicity of protein is defined based on the scales of the hydrophobicity of residue, assuming that the hydrophobicity of free amino acids is maintained. Here, we explore how the hydrophobicity of constituting amino acids in protein rely on the protein context, in particular, on the total charge and secondary structures of a protein. To this end, we calculate and investigate the hydration free energy of three short proteins based on the integral-equation theory of liquids. We find that the hydration free energy of charged amino acids is significantly affected by the protein total charge and exhibits contrasting behavior depending on the protein total charge being positive or negative. We also observe that amino acids in the β-sheets display more enhanced the hydrophobicity than amino acids in the loop, whereas those in the α-helix do not clearly show such a tendency. And the salt-bridge forming amino acids also exhibit increase of the hydrophobicity than that with no salt bridge. Our results provide novel insights into the hydrophobicity of amino acids, and will be valuable for rationalizing and predicting the strength of water-mediated interaction involved in the biological activity of proteins.

Keywords: Hydration Free Energy, B Domain of Protein A, Villin Headpiece Subdomain, Pin WW Domain, Hydrophobicity

Introduction

Water plays a decisive role in biological processes.[1,2] The water-mediated hydrophobic interaction is among the principal a factor of determination of the dynamics, function of proteins, stability and structure.[3-5] The hydrophobicity of protein is associated with behavior of protein such as protein aggregation that is related to numerous human disease.[6] Conventionally, the hydrophobicity scales defined for individual free amino acids[7,8] have been used to estimate the hydrophobic interaction between protein constituting amino acid in protein, assuming that those scales are independent of the protein context. However, recent experimental investigations on transmembrane proteins[9] and on the entire Escherichia coli proteins[10] invalidate such a conventional naive picture and claim that the systematization of the hydrophobicity requires more than sequence-based interpretation about the protein.

In this study, we explore how the hydrophobicity of constituting amino acids in protein rely on the protein context. This is done by computing the hydration free energy of amino acids constituted in a protein and comparing it with that of corresponding free amino acids. We use the integral-equation theory of liquids[12,13] for this calculation, and this is applied to three short proteins that differ in the total charge and in the secondary structures (β-sheet versus loop and salt-bridge). Thereby, we work for uncover the role of protein global factors and structural effects in determining the hydrophobicity of amino acids in a protein.

Theory and Computational Method

1. Molecular Dynamics Simulations

We studied the three short proteins (Figure 1): Pin WW domain (PDB code: 2F21[16]) of positive total charge comprising three-stranded antiparallel β-sheets, B-domain of protein A (BdpA, PDB code: 1BDC[14]) of negative total charge consisting of three α-helices and villin headpiece subdomain (HP-35, PDB code: 1YRF[15]) of positive total

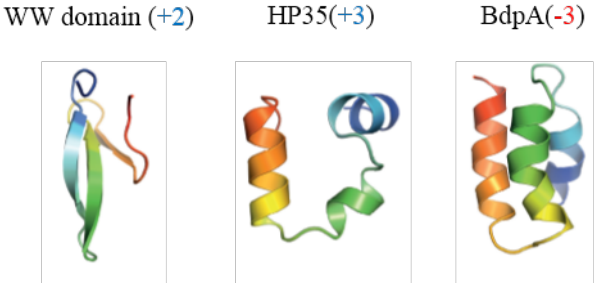

Figure 1 PDB structure of WW domain, HP35 and BdpA. Each structure is color-expressed such from blue to red at the N- and C-termini according to the sequence. Also show the total charge of each protein.

charge containing three α-helices. The PDB structure of these proteins was subjected to the energy minimization and short (200 ps) equilibration (at 300 K and 1 atm) in explicit water using the AMBER14 simulation package[17] to accommodate the protein structure to be consistent with the potential parameters (ff99SB force field[18] for protein and the TIP3P model[19] for water) employed in the hydration free energy calculation. For study the context effect of the hydrophobicity, we also calculate the hydration free energy of individual free amino acids, maintaining the conformation when they constitute the protein. We extract the capping each residue with peptide caps, which is the same side chain conformations with the residues embedded in a protein.

2. Hydration free energy calculation

We used the three-dimensional reference interaction site model (3D-RISM) theory[12,13] to compute the hydration free energy. The theory is an integral-equation theory based on the statistical mechanics for get the 3D distribution function $g_\gamma(\boldsymbol{r}) = h_\gamma(\boldsymbol{r}) + 1$ of the water site γ at position r around a protein. In this theory, the 3D-RISM equation

$$h_\gamma(\boldsymbol{r}) = \sum_{\gamma'} \int d\boldsymbol{r}' \chi_{\gamma\gamma'}(|\boldsymbol{r} - \boldsymbol{r}'|) c_{\gamma'}(\boldsymbol{r}')$$

is solved self-consistently with the closure relation

$$h_\gamma(\boldsymbol{r}) = \begin{cases} \exp\left[-\beta u_\gamma(\boldsymbol{r}) + h_\gamma(\boldsymbol{r}) - c_\gamma(\boldsymbol{r})\right] - 1 & \text{for} \quad h_\gamma(\boldsymbol{r}) \le 0 \\ -\beta u_\gamma(\boldsymbol{r}) + h_\gamma(\boldsymbol{r}) - c_\gamma(\boldsymbol{r}) & \text{for} \quad h_\gamma(\boldsymbol{r}) > 0 \end{cases}$$

In this, $c_\gamma(\boldsymbol{r})$ is the direct correlation function, $\chi_{\gamma\gamma'}(r)$ is the water susceptibility function, $u_\gamma(\boldsymbol{r})$is the protein-water interaction potential, and $\beta = 1/(k_B T)$ is the inverse temperature. The hydration free energy G_{hyd} is then obtained

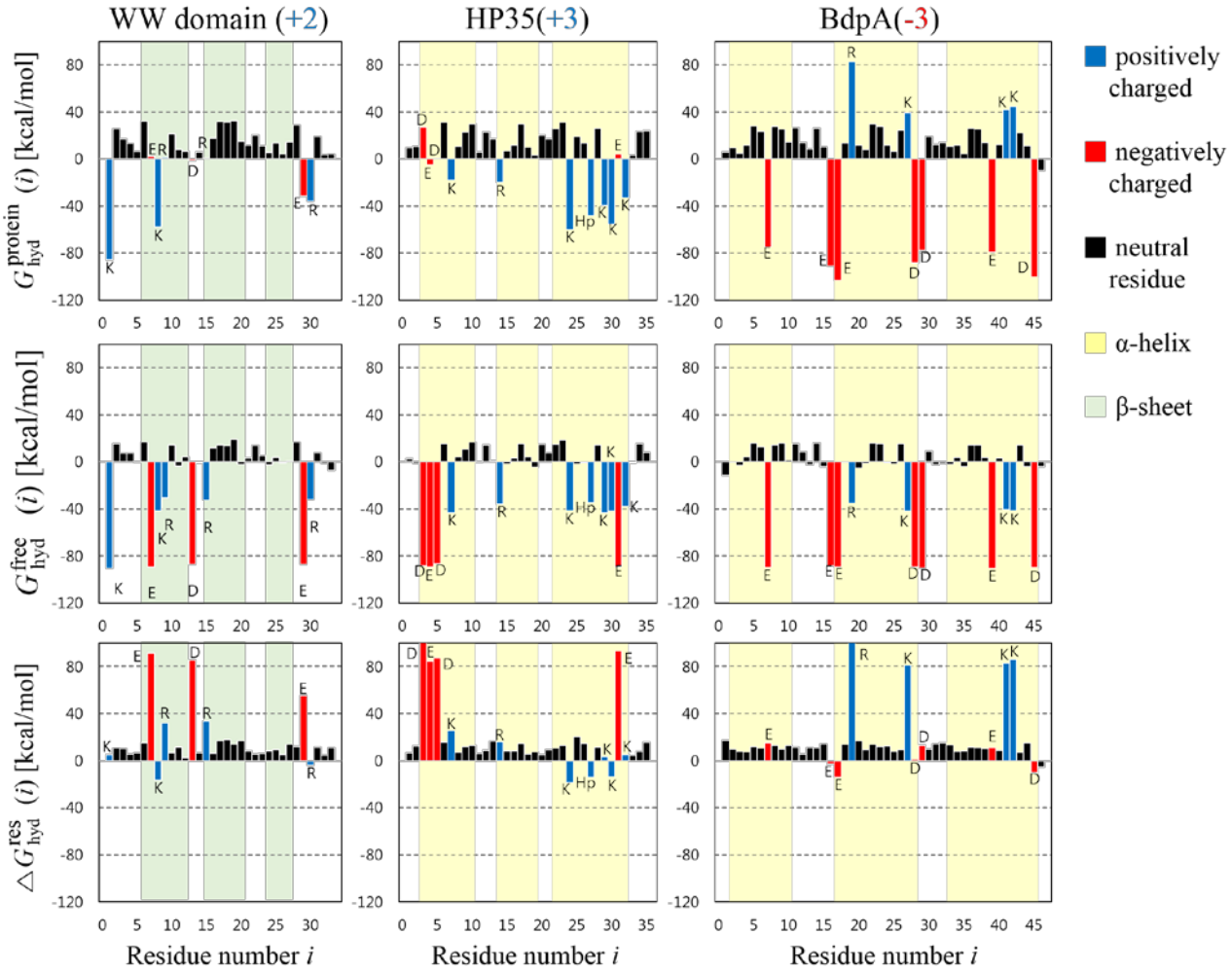

Figure 2. Hydration free energy $G_{\mathrm{hyd}}^{\mathrm{protein}}(i)$ of amino acid i embedded in a protein (top panels), $G_{\mathrm{hyd}}^{\mathrm{free}}(i)$ of free amino acid i (middle panels), and their difference $\Delta G_{\mathrm{hyd}}^{\mathrm{res}}(i) = G_{\mathrm{hyd}}^{\mathrm{protein}}(i) - G_{\mathrm{hyd}}^{\mathrm{free}}(i)$ quantifying the excess residual hydrophobicity of amino acid i (bottom panels) as a function of the residue number for WW domain, HP35 and BdpA. The contributions by negatively charged residues are shown in red, those from positively charged residues in blue, and those from neutral residues in black. The α-helix and β-sheet regions are colored by light yellow and light green, respectively.

Hanul Cho and Sihyun Ham

from

$$G_{\mathrm{hyd}} = \rho k_{\mathrm{B}} T \sum_{\gamma} \int d\boldsymbol{r} \left[\frac{1}{2} h_{\gamma}(\boldsymbol{r})^2 \Theta\left(-h_{\gamma}(\boldsymbol{r})\right) - c_{\gamma}(\boldsymbol{r}) - \frac{1}{2} h_{\gamma}(\boldsymbol{r}) c_{\gamma}(\boldsymbol{r}) \right]$$

where ρ is the average number density of water and Θ(x) is the Heaviside step function.

Results and Discussion

1. Excess residual hydrophobicity

We define the excess residual hydrophobicity to quantify how the hydrophobicity of constituting amino acids in protein depends on the protein context as follows. We first calculate the hydration free energy $G_{\mathrm{hyd}}^{\mathrm{protein}}(i)$ of amino acid *i* embedded in a protein: $G_{\mathrm{hyd}}^{\mathrm{protein}}(i)$ versus amino acid residues for the three proteins studied are displayed in the top panels of Figure 2. Secondly, using the hydration free energy for individual free amino acids obtained elsewhere,[20] we construct $G_{\mathrm{hyd}}^{\mathrm{free}}(i)$ versus the amino-acid sequence as shown in the middle panels of Figure 2. Finally, we compute the difference $\Delta G_{\mathrm{hyd}}^{\mathrm{res}}(i) = G_{\mathrm{hyd}}^{\mathrm{protein}}(i) - G_{\mathrm{hyd}}^{\mathrm{free}}(i)$ drawn in the bottom panels of Figure 2. The excess residual hydrophobicity $\Delta G_{\mathrm{hyd}}^{\mathrm{res}}(i)$ should be zero if the hydrophobicity of amino acid *i* is independent of the protein context, and therefore characterizes to what extent the hydrophobicity of amino acids is altered when they are embedded in a protein.

2. Dependence on protein total charge

We find from Figure 2 that charged residues (blue and red bars) exhibit the largest excess residual hydrophobicity ($\Delta G_{\mathrm{hyd}}^{\mathrm{res}}$). In addition, we notice a contrasting conduct between positively and negatively charged residues dependent upon the protein total charge. When the total charge of protein is positive as in WW domain (+2) and HP35 (+3), $\Delta G_{\mathrm{hyd}}^{\mathrm{res}}$ for negatively charged residues (red bars) display the large negative changes compared to positively charged residues (blue bars), whereas this trend is reversed when the protein net charge is negative as in BdpA (-3). This contrasting behavior can be rationalized in view of the protein hydration structure. When the protein net charge is negative, the equilibrium directional distribution of the periphery water molecules is such that water hydrogen is oriented toward the protein. That results in unfavorable electrostatic interactions between the water molecules and the positively charged residues, and tends to increase their hydration free energy as observed in BdpA. The contrasting results for WW domain and HP35 can be understood in a similar manner since the equilibrium orientation of water molecules around a protein of positive total charge is reversed such that water oxygen is directed toward the protein, which would lead to unfavorable electrostatic interactions between negatively water molecules and charged residues.

3. Secondary-structure effects

We examined how $\Delta G_{\mathrm{hyd}}^{\mathrm{res}}$ of amino acids are influenced by the secondary structures such as the salt-bridge formation and β-sheet regions versus loops. First, we find that the hydrophobicity of salt bridge forming amino acids in protein is exhibit enhanced than that of amino acids with no salt bridge formation (Table 1). The salt bridge is formed between charged residues, GLU exhibits most difference in that. The GLU forming salt bridge is more increase about 36 than that with no salt bridge for average $\Delta G_{\mathrm{hyd}}^{\mathrm{res}}$ of 300 conformation. Because the salt bridge interrupts the water-protein interaction on residues.[21] In other words, salt-bridge screens the water-protein interaction and reduces the strength of the interaction. This means that the thermodynamic effect of the salt bridge formation is to decrease the hydration free energy of amino acids. That is the thermodynamic effect of the salt bridge formation is to decrease the hydration free energy of the amino acids. Secondly, we find that no certain trend in $\Delta G_{\mathrm{hyd}}^{\mathrm{res}}$ for residues that belong to the α-helix comparing the loop. Because the α-helix have more twist complex structure, so there are several factors for determining the hydrophobicity of amino acids in protein such as the orientation of amino acid in twist complex and the environment of each α-helix in protein. But we find that amino acids in the β-sheets tend to

Table 1. Statistics of $\Delta G_{\mathrm{hyd}}^{\mathrm{res}}$ (kcal/mol) of salt-bridge and non-salt-bridge forming amino acid embedded in a protein for each 300 conformations (300ns MD run) WW domain, HP35 and BdpA.

	BdpA		HP35		WW domain	
	Salt Bridge	Non-SB	Salt Bridge	Non-SB	Salt Bridge	Non-SB
ARG	126.08±8.01	113.02±8.93	16.79±9.32	-0.52±8.26	26.04±16.37	-6.90±19.31
LYS	95.75±11.88	74.60±9.38	23.48±14.94	-6.78±12.30		
HIP			-7.66±4.87	-17.32±6.17		
ASP	-2.44±24.50	4.03±15.65	113.09±13.14	84.84±13.68	89.06±4.35	61.04±9.15
GLU	22.12±11.95	-3.72±14.75	99.11±8.22	78.66±14.51	92.21±6.91	56.12±7.59

display more enhanced the hydrophobicity than that of amino acids in the loop (Table 2). This means that amino acids in the loop sense the water-protein interaction is less than amino acids in the β-sheet. In this case, amino acids in the loop are exposed to the solvent. While, when forming β-sheet, the constituting amino acids are located the area that is less exposed to the solvent and thus $\Delta G_{\mathrm{hyd}}^{\mathrm{res}}$ of the β-sheet less enhanced than the loop with the result that be weak of the water-protein interaction. Especially, show that $\Delta G_{\mathrm{hyd}}^{\mathrm{res}}$ of the neutral residues in the central β-strand sandwiched by the β-sheets is more positive change than the edge β-sheets (Figure 2). The fact that such positive change in Figure 2 and Table 2 is $\Delta G_{\mathrm{hyd}}^{\mathrm{res}}$ implies that a cooperative effect is required to observe the substantial positive change in $\Delta G_{\mathrm{hyd}}^{\mathrm{res}}$ originating from the backbone dehydration. Such a cooperativity may explain highly insoluble nature (i.e., quite large hydrophobicity) of amyloid fibrils associated with many human diseases[22] since they exhibit consecutive cross-β structures and the cooperative effect is thus expected to be significant.

Conclusion

Hydrophobicity is one of the major driving factors for various biological processes, and elucidating how the hydrophobicity of a protein is determined is of fundamental importance. In this paper, we report the computational studies based on the integral-equation theory of liquids on how the hydrophobicity of constituting amino acids in protein depends on the context they are embedded in a protein. We observe that charged residues provide the largest contribution to the protein hydrophobicity, but negatively charged residues and positively charged ones play a distinct role depending on the protein total charge. We also find that the charged residues are relatively hydrophobic when the charged residue formed the salt-bridge than charged residues without the salt-bridge formation and amino acids in the β-sheets exhibit enhanced the hydrophobicity than loop because of the salt-bridge effect from decrease in surface area accessible to water and the cooperative dehydration effect. Our results supplement the traditional artless view on the hydrophobicity of a protein, and will also be valuable for understanding, predicting, and controlling the role of water involved in the biological activity of proteins.

Acknowledgments. This work has been supported by the project EDISON (EDucation-research Integration through Simulation On the Net), Chemistry.

References

1. Chaplin, M. Nat. Rev. Mol. Cell Biol. 2006, 7, 861−866.
2. Ball, P. Nature 2011, 478, 467−477.
3. Kauzmann, W. Adv. Protein Chem. 1959, 14, 1−64.
4. Tanford, C. Science 1978, 200, 1012−1018.
5. Dill, K. Biochemistry 1990, 29, 7133−7155.
6. Chiti, F.; Stefani, M.; Taddei, N.; Ramponi, G.; Dobson, C. M. Nature 2003, 424, 805−808.
7. Kyte, J.; Doolittle, R. F. J. Mol. Biol. 1982, 157, 105−132.
8. Roseman, M. A. J. Mol. Biol. 1988, 200, 513−522.
9. Hessa, T.; Meindl-Beinker, N. M.; Bernsel, A.; Kim, H.; Sato, Y.; Lerch-Bader, M.; Nilsson, I. M.; White, S. H.; von Heijne, G. Nature 2007, 450, 1026−1030.
10. Niwa, T.; Ying, B. W.; Saito, K.; Jin, W.; Takada, S.; Ueda, T.; Taguchi, H. Proc. Natl. Acad. Sci. U.S.A. 2009, 106, 4201−4206.
11. Ben-Naim, A. Hydrophobic Interactions; Plenum Press: New York, 1980.
12. Kovalenko, A. In Molecular Theory of Solvation; Hirata, F., Ed.; Kluwer: Dordrecht, Netherlands, 2003, p. 169.
13. Imai, T.; Harano, Y.; Kinoshita, M.; Kovalenko, A.; Hirata, F. J. Chem. Phys. 2006, 125, 024911.
14. Gouda, H.; Torigoe, H.; Saito, A.; Sato, M.; Arata, Y.; Shimada, I. Biochemistry 1992, 31, 9665−9672.
15. Chiu, T. K., Kubelka, J., Herbst-Irmer, R., Eaton, W. A., Hofrichter, J., & Davies, D. R. Proc.Natl.Acad.Sci.Usa. 2005, 102, 7517-7522.
16. Jäger, M.; Zhang, Y.; Bieschke, J.; Nguyen, H.; Dendle, M.; Bowman, M. E.; Noel, J.P.; Kelly, J. W. Proc. Natl. Acad. Sci. U.S.A. 2006, 103, 10648−10653.
17. DUKE, R., et al. Amber14, University of California, San Francisco, 2015.
18. Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Proteins 2006, 65, 712−725.
19. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. J. Chem. Phys. 1983, 79, 926−935.
20. Chong, S.-H.; Ham, S. Angew. Chem. Int. Ed. 2014, 53, 3961-3964.
21. Chothia, C., & Janin, J. Nature 1975, 256, 705-708.
22. Chiti, F.; Dobson, C. M. Annu. Rev. Biochem. 2006, 75, 333−366. C. Tem, G. M. Whitesides, *J. Am. Chem. Soc.* **1990**, *112*, 6409.

Table 2. Statistics of $\Delta G_{\mathrm{hyd}}^{\mathrm{res}}$ (kcal/mol) for 300 conformations of WW domain. Show the $\Delta G_{\mathrm{hyd}}^{\mathrm{res}}$ of amino acids in only both of β-sheet and loop.

	ALA	TRP	SER	ASN	ARG	GLU
β-sheet	9.29±2.14	15.16±3.41	5.27±2.25	17.03±3.13	32.08±5.84	91.65±7.41
loop	4.88±3.87	12.16±3.17	8.39±4.43	8.43±2.69	14.87±23.18	55.65±7.39