단백질 잔기의 패턴 분석을 이용한 단백질 안정화 요인 해석: 호열성 단백질과 일반 단백질간의 잔기 차이의 시스템적인 분석

백승필, 유영제
서울 대학교, 응용화학부

Elucidation of protein stabilizing factors using protein residual pattern: a systematic analysis of residual differences between thermophilic proteins and mesophilic proteins

Pack, Seung Pil and Yoo Young Je
School of Chemical Engineering, Seoul National University

Introduction

The molecular understanding of protein thermostability has been the important focus of many theoretical and experimental research efforts. Several researches have suggested that the factors that may contribute to enhanced thermostability include improved hydrogen bonding, better hydrophobic packing, enhanced secondary structure propensity, helix dipole stabilization, removal of residues sensitive to oxidation or deamination, and improved electrostatic interactions [1]. Unfortunately, these properties are not easy to engineer into any protein one want to stabilize.

Proteins from thermophilic organisms usually show substantially higher intrinsic thermal stabilities than their counterparts form mesophilic organisms, while retaining the basic fold characteristic of the particular protein family. The comparison of structures and sequences of homologous proteins from thermophilic and mesopjilic organisms could provide the important clues to stabilize proteins. Although proteins can be engineered to achieve greater stability by such a strategy, it is clear that no single and preferred mode of stabilization occurs, that is, it is difficult to derive general 'rule' for protein stabilization from the examples [2].

To reach general conclusions, systematical analyses, which are based on the computational studies analyzing various features for a group of proteins, have been tried. Argos and colleagues proposed the 'traffic rule' for preferred amino acid exchanges between mesophilic and thermophilic proteins through comparing sequences in several protein families [3]. Spassov and colleagues introduced parameters to evaluate the degree of optimization of hydrophobic and charge-charge interactions in protein structures through studying 14 thermophilic proteins, and concluded that these proteins are characterized by a higher degree of hydrophobic or electrostatic optimization than mesophilic ones [4].

Although these studies, to some extent, succeed to show several intrinsic or extrinsic factors to underlie the increased stability of thermostable protein groups, they also have a limitation in practical applications for engineering any protein to stabilize. It is still difficult to derive general methods for protein stabilization from the results.

In this study, we developed a protein residual pattern to find out the structural differences between stable proteins and less-stable proteins. The method is based on both of the conventional methods, the comparative study and systematic analysis. However, it is distinguished form other previous works in that it is a residue-based analysis, that is, it enables to analyze the structural characteristics of model protein group in the residual view.

The method categorizes the residues of proteins into eight classes under their own structure index, and could provide a residual pattern of structural properties in protein group. 20 pairs set of thermophilic and mesophilic proteins were analyzed to find out the residual differences of thermophilic proteins, which would be the residual factors applied to develop a strategy for stabilizing proteins.

Material and methods

Protein midels

The 20 pairs set of thermophilc and mesophilic protein structures was designed to be non-redundant and representative in protein structure families based on the structure data from Protein Data Bank (PDB) at the Research Collaboratory for Structural Bioinformatics (RCSB). Each set contains one thermophilic protein and its counter part mesophilic protein determined at a resolution of 0-2.0 Å and R-factor ≤ 0.25.

Residual structure index

The residual structure index was calculated by packing value based on the occluded surface algorithm was used as a local structure index describing residual conformational state in protein structure. The residual packing value for each residue is calculated by an extension of the occluded surface algorithm

Residual property calculation

Residual surface area, the ratio of side chain area to “random coil” value, and the determination of exposed/buried residue were calculated by the Getarea 1.1 program. The residual surface area computed is the locus of the center of a solvent molecule, as it rolls along the protein making maximum permitted contact. The default value of a solvent molecule size is 1.4Å, being representative of the size of a water molecule. The "random coil" value of a residue, X, is the average solvent-accessible surface area of X in the tripeptide Gly-X-Gly, a combination of 30 random conformations. Residues are considered to be solvent exposed if their ratio value exceeds 50%, and to be buried if their ratio is less than 20%.

Residual solvation free energy, residual energy and residual α-carbon flexibility were calculated by the Protable module of Biopolymer on SYBYL. The residual solvation free energies calculated are based on the atomic solvation parameters of Eisenberg and McLachan. Residual energy is determined as the sum of the energies of atoms in a residue, which contribute to atoms of each residue and to the total molecular mechanics energy of the molecule. The residuaḷ α-carbon flexibility is calculated by obtaining the temperature B value of the α-carbon atoms in the PDB data.

The number of hydrogen bonds, neighboring amino acids, and the determination of secondary structure were also calculated by the Protable module of Biopolymer on SYBYL. The secondary structure was determined by the Kabsch Sander procedure. The values are reported using the one-letter codes corresponding to the secondary structure states; B is the residue in an isolated beta-bridge, E is the extended strand that participates in the beta ladder, G is the 3/10 helix, H is the alpha helix, I is pi the helix, and T is the hydrogen bonded turn. N/d code indicates that there is no well-defined secondary structure at that residue.

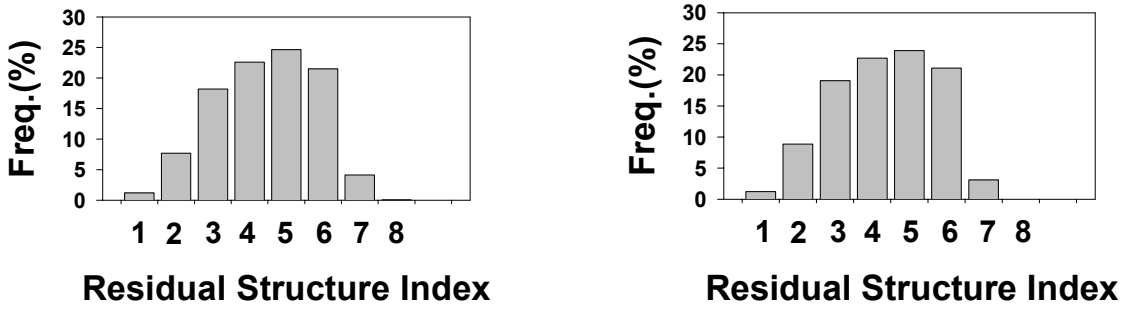
The number of ion pairs (salt bridge interaction), the number of cation-pi interactions and the number of disulfide bonds are calculated by the Protein Explorer package 1.9.

Result and Discussion

Concept of protein residual pattern

Residual structure index played major role for analyzing residual properties according to the residual structure environments and obtaining a statistical pattern of each residual properties over the residual structure environments. In this study, residual packing values based on the occluded surface algorithm were used as residual structure index such that they are representative values of all the residues in a protein indicating its local environments regardless of surface residues, buried residues, and boundary residues. In addition, the packing values are obtained as the normalized values form 0 to 1, which could make it easier to perform a statistical analysis. According to residual structure index, residues and its properties are categorized into 8 classes of range from 1 to 8, for examples, range 1 contained the residues with the values 0 to 0.1 of structure index, range 2 the values 0.1 to 0.2, and so on.

Figure 1(a) shows the relative distribution of residues of mesophilic protein groups. Figure 1 (b) shows the relative distribution of residues of thermophilic protein groups. In macroscopic view, the patterns of all the figures look similar to each other. Especially, the distribution pattern of residues in mesophilic protein groups is almost same as that of residues in non-redundant proteins. However, it shows small marginal difference compared with that of residues in thermophilic proteins. In the range 5, the residues of thermophilc protein groups show 1% higher frequencies than those of mesophilic protein group. This comparison reflected the others' conclusion that proteins from thermophiles do not differ strongly from their mesophilic counterparts, but show a bit of better hydrophobic packing of interior residues than its counterparts.



(a) thermophilic proteins (b) mesophilic proteins

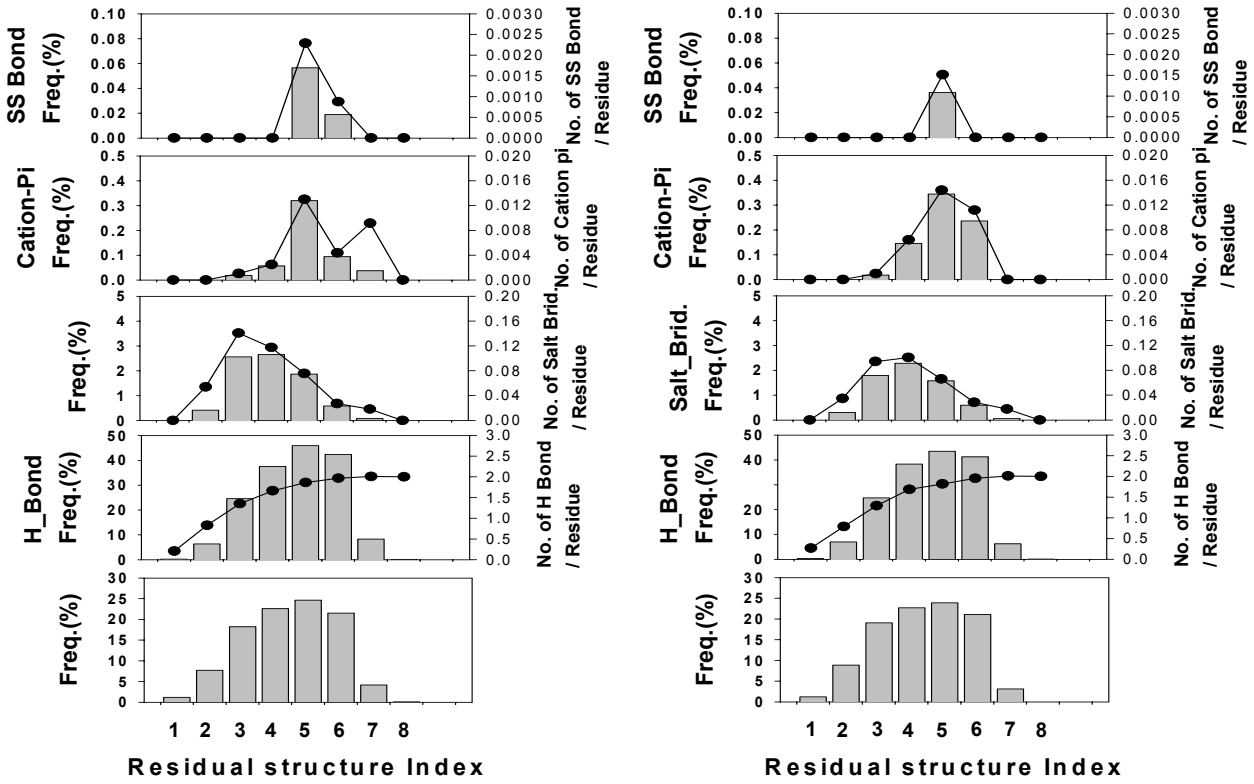
Figure 1. Comparison of structure index distribution between thermophilic proteins and mesophilic proteins.

Residual difference of molecular interactions between thermophilic proteins and mesophilic proteins

As molecular interaction or force, salt bridges (ion pair), cation pi interactions disulfide bonds, and hydrogen bonds residual were analyzed. The results of the residual difference of molecular interaction between thermophilic and mesophilic proteins are compared as shown

in Figure 2. In each range class, the relative composition of these interactions in thermophilic protein groups shows its different frequency compared with that of mesophilic protein groups. According to the interaction type, the range of distinct difference, in which the difference of both the distributions are over 20%, were provides as follows;

Salt bridges in thermophilic proteins show higher frequency in range 3 and 4 than those of mesophilic proteins. Cation pi interactions in thermophilic proteisn show higher frequency in range 7, and lower frequency in range 6. Disulfide bonds show higher frequency in range 5 and 6. Hydrogen bonds in thermophilic proteins show no different pattern compared those of mesophjilic proteins.



(a) thermophilic proteins (b) mesophilic proteins

Figure 2. Comparison of residual difference of molecular interactions between thermophilic proteins and mesophilic proteins.

Reference

1. Vilelle, C., Burdette, D.S. and Zeikus, J.G. 1996. *Biotechnol. Annu. Rev*. **2**, 1-83.
2. Vogt, G., Woell, S. and Argos, P. 1997. *J. Mol. Biol*. **269**, 631-643 4.
3. Arogos, P., Romann, M.G., Grau, U.M. Zuber, H., Frank, G. and Tratschin, J.D. 1979. *Biochemistry* **18**, 5698-5703.
4. Spassov, V.Z. Karshikoff, A.D. and Ladenstein, R. 1995. *Protein Sci*. **4**, 1516-1527.