Ecological Genetics and Genomics 1 (2016) 1–5

Contents lists available at ScienceDirect

Ecological Genetics and Genomics

journal homepage: http://www.elsevier.com/journals/ecological-genetics-and-genomics/2405-9854

# Structural adaption of extremophile proteins to the environments with special reference to hydrophobic networks

Keisuke Ueno, Martin Ibarra, Takashi Gojobori*

*Computational Bioscience Research Center, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia*

ARTICLE INFO

*Article history:*
Available online 29 October 2015

*Keywords:*
Extremophile
Functional adaptation
Hydrophobic network

ABSTRACT

Microorganisms in extreme environments have evolved with proteins and specific cell walls to survive harsh conditions. Functional proteins that enable organisms to adapt to the extreme environments are interesting targets for understanding molecular adaptation. With the discovery of functional proteins from the extremophiles, it has been known that particular amino acid sequences are preferred under an extreme environment. In the tertiary structure of a protein, hydrophobic residues of amino acids are known to play a crucial role in maintaining protein function. However, it remains unclear how the preference of amino acid sequences in relation to hydrophobicity is achieved to manifest novel protein functions. Here, we developed $R_H$ index that measures the strength of a hydrophobic network formed by residue-to-residue contacts between hydrophobic amino acids in a protein tertiary structure in order to scrutinize the adaptation mechanisms of a protein to extreme environments. We found that values of the $R_H$ index were significantly different between extremophile and mesophile proteins, and between halophile and hyperthermophile proteins. We also found that the values of the $R_H$ index vary across functional categories of proteins. Our findings will contribute to the understanding of mechanisms of molecular adaptation to various environments in an ecological system.



## 1. Introduction

Extremophiles are microorganisms that thrive in extreme environments such as high temperatures, acidic or alkaline stresses, high salinity, and nuclear radiation [16]. Those microorganisms, often cooperating with each other [17], produce extremolytes, e.g., osmolytes [10] with specific cell walls [17], to adapt to the harsh environments. Adaptation of extremophile proteins to extreme environments has been extensively investigated [6,7,13], because these proteins are actively functional, whereas vulnerable to stresses that trigger proteins denaturation and loss of biological activities.

Adaptation of extremophile proteins has attracted biologists' attention [15]. Number of the charged residues on the surface of a protein increases with the hydrophobic residues filling the cavity of the interior core; and disulfide bridges and multiple subunit structures often exist in hyperthermophile proteins [6,18]. In contrast, in halophile proteins, number of the acidic residues on the surface of a protein increases, whereas number of the disulfide bridges decreases [7,13]. However, it remains unclear how mutations were selected for their functional adaptation. For example, the mutations in hyperthermophile proteins are likely to stabilize the protein structure, while the mutations in halophile tend to destabilize the structure [15].

For understanding an adaptation mechanism of proteins in extreme environments, we focused on the interactions between hydrophobic residues of amino acids that plays a key role of protein folding and structural stabilities [4]. The hydrophobic interaction is also modulated under a high salinity condition through altering bulk water structures [12]. In this study, we developed the $R_H$ index to measure the strength of the hydrophobic network of a protein structure in the extreme environments. This $R_H$ index is a normalized number of contacts between hydrophobic residues in a protein tertiary structure by using a relationship between protein surface area and volume. We then investigated the $R_H$ index in various functional protein categories.

* Corresponding author.
*E-mail address:* takashi.gojobori@kaust.edu.sa (T. Gojobori).

http://dx.doi.org/10.1016/j.egg.2015.10.001
2405-9854/

## 2. Materials and methods

### 2.1. Protein structural data of extremophile

We obtained protein structural data from the Protein Data Bank (PDB; http://www.rcsb.org/pdb/), by excluding protein sequences with 30 amino acid residues or shorter. Using BLASTClust [1] at 90% sequence identity, we identified clusters of homologous proteins. We then selected a representative protein structure from each cluster. A total of 7839 structures were classified into halophile, mesophile, and thermophile or hyperthermophile proteins based on the annotation from JGI Genomes OnLine Database (GOLD) v.5 [14]. These classified protein structures were used for the subsequent analyses in the present study.

### 2.2. Hydrophobic networks of extremophile proteins

Hydrophobic interaction in the interior core of a protein is called a hydrophilic network. For measuring the strength of protein structure, we introduced a slight modification of the hydrophobic contact that calculates paired residues of amino acids in any atoms present within 6.5 Å [8]. In this study, an amino acid residue with a positive value of the hydropathy index [9] was treated as the hydrophobic residue.

It is difficult to compare directly the number of the hydrophobic contacts for the proteins between different families, because the number depends strongly on the size and shape of proteins. Therefore, we adopted a normalization technique based on the surface energy ($\sim\mu L^{2/3}$) for globular proteins with sequence length $L$ [5], where $\mu$ is a parameter for protein shapes, to provide a unified parameter independent of the size of a protein. We, here, introduced the $R_H$ index, which is associated with the strength of a hydrophobic network. The $R_H$ index is given by

$$R_H = N_H \Big/ L^{2/3}, \tag{1}$$

where $N_H$ is the number of the hydrophobic contacts. This is an average number of hydrophobic contacts per unit area of protein "virtual" surface.

### 2.3. Molecular functional categories associated with the $R_H$ index

We used Gene Ontology (GO) term [2] and GO slim term to examine if molecular functional categories are associated with the $R_H$ values. The molecular function enriched in upper and lower quartiles for the $R_H$ index of each extremophile protein was determined by goatools (http://github.com/tanghaibao/goatools) after Bonferroni correction.

## 3. Results and discussion

### 3.1. Local flexibility of extremophile protein structures

In hyperthermophile proteins, the change of filling the interior cavity can contribute to protein thermal stability under the high temperature condition [15]. In the study of molecular dynamic simulation and quasi-elastic neutron scattering of Initiation Factor 6 from *Methanocaldococcus jannascii*, the flexibility of a protein structure is reported to be related to the adaptation of extremophile proteins [3].

For halophile proteins, on the other hand, it is known that the mutations to small hydrophobic residues, in addition to deletion of disulfide bonds and insertion of peptides, can provide the flexibility of a structure of the halophile protein [15]. These observations suggest that the protein flexibility is a key for halophile protein structure and function [15].

Therefore, we first compared the *B*-factors of X-ray crystallography to investigate whether the flexibility of a local structure reflects the adaptation. Note that the *B*-factor is a so-called temperature factor that represents the fluctuation of atomic positions, giving a measure of the local flexibility.

For halophile proteins adapted to a high salinity environment, we analyzed proteins from *Haloarcula marismortui* found in the Dead Sea, where it grows in a saturated salt condition. On the other hand, for hyperthermophile proteins adapted to an extremely high temperature environment, we analyzed proteins from *Archeoglobus fulgidus* found in the oil field waters in the North Sea, where it grows at ~80 °C. In malate dehydrogenases from *H. marismortui* and *Bacillus anthracis str. Ames*, the *B*-factors decreased in the cores, while increased on the surfaces (Figs. 1A and 2B). The *B*-factor of

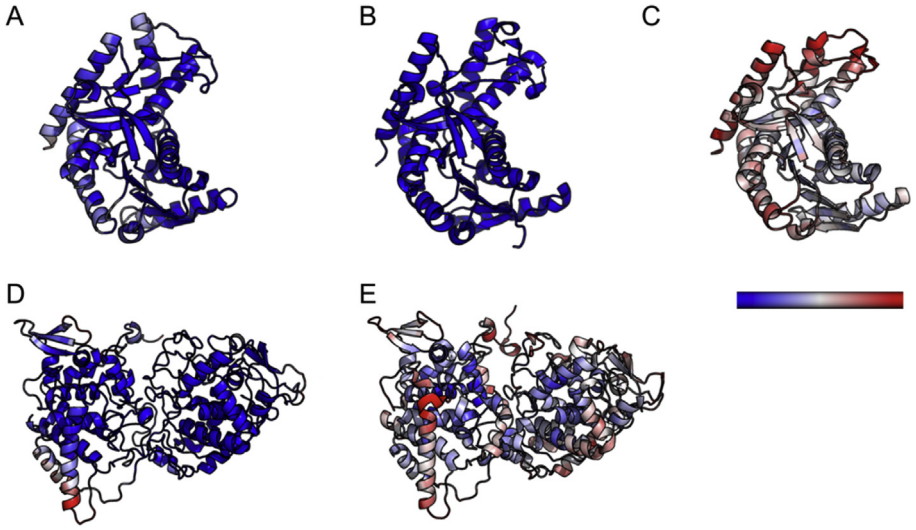


**Fig. 1.** Comparison between the *B*-factors of extremophile proteins. (A to C) Malate dehydrogenases (PDB code 1D3A, 2.9 Å resolutions; PDB code 3TL2, 1.7 Å resolutions; and PDB code 2X0J, 2.8 Å resolutions, respectively). (D and E) Catalase-peroxidases (PDB code 1ITK, 2.0 Å resolutions; and PDB code 1SJ2, 2.4 Å resolutions, respectively). (A and D) Halophiles. (B and E) Mesophiles. (C) Hyperthermophile. Each structure was colored according to the *B*-factors as indicated in the scale bar (A to C, 20 to 100; D and E, 20 to 50). Red color is flexible, and blue color is rigid. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)