Computational Biology and Chemistry 33 (2009) 137–148

Contents lists available at ScienceDirect

Computational Biology and Chemistry

journal homepage: www.elsevier.com/locate/compbiolchem

Research Article

Structural adaptation of the subunit interface of oligomeric thermophilic and hyperthermophilic enzymes

Elisa Maugini[a], Daniele Tronelli[a], Francesco Bossa[a], Stefano Pascarella[a,b,*]

[a] Dipartimento di Scienze Biochimiche "A. Rossi Fanelli", Università degli Studi di Roma "La Sapienza", Rome, Italy
[b] Centro Interdipartimentale di Ricerca per l'Analisi dei Modelli e dell'Informazione nei Sistemi Biomedici (CISB), Università degli Studi di Roma "La Sapienza", Rome, Italy

ARTICLE INFO

Article history:
Received 5 June 2008
Accepted 25 August 2008

Keywords:
Thermophiles
Hyperthermophiles
Subunit interface
Hydrophobic contact
Interface packing
Ion pairs

ABSTRACT

Enzymes from thermophilic and, particularly, from hyperthermophilic organisms are surprisingly stable. Understanding of the molecular origin of protein thermostability and thermoactivity attracted the interest of many scientist both for the perspective comprehension of the principles of protein structure and for the possible biotechnological applications through application of protein engineering. Comparative studies at sequence and structure levels were aimed at detecting significant differences of structural parameters related to protein stability between thermophilic and hyperhermophilic structures and their mesophilic homologs. Comparative studies were useful in the identification of a few recurrent themes which the evolution utilized in different combinations in different protein families. These studies were mostly carried out at the monomer level. However, maintenance of a proper quaternary structure is an essential prerequisite for a functional macromolecule. At the environmental temperatures experienced typically by hyper- and thermophiles, the subunit interactions mediated by the interface must be sufficiently stable. Our analysis was therefore aimed at the identification of the molecular strategies adopted by evolution to enhance interface thermostability of oligomeric enzymes. The variation of several structural properties related to protein stability were tested at the subunit interfaces of thermophilic and hyperthermophilic oligomers. The differences of the interface structural features observed between the hyperthermophilic and thermophilic enzymes were compared with the differences of the same properties calculated from pairwise comparisons of oligomeric mesophilic proteins contained in a reference dataset. The significance of the observed differences of structural properties was measured by a *t*-test. Ion pairs and hydrogen bonds do not vary significantly while hydrophobic contact area increases specially in hyperthermophilic interfaces. Interface compactness also appears to increase in the hyperthermophilic proteins. Variations of amino acid composition at the interfaces reflects the variation of the interface properties.

1. Introduction

Many terrestrial environments present physical and chemical conditions that can be defined extreme from an anthropocentric perspective. The organisms able to thrive in such extreme environments are called extremophiles or polyextremophiles when the environment presents combination of different extreme conditions. Among extreme conditions, high temperatures of hot permanent environments such as hydrothermal vents, volcanic areas, hot springs and the like, are relatively common (Rothschild and Mancinelli, 2001). Organism able to grow at high temperatures are defined thermophiles: in particular, organisms with an optimal growth temperature between 50 and 80 °C are named thermophiles while above 80 °C are the hyperthermophiles, mostly Archaea and Bacteria (Vieille and Zeikus, 2001). These organisms adopted several strategies to survive at high temperatures involving physiological modifications of the intracellular environment and the synthesis of thermostable biomolecules. Enzymes from thermophiles and hyperthermophiles are indeed surprisingly stable. Understanding of the molecular origin of protein thermostability and thermoactivity attracted the interest of many scientists both for the perspective comprehension of the principles of protein structure and for the possible biotechnological applications through application of protein engineering. More recently, study of the molecular and physiological properties of extremophiles became of interest for exobiology and astrobiology as well (Kounaves, 2007; Pikuta et al., 2007). Therefore, many studies were devoted to the comprehension of the etiology of molecular basis of the adaptation

* Corresponding author at: Dipartimento di Scienze Biochimiche, Università "La Sapienza", P.le A. Moro 5, 00185 Rome, Italy. Tel.: +39 06 49917694; fax: +39 06 49917566.
E-mail address: Stefano.Pascarella@uniroma1.it (S. Pascarella).

1476-9271/$ – see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compbiolchem.2008.08.003

to high temperatures (Szilágyi and Závodszky, 2000). Comparative studies at sequence and structure levels were aimed at detecting significant differences of structural parameters related to protein stability between thermophilic and hyperhermophilic structures and their mesophilic homologs. Although it now accepted that no general rule can be derived, the comparative studies were very useful in identification of a few recurrent themes which the evolution utilized in different combinations in different protein families. A plethora of factors were invoked to explain thermal stabilization of enzymes which are relevant in different families: variations of secondary structure properties including helix dipole stabilization, amount of proline in α-helices, β-strand content; degree of compactness including the number and size of cavities; fractional polar surface area, buried surface area, length of loops; aspects of general amino acid composition such as decrease in the number of thermolabile residues, increase of charged residues, greater frequency of β-branched residues and the like (Vogt et al., 1997). However, the most frequently observed variations are the increase in the number of electrostatic interactions as salt bridges and the fraction of exposed apolar surface and, in general, the optimization of hydrophobic interaction. To this respect, systematic comparative studies concluded that hyperthermophiles should be treated separately from the thermophiles for the analysis of molecular adaptations (Szilágyi and Závodszky, 2000).

Although many theoretical and experimental research efforts have been spent in the past, the molecular basis of protein thermostability remains rather elusive. Moreover, many studies were focused on the molecular adaptation at the monomer level and relatively little was carried out at the subunit interface level. This work would like to fill the gap and was focused on the subunit interface of oligomeric hyper- and thermostable enzymes. Of course, maintenance of a proper quaternary structure is an essential prerequisite for a functional macromolecule. At the environmental temperatures experienced typically by hyper- and thermophiles, the subunit interactions mediated by the interface must be sufficiently stable. Our analysis was therefore aimed at the identification of the molecular strategies adopted by evolution to enhance interface thermostability of oligomeric enzymes.

This aspect has not been previously analyzed in detail because of paucity of structural data, although studies on single families have been reported (for example, Coquelle et al., 2007). The number of oligomeric hyper- and thermophilic structures available in the structural databanks is now sufficient to undertake a systematic comparative analysis. Therefore we applied and extended our comparative approach, previously utilized for the analysis of the interfaces of psychrophilic oligomeric enzymes (Tronelli et al., 2007), to the study the molecular adaptation at the subunit interface of hyper- and thermophilic oligomeric enzymes.

2. Experimental procedures

2.1. Collection of main dataset

The crystallographic structures of the available oligomeric enzymes from thermophilic and hyperhermophilic organisms were retrieved from the Brookhaven Protein Data Bank (PDB) (Berman et al., 2000). The search was carried out with the keywords: “thermo”, “hot”, “heat tolerant”, “heat stable”, “pyro” and the like. The protein structures corresponding to the biological units were collected from the Protein Quaternary Structure (PQS) databank (Henrick and Thornton, 1998). Homologous structures from mesophilic organisms with the same oligomerization state were subsequently retrieved from PDB and PQS by means of the program BLAST (Altschul et al., 1990). To ensure structural homology, only sequences sharing ≥30% residue identity to the extremophilic sequence were considered. Only unique structures were retrieved, and in the presence of alternative structures for the same protein, only those displaying the best resolution and without point mutations, were collected. Proteins from plants were not taken into consideration owing to the ambiguous definition of “optimum temperature” for such organisms.

In order to assess the structural similarity within each collected family, we performed a structural alignment using the CE-MC program (Guda et al., 2004). Sequences of the selected proteins were aligned to each mesophilic homolog. The alignments were then manually corrected by inspection of the superimposed structures.

All the programs were written in Perl language and run under IRIX 6.5 or RED HAT ENTERPRISE LINUX 4.0 and OpenSuSE 10.3 operating systems.

2.2. Crystallographic structure quality assessment

All structures showing a resolution worse than 2.85 Å were excluded from the main dataset. All the incomplete interface sidechains were rebuilt using the program BIOPOLYMER of the InsightII package (version 2005; Accelrys, San Diego, CA, USA). The sidechain rotamer displaying the lower nonbond energy was kept and treated as experimental.

Ligands (cofactors, inhibitors, substrate analogs, etc.) and solvent molecules were always removed from the structures.

Quality check of the crystallographic structures was carried out using the PROCHECK software (Laskowski et al., 1993).

2.3. Identification of interface residues

Interface residues were defined as those residues that show a change in solvent accessibility area upon monomer association. Those residues for which the change was more than 90% were defined as composing the core interface (Bahadur et al., 2003). Solvent accessibility computation was performed with the program NACCESS (Hubbard and Thornton, 1993). The change in solvent accessibility area for each residue in the monomeric state and in the oligomeric state was calculated using a Perl script.

The structural similarity of the subunit interfaces within each protein family was evaluated on the basis of the multiple structure alignment. To ensure that the interface was structurally conserved within each family and the selected structural data comparable, the interface C_α carbons of each mesophilic member were superimposed to the equivalent atoms from the thermophilic homolog. Only interfaces showing RMSD ≤ 1.3 Å were considered similar. This threshold is within the expected structural variation corresponding to the range of sequence similarities of the multiple structure alignments (Chothia and Lesk, 1986). Indeed, the expected value of RMSD for a pair of homologous proteins whose sequence identity is 30% is equal to 1.42 Å. RMSDs were calculated using DeepView-Swiss-PdbViewer “iterative magic fit” tool (Guex and Peitsch, 1997) and the InsightII package (version 2005; Accelrys, San Diego, CA 92121, USA).

2.4. Surface characteristics

The program NACCESS (Hubbard and Thornton, 1993) was utilized to calculate the percentage of the overall surface composing the interface and the core interface, the percentage of polar and non-polar atomic contribution to the interface and the percentage of polar and non-polar atomic contribution to the core interface.

The overall hydrophobic contact area between residues of different monomers was calculated using the program PDB_NP_CONT (Drabløs, 1999) with the aid of a Perl script. The PDB_NP_CONT