Journal of Theoretical Biology 267 (2010) 388–395

Contents lists available at ScienceDirect

Journal of Theoretical Biology

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

Conformational characterization of disulfide bonds: A tool for protein classification

José Rui Ferreira Marques [a], Rute R. da Fonseca [b], Brett Drury [c], André Melo [a,*]

[a] REQUIMTE/Departamento de Química e Bioquímica, Faculdade de Ciências da Universidade do Porto, Rua do Campo Alegre, 687, 4169-007 Porto, Portugal
[b] CIMAR/CIIMAR, Centro Interdisciplinar de Investigação Marinha e Ambiental, Universidade do Porto, Rua dos Bragas, 177, 4050-123 Porto, Portugal
[c] LIAAD-INESC, Rua de Ceuta, 118, 6º, 4050-190 Porto, Portugal

ARTICLE INFO

Article history:
Received 16 June 2010
Received in revised form
29 August 2010
Accepted 8 September 2010
Available online 16 September 2010

Keywords:
Disulfide bond
Conformer
Cluster analysis
Protein classification

ABSTRACT

Background: Throughout evolution, mutations in particular regions of some protein structures have resulted in extra covalent bonds that increase the overall robustness of the fold: disulfide bonds. The two strategically placed cysteines can also have a more direct role in protein function, either by assisting thiol or disulfide exchange, or through allosteric effects. In this work, we verified how the structural similarities between disulfides can reflect functional and evolutionary relationships between different proteins. We analyzed the conformational patterns of the disulfide bonds in a set of disulfide-rich proteins that included twelve SCOP superfamilies: thioredoxin-like and eleven superfamilies containing small disulfide-rich proteins (SDP).

Results: The twenty conformations considered in the present study were characterized by both structural and energetic parameters. The corresponding frequencies present diverse patterns for the different superfamilies. The least-strained conformations are more abundant for the SDP superfamilies, while the "catalytic" +/−RHook is dominant for the thioredoxin-like superfamily. The "allosteric" −RHSaple is moderately abundant for BBI, Crisp and Thioredoxin-like superfamilies and less frequent for the remaining superfamilies. Using a hierarchical clustering analysis we found that the twelve superfamilies were grouped in biologically significant clusters.

Conclusions: In this work, we carried out an extensive statistical analysis of the conformational motifs for the disulfide bonds present in a set of disulfide-rich proteins. We show that the conformational patterns observed in disulfide bonds are sufficient to group proteins that share both functional and structural patterns and can therefore be used as a criterion for protein classification.

1. Introduction

Disulfide bonds are a common motif in Nature. These structural elements have a significant role in the thermal stability and function of proteins (Bhattacharyya et al., 2004; Creighton, 1988; Hogg, 2003; Klink et al., 2000; Sardiu et al., 2007). From an evolutionary perspective, these bonds are a relatively recent addition to protein structure (Brooks and Fresco, 2002; Brooks et al., 2002; Jordan et al., 2005; Schmidt and Hogg, 2007) According to the respective functions, the disulfide bonds can then be classified as structural, catalytic or allosteric (Schmidt et al., 2006; Schmidt and Hogg, 2007). Schmidt et al. (2006) have performed a thorough analysis of disulfides present in the X-ray structures of the PDB data base, and found that both catalytic and allosteric disulfides fell into particular structural categories. The two groups had a higher average potential energy, which reflected their functional role that implied easy bond breaking (Schmidt et al., 2006).

The disulfide three-dimensional structure is highly conserved in Nature and has been used for protein clustering (Cheek et al., 2006; Chuang et al., 2003; Harrison and Sternberg, 1996; Thangudu et al., 2007). Different schemes have been introduced to classify the disulfide conformers (Harrison and Sternberg, 1996; Hutchinson and Thornton, 1996; Ozhogina and Bominaar, 2009; Schmidt et al., 2006; Srinivasan et al., 1990) and in this work we adopted the scheme proposed by Schmidt et al. (2006). We analyzed a sample of disulfide bonds associated with a protein set extracted from *SCOP* data base (Andreeva et al., 2004, 2008; Murzin et al., 1995). The protein set included eleven superfamilies of small disulfide-rich proteins (SDP) and the thioredoxin-like superfamily. Each superfamily selected for the protein set had to fit the following criteria: (i) contain a minimum of thirty disulfide bonds, (ii) have a minimum of five PDB structures available, (iii) have X-ray structures with a resolution higher than 2.5 Å and (iv) have only uncomplexed structures. In order to understand

* Corresponding author. Tel.: +351220402503; fax: +351220402659.
E-mail addresses: zerui.marques@fc.up.pt (J.R. Marques), rute.r.da.fonseca@gmail.com (R.R. da Fonseca), brett.drury@gmail.com (B. Drury), asmelo@fc.up.pt (A. Melo).

0022-5193/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jtbi.2010.09.012

whether or not the structure of the disulfides reflected functional or evolutionary relationships between the different proteins, we grouped the disulfide from the 12 superfamilies in different clusters using a hierarchical clustering analysis (HCA) and a structural-based distance protocol. The results demonstrate that the clusters' aggregate superfamilies share both functional and structural patterns, therefore we conclude that the use of disulfide bonds conformational patterns is a valid protein classification criterion.

2. Methodology

The scheme used in this work to classify the disulfide conformers was based on five relevant torsion angles (Fig. 1). The disulfide species were treated as symmetrical. In this context, only twenty conformational categories had to be considered (Table 1). For example the –RHHook conformational category can be obtained by either combinations of torsion angles (−,+,+,−,−) or (−,−,+,+,−). This classification was based on structural patterns (Schmidt et al., 2006) that included main, orientational and peripheral motifs (Table 2).

Representative structures for the different conformational categories are presented in Tables 3–5.

The protein set under study is characterized in Table 6. We determined the five relevant torsion angles (χ_1, χ_2, χ_3, χ_2' and χ_1') for each disulfide bond. Additionally, the (C_α–C_α' and C_β–C_β') distances and the dihedral strain energy (DSE) were also evaluated.

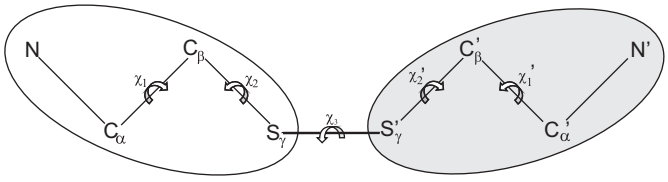

Fig. 1. Graphical representation of the five torsion angles used to classify the disulfide conformers.

Table 1
Classification of disulfide bonds in conformational categories (Schmidt et al., 2006).

Disulfide category[a]	χ_1	χ_2	χ_3	χ_2'	χ_1'
−LHSpiral	−	−	−	−	−
−RHHook	−	+	+	−	−
+/−RHSpiral	+	+	+	+	−
+/−LHSpiral	+	−	−	−	−
−RHSpiral	−	+	+	+	−
+/−RHHook	+	−	+	+	−
+RHSpiral	+	+	+	+	+
−LHHook	−	−	−	+	−
−/+RHHook	−	−	+	+	+
−RHStaple	−	−	+	−	−
+/−LHHook	+	−	−	+	−
−/+LHHook	−	−	−	+	+
+/−LHStaple	+	+	−	+	−
−LHStaple	−	+	−	+	−
+LHSpiral	+	−	−	−	+
+LHHook	+	−	−	+	+
+RHHook	+	+	+	−	+
+/−RHStaple	+	−	+	−	−
+LHStaple	+	+	−	+	+
+RHStaple	+	−	+	−	+

−: negative value for the respective torsion angle; +: positive value for the respective torsion angle.

[a] LH: left-handed oriented; RH: right-handed oriented.

Table 2
Characteristic conformational motifs used for disulfide classification.

Main motifs	χ_2	χ_3	χ_2'	Orientational motifs	χ_3	Peripheral motifs	χ_1	χ_1'
Spiral	+	+	+	**LH**	−	+	+	+
	−	−	−					
Staple	+	−	+			−	−	−
	−	+	−					
Hook	+	+	−	**RH**	+	+/−	+	−
	−	+	+					
	+	−	−			−/+	−	+
	−	−	+					

The DSE quantity was expressed, as a function of the five above-mentioned torsion angles, by the empirical equation (Katz and Kossiakoff, 1986; Weiner et al., 1984):

$$\begin{aligned} DSE(\text{kJ mol}^{-1}) = {} & 8.37(1+\cos(3\chi_1))+8.37(1+\cos(3\chi_1')) \\ & +4.18(1+\cos(3\chi_2))+4.18(1+\cos(3\chi_2')) \\ & +14.64(1+\cos(2\chi_3))+2.51(1+\cos(3\chi_3)) \end{aligned} \quad (1)$$

The DSE quantity provided a useful ranking of the most favored disulfide conformations. The minimum (2.5 kJ mol^{-1}) and the maximum (84.5 kJ mol^{-1}) values of DSE correspond to the torsion angles combinations (60^o, 60^o, $\pm 83^o$, 60^o, 60^o) and (0°, 0°, 0°, 0°, 0°), respectively (Schmidt et al., 2006). Despite its simplicity, this equation has been successfully applied for a semi-quantitative evaluation of the strain energy in disulfide bonds (Schmidt et al., 2006; Schmidt and Hogg, 2007).

Representative conformations of the different types of disulfide bonds (structural, catalytic or allosteric) are identified in Table 7. We will be referring to bonds with the conformations +/−RHHook as "catalytic", and −RHStaple as "allosteric", because these two types of bonds were found to be intimately associated with those conformational categories (Schmidt et al., 2006).

A computer program, designated by *Disulph*, was developed to perform the calculations. The disulfide bonds propensity Pr_A, for a superfamily A with np_A PDB structures, was calculated as

$$\text{Pr}_A = (1/np_A)\sum_{k=1}^{np_A} 100 \times nss_k/nres_k, \quad (2)$$

where nss_k and $nres_k$ were, respectively, the number of disulfide bonds and the number of coded residues in the PDB structure k. This quantity evaluates the frequency of the disulfide bonds within a superfamily. It is calculated as the average frequency associated with a correspondent sample of PDB structures.

The frequencies associated with all the conformational categories, defined in Table 1, were then evaluated for each superfamily and for the sample. These quantities were used to build a square Euclidean distances matrix, whose elements ($d^2_{Euclidean}(A,B)$) were defined as

$$d^2_{Euclidian}(A,B) = \sum_{i=1}^{20}(freq(i,A)-freq(i,B))^2; \quad A=1,\ldots,12 \text{ and } B=1,\ldots,12 \quad (3)$$

In Eq. (3), $freq(i,A)$ and $freq(i,B)$ are, respectively, the frequency of conformational category i in the superfamilies A and B. The square Euclidean distances matrix defines a metric for evaluating the similarities between objects in n-dimensional spaces and therefore can be used in cluster analysis.

In order to represent this matrix, we adopted the intuitive formalism introduced by Xie et al. (2000). The coordinates of the