

# Conformational characterization of disulfide bonds: A tool for protein classification

José Rui Ferreira Marques, Rute R. da Fonseca, Brett Drury, André Melo

#### ► To cite this version:

José Rui Ferreira Marques, Rute R. da Fonseca, Brett Drury, André Melo. Conformational characterization of disulfide bonds: A tool for protein classification. Journal of Theoretical Biology, 2010, 267 (3), pp.388. 10.1016/j.jtbi.2010.09.012 . hal-00634004

## HAL Id: hal-00634004 https://hal.science/hal-00634004

Submitted on 20 Oct 2011

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

### Author's Accepted Manuscript

Conformational characterization of disulfide bonds: A tool for protein classification

José Rui Ferreira Marques, Rute R. da Fonseca, Brett Drury, André Melo

PII:S0022-5193(10)00478-9DOI:doi:10.1016/j.jtbi.2010.09.012Reference:YJTBI6152

To appear in: Journal of Theoretical Biology

Received date:16 June 2010Revised date:29 August 2010Accepted date:8 September 2010



www.elsevier.com/locate/yjtbi

Cite this article as: José Rui Ferreira Marques, Rute R. da Fonseca, Brett Drury and André Melo, Conformational characterization of disulfide bonds: A tool for protein classification, *Journal of Theoretical Biology*, doi:10.1016/j.jtbi.2010.09.012

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

# Conformational characterization of disulfide bonds: a tool for protein classification

José Rui Ferreira Marques<sup>1</sup> (zerui.marques@fc.up.pt)

Rute R. da Fonseca<sup>2</sup> (rute.r.da.fonseca@gmail.com)

Brett Drury<sup>3</sup> (brett.drury@gmail.com)

André Melo<sup>1,#</sup> (asmelo@fc.up.pt)

 <sup>1</sup>REQUIMTE/Departamento de Química e Bioquímica, Faculdade de Ciências da Universidade do Porto, Rua do Campo Alegre, 687, 4169-007 Porto, Portugal
 <sup>2</sup> CIMAR/CIIMAR, Centro Interdisciplinar de Investigação Marinha e Ambiental, Universidade do Porto, Rua dos Bragas, 177, 4050-123 Porto, Portugal

<sup>3</sup> LIAAD-INESC, Rua de Ceuta, 118, 6º, 4050-190 Porto, Portugal

<sup>#</sup> Corresponding author (E-mail address: asmelo@fc.up.pt, Tel: +351220402503, Fax:

+351220402659.

#### 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.

Keywords: Disulfide bond, conformer, cluster analysis, protein classification.

.aio

#### 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 database, 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; Andreeva *et al.*, 2008; Murzin *et al.*, 1995). The protein set included eleven superfamilies of small disulfide-rich proteins (*SDP*) and the thioredoxin-like super-family. 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 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.

#### Methodology

The scheme used in this work to classify the disulfide conformers was based on five relevant torsion angles (Figure 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).

[Insert Figure 1] [Insert Table 1] [Insert Table 2]

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

[Insert Table 3]

[Insert Table 4]

[Insert Table 5]

The protein set under study is characterized in Table 6. We determined the five

relevant torsion angles ( $\chi_1$ ,  $\chi_2$ ,  $\chi_3$ ,  $\chi_2'$  and  $\chi_1'$ ) for each disulfide bond. Additionally, the

 $(C_{\alpha}-C_{\alpha}' \text{ and } C_{\beta}-C_{\beta}')$  distances and the dihedral strain energy (*DSE*) were also evaluated.

#### [Insert Table 6]

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):

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

The *DSE* quantity provided a useful ranking of the most favored disulfide conformations. The minimum (2.5 kJ mol<sup>-1</sup>) and the maximum (84.5 kJ mol<sup>-1</sup>) values of *DSE* correspond to the torsion angles combinations (60°, 60°, ±83°, 60°, 60°) 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,

$$\Pr_A = (1/np_A) \sum_{k=1}^{np_A} 100 \times nss_k / nres_k , \qquad (2)$$

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 super-family and for the sample. These quantities were used to build a square Euclidean distances matrix, whose elements  $(d_{Euclidean}^2(A, B))$  were defined as:

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

In equation (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 original objects (the twelve superfamiles) were projected in the 3D Cartesian space by minimizing the square deviation cost function *SD*:

$$SD = \sum_{A=1}^{12} \sum_{B=1}^{A-1} \left( d(A,B) - d_{Euclidian}^2(A,B) \right)^2 ,$$
(4)

where d(A,B) was the distance between the projections the superfamilies **A** and **B** in the 3D Cartesian space. We used the Newton method to carry out the iterative minimization process. The procedure associated with equation (4) was introduced for visualizing large chemical databases (Xie *et al.*, 2000). The minimization of this equation provided an appropriate representation of the original high-space of the chemical descriptors in a low dimensional space (2D or 3D).

The square Euclidean distances matrix was then used for a *HCA* procedure (Johnson and Wichern, 2007), which provided a classification of the superfamilies in different clusters. We evaluated the consistency of the *HCA* partitioning, by the evaluation of the square Euclidean distances matrix in the cluster space. The elements of this matrix were all the mean-square distances between a cluster  $C_i$  with  $n_{C_i}$  superfamilies and a

cluster  $C_i$  with  $n_{C_j}$  superfamilies (  $MSd_{Euclidean}^2(C_i, C_j)$  ) and within a cluster  $C_i$ 

$$(MSd_{Euclidean}^{2}(C_{i})):$$

$$MSd_{Euclidian}^{2}(C_{i}, C_{j}) = (1/(n_{C_{i}} \times n_{C_{j}})) \times \sum_{A=1}^{n_{C_{i}}} \sum_{B=1}^{n_{C_{j}}} d_{Euclidian}^{2}(A, B)$$
(5)

$$MSd_{Euclidian}^{2}(C_{i}) = (2/(n_{C_{i}} \times n_{C_{i}})) \times \sum_{A=1}^{n_{C_{i}}} \sum_{B=1}^{A-1} d_{Euclidian}^{2}(A,B)$$
(6)

This matrix was defined according to the mean linkage criterion within the *HCA* procedure (Johnson and Wichern, 2007). The dissimilarity between two clusters  $C_i$  and  $C_j$  increased with the increasing of the correspondent non-diagonal element  $(MSd_{Euclidean}^2(C_i, C_j))$ . On the other hand, the similarity within a cluster  $C_i$  increases with the decreasing of the correspondent diagonal element  $(MSd_{Euclidean}^2(C_i))$ . In this work we used the *HCA* divisive method which partitioned successively an initial set with *n* objects into finer clusters. The correspondent algorithm was the following:

- (i) Assign the *n* objects to a single cluster.
- (ii) Compute a distance matrix in the cluster space using an appropriate metric.
   As was mentioned above, we adopted a square Euclidean metric in this work.
- (iii) Find the least similar objects and separate them in different clusters.
- (iv) Repeat steps (ii) and (iii) until the diagonal elements of this matrix being significantly smaller than the non-diagonal ones.

#### Results

The characterization of the disulfide conformational categories found in our sample is presented in Table 7. The –LHSpiral is the most frequently observed category (28.9 %) and has the lowest *DSE* (11.5 kJ mol<sup>-1</sup>). Additionally, six least strained categories (-

LHSpiral, +/-RHSpiral, +/-LHSpiral, -RHSpiral, +RHSpiral and -/+RHHook) are clearly prevalent (63.1 %) relative to the remainder of the most strained categories (36.9 %). The representative conformations for catalytic (+/-RHHook) and allosteric (-RHStaple) disulfide bonds have moderate *DSE* values. We found the d( $C_{\alpha}$ - $C_{\alpha}$ ') distances to be more relevant for disulfide conformational specificities than the d( $C_{\beta}$ - $C_{\beta}$ ') distances (Table 7). The d( $C_{\alpha}$ - $C_{\alpha}$ ') distances were quite insensitive to the nature of conformational categories (varies from 3.3 Å to 4.0 Å), while the d( $C_{\beta}$ - $C_{\beta}$ ') distances had a significant variation over the series (from 4.4 Å to 6.0 Å). For instance, in agreement with Schmidt *et al* (2006), the –RHStaple conformational categories.

[Insert Table 7]

[Insert Table 8]

[Insert Figure 2]

The frequencies for the different conformational categories, calculated for each superfamily, are presented in Table 8 and Figure 2. From this figure, it is evident that thioredoxin-like and *SDP* superfamilies exhibit very distinct conformational patterns. The least strained conformations are significantly abundant in *SDP* superfamilies present significant abundances (from 43.4 % to 86.5 %), but occur at a very low frequency in thioredoxin-like superfamily (13.8 %). This is obvious for the most stable conformation (–LHSpiral) for which the *SDP* superfamilies present frequencies at least four times larger than the thioredoxin-like frequency (from 12.1 % to 43.8 % against 3.1 %; Table 8 and Figure 2). Most of the disulfide bonds of thioredoxin-like superfamily (50.8 %) are associated with the "catalytic" +/-RHHook conformation,

whereas this is relatively rare (from 0.0 % to 7.7 %) for the *SDP* superfamilies (Table 8 and Figure 2). On the other hand, the "allosteric" –RHSaple is moderately abundant for BBI (24.2 %), Crisp (24.1 %) and thioredoxin-like (16.9 %) superfamilies and scarce (from 0.0 % to 5.7 %) for the remainder superfamilies.

#### [Insert Figure 3]

Further insight into how the structural similarities between disulfides can reflect relationships between different proteins was obtained with a *HCA* procedure, whose dendrogram (Murtagh, 1984) is presented in Figure 3. The 3D-cartesian projection of the respective square Euclidean distances matrix is represented in Figure 4 together with the six clusters identified by this analysis. Four clusters reflect the main structural and functional motifs identified in the sample:

- $\circ~$  Cluster 1 includes the catalytic proteins of thioredoxin-like superfamily, with the lowest disulfide propensities and a dominant  $\alpha/\beta$  secondary structure;
- Cluster 4 includes most of the metabolic superfamilies (Cystine-Knot, EGF-Laminin and Plant lectins), with a dominant β secondary structure;
   Cluster 5 includes most of the toxin/defense superfamilies (Defensin-like, omega toxins, small snake toxins and scorpions-like toxins), with moderate to high disulfide propensities and a dominant β secondary structure;
- $\circ$  Cluster 2 includes the plant protease inhibitors of BBI superfamily, with high disulfide propensities and a dominant β secondary structure.

[Insert Figure 4]

The remainder two clusters reflect divergences from the mentioned motifs:

- $\circ$  Cluster 3 includes Crisp superfamily and is a divergence from cluster 5. This cluster includes toxin/defense proteins with low disulfide propensities and a dominant  $\alpha$  secondary structure.
- Cluster 6 includes BPTI-like and Kringle-like superfamilies. This cluster is the least well-characterized and includes proteins with small disulfide propensities and different biological functions. The elements of this cluster share more diffuse properties as (i) they are constrained by three disulphide bonds with the same disulfide topology (1-6, 2-4 and 3-5) and (ii) they are associated with the regulation of similar biological processes (binding mediation, proteolytic activity, blood clotting, etc.).

We represent the Euclidean distances matrix for the cluster space in Table 9. From the analysis of this Table, we can verify that the mean-square distances between the clusters are significant larger than within the clusters. These results strongly indicate that the *HCA* partitioning is consistent.

[Insert Table 9]

#### Conclusions

In this work, we carried out an extensive statistical analysis of the conformational motifs for the disulfide bonds found in set of disulfide-rich proteins from twelve SCOP superfamilies.

The frequencies of the twenty conformational categories provided a near-spectral representation of the 12-dimension hyperspace under study. The general trends observed in this sample were quite consistent with the results obtained by other authors (Schmidt *et al.*, 2006; Schmidt and Hogg, 2007) for three different protein sets. We calculated the root mean-square deviations between our and the previously obtained frequencies. The three values obtained were all lower than 2.6 %.

The *HCA* partitioning of the data using a square Euclidean distances matrix resulted in a number of clusters, the majority of which aggregates superfamilies sharing both functional and structural patterns. The only exception is cluster 6, whose elements presented more diffuse connections. We therefore suggest the use of disulfide bonds conformational patterns as a criterion in *SDP* classification, as well as to recognize main divergences between *SDP* and other disulfide-rich superfamilies. However, the generalized application of this methodology for protein classification has to be subjected to further investigation.

#### Acknowledgements:

We thank the Fundação para a Ciência e a Tecnologia (FCT) for a doctoral scholarship granted to José Rui Ferreira Marques. Rute R. da Fonseca was funded by FCT (SFRH/BPD/26769/2006). We thank the Universidade do Porto for an electric wheelchair and a TrackerPro (a computer input device that takes the place of a mouse for people with no hand movement) granted to José Rui Ferreira Marques. Accepted manuscript

#### ACCEPTED MANUSCRIPT

#### References

- Andreeva, A., Howorth, D., Brenner, S. E., Hubbard, T. J. P., Chothia, C., Murzin, A. G., 2004. SCOP database in 2004: refinements integrate structure and sequence family data. Nucleic Acids Research 32, D226-D229.
- Andreeva, A., Howorth, D., Chandonia, J. M., Brenner, S. E., Hubbard, T. J. P., Chothia,
  C., Murzin, A. G., 2008. Data growth and its impact on the SCOP database: new developments. Nucleic Acids Research 36, D419-D425.
- Bhattacharyya, R., Pal, D., Chakrabarti, P., 2004. Disulfide bonds, their stereospecific environment and conservation in protein structures. Protein Engineering Design and Selection 17, 795-808.
- Brooks, D. J., Fresco, J. R., 2002. Increased frequency of cysteine, tyrosine, and phenylalanine residues since the last universal ancestor. Molecular & Cellular Proteomics 1, 125-131.
- Brooks, D. J., Fresco, J. R., Lesk, A. M., Singh, M., 2002. Evolution of amino acid frequencies in proteins over deep time: Inferred order of introduction of amino acids into the genetic code. Molecular Biology and Evolution 19, 1645-1655.
- Cheek, S., Krishna, S. S., Grishin, N. V., 2006. Structural classification of small, disulfiderich protein domains. Journal of Molecular Biology 359, 215-237.
- Chuang, C. C., Chen, C. Y., Yang, J. M., Lyu, P. C., Hwang, J. K., 2003. Relationship between protein structures and disulfide bonding patterns. Proteins-Structure Function and Genetics 53, 1-5.

Creighton, T. E., 1988. Disulfide bonds and protein stability. Bioessays 8, 57-63.

- Harrison, P. M. and Sternberg, M. J. E., 1996. The disulphide beta-cross: From cystine geometry and clustering to classification of small disulphide-rich protein folds. Journal of Molecular Biology 264, 603-623.
- Hogg, P. J., 2003. Disulfide bonds as switches for protein function. Trends in Biochemical Sciences 28, 210-214.
- Hutchinson, E. G. and Thornton, J. M., 1996. PROMOTIF A program to identify and analyze structural motifs in proteins. Protein Science 5, 212-220.
- Johnson, R.A., and Wichern, D.W., 2007. Applied Multivariate Statistical Analysis. Prentice Hall, New Jersey.
- Jordan, I. K., Kondrashov, F. A., Adzhubei, I. A., Wolf, Y. I., Koonin, E. V., Kondrashov, A. S., Sunyaev, S., 2005. A universal trend of amino acid gain and loss in protein evolution. Nature 435, 528-528.
- Katz, B. A. and Kossiakoff, A., 1986. The crystallographically determined structures of atypical strained disulfides engineered into subtilisin. Journal of Biological Chemistry 261, 5480-5485.
- Klink, T. A., Woycechowsky, K. J., Taylor, K. M., Raines, R. T., 2000. Contribution of disulfide bonds to the conformational stability and catalytic activity of ribonuclease A. European Journal of Biochemistry 267, 566-572.
- Murtagh, F., 1984. Counting dendrograms A survey. Discrete Applied Mathematics 7, 191-199.
- Murzin, A. G., Brenner, S. E., Hubbard, T., Chothia, C., 1995. SCOP A structural classification of proteins database for the investigation of sequences and structures. Journal of Molecular Biology 247, 536-540.

- Ozhogina, O. A. and Bominaar, E. L., 2009. Characterization of the kringle fold and identification of a ubiquitous new class of disulfide rotamers. Journal of Structural Biology 168, 223-233.
- Sardiu, M. E., Cheung, M. S., Yu, Y. K., 2007. Cysteine-cysteine contact preference leads to target-focusing in protein folding. Biophysical Journal 93, 938-951.
- Schmidt, B., Ho, L., Hogg, P. J., 2006. Allosteric disulfide bonds. Biochemistry 45, 7429-7433.
- Schmidt, B. and Hogg, P. J., 2007. Search for allosteric disulfide bonds in NMR structures. BMC Structural Biology 7, 49.
- Srinivasan, N., Sowdhamini, R., Ramakrishnan, C., Balaram, P., 1990. Conformations of disulfide bridges in proteins. International Journal of Peptide and Protein Research 36, 147-155.
- Thangudu, R. R., Sharma, P., Srinivasan, N., Offmann, B., 2007. Analycys: A database for conservation and conformation of disulphide bonds in homologous protein domains. Proteins-Structure Function and Bioinformatics 67, 255-261.
- Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, S.,
  Weiner, P., 1984. A new force-field for molecular mechanical simulation of nucleic-acids and proteins. Journal of the American Chemical Society 106, 765-784.
- Xie, D. X., Tropsha, A., Schlick, T., 2000. An efficient projection protocol for chemical databases: Singular value decomposition combined with truncated-Newton minimization. Journal of Chemical Information and Computer Sciences 40, 167-177.

Accepted manuscript

#### ACCEPTED MANUSCRIPT

#### **Figures captions**

**Figure 1:** Graphical representation of the five torsion angles used to classify the disulphide conformers.

Figure 2: Frequencies for the disulfide conformational categories.

**Figure 3:** Dendrogram for the hierarchical clustering analysis. The following notation was adopted: (1) Crisp, (2) Cystine-Knot, (3) Defensin-like, (4) EGF-Laminin, (5) Omega toxins, (6) Plant lectins, (7) Small snake toxins, (8) Scorpion-like toxins, (9) BBI, (10) BPTI-like, (11) Kringle-like and (12) Thioredoxin-like.

**Figure 4:** .Projected 3-D Cartesian representation of the square Euclidean distances matrix and clusters obtained by the hierarchical clustering analysis. The following notation was adopted: (1) Crisp, (2) Cystine-Knot, (3) Defensin-like, (4) EGF-Laminin, (5) Omega toxins, (6) Plant lectins, (7) Small snake toxins, (8) Scorpion-like toxins, (9) BBI, (10) BPTI-like, (11) Kringle-like and (12) Thioredoxin-like.

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

Disulphide category <sup>#</sup>	χ1	χ2	χ3	χ2'	χ1'	
-LHSpiral	-	-	-	-	-	
-RHHook	-	+	+	-	-	
+/-RHSpiral	+	+	+	+	-	
+/-LHSpiral	+	-	-	-	-	
-RHSpiral	-	+	+	+	-	
+/-RHHook	+	-	+	+	-	
+RHSpiral	+	+	+	+	+	
-LHHook	-	-	-	+	-	
-/+RHHook	-	-	+	+	+	
-RHStaple	-	-	+	-	-	
+/-LHHook	+	-	-	+	-	
-/+LHHook	-	-	-	+	+	
+/-LHStaple	+	+	-	+		
-LHStaple	-	+	-	+	7 -	
+LHSpiral	+	-	-		+	
+LHHook	+			+	+	
+RHHook	+_	+	+	-	+	
+/-RHStaple	+	-	+	-	-	
+LHStaple	+	+	-	+	+	
+RHStaple	+	-	+	-	+	

<sup>#</sup>LH: Left-handed oriented; RH: Right-handed oriented; -: Negative value for the respective torsion angle; +: Positive value for the respective torsion angle.

Accel

		χ2	χ3	χ2'			χ3			χ1	χ1'	
	Spiral	+	+	+	otifs	LH	_	ifs	+	+	+	
n motifs	Staple	+	- +	+	ional m			eral mot	-	-	-	
Mai	Hook	+	++	- +	Orientat	рц	+	Periph	+/-	+	-	
	HUUK	+	-	- +		КП	T		<b>_/</b> +	-	+	
C		*	Ç	0	~	3		Je	ç			

Table 2. Characteristic conformational motifs used for disulphide classification.



**Table 3.** Representative structures for the spiral conformational categories.



**Table 4.** Representative structures for the staple conformational categories.



**Table 5.** Representative structures for the hook conformational categories.

Table 6. Characterization of the protein set under study. The sample used in the statistical analyses is considered to include all the disulphide bonds identified in this protein set.

Cunarfamily	Dominant secondary	Drononcitv#	No. of PDB	No. of disulphide	Function
Supertaining	structure	T TUPELISH	structures	bonds	I UICHOI
Crisp	ಶ	5.3 %	9	54	Toxins/defense
Cystine-Knot	B	3.7 %	13	112	Metabolic
Defensin-like	В	7.4 %	15	47	Toxins/defense
EGF-Laminin	β	6.4 %	27	121	Metabolic
Omega toxins	β	8.9 %	28	88	Toxins/defense
Plant lectins	β	% 6.6	8	100	Metabolic
Small snake toxins	β	6.5 %	40	209	Toxins/defense
Scorpion-like toxins	β	7.9 %	70	247	Toxins/defense
BBI (Bowman Birk Inhibitors)	β	9.6 %	5	33	Protease inhibition
BPTI-like	α+β	5.1 %	12	42	Protease inhibition
Kringle-like	β	3.7 %	12	53	Metabolic
Thioredoxin-like	α/β	0.8 %	43	66	Isomerase catalysis
<sup>#</sup> Calculated by equation 2.				Q	

25

**Table 7.** Average parameters for the disulphide bonds conformational categories in the sample under study. Representative conformations for structural (-LHSpiral), catalytic (+/-RHHook) and allosteric (-RHStaple) disulphide bonds are represented in bold.

Conformational category	Frequency	DSE/kJ mol <sup>-1</sup>	$d(C_{\alpha}-C_{\alpha}')/\text{Å}$	d(C <sub>β</sub> -C <sub>β</sub> ')/Å
-LHSpiral	28.9%	11.5	5.7	3.7
-RHHook	9.9%	25.0	5.7	4.0
+/-RHSpiral	8.6%	14.5	5.9	3.8
+/-LHSpiral	7.9%	17.9	6.0	3.7
-RHSpiral	7.0%	18.9	6.0	3.8
+/-RHHook	6.1%	19.4	5.3	3.8
+RHSpiral	6.0%	12.8	5.8	3.7
LHHook	5.2%	37.0	5.7	4.1
-/+RHHook	4.7%	17.9	5.5	3.9
-RHStaple	4.0%	21.1	4.4	4.0
+/-LHHook	2.2%	26.8	5.9	4.0
+LHHook	1.9%	32.7	6.1	4.0
+/-LHStaple	1.6%	30.3	5.0	3.7
-LHStaple	1.5%	31.4	5.5	3.9
+LHSpiral	1.4%	20.8	6.2	3.9
+LHHook	1.2%	29.3	5.9	3.8
+RHHook	0.7%	30.7	6.1	4.1
+/-RHStaple	0.6%	32.3	5.9	4.1
+LHStaple	0.4%	39.3	5.4	3.3
+RHStaple	0.1%	24.9	5.9	3.3
Least strained <sup>#</sup>	63.1%	15.6	5.8	3.8
Most strained	36.9%	28.6	5.6	3.9

<sup>#</sup>The six conformational categories with the smallest DSE have a grey background.

Accel

Superfamily Categorie	1	2	3	4	5	6	7	8	6	10	11	12	Sample
-LHSpiral	35.2%	43.8%	26.4%	46.1%	13.8%	29.0%	27.3%	28.3%	12.1%	28.6%	32.7%	3.1%	28.9%
-RHHook	0.0%	5.4%	18.9%	8.7%	20.7%	24.0%	6.7%	9.3%	0.0%	4.8%	3.8%	10.8%	9.9%
+/-RHSpiral	7.4%	14.3%	7.5%	5.2%	8.0%	20.0%	3.8%	12.1%	3.0%	4.8%	3.8%	1.5%	8.6%
+/-LHSpiral	11.1%	4.5%	3.8%	0%9.6	2.3%	22.0%	4.3%	2.8%	0.0%	19.0%	30.8%	6.2%	7.9%
-RHSpiral	14.8%	9.8%	3.8%	2.6%	3.4%	0.0%	17.2%	3.2%	21.2%	4.8%	1.9%	1.5%	7.0%
+/-RHHook	0.0%	3.6%	3.8%	3.5%	2.3%	0.0%	2.9%	7.7%	0.0%	2.4%	0.0%	50.8%	6.1%
+RHSpiral	0.0%	3.6%	1.9%	3.5%	4.6%	4.0%	12.9%	9.7%	0.0%	0.0%	1.9%	1.5%	6.0%
-LHHook	0.0%	1.8%	7.5%	6.1%	5.7%	1.0%	5.3%	10.1%	6.1%	2.4%	3.8%	1.5%	5.2%
-/+RHHook	5.6%	0.0%	0.0%	1.7%	11.5%	0.0%	2.9%	4.9%	12.1%	23.8%	15.4%	0.0%	4.7%
-RHStaple	24.1%	0.0%	5.7%	1.7%	4.6%	0.0%	2.4%	0.4%	24.2%	0.0%	0.0%	16.9%	4.0%
+/-LHHook	0.0%	0.0%	9.4%	1.7%	5.7%	0.0%	2.9%	2.0%	0.0%	0.0%	1.9%	3.1%	2.2%
-/+LHHook	0.0%	1.8%	5.7%	0.0%	3.4%	0.0%	3.8%	2.4%	0.0%	0.0%	0.0%	0.0%	1.9%
+/-LHStaple	1.9%	0.9%	0.0%	0.0%	10.3%	0.0%	1.4%	0.0%	6.1%	0.0%	1.9%	3.1%	1.6%
-LHStaple	0.0%	0.9%	3.8%	0.9%	2.3%	0.0%	1.0%	1.6%	6.1%	9.5%	0.0%	0.0%	1.5%
+LHSpiral	0.0%	8.9%	0.0%	0.0%	1.1%	0.0%	1.0%	1.2%	0.0%	0.0%	0.0%	0.0%	1.4%
+LHHook	0.0%	0.0%	1.9%	2.6%	0.0%	0.0%	2.4%	1.2%	6.1%	0.0%	0.0%	0.0%	1.2%
+RHHook	0.0%	0.9%	0.0%	1.7%	0.0%	0.0%	0.5%	1.2%	0.0%	0.0%	1.9%	0.0%	0.7%
+/-RHStaple	0.0%	0.0%	0.0%	0.9%	0.0%	0.0%	1.4%	0.8%	3.0%	0.0%	0.0%	0.0%	0.6%
+LHStaple	0.0%	0.0%	0.0%	2.6%	0.0%	0.0%	0.0%	0.8%	0.0%	0.0%	0.0%	0.0%	0.4%
+RHStaple	0.0%	0.0%	0.0%	0.9%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%
									R				
										K			

categories.
al
rmation:
9
conf
different
the
for
ncies
. freque
S
able

27

Cluster	1	2	3	4	5	6
1	0.00%	34.89%	40.29%	45.79%	32.03%	44.03%
2	34.89%	0.00%	8.77%	25.38%	13.42%	21.20%
3	40.29%	8.77%	0.00%	12.00%	11.24%	12.72%
4	45.79%	25.38%	12.00%	6.18%	19.99%	24.61%
5	32.03%	13.42%	11.24%	19.99%	3.98%	19.42%
6	44.03%	21.20%	12.72%	24.61%	19.42%	1.83%

**Table 9.** Square Euclidian distances matrix for the cluster space.



Figure 1



Disulphide conformational categories

Figure 2

Frequency







4. Figure



