OPEN ACCESS

2nd International Electronic Conference on Entropy and Its Applications
15-30 November 2015

www.sciforum.net/conference/ecea-2

Conference Proceedings Paper – Entropy

Role of disulfide bonds in stabilizing the conformation of selected enzymes - approach based on divergence entropy applied to the structure of hydrophobic core in proteins

Mateusz Banach [1,2], Barbara Kalinowska [1,2], Leszek Konieczny [3] and Irena Roterman [1,*]

[1] Department of bioinformatics and Telemedicine, Jagiellonian University – Medical College, Sw. Anny 12, 31-008 Krakow, Poland; Emails: Mateusz.banach@uj.edu.pl; Barbara.kalinowska@uj.edu.pl

[2] Faculty of Physics, Astronomy and Applied Computer Science – Jagiellonian University, Lojasiewicza 11, 30-348 Krakow, Poland

[3] Chair of Medical biochemistry, Jagiellonian University – Medical college, 31-034 Krakow, Poland; Email: mbkoniec@cyf-kr.edu.pl

* Author to whom correspondence should be addressed; E-Mail: myroterm@cyf-kr.edu.pl; Tel.: +48-12-619 96 93; Fax: +48-12-619 96 93.

Published: 13 November 2015

Abstract: One of the factors responsible for tertiary structural stabilization in proteins is the presence of the hydrophobic core – a result of hydrophobic interactions within the protein body. In some proteins (especially extracellular ones) additional stabilization is provided by covalent bonds between selected Cys residues, commonly referred to as disulfide bonds. The mutual interplay of both factors and their respective contributions to stabilization are the focus of this work. We perform an assessment of the effects of disulfide bonds by applying the Fuzzy Oil Drop (FOD) model in which individual polypeptide chain fragments (including fragments which participate in SS bonds) can be evaluated in the context of their influence upon tertiary structural stabilization by comparing their corresponding theoretical and idealized hydrophobicity density distributions. We have identified proteins where both factors reinforce each other, as well as proteins where they seem to counteract each other. We have performed our analysis for a number of enzymes, including ribonuclease, lysozyme, disulfide isomerase and phospholipase.

Keywords: Kullback-Leibler entropy; protein structure; hydrophobicity; disulfide bonds; tertiary structure stabilization; enzymes

1. Introduction

The folding process – the process by which a protein adopts a conformation which supports biological activity – is primarily driven by optimization of nonbinding interactions. In some proteins (particularly extracellular ones) covalent bonds between Cys residues, *i.e.* disulfide bonds, must also be taken into account. The main purpose of such bonds is structural stabilization. Hydrophobic interactions, resulting in the emergence of the hydrophobic core, also tend to exert a stabilizing influence upon tertiary protein structure.

The stabilizing role of disulfide bonds is well known [1,2]. Formation of structures which include disulfide bonds is significantly more complicated than in the case of polypeptide chains where such bonds are absent. From a chemical point of view, the former case calls for additional redox reactions. Not all Cys residues in the chain participate in SS bonds. A detailed study of the generation of disulfide bonds in BPTI can be found in [3]. The authors show that in BPTI folding does not proceed by way of simple sequential formation of the native SS-bonds system. Rather, the process follows a more complicated path which includes selection of a particular S-S pair from among many possible alternatives. On the one hand, the native structure may reinforce the correct set of S-S bonds, while on the other hand it may act as an inhibitor of the desired Cys/Cys combination by directing Cys residues towards the central part of the protein body, as observed in BPTI. For this reason, BPTI is regarded as a good study subject in the analysis of disulfide bond formation process. Reduction of SS bonds in BPTI causes immediate unfolding of the entire molecule, which implies that its tertiary conformation is thermodynamically linked.

Another interesting protein (in the context of disulfide bonds) is disulfide isomerase which catalyzes oxidative protein folding *in vivo*. In order to fulfill its biological role it requires access to buried thiol residues which it rearranges in stable folding intermediates (necessitating prior unfolding of these intermediates). [4].

The two domains of microcollagen-1, despite sharing an identical cystein pattern, form differing SS bond systems and therefore adopt different conformations. The N-terminal domain, while sequentially identical to the C-terminal domain, folds in a different fashion [5].

Despite unequivocal evidence of the stabilizing role of SS bonds their actual role seems varied. A fitting example is provided by a pair of proteins from the toxin subfamily, both of which exhibit similar folds and contain four SS bridges. Reduction of the fourth bridge in one protein molecule affects its 3D, structure altering the status of two stranded beta sheets from twisted to non-twisted. Reduction of the corresponding bridge in the other molecule does not affect its 3D structure despite high structural similarity [6].

Some researchers have reported strong correlation between the removal of disulfide bonds and structural stability. In the gene-3-protein of the filamentous phage fd, despite of loss of all three disulfide bonds, the midpoints of the thermal transitions were increased from 48.5 C deg to 67.0 C deg

in domain N2, and from 60.0 C deg to 78.7 C deg in domain N1. The major loss in conformational stability caused by the removal of the disulfides was thus over-compensated by strongly improved non-covalent interactions. The stabilized variants were less infectious than the wild-type protein, probably because the domain mobility was reduced [7].

Disulfides also appear to play an important role in amyloidogenesis. It has been reported that human proinsulin is markedly less susceptible to fibrillation than human insulin despite their similar thermodynamic stabilities [8]. The specific role of SS bonds in immunoglobulin domains involves joining the midpoints of antiparallel beta sheets [9]. The effects of mutations which reduce one of four SS bonds in the human lysozyme, as well as point mutations which disrupt the distribution of hydrophobic residues, are reported in [10].

This work discusses the status of disulfide bonds in the context of hydrophobic interactions. If the idealized hydrophobic core is modeled by a 3D Gaussian, we can assess to what degree actual proteins conform to theoretical predictions. Such quantitative assessment bases on Kullback-Leibler's divergence entropy criterion which expresses the relative distance between theoretical and empirical distributions and therefore reflects the structural ordering of the hydrophobic core, which is regarded as a stabilizing factor [11,12]. The divergence entropy formula can also be successfully applied to individual fragments of the polypeptide chain, including fragments bounded by Cys residues which participate in SS bonds [13,14]. It turns out that such fragments may either (1) reinforce the hydrophobic core structure by conforming to theoretical predictions regarding hydrophobicity density distribution, or (2) counteract stabilization of the chain by diverging from the FOD model. We may thus speculate that in the former case reduction of disulfide bonds should not significantly affect structural stability, while in the latter case reduction may result in major structural rearrangement as the stabilizing factor (i.e. the SS bond) is removed. An open issue concerns the relation between nonbinding interactions (whose optimization is expected to guide the folding process towards the native conformation) and processes which result in the creation of a hydrophobic core. In many cases such processes reinforce each other, producing a structure which is stabilized both by nonbinding interactions and by hydrophobic effects. In some cases, however, the opposite is true – no stable hydrophobic core emerges and fragments linked by SS bonds do not conform to the idealized hydrophobicity density distribution model.

Analysis of arbitrarily selected enzymes in which the status of fragments bounded by SS-forming Cys residues varies is a step towards determining the relations between both factors in the context of tertiary structural stabilization. In the presented model disulfide bonds are regarded as additional constraints, reducing the degree of structural freedom and therefore enforcing a specific conformation. Covalent bonds cause the protein structure to become more rigid and less adaptable to external stimuli, including the presence of water. On the other hand, the hydrophobic core is a natural response to immersion in an aqueous environment. Under ideal conditions all hydrophobic residues should be encapsulated deep within the protein body while hydrophilic residues should be exposed on its surface. Departures from this principle are often linked to biological activity (ligand/substrate binding or protein/membrane complexation) while additional SS bonds stabilize structures in which hydrophobicity density distribution differs from theoretical expectations – e.g. when a strongly hydrophobic loop is found on the surface or a binding cavity is present.

1.1. The fuzzy oil drop model as an expression of the hydrophobic stability of proteins

The so-called fuzzy oil drop (FOD) model is a modification of Kauzmann's oil drop paradigm [15]. The model likens the folding of a polypeptide chain to the behavior of a drop of oil in an aqueous environment where the contact surface between the hydrophobic substance (oil) and the polar environment (water) is minimized. In proteins hydrophobic residues are shielded from contact with water by migrating to the center of the protein body while hydrophilic residues are instead exposed on the surface. The fuzzy oil drop model introduces a quantitative description of this process, representing the resulting hydrophobicity density distribution with a 3D Gaussian. Values of this function peak at the geometric center of the molecule and then decrease along with distance from the center, reaching almost 0 on the surface. The distance between the center of the molecule and its surface is expressed using the three-sigma rule in each principal direction, yielding three coefficients: σ_x, σ_y and σ_z. The molecule can thus be encapsulated in an ellipsoid, enabling us to compute theoretical hydrophobicity density at any point within this virtual "capsule". Of course, actual (emprical) distribution of hydrophobicity density differs from theoretical expectations since it depends on the placement of each residue in the protein body as well as on its intrinsic hydrophobicity. Residues are assumed to interact with one another if their separation is below 9A (the assumed cutoff distance for hydrophobic interactions). Both theoretical and observed hydrophobicity density values are computed for the so-called effective atoms (averaged-out positions of all atoms belonging to the given residue).

The above procedure produces a list where each residue (represented by its effective atoms) is described by the following parameters: (1) intrinsic hydrophobicity (conforming to a predetermined scale); (2) expected hydrophobicity as represented by the idealized distribution; (3) observed hydrophobicity which depends on local interactions; (4) boundary hydrophobicity – another theoretical quantity calculated under the assumption that no hydrophobicity concentration exists at any point in the molecule.

The theoretical hydrophobicity distribution constitutes a limit case where the molecule is assumed to conform to the theoretical model with perfect accuracy, exhibiting a well-defined hydrophobic core along with a hydrophilic sheath which shields the core from contact with water. The other limit case corresponds to the „flat" hydrophobicity distribution profile, with equal values throughout the entire protein body. In mathematical terms these distributions can be expressed as follows.

1.1.1. Theoretical distribution:

$$\tilde{Ht}_j = \frac{1}{\tilde{Ht}_{sum}} \exp\left(\frac{-\left(x_j - \bar{x}\right)^2}{2\sigma_x^2}\right) \exp\left(\frac{-\left(y_j - \bar{y}\right)^2}{2\sigma_y^2}\right) \exp\left(\frac{-\left(z_j - \bar{z}\right)^2}{2\sigma_z^2}\right),$$

$\tilde{Ht}_j$ is the theoretical hydrophobicity density (hence the *t* designation) at the *j*-th point in the protein body. $\bar{x}, \bar{y}, \bar{z}$ correspond to the peak of the Gaussian in each of the three principal directions, while $\sigma_x, \sigma_y, \sigma_z$ denote the range of arguments for each coordinate system axis. These coefficients are selected in such a way that 99% of the Gaussian's integral is confined to a range of $\bar{x} \pm 3\sigma$. Accordingly, values of the distribution can be assumed to equal 0 beyond this range.

The above distribution is discretized using positions of effective atoms.

1.1.2. Observed distribution

Observed distribution (as proposed by M. Levitt [16]:

$$\widetilde{Ho}_j = \frac{1}{\widetilde{Ho}_{sum}} \sum_{i=1}^{N} (H_i^r + H_j^r) \begin{cases} \left[1 - \frac{1}{2} \left(7 \left(\frac{r_{ij}}{c} \right)^2 - 9 \left(\frac{r_{ij}}{c} \right)^4 + 5 \left(\frac{r_{ij}}{c} \right)^6 - \left(\frac{r_{ij}}{c} \right)^8 \right) \right] \text{ for } r_{ij} \le c \\ 0 \text{ for } r_{ij} > c \end{cases}$$

N is the number of amino acids in the protein, $\widetilde{H}_i^r$ expresses the hydrophobicity parameter of the *i*-th residue while r_{ij} expresses the distance between two interacting residues (*j*-th effective atom and *i*-th effective atom). *c* expresses the cutoff distance for hydrophobic interactions, which is taken as 9.0Å (following [16]). Observed hydrophobicity density values $\widetilde{H}o_j$ are also computes for effective atoms (geometric center of each side chain).

The $\widetilde{H}o_{sum}$ coefficient, representing the aggregate sum of all components, is needed to normalize the distribution and enable meaningful comparisons between the observed and theoretical hydrophobicity density distributions.

1.1.3. Unified distribution:

$$\widetilde{H}r_j = \frac{1}{N}$$

N is the number of amino acids in the chain.

Figure 1 provides a graphical depiction of the model. Imagine three microscopic hydrophobicity density detectors traversing the protein molecule along a predetermined axis. The first detector measures theoretical density (readings shown in dark blue); the second measures actual density which depends on the intrinsic hydrophobicity and placement of each residue (readings shown in red) while the final detector outputs a static reading determined solely by the number of residues in the chain (readings shown in green). While all distributions shown in the figure (T, O and R) are continuous, our analysis focuses on their discrete values computed for each effective atom.

1.1.4. Kullback-Leibler entropy to measure the differences between distributions

Another important issue concerns the degree of similarity between the empirical and idealized distribution (perfect hydrophobic core) as well as between the empirical and static distribution (complete lack of a hydrophobic core). Quantitative assessment of these parameters bases on Kullback-Leibler's divergence entropy formula [17]:

$$D_{KL}(p|p^0) = \sum_{i=1}^{N} p_i \log_2(p_i / p_i^0)$$

The value of D_{KL} expresses the distance between two distributions: target distribution (p_0) and analyzed distribution (p). In the fuzzy oil drop model the target distribution (T) is given by the 3D Gaussian while the observed distribution is denoted as O.

For the sake of simplicity we introduce the following notation:

$$O/T = \sum_{i=1}^{N} O_i \log_2 O_i / T_i$$

D_{KL} (O/T) expresses the "distance" between both distributions. The more divergent the distributions, the greater the value of D_{KL}. This value cannot, however, be interpreted directly since it depends on the number of points (chain length). Moreover, D_{KL} is a measure of entropy and must be compared to a reference value. In order to facilitate meaningful comparisons we introduce another boundary distribution, opposite to the idealized one – the so-called unified distribution (denoted R) which corresponds to a situation where each effective atom possesses the same hydrophobicity density (R_i = $1/N$ for each i, where N is the number of residues in the chain). The distance between the observed distribution and the unified distribution is therefore given as:

$$O/R = \sum_{i=1}^{N} O_i \log_2 O_i / R_i$$

Comparing O/T and O/R tells us whether the given protein more closely approximates the theoretical (O/T) or unified (O/R) distribution. Proteins for which O/T > O/R are regarded as lacking a well-defined hydrophobic core. To further simplify matters we introduce the following relative distance criterion:

$$RD = \frac{O/T}{O/T + O/R}$$

Here, RD < 0.5 indicates the presence of a hydrophobic core.

1.1.5. Summary of the model

It should be noted that RD may also be calculated for a selected fragment of the polypeptide chain – for example a fragment which corresponds to a known secondary fold (helix, beta strand, loop *etc.*) or a disordered fragment [18]. This, however, requires normalization (rescaling) of H_t, H_o and H_r (hydrophobicity parameter according to R distribution) so that the sum of all values assigned to a given section is equal to 1. In this work, in addition to assessment of known secondary structural folds we have also analyzed fragments bounded by disulfide bond attachment points. Our aim was to determine the degree to which a given structure is stabilized by SS bonds, as well as the involvement of a well ordered hydrophobic core in structural stabilization.

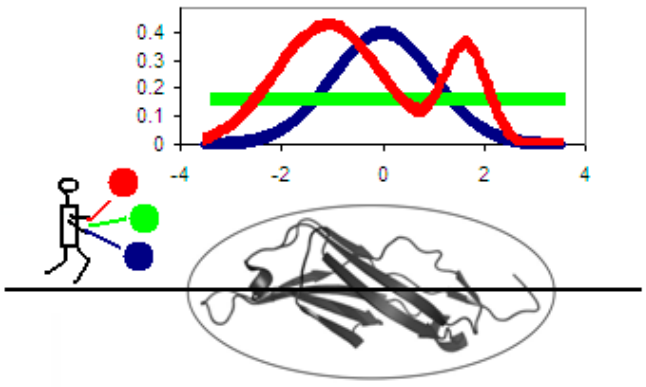


Figure 1. Schematic representation of hydrophobicity density distribution in a sample protein. The presented profiles correspond to theoretical (dark blue), observed (red) and

unified (green) distributions respectively. The "observer" shown below carries three detectors, each of which registers a single distribution. The protein body is bounded by an ellipsoid whose dimensions stretch by +-3σ in each principal dimension (for simplicity's sake the presentation is limited to a single coordinate system axis – X).

It should be noted that the value of RD (1D representation) calculated for the red distribution is only 0.215, which means that – at least in the presented case – the observed hydrophobicity profile (restricted to a single axis) approximates the theoretical distribution with high accuracy and that therefore the molecule contains a well-defined hydrophobic core along with an encapsulating hydrophilic sheath. The presented RD value means that the observed distribution lies much "closer" to the theoretical distribution (blue line) than to the unified distribution (green line).

A detailed description of the fuzzy oil drop model can be found in [19] – here we limit ourselves to a brief recapitulation of the model's core concepts.

This work assesses the status of polypeptide chain fragments stabilized by disulfide bonds in the context of their conformance to the fuzzy oil drop model. We base our research on a set of enzymes, focusing on catalytic residues as well as residues involved in binding ligands or mediating protein complexation.

2. Results and Discussion

Summary results presenting the status of sections bounded by Cys residues which form SS bonds are presented in Table 1. The table lists the status of the entire protein molecule (or domain, where appropriate) in relation to the status of its Cys-bounded fragments.

As can be seen, the relationship between the status of the entire molecule and its SS-bounded fragments varies. Below we discuss examples of proteins from each category.

Table 1. Status of proteins with regard to their hydrophobic core characteristics. Columns indicate the status of the whole molecule while rows correspond to individual fragments defined by Cys residues which form SS bonds.

		FUZZY OIL DROP ACCORDANCE WHOLE PROTEIN	
		YES	NO
STATUS OF FRAGMENTS DEFINED BY SS-BONDS	ACCORDANT – ALL	1QLL 1M6B-D3	1M6B-D1
	SOME ACCORDANT	1M6B-D2, 1M6B-D4, 1LZ1,1ANG	
	DISCORDANT - ALL	1MEK	5RSA

2.1. Hydrophobic core supported by a system of disulfide bonds

An example of a protein which conforms to the idealized hydrophobicity density distribution model is provided by a single chain of the 1QLL homodimer – phospholipase (a neurotoxin; source: *Bothrops pirajai*) [18].

Local participation of catalytic residues in the hydrophobic core may be determined by calculating RD values for a chain from which such residues have been excised. In the case of 1QLL this operation reduces the value of RD (NoE in Table 2.) which means that catalytic residues generally diverge from the idealized distribution – as seen in Figure. 1. This property is shared by catalytic residues in most proteins.

Excision of residues involved in binding ligands (NoL in Table 2.) and mediating inter-chain interactions (NoP-P in Table 2.) does not appreciably affect RD. This suggests that ligand binding and protein complexation do not require major rearrangements within the protein's hydrophobic core. 1QLL's ligand (tridecanoid acid) is bound on the surface of the protein and does require a specific cavity, while residues involved in P-P complexation are mostly hydrophilic – indicating that dimerization of 1QLL is driven primarily by electrostatic forces with scant involvement of hydrophobic effects. The inter-chain interface is comprised mostly of hydrophobic residues and does not produce a common hydrophobic core for the complex as a whole.

Figure 2. Hydrophobicity density distribution profiles: expected (T) and observed (O) for 1QLL. Cys residues which form SS bonds have been tagged (yellow triangles) as well as catalytic residues (green triangles).

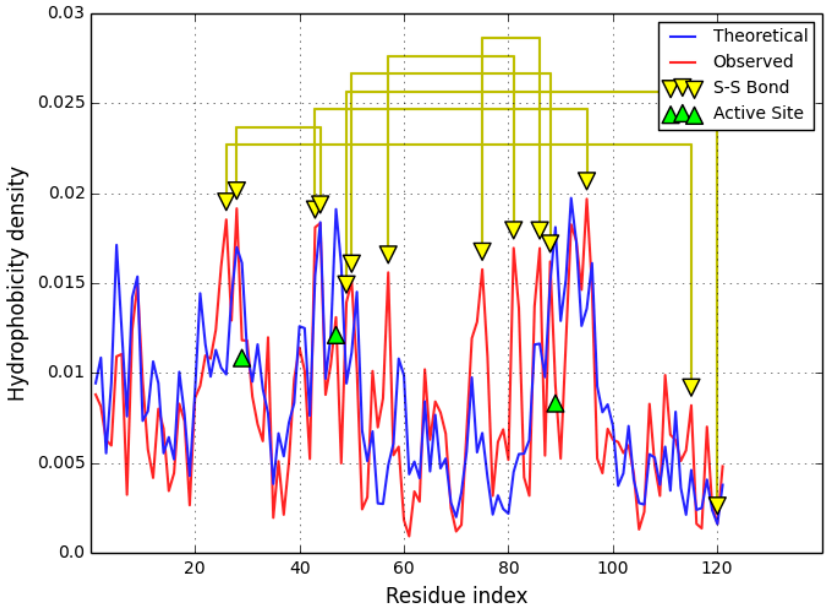
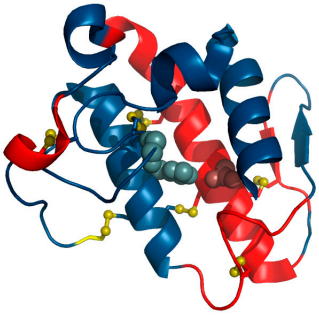


Table. 2. RD values characterizing the status of 1QLL, its secondary structural folds (with given number of Cys residues participating in disulfide bonds). The three rightmost columns list the positions of residues involved in enzymatic activity (E-Resid.), protein-protein interaction (P-P) and ligand binding (Ligand) respectively. RD values in excess of 0.5 (indicating divergence from the idealized distribution) are listed in boldface.

1QLL	Fragment	RD	E-Resid.	P-P	Ligand
COMPLEX		0.641			
CHAIN A/B		0.400/0.397			
No-E		0.393			
No-L		0.422			

1QLL	Fragment	RD	E-Resid.	P-P	Ligand
No-P-P		0.406			
SECONDARY					
HELIX	1-14	0.376		11Q,13T, 14G	2L,6G,7K,9I
HELIX	16-22	0.155			17P,21Y
LOOP Cys-2	23-37	0.317	29G		28C,29G
HELIX Cys-4	38-53	0.349	47H		44C,47H
LOOP Cys-1	**54-64**	**0.606**			
BETA	65-69	0.217			
BETA Cys-1	**72-76**	**0.619**			
HELIX Cys-4	**79-99**	**0.508**	89D	97R	92V
HELIX	**100-103**	**0.970**			
HELIX	104-108	0.247			
HELIX Cys-1	**111-115**	**0.864**			
LOOP Cys-1	116-121	0.260			
BETA-SHEET		0.394			
SS-BONDS	26-115	0.426	29,89	68,70,71,97	28,29,44,47,92
	28-44	0.198	29		28,29,44
	43-95	0.468	47,89	68,70,71,	44,47,92
	49-121	0.416	89	68,70,71,97	92
	50-88	0.483		68,70,71	
	57-81	0.500		68,70,71	
	75-86	0.444			

Figure 3. Structure of 1QLL with divergent fragments marked in red. Yellow sections correspond to Cys residues which form disulfide bonds, and dark blue balls represent the catalytic residues.



The status of individual secondary folds in 1QLL varies, despite the fact that the protein as a whole conforms to the fuzzy oil drop model (Figure 3). Divergent fragments include the loop at 54-64, the beta-helix-helix system at 72-103 and the helix at 100-103. Of particular note is the 72-103 fragment which includes a catalytic residue (although the beta sheet taken as a whole appears to match the idealized hydrophobicity density distribution profile).

The main focus of our analysis are fragments defined by the placement of Cys residues which form disulfide bonds. As it turns out, all such fragments conform to the fuzzy oil drop model. Cys residues are generally expected to migrate towards the center of the molecule as the hydrophobic core emerges – their mutual interactions may be a consequence of this phenomenon. Nevertheless, RD values remain quite high for the 72-103 fragment which includes a catalytic residue in addition to one residue responsible for ligand binding and protein interaction. We can speculate that catalysis of the substrate requires local deformations in the hydrophobicity field, which, in turn, introduce local instabilities facilitating conformational changes during catalysis. Another interesting section, 57-80, is stiffened by the presence of four SS bonds – the RD value for this section is 0.500, suggesting that hydrophobic interactions alone are not sufficient to ensure its structural stability.

We may speculate that, at least in the case of 1QLL with highly organized hydrophobic core, reduction of a disulfide bond may not result in significant conformational changes, although the local deviations from the hydrophobicity density distribution model are present.

2.2. Discordant core structure with some sections defined by SS bonds also discordant

The human lysozyme – O-glycosyl hydrolase (EC 3.2.1.17) [21] – does not follow the idealized hydrophobicity density distribution model (RD>0.530). Elimination of catalytic residues brings the observed distribution closer to theoretical values, again proving that catalytic residues generally tend to diverge from the model. Among the secondary folds present in 1LZ1 the following are found to diverge from the model: the helix at 24-37, the beta-loop-beta system at 42-55, the beta fold at 59-61 and the helix at 104-109. The beta-sheet as a whole is also discordant while all other fragments exhibit good accordance (RD<0.5), stabilizing the polypeptide chain in its native form. The 3D presentation of the status of secondary structural fragments is visualized in Figure 4.

In the case of 1ZL1 all catalytic residues are surrounded by locally discordant neighbor (Table 3). The status of fragments bounded by Cys residues which form SS bonds varies: major parts of the chain (6-128 and 30-116) remain discordant while shorter loops (65-81 and 77-95) match theoretical predictions. This suggests that the two latter bonds form spontaneously between residues brought into proximity in the process of creating a common hydrophobic core. Note that the outer SS bonds affect the placement of residues which end up participating in the inner bonds. The section at 30-65 diverges from the model.

Interpretation of the above phenomenon on the grounds of the fuzzy oil drop model suggests that poor local stabilization of the beta fragment, which includes a catalytic residue, as well as the helical fragment containing another catalytic residue, creates favorable conditions for binding and stabilizing substrates, and for the catalysis process itself. This is evidenced by the fact that the fragment bounded by Cys 30 and Cys 65 (including both catalytic residues) remains discordant from the model. Thus, both SS bonds are suspected of “enforcing” local structural instability, counteracting hydrophobic effects.

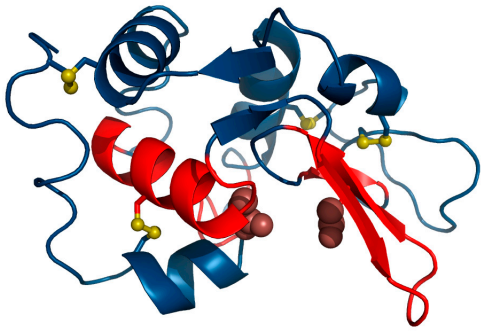
Table 3. RD values calculated for the lysozyme, its secondary folds and sections limited by Cys residues which form SS bonds. RD values in excess of 0.5 (indicating divergence from the idealized distribution) are listed in boldface.

1LZ1	FRAGMENT	RD	E-Resid.
Chain		0.530	
No E			
SECONDARY			
HELIX	4-15	0.377	
LOOP	16-23	0.234	
HELIX 1E	**24-37**	**0.566**	35E
LOOP	38-41	0.160	
BETA	**42-46**	**0.620**	
LOOP	**47-50**	**0.653**	
BETA 1E	**51-55**	**0.603**	53D
LOOP	56-58	0.473	
BETA	**59-61**	**0.858**	
LOOP	62-80	0.484	
HELIX	81-86	0.444	
HELIX	89-101	0.263	
HELIX	**104-109**	**0.502**	
HELIX	110-116	0.317	
LOOP	117-120	0.164	
HELIX	121-125	0.492	
Beta-sheet		**0.617**	
SS-BONDS	**6-128**	**0.531**	35E, 53D
	30-116	**0.543**	
	65-81	0.279	
	77-95	0.489	
Between SS-BONDS	6-30	0.415	
	30-65	**0.575**	35,53
	77-81	0.198	
	81-95	0.376	

In [23] the authors discuss the stabilizing role of selected helical fragments whose compact core can be observed even in the molten globule state, as predicted by the fuzzy oil drop model for helical fragments, with only the helix at 24-37 diverging from expectations. Other reports suggest that the lysozyme folding process occurs in stages. In particular, the alpha-helical domain is believed to fold faster than the beta-sheet domain [24]. It was concluded in [24] that folding does not become organized in a single cooperative event but that different parts of the structure become stabilized with very different kinetics. Alfa-helical domain folds faster than the Beta-sheet domain.

The observed status of beta fragments remains in good correspondence with the empirically determined role of each fragment of the human lysozyme [25,26].

Figure 4. Structure of the lysozyme. Fragments which diverge from the expected hydrophobicity density distribution are marked in red. Yellow sections indicate Cys residues which form disulfide bonds. The brown balls represent the catalytic residues.



2.3. Accordant core structure with all sections defined by SS bonds discordant

An example of this category is provided by human isomerase disulfide (EC 5.3.4.1) which contains a single, peculiar disulfide bond [20]. The bond spans a very short section – just 4 residues, all of which are enzymatically active. The molecule as a whole follows the hydrophobicity density distribution model with good accuracy (Table 4.). It comprises a beta sheet whose RD value is 0.314, even though one of its constituent parts remains discordant (RD = 0.503). Significant discordance is also observed for loops which include enzymatic residues. Elimination of catalytic residues from RD calculations brings the remainder of the molecule closer to the idealized distribution.

Analysis of secondary structural folds reveals that discordant sections of the molecule include most of its loops, with all of them exhibiting higher-than-expected hydrophobicity density (cf. Figure 5.). These loops are all exposed on the surface of the molecule and can therefore be suspected of interacting with other hydrophobic molecules (or proteins). All beta folds and helixes remain highly accordant with the model and, together, form a compact hydrophobic core, with only the beta fragment at 9-12 slightly exceeding our classification threshold.

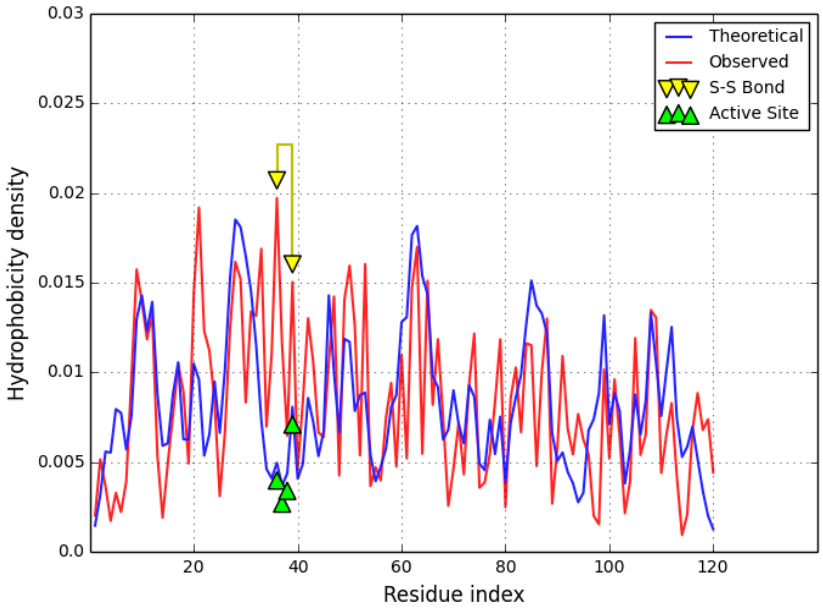
Table 4. RD values for disulfide isomerase. Also listed are RD values for fragments of the protein devoid of catalytic residues and the fragment at 86-94, which exhibits particularly strong discordance. RD values in excess of 0.5 (indicating divergence from the idealized distribution) are listed in boldface.

1MEK	FRAGMENT	RD	E-Res.	Cys
CHAIN		0.464		
NoE		0.433		
SECONDARY				
LOOP	**1-3**	**0.579**		
BETA	4-6	0.279		
BETA	**9-12**	**0.503**		
HELIX	16-24	0.497		

1MEK	FRAGMENT	RD	E-Res.	Cys
BETA	25-32	0.257		
LOOP	33-41	0.414	36C,37G, 38H,39C	36C, 39C
HELIX	42-52	0.392		
LOOP	53-61	0.464		
BETA	62-67	0.479		
LOOP	**68-72**	**0.555**		
HELIX	73-78	0.300		
LOOP	79-83	0.351		
BETA	84-90	0.380		
LOOP	**91-96**	**0.684**		
BETA	97-99	0.344		
LOOP	100-103	0.171		
HELIX	104-114	0.312		
LOOP	**115-120**	**0.747**		
BETA-SHEET	ALL BETA FRAGMENTS	0.314		
HELICES	ALL HELICES	0.397		
SS-BONDS	**36-39**	**0.510**	**37G,38H**	

Figure 5. shows a section bounded by two Cys residues which form an SS bond (36-39). This section includes catalytic residues whose hydrophobicity density diverges from idealized values.

Figure 5. Hydrophobicity density distribution in human disulfide isomerase (1MEK) – theoretical (T – dark blue) and observed (O – red). The Cys residues are distinguished by yellow triangles, catalytic residues by green triangles.



The loops at 91-96 and 115-120 are both exposed on the surface where no strong concentration of hydrophobicity is expected. This discrepancy, as seen in Figure 5, may indicate potential ligand binding site or protein complexation sites.

2.4. Multi-domain enzyme with variable number of disulfide bonds in each domain

Human serine protease (EC 2.7.10.1) [19] has been selected as an example of a multi-domain human protein. Selection of this protein was also due to its pharmacological meaning [27]. Constituent domains of this protein differ with respect to the number of SS bonds as well as their respective hydrophobicity density distribution profiles. Three of four domains in 1M6B diverge from the theoretical model and contain a large number of SS bonds (four in D1, seven in D2 and seven in D4). Domain 3 is the only one which remains accordant model – notably, it contains only two SS bonds.

According to CATH 3.80.20.20, domain D1 represents the Alpha-Beta Horseshoe architecture. Its disulfide bonds link loose fragments of the chain which do not participate in the ordered horseshoe conformation. It seems that SS bonds stabilize fragments not otherwise stabilized by nonbinding interactions (Table 5.). As such, these bonds are not directly related to the structure of the domain's hydrophobic core.

Table 5. RD values for domain D1 of 1M6B, its individual secondary folds and all fragments bounded by Cys residues which form SS bonds. RD values in excess of 0.5 (indicating divergence from the idealized distribution) are listed in boldface.

1M6B-D1	FRAGMENT	RD	SS
Domain	**8-198**	**0.610**	
Secondary			
BETA-I	8-10	0.329	10C
LOOP	11-37	0.160	37C
BETA-I	38-41	0.431	
BETA-II	**44-48**	**0.660**	
LOOP	49-54	0.346	
HELIX	55-61	0.438	
BETA II	**67-72**	**0.893**	
LOOP	73-75	0.207	
BETA-III	76-80	0.430	
LOOP	81-83	0.263	
BETA-I	84-87	0.455	
LOOP	88-94	0.177	
BETA-II	**95-102**	**0.613**	
LOOP	103-113	0.401	
BETA-III	114-116	0.401	
LOOP	117-121	0.335	
BETA-I	**122-124**	**0.547**	
BETA-II	127-132	0.237	
LOOP	133-142	0.286	137C
HELIX	143-148	0.338	
LOOP	**149-153**	**0.543**	

1M6B-D1	FRAGMENT	RD	SS
BETA-II	154-158	0.305	
LOOP	**159-198**	**0.621**	164C,167C,171C,175C,183C, 191C
BETA-SHEET	**I**	**0.755**	10C
	II	0.421	
	III	**0.539**	
HELICES		**0.558**	
SS-BOND	10-37	No data	
	137-164	0.434	
	167-175	0.397	
	171-183	0.455	
INTER-SS	**37-137**	**0.619**	
	164-167	0.374	
	175-183	0.434	
	183-191	**0.705**	

Domain D2 is classified (CATH 2.10.220.10) as a Beta Ribbon. Its loosely packed structure is made up of beta folds interspersed with locally disordered fragments. The overall conformation of the domain is stabilized by SS bonds which support a complex system of small beta sheets (Table 6.). In general, this domain is highly nonglobular and does not comprise a clear hydrophobic core (as indicated by its relatively high RD value).

Table 6. RD values for domain D2 of 1M6B, its individual secondary folds and all fragments bounded by Cys residues which form SS bonds. RD values in excess of 0.5 (indicating divergence from the idealized distribution) are listed in boldface. Right column gives the positions of Cys residues engaged in SS-bonds.

1M6B-D2	FRAGMENT	RD	SS-bonds Cys residues
Domain	**199-308**	**0.660**	
SECONDARY			
LOOP	**199-219**	**0.592**	207C,208C,212C, 216C
HELIX	**220-223**	**0.683**	
LOOP	**224-229**	**0.683**	224C,227C
BETA-I	230-232	0.308	
BETA-I	**235-237**	**0.697**	236C
LOOP	238-242	0.445	240C
BETA-II	**243-247**	**0.809**	
LOOP	**248-251**	**0.594**	
BETA-II	**252-256**	**0.518**	
LOOP	257-260	0.203	
BETA-III	261-263	0.382	
BETA-III	266-268	0.020	267C

1M6B-D2	FRAGMENT	RD	SS-bonds Cys residues
LOOP	269-275	0.182	271C
BETA-IV	**276-278**	**0.579**	
BETA-IV	**281-283**	**0.735**	282C
LOOP	284-288	0.278	286C
BETA-V	289-295	0.407	
BETA-V	**298-304**	**0.556**	301C,304C
LOOP	**305-308**	**0.884**	308C
BETA-SHEET	**I**	**0.521**	236C
	II	**0.619**	
	III	**0.558**	267C
	IV	**0.610**	282C
	V	0.456	301C,304C
SS-BONDS	195-206	Inter-domain	
	207-216	**0.671**	208C,212C
	212-224	**0.607**	216C
	227-236	0.421	
	240-267	**0.599**	
	271-282	**0.562**	
	286-301	0.316	
	304-308	**0.679**	
INTER-SS	199-207	0.286	
	216-224	**0.626**	
	224-227	**0.811**	
	236-240	**0.543**	
	267-271	0.218	
	282-286	0.224	
	301-304	**0.605**	

Domain D3, classified (CATH 3.80.20.20) as an Alpha-Beta Horseshoe, has a globular shape and contains a hydrophobic core (RD < 0.5). Despite superficial similarities to D1 most secondary folds in D3 are also accordant with the model, as are its inter-SS fragments, two beta sheets and a system of helixes (Table 7.).

Table 7. RD values for domain D3 of 1M6B, its individual secondary folds and all fragments bounded by Cys residues which form SS bonds. RD values in excess of 0.5 (indicating divergence from the idealized distribution) are listed in boldface.

1M6B-D3	FRAGMENT	RD	SS-BONDS	LIGAND
Domain	310-480	0.488		
SECONDARY				
BETA-I	310-314	0.422	312C	
LOOP	315-327	0.379		
HELIX	328-333	0.400		

1M6B-D3	FRAGMENT	RD	SS-BONDS	LIGAND
BETA-I	336-339	0.383	335C	
BETA-II	342-344	0.273		
HELIX	**345-351**	**0.649**		
LOOP	352-360	0.214		
HELIX	361-371	0.249		
BETA-I	372-376	0.317		
BETA-II	**377-381**	**0.545**		
LOOP	**382-389**	**0.635**		385-389
HELIX	**390-394**	**0.573**		
BETA-I	397-400	0.445		
LOOP	401-407	0.411		
BETA-II	**408-415**	**0.608**		
LOOP	416-428	0.340		
BETA-I	429-431	0.354		
BETA-II	434-439	0.224		
LOOP	440-449	0.436	444C	
HELIX	450-455	0.381		
LOOP	456-461	0.409		
BETA-II	462-466	0.400		
LOOP	467-469	0.390		
HELIX	**470-475**	**0.528**	474C	
LOOP	**476-480**	**0.790**		
BETA-SHEET	I	0.388		
	II	0.463		
HELIXES		0.460		
SS-BONDS	312-335	0.448		
	444-474	0.435		
INTER-SS				
	335-444	0.497		385-389
	474-481	0.460		

Domain D4, with 7 disulfide bonds, is classified (CATH 2.10.220.10) as a Beta Ribbon and exhibits a nonglobular conformation. Its high RD value suggests strong stabilizing influence of SS bonds which link loosely packed loops. Only two of its beta sheets (consisting of two fragments each) are locally accordant with the model (Table 8.).

Table 8. RD values for domain D4 of 1M6B, its individual secondary folds and all fragments bounded by Cys residues which form SS bonds. RD values in excess of 0.5 (indicating divergence from the idealized distribution) are listed in boldface.

1M6B-D4	FRAGMENT	RD	SS-BONDS
Domain	**481-580**	**0.650**	
SECONDARY			
LOOP	**481-493**	**0.571**	481C,485C,490C
HELIX	**494-497**	**0.871**	
LOOP	**498-503**	**0.557**	498C,501C
BETA-I	**504-506**	**0.559**	
BETA-I	**509-511**	**0.687**	510C
LOOP	512-521	0.352	514C
BETA-II	522-526	0.307	
BETA-II	529-533	0.434	530C,533C
LOOP	**534-559**	**0.592**	537C,544C,546C,557C
BETA-III	**560-562**	**0.693**	
BETA-III	565-567	0.242	566C
LOOP	568-580	0.482	
BETA-SHEET	**I**	**0.598**	510C
	II	0.346	530C,533C
	III	0.255	566C
SS-BONDS	**481-490**	**0.627**	485C
	485-498	0.415	490C
	501-510	**0.506**	
	514-530	0.336	
	533-546	**0.578**	537C,544C
	537-554	0.466	544C
	557-566	0.490	
INTER-SS	490-498	0.497	
	498-501	**0.531**	
	510-514	0.402	
	530-533	0.497	
	537-546	0.354	
	546-554	**0.523**	
	566-580	0.482	

Comparison of 1M6B domains points to the role of disulfide bonds as the principal stabilizer of D1, D2 and D4, while the structure of D3 owes its stability mainly to hydrophobic effects. This domain exhibits a clear hydrophobicity density distribution gradient with a well formed hydrophobic core. Much like in 1QLL, fragments bounded by SS-forming Cys residues appear to emerge as a result of conformational rearrangement leading to the creation of a hydrophobic core. Comparative analysis of D1 and D3 reveals that – despite sharing a similar topology – these domains differ significantly with

respect to the stabilizing role of the hydrophobic core and disulfide bonds. Both domains are involved in binding ligands – specifically, the epidermal growth factor. Their properties suggest that the more stable (from the point of view of hydrophobicity density distribution) domain D3 serves as the static "backbone" while domain D1 retains greater elasticity, facilitating accommodation of ligand particles. This functional and structural differentiation of similar domains justifies the comparative analysis presented in our work.

2.5. Discordant core structure with all sections defined by SS bonds also discordant

Significant deviations from the theoretical hydrophobicity density distribution model, both with regard to the entire molecule and fragments defined by SS bonds, is found in bovine ribonuclease – nucleic acid hydrolase (EC 3.1.27.5) [22]. The ribonuclease is analyzed in conjunction with the angiogenin molecule (1ANG) [28], which was found to exhibit ribonuclease A activity despite structural differences in its ribonucleolytic active center and in the putative receptor center.

We may ask – to what degree are similarities in catalytic activity reflected by the core structure and SS bond system? Table 9. privides a comparative analysis of the hydrophobic core status in both molecules and their fragments.

The structure of the hydrophobic core (which, as defined by the fuzzy oil drop model, also comprises a hydrophilic "sheath") is not evident in ribonuclease (RD = 0.550). Elimination of catalytic residues reduces the RD value, bringing the molecule closer to the theoretical model. Analysis of hydrophobicity density distribution profiles reveals local discrepancies, especially in the neighborhood of catalytic residues. We may speculate that this neighborhood retains structural properties which favor enzymatic activity.

Table 9. RD values for tyhe complete ribonuclease (5RSA) and angiogenin (1ANG) following elimination of catalytic residues, for their individual secondary folds and all fragments bounded by Cys residues which form SS bonds, along with enumeration of catalytic residues. RD values in excess of 0.5 (indicating divergence from the idealized distribution) are listed in boldface.

5RSA	FRAGMENT	RD	Enzymatic residues	RD	1ANG
Chain		**0.550**		0.479	
No E		**0.538**	←**12H,41K,119H,120F** 13H,40K,**114H**→	0.470	
Secondary					
HELIX	**3-13**	**0.568**	**12H** 13H	0.380	3-14
LOOP	**14-23**	**0.565**		0.388	15-21
HELIX	24-33	0.323		0.277	22-33
LOOP 1E	34-41	0.445	41K **40K**	**0.695**	**34-40**
BETA - I	42-48	0.443		0.411	41-47
HELIX	50-56	0.477		0.492	49-55
LOOP	57-60	0.201		0.085	58-59H
BETA - II	61-63	0.342		0.317	61-65
LOOP	**64-70**	**0.549**		0.060	66-68
BETA - II	71-75	0.459		0.255	69-73

5RSA	FRAGMENT	RD	Enzymatic residues	RD	1ANG
BETA - I	**79-87**	**0.542**		0.435	76-84
LOOP	88-95	0.123		0.430	85-92
BETA - I	**96-105**	**0.548**		0.440	93-101
BETA - II	**106-111**	**0.576**		0.445	103-108
LOOP	112-115	0.251	**114H**	**0.794**	**111-116**
BETA - II	**116-124**	**0.658**	**119H,120F**	0.418	117-121H
BETA SHEET I		**0.556**		**0.533**	
BETA SHEET II		**0.592**	**119H,120F**	0.492	
BETA SHET - ALL		**0.589**	**119H,120F**	**0.554**	
SS-BONDS					
	26-84	**0.535**		0.443	26-81
	40-95	**0.633**	**41K 40K**	**0.555**	**39-92**
	58-110	**0.536**		0.460	57-107
	65-72	**0.514**			
INTER-SS					
	1-25	**0.534**	**12H** 13H	0.310	1-26
	40-58	**0.636**	**41K 40K**	**0.575**	**39-57**
	111-124	**0.533**	**119H,120F** 114H	0.478	107-123

Comparable analysis of two enzymes in respect to their similar enzymatic activity suggests the significance of the consequences of SS-bonds limited fragments 40-95 in ribonuclease and 39-92 in angiogenin. These two fragments represent highly discordant (in respect to fuzzy oil drop model) status in respect to hydrophobic core structure (taking fuzzy oil drop model as the criterion). It is important due to the presence of catalytic residues in these fragments (41K in ribonuclease and 40K in angiogenin). Second common characteristics can be observed for fragment 40-58 in ribonuclease and analogical one 39-57 in angiogenin (these two fragments are limited by the localization of two Cys residues engaged in SS-bonds with other partners). Both these fragments represent the highest discordance in respect to fuzzy oil drop model. The two mentioned above catalytic residues are localized in these fragments.

Catalytic residues 114H in angiogenin and 119H and 120F in ribonuclease are similarly localized in the very highly discordant surrounding represented by loop 111-116 in angiogenin and 116-124 Beta-structural fragment in ribonuclease.

Some sort of switch may be observed for catalytic residues 12H and 41K in ribonuclease in comparison with 13H and 30K in angiogenin. One of them from each pair is localized in discordant neighborhood while the other one is surrounded by the fragment of ordered hydrophobic core. The similar enzymatic activity may be explained using the status of catalytic residues and their local surrounding which appear to be similar in these two enzymes.

Figure 6. Hydrophobicity density distribution profiles for bovine ribonuclease: theoretical (dark blue) and observed (red). Catalytic residues are indicated by green triangles while the yellow triangles correspond to disulfide bonds.

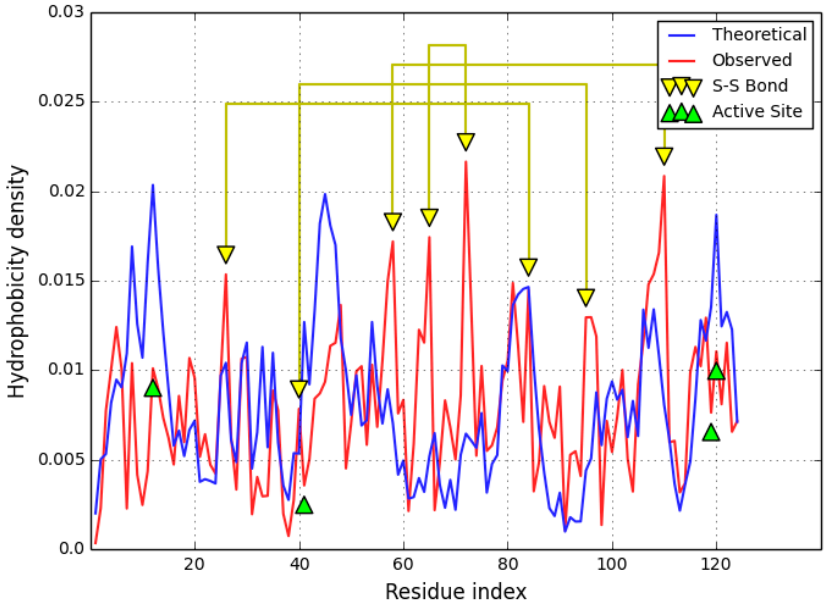
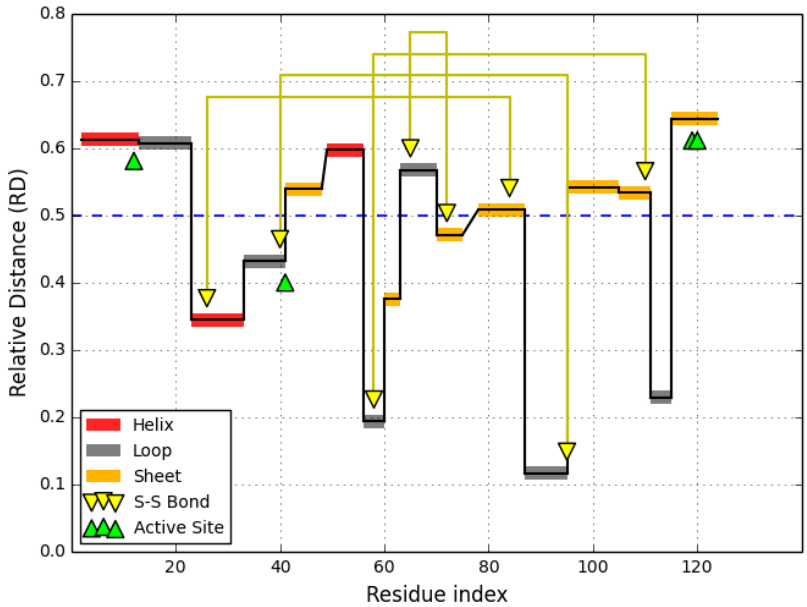


Figure 7. Diagram reveals the status of each secondary structural folds distinguished by colors (see Table 9.), as well as fragments bounded by Cys residues which form SS bonds. The yellow triangles above the horizontal dashed line (RD=0.5) distinguish fragments with RD > 0.5.



Many studies of the folding process of ribonuclease have been published over the years, including Anfisen's seminal experiment [29] which proves that, in the presence of disulfide isomerase, ribonuclease quickly achieves full *in vitro* enzymatic potency, suggesting the correct arrangement of disulfide bonds. On the basis of fuzzy oil drop model the secondary structure fragments represent

differentiated status (Figure 6.). Also in light of fuzzy oil drop analysis, structural stabilization in the presence of water is mediated by fragments which exhibit close-to-theoretical hydrophobicity density distribution (Figure 7.). We may speculate that as the polypeptide chain folds, Cys residues at positions 26, 40 and 58 reach their preferred positions, as do their complementary residues at 95 and 72. In other words, the formation of a hydrophobic „drop" in the aqueous environment promotes correct positioning of the aforementioned residues, while the remaining Cys residues (at 84, 110 and 65) require further guidance (Figure 7.).

The folding process of ribonuclease results in a complex structure comprised of a stabilizing backbone (containing a hydrophobic core) and an unstable fragment which mediates biological activity. This selective instability is supported by the correct arrangement of disulfide bonds.

2.6. Discussion

The hydrophobic core is regarded as an important factor in tertiary structural stabilization. The fuzzy oil drop model attempts to express this phenomenon in mathematical terms, describing the idealized hydrophobicity density distribution as a 3D Gaussian and applying Kullback-Leibler's divergence entropy criterion to quantitatively express departures from this theoretical profile. As a result, we can identify fragments of the polypeptide chain which do not follow the theoretical model. A hypothetical protein characterized by perfect hydrophobicity density distribution (i.e. internalization of all hydrophobic residues coupled with exposure of all hydrophilic residues on the surface) would be highly soluble – a desired property in many biological systems – but would remain incapable of interacting with any other molecules. Thus, local deviations from the theoretical model may be – and indeed usually are – connected with areas of biological activity. Such activity may include catalysis (e.g. in enzymes), ligand binding or protein complexation (for proteins with a quaternary structure). Identification of discordant fragments is therefore an interesting research topic, particularly in the context of fostering favorable conditions for intermolecular interactions in the immediate neighborhood of the catalytic center.

Another factor widely regarded as exerting a stabilizing influence upon the protein's quaternary structure, are disulfide bonds. Clearly, the introduction of additional covalent bonds produced a more rigid – and therefore more stable – molecule. The question remains: to what degree do these factors reinforce each other? In the course our work we have identified proteins where both factors appear to serve a common purpose (well defined hydrophobic core reinforced by a system of SS bonds), such as 1QLL (phospholipase neurotoxin). In other cases SS bonds appear to counteract hydrophobic effects, supporting a locally divergent structure – e.g. in the ribonuclease lysozyme, disulfide isomerase or selected domains of serine protease (1M6Q). The location of catalytic residues in particular seems to correlate with locally divergent conditions. In most cases elimination of catalytic residues produces a more accordant molecule, as evidenced by changes in its RD value [30]. Here, we focus on the effects of disulfide bonds which counteract hydrophobic forces – although this phenomenon is not evident in all proteins where such bonds are present (e.g. in 1QLL).

Also other models relate tertiary conformation to the presence of an aqueous environment, although such models do not supply any quantitative criteria. The "wet and dry area model" distinguishes regions within the protein body not available for water molecules ("dry" designation), as well as "wet"

areas which are in contact with water [31]. The model suggests the need to balance these two types of areas in protein body.

Another acknowledgement for the influence of water upon polypeptide chain folding is the nucleation model. It is suggested that the hydrophobic core coalescence around the "seed" understood as "nucleation" area [32]. The nucleation model is rather of dynamic form understood as describing the progression of the folding process. The fuzzy oil drop model represents an improvement upon this abstraction by proposing a quantitative criteria that express the status and influence of the hydrophobic core upon various types of structural units – domains, proteins and complexes [33].

The problem of hydrophobic/hydrophilic interaction between residues and its influence on the protein folding presented in [34] suggests that latter are more important than the former [35]. The idea of iceberg is treated as important conditioning of the proper folding process mediated by structuralization of water [36,37]. The "fuzzy oil drop" model solves at least one problem introducing the unification of hydrophobicity/hydrophilicity in form of mathematical model allowing the quantitative estimation of the role of hydrophobic effects in generation of native structures of proteins.

3. Experimental Section

The presented analysis focuses on a set of enzymes which contain disulfide bonds. In all selected enzymes the catalytic residues (i.e. loci of enzymatic activity) are well known. Our study set consists of proteins with varying lengths and number of disulfide bonds. We have also tried to single out enzymes whose activity profile is well known, assuming that a large body of published knowledge should facilitate validation of results.

The list shown in Table 10 includes – in addition to single-chain single-domain proteins – a multi-domain protein with a variable number of SS bonds.

Table 10. List of proteins subjected to analysis, along with the number of disulfide bonds present in each protein (numbers in parenthesis include inter-domain bonds).

PROTEIN	PDB ID	RESIDUES IN CHAIN	SS BONDS	Reference
ENZYME				
DISULFIDE ISOMERASE	1MEK	120	1	[25]
LYSOZYME	1LZ1	130	4	[20]
RIBONUCLEASE	5RSA	124	4	[27]
ANGIOGENIN	1ANG	123	3	[28]
NEUROTOXIN	1QLL	121	7	[18]
POSPHOLIPASE				
TRANSFERASE	1M6B			[19]
	Domain1	167	3 (4)	
	Domain2	110	7 (8)	
	Domain3	172	2	
	Domain4	100	7	

4. Conclusions

Taking into account the high specificity of enzymes, ligand detection mechanisms must depend upon more than just the location of catalytic residues. When studying the immediate neighborhood of the catalytic center we should expect to encounter conditions which not only promote catalysis but also facilitate accommodation of specific ligands. Fragments which diverge from the idealized hydrophobicity density distribution profile appear to fulfill this purpose. In conclusion, while SS bonds are rightly regarded as enhancing the protein's structural stability, it is worth noting that in many cases they specifically counteract the local effects of hydrophobic forces. Fragments of local instability as recognized on the basis of fuzzy oil drop model appear to be related to biological function. Disulfide bonds - generally treated as stabilization of III-order structure – seem to stabilize these local instabilities necessary for local conformational changes related to biological function.

Acknowledgments

The authors would like to express their thanks to Piotr Nowakowski and Anna Śmietańska for valuable suggestions and editorial work. This research was financially supported by the Jagiellonian University Medical College grant No. K/ZDS/001531.

Author Contributions

Leszek Konieczny designed the physical model. Irena Roterman designed the mathematical model (application of 3D Gaussian) and applied Kullback-Leibler's entropy criterion to measure the differences between theoretical, observed and unified distributions to quantitatively estimate the differences between profiles. Barbara Kalinowska assembled the study set of proteins and prepared figures with 3D presentations. Mateusz Banach prepared a program which calculates hydrophobicity density distribution profiles and performed computation of fuzzy oil drop model parameters. Irena Roterman analyzed the results, prepared the final interpretation and authored the paper. All authors have read and approved the final manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Creighton, T.E. Disulfide bonds and protein stability. *Bioessays* **1988**, *8*, 57–63.
2. Harrison, P.M.; Sternberg, M.J. Analysis and classification of disulfide connectivity in proteins. The entropic effect of cross-linkage. *J. Mol. Biol.* **1994**, *244*, 448–463.
3. Weissman, J.S.; Kim, P.S. A kinetic explanation for the rearrangement pathway of BPTI folding. *Nature Structural Biology* **1995**, *2*, 1123–1130.
4. Weissman, J.S.; Kim, P.S. Efficient catalysis of disulfide bond rearrangements by protein disulphide isomerase. *Nature* **1993**, *365*, 185–188.

5. Milbrandt, A.G.; Boulegue, C.; Moroder, L.; Renner, C. The two cysteine-rich head domains of minicollagen from Hydra nematocysts differ in their cystine frameworkand overall fold despite an identical cysteine sequence pattern. *J. Mol. Biol.* **2005**, *354*, 591–600.
6. Carrega, L.; Mosbah, A.; Ferrat, G.; Beeton, C.; Andreotti, N.; Mansuelle, P.; Darbon, H.; De Waard, M.; Sabatier, J-M. The impact of the fourth disulfide bridge in scorpion toxins of the β-KTx6 subfamily. *Proteins* **2005**, *61*, 1010–1023.
7. Kather, I.; Bippes, C.A.; Schmid, F.X. A stable disulfide-free gene-3-protein of phage fd generated by *in vitro* evolution. *J. Mol. Biol*. **2005**, *354*, 666–678.
8. Huang, K.; Maiti, N.C.; Philips, N.B.; Carey, P.R.; Weiss, M.A. Structure-specific effects of protein topology on cross-β assembly: studies of insulin fibrillation. *Biochemistry* **2006**, *45*, 10278–10293.
9. Indu, S.; Kochat, V.; Thakurela, S.; Ramakrishnan, C.; Varadarajan, R. Conformational analysis and design of cross-strand disulphides in antiparallel Beta-sheets. *Proteins* **2011**, *79*, 244–260.
10. Takano, K.; Yamagata, Y.; Yutani, K. A general rule for the relationship between hydrophobic effect and conformational stability of a protein: stability and structure of a series of hydrophobic mutants of human lysozyme. *J Mol Biol.* **1998**, *280(4),* 749–761.
11. Konieczny, L.; Brylinski, M.; Roterman, I. Gauss-Function-Based Model of Hydrophobicity Density in Proteins. *In Silico Biology*. **2006**, *2*, 15–22.
12. Kalinowska, B.; Banach, M.; Konieczny, L.; Roterman, I. Application of Divergence Entropy to Characterize the Structure of the Hydrophobic Core in DNA Interacting Proteins. *Entropy,* **2015**, *17*, 1477–1507.
13. Kalinowska, B.; Banach, M.; Konieczny, L.; Marchewka, D.; Roterman, I. Intrinsically disordered proteins-relation to general model expressing the active role of the water environment. *Adv Protein Chem Struct Biol.***2014**, *94*, 315–46.
14. Banach, M.; Konieczny, L.; Roterman, I. The fuzzy oil drop model, based on hydrophobicity density distribution, generalizes the influence of water environment on protein structure and function. *J Theor Biol*. **2014**,*359*, 6–17.
15. Kauzmann, W. Some factors in the interpretation of protein denaturation. *Adv. Protein Chem*. **1959**, *14*, 1–63.
16. Levitt, M. A simplified representation of protein conformations for rapid simulation of protein folding. *J Mol Biol,* **1976**, *104*, 59–107.
17. Kulback S, Leibler RA On information and sufficiency. *Ann. Math. Stat.* **1951**, *22*, 79–86.
18. Lee**,** W.H.**;** da Silva, Giotto, M.T.; Marangoni, S.; Toyama, M.H.; Polikarpov, I.; Garratt, R.C. Structural basis for low catalytic activity in Lys49 phospholipases A2--a hypothesis: the crystal structure of piratoxin II complexed to fatty acid. *Biochemistry.* **2001**, *40*, 28–36.
19. Cho, H.S.; Leahy, D.J. Structure of the extracellular region of HER3 reveals an interdomain tether. *Science*. **2002**, *297*, 1330–1333.
20. Artymiuk, P.J.; Blake, C.C. Refinement of human lysozyme at 1.5 A resolution analysis of non-bonded and hydrogen-bond interactions. *J Mol Biol.* **1981**, *152*, 737–762.
21. Morozova, L.A.; Haynie, D.T.; Arico-Muendel, C.; Van Dael, H.; Dobson, C.M. Structural basis of the stability of a lysozyme molten globule. *Nature Structural Biol*. **1995**, *2*, 871–875.

22. Radford, S.E.; Dobson, C.M.; Evans, P.A. The folding of hen lysozyme involves partially structured intermediates and multiple pathways. *Nature* **1992**, *358*, 302–307.
23. Dhulesia, A.; Cremades, N.; Kumita, J.R.; Hsu, S.T.; Mossuto, M.F.; Dumoulin, M.; Nietlispach, D.; Akke, M.; Salvatella, X.; Dobson, C,M. Local cooperativity in an amyloidogenic state of human lysozyme observed at atomic resolution. *J Am Chem Soc*. **2010**, *132*, 15580–15588.
24. Verma, D.; Jacobs, D.J.; Livesay, D.R. Changes in lysozyme flexibility upon mutation are frequent, large and long-ranged. *PLOS- Comp Biol*. **2012**, *8(3),* e1002409.
25. Kemmink, J.; Darby, N.J.; Dijkstra, K.; Nilges, M.; Creighton, T.E. Structure determination of the N-terminal thioredoxin-like domain of protein disulfide isomerase using multidimensional heteronuclear 13C/15N NMR spectroscopy. *Biochemistry.* **1997**, *35*, 7684–7691.
26. Yarden, Y.; Pines, G. The ERBB network: at last, cancer therapy meets systems biology. *Nat Rev Cancer* **2012**, *12*, 553–563.
27. Wlodawer, A.; Borkakoti, N.; Moss, D.S.; Howlin, B. Comparison of two independently refined models of ribonuclease-A. *Acta crystallogr, Sect b*,**1986**, *42*, 379.
28. Acharya, K.R.; Shapiro, R.; Allen, S.C.; Riordan, J.F.; Vallee, B.L. Crystal structure of human angiogenin reveals the structural basis for its functional divergence from ribonuclease. *Proc Natl Acad Sci U S A*. **1994,** *91(8),* 2915–9.
29. Anfinsen, C.B. Principles that govern the folding of protein chains. *Science*. **1973**, *181*, 223–230.
30. Prymula, K.; Jadczyk, T.; Roterman, I. Catalytic residues in hydrolases: analysis of methods designed for ligand-binding site prediction. *J Comput Aided Mol Des*. **2011**; *25(2),* 117–133.
31. Das, P.; Kapoor, D.; Halloran, K.T.; Zhou, R.; Matthews, C.R. Interplay between Drying and Stability of a TIM Barrel Protein: A Combined Simulation-Experimental Study, *J. Am. Chem. Soc*. **2013**, *135*, 1882–1890.
32. Galzitskaya, O.V.; Ivankov, D.N.; Finkelstein, A.V. Folding nuclei in proteins, *FEBS Lett.***2001**, *489*, 113–118.
33. Roterman, I.; Konieczny, L.; Banach, M.; Marchewka, D.; Kalinowska, B.; Baster, Z.; Tomanek, M.; Piwowar, M. Simulation of the Protein Folding Process. In : Computational Methods to Study the Structure and Dynamics of Biomolecules and Biomolecular Processes - From Bioinformatics to Molecular Quantum Mechanics. Springer Series in Bio-/Neuroinformatics, Vol. 1.2014. Ed Adam Liwo, Springer Pages 599–638.
34. Ben-Naim, A.Myths and verities in protein folding theories: from Frank and Evans iceberg-conjecture to explanation of the hydrophobic effect. *J Chem Phys*.**2013**, 139(16):165105.
35. Ben-Naim A. Theoretical aspects of self-assembly of proteins: a Kirkwood-Buff-theory approach. *J Chem Phys.***2013**, 138(22):224906.
36. Ben-Naim A. On the So-Called Gibbs Paradox, and on the Real Paradox *Entropy* **2007**, *9*(3), 132–136.
37. Ben-Naim A. Theoretical aspects of pressure and solute denaturation of proteins: A Kirkwood-buff-theory approach. *J Chem Phys*.**2012**, *137(23):*235102.