

Article

# Role of Disulfide Bonds in Stabilizing the Conformation of Selected Enzymes—An Approach Based on Divergence Entropy Applied to the Structure of Hydrophobic Core in Proteins

Mateusz Banach <sup>1,2</sup>, Barbara Kalinowska <sup>1,2</sup>, Leszek Konieczny <sup>3</sup> and Irena Roterman <sup>1,\*</sup>

<sup>1</sup> Department of Bioinformatics and Telemedicine, Collegium Medium, Jagiellonian University, Lazarza 16, 31-530 Krakow, Poland; mateusz.banach@uj.edu.pl (M.B.); bb.kalinowska@gmail.com (B.K.)

<sup>2</sup> Faculty of Physics, Astronomy and Applied Computer Science, Jagiellonian University, 30-059 Krakow, Poland

<sup>3</sup> Department of Medical Biochemistry, Collegium Medicum, Jagiellonian University, Kopernika 7, 31-034 Krakow, Poland; mbkoniec@cyf-kr.edu.pl

\* Correspondence: myroterm@cyf-kr.edu.pl; Tel.: +48-12-6199693

Academic Editor: Raúl Alcaraz Martínez

Received: 23 December 2015; Accepted: 18 February 2016; Published: 25 February 2016

**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. The assessment of the effects of disulfide bonds is interpreted by 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. The proteins were identified with both factors reinforcing each other, as well as proteins where they seem to counteract each other. The analysis presents 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 non-binding 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 (Bovine Pancreatic Trypsin Inhibitor) 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 cysteine 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].

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 [7,8]. 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 [9,10]. 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 (Fuzzy Oil Drop) 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 fuzzy oil drop (FOD) model is a modification of Kauzmann's oil drop paradigm [11]. The model likens the folding of a polypeptide chain to the behavior of a drop of oil immersed 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 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 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:  $\sigma_x$ ,  $\sigma_y$  and  $\sigma_z$ . 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 (empirical) 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 9 Å (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

$$\tilde{H}t_j = \frac{1}{\tilde{H}t_{sum}} \exp\left(\frac{-(x_j - \bar{x})^2}{2\sigma_x^2}\right) \exp\left(\frac{-(y_j - \bar{y})^2}{2\sigma_y^2}\right) \exp\left(\frac{-(z_j - \bar{z})^2}{2\sigma_z^2}\right).$$

$\tilde{H}t_j$  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. This distribution is called  $T$  in this paper.

### 1.1.2. Observed Distribution

Observed distribution (as proposed by Levitt [12]):

$$\tilde{H}o_j = \frac{1}{\tilde{H}o_{sum}} \sum_{i=1}^N (H_i^r + H_j^r) \begin{cases} \left[ 1 - \frac{1}{2} \left( 7 \left( \frac{r_{ij}}{c} \right)^2 - 9 \left( \frac{r_{ij}}{c} \right)^4 + 5 \left( \frac{r_{ij}}{c} \right)^6 - \left( \frac{r_{ij}}{c} \right)^8 \right) \right] & \text{for } r_{ij} \leq c \\ 0 & \text{for } r_{ij} > c \end{cases}$$

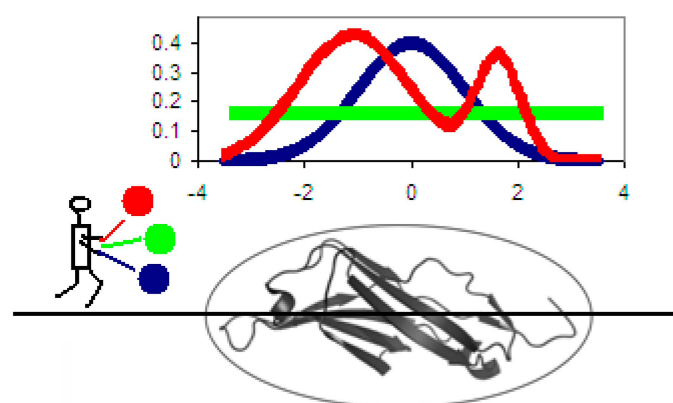
is the number of amino acids in the protein,  $\tilde{H}_i^r$  expresses the hydrophobicity parameter of the  $i$ -th residue while  $r_{ij}$  expresses the distance between two interacting residues ( $j$ -th effective atom and  $i$ -th effective atom).  $c$  expresses the cutoff distance for hydrophobic interactions, which is taken as 9 Å (following [12]). Observed hydrophobicity density values  $\tilde{H}o_j$  are also computed for effective atoms (geometric center of each side chain). Observed hydrophobicity distribution is called  $O$  in this paper.

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

### 1.1.3. Unified Distribution

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

$N$  is the number of amino acids in the chain. Figure 1 provides a graphical depiction of the model. Each residue represents equal hydrophobicity density producing the protein molecule deprived of any form of core—concentration of hydrophobicity in any place in the protein body. This distribution is called  $R$  in this paper. 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 one measures actual density which depends on the intrinsic hydrophobicity and placement of each residue (readings shown in red) while the third detector outputs a static reading, determined purely 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 (in the mathematical sense), our analysis focuses on discrete values computed for each effective atom.



**Figure 1.** Schematic representation of hydrophobicity density distribution in a sample protein. The presented profiles correspond to theoretical- $T$  (dark blue), observed- $O$  (red) and unified- $R$  (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  $\pm 3\sigma$  in each principal dimension (for simplicity’s sake the presentation is limited to a single coordinate system axis).

#### 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 (no hydrophobic core present). Quantitative assessment of these parameters bases on Kullback–Leibler’s divergence entropy formula [13]:

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

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 [9]. This, however, requires normalization (rescaling) of  $H_t$ ,  $H_o$  and  $H_r$  (hydrophobicity yielded by distribution  $R$ ) 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 and assess the involvement of a well ordered hydrophobic core in structural stabilization.

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 [8]—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. 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 1 includes—in addition to single-chain single-domain proteins—a multi-domain protein with a variable number of SS bonds.

**Table 1.** 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
Enzymes	-	-	-	-
Disulphide isomerase	1MEK	120	1	[14]
Lysozyme Hydrolase (O-glycosyl)	1LZ1	130	4	[15]
Ribonuclease Hydrolase (Nucleic acid–RNA)	5RSA	124	4	[16]
Angiogenin	1ANG	123	3	[17]
Neurotoxin Phospholipase	1QLL	121	7	[18]
Transferase	1M6B	-	-	[19]
	Domain1	167	3 (4)	-
	Domain2	110	7 (8)	-
	Domain3	172	2	-
Domain4	100	7	-	
Hydrolase (Carboxylic esterase)	1THG	544	2	[20]

Comparison of results obtained using the fuzzy oil drop model with the Accessible Solvent Area (ASA) reveals the mutual relations between both analysis methods. For this purpose we selected two proteins for which the FOD model returns differing results. ASA calculations were performed using the POPS software Version 1.0.6 (Parameter Optimised Surfaces) [21]. The resulting values indicate the percentage of each residue’s surface area which remains in contact with water.

## 3. Results and Discussion

The status of sections bounded by Cys residues which form SS bonds is summarized in Table 2. The table lists the status of the entire protein molecule (or domain, where appropriate) in relation to the status of its Cys-bounded fragments.



**Table 2.** 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.

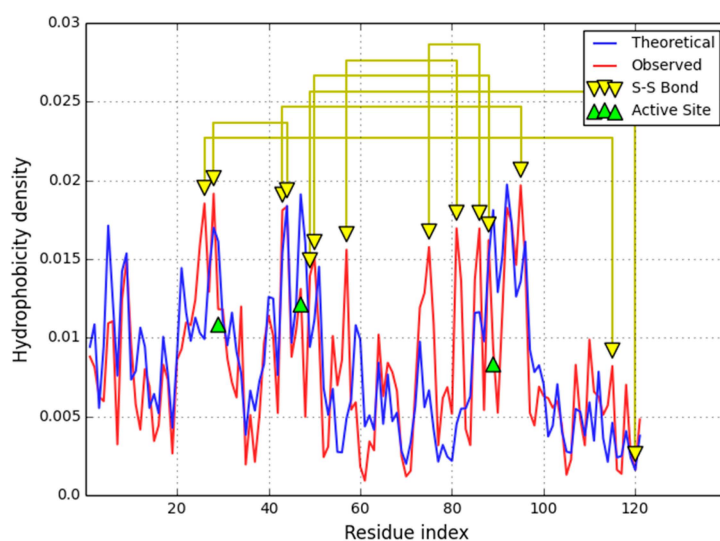
		Fuzzy oil drop accordance whole protein	
		Yes	No
Status of fragments defined by SS-bonds	Accordant—all	1QLL [14] 1M6B-D3	1M6B-D1 [15]
	Some accordant	1M6B-D2, 1M6B-D4, 1LZ1 [16], 1ANG [17], 1THG [18]	
	Discordant—all	1MEK [19]	5RSA [20]

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.

### 3.1. 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*) [14].

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 2. This property is shared by catalytic residues in most proteins.



**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).

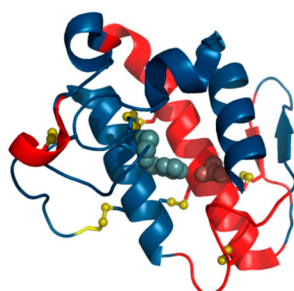
Excision of residues involved in binding ligands (NoL in Table 3) and mediating inter-chain interactions (NoP-P in Table 3) does not appreciably affect *RD*. This suggests that ligand binding and protein complexation do not require major rearrangements within the protein's hydrophobic core. 1QLL's ligand (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.

**Table 3.** *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 (Enzymatic residue), protein-protein interaction (P-P) and ligand binding (Ligand) respectively. *RD* values in excess of 0.5 (indicating divergence from the idealized distribution) are shown in boldface.

1QLL	Fragment	<i>RD</i>	Enzymatic residue	P-P	Ligand
Complex	-	0.641	-	-	-
Chain A/B	-	0.400/0.397	-	-	-
No-E	-	0.393	-	-	-
No-L	-	0.422	-	-	-
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	<b>54–64</b>	<b>0.606</b>	-	-	-
Beta	65–69	0.217	-	-	-
Beta Cys-1	<b>72–76</b>	<b>0.619</b>	-	-	-
Helix Cys-4	<b>79–99</b>	<b>0.508</b>	89D	97R	92V
Helix	<b>100–103</b>	<b>0.970</b>	-	-	-
Helix	104–108	0.247	-	-	-
Helix Cys-1	<b>111–115</b>	<b>0.864</b>	-	-	-
Loop Cys-1	116–121	0.260	-	-	-
Beta-sheet	-	0.394	-	-	-
	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
SS-bonds	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	-	-	-

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).



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



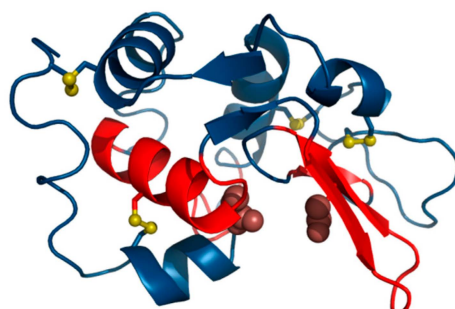
The main focus of our analysis are fragments defined by the placement of Cys residues which form disulfide bonds. As it turns out, all such fragments conform to the fuzzy oil drop model. Cys residues are generally expected to migrate towards the center of the molecule as the hydrophobic core emerges—their 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.

In 1QLL we can identify several fragments for which the actual hydrophobicity density exceeds theoretical values (shown in Figure 2). When the theoretical distribution ascribes low hydrophobicity density to a given residue, we should expect to find that residue close to the molecular surface. If the observed hydrophobicity density is greater than the corresponding theoretical value, we are dealing with an externalized hydrophobic residue. Such residues frequently mediate interaction with hydrophobic ligands or membranes (as indeed is the case in 1QLL).

### 3.2. Discordant Core Structure with Some Sections Defined by SS Bonds Also Discordant

The human lysozyme—O-glycosyl hydrolase (EC 3.2.1.17) [16]—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 tend to diverge from the model. Among the secondary folds present in 1LZ1 the following are regarded as discordant: 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.



**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.

In the case of 1LZ1 all catalytic residues are surrounded by locally discordant neighbor (Table 4). 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.

**Table 4.** *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	<i>RD</i>	Enzymatic residues
Chain	-	0.530	-
No E	-	0.515	-
Secondary	-	-	-
Helix	4–15	0.377	-
Loop	16–23	0.234	-
Helix	24–37	0.566	35E
Loop	38–41	0.160	-
Beta	42–46	0.620	-
Loop	47–50	0.653	-
Beta	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	-
	6–128	0.531	35E, 53D
SS-bonds	30–116	0.543	-
	65–81	0.279	-
	77–95	0.489	-
	6–30	0.415	-
Between	30–65	0.575	35,53
SS-bonds	77–81	0.198	-
	81–95	0.376	-

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.

In [22] 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 [23]. 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. Alpha-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].

### 3.3. 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) (1MEK) which contains a single, peculiar disulfide bond [14]. The bond spans a very short section—just 4 residues, all of which are enzymatically active. Our analysis focuses on the N-terminal domain of this four-domain

protein. Experimental studies indicate that its constituent domains are loosely linked (with just one inter-domain SS bond) and that the resulting conformation exhibits high flexibility [26].

The domain under consideration as a whole follows the hydrophobicity density distribution model with good accuracy (Table 5.). 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.

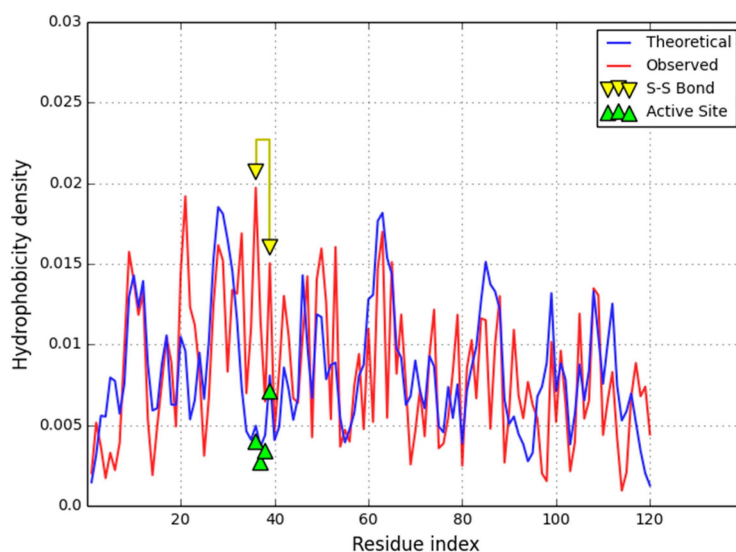
**Table 5.** *RD* values for disulfide isomerase. Also listed are *RD* values for fragments of the protein devoid of catalytic residues and the fragment at 86–94, which exhibits particularly strong discordance. *RD* values in excess of 0.5 (indicating divergence from the idealized distribution) are listed in boldface.

1MEK	Fragment	<i>RD</i>	Enzymatic residue	Cys
Chain	-	0.464	-	-
No E	-	0.433	-	-
Secondary	-	-	-	-
Loop	<b>1–3</b>	<b>0.579</b>	-	-
Beta	4–6	0.279	-	-
Beta	<b>9–12</b>	<b>0.503</b>	-	-
Helix	16–24	0.497	-	-
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	<b>68–72</b>	<b>0.555</b>	-	-
Helix	73–78	0.300	-	-
Loop	79–83	0.351	-	-
Beta	84–90	0.380	-	-
Loop	<b>91–96</b>	<b>0.684</b>	-	-
Beta	97–99	0.344	-	-
Loop	100–103	0.171	-	-
Helix	104–114	0.312	-	-
Loop	<b>115–120</b>	<b>0.747</b>	-	-
Beta-sheet	All Beta fragments	0.314	-	-
Helices	All helices	0.397	-	-
SS-bonds	<b>36–39</b>	<b>0.510</b>	<b>37G, 38H</b>	-

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 helices 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.

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.

The loops at 91–96 and 115–120 are both exposed on the surface where no strong concentration of hydrophobicity is expected. Not much information is available regarding this enzyme's interaction properties—we can only speculate that the observed discrepancy, as seen in Figure 5, indicates a potential ligand binding site or protein complexation site. Comparison with the four-domain yeast isomerase structure (3BOA—[27]) excludes involvement of the analyzed fragments in inter-domain interactions.



**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.

### 3.4. 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, and also due to its pharmacological properties [28]. Its constituent domains 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 S1 in Supplementary Material). As such, these bonds are not directly related to the structure of the domain's hydrophobic core.

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 S2 in Supplementary Material). In general, this domain is highly nonglobular and does not comprise a clear hydrophobic core (as indicated by its relatively high *RD* value).

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 helices (Table S3 in Supplementary Material).

Domain D4, with seven 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 S4 in Supplementary Material).

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.

Table 6 summarizes the structure of individual domains in 1M6B. Surprisingly, one domain (D3) out of two identified as adopting a horseshoe-like conformation remains accordant with the theoretical model while the other one (D1) diverges from it. This observation appears to correlate with the low number of SS-bonds in D3, suggesting that D3 may instead be stabilized by the presence of a hydrophobic core.

**Table 6.** Structural characterization of the complete A chain and four individual domains of transferase (1M6B). Numbers in parenthesis indicate the number of inter-domain SS-bonds. N—number of residues in the unit; NL—number of residues involved in ligand binding. The rightmost column provides a general description of the corresponding supersecondary structure.

1M6B	RD	SS-bonds	N	NL	Structure
Chain A	0.715	22	549	24	-
Domains	-	-	-	-	-
D1	0.610	5 (1)	167	2	Horseshoe
D2	0.660	7 (2)	110	5	Beta-ribbon
D3	0.488	1 (1)	172	13	Horseshoe
D4	0.650	7	100	4	Beta-ribbon

### 3.5. 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, are found in bovine ribonuclease—nucleic acid hydrolase (EC 3.1.27.5) [16]. The ribonuclease is analyzed in conjunction with the angiogenin molecule (1ANG) [17] (Table 7), which was found to exhibit ribonuclease A activity despite structural differences in its ribonucleolytic active center and in the putative receptor center.

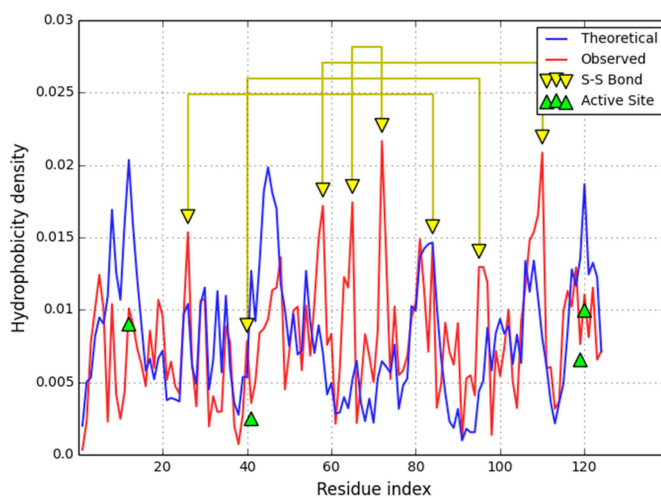
We may ask—to what degree are similarities in catalytic activity reflected by the core structure and SS bond system? Table 7 provides a comparative analysis of the hydrophobic core status in both molecules and their fragments.

The structure of the hydrophobic core (which, as defined by the fuzzy oil drop model, also comprises a hydrophilic shell Figure 6) 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.

Comparative analysis of two enzymes chosen due to their similar enzymatic activity suggests the significance of fragments delimited by SS-bonds (40–95 in ribonuclease and 39–92 in angiogenin respectively). These fragments are both highly discordant with respect to the fuzzy oil drop model. Notably, they both contain catalytic residues (41K in ribonuclease and 40K in angiogenin). Common characteristics can also be observed in fragments 40–58 in ribonuclease and 39–57 in angiogenin—both delimited by Cys residues engaged in SS-bonds and both highly discordant. The above-mentioned catalytic residues are also localized in these fragments.

**Table 7.** RD values for the 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	-	<b>0.550</b>	-	0.479	-
No E	-	<b>0.538</b>	←12H,41K,119H,120F 13H,40K,114H→	0.470	-
<b>Helix</b>	<b>3–13</b>	<b>0.568</b>	12H 13H	0.380	3–14
<b>Loop</b>	<b>14–23</b>	<b>0.565</b>	-	0.388	15–21
Helix	24–33	0.323	-	0.277	22–33
Loop 1E	34–41	0.445	41K 40K	<b>0.695</b>	<b>34–40</b>
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
<b>Loop</b>	<b>64–70</b>	<b>0.549</b>	-	0.060	66–68
Beta-II	71–75	0.459	-	0.255	69–73
<b>Beta-I</b>	<b>79–87</b>	<b>0.542</b>	-	0.435	76–84
Loop	88–95	0.123	-	0.430	85–92
<b>Beta-I</b>	<b>96–105</b>	<b>0.548</b>	-	0.440	93–101
<b>Beta-II</b>	<b>106–111</b>	<b>0.576</b>	-	0.445	103–108
Loop	112–115	0.251	114H	<b>0.794</b>	<b>111–116</b>
<b>Beta-II</b>	<b>116–124</b>	<b>0.658</b>	119H,120F	0.418	117–121H
<b>B-sheet I</b>	-	<b>0.556</b>	-	<b>0.533</b>	-
<b>B-sheet II</b>	-	<b>0.592</b>	119H,120F	0.492	-
<b>B-sheet-all</b>	-	<b>0.589</b>	119H,120F	<b>0.554</b>	-
SS-bonds	<b>26–84</b>	<b>0.535</b>	-	0.443	26–81
	<b>40–95</b>	<b>0.633</b>	<b>41K 40K</b>	<b>0.555</b>	<b>39–92</b>
	<b>58–110</b>	<b>0.536</b>	-	0.460	57–107
	<b>65–72</b>	<b>0.514</b>	-	-	-
Inter-SS	<b>1–25</b>	<b>0.534</b>	12H 13H	0.310	1–26
	<b>40–58</b>	<b>0.636</b>	41K 40K	<b>0.575</b>	<b>39–57</b>
	<b>111–124</b>	<b>0.533</b>	119H,120F 114H	0.478	107–123



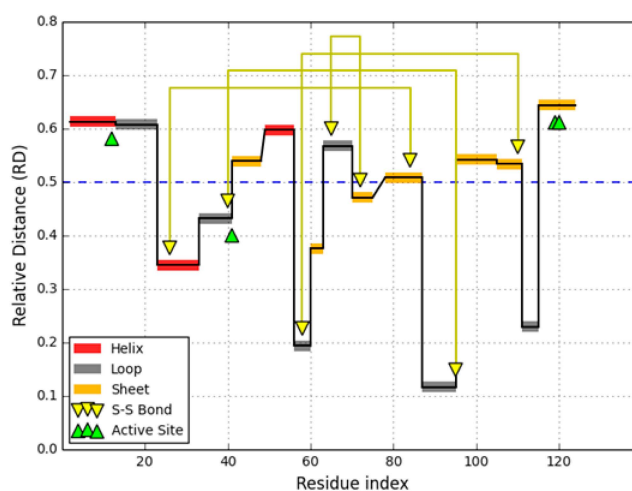
**Figure 6.** Hydrophobicity density distribution profiles for bovine ribonuclease: theoretical (dark blue) and observed (red). Catalytic residues are indicated by green triangles while the yellow triangles correspond to disulfide bonds.



Catalytic residues 114H in angiogenin and 119H and 120F in ribonuclease are similarly localized in a strongly discordant loop at 111–116 in angiogenin and a discordant Beta-structural fragment at 116–124 in ribonuclease respectively.

A sort of “switch” may be observed for catalytic residues 12H and 41K in ribonuclease in comparison with 13H and 30K in angiogenin. One residue from each pair is localized in a discordant neighborhood while the other one is surrounded by an accordant fragment. The observed similarities in enzymatic activity may be explained by the status of catalytic residues and their local neighborhoods which appear to be similar in these two enzymes.

Many studies of the folding process of ribonuclease have been published over the years, including Anfinsen’s seminal experiment [29] which proves that, in the presence of disulfide isomerase, ribonuclease quickly achieves full *in vitro* enzymatic potency, suggesting correct arrangement of disulfide bonds. The status of secondary fold is varied (Figure 7).



**Figure 7.** This diagram reveals the status of each secondary structural fold (in ribonuclease), distinguished by colors (see Table 5), as well as fragments bounded by Cys residues which form SS bonds. Yellow triangles above the horizontal dashed line ( $RD = 0.5$ ) distinguish fragments with  $RD > 0.5$ .

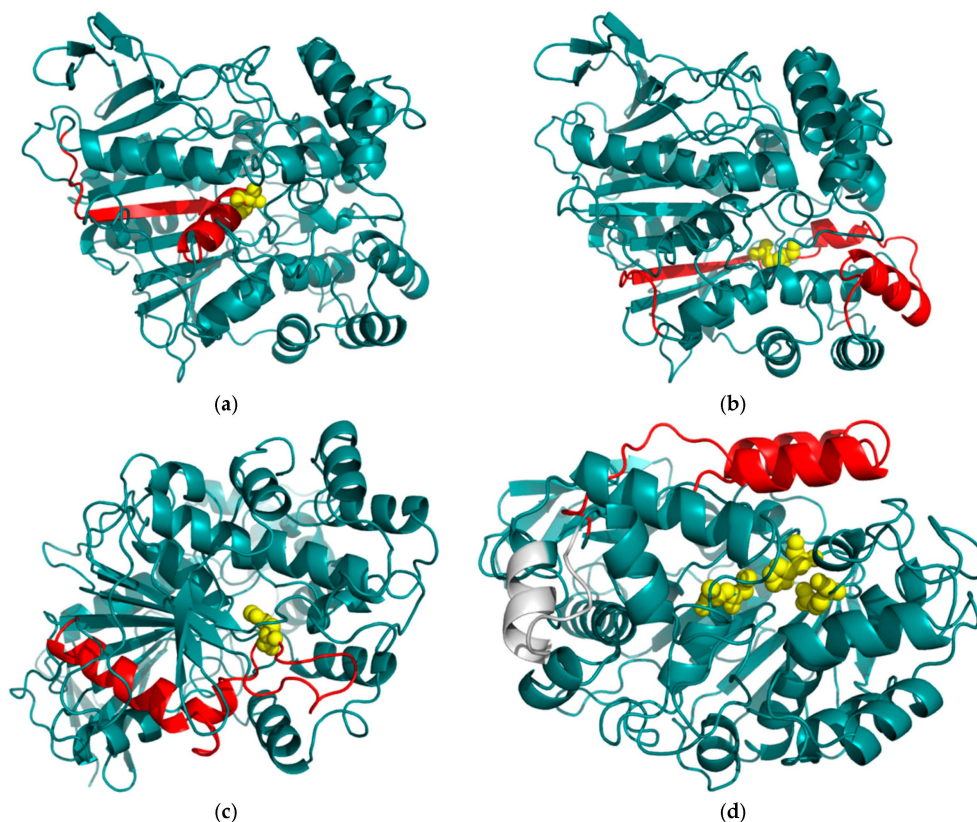
In light of fuzzy oil drop analysis, structural stabilization in the presence of water is mediated by fragments which exhibit close-to-theoretical hydrophobicity density distribution (Figure 7). We may speculate that, as the polypeptide chain folds, Cys residues at positions 26, 40 and 58 reach their 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. The folding process of ribonuclease results in a complex structure composed of a stabilizing backbone (which contains a hydrophobic core) and an unstable fragment which mediates biological activity. This selective instability is supported by the correct arrangement of disulfide bonds.

### 3.6. Discordant Section Defined by SS Bonds Related to Enzymatic Activity

Carboxylic esterase, a representative of the hydrolase group, contains two disulfide bonds. These bonds define a fragment which diverges from theoretical expectations and is located in close proximity to the active center. According to crystallographic analysis, the active center is described as “not accessible from the surface of the molecule” [18]. A single, spacious catalytic cavity can be found inside the protein body. This cavity is bounded by two helices which remain in contact with the surface. High temperature factor values have been reported for this fragment. As suggested in [18] several residues participating in each helix exhibit discrete disorders. The authors of [18] suggest that both

helices are involved in interfacial and substrate binding. On the basis of the fuzzy oil drop model, the presented fragment—a loop delimited by a disulfide bond (61–105)—diverges from the theoretical distribution, while the other similarly defined fragment (276–288) remains accordant with the model.

Figure 8 illustrates the distribution of discordant fragments. The immediate neighborhood of the catalytic center appears quite unstable, with fragments proximate to catalytic residues highly divergent from the model. A more in-depth analysis based on the fuzzy oil drop model reveals that these fragments are poorly integrated with the protein-wide hydrophobic core and can therefore undergo rearrangement and reorientation without affecting the molecule as a whole. The status of each of these fragments is listed in Figure 8a–d. As can be seen, fragments located close to catalytic residues 217S and 218A deviate from the idealized hydrophobicity density distribution profile (see Table 8). Figure 8b shows the surroundings of catalytic residue 354E, which also deviates from the idealized profile (Table 8). Figure 8c presents catalytic residue 463H which is locally discordant. The fragment which forms the “lid” of the enzymatic cavity is a discordant loop delimited by a disulfide bond—accordingly, it does not form part of the molecule-wide hydrophobic core (*i.e.*, the loop at 61–105, also bounded by a disulfide bond—see Figure 8d). On the other hand, the SS-bond-delimited loop at 276–288 remains accordant with the model and can therefore be suspected of promoting structural stability.



**Figure 8.** 3D visualization of 1THG with catalytic residues represented by yellow spheres. Fragments located in close proximity to catalytic residues are marked in red. (a) catalytic residues 217S and 218A (b) catalytic residue 354E and its discordant neighborhood ( $RD > 0.5$ ) (c) catalytic residue 463H and its discordant neighborhood ( $RD > 0.5$ ) (d) all catalytic residues marked by yellow spheres. Red: SS-bond-delimited fragment at 61–105 ( $RD > 0.5$ ); white: SS-bond-delimited fragment at 276–288 ( $RD < 0.5$ )

**Table 8.** *RD* values for 1THG (hydrolase—carboxylic esterase). Values in excess of 0.5 (indicating divergence from the idealized distribution) are listed in boldface. Fragments in close proximity to catalytic residues are underlined. The leftmost column shows the positions of catalytic residues.

	Fragment	<i>RD</i>	
1THG	-	<b>0.653</b>	
	3–7 BI	0.361	
	<b>11–15 BI</b>	<b>0.811</b>	
	16–18 BII	0.334	
	<b>20–29 BII</b>	<b>0.661</b>	
	30–33 L	0.490	
	34–38 H	0.321	
	34–38 H	0.367	
	<b>39–51 L</b>	<b>0.772</b>	
	52–54 BI	0.129	
	55–64 L	0.380	
	<b>65–78 H</b>	<b>0.627</b>	
	79–83 H	0.140	
	84–95 H	0.373	
	96–105 L	0.376	
	<b>106–114 BII</b>	<b>0.592</b>	
	115–121 L	0.436	
	<b>122–128 BII</b>	<b>0.904</b>	
	<b>129–135 L</b>	<b>0.728</b>	
	<b>136–141 H</b>	<b>0.874</b>	
	<b>143–153 H</b>	<b>0.558</b>	
	154–157 L	0.150	
	<b>158–163 BII</b>	<b>0.740</b>	
	164–165 L	0.266	
	<b>166–172 H</b>	<b>0.547</b>	
	173–181 H	0.404	
	<b>182–183 L</b>	<b>0.909</b>	
	<b>184–201 H</b>	<b>0.521</b>	
	203–204 H	0.108	
	<b>205–216 BII</b>	<b>0.545</b>	
	<u><b>217–229 H</b></u>	<u><b>0.556</b></u>	
	<u>217S, 218A</u>	231–233 H	0.431
		<b>234–236 BIII</b>	<b>0.670</b>
<b>237–238 L</b>		<b>0.915</b>	
239–241 BIII		0.211	
242–243 L		0.012	
<b>244–249 BII</b>		<b>0.682</b>	
250–264 L		0.331	
<b>265–275 H</b>		<b>0.738</b>	
276–280 L		0.380	
<b>281–292 H</b>		<b>0.515</b>	
<b>293–309 H</b>		<b>0.703</b>	
310–313 L		0.267	
<b>314–318 H</b>		<b>0.583</b>	
<b>319–330 L</b>		<b>0.532</b>	
331–338 H		0.400	
<u><b>339–344 L</b></u>		<u><b>0.516</b></u>	
<u><b>345–352 BII</b></u>		<u><b>0.530</b></u>	

Table 8. Cont.

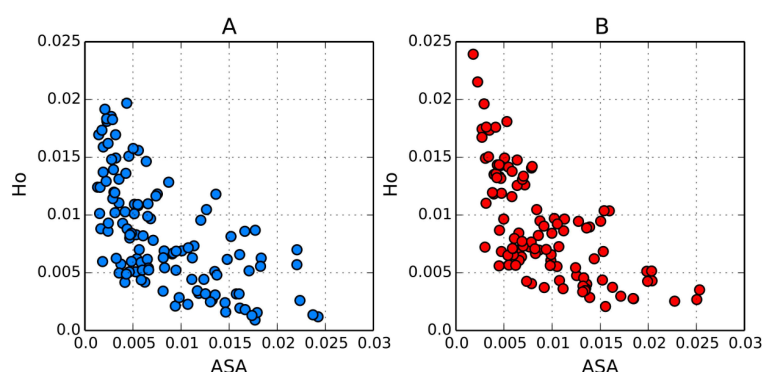
	Fragment	RD	
<b>354E</b>	<b>353–357 H</b>	<b>0.594</b>	
	<b>358–363 H</b>	<b>0.703</b>	
	<b>364–356 H</b>	<b>0.544</b>	
	<b>367–379 H</b>	<b>0.623</b>	
	<b>380–382 L</b>	<b>0.573</b>	
	383–394 H	0.367	
	395–396 L	0.101	
	397–401 H	0.419	
	<b>402–414 H</b>	<b>0.716</b>	
	<b>415–428 H</b>	<b>0.581</b>	
	<b>429–439 H</b>	<b>0.790</b>	
	440–442 BII	0.227	
	443–450 BII	0.260	
	<b>463H</b>	<b>451–465 L</b>	<b>0.744</b>
<b>466–472 H</b>		<b>0.647</b>	
<b>473–476 L</b>		<b>0.571</b>	
<b>477–492 H</b>		<b>0.617</b>	
493–510 L		0.403	
511–517 BII		0.147	
518–519 L		0.173	
<b>520–526 BII</b>		<b>0.647</b>	
527–528 L		0.066	
529–539 H		0.371	
540–543 H		0.343	
Beta I		-	0.404
<b>Beta II</b>		-	<b>0.630</b>
Beta III		-	0.269
<b>Beta I+II+III</b>	-	<b>0.705</b>	
<b>SS-bonds</b>	<b>61–105</b>	<b>0.558</b>	
	276–288	0.467	

1THG, taken as a whole, does not appear to contain a prominent hydrophobic core. This is due to the presence of numerous discordant fragments (shown in Figure 8). Nevertheless, the protein does include some fragments which promote structural stability by adhering to the theoretical hydrophobicity density distribution. The placement of red fragments suggests local flexibility which is necessary for the protein to perform its catalytic function (substrate binding, catalytic action, product release). This conclusion is in full agreement with the characteristics discovered via crystallographic analysis [20].

### 3.7. Comparison of FOD Results with Accessible Solvent Area (ASA) Descriptions

The fuzzy oil drop model determines the status of each residue by considering its expected hydrophobicity density (which is a function of the residue's placement *vis a vis* the protein's geometric center) and its actual hydrophobicity density, determined by hydrophobic interactions with other residues in a radius of 9 Å. In contrast, the Accessible Solvent Area (ASA) parameter depends on the area (Å<sup>2</sup>) or fraction of the residue's total surface area which remains in contact with water. It seems that both parameters should be correlated, with one parameter inversely proportional to the other. In order to test this hypothesis we have selected two proteins for which the FOD model yields differing results: 1QLL ( $RD = 0.400$ ) and domain 2 of 1M6B ( $RD = 0.660$ ). Figure 9 illustrates the results of our study. In order to enable comparative analysis we focused on the observed hydrophobicity density ( $O_i$ ), since this parameter determines the overall status of the residue and is affected by interactions with neighboring residues (9 Å radius). This value can be interpreted as the "likelihood" that the given

residue will act as a hydrophobic entity. As can be seen, the scatterplot depicting the mutual relation between both parameters approximates a hyperbole, with a relatively low spread.



**Figure 9.** Relation between the ASA parameter (normalized) of the surface area of each residue exposed to its environment; normalized values) and the observed hydrophobicity density calculated according to the fuzzy oil drop model (normalized values). (A) chain A of 1QLL (phospholipase); (B) domain 2 of 1M6B (transpherase).

#### 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. Local instabilities, as recognized on the basis of the fuzzy oil drop model, appear to be related to biological function. Disulfide bonds—generally treated as important for tertiary structural stabilization—appear to reinforce these local instabilities; creating suitable conditions for local conformational changes related to biological function.

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.

Disulfide bonds are widely regarded as exerting a stabilizing influence upon the protein's quaternary structure. Clearly, the introduction of additional covalent bonds produces a more rigid—and therefore more stable—molecule. The question remains: to what degree do all the listed 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. 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).

A spectacular example of local discordance found in the vicinity of catalytic residues is provided by the 1THG hydrolase—here, significant flexibility of the enzymatic cavity is required (as suggested by crystallographic analysis [20]). The presented examples illustrate the interplay between various stabilizing factors: tertiary structure, disulfide bonds and hydrophobic cores. These relations—as already indicated—remain varied; however they all serve a common purpose, enabling the protein to fulfill its biological role.

**Supplementary Materials:** The following are available online at [www.mdpi.com/1099-4300/18/3/67/s1](http://www.mdpi.com/1099-4300/18/3/67/s1), Table S1: *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, Table S2: *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, Table S3: *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, Table S4: *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.

**Acknowledgments:** The authors would like to express their thanks to Piotr Nowakowski and Anna Śmiateńska for valuable suggestions and editorial work. This research was supported by the Jagiellonian University Collegium Medicum grant No. K/ZDS/001531 and KNOW.

**Author Contributions:** Leszek Konieczny designed the physical model. Irena Roterman designed the mathematical model (based on the 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. [[CrossRef](#)] [[PubMed](#)]
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. [[CrossRef](#)] [[PubMed](#)]
3. Weissman, J.S.; Kim, P.S. A kinetic explanation for the rearrangement pathway of BPTI folding. *Nat. Struct. Biol.* **1995**, *2*, 1123–1130. [[CrossRef](#)]
4. Weissman, J.S.; Kim, P.S. Efficient catalysis of disulfide bond rearrangements by protein disulphide isomerase. *Nature* **1993**, *365*, 185–188. [[CrossRef](#)] [[PubMed](#)]
5. Milbrandt, A.G.; Boulegue, C.; Moroder, L.; Renner, C. The two cysteine-rich head domains of minicollagen from Hydra nematocysts differ in their cystineframework and overall fold despite an identical cysteine sequence pattern. *J. Mol. Biol.* **2005**, *354*, 591–600. [[CrossRef](#)] [[PubMed](#)]
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  $\beta$ -KTx6 subfamily. *Proteins* **2005**, *61*, 1010–1023. [[CrossRef](#)] [[PubMed](#)]
7. Konieczny, L.; Brylinski, M.; Roterman, I. Gauss-Function-Based Model of Hydrophobicity Density in Proteins. *Silico Biol.* **2006**, *2*, 15–22.
8. 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. [[CrossRef](#)]



9. 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*, 315–346. [PubMed]
10. 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. [CrossRef] [PubMed]
11. Kauzmann, W. Some factors in the interpretation of protein denaturation. *Adv. Protein Chem.* **1959**, *14*, 1–63. [PubMed]
12. Levitt, M. A simplified representation of protein conformations for rapid simulation of protein folding. *J. Mol. Biol.* **1976**, *104*, 59–107. [CrossRef]
13. Kulback, S.; Leibler, R.A. On information and sufficiency. *Ann. Math. Stat.* **1951**, *22*, 79–86. [CrossRef]
14. 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 <sup>13</sup>C/<sup>15</sup>N NMR spectroscopy. *Biochemistry* **1997**, *35*, 7684–7691. [CrossRef] [PubMed]
15. Artymiuk, P.J.; Blake, C.C. Refinement of human lysozyme at 1.5 Å resolution analysis of non-bonded and hydrogen-bond interactions. *J. Mol. Biol.* **1981**, *152*, 737–762. [CrossRef]
16. Wlodawer, A.; Borkakoti, N.; Moss, D.S.; Howlin, B. Comparison of two independently refined models of ribonuclease-A. *Acta Crystallogr. Sect. B* **1986**, *42*, 379–387. [CrossRef]
17. 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. USA* **1994**, *91*, 2915–2919. [CrossRef] [PubMed]
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. [CrossRef] [PubMed]
19. Cho, H.S.; Leahy, D.J. Structure of the extracellular region of HER3 reveals an interdomain tether. *Science* **2002**, *297*, 1330–1333. [CrossRef] [PubMed]
20. Schrag, J.D.; Cygler, M. 1.8 Å refined structure of the lipase from *Geotrichum candidum*. *J. Mol. Biol.* **1993**, *230*, 575–591. [CrossRef] [PubMed]
21. POPS, Version 1.0.6, software for calculate solvent accessible surface areas of proteins; The Francis Crick Institute: London, UK; Available online: <http://mathbio.nimr.mrc.ac.uk/wiki/POPS> (accessed on 23 February 2016).
22. 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. *Nat. Struct. Biol.* **1995**, *2*, 871–875. [CrossRef]
23. 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. [CrossRef] [PubMed]
24. 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. [CrossRef] [PubMed]
25. Verma, D.; Jacobs, D.J.; Livesay, D.R. Changes in lysozyme flexibility upon mutation are frequent, large and long-ranged. *PLoS Comput. Biol.* **2012**, *8*, e1002409.
26. Wang, C.; Chen, S.; Wang, X.; Wang, L.; Wallis, A.K.; Freedman, R.B.; Wang, C.C. Plasticity of human protein disulfide isomerase: Evidence for mobility around the X-linker region and its functional significance. *J. Biol. Chem.* **2010**, *285*, 26788–26797. [CrossRef] [PubMed]
27. Tian, G.; Kober, F.X.; Lewandrowski, U.; Sickmann, A.; Lennarz, W.J.; Schindelin, H. The Catalytic Activity of Protein-disulfide Isomerase Requires a Conformationally Flexible Molecule. *J. Biol. Chem.* **2008**, *283*, 33630–33640. [CrossRef] [PubMed]
28. Yarden, Y.; Pines, G. The ERBB network: At last, cancer therapy meets systems biology. *Nat. Rev. Cancer* **2012**, *12*, 553–563. [CrossRef] [PubMed]
29. Anfinsen, C.B. Principles that govern the folding of protein chains. *Science* **1973**, *181*, 223–230. [CrossRef] [PubMed]

