Symmetry and Dissymmetry in Protein Structure—System-Coding Its Biological Specificity

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

[1] Department of Bioinformatics and Telemedicine, Medical College, Jagiellonian University, Łazarza 16, 31-530 Krakow, Poland; mateusz.banach@uj.edu.pl
[2] Medical Biochemistry, Medical College, Jagiellonian University, Kopernika 7, 31-034 Kraków, Poland; mbkoniec@cyf-kr.edu.pl
* Correspondence: myroterm@cyf-kr.edu.pl

check for
updates

Abstract: The solenoid is a highly ordered structure observed in proteins, characterized by a set of symmetries. A group of enzymes—lyases containing solenoid fragments—was subjected to analysis with focus on their distribution of hydrophobicity/hydrophilicity, applying the fuzzy oil drop model. The model differentiates between a monocentric distribution hydrophobic core (spherical symmetry—mathematically modeled by a 3D Gaussian) and linear propagation of hydrophobicity (symmetry based on translation of structural units, i.e., chains—evident in amyloids). The linearly ordered solenoid carries information that affects the structure of the aqueous solvent in its neighborhood. Progressive disruption of its symmetry (via incorporation of asymmetrical fragments of varying size) appears to facilitate selective interaction with the intended substrate during enzymatic catalysis.

1. Introduction

The structure referred to as "solenoid" was, on initial recognition, assigned to the "new fold" category in Protein Data Bank (PDB). The solenoid is defined as the structure constructed by tandem repeated units of similar secondary structure. The specific solenoid form, constructed by short beta-fragments generating elongated tubes, are selected for analysis in this paper. This was due to the presence of a heretofore unknown beta-helix conformation. This name itself pertains to a sequence of short beta-strands, arranged in a tubular frame that superficially resembles a helix. This effect is so peculiar that pectate lyase (2PEC) was even described as a "surprising new fold" [1]. In a solenoid, a short structural motif composed of beta-strands is cyclically replicated along an axis perpendicular to its surface, producing a tubule [2–5]. The cross-section of this tubule may be circular, flattened (oval), square-like, or rectangular, possibly with deformed edges. Regardless, the structure adopts an elongated, regular form which may be described as a cylindrical micelle [6]. Solenoid-like fragments have been identified in antifreeze proteins [7] responsible for counteracting the formation of ice crystals, and in lyases—a group of enzymes that cleave chemical bonds. This can be readily observed by searching PDB for occurrences of the "solenoid" keyword [8] (although note that some lyases lack solenoid fragments and will not be included in the presented analysis).

Our study of antifreeze proteins [7] and lyases is based on the fuzzy oil drop model, which categorizes the distribution of hydrophobicity in protein molecules [9,10]. Its predecessor—the classic oil drop model [11]—asserts the presence of a centrally located hydrophobic core, which is sheathed by hydrophilic residues facing the water environment. In the fuzzy oil drop model this

discrete (two-layer) distribution is replaced with a continuous hydrophobicity gradient, stretching from the center all the way to the molecular surface and mathematically modeled by a 3D Gaussian. The Gaussian form is then assumed to represent the theoretical (or "idealized") distribution of hydrophobicity for a given protein. Its values peak at the center and then decrease along with distance from the center, reaching almost 0 on the surface (at a distance of 3σ in each principal direction). Of course, the actual (observed) distribution of hydrophobicity does not always correspond to this model with perfect accuracy. Comparing both distributions (theoretical vs. observed) provides clues regarding the structural order (or disorder) present in the protein. Discordances may be either global or local. In the former case, the protein is said to lack a hydrophobic core in the sense of the fuzzy oil drop model. In the latter case, areas where both distributions diverge from each other often coincide with sites of biological activity. Excess hydrophobicity on the protein surface marks complexation sites [12,13], while local hydrophobicity deficiencies are often found in enzymatic active sites [14]. An entirely different distribution of hydrophobicity, lacking a monocentric core, is present in amyloids [6,15], where local minima and maxima propagate along the axis of the fibril. This paper presents the continuation of the discussion concerning the linear distribution (in contract to centric one) which is observed in amyloids [6,15]. The linear propagation as observed in amyloids lacks a natural "stop" signal and may continue indefinitely, which explains the unrestricted length of amyloids. The detailed discussion focused on amyloids can be found in [16]. This paper can be treated also as supplementary to the discussion of 3D structure of antifreeze proteins, where solenoid is also present [7]. The common background for these groups of proteins is the linear propagation of bands of different hydrophobicity.

As it turns out, linear propagation of hydrophobicity can also be observed in biologically active proteins, and specifically in various solenoid fragments, where consistent "bands" of hydrophobicity propagate along a given axis. Given the lack of a natural mechanism to arrest this propagation, a solenoid must be equipped with a specific "cap" [17]. Such fragments are indeed observed in proteins, and their role is to adapt the emerging structure to the 3D Gaussian by exposing hydrophilic residues on its water-facing side, while maintaining compatibility with the solenoid on the underside. Antifreeze proteins which contain solenoids provide examples of such structures [18].

The analysis presented in this work focuses on lyases which contain solenoid fragments. Unlike antifreeze proteins, lyases also include enzymatic active sites. Considering the fuzzy oil drop model, antifreeze proteins—by virtue of their presence—disrupt the structural ordering of the aqueous medium which would be required for an ice crystal to emerge. It can be said that an antifreeze protein generates a distinct "field" which affects the structuralization of water. By analogy, the solenoid fragment present in lyases may be treated as the source of a "field" which promotes enzymatic catalysis. Our work provides a description of this field and its relationship with the active site of solenoid-containing lyases.

2. Materials and Methods

The fuzzy oil drop model was applied in the analysis of proteins belonging to the lyase class, whose structure includes solenoid fragments.

2.1. Proteins Subjected to Analysis

Table 1 lists proteins which have been selected for analysis. From among solenoid-containing lyases, we have selected those for which PDB provides a description of catalytic residues [8]. Lyases have been divided into groups depending on their substrate.

Table 1. List of proteins subjected to analysis. Proteins listed in boldface are discussed in detail in this paper.

PDB ID	Organism	Ref.
EC 4.2.1.1. Carbonic anhydrase $H_2CO_3 = CO_2 + H_2O$		
1QRG Zn2+	Methanosarcina thermophila	[18]
1QRM Zn2+	Methanosarcina thermophila	[18]
2FKO Zn2+	Pyrococcus horikoshi	[19]
EC 4.2.2.2. Pectate lyaseEliminative cleavage of pectate to give oligosaccharides with 4-deoxy-alpha-D-gluc-4-enuronosyl groups at their non-reducing ends		
1PLU Lu	Erwinia chrysanthemiplant	[20]
1BN8 Ca2+	Bacillus subtilis	[21]
2BSP Ca2+	Bacillus subtilis	[22]
1JTA	Erwinia chrysanthemi	[23]
1JRG	Erwinia chrysanthemi	[23]
1O88 Ca2+	Erwinia chrysanthemi	[24]
1OOC	Erwinia chrysanthemi	[25]
1PXZ	Juniperus ashei	[26]
1RU4 2Ca2+	Erwinia chrysanthemi	[27]
EC 4.2.2.10. Pectin lyase Eliminative cleavage of pectin to give oligosaccharides with terminal 4-deoxy-6-methyl-alpha-D-galact-4-enuronosyl groups		
1IDJ	Aspergillus niger	[28]
EC 4.2.1.59 3-hydroxyacyl-[acyl-carrier-protein] dehydratase $(3R)$-3-hydroxyacyl-[acyl-carrier protein] = a *trans*-2-enoyl-[acyl-carrier protein] + H_2O		
3DP3	Helicobacter pylori	[29]
3ED0	Helicobacter pylori	[30]
EC 3.4.11.18 Methionyl aminopeptidase Pectate II		
2QXZ	xanthomonas campestris	[31]
2QY1	xanthomonas campestris	[31]

2.2. Model Description

The fuzzy oil drop model has been thoroughly described in numerous publications [9,10]. Consequently, we will limit our description to a brief recapitulation of its core aspects, relevant from the point of view of the presented study.

The target protein molecule is encapsulated by a 3D Gaussian form, whose dimensions (σ coefficients) are each equal to 1/3 of the distance from the molecule's center to the most distal atom along each principal axis. The center of the coordinate system coincides with the geometric center of the molecule. Under such conditions, values of the 3D Gaussian are assumed to reflect theoretical hydrophobicity at any point within the protein body. This distribution is referred to as the theoretical distribution (denoted T).

Actual proteins do not always conform to the theoretical distribution. The observed distribution (denoted O) is produced by inter-residual interactions, which, in turn, depend on the distance between residues and on their intrinsic hydrophobicity [32]. Each amino acid is represented by its so-called effective atom (averaged-out position of all atoms comprising the given residue). Computing values of the aforementioned Gaussian for all effective atoms enables us to directly compare both distributions (T and O). The distance between these distributions can be quantitatively expressed using the concept of divergence entropy (distance entropy) proposed by Kullback and Leibler [33]. The value of D_{KL} represents the distance between T and O (denoted O/T). Like any other measure of entropy, this value cannot be interpreted on its own—it requires a reference, which in our case is the so-called unified distribution (R). This distribution assigns a value of 1/N to each residue in the protein chain (N being the total number of residues), indicating that no concentration of hydrophobicity exists at any point in the protein body. Comparing T/O with O/R tells us whether the observed distribution more closely approximates the theoretical distribution or the random distribution. To avoid dealing with two separate values, we introduce the following relative distance (RD) coefficient:

$$RD_{TR} = \frac{O/T}{(O/T + O/R)}$$

Another potential reference (other than R) is a distribution which corresponds to the intrinsic hydrophobicity of each amino acid. This distribution, denoted H, is used to determine whether the observed distribution is dominated by the intrinsic properties of each participating residue. Since relative distance may be measured in two different systems (T–O–R and T–O–H), we further distinguish two distinct values of RD: RD_{TR} and RD_{TH}. Visual depiction of the former case is presented in Figure 1 (reduced to a single dimension for the sake of clarity).

$$RD_{TH} = \frac{O/T}{(O/T + O/H)}$$

D_{KL} and RD may be calculated for any arbitrarily selected structural unit: protein complex, individual chain, or selected domain. In each case, a suitable 3D Gaussian is plotted to encapsulate the target structure. It is also possible to analyze the status of chain fragments (e.g., a single beta-sheet or an individual secondary structure)—following prior normalization of T_i, O_i, R_i, and H_i values for all residues comprising the target fragment. When studying individual fragments, no distinct ellipsoid capsule is formed. The resulting value of RD then tells us how the given fragment affects the larger structure for which T_i, O_i, R_i, and H_i had previously been computed.

In our work we characterize complete structural units as well as selected fragments of such units. Additionally, we compute correlation coefficients for pairs of values: observed vs. theoretical hydrophobicity; observed vs. intrinsic hydrophobicity and theoretical vs. intrinsic hydrophobicity. The resulting values express the extent to which intrinsic hydrophobicity determines the theoretical hydrophobicity for the target segment, as well as the involvement of each residue in generating a monocentric hydrophobic core. High values of the observed vs. theoretical coefficient indicates that the given residues align themselves to the hydrophobic core, while high values of the observed vs. intrinsic coefficient suggest that intrinsic hydrophobicity dominates the conformation of the target segment.
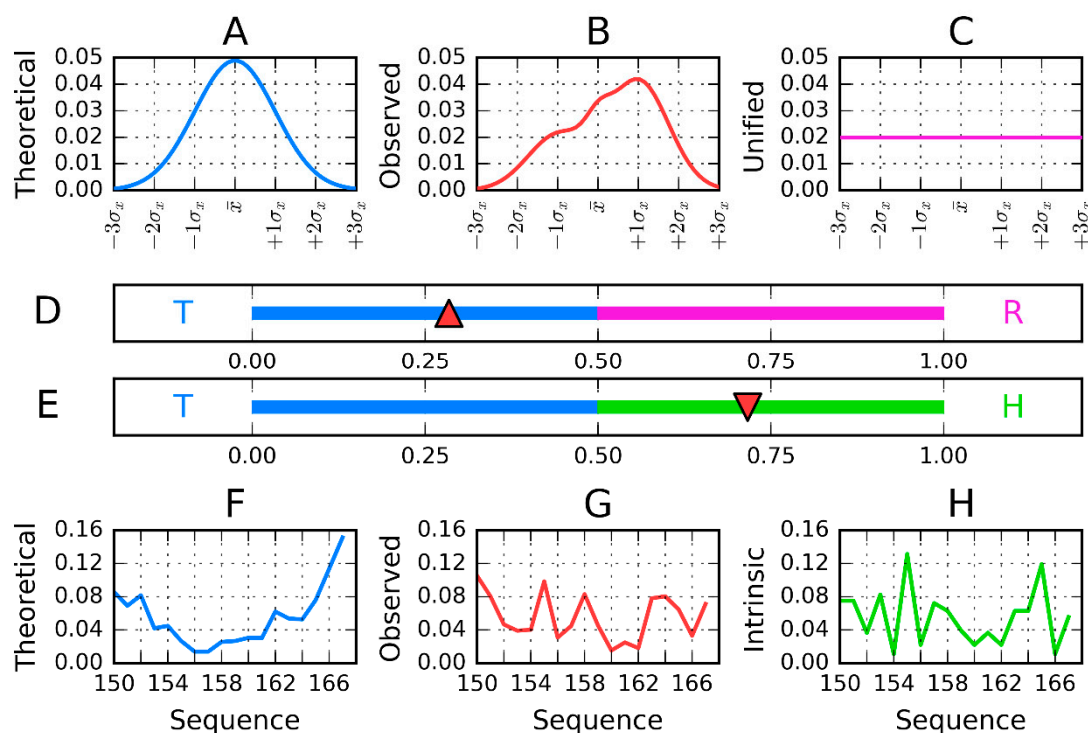
Figure 1. Theoretical—called T (**A**)—and unified—called R (**C**) distributions applied as reference boundary cases for the observed distribution—called O (**B**). The computed value of RD for the T–O–R system is plotted on the horizontal axis (**D**). In this case, RD = 0.284 (T–O–R). Plotting the observed—called O (**F**), theoretical—T called (**E**), and intrinsic hydrophobicity—called H (**G**) distributions results in RD = 0.716 (called T–O–H) as shown on the horizontal axis (**H**), indicating that the observed distribution is distant from the theoretical distribution and close to the H distribution (intrinsic hydrophobicity of each residue). Plots F, G, and H depict the 150−167 fragment of 1PLU. The profiles shown in A, B, and C—generated individually for this figure.

2.3. Calculation Procedure

The detailed description of fuzzy oil drop model is available in [10]. The short presentation of the procedure aimed on RD calculation is given:

a. the representation of the molecule with X, Y, Z coordinates (as available in PDB) given for each atom is changed to the effective atom representation. Effective atom—position of averaged coordinates of atoms belonging to certain amino acids;

b. the molecule under consideration shall be oriented in coordinate system as follows: A—the geometric center localized in origin of coordinate system; B—the X-axis oriented according to two most distant effective atoms in molecule; C—the Y-axis oriented according to most distant effective atoms—distance limited to Y–Z-plane;

c. the T-distribution–sigma parameters calculated as 1/6 of largest distance along each axis independently (according to 3-sigma rule for normal distribution);

d. the second parameter (position of maximum) is calculated as a mean value of effective atoms positions;

e. the Ti values are calculated as values of the 3D Gauss function in the position of each effective atom;

f. the Oi values express the hydrophobic interaction between effective atoms—thus its value depends on the distance between two interacting residues and their intrinsic hydrophobicity taken according to the selected scale. The scale presented in [10] was taken in our calculations. The function expressing this interaction was taken according to Levitt function [32].

g.   Both T and O distribution shall be normalized—sum of all Ti and Oi values shall be equal to 1.0. This operation allows the comparison of T and O distributions;

h.   The Kullback Leibler divergence entropy (D_{KL}) is applied to measure the difference between two profiles T and O [33]. Since the value of D_{KL} entropy cannot be interpreted (entropy category) the second reference distribution is defined (Figure 1)—the unified one with Ri values for all residues equal to 1/N. The comparison of O distribution versus two reference distributions T and R may suggest the similarity to one of them. The closeness to one of them is expressed by RD parameter, the calculation of which is described above.

The procedure of (D_{KL}) (and in consequence RD) calculation may be applied for structural units: protein complexes, one-chain proteins, domains. The independent 3D Gauss function shall be generated for each structural unit. The (D_{KL}) values (and in consequence RD) may be calculated for selected fragments of polypeptide chain after prior normalization of distributions for this selected chain fragment. It allows declare the status of individual units as well as the status of selected chain fragment.

2.4. Antifreeze and Amyloids Versus Lyases

The three groups of proteins—antifreeze type III [7], amyloids [16] and lyases (discussed in this paper)—represent common structural characteristics. This is the presence of structural forms with linear propagation of bands of different hydrophobicity level. This similarity is discussed in the Section 4 in this paper.

3. Results

For detailed analysis, we have selected proteins representative of various enzyme classes within EC 4.2.2. These proteins, in addition to solenoid fragments, also include disordered fragments in their N- and C-terminal sections, helical "stoppers" preventing further propagation of solenoids, and loops which appear to form part of their active sites. Some of these proteins (EC 4.2.1.1) are further characterized by the presence of a helix which runs parallel to the solenoid (replacing the disordered N- and C-terminal fragments). Proteins which belong to the Lyase II family contain solenoids which do not form a full twist of the beta-helix, but merely half of it. These proteins are also included in our analysis since they include repeating, coherent beta-strands.

Recurring beta-strands form a beta-sheet which constitutes one of the side walls of the solenoid tubule. As shown in [15], local maxima and minima of hydrophobicity propagate linearly along the axis of the solenoid (or amyloid fibril) [6,15]. In accordance with PDBSum data [34], we define the solenoid as a collection of beta sheets which represent its walls.

In principle, such linear propagation may continue indefinitely and therefore must be arrested by a custom "stopper". In most of the presented proteins, this role falls to a short helical fragment positioned athwart the solenoid [17]. This helix is amphipathic, with its hydrophilic residues facing the aqueous environment and the "underside" directly adjacent to the solenoid. Comprehensive results reveal the status of the solenoid itself as well as of the surrounding "caps". In proteins which include a parallel helix we also provide a characterization of this helix, or of any corresponding external fragments parallel to the solenoid.

The solenoid is surrounded by helices, loops, or unstructured fragments. We jointly refer to such fragments as the "envelope" and theorize that they act to increase the solubility of the protein under consideration. Proteins selected for our study exhibit a gradual increase in the size and sophistication of the envelope structure.

The catalytic center is defined as the immediate neighborhood of catalytic residues (if a catalytic residue is present at position *i*, we define "immediate neighborhood" as the fragment between $i - 10$ and $I + 10$). Additionally, extended loops above the catalytic center are also considered to belong to the catalytic center due to their close proximity and stepwise development. Table 2 provides a summary of all analyzed enzymes.

Table 2. RD values computed for complexes, proteins, and solenoid fragments. RD values are also listed for terminal sections (caps) bracketing the solenoid, as well as for any other notable structural features (e.g., parallel helix). The description refers to monomers of each listed protein. Underlined items correspond to proteins selected for detailed analysis in this paper. To distinguish the fragments included/excluded from calculations in ALL position the and * symbols were used.

PDB ID IONS	CHAIN	SOLENOID	ENVELOPE	STOP
EC 4.2.1.1				
1QRG	0.418	0.506	Helix—0.373	0.298
1QRM	0.476	0.601	Helix—0.426	0.313
2FKO	0.457	0.526	Helix—0.348	0.399
EC 4.2.2.2				
1PLU	0.654	0.627	0.594	0.362
1BN8	0.684	0.416	0.469 0.607	0.314
2BSP	0.688	0.729	0.667	0.304
1JTA	0.632	0.603	0.457	0.376
1JRG	0.649			0.310
1O88	0.650	0.608	0.667	0.304
1OOC	0.643	0.719	0.437	0.369
1PXZ	0.591	0.620	0.509	0.503
1RU4	0.685	0.705	0.724 *	0.752
EC 4.2.2.10				
1IDJ	0.722	0.769	0.634	0.255
EC 4.2.2.59				
3DP3	0.463	0.	0.770	0.343
3ED0	0.457	0.431	0.778	0.333
Lyase II				
2QXZ	0.675	0.693	0.442	0.532
2QY1	0.674	0.665	0.444	0.494

Enzymes which comprise the EC 4.2.1.1 class are all similar in status, despite being derived from various organisms. The whole molecule is characterized by RD < 0.5, interpreted as the presence of a monocentric hydrophobic core, even though the protein also contains a solenoid fragment. When considering the solenoid by itself, we arrive at RD > 0.5—as expected given that within this fragment the monocentric core is replaced by linear propagation of local maxima and minima. In contrast, the "stop" sequence and the parallel helix both comply with the model (RD < 0.5), indicating that they mediate contact with the aqueous environment, and that the molecule as a whole contains a prominent hydrophobic core even though a solenoid fragment is also present.

Lyases which belong to class EC 4.2.2.2 diverge from the theoretical model. In nearly all these molecules the solenoid does not conform to the theoretical distribution (1BN8 being the sole exception). "Stop" sequences generally provide hydrophobic closure for the solenoid, except in 1PXZ where their RD values slightly exceed 0.5. Parallel fragments—corresponding to the location of the parallel helix in EC 4.1.1.1—adopt variable conformations, with a short helix and several disordered fragments in evidence. The RD values listed in Table 2 are computed for all these fragments treated as a single unit.

The representative of class EC4.2.2.10 exhibits accordance with the model only within its "stopper" helix.

Proteins which belong to class EC4.2.2.59 are derived from the same source organism, but in their crystal structure they interact with various ligands acting as inhibitors [29,30]. From the point of view of FOD analysis, these proteins are not structurally different. Both remain consistent with the model; additionally, they are the only two proteins in the study set where the solenoid section itself has been identified as consistent with the model. As expected, the "stopper" sequences are also accordant. The reason behind the unusual status of the solenoid section (compared to other proteins discussed in

this paper) may be related to the fact that the repetitive beta-sheet forms only one half of the "tubule" which sheathes a centrally located helix. Additionally—unlike other presented proteins the catalytic residue forms part of the central helix.

Two proteins classified under the "Lyase II" header differ only with respect to their "stopper" sequences. Surprisingly, in 2QXZ this sequence diverges from the theoretical distribution, which may be due to the presence of an adjacent chain (the crystal structure is made up of dimers). Both helices contain residues which mediate complexation, and this may affect their FOD status.

The following sections characterize the status of proteins representing each class of enzymes.

3.1. 2FKO—Representative of EC 4.2.1.1

2FKO, a lyase and a member of the carbonic anhydrase family, contains a solenoid section with a characteristic parallel helix, very similar to the one observed in antifreeze proteins. By analyzing the parameters listed in Table 3, we may conclude that the helix introduces a distribution of hydrophobicity which favors contact with the aqueous environment, thereby ensuring solubility. On the other hand, each beta-sheet as well as the entire solenoid (i.e., both sheets taken together) remain divergent from the model. Helical "caps" at either end of the solenoid are consistent with the theoretical distribution of hydrophobicity, which can be explained by their amphipathic properties (shown in Figure 2C—pink fragment, where the red line is closely aligned with the blue line, as well as in Table 3).

Table 3. FOD characterization of pectate lyase (2FKO). RD coefficients determine the relative distance between the observed distribution (O) and two pairs of boundary cases (theoretical vs. unified—T–R—and theoretical vs. intrinsic—T–H). Fragments which diverge from the monocentric theoretical model are listed in boldface. To identify beta sheets, the table lists the earliest fragments belonging to each sheet. Identification of secondary structure follows PDBSum criteria [34]. *—only numbers of residues present in first beta-strand in beta-sheet are given to identify particular beta-sheet.

| Fragment	Structural Properties	RD		Correlation Coefficient		
		T–O–R	T–O–H	HvT	TvO	HvO
2–144	SOLENOID	0.526	0.468	0.295	0.516	0.775
2–4	Beta-sheet	0.542	0.703	0.408	0.462	0.804
12–14 *	Beta-sheet	0.386	0.297	0.374	0.639	0.848
18–20	Beta-sheet	0.643	0.544	0.145	0.516	0.658
STOP						
1–11	STOP N-term	0.305	0.143	0.427	0.790	0.767
137–144	STOP signal	0.399	0.163	0.531	0.652	0.794
ENVELOPE						
144–173	Helix C-term	0.348	0.160	0.423	0.694	0.785
CATAL. Center						
60–80 Y70	Catalytic residue	0.361	0.423	0.392	0.733	0.766
38–94	surrounding	0.376	0.299	0.363	0.668	0.750
ION BINDING						
41–51		0.372	0.441	0.588	0.682	0.885
60–70	Zn2+ (65)	0.367	0.380	0.289	0.728	0.609
82–87	Zn2+ (87)	0.419	0.246	0.544	0.552	0.885
100–110		0.502	0.283	0.425	0.308	0.825
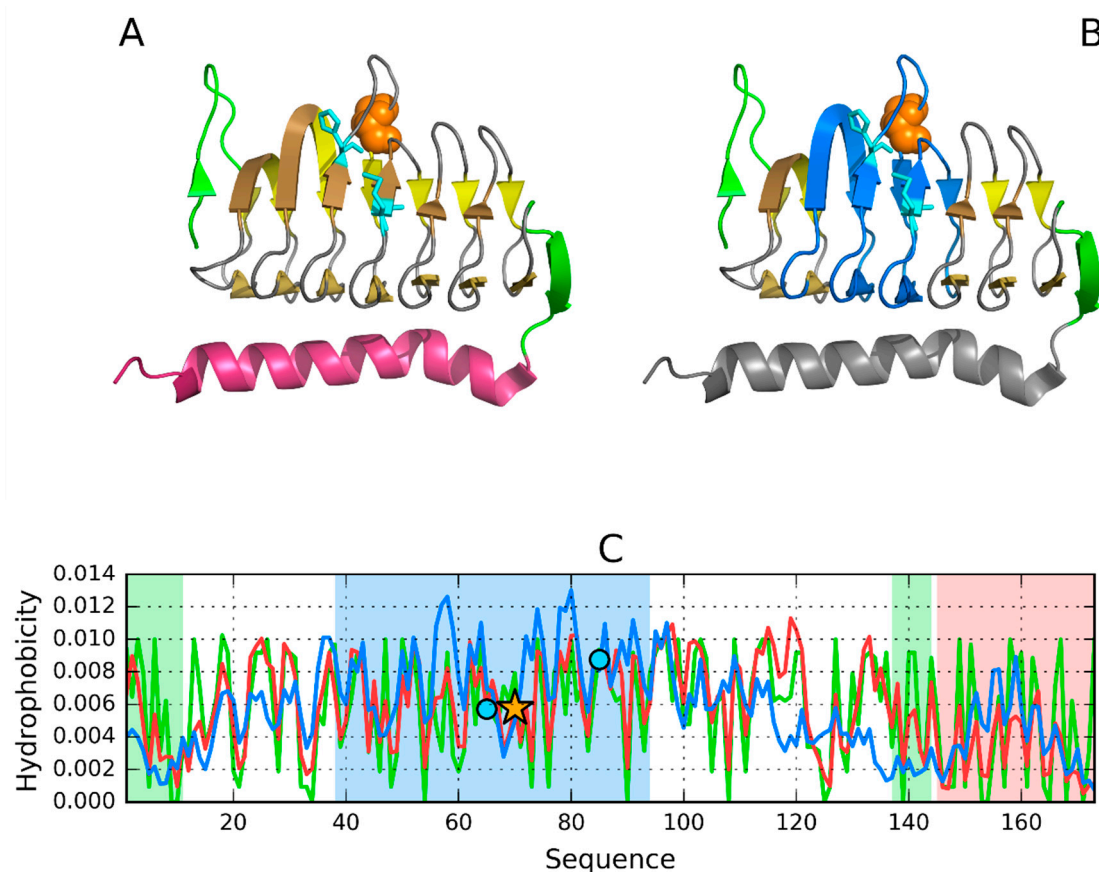
Figure 2. Analysis of 2FKO structure. Differentiation of fragments is consistent with the data presented in Table 3. Catalytic residues—orange spheres and an orange star in Figure 2C. Colors distinguishing fragments correspond to those shown in Figure 2C. Pink—C-terminal fragment, identified as accordant with the fuzzy oil drop model. Green—"stopper" fragments also accordant with the fuzzy oil drop model. The solenoid fragment in Figure 2A is yellow; turquoise—positions of Zn ions shown in Figure 2C as blue dots. In Figure 2B the blue fragment represents catalytic residue environment; C—T (blue), O (red), and H (green) profiles in 2FKO. Yellow star in C—catalytic residue. Blue circles in C—position of Zn2+ complexation.

When interpreting the distributions shown in Figure 2 (fragments with green background in C and green fragments in A and B) we can observe accordance in the N- and C-terminal sections. Within the solenoid itself the observed distribution follows a sinusoid pattern, which is due to cyclic repetition of a specific unit sequence. The beginning and end of the solenoid diverge significantly from the theoretical distribution. The placement of Zn^{2+} ions, as shown in Table 3, likely affects the distribution of hydrophobicity producing locally accordant conditions.

The neighborhood of the enzymatically active residue conforms to the aforementioned sinusoid pattern and does not form a central hydrophobic core. This distribution corresponds to linear propagation of alternating minima and maxima.

In this particular case, linear distribution is not in conflict with the theoretical distribution due to the hydrophilic nature of exposed loops, along with the central (within each beta-strand) location of hydrophobic residues. The effect is further reinforced by the amphipathic properties of the adjacent helix, producing a molecule which is largely consistent with the fuzzy oil drop model (Table 3), even though the solenoid itself does not contain a centralized hydrophobic core. When considered on its own, the solenoid diverges from the model in favor of a linear distribution of hydrophobicity, as shown in Figure 2C. (fragment 10–145) and in Figure 2A,B—yellow fragments.

The 60–80 fragment, which comprises two beta-strands participating in two distinct beta sheets, along with the catalytic residue, remains locally accordant.

3.2. 1PLU—Representative of EC 4.2.2.2

The pectate lyase (EC 4.2.2.2) group is represented by 1PLU. As a whole, this protein lacks a central hydrophobic core (Table 4); however, its helical "stopper" remains accordant, likely due to its role as a means of preventing unchecked propagation of local hydrophobicity maxima.

Table 4. FOD characterization of pectate lyase (1PLU). RD coefficients determine the relative distance between the observed distribution (O) and two pairs of boundary cases (theoretical vs. unified—T–R—and theoretical vs. intrinsic—T–H). Fragments which diverge from the monocentric theoretical model are listed in boldface. To identify beta sheets (*), the table lists the earliest fragments belonging to each sheet. Identification of secondary structure follows PDBSum criteria [34]. Envelope positions marked by double asterisks represent fragments with residues engaged in SS-bonds eliminated, or N-terminal fragment (usually very flexible). To distinguish the fragments included/excluded from calculations in ALL position the * and ** symbols were used.

Fragment	Structural Properties	RD		Correlation Coefficient		
		T–O–R	T–O–H	H/T	T/O	H/O
SOLENOID		0.627	0.543	0.250	0.214	0.800
19–23 *	Beta-sheet	0.756	0.590	0.210	0.278	0.700
80–86 *	Beta-sheet E	0.356	0.335	0.648	0.720	0.853
118–122 *	Beta-sheet	0.521	0.523	0.431	0.278	0.772
STOP						
25–37	Helix	0.371	0.314	0.775	0.664	0.861
286–297		0.348	0.306	0.583	0.710	0.782
ENVELOPE						
1–17	Random Coil	0.455	0.515	0.031	0.590	0.456
37–50	Random Coil	0.432	0.251	0.252	0.642	0.754
61–70	Helix	0.480	0.604	0.564	0.623	0.774
282–358	Random Coil	0.511	0.280	0.326	0.523	0.763
286–297	C-term stop	0.348	0.306	0.583	0.710	0.782
327–336	helix	0.428	0.273	0.498	0.489	0.965
All above		0.571	0.388	0.265	0.517	0.706
1–26		0.536	0.413	0.105	0.529	0.543
6–26 **		0.495	0.330	0.165	0.646	0.513
38–50		0.393	0.225	0.351	0.717	0.761
57–73		0.600	0.614	0.096	0.315	0.808
57–73 **		0.441	0.501	0.536	0.724	0.654
313–353		0.587	0.434	0.243	0.414	0.795
313–345 *		0.467	0.285	0.266	0.523	0.791
All above		0.626	0.439	0.186	0.426	0.726
All ** above		0.467	0.285	0.656	0.844	0.852
CATAL. C.						
96–248		0.617	0.556	0.330	0.380	0.780
58–80	loop	0.679	0.660	−0.179	−0.185	0.848
123–132	loop	0.596	0.430	0.239	0.223	0.871
150–167	loop	0.671	0.716	−0.047	0.205	0.696
195–202	loop	0.111	0.407	0.029	0.725	0.487
126–136	131D—Beta fr.	0.572	0.436	0.321	0.491	0.897
213–223	218R—Beta fr.	0.471	0.384	0.498	0.549	0.930
117–224		0.621	0.525	0.314	0.402	0.774
117–224 58–80		0.720	0.650	0.239	0.372	0.782

Compared to the previously discussed protein, in this case the long, parallel helix is replaced by disordered N- and C-terminal sections, which contain only short helical fragment. Additionally, the active site is not exposed and appears shielded by two protruding loops. These loops are referred to as "catalytic" due to the presence of catalytic residues (Table 4). One of them diverges significantly from the theoretical model, while another is locally accordant. Each contains one catalytic residue. Taken as a whole, the section which includes both catalytic residues along with a fragment of the solenoid is highly discordant from the model.

From among four beta sheets (parts of solenoid), one is consistent with the model while the remaining ones exhibit rather high RD values. The solenoid as a whole is also discordant.

The "stop" fragment, which is helical, follows the theoretical distribution—as expected, given its role (exposure of hydrophilic residues on the outside, and internalization of hydrophobic residues to ensure compatibility with the solenoid). The set of loose fragments surrounding the solenoid is also discordant; however certain elements of the solenoid's "sheath" exhibit local accordance, as demonstrated by their high correlation coefficients.

Catalytic residues are found in the beta-strands which comprise the solenoid. Their status varies. The loops which bracket the catalytic center are characterized by high RD values, except for the one at 195–202. The entire neighborhood of the catalytic center, treated as a single unit (58–80 and 117–224), is highly disordered (in the sense of the fuzzy oil drop model). Figure 3 and Figure 5 provide a visual depiction of the catalytic loops and the so-called outlying loops.

In summary, it is evident that the solenoid does not follow the distribution of hydrophobicity stipulated by the fuzzy oil drop model: instead of a monocentric core we are faced with a linear distribution. The catalytic active site is also discordant, whereas the collection of outlying loops which "sheathe" the solenoid remains consistent with the model and may be regarded as an "envelope", ensuring solubility of the protein. Their role therefore corresponds to that of the parallel helix in 2FKO.

Theoretical and observed hydrophobicity distribution profiles (Figure 3) reveal that the N- and C-terminal fragments remain accordant. In contrast, the fragments directly adjacent to the active site diverge from the model (Figure 3C—green background and green fragments in Figure 3A,B).

The solenoid fragment (Figure 3A—yellow fragments) corresponds to a sinusoid distribution of hydrophobicity, lacking a monocentric core. Both distributions diverge greatly in fragments which correspond to catalytic loops (differences between blue and red lines in Figure 3C).

The linear distribution of hydrophobicity in the beta-sheet which begins with the fragment at 19–23 is the source of high values of RD given in Table 4. Clear differences between O and T can be observed, with the observed distribution approximating intrinsic hydrophobicity values. Additionally, local minima are arranged in a periodic pattern, reflecting the conformation of successive beta-strands, with each local minimum located in an area where the model predicts a hydrophobicity peak instead.

It is interesting to note the structure of the active site. We have singled out fragments which contain catalytic residues—beta-strands which form part of the solenoid—as well as loose coils bracketing the active site. The entire system, composed of catalytic beta-strands and the outlying loops, is significantly discordant and dominated by the intrinsic properties of each residue (Figure 3).
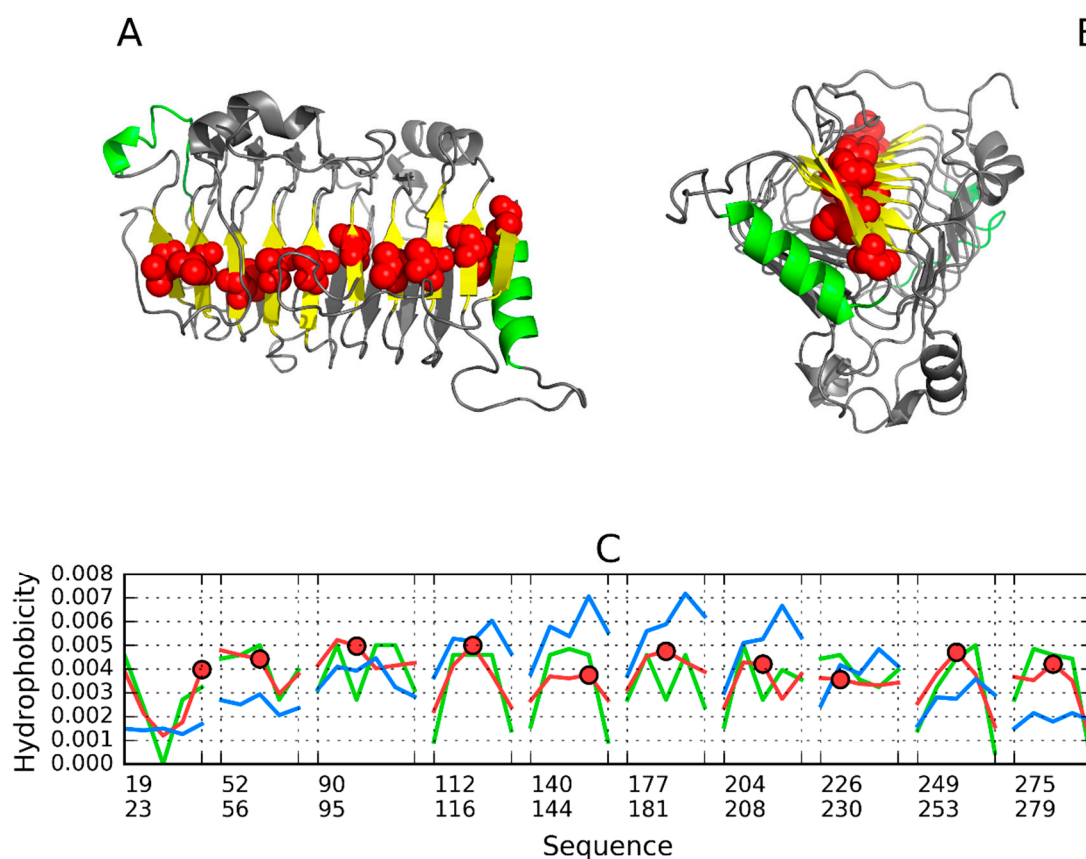
Figure 3. Characteristics of 1PLU. A and B—3D presentation of 1PLU. Residues distinguished as red reveal linear propagation of a highly hydrophobic band, which strongly diverges from the expected hydrophobicity. This is further visualized in the profiles shown below (T—blue, O—red, and H—green hydrophobicity distributions). Hydrophobicity distributions are plotted for beta-strands which comprise the beta-sheet of 1PLU (the first such fold is located at 19–23). Notably, a local minimum is present where a local maximum would be expected. Good alignment between O and H is also evident for the beta portion of the solenoid, highlighting the discordance of residues distinguished as red dots. Green fragments—stoppers preventing infinite elongation of the solenoid.

3.3. Pectin Lyase (1IDJ)—Representative of EC 4.2.2.10

Table 5 lists RD values for 1IDJ (pectin lyase), along with the corresponding correlation coefficients. This protein contains four disulfide bonds. As expected, the solenoid exhibits a linear distribution of hydrophobicity (sinusoidal form of the observed distribution highly different in respect to theoretical distribution, additionally following the H distribution). The structure is dominated by the intrinsic hydrophobicity of each participating residue rather than by the tendency to form a shared core. The solenoid terminates in a helical "stopper", which is consistent with the model and exposes its hydrophilic residues towards the aqueous environment (fragments of the profiles shown in Figure 4C. distinguished by green background and green fragments in Figure 4A,B).

The "envelope" (which is equivalent to the parallel helix in 2FKO—pink fragments in Figure 4A—profiles of these fragments distinguished in Figure 4C. by the green background) diverges from the model; however its constituent fragments exhibit high accordance, which can be interpreted as energetically favorable contact with the environment.

The neighborhood of the catalytic active site, including the outlying loops as well as beta-strands which form part of the solenoid and contain catalytic residues, exhibits significant deviations from the model, with only the local neighborhood of residue D154 (145–152) seen as accordant. Catalytic center

(together with the close neighborhood) is distinguished as blue fragments (Figure 4B.) and by blue background in Figure 4C.

Figure 7 provides a depiction of the presented fragments.

Correlation coefficients calculated for fragments directly adjacent to the active site exhibit negative values. This means that the given fragment not only diverges from the theoretical distribution, but in fact adopts a distribution which can be regarded as opposite to it. This, in turn, suggests a high degree of structurally encoded functionality, since—given the variable nature of the peptide bond—a structure which opposes the natural alignment of T and O would be unlikely to emerge randomly.

As seen, in Figure 4, the observed distribution in the neighborhood of the active site is dominated by a sinusoid pattern, suggesting linear arrangement of identical beta-strands which comprise the solenoid (compare Figure 4 and Table 5). The loose loops which contain catalytic residues are also highly discordant.

Comparing 1PLU with 1IDJ reveals greater structural complexity of the latter protein (in the sense of the fuzzy oil drop model).

Table 5. FOD characterization of pectin lyase (1IDJ). RD coefficients determine the relative distance between the observed distribution (O) and two pairs of boundary cases (theoretical vs. unified—T–R—and theoretical vs. intrinsic—T–H). Fragments which diverge from the monocentric theoretical model are listed in boldface. To identify beta sheets, the table lists the earliest fragments belonging to each sheet. Identification of secondary structures follows PDBSum criteria [34]. Positions marked with the double asterisks correspond to SS-bonds which disrupt hydrophobicity-based ordering. Only numbers of residues belonging to stating beta-strand in beta-sheet are used for identification.

| Fragment	Structural Properties	RD		Correlation Coefficient		
		T–R	T–H	H/T	T/O	H/O
SOLENOID		0.769	0.635	0.167	0.205	0.788
56 … 97	Beta-sheet	0.584	0.766	0.464	0.124	0.770
74 … 150	Beta-sheet	0.567	0.162	0.365	0.144	0.729
264 … 272	Beta-sheet	0.591	0.614	0.238	0.271	0.784
STOP						
27–36	Helix	0.255	0.258	0.727	0.868	0.937
317–326		0.682	0.705	0.094	−0.096	0.676
ENVELOPE						
1–23		0.659	0.683	0.016	0.214	0.572
298–311	loop	0.647	0.632	0.159	0.250	0.728
327–359		0.553	0.484	0.148	0.494	0.632
295–307	helix	0.345	0.445	0.640	0.668	0.742
293–312		0.489	0.611	0.509	0.553	0.810
1–26 295–359	Together	0.662	0.656	0.082	0.258	0.664
CATAL. C.						
51–103		0.609	0.766	0.321	0.321	0.761
120–272		0.715	0.476	0.264	0.226	0.762
144–164	D154	0.601	0.363	0.334	0.375	0.858
166–186	R176	0.806	0.661	0.354	0.167	0.771
226–246	R236	0.672	0.319	-0.103	0.043	0.760
51–103	Loop	0.609	0.766	0.321	0.321	0.761
145–152	Loop D154	0.282	0.073	0.648	0.808	0.775
211–235	Loop R236	0.760	0.424	0.344	0.172	0.606
146–154	Loop	0.544	0.176	−0.003	0.288	0.691
199–210	Loop	0.701	0.731	0.132	−0.066	0.742
103–242	All Eres includ.	0.732	0.461	0.289	0.239	0.781
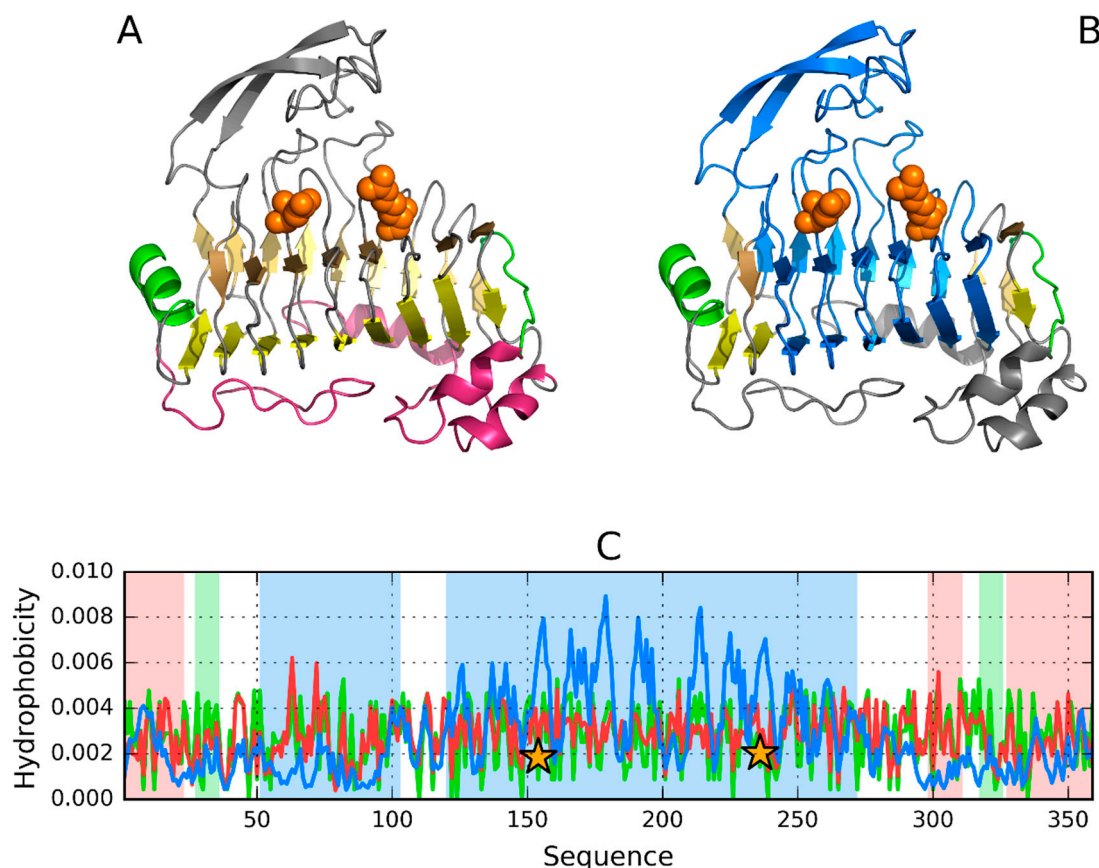
Figure 4. Characteristics of 1IDJ. (**A**,**B**)—3D presentation, (**C**)—T (blue), O (red), and H (green) hydrophobicity distribution profiles. Orange circles mark catalytic residues, whose position is represented by orange stars in C. Fragments distinguished in C as pink correspond to the envelope shown in A. Green fragments in C correspond to "stoppers" shown in A and B (also in green). Blue fragment in C represent the catalytic environment shown in B (also in blue). Orange stars—catalytic residues; gray—envelope; green—"stop" fragments; turquoise—catalytic center together with its neighborhood. Yellow fragments in A comprise the solenoid section. The yellow stars in C—catalytic residues.

3.4. 3DP3—Representative of EC 4.2.2.59

EC 4.2.2.59 is represented by 3DP3, a protein which differs structurally from the previously discussed examples due to its hexameric crystal structure. In the literature, the functional unit is characterized as a dimer [29]; however, our analysis includes monomer, dimer, and hexamer.

3DP3 contains a somewhat different type of solenoid, where the beta-sheet forms only part of the "tubule" and runs parallel to a central helix. Nevertheless, the specific ordering present in the beta-sheet suggests solenoid-like properties. Due to the biological activity of this protein (a lyase) we decided to include it in our study set.

Since the complex listed in PDB is tightly packed (large intermolecular contact area), our analysis also considers the effect of inter-chain interactions upon individual fragments of the chain. A marked decrease in RD is observed following elimination of residues which mediate such interaction.

Our analysis of the monomer indicates higher (than in former proteins) match between T and O (Figure 5C). No characteristic linear propagation (sinusoid pattern) is discernible; however, the catalytic residue diverges from the model, along with its whole neighborhood—in favor of intrinsic hydrophobicity (refer to correlation coefficients in Table 6). This means that the structure of the catalytic site is dominated by the intrinsic properties of each residue rather than by broad hydrophobic forces (Figure 5C—blue background and blue fragments in Figure 5A,B.)

Localization of fragments described in Table 6 can be seen in Figure 5.

Table 6. FOD characterization of tryptophanase (3DP3). RD coefficients determine the relative distance between the observed distribution (O) and two pairs of boundary cases (theoretical vs. unified—T–R—and theoretical vs. intrinsic—T–H). Fragments which diverge from the monocentric theoretical model are listed in boldface. To identify beta sheets, the table lists the earliest fragments belonging to each sheet. Identification of secondary structure follows PDBSum criteria [34]. "E" corresponds to the presence of a catalytic residue, with the preceding number listing the number of such residues. Here, the reference to a solenoid (*) extends the traditional definition of solenoid fragments. The protein in question contains a repeating beta structure; however, this structure forms only half of the characteristic solenoid "tubule".

| Fragment	Structural Properties	RD		Correlation Coefficient		
		T–R	T–H	H/T	T/O	H/O
MONOMER						
Chain A		0.463				
	SOLENOID *	0.433	0.208	0.277	0.547	0.728
32–37	Beta-sheet	0.414	0.185	0.333	0.566	0.725
66–88	Helix 72E	0.770	0.555	−0.245	0.259	0.557
	Beta-sheet	0.433	0.208	0.276	0.547	0.728
	No P-P	0.429	0.295	0.350	0.660	0.736
	P-P	0.625	0.351	0.006	0.278	0.634
STOP						
90–95	Helix	0.333	0.202	0.529	0.709	0.931
8–21	STOP	0.244	0.212	0.549	0.880	0.664
14–21	helix	0.346	0.281	0.705	0.876	0.872
53–58	Helix-interface	0.450	0.849	0.633	0.481	0.906
CATAL. Center						
22–30		0.095	0.127	0.669	0.966	0.647
48–88		0.533	0.521	0.128	0.635	0.674
Envelope		0.534	0.527	0.188	0.586	0.634
DIMER						
Dimer		0.588				
66–88	Helix 72E	0.906	0.783	−0.159	0.243	0.582
	Beta-sheet	0.609	0.318	0.152	0.224	0.780
	No P-P	0.580	0.432	0.303	0.589	0.739
	P-P	0.512	0.291	0.263	0.562	0.729
90–95	Stop-helix	0.421	0.277	0.385	0.529	0.940
8–21	Stop-helix	0.473	0.449	0.331	0.651	0.683
53–58	Helix-interface	0.251	0.600	0.601	0.836	0.858
COMPLEX						
Six units		0.745				
66–88	Helix 72E	0.908	0.786	−0.237	0.047	0.610
	Beta-sheet	0.637	0.366	0.112	0.185	0.779
	No P-P	0.659	0.516	0.257	0.305	0.745
	P-P	0.712	0.461	−0.058	−0.042	0.774
90–95	Stop-helix	0.598	0.395	−0.284	−0.225	0.897
14–21	Stop-helix	0.538	0.493	0.333	0.216	0.906
53–58	Helix-interface	0.620	0.844	−0.605	−0.880	−0.856
	SOLENOID	0.634	0.363	0.111	0.185	0.785
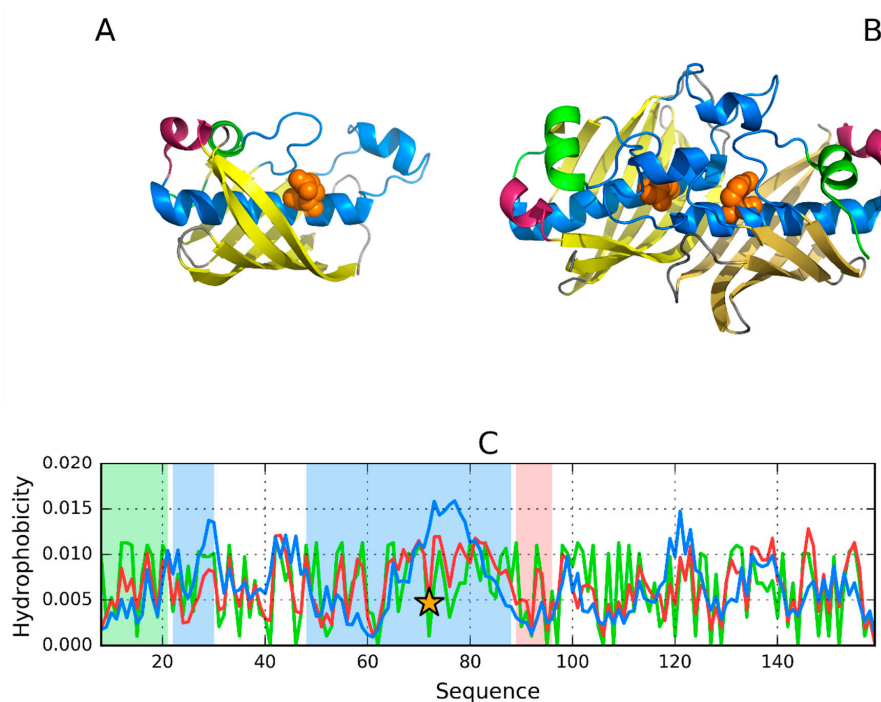
Figure 5. Characteristics of 3DP3. (**A**)—monomer and (**B**)—dimer in 3D presentation. (**C**)—T (blue), O (red) and H (green) hydrophobicity distribution profiles plotted for the monomer. The green fragment in C represents the "stopper", the pink fragment comprises the envelope shown in A and B (also as pink), while the blue fragments represents the catalytic center environment shown in A and B (also in blue). Yellow fragments in A and B—solenoids. Yellow star—position of catalytic residue.

The detailed status of the catalytic residue and its close neighborhood is presented in Figure 6. shown in the form of profiles (Figure 6A—status in monomer; Figure 6B—status in dimer).
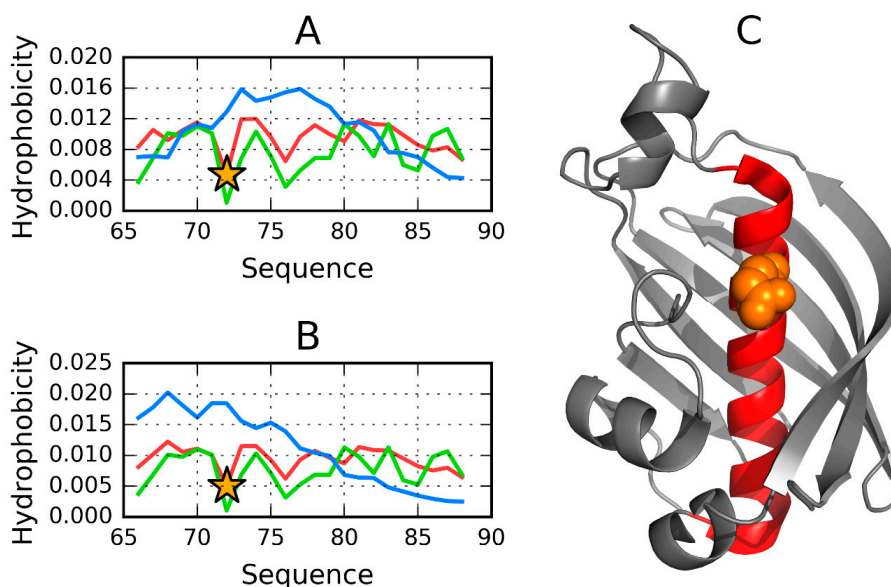
Figure 6. Characteristics of 3DP3. (**A**)—monomer and (**B**)—dimer profiles of hydrophobicity distribution for fragments proximate to catalytic residues (orange stars). T—blue line, O—red line, H—green line. (**C**)—3D presentation of monomer with the catalytic residue shown as orange spheres.

3.5. Pectate Lyase—2QY1—Lyase II

The characteristics of this enzyme (as given in PDB) is as follows: eliminative cleavage of (1→4)-α-D-galact-4-enuronosyl groups at their non-reducing ends. This enzyme is classified as Lyase II.

2QY1 is a representative of the fairly numerous pectate lyase class. It crystallizes as a dimer; however, our analysis focuses on its monomeric form. The protein itself is a mutant (R236F and A31G). The former mutation is near the catalytic residue (R230) and likely affects its enzymatic activity due to structural differences. Unlike R236F, the latter mutation (A31G) does not appear significant from the point of view of FOD analysis.

2QY1 is conformationally similar to EC 4.2.2.2 group proteins, with a solenoid fragment surrounded by loose fragments, a "stopper" helix and exposed loops surrounding the active site—although these loops are shorter than in EC 4.2.2.2 enzymes.

Table 7 summarizes the properties of the presented fragments of 2QY1 (monomer).

The RD value (Table 7) for the complete 2QY1 molecule suggests that no molecule-wide hydrophobic core is present. Results listed in Table 7 indicate that the monocentric distribution is heavily disrupted by the presence of a solenoid fragment whose own hydrophobicity profile is determined by the intrinsic properties of each residue. Beta sheets are similarly discordant versus the theoretical distributions (as seen in Figure 7). Linear arrangement of local hydrophobicity peaks can be observed along the axis of the solenoid. The N- and C-terminal sections remain accordant, likely due to the need for solubility in an aqueous environment.
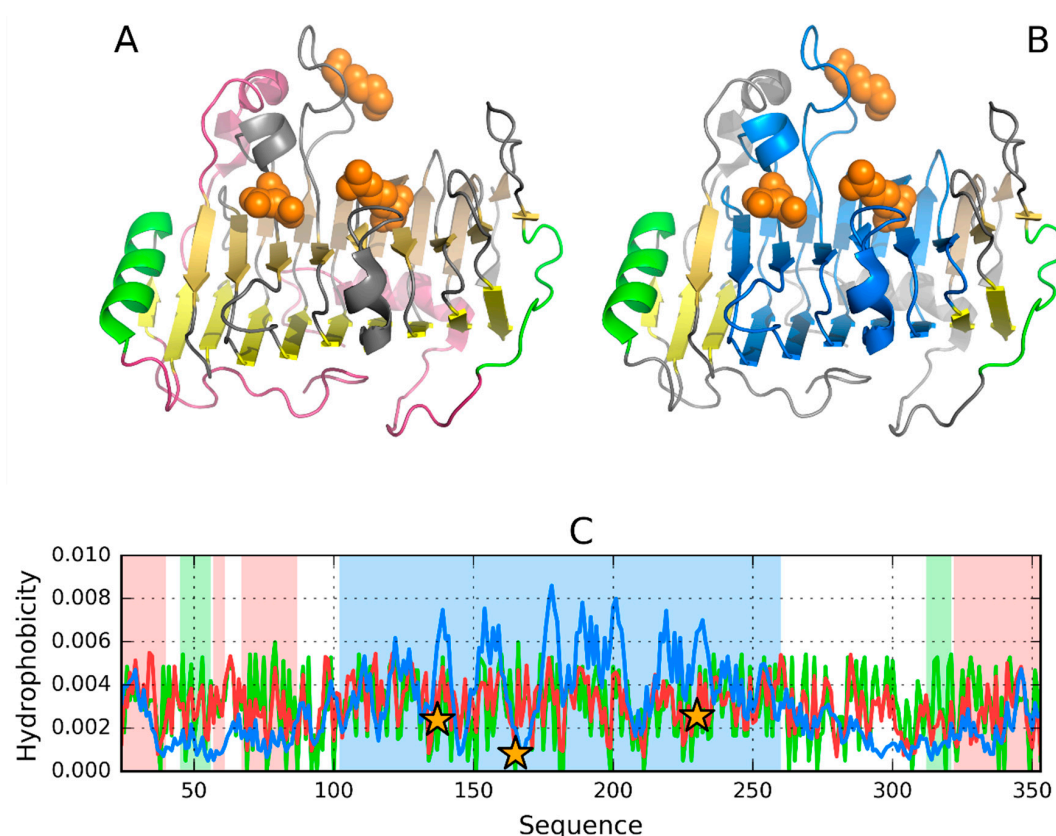

Figure 7. 3D presentation of 2QY1. Orange spheres—catalytic residues in (**A**,**B**), and orange stars in (**C**); green—"stop" fragments; blue—catalytic center together with its neighborhood; pink—envelope.

Table 7. FOD characterization of pectate lyase (2QY1). RD coefficients determine the relative distance between the observed distribution (O) and two pairs of boundary cases (theoretical vs. unified—T–R—and theoretical vs. intrinsic—T–H). Fragments which diverge from the monocentric theoretical model are listed in boldface. To identify beta sheets, the table lists the earliest fragments belonging to each sheet. Identification of secondary structure follows PDBSum criteria [34]. As with other Tables the numbers of residues belonging to starting beta-strand in beta-sheet are used for identification.

Fragment	Structural Properties	RD		Correlation Coefficient		
		T–R	T–H	H/T	T/O	H/O
SOLENOID		0.665	0.477	0.205	0.321	0.761
41–43	Beta-sheet	0.762	0.453	0.077	0.153	0.735
88–92	Beta-sheet	0.694	0.550	0.282	0.381	0.829
125–129	Beta-sheet	0.672	0.477	0.184	0.232	0.840
201–203	Beta-sheet	0.824	0.756	0.022	0.143	0.874
STOP						
45–55	Helix	0.444	0.391	0.654	0.669	0.838
ENVELOPE						
24–40	N-terminal loop	0.475	0.314	0.374	0.574	0.316
57–61		0.244	0.329	0.786	0.866	0.580
67–87		0.525	0.178	0.295	0.272	0.878
322–353		0.497	0.240	0.227	0.556	0.589
39–61		0.488	0.459	0.474	0.536	0.654
303–343		0.191	0.437	0.362	0.611	0.574
344–353	C-terminal loop	0.133	0.288	0.471	0.857	0.622
Loops complete		0.541	0.329	0.167	0.542	0.504
CATAL. Center						
102–260		0.639	0.518	0.246	0.317	0.747
127–240		0.652	0.472	0.268	0.392	0.764
132–142	Cat. Res D137	0.624	0.510	0.477	0.456	0.901
160–170	Cat. Res. K165	0.697	0.598	0.251	0.273	0.734
225–235	Cat. Res R230	0.246	0.294	0.469	0.745	0.858
67–87	loop	0.525	0.178	0.295	0.272	0.878
129–137	loop	0.553	0.721	0.044	0.115	0.877
161–177	loop	0.716	0.650	−0.005	0.349	0.654
204–215		0.230	0.153	0.413	0.852	0.626
129–237		0.629	0.468	0.271	0.400	0.764
67–222	total	0.652	0.441	0.262	0.404	0.753

The solenoid also contains "stop" fragments which conform to the model by exposing their hydrophilic residues towards the environment while internalizing hydrophobic residues.

As already remarked, the N- and C-terminal sections remain accordant and can be found on the surface of the protein alongside the solenoid. Their role is to ensure solubility, while the function of "stoppers" (helix at 45–55) is to counteract unchecked propagation of the solenoid fragment itself. This observation is supported by high values of all three correlation coefficients, indicating that the fragments in question are well adapted to their role as stoppers (i.e., they mediate contact with the aqueous environment) (Figure 7A,B). It is interesting to note the status of the catalytic active site, which is significantly discordant vs. the theoretical distribution. The specific role of R236F is difficult to determine; however, the neighborhood of the active site appears similar to the structure of EC 4.2.2.2 enzymes.

4. Discussion

Enzymatic activity requires a suitably structured active site, adapted to the target substrate which must be recognized by the enzyme. Analysis of hydrophobicity distribution in antifreeze proteins, with evidence of linear propagation (alternating local minima and maxima), may provide

hints regarding the structuralization of water near the protein surface. This effect is believed to counteract ice crystal formation by altering the structural properties of the aqueous medium [7,17]. Similarly, in the case of lyases which include solenoid fragments, the structural ordering of water may play an important role in catalysis. This hypothesis is based on the interpretation of the presence of similar (linear propagation of different level of hydrophobicity bands) ordering of hydrophobicity in antifreeze proteins. In this case, this linearity appears to be the signal for surrounding water to differentiate the ordering of water molecules against the ice generation. In consequence it may be assumed that the presence of solenoid in protein of much higher complication as the enzymes is aim-oriented. If the influence of solenoid structure in antifreeze proteins can be accepted, similar influence of solenoid structure in discussed lyases can be interpreted as the source of similar force field addressed to water environment.

Even without detailed knowledge regarding the specific structural properties of the medium, we note that the conditions encountered near hydrophilic surfaces differ markedly from those, which can be observed near hydrophobic areas—particularly in light of reports of water levitation above hydrophobic surfaces [35]. The force field addressed to water environment by linear propagation of different hydrophobicity bands modified by the parts of polypeptide chains called here as envelope supply the additional modification of the force field spread around particular lyase. The gradual complication of the "envelope" part in discussed lyases may be interpreted as gradual modification of specificity of the enzyme—it means substrate recognition. All other interactions between enzyme and substrate take part in the recognition process (electrostatic, vdW H-Bonds etc.). The hydrophobic interaction-based approach is presented in this paper. Comparison of the 3D-structure of lyases with various substrates indicates that the structuralization of water enforced by the protein's own hydrophobicity distribution, may generate a force field which promotes enzymatic activity. One example is the environment of the catalytic active site in 2KFO, whose immediate neighborhood differs from other presented lyases. We suspect that the resulting force field may play a role in guiding the substrate to the active site and that furthermore it may benefit the catalysis process itself.

The structure of any field, including the hydrophobic force field, is a way to encode information. The more complex the field, the more information it carries, and this may prove important from the point of view of catalytic processes. The quantity of information carried by 2FKO seems significantly lower than e.g., in 3DP3, which may be due to the specific properties of the latter protein's substrate, requiring a finely tuned force field. Additional information-carrying factor is the presence of ions (Table 1.) which are localized in specific positions delivering additional signals for water ordering in their neighborhood. One can speculate that they take part in specification of substrate directing to the catalytic center.

Reports regarding the possible involvement of amyloid structures at early stages of protein evolution underscore the importance of linear hydrophobic force fields [36]. The fact that similar conditions exist in enzymes—representing the apex of biological activity—seems consistent with the observations presented in [37–41].

A highly symmetrical system carries relatively little information due to its repeatability and symmetry (both of which render it structurally predictable). Here, the notion of symmetry refers not just to the geometric properties of the folded polypeptide chain, but also to the highly ordered placement of fragments which interact with the environment. Specifically, a clear balance emerges between the placement of polar and nonpolar residues. By exposing such structures on its surface, the protein effectively transmits a signal into its environment, causing the surrounding water molecules to align themselves with the protein body as discussed already above. This "signal" is likely recognized by other solvated molecules. Thus, the symmetry of the protein itself extends to its environment, ensuring the formation of a specific local force field.

One example of this phenomenon is provided by the set of lyases in which a symmetrical structure (the solenoid) is offset by dissymmetrical fragments, producing a balance between symmetry and

dissymmetry. This balance affects the structure of the aqueous solvent and likely plays a crucial role in attracting potential partner molecules for chemical reactions catalyzed by lyases.

5. Conclusions

In conclusion, we can state that protein folding represents the search for a balance between structural ordering consistent with the fuzzy oil drop model and local deviations from the model. Such deviations often correspond to the protein's biological activity [37]. Antifreeze proteins, whose activity does not require complexation of ligands or other proteins, appear to work by generating a force field which obstructs formation of ice crystals [7]. In the case of lyases, the same event may be exploited locally to identify and attract a specific ligand (substrate)—as demonstrated by the structure of active sites of the presented enzymes. The lysozyme, an enzyme which belongs to a different class (hydrolase), exhibits a markedly different secondary conformation; however, it appears to rely on the same general effect, i.e., balance between structural order (adherence to the 3D Gaussian) and local deviations from that order. It would be difficult to contradict the observation that a protein in perfect agreement with the 3D Gaussian would act upon its environment in a different manner, by exposing only hydrophilic residues and therefore affecting the arrangement of nearby water dipoles. In particular, local exposure of hydrophobic residues may play a critical role in the protein's biological profile. Linear propagation of hydrophobicity/hydrophilicity promotes unlimited elongation (e.g., in amyloids) [6,15], but it may also locally disrupt the aqueous environment in a way which attracts specific ligands and promotes catalysis (in proteins which contain solenoid fragments).

The set of proteins selected for analysis was aimed to show the gradual complication of the structure and the increasingly specific influence on the surrounded water. This stepwise increase of complication is interpreted as increase of the specificity of the enzyme what is shown using different substrates for enzymes under consideration.

References

1. Sternberg, M.J.E. *Protein Structure Prediction—Principles and Approaches in Protein Structure Prediction*; Ed Sternberg. IRL Press Oxford University Press: Oxford, UK; New York, NY, USA; Tokyo, Japan, 1996; p. 11.
2. Andrade, M.A.; Perez-Iratxeta, C.; Ponting, C.P. Protein repeats: Structures, functions, and evolution. *J. Struct. Biol.* **2001**, *134*, 117–131. [CrossRef] [PubMed]
3. Kobe, B.; Kajava, A.V. The leucine-rich repeat as a protein recognition motif. *Curr. Opin. Struct. Biol.* **2001**, *11*, 725–732. [CrossRef]
4. Kajava, A.V. Structural diversity of leucine-rich repeat proteins. *J. Mol. Biol.* **1998**, *277*, 519–527. [CrossRef] [PubMed]
5. Bazan, J.F.; Kajava, A.V. Design on a curve. *Nat. Struct. Mol. Biol.* **2015**, *22*, 103–105. [CrossRef] [PubMed]
6. Roterman, I.; Banach, M.; Konieczny, L. Application of the Fuzzy Oil Drop Model Describes Amyloid as a Ribbonlike Micelle. *Entropy* **2017**, *19*, 167. [CrossRef]
7. Banach, M.; Konieczny, L.; Roterman, I. Why do antifreeze proteins require a solenoid? *Biochimie* **2018**, *144*, 74–84. [CrossRef] [PubMed]

8.   Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242. [CrossRef]

9.   Konieczny, L.; Bryliński, M.; Roterman, I. Gauss-function-Based model of hydrophobicity density in proteins. *Silico Biol.* **2006**, *6*, 15–22.

10.   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]

11.   Kauzmann, W. Some factors in the interpretation of protein denaturation. *Adv. Protein Chem.* **1959**, *14*, 1–63.

12.   Banach, M.; Konieczny, L.; Roterman, I. Can the Structure of Hydrophobic Core Determine the Complexation Site? In *Identification of Ligand Binding Site and Protein-Protein Interaction Area*; Roterman-Konieczna, I., Ed.; Springer: Dordrecht, The Netherlands, 2013; pp. 41–54.

13.   Banach, M.; Konieczny, L.; Roterman, I. Use of the "Fuzzy Oil Drop" Model to Identify the Complexation Area in Protein Homodimers. In *Protein Folding*; Roterman-Konieczna, I., Ed.; Silico, Woodhead Publishing (currently Elsevier): Oxford, UK; Cambridge, UK; Philadelphia, PA, USA; New Dehli, India, 2012; pp. 95–122.

14.   Banach, M.; Konieczny, L.; Roterman, I. Ligand-binding-site Recognition. In *Protein Folding*; Roterman-Konieczna, I., Ed.; Silico, Woodhead Publishing (currently Elsevier): Oxford, UK; Cambridge, UK; Philadelphia, PA, USA; New Dehli, India, 2012; pp. 78–94.

15.   Roterman, I.; Banach, M.; Kalinowska, B.; Konieczny, L. Influence of the Aqueous Environment on Protein Structure—A Plausible Hypothesis Concerning the Mechanism of Amyloidogenesis. *Entropy* **2016**, *18*, 351. [CrossRef]

16.   Fabian, P.; Stapor, K.; Banach, M.; Ptak-Kaczor, M.; Konieczny, L.; Roterman, I. Different synergy in amyloids and biologically active forms of proteins. *Int. J. Mol. Sci.* **2019**, *20*, 4436. [CrossRef] [PubMed]

17.   Roterman, I.; Banach, M.; Konieczny, L. Propagation of Fibrillar Structural Forms in Proteins Stopped by Naturally Occurring Short Polypeptide Chain Fragments. *Pharmaceuticals* **2017**, *10*, 89. [CrossRef] [PubMed]

18.   Iverson, T.M.; Alber, B.E.; Kisker, C.; Ferry, J.G.; Rees, D.C. A closer look at the active site of gamma-class carbonic anhydrases: High-resolution crystallographic studies of the carbonic anhydrase from Methanosarcina thermophila. *Biochemistry* **2000**, *39*, 9222–9231. [CrossRef] [PubMed]

19.   Jeyakanthan, J.; Rangarajan, S.; Mridula, P.; Kanaujia, S.P.; Shiro, Y.; Kuramitsu, S.; Yokoyama, S.; Sekar, K. Observation of a calcium-binding site in the gamma-class carbonic anhydrase from Pyrococcus horikoshii. *Acta Crystallogr. D Biol. Crystallogr.* **2008**, *64 Pt 10*, 1012–1019. [CrossRef]

20.   Yoder, M.D.; Jurnak, F. The Refined Three-Dimensional Structure of Pectate Lyase C from Erwinia chrysanthemi at 2.2 Angstrom Resolution (Implications for an Enzymatic Mechanism). *Plant Physiol.* **1995**, *107*, 349–364. [CrossRef] [PubMed]

21.   Pickersgill, R.; Jenkins, J.; Harris, G.; Nasser, W.; Robert-Baudouy, J. The structure of Bacillus subtilis pectate lyase in complex with calcium. *Nat. Struct. Biol.* **1994**, *1*, 717–723. [CrossRef] [PubMed]

22.   Pickersgill, R.; Worboys, K.; Scott, M.; Cummings, N.; Cooper, A.; Jenkins, J.; Smith, D. The Conserved Arginine Proximal to the Essential Calcium of Bacillus Subtilis Pectate Lyase Stabilizes the Transition State. Available online: https://www.rcsb.org/structure/2BSP (accessed on 28 September 2019).

23.   Thomas, L.M.; Doan, C.N.; Oliver, R.L.; Yoder, M.D. Structure of pectate lyase A: Comparison to other isoforms. *Acta Crystallogr. D Biol. Crystallogr.* **2002**, *58*, 1008–1015; [CrossRef]

24.   Herron, S.R.; Scavetta, R.D.; Garrett, M.; Legner, M.; Jurnak, F. Characterization and implications of Ca2+ binding to pectate lyase C. *J. Biol. Chem.* **2003**, *278*, 12271–12277; [CrossRef]

25.   Dehdashti, S.J.; Doan, C.N.; Chao, K.L.; Yoder, M.D. Effect of mutations in the T1.5 loop of pectate lyase A from Erwinia chrysanthemi EC16. *Acta Crystallogr. D Biol. Crystallogr.* **2003**, *59*, 1339–1342. [CrossRef]

26.   Czerwinski, E.W.; Midoro-Horiuti, T.; White, M.A.; Brooks, E.G.; Goldblum, R.M. Crystal structure of Jun a 1, the major cedar pollen allergen from Juniperus ashei, reveals a parallel beta-helical core. *J. Biol. Chem.* **2005**, *280*, 3740–3746. [CrossRef] [PubMed]

27.   Jenkins, J.; Shevchik, V.E.; Hugouvieux-Cotte-Pattat, N.; Pickersgill, R.W. The crystal structure of pectate lyase Pel9A from Erwinia chrysanthemi. *J. Biol. Chem.* **2004**, *279*, 9139–9145. [CrossRef]

28.   Mayans, O.; Scott, M.; Connerton, I.; Gravesen, T.; Benen, J.; Visser, J.; Pickersgill, R.; Jenkins, J. Two crystal structures of pectin lyase A from Aspergillus reveal a pH driven conformational change and striking divergence in the substrate-binding clefts of pectin and pectate lyases. *Structure* **1997**, *5*, 677–689. [CrossRef]

29.   He, L.; Zhang, L.; Liu, X.; Li, X.; Zheng, M.; Li, H.; Yu, K.; Chen, K.; Shen, X.; Jiang, H.; et al. Discovering potent inhibitors against the beta-hydroxyacyl-acyl carrier protein dehydratase (FabZ) of Helicobacter pylori:

Structure-based design, synthesis, bioassay, and crystal structure determination. *J. Med. Chem.* **2009**, *52*, 2465–2481. [CrossRef]

30. Chen, J.; Zhang, L.; Zhang, Y.; Zhang, H.; Du, J.; Ding, J.; Guo, Y.; Jiang, H.; Shen, X. Emodin targets the beta-hydroxyacyl-acyl carrier protein dehydratase from Helicobacter pylori: Enzymatic inhibition assay with crystal structural and thermodynamic characterization. *BMC Microbiol.* **2009**, *9*, 91. [CrossRef] [PubMed]

31. Xiao, Z.; Bergeron, H.; Grosse, S.; Beauchemin, M.; Garron, M.L.; Shaya, D.; Sulea, T.; Cygler, M.; Lau, P.C. Improvement of the thermostability and activity of a pectate lyase by single amino acid substitutions, using a strategy based on melting-temperature-guided sequence alignment. *Appl. Environ. Microbiol.* **2008**, *74*, 1183–1189. [CrossRef]

32. Levitt, M. A simplified representation of protein conformations for rapid simulation of protein folding. *J. Mol. Biol.* **1976**, *104*, 59–107. [CrossRef]

33. Kullback, S.; Leibler, R.A. On Information and Sufficiency. *Ann. Math. Stat.* **1951**, *22*, 79–86. [CrossRef]

34. de Beer, T.A.P.; Berka, K.; Thornton, J.M.; Laskowski, R.A. PDBsum additions. *Nucleic Acids Res.* **2014**, *42*, D292–D296. [CrossRef]

35. Schutzius, T.M.; Jung, S.; Maitra, T.; Graeber, G.; Köhme, M.; Poulikakos, D. Spontaneous droplet trampolining on rigid superhydrophobic surfaces. *Nature* **2015**, *527*, 82–85. [CrossRef]

36. 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*, 117–133. [CrossRef] [PubMed]

37. Greenwald, J.; Riek, R. On the possible amyloid origin of protein folds. *J. Mol. Biol.* **2012**, *421*, 417–426. [CrossRef] [PubMed]

38. Rufo, C.M.; Moroz, Y.S.; Stöhr, J.; Smith, T.A.; Hu, X.; DeGrado, W.F.; Korendovych, I.V. Short peptides self-assemble to produce catalytic amyloids. *Nat. Chem.* **2014**, *6*, 303–309. [CrossRef] [PubMed]

39. Seelig, B.; Szostak, J.W. Selection and evolution of enzymes from a partially randomized non-catalytic scaffold. *Nature* **2007**, *448*, 828–831. [CrossRef] [PubMed]

40. Smith, B.A.; Hecht, M.H. Novel proteins: From fold to function. *Curr. Opin. Chem. Biol.* **2011**, *15*, 421–426. [CrossRef] [PubMed]

41. Murphy, G.S.; Greisman, J.B.; Hecht, M.H. De Novo proteins with life-sustaining functions are structurally dynamic. *J. Mol. Biol.* **2016**, *428*, 399–411. [CrossRef] [PubMed]