Entropy **2015**, *17*, 1477-1507; doi:10.3390/e17031477

OPEN ACCESS

entropy

ISSN 1099-4300

www.mdpi.com/journal/entropy

Article

Application of Divergence Entropy to Characterize the Structure of the Hydrophobic Core in DNA Interacting Proteins

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

[1] Department of Bioinformatics and Telemedicine, Collegium Medium, Jagiellonian University, Lazarza 16, 31-530 Krakow, Poland; E-Mails: malijka@gmail.com (B.K.); mateusz.banach@uj.edu.pl (M.B.)

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

[3] Chair of Medical Biochemistry, Collegium Medicum, Jagiellonian University, Kopernika 7, 31-034 Krakow, Poland; E-Mail:mbkoniec@cyf-kr.edu.pl

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

Academic Editor: Wayne Dawson

Received: 4 September 2014 / Accepted: 17 March 2015 / Published: 23 March 2015

Abstract: The fuzzy oil drop model, a tool which can be used to study the structure of the hydrophobic core in proteins, has been applied in the analysis of proteins belonging to the jumonji group—JARID2, JARID1A, JARID1B and JARID1D—proteins that share the property of being able to interact with DNA. Their ARID and PHD domains, when analyzed in the context of the fuzzy oil drop model, are found to exhibit structural variability regarding the status of their secondary folds, including the β-hairpin which determines their biological function. Additionally, the structure of disordered fragments which are present in jumonji proteins (as confirmed by the DisProt database) is explained on the grounds of the hydrophobic core model, suggesting that such fragments contribute to tertiary structural stabilization. This conclusion is supported by divergence entropy measurements, expressing the degree of ordering in each protein's hydrophobic core.

Keywords: divergence entropy; hydrophobicity; fuzzy oil drop; disordered proteins; JARID2; JARID1A; JARID1B; JARID1D; ARID; PHD; jumonji

1. Introduction

Hydrophobic interactions are traditionally regarded as responsible for tertiary structural stabilization [1–10]. The original purpose of the "oil drop" model was to explain the presence of a hydrophobic core which aggregates hydrophobic residues while polar residues are exposed on the protein's surface. This behavior is akin to that of a drop of oil which avoids contact with water by minimizing its surface area. We have expanded Kauzmann's original qualitative description [11] creating a new model which we refer to as the "fuzzy oil drop" (FOD) [12]. This model provides a comprehensive formal description of a centrally located hydrophobic core as well as a hydrophilic shell which insulates it from contact with water. Additionally, our model proposes quantitative measures for assessing the status of the hydrophobic core and therefore of the protein's structural stability. This assessment can be performed by invoking the concept of divergence entropy, which expresses similarities between distributions of probability, or—in our case—distributions of hydrophobic density in target proteins [13]. The concept of divergence entropy permits us to quantitatively measure the contribution from the individual polypeptide chain fragments to the formation of a common hydrophobic core, as well as (indirectly) elucidate the protein's biological properties. In order to achieve this goal, we propose an "idealized" hydrophobic density distribution, which is approximated—to a greater or lesser extent—by the actual proteins. We assume that the idealized hydrophobic core structure, by virtue of its ordered form, enhances the protein's structural stability. At this point we should emphasize that, by referring to a "hydrophobic core", we actually mean the entire distribution of hydrophobic density throughout the protein body, including its outer hydrophilic layers, which render the protein water-soluble. Such quantitative assessment of the hydrophobic core structure, as well as of the contribution of selected secondary folds (outer regions of the protein), is facilitated by applying Kullback-Leibler's divergence entropy criterion [13]. The assessment may apply to structural units of varying scope (complexes, chains, domains), to selected fragments of a single polypeptide chain (such as individual secondary folds) or indeed to any arbitrarily selected arrangement of peptides.

In this work, we focus on well-defined secondary folds and fragments identified as "intrinsically disordered" (as listed in the DisProt database) [14–16]. More specifically, the subject of our analysis will be a set of proteins encoded by the jumonji gene [17]. These proteins generally fall into two groups: AT-rich-ARID and PHD. Jumonji proteins are functionally related to epigenetic modifications of histone H3K4 and include JARID1A (lysine-specific demethylase 5A, coded for by KDM5A); JARID1B (lysine-specific demethylase 5B, coded for by KDM5B) and JARID1D (lysine-specific demethylase 5D, coded for by KDM5D). All these proteins are capable of interacting with DNA. They also form parts of the polycomb-repressive complex 2 (PRC2). Domains selected for analysis include fragments responsible for DNA interactions; specifically, the β-hairpin which, in the case of PHD domains, is characterized by the stabilizing presence of zinc (Zn^{2+}) ions. Our selection is motivated by the fact that these domains participate in epigenetic phenomena (histone demethylase activity) [18] as well as in polycomb complexation [19–21], both of which are important in the context of pathological (disease-related) processes [22,23].

The fuzzy oil drop model has been used to determine the structure of the hydrophobic core in various types of proteins, including antifreeze proteins and their mutants [24], downhill proteins [25] and enzymes (particularly hydrolases [26]). It can also be used to identify the location of ligand binding

cavities [27,28], protein complexation areas [29,30], the structure of large protein complexes (such as chaperonin) [31] and the properties of proteins which include intrinsically disordered fragments [32] or share certain structural characteristics (e.g., immunoglobulin-like domains) [33]. In this work we apply the model to a set of relatively small proteins whose biological function involves interaction with DNA via a specific loop. Additionally, we also show how the fuzzy oil drop model may be applied to study the status of disordered fragments and determine their contribution to the formation of a common hydrophobic core.

The issue addressed in this work is a crucial aspect of a wider problem, namely protein folding. We expect that observations based on the fuzzy oil drop model will yield further insight into the folding process. Protein folding simulations based on the optimization of nonbonding interactions (internal force fields) should be extended with components which reflect the presence of an external force field (*i.e.*, the protein's water environment) and its role in generating a hydrophobic core—a key determinant of structural stability and biological function in many different proteins.

2. Theory

The fuzzy oil drop model classifies protein structures with regard to the presence and structure of their hydrophobic cores. For readers unfamiliar with modeling hydrophobicity and the fuzzy oil drop model presented here, the basic concepts are briefly explained in the subsections that follow.

2.1. Background on Modeling the Hydrophobic Density

As already mentioned, we have extended Kauzmann's original "oil drop" abstraction with a mathematical formalism which models the idealized "droplike" protein structure using a 3D Gaussian function [12]. This function peaks at the central point of its independent variable range, with values decreasing with distance from the center (bell curve) and approaching 0 at a distance of 3σ in each direction (the so-called three-sigma rule). We assume that the theoretical distribution of hydrophobic density in an "idealized" hydrophobic core corresponds to the 3D Gaussian with perfect accuracy. This can be mathematically expressed as follows:

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

where $\tilde{H}t_j$ describes the theoretical hydrophobic density (hence the *t* index) at point *j*. The position of the point *j* can be selected arbitrarily (the Gauss function is continuous); however, in calculating a particular protein, these points are selected according to the positions of residues present in the protein under consideration. In particular, the positions of the "effective side chains" (averaged position of atoms belonging to particular residue) are taken to represent the hydrophobic density as it appears in protein under consideration. The parameters $\bar{x},\bar{y},\bar{z}$ are the peak of the Gaussian function and $\sigma_x,\sigma_y,\sigma_z$ represent the Gaussian function on each of the three principal directions: according to the three-sigma rule, over 99% of the function's integral is contained in an area given by $(\bar{x}\pm 3\sigma)$ .When considering three dimensions, the corresponding assumption is that 99% of the protein's hydrophobic density is

contained in an ellipsoid bounded by $(\bar{x} \pm 3\sigma_x, \bar{x} \pm 3\sigma_y, \bar{x} \pm 3\sigma_z)$ and that the Gaussian can be safely truncated beyond these limits.

Thus, $\tilde{H}t_j$ is the hydrophobic density of the *j*-th peptide, measured at coordinates (x_j, y_j, z_j), which represent that peptide's effective side chain.

The values $\tilde{H}t_j$ are calculated for selected positions (points in the ellipsoid). The next step is to calculate the sum of all of them. To make the final $\tilde{H}t_j$ values normalized, all of them are multiply by the coefficient $\frac{1}{\tilde{H}t_{sum}}$. Its presence ensures normalization of $\tilde{H}t_j$ (its total value becomes equal to 1.0 regardless of the protein being analyzed).

As a result, the values of $\tilde{H}t_j$ express the *theoretical* hydrophobic density which should characterize each amino acid if the resulting hydrophobic core is to match the theoretical expectations. The greater the distance between the effective side chain and the center of the molecule the lower its expected hydrophobic density. For residues exposed on the protein's surface, this value becomes close to 0. The simplified graphic presentation of the encapsulation of the protein molecule in a 3D Gaussian function "drop" is shown in Figure 1.

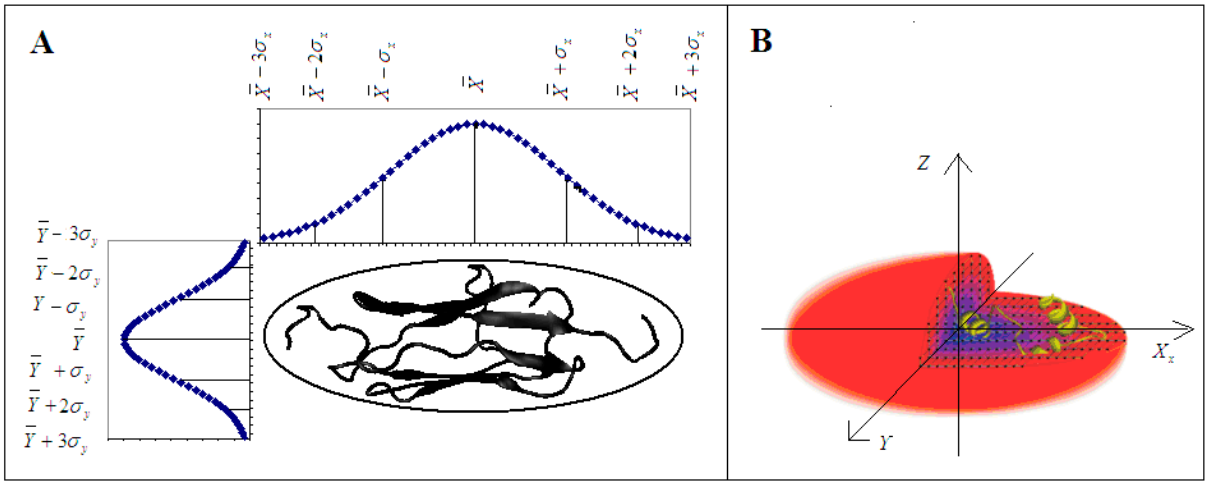

Figure 1. Visualization of protein molecule encapsulation in 3D Gaussian function shaped "fuzzy drop". (**A**): The 3D Gaussian function reduced to two-dimensional form for simplicity. The position of the center is denoted as $\bar{x}, \bar{y}$ for each axis. The parameters σ_x, σ_y represent the standard deviation. According to the "three sigma rule" about 99% of all values of Gauss function are incorporated. This is why the space enclosed in $(\bar{x} \pm 3\sigma_x, \bar{y} \pm 3\sigma_y, \bar{z} \pm 3\sigma_z)$ can be treated as the complete capsule. One can see that the σ_x, σ_y can be of different dimensions; (**B**): The 3D projection of the protein molecule (yellow ribbon) encapsulated in the ellipsoid (red). The intensity of the blue color represents the gradual increase of the hydrophobic density with its maximum at the center of the ellipsoid.

2.2. The Real Hydrophobic Distribution in a Protein Molecule

The formalism presented above expresses the theoretical hydrophobic density distribution in a target protein. Real-life proteins, however, do not follow this distribution with perfect accuracy. Actual

(observed) hydrophobic density distribution may either approximate the theoretical model, or diverge from it (see central diagram in Figure 2). In a real polypeptide chain, the hydrophobic density distribution depends on the placement of side chains (represented by their effective atoms) and on their own hydrophobicity which, in turn, depends on the type of peptide and on its interactions with neighboring residues. In order to formally express this relation we apply the following formula, originally proposed by Levitt [34]:

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

where N is the number of amino acids in the protein, $\tilde{H}_i^r$ expresses the hydrophobic parameter of the i-th residue, while r_{ij} expresses the distance between two interacting residues (j-th "effective side chain" and i-th "effective side chain"). There are many scales measuring the intrinsic hydrophobicity of each amino acids [35–41]—in our work we apply the scale proposed in [42]. The parameter c expresses the cutoff distance for hydrophobic interactions, which is taken as 9.0 Å (following [34]). The $\tilde{Ho}_{sum}$ coefficient (calculated for all $\tilde{Ho}$) represents the aggregate sum of all the components and normalizes the distribution. Such normalization, when performed for both the theoretical and observed distributions, allows us to meaningfully compare these distributions. The parameters expressing the hydrophobicity scale are given in Table 1.

Table 1. Intrinsic hydrophobicity of each amino acid as applied in our calculations. The values are based on the relative distance of the residue position (it depends on the size of molecule) from the molecule's center: the farther away from the center, the lower the value. In deriving these coefficients we relied on a non-redundant PDB set.

AA	Intrinsic Hydrophobicity	AA	Intrinsic Hydrophobicity
LYS	0.001	ALA	0.572
GLU	0.083	HIS	0.628
ASP	0.167	TYR	0.700
GLN	0.250	LEU	0.783
ARG	0.272	VAL	0.811
ASN	0.278	MET	0.828
PRO	0.300	TRP	0.856
SER	0.422	ILE	0.883
THR	0.478	PHE	0.906
GLY	0.550	CYS	1.000

The applied hydrophobicity scale closely approximates the one proposed by Kyte-Doolittle [35]. The correlation coefficient measuring the accordance of these two scales equals 0.83. In-depth comparative FOD analysis applying each of the presented scales points to quantitative differences which, however, do not translate into qualitative changes. The parameters based on FOD calculated for different scales differ but the overall status of the molecule remains the same for all analyzed proteins. In conclusion,

the choice of intrinsic hydrophobicity scale [31–41] does not appear to affect the outcome of fuzzy oil drop analysis. The comparative analysis was performed for hydrophobic parameters applying the Kyte-Doolittle scale [35] and shown in appropriate tables.

2.3. Measuring Differences between the Theoretical and Observed Distributions—the Role of Divergence Entropy

Formal assessment of the differences between the theoretical and observed distributions is facilitated by Kullback and Leibler's divergence entropy criterion (also referred to as distance entropy) [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 entropy between the empirical (p) and target (p^0) distributions. In our case, the empirical distribution is the observed one while the target distribution is supplied by the 3D Gaussian function. N is the number of residues in the protein chain (number of effective atoms). According to its definition, D_{KL} is a measure of the entropy and thus cannot be interpreted on its own. An independent target distribution is required—one in which no concentration of hydrophobic density can be discerned at the center of the molecule. In this so-called *uniform* distribution, each residue is assigned a hydrophobic density value of 1/N, where N is the number of residues in the polypeptide chain (see the rightmost diagram in Figure 2).

To simplify matters, the following notation can be applied:

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

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

where $O|T$ is the divergence entropy expressing the distance between the observed (O) and theoretical (T) distributions, while $O|R$ is the divergence entropy expressing the distance between the observed distribution (O) and a distribution in which each residue carries the same hydrophobic density value, *i.e.*, no hydrophobic core is present (R).This simplification introduces the following notation: $\tilde{H}o_j = O_j$ and $\tilde{H}t_j = T_j$.

The presented selection of targets enables us to compare the observed distribution with two limiting cases—the 3D Gaussian (perfect hydrophobic core) and the uniform distribution (no hydrophobic core of any kind). Comparing $O|T$ and $O|$R profiles reveals the "closeness" between the observed and theoretical distributions for a given protein. A binary predicate can be adopted at this stage: we assume that $O|T<O|R$ indicates the presence of a hydrophobic core.

In order to avoid continuously comparing pairs of values, we introduce the following notation:

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

where RD stands for the *relative distance* between the observed and theoretical distributions. The lower its value, the more closely a given polypeptide chain approximates the corresponding theoretical optimum. $RD = 0.5$ is used as the threshold for distinguishing between a hydrophobic core ($RD < 0.5$)

and the absence of a hydrophobic core for $RD > 0.5$. (The interpretation is that in the first case the observed distribution is "closer" to the theoretical.) Figure 2 provides a graphical depiction of this relationship. Proteins that satisfy $RD < 0.5$ are called "accordant".

The *RD* value computed for the sample protein in Figure 2 (central diagram) is 0.54, indicating that the observed deformation of the central peak is sufficient to classify this protein as "discordant". How can this type of result be interpreted? As shown in the diagram, the distribution is distinctly lopsided, with a shallow left-hand slope and a steep drop off beyond the peak. This can be taken as an indication that the "left-hand section" of the protein is more susceptible to penetration by polar water molecules, potentially reducing its stability. While structural stability can be achieved in other ways (e.g., with disulfide bonds or a suitable arrangement of ionic bridges), the presented protein does not benefit from the stabilizing influence of a well-defined hydrophobic core.

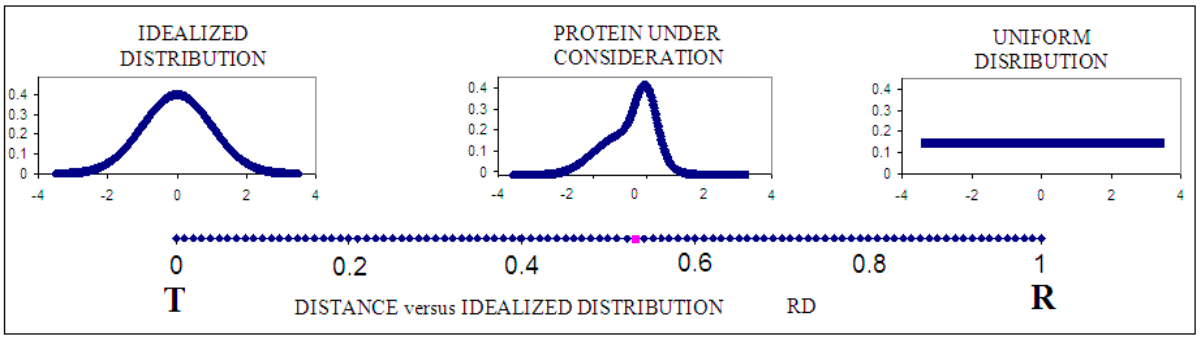

Figure 2. Visualization of a protein structural assessment (reduced to a single dimension for the sake of clarity). The left-hand diagram represents the idealized Gaussian distribution (T) while the right-hand diagram corresponds to the uniform distribution (R). The observed hydrophobic distribution for a sample is shown in the center, with its corresponding *RD* value marked on the horizontal axis with a pink dot. According to fuzzy oil drop criteria, this protein's hydrophobic core does not correspond to theoretical expectations ($RD = 0.54$).

An interesting question which can be asked at this point is whether the status of two (hypothetical) proteins with $RD = 0.499$ and $RD = 0.501$ respectively is, in fact, quantitatively different. While we do not expect this small difference to matter, accurate values of *RD* enable us to rank proteins in the order of adherence to the theoretical hydrophobic core model.

In terms of its applicability, the fuzzy oil drop model is best suited to globular proteins that can be readily encapsulated in an ellipsoid. Applying it to elongated structures, such as a single long helix, is not recommended as it typically leads to very high *RD* values (although this type of classification may make sense, for example performing blind analysis of a large number of dissimilar proteins in order to prepare a ranking list expressing their accordance with the model).

Figure 2 depicts the distribution of hydrophobic density for a sample protein, prepared under the assumption that we possess a "hydrophobic detector" with which we can measure each residue belonging to the protein. This leads to a chart where each residue is assigned a pair of attributes—T_i and O_i. The following diagram presents two sample proteins which differ with respect to their accordance with the fuzzy oil drop model.

Visual inspection of the diagrams shown in Figure 3 reveals differences in the degree of accordance observed in both proteins. Formal quantitative assessment based on the concept of divergence entropy will be presented later on in this work; meanwhile even a cursory glance at Figure 3 shows that the status of the entire domain—as well as the β-hairpin—depends on the protein in question: 2MA5 is seen as accordant with the theoretical hydrophobic density distribution, while 3GL6 diverges from it.

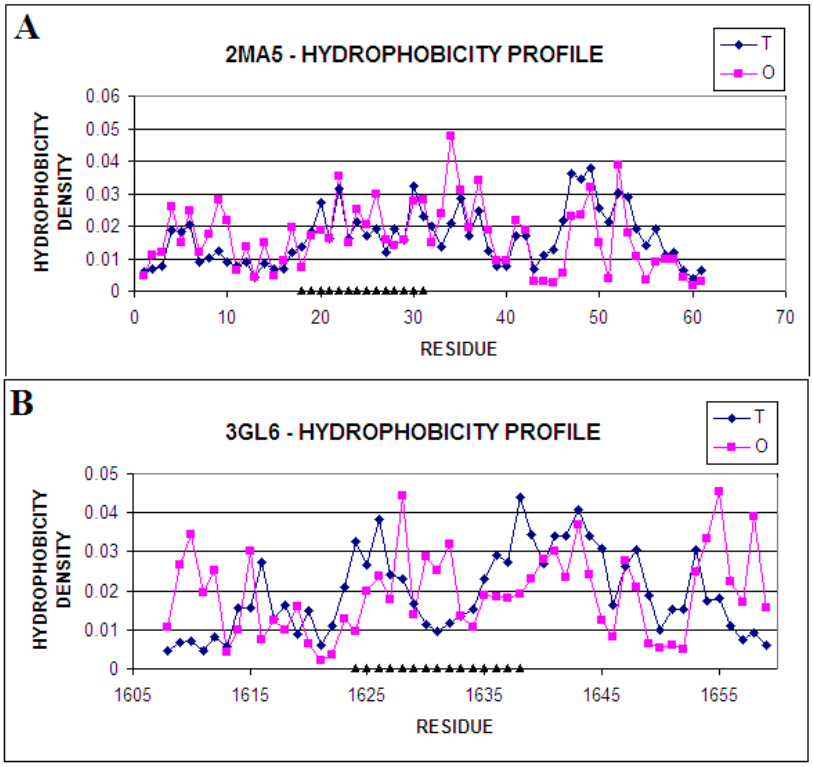

Figure 3. Two distributions with dissimilar interpretations: (**A**): good accordance between the theoretical (T) and observed (O) distribution for 2MA5, with RD = 0.368. (**B**): poor accordance between T and O for 3GL6 (1608–1660 domain) with RD = 0.698. The biologically active β-hairpin fold is marked on the horizontal axis.

2.4. Determining the Status of Individual Fragments of the Polypeptide Chain

The value of divergence entropy (D_{KL}), as well as the $O|T$, $O|$R and RD coefficients, can be calculated for various structural units: protein complexes, individual chains and separate domains. In each of these cases an appropriate ellipsoid must be defined and the appropriate calculations performed (as described above). Our to-date experience, based on studying hundreds of different proteins, complexes and domains, indicates that a vast majority of the domains remain accordant with the model, and that excellent accordance is also observed for certain classes of proteins such as antifreeze [24] and downhill [25] proteins.

The value of the divergence entropy may also help determine the status of fragments of the polypeptide chain which correspond to known secondary structural motifs, including loops and disordered fragments. In such cases, no separate ellipsoid is defined—instead, the value of O_i, T_i and R_i (where i belongs to the selected fragment) are subjected to renormalization. In the next step, a new pair of coefficients ($O|T$ and $O|$R) is derived. This, in turn, leads to a value of RD that represents the status

of the selected chain fragment. Fragments for which $RD < 0.5$ are regarded as accordant with the theoretical model: participating in the formation of a common hydrophobic core and therefore increasing the structural stability of the given unit (complex, protein or domain).

2.5. Determining the Biological Properties of Proteins Based on the Fuzzy Oil Drop Model

What properties would a perfectly accordant protein possess? In such a protein, all hydrophobic residues would be buried inside the molecule and insulated by a strongly hydrophilic shell. The protein would be highly water-soluble but also incapable of interacting with any external molecules, except perhaps ions or polar ligands (and even these complexes would be unstable in the presence of water).

It turns out that real proteins exhibit certain discrepancies versus the idealized hydrophobic density distribution. These discrepancies may manifest themselves as either an excess of hydrophobic density or as a deficiency of same. Local excess—if present on the surface of the protein—suggests a potential complexation site for molecules that exhibit similar properties [29,30], while a hydrophobicity deficiency is usually associated with the presence of a binding pocket, capable of housing a ligand or substrate [28,29]. In this way, the fuzzy oil drop model may be applied to determine the biological activity of a given protein (or protein fragment). For example, we have been able to explain the high variability (in terms of biological properties) of immunoglobulin-like domains despite their structural uniformity. One of the proteins that share these properties is titin (1TIT), which remains accordant with the model as a unit as well as a collection of individual secondary folds ($RD = 0.382$). It should be noted that titin is found in muscle tissue where it is subjected to repeating cycles of stretching and relaxation. When the external stretching force disappears, the protein spontaneously reverts to its native form, ensuring proper operation of the muscle. This property is a consequence of titin's good accordance with the idealized hydrophobic density distribution, and our results remain fully consistent with molecular dynamics simulations carried out in the presence of an external force field [43]. Good accordance with the model enables titin to spontaneously revert to its original state in the absence of stretching forces, purely via hydrophobic interactions.

A similar relation between the hydrophobic core status and the amyloidogenic properties of transthyretin (1DVQ) is presented in [33]. Experiments have revealed significant differentiation of its N- and C-terminal halves, and the fuzzy oil drop model confirms functional differences between these sections [33].

3. Materials and Methods

This work focuses on the properties of the ARID and PHD domains which comprise JARID2 proteins (also including JARID1A, JARID1B and JARID1D). All these proteins share similar biological properties—especially the ability to interact with DNA. We attempt to identify the fuzzy oil drop status of the β-hairpin loop which mediates this activity.

3.1. Data

Table 2 lists proteins which have been subjected to analysis, along with a brief description of each of these proteins.

Table 2. Summary of proteins subjected to analysis. The leftmost column provides a brief description of each protein, based on PDBSum data [44] along with the corresponding CATH classification [45]. HLSD: histone lysine-specific demethylase; KDM: histone identification; "H" indicates helixes (along with the number of helical fragments). The table also provides information on zinc ion complexation capabilities and inclusion in the DisProt database.

	JARID1A	**JARID1B**	**JARID1D**	**JARID2**
ARID 1.10.150.60 Mainly Orthogonal α-Bundle	2JXJ [46] HLSD-5A KDM5A_HUMAN 1H + β-hairpin + 5H DisProt (103-116)	2EQY HLSD-5B KDM5B_MOUSE 1H + β-hairpin + 5H DisProt (115-128)	2YQE HLSD-5D KDM5D_HUMAN 1H + β-hairpin + 5H	2RQ5 Transcription Jumonji protein JARID2_MOUSE 1H + β-hairpin + 6H
PHD 2 x Zn^{2+} 3.30.40.10 α-2-Layer β-Sandwich	2KGG [47] HLSD-5A KDM5A_HUMAN β-hairpin + 1H 2KGI [47] HLSD-5A KDM5A_HUMAN C-terminal β-hairpin + 1H 2 x Zn^{2+} Ligand (fragment of histone) 3GL6 PHD3 HLSD-5A KDM5A_HUMAN β-hairpin + 2H 3 x Zn^{2+} DisProt (1609-1659)	2MA5 HLSD-5B KDM5B_HUMAN C-terminal β-hairpin + 3H	2E6R HLSD-5D KDM5D_HUMAN 1H + β-hairpin + 1H	

The study group consists of jumonji proteins, including JARID2 JARID1A, JARID1B, and JARID1D. Three of them (2JXJ, 2EQY and 3GL6) are also listed in the DisProt database. Fragments identified as disordered comprise (among others) the β-hairpin loop which mediates the biological properties of the target proteins. Our goal is to determine the status of these loops with respect to the presented hydrophobic core construction model.

Proteins under consideration in this work listed in Table 2 are mostly represented by structures determined by NMR with 20 structural models deposited in PDB. Such models capture the dynamic conformational properties of proteins and therefore some fragments of polypeptide chains exhibit structural changes. In order to ensure the validity of FOD analysis results, each protein is assessed in terms of its structural variability shown in Table 3.

Table 3. RMS-D values for 20 structural models stored in PDB files. For each protein the corresponding structures were determined by NMR. Due to the presence of loose (N-terminal and C-terminal) fragments, the compact fragment was analyzed as a separate unit, with the rightmost column listing its RMS-D values. No structural differentiation is observed in 2KGG and 2KGI.

Complete Molecule	Protein	Compact Part	
RMS-D		RMS-D	Fragment
1.67	2JXJ	0.804	88-175
11.74	2EQY	0.741	97-187
3.08	2YQE	0.598	83-168
3.22	2RQ5	0.577	618-728
0.45	2KGG		
0.78	2KGI		
3.72	2MA5	1.679	17-56
14.05	2E6R	0.116	16-68

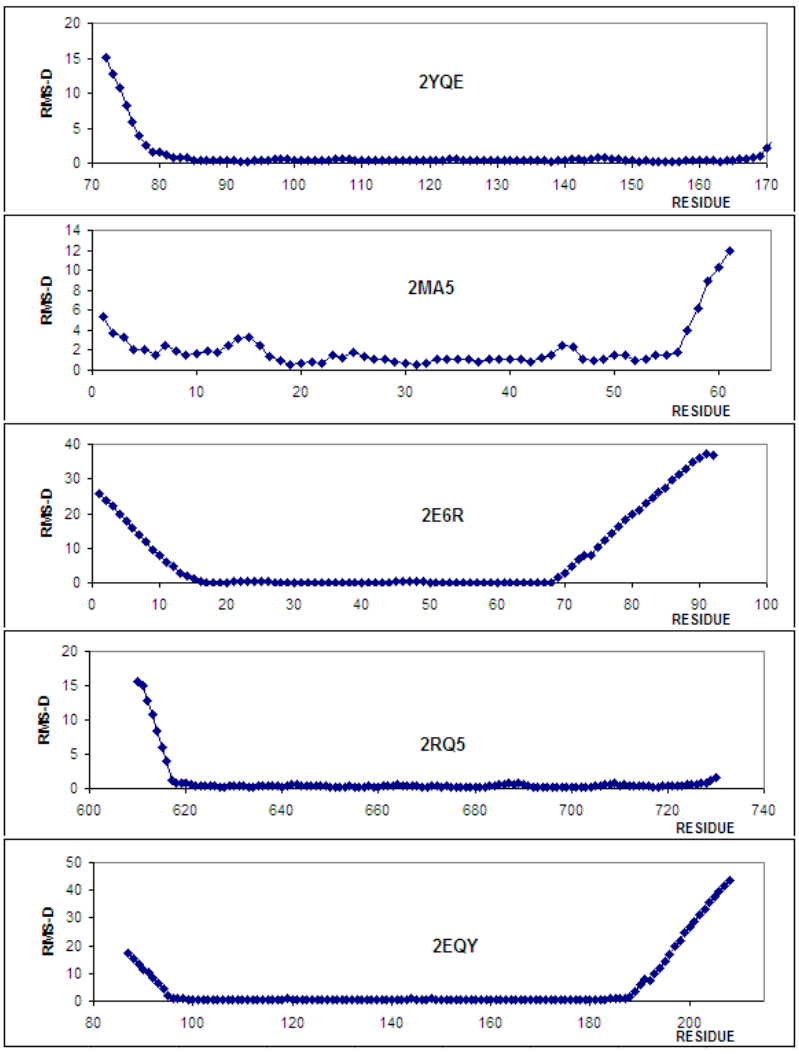

Figure 4. The RMS-D values for Cα atoms for proteins (20 models) of high global RMS-D values revealing the high stability of packed part of protein and relatively high RMS-D values for outstanding fragments.

Data supplied in Table 3 indicates low structural variability for 2KGG and 2KGI. The remaining proteins include fragments loosely bound to the compact fragment; however, the compact fragment itself remains mostly unchanged. For this reason both parts of the molecule will be analyzed separately further on in this work. The RMS-D values calculated for 20 models for Cα atoms visualise the specific flexibility of N- and C-terminal fragments. The compact part is easy to be distinguished (Figure 4).

3.2. DisProt Database

As already mentioned above, from among the proteins selected for analysis, three—2JXJ, 2EQY and 3GL6—also appear in the DisProt database [23]). This database specifically lists "protein sequences that do not assume a defined structural motif such as a β-pleated sheet or an α-helix in isolation, but may assume many conformations in association with other proteins or factors" (citation from Reference 48). Proteins which include intrinsically disordered fragments (according to DisProt criteria) will be further tagged with the "DisProt" keyword.

The aim of this work is to determine the status of such variable fragments with regard to the fuzzy oil drop model. In addition, based on structural resemblance to DisProt proteins and the presence of loose, poorly ordered secondary folds, we have identified certain other fragments as "Unstructured" in order to study their involvement in shaping the common hydrophobic core. This classification is based on the RMS-D profile shown in Figure 4.

4. Results

The proteins comprising this current study group have been subjected to analysis focusing on the status of their secondary folds and disordered fragments (including, in particular, those listed in the DisProt database). The fuzzy oil drop model posits a hydrophobic density peak near the center of the molecule, along with an encapsulating hydrophilic "shell", with near-zero hydrophobic density values on the surface. Fragments identified as accordant are those for which the actual (observed) hydrophobic density matches theoretical (idealized) values, regardless of their placement in the protein body.

Our analysis singles out well-defined secondary folds and loops, including the β-hairpin loop responsible for interactions with DNA. Intrinsically disordered fragments (as listed in DisProt), along with those deemed unstructured on the basis of the authors' subjective opinion, are analyzed in the context of the fuzzy oil drop model.

The fuzzy oil drop model is intended for analysis of globular proteins—thus, in the case of proteins which comprise elongated appendages or protrusions, our analysis will consider both the complete domain (tagged "C" in the result tables) as well as its globular subsection (tagged "P").

4.1. ARID Domains

This group of proteins (AT-rich domain) comprises 2EQY, 2JXJ, 2RQ5 and 2YQE, two of which (2JXJ and 2EQY) are also listed in the DisProt database. Each of these domains will be characterized by determining its $O|T$, $O|\mathrm{R}$ and RD coefficients.

The ARID domain in 2EQY appears accordant with the theoretical model despite the presence of three helixes (out of six) in which the hydrophobic density profile differs from idealized values (Table 4).

Good agreement between the observed and idealized hydrophobic density distributions can be observed for the complete domain as well as for its portion deprived of unstructured fragments. The fragment which the DisProt database identifies as disordered turns out to match the expected hydrophobic density values with good accuracy. On the other hand, two additional fragments, which have been arbitrarily classified as unstructured, do not correspond to the model (in the complete domain)—probably as a result of their conformation which significantly distorts the domain's globular structure.

Due to the presence of long, loosely packed fragments (the N- and C-terminal fragments), we have restricted our analysis to the globular portion of the molecule (96-188 aa) in order to determine the influence of such loose fragments upon the protein's hydrophobic core. As it turns out, elimination of the unstructured fragments only alters the status of the helix at 99-115 and the loop at 116-120.

Table 4. $O|T$, $O|$R and *RD* values for 2EQY. The results are also listed for selected secondary folds and unstructured fragments (while 2YQE is not present in the DisProt database, we have identified unstructured fragments as described in the Methods section): α—α-helix, β—β-twist, and L—loop. "C" denotes the complete molecule, as listed in PDB (87-208 aa), while "P" corresponds to the fragment subjected to fuzzy oil drop analysis (95-190 aa) following elimination of N- and C-terminal fragments. Values listed in boldface indicate departures from the idealized distribution. The values given in italics are the results of calculation using Kyte-Doolittle hydrophobicity scale.

	Secondary	***O|T***		***O|R***		***RD***	
	Fragment	**C**	**P**	**C**	**P**	**C**	**P**
2EQY		0.283	0.223	0.305	0.251	0.481 *0.497*	0.471 *0.471*
Unstructured	**87-95**	**0.168**		**0.031**		**0.843**	
	α 96-115	**0.228**	0.159	**0.187**	0.182	**0.549**	0.466
	L 116-120	0.172	**0.217**	0.276	**0.201**	0.383	**0.519**
	β 121-124	0.041	0.034	0.107	0.107	0.278	0.243
	β 125-128	0.128	0.158	0.642	0.642	0.166	0.198
	α 129-141	**0.210**	**0.189**	**0.176**	**0.176**	**0.544**	**0.517**
	α 142-149	0.158	0.142	0.257	0.257	0.381	0.355
	α 151-159	0.145	0.127	0.297	0.366	0.328	0.258
	L 160-164	0.072	0.028	0.133	0.160	0.350	0.148
	α 165-178	**0.262**	**0.206**	**0.190**	**0.195**	**0.579**	**0.514**
	α 179-188	**0.194**	**0.159**	**0.047**	**0.120**	**0.805**	**0.570**
Unstructured	**189-208**	**0.375**		**0.321**		**0.538**	
β-hairpin	120-129		0.103		0.298		0.258
DisProt	115-128	0.172	0.225	0.276	0.276	0.384	0.449

We should also note the presence of a β-hairpin in the DisProt area. The β-hairpin itself remains accordant with the model both when analyzed as part of the complete molecule and in the scope of its globular portion (without unstructured fragments).

The second protein listed in DisProt is 2JXJ. Its properties are briefly characterized in Table 5. The 2JXJ structure contains an AT-rich DNA binding domain required for RBP2 demethylase activity that, in turn, calls for specific identification of DNA strands in order to regulate transcription.

According to the DisProt database, the 103-116 fragment in 2JXJ is intrinsically disordered. Table 5 also reveals the status of this protein's exon-encoded fragments (only one such fragment is present in the structure under consideration) in relation to the fuzzy oil drop model.

In this protein, the DisProt fragment comprises a portion of the β-hairpin. When considered as a separate unit, this fragment exhibit good accordance with the model despite discrepancies affecting the individual β-structural fragments which form the β-hairpin loop, and, particularly, the 108-110 fragment (the other β-structural fragment at 113-115 more closely approximates theoretical values). The β-hairpin by itself is highly accordant with the idealized hydrophobic density distribution, indicating that it participates in this protein's hydrophobic core (Table 5).

The ARID domain in 2JXJ is also accordant with the model, even though three of its six helixes diverge from it. Additionally, one of two β-structural fragments (part of the β-hairpin) is identified as discordant.

Table 5. Status of selected ordered and disordered fragments in 2JXJ. Values listed in boldface indicate departures from the idealized distribution: α—α-helix, β—β-structural fragment, and L—loop. The values given in italics are the results of calculation using Kyte-Doolittle hydrophobicity scale.

	Secondary Fragment	***O|T***	***O|R***	***RD***
2JXJ	Complete Molecule	0.206	0.248	0.453 *0.473*
	α 84-102	**0.188**	**0.184**	**0.505**
	β 108-110	**0.046**	**0.033**	**0.577**
	β 113-115	0.068	0.372	0.156
	α 117-129	**0.192**	**0.125**	**0.606**
	α 130-137	0.232	0.465	0.333
	α 138-147	0.114	0.333	0.256
	α 153-163	**0.161**	**0.115**	**0.582**
	α 166-174	0.080	0.088	0.477
β-hairpin	107-115	0.081	0.269	0.233
DisProt	103-116	0.176	0.235	0.428
EXON	**81-121**	**0.207**	**0.201**	**0.507**

The exon fragment is interesting due to including a complete DisProt fragment with an *RD* value narrowly in excess of 0.5. Given that this fragment is recognized as "intrinsically disordered", such discordance is relatively unremarkable and the fragment as a whole can be described as contributing to the protein's hydrophobic core.

The next ARID protein subjected to analysis is 2YQE, presented in Table 6. The ARID domain in 2YQE diverges from the theoretical hydrophobic density distribution, most likely due to the presence of the highly discordant fragment at 72-79 (which is also listed as unstructured). Additionally, three out of six helical fragments diverge from the model.

Table 6. *O*|*T*, *O*|*R* and *RD* values for 2YQE. The results are also listed for selected secondary folds and unstructured fragments (while 2YQE is not present in the DisProt database, we have identified fragments that do not participate in its globular structure): α—α-helix, β—β-structural fragment and L—loop. Values listed in boldface indicate departures from the idealized distribution. The values given in italics are the results of calculation using Kyte-Doolittle hydrophobicity scale.

	Secondary Fragment	***O*|*T***	***O*|*R***	***RD***
2YQE	Complete Molecule	**0.255**	**0.231**	**0.524** ***0.500***
Unstructured	**72-79**	**0.231**	**0.025**	**0.903**
	α 80-97	**0.422**	**0.223**	**0.654**
	β103-106	0.036	0.108	0.264
	β 107-110	0.138	0.637	0.179
	α 111-123	**0.146**	**0.109**	**0.573**
	α 124-131	0.125	0.327	0.277
	α 133-141	0.139	0.244	0.362
	L 142-147	0.123	0.214	0.364
	α 148-160	**0.183**	**0.180**	**0.504**
	α 161-168	0.088	0.090	0.493
β-hairpin	103-110	0.098	0.349	0.219
Unstructured	98-110	0.149	0.283	0.344

Another representative of the ARID group in the study set is the protein designated 2RQ5 (Table 7).

The ARID domain, as it appears in 2RQ5, exhibits good accordance with the model. Unlike the proteins discussed above, this domain includes a β-hairpin loop. Its individual β-structural fragments diverge from the theoretical distribution, although the structure as a whole remains accordant—despite the presence of an unstructured fragment which is a poor match for the fuzzy oil drop model.

One of the differences between 2RQ5 and 2JXJ involves the presence of a helix in the C-terminal fragment of this domain. In order to facilitate comparative analysis we have performed calculations for the 2RQ5 domain truncated to 610-710 aa. The results are listed in Table 7. ("P" column). This fragment exhibits better accordance with the model compared with the whole domain. The β-hairpin in 2RQ5 differs from the theoretical distribution, suggesting that it does not contribute to stabilization of the complete molecule or its packed fragment (610-710 aa).

Summarizing the status of ARID proteins, we should note the differences in the β-hairpin fragments; *i.e.*, two accordant folds in 2YQE and 2EQY, one accordant fold in 2JXJ and no accordant folds in 2RQ5. In the case of 2YQE, the entire unstructured fragment remains accordant, much like the DisProt fragment in 2JXJ—this means that fragments classified as disordered on the basis of structural (geometric) criteria may still contribute to the creation of a stable hydrophobic core. Of all the analyzed proteins, only 2RQ5 comprises a β-hairpin that does not enhance structural stabilization. Figure 5 provides a comparative overview of ARID proteins. We should note similarities in the conformation of helices identified as discordant from the fuzzy oil drop model—such similarities are due to the identical structural arrangement of the β-hairpin (even though not all software packages recognize this fragment as a random coil—as illustrated in Figure 5).

Table 7. *O|T*, *O|R* and *RD* values for 2RQ5. "C" and "P" correspond to the complete molecule (610-730 aa) and its part (610-710 aa) respectively. The results are also listed for individual secondary folds and for disordered fragments: α—α-helix, β—β-structural fragment, and L—loop. Values listed in boldface indicate departures from the idealized distribution. The values given in italics are the results of calculation using Kyte-Doolittle hydrophobicity scale.

	Secondary	**O|T**		**O|R**		**RD**	
	Fragment	**C**	**P**	**C**	**P**	**C**	**P**
2RQ5	Chain	0.207	0.182	0.259	0.254	0.444 *0.475*	0.417 *0.479*
	α 620-636	0.126	0.147	0.169	0.174	0.427	0.426
Unstructured	**637-643**	**0.300**		**0.257**		**0.539**	
	β644-646	**0.092**	**0.129**	**0.041**	**0.041**	**0.690**	**0.628**
	β649-651	**0.241**	**0.069**	**0.126**	**0.126**	**0.656**	**0.642**
	α 652-663	**0.174**	**0.227**	**0.138**	0.139	**0.556**	0.424
	α 665-672	0.073	0.102	0.203	0.204	0.263	0.332
	α674-682	0.102	0.101	0.391	0.391	0.207	0.246
	L 683-689	0.355	0.341	0.458	0.458	0.437	0.427
	α690-701	0.097	0.128	0.189	0.188	0.339	0.304
	α 702-709	**0.117**	**0.083**	**0.078**	**0.056**	**0.600**	**0.601**
	α710-727	0.130		0.191		0.405	
β-hairpin	643-652	**0.268**	**0.233**	**0.106**	**0.106**	**0.716**	**0.687**

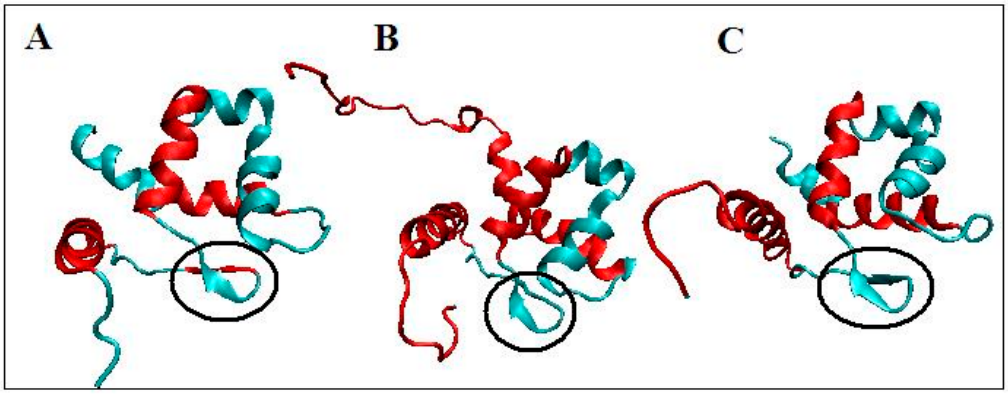

Figure 5. ARID proteins: (**A**): 2JXJ, (**B**): 2EQY, (**C**): 2YQE. The red fragments exhibit departures from the idealized hydrophobic density distribution. The black ellipse marks the β-hairpin fragment in 2EQY and 2YQE respectively. In 2EQY and 2YQE both folds comprising the β-hairpin loop remain consistent with the model, while in 2JXJ only one fragment retains this property.

4.2. PHD Domains

This group comprises proteins associated with the so-called PHD domain, including 2E6Q and 2MA5. In the case of 2E6Q, we have performed the fuzzy oil drop analysis both for the entire molecule and for its globular fragment (25-69 aa), eliminating the unstructured N- and C-terminal fragments.

Table 8. $O|T$, $O|R$ and RD values for 2E6R. The results are also listed for selected secondary folds and unstructured fragments (while 2E6R is not present in the DisProt database, we have identified fragments not associated with the globular structure): α—α-helix, β—β-structural fragment. Values listed in boldface indicate departures from the idealized distribution. "P" denotes the fragment 25-69 aa. The values given in italics are the results of calculation using Kyte-Doolittle hydrophobicity scale.

	Secondary	***O|T***		***O|R***		***RD***	
	Fragment	**C**	**P**	**C**	**P**	**C**	**P**
2E6R		**0.513**	0.164	**0.227**	0.227	**0.649** ***0.580***	0.420 *0.412*
Unstructured	**1-19**	**0.640**		**0.237**		**0.730**	
	α 25-31	0.088	0.111	0.318	0.267	0.216	0.293
Zn^{2+} (34)	**β 32-34**	**0.154**	0.153	**0.147**	0.159	**0.510**	0.490
	β 39-41	**0.089**	**0.023**	**0.014**	**0.016**	**0.861**	**0.584**
Unstructured	**42-60**	**0.402**	0.065	**0.229**	0.161	**0.637**	0.289
Zn^{2+} (60, 63)	α 61-69	0.179	0.180	0.246	0.224	0.421	0.445
β-hairpin	**31-43**	**0.356**	**0.161**	**0.085**	**0.105**	**0.807**	**0.606**
Unstructured	**79-92**	**0.551**		**0.142**		**0.795**	

2E6R is characterized by an abundance of loose fragments, with only one short fragment of the chain representing a tightly packed domain (Table 8). This low degree of packing is characteristic of proteins that cannot be encapsulated in a simple 3D Gaussian. Even so, the β-hairpin is identified as discordant while the two helical fragments remain accordant. Calculating the *RD* values for individual secondary folds following the elimination of the unstructured N- and C-terminal fragments confirms a change in status of the β-structural at fragment 32-34 (which forms part of the β-hairpin), as well as the unstructured fragment at 42-60. The packed portion of the complete molecule is notably discordant, in a way similar to the other molecules under consideration. The relatively long unstructured fragment exhibits good accordance with the theoretical model. It seems that even "unstructured" folds may still remain "structured" in the sense of the fuzzy oil drop model. However, the packed fragment of the domain under consideration is shown to conform to theoretical expectations regarding the hydrophobic density distribution.

Particularly good agreement between the actual and theoretical hydrophobic density distribution is observed in 2MA5. Values listed in Table 9 indicate a high level of accordance both with respect to the molecule as a whole and its individual secondary folds, including loose fragments. This is an interesting example of a domain in which each part contributes to a common hydrophobic core (note that in the sense of the fuzzy oil drop model, the concept of a "hydrophobic core" includes a hydrophilic shell which protects the central portion of the molecule from contact with water). According to this interpretation, both the highly hydrophobic center and the external shell are equally important in

ensuring a distribution of hydrophobic residues, which is consistent with theoretical expectations. Thus, we can conclude that 2MA5 should be characterized by a high solubility and a low propensity for interaction with other molecules.

Table 9. *O*|*T*, *O*|*R* and *RD* values for 2MA5 (2MA5 is not present in DisProt database): α—α-helix, β—β-structural fragment, and L—loop. The values given in italics are the results of calculation using Kyte-Doolittle hydrophobicity scale.

	Secondary Fragments	***O*|*T***	***O*|*R***	***RD***
2MA5	Complete Molecule	0.164	0.281	0.368 *0.333*
Unstructured Zn^{2+} 4, 9	1-17	0.079	0.193	0.290
Zn^{2+} 22	β 18-22	0.043	0.167	0.206
Zn^{2+} 26	L 23-27	0.119	0.400	0.230
Zn^{2+} 31	β 28-31	0.026	0.068	0.279
Zn^{2+} 34	α 32-36	0.113	0.119	0.486
	α 38-44	0.132	0.286	0.316
	L 45-48	0.022	0.051	0.301
Zn^{2+} 49, 52	α 49-55	0.190	0.389	0.327
β-hairpin	18-32	0.048	0.096	0.333

Both proteins share a distinct capability to bind zinc ions. The presence of (abundant) ions does not seem to significantly alter the hydrophobic core structure in 2MA5. From the point of view of the fuzzy oil drop model, ions do not register as deformations in the core structure (e.g., unlike large ligands or external proteins). We refer to this as "static" binding, where the ion aligns itself with the existing structure without deforming it. (In contrast, "dynamic" binding refers to a situation where the ligand alters the conformation of the target protein, affecting its hydrophobic core).

It should be noted that an *RD* value under 0.4 is fairly rare—this suggests that the protein reaches its final conformation before it has bound any ions, and that ions take no part in the folding process (note, however, that this observation is only speculative and follows from analysis of data obtained by applying the fuzzy oil drop model). The presented conclusion is based on the assumption that the water environment alone enables the protein to reach its native form. If the presence of ions was a requirement in this process, the final structure would be expected to diverge from theoretical expectations.

Regarding the PHD group, 2MA5 is unique in that it entirely conforms to the fuzzy oil drop model—despite its large disordered N-terminal fragment. This molecule is also characterized by a relatively long loop between the individual folds comprising its β-hairpin. Regarding 2E6R, two long disordered N- and C-terminal fragments cause the molecule (as a whole) to diverge from the model, although its small compact fragment remains accordant (Figure 6).

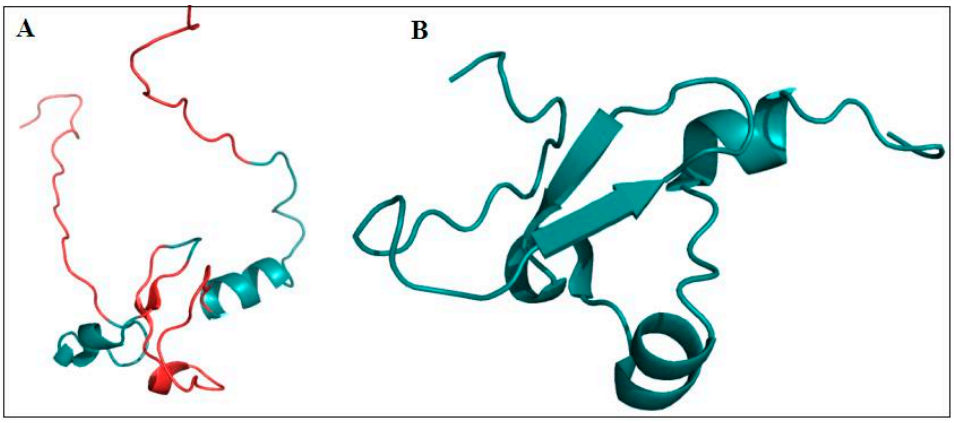

Figure 6. Proteins representing the PHD group: (**A**): 2E6R; (**B**): 2MA5. The red fragments diverge from the theoretical hydrophobic density distribution.

4.3. C-terminal PHD Finger

This group comprises proteins designated 2KGG, 2KGI and 3GL6. The two latter molecules include a peptide responsible for histone complexation (which corresponds to their biological function). As such, it is possible to analyze the status of individual fragments both in the isolated molecule and in the protein-histone complex.

Table 10. *O*|*T*, *O*|*R* and *RD* values for 2KGG. The results are also listed for selected secondary folds and unstructured fragments (2KGG is not present in the DisProt database): α—α-helix, β—β-structural fragment, and L—loop. Values listed in boldface indicate departures from the idealized distribution. The values given in italics are the results of calculation using Kyte-Doolittle hydrophobicity scale.

	Secondary Fragment	***O*|*T***	***O*|*R***	***RD***
2KGG	Complete Molecule	0.226	0.292	0.436 *0.388*
Unstructured	1-16	0.154	0.398	0.279
	β 17-21	**0.293**	**0.267**	**0.523**
Zn^{2+} 25	**L 22-26**	**0.125**	**0.093**	**0.572**
Zn^{2+} 30	β 27-31	0.030	0.096	0.238
	L 32-36	0.039	0.069	0.366
	α 37-44	0.115	0.283	0.288
β-hairpin	**17-31**	**0.283**	**0.153**	**0.649**

Visual inspection of 2KGG suggests a very low degree of packing (Table 10). Nevertheless, the protein conforms to the theoretical hydrophobic density distribution, with only two exceptions (one β-structural fragment comprising the β-hairpin and a loop which interacts with one of two zinc ions present in the structure). In particular, the disordered fragment remains accordant with the model despite its considerable length (16 aa).

The second representative of this group, 2KGI, was analyzed both as a standalone molecule and as a complex (Table 11). Chain B is the histone fragment to which 2KGI specifically binds.

Table 11. *O*|*T*, *O*|*R* and *RD* values for 2KGI. The results are also listed for selected secondary folds and unstructured fragments (2KGI is not present in the DisProt database): α—α-helix, β—β-structural fragment, and L—loop. Values listed in boldface indicate departures from the idealized distribution. The values given in italics are the results of calculation using Kyte-Doolittle hydrophobicity scale.

	Secondary Fragment	***O*|*T***	***O*|*R***	***RD***
2KGI-AB	Complex	0.210	0.273	0.435 *0.388*
Unstructured	1-14	0.052	0.091	0.365
L 17-21	β 17-21	0.141	0.301	0.319
L 28, 31 Zn^{2+} 30	**β 27-31**	**0.222**	**0.155**	**0.589**
	L 32-36	0.078	0.245	0.243
L 41, 44-45	α 37-44	0.032	0.081	0.285
	L 45-52	0.143	0.166	0.463
Chain B		0.105	0.191	0.355 *0.431*
β-hairpin	14-31 301-308	0.270	0.283	0.488
β-hairpin	**14-31**	**0.279**	**0.196**	**0.801**
2KGI-A	CHAIN	0.261	0.271	0.491 *0.364*
Unstructured	1-14	0.144	0.323	0.308
L 17-21	**β 17-21**	**0.318**	**0.223**	**0.587**
L 28, 31 Zn^{2+} 30	β 27-31	0.048	0.091	0.347
	L 31-36	0.110	0.143	0.435
L 41, 44-45	α 37-44	0.116	0.221	0.344
	L 45-52	0.234	0.268	0.466
β-hairpin	**14-31**	**0.359**	**0.258**	**0.582**

Further analysis of 2KGI points to one of the β-structural fragments comprising the β-hairpin loop as discordant from the model. This recurring phenomenon indicates that the delicate balance between the stability of one part of the β-hairpin and the instability of the other part is a prerequisite of its biological activity (in this case—interaction with a histone). While both β-structural fragments interact with the histone, only one of them exhibits accordance with the fuzzy oil drop model. The situation changes in the protein complex where the status of both fragments is reversed.

The unstructured fragments again prove consistent with the model.

Much like 2KGI, 3GL6 can be studied either as a standalone molecule or in complex with a histone. A large portion of this protein consists of discordant fragments, which also appear in its tightly packed section (see Tables 12 and 13).

Table 12. *O*|*T*, *O*|*R* and *RD* values for 3GL6, calculated for the protein-histone complex, for each of its components and for selected secondary folds. Values listed in boldface indicate departures from the idealized distribution. The DisProt database qualifies the entire A chain as disordered: α—α-helix, β—β-structural fragment, and L—loop. The "L" tags in the leftmost column signify interaction with the ligand (polypeptide chain contributed by the histone). The associated number corresponds to the number of residues involved in interaction for each fragment.

	Secondary Fragment	***O*|*T***	***O*|*R***	***RD***
3GL6-COMPLEX	Complete Molecule	**0.367**	**0.208**	**0.638**
	Chain A	**0.362**	**0.202**	**0.641**
	Chain B	**0.282**	**0.082**	**0.772**
Unstruct. 4 Zn^{2+}	**1608-1623**	**0.547**	**0.259**	**0.678**
1 Zn^{2+}4 L	**β 1624-1628**	**0.227**	**0.125**	**0.643**
1 L	**β 1634-1638**	**0.039**	**0.036**	**0.519**
	α 1639-1642	**0.025**	**0.009**	**0.729**
3 L	α 1644-1651	0.091	0.251	0.267
2 Zn^{2+}	**L 1652-1659**	**0.290**	**0.170**	**0.618**
β-hairpin	**1621-1638**	**0.023**	**0.014**	**0.629**
β-hairpin + ligand	**1621-1638 1-8**	**0.277**	**0.177**	**0.610**

Table 13. *O*|*T* ,*O*|*R* and *RD* values for 3GL6 (chain A), for its selected secondary folds and for disordered fragments. This protein is listed in the DisProt database, with the entire molecule flagged as disordered. Two unstructured fragments are also distinguished: α—α-helix, β—β-structural fragment, and L—loop. "C" and "P" correspond to the complete molecule (1608-1659 aa) and its part (1608-1644 aa) respectively. Values listed in boldface indicate departures from the idealized distribution. The values given in italics are the results of calculation using Kyte-Doolittle hydrophobicity scale.

	SecondaryFragment	***O*|*T***		***O*|*R***		***RD***	
	Chain	C	P	C	P	C	P
3GL6	Complete Molecule	**0.381**	**0.439**	**0.231**	**0.190**	**0.623** ***0.614***	**0.698** ***0.677***
Unstruct.4 Zn^{2+}	**1608-1623**	**0.605**	**0.570**	**0.320**	**0.320**	**0.654**	**0.641**
1 Zn^{2+}; 4 L	**β 1624-1628**	**0.264**	**0.339**	**0.173**	**0.103**	**0.604**	**0.766**
1 L	**β 1634-1638**	**0.036**	**0.080**	**0.028**	**0.026**	**0.563**	**0.755**
	α 1639-1642	**0.023**	**0.057**	**0.009**	**0.008**	**0.708**	**0.872**
3 L	α 1644-1651	0.097		0.265		0.269	
2 Zn^{2+}	**L 1652-1659**	**0.275**		**0.189**		**0.593**	
β-hairpin	**1608-1644**	**0.311**	**0.507**	**0.215**	**0.074**	**0.591**	**0.743**
DisProt	**1609-1659**	**0.357**	**0.438**	**0.183**	**0.190**	**0.662**	**0.698**

To enable meaningful comparative analysis, we have selected the 1608-1644 fragment, thus adapting the structure of 3GL6 to that of 2KGG and 2KGI. Surprisingly, despite this modification, significant differences persist, both with respect to the domain as a whole and to each of its fragments: 3GL6 continues to diverge from the model whereas 2KGI and 2KGG satisfy $RD < 0.5$, with only one β-fold (part of the β-hairpin loop) identified as discordant.

According to the fuzzy oil drop model, interaction with an external protein may locally distort the hydrophobic core. This is not evident in 3GL6. Elimination of all residues involved in external interactions (ions + polypeptide contributed by the histone) does not change the status of the remainder of the molecule. Thus, any discrepancies between the idealized and observed hydrophobic density distribution are not due to interaction with ligands.

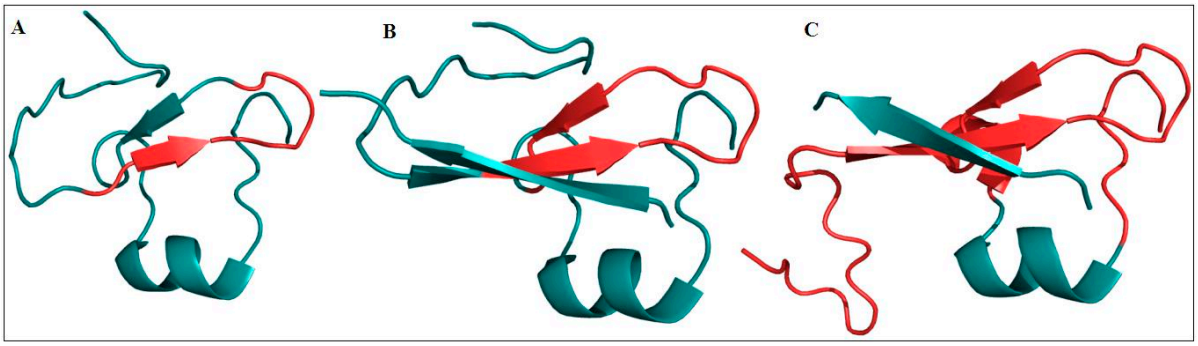

Figure 7. Proteins which include the C-terminal PHD finger: (**A**): 2KGG; (**B**): 2KGI; (**C**): 3GL6. The red fragments diverge from the theoretical hydrophobic density distribution as predicted by the fuzzy oil drop model. The fold seen in the foreground of Figure 6B and 6C is contributed by the interacting histone.

In summary, proteins which contain C-terminal PHD fingers seem to exhibit variations in their status: in 2KGG and 2KGI the entire molecule conforms to the theoretical distribution while, in 3GL6, notable differences can be observed. Similar variability applies to individual twists which comprise the β-hairpin—monomers often differ from complexes in this regard, suggesting that the β-hairpin plays an important role in protein-ligand interactions (Figure 7).

Analysis of fragments associated with the biological function of mutants of lysozymes which either retain or lose enzymatic activity (with a publication currently underway) points to the need for a subtle equilibrium regarding the status of individual folds. This equilibrium is particularly important for dynamic fragments which comprise the active site. Stabilization of one fold coupled with destabilization of another introduces a degree of variability, which—in turn—determines the protein's biological properties (it should be noted, however, that this observation is only a speculative consequence of data obtained by applying the fuzzy oil drop model.)

4.4. Comparative Analysis

4.4.1. Sequence Analysis

To complement our comparative study we have also analyzed various types of β-hairpins present in the listed proteins.

Table 14. Percentage of values expressing the sequential similarity of β-hairpins present in ARID domains. Results obtained using the ClustalW 2.1 package [49].

	2JXJ	2EQY	2YQE
2EQY	88.89		
2YQE	87.50	87.50	
2RQ5	11.11	30.00	12.50

Analysis of the sequential similarity of β-hairpins in ARID domains indicates that only 2RQ5 substantially differs from other proteins in this regard (Table 14). This is likely due to the fact that 2RQ5 comes from a different species.

Table 15. Degree of sequential similarity (identity) of residue sequences in the PHD domain. Results obtained using the ClustalW 2.1 package [49].

	2KGG-complex 2KGI 3GL6-complex	2E6R
2MA5	80.00	23.08
2E6R	30.77	

Regarding the PHD domains, 2KGG, 2KGI and 3GL6 (identical sequences) are all highly similar to 2MA5 (80.0% similarity) despite the fact that 2MA5 is the only protein in this table which is derived from *Mus musculus* (Table 15).

4.4.2. Different Hydrophobic Scales in the Fuzzy Oil Drop Model

The hydrophobicity scale used in the calculation was defined using the distance between the position of particular amino acid and the center of the ellipsoid ([42], also see Table 1). However the fuzzy oil drop model can be used with any hydrophobicity scale. A comparative analysis was performed to evaluate the influence of the hydrophobicity scale defined by Kyte and Doolittle [35]. Tables 4–12 give the values of RD calculated according to the Kyte-Doolittle scale [35]. A comparison of these values suggests sufficient high accordance in the final results. The most important is the common interpretation of the status of the protein under consideration despite using different hydrophobicity scales to calculate the RD. The Supplementary Tables (Tables S1–S9) show this observation also with respect to the status of the selected fragments of polypeptides. The comparison of these values also suggests consistency of the results based on different hydrophobicity scales applied to fuzzy oil drop model.

The value of the hydrophobic parameter that is constant for a particular residue appears to be modified in the protein molecule. It is a result of the influence of the local environment. The hydrophobic parameters can be treated as primary (I-order) hydrophobic structure in analogy to primary (I-order) structure. The values of O_i can be interpreted as tertiary (III-order) hydrophobic structure of the protein. It is like the tertiary structure of the hydrophobic structure. The values called "observed (O_i)" expresses the collection of hydrophobic interactions with the surrounding residues. The extent to which the parameters are changed is shown in Figure 8.

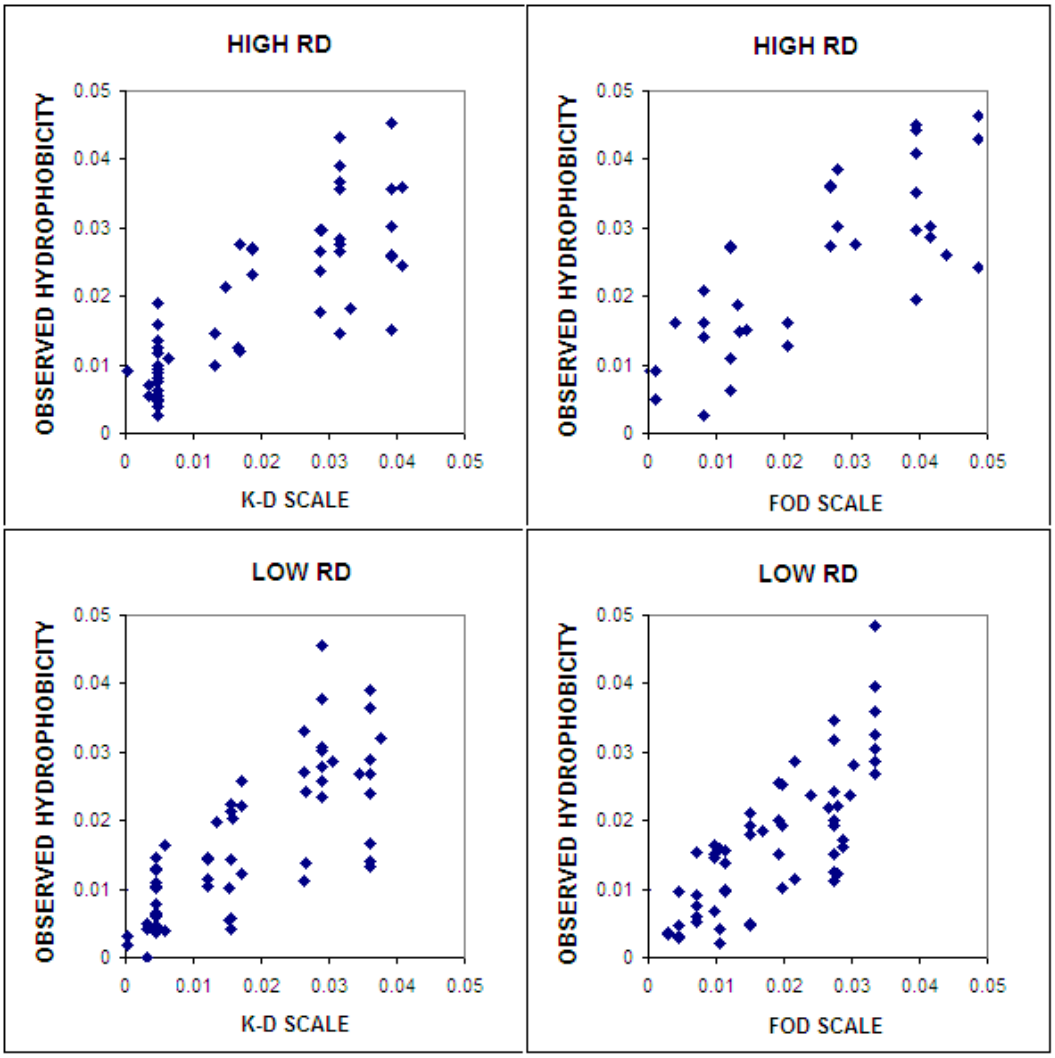

Figure 8. The relation between hydrophobic parameters (I-order of hydrophobic structure) and the status of residue in the fuzzy oil drop environment for protein (III-order of hydrophobic structure) with high (2MA5) and low (3GL6) accordance of hydrophobic core with the fuzzy oil drop model.

4.4.3. Structural Analysis

Table 16 presents a summary of the status of the individual domains, listing the specific features of their hydrophobic cores. Computing the status of the complete chain indicates that JARID1D (both ARID and PHD) domains appear to represent the status of discordance with respect to the fuzzy oil drop model as well as 3GL6 JARID1A) in the group of PHD domains. The β-hairpin loop is identified in ARID domain of 2RQ5 as discordant with respect to fuzzy oil drop model. This amino acid sequence in this domain is significantly different with respect to all the others in the group of ARID domains. Two domains in the PHD group represent the presence of an ordered hydrophobic distribution in the β-hairpin loop—they are 2KGG and 2MA5. One can conclude that the ARID domains represent a more frequent distribution of hydrophobic density of regular ordered structure in form of a hydrophobic core.

Two of the three fragments recognized as DisProt appeared to represent the hydrophobic density distribution of the ordered form. It means that their unstructured form fits well to the ordered form of

hydrophobic distribution and in consequence one may conclude they participate in the construction of a stable hydrophobic core in this domain.

The presence of Zn^{2+} ions is expected to disturb the regular distribution of hydrophobic density due to a significant introduction of strong electrostatic interactions. However, in the analyzed domains, this is not always the case (bottom part of Table 16). A very high accordance with the hydrophobic distribution is observed for 2E6R (fragment 25-69). Additionally, the positions of the residues engaged binding of the ions (despite of their engagements in strong electrostatic interaction) represent a hydrophobic distribution accordant with the 3D Gauss distribution. It may be that the stability of this domain is supported by both—electrostatic as well as hydrophobic interactions.

Table 16. Summary of the status of ARID and PHD domains in jumonji proteins as determined using fuzzy oil drop model. Symbol "+" denotes presence or ordered hydrophobic core, symbol "−" absence of hydrophobic core. Symbol in parenthesis describes the compact part (the outstanding fragments eliminated form calculations).

	Complete Domain			
	JARID1A	JARID1B	JARID1D	JARID2
ARID	2JXJ +	2EQY +	2EQY −	2RQ5 +
PHD	2KGG + 2KGI + 3GL6 −	2MA5 +	2E6R −(+)	
	β-HAIRPIN LOOP			
	JARID1A	JARID1B	JARD1D	JARID2
ARID	2JXJ +	2EQY +	2YQE +	2RQ5 −
PHD	2KGG + 2KGI − 3GL6 −	2MA5 +	2E6R −	
	DisProt FRAGMENTS			
	JARID1A	JARID1B	JARID1D	JADIR2
ARID	2JXJ +	2EQY +		
PHD	3GL6 −			
	PRESENCE of Zn^{2+} ions			
	JARID1D	JARID1B	JARID1D	JARID2
PHD	2KGG −(+) 2KGI −(+) 3GL6 −	2MA5 +	2E6R −(+)	

One may also note that the complexation of the ligand (peptide contributed by the histone) further stabilizes the domain, with *RD* values of 0.491 and 0.436 for the individual protein and for the protein-ligand complex respectively.

Figure 6 depicts two fringe cases by visualizing the hydrophobic density profiles for 2MA5 and 3GL6. It reveals the differences between the theoretical and observed hydrophobic density distributions (particularly in the scope of the β-hairpin, which is marked by the sequence of dots on the horizontal axis). The *RD* values for these fragments are listed in Tables 9, 11 and 12.

Table 17. *RD* (average and standard deviation) values for all 20 models for proteins (structure determined by NMR technique) to measure the structural differentiation.

Complete Molecule		Protein	Compact Fragment		
Average	ST.DEV.		# AA	Average	ST.DEV.
0.544	0.062	2E6R	16-68	0.400	0.010
0.498	0.025	2EQY	97-187	0.450	0.015
0.484	0.018	2JXJ	88-175	0.428	0.014
0.298	0.027	2MA5	17-56	0.361	0.041
0.467	0.018	2RQ5	618-728	0.412	0.011
0.516	0.025	2YQE	83-168	0.408	0.010

The values of the averaged *RD* parameters and their standard deviation given in Table 17 show quantitative differences for individual models; however, qualitative interpretation of results is common (low values of standard deviation). It suggests that the overall status of the hydrophobic distribution is rather common independent of the local rearrangements (particularly for loose N- and C-terminal fragments). The encapsulation (size and shape fit to the particular structure) of the entire molecules makes the relative hydrophobic distribution comparable.

5. Discussion and Conclusions

In our to-date research, the fuzzy oil drop has been applied to determine the status of the hydrophobic core in various proteins and protein domains. As already suggested, individual domains typically match the theoretical model, indicating that—for the most part—they fold on their own. Local deformations in the hydrophobic core structure often correspond to highly specific ligand binding cavities or complexation sites. We have furthermore determined that structurally similar domains (such as immunoglobulin-like folds [33]) may exhibit a variable hydrophobic core status. Likewise, the status of disordered fragments varies: some of them exhibit good accordance despite their chaotic structural nature (confirmed by inclusion in the DisProt database).

The presented work shows that the β-hairpin which is responsible for DNA interactions remains accordant with the model in most ARID domains, while the opposite is usually true in PHD domains.

The strong stabilizing effect exerted by Zn^{2+} ions seems sufficient to ensure a proper conformation of the β-hairpins in the PHD domains. Surprising, however, is the residues engaged in ion binding (and stabilized by this ion) represents also a highly ordered hydrophobic density distributed accordant with the assumed model (according to 3D Gauss function for 25-69 aa in 2E6R $RD = 0.235$). The high accordance of observed hydrophobic density distribution in Zn^{2+} binding is due to the presence of three Cys residues. This residue is recognized by many hydrophobic scales as highly hydrophobic. Their symmetrical orientation in the central part of the loop results in a highly accordant order of the hydrophobic density distribution. It suggests possible support from the water environment to generate the stable system for Zn^{2+} complexation.

Fragments listed as disordered in the DisProt database comprise β-hairpin loops that undergo structural changes as a result of their biological activity (this is, in fact, one of the prerequisites for inclusion in DisProt). In the crystal structures of the presented proteins (except for 3GL6), these loops participate in the formation of a common hydrophobic core. Of particular note is the status of the β-hairpin

in 2KGI. When analyzing the domain as an individual unit, the loop diverges from the model, whereas, in the protein-ligand complex (the ligand being a peptide contributed by a histone), the loop remains accordant. This indicates that the β-hairpin adjusts its structure to accommodate the ligand. Such conclusions would not be possible without invoking Kullback-Leibler's divergence entropy criterion which enables us to determine the extent to which a given structure (in this case—the protein-ligand complex) approximates the theoretical distribution of hydrophobic density. This comparison seems to be critical for determining the applicability of the fuzzy oil drop model (as well as assessing the structural stability of target proteins).

Ongoing analysis provides further proof of the relation between accordance with the fuzzy oil drop model and the biological properties of a given structural unit (be it the β-hairpin or the entire domain).

One may consider also other models that relate tertiary conformation to the presence of an aqueous environment, although such models do not supply any quantitative criteria. For example, the "wet and dry area model" distinguishes places within the protein body where water cannot penetrate (hence the "dry" designation), as well as the "wet" areas which are exposed to water [50]. The model also underscores the need for a balance between both types of areas.

Another model which acknowledges the influence of water upon polypeptide chain folding is the nucleation model. "Nucleation" is understood as the emergence of a "seed" around which a hydrophobic core can coalesce [51]. Unlike the wet and dry area model, the nucleation model is dynamic in scope, describing the progression of the folding process. The fuzzy oil drop model represents an improvement upon this abstraction by proposing quantitative criteria that express the status and influence of the hydrophobic core upon various types of structural units—domains, proteins and complexes [52].

The discussion on the problem of hydrophobic/hydrophilic interaction between residues and its influence on the protein folding presented in [53] suggests that latter are more important than the former [54]. The structuralization of water in form of an iceberg is treated as important conditioning of the proper folding process [55,56]. The "fuzzy oil drop" model solves at least one problem introducing the unification of hydrophobicity/hydrophilicity in form of mathematical model.

Acknowledgments

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

Author Contributions

Leszek Konieczny designed the physical model. Irena Roterman designed mathematical model (application of 3D Gauss function), applied Kullback-Leibler entropy to measure the differences between theoretical, observed and unified distribution to quantitatively measure the differences between profiles. Barbara Kalinowska performed the search for disordered proteins and calculation using ClustalW program and prepared figures with 3D presentation of proteins under consideration (Figures 5–7). Mateusz Banach prepared the program to calculate the hydrophobic distributions and performed calculation of parameters based on fuzzy oil drop model for both applied hydrophobic scales.

Irena Roterman analyzed the results and prepared the final interpretation. Irena Roterman prepared Figures 1–4,8 and wrote the paper. All authors have read and approved the final manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Matsumura, M.; Wozniak, J.A.; Sun, D.P.; Matthews, B.W. Structural studies of mutants of T4 lysozyme that alter hydrophobic stabilization. *J. Biol. Chem.* **1989**, *264*, 16059–16066.
2. Van den Burg, B.; Dijkstra, B.W.; Vriend, G.; van der Vinne, B.; Venema, G.; Eijsink, V.G. Protein stabilization by hydrophobic interactions at the surface. *Eur. J. Biochem.* **1994**, *220*, 981–985.
3. Chang, B.S.; Mahoney, R.R. Enzyme thermostabilization by bovine serum albumin and other proteins: Evidence for hydrophobicinteractions. *Biotechnol. Appl. Biochem.* **1995**, *22*, 203–214.
4. Ventura, S.; Serrano, L. Designing proteins from the inside out. *Proteins* **2004**, *56*, 1–10.
5. Chattopadhyay, K.; Mazumdar, S. Stabilization of partially folded states of cytochrome c in aqueous surfactant: Effects of ionic and hydrophobic interactions. *Biochemistry* **2003**, *42*, 14606–14613.
6. Vondrásek, J.; Bendová, L.; Klusák, V.; Hobza, P. Unexpectedly strong energy stabilization inside the hydrophobic core of small protein rubredoxin mediated by aromatic residues: Correlated ab initio quantum chemical calculations. *J. Am. Chem. Soc.* **2005**, *127*, 2615–2619.
7. Gerstman, B.S.; Chapagain, P.P. Self-organization in protein folding and the hydrophobic interaction. *J. Chem. Phys.* **2005**, *123*, 054901.
8. Arunachalam, J.; Gautham, N. Hydrophobic clusters in protein structures. *Proteins* **2008**, *71*, 2012–2025.
9. Dong, H.; Mukaiyama, A.; Tadokoro, T.; Koga, Y.; Takano, K.; Kanaya, S. Hydrophobic effect on the stability and folding of a hyperthermophilic protein. *J. Mol. Biol.* **2008**, *378*, 264–272.
10. Dill, K.A.; MacCallum, J.L. The protein-folding problem, 50 years on. *Science* **2012**, *338*, 1042–1046.
11. Kauzmann, W. Some factors in the interpretation of protein denaturation. *Adv. Protein Chem.* **1959**, *14*, 1–63.
12. Konieczny, L.; Brylinski, M.; Roterman, I. Gauss function based model of hydrophobicity density in proteins. *In Silico Biol.* **2006**, *6*, 15–22.
13. Kullback, S.; Leibler, R.A. On Information and Sufficiency. *Ann. Math. Stat.* **1951**, *22*, 79–86.
14. Vucetic, S.; Obradovic, Z.; Vacic, V.; Radivojac, P.; Peng, K.; Iakoucheva, L.M.; Cortese, M.S.; Lawson, J.D.; Brown, C.J.; Sikes, J.G.; *et al*. DisProt: A database of protein disorder. *Bioinformatics* **2005**, *21*, 137–140.
15. Homepage of DisProt. Available online: http://www.disprot.org (accessed on 18 March 2015).
16. Uversky, V.N.; Dunker, A.K. Understanding protein non-folding. *Biochimica Biophysica Acta* **2010**, *1804*, 1231–1264
17. Bergé-Lefranc, J.L.; Jay, P.; Massacrier, A.; Cau, P.; Mattei, M.G.; Bauer, S.; Marsollier, C.; Berta, P.; Fontes, M. Characterization of the human jumonji gene. *Hum. Mol. Genet.* **1996**, *5*, 1637–1641.

18. Li, L.; Greer, C.; Eisenman, R.N.; Secombe, J. Essential Functions of the Histone Demethylase Lid. *PLoS Genet.* **2010**, *6*, e1001221
19. Liefke, R.; Oswald, F.; Alvarado, C.; Ferres-Marco, D.; Mittler, G.; Rodriguez, P.; Dominguez, M.; Borggrefe, T. Histone demethylase KDM5A is an integral part of the core Notch-RBP-J repressor complex. *Genes Dev.* **2010**, *24*, 590–601.
20. Sauvageau, M.; Sauvageau, G. Polycomb group proteins: Multi-faceted regulators of somatic stem cells and cancer. *Cell Stem Cell* **2010**, *7*, 299–313.
21. Pasini, D.; Cloos, P.A.; Walfridsson, J.; Olsson, L.; Bukowski, J.P.; Johansen, J.V.; Bak, M.; Tommerup, N.; Rappsilber, J.; Helin, K. JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells. *Nature* **2010**, *464*, 306–310.
22. Zhou, X.; Sun, H.; Chen, H.; Zavadil, J.; Kluz, T.; Arita, A.; Costa, M. Hypoxia induces trimethylated H3 lysine 4 by inhibition of JARID1A demethylase. *Cancer Res.* **2010**, *70*, 4214–4221.
23. Takeuchi, T.; Kojima, M.; Nakajima, K.; Kondo, S. *jumonji* gene is essential for the neurulation and cardiac development of mouse embryos with a C3H/He background. *Mech. Dev.* **1999**, *86*, 29–38.
24. Banach, M.; Prymula, K.; Jurkowski, W.; Konieczny, L.; Roterman, I. Fuzzy oil drop model to interpret the structure of antifreeze proteins and their mutants. *J. Mol. Model.* **2012**, *18*, 229–237.
25. Roterman, I.; Konieczny, L.; Jurkowski, W.; Prymula, K.; Banach, M. Two-intermediate model to characterize the structure of fast-folding proteins. *J. Theor. Biol.* **2011**, *283*, 60–70.
26. 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.
27. Bryliński, M.; Kochańczyk, M.; Broniatowska, E.; Roterman, I. Localization of ligand binding site in proteins identified in silico. *J. Mol. Model.* **2007**, *13*, 665–675.
28. Banach, M.; Konieczny, L.; Roterman, I. Lignad-binding site recognition. In *Protein Folding in Silico: Protein Folding Versus Protein Structure Prediction*; Roterman-Konieczna, I., Ed.; Woodhead Publishing: Oxford, Cambridge, UK & Philadelphia, PA, USA & New Dehli, India, 2012; pp. 79–94.
29. Banach, M.; Konieczny, L.; Roterman, I. Use of the "fuzzy oil drop" model to identify the complexation area in protein homodimers. In *Protein Folding in Silico: Protein Folding Versus Protein Structure Prediction*; Roterman-Konieczna, I., Ed.; Woodhead Publishing: Oxford, Cambridge, UK & Philadelphia, PA, USA & New Dehli, India, 2012; pp. 95–122.
30. Marchewka, D.; Jurkowski, W.; Banach, M.; Roterman, I. Prediction of protein-protein binding interfaces. In *Identification of Ligand Binding Site and Protein-Protein Interaction Area*; Roterman-Konieczna, I., Ed.; Springer: Heidelberg, Germany & New York, NY, USA & London, UK, 2013; pp. 105–134.
31. Banach, M.; Stapor, K.; Roterman, I. Chaperonin structure: The large multi-subunit protein complex. *Int. J. Mol. Sci.* **2009**, *10*, 844–861.
32. 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.
33. 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.

34. Levitt, M. A simplified representation of protein conformations for rapid simulation of protein folding. *J. Mol. Biol.* **1976**, *104*, 59–107.
35. Kyte, J.; Doolittle, R.F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **1982**, *157*, 105–132.
36. Eisenberg, D.; Weiss, R.M.; Terwilliger, T.C.; Wilcox, W. Hydrophobic moments and protein structure. *Faraday Symp. Chem. Soc.* **1982**, *17*, 109–120.
37. Engelman, D.M.; Zaccai, G. Bacteriorhodopsin is an inside-out protein. *Proc. Natl. Acad. Sci. USA* **1986**, *77*, 5894–5898.
38. Hopp, T.P.; Woods, K.R. Prediction of protein antigenic determinants from amino acid sequences. *Proc. Natl. Acad. Sci. USA* **1981**, *78*, 3824–3828.
39. Rose, G.D.; Geselowitz, A.R.; Lesser, G.J.; Lee, R.H.; Zehfus, M.H. Hydrophobicity of amino acid residues in globular proteins. *Science* **1985**, *229*, 834–838.
40. Wimley, W.C.; White, S.H. Experimentally determined hydrophobicity scale for proteins at membrane interfaces. *Nat. Struct. Biol.* **1996**, *33*, 842–848.
41. Wolfender, R.; Anderson, L.; Cullis, P.M.; Soulhgate, C.C. Affinities of amino acids side chains for solvent water. *Biochemistry* **1981**, *20*, 846–855.
42. Bryliński, M.; Konieczny, L.; Roterman, I. Is the protein folding an aim-oriented process? Human haemoglobin as example. *Int. J. Bioinform. Res. Appl*. **2007**, *3*, 234–260.
43. Rico, F.; Gonzales, L.; Casuso, I.; Puig-Vidal, M.; Scheuring, S. High-speed force spectroscopy unfolds titin at the velocity of molecular dynamics simulation. *Science* **2013**, *342*, 741–743.
44. Laskowski, R.A. PDBsum new things. *Nucleic Acids Res.* **2009**, *37*, D355–D359.
45. Orengo, C.A.; Michie, A.D.; Jones, S.; Jones. D.T.; Swindells, M.B.; Thornton, J.M. CATH-a hierarchic classification of protein domain structures. *Structure* **1997**, *5*, 1093–1108.
46. Tu, S.; Teng, Y.C.; Yuan, C.; Wu, Y.T.; Chan, M.Y.; Cheng, A.N.; Lin, P.H.; Juan, L.J.; Tai, M.D. The ARID domain of the H3K4 demethylase RBP2 binds to a DNA CCGCCC motif. *Nat. Struct. Biol*. **2008**, *15*, 419–421.
47. Wang, G.G.; Song, J.; Wang, Z.; Dormann, H.L.; Casadio, F.; Li, H.; Luo, J.L.; Patel, D.J.; Allis, C.D. Haematopoietic malignancies caused by dysregulation of a chromatin-binding PHD finger. *Nature* **2009**, *459*, 847–851.
48. Kedersha, N.; Ivanov, P.; Anderson, P. Stress granules and cell signaling: More than just a passing phase? *Trends Biochem. Sci.* **2013**, *38*, 494–506.
49. McWilliam, H.; Li, W.; Uludag, M.; Squizzato, S.; Park, Y.M.; Buso, N.; Cowley, A.P.; Lopez, R. Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acids Res.* **2013**, *41*, W597–W600.
50. Das, P.; Kapoor, D.; Halloran, K.T.; Zhou, R.; Matthews, C.R. Interplay between Drying and Stability of a TIM Barrel Protein: A Combined Simulation-Experimental Study. *J. Am. Chem. Soc*. **2013**, *135*, 1882–1890.
51. Galzitskaya, O.V.; Ivankov, D.N.; Finkelstein, A.V. Folding nuclei in proteins, *FEBS Lett.* **2001**, *489*, 113–118.

52. Roterman, I.; Konieczny, L.; Banach, M.; Marchewka, D.; Kalinowska, B.; Baster, Z.; Tomanek, M.; Piwowar, M. Simulation of the Protein Folding Process. In *Computational Methods to Study the Structure and Dynamics of Biomolecules and Biomolecular Processes: From Bioinformatics to Molecular Quantum Mechanics*; Liwo, A., Ed.; Springer Series in Bio-Neuroinformatics; Springer: Berlin, Germany, 2014; Volume 1, pp. 599–638.
53. Ben-Naim, A. Myths and verities in protein folding theories: From Frank and Evans iceberg-conjecture to explanation of the hydrophobic effect. *J. Chem. Phys*. **2013**, *139*, 165105.
54. Ben-Naim, A. Theoretical aspects of self-assembly of proteins: A Kirkwood-Buff-theory approach. *J. Chem. Phys.* **2013**, *138*, 224906.
55. Ben-Naim, A. On the So-Called Gibbs Paradox, and on the Real Paradox. *Entropy* **2007**, *9*, 132–136.
56. Ben-Naim, A. Theoretical aspects of pressure and solute denaturation of proteins: A Kirkwood-buff-theory approach. *J. Chem. Phys.* **2012**, *137*, 235102.