Biochimie 144 (2018) 74–84

Contents lists available at ScienceDirect

Biochimie

journal homepage: www.elsevier.com/locate/biochi

Research paper

Why do antifreeze proteins require a solenoid?

M. Banach [a], L. Konieczny [b], I. Roterman [a, *]

[a] Department of Bioinformatics and Telemedicine, Jagiellonian University, Medical College, Lazarza 16, 31-530, Krakow, Poland
[b] Chair of Medical Biochemistry, Jagiellonian University, Medical College, Kopernika 7, 31-034, Krakow, Poland

ARTICLE INFO

Article history:
Received 18 August 2017
Accepted 12 October 2017
Available online 17 October 2017

Keywords:
Hydrophobicity
Antifreeze proteins
Solenoid
Amyloid

ABSTRACT

Proteins whose presence prevents water from freezing in living organisms at temperatures below 0 °C are referred to as antifreeze proteins. This group includes molecules of varying size (from 30 to over 300 aa) and variable secondary/supersecondary conformation. Some of these proteins also contain peculiar structural motifs called solenoids. We have applied the fuzzy oil drop model in the analysis of four categories of antifreeze proteins: 1 – very small proteins, i.e. helical peptides (below 40 aa); 2 – small globular proteins (40–100 aa); 3 – large globular proteins (>100 aa) and 4 – proteins containing solenoids. The FOD model suggests a mechanism by which antifreeze proteins prevent freezing. In accordance with this theory, the presence of the protein itself produces an ordering of water molecules which counteracts the formation of ice crystals. This conclusion is supported by analysis of the ordering of hydrophobic and hydrophilic residues in antifreeze proteins, revealing significant variability – from perfect adherence to the fuzzy oil drop model through structures which lack a clearly defined hydrophobic core, all the way to linear arrangement of alternating local minima and maxima propagating along the principal axis of the solenoid (much like in amyloids). The presented model – alternative with respect to the ice docking model – explains the antifreeze properties of compounds such as saccharides and fatty acids. The fuzzy oil drop model also enables differentiation between amyloids and antifreeze proteins.

1. Introduction

The role of antifreeze proteins cannot be properly analyzed without a discussion of the structuralization of water itself. Numerous publications exist where the structure of ice is discussed, starting with Bernal-Fowler rules [1–9]. In fact, structuralization of ice is a far more popular study subject than that of liquid water in the presence of dissolved compounds [10–17]. Ben Naim in Ref. [18] proposes an iceberg model to explain the ordering of water molecules. Our work approaches structuralization of water from the perspective of its effects on other molecules. In particular, surfactant micelles where hydrophobicity is concentrated in the central portion of the micelle while polar structures remain exposed, can only emerge in the presence of water [19]. Polypeptide chain folding appears to result from a similar active influence of the water environment. Altering the properties of this environment triggers structural changes, affecting the ultimate conformation of the protein – for example, in the case of elastin [20]. In contrast to individual surfactant molecules, the polypeptide chain exhibits variable hydrophobicity. The oil drop model [21] predicts concentration of hydrophobic residues at the core of the protein, along with exposure of hydrophilic residues on the surface (where they remain in contact with water). This is regarded as a consequence of water acting on individual fragments of the target chain. This simplistic model has since been extended, resulting in the so-called fuzzy oil drop model, where the distribution of hydrophobicity in a protein is modeled by a 3D Gaussian. A detailed description of the fuzzy oil drop model can be found in Ref. [22], where the authors show consistent results regardless of the applied intrinsic hydrophobicity scale. The β-strand has long been known for its association with amyloid-like and amyloid forms, including solenoids [23–25]. Solenoids themselves are categorized on the basis of such parameters as handedness, twist angle, oligomerization state and coil shape [26]. Proteins which contain repeated sequences are strongly predisposed towards generation of solenoid (or solenoid-like) conformations. This fact has led researchers to assemble a database of tandem repeated structure proteins [27,28]. Factors which determine linear deformations in solenoids compared to the so-called horseshoe structure are discussed in Ref. [29]. Antifreeze

* Corresponding author.
E-mail address: myroterm@cyf-kr.edu.pl (I. Roterman).

https://doi.org/10.1016/j.biochi.2017.10.011

proteins are described in numerous publications, some of which suggest that their mechanism of action is based on docking of ice crystals [30–32].

2. Materials and methods

2.1. Data

Table 1 lists proteins which have been subjected to analysis. This set was obtained by scanning the PDB database using the "antifreeze" keyword. Proteins were divided into groups depending on their size. Particular attention was devoted to solenoid structures which are present in some of the analyzed proteins.

Since all proteins selected for analysis are the antifreeze proteins, it is difficult to distinguish each of them. This is why the PDB IDs are treated as "names" of proteins.

2.2. Programs

All parameters were calculated using custom software. Identification of domains and secondary folds follows the PDBSum classification [60].

2.3. Fuzzy oil drop model

A detailed introduction to the fuzzy oil drop model can be found in Ref. [22] Below we provide a brief recapitulation of the model's basic concepts insofar as they relate to the presented work. The general principle is that the observed hydrophobicity distribution in the target molecule (denoted O), which results from inter-residue interactions [61], is compared to the so-called idealized (or theoretical) distribution (denoted T), mathematically expressed by a 3D Gaussian. Quantitative comparison of both distributions is based on the concept of divergence entropy [62]. The resulting similarity measure depends on the length of the input chain. Additionally, since the obtained value represents entropy, it may not be interpreted on its own, but must instead be compared to another boundary distribution, which we refer to as unified (R). In this distribution, each residue is ascribed the same hydrophobicity value of 1/N, where N is the number of residues in the input chain. To avoid having to work with two distinct parameters, i.e. observed-vs.-theoretical and observed-vs.-unified entropy (O|T and O|R respectively), we derive an additional coefficient referred to as Relative Distance (RD):

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

$O|T$, $O|R$ and RD may be computed for any arbitrarily selected structural unit: protein complexes, individual proteins and specific domains. In each case, a different encapsulating ellipsoid, custom-tailored for the target unit, must be prepared. Additionally, when considering specific parts of the protein chain, O_i, T_i and R_i values must be renormalized so that their sum is always equal to 1. This process enables us to identify regions which exhibit good accordance with the model and therefore contribute to tertiary structural stabilization.

Fig. 1 provides a visual description of a representative case.

RD tells us whether the molecule contains a well-ordered hydrophobic core (RD < 0.5) or lacks such a core (RD > 0.5). The threshold value of 0.5 was selected since the distance comparison is relative in scope. Simply speaking, RD < 0.5 means that the molecule more closely resembles the idealized Gaussian distribution than the unified distribution, while the opposite is true when RD > 0.5.

It should be noted that similar analysis can be performed for selected fragments of the protein. In such cases, the 3D Gaussian is plotted for the specific unit (domain, chain, complex) and a new value of RD is calculated following prior normalization of O_i, T_i and R_i. This value expresses the status of the given unit within the framework of the larger structure to which it belongs. For example, in this work we compute RD coefficients for solenoid fragments and for selected secondary folds.

3. Results

The results of our analysis are summarized by a set of RD values calculated for complexes, individual proteins and selected domains (where applicable). Since antifreeze proteins vary in length, we further subdivided this class into groups, as shown in Table 2. For each group, several representative cases were singled out for detailed analysis, which involved computing RD values for individual secondary folds, plotting T and O distributions in a manner which enables visual comparison, and presenting 3D images of each target protein.

The presented proteins exhibit significant conformational variability – this is reflected by variable presentation of results, depending on the complexity of the given structure.

3.1. Peptides – length below 40 aa

This set of proteins represents structural forms which are essentially helical. The chain length is too short to enable generation of tertiary structures. Applying the hydrophobic core drop model to individual helices is questionable on theoretical grounds; however, mindful of the aim of presenting a holistic description of all types of antifreeze proteins, we have calculated FOD parameters for these proteins as well. Table 2 lists the corresponding RD values. Only 2LQ0, a *de novo* protein, remains accordant with the theoretical distribution – however, it should be noted that this protein was synthesized with the specific goal of retaining a centralized hydrophobicity peak. No naturally occurring category I antifreeze proteins exhibit similar properties. Fig. 2 highlights the differences between T and O for two representative polypeptides.

The *de novo* protein (2LQ0) contains hydrophilic residues in its

Table 1
Set of proteins subjected to analysis, assigned to distinct groups depending on their size. Brief structural characteristics are also listed for each group. Underlined identifiers correspond to proteins which have been selected for detailed presentation.

Length	Structure	PDB ID and references
<40 aa	Loose helices	1J5B [33], 1Y03 [34], 2LQ0 [35]
40 aa< <100 aa	Globular	1B7I [36], 3NLA [37], 1KDE [38], 2LX2 [39]
100 aa <	Globular, complexes	2PY2 [40], 1C3Y [41], 4KDV [42], 1C89 [43], 2ZIB [44]
100 aa <	Solenoid complexes	1EWW [45],1L0S [46],1L1I [47],1M8N [48], 1N4I [49], 1Z2F [50],2PNE [51], 3BOG [51], 3P4G [52], 3ULT [53]
100 aa <	Solenoid with helix complexes	3UYU [54],3VN3 [55],3WP9 [56],4NU2 [57],4NUH [57],5B5H [58]
100 aa <	Amyloid-like	4DT5 [59]