Is the hydrophobic core a universal structural element in proteins?

Barbara Kalinowska1,2 · Mateusz Banach1,2 · Zdzisław Wiśniowski1 · Leszek Konieczny3 · Irena Roterman1

Abstract The hydrophobic core, when subjected to analysis based on the fuzzy oil drop model, appears to be a universal structural component of proteins irrespective of their secondary, supersecondary, and tertiary conformations. A study has been performed on a set of nonhomologous proteins representing a variety of CATH categories. The presence of a well-ordered hydrophobic core has been confirmed in each case, regardless of the protein’s biological function, chain length or source organism. In light of fuzzy oil drop (FOD) analysis, various supersecondary forms seem to share a common structural factor in the form of a hydrophobic core, emerging either as part of the whole protein or a specific domain. The variable status of individual folds with respect to the FOD model reflects their propensity for conformational changes, frequently associated with biological function. Such flexibility is expressed as variable stability of the hydrophobic core, along with specific encoding of potential conformational changes which depend on the properties of helices and \( \beta \)-folds.

Keywords Protein folding · Hydrophobicity · Hydrophobic core

Introduction

The traditional classification of secondary and supersecondary protein structures covers helical and \( \beta \)-fragments along with their mutual arrangement within the protein body. This \( \alpha/\beta \) classification varies with respect to the participation of both folds in proteins, as well as their orientation. Purely helical (hemoglobin, cytochrome) or purely \( \beta \)-shaped (immunoglobulins) structures, as well as proteins characterized by near-equal participation of both types (lactate dehydrogenase or carboxypeptidase), can be identified in protein databases (e.g., PDB). The supersecondary structure, which determines the mutual arrangement of secondary folds can be expressed as follows (the following list is derived from a commonly used biochemistry textbook [1]):

1. \( \beta\alpha\beta \) – helical fragment linking two parallel \( \beta \)-strands
2. \( \beta \) hairpin – two anti-parallel \( \beta \)-strands linked by a tight U-turn loop
3. \( \alpha\alpha \) motif – two successive helices linked by a tight U-turn loop
4. Greek key motif – \( \beta \)-strands following one another in an arrangement reminiscent of classic Greek ornaments
5. \( \beta \) barrels – coaxial \( \beta \)-strands forming a cyclical system reminiscent to the arrangement of planks forming a wooden barrel.

Attempts to introduce a detailed structural classification of domain units in large proteins have resulted in the creation of the CATH/Gene3D (http://www.cathdb.info/ 27 Apr 2016) database, which currently (as of 27 April 2016) contains 26 million domains arranged into 2738 families. Structural classification is performed automatically upon inclusion of a new protein in PDB, via an exhaustive search for proteins homologous to the new entrant [2]. It is immediately apparent that the number of distinct families is quite high,
notwithstanding the overall size of the database. This suggests that protein structures are diverse and difficult to classify.

The concept of biological activity likewise remains a mystery [3], despite successful attempts to program it into synthetic proteins. An in-depth analysis of the stability of the backbone and side chains facilitating stabilization of one of two possible rotametric forms, is presented in [4]. The great diversity of geometric forms makes it difficult to propose a single, common classification based on a unified criterion. This, however, does not imply that attempts to identify such a criterion are futile.

Our work focuses on proteins characterized by major supersecondary variability and proposes a coherent classification, covering all cases. The starting point is to acknowledge the common environment in which all proteins undergo folding and gain biological function — water. Water is the immanent condition ensuring proper folding and biological activity. This is why the water participation and its influencing on protein structure is the basis for the fuzzy oil drop model applied to interpret the different structures of proteins. We assume that the proposed model visualizes the effects of the influence of water environment despite the lack of knowledge of its own structuralization.

Table 1  Summary of proteins subjected to analysis

| Protein       | Name                      | Source   | Class                          | Ref.                           |
|---------------|---------------------------|----------|-------------------------------|--------------------------------|
| 1A5Z          | Lactate dehydrogenase     | Bacteria | αβ complex                     | (Auerbach et al. [10])         |
| 1FW8          | Phosphoglycerate kinase   | Yeast    | αβ 3-layer (aba)               | (Tougard et al. [11])          |
| 1AMK          | Triose phosphate isomerase| Bacteria | α β barrel                    | (Williams et al. [12])         |
| 4DRS          | Pyruvate kinase           | Bacteria | αβ 3-layer (aba)               | (Cook et al. [13])             |
|               | All α -domain super-fold  |          | αβ barrel mainly β barrel     |                                |
| 1B4L          | Superoxide dismutase      | Yeast    | Mainly β sandwich             | (Hart et al. [14])             |
| 1CON          | Concanavalin              | Plant    | Mainly β sandwich             | (Naismith et al. [15])         |
| 1RBP          | Retinol binding           | Human    | Mainly β barrel               | (Cowan et al. [16])            |
| 1PNG          | Asparagine amidase        | Bacteria | Mainly β sandwich             | (Kuhn et al. [17])             |
| 1TIM          | Triose phosphatase isomerase| Chicken | α β barrel                    | (Banner et al. [18])           |
|               | Miscellaneous             |          |                               |                                |
| 256B          | Cytochrome                | E-coli   | α Up-down bundle              | (Lederer et al. [19])          |
| 7FAB          | Immunoglobulin            | Human    | Mainly β sandwich             | (Saul and Poljak [20])         |
| 6LDH          | Lactate dehydrogenase     | Bacteria | α β complex                   | (Abad-Zapatero et al. [21])    |
|               | Cytochrome fold           |          |                               |                                |
| 155C          | Cytochrome                | Bacteria | α orthogonal bundle           | (Timkovich and Dickerson [22]) |
| 1JDL          | Cytochrome                | Bacteria | α orthogonal bundle           | (Camara-Artigas et al. [23])   |
| 2C2C          | Cytochrome                | Bacteria | α orthogonal bundle           | (Bhatia –PDB [24])            |
| 5CYT          | Cytochrome                | Fish     | α orthogonal bundle           | (Takano [25])                  |
| 4 J20         | Cytochrome                | Bacteria | α orthogonal bundle           | (Yu et al. [26])               |

The selection is based on supersecondary structural order. The study set is consistent with the one proposed in (Devlin 2011).
values. Such parts are thought to mediate structural stabilization of the protein (or domain) as a whole [5–9].

In this work we show that, regardless of secondary and supersecondary conformational properties, the presence of a more or less prominent hydrophobic core is a common phenomenon. We identify proteins where the core is highly consistent with theoretical values (3D Gaussian) as well as those which exhibit local deviations from the theoretical model, typically associated with the capability to bind ligands or attract complexation partners. The study set is very diverse, yet exhibits common characteristics which point to the necessary presence of water in their environment — this applies to many different molecules encountered in the cell, but particularly to proteins.

Restricting protein structure analysis to topological aspects would disregard the most important factor — the influence of the water environment, which plays an active and often decisive role in the folding process. Hence, our study of the hydrophobic core — a product of the environment, which determines the biological activity of proteins.

Water environment is treated as the external force field which together with the internal force field (non-bonding interaction) participates in structuralization process and ensures the biological activity. The presented paper is an attempt to show the possible interpretation of structural effects which may express the role of water.

Materials and methods

Protein database

Our analysis is based on a set of proteins derived from a popular biochemistry textbook [1]. Table 1 reveals the structural variability observed at the supersecondary level.

The fuzzy oil drop model

An in-depth presentation of the fuzzy oil drop model can be found in an Open Access publication [27, 28]. In the interest of saving space, we will limit ourselves to recapitulating its basic tenets.

The fuzzy oil drop model is a modification of the oil drop model originally introduced by Kauzmann [29]. The original model was discrete, distinguishing a central hydrophobic part and an outer hydrophilic shell. In contrast, the fuzzy oil drop model introduces a 3D Gaussian which peaks at the center of the encapsulating ellipsoid, with values decreasing along with distance from the center, reaching near 0 at a distance of 3σ (where σ is a parameter of the Gaussian expressing the size of the ellipsoid). Distribution along each axis can be characterized by an appropriate sigma parameter. The values of such a function are assumed to represent theoretical (idealized) distribution of hydrophobicity in a protein molecule.

The actual (observed) hydrophobicity distribution depends on the location of each residue in the protein body (in our calculations we apply the function proposed by Levitt [30]). Each residue (represented by its effective atom — averaged-out positions of all atoms belonging to a particular residue) aggregates hydrophobic interactions with other residues separated from it by not more than 9 Å. Clearly, the observed distribution may differ from the idealized (Gaussian) profile. The scope of such interactions is dependent on the relative proximity between both residues and on their intrinsic hydrophobicity, which can be measured experimentally or predicted on theoretical grounds. Different hydrophobicity scales can be

![Graphical representation of fuzzy oil drop model hydrophobicity distributions](image)

Fig. 1 Graphical representation of fuzzy oil drop model hydrophobicity distributions obtained for a hypothetical protein reduced to a single dimension for simplicity. A is the theorized Gaussian distribution (blue) while chart C corresponds to the uniform distribution (green). Actually observed (red) hydrophobicity density distribution in the target protein B, while its corresponding value of RD (relative distance) in D is marked on the horizontal axis with a red diamond. According to the fuzzy oil drop model this protein does not contain a well-defined hydrophobic core, because its RD value, equal to 0.619, is above the 0.5 threshold (or, generally, closer to R than to T).
applied (a comparison of results obtained for two different scales is presented in [31]). Normalization of both distributions (theoretical and observed) facilitates meaningful comparisons: in particular, we can identify residues for which

Table 2 Structural properties of the hydrophobic core in two proteins representing the α/β-domain fold twisted β-sheet category

| Structure          | Fragment       | RD   | Structure          | Fragment       | RD   |
|--------------------|----------------|------|--------------------|----------------|------|
| Lactate dehydrogenase (1A5Z-D1) 3.40.50.720 α-β 3-layer (aba) sandwich | Domain 1 | 0.416 | Domain 2 | 0.523 |
| β-sheet            | Parallel       | 0.298 | β-sheet I and II   | Parallel       | 0.516 |
| Helices            | 0.530          |      | Helices            | 0.502          |
| No 68–84 β-conform.|                |      | β-sheet I          | 0.492          |
|                    | 0.449          |      | Parallel           | 0.399          |
| 22–28              | 0.198          |      | 133–140            | 0.127          |
| 47–52              | 0.259          |      | 158–163            | 0.145          |
| 77–83              | 0.331          |      | 205–207            | 0.145          |
| 92–97              | 0.180          |      | 210–216            | 0.706          |
| 134–138            | 0.243          |      | 223–229            | 0.455          |
| 159–163            | 0.307          |      | 238–242            | 0.763          |
| 258–263            |                |      | 258–263            | 0.426          |
| 294–298            | 0.144          |      | 294–298            | 0.144          |
| 30–44              | 0.498          |      | 30–44              | 0.589          |
| 55–70              |                |      | 55–70              | 0.980          |
| 71–73              | 0.980          | Helices | 146–155            | 0.353          |
| 84–89              | 0.366          |      | 166–175            | 0.244          |
| 108–131            | 0.444          |      | 185–202            | 0.347          |
| 141–154            | 0.365          |      | 243–257            | 0.358          |
|                    |                |      | 271–275            | 0.363          |
|                    |                |      | 276–291            | 0.721          |
|                    |                |      | 300–309            | 0.378          |
|                    |                |      | 311–315            | 0.228          |
|                    |                |      | Domain 2           | 0.492          |
|                    |                |      | No β-II            | 0.320          |
|                    |                |      | β-sheet II         | 0.625          |
|                    |                |      | Anti-parallel      |                |
|                    |                |      | No ligand binding  |                |
|                    |                |      |                   |                |
|                    |                |      |                   |                |

Values listed in boldface correspond to RD > 0.5. Items labeled “No (…)” represent shortened fragments where elimination of the indicated residues changes the relation to RD < 0.5. “Ligand binding” stands for fragments directly involved in ligand binding. Our experience with the fuzzy oil drop model indicates that such fragments often exhibit significant deviations from the theoretical hydrophobicity distribution.
theoretical (T) values diverge significantly from their observed (O) counterparts.

Both hydrophobicity distribution profiles — expected (T) and observed (O) — can be compared quantitatively. Quantitative expression of the differences between the expected (T) and observed (O) distribution is possible by using the Kullback-Leibler divergence entropy formula [32]:

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

The value of $D_{KL}$ expresses the distance between the observed ($p$) and target ($p^0$) distributions, the latter of which is given by the 3D Gaussian (T). The observed distribution ($p$) is referred to as O.

For the sake of simplicity, we introduce the following notation:

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

**Table 3**  
RD values obtained for 1B4L and 1CON, for their respective secondary folds, for the entire β-sheet and for the fragment bounded by CYS residues which form a disulfide bond

| Characteristics | Fragment  | RD |
|-----------------|----------|----|
| Cu, Zn superoxide dismutase (1B4L) 2.60.40.200 – mainly β-sandwich | Protein 153aa | 0.433 |
|                 | 1–7 BI   | 0.065 |
|                 | 14–21 BI | 0.403 |
|                 | 28–35 BI | 0.468 |
|                 | I 41–48 BI | 0.246 |
|                 | E 63H, I57–61 H | **0.724** |
|                 | I 82–89 BI | 0.433 |
|                 | 95–101 BI | 0.343 |
|                 | I 115–120 BI | 0.257 |
|                 | 131–135 H | 0.349 |
|                 | E 143R | 0.317 |

| Characteristics | Fragment  | RD |
|-----------------|----------|----|
|                 | Protein 237aa | 0.416 |
|                 | I 3–11 B | 0.407 |
|                 | I 14–18 H | 0.421 |
|                 | 23–30 B | 0.355 |
|                 | I 36–40 B | 0.295 |
|                 | E 63H, I57–61 H | 0.333 |
|                 | I 59–66 B | 0.338 |
|                 | 73–79 B | 0.335 |
|                 | I 80–84 H | 0.218 |
|                 | I 87–97 B | 0.330 |
|                 | E 143R | 0.135 |
|                 | 103–118 B | **0.601** |
|                 | 122–130 B | 0.135 |
|                 | 139–145 B | 0.438 |
|                 | 147–149 B | **0.851** |
|                 | 150–152 H | 0.365 |
|                 | 153–156 B | 0.384 |
|                 | 169–176 B | 0.336 |
|                 | 179–181 B | **0.643** |
|                 | I 185–200 B | **0.762** |
|                 | 208–217 B | 0.337 |
|                 | 226–230 H | 0.324 |

Values listed in boldface satisfy RD > 0.5. “B” and “H” stand for β-structural and helical forms respectively; “I” indicates an ion-binding fragment while “E” denotes that the given fragment contains an enzymatically active residue (listing its number and type).
Since $D_{KL}$ is a measure of entropy it must be compared to a reference value. In order to facilitate meaningful comparisons, we have introduced another opposite boundary distribution (referred to as “uniform” or $R$) which corresponds to a situation where each effective atom possesses the same hydrophobicity density ($1/N$, where $N$ is the number of residues in the chain). This distribution is deprived of any form of hydrophobicity concentration at any point in the protein body:

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

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

$$RD = O|T/(O|T + O|R)$$

RD < 0.5 is understood to indicate the presence of a hydrophobic core. Figure 1 presents a graphical representation of RD values, restricted (for simplicity) to a single dimension.

$D_{KL}$ (as well as $O|T$, $O|R$ and RD) may be calculated for specific structural units (protein complex, single molecule, single chain, selected domain etc.) In such cases the bounding ellipsoid is restricted to the selected fragment of the protein. It is also possible to determine the status of polypeptide chain fragments within the context of a given ellipsoid. This procedure requires prior normalization of $O$ and $R$ distributions describing the analyzed fragment, whose length is denoted as $N$. Note that any selected fragment must be of a reasonable length — calculations cannot be performed for an individual residue.

Generally the idealized distribution (T) is treated as the target for $O|T$ calculation, while the unified distribution (R) is used when computing $O|R$.

The above procedure will be applied in the analysis of proteins described in this paper. By restricting our analysis to individual fragments, we can determine whether a given fragment participates in the formation of a hydrophobic core. In particular, fragments representing well-defined secondary folds which satisfy $RD < 0.5$ are thought to contribute to structural stabilization, while fragments for which $RD > 0.5$ are less stable. Such fragments, if present on the surface of the protein, may potentially form complexation sites.

Specific fragments are derived by analyzing the protein’s secondary conformation. Identification of secondary folds and the composition of protein domains follows the CATH [33] and PDBsum [34] classifications. Likewise, inter-domain/inter-chain contacts have been identified on the basis of the PDBsum distance criteria [34].

A graphical presentation of RD interpretation is shown in Fig. 1.

## Results

### α/β-domain fold twisted β-sheet

This category is represented by the nonhomologous domain 1 of lactate dehydrogenase (1A5Z – domain 1) and
phosphoglycerate kinase (1FW8 – domain 2). The α,β-superfold can be described as a central twisted β-sheet surrounded by a ring of helices. RD values for both domains and their individual secondary folds are listed in Table 2. 

From the structural point of view both proteins provide examples of the so-called flavodoxin fold [35] with a centrally located β-propeller. Domain D1 (1A5Z) appears to include a well-ordered hydrophobic core while domain D2 (1FW8) lacks such a core, as indicated by its RD value, which is in excess of 0.5 (Table 2). It moreover turns out that the extra β-sheet (comprising three separate folds), which is not present in 1A5Z, diverges from the idealized distribution (Fig. 2). Eliminating this fragment from computations (an independent droplike capsule constructed for the remainder of the domain) produces a structure which is a good match for the theoretical model. Both domains exhibit similar properties in the scope of their central β-sheet, suggesting that the sheet contributes to structural stabilization (this is based on the assumption that the presence of a prominent hydrophobic core promotes tertiary stabilization). Both domains also contain a single helix (located at a similar distance from the propeller) which diverges from the model (Fig. 3 and Table 2).

### Table 4  
RD values obtained for 1AMK and 4DRS and for their respective secondary folds

| Fragment       | Characteristics | RD    | Fragment       | Characteristics | RD    |
|----------------|-----------------|-------|----------------|-----------------|-------|
| Protein 250    |                 | 0.506 | Protein 248    |                 | 0.559 |
| 6–12 B         | E 11 N, 13 K,   | 0.401 | 47–52 B        |                 | 0.330 |
| 17–30 H        |                 | 0.402 | 53–57 H        |                 | 0.278 |
| 38–42 B        |                 | 0.452 | 58–69 H        |                 | 0.475 |
| 44–47 H        |                 | 0.747 | 71–77 B        |                 | 0.497 |
| 48–55 H        |                 | 0.336 | 81–98 H        |                 | 0.432 |
| 59–63 B        |                 | 0.263 | 105–110 B      |                 | 0.580 |
| 79–86 H        |                 | 0.330 | 214–224 H      |                 | 0.404 |
| 89–93 B        |                 | 0.148 | 231–235 B      |                 | 0.378 |
| 95–102 H       | E 95H           | 0.634 | 240–253 H      |                 | 0.387 |
| 105–119 H      |                 | 0.366 | 266–271 B      |                 | 0.477 |
| 122–127 B      |                 | 0.321 | 273–279 H      |                 | 0.551 |
| 130–136 H      |                 | 0.541 | 280–287 H      |                 | 0.329 |
| 138–152 H      |                 | 0.364 | 289–294 B      |                 | 0.390 |
| 153–154 H      |                 | 0.319 | 296–301 H      |                 | 0.647 |
| 155–160 H      |                 | 0.340 | 302–305 H      |                 | 0.534 |
| 161–167 B      | E 167E          | 0.317 | 306–321 H      |                 | 0.569 |
| 168–172 H      |                 | 0.571 | 322–327 B      |                 | 0.395 |
| 179–198 H      | E 173G          | 0.423 | 331–336 H      |                 | 0.440 |
| 199–206 H      |                 | 0.654 | 341–355 H      |                 | 0.418 |
| 207–212 B      |                 | 0.542 | 357–361 B      |                 | 0.306 |
| 218–224 H      |                 | 0.382 | 362–367 H      |                 | 0.603 |
| 229–233 B      |                 | 0.524 | 370–387 H      |                 | 0.553 |
| 234–239 H      |                 | 0.222 |                |                 |       |
| 240–241 H      |                 | 0.600 |                |                 |       |
| 242–248 H      |                 | 0.160 |                |                 |       |
| B-sheet        |                 |       | β-sheet        |                 | 0.429 |
| Helices        |                 |       | Helices        |                 | 0.482 |
| Eliminated     | 95–100,198–203  | 0.479 | No ligand      |                 | 0.546 |
| No ligand      | 0.498           | No residues | 74–80 L, 229–236 L, 259–262, 264–272 L, 293–296, 319–321 L | 0.486 |
| NoE            | 11,13,95,167,173| 0.500 |                |                 |       |

Values listed in boldface satisfy RD > 0.5. “B” and “H” stand for β- and helical forms respectively; “L” indicates a ligand-binding fragment while “E” denotes that the given fragment contains an enzymatically active residue (listing its number and type). The row labeled “NoE” represents the status of each domain following elimination of catalytic residues — the observed reduction in RD values indicates that catalytic residues diverge from the theoretical model.
Lactate dehydrogenase (1A5Z) is derived from the hyperthermophilic bacterium *Thermotoga maritima* and folds near the boiling point of water. Thermostable proteins are the focus of a separate study based on the fuzzy oil drop model (publication currently in preparation).

The catalytic residues in 1A5Z, as well as its disulfide bonds, are all located in domain 2—thus, we will not consider them in the presented analysis.

Domain 2 of phosphoglycerate kinase (1FW8-D2) exhibits $RD > 0.5$, which means that its hydrophobic core is deformed. The domain contains two $\beta$-sheets—a parallel sheet (labeled “I” in Table 2—133-140, 158-163, 205-207, 258-263, 294-298) and an antiparallel sheet (labeled “II” — 210, 223-229, 238-242). Such antiparallel conformation is not present in lactate dehydrogenase (1A5Z-D1). The $\beta$-structure which is analogous to 1A5Z is characterized by $RD = 0.492$, which implies local consistency with the theoretical hydrophobic core model.

As posited by the fuzzy oil drop model, ligand binding residues exhibit deviations from the theoretical status; this is because such residues interact with the ligand, which also participates in the formation of a shared core. In the case of strongly polar ligands, deviations may also be associated with the presence of additional structural elements not directly involved in forming the hydrophobic core, but instead distorting its structure.

In both presented proteins eliminating residues which participate in ligand binding results in a lower $RD$ value. This means that the remainder of the molecule fulfills a stabilizing role by presenting a well-ordered hydrophobic core.

The exposed $\beta$-hairpin in D2 of 1FW8 may mediate interaction with other molecules present in the protein’s environment, or it may constitute a dynamic and unstable element (no prominent hydrophobic core) which also potentially affects interaction.

It furthermore appears that the $\alpha,\beta$-superfold — particularly the twisted $\beta$-sheet — contains a hydrophobic core and therefore improves the stability of its parent domain.

**$\alpha,\beta$-domain fold**

The all-$\beta$-sheet superfold is represented by Cu, Zn superoxide dismutase and concanavalin A. This type of all-$\beta$ domain has the shape of a $\beta$-barrel. Both proteins appear to contain hydrophobic cores ($RD < 0.5$ in each case).

In 1B4L enzymatic and heavy metal ion binding residues are all located in loops. Table 3 lists positions adjacent to each fragment. The helical fold at 57–61 is locally discordant and contains a catalytic residue which binds Cu$^{2+}$. Similarly, in 1CON the cadmium ion (Cd$^{2+}$) is bound by a locally discordant $\beta$-fragment.

Cu, Zn superoxide dismutase adheres to the theoretical hydrophobicity distribution model as a whole, although some secondary folds diverge from the model. The helical fragment at 57–61 is an example ($RD > 0.5$). Similarly, in 1CON, despite the overall adherence of the molecule and its $\beta$-sheet, four individual $\beta$-folds are found to be discordant.
Both proteins (Fig. 4 and Fig. 5) exhibit highly ordered hydrophobic cores in spite of their biological diversity. Note that under the fuzzy oil drop model the concept of a hydrophobic core refers to a concentration of hydrophobicity density at the center of the molecule along with the presence of an encapsulating hydrophilic shell.

**Table 5**  
RD values representing individual secondary folds in 1RBP, 1PNG, and 1TIM

| Fragment | RD |
|----------|----|
| Protein 175aa | 0.553 |
| Domain 2 135aa | 0.694 |
| 142–151 BI | 0.771 |
| 154–158 H | 0.472 |
| 159–161 BI | 0.474 |
| 5–9 H | 0.332 |
| 162–168 L | 0.605 |
| 22–30 B | 0.567 |
| 169–175 BI | 0.610 |
| 39–48 B | 0.431 |
| 176–180 L | 0.427 |
| 52–63 B | 0.406 |
| 181–198 BI | 0.688 |
| 67–80 B | 0.478 |
| 199–201 L | 0.560 |
| 84–93 B | 0.589 |
| 202–206 BI | 0.187 |
| 100–110 B | 0.506 |
| 207–209 L | 0.415 |
| 113–124 B | 0.801 |
| 210–217 BI | 0.436 |
| 128–139 B | 0.787 |
| 219–226 BI | 0.649 |
| 145–159 H | 0.501 |
| 227–229 L | 0.511 |
| 166–168 B | 0.404 |
| 230–234 H | 0.209 |
| 235–259 L | 0.654 |
| 260–264 BI | 0.832 |
| 265–269 H | 0.381 |
| 271–278 BI | 0.742 |
| 279–289 L | 0.800 |
| 290–303 BI | 0.739 |
| 304–310 L | 0.578 |
| 311–313 BI | 0.508 |
| β-sheet | 0.615 |
| α-Helixes | 0.522 |
| SS 4–160 | 0.566 |
| SS 70–174 | 0.582 |
| Removed: | 0.467 |
| 29–31, 69–71 SS, 91–97 Loop, 117, 127–130 SS, 133–134 L | |

Values listed in boldface satisfy RD > 0.5. H = helix; B = β-fold; E # = placement of catalytic residue (with # corresponding to its code); L = ligand present as part of the described fragment.

All β-domain superfold

This category is represented by triose phosphate isomerase and pyruvate kinase domain 1. The core structure comprises a centrally located β-barrel where individual β-folds are interconnected by α-helices located outside of the fold. The sample proteins differ with respect to their composition and biological activity.

Elimination of catalytic residues (Table 4, Fig. 6, and Fig. 7) lowers the RD value to 0.479. This indicates local disorder in the area of the enzymatic active site.

Local deviations from the theoretical model caused by a ligand are commonplace. The ligand — by virtue of its presence — enforces a conformation which represents a balance between the protein’s own folding tendencies and the altered...
conditions introduced by the ligand’s polarized atoms. The end result of this process depends strongly on the ligand’s hydrophobicity [36].

Residues which diverge from the model, despite not taking part in ligand binding, may represent structurally encoded “readiness” for structural changes in the α/β domain.

Despite the observed differences, both domains share certain similarities. The helical fold at 95–100 in 1AMK exhibits a similar deviation from theoretical values when compared with the folds at 319–321 and 362–364 in 4DRS D2. It seems that local instability in all these areas may promote structural rearrangement.

β-barrel

Three proteins representative of the β-barrel domain are characterized in Table 5 and Fig. 8.

Each of the above listed proteins represents a different type of β-barrel. The assessment of hydrophobic core structure is, likewise, different in each case. It should be noted that the

Table 6  RD values calculated for fragment domains of the Fab fragment of immunoglobulin G, revealing variable stability of individual secondary folds

| Chain L | Domain VL | Domain CL | Chain H | Domain VH | Domain CH1 |
|---------|-----------|-----------|---------|-----------|-----------|
| Fragment | RD | Fragment | RD | Domain | Fragment | RD | Domain | Fragment | RD |
| 0.584 | 0.584 | 0.584 | 0.649 | 0.629 | 0.369 | 0.703 | 0.703 | 0.675 | 0.675 | 0.468 | 0.358 | 0.710 | 0.600 | 0.613 | 0.477 |
| 8–12 BI | 8–12 BI | 18–24 BI | 28–32 H | 36–40 BI | 47–49 BI | 56–62 BI | 64–71 BI | 74–78 H | 79–87 BI | 90–94 BI | 97–102 BI | SS 22–86 | SS 22–95 | B - I | B - II |
| 0.320 | 0.073 | 0.388 | 0.254 | 0.047 | 0.281 | 0.369 | 0.083 | 0.332 | 0.353 | 0.394 | 0.482 | 0.332 | 0.331 | 0.482 | 0.331 |
| 3–7 BI | 110–114 BI | 117–123 H | 126–135 BI | 141–147 BI | 149–151 BI | 154–158 BI | 161–163 BI | 167–176 BI | 177–183 H | 186–193 BII | 196–202 BII | SS 130–189 | SS 144–200 | B - I | B - II |
| 0.555 | 0.554 | 0.545 | 0.525 | 0.083 | 0.332 | 0.453 | 0.555 | 0.332 | 0.354 | 0.394 | 0.482 | 0.568 | 0.638 | 0.481 | 0.703 |
| 0.320 | 0.073 | 0.388 | 0.254 | 0.047 | 0.281 | 0.369 | 0.083 | 0.332 | 0.353 | 0.394 | 0.482 | 0.332 | 0.331 | 0.482 | 0.331 |
| Domain | Domain | Domain | Domain | Domain | Domain | Domain | Domain | Domain | Domain | Domain | Domain | Domain | Domain | Domain | Domain |
| 0.601 | 0.682 | 0.551 | 0.702 | 0.702 | 0.702 | 0.702 | 0.702 | 0.702 | 0.702 | 0.702 | 0.702 | 0.702 | 0.702 | 0.702 | 0.702 |

Values listed in boldface satisfy RD > 0.5.
protein exists to fulfill a specific biological role and that its structure is a means to this end.

1RBP is a retinol binding protein. The presented domain contains three disulfide bonds which promote tertiary structural stability. Two of these bonds stabilize the domain in a “discordant” state, which is very likely their principal role.

In order to facilitate comparisons with other types of β-dominated supersecondary structures, we have included triole phosphate isomerase (1TIM) in our study. The corresponding form is an α/β barrel consisting of eight pairs of alternating β-strands and α-helices rolled up to form an inner barrel of eight parallel β-strands surrounded by an outer barrel of eight parallel helices. Analysis based on the fuzzy oil drop model indicates the presence of a well ordered hydrophobic core. Stabilization is provided by the arrangement of β-folds (RD = 0.493), while catalytic residues deform the domain (their elimination results in RD = 0.478).

Discordant helixes include the fragments at 95–103 (RD = 0.617), 130–137 (RD = 0.725), 138–154 (RD = 0.527), 168–170 (RD = 0.671), 177–197 (RD = 0.551), 198–204 (RD = 0.590). Catalytic residues are located in direct proximity to the active site. Experience with the fuzzy oil drop model indicates that enzymatic active sites are, in most cases, deficient in terms of hydrophobicity and diverge significantly from theoretical values. In the presented cases catalytic residues belong to the helices at 95–103, 168–170, 177–197, and 198–204, with RD values in excess of 0.5 in all cases.

**Miscellaneous**

The arrangement of helical and β-folds in higher-order structures varies greatly both with respect to the quantity of each and their conformation. An example of a purely α protein is provided by the b562 cytochrome (1APT) from *E. coli*. Topologically, this protein is a sequence of antiparallel helical fragments described as an up-down-up-down 4-helix bundle.

In contrast, the Fab IgG immunoglobulin domain (7FAB) is comprised almost entirely of β-folds, with its structure characterized as a sandwich of 3- and 4-stranded antiparallel β-sheets.

A combination of helical and β-motifs is found in the 163-residual domain of dogfish lactate dehydrogenase (6LDH), which contains a 6-stranded parallel β-sheet with crossovers between β-strands and α-helices forming a right-handed helical turn with flanking β-strands.

The cytochrome is characterized by RD = 0.426, suggesting the presence of a well-ordered hydrophobic core. All its helical fragments also satisfy RD < 0.5 (3–20: RD = 0.355; 23–43: RD = 0.481; 45–49: RD = 0.342; 57–81: RD = 0.402; 82–94: RD = 0.408). It seems that this protein depends strongly on the stabilizing influence of hydrophobic interactions.

Given that the protein’s biological role is to bind heme, its structure provides suitable conditions for firm anchoring of the ligand at its center, with the polypeptide chain forming an “envelope” around the heme molecule.

---

**Table 7** RD values calculated for the N-terminal domain of dogfish lactate dehydrogenase (6LDH)

| Fragment | RD  |
|----------|-----|
| Domain 20–162 | 0.428 |
| 21–27 B | 0.146 |
| 29–41 H | 0.241 |
| 47–52 B | 0.293 |
| **54–68 H** | **0.525** |
| 76–80 B | 0.337 |
| 82–87 H | 0.447 |
| 90–95 B | 0.214 |
| 105–128 H | 0.447 |
| **132–136 B** | **0.559** |
| 139–152 H | 0.349 |
| 157–161 B | 0.332 |
| β-sheet | 0.366 |
| HELICES | 0.396 |

Values listed in boldface satisfy RD > 0.5.

The asterisk marks the right-handed helical fold which causes β-strand flanking.
The Fab IgG domain fragment consists of two chains: the light chain (L) and the heavy chain (H), each of which is further composed of two domains (labeled V and C). All these structures exhibit a similar conformation, often referred to as an immunoglobulin-like domain: a sandwich comprising two β-sheets which we will refer to as the upper core and lower core respectively.

Under the CATH classification the Fab fragment is characterized as 2.60.40.10 mainly β-sandwich. Both cores are linked by a centrally placed disulfide bond. The biological role of the Fab fragment is to recognize antigens via so-called CDRs—short variable loops which interact with the antigen, triggering a process called immunological signal transduction. This produces structural changes in the Fc fragment of immunoglobulin, priming it for complexation of the C1q complement, which is a first step in a cascade of reactions resulting in destruction of the cell which originally presented the antigen.

The role of the Fab fragment is therefore to bind the antigen and trigger a signaling pathway. Both processes require a flexible structure, capable of accommodating the antigen and producing structural changes elsewhere in the molecule. The flexibility must, however, remain selective so as to ensure that the correct ligand is recognized and the correct signal sent.

This situation is evidenced by the RD status of individual β-folds in all four domains: VH, CH1, VL, and CL (see Table 6 and Fig. 9).

The variable status of individual secondary folds comprising Fab domains suggests local instabilities required for complexation of antigens. One of the domains (CL) is stable as a whole and includes a stable β-I sheet, while all other domains are characterized as relatively unstable and therefore flexible. The stabilizing influence of SS bonds likewise varies: only in the CL domain does the affected fragment remain consistent with theoretical predictions.

Immunoglobulins work in an unpredictable environment—the antigen is never known a priori. Thus, a system characterized by variable local stability creates suitable conditions for targeted signaling while preserving the immunoglobulins' ability to interact with a variety of antigens.

A review of immunoglobulin-like domains (which includes domains not associated with immunoglobulin activity, including enzymes and structural proteins) can be found in [8]. The authors reveal a specific arrangement of accordant and discordant fragments comprising the domain, which predispose it toward specific structural changes. Overall, the domain is characterized by poor stability, evidenced by locally high RD values.

The dogfish lactate dehydrogenase N-terminal 163-aa fragment is an example of a 6-strand parallel β-structure linked by a right-handed helical turn. The domain as a whole is characterized by RD = 0.489, while elimination of its N-terminal fragment (1–19 aa), which significantly disrupts the globular form, further reduces RD to 0.428. Evidently, the domain contains a prominent hydrophobic core. Table 7 lists the RD status of its individual folds (compare also Fig. 10 and Fig. 11).

**Cytochrome fold**

The author of [1] also addresses the concept of protein families, with cytochromes discussed as a representative example. This family comprises both eukaryotic cytochromes and c-type cytochromes present in prokaryotes. Both fulfill a similar biological role, acting as electron
carriers. Cytochromes derived from various organisms are sequentially dissimilar despite retaining similar 3D forms. The conclusion presented in [1] is that evolution preserves the structural and functional properties of proteins rather than their sequential arrangement. We further propose that the subject of evolutionary conservation is the structure of the proteins hydrophobic core — as we attempt to substantiate in this part.

### Table 9 RD values calculated for cytochromes b_{562}, c_{2} and c

| Protein             | Fragment | RD   | Fragment | RD   |
|---------------------|----------|------|----------|------|
| 256B Ecoli 1.20.120.10 mainly α up-down bundle | 2–20 H   | 0.411| 2–20 H   | 0.341|
| Cytochrome b_{562}  | 22–41 H* | 0.543| 13–15 H* | 0.295|
|                     | 45–49 H* | 0.371| 16–49 L* | 0.402|
|                     | 55–81 H* | 0.424| 50–59 H* | 0.506|
|                     | 83–92 H  | 0.300| 60–63 L* | 0.777|
|                     | 93–106*  | 0.377| 64–72 H* | 0.211|
|                     |          |     | 74–86 H | 0.661|
|                     |          |     | 88–93 H | 0.446|
|                     |          |     | 94–103 L* | 0.487|
| Ligand              | 0.225    |     | Ligand   | 0.549|
| No ligand           | 0.408    |     | No ligand| 0.457|

Values listed in boldface satisfy RD > 0.5. Asterisks mark fragments involved in interaction with the ligand. The row labeled “Ligand” corresponds to ligand-binding residues, while the row labeled “No ligand” presents residues not involved in ligand interaction.
Table 8 and Table 9 present the properties of cytochromes under consideration. The FOD status of the entire molecule appears preserved despite differences in chain length (88–135 aa) along with the placement and conformation of individual secondary folds. “No ligand” indicates parts of the molecule which do not contact the ligand. Evidently, such fragments are also characterized by consistent RD values.

Cytochromes are proteins which act as “envelopes” for their ligand (heme – see Fig. 12). This ligand is a large, flat molecule dominated by hydrophobic interactions. Comparative analysis of cytochromes boils down to comparing various forms of “envelopes” which create suitable conditions for maintaining the intended level of oxidation of iron ions and expose them for participation in biological processes. These forms are briefly outlined in Tables 8 and 9. The status of the fragment which does not participate in binding heme appears not to change. This is due to its primary role — structural stabilization. In contrast, heme-binding residues exhibit variable FOD status, which may be related to the process of exposing the ligand for various types of interactions.

The conclusions of [1] also highlight the high structural similarity of the NAD⁺ binding fragment in enzymes which participate in many different metabolic pathways. These proteins are more widely discussed in [8].

**Discussion and conclusions**

Applying the fuzzy oil drop model to structures classified using the supersecondary fold criterion reveals similar FOD characteristics, with hydrophobic cores observed either on the domain or on the protein level. A global analysis of a broad nonredundant set of proteins derived from the PDB database indicates that most domains generate prominent hydrophobic cores (RD < 0.5) [37]. It seems that, at least on the level of domains, the water environment greatly influences the folding process.

It is up to each researcher to select fragments of interest, depending on the problem being studied. In our experience, fragments which diverge from theoretical predictions produced by the FOD model are typically associated with biological function. The status of particular secondary fragments in supersecondary structures may quantitatively reveal their contribution to overall supersecondary stabilization.

In summary, we can state that the targeted stability required in many biological processes can be supplied by a combination of nonbinding interactions, structural stabilization mediated by hydrophobic forces, and the presence of disulfide bonds. As shown in [38], SS-bonds may either reinforce or disrupt hydrophobic core stability. The influence of the water environment on protein folding also explains the mechanism behind hydrophobic collapse, which reduces the protein’s conformational space sufficiently to enable a random search for the native fold [39]; this shows that hydrophobic collapse is sufficient to solve Levinthal’s paradox [40].

Discussions concerning the fundamental role of the hydrophobic core have a long history [41] with both experimental [42] and theoretical approaches [43].

In the context of these observations and the analysis presented in this paper we can conclude that the answer to the question posed in the title is “yes”. We can furthermore state that the fuzzy oil drop model is able to describe the hydrophobic core structure in both qualitative and quantitative terms including the substantial and specific difference of hydrophobicity distribution observed in amyloids [31].
References

1. Devlin TM (2011) Textbook of biochemistry with clinical correlations, 7th edn. Wiley, New York
2. Sillitoe I, Lewis TE, Cuff AL, Das S, Ashford P, Dawson NL, Furnham N, Laskowski RA, Lee D, Lees J, Lehitten S, Studer R, Thornton JM, Oregen CA (2015) CATH: comprehensive structural and functional annotations for genome sequences Nucleic Acids Res. doi:10.1093/nar/gku947
3. Skolnick J, Gao M, Zhou H (2016) How special is the biochemical function of native proteins? F1000Res 2016:5
4. Davis IW, Arendall 3rd WB, Richardson DC, Richardson JS (2006) The backrub motion: how protein backbone shuffles when a sidechain dances Structure 14(2):265–274
5. Banach M, Przynula K, Jurkowski W, Konieczny L, Roterman I (2012) Fuzzy oil drop model to interpret the structure of antifreeze proteins and their mutants J Mol Model 18:229–237
6. Banach M, Konieczny L, Roterman I (2012) Use of the “fuzzy oil drop” model to identify the complexation area in protein homodimers. In: Roterman-Konieczna I (ed) Protein folding in silico. Woodhead, Oxford, pp. 95–122
7. Banach M, Konieczny L, Roterman I (2012) Ligand-binding-site recognition. In: Roterman-Konieczna I (ed) Protein folding in silico. Woodhead, Oxford, pp. 78–94
8. Banach M, Konieczny L, Roterman I (2013) Can the structure of hydrophobic core determine the complexation site? In: Roterman-Konieczna I (ed) Identification of ligand binding site and protein–protein interaction area. Springer, Berlin, pp. 41–54
9. Przynula K, Jadczyk T, Roterman I (2011) Catalytic residues in hydrolases: analysis of methods designed for ligand-binding site prediction J Comput Aided Mol Des 25:117–133
10. Auerbach G, Ostendorp R, Prade L, Komodör I, Dams T, Huber R, Jaenicke R (1998) Lactate dehydrogenase from the hyperthermic bacterium thermotoga maritima: the crystal structure at 2.1 A resolution reveals strategies for intrinsic protein stabilization Structure 6:769–781
11. Tougaard P, Bizebard T, Ritco-Vonovitch M, Minard P, Desmaril M (2002) Structure of a circularly permuted phosphoglycerate kinase Acta Crystallogr D Biol Crystallogr 58:2018–2023
12. Williams JC, Zeelen JP, Neubauer G, Vriend G, Backmann J, Michels PA, Lambeir AM, Wierenga RK (1999) Structural and mutagenesis studies of leishmania triosephosphate isomerase: a point mutation can convert a mesophilic enzyme into a superstable enzyme without losing catalytic power Protein Eng 12:243–250
13. Cook WJ, Senkovich O, Aleem K, Chattopadhyay D (2012) Crystal structure of Cryptosporidium parvum pyruvate kinase PLoS One 7:e46875
14. Hart PJ, Balbimie MM, Ogihara NL, Nersissian AM, Weiss MS, Valentine JS, Eisenberg D (1999) A structure-based mechanism for copper-zinc superoxide dismutase Biochemistry 38:2167–2178
15. Naismith JH, Habash J, Harrop S, Holliswell JR, Hunter WN, Wan TC, Weisgerber S, Kalb AJ, Yariv J (1993) Refined structure of cadmium-substituted concanavalin A at 20 resolution Acta Crystallogr D Biol Crystallogr 49:561–571
16. Cowan SW, Newcomer ME, Jones TA (1990) Crystallographic refinement of human serum retinol binding protein at 2A resolution Proteins 8:44–61
17. Kuhn P, Tarentino AL, Plummer Jr TH, Van Roey P (1994) Crystal structure of peptide-N4-(N-acetyl-beta-D-glucosaminyl)asparagine amide F at 2.2-A resolution Biochemistry 33:11699–11706
18. Banner DW, Bloomer A, Petsko GA, Phillips DC, Wilson IA (1976) Atomic coordinates for triose phosphate isomerase from chicken muscle Biochem Biophys Res Commun 72:146–155
19. Lederer F, Glatigny A, Bethge PH, Bellamy HD, Matthew FS (1981) Improvement of the 2.5 A resolution model of cytochrome c562 by redefining the primary structure and using molecular graphics J Mol Biol 148:427–448
20. Saul FA, Poljak RJ (1992) Crystal structure of human immunoglobulin fragment Fab New refined at 2.0 A resolution Proteins 14:363–371
21. Abd-Zapatero C, Griffith JP, Sussman JL, Rossman MG (1987) Refined crystal structure of dogfish M4 apo-lactate dehydrogenase J Mol Biol 198:445–467
22. Timkovich R, Dickerson RE (1976) Refined crystal structure of dogfish M4 apo-lactate dehydrogenase J Biol Chem 251:4033–4046
23. Camara-Artigas A, Williams JC, Allen JP (2001) Structure of cytochrome c2 from Rhodospirillum centenum Acta Crystallogr D Biol Crystallogr 57:1498–1505
24. Bhagia G, Finzel BC, Kraut J (1984) Electron transport protein (cytochrome) from Rhodospirillum rubrum PDB Depositions Authors: of 2C2C. http://www.rcsb.org/pdb/explore/explore.do?structureId=2c2c
25. Takano T (1984) Methods and applications in crystallographic computing. Clarendon, Oxford, p. 262
26. Yu LJ, Umno M, Kimura Y, Yanagimoto K, Oh-oka H, Wang-otomo ZY (2013) Structure analysis and characterization of the cytochrome c544 from thermophilic green sulfur photosynthetic bacterium Chlorobaculum tepidum Photosynth Res 118:249–258
27. Konieczny L, Breyúlskí M, Roterman I (2006) Gauss-function-based model of hydrophobicity density in proteins In Silico Biol 6:15–22
28. Kalinowska B, Banach M, Konieczny L, Roterman I (2015) Application of divergence entropy to characterize the structure of the hydrophobic core in DNA interacting proteins Entropy 17(3): 1477–1507
29. Kauzmann W (1959) Some factors in the interpretation of protein denaturation Adv Protein Chem 14:1–63
30. Levitt MA (1976) A simplified representation of protein conformations for rapid simulation of protein folding J Mol Biol 104:59–107
31. Roterman I, Banach M, Kalinowska B, Konieczny L (2016) Influence of the aqueous environment on protein structure—a plausible hypothesis concerning the mechanism of amyloidogenesis Entropy 18(10):351
32. Kullback S, Leibler RA (1951) On information and sufficiency Ann Entropy 18(10):351
prediction of the fold code: application to immunoglobulin and flavodoxin cases PLoS One 10(4):e0125098

36. Banach M, Konieczny L, Roterman I (2014) The fuzzy oil drop model, based on hydrophobicity density distribution, generalizes the influence of water environment on protein structure and function J Theor Biol 359:6–17

37. Sałapa K, Kalinowska B, Jadczyk T, Roterman I (2012) Measurement of hydrophobicity distribution in proteins – complete redundant protein data bank Bio-Algorithm and Med-Systems 8(2):195–206

38. Lin MM, Zewail AH (2012) Hydrophobic forces and the length limit of foldable protein domains Proc Natl Acad Sci U S A 109(25):9851–9856

39. Levinthal C (1968) Are there pathways for protein folding? J Chem Phys 65:44–45

40. Tanford C (1968) Protein denaturation Adv Protein Chem 23:121–282

41. Agashe VR, Shastry MC, Udgaonkar JB (1995) Initial hydrophobic collapse in the folding of barstar Nature 377(6551):754–757

42. Duan Y, Kollman PA (1998) Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution Science 282(5389):740–744

43. Burke MG, Woscholski R, Yaliraki SN (2003) Differential hydrophobicity drives self-assembly in Huntington’s disease Proc Natl Acad Sci U S A 100(24):13928–13933