STRUCTURAL SIMILARITY OF CheY-LIKE PROTEINS

MATEUSZ BANACH[1], LESZEK KONIECZNY[2] AND IRENA ROTERMAN[1]

[1]*Department of Bioinformatics and Telemedicine*
Jagiellonian University – Medical College
Lazarza 16, 31-530 Krakow, Poland

[2]*Chair of Medical Biochemistry, Jagiellonian University – Medical College*
Kopernika 7, 31-034 Krakow, Poland

(received: 7 September 2016; revised: 5 October 2016;
accepted: 11 October 2016; published online: 24 October 2016)

Abstract: The problem of structural similarity of polypeptide chains of low sequence similarity representing a similar 3D structural form has been the object of analysis of researchers engaged in the protein folding problem. Three homologous proteins of similar biological function with low sequence similarity are the objects of analysis presented in this paper. The structure of a hydrophobic core is used as the criterion for structural similarity assessment of these three proteins. The applied method allows recognition of differentiation in topologically similar structures.

Keywords: hydrophobicity, tertiary structure, hydrophobic core
DOI: https://doi.org/10.17466/tq2016/20.4/j

1. Introduction

Three proteins: the sporulation response regulator Spo0F [1], the N-terminal receiver domain of nitrogen regulation proteins NtrC [2] and the bacterial chemotaxis protein Che-Y [3] were taken as examples representing a similar fold with low sequence identity to show the differentiation of the structure of their hydrophobic core. The biological activity is coded in the 3D structure of the native form of proteins. No activity is possible without local instability allowing local flexibility. The instability may be identified using the simulation of molecular dynamics. However, the differentiation of possible local flexibility is expected to be also coded in the 3D structure of protein. The local instability may be a consequence of lower non-bonding interaction. The system of SS-bonds is treated commonly as responsible for tertiary structure stabilization. These bonds are not

present in selected proteins. The second important factor for the tertiary structure stabilization is the presence of a hydrophobic core, the structure of which may differ locally.

The "fuzzy oil drop" model [4] allows identification of the stability of the hydrophobic core as a whole. It allows also recognition of particular fragments of the polypeptide chain of lower participation in the hydrophobic core formation.

Since the proteins selected for analysis represented a highly similar 3D structure, fragments of a well defined secondary structure were selected as structural units for their participation in the hydrophobic core formation.

The "fuzzy oil drop" model is described in detail in [4] and recently in [5]. A short general description this model only is given in this paper.

2. Materials and Methods

2.1. Data

As mentioned above, three proteins were selected for analysis. Their short characteristics is given in Table 1.

Table 1. The proteins selected for analysis; Each proteins is briefly characterized and the reference publication with a detailed description of the specific protein is quoted

	Function	Length	Origin	CATH classification	Ref.
1SRR	phosphatase resistant mutant of sporulation response regulator spo0f	119	Bacillus subtilis	3.40.50.2300 Alpha3 – Beta Layer (aba) Sandwich	[1]
1DC7	transiently phosphorylated "switch" in bacterial signal transduction	124	Salmonella typhimurium	3.40.50.2300 Alpha3 – Beta Layer (aba) Sandwich	[2]
3CHY	Bacterial chemotaxis protein CheY signal transduction protein	128	Ecoli	3.40.50.2300 Alpha3 – Beta Layer (aba) Sandwich	[3]

As can be seen in Table 1 all the proteins under consideration represent the same class according to the CATH classification [6]. Nevertheless, the sequence similarity is very low [7]. The sequence similarity can be expressed as follows: in the polypeptide chains (128 aa in 3CHY, 124 aa in 1DC7 and 124 aa in 1SRR) 23 residues were found to be identical, 26 positions occupied by similar residues and 12 positions representing similar hydrophobicity status expressed by the hydrophobicity parameter. It means that about half of the chain length is different.

2.2. Fuzzy oil drop model description

According to the "fuzzy oil drop" model it is assumed that the idealized hydrophobicity distribution in the ordered hydrophobic core may be expressed by the 3D Gauss function [8]. The hydrophobic core in the encapsulated protein molecule in ellipsoid is assumed to follow the distribution expressed by this

function. It means: the maximum hydrophobicity density in the center of ellipsoid (center of molecule), a decrease in hydrophobicity with an increase in the distance versus the center reaching the zero level on the surface (in the 3sigma distance versus the center). Such idealized hydrophobicity distribution produces a protein molecule of perfect solubility – what is expected in the majority of proteins. Nonetheless, on the other hand, such a molecule is deprived of any form of activity since it is not interested in any form of interaction with other molecules except water. This is why a local discrepancy is observed in many proteins. The local excess of hydrophobicity (specially on the surface) suggests potential possibility for protein complexation. The local hydrophobicity deficiency (usually occurring in a cavity) is assumed to be ready to bind the ligand molecule.

This is why the search for local discrepancy versus the idealized hydrophobicity distribution is an important factor for similarity search in proteins of low sequence similarity.

The quantitative measurement of accordance/discordance of the observed hydrophobicity distribution versus the idealized one is expressed using the Kullback-Leibler entropy [9]. The reference distribution for the observed distribution are: the theoretical (accordant with 3D Gauss function) and unified distributions (hydrophobicity equal for each residue in a molecule) representing the status of hydrophobicty concentration entirely deprived of any form. RD – the relative distance parameter was introduced to quantitatively measure the status of the observed distribution. Its value is calculated as follows:

The Kullbach-Leibler entropy is calculated for the relation of the observed profile taking the idealized distribution as the target. In the next step the distance between the observed distribution is calculated taking the unified distribution as the target. To avoid the two values to express the status of the observed distribution, RD – the relative distance was introduced. Its value expresses the distance between the theoretical and observed distributions in relation to the sum of the two distances: versus the theoretical and unified ones. This is why the RD lower than 0.5 suggests the presence of a well defined hydrophobic core, while $RD > 0.5$ suggests higher proximity versus a unified distribution – it means that the distribution is deprived of a well defined core in the center of the molecule.

The RD values calculated for selected polypeptide chain fragments (for example: secondary fragments) measure the degree to which a particular fragment is participating (or not participating) in the generation of the global hydrophobic core formation.

2.3. Sequence similarity measurement

The sequence similarity was measured using the Clustal W Omega program available at http://www.ebi.ac.uk/Tools/services/web/toolresult.ebi?tool=clustalo&jobId=clustalo-E20161128-183836–0113–44133942-es .

3. Results

The results of the analysis are given in Table 2.

Table 2. The RD values for three proteins selected for analysis; Values given in bold – fragments characterised by RD > 0.5

3CHY		0.417	1DC7		0.393	1SRR		0.362
Secondary form	fragment	RD	Secondary form	fragment	RD	Secondary form	fragment	RD
Loop	2–6	0.192	Loop	1–5	0.149	Beta	4–10	0.188
Beta	7–12	0.208	Beta	6–9	0.314	Helix	12–25	0.476
Helix	14–19	0.427	**Loop**	10–13	**0.665**	Loop	26–28	0.140
Beta	32–36	0.410	Helix	14–24	0.446	Beta	29–33	0.302
Helix	38–49	0.400	**Loop**	25–30	**0.549**	Helix	35–47	0.315
Loop	50–52	**0.559**	Beta	31–33	0.422	Beta	49–55	0.203
Beta	53–58	0.226	Loop	34–36	0.403	Loop	56–60	0.450
Loop	59–63	0.386	Helix	37–43	0.424	Helix	61–73	0.303
Beta	64–75	0.293	**Loop**	44–49	**0.666**	Loop	74–76	0.073
Loop	76–81	**0.755**	**Beta**	50–54	**0.515**	Beta	77–82	0.182
Beta	82–87	0.351	Loop	55–64	0.458	Loop	83–85	0.192
Loop	88–90	0.105	Helix	65–73	0.274	Helix	86–97	0.295
Helix	91–101	0.252	Loop	74–106	0.324	Beta	100–104	0.113
Loop	102–104	**0.646**	Helix	107–122	0.219	Helix	107–119	0.206
Beta	105–109	0.390						
Helix	112–128	**0.507**						
Beta-sheet		0.246	**Beta-sheet**		**0.629**	Beta-sheet		0.230

According to the analysis based on the "fuzzy oil drop" model all three proteins represent the status with a well defined hydrophobic core (the RD values calculated for the entire domain are below 0.5).

Nonetheless, the status of selected fragments of secondary forms differs. The 1SRR appears to represent the lowest value of the RD for the entire molecule and for all the identified secondary fragments. In 3CHY and in 1DC7 some fragments appear to be discordant versus the assumed idealized hydrophobicity distribution.

Three loops and one helical fragment in 3CHY were recognized as discordant versus the assumed distribution. In 1DC7 also short loops and one β-fragment were identified as discordant.

The comparison of the β-sheet status suggests a similar status for 3CHY and 1SRR while in 1DC7 the beat-sheet appears to be discordant.

The analysis of the results expressed by the RD with respect to the biological activity of proteins under consideration is as follows.

3.1. 1SRR – Spo0F, a phosphotransferase containing an aspartyl pocket

Spo0F is involved in the signaling pathway (phosphorelay) controlling sporulation in Bacillus subtilis. This is the phosphotransferase containing an

aspartyl pocket. This protein belongs to the superfamily of bacterial response regulatory proteins, which requires phosphorylation of an invariant aspartate residue.

The Spo0F protein represents a structure highly accordant versus the idealized distribution of hydrophobicity. It can be seen in Figure 1 and Table 2. All fragments representing secondary structural forms also represent the status close to the expected one. The only fragment recognized on the profile (Figure 1) is a fragment of helix (residues 13–19). The RD for this fragment is equal to 0.913. Figure 1 visualizes the opposite distribution of the observed versus expected hydrophobicity. It should be emphasized that the Kullback-Leibler entropy recognizes them as different distributions which are opposite to each other.

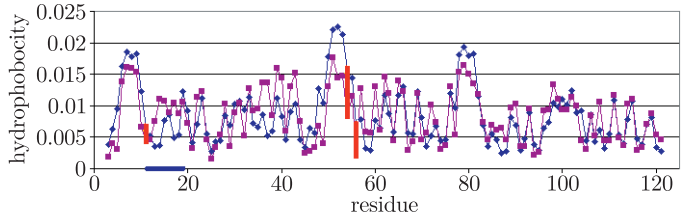

Figure 1. Hydrophobicity profile: blue – theoretical distribution (T), magenta – observed (O) hydrophobicity distribution in 1SRR; The residues engaged in Ca^{2+} complexation are marked by red vertical lines; The black line on the x-axis shows the fragment of local RD > 0.5; The red vertical lines show residues engaged in biological activity

This identified fragment is closely located with respect to the residues engaged in the biological activity (Figure 2). It can be speculated that the local elasticity (assuming that the hydrophobic core is responsible for the tertiary structure stabilization) is necessary for the required structural changes which, as natural, should be present in the process related to this activity.

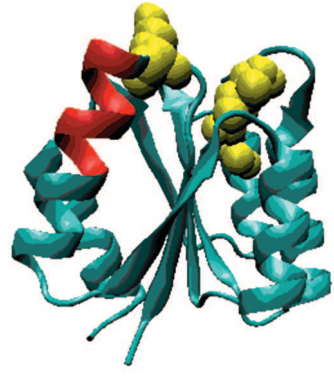

Figure 2. 3D structure of 1SRR protein with a red fragment visualizing the region of discordance versus the idealized distribution of hydrophobicity. The yellow residues (VDW presentation) visualize the location of residues related to biological activity (according to [1])

The distributions (T and O) in this fragment appear to be opposite to each other. The proximity of this fragment versus the active center may suggest to be responsible for the local force field for the Ca^{2+} binding. So far, this is speculation only, however, this set of conditions (ligand binding and local discordance in close proximity) is observed rather frequently. The highly discordant fragment is the part of helix. Opposite distribution in relation to the expected one may suggest possible rotation of the helix or swinging in search for a better fit to the ordered hydrophobic core in this area.

3.2. 1DC7 – N-terminal receiver domain of nitrogen regulation protein NtrC

This domain represents the status of RD = 0.393 – higher than 1SRR, nonetheless still classifying this domain as a highly ordered one.

The analysis of the profiles of the T and O hydrophobicity distribution shown in Figure 3 (and Table 2) reveals the loops 10–13, 25–30 and 44–49 as well as the β-fragment 50–54 as discordant versus the expectations. The β-fragment 50–54 which is central introduces local instability (assuming that the hydrophobic core is responsible for the tertiary structure stabilization) (Figure 4). This is why the entire β-sheet appears to represent the RD above 0.5. In contrast to 1SRR the potential flexibility may occur in the central part of the molecule since the highest expected hydrophobicity appears not to be present.

It should be emphasized that the hydrophobic core – as it is understood in the "fuzzy oil drop" model – includes also the hydrophilic external shell as an integral part of the ordered hydrophobicity distribution in the protein molecule.

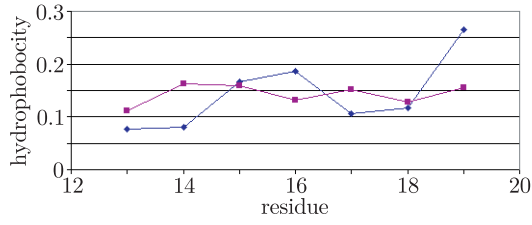

Figure 3. Hydrophobicity distribution in 1DC7: blue – theoretical (T) and magenta – observed (O) for fragment 13–19 to visualize the opposite status of residues present in the fragment under consideration

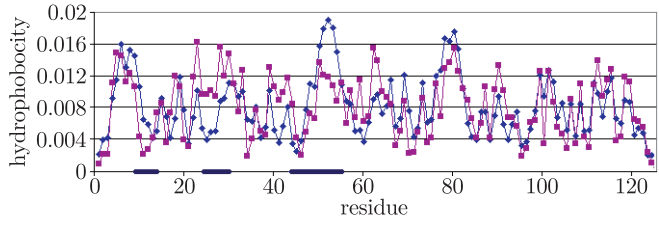

Figure 4. Hydrophobicity profile: blue – theoretical (T), magenta – observed (O) hydrophobicity in 1DC7. The fragments of RD > 0.5: loops 10–13, 25–30 44–49 and β-fragment 50–54 are marked by the dark line on the x-axis

This is why the entire molecule is treated by a fuzzy oil drop based analysis as accordant with the expectation.

The structural characteristics (as reported in [2]) suggest a shift of β-strands 4 and 5 and α-helices 3 and 4 away from the active site and the axial rotation in helix 4 seems to create an exposed hydrophobic surface that is likely to transmit the signal to the transcriptional activation domain. The analysis of the hydrophobic core of the 1DC7 molecule suggests lower stabilization of one β-fragment (in 3-β-fragment β-sheet). The localization of less stable loops suggests possible potential conformational changes (Figure 4).

The T and O distribution for the fragment 13–19 shown in Figure 5 visualizes the opposite tendency for each residue in this fragment. This is an example of the Kullback-Leibler entropy to identify this fragment as highly discordant versus an idealized one.

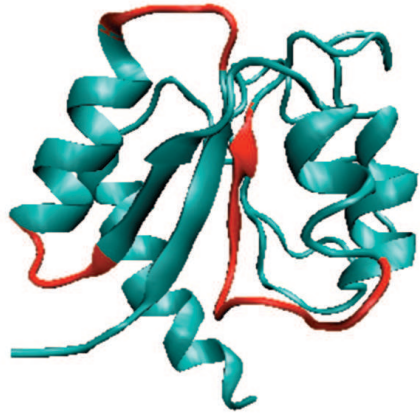

Figure 5. 3D presentation of the 1DC7 structure with fragments (red) recognized by the "fuzzy oil drop" model as discordant versus the idealized hydrophobic core structure

3.3. 3CHY – bacterial chemotaxis protein CheY

The discordant status versus the expected ordered profile is visualized in Figure 6. According to [3] the active site area is bordered by the carboxyl termini of the three central β-strands, by alpha 1, and by the loop connecting β-fragment 5 to alpha 5 marked as red fragments in Figure 7. The epsilon-amino group of Lys-109 is in a close bonding contact with the carboxyl group of Asp-57. These two residues marked in Figure 7 in yellow visualize the position versus the unstable fragments.

The details of the hydrogen bonding network in the phosphorylation region indicate that the phosphorylation of Asp-57 must be accompanied by structural rearrangements.

An analysis of the T and O profiles reveals the fragment 112–115 as highly discordant (RD = 0.697). This fragment of helix (112–128) appears discordant, however, not as much as in 1SRR.

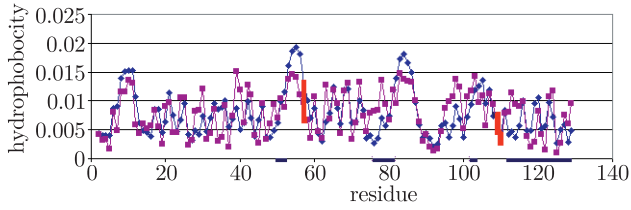

Figure 6. Hydrophobicity profile: blue – theoretical (T), magenta – observed (O) hydrophobicity in 3CHY; The fragments of RD > 0.5 are marked by the black line on the x-axis. The residues: Asp57 and Lys109 are marked in red

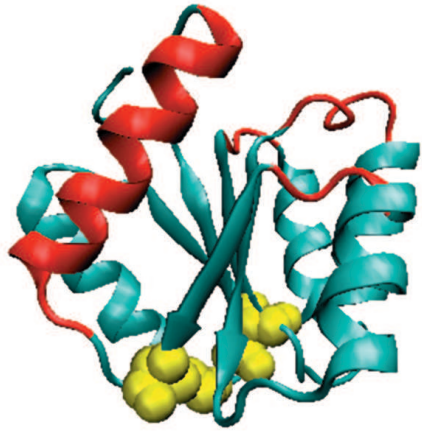

Figure 7. 3D structure of 3CHY with red fragments visualizing regions representing the hydrophobicity distribution discordant versus the idealized distribution; The residues are marked as yellow: Asp57 and Lys109

The positions of loops of lower stability (taking the hydrophobic core as a criterion) are localized versus 57Asp and 109 Lys on the opposite site of the molecule, revealing possible cooperation in conformational changes required for the biological activity of this molecule.

4. Conclusions

The 3D structure of protein molecules is an he effect of the non bonding interaction and the influence of the water environment which directs the holding polypeptide to generate the hydrophobic core (concentration of hydrophobic residues in the center with the exposed hydrophilic residues on the surface). The hypothetical protein folded according to the idealized hydrophobicity distribution according to the 3D Gauss function is perfectly well soluble – what is expected for proteins, however, on the other hand, the protein is not able to interact with any other molecule except water. The examples of proteins with a structure following the 3D Gauss function distribution of hydrophobicity are antifreeze proteins [10] and down-hill proteins [11]. It seems that these two groups follow the "fuzzy oil drop model" mechanism. However, they are exceptions in the world of proteins.

All other proteins prepared by the Nature to be active will represent controlled discordance versus an idealized model which seems to be related to the biological activity.

Three proteins the structure of which is discussed in this paper represent a highly similar topology. The surprise is that their sequence is quite different (15% identity). The differentiation is necessary to ensure the local discordance which – as is shown – may be related to their biological activity. The subtle balance between highly stable (a fragment of a molecule accordant with the expected hydrophobic core structure) and relatively locally differentiated lower stability (fragments discordant versus the expected distribution) can be observed.

The different distribution of fragments of low accordance with the idealized distribution seems to differentiate the potential action coded in the very similar topology of the protein structure.

The three proteins are responsible regulators involved in various signal transduction pathways [4–7]. This process requires phosphorylation which induces a large conformational change involving a displacement of secondary fragments of polypeptide chain in these proteins. This process creates an exposed hydrophobic surface that is likely to transmit the signal to the transcriptional activation domain as reported in [2]. The analysis presented in this paper is an attempt to evaluate the relative stabilization of specific polypeptide chain fragments. The differentiation observed using the fuzzy oil drop model suggests different localization of fragments of potential conformational changes.

A large set of proteins representing the flavodoxin fold with different sequence similarity was the object of analysis in [12]. The results presented there reveal the differentiation of the status of secondary fragments despite a highly similar topology of compared proteins.

The reliability of the fuzzy oil drop model seems to be proved using the set of antifreeze proteins [10] and down-hill proteins [11]. These two groups of proteins appear to represent a distribution of hydrophobicity highly accordant with the expected one. The case of a similar topology of proteins generated by sequences of low similarity was also discussed taking the immunoglobulin fold as the object of analysis [13], revealing significant differences in the order of the hydrophobic core.

Acknowledgements

The work was financially supported by a grant system in Collegium Medicum – Jagiellonian University – Grant # 006363.

References

[1] Madhusudan, Zapf J, Whiteley J M, Hoch J A, Xuong N H, Varughese K I 1996 *Structure* **4** 679
[2] Kern D, Volkman B F, Luginbühl P, Nohaile M J, Kustu S, Wemmer D E 1999 *Nature* **402** 894
[3] Volz K, Matsumura P 1991 *J Biol Chem* **266** 15511
[4] Kalinowska B, Banach M, Konieczny L, Roterman I 2015 *Entropy* **17** (3) 1477

[5] Roterman I, Banach M, Kalinowska B, Konieczny L *Entropy* (in press)
[6] Sillitoe I, Lewis T E, Cuff A L, *et al.* 2015 *Nucleic Acids Res* doi: 10.1093/nar/gku947
[7] Hills R D Jr, Kathuria S V, Wallace L A, Day I J, Brooks C L 3rd, Matthews C R 2010 *J Mol Biol* **398** (2) 332
[8] Konieczny L, Brylinski M, Roterman I 2006 *In Silico Biol* **6** (1-2) 15
[9] Kullback S, Leibler R A 1951 *Ann. Math. Stat.* **22** 79
[10] Banach M, Prymula K, Jurkowski W, Konieczny L, Roterman I 2012 *J Mol Model* **18** (1) 229
[11] Roterman I, Konieczny L, Jurkowski W, Prymula K, Banach M 2011 *J Theor Biol* **283** (1) 60
[12] Banach M, Prudhomme N, Carpentier M, Duprat E, Papandreou N, Kalinowska B, Chomilier J, Roterman I 2015 *PLoS One* **10** (4), e0125098
[13] Banach M, Konieczny L, Roterman I 2014 *J Theor Biol* **359** 6