Bioinformation
www.bioinformation.net

open access
Hypothesis

Recognition of protein complexation based on hydrophobicity distribution

Mateusz Banach[1,2], Irena Roterman[1,*]

[1]Department of Bioinformatics and Telemedicine, Collegium Medium – Jagiellonian University, Lazarza 16, 31-530 Krakow, Poland; [2]Faculty of Physics, Astronomy and Applied Computer Science – Jagiellonian University, Reymonta 4, 30-059 Krakow, Poland; Irena Roterman - Email: myroterm@cyf-kr.edu.pl; [*]Corresponding author

Received August 04, 2009; Accepted August 18, 2009; Published September 30, 2009

Abstract:
The identification of the surface area able to generate the protein-protein complexation ligand and ion ligation is critical for the recognition of the biological function of particular proteins. The technique based on the analysis of the irregularity of hydrophobicity distribution is used as the criterion for the recognition of the interaction regions. Particularly, the exposure of hydrophobic residues on the surface of protein as well as the localization of the hydrophilic residues in the hydrophobic core is treated as potential area ready to interact with external molecules. The model based on the "fuzzy oil drop" approach treating the protein molecule as the drop of hydrophobicity concentrated in the central part of structure with the hydrophobicity close to zero on the surface according to 3-dimensional Gauss function. The comparison with the observed hydrophobicy in particular protein reveals some irregularities. These irregularities seem to represent the aim-oriented localization.

Keywords: hydrophobicity distribution, protein complexation, fuzzy-oil-drop model

Background:
The recognition of potential areas in protein ready to interact with other molecules (protein, ligand, ion, nucleic acids) may be useful for the identification of the biological function of the protein under consideration. The interaction of proteins with other molecules seems to represent the deterministic form since only specific complexes are observed in the organisms. The multi-protein complex structure is equally important for proper functioning as the three-dimensional structure of the protein. The recognition of the protein-protein complexation area and definition of the possible mechanism responsible for complexation is the aim of the project called CAPRI **[1-3].** The interaction of protein with ligands determines also their biological activity. The search for techniques allowing prediction of the ligation site in proteins is also in the focus of the research **[4, 5]** as well as the interaction with ions. The general idea aimed on the recognition of potential areas on the proteins surface engaged in different types of interaction is presented in this paper.

Methodology:
The protein 1KFV (formamido-pyrimidine DNA glycosylase EC#:3.2.2.23, mutation: P1G, polypeptide dimer complexed to DNA **[6])** deposited in PDB was selected to verify the applicability of the assumed model. This molecule appears in form of dimer, complexes DNA molecule, interacts with ligands and ions. Thus is very good example to analyze the potential complexation areas specific for particular "partner" in interaction. The specific areas identification is based on hydrophobicity irregularity versus the expected, idealized one.

See **supplementary material** for additional methodology section.

Discussion:
The $\Delta\widetilde{H}_i$ profile (**Figure 1A**) reveals the differentiation of the irregularity along the polypeptide chain suggesting specificity of particular fragments to be involved in the specific interaction with other molecules. The spatial distribution (**Figure 1B**) visualizes the localization of the irregularities together with the residues engaged in interaction. The structure of complex shown in the **Figure 1B** visualizes the spatial distribution of hydrophobicity irregularity. The $\Delta\widetilde{H}_i$ profile reveals the engagement of the residue of the highest $\Delta\widetilde{H}_i$ in catalytic activity. The hydrophobicity deficiency appears when the low hydrophobicity residue is localized in the cavity. This is the case in this example. This is also observed for residues of local maxima of $\Delta\widetilde{H}_i$ which are engaged in the interaction with DNA molecule – polar residues localized in the deep cavity. Metal ions are localized on the surface (the hydrophobicity excess is observed usually on the surface) in the local $\Delta\widetilde{H}_i$ minima. The residues representing the local maxima of $\Delta\widetilde{H}_i$ are engaged in protein-protein interaction although their compact localization in the well defined area on the protein surface suggests the localization of the protein-protein complexation. The relation of the $\Delta\widetilde{H}_i$ values on the $\Delta\widetilde{H}_i$ profile (**Figure 1A**) to the function in form of different complexation is also shown. The qualitative summary of the characteristics of the $\Delta\widetilde{H}_i$ values for different forms of complexation is given in **Table 1** (**see supplementary material**). The comparison of the values expressing average $\Delta\widetilde{H}_i$ for residues interacting with different molecules reveals the possible mechanism of the ligation. The protein and ligand molecule complexed engaged the residues of relatively high positive $\Delta\widetilde{H}_i$ (see also **Figure 1A** – local maxima). The residues representing the deficiency of hydrophobicity localized in the cavity attract the glycerol molecule engaging its oxygen atoms to interact with ARG (74), TYR (58) and THR (113).

Service description:
Results were obtained using our own software in form of web application written in Python programming language with help of third-party extensions and being served by web2py enterprise framework **[9].** Main idea of the application is to allow automatic evaluations of $\Delta\widetilde{H}_i$ values and non-bonded contact maps for arbitrary numbers of amino-acid residues from given PDB input. These maps describe interactions between protein and protein, ligand, ion or nucleic residues with close accordance to PDBsum service output. User is permitted to choose any model present in the submitted file. Ten jobs analyzing up to ten structures are available after successful parsing and correct identification of all chains present in the file (protein, nucleic acids, ligands). The options oriented on selection of the object under consideration (protein, selected fragment, set of chains) is available. The self-defined orientation of the molecule in the coordinate system is also possible particularly useful for molecules of axial symmetry. For selected residue set, $\Delta\widetilde{H}_i$ values and non-bonded contact map are evaluated and returned together in three possible result files: textual, with complete numeric output, PDB, for 3D visualization and vector (PDF), containing $\Delta\widetilde{H}_i$ plot, contact map scatter and various histograms. Jmol applet allows the visualization of all PDB files directly in the browser. **[10].**

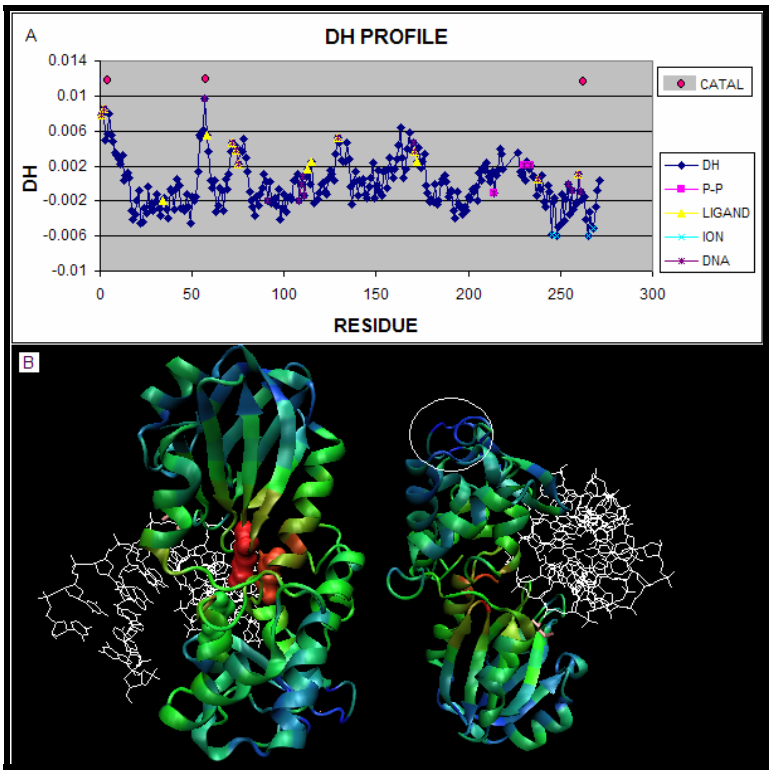

Figure 1: The structure of 1KFV characterized according to hydrophobicity irregularity. A – the $_{\Delta\tilde{H}_i}$ profile with residues engaged in particular form of complexation according to legend. B – the 3-D presentation of the hydrophobicity irregularity distribution colored as follows: the more blue color the lower negative $\Delta\tilde{H}_i$ value, the more red color the higher $_{\Delta\tilde{H}_i}$ value. The dimer structure complexed to DNA (representation using white lines). The white circle distinguishes the metal ion localization of high hydrophobicity excess area on the protein surface. The residues engaged in enzymatic activity given in space-filling graphic version colored according to $\Delta\tilde{H}_i$ values.

Conclusion:

The selected protein appeared to be very good example to verify the applicability of the "fuzzy oil drop" model to recognize and to predict specific characteristics of particular fragment of protein. The interaction between molecules present in the cell is mediated by water molecules. Their organization form was interpreted to influence the interaction between molecules present in the cell. One of the important components for these interactions generating the environment for all processes is the water-dependent hydrophobic interaction. The "fuzzy oil drop" model seems to simulate the form of environment pointing out the specific irregularities and suggesting possible aim-oriented forms in the protein body and on the surface specifically. The example protein presented in this paper encourages its applicability for the large scale of calculation (all proteins present in PDB are going to be analyzed using "fuzzy oil drop" model as the criterion). Particularly the applicability to the CAPRI experiment of the fuzzy oil drop model seems to be promising. The "fuzzy oil drop" model was applied to the identification of the active site and ligation site **[8]** as well as in simulation of protein folding process representing folding environment oriented on concentration of hydrophobic residues in the protein core and the exposure of hydrophilic residues on the surface **[11-13].** The specificity of the distribution of hydrophobicity irregularity suggested also the necessary presence of the specific ligand in the folding environment ensuring the generation of specific ligand binding cavity **[14, 15].** The large scale analysis as well as the search for specificity of the hydrophobicity irregularity in relation to biological activity of the proteins is under consideration.

References:

[1] J Janin *Proteins: Structure, Function, and Genetics* **47**:257 (2002).
[2] SJ Wodak, *Proteins: Structure, Function, and Bioinformatics* **69**:697 (2007).
[3] S Vajda *et al., Proteins: Structure, Function, and Genetics* **47**:444 (2002).
[4] MR Chance *et al., Protein Sci.* **11**:723 (2002).
[5] SK Burley *et al., Nat Genet.* **23**:151 (1999).
[6] L Serre *et al.*, EMBO J. **21:**2854 (2002).
[7] M. Levitt *J Mol Biol* **104**:59 (1976) [PMID: 957439].
[8] M. Brylinski *et al., Bioinformation* **1**:127 (2006).
[9] http://web2py.com
[10] http://jmol.sourceforge.net
[11] L Konieczny *et al., In Silico Biol.* **6**:15 (2006).
[12] M Brylinski *et al., J Biomom Struct Dynam.* **23**:517 (2006).
[13] M Bryliński *et al., Biochemie* **88**:1229 (2006).
[14] M Bryliński *et al., Int J Bioinform Res Applic.* **3**:234 (2007).
[15] M Bryliński *et al., Comp Biol Chem* **30**:255 (2006).

Edited by P. Kangueane
Citation: Banach & Roterman, Bioinformation 4(3): 98-100 (2009)

Bioinformation
www.bioinformation.net
open access
Hypothesis

Supplementary material:

Methodology (additional data)

Expected hydrophobicity distribution:
The theoretical distribution is assumed to follow the three-dimensional Gauss function as described in equation 1.

$$\widetilde{Ht}_j = \frac{1}{\widetilde{Ht}_{sum}} \exp\left(\frac{-(x_j-\bar{x})^2}{2\sigma_x^2}\right)\exp\left(\frac{-(y_j-\bar{y})^2}{2\sigma_y^2}\right)\exp\left(\frac{-(z_j-\bar{z})^2}{2\sigma_z^2}\right) \quad \text{(Equation 1)}$$

where $\sigma_X, \sigma_Y, \sigma_Z$ express the size of the drop, the point $(\bar{x}, \bar{y}, \bar{z})$ (which in traditional interpretation of the Gauss function represents the position of mean values) which is localized in the center of the ellipsoid carries the highest hydrophobicity density. Coefficient $\widetilde{Ht}_{sum}$ expresses the sum of the hydrophobicity of all analyzed grid points (positions of the effective atoms – averaged position of side chains). In consequence the value of $\widetilde{Ht}_j$ is the relative (standardized) hydrophobicity density in *j*-th point. The $\bar{x}, \bar{y}, \bar{z}$ is the geometric center of the molecule localized in the origin of the coordinate system. This is why these values can be taken as equal to zero. The size of molecule is expressed by the σ_x, σ_y, σ_z which is calculated for each molecule individually on the condition of the defined orientation of the molecule with longest possible inter-effective atoms distances oriented according to coordinate system axis orientation. The σ values are calculated as the 1/3 of the longest distance between two effective atoms calculated along each axis. The value of Gauss function in any point of protein body is treated as idealized hydrophobicity density defining the hydrophobic core.

Observed hydrophobicity density:
The hydrophobicity distribution as it appears in real proteins differs versus the expected one. The observed hydrophobicity distribution may be calculated according to the position of hydrophobic residues in the protein body following the Levitt **[7]** function described in equation 2:

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

Where *N* expresses the number of amino acids in protein (number of grid points), $\widetilde{H}_i^r$ expresses the hydrophobicity of The *i-th* residue according to accepted hydrophobicity scale, r_{ij} expresses the distance between *i-th* and *j-th* interacting residua, *c* expresses the cutoff distance which according to the original paper **[7]** is assumed to be 9 Å. Division by the coefficient $\widetilde{Ho}_{sum}$, which is the sum of all hydrophobicities attributed to all grid points makes the values of $\widetilde{Ho}_j$ standardized.

Irregularity in hydrophobicity distribution:
Since both distributions (theoretical and empirical) are standardized the difference between expected and observed distribution (for the position of effective atoms) can be calculated.

$$\Delta\widetilde{H}_j = \widetilde{Ht}_j - \widetilde{Ho}_j \quad \text{(Equation 3)}$$

where $\widetilde{Ht}_j$ and $\widetilde{Ho}_j$ express expected and observed hydrophobicity density respectively for *N* = total number of grid points.

The $\Delta\widetilde{H}_i$ values represent the deviation of the observed distribution versus the idealized one. The sign and quantity of the $\Delta\widetilde{H}_i$ value is assumed to measure the irregularity of hydrophobocity distribution. Many cases reveal these irregularities as carrying the aim-oriented character **[8].** The positions (residues) of high $\Delta\widetilde{H}_i$ values are interpreted as the positions of the hydrophobicity deficiency, while low negative values of $\Delta\widetilde{H}_i$ are interpreted as hydrophobicity excess.

Table 1: The averaged $\Delta\widetilde{H}_i$ values calculated for residues interacting with protein, ligand, metal ion, nucleic acid respectively and residues carrying the enzymatic activity. For simplicity the values are multiplied by 10^3. The value in parenthesis is calculated for four residues including the mutated residue (P1G)

COMPLEX WITH	AVERAGE	NUMER OF RESIDUES
PROTEIN	1.000	3
LIGAND	3.407	14
ION	-5.771	4
NUCLEIC ACID	2.194	22
CATALYTIC RESIDUES	6.457 (6.814)	3 (4)