BIOINFORMATION
Discovery at the interface of physical and biological sciences

open access www.bioinformation.net Hypothesis

Volume 6(8)

Internal force field in proteins seen by divergence entropy

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

[1]Department of Bioinformatics and Telemedicine, Jagiellonian University, Medical College Krakow, Poland; [2]Faculty of Physics, Astronomy and Applied Computer Science, Jagiellonian University, Krakow, Poland; Irena Roterman - Email: myroterm@cyf-kr.edu.pl; *Corresponding author

Received May 28, 2011; Accepted June 27, 2011; Published July 06, 2011

Abstract:
The characteristic distribution of non-binding interactions in a protein is described. It establishes that hydrophobic interactions can be characterized by suitable 3D Gauss functions while electrostatic interactions generally follow a random distribution. The implementation of this observation suggests differentiated optimization procedure for these two types of interactions. The electrostatic interaction may follow traditional energy optimization while the criteria for convergence shall measure the accordance with 3-D Gauss function.

Background:
The "oil drop" model proposed by Kauzmann **[1]** suggested the protein molecule to be treated as oil drop in hydrophobic environment directing the hydrophobic residues toward the central part of protein body with simultaneous exposure of hydrophilic residues on the protein surface **[2]**. The model applied here uses the 3-D Gauss function as the function expressing the distribution of hydrophobic interaction in protein body. The highest concentration of hydrophobic interaction is expected in central part of protein body, as it is for Gauss function, which decreases towards the surface reaching zero (or close to zero) hydrophobicity level, according to its bell-shaped form. The accordance of hydrophobic interaction in protein with the idealized distribution is shown, in contrast to the electrostatic interactions distribution, which appeared to represent the random distribution.

Methodology:
Data:
The protein 2I5M was selected to demonstrate the example revealing the differences between distributions of electrostatic, vdW-type and hydrophobic interactions **[3]**. It is the protein of length 66 amino acid residues, which is participating in gene regulation processes. This protein represents bacterial cold shock proteins (Csp) which are widely used as models for the experimental and computational analysis of protein stability. The 2I5M is the mutant A46K and S48R, produced to reveal the particular role in structure stabilization of the original protein (1CSB).

Model:
The tertiary structure of the protein is assumed to include a hydrophobic core and involve optimization of all other non-bonding interactions (electrostatic, van der Waals and torsion potential). In contrast to many force fields incorporating the hydrophobic interaction as the component of internal force field, the presence of an external force field is expressed via a three-dimensional Gauss function **[2] (see Supplementary material, equation 1)**. The value of the Gauss function at any point within the protein body can be treated as the idealized hydrophobicity density, determining the structure of the protein's hydrophobic core.

According to the "fuzzy oil drop" model, theoretical hydrophobicity can be calculated with the use of the Gauss function, assuming that the molecule's geometric center coincides with the origin of the coordinate system. Empirical hydrophobicity distribution is given by Levitt's function **[4] (see Supplementary material, equation 2)**. Further normalization of both distributions enables quantitative comparison (observed vs. theoretical values), which in turn, allows us to analyze the emerging discrepancies.

Kullback-Leibler information entropy:
The agreement between the idealized and observed hydrophobicity distribution is measured according to the Kullback-Leibler relative (divergence) entropy **[6]**, which quantifies the distance between both distributions. The distance between the observed and theoretical distribution (O/T) was calculated in this study. Since this value can only be interpreted in comparison to other potential solutions, random distribution of hydrophobicity (O/R) was also estimated by assigning equal probability of hydrophobicity to each amino acid (R_j equal to 1/N, where N is the number of amino acids) The relation O/T < O/R was taken as evidence of non-random distribution, approximating theoretical values **(see Supplementary material, equation 3)**. The calculation O/T and O/R for hydrophobic, electrostatic and vdW interactions is presented in this paper.

Nonbonding interactions:
In order to determine the distribution of non-hydrophobic (i.e. electrostatic and van der Waals) interactions, structures derived from PDB were processed using an energy optimization algorithm to eliminate any local collisions. Electrostatic and van der Waals forces (calculated for each atom separately) were then aggregated over whole amino acids, mimicking the distribution of hydrophobic interactions. The distribution of non-bonding interaction was calculated using Gromacs force field applying the group procedure **[7, 8, 9, 10, and 11]**. Each amino acid on each side of the protein molecule was sequentially defined as a group, to calculate its interaction with the rest of the protein molecule.

Discussion:

Distribution of nonbinding interactions in proteins:

The characteristics of hydrophobic, electrostatic and vdW interaction is shown in **Table 1 (see Supplementary material)**. The results expressing the relation O/T < O/R treated as accordant with assumed "fuzzy oil drop" model are given in bold. Good accordance between the observed and theoretical distribution of hydrophobic interaction can be recognized interpreting the values shown in **Table 1 (see Supplementary material)**. The random coil character was recognized for electrostatic interaction. The "fuzzy oil drop" structure for the vdW interactions has been found for this protein. The graphic presentation of the results is shown in **Figure 1.**

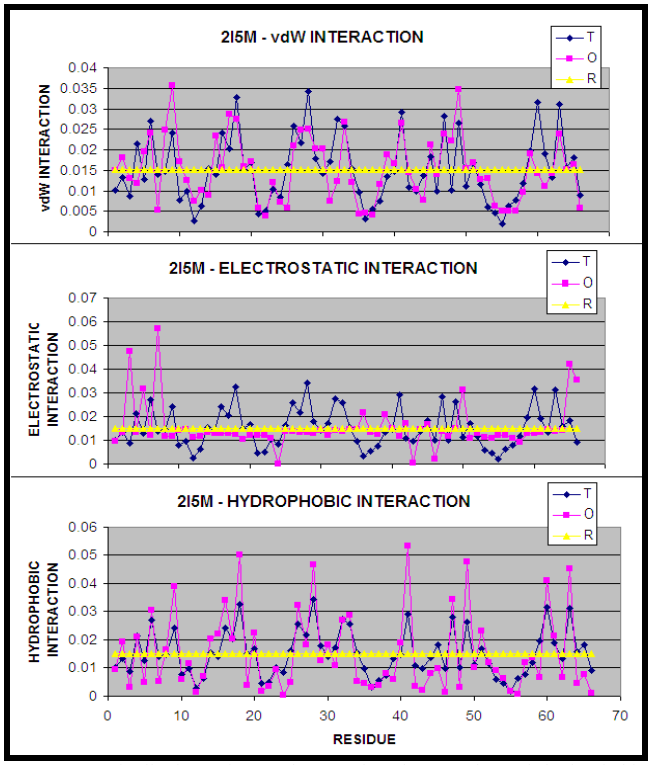

Figure 1: Distribution of van der Waals, electrostatic and hydrophobic interactions (top, middle and bottom respectively) in the 2I5M molecule. Dark blue line indicates theoretical distribution (accordant with the 3D Gauss function); pink line shows actual distribution of hydrophobicity in 2I5M; yellow line corresponds to random distribution.

Conclusion:

It should be noted that only one protein was presented in this study although the larger group of proteins was presented elsewhere **[12]**. It was shown that the accordance of vdW interaction is observed rather rarely, although no protein has been shown to represent other than random distribution for electrostatic interaction. These observations which can be extended to larger number of proteins **[12, 13]** suggest that the optimization procedure applied for pair-wise interaction is proper only for electrostatic interaction (and vdW interaction). The result of such optimization produces the structure of low energy (large stabilization) independently on the localization in the protein body. The optimization of hydrophobic interaction should be treated in different way. This optimization shall be oriented to fit the expected hydrophobic interaction accordant with 3-D Gauss function. The pair-wise interaction for this type of interaction produces different than expected interaction distribution. Sample folding simulations acknowledging the presence of an external field (described by a 3D Gauss function which steers the process towards the generation of a hydrophobic core) were conducted for BPTI **[14]**, lysozyme **[15]** and T0215 **[2]**. Similar computations involving the presence of a ligand were performed for human hemoglobin **[5]** and ribonuclease **[16]**. The ligand was intended to affect the folding process in such a way, as to ensure the creation of a suitable binding pocket. The influence of the external environment (the cellular membrane) upon the dynamic properties of transmembrane proteins was verified with the use of the "fuzzy oil drop" model **[17]**. The applicability of "fuzzy oil drop" model to mutation influence on the structure of antifreeze proteins is presented elsewhere **[18]**.

The "fuzzy oil drop" model applied for the identification of ligand binding sites **[19]** and active sites in hydrolases **[20]** yields insights into the properties of individual proteins. Deformations in the structure of the hydrophobic core resulting from interaction with external molecules can be used to identify binding sites (associated with areas of biological activity). Liang MP *et al.* (2003) further elaborated this issue in a comparative study of various software packages used for identification of ligand binding cavities, taking into account their efficiency and correctness of results **[20]**. A similar comparative analysis of several packages WebFEATURE **[21]**, SuMo **[22]**, ConSurf **[23]**, CASTp **[24, 25]**, PASS **[26]** and QSiteFinder **[27]** using methods based on the "fuzzy oil drop" model, can be found elsewhere **[20]**. The 3-D Gauss function introduced to represent the structure of hydrophobic core is able to describe only the static form of protein molecule. However the dynamics of protein structure plays substantial and critical role in respect to its biological function. This is why the dynamic form of the 3-D Gauss function of different shapes even being far in respect to the regular ordered form of 3-D Gauss function is planned to be applied for simulation of dynamic forms of protein. The structural dynamics of a protein molecule seems to be the effect of external force field deformation. The source of this deformation is the influence of molecules present in close neighbourhood of that protein molecule. Simulation of structural changes of protein molecule as the effect of external force field deformation is under consideration.

Acknowledgements:

The work was financially supported by Jagiellonian University – Medical College – project K/ZDS/001531. We would also like to thank Piotr Nowakowski from ACC CYFRONET AGH for proofreading the document.

References:

[1] Kauzmann W. *Adv Protein Chem.* 1959 **14**: 1 [PMID: 14404936]
[2] Konieczny L *et al. In Silico Biol.* 2006 **6**: 0002
[3] Max KE *et al. J Mol Biol.* 2007 **369**: 1087 [PMID: 17481655]
[4] Levitt M. *J Mol Biol.* 1976 **104**: 59 [PMID: 957439]
[5] Brylinski M *et al. Int J Bioinform Res Appl*. 2007 **3**: 234 [PMID: 18048191]
[6] Kullback S & Leibler RA. *Annals of Mathematical Statistics.* 1951 **22**: 79
[7] Van Der Spoel D *et al. J Comput Chem.* 2005 **26**: 1701 [PMID: 16211538]
[8] Van Der Spoel D *et al*. Gromacs User Manual version 3.3.
[9] Berendsen HJC. *Comput Phys Commun.* 1995 **91**: 43
[10] Lindahl E *et al. J Mol Model.* 2001 **7**: 306
[11] Berendsen HJ *et al. Ann N Y Acad Sci*. 1986 **482**: 269 [PMID: 3471111]
[12] Marchewka D & Roterman I. ThinkMind BIOTECHNO 2011, The Third International Conference on Bioinformatics, Biocomputational Systems and Biotechnologies, 110-115.
[13] Roterman I *et al. J Theor Biol*. 2011 **283**: 60
[14] Brylinski M *et al. Biochimie* 2006 **88**: 1229 [PMID: 16647798]
[15] Brylinski M *et al. J Biomol Struct Dyn.* 2006 **23**: 519 [PMID: 16494501]
[16] Brylinski M *et al. Comp Biol Chem.* 2006 **30**: 255 [PMID: 16798094]
[17] Zobnina V & Roterman I. *Proteins* 2009 **77**: 378 [PMID: 19455711]
[18] Banach M *et al. J Mol Model.* 2011 [PMID: 21523554]
[19] Brylinski M *et al. PLoS Comput Biol.* 2007 **3**: e94 [PMID: 17530916]
[20] Prymula K *et al. J Comput Aided Mol Des.* 2011 **25**: 117 [PMID: 21104192]
[21] Liang MP *et al. Nucleic Acids Res.* 2003 **31**: 3324 [PMID: 12824318]
[22] Jambon M *et al. Proteins* 2003 **52**: 137 [PMID: 12833538]
[23] Landau M *et al. Nucleic Acids Res.* 2005 **33**: W299 [PMID: 15980475]
[24] Binkowski A *et al. Nucleic Acids Res.* 2003 **31**: 3352 [PMID: 12824325]
[25] Dundas J *et al. Nucleic Acids Res*. 2006 **34**: W116 [PMID: 16844972]
[26] Brady GP & Stouten PF. *J Comput Aided Mol Des*. 2000 **14**: 383 [PMID: 10815774]
[27] Laurie AT & Jackson RM. *Bioinformatics* 2005 **21**: 1908 [PMID: 15701681]

Edited by P Kangueane
Citation: Marchewka *et al.* Bioinformation 6(8): 300-302 (2011)

BIOINFORMATION open access

Supplementary material:

Equation 1:
Gauss function:

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

where $\bar{x}, \bar{y}, \bar{z}$ are the coordinates of the geometric center of the molecule (usually located at the origin of the coordinate system). The size of the molecule is expressed by the triplet σ_x, σ_y, σ_z, which is calculated for each molecule individually, assuming that the longest possible distance between effective atoms within the molecule coincides with an axis of the coordinate system. σ values are calculated as 1/3 of the longest distance between effective atom and the origin of coordinate system (along each axis).

Equation 2:
Levitt's function:

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

where *N* expresses the number of amino acids in the protein body (number of grid points), $\widetilde{H}_i^r$ expresses the hydrophobicity of the *i-th* residue according to the accepted hydrophobicity scale (Brylinski scale as described in **[5]**), r_{ij} expresses the distance between the *i-th* and *j-th* interacting residues, and *c* expresses the cutoff distance which, according to the original paper **[4]**, is assumed to be 9Å. The values of $\widetilde{Ho}_j$ are standardized through division by the $\widetilde{Ho}_{sum}$ coefficient, which represents the aggregate sum of all hydrophobicity values assigned to grid points.
$\widetilde{Ho}_j$ and $\widetilde{Ht}_j$ values are calculated for identical points within each amino acid, corresponding to the position of their effective atoms.

Equation 3:

$$O/T = \sum_{i=1}^{N} O_i * \log_2(O_i / T_i) \quad [3]$$

where: O/T– distance entropy, O_i – probability of occurrence of a particular event in the observed distribution, T_i – corresponding probability in the reference distribution. The index *i* corresponds to a specific amino acid, while *N* denotes the total number of amino acids in the polypeptide chain. The probability values O_i and T_i shall satisfy the condition expressed as sum of all O_i and T_i values shall be equal to 1.

Table 1: O/T and O/R values calculated for hydrophobic, electrostatic and van der Waals interactions

PDB ID	HYDROPHOBICITY		ELECTROSTATIC		vdW	
	O/T	O/R	O/T	O/R	O/T	O/R
2I5M	**0.139**	**0.253**	0.444	0.215	**0.116**	**0.188**

Values accordant with the presented model are highlighted.