Bioinformation by Biomedical Informatics Publishing Group *open access*

www.bioinformation.net **Current Trends**

Ligation site in proteins recognized *in silico*

Michal Brylinski[1,2], Leszek Konieczny[3], Irena Roterman[1,4,*]

[1]Department of Bioinformatics and Telemedicine, Collegium Medicum – Jagiellonian University, Kopernika 17, 31-501 Krakow, Poland; [2]Faculty of Chemistry, Jagiellonian University, Ingardena 3, 30-060 Krakow, Poland; [3]Institute of Medical Biochemistry, Collegium Medicum – Jagiellonian University, Kopernika 7, 31 034 Krakow, Poland; [4]Faculty of Physics, Jagiellonian University, Reymonta 4, 30-060 Krakow, Poland

E-mail contacts: myroterm@cyf-kr.edu.pl, mybrylin@cyf-kr.edu.pl, mbkoniec@cyf-kr.edu.pl

received March 27, 2006; accepted April 5, 2006; published online April 11, 2006

Abstract:

Recognition of a ligation site in a protein molecule is important for identifying its biological activity. The model for in silico recognition of ligation sites in proteins is presented. The idealized hydrophobic core stabilizing protein structure is represented by a three-dimensional Gaussian function. The experimentally observed distribution of hydrophobicity compared with the theoretical distribution reveals differences. The area of high differences indicates the ligation site.

Availability: http://bioinformatics.cm-uj.krakow.pl/activesite

Keywords: hydrophobicity; active site; function recognition; protein structure

Background:

The classic model of an oil drop representing the hydrophobic core in proteins given by Kauzmann **[1]** was intended to visualize the importance of hydrophobic interactions responsible for forming and stabilizing the protein tertiary structure **[2, 3, 4]**. The hydrophilic surface with the hydrophobic center of the molecule is generally accepted **[5, 6]** as the model according to which the amino acid sequence partitions a protein into its inside and outside **[7]**.

The model oriented on localization of the area responsible for ligand binding, based on characteristics of spatial distribution of hydrophobicity which changes from protein interior (maximal hydrophobicity) to exterior (close to zero level of hydrophobicity), can be represented by a three-dimensional Gaussian function **[8, 9, 10].** The simple comparison of theoretical (Gaussian function) and empirical spatial distributions of hydrophobicity in protein allows identification of the areas of high discrepancy, which, as observed in crystal forms of protein-ligand complexes, can be recognized as ligation sites in proteins.

Methodology:

Data: Complexes selected for analysis presented in this paper are: cAMP-dependent protein kinase (PDB ID: 1CDK), cyclin-dependent protein kinase 2 (PDB ID: 1E1V), proto-oncogene tyrosine-protein kinase ABL (PDB ID: 1IEP), S-lectin (PDB ID: 1SLT).

Grid system: The grid system (with constant step size) is constructed for the protein molecule localized with its geometrical center in the origin of the coordinate system $(0{,}0{,}0)$ and oriented as follows: longest inter-effective atoms (side chains represented by the geometrical centers) distance along the X-axis and longest distance between projections (on YZ plane) of effective atoms along the Y-axis. The size of the ellipsoid can be calculated by taking the maximum and minimum values of the X, Y and Z coordinates found in the molecule, oriented as above.

Theoretical hydrophobicity distribution: The theoretical hydrophobicity value for each grid point can be calculated according to a three-dimensional Gaussian function:

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

where: $\widetilde{Ht}_j$ denotes the hydrophobicity for *j*-th grid point (x_j, y_j, z_j), the $(\bar{x}, \bar{y}, \bar{z})$ - the origin of coordinate system $(0{,}0{,}0)$ and $\sigma_x, \sigma_y, \sigma_z$ - the ellipsoid size (⅓ of the maximum length along each axis, respectively). The coefficient $\widetilde{Ht}_{sum}$ (sum of hydrophobicity values attributed to all grid points) makes the $\widetilde{Ht}_j$ standardized (the sum of $\widetilde{Ht}_j$ over all grid pints equal to 1.0).

Empirical hydrophobicity distribution: The empirical hydrophobicity distribution can be calculated using the original function introduced by Levitt **[11]:**

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

where $\widetilde{Ho}_j$ - the empirical hydrophobicity attributed to *j*-th grid point being the result of hydrophobic interaction of side chains of individual $\widetilde{H}_i^r$ hydrophobicity. $\widetilde{Ho}_{sum}$ - sum of all grid points hydrophobicity, which makes the distribution of empirical hydrophobicity standardized. The r_{ij} is the distance between *i*-th effective atoms and *j*-th grid point characterized by zero hydrophobicity. Each grid point collects the observed hydrophobicity $\widetilde{Ho}_j$ from effective atoms localized closer than 9Å (cut-off distance for hydrophobic interaction according to Levitt **[11]**). More details concerning the presented model can be found in recently published papers **[8, 9, 10].**

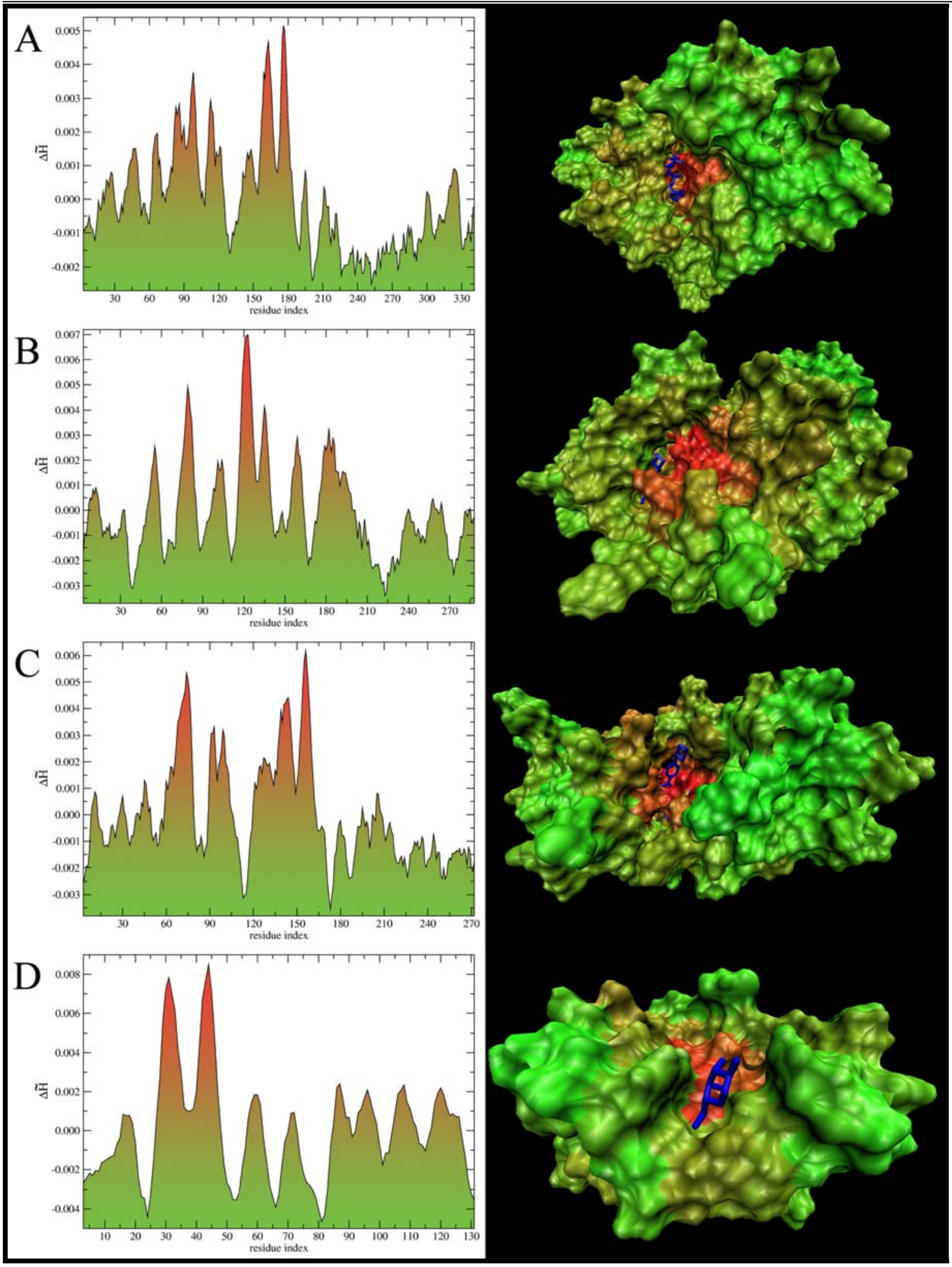

Figure 1: One-dimensional profiles of $\Delta\widetilde{H}$ per amino acid (color scale) (left column) and three-dimensional distribution of $\Delta\widetilde{H}$ on protein surface (right column): A – AMP-dependent protein kinase complexed with 5'-adenyly-imido-triphosphate, B – cyclin-dependent protein kinase 2 complexed with 6-O-cyclohexylmethyl guanine, C – proto-oncogene tyrosine-protein kinase ABL complexed with STI-571, D – S-lectin complexed with D-galactose. The ligands (dark blue thick line) are localized at their binding sites according to crystal structure.

Prediction results:

Theoretical versus empirical hydrophobicity distribution:

Since theoretical (Equation 1) and empirical (Equation 2) hydrophobicity distributions are standardized, the hydrophobicity values attributed to each grid point can be compared by a simple subtraction:

$$\Delta\widetilde{H}_i = \widetilde{H}t_i - \widetilde{H}o_i \quad \text{(Equation 3)}$$

The color scale introduced to express the magnitude of difference $\Delta\widetilde{H}$ in a particular protein (Figure 1) area enables the visualization of the localization of these discrepancies in the protein molecule. The profile of $\Delta\widetilde{H}_i$ along the polypeptide chain (also in color scale) reveals the fragments of polypeptide of high difference between idealized and empirical hydrophobicity density. The same color scale applied to a three-dimensional representation of protein molecule allows for the localization of the ligation site in the protein molecule. The results of analysis of selected protein molecules are shown in Figure 1.

Conclusion:

The many proteins of unknown biological function, identified on the basis of genome analysis, await a unified automated method for determining their biological activity **[12].** The next step is to develop methods able to predict a protein's function from an examination of its structure. Some of the techniques used to identify functionally important residues from the sequence or structure are based on searching for homologues of proteins of known function **[13, 14].** However, homologues need not have related activity, particularly when the sequence identity is below 25% **[15].** The model presented in this paper is oriented on localizing the area responsible for ligand binding, based on the characteristics of the spatial distribution of hydrophobicity in a protein molecule. It is generally accepted that the core region is not well described by a spheroid of buried residues surrounded by surface residues due to hydrophobic channels that permeate the molecule **[16, 17].** This being so, we should be able to identify regions with high deviation versus the ideal model by making a simple comparison of the theoretical (idealized according to the Gaussian function) and empirical spatial distribution of hydrophobicity in a protein. The regions recognized by high hydrophobicity density differences seem to reveal functionally important sites in proteins.

Acknowledgement:

Many thanks to Prof. Marek Pawlikowski (Faculty of Chemistry, Jagiellonian University) for fruitful discussions. This research was supported by the Polish State Committee for Scientific Research (KBN) grant 3 T11F 003 28 and Collegium Medicum grants 501/P/133/L and WŁ/222/P/L.

References:

[1] W. Kauzmann, *Adv Protein Chem*, **14**, 1-63, (1959). [PMID: 14404936]

[2] M.H. Klapper, *Biochim Biophys Acta*, **229**, 557-566, (1971). [PMID: 5555208]

[3] I.M. Klotz, *Arch Biochem Biophys*, **138**, 704-706, (1970). [PMID: 4988452]

[4] H. Meirovitch & H.A. Scheraga, *Macromolecules*, **13**, 1398-1405, (1980).

[5] J. Kyte & R.F. Doolittle, *J Mol Biol*, **157**, 105-132, (1982). [PMID: 7108955]

[6] H. Meirovitch & H.A. Scheraga, *Macromolecules*, **14**, 340-345, (1981).

[7] G.D. Rose & S. Roy, *Proc Natl Acad Sci U S A*, **77**, 4643-4647, (1980). [PMID: 6933513]

[8] M. Brylinski et al., *J Biomol Struct Dyn*, **23**, 519-528, (2006). [PMID: 16494501]

[9] M. Brylinski et al., *Biochimie*, in press, (2006).

[10] L. Konieczny et al., *In Silico Biol*, **6**, 0002, (2006).

[11] M. Levitt, *J Mol Biol*, **104**, 59-107, (1976). [PMID: 957439]

[12] S.K. Burley et al., *Nat Genet*, **23**, 151-157, (1999). [PMID: 10508510]

[13] P. Bork et al., *J Mol Biol*, **283**, 707-725, (1998). [PMID: 9790834]

[14] J. Skolnick & J.S. Fetrow, *Trends Biotechnol*, **18**, 34-39, (2000). [PMID: 10631780]

[15] D. Devos & A. Valencia, *Proteins*, **41**, 98-107, (2000). [PMID: 10944397]

[16] G.M. Crippen & I.D. Kuntz, *Int J Pept Protein Res*, **12**, 47-56, (1978). [PMID: 681085]

[17] I.D. Kuntz & G.M. Crippen, *Int J Pept Protein Res*, **13**, 223-228, (1979). [PMID: 429098]

Citation: Brylinski et al., Bioinformation 1(4): 127-129 (2006)
Edited by P. Kangueane