Internal Force Filed in Proteins

Marchewka Damian
Department of Bioinformatics and Telemedicine
Medical College Jagiellonian University
Lazarza 16, 31-530 Krakow, Poland
damian.a.marchewka@gmail.com

Roterman Irena*
Department of Bioinformatics and Telemedicine
Medical College Jagiellonian University
Lazarza 16, 31-530 Krakow, Poland
corresponding Author: myroterm@cyf-kr.edu.pl

***Abstract*— The proteins representing the structures of ordered form in respect to their tertiary structure are discussed. The hydrophobic core of 3-D Gauss function ("fuzzy oil drop" model) appeared to be present in some proteins. The proteins of the structure with ordered form of vdW and/or electrostatic internal interaction in protein body are discussed in the paper. The vdW interaction distribution in proteins appeared to be accordant with assumed 3-D Gauss function while the electrostatic represents the distribution of random form. This characteristics allows interpretation of the tertiary structure as the effect of external/internal force field influence expressed by 3-D Gauss function in respect to hydrophobic and/or vdW interaction. It is postulated that the additional introduction of 3-D Gauss function representing the influence of external environment (in contrast to internal force field of protein body) in the simulation of protein folding process in silico may simplify the optimization procedure leading to the appropriate order on the level of tertiary structure of protein molecule and directing the hydrophobic residues toward the center of the protein body with the exposure of hydrophilic residues on the surface. The procedure minimize the differences between internal interactions and the idealized one expressed by external force field.**

Internal force field; External force field; Interactions in protein; 3-D Gauss function; Information theory

I. INTRODUCTION

The procedure of protein structure prediction is based on the search for polypeptide structure of low internal energy expressed mostly by side chain - side chain interaction of electrostatic, vdW forms. Thus the optimization procedure (minimization of the energy) is the most important element of the procedure oriented on the protein structure prediction especially in ab initio (new fold – according to CASP nomenclature – Critical Assessment of Protein Structure Prediction) as well as homology search based (comparative modeling - according to CASP nomenclature) computational techniques. The CASP initiative organized every second year is the event assumed to monitor the progress in protein structure prediction [1].

The groups of proteins representing highly ordered structure in respect to the hydrophobic density distribution have been found: downhill proteins [2], antifreeze proteins [3] and some proteins acting in form of homodimers [4]. The high accordance with the 3-D Gauss function with the empirically observed hydrophobicity density distribution [5] in protein body suggested the search for possible ordered force field of other character like electrostatic, dipole-interaction oriented etc.

This is why the analysis of the distribution of the energy components like: electrostatic interaction and vdW interaction additionally to hydrophobic interaction in protein body was undertaken. The question was, whether other than hydrophobic interaction represents the ordered character on the tertiary structure level.

The search for the order system on the level of tertiary structure of proteins is presented in this work based on earlier observed organization of hydrophobic core organization accordant with 3-D Gauss function.

II. MATERIAL AND METHODS

A. Data

The proteins representing different status of hydrophobic core organization were selected for analysis (Tab. I). The selection of proteins was done according to the analysis of down-hill proteins presented in [2].

TABLE I. THE LIST OF PROTEINS UNDER CONSIDERATION. THE PDB ID, LENGTH OF POLYPEPTIDE CHAIN, SOURCE ORGANISM, SECONDARY STRUCTURE DESCRIPTION, BIOLOGICAL FUNCTION AND REFERENCES ARE GIVEN

Protein	N	Source	Structure description	Biological function	Reference
1HZC	66	bacteria	β-barrel	Cold shock protein	[6]
1BDC	60	bacteria	Mainly helical	Immunoglobulin binding domain	[7]
1VII	36	chicken	Mainly helical	Villin subdomain Actin binding	[8]
2I5M	66	bacteria	Mainly β-structural	Cold shock protein	[9]
1CSP	67	bacteria	Mainly β-structural	Cold shock protein	[10]
1RIJ	23	De novo design	Mainly helical	De novo design	[11]

B. Energy optimization using Gromacs

The Gromacs program was applied to run the energy optimization procedure to relax the crystal structure.
All EMs (energy minimization) have been performed with Gromacs software package v4.0.3 and Gromos96 43a1 force field [12-16]. The coordinates for starting structures have been taken from the Protein Data Bank. In first step, all EMs have been compared an in vacuo model to a solvated model.

 ISBN: 978-1-61208-137-3

Default protonation states and hydrogen positions were generated by pdb2gmx utility of the Gromacs package. The energy optimization procedure was performed in water solvent. SPC water model was used [16]. The total charge of the molecule was null.

The parameters for energy minimization procedure were as follows:

Maximum number of iterations - 100 steps; The minimization was converged when the max force was smaller than 1000.0 kJ mol-1 nm-1; Initial step size - 0,01 nm; Method to determine neighbour list – Grid; Treatment of long range electrostatic interactions – cut-off; Long range electrostatic cut-off - 1.0 nm; Long range van der Waals cut-off – 1.0nm; Cut-off distance for short-range neighbour list - 1.0nm; Constraint algorithm used to restrain bond lengths – none;

Frequency to update the neighbour list -10 steps;

The individual interactions of particular residues with the rest of the protein molecule was performed using the make_ndx procedure defining the "group" under consideration and g_energy program in order to extract data from output energy files.

Each residue was taken as one group while the rest of protein molecule was defined as second group. The energy calculation was performed for each amino acid in this system. The set of individual interactions was standardised to the unit making the interaction distribution unified allowing the comparison with other distributions (theoretical and random one).

C. *"Fuzzy oil drop" model*

The assumption of this model is the accordance of hydrophobic (and possible other) interactions in protein with the idealized one expressed by 3-D Gauss function [5]. The procedure allowing generation of this type of force field is shown below.

1) The theoretical hydrophobicity density distribution: The geometric center of the protein molecule is localized in the origin of coordinate system. The longest distance between two effective atoms (averaged position of atoms belonging to side chain) determines the orientation of the X-axis. The longest distance between two projections (on the YZ-plane) of effective atoms determines the orientation of the Y-axis. The longest distance between elements along each axis in coordinate system is expressed by 3σ. The hydrophobicity density in each position of effective atom can be calculated as follows:

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

where $\bar{x},\bar{y},\bar{z}$ are the coordinates of the geometric center of the molecule (usually located in the origin of the coordinate system). This is why these values can be considered equal to zero. The size of the molecule is expressed by the triple σ_x, σ_y, σ_z, which is calculated for each molecule individually provided that the orientation of the molecule with the longest possible inter-effective atoms distance is determined according to the appropriate coordinate system axis. The σ values are calculated as the 1/3 of the longest distance between two effective atoms calculated along each axis. The value of the Gauss function at any point of protein body is treated as the idealized hydrophobic density defining the hydrophobic core.

2) Observed distribution: On the other hand, the empirical hydrophobicity distribution is calculated according to the function presented by Levitt [17]:

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

where N expresses the number of amino acids in the protein (number of grid points), $\widetilde{H}_i^r$ expresses the hydrophobicity of the i-th residue according to the accepted hydrophobicity scale (the scale presented in [18]) was applied in this work, 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 [17] is assumed to be 9 Å. The values of $\widetilde{Ho}_j$ are standardized by dividing them by the coefficient $\widetilde{Ho}_{sum}$, which is the sum of all hydrophobicities attributed to grid points.

3) Electrostatic and vdW interactions: The individual interactions of particular residues with the rest of the protein molecule was performed using the make_ndx procedure defining the "group" under consideration. Each residue was taken as one group while the rest of protein molecule was defined as second group. The energy calculation was performed for each amino acid in this system. The set of individual interactions was standardised to the unit making the interaction distribution unified allowing the comparison with other distributions (theoretical and random one).

4) The analysis of distributions: To evaluate quantitatively the accordance between the idealized and empirically observed distribution of the density of selected parameter (interaction), divergence entropy (also known as Kullback-Leibler entropy [19] was calculated:

$$D_{KL}(P \parallel Q) = \sum_i P(i) \log_2 \frac{P(i)}{Q(i)}$$

where $D_{KL}(P \parallel Q)$ denotes the distance entropy (also called deficiency/divergence entropy), which is a measure of the distance between P(i) and Q(i) distributions (probabilities), where Q(i) plays the role of target distribution.

The values of Q(i) were taken according to the 3-G values for the ellipsoid of particular protein. This target (reference)

function was commonly used for all types of interaction (electrostatic and vdW) under consideration. The values of P(i) expressed the density of particular type of interaction calculated in relation to the sum of interaction of each residue with the entire protein molecule. The values expressing particular type of interaction were standardized to make the sum of all values equal to 1 after the unified rescaling of negative and positive values.

Since the entropy values can be interpreted only in the relative scale the comparison of observed distribution with the random one was performed. The protein of the distance between observed distribution (O) and theoretical one (T) expressed as O/T lower in relation to the distance between O and random distribution (R) (O/R) was treated as the protein of distribution accordant with expected one.

III. RESULTS

Hydrophobicity distribution: the 3-D Gauss function was taken as the target distribution for hydrophobic force field. The Kulback Leibler entropy values are given in Tab. II. All the proteins classified as downhill proteins appeared to represent the structure accordant with the idealized hydrophobic core.

A. Density distribution in proteins under consideration

The distribution profile of each component of the force field for selected proteins: 1HZC and1BDC are shown in Fig. 1 and Fig.2 respectively.

The profiles visualize the range of similarity/discrepancy between expected and observed distribution. The high accordance between idealized hydrophobic distribution and observed one can be seen except 1HZC. The high accordance between random and observed electrostatic density can be seen in all cases.

B. Summary of the internal interaction in proteins

The summary characterizing the structure of internal force field is given in Tab.II.

TABLE II. THE O/T AND O/R ENTROPY VALUES CALCULATED FOR INDIVIDUAL TYPES OF INTERACTIONS (HYDROPHOBIC, ELECTROSTATIC AND VDW) TAKING THE IDEALIZED 3-D GAUSS FUNCTION (T) AS THE TARGET AND THE RANDOM DISTRIBUTION (R) TO MAKE POSSIBLE INTERPRETATION OF THE ENTROPY VALUES. THE VALUES FOR STRUCTURES ACCORDANT WITH ASSUMED MODEL ARE GIVEN IN BOLD.

Protein	Hydrophobicity		Electrostatic		vdW	
	O/T	O/R	O/T	O/R	O/T	O/R
1HZC	0.213	0.207	0.323	0.057	0.307	0.252
1BDC	**0.121**	**0.141**	0.348	0.097	**0.149**	**0.186**
1VII	**0.223**	**0.568**	0.336	0.115	0.226	0.101
2I5M	**0.188**	**0.559**	0.444	0.215	**0.116**	**0.188**
1CSP	**0.134**	**0.466**	0.266	0.035	**0.103**	**0.190**
1RIJ	**0.171**	**0.583**	0.244	0.110	**0.079**	**0.087**

The proteins characterized in Tab. II. were selected to represent different status in respect to the order of energy components distribution.

The protein 1HZC (cold shock protein) of the form of compact β-barrel proteins without disulfide bonds and cis-proline residues has been recognized as the molecule of low stability [6]. The absence of ordered form of hydrophobic core as well as absence of any ordered form of internal force field seems to explain the low stability of this molecule (Fig.1.).

The protein 1BDC (immunoglobulin binding domain) represents the mainly helical structural form [7]. According to the analysis presented in this paper its stability may the result of the ordered structure of hydrophobic core as well as ordered vdW internal force field (Fig. 2.).

IV. CONCLUSIONS

The results presented in this paper suggest that the tertiary level organization is expressed by the ordered form of hydrophobic as well as vdW interaction force field. No regularity (random distribution) was found for electrostatic interaction. The local, biological function related charge presence (enzymatic active site) was not taken into account. It was aimed to analyze the non-specific distribution of charges (the electrostatic interaction).

The regularity of the hydrophobicity distribution identified in downhill proteins (as well as in antifreeze proteins [3] and some homodimers [4]) suggests that the folding process is directed by the hydrophobic interaction in the form accordant with the 3-D Gauss function. The introduction of the external force field of 3-D Gauss function during the folding process simulation may facilitate the structure optimization process in silico. The presence of external force field may direct the hydrophobic residues toward the center of the protein body with the exposure of hydropholic residues on the surface [18]. The folding process accordant with high density of vdW interactions in the center of the protein molecule may additionally introduce the expected order of residues in the space. The “fuzzy oil drop” model was proved performing the molecular dynamics simulation of trans-membrane protein. The simulation performed in the presence of external force field in form of 3-D Gauss function for hydrophobic interaction revealed high accordance of results with those received using the traditional simulation performed in the presence of membrane and water molecules [20]. The regression function comparing the results received using the explicite water molecules and “fuzzy oil drop” model was of the form y=1*x. The time consumption for “fuzzy oil drop” model was significantly lower in comparison with traditional molecular dynamics simulation in all-atoms form [20].

The proteins representing different secondary structures and different biological function appeared to represent also different accordance with the assumed model expecting the density distribution of particular type of interaction accordant with 3-D Gauss function. The ordered distribution of particular type of interaction seems to generate the ordered internal force field probably responsible for tertiary stabilization. The differences between proteins of different

 ISBN: 978-1-61208-137-3

structural (secondary structure) characteristics of the internal force field suggest different mechanism of the structure generation. The absence of the accordance of the assumed order in respect to electrostatic interactions suggests low influence of external force field of electrostatic character. The discordance between the expected (3-D Gauss distribution) and the observed one was recognized to appear due to the presence of ligand (including also the protein-protein complexation interaction area) [5]. It may suggests the aim-oriented local disorder related to specific biological function [18].

ACKNOWLEDGMENT

The work was financially supported by Jagiellonian University – Medical College grant K/ZDS/001531.

REFERENCES

[1] CASP http://predictioncenter.org/

[2] Roterman I., Konieczny L., Jurkowski W., Prymula K. and Banach M. (2011) Two-intermediate model to characterise the structure of fast-folding proteins – submitted

[3] Banach M., Prymula K., Jurkowski W., Konieczny L. and Roterman I. (2011) Fuzzy oil drop model to interpret the structure of antifreeze proteins and their mutants. J. Mol. Mod. – in press

[4] Banach M., Konieczny L. and Roterman I. (2011) "Fuzzy oil drop" model to identify the complexation area in protein homodimers. – submitted

[5] Konieczny L., Brylinski M. and Roterman I. (2006) Gauss-function-based model of hydrophobicity density in proteins In Silico Biol 6, 0002.

[6] Delbruck H., Mueller U., Perl D., Schmid F.X. and Heinemann U. (2001) Crystal structures of mutant forms of the Bacillus caldolyticus cold shock protein differing in thermal stability. J.Mol.Biol. 313: 359-369

[7] Gouda H., Torigoe H., Saito A., Sato M., Arata Y. and Shimada I. (1992) Three-dimensional solution structure of the B domain of staphylococcal protein A: comparisons of the solution and crystal structures. Biochemistry. 31(40): 9665-72

[8] McKnight C.J., Matsudaira P.T. and Kim P.S. (1997) NMR structure of the 35-residue villin headpiece subdomain. Nat.Struct.Biol. 4: 180-184

[9] Max K.E., Wunderlich M., Roske Y., Schmid F.X. and Heinemann U. (2007) Optimized variants of the cold shock protein from in vitro selection: structural basis of their high thermostability. J.Mol.Biol. 369: 1087-1097

[10] Schindelin H., Marahiel M.A. and Heinemann U. (1993) Universal nucleic acid-binding domain revealed by crystal structure of the B. subtilis major cold-shock protein. Nature 364: 164-168

[11] Liu Y., Liu Z., Androphy E., Chen J. and Baleja J.D. (2004) Design and characterization of helical peptides that inhibit the E6 protein of papillomavirus. Biochemistry 43: 7421-7431

[12] Berendsen H.J., van der Spoel D. and van Drunen R. (1995) GROMACS: a message-passing parallel molecular dynamics implementation. Comput Phys Commun 91: 43-56

[13] Berendsen H.J.C., Postma J.P.M., van Gunsteren W.F. and Hermans J. (1981) Interaction models for water in relation to protein hydration. In: Intermolecular Forces. Pullman, B. ed. . D. Reidel Publishing Company Dordrecht pp 331–342

[14] van der Spoel D., Lindahl E., Hess B., Groenhof G., Mark A.E. and Berendsen H.J. (2005) GROMACS: fast, flexible, and free. J Comput Chem 26: 1701-1718

[15] Lindahl E., Hess B. and van der Spoel D. (2001) GROMACS 3.0: a package for molecular simulation and trajectory analysis. J Mol Model 7: 306-317

[16] van der Spoel D., Lindahl E., Hess B., van Buuren A.R., Apol E., Meulenhoff P.J., Tieleman D.P., Sijbers A.L., Feenstra K.A., van Drunen R. and Berendsen H.J. Gromacs User Manual version 3.3. www.gromacs.org; 2005b

[17] Levitt M. (1976) A simplified representation of protein conformations for rapid simulation of protein folding. J Mol Biol 104: 59-107

[18] Brylinski M., Konieczny L. and Roterman I. (2007) Is the protein folding an aim-oriented process? Human haemoglobin as example. Int J Bioinform Res Appl. 3: 234-260

[19] Nalewajski R. F. (2006) Information theory of molecular systems. Amsterdam: Elsevier

[20] Zobnina V. and Roterman I. (2009) Application of the fuzzy-oil-drop model to membrane protein simulation Proteins 77: 378-394

 ISBN: 978-1-61208-137-3

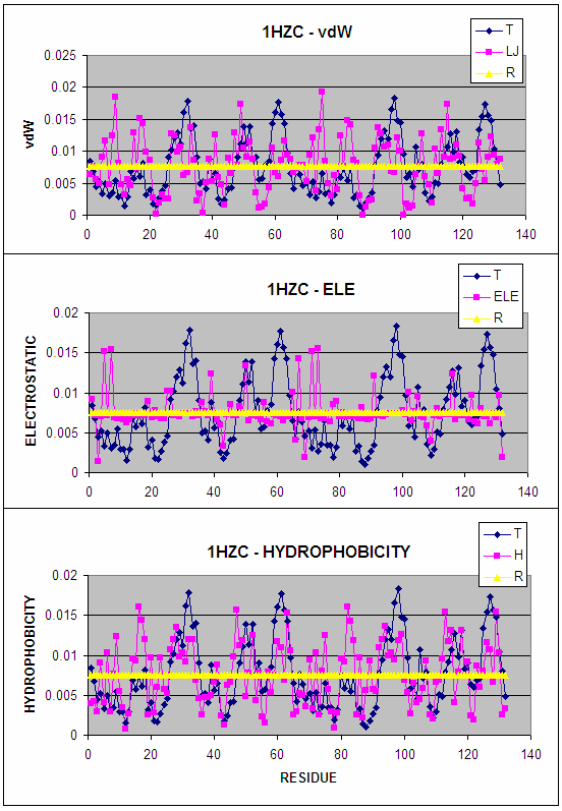

Figure 1. The profiles representing the density distribution of electrostatic, vdW and hydrophobic interaction in protein body in 1HZC. The lack of accordance can be seen in all profiles.

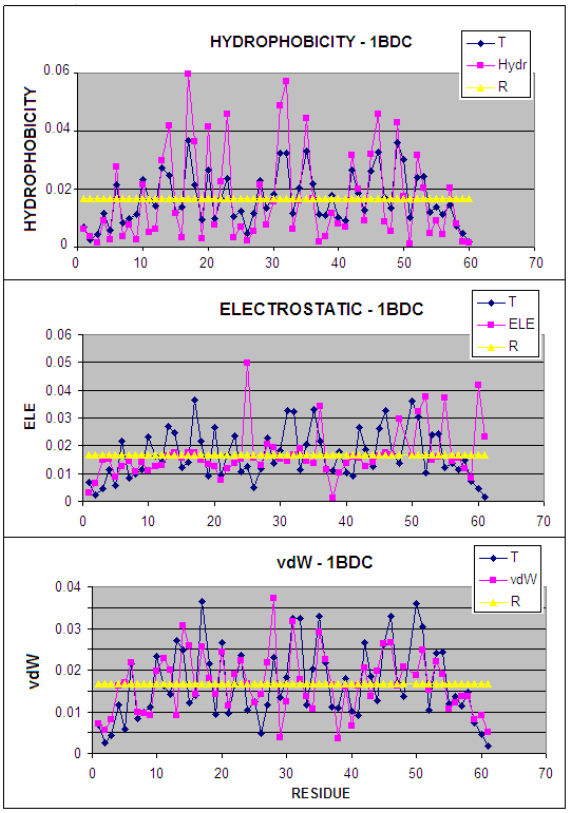

Figure 2. The profiles representing the density distribution of electrostatic, vdW and hydrophobic interaction in protein body in 1BDC. The lack of accordance can be seen in profile of electrostatic interaction.

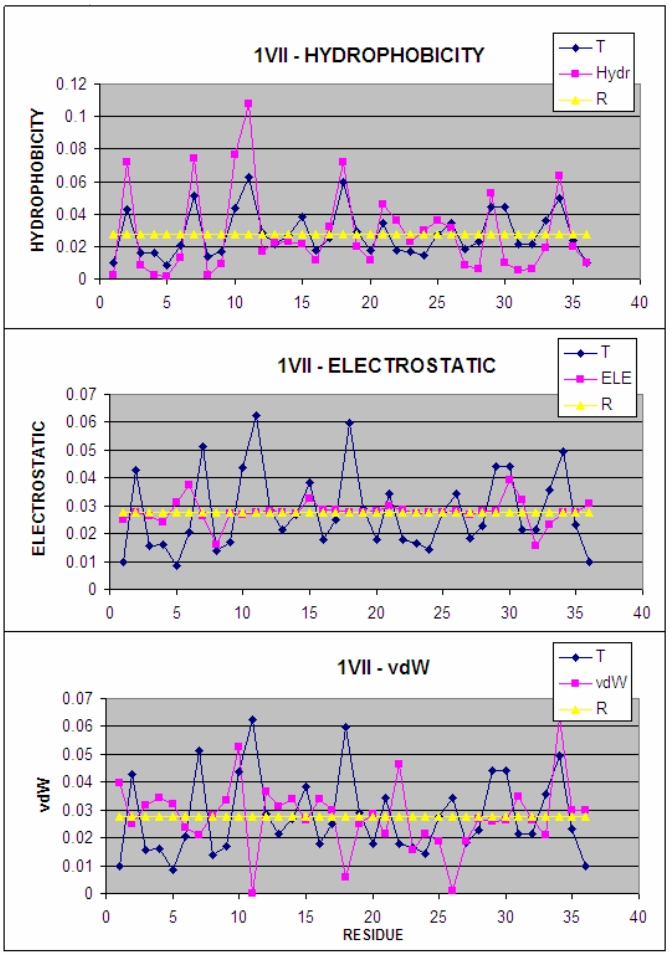

Figure 3. The profiles representing the density distribution of electrostatic, vdW and hydrophobic interaction in protein body in 1RIJ. The high accordance between idealized and observed distribution can be seen for hydrophobic and vdW interactions.

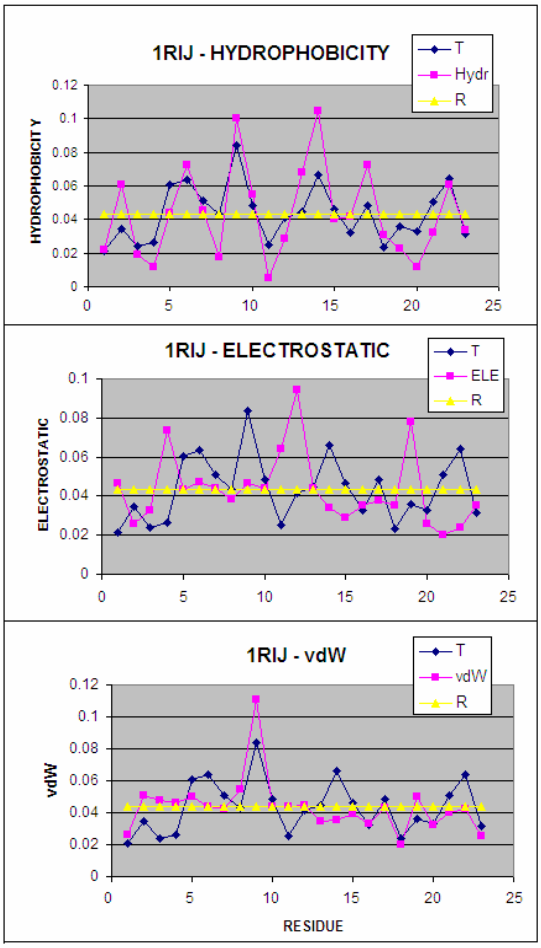

Figure 4. The profiles representing the density distribution of electrostatic, vdW and hydrophobic interaction in protein body in 1VII. The lack of accordance can be seen in profile of electrostatic interaction.

 ISBN: 978-1-61208-137-3