

Tight-binding ‘dihedral orbitals’ approach to electronic communicability in macromolecular chains

Ernesto Estrada ^{a,*}, Naomichi Hatano ^b

^a *Complex Systems Research Group, X-rays Unit, RIAIDT, Edificio CACTUS, University of Santiago de Compostela, 15706 Santiago de Compostela, Spain*

^b *Institute of Industrial Science, University of Tokyo, Komaba 4-6-1, Meguro, Tokyo, Japan*

Received 3 September 2007; in final form 5 October 2007

Available online 12 October 2007

Abstract

An electronic orbital of a dihedral angle of a molecular chain is introduced. A tight-binding Hamiltonian on the basis of the dihedral orbitals is defined. This yields the Green’s function between two dihedral angles of the chain. It is revealed that the Green’s function, which we refer to as the electronic communicability, is useful in differentiating protein molecules of different types of conformation and secondary structure.

© 2007 Elsevier B.V. All rights reserved.

1. Introduction

The main challenge for the current post-genomic research consists of starting from the gene sequence, producing the protein, then determining its three-dimensional structure and finally extracting useful biological information about the biological role of the protein in the organism [1]. Due to the tremendous amount of structural data existing today, it is necessary to develop and use new theoretical tools to extract the maximum structural information from protein structures.

One of the most important characteristics of the three-dimensional structure of a protein is its degree of folding (DOF). The first attempt to assign a quantitative measure to DOF was carried out by Randić and Krilov [2,3]. Balaban and Rucker [4] introduced ‘protochirons’, and more recently, Liu and Wang [5] extended this approach by including four new kinds of three-steps path conformations for studying DOF of protein chains. One of the present authors (E.E.) proposed to quantify DOF by using graph spectral theory [6–12]. This index of DOF (called the I_3 index) has been very useful in structure–function studies

of proteins as well as in protein secondary structure classification. All these methods are phenomenological approaches based on chemical intuition. However, it is possible to derive the I_3 index on the basis of a tight-binding Hamiltonian based on orbitals centered at the dihedral angles of a linear chain. We can thereby use a Hückel-like approach to calculate an ‘electronic dihedral energy’ of the (protein) chain as well as an ‘electronic dihedral partition function’ [13]. Here we extend this approach by considering the ‘electronic communicability’ between the dihedral orbitals of a macromolecular chain by defining its electronic dihedral Green’s function.

2. The tight-binding ‘dihedral orbitals’ approach

One of the present authors (E.E.) proposed to consider a set of orbitals situated in each dihedral angle of the chain instead of using atomic-centered orbitals as the basis of the molecular wave functions [13]. This approach is similar to the one developed by Lennard-Jones and Hall, who defined ‘equivalent orbitals’ as orbitals centered on each bond between two atoms [14–17]. They used the bond orbitals as a basis set to give molecular orbitals (LCBO–MO) [18].

If we represent the atomic orbitals as the nodes of a chain, the bond orbitals can be represented by means of

* Corresponding author. Fax: +34 981 547 077.

E-mail address: estrada66@yahoo.com (E. Estrada).

the so-called line graph of the chain. The line graph $L(G)$ is the graph in which the bonds of the chain G are represented as the nodes of $L(G)$. Two nodes of $L(G)$ are adjacent if the corresponding bonds in G shares an atom [19]. We can extend this approach to consider the second line graph of the molecular chain in which every node represents a bond angle. This approach is known as the iterated line graph sequence and it has been used in the study of organic molecules and macromolecules [20–22]. The corresponding orbitals located on the plane formed by the bond angle are designated as the plane orbital. Finally, what we consider here is the third line graph $L^3(G)$. The resulting orbitals are ‘dihedral’ orbitals localized at two planes formed by two bond angles.

The molecular orbitals for the dihedral electrons in the atomic chain can be written as

$$|\Psi_n^D\rangle = \sum_{i=1}^N C_n^D(i) |\phi_i^D\rangle, \quad (1)$$

where ϕ_i^D is a dihedral orbital located in the i th dihedral angle of the chain. Eq. (1) represents the basis of the linear combination of dihedral orbitals to give molecular orbitals (LCDO–MO). Then we assume that the ‘electronic dihedral energy’ of a chain can be obtained by solving a dihedral version of the Schrödinger equation:

$$\mathbf{H}^D |\Psi^D\rangle = \varepsilon^D |\Psi^D\rangle, \quad (2)$$

where the superscript **D** is used to designate the dihedral angles. Substituting (1) into (2), we have

$$\sum_i C_n^D(i) \mathbf{H}^D |\phi_i^D\rangle = \varepsilon^D \sum_i C_n^D(i) |\phi_i^D\rangle. \quad (3)$$

Multiplying both sides of this expression from the left by $\langle \phi_j |$ yields the secular equation

$$\sum_i C_{ni} (H_{ji} - \varepsilon S_{ji}) = 0, \quad (4)$$

where $H_{ji} = \langle \phi_j | \mathbf{H} | \phi_i \rangle$, $S_{ji} = \langle \phi_j | \phi_i \rangle$ and we have removed, for the sake of simplicity, the superscript **D** in the dihedral Hamiltonian. We consider that the dihedral orbitals are orthonormal, thus $S_{ij} = \delta_{ij}$ hereafter. The nontrivial solutions of the secular equation (4) are obtained by solving the determinant equation

$$|\mathbf{H} - \varepsilon \mathbf{I}| = 0. \quad (5)$$

We assume that the Coulomb integral H_{ii} of a dihedral orbital ϕ_i depends only on the angle between the two planes forming the dihedral orbital. We set the Coulomb integral in the form $H_{ii} = p - V_i q$, where the effective potential V_i is some function of the i th dihedral angle φ_i of the linear chain. The constant p sets the origin of the energy and the constant q sets the energy scale. The resonance integral H_{ij} between dihedral orbitals ϕ_i and ϕ_j is, for the moment, assumed to be zero, unless i and j are adjacent dihedrals in the chain, in which case we set $H_{ij} = q$. We

will argue an extension in Section 4, but for the moment, we obtain the Hamiltonian for a linear chain having N dihedral angles in the matrix form

$$H = pI - q \begin{pmatrix} V_1 & -1 & 0 & & \\ -1 & V_2 & \ddots & & 0 \\ & -1 & \ddots & & \\ & & \ddots & & \\ 0 & & & & -1 \\ & & & & -1 & V_N \end{pmatrix}. \quad (6)$$

This matrix is a tri-diagonal matrix $\mathbf{H} = [H_{ij}]$, which means that $H_{ij} = 0$ whenever $|i - j| > 1$ [23]. Below, we will use this Hamiltonian in order to obtain the minimum dihedral energy for the most folded conformer.

The orbital energy is determined by the eigenvalues of \mathbf{H}

$$\varepsilon_j = p - q\mu_j, \quad (7)$$

where μ_j is an eigenvalue of the matrix on the r.h.s. of Eq. (6). When all dihedral orbitals are fully occupied, the total electronic ‘dihedral’ energy (dihedral energy for brief) is given by

$$E_{\text{dih}} = \sum_{j=1}^N 2\varepsilon_j = Np - 2q \sum_{j=1}^N \mu_j, \quad (8)$$

where N is the number of dihedral angles in the linear chain and $q < 0$. We consider that the most folded conformation permits the largest overlap between dihedral orbitals, which reduces the dihedral energy to a minimum.

From now on we set $p \equiv 0$ without loss of generality, since p simply sets the origin of the energy scale. This makes the Hamiltonian

$$H = \begin{pmatrix} V_1 & -q & 0 & & \\ -q & V_2 & \ddots & & 0 \\ & -q & \ddots & & \\ & & \ddots & & \\ 0 & & & & -q \\ & & & & -q & V_N \end{pmatrix} \quad (9)$$

and the dihedral energy [13]

$$E_{\text{dih}} = 2|q| \sum_{j=1}^N \mu_j, \quad (10)$$

where we assumed $q < 0$.

3. Effective potential for a dihedral angle

We now determine the functional form of the effective potential V_i [13]. Our approach assumes that this potential depends only on the angle φ_i formed between the two planes determining the dihedral angle. A natural way of selecting this function is to make it the cosine of the dihe-

dral angle. This function satisfies our intuition that when $\varphi_i = 90^\circ$ there is no overlapping between the plane angles forming the dihedral and V_i should vanish. We also consider that for $90^\circ < \varphi_i \leq 180^\circ$ there should be no overlapping between the dihedral angles, which means $V_i = 0$ in this region. This condition is not satisfied by the function $V_i = \cos\varphi_i$, which takes negative values for $90^\circ < \varphi_i \leq 180^\circ$. Thus we select as the effective potential the half-cosine function in the form [13]

$$V_{ii} = \frac{1}{2} [1 + \text{sgn}(\cos \varphi_i)] \cos \varphi_i, \quad (11)$$

where φ_i is the i th dihedral angle of a particular configuration of the molecular chain and $\text{sgn}(x)$ is the sign. This function is equal to cosine of the angle for $0^\circ \leq \varphi_i \leq 90^\circ$ and is zero for $90^\circ < \varphi_i \leq 180^\circ$ as desired for our effective potential. Of course, other more sophisticated potentials may be used without modifying the significance of the current approach.

4. Interaction between non-adjacent dihedral angles

We now introduce an extension of the tight-binding LCDO–MO Hamiltonian (9). Depending on the three-dimensional conformation of the molecule, dihedral angles not adjacent on the linear chain may be geometrically close in the three-dimensional space. In order to take this effect into account, we introduce a coupling between non-adjacent dihedral angles. That is, we introduce a function V_{ij} for the non-diagonal entries of the Hamiltonian.

Here we simply consider that the through-space jump of an electron is dependent on the geometrical (not topological) distance separating the dihedral angles. Hence we assume the following potential [24,25]

$$V_{ij} = qe^{-k(d_{ij}-d_{\min})}, \quad (12)$$

where k is the coupling constant and d_{\min} is the minimal distance between a pair of non-adjacent dihedral angles. Then the LCDO Hamiltonian is modified to

$$H = \begin{pmatrix} V_1 & -q & -qe^{-k(d_{13}-d_{\min})} & \dots & -qe^{-k(d_{1N}-d_{\min})} \\ -q & V_2 & \ddots & \dots & -qe^{-k(d_{2N}-d_{\min})} \\ -qe^{-k(d_{31}-d_{\min})} & -q & \ddots & \dots & \dots \\ \vdots & \vdots & \ddots & \ddots & \dots \\ -qe^{-k(d_{N1}-d_{\min})} & -qe^{-k(d_{N2}-d_{\min})} & \dots & -q & V_N \end{pmatrix}. \quad (13)$$

In order to define the distance between two non-adjacent dihedral angles we consider the distance between the centers of gravity of both angles. In further calculations of the electronic communicability in proteins we put $k = 1$ and $d_{\min} = 0$ for simplicity. These assumptions are

justified by the fact that all protein backbones have the same chemical composition and that the possible minimal separation between dihedrals in such chains is always the same.

5. Green's function of the dihedral chain

Once the electronic Hamiltonian is given as (13), we can define the Green's function [25]. Different approaches based on Green's function to the electron transfer in proteins have been reported [26–28]. The thermal Green's function is defined by

$$G_{ji}(\beta) = \frac{1}{Z(\beta)} \langle j | e^{-\beta H} | i \rangle = \langle j | e^{-\beta(H-F)} | i \rangle, \quad (14)$$

where the partition function is [29]

$$Z(\beta) = \text{Tr} e^{-\beta H} \quad (15)$$

and the free energy is [29]

$$F = -\frac{1}{\beta} \log Z(\beta) \quad (16)$$

at the inverse temperature β . The Green's function describes how well the dihedral angles i and j are electronically connected. Hereafter we refer to the Green's function between two dihedral angles as the communicability between them, motivated by another work of the present authors in which the Green's function is identified as a graph theoretic invariant related to the number of walks linking two nodes in the graph [30].

Suppose that the Hamiltonian (13) has the eigenvalues $\{\varepsilon_v = |q|\mu_v\}$ with the eigenfunctions $\{|\psi_v\rangle\}$. The Green's function (14), or the communicability between the i th and j th dihedral angles is written in the form

$$G_{ji}(\beta) = \frac{1}{Z(\beta)} \sum_{v=1}^N \langle j | \psi_v \rangle \langle \psi_v | i \rangle e^{-\beta|q|\mu_v} \quad (17)$$

with the partition function

$$Z(\beta) = \sum_{v=1}^N e^{-\beta|q|\mu_v}. \quad (18)$$

In the low-temperature limit $\beta \rightarrow \infty$, the Green's function is reduced to

$$G_{ji}(\infty) = \langle j | \psi_1 \rangle \langle \psi_1 | i \rangle, \quad (19)$$

where $v = 1$ denotes the ground state, while in the high-temperature limit $\beta \rightarrow 0$, it is reduced to

$$G_{ji}(0) = \frac{1}{N} \sum_{v=1}^N \langle j | \psi_v \rangle \langle \psi_v | i \rangle. \quad (20)$$

Hereafter we set $|q|\beta \equiv 1$ for simplicity, since $|q|$ specifies an energy scale chosen arbitrarily.

6. Computational results

Here we calculate the communicability function for the human transcriptional elongation factor TFIIS, a small protein of 50 amino acids that contains a Zn(2+)-binding site. The structure of this protein, determined by complete ^1H and ^{15}N NMR [31], is used to build the LCDO Hamiltonian for the backbone chain.

We compared the result for the Hamiltonian (13) with the result for the simpler version of the Hamiltonian, Eq. (9), which ignores the terms $-e^{-k(d_{ij}-d_{\min})}$ in the expression (13). In Fig. 1a and b, we illustrate the contour plots representing the dihedral angles as the x and y -axes and the values of the electronic communicability between the dihedrals as the z -axis. We then fit the data points by using

the weighted least square method [32] implemented in the STATISTICA package [33].

As can be seen in Fig. 1a and b, there are small but observable differences between the two contour plots due to the consideration of the inter-dihedral distances. There is an increase in the communicability between the dihedral angles located at the top-left corner of this plot (red contour) and decrease of the communicability at the center of the plot. This observation indicates that we need to consider the inter-dihedral distances in any further calculation of the electronic communicability in protein chains.

The electronic communicability between non-adjacent bonds can be carried out by both through-space and through-bond interactions. Consequently, the communicability between two dihedral angles separated to each other in the linear chain at a large topological distance can be influenced by their geometrical separation in the space. In order to illustrate this situation, we build the contour plot of the electronic communicability in terms of the amino acid numbers in the protein chain. In Fig. 1c, we illustrate this plot for 1TFI by considering all values of G_{ji} from zero to its maximum (9.490) and in Fig. 1d, we plot only the values in the range $0 \leq G_{ji} \leq 0.01$.

In Fig. 1c, we can see that the largest communicability takes place between pairs of adjacent dihedrals (main diagonal of the plot). We also note that the dihedral angles which are close to each other in the sequence also display significant communicability. The absolutely largest communicability is observed for the terminal amino acids of the chain, only because they form two of the four strand present in this chain [31]. The large communicability between the amino acids in these regions is better observed in Fig. 1d, where the upper-right and bottom-left parts of the contour are displayed in red color. Another region with large communicability in this plot is that formed by amino acids 39–47 and amino acids 5–20. The center of this region, which has the largest communicability, is given by the interaction of amino acids 15 and 43. This is an extremely important region of the human transcriptional elongation factor TFIIS protein as it represents the binding site of the Zn(2+), which is formed by Cys40, Cys43, Cys15 and Cys12 [28]. The other two regions at the center of the contour plot which display large communicability correspond to β -turns [31].

An important part of the conformational arrangement of the dihedral angles is contained in the secondary structure of proteins [34]. Consequently, the electronic communicability should be affected by the differences in the protein secondary structure. In Fig. 2 we illustrate the contour plots of six proteins having differences in their secondary structure according to the CATH protein structure classification [35]. As can be seen in the plots a and b (mainly α protein) the largest communicability takes place for residues which have dihedral angles centered about 60° , which corresponds to the ψ and ϕ dihedral angles of left-handed α helices. In Fig. 2c and d (α - β protein) the largest communicability appears for the amino acids forming left-handed α -helices as well as for the regions around 240° ,

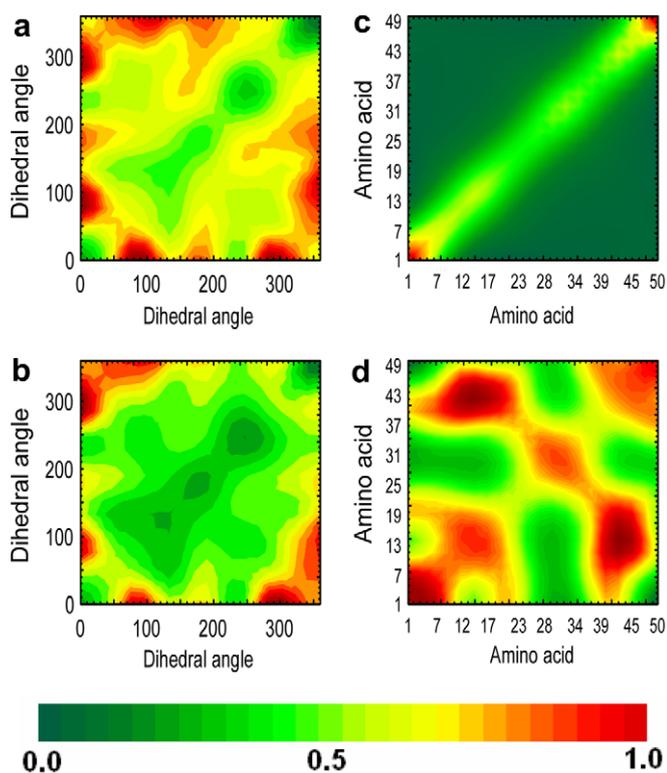


Fig. 1. Plot of the electronic communicability between dihedral angles for the human transcriptional elongation factor TFIIS (1TFI) without considering the distances in the Hamiltonian (a) and considering them (b). Electronic communicability between amino acids for the same protein considering the whole range of communicability values (c) and by considering only those values between 0 and 0.01 (d). In all cases the values of communicability are normalized.

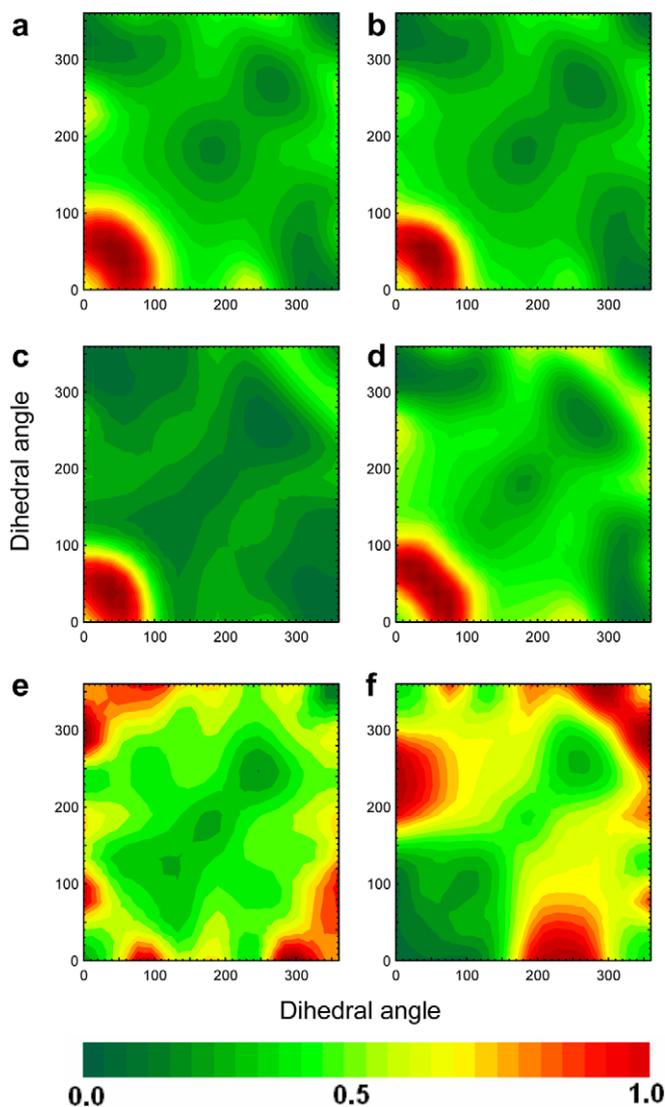


Fig. 2. Contour plot of six proteins having differences in their secondary structures. Plots (a) and (b) correspond to proteins with mainly α structures (PDB codes: 1BGC and 1RIB). Plots (c) and (d) correspond to proteins with α - β structures (PDB codes: 1RKR and 1HGE chain B). Plots (e) and (f) correspond to mainly β structures (PDB codes: 1TFI and 1SGH). In all cases the values of communicability are normalized.

which correspond to β -strands. This change is more evident in the plots e and f, (mainly β proteins) where the largest communicability appears in the regions corresponding to β -strands and turns.

7. Summary

We defined the Green's function of a linear chain on the basis of dihedral orbitals. It indicates the mobility, or the 'communicability' of an electron between the dihedral orbitals. We demonstrated the application of the above idea to real molecules. The electronic communicability was useful in differentiating molecules of different types of conformation and secondary structure. The current approach is not

limited to the consideration of the backbone dihedrals only. It can be straightforwardly extended for considering all dihedral angles in a molecule. In such a case the number of dihedrals increases dramatically but the interpretation of electronic communicability function remains unaltered. These results can also be extended to other macromolecular chains, such as DNA.

Acknowledgement

E.E. thanks the program 'Ramón y Cajal', Spain for partial financial support.

References

- [1] J.M. Thornton, A.E. Todd, N. Borkakoti, C. Orengo, *Nature Struct. Biol. Struct. Genom. Suppl.* (2000) 991.
- [2] M. Randić, G. Krilov, *Chem. Phys. Lett.* 272 (1997) 115.
- [3] M. Randić, G. Krilov, *Int. J. Quantum Chem.* 75 (1999) 1017.
- [4] A.T. Balaban, C. Rucker, *J. Chem. Inf. Comput. Sci.* 41 (2001) 1145.
- [5] L. Liu, T. Wang, *Int. J. Quantum Chem.* 107 (2007) 1970.
- [6] E. Estrada, *Chem. Phys. Lett.* 319 (2000) 713.
- [7] E. Estrada, *Bioinformatics* 18 (2002) 697.
- [8] E. Estrada, *Comput. Biol. Chem.* 27 (2003) 305.
- [9] E. Estrada, *Proteins: Struct., Funct. Bioinf.* 54 (2004) 727.
- [10] E. Estrada, *J. Chem. Inf. Comput. Sci.* 44 (2004) 1238.
- [11] E. Estrada, E. Uriarte, *Comput. Biol. Chem.* 29 (2005) 345.
- [12] E. Estrada, E. Uriarte, S. Vilar, *J. Proteome Res.* 5 (2006) 105.
- [13] E. Estrada, A tight-binding 'dihedral orbitals' approach to the degree of folding of macromolecular chains, *J. Phys. Chem.*, in press. doi:10.1012/jp074595x.
- [14] J.E. Lennard-Jones, *Proc. Roy. Soc. A* 198 (1949) 1.
- [15] G.G. Hall, J.E. Lennard-Jones, *Proc. Roy. Soc. A* 202 (1950) 155.
- [16] G.G. Hall, J.E. Lennard-Jones, *Proc. Roy. Soc. A* 205 (1951) 357.
- [17] J.E. Lennard-Jones, J. Pople, *Proc. Roy. Soc. A* 202 (1950) 166.
- [18] R.D. Brown, *J. Chem. Soc.* (1953) 2615.
- [19] N. Trinajstić, *Chemical Graph Theory*, CRC Press, Boca Raton, 1992, p. 20.
- [20] E. Estrada, N. Guevara, I. Gutman, L. Rodríguez, *SAR & QSAR Environ. Res.* 9 (1998) 229.
- [21] E. Estrada, *J. Chem. Inf. Comput. Sci.* 39 (1999) 90.
- [22] D.B. Dix, An application of iterated line graphs to biomolecular conformation. <<http://www.math.sc.edu/~dix/graph.pdf>>.
- [23] R.A. Horn, C.R. Johnson, *Matrix Analysis*, Cambridge University Press, 1985.
- [24] D.N. Beratan, J.N. Onuchi, J.J. Hopfield, *J. Chem. Phys.* 86 (1987) 4488.
- [25] C. Goldman, *Phys. Rev. A* 43 (1991) 4501.
- [26] A.A. Stuchebrukhov, *Theor. Chem. Acc.* 110 (2003) 291.
- [27] T. Shimazaki, H. Maruyama, Y. Asai, K. Yamashita, *J. Chem. Phys.* 123 (2005) 164111.
- [28] P.M. Morse, H. Feshbach, *Methods of Theoretical Physics*, McGraw-Hill, New York, 1953 (Chapter 7).
- [29] E. Estrada, N. Hatano, *Chem. Phys. Lett.* 439 (2007) 247.
- [30] E. Estrada and N. Hatano, Communicability in complex networks. ArXiv: 0707.0756.
- [31] X. Qian, S. Gozani, H.S. Yoon, C.J. Jeon, K. Agarwal, M.A. Weiss, *Biochemistry* 32 (1993) 9944.
- [32] D.H. McLain, *Comput. J.* 17 (1974) 318.
- [33] STATISTICA 6.0., StatSoft Inc., 2002.
- [34] W.R. Taylor, A. Aszódi, *Protein Geometry, Classification, Topology and Symmetry. A Computational Analysis of Structure*, Taylor & Francis, New York, 2005.
- [35] F.M. Pearl et al., *Nucleic Acids Res.* 31 (2003) 452.