Genetics and Molecular Biology, 29, 2, 408-412 (2006)

www.sbg.org.br

Short Communication

Molecular surface directionality of the DNA-binding protein surface on the earth map

Wei-Po Lee[1] and Wen-Shyong Tzou[2,3,4]

[1]*National University of Kaohsiung, Department of Information Management, Taiwan.*
[2]*National Taiwan Ocean University, Institute of Bioscience and Biotechnology, Keelung, Taiwan.*
[3]*National Taiwan Ocean University, Department of Life Science, Keelung, Taiwan.*
[4]*National Taiwan Ocean University, Center for Marine Bioscience and Biotechnology, Keelung, Taiwan.*

Abstract

Protein-DNA interactions play a pivotal role in both the transcriptional control and the maintenance of genome integrity, and these are two properties that are closely linked to the development of an organism, differentiation, physiology and to the progression of diseases. Chemical and geometric properties are typically two of the key components in any analysis that aims to understand the precise origin of specificity and elucidate the atomic features of a protein-DNA interface. In this study, we have developed a unique representation of the directionality of the molecular surface of a DNA-binding protein. The stereo-orientation of the normal vector that signifies the geometric properties of a protein surface was projected onto a two-dimensional surface (referred to here as an earth map). We identified considerably diverse patterns of the vector distribution of the protein surface, and besides this, the DNA-contact surface, a subset of an entire protein surface, has also been found to contain diverse patterns. At the same time, the direction of the DNA-contact surface was also tracked onto the earth map on a base-pair basis and distinct intertwining properties particular to the specific family of that DNA-binding protein are revealed.

Key words: normal vector, molecular surface, protein-DNA interaction, earth map.

Received: June 1, 2005; Accepted: November 18, 2005.

One of the ultimate goals of functional genomics is to identify and record the entire transcription map of any given species among all living organisms. The interactions between DNA-binding proteins and DNA are, after all, responsible for the initiation, elongation and the termination of transcription processes. Aside from this, the maintenance of genome integrity, the control of epigenetics and the recombination of DNA are all molecular processes that involve protein-DNA interactions. Both the manner in which specificity is born and in which proteins search and anchor onto DNA during protein-DNA interactions have been the focus of a great deal of intensive research for much of the past decades (Pabo and Sauer, 1984; Choo and Klug, 1997; Garvie and Wolberger, 2001; Jayaram and Jain, 2004).

While static images of protein-DNA complexes can be obtained by means of X-ray crystallography and nuclear magnetic resonance, the actual realization of a dynamic portrait of a protein-DNA recognition process has yet to be fully achieved. In search of specific atomic details, protein-DNA interfaces have been well examined, but thus far, only two properties have been identified to explain the complementary properties on the interfaces: chemical and geometric complementarity. As concerns the former, it has been established that most protein-DNA interfaces are more polar and that they contain more hydrogen bonds than do protein-protein interfaces (Jones *et al.*, 1999; Nadassy *et al.*, 1999) despite the comparable interfacial gap volume between the two (Jones *et al.*, 1999). And certainly not to be ignored, given the importance of specific DNA sequences and the flexibility of DNA to protein-DNA interactions, several other studies have been centered on changes in DNA conformation (Suzuki and Yagi, 1994; Meierhans *et al.*, 1997; Dickerson, 1998; Mandel-Gutfreund and Margalit, 1998; Segal and Barbas, 2000; Maris *et al.*, 2002; Ahmad *et al.*, 2004; Havranek *et al.*, 2004; Paillard *et al.*, 2004).

Turning to geometric complementarity, what is striking is the way protein shape "follows" DNA conformation, or conversely, the way DNA "adjusts" its conformation such that it closely parallels the shape of a protein surface. To quantitate exactly how the shape of a protein surface

Send correspondence to Wen-Shyong Tzou. National Taiwan Ocean University, Institute of Bioscience and Biotechnology, 2 Pei-Ning Road, 20224 Keelung, Taiwan. E-mail: wstzou@ntou.edu.tw.

stays in line with that of a DNA molecule, we have previously investigated the relationship between the "direction" of a protein surface (*i.e.*, the "normal vector" of a protein surface) and that of a DNA molecule (*i.e.*, the "axis" of a base-pair plane) (Yeh *et al.*, 2003). In the analysis of a set of non-redundant protein-DNA complexes with known three-dimensional structures, strong evidence has substantiated that a significant correlation exists between the direction of a protein surface and the conformation of a DNA molecule. Thus, in that research (Yeh *et al.*, 2003), a new geometric property of protein-DNA interfaces was determined, showing that the 'shape complementarity' of protein-DNA recognition unambiguously bears the property of 'directionality'.

Equally important, our goal in this study was to investigate the distribution of the orientation that best depicts the direction of the molecular surface of a DNA-binding protein. To achieve this we, first of all, moved the root of each normal vector to the same origin so that we could compare all of the vectors by only seeing the arrow heads of the vector from a shared origin. To facilitate the examination of the vector distribution, we employed the Hammer-Aitoff projection method to map a projection of the vector tips from the three-dimensional unit sphere to a two-dimensional plane. The transformation formulas are as follows (Snyder, 1993):

$$s = \frac{2R\sqrt{2}\ \cos\ \sin(\ /2)}{\sqrt{1\ \cos\ \cos(\ /2)}}$$

$$t = \frac{R\sqrt{2}\ \sin}{\sqrt{1\ \cos\ \cos(\ /2)}}$$

where ϕ is the latitude of each vector tip (+ if north and - if south); λ is the longitude (+ if eastward and - if westward); R is the radius of the sphere ($R = 1$); and s and t are the positions on the projected two-dimensional plane (*i.e.*, the earth map). The results from the two-dimensional perspective of the normal vectors of the protein surface were enlightening. Take the human YY1 zinc finger protein-DNA complex structure as an example (Figure 1A). It is readily observed that the normal vectors are unevenly distributed on the earth map, with some strips and zones showing a clustering of vectors (Figure 2A). The pattern of this map is specific to the YY1 zinc finger protein surface. That is, it has its own special pattern as opposed to, say, being P53 specific with its own special pattern (Figure 1B, 2B).

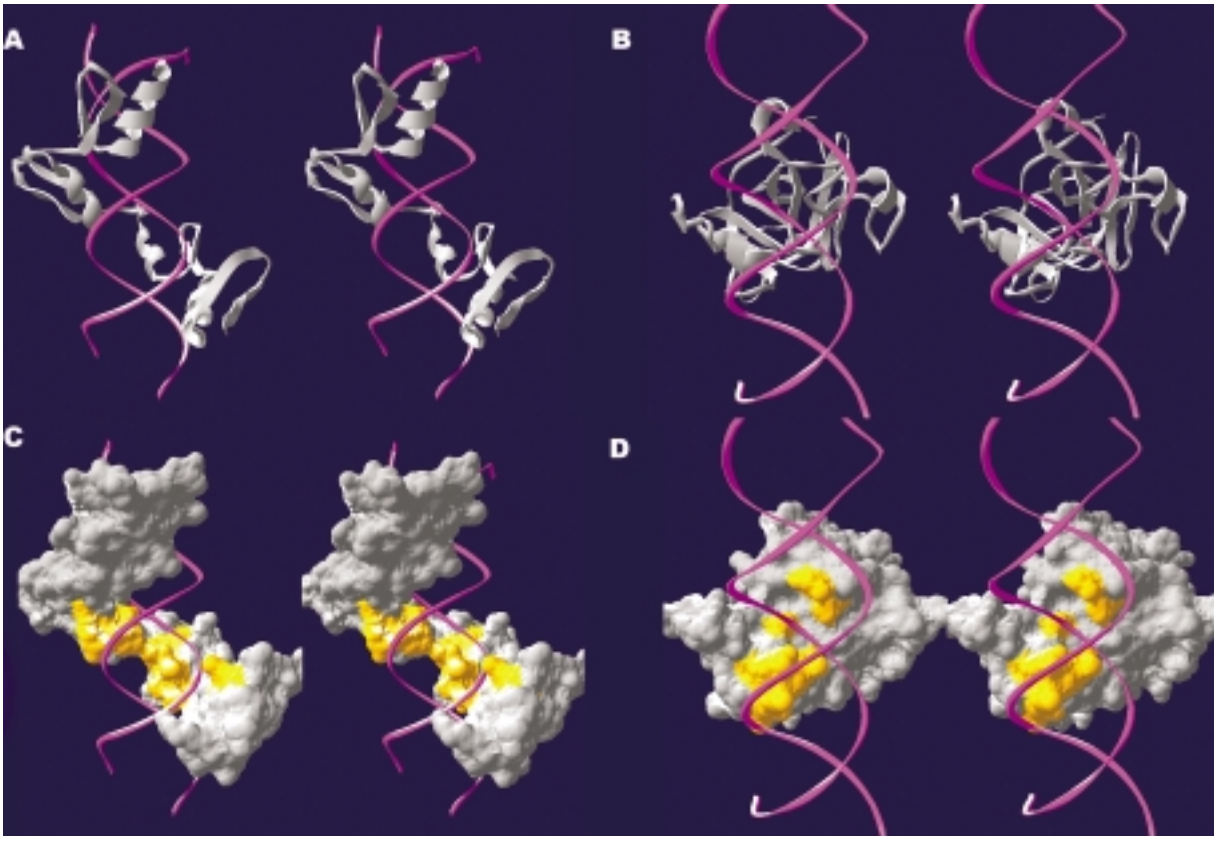

Figure 1 - Stereo image of the structure (A and B) and molecular surface (C and D) of the YY1 zinc finger-DNA complex (A and C, PDB code 1ubd) and the P53-DNA complex (B and D, PDB code 1tsr). The molecular surface of the protein is marked in grey, and the surface in contact with the DNA is in yellow. For clarity, the DNA is shown as a ribbon. (For the views of different orientations, please see http://140.121.200.163/molecule.asp?pathn=1ubd and http://140.121.200.163/molecule.asp?pathn=1tsr).

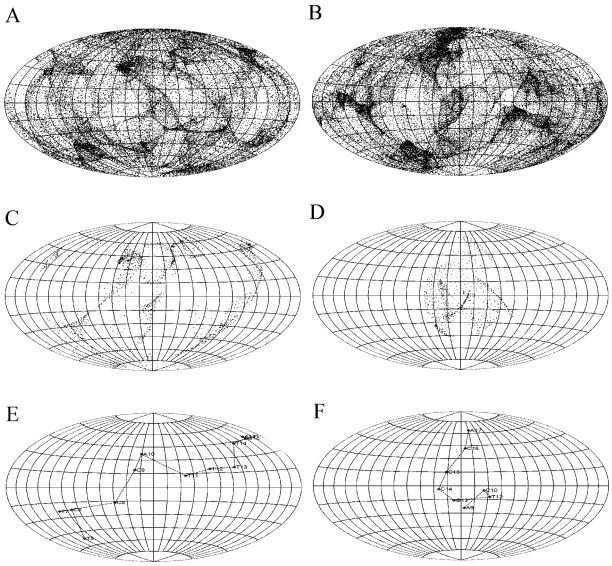

Figure 2 - Orientation distribution of the normal vectors of a protein surface shown on the earth map (A, C and E for the YY1 zinc finger-DNA complex; B, D and F for the P53-DNA complex). The normal vectors for the whole protein are shown in A and B, while the normal vectors involved in the DNA contact are shown in C and D. The normal vectors of the protein surface involved in DNA contact are averaged on a base-pair unit and are shown on the earth map (E, F). Each diamond-shaped symbol represents a base-pair, and they are linked with a dashed line to trace the overall orientation of the average of the normal vectors. Also shown is the nucleotide sequence of the YY1 zinc finger protein-bound DNA (nucleotide sequence T5-C6-T7-C8-C9-A10-T11-T12-T13-T14-G15-A17, PDB ID: 1ubd) (E) and the P53-bound DNA (A9-C10-T12-G13-C14-C16-A17, PDB ID: 1tsr) (F).

We also explored the normal vectors of the DNA-binding proteins that are involved in DNA contact. Since the molecular surface of the protein involved in each DNA contact is but a subset of the whole surface of all the proteins, only some of the vectors in Figure 2A are shown in Figure 2C. Characteristic of the vector of the YY1 protein surface in contact with the DNA are three strips; however, there is an isolated patch of vectors which features the normal vector of the P53 protein surface in contact with the DNA (Figure 2D).

We then examined the molecular surface contact between the protein and the DNA in detail by profiling the normal vectors of the protein surface averaged on a base-pair unit. We calculated the vector sum of the normal vectors of the protein surface in contact with each DNA base-pair and also illustrated the vector sum on the earth map based upon the base-pair unit. As shown in the example of the YY1 zinc finger protein, the normal vectors of the protein surface intertwining through the major groove of the DNA (Figure 1A and 1C) traverses longitudinally through a circle on the map (almost 360°) (Figure 2E). This particular "intertwining" feature is dramatically different from that shown for P53, which zigzags in a latitudinal fashion (Figure 2F).

We also employed the clustering of the normal vectors of the DNA-contact protein surface averaged on a base-pair unit on the earth map using the dynamic programming method. We found that the directional features of the molecular surface of DNA-binding proteins, including helix-loop-helix, zinc finger, β hairpin/ribbon and helix-turn-helix families, formed distinct groups in the dendrogram (Figure 3), demonstrating that DNA-binding proteins of the same family have similar directionality patterns.

We have also constructed a website (named "ShapeCom" at http://140.121.200.163/shapecom.htm) to present the details of different protein-DNA interactions. On the website, the geometric properties of the interface from various protein-DNA complexes are visually represented, and valuable links to other web-sites are provided for the examination of the coordinates and the chemical properties of protein-DNA interfaces. The web contents for each protein-DNA complex include:

Data & links

1) links to the Protein Data Bank (PDB, http://www.rcsb.org/pdb/) (Deshpande *et al.*, 2005);

2) the coordinates of the protein part, the DNA part and the protein-DNA complex(Deshpande *et al.*, 2005);

3) links to the web-site of the Biomolecular Structure and Modelling (BSM) group, University College, London (http://www.biochem.ucl.ac.uk/bsm/prot_dna/prot_dna_cover.html) ; and

4) links to PDBsum (Laskowski *et al.*, 2005).

Surface properties

5) stereo images of the molecular surfaces of various rotations of molecules, featuring the topography of a protein-DNA interface (based on the Swiss-PdbViewer,

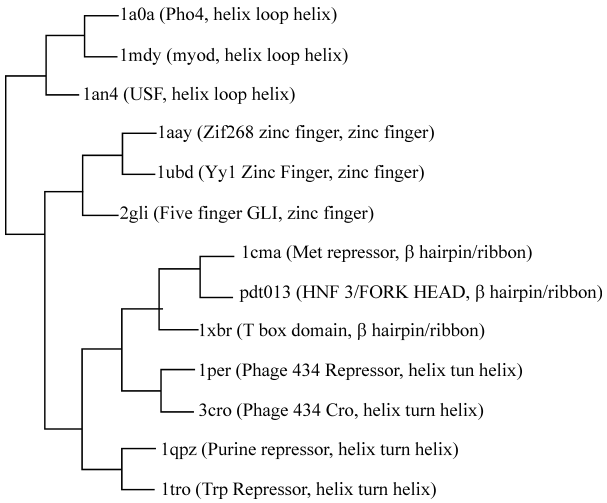

Figure 3 - A dendrogram for the clustering of the normal vectors of the DNA-contact protein surface averaged on a base-pair unit on the earth map. For each leaflet, PDB ID, protein name and the DNA-binding protein family name are shown.

http:// swissmodel.expasy.org/spdbv/ (Guex and Peitsch, 1997));

6) stereo images of the normal vectors of a protein surface and the axes of a DNA base-pair plane (Yeh *et al.*, 2003);

7) the angles between the normal vectors of a protein surface and the axes of a DNA base-pair plane(Yeh *et al.*, 2003);

8) a two-dimensional projection of the normal vectors of a protein surface (this study, Figures 2A, 2B, 2C and 2D); and

9) a two-dimensional projection of the normal vectors of a protein surface in contact with DNA, averaged on a base-pair unit (this study, Figures 2E and 2F).

On the weight of the evidence from our previous investigation, we concluded that the normal vectors of a DNA-contacting protein surface distinctly prefer certain angles. This enables them to align with certain axes that characterize the conformation of the DNA. We have now extended our 'shape complementarity' studies of protein-DNA recognition to encompass the topographic properties of the 'directionality' of the protein surface. By employing the two-dimensional projection techniques presented here, we found that the distribution of the normal vectors on the earth map is uneven and, at the same time, that it varies among different DNA-binding proteins. Beyond this, most of the vectors in contact with DNA also have their own distinct pattern depending on the protein under investigation. Our two-dimensional representation of the normal vectors of a protein surface can be regarded as a natural extension of the current trends in protein research. The protein surface has long been used to categorize and predict the functions of proteins and their interaction with other biological molecules (Lichtarge *et al.*, 1996; Jones and Thornton, 1997). To simplify the representation and comparison, the protein surface was approximated using the spherical harmonic function (Duncan and Olson, 1993). More specifically, a spherical approximation of the protein surface was used to analyze the surface features within homologous families and to predict the conservation and divergence of protein functions and protein-protein interactions (Pawlowski and Godzik, 2001). Worth bearing in mind is that the normal vectors of a protein surface were previously used in the protein-protein and protein-ligand docking problem (Norel *et al.*, 1995; Norel *et al.*, 1999). To the best of our knowledge, however, never before have the normal vectors been represented on a two-dimensional projection (earth map); nor have the distinct patterns for the "directionality" of proteins and their DNA-contact patches been visualized on an earth map. As further detailed in our web-site (ShapeCom), DNA-binding proteins in different families utilize the family-specific protein surface in DNA contacts. The specific directionality of each family of DNA-binding proteins may very well play an important role when it comes to understanding the recognition process of protein-DNA and protein-protein interactions.

Acknowledgements

We are grateful for the discussions with Dr. Ming-Jing Hwang and the help of Mr. Edward S. C. Shih in the clustering work. This work was supported by the National Science Council of Taiwan (92-2113-M-019-001, NSC 93-2113-M-019-001, NSC 93-2213-E-390-002).

References

Ahmad S, Gromiha MM and Sarai A (2004) Analysis and prediction of DNA-binding proteins and their binding residues based on composition, sequence and structural information. Bioinformatics 20:477-486.

Choo Y and Klug A (1997) Physical basis of a protein-DNA recognition code. Curr Opinion Struct Biol 7:117-125.

Deshpande N, Addess KJ, Bluhm WF, Merino-Ott JC, Townsend-Merino W, Zhang Q, Knezevich C, Xie L, Chen L, Feng Z, Green RK, Flippen-Anderson JL, Westbrook J, Berman HM and Bourne PE (2005) The RCSB protein data bank: A redesigned query system and relational database based on the mmCIF schema. Nucleic Acids Res 33:D233-D237.

Dickerson RE (1998) DNA bending: The prevalence of kinkiness and the virtues of normality. Nucleic Acids Res 26:1906-1926.

Duncan BS and Olson AJ (1993) Approximation and characterization of molecular surfaces. Biopolymers 33:219-229.

Garvie CW and Wolberger C (2001) Recognition of specific DNA sequences. Mol Cell 8:937-946.

Guex N and Peitsch MC (1997) SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. Electrophoresis 18:2714-2723.

Havranek JJ, Duarte CM and Baker D (2004) A simple physical model for the prediction and design of protein-DNA interactions. J Mol Biol 344:59-70.

Jayaram B and Jain T (2004) The role of water in protein-DNA recognition. Annu Rev Biophys Biomol Struct 33:343-361.

Jones S, van Heyningen P, Berman HM and Thornton JM (1999) Protein-DNA interactions: A structural analysis. J Mol Biol 287:877-896.

Jones S and Thornton JM (1997) Prediction of protein-protein interaction sites using patch analysis. J Mol Biol 272:133-143.

Laskowski RA, Chistyakov VV and Thornton JM (2005) PDBsum more: New summaries and analyses of the known 3D structures of proteins and nucleic acids. Nucleic Acids Res 33:D266-D268.

Lichtarge O, Bourne HR and Cohen FE (1996) An evolutionary trace method defines binding surfaces common to protein families. J Mol Biol 257:342-358.

Luscombe NM, Austin SE, Berman HM and Thornton JM (2000) An overview of the structures of protein-DNA complexes. Genome Biol 1:REVIEWS 001.1-001.37.

Mandel-Gutfreund Y and Margalit H (1998) Quantitative parameters for amino acid-base interaction: Implications for prediction of protein-DNA binding sites. Nucleic Acids Res 26:2306-2312.

Maris AE, Sawaya MR, Kaczor-Grzeskowiak M, Jarvis MR, Bearson SM, Kopka ML, Schroder I, Gunsalus RP and Dickerson RE (2002) Dimerization allows DNA target site recognition by the NarL response regulator. Nat Struct Biol 9:771-778.

Meierhans D, Sieber M and Allemann RK (1997) High affinity binding of MEF-2C correlates with DNA bending. Nucleic Acids Res 25:4537-4544.

Nadassy K, Wodak SJ and Janin J (1999) Structural features of protein-nucleic acid recognition sites. Biochemistry 38:1999-2017.

Norel R, Lin SL, Wolfson HJ and Nussinov R (1995) Molecular surface complementarity at protein-protein interfaces: The critical role played by surface normals at well placed, sparse, points in docking. J Mol Biol 252:263-273.

Norel R, Petrey D, Wolfson HJ and Nussinov R (1999) Examination of shape complementarity in docking of unbound proteins. Proteins 36:307-317.

Pabo CO and Sauer RT (1984) Protein-DNA recognition. Annu Rev Biochem 53:293-321.

Paillard G, Deremble C and Lavery R (2004) Looking into DNA recognition: Zinc finger binding specificity. Nucleic Acids Res 32:6673-6682.

Pawlowski K and Godzik A (2001) Surface map comparison: Studying function diversity of homologous proteins. J Mol Biol 309:793-806.

Segal DJ and Barbas CF III (2000) Design of novel sequence-specific DNA-binding proteins. Curr Opinion Chem Biol 4:34-39.

Snyder JP (1993) Flattening the earth: Two thousand years of map projections. University of Chicago Press, Chicago, 133 pp.

Suzuki M and Yagi N (1994) DNA recognition code of transcription factors in the helix-turn-helix, probe helix, hormone receptor, and zinc finger families. Proc Natl Acad Sci USA 91:12357-12361.

Yeh CS, Chen FM, Wang JY, Cheng TL, Hwang MJ and Tzou WS (2003) Directional shape complementarity at the protein-DNA interface. J Mol Recognit 16:213-222.

Associate Editor: Sandro José de Souza

Genetics and Molecular Biology, 29, 4, 771-771 (2006)

www.sbg.org.br

Erratum

Erratum: 29(2) p 408-412
Molecular surface directionality of the DNA-binding protein surface on the earth map

Wei-Po Lee[1] and Wen-Shyong Tzou[2,3*]

[1]*Department of Information Management, National University of Kaohsiung, Taiwan.*
[2]*Institute of Bioscience and Biotechnology, National Taiwan Ocean University 2, Pei-Ning Road, Keelung, Taiwan.*
[3]*Department of Life Science, National Taiwan Ocean University 2, Pei-Ning Road, Keelung, Taiwan.*

The size of Figure 2 was over-reduced. The Publisher regrets the error and publishes this figure in larger format.

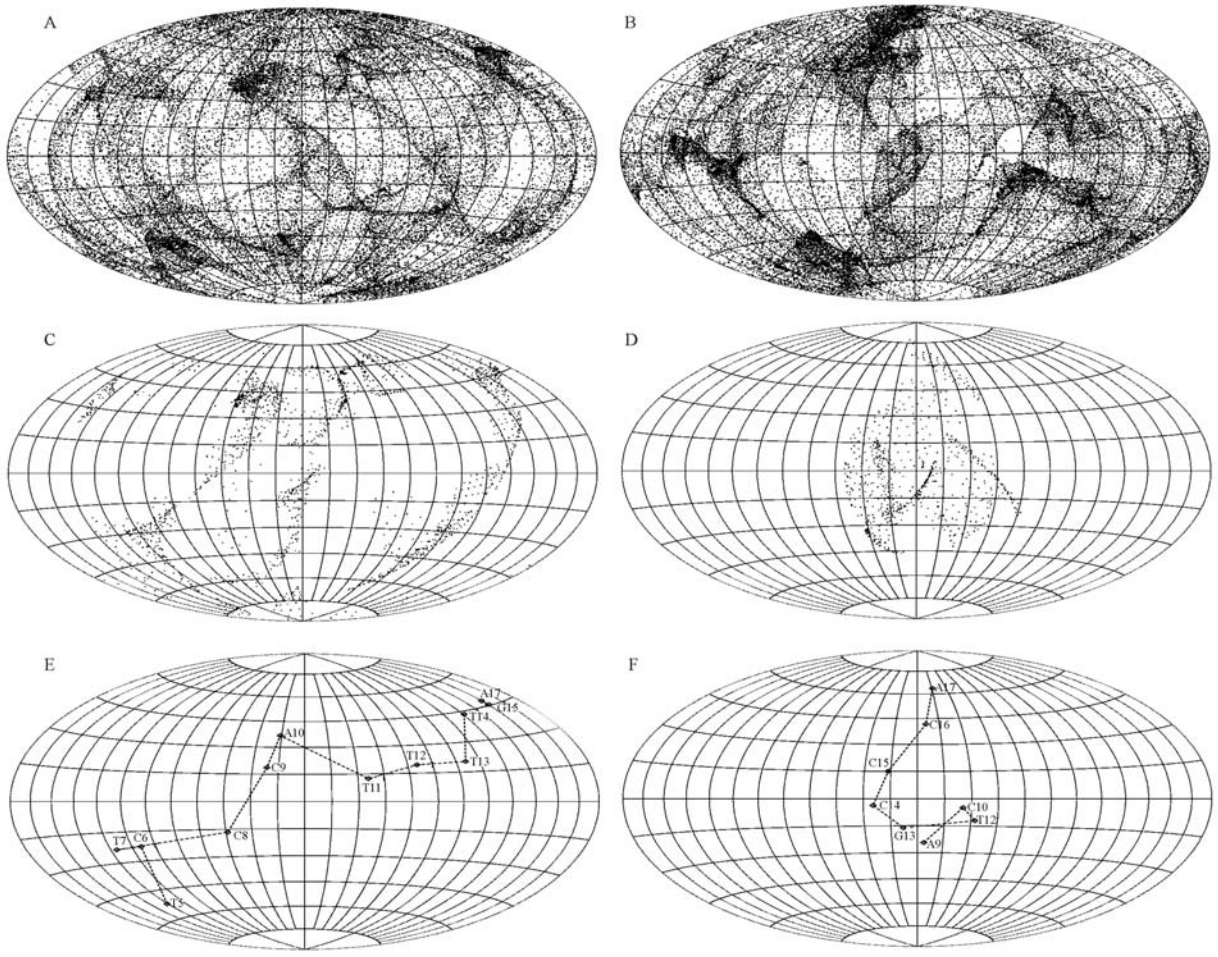

Figure 2 - Orientation distribution of the normal vectors of a protein surface shown on the earth map (A, C and E for the YY1 zinc finger-DNA complex; B, D and F for the P53-DNA complex). The normal vectors for the whole protein are shown in A and B, while the normal vectors involved in the DNA contact are shown in C and D. The normal vectors of the protein surface involved in DNA contact are averaged on a base-pair unit and are shown on the earth map (E, F). Each diamond-shaped symbol represents a base-pair, and they are linked with a dashed line to trace the overall orientation of the average of the normal vectors. Also shown is the nucleotide sequence of the YY1 zinc finger protein-bound DNA (nucleotide sequence T5-C6-T7-C8-C9-A10-T11-T12-T13-T14-G15-A17, PDB ID: 1ubd) (E) and the P53-bound DNA (A9-C10-T12-G13-C14-C16-A17, PDB ID: 1tsr) (F).

This article has received corrections in agreement with the ERRATUM published in Volume 29 Number 4.