BIOINFORMATICS APPLICATIONS NOTE Vol. 19 no. 2 2003
Pages 313–314

DPX: for the analysis of the protein core

Alessandro Pintar [1,*], *Oliviero Carugo* [1,2] *and Sándor Pongor* [1]

[1]*Protein Structure and Bioinformatics Group, International Center for Genetic Engineering and Biotechnology (ICGEB), AREA Science Park, Padriciano 99, 34012 Trieste, and* [2]*Department of General Chemistry, University of Pavia, Viale Taramelli 12, 27100 Pavia, Italy*

Received on May 23, 2002; revised on July 22, 2002; accepted on September 13, 2002

ABSTRACT

Summary: In order to obtain an accurate description of the protein interior, we describe a simple and fast algorithm that measures the depth of each atom in a protein (dpx), defined as its distance (Å) from the closest solvent accessible atom. The program reads a PDB file containing the atomic solvent accessibility in the B-factor field, and writes a file in the same format, where the B-factor field now contains the dpx value. Output structure files can be thus directly displayed with molecular graphics programs like RASMOL, MOLMOL, Swiss-PDB View and colored according to dpx values.

Availability: The algorithm is implemented in a standalone program written in C and its source is freely available at ftp.icgeb.trieste.it/pub/DPX or on request from the authors.

Contact: pintar@icgeb.org; carugo@icgeb.org; pongor@icgeb.org

In our search for fast algorithms that can be applied efficiently to entire structure databases (Carugo and Pongor, 2002; Pintar *et al.*, 2002), we addressed the following problem: is there a simple way to measure to what extent an atom is buried in the interior of a protein, or, in other words, to measure its depth? This question has several biological implications. First, residues that are most deeply buried in the protein interior in the native structure might play a key role in the folding process. Second, deeply buried residues might contribute more to the thermodynamic stability of a protein. Third, the characteristic depth of each atom in a protein can be used to classify protein atoms independently of their explicit physico-chemical properties. Finally, atoms that are buried, but close to the protein surface, might become accessible through the internal dynamics movements of the protein, or through small, energetically inexpensive conformational changes; these atoms might then be involved in interactions with other molecules, or become targets for post-translational modifications.

*To whom correspondence should be addressed.

While the solvent accessible area (Lee and Richards, 1971) has been widely used, and is very useful in the analysis of atoms and residues at the protein surface, it provides little information on atoms and residues that are buried. To get insight into the protein interior, we thus introduce a new physical descriptor, atom depth. Although the most accurate definition of atom depth would be the distance between a buried atom and the solvent accessible or the contact surface, or even the distance between a buried atom and its closest water molecule surrounding the protein, we opted for a simpler algorithm, which is then very fast in its execution.

In our approach, the depth of an atom i (dpx_i) is defined as its distance (Å) from the closest solvent accessible atom j (atomic solvent accessibility, $asa_j > 0$):

$$\mathrm{dpx}_i = \min(d1, d2, d3, \ldots, dn)$$

where $d1, d2, d3, \ldots, dn$ are the distances between the atom i and all solvent accessible atoms. The depth (dpx) is thus zero for solvent accessible atoms, and >0 for atoms buried in the protein interior, deeply buried atoms thus having higher dpx values.

The DPX principle is implemented in a simple standalone program written in C. The program reads a coordinate file in the PDB format, where the B-factor field contains the atomic accessible surface (Å^2). Routinely, we used NACCESS (Hubbard and Thornton, 1993) with a probe radius of 1.4 Å to calculate atomic solvent accessibility. The .asa output file from NACCESS can then be directly read by the program. As in the default settings of NACCESS, DPX ignores HETATOM entries, and multiple chain PDB files are treated as a single molecule. The output is a coordinate file in PDB format in which the B-factor field contains the dpx value. The output files can be thus read by molecular graphics programs like RASMOL (Sayle and Milner-White, 1995), MOLMOL (Koradi *et al.*, 1996), and Swiss-PdbView (Guex and Peitsch, 1997) and atoms colored by their dpx value. The dpx value is an atomic property with a straightforward physical meaning (it is a distance in Å) and it can be thus

 © Oxford University Press 2003; all rights reserved.

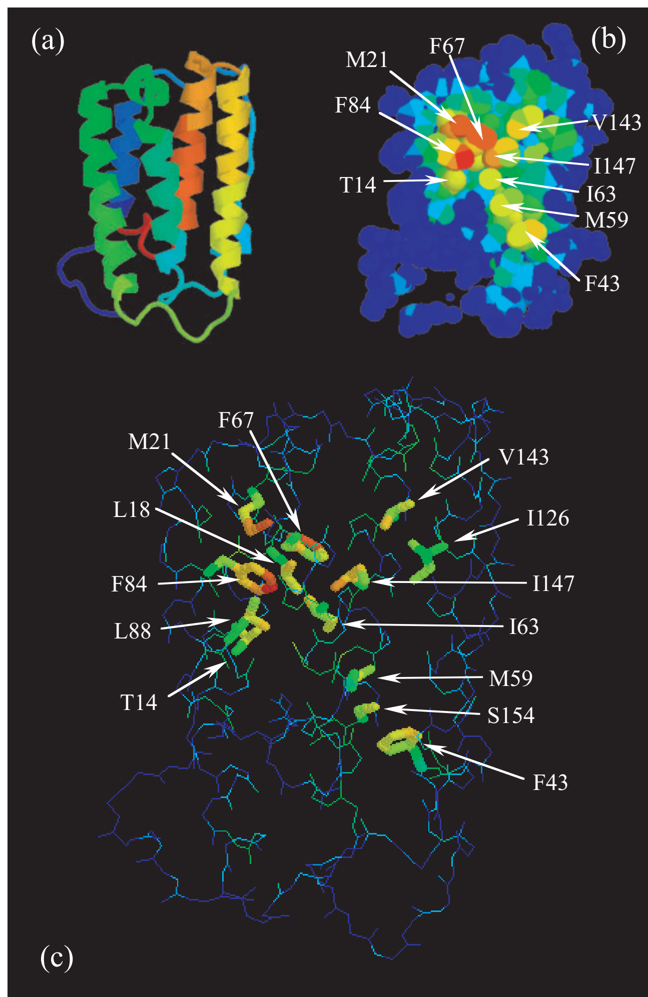

Fig. 1. Analysis of the protein interior. The structure of human interferon-α2a (PDB: 1ITF, model 1; Klaus *et al.*, 1997) shown in the same orientation as **(a)** a cartoon model, **(b)** a slice view of the CPK model, **(c)** a backbone model, with the side chains of the most deeply buried residues shown as sticks. In (a) the chain is colored according to residue number, from blue (N-terminus) to red (C-terminus); in (b) and (c) the model is colored according to dpx values, from blue (dpx $=$ 0) to red (high dpx values). The figure was prepared using RASMOL (Sayle and Milner-White, 1995).

handled easily: for example, main chain, side chain, and residue mean values can be calculated.

On a SGI R10000 (195 MHz) processor, the program required $\sim$0.02 s c.p.u. time for our test molecule (PDB: 1ITF, $\sim$1350 atoms). On the same molecule, NACCESS required $\sim$0.90 s c.p.u. time.

To show the potential of DPX, we calculated the dpx values for the NMR structure of human interferon-α2a (PDB: 1ITF, model 1; Klaus *et al.*, 1997). Different views of this molecule are shown in Figure 1. DPX allows for the prompt identification of the side chains that are most deeply buried in the protein interior. It can be remarked that, despite the apparently high symmetry of the molecule, the hydrophobic core is highly polarized, with the most deeply buried residues clustered around L18 (Figure 1c), as shown by dpx values.

REFERENCES

Carugo,O. and Pongor,S. (2002) Protein fold similarity estimated by a probabilistic approach based on C(alpha)–C(alpha) distance comparison. *J. Mol. Biol.*, **315**, 887–898.

Guex,N. and Peitsch,M.C. (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis*, **18**, 2714–2723.

Hubbard,S.J. and Thornton,J.M. (1993) *NACCESS*. Department of Biochemistry and Molecular Biology, University College, London.

Klaus,W., Gsell,B., Labhardt,A.M., Wipf,B. and Senn,H. (1997) The three-dimensional high resolution structure of human interferon alpha-2a determined by heteronuclear NMR spectroscopy in solution. *J. Mol. Biol.*, **274**, 661–675.

Koradi,R., Billeter,M. and Wuthrich,K. (1996) MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.*, **14**, 51–55, 29–32.

Lee,B. and Richards,F.M. (1971) The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.*, **55**, 379–400.

Pintar,A., Carugo,O. and Pongor,S. (2002) CX, an algorithm that identifies protruding atoms in proteins. *Bioinformatics*, **18**, 980–984.

Sayle,R.A. and Milner-White,E.J. (1995) RASMOL: biomolecular graphics for all. *Trends Biochem. Sci.*, **20**, 374.