eF-site and PDBjViewer: database and viewer for protein functional sites

Kengo Kinoshita¹,² and Haruki Nakamura³,*

¹Graduate School of Integrated Science, Yokohama City University, 1-7-29 Suehiro-cho, Tsurumi-ku, Yokohama, 230-0045, Japan, ²Structure and Function of Biomolecules, PRESTO, JST, 4-1-8 Honcho, Kawaguchi, Saitama, 332-0012, Japan and ³Institute for Protein Research, Osaka University, 3-2 Yamadaoka, Suita, Osaka, 565-0871, Japan

Received on September 2, 2003; revised on October 27, 2003; accepted on November 13, 2003
Advance Access publication February 10, 2004

ABSTRACT
Summary: The electrostatic-surface of functional site (eF-site) is a database for the molecular surfaces of protein functional sites. To enable browsing of each molecular surface along with the atomic model, we have developed a new three-dimensional interactive viewer, PDBjViewer, that can be used both as an applet and as a stand-alone program.

Availability: The eF-site database and PDBjViewer are freely available from http://www.pdbj.org/eF-site/

Contact: harukin@protein.osaka-u.ac.jp

INTRODUCTION

The goal of current structural genomics and structural proteomics studies is to reveal biological and biochemical functions of individual protein molecules based on their three-dimensional (3D) structures. For that purpose, we have focused on the protein molecular surfaces, where most of the functional activity occurs. Along with surface geometry, electrostatic properties are considered essential for molecular recognition (Honig and Nicholls, 1995; Nakamura, 1996).

A database named electrostatic-surface of functional site (eF-site: http://www.pdbj.org/eF-site/) has been constructed such that the electrostatic information of the surface of each functional site is computed, registered in extensible markup language (XML) format and displayed by our original 3D viewer program, PDBjViewer. All the registered data as well as the PDBjViewer program can be freely downloaded from the above Web site.

One advanced feature of the eF-site database is the ability to carry out similarity searches of electrostatic molecular surfaces in order to identify biochemical function and locate active sites of hypothetical proteins, once a 3D structure is determined. Recently, several methods have been proposed for surface comparison (Kinoshita et al., 2002; Kinoshita and Nakamura, 2003a), and these methods are now applicable to actual functional predictions (Kinoshita and Nakamura, 2003b; Handa et al., 2003).

FUNCTIONAL SITE INFORMATION AND ELECTROSTATIC MOLECULAR SURFACES OF PROTEINS

We organize the functional sites of proteins into five categories, eF-site/BindingSite, eF-site/ActiveSite, eF-site/prosite, eF-site/antibody and eF-site/membrane. This classification is the same as in the previous report (Kinoshita et al., 2002), but the number of entries in each category have been doubled, which covers almost all entries appearing in the recent Protein Data Bank (PDB; Berman et al., 2000). A major enhancement has been done for eF-site/BindingSite. First, a total of 10 705 protein 3D complexes are taken from PDB for which the resolution is better than 2.5 Å, and the HETATM groups do not consist of metal ions, phosphate ions or sulphate ions. Based on these complexes, 15 480 functional sites were identified so that the corresponding amino acid residues are less distant than 5.0 Å from the HETATM groups, and they were entered into the eF-site/BindingSite category along with the corresponding ligand molecules.

In the eF-site/ActiveSite category, a total of 5 042 have been registered that were taken from SWISS-PROT (http://www.expasy.org/sprot/) (Boeckmann et al., 2003), with the features keyword ‘ACT_SITE’. In eF-site/prosite, 5399 motif surfaces were classified by their PROSITE consensus patterns and tagged with the corresponding prosite keyword (Sigrist et al., 2002). And for eF-site/Antibody and eF-site/membrane, molecules were manually extracted and prepared. The importance of these two groups in ligand recognition is great. Moreover, antibodies are good examples of proteins that have the same 3D framework but recognize a large variety of different molecules. More than 100 antibodies whose tertiary structures were determined with resolution better than 2.0 Å were included (Shirai et al., 1999).

A total of 19 121 functional sites are now registered for all the above categories in the eF-site database without duplication. Currently, the PDB files involve some crystallographic packing artefacts, which may be corrected by referring to the biological unit information in the beta Web site of PDB (http://beta.rcsb.org/).

*To whom correspondence should be addressed.
Every molecular surface was calculated by the program MSP, developed by Connolly (1983), based on the atomic coordinates of the protein. The electrostatic potential of every protein molecule was calculated by solving the Poisson–Boltzmann equations using the program SCB (Nakamura and Nishida, 1987), which used the self-consistent boundary algorithm to eliminate the effect of the boundary in the finite-difference method with a 1 Å grid. Dielectric constants of 2 and 80 were used for the protein and the solvent region, respectively, and an ionic strength of 0.1 M was assumed for every case. Changing these parameters could give different electrostatic potentials quantitatively, but the qualitative features are not very sensitive to the parameters (Nakamura, 1996). For the purpose of thorough comparison of molecular surfaces, we fixed the parameter values.

The electrostatic potential at each individual point, which is 1.4 Å away from each vertex on the surface triangle along the normal vector of the surface, was interpolated from the potential values at the grid points, as calculated by the continuum method. The hydrophobic properties of each vertex at the molecular surface were also provided according to the hydrophathy value (Kyte and Doolittle, 1982) of the corresponding side-chain of the residue.

DATABASE FRAMEWORK AND DATA FORMAT

The information for each functional site of a protein is registered in a file named ‘seqinfo’, which contains descriptions of the function and the location of the site with the corresponding residue and atom identifiers. In contrast, the molecular surface information on the physicochemical properties such as the electrostatic potentials and hydropathy values is in a file named ‘efvet’, which was extended from the VET file used in MSP (Connolly, 1983).

Both the seqinfo and efvet files are described in XML, and they can be downloaded from the eF-site Web page. Their schemas are also provided on the web. The eF-site database is maintained by a relational database using MySQL (http://www.mysql.com/), and simple text-based searching is available. A static surface picture is provided for each functional site in a JPEG-formed file, and browsing the molecular surface and the corresponding 3D atomic geometry is also available with our own viewer, PDBjViewer.

PDBjViewer

We have developed our own 3D molecular viewer, named PDBjViewer, as a product of PDB Japan (PDBj: http://www.pdbj.org/), one of the members of Worldwide PDB (wwPDB), which begins its international activity in 2003. In addition to drawing several standard molecular models like the popular RasMol program (http://www.OpenRasMol.org/), PDBjViewer can illustrate any polygon objects in XML format. The molecular surface displayed in eF-site is an example of a polygon object. Furthermore, it can parse atomic coordinates from both the flat PDB files and the PDB XML files (a combination of ftp://beta.rcsb.org/pub/pdb/uniformity/data/XML/all-noatom/ and ftp://beta.rcsb.org/pub/pdb/uniformity/data/XML/all-extatomi/). PDBjViewer is implemented with Java and Java3D technologies, and it can be used both as an applet for a Web browser and as a stand-alone viewer program.

ACKNOWLEDGEMENTS

The authors thank Dr Hiroyuki Sato (Information and Mathematical Science Laboratory, Inc.) for his technical support. Development of the eF-site database has been supported by a grant-in-aid from Institute for Bioinformatics Research and Development, Japan Science and Technology Agency (BIRD-JST).

REFERENCES


