BIOINFORMATICS APPLICATIONS NOTE Vol. 21 no. 20 2005, pages 3929–3930
doi:10.1093/bioinformatics/bti645

Structural bioinformatics

The SuMo server: 3D search for protein functional sites

Martin Jambon[1,*], Olivier Andrieu[2], Christophe Combet[1], Gilbert Deléage[1], François Delfaud[2] and Christophe Geourjon[1]

[1]Pôle BioInformatique Lyonnais, Institut de Biologie et Chimie des Protéines, UMR 5086 CNRS/UCBL, IFR 128, 7 passage du Vercors, 69367 Lyon cedex 07, France and [2]MEDIT SA, 2 rue du Belvédère, 91120 Palaiseau, France

Received on May 24, 2005; revised on August 2, 2005; accepted on August 24, 2005
Advance Access publication September 1, 2005

ABSTRACT

Summary: We provide the scientific community with a web server which gives access to SuMo, a bioinformatic system for finding similarities in arbitrary 3D structures or substructures of proteins. SuMo is based on a unique representation of macromolecules using selected triplets of chemical groups having their own geometry and symmetry, regardless of the restrictive notions of main chain and lateral chains of amino acids. The heuristic for extracting similar sites was used to drive two major large-scale approaches. First, searching for ligand binding sites onto a query structure has been made possible by comparing the structure against each of the ligand binding sites found in the Protein Data Bank (PDB). Second, the reciprocal process, i.e. searching for a given 3D site of interest among the structures of the PDB is also possible and helps detect cross-reacting targets in drug design projects.

Availability: The web server is freely accessible to academia through http://sumo-pbil.ibcp.fr and full support is available from MEDIT (http://www.medit.fr).

Contact: mjambon@burnham.org

1 BACKGROUND

We are interested in predicting ligand binding sites in protein structures, and the nature of the ligands. This problem has become more and more important since structural genomics provides hundreds of protein structures of unknown function. This problem (Jones and Thomton, 2004) can be treated by comparing protein structures against known 3D ligand binding sites. Several approaches exist in this domain and some are accessible through web servers (Kleywegt, 1999; Barker and Thomton, 2003; Russell, 1998; Ivanisenko, *et al*., 2004; Shulman-Peleg *et al*., 2004; Ausiello, *et al*., 2005). In the present study, we present the web interface of SuMo, a method that has been developed earlier [Jambon *et al*. (2003); Jambon (2003); (http://martin.jambon.free.fr/phd.html)] for searching 3D ligand binding sites in protein structures.

2 METHOD

A protein structure is represented by a set of functional groups: unbound hydrogen bond donors or acceptors, accessible sides of aromatic rings and carboxylate, primary amide, guanidinium, hydroxyl, imidazole, thioether and thiol groups. Each of these chemical groups are represented by a position and vectors that represent the orientation and the symmetry of the object.

The comparison procedure actually compares triplets of chemical groups because triplets are the minimum entity which can bind a ligand rigidly and because they carry more information than chemical groups alone. In addition to the information previously mentioned, the shape given by the position the atoms that are close to the triplet is also taken into account. The final representation of a protein or a site is given by the graph made of triplets which are connected when they have two functional groups in common.

The comparison heuristic consists first in matching compatible triplets according to the relative position and orientation of the chemical groups. Then each pair (A,A') of matched triplets is considered to belong to the same matched site as the matched triplets (B,B'), if A and B are neighbors in the first protein, and if A' and B' are connected similarly in the other protein. This results in an arbitrary number of sites of arbitrary size.

The final score is simply given by the size of the matched sites, indicating whether a match is partial or not.

3 MAJOR STEPS OF A SESSION

3.1 Target structure

The target 3D structure of protein has to be specified. A selection within this structure may be performed: the whole protein can be selected or only some chains or some sites which are bound to ligands. A non-interactive interface using a simple query language (SuMoQ) is also proposed to advanced users and allows more flexibility in the queries—including arbitrary selection in the 3D structure—and an easy automation of the submission process.

3.2 Target database

The next step is to select the database to scan. The first database available consists of all the PDB structures in which only redundant chains have been removed, while keeping the chain–chain boundaries intact. The second database consists of the ligand binding sites that are found in the former database. Unlike in many bioinformatic servers, the variety of structures of proteins of identical or very close sequences is preserved, in order to represent a large panel of conformational variants of the proteins under different conditions.

3.3 Global results

The results of comparing the selected structure against the database are displayed as a list of potentially interesting similarities

*To whom correspondence should be addressed at The Burnham Institute, 10901 North Torrey Pines Road, La Jolla, CA 92037, USA

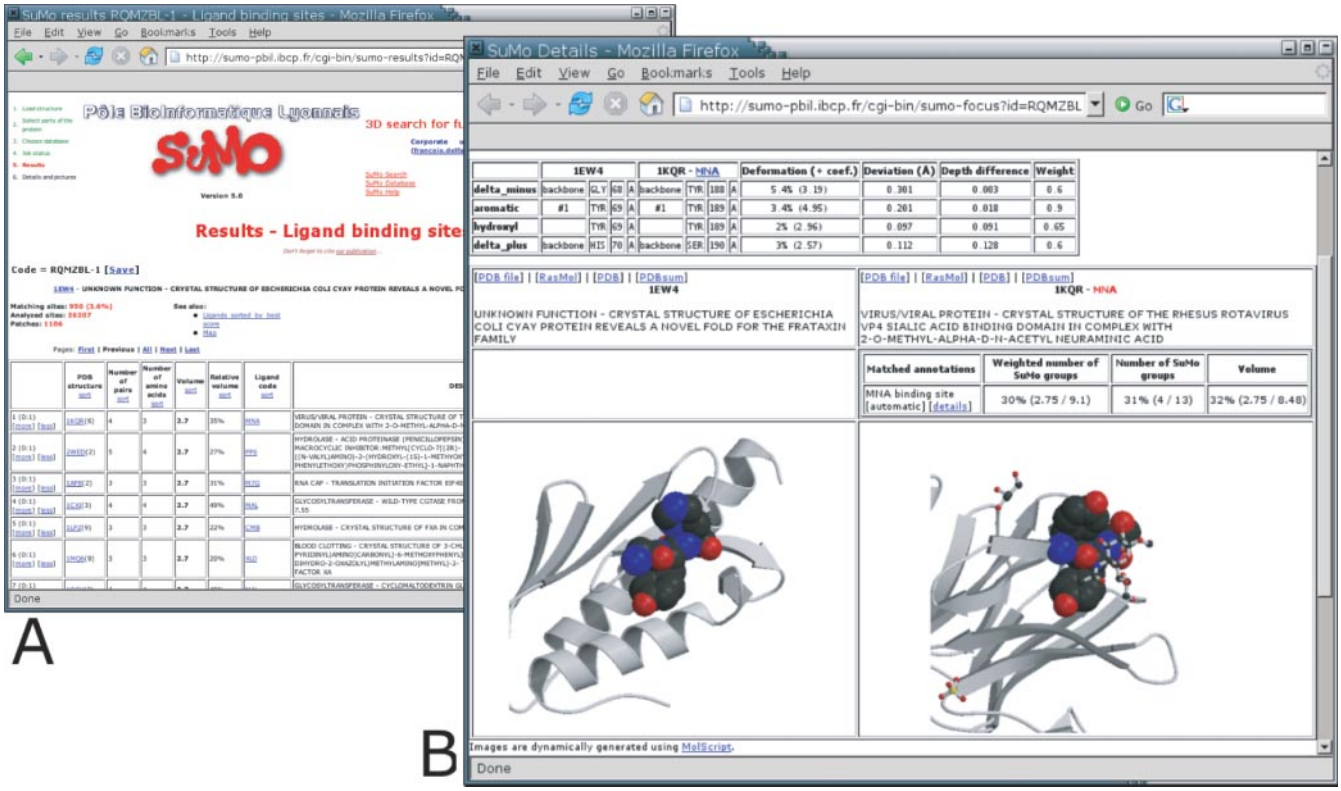

Fig. 1. Prediction of ligand binding sites with SuMo. A crystal structure of a protein of new fold and unknown function was compared with each site of the database of ligand binding sites. (A) Results displayed as a list of matched sites. (B) Detailed result showing the equivalent chemical groups in the first hit.

(Fig. 1A). According to the specific problem being dealt with, results can be sorted according to different criteria, such as the volume of the matched sites, the number of matched pairs of chemical groups or the number of amino acids involved. The query specification and the results can be saved or exported as text for further analysis.

In the case of scanning the database of ligand binding sites, results are also summarized as a mapping of the potential ligand binding chemical groups in the query structure and as a list of potential ligands sorted by maximal score (volume of the site).

3.4 Detailed results

For each pair of matched 3D sites, detailed information is given (Fig. 1B): description of the chemical groups that are matched, parallel view of both sites in the same orientation, direct view in RasMol, links to other resources. Also, various numerical parameters are given as a support for extended analyses of the results.

ACKNOWLEDGEMENTS

We thank Iddo Friedberg for improving this manuscript. This work was supported by the CNRS and the French Ministère de la Recherche. MolScript (Kraulis, 1991) is being used in the process of producing images of molecules by the server.

Conflict of Interest: none declared.

REFERENCES

Ausiello,G. *et al*. (2005) pdbFun: mass selection and fast comparison of annotated PDB residues. *Nucleic Acids Res*., **33** (Web Server issue), W133–W137.

Barker,J.A. and Thornton,J.M. (2003) An algorithm for constraint-based structural template matching: application to 3D templates with statistical analysis. *Bioinformatics*, **19**, 1644–1649.

Ivanisenko,V.A. *et al*. (2004) PDBSiteScan: a program for searching for active, binding and posttranslational modification sites in the 3D structures of proteins. *Nucleic Acids Res*., **32** (Web Server issue), W549–W554.

Jambon,M. *A bioinformatic system for searching functional similarities in 3D structures of proteins*. Université Claude Bernard Lyon 1, Ph.D thesis.

Jambon,M. *et al*. (2003) A new bioinformatic approach to detect common 3D sites in protein structures. *Proteins*, **52**, 137–145.

Jones,S. and Thornton,J.M. (2004) Searching for functional sites in protein structures. *Curr. Opin. Chem. Biol*., **8**, 3–7.

Kleywegt,G.J. (1999) Recognition of spatial motifs in protein structures. *J. Mol. Biol*., **285**, 1887–1897.

Kraulis,P.J. (1991) Molscript: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr*., **24**, 946–950.

Russell,R.B. (1998) Detection of protein three-dimensional side-chain patterns: new examples of convergent evolution. *J. Mol. Biol*., **279**, 1211–1227.

Shulman-Peleg,A. *et al*. (2004) Recognition of functional sites in protein structures. *J. Mol. Biol*., **339**, 607–633.