BIOINFORMATICS APPLICATIONS NOTE Vol. 20 no. 13 2004, pages 2145–2147
doi:10.1093/bioinformatics/bth203

STING Contacts: a web-based application for identification and analysis of amino acid contacts within protein structure and across protein interfaces

A. L. Mancini[1], R. H. Higa[1], A. Oliveira[2], F. Dominiquini[1], P. R. Kuser[1], M. E. B. Yamagishi[1], R. C. Togawa[2] and G. Neshich[1,*]

[1]Núcleo de Bioinformática, Centro Nacional de Pesquisa Agropecuária, Empresa Brasileira de Pesquisa Agropecuária, Campinas, SP, Brazil and [2]Laboratório de Bioinformática, Embrapa/Recursos Genéticos e Biotecnologia

Received on September 25, 2003; revised on January 27, 2004; accepted on February 5, 2004
Advance Access publication April 8, 2004

ABSTRACT

Amino acid contacts in terms of atomic interactions are essential factors to be considered in the analysis of the structure of a protein and its complexes. Consequently, molecular biologists do require specific tools for the identification and visualization of all such contacts. Graphical contacts (GC) and interface forming residue graphical contacts (IFRgc) presented here, calculate atomic contacts among amino acids based on a table of predefined pairs of the atom types and their distances, and then display them using number of different forms. The inventory of currently listed contact types by GC and IFRgc include hydrogen bonds (in nine different flavors), hydrophobic interactions, charge–charge interactions, aromatic stacking and disulfide bonds. Such extensive catalog of the interactions, representing the forces that govern protein folding, stability and binding, is the key feature of these two applications. GC and IFRgc are part of STING Millennium Suite.

Availability: http://sms.cbi.cnptia.embrapa.br/SMS, http://trantor.bioc.columbia.edu/SMS, http://mirrors.rcsb.org//SMS, http://www.es.embnet.org/SMS and http://www.ar.embnet.org/SMS (Options: Graphical Contacts and IFR Graphical Contacts).

Contact: neshich@cbi.cnptia.embrapa.br

Amino acid interactions are indispensable structural elements crucially contributing to the protein stability and binding specificity. Graphical contacts (GC) and interface forming residue graphical contacts (IFRgc) are integrated tools for the identification and graphical visualization of such contacts. In addition, GC and IFRgc present a complete table of contacts, together with the accompanying descriptors of the structure and sequence for amino acids to which the identified contacting pair of atoms belong. Reported contacts might be established either within the given protein chain or across the protein interfaces.

Together with the other Sting Millennium Suite (SMS) components [Scorpion and Formiga in particular Neshich *et al*. (2003)], GC and IFRgc offer a very rich environment for deciphering mechanisms responsible for the observed protein stability and binding specificity.

A method which we have adopted for the identification of all POTENTIAL intra- and inter-protein chain contacts, consists of (a) classifying the atoms in groups according to both their electrostatic behavior and position in the amino acid (main or side chain); and (b) the atoms are then further selected based on the type of the contacts they potentially can make and also based on the experimentally defined distance restrictions (Harris and Mildvan, 1999; Sobolev *et al*., 1999; Swindells, 1995). The user can customize the default distance parameters used by GC and IFRgc.

The resulting list of possible pairs of interacting atoms is integrated to GC and IFRgc and can be examined using the GC graphical interface. The type and number of contacts established by the interacting atoms are listed in the HTML formatted table (Fig. 1C). Also, all contacts can be mapped in the three-dimensional (3D) structure representation within the SMS 3D (chime) window. Special attention is given to the inter-chain interactions as they represent the essential factor defining the protein binding and specificity.

The inclusion of inter-chain interactions is one of the outstanding features of GC and IFRgc. Indeed they are among the few products available, if any, with such a degree of

*To whom correspondence should be addressed.

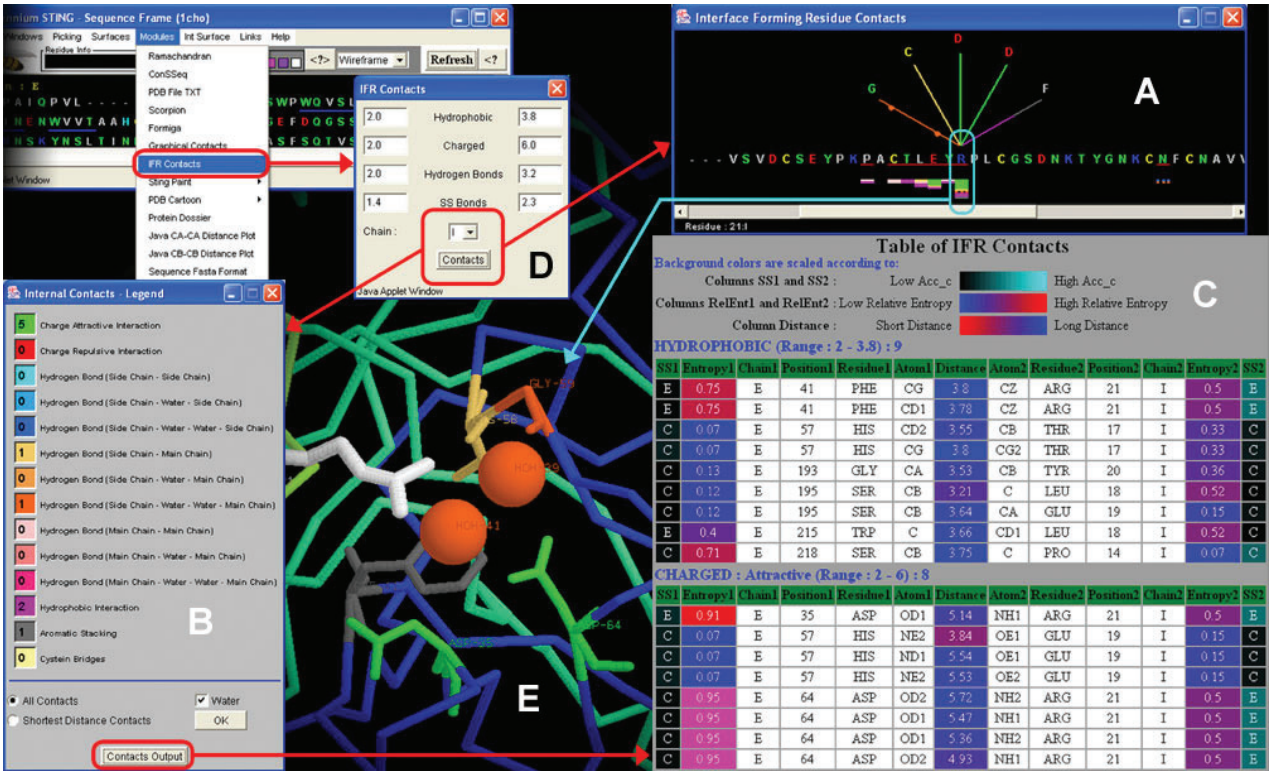

Fig. 1. Output generated by IFRgc within the SMS environment: the PDB entry used is 1cho.pdb. The complete profile shows (**A**) a primary sequence of the 'I' chain and below it a histogram containing the number of IFR contacts made by the respective residue. Such contacts are divided in number of classes: hydrogen bonds (main chain/main chain, side chain/main chain, side chain/side chain), hydrogen bonds with intermediary water molecules, hydrophobic contacts, aromatic ring stacking contacts, electrostatic (attractive and repulsive) contacts and finally, disulphide bridges. (**B**) To each contact type, a specific color is attributed (both in the histogram and in virtual contact lines presentation). Virtual contact lines connect respective residues in a fan fashion. (**C**) The IFR contacts are presented in a tabular form with the color codes used to identify the accessibility [as calculated by the surfv program, Sridharan *et al*. (1992)] and entropy [as found in HSSP, Sander and Schneider (1991)] for each amino acid. The secondary structure element of each residue involved in the contact, as well as the distance between contacting atoms, is also shown. (**D**) The default distance restraints might be modified by the user. (**E**) 3D constellation of contacting residues (painted according to contact color code) can be examined in details in the SMS 3D window with the base residue always colored in white color and the participating water molecules shown in CPK.

completeness in terms of the types of contacts listed and with such graphic quality.

In order to use GC and/or IFRgc, the user needs to supply a four-character PDB-ID (Berman *et al*., 2000). GC and IFRgc are implemented using C++, Perl and Java.

FUTURE DEVELOPMENTS

GC and IFRgc will be including more information starting from the new SMS release, now named Gold STING. Namely, our laboratory is making the final tests on the module that identifies the contacts between the protein and DNA chains.

ACKNOWLEDGEMENTS

Work supported by FAPESP, Fundação de Amparo a Pesquisa do Estado de São Paulo—Project no. 1945/01, CNPq, Conselho Nacional de Desenvolvimento Científico e Tecnológico—Project no. 521093/2001–5 (NV) and FINEP, Financiadora de Estudos e Projetos—Project no. 01/08895–0.

REFERENCES

Berman,H.M., Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H., Shindyalov,I.N. and Bourne,P.E. (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242.

Harris,T.K. and Mildvan,A.S. (1999) High-Precision Measurement of Hydrogen Bond Lengths in Proteins by Nuclear Magnetic Resonance Methods. *Prot. Struc. Func. Genet.*, **35**, 275–282.

Neshich,G., Togawa,R.C., Mancini,A.L., Kuser,P.R., Yamagishi,M.E.B., Pappas,G. Jr., Torres,W.V., Campos,T.F., Ferreira,L.L., Luna,F.M., *et al*. (2003) STING millennium: a web based suite of programs for comprehensive and simultaneous analysis of protein structure and sequence. *Nucleic Acids Res.*, **31**, 3386–3392.

Sander,C. and Schneider,R. (1991) Database of homology-derived protein structures and the structural meaning of sequence alignment. *Prot. Struc. Func. Genet.*, **9**, 56–68.

Sobolev,V., Sorokine,A., Prilusky,J., Abola,E.E. and Edelman,M. (1999) Automated analysis of interatomic contacts in proteins. *Bioinformatics*, **15**, 327–332.

Sridharan,S., Nicholls,A. and Honig,B. (1992) A new vertex algorithm to calculate solvent accessible surface areas. *Biophys. J.*, **61**, 174a.

Swindells,M.B., (1995) A procedure for the automatic determination of hydrophobic cores in protein structures. *Protein Sci.*, **4**, 93–102.