

# SPACE: a suite of tools for protein structure prediction and analysis based on complementarity and environment

Vladimir Sobolev\*, Eran Eyal, Sergey Gerzon, Vladimir Potapov, Mariana Babor, Jaime Prilusky<sup>1</sup> and Marvin Edelman

Department of Plant Sciences and <sup>1</sup>Department of Biological Services, Weizmann Institute of Science, Rehovot 76100, Israel

Received February 13, 2005; Revised and Accepted March 14, 2005

## ABSTRACT

**We describe a suite of SPACE tools for analysis and prediction of structures of biomolecules and their complexes. LPC/CSU software provides a common definition of inter-atomic contacts and complementarity of contacting surfaces to analyze protein structure and complexes. In the current version of LPC/CSU, analyses of water molecules and nucleic acids have been added, together with improved and expanded visualization options using Chime or Java based Jmol. The SPACE suite includes servers and programs for: structural analysis of point mutations (MutaProt); side chain modeling based on surface complementarity (SCCOMP); building a crystal environment and analysis of crystal contacts (CryCo); construction and analysis of protein contact maps (CMA) and molecular docking software (LIGIN). The SPACE suite is accessed at <http://ligin.weizmann.ac.il/space>.**

## INTRODUCTION

The Protein Data Base (PDB) (1,2) has become a major source for the analysis of biological processes at the molecular level, and allows analysis of interactions in proteins and their complexes. A number of web-based servers, including our own [LPC/CSU (3)], provide information on inter- and intramolecular contacts in proteins [PDBsum (4,5), GRASS (6), Relibase (7), RankViaContact (8), STING CONTACTS (9) and Monster (10)]. The LPC/CSU approach differs from others mainly in the definition of contacting atoms (11) and in the provision of a more detailed description of contacts founded on an atom classification (12). Atoms are considered to be in contact with one another based on inter-atomic distances and the extent of crowding in the environment. For example, in

non-packed regions two atoms could be listed as hydrogen bonded at distances up to 5 Å (assuming water mediation), while in packed regions they would not. In addition, a measure of contact surface area is provided. As a result, the LPC/CSU approach was applied not only for detailed structure analysis (13–15) but also for the derivation and application of knowledge-based functions to the protein folding problem (16–19) and to molecular docking (20,21).

In this communication, we describe a suite of SPACE tools designed to assist in the analysis and prediction of biomolecular structures and their complexes. A shared feature of all SPACE tools is the application of the LPC/CSU definition for inter-atomic contacts and surface complementarity. Inter-atomic contacts are calculated either numerically (11) or analytically (22). Complementarity is estimated based on the deviation of atoms into eight classes according to their physicochemical properties (12).

## SPACE WEB TOOLS

### LPC/CSU: contact analysis of biomolecules

The LPC/CSU server in its current version analyzes and visualizes (either with Chime plug-in or Java based Jmol) atomic interactions within a protein or protein complex, including resolved water molecules and attached ligands, and nucleic acids. Different levels of analysis can be chosen: contacts can be grouped and sorted by atom, residue or contact type (H-bond, hydrophobic–hydrophobic and aromatic–aromatic). The output provides characteristics for every atom–atom contact (atom properties, distance and contact area). A typical output is illustrated in Figure 1.

### CryCo: analysis of crystal contacts

The CryCo server builds coordinate files in a PDB format for the unit cell as well as for the complete crystal environment

\*To whom correspondence should be addressed. Tel: +972 8 9344434; Fax: +972 8 9469124; Email: [vladimir.sobolev@weizmann.ac.il](mailto:vladimir.sobolev@weizmann.ac.il)

**MENU****CONTACTS GROUPED BY**

- [Residue atoms](#)
- [Residues](#)
- [Contact types](#)

**CONTACTS SORTED BY**

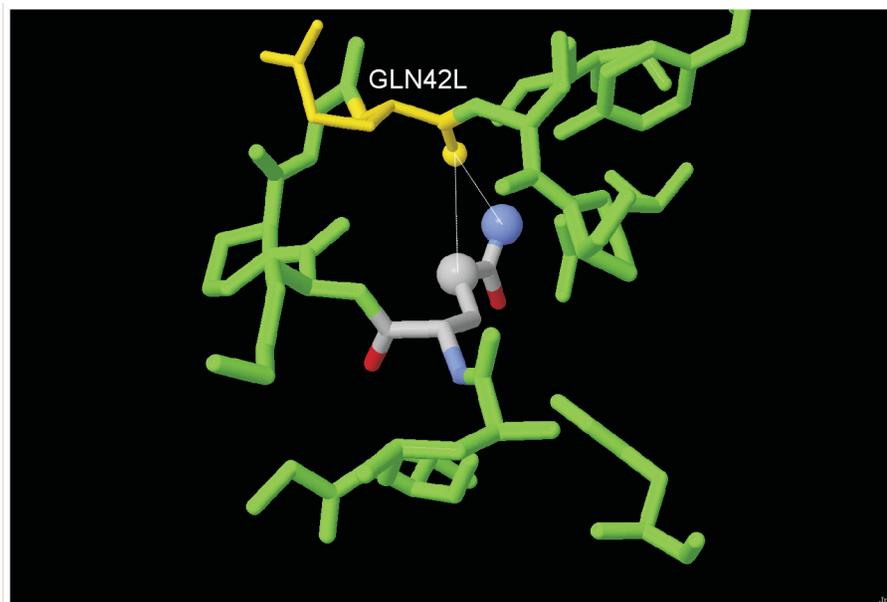
- [Residue atoms](#)
- [Residues](#)
- [Contact types](#)

**GO TO**

- [Another residue](#)
- [Another PDB entry](#)

**HELP**

- [Description approach](#)

**RESET**

List of residues forming contacts with the Residue GLN 38 (chain L) in PDB entry 1DLF

Select	Add	Residue		Dist(Å)	Surf(Å <sup>2</sup> )	Number of contacts
<input type="button" value="SELECT"/>	<input type="button" value="ADD"/>	LEU	37 L	1.3	76.1	9
<input type="button" value="SELECT"/>	<input type="button" value="ADD"/>	LYS	39 L	1.3	68.7	11
<input type="button" value="SELECT"/>	<input type="button" value="ADD"/>	PRO	40 L	4.7	0.3	1
<input type="button" value="SELECT"/>	<input type="button" value="ADD"/>	GLY	41 L	5.9	0.4	2
<input type="button" value="SELECT"/>	<input type="button" value="ADD"/>	GLN	42 L	2.8	30.0	2
<input type="button" value="SELECT"/>	<input type="button" value="ADD"/>	SER	43 L	4.1	5.4	2

**Figure 1.** Graphical output of the CSU program. Interactions of residue Gln38L from pdb entry 1DLF (anti-dansyl Fv fragment) are illustrated. All residues and water oxygens in contact with Gln38L are evoked by clicking on 'contacts grouped by residues'. Pressing 'Select' for Gln42L in the bottom frame results in labeling and highlighting (in yellow) of the residue selected. The visualization screen shows that only two contacts are formed and one of them is a H-bond to the backbone oxygen of Gln42L.

of one molecule. The structural environment is built in several steps. First, symmetry related molecules are created using the PDBSET program from the CCP4 suite (23). When necessary, all molecules are translocated to one unit cell. The 26 adjacent cells in the crystal lattice are then constructed by translation and finally any atom farther than a chosen threshold from the closest atom of the central molecule is removed. Detailed analyses of atomic contacts are based on CSU software. Interactive visualization options and coordinate output files are provided, and also an option to submit a structural file for analysis. An example of the output from the CryCo server is provided in Figure 2. CryCo differs from existing tools such as the WHAT IF web server (24) and the xpack VRML-based program (25) in providing visualization options, detailed contact analyses and several files with new features for downloading.

**CMA: contact map analysis**

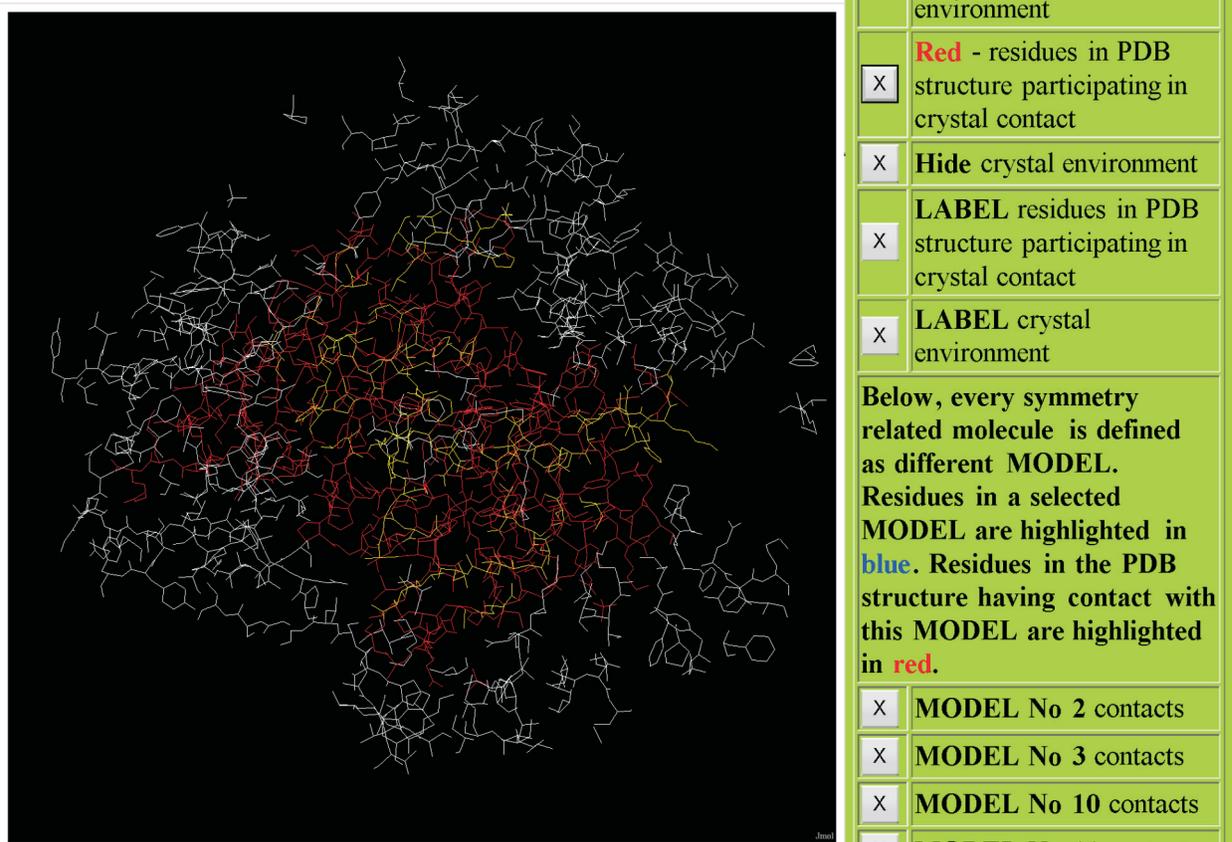
For a given PDB file, the 'Contact Map Analysis' server (CMA) evaluates residue-residue contacts between two chains or within a single one. In the example illustrated in Figure 3, the interface contacts between chains L and H in PDB file 1DLF are considered. Residue-residue contacts are

represented as an interactive contact map, where a square at the crossing of two residues indicates a contact (Figure 3a). Positioning the cursor over the square highlights summary information about the contacting residues and their total contact area. Clicking on the square reveals a table with more detailed contact information based on LPC/CSU software, including names of the contacting atoms, distances and atom-atom contact areas (Figure 3b). Links for analysis and visualization of contact residues are likewise provided (Figure 3c). The CMA server was extensively used to analyze inter domain contacts in sandwich-like proteins (26). It differs from existing servers, such as WebMol (27), iMolTalk (28) and Stride (29), by providing detailed visualization and detailed contact analysis.

**MutaProt: structural rearrangements upon point mutations**

MutaProt contains a database of pairs of PDB files whose members differ in one or two amino acids (30). The software examines the microenvironment of the mutated residues. The database is accessed by specifying a PDB file, keyword or a pair of amino acids. Accessibility and atomic contacts of the mutated residue are provided by CSU software.

## Crystal Contact in PDB entry 1DLF



**Figure 2.** Graphical output of the CryCo server. The PDB structure under analysis is highlighted in yellow while its crystal environment is in white. Residues of the structure in crystal contact have been highlighted in red by clicking on the second panel. The atomic coordinates for every molecule in contact with the PDB structure are formatted in the output file as a different MODEL. Contacts formed by every such molecule can be visualized separately.

The current version of the server has a number of significant improvements. MutaProt now extracts pairs based on differences at the chain level. This dramatically increases the database to ~200 000 pairs. Wild-type structures are distinguished from mutant ones where information is available. An option has been included for user submission of a structural file for pairing up with PDB entries and MutaProt analysis. The interactive graphics have been expanded to include the entire PDB structure and presentation of the protein sequence is included along with secondary structure assignment based on DSSP (31). In addition, superposition of the two pair members is now done analytically (32). A list of publicly available mutation databases is provided. MutaProt is unique in providing detailed on line analysis of atomic contacts and offering a superimposed 3D presentation of regions being compared.

### SCCOMP: side chain modeling

SCCOMP is a server for side chain modeling. It uses a scoring function (33) that includes terms for complementarity (CSU definitions of geometric and chemical compatibility), excluded volume, internal energy based on rotamer probability

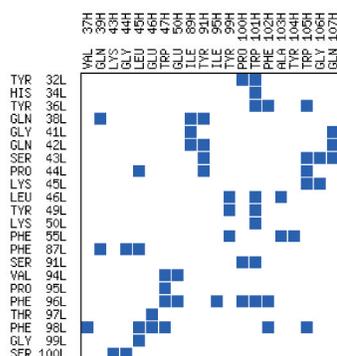
and solvent accessible surface. The input for the program is a coordinate file in the PDB format with or without side chain coordinates. The output is the file with predicted coordinates for the side chains. The program has an accuracy of ~93% for  $\chi_1$  prediction ( $\pm 40^\circ$ ) of buried residues, ~71% for exposed residues, ~83% for all residues and an overall RMSD of 1.7 Å (not including  $C_\beta$ ). A fast iterative search takes ~1 min for a typical protein; the slower stochastic search takes ~12 min and improves prediction by ~2% and 0.1 Å RMSD. SCCOMP permits modeling a subset of residues, introducing any number of mutations, and using homologous structures as templates. It complements another publicly available server (<http://www1.jcsg.org/scripts/prod/scwrl/serve.cgi>) uses a less sophisticated scoring function (34). Although our program is slower, it more accurately predicts  $\chi_{1+2}$  and returns a lower RMSD for the overall structure. Furthermore, SCCOMP is convenient for performing *in silico* mutagenesis.

### SPACE PROGRAMS

The SPACE suite provides an option to download a number of programs. Source codes for the LIGIN (12) (molecular

**(a) Contact Map**

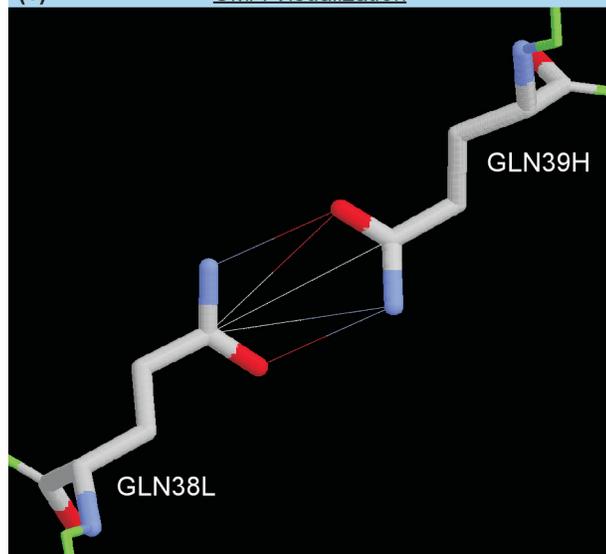
Interface contacts between chains L and H in PDB entry 1DLF  
(click at any square to see detailed contact information)

**(b) CMA Analysis**

Interface contacts between residues GLN 38L and GLN 39H  
in PDB entry 1DLF

GLN 38L	GLN 39H	Distance	Cont. area
CD	OE1	3.7	0.7
CD	NE2	3.8	0.4
CD	CD	4.2	0.2
OE1	NE2	3.0	19.4
NE2	OE1	2.9	21.2

Full list of atomic contacts [[GLN 38L](#)] [[GLN 39H](#)]  
[Visualise contact residues](#)

**(c) CMA Visualization**

**Figure 3.** Server for construction and analysis of protein contact maps. (a) Contact map (b) CMA analysis and (c) CMA visualization.

docking) and SCCOMP (33) (side chain modeling) programs are available at the SPACE website. To enable the analysis of a large number of PDB files, LPC and CSU programs with output as simple text files are also provided.

**ACKNOWLEDGEMENTS**

Funding to pay the Open Access publication charges for this article was waived by Oxford University Press.

*Conflict of interest statement.* None declared.

**REFERENCES**

- Bernstein,F.C., Koetzle,T.F., Williams,G.J.B., Meyer,E.F., Jr, Brice,M.D., Rodgers,J.R., Kennard,O., Shimanouchi,T. and Tasumi,M. (1977) The Protein Data Bank: a computer based archival file for macromolecular structures. *J. Mol. Biol.*, **112**, 535–542.
- 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.
- Sobolev,V., Sorokine,A., Prilusky,J., Abola,E.E. and Edelman,M. (1999) Automated analysis of interatomic contacts in proteins. *Bioinformatics*, **15**, 327–332.
- Laskowski,R.A., Hutchinson,E.G., Michie,A.D., Wallace,A.C., Jones,M.L. and Thornton,J.M. (1997) PDBsum: a Web-based database of summaries and analyses of all PDB structures. *Trends Biochem. Sci.*, **22**, 488–490.
- Laskowski,R.A., Chistyakov,V.V. and Thornton,J.M. (2005) PDBsum more: new summaries and analyses of the known 3D structures of proteins and nucleic acids. *Nucleic Acids Res.*, **33**, D266–D268.
- Nayal,M., Hitz,B.C. and Honig,B. (1999) GRASS: A server for the graphical representation and analysis of structures. *Protein Sci.*, **8**, 676–679.
- Hendlich,M., Bergner,A., Gunther,J. and Klebe,G. (2003) Relibase: design and development of a database for comprehensive analysis of protein–ligand interactions. *J. Mol. Biol.*, **326**, 607–620.
- Shen,B.R. and Vihinen,M. (2003) RankViaContact: ranking and visualization of amino acid contacts. *Bioinformatics*, **19**, 2161–2162.
- Mancini,A.L., Higa,R.H., Oliveira,A., Dominiquini,F., Kuser,P.R., Yamagishi,M.E.B., Togawa,R.C. and Neshich,G. (2004) STING Contacts: a web-based application for identification and analysis of amino acid contacts within protein structure and across protein interfaces. *Bioinformatics*, **20**, 2145–2147.
- Salerno,W.J., Seaver,S.M., Armstrong,B.R. and Radhakrishnan,I. (2004) MONSTER: inferring non-covalent interactions in macromolecular structures from atomic coordinate data. *Nucleic Acids Res.*, **32**, W566–W568.
- Sobolev,V. and Edelman,M. (1995) Modeling the quinone-B binding site of the photosystem-II reaction center using notions of complementarity and contact-surface between atoms. *Proteins*, **21**, 214–225.
- Sobolev,V., Wade,R.C., Vriend,G. and Edelman,M. (1996) Molecular docking using surface complementarity. *Proteins*, **25**, 120–129.
- Amitai,G., Shemesh,A., Sitbon,E., Shklar,M., Netanel,D., Venger,I. and Pietrokovski,S. (2004) Network analysis of protein structures identifies functional residues. *J. Mol. Biol.*, **344**, 1135–1146.
- Swint-Kruse,L. (2004) Using networks to identify fine structural differences between functionally distinct protein states. *Biochemistry*, **43**, 10886–10895.
- Reichmann,D., Rahat,O., Albeck,S., Meged,R., Dym,O. and Schreiber,G. (2005) The modular architecture of protein–protein binding interfaces. *Proc. Natl Acad. Sci. USA*, **102**, 57–62.
- McConkey,B.J., Sobolev,V. and Edelman,M. (2003) Discrimination of native protein structures using atom-atom contact scoring. *Proc. Natl Acad. Sci. USA*, **100**, 3215–3220.
- Kaya,H. and Chan,H.S. (2003) Solvation effects and driving forces for protein thermodynamics and kinetic cooperativity: how adequate is native-centric topological modeling? *J. Mol. Biol.*, **326**, 911–931.
- Yang,S.C., Cho,S.S., Levy,Y., Cheung,M.S., Levine,H., Wolynes,P.G. and Onuchic,J.N. (2004) Domain swapping is a consequence of minimal frustration. *Proc. Natl Acad. Sci. USA*, **101**, 13786–13791.
- Ollerenshaw,J.E., Kaya,H., Chan,H.S. and Kay,L.E. (2004) Sparsely populated folding intermediates of the Fyn SH3 domain: matching native-centric essential dynamics and experiment. *Proc. Natl Acad. Sci. USA*, **101**, 14748–14753.

20. Sobolev, V., Niztaev, A., Pick, U., Avni, A. and Edelman, M. (2002) A case study in applying docking prediction: modeling the tentoxin binding sites of chloroplast F1-ATPase. *Curr. Sci.*, **83**, 857–867.
21. Lloyd, D.G., Hughes, R.B., Zisterer, D.M., Williams, D.C., Fattorusso, C., Catalonotti, B., Campiani, G. and Meegan, M.J. (2004) Benzoxepin-derived estrogen receptor modulators: a novel molecular scaffold for the estrogen receptor. *J. Med. Chem.*, **47**, 5612–5615.
22. McConkey, B.J., Sobolev, V. and Edelman, M. (2002) Quantification of protein surface, volumes and atom–atom contacts using a constrained Voronoi procedure. *Bioinformatics*, **18**, 1365–1373.
23. Collaborative Computational Project, Number 4 (1994), The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.*, **50**, 760–763.
24. Rodriguez, R., China, G., Lopez, N., Pons, T. and Vriend, G. (1998) Homology modeling, model and software evaluation: three related resources. *Bioinformatics*, **14**, 523–528.
25. Fu, T.Y. and Chen, Y.W. (1996) Visualization of macromolecular crystal packing using Virtual Reality Modelling Language (VRML). *J. Appl. Crystallogr.*, **29**, 594–597.
26. Potapov, V., Sobolev, V., Edelman, M., Kister, A. and Gelfand, I. (2004) Protein–protein recognition: juxtaposition of domain and interface cores in immunoglobulins and other sandwich-like proteins. *J. Mol. Biol.*, **342**, 665–679.
27. Walther, D. (1997) WebMol—a Java based PDB viewer. *Trends Biochem. Sci.*, **22**, 274–275.
28. Diemand, A.V. and Scheib, H. (2004) iMolTalk: an interactive, internet-based protein structure analysis server. *Nucleic Acids Res.*, **32**, W512–W516.
29. Frishman, D. and Argos, P. (1995) Knowledge-based protein secondary structure assignment. *Proteins*, **23**, 566–579.
30. Eyal, E., Najmanovich, R., Sobolev, V. and Edelman, M. (2001) MutaProt: a web interface for structural analysis of point mutations. *Bioinformatics*, **17**, 381–382.
31. Kabasch, W. and Sander, C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577–2637.
32. Arun, K.S., Huang, T.S. and Blostein, S.D. (1987) Least-squares fitting of two 3-D point sets. *IEEE Trans. Pattern Anal. Mach. Intel.*, **9**, 698–700.
33. Eyal, E., Najmanovich, R., McConkey, B.J., Edelman, M. and Sobolev, V. (2004) Importance of solvent accessibility and contact surfaces in modeling side-chain conformations in proteins. *J. Comput. Chem.*, **25**, 712–724.
34. Canutescu, A.A., Shelenkov, A.A. and Dunbrack, R.L., Jr (2003) A graph-theory algorithm for rapid protein side-chain prediction. *Protein Sci.*, **12**, 2001–2014.