

This Provisional PDF corresponds to the article as it appeared upon acceptance. The fully-formatted PDF version will become available shortly after the date of publication, from the URL listed below.

## **SCOWLP: a web-based database for detailed characterization and visualization of protein interfaces.**

*BMC Bioinformatics* 2006, 7:104 doi:10.1186/1471-2105-7-104

Joan Teyra ([joan.teyra@biotec.tu-dresden.de](mailto:joan.teyra@biotec.tu-dresden.de))  
Andreas Doms ([andreas.doms@biotec.tu-dresden.de](mailto:andreas.doms@biotec.tu-dresden.de))  
Michael Schroeder ([michael.schroeder@biotec.tu-dresden.de](mailto:michael.schroeder@biotec.tu-dresden.de))  
M. Teresa Pisabarro ([mayte.pisabarro@biotec.tu-dresden.de](mailto:mayte.pisabarro@biotec.tu-dresden.de))

**ISSN** 1471-2105

**Article type** Database

**Submission date** 11 Nov 2005

**Acceptance date** 2 Mar 2006

**Publication date** 2 Mar 2006

**Article URL** <http://www.biomedcentral.com/1471-2105/7/104>

Like all articles in BMC journals, this peer-reviewed article was published immediately upon acceptance. It can be downloaded, printed and distributed freely for any purposes (see copyright notice below).

Articles in BMC journals are listed in PubMed and archived at PubMed Central.

For information about publishing your research in BMC journals or any BioMed Central journal, go to

<http://www.biomedcentral.com/info/authors/>

***SCOWLP: a web-based database for detailed characterization and visualization of protein interfaces.***

*Joan Teyra*<sup>§</sup>, *Andreas Doms*, *Michael Schroeder* and *M. Teresa Pisabarro*<sup>§</sup>

Department of Bioinformatics, BIOTEC TU Dresden, Tatzberg 47-51, 01307  
Dresden, Germany

<sup>§</sup>Corresponding authors

Joan Teyra ([joan.teyra@biotec.tu-dresden.de](mailto:joan.teyra@biotec.tu-dresden.de))

Andreas Doms ([andreas.doms@biotec.tu-dresden.de](mailto:andreas.doms@biotec.tu-dresden.de))

Michael Schroeder ([michael.schroeder@biotec.tu-dresden.de](mailto:michael.schroeder@biotec.tu-dresden.de))

M. Teresa Pisabarro ([mayte.pisabarro@biotec.tu-dresden.de](mailto:mayte.pisabarro@biotec.tu-dresden.de))

## **Abstract**

### **Background**

Currently there is a strong need for methods that help to obtain an accurate description of protein interfaces in order to be able to understand the principles that govern molecular recognition and protein function. Many of the recent efforts to computationally identify and characterize protein networks extract protein interaction information at atomic resolution from the PDB. However, they pay none or little attention to small protein ligands and solvent. They are key components and mediators of protein interactions and fundamental for a complete description of protein interfaces. Interactome profiling requires the development of computational tools to extract and analyze protein-protein, protein-ligand and detailed solvent interaction information from the PDB in an automatic and comparative fashion. Adding this information to the existing one on protein-protein interactions will allow us to better understand protein interaction networks and protein function.

### **Description**

SCOWLP (Structural Characterization Of Water, Ligands and Proteins) is a user-friendly and publicly accessible web-based relational database for detailed characterization and visualization of the PDB protein interfaces. The SCOWLP database includes proteins, peptidic-ligands and interface water molecules as descriptors of protein interfaces. It contains currently 74,907 protein interfaces and 2,093,976 residue-residue interactions formed by 60,664 structural units (protein domains and peptidic-ligands) and their interacting solvent.

The SCOWLP web-server allows detailed structural analysis and comparisons of protein interfaces at atomic level by text query of PDB codes and/or by navigating a SCOP-based tree. It includes a visualization tool to interactively display the interfaces and label interacting residues and interface solvent by atomic physicochemical properties. SCOWLP is automatically updated with every SCOP release.

### **Conclusions**

SCOWLP enriches substantially the description of protein interfaces by adding detailed interface information of peptidic-ligands and solvent to the existing protein-protein interaction databases. SCOWLP may be of interest to many structural bioinformaticians. It provides a platform for automatic global mapping of protein interfaces at atomic level, representing a useful tool for classification of protein interfaces, protein binding comparative studies, reconstruction of protein complexes and understanding protein networks. The web-server with the database and its additional summary tables used for our analysis are available at <http://www.scowlp.org>.

## Background

One of the most interesting and important challenges in the so-called “Post-genomic Era” is the understanding of protein networks. Protein-protein interactions have been extensively investigated using a variety of methods [1], and many databases have been built becoming very helpful tools for the analysis of protein networks [2-4].

Protein interfaces have long been studied at protein chain and domain interface levels [5-12]. Furthermore, numerous analyses have used datasets of protein chain interfaces to investigate residue type propensities, sequence and structure conservation at protein interfaces [8, 11, 13-16]. Databases containing structural domain-domain interactions have also been recently created: 3did [17], PiBase [18], iPfam [19], PSIBase [20], InterPare [21], PRISM [22]. However, in these methods still many protein residues are not taken into account as “interfacial” or “interacting” because of peptidic-ligands and also solvent being frequently ignored from the protein interaction analysis.

Peptidic-ligands and solvent mediate protein interactions and are fundamental components for a complete description of protein interfaces. Proteins can interact with peptides to perform their biological function. Besides, peptides have been used to mimic protein binding interfaces, and their complexes with proteins have been used to study protein binding affinity/specificity properties in a simplified way [23-25]. For these reasons, many protein-peptide complexes have been experimentally studied by X-ray crystallography and/or NMR studies, providing additional information on protein interfaces [25]. Moreover, protein interactions take place in an aqueous solution. Solvent molecules can bridge binding partners via hydrogen bonds contributing significantly to molecular recognition and function [23, 26-31].

Most current methods do not provide an accurate description of protein interfaces, which is required to be able to establish the bases for understanding the principles that govern molecular recognition and protein function.

Here we present SCOWLP (Structural Characterization Of Water, Ligands and Proteins), a platform for complete and detailed characterization and visualization of protein interfaces. Our database includes all protein-interacting components of the PDB including peptides and solvent, which until now have been excluded from systematic protein interface analysis and databases. In our database all interface interactions are described at atom, residue and domain level by using interacting rules based on atomic physicochemical criteria. This complete characterization makes SCOWLP useful for comparative structural analysis of molecular interfaces. The web application allows the user to access all the atomic interaction information by querying the PDB or the SCOP hierarchy. All interface information characterized by different interaction descriptors can be interactively visualized by using a Jmol 3D applet [32].

## Construction and content

SCOWLP is a web-based relational database formed by eleven tables describing PDB interface interactions at atom, residue and domain level. The database contains 74,907 protein interfaces and 2,093,976 residue-residue interactions formed by 60,664 *structural units* and interacting solvent. For the creation of the SCOWLP, we extract 3D data of protein domains, peptidic-ligands and interface solvent from the PDB [33], and we define protein domains from the SCOP 1.69 [34]. We compute protein interactions at atom, residue and domain level by using bounding shape-based

algorithms [35]. We also have developed a web application to handle and navigate through the interfacial data in an automatic and user-friendly fashion. We designed the SCOWLP methodology based on the following steps:

### **SCOL-Ligand (StrUctural Characterization Of Peptidic-Ligands)**

The first step of our methodology consists of creating the SCOL table. Each *structural unit* in a PDB file is represented by a different chain name. We extract all *structural units* of the PDB and compare them with the domain definitions of SCOP. Although SCOP has a “Peptide” class containing functional peptides, it does not contain all peptidic-ligands complexed in the PDB. For this reason, *structural units* bigger than two and smaller than one hundred residues not defined in SCOP are considered *peptidic-ligands*. We stored this information in the SCOL table (Fig. 1). Heteroatoms and modified residues that form part of the same polypeptide chain are included, and DNA residues are excluded. We characterize each SCOL *peptidic-ligand* by resolution, sequence length and secondary structure. SCOWLP contains 2,739 peptidic-ligands, which add 3,413 new interfaces (Fig. 2).

### **Interacting Structural Unit Pairs**

We label all *structural units* of the PDB with the SCOL-peptide and the SCOP-domain definitions in order to compute their interactions. We consider a contact distance cut-off of 9Å between two residues in order to allow up to two bridging water molecules in the shortest axes defining the interface. We use bounding shape-based algorithms to compute a 9Å convex hull (the smallest convex set containing all atoms at 9Å) for each *structural unit* of each PDB entry. Convex hull algorithms have been proved to reduce the computational time required for an interface calculation by both, reducing the search space to decrease the number of residues checked for the calculation and allowing distributed computations [35]. *Structural units* with intersecting shapes and having at least one residue-residue interaction are considered *interacting pairs* (Fig. 1).

### **SCOW-Water (StrUctural Characterization Of Water)**

We consider a water molecule as part of an interface when it is located in the shape intersection of two interacting *structural units*. All interface water molecules are stored in the SCOW table and are then included in the atomic interface computation. We also consider an interaction when two residues are bridging through one or two water molecules. Residue contacts are defined as only water-mediated (*OWM*), non water-mediated or direct (*D*), and mixed (*M*). Residues that only interact through water are defined as *wet spots* (Fig. 3). SCOWLP contains 435,086 new water-mediated interactions thanks to the implementation of SCOW. This represents 20% of the SCOWLP database (Fig. 2).

### **Interaction Rules for Interface Computation**

Only amino acid residues and water molecules placed in the intersection of *structural unit* shapes are potential interactors. We apply atom type and distance criteria to compute interactions between *structural unit* pairs at physicochemical level. For hydrogen bonds we apply a  $\leq 3.2$  Å donor-acceptor distance. For salt bridges, we apply a  $\leq 4$  Å distance criteria. Van der Waals energies are defined by hydrophobic atoms at van der Waals radii distance. At atomic level, we characterize the interactions by: i) nature: hydrophilic, hydrophobic; ii) contact type: main chain, side chain, mixed; iii) number of bridging water molecules. At residue level, we characterize the interactions by: i) nature: hydrophilic, hydrophobic, dual; ii) contact

type: main chain, side chain, mixed; iii) number of bridging water molecules; iv) total number of atoms contacting. At *structural unit* level, we characterize the interactions by: i) contact volume; ii) surface area from convex hull surface; iii) number of interacting atoms/residues per unit; iv) type of interaction: intra-/inter-molecular. All interfacial interaction information is stored in the SCOWLP database (Fig. 1).

### Summary Tables

We have created the following additional tables for the filtering and comparative analysis of the information contained in the database:

*Interface description.* This table summarizes all interfaces of the SCOWLP database. It contains 74,907 interfaces constituted by SCOP domains labelled with the attributes: PDB Id code, atomic resolution, contact type (intra-/inter-molecular) and SCOP Id code. All interfaces are also labelled by number of interactions (total, all water-mediated and only water-mediated) and number of interacting residues per binding partner. Each interaction is classified by type (side-/main-chain or both) and by number of bridging water molecules.

*Wet interfaces selection.* This table stores interfaces of complexes at resolution  $\leq 2.5$  Å from the *Interface description table* for interfacial solvent analysis. This table does not include homodimer interfaces because of their patchy, poorly packed and highly hydrated nature [36]. With the resultant dataset, we create three tables:

- *Content.* This table can be used to rank superfamilies based on their content in water mediating interface interactions. For each interface, it contains the average of total interactions, all water-mediated interactions and the ratio from the percentage of water-mediated interactions at superfamily level.
- *Morphology.* This table can be used to rank the interfaces by number of *wet spots*. In this table each family is represented by the complex with the highest number of *wet spots*, labelled with the total number of interacting residues and *wet spots*.
- *Comparative.* This table can be used to monitor solvent variations in interfaces and compare them at family level. It contains interfaces sorted out by domain, and then by their respective ligands (protein or peptide). Because a protein-ligand interface can be found in different PDBs, we select the interfaces that appear more than once and contain *wet spots*. When the same PDB file contains a repeated interface of two binding partners, we select as a representative the one with more *wet spots*.

### Implementation

We used MySQL and the Java programming language to generate and analyze the SCOWLP database. Interface calculations are performed on a 2.6 GHz Pentium IV in approximately 36 hours. SCOWLP is automatically updated with every SCOP release.

## Utility and Discussion

SCOWLP database contains detailed information of protein interfaces including peptidic-ligands and solvent in the PDBs, and classifies protein interfaces by using specific physicochemical atomic criteria. The database can be accessed through a user-friendly web application.

### Interaction rules

The use of atom type and distance rules allows us to characterize and classify interactions at physicochemical level. Other existing methods adopt exclusively a general distance criterion. PSIMAP [35], for example, considers as an interacting pair

any atom distance at  $\leq 5 \text{ \AA}$ . For this reason, the total number of residue-residue and *structural unit* interactions we obtain by applying our interaction rules is reduced in comparison to PSIMAP (Fig. 2). This reduction translates into more accurate interface definitions.

### Peptidic-ligand contribution

Some proteins have been subject of many structural studies complexed with peptides (e.g. Proteases, b.47.1). Besides, the superfamilies that have the higher occurrence of peptides are not necessarily those with higher domain-domain representation (e.g. Cyclophilin, b.62.1). By taking into account information about protein-peptide complexes SCOWLP contributes interfacial information of 8 SCOP superfamilies uniquely represented by protein-peptide complexes (a23.4, a.50.1, d.76.1, a.8.5, d195.1, g.33.1, a.144.1, a.12.1). In addition, it contributes with more than 50% of the interacting information in other superfamilies. Our results show the importance of including protein-peptide interfacial information in order to enrich considerably the description of protein interfaces.

Proteins can bind to peptides in places that do not exactly correspond to binding sites in their known protein-protein complexes. As an example, we show the BTB/POZ (Poxvirus and Zinc finger) family. The twelve BTB/POZ complexes in the PDB present five domain-binding regions, two of them described by the protein-peptide complexes (Fig. 4A). The POZ-peptide interfacial information is functionally relevant. It may help to propose new POZ contacts when reconstructing multi-protein complexes and modelling signalling pathways where the POZ domain-containing proteins are involved. Our results show that the addition of peptidic information can help to complete the view on how a protein recognizes its binding partners.

### Solvent contribution

All superfamilies of the *Content table* contain solvent mediating interactions. Furthermore, in some of these superfamilies water-mediated interactions represent up to 75% of the total interfacial interactions (e.g. d.250.1). Relating to the “only water-mediated” interactions, we observe from the *Morphology table* that 43 is the maximum number of *wet spots* found. Figures 4B and 4C illustrate how solvent, in particular *wet spots*, may play an important role in the morphological description of protein interfaces (shape and size). Considering the solvent, a discontinuous surface formed by several small isolated patches changes to a bigger and rounded patch. These observations show that we can enrich the description of protein interfaces by considering interfacial solvent.

Although solvent molecules mediating protein interactions can be conserved in a protein family, variations may occur due to different facts: i) atomic resolution and/or quality of the structural data, ii) conformational changes upon ligand binding, iii) protein flexibility, iv) new interacting regions (e.g. loop insertions and deletions), v) residue mimicry. *Wet spots* variations may be used as indicators in these cases. The *Comparative table* allows us to compare the interfaces of 127 families in 751 complexes based on *wet spots* variations.

Solvent molecules play an important role in the replacement of residues in protein interfaces. Sometimes the atomic resolution, the existence of different rotamers or even small differences in contact distances defining the interaction may influence the number of *wet spots*. Nevertheless, small variations of *wet spots* in complexes of the same family that do not present changes in *total number of interactions* can be used to locate residue mimicry cases (e.g. Lys+H<sub>2</sub>O $\approx$ Arg). Making use of this information

may be very useful in analysis of protein interfacial evolution and in protein engineering/rational design when designing affinity and specificity of a protein for its ligands.

### Web application

SCOWLP contains atomic interfacial information of all the PDB entries structured by the SCOP hierarchy. There are two ways to query our database: SCOP or PDB. The user can query SCOP by keywords, SCOP/PDB Ids, or by simply navigating the SCOP hierarchical tree (Fig 5.1). When the user selects a family from the tree (labelled as FA), SCOWLP retrieves a list of the PDBs containing interfaces of that family in one frame. A second frame shows a summary table listing all the interfaces of that family with PDB id, type of contact, superfamily description of binding partners, interfacial area, total interacting residues and number of *wet spots*. This summary table gives a good overview over the interacting partners and interfacial variations at family level. By selecting any of the PDB IDs in this table, the user retrieves a list of all the interfaces of that PDB organized in two interactive tables: *Interfaces* and *Interactions*. We obtain the same tables querying SCOWLP by PDB ID (Fig 5.2). The “Interfaces” table shows binding partners, interfacial area, total number of interfacial residues and *wet spots*. The *Interaction Types* table classifies the interactions based on their water mediation, nature and type. The user can select the interfaces in a master/slave way to display a 3D molecular viewer and the selected domain contacts. We have implemented Jmol scripts [32] to allow the user to display and interactively analyze interfaces by using two control panels (Fig 5.3). The first one (on the right; Fig 5.3a; *Domain Contact Selection*) controls the interface display in the 3D viewer, allowing the user to highlight the residues forming part of each interface. The second panel (bottom left; Fig 5.3.c) controls: *Molecule View*: ON/OFF residue labelling, water mediators and spinning; *Interacting Descriptions*: interfacial residues colouring based on *wet spots*, nature and type. Fig. 5.3 shows a protein domain (red) interacting with a peptidic-ligand (yellow) and their respective interacting residues (*wet spots* in blue).

By using SCOWLP, the user can achieve specific queries, SCOP family analysis, interface comparisons and a detailed 3D display of the atomic interaction data contained in PDBs.

## Conclusions

Detailed analysis of the interfacial information contained in the PDB is very useful to obtain more accurate descriptions of protein interfaces. We have created SCOWLP to have a platform for the characterization and 3D visualization of protein interfaces. SCOWLP enlarges the available information on protein-protein interactions by introducing 3,413 new protein-peptide interfaces and 435,086 additional water-mediated interactions. All interactions contained in SCOWLP are characterized and classified at physicochemical level instead of using general distance criteria. This allows a more appropriate definition and enhanced comparison of the interfaces contained in our database.

As the origin of specificity and affinity in molecular recognition can be partially explained in terms of solvent's contribution to the interaction, our database constitutes a very useful tool to facilitate rational ligand design. In particular *wet spots* can be used as indicators of interfacial solvent variations, being helpful in comparison of protein family interfaces, and perhaps guiding docking experiments.

SCOWLP may be of interest to many structural bioinformaticians, representing a useful tool for classification of protein interfaces, protein binding comparative studies, reconstruction of protein complexes and understanding protein networks.

## Availability and requirements

SCOWLP is available at <http://www.scowlp.org>. The database and all summary tables used in this paper can be freely downloaded for independent studies.

## Authors' contributions

JT developed the new methodology to create the SCOWLP database. JT and AD designed the web application. MS provided the PSIMAP software code. MTP coordinated and supervised the project.

## Acknowledgements

We thank Gerd Anders and Jens Lättig for useful comments on the manuscript. We thank Christof Winter for helping with the Java programming. M.T.P. group is funded by Klaus Tschira Stiftung GmbH.

## References

1. Phizicky EM, Fields S: **Protein-protein interactions: methods for detection and analysis**. *Microbiol Rev* 1995, **59**(1):94-123.
2. Bader GD, Betel D, Hogue CW: **BIND: the Biomolecular Interaction Network Database**. *Nucleic Acids Res* 2003, **31**(1):248-250.
3. Zanzoni A, Montecchi-Palazzi L, Quondam M, Ausiello G, Helmer-Citterich M, Cesareni G: **MINT: a Molecular INTeraction database**. *FEBS Lett* 2002, **513**(1):135-140.
4. Xenarios I, Salwinski L, Duan XJ, Higney P, Kim SM, Eisenberg D: **DIP, the Database of Interacting Proteins: a research tool for studying cellular networks of protein interactions**. *Nucleic Acids Res* 2002, **30**(1):303-305.
5. Argos P: **An investigation of protein subunit and domain interfaces**. *Protein Eng* 1988, **2**(2):101-113.
6. Janin J, Miller S, Chothia C: **Surface, subunit interfaces and interior of oligomeric proteins**. *J Mol Biol* 1988, **204**(1):155-164.
7. Tsai CJ, Lin SL, Wolfson HJ, Nussinov R: **Protein-protein interfaces: architectures and interactions in protein-protein interfaces and in protein cores. Their similarities and differences**. *Crit Rev Biochem Mol Biol* 1996, **31**(2):127-152.
8. Jones S, Thornton JM: **Principles of protein-protein interactions**. *Proc Natl Acad Sci U S A* 1996, **93**(1):13-20.
9. Lo Conte L, Chothia C, Janin J: **The atomic structure of protein-protein recognition sites**. *J Mol Biol* 1999, **285**(5):2177-2198.
10. Park J, Lappe M, Teichmann SA: **Mapping protein family interactions: intramolecular and intermolecular protein family interaction repertoires in the PDB and yeast**. *J Mol Biol* 2001, **307**(3):929-938.
11. Aloy P, Ceulemans H, Stark A, Russell RB: **The relationship between sequence and interaction divergence in proteins**. *J Mol Biol* 2003, **332**(5):989-998.

12. Keskin O, Tsai CJ, Wolfson H, Nussinov R: **A new, structurally nonredundant, diverse data set of protein-protein interfaces and its implications.** *Protein Sci* 2004, **13**(4):1043-1055.
13. Valdar WS, Thornton JM: **Conservation helps to identify biologically relevant crystal contacts.** *J Mol Biol* 2001, **313**(2):399-416.
14. Ofra Y, Rost B: **Analysing six types of protein-protein interfaces.** *J Mol Biol* 2003, **325**(2):377-387.
15. Caffrey DR, Somaroo S, Hughes JD, Mintseris J, Huang ES: **Are protein-protein interfaces more conserved in sequence than the rest of the protein surface?** *Protein Sci* 2004, **13**(1):190-202.
16. Jones S, Marin A, Thornton JM: **Protein domain interfaces: characterization and comparison with oligomeric protein interfaces.** *Protein Eng* 2000, **13**(2):77-82.
17. Stein A, Russell RB, Aloy P: **3did: interacting protein domains of known three-dimensional structure.** *Nucleic Acids Res* 2005, **33**(Database issue):D413-417.
18. Davis FP, Sali A: **PIBASE: a comprehensive database of structurally defined protein interfaces.** *Bioinformatics* 2005, **21**(9):1901-1907.
19. Finn RD, Marshall M, Bateman A: **iPfam: visualization of protein-protein interactions in PDB at domain and amino acid resolutions.** *Bioinformatics* 2005, **21**(3):410-412.
20. Gong S, Yoon G, Jang I, Bolser D, Dafas P, Schroeder M, Choi H, Cho Y, Han K, Lee S *et al*: **PSIbase: a database of Protein Structural Interactome map (PSIMAP).** *Bioinformatics* 2005, **21**(10):2541-2543.
21. Gong S, Park C, Choi H, Ko J, Jang I, Lee J, Bolser DM, Oh D, Kim DS, Bhak J: **A protein domain interaction interface database: InterPare.** *BMC Bioinformatics* 2005, **6**:207.
22. Ogmen U, Keskin O, Aytuna AS, Nussinov R, Gursoy A: **PRISM: protein interactions by structural matching.** *Nucleic Acids Res* 2005, **33**(Web Server issue):W331-336.
23. Zeng J: **Mini-review: computational structure-based design of inhibitors that target protein surfaces.** *Comb Chem High Throughput Screen* 2000, **3**(5):355-362.
24. Pawson T: **Specificity in signal transduction: from phosphotyrosine-SH2 domain interactions to complex cellular systems.** *Cell* 2004, **116**(2):191-203.
25. Castagnoli L, Costantini A, Dall'Armi C, Gonfloni S, Montecchi-Palazzi L, Panni S, Paoluzi S, Santonico E, Cesareni G: **Selectivity and promiscuity in the interaction network mediated by protein recognition modules.** *FEBS Lett* 2004, **567**(1):74-79.
26. Levy Y, Onuchic JN: **Water and proteins: a love-hate relationship.** *Proc Natl Acad Sci U S A* 2004, **101**(10):3325-3326.
27. Palencia A, Cobos ES, Mateo PL, Martinez JC, Luque I: **Thermodynamic dissection of the binding energetics of proline-rich peptides to the Abl-SH3 domain: implications for rational ligand design.** *J Mol Biol* 2004, **336**(2):527-537.
28. Janin J: **Wet and dry interfaces: the role of solvent in protein-protein and protein-DNA recognition.** *Structure Fold Des* 1999, **7**(12):R277-279.
29. Levitt M, Park BH: **Water: now you see it, now you don't.** *Structure* 1993, **1**(4):223-226.

30. Papoian GA, Ulander J, Wolynes PG: **Role of water mediated interactions in protein-protein recognition landscapes.** *J Am Chem Soc* 2003, **125**(30):9170-9178.
31. Petukhov M, Rychkov G, Firsov L, Serrano L: **H-bonding in protein hydration revisited.** *Protein Sci* 2004, **13**(8):2120-2129.
32. [<http://jmol.sourceforge.net>]
33. [<http://www.rcsb.org/pdb>]
34. [<http://scop.mrc-lmb.cam.ac.uk>]
35. Dafas P, Bolser D, Gomoluch J, Park J, Schroeder M: **Using convex hulls to extract interaction interfaces from known structures.** *Bioinformatics* 2004, **20**(10):1486-1490.
36. Rodier F, Bahadur RP, Chakrabarti P, Janin J: **Hydration of protein-protein interfaces.** *Proteins* 2005, **60**(1):36-45.

## Figure Legends

### Figure 1 - Schematic overview of the methodology.

SCOWLP uses information from PDB, SCOP, SCOL and SCOW for the computation of atomic interface interactions.

### Figure 2 - Comparative histogram of SCOWLP vs. PSIMAP database.

Representation of the number of residue-residue (left *y* axis) and *structural unit* interactions (right *y* axis) contained in SCOWLP and comparison with PSIMAP.

### Figure 3 - Schematic representation of the interface interaction of two molecules and definition of *wet spots*.

Molecules A and B form an interface. Interacting residues and water molecules are represented as black and open circles, respectively.

### Figure 4 - Enrichment of the interface definitions by peptidic-ligands and solvent.

**A)** Enrichment in the description of protein interfaces by peptidic-ligands. The molecular recognition features of the BTB/POZ domain family are summarized. A representative POZ domain (green) is surrounded by five different ligands representing all possible BTB/POZ binding zones. Peptidic-ligands are represented in grey (PDB codes and chains used: 1ldk\_A, 1lqb\_ACD, 1r2b\_D). **B)** Enrichment in the description of protein interfaces by *wet spots*. The complex of TEM1  $\beta$ -lactamase (orange surface) with the inhibitor BLIPII (green ribbon; PDB 1jtd) is shown. White represents residues forming the interface before taking into account *wet spots* (in blue). **C)** The  $\alpha$  and  $\beta$  (orange and green ribbon, respectively) chains of the Respiratory nitrate-reductase 1 (PDB 1q16) are shown. White represents residues forming the interface before taking into account *wet spots* (in blue). Figures created with *InsightII*, Accelrys.

### Figure 5 – Screenshots and legends showing the structure of the SCOWLP website.

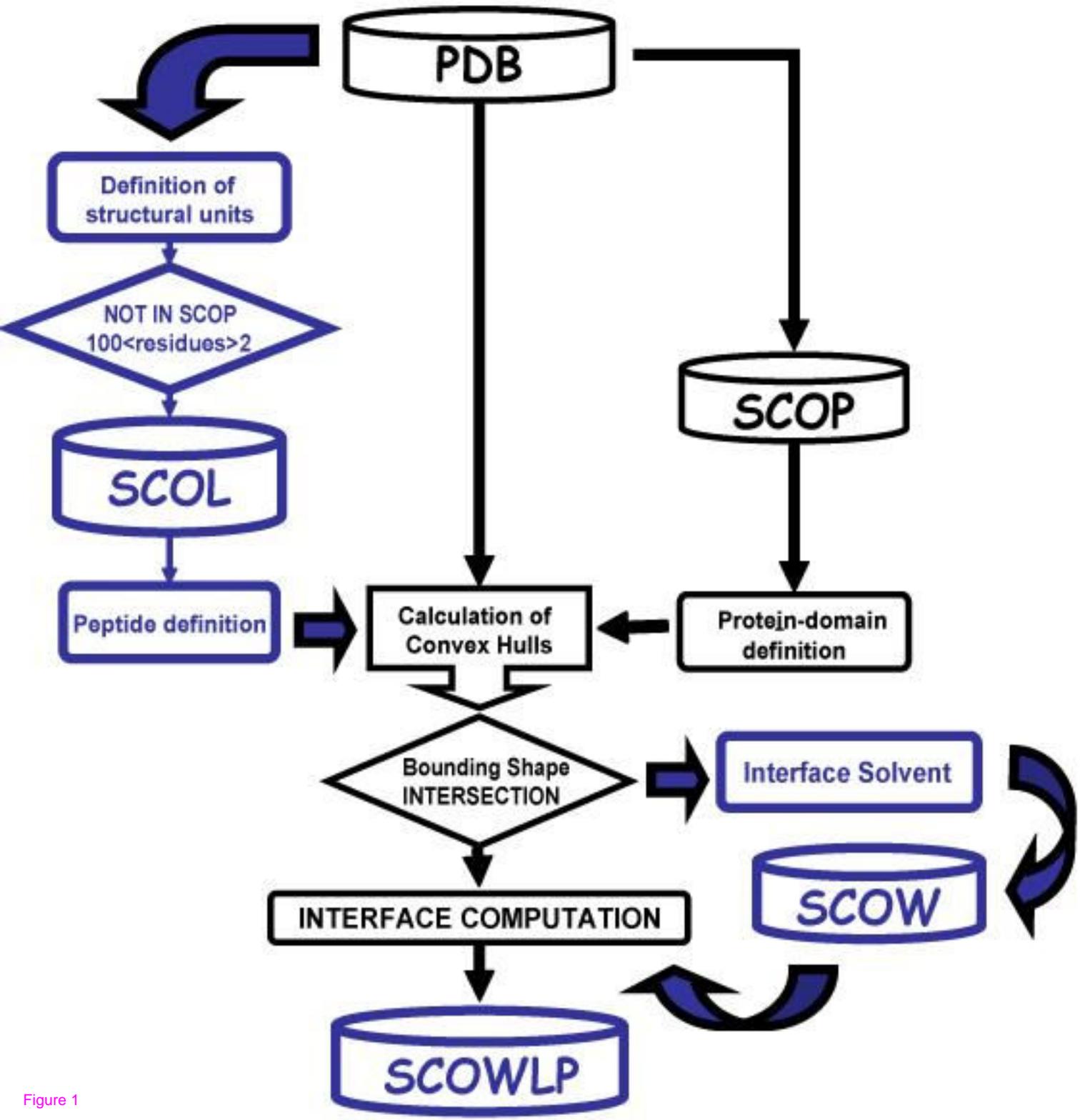
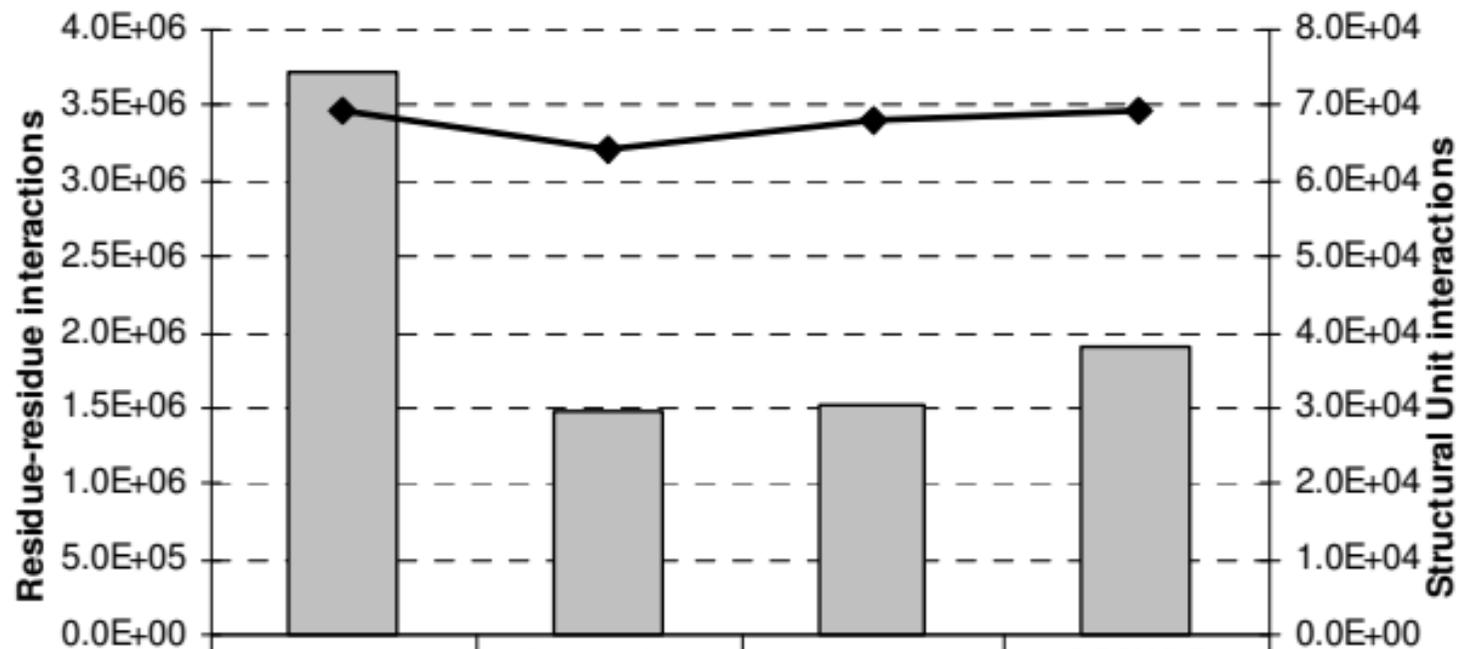


Figure 1



■ residue

◆ Unit

PSIMAP

Interacting  
Rules (IR)

SCOL (+IR)

SCOWLP  
(SCOW+SCOL)

3724525

1465882

1516575

1909434

69387

64364

67897

69103

Figure 2

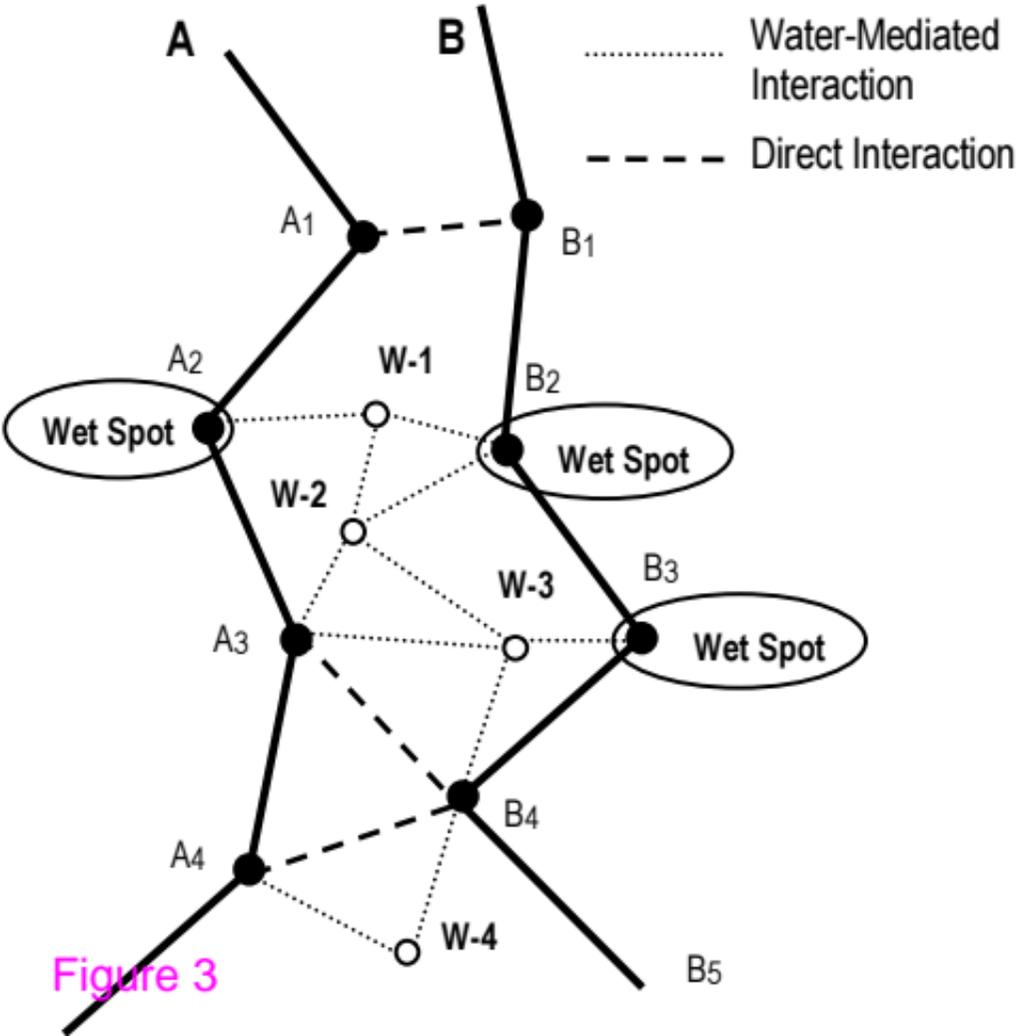


Figure 3

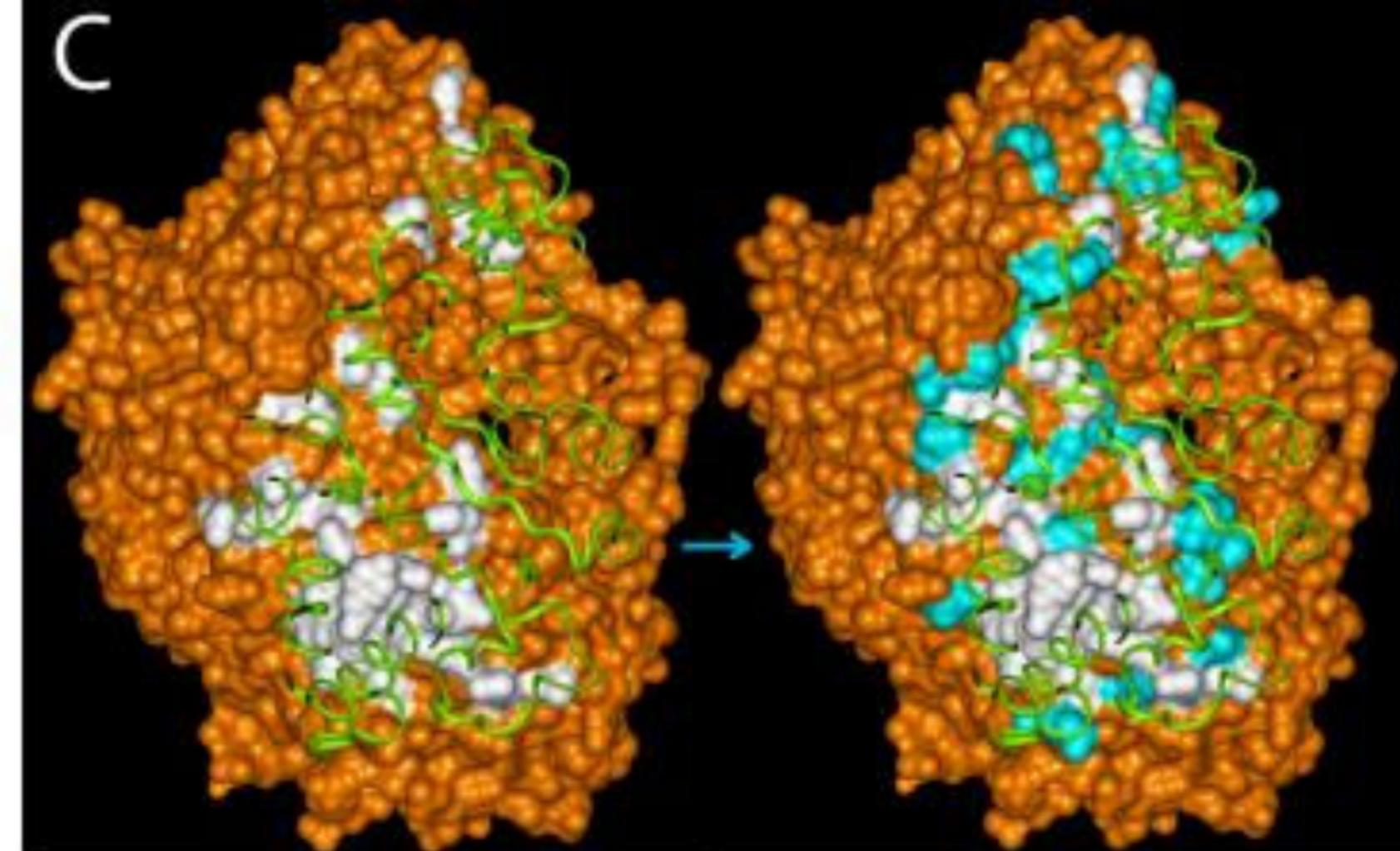
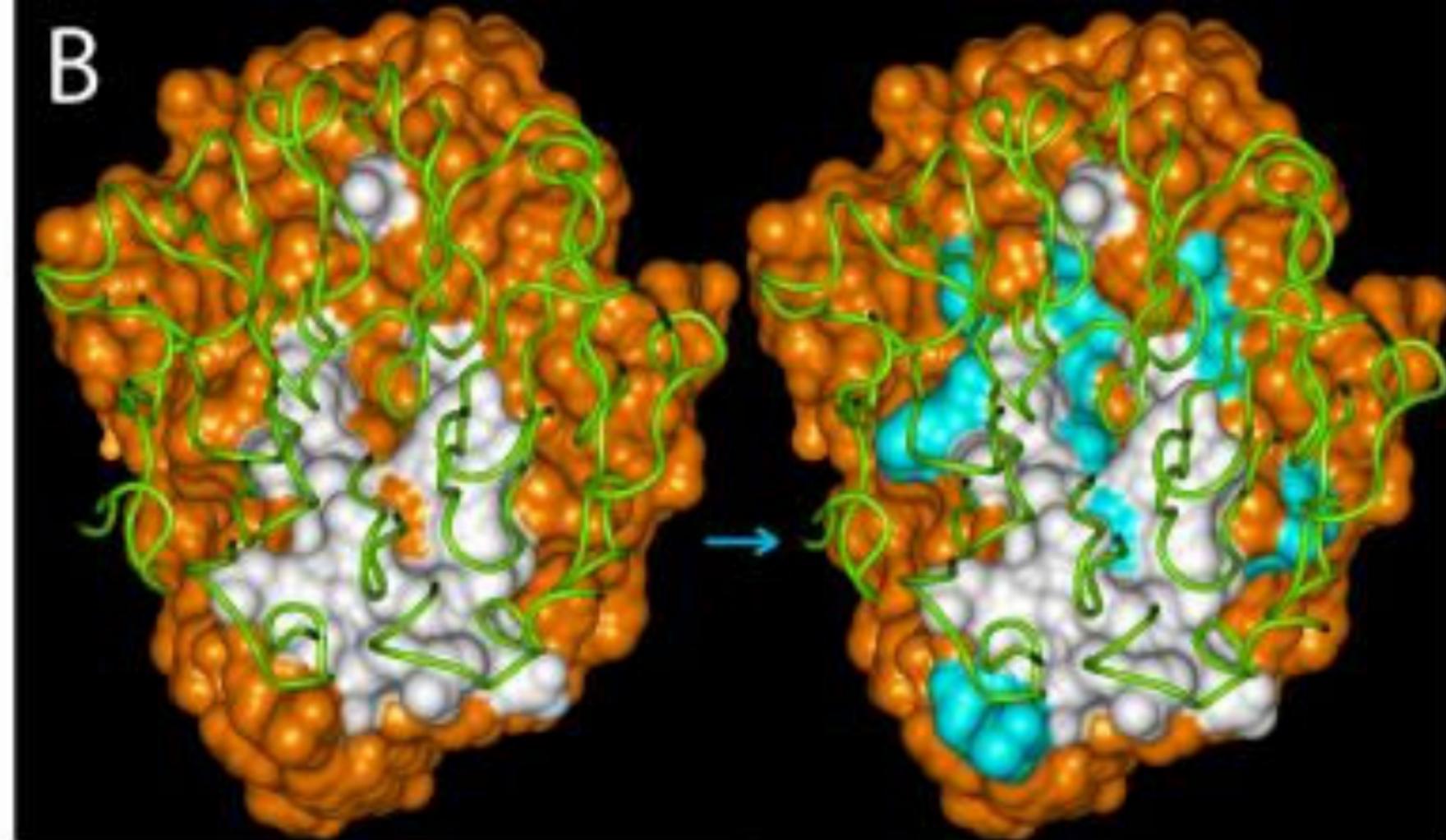
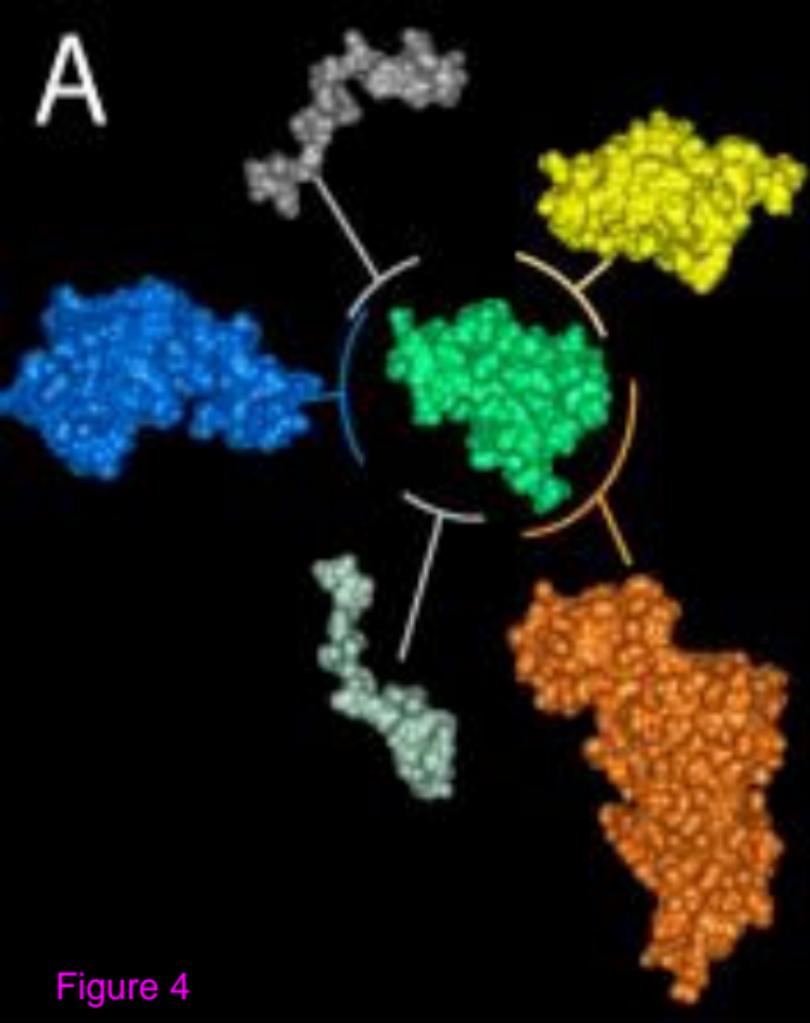


Figure 4

**1-SCOP id and name queries:**

**2- PDB id query:**

**3- Interface viewer:**

### SCOP navigation window:

1a: SCOP hierarchy tree.

1b: Family PDB Ids.

1c: Family interactions table.

### Interface selection window:

2a: Interface summary table.

2b: Interaction type table.

### 3D interface window:

3a: Residue contact tables and interface selection button.

3b: PDB graphical representation highlighting the selected interface

3c: Interactive colouring and ON/OFF buttons.

**Additional files provided with this submission:**

Additional file 1: List\_of\_changes.pdf : 35KB

<http://www.biomedcentral.com/imedia/1836517130937513/sup1.pdf>