ABC database – a database for analyzing biomolecular contacts

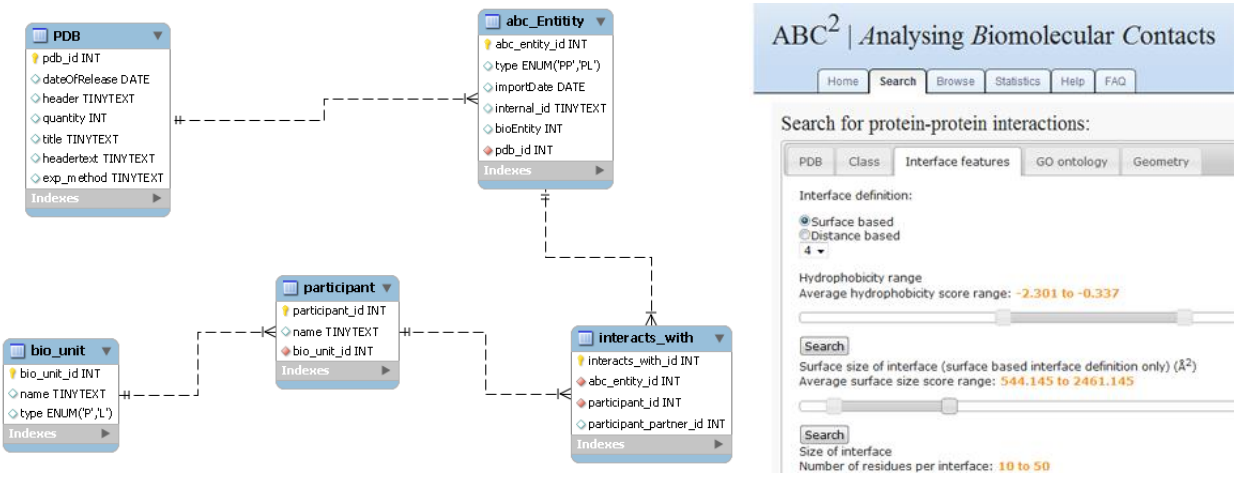
Peter Walter, Volkhard Helms

Chair for Computational Biology, Saarland University

In the last years modern proteomic methods have helped to get a better understanding of the complexity within living cell and organism. A growing number of sequences of unknown proteins are deciphered, their function is revealed, structural details are detected and the interaction in the complex network of biological processes is uncovered. Also, more and more details about biomolecular contacts are known, which are defined as any non-covalent interaction between proteins, DNA/RNA strands or small molecules. In particular, protein-protein interactions play a crucial role in most biological processes such as transcription, signal transduction and metabolism. This makes them an important target for therapeutic use and establishes a close relationship to the field of protein-small molecule interactions.

The vast increase of such information necessitates the application of databases for easy handling and analysis. The latest release of the ABC database stores data about protein-protein and protein-small molecule interfaces [1]. A webinterface allows for queries such as search for hydrophobic interfaces or interfaces exhibiting a certain function according to gene ontology.

Figures: The left figure illustrates the database schema of ABC database. On the right hand side, the web interface for accessing the database is shown.



The database schema consists of about 100 relations. The basic structure is depicted above. An *abcEntity* stands for a certain interaction between two different molecules which is originated from a PDB file that is represented by the *PDB* relation. The *interacts_with* provides more details about the interface region. The structures that are involved in an interaction are represented by the *participant* relation. Different participants may share a common feature which is covered by the *bio_unit* relation. As an example, different protein chains which participate in an interaction may be based on the same sequence, so they have the same *bio_unit*-id.

Programming was mainly done using Java, including servlets/JSP for web-specific tasks. The website is running under a Tomcat webserver, MySQL was used as database engine. A number of external programs and tools was used among them Marvin, JChem, Jmol, JFreeChart.

[1] P. Walter, S. Ansari, V. Helms, *Journal of integrative bioinformatics*, **2007**, *4*.