Ligand Depot: a data warehouse for ligands bound to macromolecules

Zukang Feng, Li Chen, Himabindu Maddula, Ozgur Akcan, Rose Oughtred, Helen M. Berman and John Westbrook*

Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Department of Chemistry and Chemical Biology, Rutgers, The State University of New Jersey, 610 Taylor Road, Piscataway, NJ 08854, USA

Received on December 12, 2003; revised on March 3, 2004; accepted on March 3, 2004
Advance Access publication April 1, 2004

ABSTRACT

Summary: Ligand Depot is an integrated data resource for finding information about small molecules bound to proteins and nucleic acids. The initial release (version 1.0, November, 2003) focuses on providing chemical and structural information for small molecules found as part of the structures deposited in the Protein Data Bank. Ligand Depot accepts keyword-based queries and also provides a graphical interface for performing chemical substructure searches. A wide variety of web resources that contain information on small molecules may also be accessed through Ligand Depot.

Availability: Ligand Depot is available at http://ligand-depot.rutgers.edu/. Version 1.0 supports multiple operating systems including Windows, Unix, Linux and the Macintosh operating system. The current drawing tool works in Internet Explorer, Netscape and Mozilla on Windows, Unix and Linux. Contact: jwest@rcsb.rutgers.edu

The Protein Data Bank (PDB, http://www.pdb.org/) (Berman et al., 2000) is the single worldwide repository of structural data on biological macromolecules. About 10 years ago, the PDB contained only ~2000 structures, but this number has grown dramatically to over 24,000 entries today. Many of these structures are complexes of proteins or nucleic acids with other macromolecules or with small molecules. In the PDB, small molecules include ligands, drugs, cofactors, inhibitors, ions and solvent molecules. Structures of small molecules bound to proteins or nucleic acids are a valuable source of information for understanding biological systems and drug design.

With the accelerated accumulation of structural data, there is a pressing need for tools that help researchers effectively navigate, deposit and retrieve macromolecular structures containing small molecules. Several existing web services allow for the search and retrieval of information on small molecules present in the PDB. These include CHEM-PDB (Boutselakis et al., 2003), HIC-up (Kleywegt and Jones, 1998), PDBsum (Laskowski, 2001), the IMB Jena Image Library of Biological Macromolecules (Reichert and Suhnel, 2002) and Relibase (Hendlich, 1998). An enhanced version of Relibase, known as Relibase+ (Hendlich et al., 2003), is available from the Cambridge Crystallographic Data Centre (CCDC). The CCDC also maintains the Cambridge Structural Database (CSD) (Allen, 2002) that houses information on over 250,000 small molecule crystal structures.

Here we describe Ligand Depot, a data warehouse that integrates databases, services and tools related to small molecules bound to macromolecules. It allows users to extract ligand information quickly and easily from the PDB, to perform chemical substructure searches of PDB ligands using a graphical interface and also to browse other relevant small molecule resources on the Web. Ligand Depot can also perform keyword searches on a subset of these relevant websites so that they may be searched simultaneously using a single query. One of the distinguishing features of Ligand Depot is that it allows users to retrieve the coordinates of any small molecule found within the structure entries of the PDB. Ligand Depot is also updated daily and therefore provides the most current information on small molecules present in the PDB.

Ligand Depot is designed to support three modes of information retrieval: keyword search, chemical substructure search and browsing by category. These are described in the following sections.

KEYWORD SEARCH

Ligand Depot currently includes chemical descriptions for the ~4600 small molecules that are part of the structures deposited with the PDB and it offers various search options for obtaining information on these ligands. It accepts queries based on PDB ligand code, compound name and chemical formula (Fig. 1). When searching for ligands by chemical name or formula, flexibility is provided by allowing either exact or partial matches. Query results are summarized as a list of ligands with active links to more detailed ligand reports.

*To whom correspondence should be addressed.
Fig. 1. Screenshot of a Ligand Depot session. A simple user interface allows for keyword searches (top). A graphical interface permits chemical substructure searches (bottom left). Resulting ligand reports include a chemical description, links to relevant PDB entries and downloadable coordinates for the ligand of interest (bottom right).

Each ligand report includes a chemical description, a two-dimensional chemical diagram and a link to the ligand’s macromolecular Crystallographic Information File (mmCIF) (Bourne et al., 1997), which is found in the PDB ligand dictionary (http://deposit.pdb.org/public-component-erf.cif). A ligand’s mmCIF dictionary description specifies the chemical identity of the molecule and lists its atom types along with their connectivity and corresponding bond orders.

The Protein Data Bank entries that contain a small molecule of interest are also listed in the ligand report and the PDB IDs are actively linked to their respective Structure Summary reports on the PDB website. The coordinates for each PDB structure containing the small molecule may be downloaded from the Structure Summary page. The list of PDB structures containing the ligand is conveniently sorted by resolution and the coordinates for each ligand may also be downloaded in PDB, mmCIF and MOL2 formats for viewing in various molecular visualization tools.

CHEMICAL SUBSTRUCTURE SEARCH

The capability of searching for ligands based on their chemical structure is a powerful feature offered by Ligand Depot. Using a simple graphical interface, a structural comparison may be performed between a small molecule and all of the ligands present in the PDB. In version 1.0, the Marvin Sketch drawing tool (ChemAxon Ltd., http://www.chemaxon.com/marvin) (Fig. 1) is provided for users to draw a small molecule diagram from scratch, or alternatively, to upload a graphical file of a ligand in either mmCIF or MOL2 format. These molecular diagrams may be modified using the drawing tool, if desired and structural searches may then be submitted. Thus, the graphical interface provides an efficient means
Ligand Depot performs the structural search by comparing the atoms and bonds of both the queried and target molecules. The results may include either the actual ligand being queried or else larger ligands that contain the queried molecule as a substructure. Therefore, the structural comparison will return a list of PDB ligands that are structurally similar, either in part or in whole. The resulting list of ligands is sorted by atom number with the closest match to the substructure of interest being returned first. The query results are also actively linked to detailed ligand reports for each matching small molecule.

**BROWSING BY CATEGORY**

Ligand Depot not only provides direct access to ligands in the PDB, it can also be used to browse several other small molecule resources on the Web. A manually intensive process was carried out to identify these relevant sites. Conventional search engines and a third party search utility (Web Ferret: http://www.ferretsoft.com/) were used and any matches returned were validated.

Currently, Ligand Depot stores information from 70 small molecule sites. The different resources are organized into four categories including nomenclature sites, molecular visualization sites, commercial sites and chemical databases. Selecting one of these categories returns a list of web resources with a brief description of what each one has to offer. Each annotated list of websites allows the user to make an informed decision about which site to explore further and thereby retrieve information on small molecules more efficiently.

Another valuable feature of Ligand Depot is that it allows keyword searches to be performed on these external websites if they are search-enabled. A mechanism is provided to facilitate the searching of multiple websites for any given chemical name or formula. Thus, information on small molecules can be extracted quickly and easily from a diverse collection of web resources using a single query.

**ARCHITECTURE**

Ligand Depot has been implemented as a three-tiered web-based client/server application. It has the user’s web browser on the client side, a MySQL database server as the back-end and a Tomcat (http://jakarta.apache.org/tomcat/) application server as the middle tier. The back-end has a set of normalized tables that store URLs and other information about websites related to small molecules. The processing logic happens in the middle-tier and is handled by the Tomcat application server using Java servlets.

**FUTURE WORK**

Ligand Depot provides a simple user interface for searching and retrieving information on small molecules present in structures deposited with the PDB. It offers flexible query capabilities, a drawing tool for carrying out substructure searches and a means for importing and exporting graphical files of small molecules. This integrated data resource also delivers, in a timely manner, the most up-to-date information on small molecules present in PDB entries. As a data warehouse, Ligand Depot optimizes the query and reporting of ligand information present in the PDB and in a number of relevant small molecule sites. Future implementations will include enhanced searching capabilities and the incorporation of a more sophisticated graphical user interface. These and other future developments will ensure that Ligand Depot continues to be a valuable information resource on small molecules for a wide variety of users.

**ACKNOWLEDGEMENTS**

This work was sponsored by the State of New Jersey Commission on Higher Education, Rutgers Strategic Resources and Opportunity Analysis project and by funds that support the RCSB PDB from the NSF, NIGMS, DOE, NLM, NCI, NCRR, NIBIB and NINDS.

**REFERENCES**


