

Abstract

Visualization of protein structural data is an important aspect of protein research. Incorporation of genomic annotations into a protein structural context is a challenging problem, because genomic data is too large and dynamic to store on the client and mapping to protein structures is often nontrivial. To overcome these difficulties we have developed a suite of SOAP-based Web services and extended the commonly used structural visualization tools UCSF Chimera and Delano Scientific PyMOL via plugins. The initial services focus on (1) displaying both polymorphism and disease associated mutation data mapped to protein structures from arbitrary genes and (2) structural and functional analysis of protein structures using residue environment vectors. With these tools, users can perform sequence and structure based alignments, visualize conserved residues in protein structures using BLAST, predict catalytic residues using an SVM, predict protein function from structure, and visualize mutation data in SWISS-PROT and dbSNP. The plugins are distributed to academics, government and nonprofit organizations under a restricted open source license. The Web services are easily accessible from most programming languages using a standard SOAP API. Our services feature secure communication over SSL and high performance multi-threaded execution. They are built upon a mature networking library, Twisted, that allow for new services to easily be integrated. Services are self-described and documented automatically enabling rapid application development. The plugin extensions are developed completely in the Python programming language and are distributed at

<http://www.lifescienceweb.org/>

The LSW Website contains developer tools and mailing lists, and we encourage other developers to extend their applications using our services.

Project Goals

Web services are an efficient way to provide genomic data in the context of protein structural visualization tools. Our goal is to define a series of bioinformatic web services that can be used to extend protein structural visualization tools, and other extensible computational biology desktop applications. Our current focus is on extending UCSF Chimera (<http://www.cgl.ucsf.edu/chimera/>) and Delano Scientific PyMOL (<http://pymol.sourceforge.net>).

Figure 1: Screen grab of the current services list from <http://www.lifescienceweb.org/>.

Services currently offered include:

- ClustalW alignments
- Mutation <-> PDB mapping
- SVM based catalytic residue prediction
- Sequence conservation based on PSI-BLAST PSSM



Web Services

AlignmentService (wsdl)
A collection of methods for structural and sequence alignment.

CSPService (wsdl)
Web service for catalytic site prediction. Use the SBlest service for password management.

LSWService (wsdl)
Methods provided in this service are meant to cater to needed functionality for clients. News and update information can be pushed through to clients.

MutDBService (wsdl)
This service is an interface to the Mutation Database as maintained by Sean Mooney (mutdb.org).

PDBChainService (wsdl)
A service for fetching pdb structures. This is meant to be used by the LSW clients only. It provides advanced usage such as fetching compressed structures or just title information for example. All methods accept standard pdb codes, or pdb codes with a chain specifier.

PDBGeneService (wsdl)
This service provides a mapping for finding pobs related to a gene, or vice-versa.

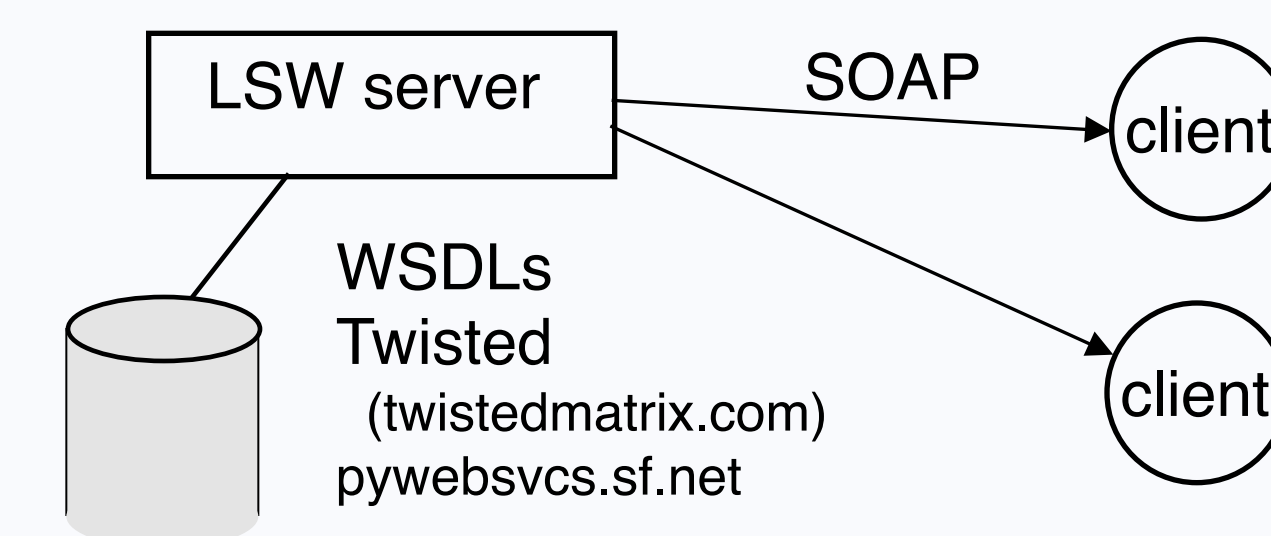
PSSMService (wsdl)
Sequence conservation service.

SBlestService (wsdl)
Web service for SBlest (www.sblest.org).

Copyright 2005, The Trustees of Indiana University. All rights reserved. This file is part of the LSW Web Services source distribution. For license information, see the LICENSE file in the top level directory of the LSW Web Services source distribution. Pervasive Technology Labs

Services Model

Web services are an efficient way to provide genomic data in the context of protein structural visualization tools. Our goal is to define a set of bioinformatic web services that can be used to extend protein structural visualization tools, and other extensible computational biology desktop applications. We are currently focused on extending UCSF Chimera (<http://www.cgl.ucsf.edu/chimera/>) and Delano Scientific PyMOL (<http://pymol.sourceforge.net>). Our services use the SOAP protocol and are currently developed using open source Python-based projects.



(We will address service discovery in the future)

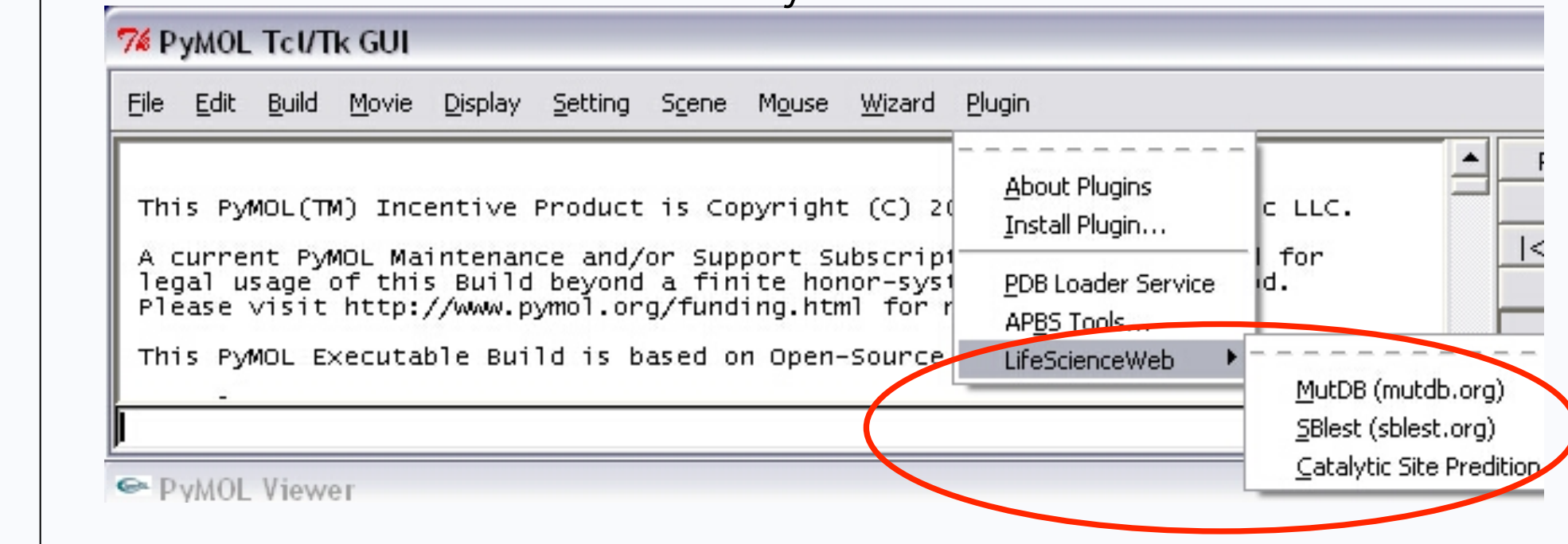
Software Plugin Extensions

We have extended UCSF Chimera and Delano Scientific PyMOL to access our services. The three primary services we provide now are:

1. Disease associated mutation and SNP to protein structure mapping and visualization
2. Protein sequence and structure residue analysis with PSI-BLAST and S-BLEST
3. Catalytic residue prediction using a support vector machine (Youn, E., *et al. submitted*)

Installation Plugin installation is easy and can be performed for a user without root privileges. Currently, all platforms supported by UCSF Chimera and PyMOL are supported and include UNIX platforms, LINUX, Mac OS X and Windows XP. For either of the two clients supported (PyMOL or UCSF Chimera), simply follow the directions linked on the download page at <http://www.lifescienceweb.org/>. They will thereafter be available from the menu, as shown below.

Figure 2: Running our tools from the client application, shown using PyMOL.



Updates

The annotations are currently updated every 2-3 months. Internally, we provide services for annotating genes or coordinates not in the PDB usually through a collaboration. For information on how to do this please contact Sean Mooney, sdmooney@iupui.edu.

Acknowledgements

CM and RH are funded through the IPCRES Initiative grant from the Lilly Endowment. SDM is funded from a grant from the Showalter Trust, an Indiana University Biomedical Research Grant and startup funds provided through INGEN. The Indiana Genomics Initiative (INGEN) is funded in part by the Lilly Endowment.

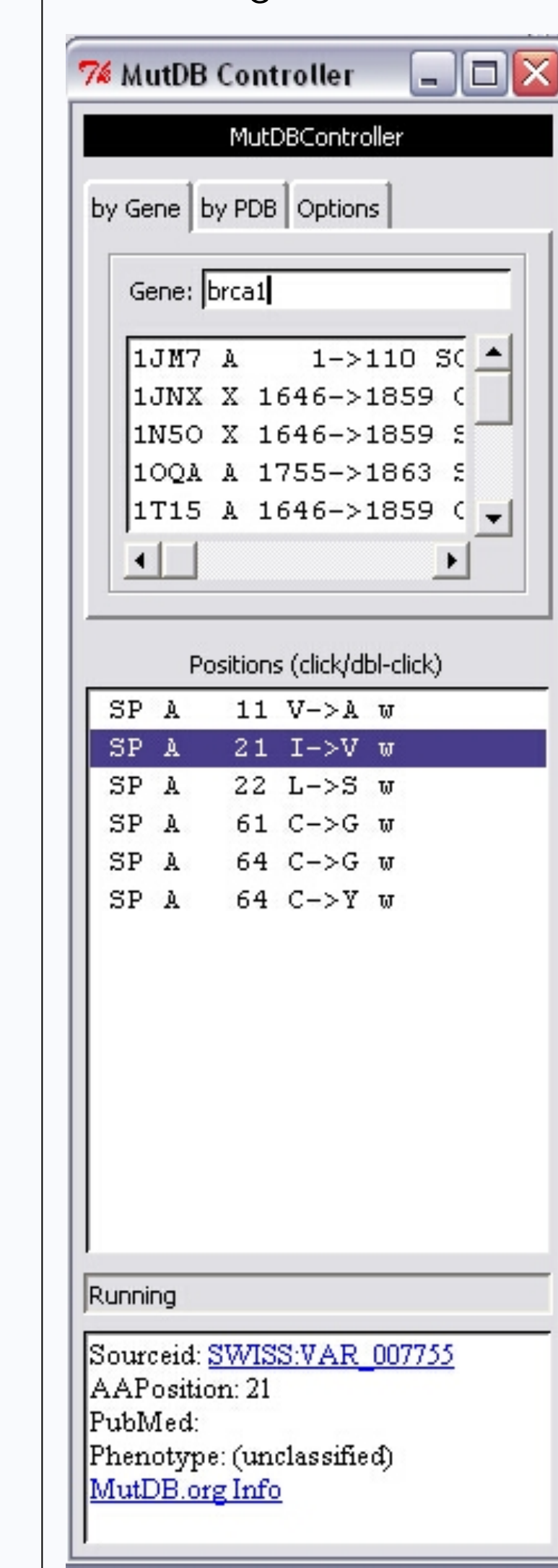
The authors would like to thank the authors of UCSF Chimera and PyMOL for their help in extending their applications. You can download these tools from the following:

- UCSF Chimera: <http://www.cgl.ucsf.edu/chimera/>
- Delano Scientific PyMOL: <http://pymol.sourceforge.net>

Visualization of Mutations on Protein Structures

We provide mapping between mutations and SNPs and protein structures. The mutations are mapped using Smith-Waterman based alignments. Swiss-Prot mutations and nonsynonymous SNPs in dbSNP are currently supported. See <http://mutdb.org/> for a current list of the versions of each dataset we provide.

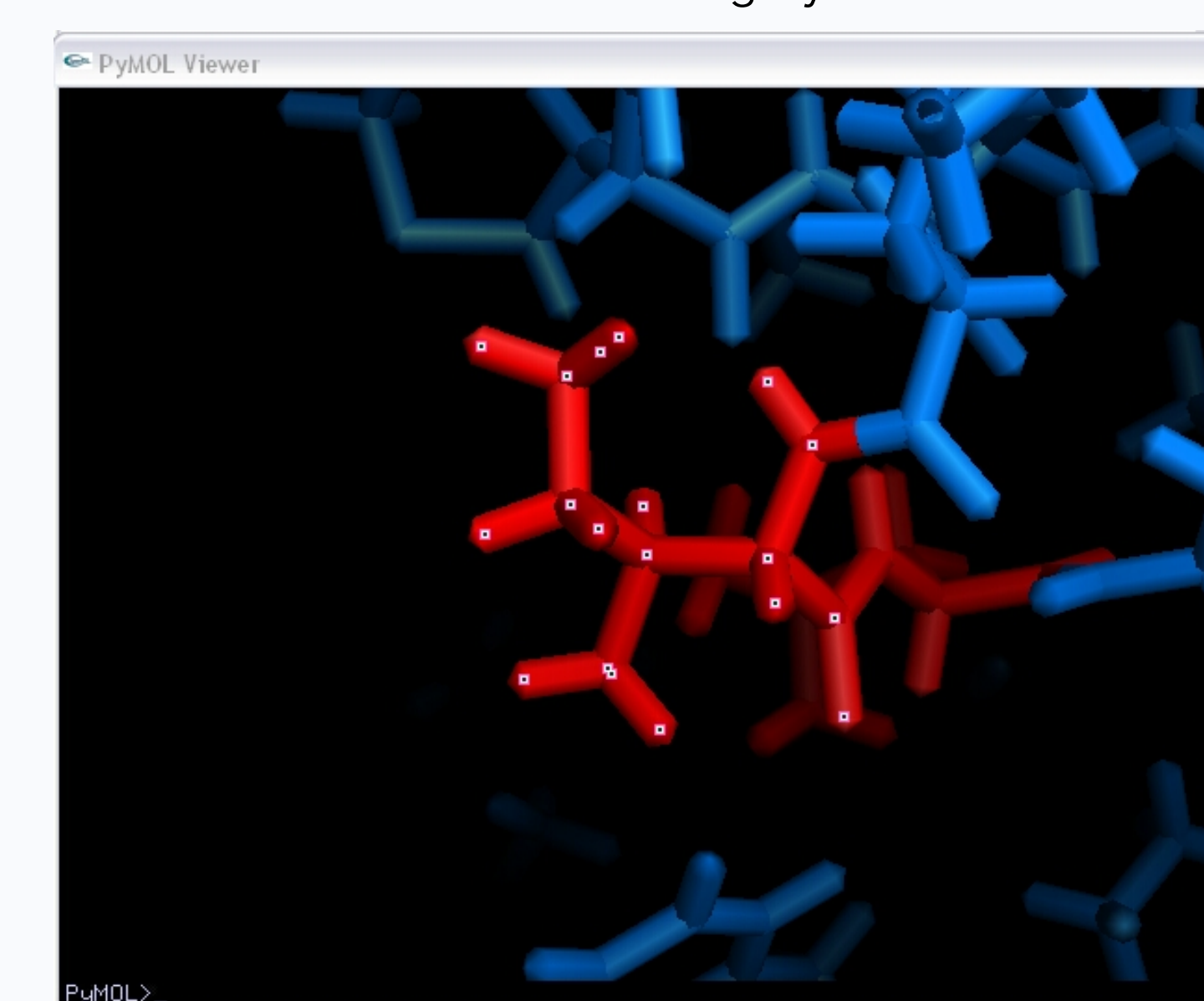
Figure 3: MutDB controller window, shown using PyMOL.



Controller features include (from the top):

- Tabbed selection of query type and controller options.
- Query entry text box and resulting hits from PDB shown below, with PDB ID, chain, residues, and TITLE of PDB.
- Once a PDB ID above is selected, the coordinates are downloaded and the mutations from Swiss-Prot (SP) and dbSNP (SNP) are retrieved. The database source, type, position, mutation and wildtype flag are displayed. Upon selection, the mutation is highlighted in the coordinate visualization window.
- Status window that displays the number of mutations or PDB coordinates found.
- Mutation information window displays a link to the source (which opens in the browser), the position and annotations in that may be available, including PubMed ID (as link), phenotype and a link to MutDB.org.

Figure 4: MutDB structure visualization window showing a highlighted mutation using PyMOL.



Citations

Dantzer J, Moad C, Heiland R, Mooney S. (2005) "MutDB services: interactive structural analysis of mutation data". *Nucleic Acids Res.*, 33, W311-4.

Peters B, Moad C, Youn E, Buffington K, Heiland R, Mooney S, "Identification of Similar Regions of Protein Structures Using Integrated Sequence and Structure Analysis Tools". Submitted.

Mooney, S.D., Liang, H.P., DeConde, R., Altman, R.B., *Structural characterization of proteins using residue environments*. *Proteins*, 2005. 61(4): p. 741-7.

Automated Sequence and Structural Analysis of Protein Structures

Using PSI-BLAST and S-BLEST, we provide analysis of residue environments that match between protein structures in a queried database. Additionally, if the found environments represent similar structure or function classes, the environments that are most structurally associated to those environments are returned. This service is authenticated and SSL encrypted, and all coordinate data and analysis data are stored on our servers. Currently, users can query the ASTRAL 40 v1.69 and ASTRAL 95 v1.69 nonredundant domain datasets, as well as other commonly used nonredundant protein structure databases.

Figure 5: S-BLEST controller window shown using UCSF Chimera.

On the right, the control box has (from top):

- Tabs for selecting hits in database with matching environments (or significant sequence similarity using PSI-BLAST) or common functional annotations in the hits.
- A pull down selection box showing the PDB ID's with matching environments and the Z-score between the best environments. Upon selection the hit is downloaded and displayed in the visualization window (left).
- A button to retrieve a ClustalW alignment between the the selected hit structure and the query.
- The most significantly matched residue environments between the query and the hit. Displays Z-score, the matched residues, the ranking of that match (overall for that query residue environment) and the Manhattan distance. When residues are selected from this list, the coordinates in the visualization window are aligned using a the Chimera match command.

• Below the windows a ClustalW alignment is shown

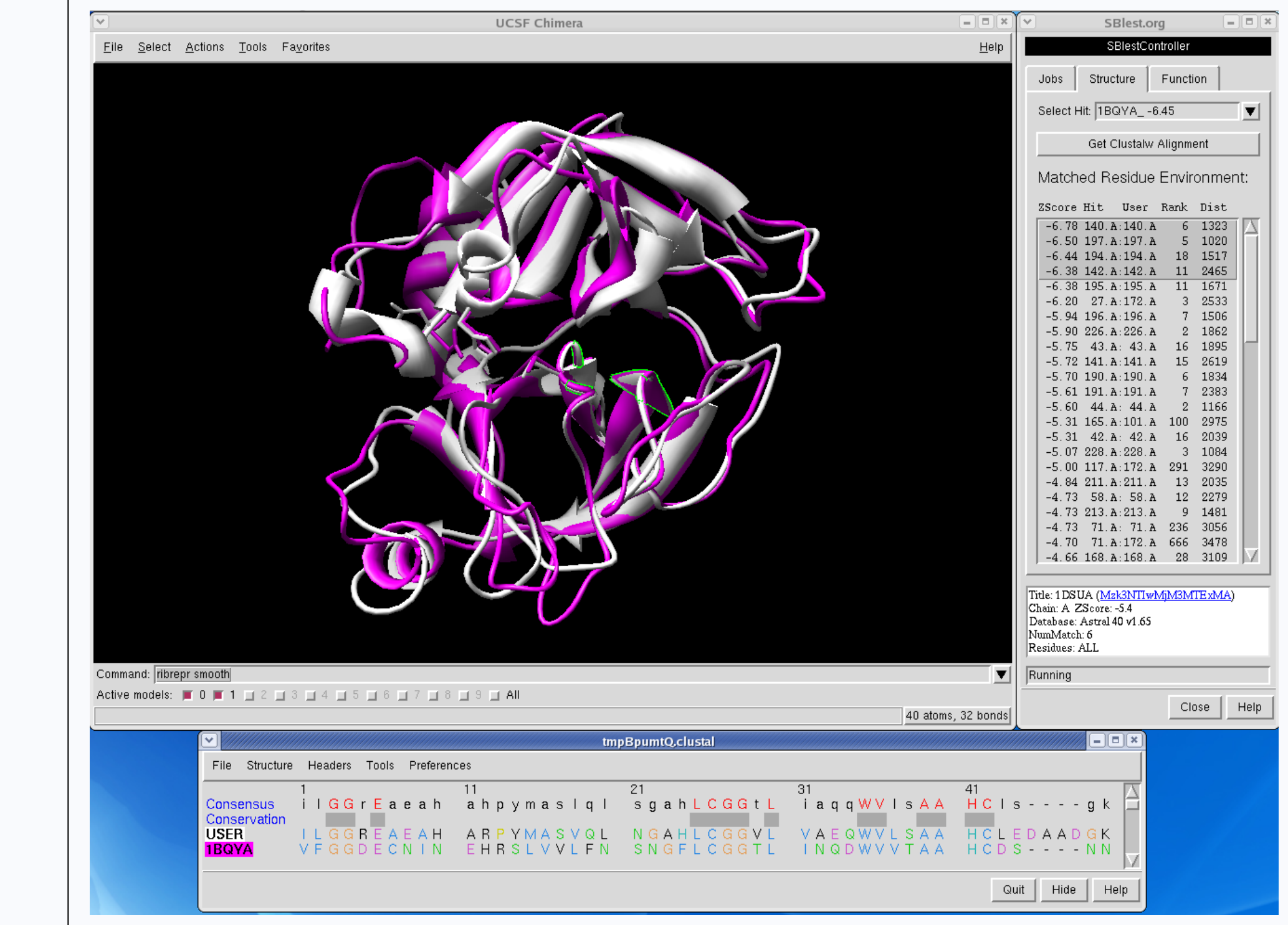


Figure 6: S-BLEST controller window showing the function analysis tab using UCSF Chimera.

