soaPDB: a web application for searching the Protein Data Bank, organizing results, and receiving automatic email alerts

Charles A. Lesburg* and José S. Duca

Department of Drug Design, Schering-Plough Research Institute, 2015 Galloping Hill Road, Kenilworth, NJ 07033, USA

ABSTRACT

soaPDB is a web application that allows generation and organization of saved PDB searches, and offers automatic email alerts. This tool is used from a web interface to store PDB searches and results in a backend relational database. Written using the Ruby on Rails open-source web framework, soaPDB is easy to deploy, maintain and customize. soaPDB is freely available upon request for local installation and is also available at http://soapdb.dyndns. org:3000.

INTRODUCTION

The goal of this effort is to provide a tool that facilitates the creation, organization and reproducibility of RCSB Protein Data Bank (1–3) searches. The PDB is typically accessed via their website (http://www.pdb.org), which offers a rich suite of searching and reporting tools for the vast database of macromolecular structures, including a flexible and powerful XML-based query syntax. Currently, advanced search queries are saved only for the duration of the browser session, often no longer than a few hours. To automate routine and reproducible PDB searches—especially useful for both simple and complex compound queries alike—we have created a web-based tool, which accesses the PDB's Web Services API (http://www.pdb.org/robohelp/#webservices/summary.htm). Compared with existing web services, which provides email alerts based on sequence comparisons or other keywords (4,5), the tool presented herein allows for construction of arbitrarily detailed PDB queries as well as the ability to run a *soaPDB* server privately within one's own institution to enable collaboration among local users without public exposure.

DESCRIPTION AND IMPLEMENTATION

We present *soaPDB*, a cross-platform web-based application that allows authenticated users to save their PDB searches and organize their results in a relational database. A public server (http://soapdb.dyndns.org:3000) is available for testing, and it is encouraged to install soaPDB locally or departmentally to take advantage of local collaboration. Email alerts are automatically sent when a user's saved search yields new results. Interaction with the PDB is performed using the Simple Object Access Protocol (SOAP, http://www.w3.org/TR/soap). Since the interface is a web browser, a number of which were tested, no installation is required for the end user thus minimizing setup and maintenance costs. The server side of this application was written using the Ruby on Rails (6) open-source web abstraction framework. This high-level language extension allows rapid and efficient web application prototyping and development with built-in support for many computer operating systems and relational database types. The *soaPDB* application server was developed using Microsoft Windows XP and linux. The database server tested was MySQL and Rails supports many others. The Rails architecture assures portability to the user's preferred operating system, web server and database server application including various linux distributions, Mac OS X and Windows XP.

We created this tool based on the following principles:

- *Persistence, reproducibility and automation.* By adding the ability to save and organize preformed searches, we aim to enhance the rich functionality available at the PDB, not to supplant it. Therefore, all result reporting is expected to be performed using the tools available from the PDB. A supporting database facilitates search persistence and reproducibility. Notification of new search results is performed using a query runner, executed automatically on a regular basis.

---

*To whom correspondence should be addressed. Tel: +1 908 740 3515; Fax: +1 908 740 7664; Email: charles.lesburg@spcorp.com

- *Simplicity*. A simple interface to one's saved searches, with an emphasis on the search results themselves, keeps this tool focused. There are very few necessary configuration options. For instance, email updates occur by default on every Wednesday, the day after weekly PDB updates. Relying on XML-based queries of any complexity, generated using the advanced search tools available via the PDB website, this simple tool should appeal to the novice and expert alike. While the *soaPDB* saved searches may be as complex as desired, the default options are tailored to the needs of a typical user. We envision that such a user of this web application is a member of the macromolecular structural biology, structural chemistry or structure-based drug design communities. Moreover, it is anticipated that the server would be installed locally or departmentally in order to share saved searches among coworkers via the cloning and editing functionality of *soaPDB*. A list of frequently-asked questions and an extensive tutorial with worked examples is provided.
- *Portability and adaptability*. The Rails web application framework was chosen for its cross-platform support. Moreover, since the code is not compiled, it is readily customizable to fit any circumstance. For instance, user authentication may be performed using an internal user database or via secure access to a preferred LDAP server.

MODUS OPERANDI

The *soaPDB* usage workflow typically requires the following steps, shown schematically in Figure 1.

Login

Upon authentication, the user is presented with a main page where his/her own and other's saved searches are listed along with number of search results (Figure 1b). The user may run, edit or delete current queries, examine other's queries or generate a new query, either by cloning existing queries or creating new ones.
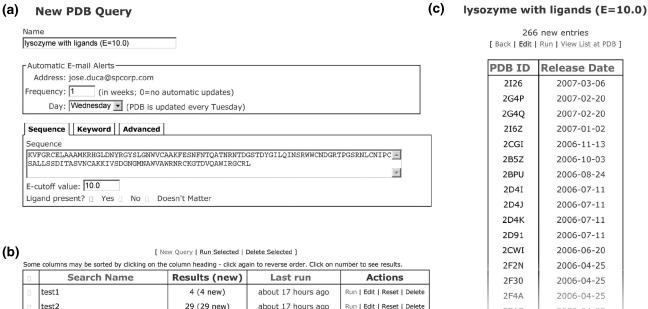
New query generation

To create a new query (Figure 1a), the user may use one of three choices:

(1) The *Sequence* tab where an amino acid sequence is entered, along with *E*-value cutoff and optionally in the presence or absence of a ligand. The search is then performed using BLAST (7), as implemented by the PDB.
(2) The *Keyword* tab where any text can be entered, optionally in combination with the presence or absence of a ligand.
(3) The *Advanced* tab, which accepts any XML-based query from the PDB. There is a direct link to the PDB Advanced Search web page to ease XML creation.

Finally, the user may elect to receive an email alert containing any new search results.

Alternatively, queries may be cloned from the list saved searches, including those from other *soaPDB* users. Saved searches may be edited at any time to adjust the name, the email notification settings or the XML query itself.

Figure 1. Schematic of a typical *soaPDB* work flow. (**a**) Creation of a new saved search. (**b**) List of saved searches, with links to lists of results. (**c**) List of PDB ID codes returned by a saved search. Each link can be viewed individually or the list may be viewed at the PDB website.

Query execution and results

Upon running a query, the number of new hits is presented to the user along with the total number of hits (Figure 1b). Navigating to a list of results offers a list PDB ID codes and their release dates (Figure 1c). Results may be explored either individually or collectively by redirecting one or more resulting entries to the PDB. At that point, all of the tools at the PDB are available for use. It is noteworthy that all searching is performed by the PDB via a SOAP request, and results are stored locally by the *soaPDB* server.

CONCLUSIONS

soaPDB is a tool, which enhances the search capability of the Protein Data Bank by storing searches and allowing for their routine and reproducible execution. As a web application, usage is straightforward and platform-independent. Since this application was created using the Ruby on Rails web abstraction framework and supports a variety of database servers, the server may be easily deployed and readily customized to fit into many situations. The *soaPDB* code is freely available upon request from the authors.

REFERENCES

1. 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.
2. Berman,H.M., Henrick,K. and Nakamura,H. (2003) Announcing the worldwide Protein Data Bank. *Nat. Struct. Biol.*, **10**, 980.
3. Deshpande,N., Addess,K.J., Bluhm,W.F., Merino-Ott,J.C., Townsend-Merino,W., Zhang,Q., Knezevich,C., Xie,L., Chen,L., Feng,Z. *et al.* (2005) The RCSB Protein Data Bank: a redesigned query system and relational database based on the mmCIF schema. *Nucleic Acids Res.*, **33**, D233–D237.
4. Prilusky,J. (2003) SeqAlert web site: http://bip.weizmann.ac.il/salertb/main (5 May 2008, date last accessed).
5. National Center for Biotechnology Information web site: http://www.ncbi.nlm.nih.gov (5 May 2008, date last accessed).
6. Thomas,D. and Hansson,D.H. (2006) *Agile Web Development with Rails* Pragmatic Bookshelf, Lewisville, TX.
7. Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.