PDB NEWSLETTER

NUMBER 1 ◆ JANUARY 1999 ◆ RCSB

PUBLISHED QUARTERLY BY THE RESEARCH COLLABORATORY FOR STRUCTURAL BIOINFORMATICS

CONTENTS

Welcome from Helen M. Berman 1
Members of the RCSB 1
Current and Future Status of the PDB. 2
mmCIF and PDB Formats in the RCSB PDB. 3
Open Discussion of the RCSB PDB and Its Role in Modeling and Structure Prediction. 3
New PDB Web Tools. 4
PDB E-mail Discussion Group 5
Linking to the PDB through Your Browser 6
OMG Requests User Input Regarding CORBA Protocol Specification for Macromolecular Structure. 6
RCSB PDB Meetings and Presentations. 7
PDB Staff. 8
Statement of Support 8

PDB SNAPSHOT

9176 released atomic coordinate entries

MOLECULE TYPE

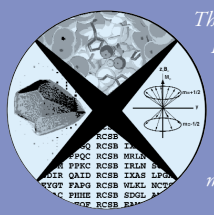
8143 proteins, peptides, and viruses
381 protein/nucleic acid complexes
643 nucleic acids
12 carbohydrates

EXPERIMENTAL TECHNIQUE

206 theoretical modeling
1452 NMR
7521 diffraction and other
2424 structure factor files
522 NMR restraint files

RCSB

SDSC: www.rcsb.org
RUTGERS: rcsb.rutgers.edu
NIST: rcsb.nist.gov
E-MAIL: info@rcsb.org
FTP: ftp.rcsb.org
BNL: www.pdb.bnl.gov



The Research Collaboratory for Structural Bioinformatics (RCSB) is a non-profit consortium dedicated to improving our understanding of the function of biological systems through an understanding of 3-D biological macromolecular structure.

WELCOME FROM HELEN M. BERMAN

Welcome to the first newsletter of the Research Collaboratory for Structural Bioinformatics (RCSB) for the Protein Data Bank (PDB). As you all know, the management of the PDB has changed from Brookhaven National Laboratory (BNL) to the RCSB. This change in management will take place during a transition period that will end October 31, 1999. In this newsletter, we introduce ourselves, our plans for the transition, and our plans for the future.

We have received a variety of responses to this news. Thanks to all of you who have sent your support. We are aware that there are some concerns over the change in management; if you are interested in a particular area that is not addressed in this newsletter or on our Web site at **http://www.rcsb.org/**, please let us know by sending e-mail to **info@rcsb.org**.

We are looking forward to introducing the RCSB system and its many new features. We are particularly excited about a higher, faster throughput of deposited data; a greater number of query capabilities, including more complex and more accurate queries; a uniform archive; dynamic cross-links to other databases; and the availability of structure and sequence neighboring. The PDB data will be stored and mirrored at all three RCSB sites. The RCSB has entered several different collaborations and will be highlighting these projects and other developments in newsletters to come.

Helen M. Berman ✦

MEMBERS OF THE RCSB

The Research Collaboratory for Structural Bioinformatics (RCSB) is a non-profit consortium dedicated to improving our understanding of the function of biological systems through an understanding of 3-D biological macromolecular structure. The PDB is managed by an RCSB Project Team directed by Dr. Helen M. Berman. The Project Team consists of:

HELEN M. BERMAN, a structural biologist, is a Professor II in the Department of Chemistry and a member of the Waksman Institute at Rutgers, the State University of New Jersey. Dr. Berman is the Director Designee of the Protein Data Bank. She has founded and directs the Nucleic Acid Database Project and has been a leader in the national and international mmCIF effort. She is the chair of the International Union of Crystallography Database Committee. Dr. Berman was President of the American Crystallographic Association (ACA) and has served on numerous advisory boards. Dr. Berman was one of the founders of the PDB.

JOHN WESTBROOK, a bioinformaticist, is a Research Associate Professor in the Department of Chemistry at Rutgers, the State University of New Jersey. Dr. Westbrook is the principal architect of the Nucleic Acid Database. He is the author of the Dictionary Definition Language used

by mmCIF and is one of the two Technical Editors for the mmCIF Dictionary. Dr. Westbrook is also a key member of the imgCIF project, which is creating an exchange format for diffraction data. Dr. Westbrook was formerly the Director of Computational Chemistry at Rutgers.

Gary Gilliland, a structural biologist, is the Chief of the Biotechnology Division of the Chemical Science and Technology Laboratory of the National Institute of Science and Technology (NIST) and an Adjunct Professor of the University of Maryland Biotechnology Institute at the Center for Advanced Research in Biotechnology (CARB). Dr. Gilliland founded the Biological Macromolecular Crystallization Database (BMCD) and, for the past 10 years, has organized and been an instructor for the macromolecular crystallography course offered by Cold Spring Harbor Laboratory. He is a former Associate Director of CARB.

Phoebe Fagan, a scientific database developer, has extensive experience in related activities carried out during her tenure in NIST's Standard Reference Data Program. Ms. Fagan is a program manager and has experience in managing database review and quality control. Previously, Fagan headed the database development group within the NIST Standard Reference Data Program and initiated its archive.

Peter Arzberger, a computational biologist, is Executive Director of the National Partnership for Advanced Computational Infrastructure (NPACI), which involves the coordination of software development and implementation activities among 46 institutions via 10 applications and technology thrust areas. Dr. Arzberger was formerly Deputy NSF HPCC Coordinator and Program Director of the Computational Biology and of the Statistics and Probability programs.

Phil Bourne, a structural biologist, is a Senior Principal Scientist at the San Diego Supercomputer Center and an Adjunct Professor in the Department of Pharmacology, UCSD and the Burnham Institute. He has more than 20 years' experience as a small molecule and macromolecular crystallographer and, in the last five years, in bioinformatics. He is chair of the International Union of Crystallography Computing Commission and the American Crystallography Association Data and Computing Committee. He oversees a major NPACI molecular sciences project which is charged with providing the capability to query biological data. ✦

Current and Future Status of the PDB

Status of Data Processing during Transition Period

On January 27, 1999, the RCSB became responsible for the processing of all data received by the PDB. Brookhaven National Laboratory (BNL) will be responsible for processing the files received prior to January 27, 1999.

For now, the process of depositing data will not change. Crystal structures of proteins and all NMR structures should be deposited using either AutoDep or the PDB Deposition Form, which are available at **http://www.pdb.bnl.gov/** and **http://www2.ebi.ac.uk/pdb/**. Crystal structures of nucleic acids should continue to be submitted directly to the Nucleic Acid Database.

After data has been deposited, it will be processed immediately by the RCSB and returned to the author. For most cases, files will be released within one week of their arrival at PDB or immediately following their release from a HOLD status. This time is based upon the extensive testing done by the RCSB over the past three months.

The files released by the RCSB will be in a fully processed and final format.

Structures deposited prior to January 27, 1999, will be processed and released according to the Layered Release Protocol previously described in the PDB Newsletter of October 1997, January 1998, and April 1998. BNL will be responsible for the completion of the processing of all files received up to that point.

In the near future, the RCSB will introduce a new Web-based tool for deposition called the AutoDep Input Tool (ADIT) at the RCSB Web site. The features of ADIT are described below. During the transition period, both AutoDep and ADIT will be available to depositors.

Questions about deposition should be sent to **deposit@rcsb.rutgers.edu**.

Status of Query during Transition Period

During the transition period, which ends in October 1999, the two query engines currently available from BNL (PDBLite and 3DB Browser) will continue to be maintained by BNL and made available from the BNL PDB Web sites. Additionally, new query engines will be introduced at the RCSB Web sites. Anyone using the PDBLite or 3DB Browser query engines at the BNL Web site or any of its mirrors should be able to continue to use those services.

New query engines have been developed by the RCSB, and these will be introduced during the transition year. The aim of these query engines is to provide basic and more complex query capabilities for the PDB in convenient forms. The basic level query engine is called SearchLite. This keyword-based query system is the subject of an article in this newsletter. (See "New PDB Web Tools.") Although the scope of the SearchLite query engine is to provide basic query capability, this engine nevertheless exposes the information content and accompanying services provided by the RCSB PDB. The SearchLite query engine is available from the RCSB PDB Web site, and we encourage users to try it out.

A prototype version of a query engine that supports more advanced queries has been developed, and alpha testing is beginning. This engine will be made available in two phases later this year after a thorough testing and evaluation process.

Status of Distribution during Transition Period

During the transition period, the PDB will be accessible from two Web sites: one reflecting the RCSB PDB and one reflecting the BNL PDB. Both sites will maintain the current version of the PDB FTP archive. The FTP archive will remain in its current form.

The RCSB PDB Web site provides the PDB FTP archive, up-to-

date transition information, and access to new features of the RCSB PDB system as they become available. These new features, described below, will include more complex query searching, a new data deposition tool, and a structure validation server.

The BNL PDB Web site will provide the PDB FTP archive, AutoDep, PDB Lite, and 3DB Browser and its related resources.

Structures will be released weekly by the RCSB to all PDB Web sites.

In the future, additional RCSB mirrors will be made available. ✦

mmCIF and PDB Formats in the RCSB PDB

The RCSB will continue to accept and distribute coordinates in the mmCIF and PDB formats. The PDB format released by the RCSB will be uniform and will continue to use the four-character PDB identifier.

Internally, the RCSB uses the macromolecular Crystallographic Information File (mmCIF)[1] format to process data. mmCIF is the IUCr-approved data representation for macromolecular structures. The mmCIF dictionary, a collection of more than 1,600 definitions, provides the basis for the RCSB's integrated data processing and query system.

Structures can be deposited in PDB format and then are converted to an mmCIF format file using software developed by the RCSB.

mmCIF information and resources are available at **http://ndbserver.rutgers.edu/mmcif/**. ✦

Open Discussion of the RCSB PDB and Its Role in Modeling and Structure Prediction

December 13, 1998, Asilomar, CA

Leveraging the CASP3 meeting in Asilomar, California, the RCSB held a workshop on December 13, 1998. The aim of this workshop was to report on the status and plans for the PDB and to engage a large subset of users of the PDB in open discussions on future direction and policy. The meeting was chaired by Helen Berman and Phil Bourne and was attended by key representatives from Glaxo Wellcome, MRC, Columbia University, NIH, University College London, University of Wisconsin, Rutgers University, UC San Diego, EBI, Molecular Simulations, Inc. and Stanford University. The meeting began with introductions of all participants and a statement of purpose to provide information, address concerns, and gain input.

Helen Berman described the two grand challenges for the RCSB as the goals of (1) relating sequence, structure, and function and (2) helping researchers, educators, and industry. These challenges need to be met in the face of an increase in both the number and variety of users.

[1] *P.E. Bourne, H.M. Berman, B. McMahon, K.D. Watenpaugh, J. Westbrook, and P.M.D. Fitzgerald. The Macromolecular Crystallographic Information File (mmCIF).* Meth. Enzymol. *(1997) 277, 571-590.*

Dr. Berman then provided an overview of all aspects of RCSB PDB. The key features of the RCSB resource were described as rapid and reliable data processing, a commitment to uniform data, versatile query and reporting capability for individual structures across the archive, and links to other data.

RCSB data processing is a single integrated system for deposition, annotation, validation, database management, and archiving. The system is based on a community reviewed standard (CIF) and should scale automatically with new content and technology. The data-processing system is adaptable for different types of molecule (protein, nucleic acid), handles multiple input and output formats, can be used for primary deposition and for creating data uniformity, and can be customized according to the type of experiment (e.g., X-ray, NMR). The deposition system is designed to accept Web input and rapidly return annotated and validated entries. Information provided to the depositor includes information on sequence neighbors, structural neighbors, experimental validation, and a validation summary (ProCheck, Nucheck, SFCheck). In the future, it is expected that the deposition component will be distributed, the ligand database will be integrated, data harvesting approaches will be integrated, and more links to other resources will be added.

In discussion, it was pointed out that there is a strong synergy between data quality and obtaining reliable results from queries. A great deal of difficulty in obtaining reliable information from searches across the database can be avoided if the structure files are processed in a consistent manner. Query features supported by the new search engines include iterative queries, multiple structure analysis, and resources including structure alignments and neighboring.

Electronic distribution of the PDB will continue with SDSC as the primary distribution site and a network of mirrors and partial mirrors. A CD-ROM of the PDB will be distributed from NIST, which will maintain the master archive. Outreach efforts will continue and involve the community of depositors, users, and software developers via the help desk, newsletter, and the Web.

John Westbrook described data-processing requirements and the system being established at the RCSB. Requirements of the data-processing system include the ability to capture information in a flexible and efficient manner, store information in a well defined and uniform manner (facilitating query and exchange), and use methods of information technology that will scale well with volume and content.The data representation is the mmCIF standard, for which the dictionary has currently more than 1,700 definitions. This representation defines relationships between data items, types, range restrictions, and allowed values. The representation uses a simple table-like organization of data and data definitions. Furthermore, the dictionary is fully software-accessible. This standard representation received a detailed community expert review, is maintained by the IUCr, evolves with science, and is a foundation for data exchange and interoperability. A point that was stressed is that dictionaries are key to data processing: They provide a standard electronic description of all terminology, they make software extensible to changes and content in data, and they provide a framework in which to handle reference data (e.g., ligands, modified amino acids).

Phil Bourne provided an overview of the PDB query systems that are being implemented. They include systems that support basic and complex queries. Helge Weissig gave a computer demonstration of the SearchLite system that will be available in the first phase of the RCSB implementation. ✦

New PDB Web Tools

SearchLite: Version 1.0 of the PDB Web Interface

The RCSB has developed a set of query and structure reporting tools for use with the PDB. These tools will continue to evolve with each release of the RCSB PDB Web site. Two sets of upwardly compatible enhancements to the query capability are planned for release later in 1999.

A user accesses the PDB through the Web by making a query and receiving a result. These actions are shown on the left- and right-hand side of Figure 1, respectively. Each of these two components will evolve with every new release of the Web interface. Two further upwardly compatible enhancements are planned for 1999. This article discusses only version 1.0, which is available now. Future newsletters will discuss new versions as they become available. To simplify the discussion, the query and subsequent result are discussed separately.

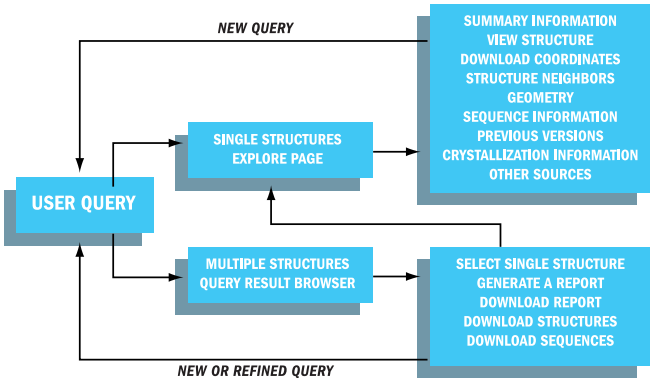


Figure 1. PDB Query and Result Overview

Query

Version 1.0 of the query interface is a keyword query: A user supplies one or more keywords, and all structures that contain those keywords are returned. The search for keywords is performed on the contents of the PDB files. To make a more specific query, a user may pose the query to a subset of PDB record types. For example, a search for Jones will find many structures for which Jones is not an author but which, for example, were solved using the map-fitting programs developed by Alwyn Jones. Restricting the search field to just author returns structures for which Jones was an author according to PDB JRNL and REMARK 1 records.

Multiple keywords can be used as part of the same query. A query term of protein kinase will return all references to structures that contain references to the terms "protein" and "kinase." Since "protein" is a general term and appears in nearly all structures, this becomes a search for "kinase" and will return, for example, "histidine kinase," which is not a member of the protein kinase family. Using the search term "protein kinase" (in quotes) requires the keywords to be contiguous, which will return a set of structures closer to those belonging to the protein kinase family. However, it will also return structures that contain the phrase "protein kinase inhibitor," which may not be desired. There is no NOT clause supplied at present to refine this query to exclude the keyword "inhibitor." Results from such queries can be manually trimmed or added to as desired.

Results

Figure 1 shows two types of results presented for a given query—a single structure result or a multiple structure result. For example, entering a specific PDB identifier will return a single structure, whereas a phrase such as "protein kinase" will return multiple structures. Each case is discussed separately.

Single Structure

A single structure result will return an Explore page. As the name suggests, the page provides the opportunity to further explore several different aspects of the structure, both from the PDB and elsewhere on the Internet. The Explore page presents summary information about the structure, including the release date, author names, compound name, and primary citation, and a dynamic list of options based upon what information is available for the specific structure. For example, if the structure is a crystal structure of a nucleic acid or nucleic acid complex, a link will appear to the Nucleic Acid Database Atlas page (NDB; Berman et al. (1992) *Biophys J.* 63(3), 751-9). The NDB Atlas Entry provides further summary information, hand-curated images, and highly curated coordinate files. Similarly, if a previous version or versions of a structure exist, the page links to a review of the chronology of that particular structure. If no previous versions exist, the link will not appear. Other options that appear on the Explore page are described below.

- **View Structure** provides still and interactive views of the structure. Still images at different sizes and resolution are produced on the fly with Molscript (Kraulis (1991) *J. App. Cryst.* 24, 946-950) and Raster-3D (Merritt and Bacon (1997) *Methods in Enzymology* 277, 505-524). Images are displayed in a standard orientation (right-hand frame x-axis horizontal, y-axis vertical, looking down the z-axis) and use the author-deposited secondary structure assignments to denote secondary structure as ribbons or cylinders. Interactive views use a standard VRML browser and the Molscript VRML output option (Kraulis loc. cit.), Rasmol (R. Sayle and E. Milner-White (1995) *TIBS* 20(9), 374), or the QuickPDB Java applet (Shindyalov and Bourne, unpublished).
- **Download Coordinates** lets users download atomic coordinates in PDB or mmCIF format either as uncompressed ASCII files or compressed with UNIX compress, gzip, or pkzip.
- **Structure Neighbors** provides direct access to reports of other structures exhibiting 3-D structure homology to the structure being explored. Access is provided to the databases of the common classification methods: CATH (Orengo, Michie, Jones, Jones, Swindells, and Thornton (1997) *Structure* 5(8),

1093-1108); CE (Shindyalov and Bourne (1998) *Protein Engineering* 11(9), 739-747); FSSP (Holm and Sander (1998) *Nucl. Acids Res.* 26, 316-319); SCOP (Murzin, Brenner, Hubbard, and Chothia (1995) *J. Mol. Biol.* 247, 536-540) and VAST (Gibrat, Madej and Bryant (1996) *Current Opinion in Structural Biology* 6, 377-385).

- **Geometry** provides a tabular and graphical (requires Rasmol) representation of the stereochemistry of the structure. Both are color-coded to indicate deviation from the standard values reported by Engh and Huber (1991) *Acta Cryst.* A47, 392-400.
- **Sequence Information** reports the size and molecular weight of each chain in the macromolecule as well as the sequence, and in the case of protein chains, the secondary structure assignment according to Kabsch and Sander (1983) *Biopolymers* 22(12), 2577-2607).
- **Previous Versions** provides a graphical summary of the release and withdrawal dates of previous versions of a structure and a tabular comparison of the different features within each version. Stereochemical comparisons between each version of the structure are also presented.
- **Crystallization Information** reports data where available from the Biological Macromolecular Crystallization Database (BMCD; Gilliland et al. (1994) *Acta Cryst.* D50 408-413). These data includes details of the crystal(s), crystallization conditions, and references.
- **Other Sources** provides pointers to other relevant information available via the Web. Again, this is a dynamic list. What appears depends on the structure in question.

Multiple Structures

Queries that return multiple structures are subject to three actions: filtering, downloading, and summarizing. Each is discussed separately.

- **Filtering**—A single structure, a subset of structures, or all structures can be chosen for downloading or summarizing. Alternatively, the selected list of structures can be used as input to a subsequent query, refining the search.
- **Downloading**—The selected list of structures can be downloaded as a set of PDB files or mmCIFs in compressed or uncompressed formats. Sequences only, taken from the PDB SEQRES records, can be downloaded in FASTA format.
- **Summarizing**—Reports on the selected list of structures can be generated. They can be presented as formatted Web pages (HTML) for printing or as tables with delimited fields suitable for loading into a spreadsheet or user-provided program (TEXT). Reports are available for cell constants and space group, primary citation, sequence, experimental technique, and refinement details (where applicable).

Summary

The basic structure for making a query and interpreting results described here will form the basis of more powerful query capabilities in the future. Examples of such queries will be reported in future issues.

Acknowledgements

The RCSB is grateful to Drs. Steven Brenner and Paula Fitzgerald for beta testing version 1.0 of SearchLite.

ADIT: AutoDep Input Tool

A new deposition software tool, called the AutoDep Input Tool (ADIT), has been developed by the RCSB. This tool will be made available at the RCSB Web site for testing during the initial part of the transition period and subsequently made available as an alternative to the existing AutoDep system.

ADIT is a Web-based deposition tool that builds a collection of HTML forms. Each form presents data items selected from a single mmCIF category. The scope of all possible data items available to ADIT is determined by the content of an underlying mmCIF data dictionary. Because this dictionary contains more than 1,600 definitions, the most important function of the view is to present to the ADIT user those data items that are relevant to deposition.

ADIT is undergoing beta testing by members of the crystallographic and NMR spectrographic communities. The test period will extend until the system has been found to be robust by depositors and the archive. We expect this test period to last two to four months. Following testing, ADIT will be made available for general use. ADIT, used with a specialized view for annotation, is used by the RCSB to process structures deposited to the PDB.

ADIT: Validation Server

The RCSB has released its validation server (**http://pdb.rutgers.edu/validate/**) at the RCSB Web site. This validation server can be used to check the format consistency of the coordinates and perform a validation pre-check of the structural features and structure factors before the structure is deposited. The validation server will accept coordinate and structure factor data and produce a report of geometrical and experimental checks. The content and presentation of the validation report is the same as the report produced during the deposition process and can be used by the depositor prior to deposition or at any time during a structure refinement.

Tutorials are available online. ✦

PDB E-mail Discussion Group

As of February 3, 1998, the PDB e-mail discussion group (the PDB listserver) will be maintained by the RCSB PDB staff. The purpose of this discussion group is to facilitate open discussion of topics related to macromolecular structure. These topics are not necessarily limited to the PDB. For example, problems in structure analysis, job ads, and requests for information relating to protein structure are legitimate postings. Personal messages, abuse, and messages of unrelated commercial nature are not acceptable.

At this time, all messages should be posted to the RCSB listserver and postings to the BNL listserver should cease—i.e., people should discontinue using the BNL PDB list and listserver addresses. To ensure uninterrupted service, a grace period of two months

is being provided in which any messages that are inadvertently posted to the BNL PDB listserver will be transferred to the RCSB PDB listserver. The complete mailing list archive will be maintained by the RCSB at **http://www.rcsb.org/pdb/lists/pdb-l archive**.

Current subscribers to the BNL PDB listserver will be automatically transferred to the RCSB PDB listserver, and instructions on posting to the new listserver will be provided by e-mail. New users may subscribe by sending the command

```
subscribe pdb-l
```

in the body of an e-mail message to **majordomo@rcsb.org** Instructions on mailing to the list will then be returned by e-mail. Any user who wishes to be removed from the listserver should send the message

```
unsubscribe pdb-l
```

to **majordomo@rcsb.org**.

The current RCSB operating policy for the discussion group is that it will not be moderated, but only subscribers can to post to it. The purpose of this policy is to permit uncensored access to the group while avoiding external junk mail. Our mechanism for checking that a sender is subscribed is to check the complete e-mail address against the subscription list. This mechanism means that senders must be sure that the address from which they are posting is the same as the one from which they subscribed. In cases where persistent delivery problems are detected, the RCSB reserves the right to remove addresses from the list. ✦

Linking to the PDB through Your Browser

Articles in this newsletter have described the search query capabilities at the RCSB PDB site and have given a glimpse of extended capabilities that will become available in the future. However, a PDB user with a Web resource may wish to build structure retrieval mechanisms into the resource Web pages. To facilitate this type of activity, the CGI scripts at the RCSB PDB Web site may be used directly from a user's own Web page. For single structure retrieval, it is only necessary to provide a URL of the form **http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=XXXX** with the (case-insensitive) XXXX replaced by the four letter identification code for the respective entry. The URL will return to the browser the "Structure Explorer" page corresponding to that entry and this page allows a direct download of the corresponding coordinate file. For example, the URL **http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=2cpk** will return the page on structure entry **2cpk.**

An illustrative html code fragment that shows how to incorporate a CGI script at the RCSB PDB site for returning multiple entries to the browser is

```
<FORM ACTION="http://www.rcsb.org/pdb/cgi/import.cgi"
        METHOD="GET">
<INPUT TYPE="HIDDEN" VALUE="XXXX,XXXX,...."
        NAME="pdbIds">
<INPUT TYPE="SUBMIT" VALUE="Import to PDB">
</FORM>
```

in which the series of XXXX's may be replaced on the Web page by the four-letter identification codes for the required entries. This code fragment will create an "Import to PDB" button with an associated parameter field in which you may enter the PDB entry codes.

As a concrete example, the code fragment

```
<FORM ACTION="http://www.rcsb.org/pdb/cgi/import.
        cgi" METHOD="GET">
<SELECT NAME="pdbIds" SIZE="6" MULTIPLE>
<OPTION VALUE="2hhb">2hhb
<OPTION VALUE="3hhb">3hhb
<OPTION VALUE="4hhb">4hhb
<OPTION VALUE="1ber">1ber
<OPTION VALUE="9icg">9icg
<OPTION VALUE="1a02">1a02
</SELECT><INPUT TYPE="SUBMIT" VALUE="Import to PDB">
```

will create a utility function for selecting and returning information on any or all of the six defined structures.

More information on linking to the PDB may be found at **http://www.rcsb.org/pdb/linking.html**. ✦

OMG Requests User Input Regarding CORBA Protocol Specification for Macromolecular Structure

On January 15, 1999, the Object Management Group (OMG), a non-profit industry consortium that oversees CORBA specifications, voted to issue a Request For Information (RFI) in the area of macromolecular structure. This is the first step in a process to create an official CORBA specification detailing Internet access methods for PDB data. At their upcoming May meeting, interested OMG members, which include representatives from RCSB, will review the RFI responses to evaluate recommendations and requirements for specification proposals.

The OMG consists of more than 800 member institutions, including all major computer companies and many large corporate users of information technology. The OMG developed the Common Object Request Broker Architecture (CORBA) as an open standard to support the development of distributed, object-oriented interfaces independent of hardware platform, operating system, or implementation language.

The OMG relies on domain task forces in areas such as finance, manufacturing, and transportation to create interfaces specific to their industry. By issuing this RFI, the OMG Life Sciences Research Domain Task Force has recognized the potential value of an industry standard for network access to macromolecular structure information.

One benefit of a CORBA macromolecular structure specification would be to allow efficient application access to PDB data via the Internet. Client programs written by end users or third-party software vendors could use the OMG specified interface to obtain the most up-to-date information from PDB CORBA servers.

The complete text of the RFI is available in PDF format. Interested parties are encouraged to respond by the April 26, 1999, deadline.

Further information about CORBA and the OMG is available at **http://www.omg.org/**.✦

RCSB PDB Meetings and Presentations

Planned RCSB Presentations 1999

MEETING/SEMINAR	DATE/LOCATION/ATTENDEE
Finding the Path: Issues of Access to Research Resources *National Science and Technology Council Committee on Science, Subcommittee on Biotechnology*	January 27-28, 1999 *Washington, DC* Helen M. Berman
Argonne National Laboratory	February 11, 1999 *Chicago, IL* Helen M. Berman
NIGMS Structural Genomics Targets Workshop	February 11-12, 1999 *Washington, DC* Helen M. Berman
Biophysical Society Annual Meeting	February 13-17, 1999 *Baltimore, MD* Helen M. Berman
The Association of Biomolecular Resource Facilities '99: Bioinformatics and Biomolecular Technologies	March 19-22, 1999 *Durham, NC* Helen M. Berman and Phil Bourne
The Sealy Center Fourth Symposium on Structural Biology	March 19-21, 1999 *Galveston, TX* Helen M. Berman
Data Mining Course	May 12-23, 1999 *Erice, Italy* Helen M. Berman and Phil Bourne
American Crystallographic Association Annual Meeting	May 22-27, 1999 *Buffalo, NY* Helen M. Berman, Phil Bourne, Gary Gilliland, and John Westbrook
Eleventh Conversation in Biomolecular Stereodynamics	June 15-16, 1999 *Albany, NY* Helen M. Berman
International Union of Crystallography Congress and General Assembly	August 4-13, 1999 *Glasgow, Scotland* Helen M. Berman, Phil Bourne, and John Westbrook

RCSB Workshops

MEETING	DATE/LOCATION/ATTENDEE
Data Processing and Annotation Workshop	October 15-16, 1998 *Piscataway, NJ* Helen M. Berman, Gary Gilliland, and John Westbrook
Meetings with protein modeling groups	December 8-9, 1998 *Piscataway, NJ* Helen M. Berman, John Westbrook, and Phil Bourne December 11, 1998 *San Diego, CA* John Westbrook and Phil Bourne December 12-13, 1998 *Asilomar, CA* Helen M. Berman, Phil Bourne, and John Westbrook

RCSB Presentations (September 1, 1998-January 22, 1999)

MEETING/SEMINAR	DATE/LOCATION/ATTENDEE
CCP4 Data Harvesting Meeting	September 16-19, 1998 *Cambridge, United Kingdom* Helen M. Berman
Structure-based Functional Genomics	October 4-7, 1998 *Avalon, NJ* Helen M. Berman, John Westbrook, Phil Bourne, and Gary Gilliland
1998 Cold Spring Harbor Laboratory (CSHL) Macromolecular Crystallography	October 14-27, 1998 *Cold Spring Harbor, NY* Gary Gilliland, Phil Bourne, and Helen M. Berman
United States National Committee for Crystallography Meeting	November 8, 1998 *Washington, DC* Helen M. Berman
SC98: High Performance Networking and Computing Conference	November 7-13, 1998 *Orlando, FL* Phil Bourne
Molecular Modeling in the Large	December 6-10, 1998 *San Diego, CA* Phil Bourne
RNA Informatics Workshop	December 8-11, 1998 *Arlington, VA* Helen M. Berman
Bristol-Myers Squibb	December 17, 1998 *Princeton, NJ* Helen M. Berman
DOE Computational Structural Biology Research Meeting	December 18-19, 1998 *Monterey, CA* Gary Gilliland
Pacific Symposium on Biocomputing '99	January 4-9, 1999 *Big Island, HI* Phil Bourne
Keystone Symposium: Frontiers of NMR in Molecular Biology VI	January 9-15,1999 *Breckenridge, CO* Gary Gilliland, Diane Hancock
International Symposium on Structural Biology and Genomics	January 11-12, 1999 *Tsukuba, Japan* Helen M. Berman
DOE Contractor and Grantee Genome Progress Workshop	January 12-16, 1999 *Oakland, CA* Gary Gilliland

PDB Staff

RCSB Project Team

Dr. Helen M. Berman
Department of Chemistry
Rutgers University
610 Taylor Road
Piscataway, NJ 08854-8087
732-445-4667
Fax: 732-445-4320
berman@adenine.rutgers.edu

Dr. John Westbrook
Department of Chemistry
Rutgers University
610 Taylor Road
Piscataway, NJ 08854-8087
732-445-4290
Fax: 732-445-4320
jwest@ndb.rutgers.edu

Dr. Gary Gilliland
Center for Advanced Research in Biotechnology
National Institute of Standards and Technology
9600 Gudelsky Drive
Rockville, MD 20850
301-975-2629
Fax: 301-330-3447
gary.gilliland@nist.gov

Phoebe Fagan
Center for Advanced Research in Biotechnology
National Institute of Standards and Technology
9600 Gudelsky Drive
Rockville, MD 20850
301-738-6285
Fax: 301-738-6280
phoebe.fagan@nist.gov

Dr. Peter Arzberger
San Diego Supercomputer Center
University of California, San Diego
9500 Gilman Drive
La Jolla, CA 92093-0505
619-534-5079
Fax: 619-822-0948
parzberg@sdsc.edu

Dr. Phil Bourne
San Diego Supercomputer Center
University of California, San Diego
9500 Gilman Drive
La Jolla, CA 92093-0537
619-534-8301
Fax: 619-822-0873
bourne@sdsc.edu

RCSB Members

NIST

CARB
9600 Gudelsky Drive
Rockville, MD 20850

Dr. Hillary S. Gilson
301-738-6106
Fax: 301-738-6280
hillary.gilson@nist.gov

Dr. Diane Hancock
301-738-6110
Fax: 301-738-6280
diane.hancock@nist.gov

Dr. Narmada Thanki
301-738-6204
Fax: 301-738-6280
narmada@dakota.carb.nist.gov

Dr. Michael Tung
301-738-6284
Fax: 301-738-6280
michael.tung@nist.gov

Dr. Greg Vasquez
301-738-6209
Fax: 301-738-6280
gregory.vasquez@nist.gov

Dr. Lynn Young
301-738-6283
Fax: 301-738-6280
lynn.young@nist.gov

UCSD/SDSC

SDSC
UC San Diego
9500 Gilman Drive
La Jolla, CA 92093

Dr. John Badger
619-822-0936
Fax: 619-822-0873
badger@sdsc.edu

Dr. Douglas S. Greer
619-534-8357
Fax: 619-534-5113
dsg@sdsc.edu

Dr. Michael Gribskov
619-534-8312
Fax: 619-822-0873
gribskov@sdsc.edu

Dr. Arcot K. Rajasekar
619-534-8378
Fax: 619-534-5077
sekar@sdsc.edu

Dr. Lynn F. Ten Eyck
619-534-5141
Fax: 619-534-5113
teneyckl@sdsc.edu

Dr. Helge Weissig
619-534-8399
Fax: 619-822-0873
helgew@sdsc.edu

RUTGERS

Department of Chemistry
Rutgers University
610 Taylor Road
Piscataway, NJ 08854-8087

Kyle Burkhardt
732-445-2200
Fax: 732-445-4320
kburkhar@ndb.rutgers.edu

Dr. Zukang Feng
732-445-0103
Fax: 732-445-4320
zfeng@ndb.rutgers.edu

Dr. Anke Gelbin
860-685-2777
Fax: 860-685-2211
ankeg@ndb.rutgers.edu

Dr. Shri Jain
732-445-0103
Fax: 732-445-4320
sjain@ndb.rutgers.edu

Dr. Rachel Kramer
732-445-2200
Fax: 732-445-4320
kramer@ndb.rutgers.edu

Dr. Bohdan Schneider
732-445-0103
Fax: 732-445-4320
bohdan@ndb.rutgers.edu

Dr. Kata Schneider
732-445-0103
Fax: 732-445-4320
kata@ndb.rutgers.edu

Christine Zardecki
732-445-0103
Fax: 732-445-4320
zardecki@ndb.rutgers.edu

BNL PDB Staff

Joel L. Sussman, *Head*
Enrique E. Abola, *Deputy Head and Head of Scientific Content/Archive Management*
Otto Ritter, *Head of Informatics*

Betty R. Deroski
Arthur Forman
Sabrina Hargrove
Jiansheng Jiang
Jiri Koutnik
Patricia A. Langdon
Michael D. Libeson
Dawei Lin
Nancy O. Manning
John E. McCarthy
Christine Metz
Michael J. Miley
Steve Murtagh
Keith Peters
Janice Picarelli
Jeremy Praissman
Regina K. Shea
Janet L. Sikora
Lu Sun
S. Swaminathan
Latrice Turpin
Wenjuan Xu
Dejun Xuemorphy

Statement of Support *The RCSB's PDB is supported by funds from the National Science Foundation, Department of Energy, and two units of the National Institutes of Health: The National Institute of General Medical Sciences (NIGMS) and the National Library of Medicine (NLM), in addition to resources and staff made available by the host institutions.*

Protein Data Bank
University of California, San Diego
San Diego Supercomputer Center, MC 0537
9500 Gilman Drive
La Jolla, CA 92093-0537

ADDRESS SERVICE REQUESTED