J Struct Funct Genomics (2011) 12:45–54
DOI 10.1007/s10969-011-9106-2

The Structural Biology Knowledgebase: a portal to protein structures, sequences, functions, and methods

Margaret J. Gabanyi · Paul D. Adams · Konstantin Arnold · Lorenza Bordoli · Lester G. Carter · Judith Flippen-Andersen · Lida Gifford · Juergen Haas · Andrei Kouranov · William A. McLaughlin · David I. Micallef · Wladek Minor · Raship Shah · Torsten Schwede · Yi-Ping Tao · John D. Westbrook · Matthew Zimmerman · Helen M. Berman

Received: 22 December 2010 / Accepted: 21 March 2011 / Published online: 7 April 2011

Abstract The Protein Structure Initiative's Structural Biology Knowledgebase (SBKB, URL: http://sbkb.org) is an open web resource designed to turn the products of the structural genomics and structural biology efforts into knowledge that can be used by the biological community to understand living systems and disease. Here we will present examples on how to use the SBKB to enable biological research. For example, a protein sequence or Protein Data Bank (PDB) structure ID search will provide a list of related protein structures in the PDB, associated biological descriptions (annotations), homology models, structural genomics protein target status, experimental protocols, and the ability to order available DNA clones from the PSI:Biology-Materials Repository. A text search will find publication and technology reports resulting from the PSI's high-throughput research efforts. Web tools that aid in research, including a system that accepts protein structure requests from the community, will also be described. Created in collaboration with the Nature Publishing Group, the Structural Biology Knowledgebase monthly update also provides a research library, editorials about new research advances, news, and an events calendar to present a broader view of structural genomics and structural biology.

Keywords Protein · Protein production · Structural biology · Structural databases · Structural genomics · Theoretical models

Abbreviations

3D	Three-dimensional
CGI	Common Gateway Interface
CSMP	PSI Center for Structures of Membrane Proteins
JCMM	PSI Joint Center for Molecular Modeling
JCSG	PSI Joint Center for Structural Genomics
MCSG	PSI Midwest Center for Structural Genomics
NESG	PSI Northeast Structural Genomics Consortium
NMHRCM	PSI New Methods in High-Resolution Comparative Modeling
NIGMS	National Institute of General Medical Sciences

M. J. Gabanyi · J. Flippen-Andersen · A. Kouranov · D. I. Micallef · R. Shah · Y.-P. Tao · J. D. Westbrook · H. M. Berman
Department of Chemistry and Chemical Biology, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA

P. D. Adams · L. G. Carter · L. Gifford
Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720-8110, USA

K. Arnold · L. Bordoli · J. Haas · T. Schwede
Swiss Institute of Bioinformatics and Biozentrum, University of Basel, 4056 Basel, Switzerland

W. A. McLaughlin
Dept of Basic Sciences, The Commonwealth Medical College, Scranton, PA 18510, USA

W. Minor · M. Zimmerman
Deparment of Molecular Physiology and Biology Physics, University of Virginia, Charlottesville, VA 22908, USA

Present Address:
A. Kouranov
Monsanto Company, Chesterfield, MO 63017, USA

H. M. Berman (✉)
Rutgers, The State University of New Jersey, Center for Advanced Biotechnology and Medicine,
679 Hoes Lane, Piscataway, NJ 08854, USA
e-mail: berman@rcsb.rutgers.edu

NPG	Nature Publishing Group
NYSGXRC	PSI New York SGX Research Center for Structural Genomics
PepcDB	Protein expression purification and crystallization database
PDB	Protein Data Bank
PMP	Protein Model Portal
PSI	Protein Structure Initiative
RSS	Really Simple Syndication
SBKB	Structural Biology Knowledgebase
SG	Structural Genomics
SOAP	Simple Object Access Protocol
WSDL	Web Services Description Language

Introduction

When the Protein Structure Initiative (PSI) began in 2000, project results were communicated independently through publications [1], the Protein Data Bank (PDB) [2, 3] structure depositions, individual PSI websites, and presentations at meetings. The PSI-1 and -2 centers had focused their efforts on rapidly determining the structures of proteins on a genomic scale with the emphasis on covering sequence-structure space. To accomplish this, the centers developed a wide variety of tools to carefully select targets and many new technologies to determine and annotate the structures. However, there was no centralized access point for this information. It became clear that in order for the biological community to get the maximum benefit from the products of the PSI effort, a single centralized website was needed. In February 2008, the Structural Genomics Knowledgebase [4] was created which combined the PSI products with data from publicly available biological resources in order to show comprehensive information, including experimental data prior to publication, to jumpstart biological research. In September 2008, the Structural Genomics Knowledgebase entered into a collaboration with the Nature Publishing Group (NPG) to become a "Gateway" site, delivering editorial content about the latest structure and technology reports, a research library, events calendar and science news in addition to the searchable protein database.

In keeping with the recent start of the third phase of the Protein Structure Initiative, PSI:Biology, the website changed its name in August 2010 to the Structural Biology Knowledgebase (SBKB, URL http://sbkb.org), and developed more tools to aid in protein research design. It also improved features to foster collaborations between the biological community and the PSI. Editorial content, displayed on the SBKB as the "Structural Biology Update", is updated on the third Thursday of each month, with database updates occurring on a weekly basis. In this article, we describe all of the features available on the SBKB that can be used to enable biological research, and present some examples of its use.

Navigating the SBKB

The SBKB homepage provides entry points to the following content and functionalities:

Left navigation menu

These menu items give facile access to topics relevant to the biological and biomedical research, such as resources for target, structure, methods, models, and publication information. It also contains menu items for PSI and SBKB information such as FAQs, classroom tutorials, PSI Policies and Reports, links to PSI administrative and Center sites, PSI funding opportunities, and links to NPG online resources.

Structural biology update

This section of the SBKB provides a collection of recent research and technical highlights from the PSI and broader structural biology community, news, upcoming events, and a research library of PSI and other structural biology articles provided by the NPG.

Featured molecule

This highlight [5] consists of simplified explanations and illustrations of interesting PSI protein structures with molecular graphics that allow the user to interact and learn about the molecule's biological role. A new molecule is selected each month.

Query capabilities

A central search box queries the SBKB database for all sequence, structure, functional or technological data related to a sequence, PDB ID code or text string. Text searches will also return matching highlights published as part of the Structural Biology Update.

Nature e-alerts and RSS feeds

Users can subscribe to a monthly electronic Table of Contents alert service (e-alert) on Nature.com with links to the SBKB's latest content. Alternatively, two RSS (Really Simple Syndication) feeds, which can be managed by a user's web browser, keeps readers apprised of (a) the month's Structural Biology Update content and (b) the latest PSI structures released by the PDB.

Community-Nominated Targets portal

The PSI:Biology Network invites the biological community to nominate proteins of biological relevance for structural determination. This proposal system begins the process of matching your project with one of 13 high-throughput and membrane protein structure determination centers to carry out the study. Access is available from either the SBKB homepage or at http://cnt.sbkb.org/CNT.

Sequence Comparison and Analysis tool

The Sequence Comparison and Analysis (SCA) tool consists of the same functionality as the Community-Nominated Target proposal system described previously, but provides an evaluation report to the author only rather than forwarding it to a selection committee. It also supports batch submission to evaluate 100s of sequences. This tool can also be found at http://cnt.sbkb.org/CNT.

Functional Sleuth

Functional Sleuth enables further research for proteins in the Protein Data Bank archive whose functions are unknown or minimally characterized. These "structures of unknown function" (SUFs) are currently organized by source organism, and users can choose to display the SUFs from any phylogenetic level from domain to species. Making a selection on a "tree-of-life" image will launch an interactive tree browser which will further filter your list of SUFs, and right clicking on the name will display the gallery of protein structures. You can download a comma-separated-variable (.csv) file of PDB IDs from your particular gallery, and a full list of all SUF PDB IDs is available at http://sbkb.org/KB/unkstrucs.txt. This feature is updated weekly in conjunction with the weekly PDB release.

Latest PSI statistics

The PSI Network tracks its measure of success and progress though a series of agreed metrics. These metrics, including statistics such as the number of protein structures solved and calculated modeling leverage, were defined by the Goals and Metrics Committee for the PSI-2 Network.

BioSync

BioSync provides technical details about structural biology beamlines at synchrotron radiation facilities. Originally maintained by the RCSB PDB, it is now a part of the SBKB. Progress on future facilities is tracked and information on decommissioned sites is maintained for historical purposes. Links are also provided to related external resources. Summary statistics, based on PDB depositions, are produced and updated weekly. At the beamline level, galleries of structures, tables of citations and general information are also available. Separate statistics are provided for structures solved by structural genomics efforts. This site can be found on the methods hub page from the left navigation menu.

Tutorial and educational resources

The SBKB also has tutorials and classroom exercises to explain how to use the SBKB. These tutorials not only introduce new users to the SBKB, but also give lessons on how to interpret the data being presented. See http://sbkb.org/about/getting_started.html or contact us at comments@sbkb.org for more specific requests.

Content in the Structural Biology Knowledgebase

The SBKB is a web portal [6] that collects different types of structural and methodological information about proteins from the PSI centers and publicly available resources to provide facile access for biomedical researchers and students. The SBKB has also established additional portals (Table 1) to capture data produced by the PSI including protein selection and status (TargetDB [7]), their trial history and protocols (PepcDB [8]), homology models (Protein Model Portal [9]), developed technologies (PSI Technology Portal), and publications (PSI Publications Portal). To goal of these portals is to organize the PSI results and methodological data into a learning and research design optimization resource, especially since much of data is released prior to publication. This information is accessible by searching the SBKB by protein or DNA sequence, PDB structure ID, or text, and more specific queries can be made using the individual portals' websites.

Searching the SBKB

We describe the resources used in SBKB query and reporting mechanism in the context of commonly used examples.

Finding sequence-level or structure-level information about a protein of interest

Conducting a sequence or PDB ID search of the SBKB will yield the following:

Table 1 List of web addresses to access the features and underlying portals of the Structural Biology Knowledgebase

SBKB and portal sites	Data	Web address
Structural Biology Knowledgebase	Query and reporting; editorials and news	http://sbkb.org
TargetDB	Protein sequence selection and progress	http://targetdb.sbkb.org
PepcDB	Target trial information and protocols	http://pepcdb.sbkb.org
Protein Model Portal	Theoretical models	http://www.proteinmodelportal.org
PSI Technology Portal	Technology reports	http://technology.lbl.gov/portal/home
PSI Publications Portal	PSI publications	http://olenka.med.virginia.edu/psi/
BioSync	X-ray crystallography methods	http://biosync.sbkb.org
Community-Nominated Target proposal system	Community requests for protein structures	http://cnt.sbkb.org/CNT

Links to information from publicly available biological resources

The SBKB web portal draws links, identifiers, and annotation values from over 100 PSI and other public resources for protein sequences found in TargetDB and the PDB archive. From the PSI network, information derived from structural and functional annotation tools developed by the four PSI Large-Scale centers such as TOPSAN and others [10–13] is presented. In addition, the external resources include genomic databases [14–18] and model organism databases [19–27], protein primary sequence resources such as Pfam [28], InterPro [29], and others [30–41], structural databases such as the PDB and others [3, 42–52] and structure comparison resources [53–56], functional annotation resources [57–65] evolutionary relationships [66], interactions and pathways [64, 65, 67–78], protein expression profiles [79–98], genetic variations [99–102], disease and pharmacological relationships [103–106], and PubMed [107] for reference information. Access to this collected data is available from a central list appearing next to each target sequence or 3D structure in the SBKB search results. Available for each structure or target results, an example is shown for (PDB + 3cqw) in Fig. 1. By combining all of the total evidence in one view, the user not only reviews what knowledge exists but also can recognize gaps in the literature on that particular topic to enable future hypotheses and studies.

Visualizing protein structures in an interactive viewer

If a 3D structure exists for a protein of interest, the SBKB currently utilizes the molecular viewer FirstGlance [108] to explore the molecule and binding partners (other proteins, ligands, nucleic acids, etc.). This Java-based viewer also provides explanations of the images and representations for the novice user, with display options (such as "color by B-factors/uncertainty") for advanced users as well.

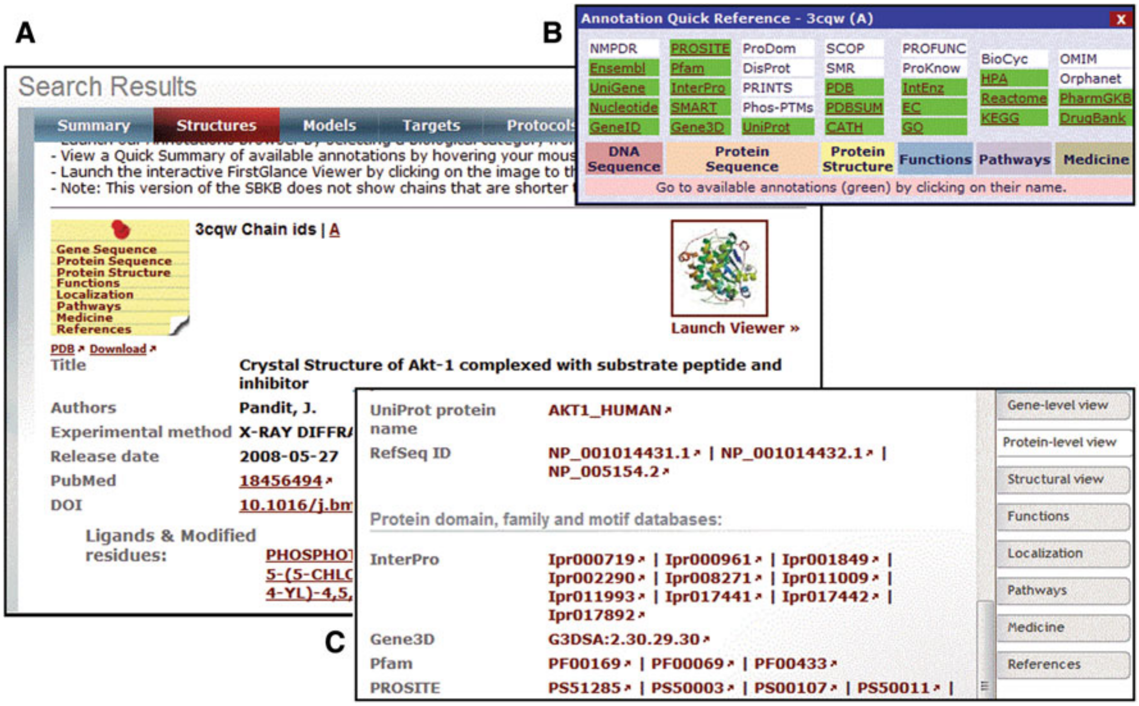
Experimental target tracking databases, TargetDB and PepcDB

The SBKB manages two databases that track structural genomics efforts. TargetDB [7] tracks information on over a quarter million protein sequences, or "targets", that have been selected for structural determination by worldwide structural genomics projects. This information includes sequence and site information, target experimental status, and timestamps for the latest experimental step (cloned, expressed, purified, *etc.*). The Protein Expression, Purification, and Crystallization database [8], PepcDB, provides more detailed information about PSI targets registered in TargetDB (275,000 as of April 2011) by reporting on progress for each experimental trial. Information in PepcDB includes the sequence and site information, rationale for the target's selection (biomedical, community-nominated, technology development, PSI:Biology partnership selection, etc.), each target's experimental histories including individual trial details, the protocols used for protein production and structure determination, and reasons why work was terminated if a 3D structure was not determined. TargetDB and PepcDB are updated weekly with data provided by the PSI centers, and are searchable by TargetID or other popular accession ID (UniProt [30], GenBank [16], PDB [3], see site for details), or filter results by the discussed data attributes. Regular checks are made to ensure that these databases are consistent with relevant information in the PDB and the PSI:Biology-Materials Repository (psimr.asu.edu).

In 2011, a new target history tracking resource that merges TargetDB and PepcDB will be released. Please visit the TargetDB and PepcDB websites for more information on this transition.

Links to the PSI:Biology-Materials Repository

The SBKB also searches its partner PSI resource center, the PSI:Biology-Materials Repository, which stores, maintains and distributes protein expression plasmids and vectors



Go to available annotations (green) by clicking on their name.

- View a Quick Summary of available annotations by hovering your mous
- Launch the interactive FirstGlance Viewer by clicking on the image to th
- Note: This version of the SBKB does not show chains that are shorter

3cqw Chain ids | A

Title	Crystal Structure of Akt-1 complexed with substrate peptide and inhibitor
Authors	Pandit, J.
Experimental method	X-RAY DIFFRA
Release date	2008-05-27
PubMed	18456494
DOI	10.1016/j.bm
Ligands & Modified residues:	PHOSPHO 5-(5-CHLC 4-YL)-4,5,

UniProt protein name	AKT1_HUMAN
RefSeq ID	NP_001014431.1 | NP_001014432.1 | NP_005154.2

Protein domain, family and motif databases:

InterPro	Ipr000719 | Ipr000961 | Ipr001849 | Ipr002290 | Ipr008271 | Ipr011009 | Ipr011993 | Ipr017441 | Ipr017442 | Ipr017892
Gene3D	G3DSA:2.30.29.30
Pfam	PF00169 | PF00069 | PF00433
PROSITE	PS51285 | PS50003 | PS00107 | PS50011 |

Fig. 1 Example of a structure result summary for (PDB + 3cqw). **a** Search results that contain structural matches will present an annotations panel and molecular visualization tool. **b** Hovering the mouse over the Chain ID will launch an annotations quick table, summarizing if popular resources have an annotation for the protein sequence in that structure's chain. **c** All annotations are accessible from a "post-it" with biological subjects on it—clicking on any subject will take you directly to that page of the "annotations notebook"

created by the PSI centers. It currently holds over 40,000 PSI plasmids and nearly 100 empty vectors available for request with an additional PSI plasmids added on a monthly basis.

If a sequence search yielded no experimental structures for the protein of interest, there are other sources of information to help design future experiments:

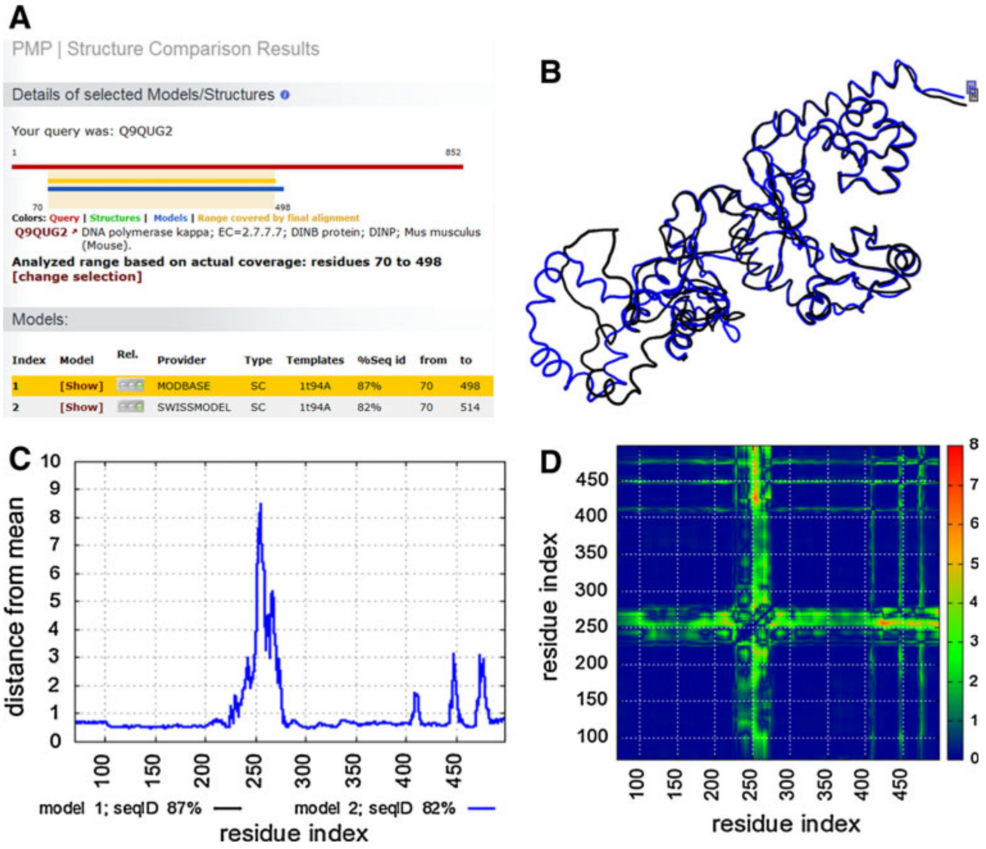
Theoretical models from the Protein Model Portal

The Protein Model Portal (PMP) was created in 2007 as a single entry point to federate structure information from different resources: theoretical structure models from several modeling resources, i.e. the PSI centers CSMP, JCSG, MCSG, NESG, NMHRCM, NYSGXRC, JCMM, and the large scale comparative modeling resources ModBase [51] and SWISS-MODEL Repository [52], and experimental protein structure information from the PDB. One of the challenges in using model information effectively has been to access all models available for a specific protein in heterogeneous formats at different sites using various incompatible accession code systems. To overcome this problem, protein sequences of the UniProt knowledgebase are used as a unified reference system to organize heterogeneous structural information from the different sites. For the first time, it was now possible to access several modeling resources simultaneously, and be able to compare the results computed by different methods in a comprehensive way. An example is shown in Fig. 2 of two theoretical models of a DNA polymerase.

The current release (Mar 2011) consists of 15.1 million comparative protein models for 3.8 million distinct UniProt sequences. Queries can be entered as protein sequences in one-letter code or UniProt accession IDs. Results are presented in a graphical and intuitive way, indicating regions of the proteins where structural information (experimental or theoretical) is available, complemented with functional and domain annotation. Detailed technical information about each model is also provided, such as the date of creation and "latest verification", the template structure and the target-template alignment the model was based on, and the expected model accuracy based on the evolutionary distance between the target and the template. Graphs and images that display and assess model quality and reliability are available, and are an essential component to allow users to select the best available structural information for a specific application.

Community requests to the PSI

Individual investigators are encouraged to nominate protein sequences for structural determination by the PSI:Biology network's high-throughput and membrane protein centers (sbkb.org/KB/psi_centers.html). These sequences should match current PSI goals of biological/



PMP | Structure Comparison Results

Details of selected Models/Structures

Your query was: Q9QUG2

Colors: Query | Structures | Models | Range covered by final alignment

Q9QUG2 DNA polymerase kappa; EC=2.7.7.7; DINB protein; DINP; Mus musculus (Mouse).

Analyzed range based on actual coverage: residues 70 to 498
[change selection]

Models:

Index	Model	Rel.	Provider	Type	Templates	%Seq id	from	to
1	[Show]	●●●	MODBASE	SC	1t94A	87%	70	498
2	[Show]	●●●	SWISSMODEL	SC	1t94A	82%	70	514

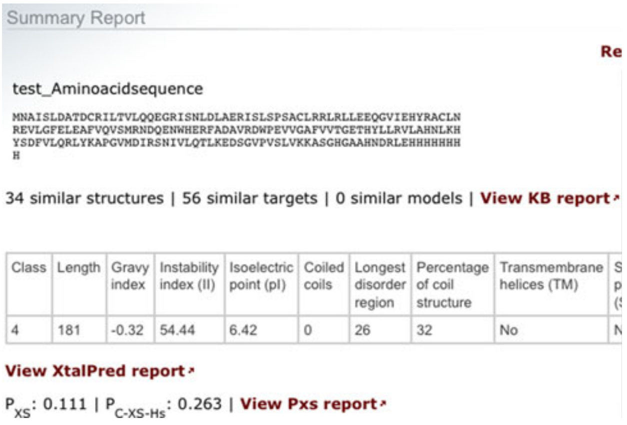
Fig. 2 Example of a Protein Model Portal model comparison analysis. Two theoretical models of a DNA polymerase from mouse (UniProt + Q9QUG2) have been selected for further structural variability analysis by the PMP. **a** The graphical representation indicated the overlapping residue range that can be compared. **b** The superposition of the two models shows that ModBase (*black*) and Swiss-model (*blue*) predict a different structure on the left side of the depiction. **c**. A graph showing the local (per residue) deviation of individual models/ structures from mean of the ensemble of models/structures based on a distance RMSD (dRMSD). **d** The *colors of the spectrum* indicate the degree of variability (based on a weighted dRMSD) among the structure models (*blue* = low, *green* = medium, *red* = high)

biomedical relevance and sequence novelty. The Community-Nominated Target (CNT) proposal system performs an analysis of each proposed target sequence and reports information including crystallization propensities from the PXS [109–111] and XtalPred [112] prediction servers. It also provides a list of similar structures, targets, and homology models found in the SBKB. An example report is shown in Fig. 3. All results are presented to the contributing investigator prior to final submission. Once submitted, the CNT proposals are reviewed by a target selection committee using procedures defined by the PSI and NIGMS [113]. Access to the proposal system is available from the right column of the SBKB homepage or at http://cnt.sbkb.org/CNT/.

Finding new methods and technologies

The PSI centers have developed and utilized many technologies that facilitated structural determination and analysis. A text search of the SBKB allows a user to find these methods.

The PSI Technology Portal



Summary Report

Re

test_Aminoacidsequence

```
MNAISLDATDCRILTVLQQEGRISNLDLAERISLSPSACLRRLRLLEEQGVIEHYRACLN
REVLGFELEAFVQVSMRNDQENWHERFADAVRDWPEVVGAFVVTGETHYLLRVLAHNLKH
YSDFVLQRLYKAPGVMDIRSNIVLQTLKEDSGVPVSLVKKASGHGAAHNDRLEHHHHHHH
H
```

34 similar structures | 56 similar targets | 0 similar models | **View KB report**

Class	Length	Gravy index	Instability index (II)	Isoelectric point (pI)	Coiled coils	Longest disorder region	Percentage of coil structure	Transmembrane helices (TM)	S p (
4	181	-0.32	54.44	6.42	0	26	32	No	N

View XtalPred report

P_{XS}: 0.111 | $P_{C\text{-}XS\text{-}Hs}$: 0.263 | **View Pxs report**

Fig. 3 Example of a Sequence Analysis report. The CNT system and SCA-tool will both perform an SBKB query on given sequences to find the number of similar structures, models, and targets. It will also submit the sequence to the PSI-developed XtalPred and Pxs servers to calculate biophysical parameters and crystallization propensities. In this example, the given sequence was calculated to be a class 4 difficulty (hard), so further construct optimization may be required to ensure a successful structure determination

The PSI Technology Portal provides access to over 200 summaries of key PSI technologies with links to the responsible PSI center or related publication. Reports can be queried by text, by PSI center, or by clicking an image of an experimental pipeline diagram. All tech reports can be "forwarded" to display on a user's social networking sites, such as Facebook, del.icio.us, and others. A YouTube channel to show videos of PSI technologies in action has also been created (www.youtube.com/user/sbkbtech) and a Nature Network discussion forum (network.nature.com/groups/psikb_tech/) was also created as a means of communicating with the user directly.

The PSI Publications Portal

The PSI has published over 1,500 peer-reviewed articles over the past 10 years. These articles are organized by topic (structural or methodological) or by PSI center, and can be searched by author, title, journal, PDB ID, or other attributes. Information presented includes links to the PubMed abstract, related PDB ID, number of times cited, and can also be downloaded in EndNote format. Text searches of the Publications Portal can also be conducted from the central SBKB search box.

Using the Structural Biology update—research library

As part of the Structural Biology update, SBKB and Nature editors add articles published by leading scientific journals that relate to structural biology and structural genomic. Articles are organized into subjects that range from cloning to automated annotation, it should be noted that the articles listed here will require subscription to the relevant journals.

Tools for the Home Lab: a booklet about methods developed by the PSI

The SBKB presented a four-part “PSI-2 Achievements” series from July 2010 to October 2010 detailing advances in the areas of methods, modeling, structures, and outreach. The methods articles from this series summarized the latest strategies developed for protein isolation and structural determination methods. We have assembled these articles into a handy PDF booklet called “Tools for the Home Lab: New Methods from the Protein Structure Initiative”. This book of ideas, which can be downloaded from http://sbkb.org/pdf/PSI-2methods.pdf, can enable any level of wet-bench scientist interested in protein research.

Architecture

The SBKB uses an architecture that collects IDs and selected annotations into a central portal database that is organized to facilitate queries and the addition of new annotations. Each of the SBKB portals developed their underlying architectures with guidelines for common protocols that would allow for easy data exchange with the central SBKB. With this structure, it is possible to query the information from the SBKB home page or from an individual portal module for more specialized queries.

Remotely accessing the SBKB

The content and functionality of the SBKB may be accessed remotely in a variety of ways. Editorial content may be accessed by linking to SBKB pages that are regularly updated (e.g. http://sbkb.org/update/). The RSS feeds are provided to describe both new editorial content and new structure data. Web services are provide to enable program level access to search features of the SBKB. Remote access is provided via both Common Gateway Interface (CGI) and Simple Object Access Protocol (SOAP) protocols. The SBKB also provides an embeddable widget which can be incorporated into a web page at a remote site. The widget provides a dynamically updated display of new articles, features and structure data on the KB site. The protocol interface details and widget installation instructions are described in http://sbkb.org/about/webservices.html.

Conclusions

With the overarching goal of creating new knowledge about the interrelationships of sequence, structure and function, the Structural Biology Knowledgebase captures and highlights the products of the PSI projects for use by the broader biological communities, creates an information repository and web portal that integrates the products of the PSI with publicly available biological information, and encourages collaborative interactions between the PSI and the biological communities. We welcome feedback from the community; users with questions may contact the SBKB at comments@sbkb.org.

Acknowledgments The SBKB is a resource center within the Protein Structure Initiative and is supported by grant U01 GM093324-01 from the National Institute of General Medical Sciences.

References

1. Smith TL (ed) (2000) Structural Genomics Supplement Issue. Nat Struct Biol 7(11s):927–994
2. Berman HM, Henrick K, Nakamura H (2003) Announcing the worldwide Protein Data Bank. Nat Struct Biol 10(12):980
3. Berman HM et al (2000) The protein data bank. Nucleic Acids Res 28:p235–p242

4. Berman HM et al (2009) The protein structure initiative structural genomics knowledgebase. Nucleic Acids Res 37(Database issue):D365–D368
5. Goodsell D (2009) PSI featured molecule series. Available from: http://sbkb.org/KB/structures.jsp
6. Reddy P (2004) In: Bidgoli H (ed) The internet encyclopedia, vol 2 G-O. Wiley, Hoboken, NJ, pp 298–310
7. Chen L et al (2004) TargetDB: a target registration database for structural genomics projects. Bioinformatics 20:2860–2862
8. Kouranov A et al (2006) The RCSB PDB information portal for structural genomics. Nucleic Acids Res 34:D302–D305
9. Arnold K et al (2009) The protein model portal. J Struct Funct Genomics 10(1):1–8
10. The Open Protein Structure Annotation Network (2009). Available from: http://www.topsan.org/
11. Binkowski A (2009) Global protein surface survey. Available from: http://gpss.mcsg.anl.gov/
12. Fischer M (2009) NESG function annotation server. Available from: http://luna.bioc.columbia.edu/honiglab/nesg/cgi-bin/browse.pl
13. Functional Analysis Server at the NYSGXRC (2009). Available from: http://www.nysgxrc.org/functional/
14. Hubbard T et al (2005) Ensembl 2005. Nucleic Acids Res 33(Database issue):D447–D453
15. Flicek P et al (2008) Ensembl 2008. Nucleic Acids Res 36(Database issue):D707–D714
16. Benson DA et al (2008) GenBank. Nucleic Acids Res 36(Database issue):D25–D30
17. Kanehisa M et al (2004) The KEGG resource for deciphering the genome. Nucleic Acids Res 32(Database issue):D277–D280
18. Perriere G, Duret L, Gouy M (2000) HOBACGEN: database system for comparative genomics in bacteria. Genome Res 10(3):379–385
19. Rhee SY et al (2003) The Arabidopsis Information Resource (TAIR): a model organism database providing a centralized, curated gateway to Arabidopsis biology, research materials and community. Nucleic Acids Res 31(1):224–228
20. Guldener U et al (2005) CYGD: the Comprehensive Yeast Genome Database. Nucleic Acids Res 33(Database issue): D364–D368
21. Chisholm RL et al (2006) dictyBase, the model organism database for Dictyostelium discoideum. Nucleic Acids Res 34(Database issue):D423–D427
22. Rudd KE (2000) EcoGene: a genome sequence database for *Escherichia coli* K-12. Nucleic Acids Res 28(1):60–64
23. Crosby MA et al (2007) FlyBase: genomes by the dozen. Nucleic Acids Res 35(Database issue):D486–D491
24. Eppig JT et al (2007) The mouse genome database (MGD): new features facilitating a model system. Nucleic Acids Res 35(Database issue):D630–D637
25. Twigger SN et al (2007) The Rat Genome Database, update 2007–easing the path from disease to data and back again. Nucleic Acids Res 35(Database issue):D658–D662
26. Bieri T et al (2007) WormBase: new content and better access. Nucleic Acids Res 35(Database issue):D506–D510
27. Sprague J et al (2006) The Zebrafish Information Network: the zebrafish model organism database. Nucleic Acids Res 34(Database issue):D581–D585
28. Finn RD et al (2010) The Pfam protein families database. Nucleic Acids Res 38(Database issue):D211–D222
29. Apweiler R et al (2001) The InterPro database, an integrated documentation resource for protein families, domains and functional sites. Nucleic Acids Res 29(1):37–40
30. The UniProt Consortium (2007) The Universal Protein Resource (UniProt). Nucleic Acids Res 35(Database issue):D193–D197
31. Finn RD et al (2008) The Pfam protein families database. Nucleic Acids Res 36(Database issue):D281–D288
32. Wu CH et al (2001) iProClass: an integrated, comprehensive and annotated protein classification database. Nucleic Acids Res 29(1):52–54
33. Attwood TK et al (2003) PRINTS and its automatic supplement, prePRINTS. Nucleic Acids Res 31(1):400–402
34. Gattiker A et al (2003) Automated annotation of microbial proteomes in SWISS-PROT. Comput Biol Chem 27(1):49–58
35. Haft DH et al (2001) TIGRFAMs: a protein family resource for the functional identification of proteins. Nucleic Acids Res 29(1):41–43
36. Haft DH, Selengut JD, White O (2003) The TIGRFAMs database of protein families. Nucleic Acids Res 31(1):371–373
37. Bru C et al (2005) The ProDom database of protein domain families: more emphasis on 3D. Nucleic Acids Res 33(Database issue):D212–D215
38. Hulo N et al (2006) The PROSITE database. Nucleic Acids Res 34(Database issue):D227–D230
39. Mihalek I, Res I, Lichtarge O (2006) Evolutionary trace report_maker: a new type of service for comparative analysis of proteins. Bioinformatics 22(13):1656–1657
40. Pruitt KD, Tatusova T, Maglott DR (2007) NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res 35(Database issue):D61–D65
41. Yeats C et al (2008) Gene3D: comprehensive structural and functional annotation of genomes. Nucleic Acids Res 36(Database issue):D414–D418
42. Laskowski RA et al (1997) PDBsum: a Web-based database of summaries and analyses of all PDB structures. Trends Biochem Sci 22:488–490
43. Orengo CA et al (1999) The CATH Database provides insights into protein structure/function relationships. Nucleic Acids Res 27(1):275–279
44. Cuff AL et al (2009) The CATH classification revisited–architectures reviewed and new ways to characterize structural divergence in superfamilies. Nucleic Acids Res 37(Database issue):D310–D314
45. Andreeva A et al (2004) SCOP database in 2004: refinements integrate structure and sequence family data. Nucleic Acids Res 32(Database issue):D226–D229
46. Murzin AG et al (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures. J Mol Biol 247:536–540
47. Huhne R, Koch FT, Suhnel J (2007) A comparative view at comprehensive information resources on three-dimensional structures of biological macro-molecules. Brief Funct Genomic Proteomic 6(3):220–239
48. Schultz J et al (1998) SMART, a simple modular architecture research tool: identification of signaling domains. Proc Natl Acad Sci U S A 95(11):5857–5864
49. Letunic I, Doerks T, Bork P (2009) SMART 6: recent updates and new developments. Nucleic Acids Res 37(Database issue): D229–D232
50. Ulrich EL et al (2008) BioMagResBank. Nucleic Acids Res 36(Database issue):D402–D408
51. Pieper U et al (2009) MODBASE, a database of annotated comparative protein structure models and associated resources. Nucleic Acids Res 37(Database issue):D347–D354
52. Kiefer F et al (2009) The SWISS-MODEL Repository and associated resources. Nucleic Acids Res 37(Database issue):D387–D392
53. Ye Y, Godzik A (2003) Flexible structure alignment by chaining aligned fragment pairs allowing twists. Bioinformatics 19(Suppl 2): ii246–ii255

54. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32(5):1792–1797
55. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22(22):4673–4680
56. Waterhouse AM et al (2009) Jalview Version 2–a multiple sequence alignment editor and analysis workbench. Bioinformatics 25(9):1189–1191
57. Bairoch A (2000) The ENZYME database in 2000. Nucleic Acids Res 28(1):304–305
58. Hodis E et al (2008) Proteopedia—a scientific 'wiki' bridging the rift between three-dimensional structure and function of biomacromolecules. Genome Biol 9(8):R121
59. Pal D, Eisenberg D (2005) Inference of protein function from protein structure. Structure 13(1):121–130
60. The Gene Ontology Consortium (2000) Gene Ontology: tool for the unification of biology. Nat Genet 25:25–29
61. Laskowski RA, Watson JD, Thornton JM (2005) ProFunc: a server for predicting protein function from 3D structure. Nucleic Acids Res 33(Web Server issue):W89–W93
62. Barthelmes J et al (2007) BRENDA, AMENDA and FRENDA: the enzyme information system in 2007. Nucleic Acids Res 35(Database issue):D511–D514
63. Chen X, Liu M, Gilson MK (2001) BindingDB: a web-accessible molecular recognition database. Comb Chem High Throughput Screen 4(8):719–725
64. Liu T et al (2007) BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities. Nucleic Acids Res 35(Database issue):D198–D201
65. Schaefer CF et al (2009) PID: the Pathway Interaction Database. Nucleic Acids Res 37(Database issue):D674–D679
66. Nikolskaya AN et al (2006) PIRSF family classification system for protein functional and evolutionary analysis. Evol Bioinform Online 2:197–209
67. Salwinski L et al (2004) The database of interacting proteins: 2004 update. Nucleic Acids Res 32(Database issue):D449–D451
68. Shannon P et al (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13(11):2498–2504
69. Chatr-aryamontri A et al (2007) MINT: the Molecular INTeraction database. Nucleic Acids Res 35(Database issue):D572–D574
70. Chautard E et al (2009) MatrixDB, a database focused on extracellular protein-protein and protein-carbohydrate interactions. Bioinformatics 25(5):690–691
71. Goll J et al (2008) MPIDB: the microbial protein interaction database. Bioinformatics 24(15):1743–1744
72. Mewes HW et al (2006) MIPS: analysis and annotation of proteins from whole genomes in 2005. Nucleic Acids Res 34(Database issue):D169–D172
73. Stark C et al (2006) BioGRID: a general repository for interaction datasets. Nucleic Acids Res 34(Database issue):D535–D539
74. Wang R et al (2005) The PDBbind database: methodologies and updates. J Med Chem 48(12):4111–4119
75. Brown KR, Jurisica I (2007) Unequal evolutionary conservation of human protein interactions in interologous networks. Genome Biol 8(5):R95
76. Joshi-Tope G et al (2005) Reactome: a knowledgebase of biological pathways. Nucleic Acids Res 33(Database issue):D428–D432
77. Karp PD et al (2005) Expansion of the BioCyc collection of pathway/genome databases to 160 genomes. Nucleic Acids Res 33(19):6083–6089
78. Kerrien S et al (2007) IntAct–open source resource for molecular interaction data. Nucleic Acids Res 35(Database issue): D561–D565
79. NextBio (2009). Available from: http://www.nextbio.com/
80. Oxford GlycoProteomics 2-DE database (2009). Available from: http://proteomewww.bioch.ox.ac.uk/2d/2d.html
81. Human Cornea 2-DE database (2009). Available from: http://www.cornea-proteomics.com/
82. DOSAC-COBS 2D-PAGE database (2009). Available from: http://www.dosac.unipa.it/2d/
83. Parasite host cell interaction 2D-PAGE database (2009). Available from: http://www.gram.au.dk/2d/2d.html
84. Purkyne Military Medical Academy 2D-PAGE database (2009). Available from: http://www.pmma.pmfhk.cz/2d/2d.html
85. Reproduction 2D-PAGE (2009). Available from: http://reprod.njmu.edu.cn/cgi-bin/2d/2d.cgi
86. Bini L et al (2009) 2D-PAGE database from the Department of Molecular Biology, University of Siena, Italy. Available from: http://www.bio-mol.unisi.it/2d/2d.html
87. Celis JE et al (1998) Human and mouse proteomic databases: novel resources in the protein universe. FEBS Lett 430(1–2): 64–72
88. Evans G et al (1997) Construction of HSC-2DPAGE: a two-dimensional gel electrophoresis database of heart proteins. Electrophoresis 18(3–4):471–479
89. Hoogland C et al (2008) The World-2DPAGE Constellation to promote and publish gel-based proteomics data through the ExPASy server. J Proteomics 71(2):245–248
90. Hoogland C et al (2004) SWISS-2DPAGE, ten years later. Proteomics 4(8):2352–2356
91. Imin N et al (2001) Characterisation of rice anther proteins expressed at the young microspore stage. Proteomics 1(9):1149–1161
92. Li XP et al (1999) A two-dimensional electrophoresis database of rat heart proteins. Electrophoresis 20(4–5):891–897
93. Parkinson H et al (2007) ArrayExpress–a public database of microarray experiments and gene expression profiles. Nucleic Acids Res 35(Database issue):D747–D750
94. Pitarch A et al (2003) Analysis of the Candida albicans proteome. II. Protein information technology on the Net (update 2002). J Chromatogr B Analyt Technol Biomed Life Sci 787(1): 129–148
95. Praz V, Jagannathan V, Bucher P (2004) CleanEx: a database of heterogeneous gene expression data based on a consistent gene nomenclature. Nucleic Acids Res 32(Database issue):D542–D547
96. Uhlen M et al (2005) A human protein atlas for normal and cancer tissues based on antibody proteomics. Mol Cell Proteomics 4(12):1920–1932
97. VanBogelen RA et al (1997) *Escherichia coli* proteome analysis using the gene-protein database. Electrophoresis 18(8): 1243–1251
98. Vijayendran C et al (2007) 2DBase: 2D-PAGE database of *Escherichia coli*. Biochem Biophys Res Commun 363(3): 822–827
99. Thorisson GA et al (2005) The International HapMap Project Web site. Genome Res 15(11):1592–1593
100. Sherry ST et al (2001) dbSNP: the NCBI database of genetic variation. Nucleic Acids Res 29(1):308–311
101. Packer BR et al (2004) SNP500Cancer: a public resource for sequence validation and assay development for genetic variation in candidate genes. Nucleic Acids Res 32(Database issue): D528–D532
102. Karchin R et al (2005) LS-SNP: large-scale annotation of coding non-synonymous SNPs based on multiple information sources. Bioinformatics 21(12):2814–2820

103. Hamosh A et al (2005) Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. Nucleic Acids Res 33(Database issue):D514–D517
104. Thorn CF, Klein TE, Altman RB (2005) PharmGKB: the pharmacogenetics and pharmacogenomics knowledge base. Methods Mol Biol 311:179–191
105. Wishart DS et al (2006) DrugBank: a comprehensive resource for in silico drug discovery and exploration. Nucleic Acids Res 34(Database issue):D668–D672
106. Liem SL (2008) Orphanet and the Dutch Steering Committee Orphan Drugs. A European and Dutch databank of information on rare diseases. Ned Tijdschr Tandheelkd 115(11):621–623
107. Wheeler DL et al (2008) Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 36(Database issue):D13–D21
108. Martz E (2009) FirstGlance in Jmol. Available from: http://firstglance.jmol.org
109. Price WN 2nd et al (2009) Understanding the physical properties that control protein crystallization by analysis of large-scale experimental data. Nat Biotechnol 27(1):51–57
110. Rost B, Yachdav G, Liu J (2004) The PredictProtein server. Nucleic Acids Res 32(Web Server issue):W321–W326
111. Ward JJ et al (2004) Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. J Mol Biol 337(3):635–645
112. Slabinski L et al (2007) XtalPred: a web server for prediction of protein crystallizability. Bioinformatics 23(24):3403–3405
113. Framework for Handling PSI-2 Community Nominated Targets (2008). Available from: http://sbkb.org/KB/index1.jsp?pageshow=62