

## Original article

# PDBWiki: added value through community annotation of the Protein Data Bank

Henning Stehr<sup>1,\*</sup>, Jose M. Duarte<sup>1</sup>, Michael Lappe<sup>1</sup>, Jong Bhak<sup>2,\*</sup> and Dan M. Bolser<sup>1</sup>

<sup>1</sup>Otto-Warburg-Laboratory, Max Planck Institute for Molecular Genetics, Berlin, Germany and <sup>2</sup>Theragen BiO Institute, Suwon, South Korea

\*Corresponding author: Tel: +49 30 804093 14; Fax: +49 30 8413 1960; Email: [stehr@molgen.mpg.de](mailto:stehr@molgen.mpg.de)  
Correspondence may also be addressed to Jong Bhak. Tel: +82 31 888 9311; Fax: +82 31 888 9314; Email: [jongbhak@yahoo.com](mailto:jongbhak@yahoo.com)

Present addresses:

Jose M. Duarte, Paul Scherrer Institut, OFLC/110, 5232 Villigen PSI, Switzerland.

Dan M. Bolser, School of Life Sciences Research, University of Dundee, Dundee, DD1 5EH, UK.

Submitted 2 December 2009; Revised 12 March 2010; Accepted 20 March 2010

The success of community projects such as Wikipedia has recently prompted a discussion about the applicability of such tools in the life sciences. Currently, there are several such 'science-wikis' that aim to collect specialist knowledge from the community into centralized resources. However, there is no consensus about how to achieve this goal. For example, it is not clear how to best integrate data from established, centralized databases with that provided by 'community annotation'. We created PDBWiki, a scientific wiki for the community annotation of protein structures. The wiki consists of one structured page for each entry in the the Protein Data Bank (PDB) and allows the user to attach categorized comments to the entries. Additionally, each page includes a user editable list of cross-references to external resources. As in a database, it is possible to produce tabular reports and 'structure galleries' based on user-defined queries or lists of entries. PDBWiki runs in parallel to the PDB, separating original database content from user annotations. PDBWiki demonstrates how collaboration features can be integrated with primary data from a biological database. It can be used as a system for better understanding how to capture community knowledge in the biological sciences. For users of the PDB, PDBWiki provides a bug-tracker, discussion forum and community annotation system. To date, user participation has been modest, but is increasing. The user editable cross-references section has proven popular, with the number of linked resources more than doubling from 17 originally to 39 today.

Database URL: <http://www.pdbwiki.org>

## Introduction

The number of protein structures deposited in the Protein Data Bank (PDB) (1) has recently surpassed 60 000 and continues to grow at an increasing rate (2, 3). The PDB was one of the first central repositories of biological experimental data and has enabled progress in many areas of structural biology. The atomic resolution data in the PDB is the basis for the analysis of enzyme catalysis (4–6), the study of protein folding (7) and the evolution of protein structure (8–10). Protein structure prediction, for example, has led to functional insights with direct bearing on human health and disease (11).

The PDB provides a primary source of protein structure reference data for the community. Its central importance is

reflected in the number of derived databases that stem from and augment the central archive (see [http://pdbwiki.org/index.php/Template:PDB\\_search](http://pdbwiki.org/index.php/Template:PDB_search)). However, for most protein structure studies, it is not feasible to create a 'database' of annotations to present the findings. Therefore, the results of most studies are typically only found in the literature. For this reason, much of the information about the structures in the PDB cannot be easily integrated back into the central archive.

Although centrally controlled databases of biological data have proven crucially important for research, growth in the volume of data has led to several problems. The data they contain are often static, not permitting user-contributed edits. This means that changes to the

data must be coordinated by typically just a few database curators, and annotations are prone to becoming out of date as new algorithms are developed and new discoveries are made. Similarly, known errors may go uncorrected and special cases in the data may not be deposited adequately. These problems emphasize the growing need for usable community annotation databases.

Scientific wikis have the potential to allow for the community annotation of important biological data (12–14). Such projects aim to integrate specialist knowledge from many disparate groups of researchers, ensuring annotations are accurate and up-to-date. Currently, several such projects exist, either based within Wikipedia (15, 16) or on distinct wikis (17–19). Although there is clearly considerable interest in developing such community annotation systems, there is no general consensus about how best to combine the strengths of community annotation with those of more traditional biological databases (20–23).

Here, we present a new scientific wiki that addresses some of the issues described above. PDBWiki is a system for the annotation of the protein structures deposited in the PDB (1). The system combines a large, static database of heterogeneous biological data with a wiki system for community annotation. The resulting database can serve as a bug-tracker and discussion forum for the structures in the PDB. The database runs in parallel to the central archive, and incorporates user comments in a semi-structured way, allowing for the possibility of incorporating annotations back into the PDB.

## Methods

PDBWiki has been implemented using MediaWiki <http://www.mediawiki.org>, the same software that runs Wikipedia. MediaWiki provides a very stable framework that allowed us to create a very large set of pages with many inbuilt features for community annotation.

Briefly, some of the key features of MediaWiki include the following. ‘Versioning’, every edit made to a page is stored and can be reviewed at any time. Previous versions can be easily reviewed or recovered. ‘Notifications’, pages of interest to the user can be added to a ‘watch-list’, whereby they are automatically notified of all changes by email or RSS. ‘Transparency’, any change to the data in the wiki can be tracked, and the user responsible for a given change can be identified.

Our MediaWiki installation has been extended with several standard and two newly developed extensions. The first new extension provides the ‘user comments’ form that is used on each PDB entry page. The second new extension adjusts the default image functionality of the wiki to manage the image provided for each PDB entry. Among the standard extensions, we use the Dynamic Page

List extension [http://semeb.com/dpldemo/index.php?title=Dynamic\\_Page\\_List](http://semeb.com/dpldemo/index.php?title=Dynamic_Page_List), to generate the structured reports.

Data from the PDB is automatically synchronized with PDBWiki every week. In summary, new PDB entries are added as new pages, updated entries are recreated and obsolete entries are ‘retired’. User comments for obsolete entries remain in the system and are automatically linked to the updated PDB entry (where applicable).

To facilitate the update process, we use the OpenMMS software package <http://openmms.sdsc.edu> to create a relational database from the data in the PDB. Building the relational database from the complete set of entries in the PDB takes roughly 24 h. Using this relational database, we collect the data needed to update PDBWiki using a series of SQL queries. Finally, the wiki is updated using the Python Wikipedia Robot framework <http://pywikipediaibot.sourceforge.net>.

All the source code developed as part of the PDBWiki project has been made available in an open source project repository, hosted by The Bioinformatics Organization ([http://Bioinformatics.Org/project/?group\\_id=936](http://Bioinformatics.Org/project/?group_id=936)).

The article has been written collaboratively using a separate MediaWiki installation.

## Results

### Overview

PDBWiki is centered around the macromolecular structures deposited into the PDB. Each structure in the PDB has a separate page in the wiki, giving a total of over 60 000 structure pages. A structure page is split into three main sections: data, user comments and links (Figure 1). The three main sections are described in the following three paragraphs.

**Data.** The first section of a structure page contains basic information about the macromolecule. This includes the title, author, deposition date and the sequence for each unique chain. Where possible, sequences are linked to their appropriate entries in UniProt or GenBank. The data in this section are obtained directly from the PDB and is automatically updated when the underlying data are changed. These structured data allow for searching and navigation of the archive just like any other web-based database.

**User comments.** Below the static data are the user editable section for community annotation. The user can create or update annotations using the same wiki syntax as Wikipedia or by using a simple form. The semi-structured approach of PDBWiki allows free text comments without restrictions to a particular format, but provides a system for classifying annotations in a hierarchical way (Table 1). For example, the user can tag a comment with the category ‘secondary structure annotation error’. This category is

**1knb**

**CRYSTAL STRUCTURE OF THE RECEPTOR-BINDING DOMAIN OF ADENOVIRUS TYPE 5 FIBER PROTEIN AT 1.7 ANGSTROMS RESOLUTION**

Authors Xia, D. Henry, L.J. Gerard, R.D. Deisenhofer, J.  
 Citation Crystal structure of the receptor-binding domain of adenovirus type 5 fiber protein at 1.7 A resolution.  
 Release date 1995-03-31  
 Exp. Method X-RAY DIFFRACTION  
 Resolution 1.7 Å  
 Classification CELL RECEPTOR RECOGNITION



**Sequence**  
 Chain(s) A (196 residues): [Blast](#) [Uniprot](#)

**User comments**

**Errors in PDB file**  
 Source: [1] [2]  
 In this structure, all the strands are anti-parallel. But in its PDB files:

SHEET	1	V	6	THR	400	TRP	402	0
SHEET	2	V	6	ASP	418	LYS	427	1
SHEET	3	V	6	SER	430	ALA	440	1
SHEET	4	V	6	ASN	479	ASN	482	1
SHEET	5	V	6	LEU	485	THR	486	1
SHEET	6	V	6	TYR	573	TYR	577	1
SHEET	1	R	4	SER	454	PHE	461	0
SHEET	2	R	4	ASN	515	TYR	521	1
SHEET	3	R	4	LYS	528	THR	535	1
SHEET	4	R	4	TYR	550	TRP	556	1

The senses of the strands are all incorrectly labelled as parallel as indicated in col 39-40. They should be labelled as "-1".

Name (required):   
 Comment title:   
 Similar cases:   
 Comment:   
 Category for this comment:

**Links**

Search for **1knb** in: [Edit me!](#)

Databases	Visualization tools	Analysis tools	Quarternary structure
<ul style="list-style-type: none"> <li>Protein Data Bank</li> <li>PDBsum</li> <li>SCOP</li> <li>CATH</li> <li>Pfam</li> <li>JenaLib</li> <li>OCA</li> <li>Proteopedia</li> <li>TOPSAN</li> <li>Electron Density Server</li> <li>add database</li> </ul>	<ul style="list-style-type: none"> <li>RCSB Jmol viewer</li> <li>RCSB KING viewer</li> <li>FirstGlance (Jmol)</li> <li>AstexViewer</li> <li>Polyview 3D (Rendered images)</li> <li>PDBCartoon (secondary structure)</li> <li>FaceView</li> <li>Blue Star Sting</li> <li>Surface topography (CastP)</li> <li>Domain information (XDom)</li> <li>add visualization tool</li> </ul>	<ul style="list-style-type: none"> <li>Structural Neighbours (Dali Database)</li> <li>Likely biochemical function (ProFunc)</li> <li>Functionally important residues (FireDB)</li> <li>Predicted functional sites (WebFeature)</li> <li>Catalytic sites (CSA)</li> <li>WHAT IF quality report</li> <li>ProSA-web quality analysis</li> <li>MolProbity quality analysis</li> <li>NO-Flipper</li> <li>Contact Analysis (CSU)</li> <li>add analysis tool</li> </ul>	<ul style="list-style-type: none"> <li>PDB (Download coordinates)</li> <li>PQS</li> <li>ProtBuD</li> <li>PIQSi</li> <li>3DComplex</li> <li>PISA interfaces</li> <li>PISA monomers</li> <li>PISA assemblies</li> <li>add quarternary structure link</li> </ul>

Categories: PDB annotation error | Annotated PDB entry | KN | PDB entry | X-RAY DIFFRACTION | CELL RECEPTOR RECOGNITION | HUMAN ADENOVIRUS 5 | 1995

**Figure 1.** A typical PDBWiki 'structure page'. There is one structure page for each structure in the Protein Data Bank, with three main sections; data (A), user comments (B) and links (C). See the Overview section of the 'Results' section for a detailed description of these sections.

used to mark structures that have mis-assigned secondary structure elements. Subsequently, this label can be used as a filter to find or exclude structures with this annotation.

**Links.** The third section contains links to external sources of information about the structure. The links are grouped into four subsections: databases, visualization tools, analysis tools and quarternary structure (for websites providing information about the protein's likely quarternary structure). Wherever possible, the links directly point to the relevant data pages or analysis results for the currently viewed structure. In contrast to similar link collections in databases,

such as PDBsum or OCA, in PDBWiki this section is fully user editable. If the user feels that a particular analysis tool is missing she/he can simply add it to the list. Similarly, instead of relying on a particular visualization method chosen by the authors, the user can add a favorite visualization tool and have it available for every structure page visited. Many of the main biomolecular structure web sites carry at the moment reciprocal links back to PDBWiki. Currently these are: RCSB PDB, JenaLib, CSA and PDBsum.

Quarternary structure was regarded as important enough to deserve its own section in the external resource list.

**Table 1.** Part of the hierarchical category system for the pages in PDBWiki (for an up to date list, see: [http://pdbwiki.org/index.php/Category:PDB\\_entry\\_annotation](http://pdbwiki.org/index.php/Category:PDB_entry_annotation))

Category	No. of PDB entries
<i>Biological or biochemical detail:</i> annotations with biological significance	–
Splice variant	12
–	9
<i>Error report:</i> technical errors related to the PDB file	–
Format Error	13
Format in Consistency	11
C $\alpha$ only structure	6
Biounit error	6
Annotation error	3
Data deposition error	3
Experimental error	3
<i>Related work:</i> related articles and resources.	–
–	2
<i>External resource:</i> databases, websites, movies, etc.	–
Movie	4
<i>General comment:</i> anything that does not fit in any other category	–
–	6

The categories shown are those used to classify the ‘user comments’ that have been added to the entries in the PDB. The category system is handled by the MediaWiki software, and is therefore fully user editable.

For example, novice users of PDB structures are often not aware that the coordinates of crystal structures generally represent an asymmetric unit rather than the biological assembly, and even for experts it is often not clear which biological unit is correct (24). The quaternary structure subsection contains links to various resources with information about the putative biological unit, giving users easy access to this kind of information.

### Use cases

PDBWiki provides broadly different functionality to several different target audiences, such as:

#### (1) Bioinformaticians

- Improved navigation of PDB contents by standard categories and customized reports.
- A collection of links to PDB-related databases and tools kept up-to-date by PDBWiki administrators and the user community.
- A central resource for tracking technical problems with individual structures that are relevant for bioinformatics analyses.

#### (2) (Structural) Biologists and Biochemists

- A convenient interface for looking up basic information about PDB structures with additional annotation from the community.

- A forum for discussion about molecular structures.
- Learning about problems with the data quality of individual PDB entries (see for example, user annotations for entry 2hr0).
- Discovering additional information submitted by original authors of a structure (see for example, user annotations for entry 2cme).
- Learning about the studies that a particular structure has been used in (see for example, user annotations for entry 1a32).

#### (3) Crystallographers

- Raising awareness of publications demonstrating important work to a wider community.
- Keeping track of how published structures are being used.
- Publicizing information about structures that do not fit into the standard database format.

### Navigation, querying, and reporting

To help browsing through the structures in the PDB, PDBWiki provides navigational categories that group together structures by name, functional classification, experimental method, host organism, enzyme classification and deposition year. These categories are kept synchronized with those in the PDB. In PDBWiki, the categories to

which a structure belongs appear at the bottom of every structure page. Clicking on a category takes the user to the category page, where all the structures in the category are listed. In addition, structures are categorized by the type of user annotation that has been added. For example all structures with user annotations related to biological units can be found in the category 'Biounit discussion'. Further categories are listed in Table 1.

In addition to free text searches, the user can create custom reports similar to database queries. Both the selection criteria and the displayed information about the structures can be customized. Reports are generated based on combinations of categories or specific lists of protein structures. The result is a wiki page that can be downloaded for further analysis. Figure 2 shows a custom report for proteins in the taxonomic category *Escherichia coli* that are functionally classified as antibiotics. The report includes the PDB code, title, resolution and a thumbnail. An advantage of these reports is that they can be embedded in any wiki page while being always automatically up to date, since they are based on the current state of the data in the wiki. A detailed tutorial of how to create custom reports including examples can be found in the website ([http://pdbwiki.org/index.php/Structure\\_report](http://pdbwiki.org/index.php/Structure_report)).

Unlike the typical search results pages common to most web-based databases, custom reports in PDBWiki are saved as distinct pages in the wiki. In this way, the results of a specific report can be easily linked, shared or discussed. For example, a hand curated list of structures could be created on PDBWiki and used in a custom report. A link to the report could then be posted on a mailing list for discussion. Additionally, reports created using categories will remain up-to-date as new structures in the categories are released, updated or obsoleted.

### Added value provided by PDBWiki

In software engineering, bug-tracking systems have proven to be essential tools for quality management. Despite best efforts during development, effective quality control can only be achieved by contributions from the user community. Such systems are typically implemented as public databases, where users can submit problem reports and feature requests. PDBWiki has been designed to provide similar functionality for the PDB community.






Having a central community resource with quality-related information about protein structures can help bioinformaticians and structural biologists to choose proteins for experiments and to avoid pitfalls with well-known 'problem-structures'. To this end, PDBWiki provides comment features in combination with categories for labeling comments related to quality issues (Table 1), as well as easy access to external analysis tools (Figure 1).

To make full use of the collaborative effort, there should be a way of integrating relevant user submissions back to the original database. We think that this should be carefully done by expert curators. The semi-structured nature of PDBWiki user comments can help the curators of the PDB to monitor new annotations and possibly incorporating them into the original data.

To provide an additional resource for the structural biology community, we have summarized over 2 years worth of posts to a popular community mailing list <http://lists.sds-c.edu/mailman/listinfo.cgi/pdb-l>. These posts have been summarised in the form of a 'Frequently Asked Questions' (FAQ) page. Specifically, we wanted to provide a useful resource that would attract users to the site. The FAQ has been visited over 5000 times and edited 15 times by users originally unknown to the present authors. The FAQ contains over 30 categorized questions, and like the rest of

### Custom report: ANTIBIOTIC and ESCHERICHIA COLI

The query Category:ANTIBIOTIC and Category:ESCHERICHIA COLI returned 5 structures

#	PDB code	Image	Title	Method	Resolution
1	1pp5		Structure of Antibacterial Peptide Microcin J25: a 21-Residue Lariat Protoknot	SOLUTION NMR	0.0
2	1rh1		crystal structure of the cytotoxic bacterial protein colicin B at 2.5 Å resolution	X-RAY DIFFRACTION	2.5
3	2ixw		STRUCTURAL BASIS FOR OMPC MEDIATED ANTIBIOTIC RESISTANCE IN CLINICAL ISOLATES OF ESCHERICHIA COLI	X-RAY DIFFRACTION	2.7
4	3da3		Crystal Structure of Colicin M, A Novel Phosphatase Specifically Imported by Escherichia Coli	X-RAY DIFFRACTION	2.5
5	3da4		Crystal Structure of Colicin M, a Novel Phosphatase Specifically Imported by Escherichia Coli	X-RAY DIFFRACTION	1.7

[Download list of PDB codes](#)

Category: PDBWiki tabular report

**Figure 2.** A custom structure report. The report shows data from proteins in *E. coli* that are classified as antibiotics. The report was generated from a combination of categories and displays a selection of the data available for each structure. A detailed tutorial of how to create custom reports including examples can be found in the website ([http://pdbwiki.org/index.php/Structure\\_report](http://pdbwiki.org/index.php/Structure_report)).

PDBWiki, is free for anyone to edit. One indication of the community value added by this resource is that performing a Google keyword search for the terms 'PDB' and 'FAQ' returns the PDBWiki FAQ page as the first hit.

### Usage

Since November 2007, there have been over 50 000 visits to PDBWiki. Excluding the present authors, there are 11 registered, contributing users and another 20 registered but inactive users. In total, we have collected over 100 user comments, each providing additional information about one of the structures in the PDB. All user comments are assigned to one of several broad categories. The most frequently used categories on the site are: format errors (11), format inconsistencies (16), C $\alpha$  only structures (10) and errors in the proposed biological unit (7). A number of these reported issues have been resolved as part of the ongoing remediation efforts of the PDB. These cases are additionally categorized as 'solved' (4).

One example of an interesting user annotation is that of PDB entry 2hr0. The structure was annotated in December 2008, indicating that a study published the year before in the journal *Nature* had deemed it incorrect (25). In December 2009, it was announced by the University of Alabama that the structure was one of a number of fabricated crystal structures <http://main.uab.edu/Sites/reporter/articles/71570>. PDB has since retracted one of them (1bef). This kind of information is invaluable for structural biologists or bioinformaticians using that structure in their studies.

Further examples of user annotations exemplifying the kinds of comments being submitted and the added value provided by such annotations can be found in the wiki ([http://pdbwiki.org/index.php/Example\\_annotations](http://pdbwiki.org/index.php/Example_annotations)).

One of the most commonly edited sections of the wiki is the collection of user-contributed resources. This section is included on every structure page and provides specific cross-references to resources for the structure (as described above). To date, there have been 13 user edits to this section. Including additional resources that were added by request, the number of cross-references has grown from 17 originally to 39 to date. Resources that have been added include STRAP (26), MolProbity (27), TOPSAN (28) and Polyview 3D (29).

Although the above counts of user and edits may seem low, they are encouraging relative to similar projects where user contributions can be rare. The number of contributions to PDBWiki is continuing to grow, and we believe will continue to do so as the site becomes increasingly well known. We expect to see a positive feedback effect on the amount of user contribution, as increasing contributions will in turn increase the value of the service the site provides.

## Discussion

Community data curation promises to be a solution to the problem of coping with the increasing size and complexity of biological data. The challenge is to make use of the 'wisdom of the many' without compromising the advantages of central, trusted and manually curated databases. In Wikipedia, most edits are small and come from many different users, yet this activity accounts for the majority of the content added to the site. However, this situation relies on a well-established infrastructure on which the community can build (30).

Here, we present an easy-to-use system that combines the data from the PDB with community annotation features. The design, based on the well-known MediaWiki software, keeps the barrier for user contribution low. We address the issues of data integrity and trust by clearly separating the original database content from the user annotations. To organize the user-contributed data, we provide a category system that will help curators to incorporate community suggestions back into the original database.

PDBWiki shares many similarities with Wikipedia. Although the common user interface and reuse of open source code have clear benefits (see the 'Methods' section), the value of PDBWiki lies in its differences from Wikipedia. Specifically, PDBWiki combines semi-structured comments with data from an external, authoritative database.

PDBWiki was always envisioned as a community project. All the source code for the site is freely available in an SVN repository <svn://bioinformatics.org/svnroot/pdbwiki> and developer discussion is publicly archived on the [pdbwiki-devel](http://www.bioinformatics.org/pipermail/pdbwiki-devel) mailing list <http://www.bioinformatics.org/pipermail/pdbwiki-devel>. The wiki includes a page dedicated to 'feature requests' <http://pdbwiki.org/index.php/Development>, where users can add and review suggestions for new features. We plan to keep working on the website, improving the existing functionality and implementing new ideas.

### Related work

Similar projects to PDBWiki have recently appeared. Proteopedia (19) is a wiki-based system for the dissemination of knowledge about macromolecular structures to a broad scientific audience. As with PDBWiki, Proteopedia consists of one page per PDB entry that can be edited by users to contribute information about the particular structure. A key feature of Proteopedia is the Scene Authoring Tool, that can be used to easily create so-called molecular scenes that contribute by illustrating specific points about the structures. They focus on being an educational resource and dissemination tool for structural biology, while in the current project, our main focus is on annotation for the 'end users' of the structural data. For example, user comments in PDBWiki are linked to entries via categories.

Proteopedia, in contrast, has longer, more free ranging educational articles that span several entries.

Although the initial motivation for PDBWiki and Proteopedia was similar, it has become apparent that the two projects have developed in different directions. In broad terms, PDBWiki functions as a 'community discussion forum' and Proteopedia as a 'community education portal'. In future we expect this division to continue, with PDBWiki tending towards integration with the PDB and Proteopedia tending towards integration with Wikipedia.

The Open Protein Structure Annotation Network (TOPSAN) (28) can be viewed as another similar project. TOPSAN stems from the Protein Structure Initiative (a Structural Genomics project) and was created with the intention of enriching the knowledge of poorly annotated protein structures solved by Structural Genomics projects. Structures from both the PDB and the Structural Genomics projects are annotated using a wiki-style system. Thus, the key difference from PDBWiki is the scope: targets from Structural Genomics versus the whole of the PDB.

There exists a number of other community-based biological curation efforts, for example, Human Proteinpedia (23), Wikiproteins (18) and Gene Wiki (16). Although they are all based on the idea of user-contributed annotations they differ significantly in their goals and scopes from the present project.

### Community annotation

The life sciences community is still in an early stage of experimenting with different approaches to community annotation. Some databases have already started to provide user feedback systems into their interfaces (31, 32). Other projects seek to build on the established authority and user base of Wikipedia by creating sub-projects within Wikipedia itself (16).

The size of the user base and the amount of content varies for the different projects. The question arises as to why some projects attract more participating users than others. We can not give a final answer to this question, but we have observed that for collaborative projects such as wikis to be successful, there are two key factors. (i) The project should be focused, such that the potential benefit of the collaborative effort is apparent for the target audience. (ii) The project should provide an immediate value for the users to encourage them to visit and eventually contribute.

There are clear examples of successful scientific wiki projects. To mention just two examples, WikiPathways (17) currently contains 1269 annotated biological pathways, maintained by 810 users, and the Molecular and Cellular Biology project within Wikipedia maintains nearly 20 000 articles with 260 registered participants.

The success of these projects shows that there are enough scientists who are willing to contribute to

collaborative efforts. These projects are able to sustain contribution, even without traditional mechanisms of reward.

## Conclusion

Here, we present a wiki system that allows community annotation of PDB structures. Unlike a database, PDBWiki allows free text comments and unlike traditional wikis, it provides structured classification of the contents. So far, our focus has been on technical comments for users working with PDB structures but the system is not restricted to this use.

The positive feedback we have received about PDBWiki and the growing number of users and annotations are encouraging. We continue to make the site easier to use and to provide value in the form of content to attract more users and contributors. The forms extension we developed, allows users to enter comments with minimal effort through a webform without having to know wiki syntax as in other wikis. This extension also allows easy semantic classification of the submitted comments.

We believe that ultimately, the development of new software, specifically dedicated to community annotation of biological databases combined with the authority of the established resources, such as the PDB, will provide the impetus to achieve the full potential of community annotation.

Many newly developed databases already incorporate explicit mechanisms for user contribution, beyond simple email feedback (32). This shows that wikis or some variation of the wiki idea will play an important role in computational biology in the future. In the present work, we not only demonstrate what can be done in this regard with current technology but also present a valuable resource for the structural community that provides unique functionality and unique user-submitted content that can currently not be found anywhere else.

## Acknowledgements

The article was collaboratively written using the MediaWiki software package <http://www.mediawiki.org>. The authors would like to thank Maryana Bhak for help proofreading and correcting a draft of the article. The authors would also like to acknowledge the Max Planck Institute for Molecular Genetics for continued support.

## Funding

Funding for open access charge: Otto-Warburg-Laboratory, Max Planck Institute for Molecular Genetics, Berlin, Germany.

Conflict of interest. None declared.

## References

1. Berman, H.M., Westbrook, J., Feng, Z. et al. (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242.
2. Holm, L. and Sander, C. (1998) Dictionary of recurrent domains in protein structures. *Proteins*, **33**, 88–96.
3. Levitt, M. (2007) Growth of novel protein structural data. *Proc. Natl Acad. Sci. USA*, **104**, 3183–3188.
4. Blake, C.C., Koenig, D.F., Mair, G.A. et al. (1965) Structure of hen egg-white lysozyme. A three-dimensional Fourier synthesis at 2 Angstrom resolution. *Nature*, **206**, 757–761.
5. Johnson, L.N. and Phillips, D.C. (1965) Structure of some crystalline lysozyme-inhibitor complexes determined by X-ray analysis at 6 Angstrom resolution. *Nature*, **206**, 761–763.
6. Monod, J., Wyman, J. and Changeux, J.P. (1965) On the nature of allosteric transitions: a plausible model. *J. Mol. Biol.*, **12**, 88–118.
7. Levitt, M. and Warshel, A. (1975) Computer simulation of protein folding. *Nature*, **253**, 694–698.
8. Perutz, M.F. (1962) Relation between structure and sequence of haemoglobin. *Nature*, **194**, 914–917.
9. Perutz, M.F., Kendrew, J.C. and Watson, H.C. (1965) Structure and function of haemoglobin ii. some relations between polypeptide chain configuration and amino acid sequence. *J. Mol. Biol.*, **13**, 669–678.
10. Murzin, A.G., Brenner, S.E., Hubbard, T. et al. (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.*, **247**, 536–540.
11. Zhou, H.X. (2004) Improving the understanding of human genetic diseases through predictions of protein structures and protein-protein interaction sites. *Curr. Med. Chem.*, **11**, 539–549.
12. Gong, S., Kim, T.H., Oh, J.S. et al. (2006) BioCC: an openfree hyper-text bio community cluster for biology. *Genomics Informatics*, **4**, 125–128.
13. Salzberg, S.L. (2007) Genome re-annotation: a wiki solution? *Genome Biol.*, **8**, 102.
14. Waldrop, M. (2008) Big data: Wikiomics. *Nature*, **455**, 22–25.
15. Daub, J., Gardner, P.P., Tate, J. et al. (2008) The RNA WikiProject: community annotation of RNA families. *RNA*, **14**, 2462–2464.
16. Huss, J.W., Orozco, C., Goodale, J. et al. (2008) A gene wiki for community annotation of gene function. *PLoS Biol.*, **6**, e175.
17. Pico, A.R., Kelder, T., van Iersel, M.P. et al. (2008) WikiPathways: pathway editing for the people. *PLoS Biol.*, **6**, e184.
18. Mons, B., Ashburner, M., Chichester, C. et al. (2008) Calling on a million minds for community annotation in WikiProteins. *Genome Biol.*, **9**, R89.
19. Hodis, E., Prilusky, J., Martz, E. et al. (2008) Proteopedia a scientific 'wiki' bridging the rift between three-dimensional structure and function of biomacromolecules. *Genome Biol.*, **9**, R121.
20. Hubbard, T. and Birney, E. (2000) Open annotation offers a democratic solution to genome sequencing. *Nature*, **403**, 825.
21. Arshinoff, B.I., Suen, G., Just, E.M. et al. (2007) Xanthusbase: adapting wikipedia principles to a model organism database. *Nucleic Acids Res.*, **35**, D422–426.
22. Pennisi, E. (2008) DNA data. Proposal to 'Wikify' GenBank meets stiff resistance. *Science*, **319**, 1598–1599.
23. Mathivanan, S., Ahmed, M., Ahn, N.G. et al. (2008) Human Proteinpedia enables sharing of human protein data. *Nat. Biotechnol.*, **26**, 164–167.
24. Henrick, K. and Thornton, J.M. (1998) PQS: a protein quaternary structure file server. *Trends Biochem. Sci.*, **23**, 358–361.
25. Janssen, B.J., Read, R.J., Brunger, A.T. et al. (2007) Crystallography: crystallographic evidence for deviating c3b structure. *Nature*, **448**, 1–2.
26. Gille, C. and Frmmel, C. (2001) STRAP: editor for STRuctural Alignments of Proteins. *Bioinformatics*, **17**, 377–378.
27. Davis, I.W., Leaver-Fay, A.V., Chen, B. et al. (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.*, **35**, W375–W383.
28. Krishna, S.S., Weekes, D., Edwards, C.X. et al. (2006) The open protein structure annotation network. In: *Automated Function Prediction Conference 2006 Abstract Book*, San Diego, USA, pp. 89–90.
29. Porollo, A. and Meller, J. (2007) Versatile annotation and publication quality visualization of protein complexes using POLYVIEW-3D. *BMC Bioinformatics*, **8**, 316.
30. Kittur, A., Chi, E.H., Pendleton, B.A. et al. (2007) Power of the few vs. wisdom of the crowd: Wikipedia and the rise of the bourgeoisie. In: *25th Annual ACM Conference on Human Factors in Computing Systems (CHI 2007)*, San Jose, USA.
31. Bairoch, A., Bougueleret, L., Altairac, S. et al. (2009) The Universal Protein Resource (UniProt) 2009. *Nucleic Acids Res.*, **37**, D169–D174.
32. Menda, N., Buels, R.M., Tecle, I. et al. (2008) A community-based annotation framework for linking solanaceae genomes with phenomes. *Plant Physiol.*, **147**, 1788–1799.