RCSB Protein Data Bank: A Resource for Chemical, Biochemical, and Structural Explorations of Large and Small Biomolecules

RCSB Protein Data Bank: a resource for chemical, biochemical, and structural explorations of large and small biomolecules

Christine Zardecki,*[1] Shuchismita Dutta,[1] David S. Goodsell,[2] Maria Voigt,[1] Stephen K. Burley[1]

[1]Center for Integrative Proteomics Research and Department of Chemistry and Chemical Biology, 174 Frelinghuysen Road, Piscataway, NJ

[2]Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, CA

ABSTRACT

The RCSB Protein Data Bank supports scientific research and education worldwide by providing access to annotated information about three-dimensional (3D) structures of macromolecules (*e.g.* nucleic acids, proteins), and associated small molecules (*e.g.* drugs, cofactors, inhibitors) in the Protein Data Bank (PDB) archive. Researchers, educators, and students use RCSB PDB resources to study the shape and interactions of biological molecules and their implications in molecular biology, medicine, biotechnology, and beyond.

RCSB PDB supports development of standards for data deposition, representation, annotation, and validation of atomic structural data obtained from various experimental methods. Uniform representation of PDB data is essential for providing consistent search and analysis capabilities for all PDB users, from beginning students to domain experts.

The RCSB PDB website provides tools for searching, visualizing, and analyzing PDB data, including easy exploration of chemical interactions that stabilize macromolecules and play important roles in their interactions and functions. In addition, educational resources are available for free and unrestricted use in the classroom for exploring chemistry and biology at the molecular level.

KEYWORDS

Audience: General Public; High School / Introductory Chemistry; First-Year Undergraduate / General; Graduate Education / Research

Domain: Biochemistry; Interdisciplinary / Multidisciplinary

Pedagogy: Internet / Web-Based Learning

Topic: Nucleic Acids / DNA / RNA; Proteins / Peptides; X-ray Crystallography

The Protein Data Bank (PDB) is the first open access digital resource in biology for sharing three-dimensional (3D) protein structures.[1] The PDB was established in 1971 with 7 structures, and has grown exponentially to provide access to more than 113,000 entries of natural and designed macromolecules (proteins, nucleic acids and carbohydrates), more than 84,000 of which are complexed with small chemical components (solvent molecules, ions, cofactors, inhibitors, and drugs). Originally, PDB was a resource designed for the structural biology community, but through the years, its utility has grown and the PDB users now include biologists, software developers, computational and other scientists, bioinformaticians, students, educators and the general public.

The PDB archive of data files is one of the most heavily used biological data resources worldwide. In 2014, more than 505,000,000 atomic coordinate and experimental data files were downloaded for research and education. These downloads also include routine downloads by pharmaceutical and biotechnology companies for use in proprietary drug discovery efforts. A huge number of free resources and tools utilize PDB data to serve their users. These range from educational resources such as the NIH 3D Print Exchange[2] and Proteopedia;[3] molecular viewers including Jmol/JSmol,[4] Chimera,[5] Pymol;[6] and many scientific research tools and databases.[7]

The PDB archive is managed by a collection of regional data centers, called the Worldwide Protein Data Bank (wwPDB)[8], spread across the United States, Europe, and Japan. wwPDB centers collaborate on data deposition and annotation/validation practices. Each member hosts a distribution center, and provides tools for access and usage. RCSB PDB, based at Rutgers, The State University of New Jersey and the University of California, San Diego, is focused on providing resources for research and education.[9] As part of the wwPDB, RCSB PDB members curate PDB data and develop data standards and software for the deposition and annotation pipeline. RCSB PDB also aims to enable breakthroughs in scientific inquiry, medicine, drug discovery, and technology by offering tools that provide rich structural views of biological molecules and systems. In addition to supporting biological and chemical learning, RCSB PDB is an exemplar of the new discipline of data science; it provides a glimpse into science history, and serves as a resource for developing database query and analysis skills.

In this paper, we describe data annotation practices, and highlight RCSB PDB resources available for query and analysis, and education.

ARCHIVING PDB DATA: DEPOSITION, ANNOTATION, AND VALIDATION

The PDB archive includes 3D structures of macromolecules (primarily proteins, DNA and RNA) as determined by experiments using X-ray crystallography, nuclear magnetic resonance (NMR), and/or 3D electron microscopy (3DEM). For each new structure, researchers submit atomic coordinates, experimental data, and molecular information using specialized tools, and wwPDB biocurators then review, annotate and validate the entry. Atomic coordinates are checked for consistency with the known sequence of the macromolecule and chemical structure of small molecules, and biological assemblies are defined and annotated. The entry is also extensively annotated with experimental information and cross-referenced to related entries and external resources. The wwPDB collaborates closely with archives that maintain related data,

including the Cambridge Structural Database of small molecule crystal structures,[10] and the EMDataBank for 3D Electron Microscopy maps and models.[11]

Uniform annotation of PDB data enables consistent searching and analysis across the archive. To facilitate uniformity, PDB curation relies upon standard data dictionaries that define the representation of all components in the entry (Table 1). While the PDB Exchange (PDBx)[12] and macromolecular Crystallographic Information File (mmCIF)[13] data dictionaries provide the bases for internal data cross-referencing, processing, annotation, validation and database management operations, the Chemical Component Dictionary[14] describes all standard and modified amino acids/nucleotides, small molecule ligands, and solvent molecules. All chemical components are checked against this dictionary during annotation. Protein and nucleic acid polymers can be built by linking together individual chemical components in a specified order denoted by the polymer sequence. Specialized molecules with unusual chemistries and interesting biological and pharmaceutical functions, such as peptide-like inhibitors and many antibiotics, are included in the Biologically Interesting molecule Reference Dictionary.[15] Use of these dictionaries enables specialized query and access to small molecule information, from specialized information pages for all ligands in the PDB to tools for visualizing ligand-protein interactions.

Table 1. Data Dictionaries used in PDB biocuration

Resource	Description
PDB Exchange (PDBx)/macromolecular Crystallographic Information File (mmCIF) Dictionary http://mmcif.wwpdb.org (accessed 6 Nov 2015)	Crystallographic data dictionaries with extensions describing NMR, 3DEM, and protein production, contains >4300 data items
Chemical Component Dictionary (CCD)	Each chemical definition includes descriptions of chemical properties such as

http://www.wwpdb.org/data/ccd (accessed 6 Nov 2015)	stereochemical assignments, chemical descriptors (SMILES[16] & InChI[17]), systematic chemical names, and idealized coordinates (generated using Molecular Networks' Corina, and if there are issues, OpenEye's OMEGA). Contains > 18,000 small molecules
Biologically Interesting molecule Reference Dictionary (BIRD) http://www.wwpdb.org/data/bird (accessed 6 Nov 2015)	Molecular weight and formula, polymer sequence and connectivity, descriptions of structural features and functional classification, natural source (if any), and external references to corresponding UniProt or Norine entries. Contains ~750 small molecules

wwPDB uses community-accepted standards to "validate" deposited data, and produces reports that provide an assessment of structure quality based upon geometric and experimental data validation. wwPDB has convened method-specific Validation Task Forces[18] to develop recommendations for validation standards and software for use in annotation.[19] During annotation, validation reports are provided to the depositor to highlight any areas of concern. Journal editors may request these reports from authors to inform manuscript review.

The PDB archive includes structural information with a wide range of quality, due to the many challenges inherent in the experimental methods, and the nature of the molecule(s) or complex(es) being studied. Validation reports for all PDB structures determined by X-ray crystallography include a "slider" graphic (Figure 1) to summarize the quality of the determined structure as compared with other structures in the archive. These graphics are displayed on the RCSB PDB website to help users find the structures of highest quality and to provide a warning to be critical when using structures with less experimental support.

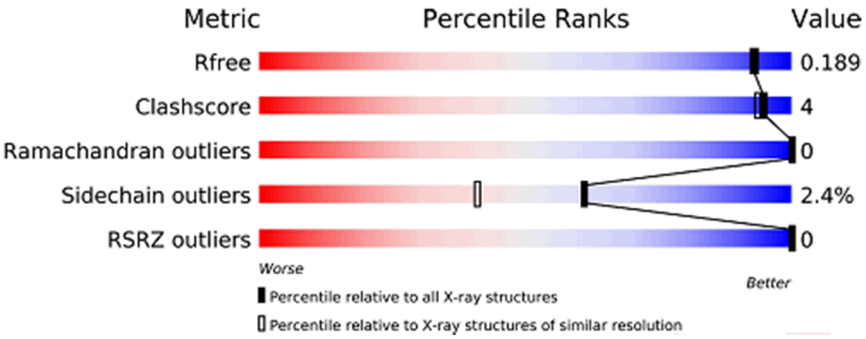

Figure 1. Validation Report slider graphic indicates the quality of an entry as compared with other PDB entries. Shown is the slider image for an entry with better overall quality (PDB entry 1CBS, a small protein with a ligand at 1.8 Å resolution).[20]

QUERY, REPORTING, AND ANALYSIS

The RCSB PDB website[21] integrates PDB data, related information about the structure from external scientific resources, and pre-calculated comparative and statistical information for query, analysis, and visualization.[22] On average, the website is accessed by ~325,000 unique users every month from ~190 countries. The top search bar supports simple keyword searches (ID, author, molecule name, chemical name), and suggests results options organized by different categories, including organism, molecule name, or experimental technique. Advanced searching allows users to combine searches for many specific data items, such as molecule name, authors, experimental techniques, and resolution. Browsers are available to find PDB structures organized using data annotations from external resources (e.g., Gene Ontology terms describing biological process, cellular component, and molecular function;[23] Enzyme Classification;[24] the World Health Organization Collaborating Centre's Anatomical Therapeutic Chemical (ATC) Classification System); or by exploring drill-down distributions of standard characteristics (e.g., polymer type, organism, resolution)). Since general searches, such as for "hemoglobin," can return hundreds of matching structures, a variety of tools are available to help narrow the focus of the inquiry. For example, search autosuggestions, query refinement options, and sorting results by most

recently solved entry, molecular weight, and resolution can help guide users to structures of higher quality and relevance. When available, searches will also return corresponding educational *Molecule of the Month* features, which are described in more detail below.

Every entry has a Structure Summary page that provides access to many aspects of the structure (Figure 2). Interactive 3D viewers, including Jmol/JSmol and Ligand Explorer[25] can be used to rotate the molecule, select specific residues, and highlight ligand-protein interactions.[4] Many of the data items shown can be used to query for other entries with the same data (e.g., sequence database reference, specific chemical component). The entry's corresponding validation slider described above (Figure 1) is also displayed. Additional links provide related information from external scientific sources, such as functional annotations from CATH[26] and SCOP[27], and sequence information from UniProt.[28]

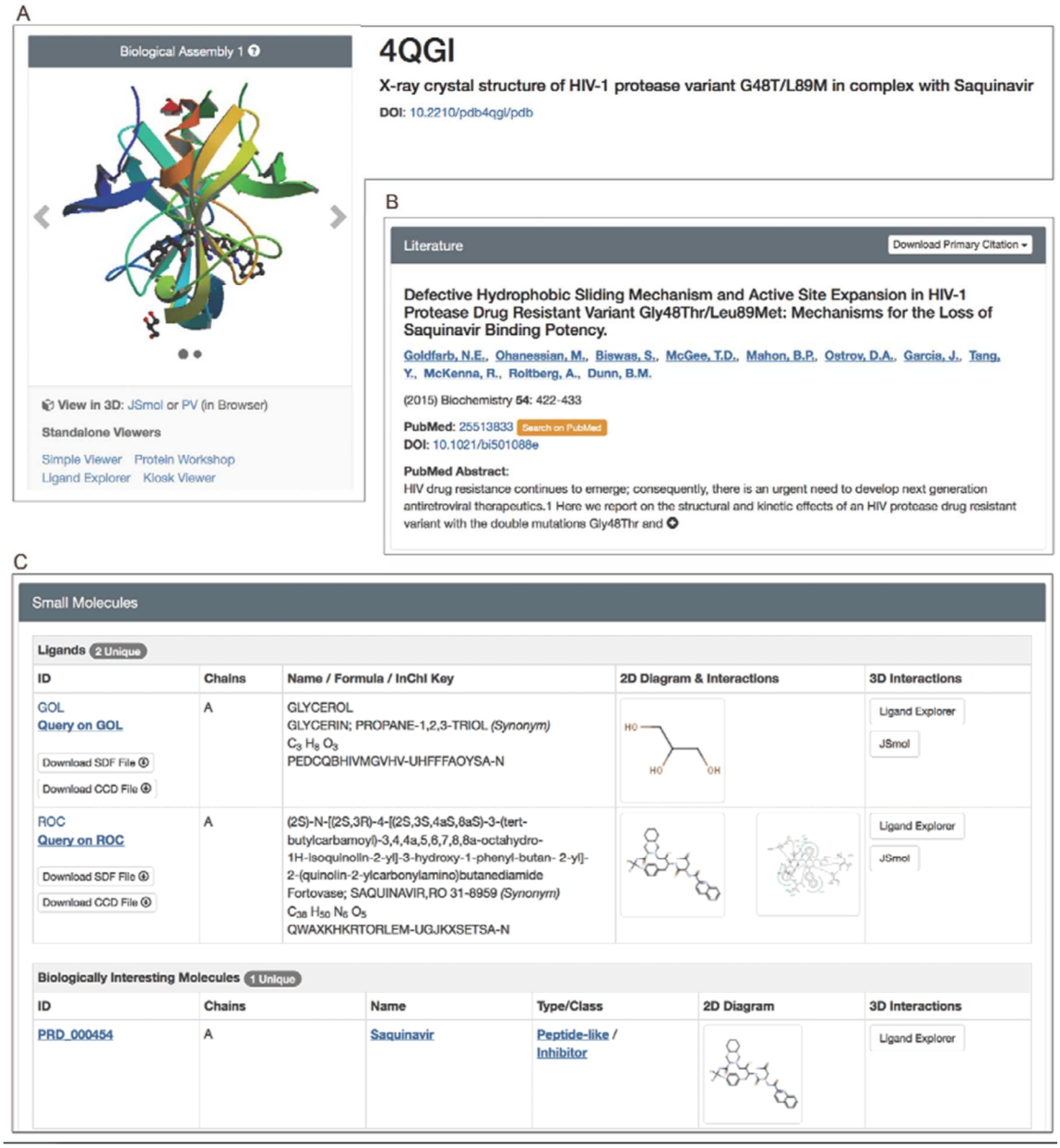

Figure 2. Example Structure Summary page highlights for PDB entry 4QGI, an HIV protease complexed with the drug saquinavir.[29] A) PDB ID, title, 2D image with links to interactive 3D viewers; B) information about the publication describing the entry, with links to PubMed and reference information; C) small molecule information with links to summary information and 3D views. Structure Summary pages also include links to the atomic coordinates, sequence information, experimental data, and validation information.

The RCSB PDB website also supports small molecule searching by ID, name, formula, or chemical drawing. Summary pages are available for each chemical component to provide 2D and 3D visual representations, any subcomponent information, corresponding DrugBank[7d] information, and access to atomic coordinates (Figure 3).

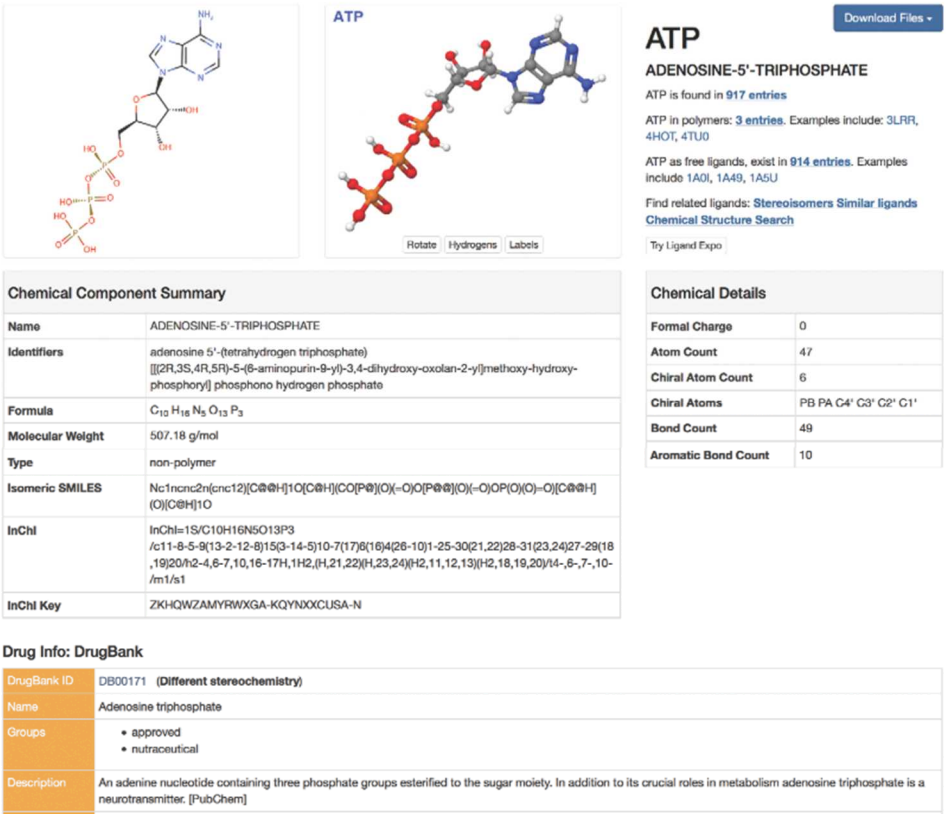

Figure 3. Ligand Summary for ATP. The information shown on this page is built using the corresponding entry for ATP contained in the Chemical Component Dictionary. Blue text links to "query-by-example" searches of the archive, and may be used to find entries that include the ligand. Data highlighted in orange are integrated from external resources such as DrugBank to provide any pharmaceutical context to the structure not included in the deposited PDB entry.[7d] The top ligand image can be downloaded, and the 3D view of the molecule rotated using Jmol.[4]

Additionally, the RCSB PDB *Mobile* app supports access to molecular data on mobile devices (Android and iOS).[30] App users can search the entire PDB database by ID, molecule name, author name and view a summary report for corresponding

structures. The 3D visualization program NDKMol[31] allows app users to interactively view structures and save views as images. The app can be used to search and explore structures during lectures, symposia, and poster sessions.

EDUCATIONAL RESOURCES

A recent survey revealed that the RCSB PDB website is used by a wide range of communities, including educators and students at the high school and university levels. For the educational community, RCSB PDB tools take a subject-based approach, allowing chemistry students to visualize the chemical and structural basis of biological processes, such as *how is oxygen stored and transported to different cells in the human body* or *how do specific drugs act on their target proteins*. Tools are available to find and visualize PDB molecules of interest and explore their interactions. In addition, a number of resources provide non-experts with information and examples of how to interpret the molecule's functions in the context of chemical interactions.

To enable broader access by educators and students, RCSB PDB established an education-focused portal to PDB data. The PDB-101 website[32] hosts regularly published articles, educational materials, and introductions to information specific to PDB data and their representations within the archive.

The *Molecule of the Month* column[33] serves as the foundation of many PDB-101 resources. Since 2000, this feature has highlighted selected biological structures with text, images, and interactive views (Figure 4). The column provides a curated set of example structures from the archive to illustrate key points about a molecule. The collection of Molecule of the Month articles has been organized by biological concept into the *Structural View of Biology* browser to enable top-down searching by functional category. Rather than searching by a particular molecule, users can browse articles about specific topics (such as viruses or the immune system).

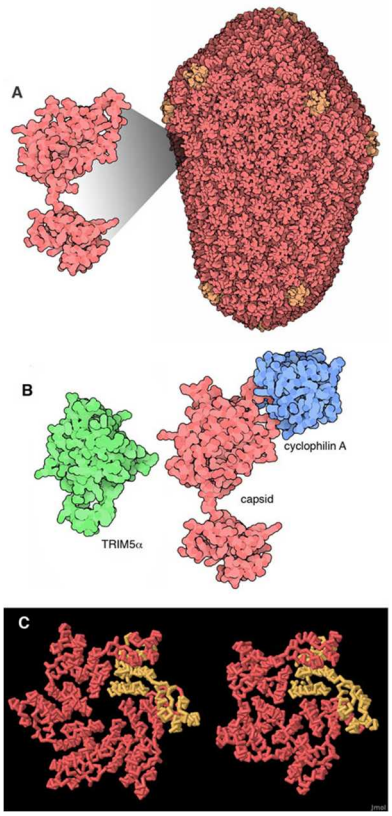

Figure 4. Images from the Molecule of the Month column on HIV Capsid[34] A) The feature begins with a description of the general biology, with an illustration of a single capsid protein (left; PDB entry 1e6j)[35] and the full capsid (right; 3j3q);[36] B) a description of molecules that interact with the HIV capsid follows, showing TRIM5 (2lm3)[37] and cyclophilin (1ak4).[38] C) The final section presents a JSmol[39] view to allow interactive exploration of the HIV capsid hexamer (3mge)[40] and pentamer (3p05).[41]

Beyond *Molecule of the Month*, PDB-101 offers a wide range of educational materials to explore biomolecular structure and function. Videos and animations explore specific topics, from the biology of HIV to protein folding. *Understanding PDB Data* offers a general introduction to structural biology and PDB data files, with topics such as crystallographic resolution and biological assemblies. Hands-on model activities can be used to explore the folding of proteins and nucleic acids.

Resources have been compiled to provide activities, lesson plans, and curricula. For example, the Green Fluorescent Protein (GFP) activity (Figure 5) references the GFP

Molecule of the Month, and uses a paper model to provides a hands-on understanding of the polymer nature of the protein, overall shape and folding of the protein, and the assembly of the GFP chromophore from chemical interactions between three specific amino acids in the core of the protein. An interactive JSmol view demonstrates the chemistry involved in the creation of chromophore.

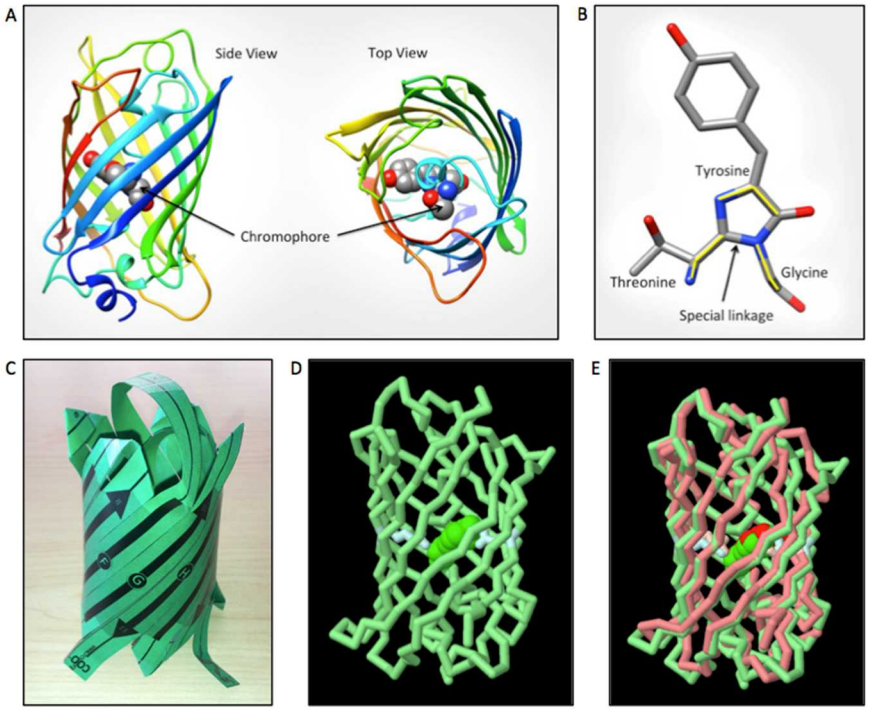

Figure 5. Exploring the Structure of Fluorescent Proteins. Images from the GFP activity at PDB-101. A) ribbon view of GFP with the chromophore highlighted in its core, based on entry 1ema;[42] B) close-up of specific residues chemically linked to form the chromophore; C) a downloadable PDF can be used to create a paper model of GFP; curated JSmol view of GFP highlights D) conserved residues in the protein core that play a role in the chromophore formation and E) the same structure superimposed with a distant relative DsRed (PDB entry 1g7k)[43], showing high structure conservation despite limited sequence similarity.

In December 2014, high school curricula were launched at PDB-101 to combine a variety of PDB-101 and external resources (videos, animations, games, activities and exercises) for comprehensive learning about the biology of HIV/AIDS at introductory

and advanced high school levels. Using these materials, classes studied the HIV lifecycle, interactions with the immune system, and the basis of current infection treatments. Website materials were accessed more than 8000 times during this pilot session. Based on feedback from high school instructors who participated in the pilot, the curricula are being re-organized into individual modules (Biomolecular Structures and Models, Molecular Immunology, Molecular View of HIV/AIDS) that can be implemented in a variety of other lesson plans. Development of similar modules focusing on the structural basis of diabetes is underway.

PDB-101 is also used as a resource in college education. In a 2014 survey of PDB-101 usage, 28% of respondents were undergraduate students, and 33% were graduate students. The most popular area/s of research selected were the life sciences (66%), chemistry (34%), and computational sciences (20%). Examples of how these PDB-related materials have been incorporated at the collegiate level are frequently highlighted in our *Education Corner*, a guest column published in our quarterly newsletter. Examples have included drug discovery projects,[44] interesting molecular visualizations,[45] and cell biology.[46] The usage of PDB data and the utility of accurate molecular visualizations in undergraduate education has been a topic of much study.[47] Other published examples of undergraduate classroom usage include 3D printing and models,[48] molecular modeling,[49] pharmaceutical and medicinal chemistry,[50] and beyond.[51]

CONCLUSION

The RCSB PDB offers free access to a broad range of primary research data and educational materials for all users. Structural entries in the PDB are extensively annotated and validated according to current community standards, providing a rich resource for chemical exploration. Using the RCSB PDB tools and resources, users may

explore detailed information about small and large biomolecules, their chemical interactions, as well as study broader structural and functional concepts in biology.

ACKNOWLEDGEMENTS

RCSB PDB is funded by the NSF (DBI-1338415), NIH, and DOE. The RCSB PDB is a member of the Worldwide PDB along with PDB in Europe, PDB Japan, and BioMagResBank.

AUTHOR INFORMATION

Corresponding Author

*E-mail: Zardecki@rcsb.rutgers.edu

REFERENCES

1. (a) Protein Data Bank, Protein Data Bank. *Nature New Biology* **1971,** *233* (42), 223; (b) Berman, H. M.; Kleywegt, G. J.; Nakamura, H.; Markley, J. L., How community has shaped the Protein Data Bank. *Structure* **2013,** *21* (9), 1485-91.
2. NIH 3D Print Exchange. http://3dprint.nih.gov/ (accessed 6 Nov 2015).
3. Prilusky, J.; Hodis, E.; Canner, D.; Decatur, W. A.; Oberholser, K.; Martz, E.; Berchanski, A.; Harel, M.; Sussman, J. L., Proteopedia: a status report on the collaborative, 3D web-encyclopedia of proteins and other biomolecules. *J Struct Biol* **2011,** *175* (2), 244-52.
4. Jmol: an open-source Java viewer for chemical structures in 3D. http://www.jmol.org/.
5. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E., UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **2004,** *25* (13), 1605-12.
6. DeLano, W. The PyMOL molecular graphics system, http://www.pymol.org. (accessed 6 Nov 2015).
7. (a) Kirchmair, J.; Markt, P.; Distinto, S.; Schuster, D.; Spitzer, G. M.; Liedl, K. R.; Langer, T.; Wolber, G., The Protein Data Bank (PDB), its related services and software tools as key components for in silico guided drug discovery. *J Med Chem* **2008,** *51* (22), 7021-40; (b) Berman, H. M.; Kleywegt, G. J.; Nakamura, H.; Markley, J. L., The Protein Data Bank archive as an open data resource. *J Comput Aided Mol Des* **2014,** *28*, 1009-1014; (c) Fernandez-Suarez, X. M.; Rigden, D. J.; Galperin, M. Y., The 2014 Nucleic Acids Research Database Issue and an updated NAR online Molecular Biology Database Collection. *Nucleic Acids Res* **2014,** *42* (Database issue), D1-6; (d) Law, V.; Knox, C.; Djoumbou, Y.; Jewison, T.; Guo, A. C.; Liu, Y.; Maciejewski, A.; Arndt, D.; Wilson, M.; Neveu, V.; Tang, A.; Gabriel, G.; Ly, C.; Adamjee, S.; Dame, Z. T.; Han, B.; Zhou, Y.; Wishart, D. S., DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Res* **2014,** *42* (Database issue), D1091-7; (e) Wassermann, A. M.; Bajorath, J., BindingDB and ChEMBL: online compound databases for drug discovery. *Expert opinion on drug discovery* **2011,** *6* (7), 683-7; (f) Hu, L.; Benson, M. L.; Smith, R. D.; Lerner, M. G.; Carlson, H. A., Binding MOAD (Mother Of All Databases). *Proteins* **2005,** *60* (3), 333-40; (g) Galperin, M. Y.; Rigden, D. J.; Fernandez-Suarez, X. M., The 2015 Nucleic Acids Research Database Issue and molecular biology database collection. *Nucleic Acids Res* **2015,** *43* (Database issue), D1-5.
8. Berman, H. M.; Henrick, K.; Nakamura, H., Announcing the worldwide Protein Data Bank. *Nat Struct Biol* **2003,** *10* (12), 980.
9. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E., The Protein Data Bank. *Nucleic Acids Res* **2000,** *28* (1), 235-42.

10. Allen, F. H., The Cambridge Structural Database: a quarter of a million crystal structures and rising. *Acta crystallographica. Section B: Structural crystallography and crystal chemistry* **2002,** *58* (Pt 3 Pt 1), 380-8.
11. Lawson, C. L.; Baker, M. L.; Best, C.; Bi, C.; Dougherty, M.; Feng, P.; van Ginkel, G.; Devkota, B.; Lagerstedt, I.; Ludtke, S. J.; Newman, R. H.; Oldfield, T. J.; Rees, I.; Sahni, G.; Sala, R.; Velankar, S.; Warren, J.; Westbrook, J. D.; Henrick, K.; Kleywegt, G. J.; Berman, H. M.; Chiu, W., EMDataBank.org: unified data resource for CryoEM. *Nucleic Acids Research* **2011,** *39* (Database issue), D456-64.
12. Westbrook, J.; Henrick, K.; Ulrich, E. L.; Berman, H. M., 3.6.2 The Protein Data Bank exchange data dictionary. In *International Tables for Crystallography*, Hall, S. R.; McMahon, B., Eds. Springer: Dordrecht, The Netherlands, 2005; Vol. G. Definition and exchange of crystallographic data, pp 195-198.
13. Fitzgerald, P. M. D.; Westbrook, J. D.; Bourne, P. E.; McMahon, B.; Watenpaugh, K. D.; Berman, H. M., 4.5 Macromolecular dictionary (mmCIF). In *International Tables for Crystallography G. Definition and exchange of crystallographic data*, Hall, S. R.; McMahon, B., Eds. Springer: Dordrecht, The Netherlands, 2005; pp 295-443.
14. Westbrook, J. D.; Shao, C.; Feng, Z.; Zhuravleva, M.; Velankar, S.; Young, J., The chemical component dictionary: complete descriptions of constituent molecules in experimentally determined 3D macromolecules in the Protein Data Bank. *Bioinformatics* **2014,** *31* (8), 1274-8.
15. Dutta, S.; Dimitropoulos, D.; Feng, Z.; Persikova, I.; Sen, S.; Shao, C.; Westbrook, J.; Young, J.; Zhuravleva, M. A.; Kleywegt, G. J.; Berman, H. M., Improving the representation of peptide-like inhibitor and antibiotic molecules in the Protein Data Bank. *Biopolymers* **2014,** *101* (6), 659-68.
16. Weininger, D., SMILES 1. Introduction and encoding rules. *J. Chem. Inf. Comput. Sci.* **1988,** *28*, 31-36.
17. © The International Union of Pure and Applied Chemistry *IUPAC International Chemical Identifier (InChI)*, (contact: secretariat@iupac.org): 2005.
18. (a) Read, R. J.; Adams, P. D.; Arendall, W. B., 3rd; Brunger, A. T.; Emsley, P.; Joosten, R. P.; Kleywegt, G. J.; Krissinel, E. B.; Lutteke, T.; Otwinowski, Z.; Perrakis, A.; Richardson, J. S.; Sheffler, W. H.; Smith, J. L.; Tickle, I. J.; Vriend, G.; Zwart, P. H., A new generation of crystallographic validation tools for the protein data bank. *Structure* **2011,** *19* (10), 1395-412; (b) Henderson, R.; Sali, A.; Baker, M. L.; Carragher, B.; Devkota, B.; Downing, K. H.; Egelman, E. H.; Feng, Z.; Frank, J.; Grigorieff, N.; Jiang, W.; Ludtke, S. J.; Medalia, O.; Penczek, P. A.; Rosenthal, P. B.; Rossmann, M. G.; Schmid, M. F.; Schroder, G. F.; Steven, A. C.; Stokes, D. L.; Westbrook, J. D.; Wriggers, W.; Yang, H.; Young, J.; Berman, H. M.; Chiu, W.; Kleywegt, G. J.; Lawson, C. L., Outcome of the first electron microscopy validation task force meeting. *Structure* **2012,** *20* (2), 205-14; (c) Montelione, G. T.; Nilges, M.; Bax, A.; Guntert, P.; Herrmann, T.; Richardson, J. S.; Schwieters, C. D.; Vranken, W. F.; Vuister, G. W.; Wishart, D. S.; Berman, H. M.; Kleywegt, G. J.; Markley, J. L., Recommendations of the wwPDB NMR Validation Task Force. *Structure* **2013,** *21* (9), 1563-70; (d) Trewhella, J.; Hendrickson, W. A.; Kleywegt, G. J.; Sali, A.; Sato, M.; Schwede, T.; Svergun, D. I.; Tainer, J. A.; Westbrook, J.; Berman, H. M., Report of the wwPDB Small-Angle Scattering Task Force: data requirements for biomolecular modeling and the PDB. *Structure* **2013,** *21* (6), 875-81.
19. Gore, S.; Velankar, S.; Kleywegt, G. J., Implementing an X-ray validation pipeline for the Protein Data Bank. *Acta Crystallogr D Biol Crystallogr* **2012,** *68* (Pt 4), 478-83.
20. Kleywegt, G. J.; Bergfors, T.; Senn, H.; Le Motte, P.; Gsell, B.; Shudo, K.; Jones, T. A., Crystal structures of cellular retinoic acid binding proteins I and II in complex with all-trans-retinoic acid and a synthetic retinoid. *Structure* **1994,** *2* (12), 1241-58.
21. Heller, S.; McNaught, A.; Stein, S.; Tchekhovskoi, D.; Pletnev, I., InChI - the worldwide chemical structure identifier standard. *J Cheminform* **2013,** *5* (1), 7.
22. Rose, P. W.; Prlic, A.; Bi, C.; Bluhm, W. F.; Christie, C. H.; Dutta, S.; Green, R. K.; Goodsell, D. S.; Westbrook, J. D.; Woo, J.; Young, J.; Zardecki, C.; Berman, H. M.; Bourne, P. E.; Burley, S. K., The RCSB Protein Data Bank: views of structural biology for basic and applied research and education. *Nucleic Acids Res* **2015,** *43* (Database issue), D345-56.
23. The Gene Ontology Consortium, Gene Ontology: tool for the unification of biology. *Nature Genetics* **2000,** *25*, 25-29.

24. Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB) Enzyme Nomenclature: Recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology on the Nomenclature and Classification of Enzymes by the Reactions they Catalyse. http://www.chem.qmw.ac.uk/iubmb/enzyme (accessed 6 Nov 2015).
25. Moreland, J. L.; Gramada, A.; Buzko, O. V.; Zhang, Q.; Bourne, P. E., The Molecular Biology Toolkit (MBT): a modular platform for developing molecular visualization applications. *BMC bioinformatics* **2005,** *6* (1), 21.
26. Sillitoe, I.; Cuff, A. L.; Dessailly, B. H.; Dawson, N. L.; Furnham, N.; Lee, D.; Lees, J. G.; Lewis, T. E.; Studer, R. A.; Rentzsch, R.; Yeats, C.; Thornton, J. M.; Orengo, C. A., New functional families (FunFams) in CATH to improve the mapping of conserved functional sites to 3D structures. *Nucleic Acids Res* **2013,** *41* (Database issue), D490-8.
27. Fox, N. K.; Brenner, S. E.; Chandonia, J. M., SCOPe: Structural Classification of Proteins--extended, integrating SCOP and ASTRAL data and classification of new structures. *Nucleic Acids Res* **2014,** *42* (Database issue), D304-9.
28. The UniProt Consortium, Activities at the Universal Protein Resource (UniProt). *Nucleic Acids Res* **2014,** *42* (Database issue), D191-8.
29. Goldfarb, N. E.; Ohanessian, M.; Biswas, S.; McGee, T. D., Jr.; Mahon, B. P.; Ostrov, D. A.; Garcia, J.; Tang, Y.; McKenna, R.; Roitberg, A.; Dunn, B. M., Defective hydrophobic sliding mechanism and active site expansion in HIV-1 protease drug resistant variant Gly48Thr/Leu89Met: mechanisms for the loss of saquinavir binding potency. *Biochemistry* **2015,** *54* (2), 422-33.
30. Quinn, G. B.; Bi, C.; Christie, C. H.; Pang, K.; Prlic, A.; Nakane, T.; Zardecki, C.; Voigt, M.; Berman, H. M.; Bourne, P. E.; Rose, P. W., RCSB PDB Mobile: iOS and Android mobile apps to provide data access and visualization to the RCSB Protein Data Bank. *Bioinformatics* **2015,** *31* (1), 126-7.
31. Nakane, T., Molecular Visualization, Online and On-the-Go *RCSB PDB Newsletter* **2013,** *56*, http://bit.ly/1WM3XTv (accessed 6 Nov 2015).
32. RCSB PDB pdb101.rcsb.org. (accessed 6 Nov 2015).
33. Goodsell, D.; Dutta, S.; Zardecki, C.; Voigt, M.; Berman, H.; Burley, S., The RCSB PDB "Molecule of the Month": Inspiring a Molecular View of Biology. *PLoS Biol* **2015,** *13*, e1002140.
34. Goodsell, D. S., HIV Capsid. *RCSB PDB Molecule of the Month* **2013,** *163*, doi: 10.2210/rcsb_pdb/mom_2013_7 (accessed 6 Nov 2015).
35. Monaco-Malbet, S.; Berthet-Colominas, C.; Novelli, A.; Battai, N.; Piga, N.; Cheynet, V.; Mallet, F.; Cusack, S., Mutual conformational adaptations in antigen and antibody upon complex formation between an Fab and HIV-1 capsid protein p24. *Structure* **2000,** *8* (10), 1069-77.
36. Zhao, G.; Perilla, J. R.; Yufenyuy, E. L.; Meng, X.; Chen, B.; Ning, J.; Ahn, J.; Gronenborn, A. M.; Schulten, K.; Aiken, C.; Zhang, P., Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. *Nature* **2013,** *497* (7451), 643-6.
37. Biris, N.; Yang, Y.; Taylor, A. B.; Tomashevski, A.; Guo, M.; Hart, P. J.; Diaz-Griffero, F.; Ivanov, D. N., Structure of the rhesus monkey TRIM5alpha PRYSPRY domain, the HIV capsid recognition module. *Proc Natl Acad Sci U S A* **2012,** *109* (33), 13278-83.
38. Gamble, T. R.; Vajdos, F. F.; Yoo, S.; Worthylake, D. K.; Houseweart, M.; Sundquist, W. I.; Hill, C. P., Crystal structure of human cyclophilin A bound to the amino-terminal domain of HIV-1 capsid. *Cell* **1996,** *87* (7), 1285-94.
39. Hanson, R. M.; Prilusky, J.; Renjian, Z.; Nakane, T.; Sussman, J. L., JSmol and the Next-Generation Web-Based Representation of 3D Molecular Structure as Applied to Proteopedia. *Isr J Chem* **2013,** *53* (3-4), 207-216.
40. Pornillos, O.; Ganser-Pornillos, B. K.; Kelly, B. N.; Hua, Y.; Whitby, F. G.; Stout, C. D.; Sundquist, W. I.; Hill, C. P.; Yeager, M., X-ray structures of the hexameric building block of the HIV capsid. *Cell* **2009,** *137* (7), 1282-92.
41. Pornillos, O.; Ganser-Pornillos, B. K.; Yeager, M., Atomic-level modelling of the HIV capsid. *Nature* **2011,** *469* (7330), 424-7.
42. Ormo, M.; Cubitt, A. B.; Kallio, K.; Gross, L. A.; Tsien, R. Y.; Remington, S. J., Crystal structure of the Aequorea victoria green fluorescent protein. *Science* **1996,** *273* (5280), 1392-5.

43. Yarbrough, D.; Wachter, R. M.; Kallio, K.; Matz, M. V.; Remington, S. J., Refined crystal structure of DsRed, a red fluorescent protein from coral, at 2.0-A resolution. *Proc Natl Acad Sci U S A* **2001,** *98* (2), 462-7.
44. Beckham, J., The Virtual Drug Screening Stream: research and discovery as undergraduates. *RCSB PDB Newsletter* **2014,** *61*, http://bit.ly/R2PPe9 (accessed 6 Nov 2015).
45. (a) Hanson, R., Molecular Visualization for the Masses – PDB Goes to Disneyland. *RCSB PDB Newsletter 63*, http://bit.ly/1odvuwo (accessed 6 Nov 2015); (b) McFadden, P., Undergraduate Course: Protein Portraits. *RCSB PDB Newsletter* **2010,** *46*, http://bit.ly/1JnPRQW (accessed 6 Nov 2015).
46. Kramer, I. M., A Search for Best Methods to Illustrate Complex Information. *RCSB PDB Newsletter* **2010,** *44*, http://bit.ly/1EM18JD (accessed 6 Nov 2015).
47. (a) Bateman Jr., R. C.; Booth, D.; Sirochman, R.; Richardson, J.; Richardson, D. C., Teaching and assessing three-dimensional molecular literacy in undergraduate biochemistry. *J. Chem. Educ.* **2002,** *79*, 551; (b) Dahmani, H. R.; Schneeberger, P.; Kramer, I. M., Analysis of students' aptitude to provide meaning to images that represent cellular components at the molecular level. *CBE Life Sci Educ* **2009,** *8* (3), 226-38; (c) Kimberly J. Linenberger; Holme, T. A., Results of a National Survey of Biochemistry Instructors To Determine the Prevalence and Types of Representations Used during Instruction and Assessment. *J. Chem. Ed.* **2014,** *91*, 800-806.
48. (a) Violante, L. E. H.; Nunez, D. A.; Ryan, S. M.; Grubbs, W. T., 3D printing in the chemistry curriculum: inspiring millennial students to be creative innovators. *ACS Symposium Series* **2014,** *1180*, 125-146; (b) Sergio Rossi, M. B., Davide Brenna, Riccardo Porta, Manuel Orlandi, Three Dimensional (3D) Printing: A Straightforward, User-Friendly Protocol To Convert Virtual Chemical Models to Real-Life Objects. *J. Chem. Ed.* **2015,** *92*, 1398-1401; (c) Chakraborty, P.; Zuckermann, R. N., Coarse-grained, foldable, physical model of the polypeptide chain. *Proc Natl Acad Sci U S A* **2013,** *110* (33), 13368-73.
49. Hayes, J. M., An Integrated Visualization and Basic Molecular Modeling Laboratory for First-Year Undergraduate Medicinal Chemistry. *J Chem Educ* **2014,** *91*, 919-923.
50. (a) C. Stan Tsai, Using Computer Applications and Online Resources To Teach and Learn Pharmaceutical Chemistry. *J. Chem. Ed.* **2007,** *84*, 2019-2023; (b) Ricardo P. Rodrigues, S. F. A., Susimaire P. Mantoani, Vera L. Eifler-Lima, Vinicius B. Silva, Daniel F. Kawano, Using Free Computational Resources To Illustrate the Drug Design Process in an Undergraduate Medicinal Chemistry Course. *J. Chem. Ed.* **2015,** *92*, 827-835; (c) Ivone Carvalho; Áurea D. L. Borges; Bernardes, L. S. C., Medicinal Chemistry and Molecular Modeling: An Integration To Teach Drug Structure–Activity Relationship and the Molecular Basis of Drug Action. *J. Chem. Ed.* **2005,** *82*, 588-596.
51. (a) Maikel Wijtmans, L. v. R., Jacqueline E. van Muijlwijk-Koezen, Activating Students' Interest and Participation in Lectures and Practical Courses Using Their Electronic Devices. *J. Chem. Ed.* **2014,** *91*, 1830-1837; (b) Johnson, R. J., Teaching foundational topics and scientific skills in biochemistry within the conceptual framework of HIV protease. *Biochemistry and molecular biology education : a bimonthly publication of the International Union of Biochemistry and Molecular Biology* **2014,** *42* (4), 299-304.