The RCSB PDB “Molecule of the Month”: Inspiring a Molecular View of Biology

David S. Goodsell1,2*, Shuchismita Dutta1,3, Christine Zardecki1,3, Maria Voigt1,3, Helen M. Berman1,3, Stephen K. Burley1,3,4,5

1 RCSB Protein Data Bank, 2 Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, California, United States of America, 3 Center for Integrative Proteomics Research and Department of Chemistry and Chemical Biology, Rutgers, The State University of New Jersey, New Brunswick, New Jersey, United States of America, 4 BioMaPS Institute for Quantitative Biology and Cancer Institute of New Jersey, Rutgers, The State University of New Jersey, New Brunswick, New Jersey, United States of America, 5 San Diego Supercomputer Center and Skaggs School of Pharmaceutical Sciences, University of California, San Diego, San Diego, California, United States of America

Abstract

The Research Collaboratory for Structural Bioinformatics (RCSB) Molecule of the Month series provides a curated introduction to the 3-D biomolecular structures available in the Protein Data Bank archive and the tools that are available at the RCSB website for accessing and exploring them. A variety of educational materials, such as articles, videos, posters, hands-on activities, lesson plans, and curricula, build on this series for use in a variety of educational settings as a general introduction to key topics, such as enzyme action, protein synthesis, and viruses. The series and associated educational materials are freely available at www.rcsb.org.

Introduction

Three-dimensional (3-D) structures of biological macromolecules are important because “function follows form” in biology. Knowledge of the shape, structure, and interactions of biological macromolecules (proteins, nucleic acids, and carbohydrates) helps us understand and “see” biology at the molecular level in atomic detail [1]. Changes in shape, chemical properties, and assemblages of biomolecules are critical to understanding disease processes and in discovering and developing targeted treatments for diseases. Since 2000, a monthly feature at the Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB) [2], called “Molecule of the Month” (MotM), has highlighted structures and functions of selected biological macromolecules and helped explain their importance in biology and medicine. Today, more than 180 of these monthly features provide users with a curated sampling of the Protein Data Bank (PDB).

Interactive tutorials describing biomolecular subjects were pioneered by Eric Martz [3], initially using Chime and Protein Explorer and more recently using Jmol/JSmol. Concurrent with development of the Molecule of the Month at the RCSB PDB, several projects have developed...
related resources that allow users to view and explore PDB structures, including the InterPro "Protein Focus" (formerly known as "Protein of the Month") and "Quips" [4], both of which come from the European Bioinformatics Institute (EMBL-EBI), and Proteopedia [5] from the Weizmann Institute, which uses a crowdsourcing approach based on Wikipedia.

Established in 1971 as the first open access digital resource in biology, the PDB [6] provides free, global access to more than 100,000 experimentally determined, 3-D, atomic structures of biological macromolecules. These data are widely used by structural biologists and other scientists working in basic and applied sciences. The same data are also made available to students, educators, and general audiences to promote understanding and discovery of new science. The PDB archive is maintained by the worldwide PDB organization (wwPDB), which today encompasses regional data centers in the United States (RCSB PDB; http://www.rcsb.org) at Rutgers University and University of California, San Diego; the Biological Magnetic Resonance Data Bank at University of Wisconsin, Madison; and centers in Europe (Protein Data Bank in Europe, PDBe; http://pdbe.org) and Japan (Protein Data Bank in Japan, PDBj; http://pdbj.org) [6]. The wwPDB data centers collaboratively manage data deposition and validation. Thereafter, each data center provides online access to contents of the PDB archive. The primary data distributed by all centers is the same, but each center offers different views, tools, and resources for accessing and using the information. The RCSB PDB [7] offers both expert (http://www.rcsb.org) and introductory/educational portals (http://www.rcsb.org/pdb-101) to the PDB archive [8].

The PDB-101 educational portal allows non-expert users to explore biology through a structural lens and learn how to use and understand PDB data. This resource presents a wide range of educational materials, such as articles, videos, posters, hands-on activities, lesson plans, and curricula (Box 1). Since 2000, the MotM has profiled molecular structures, together with RCSB PDB tools to explore them. Herein, we describe the conceptual design of these features, their use in research and education, and their ability to bring a structural view of biology and medicine to a general audience.

Box 1. PDB-101 at a Glance

PDB-101 (http://www.rcsb.org/pdb-101) provides an introduction to key structural biology topics and access to relevant biomolecular structures. Monthly MotM features serve as a bridge between traditional teaching materials and the primary scientific data, which is freely available without restrictions on usage from the PDB archive.

Key components

- Molecule of the Month—Short, illustrated features on selected molecules from the archive provide a sampling of the PDB holdings.
- Structural View of Biology Browser—MotM features are organized by theme for exploration by all users.
- Educational Resources—Various materials, including posters, articles, hands-on activities, images, and animations, are available for educational use.
- Video Challenge—Instructions, rules, resources, and information related to a competition for high school students are available here.
- Curricula—Curricula have been developed through collaborations and participation of scientists, curriculum design experts, educators, clinicians, and local teachers.
Box 2. Learning about Biology from MotM Features

Topics presented in MotM features can be organized by user educational background.

Elementary and Middle School

- Biological Information: The primary feature on DNA highlights the double helical structure of B-DNA and the nature of Watson-Crick base pairing. Other features explain how DNA is read, duplicated, regulated, and repaired.
- Viruses: Many features on well-known viruses, such as Ebola virus, HIV, influenza, and rhinovirus help to show the science behind the news.

High School (General Biology)

- Mechanism of Protein Synthesis: A feature on ribosomes explains the multi-step process of protein synthesis. Other features describe supporting players, such as the transfer RNAs and elongation factors.
- Molecular Recognition: An introductory feature on hemoglobin presents an example of molecular recognition, highlighting protein motions that regulate oxygen binding in the lung and release in the blood stream.

High School (AP Biology), Undergraduate, and Beyond

- Enzyme Action: Features describe various enzymes commonly used in research, including carbonic anhydrase and catalase, revealing how structure underpins enzymatic activity.
- Molecular Basis of Disease: Many features describe molecules involved in cancer, explaining the connection between human diseases and gene mutations affecting protein structure and function.
- Biotechnology and Nanotechnology: Several features highlight molecular structures designed by scientists, such as artificial DNA lattices and protein cages, revealing how understanding of molecular structure and function can be put to practical use.

Design of the Molecule of the Month

MotM features introduce 3-D structures of proteins, nucleic acids, and carbohydrates in the context of basic and applied biological and medical research. Themes for the features are selected from topics of general interest (Box 2, Fig 1), including molecules related to human health and disease (for example, hemoglobin and collagen), biological machines (ribosomes), and current events (Ebola virus, ricin, anthrax, and HIV/AIDS). Non-technical text explains the overall shapes of relevant biomolecules and highlights structural features important for their functions while illustrations highlight molecular structural features described in the text (Fig 2). Accompanying interactive JSmol views present curated explorations of one or more relevant
As of the end of 2014, 180 different molecules have been presented in Molecule of the Month features. Topics are selected to promote understanding of biomolecular structure and function and connections to human health and disease as well as biotechnology. Some of the more popular subjects are depicted herein. From top to bottom: DNA polymerase [16], ribosome [17,18], hemoglobin [19], ATP synthase [20,21], a G-protein bound to a G-protein-coupled receptor (GPCR) [22], and the collagen triple helix [23].

Fig 1. Molecules highlighted in the MotM features. As of the end of 2014, 180 different molecules have been presented in Molecule of the Month features. Topics are selected to promote understanding of biomolecular structure and function and connections to human health and disease as well as biotechnology. Some of the more popular subjects are depicted herein. From top to bottom: DNA polymerase [16], ribosome [17,18], hemoglobin [19], ATP synthase [20,21], a G-protein bound to a G-protein-coupled receptor (GPCR) [22], and the collagen triple helix [23].

doi:10.1371/journal.pbio.1002140.g001
structures allowing users to view them in 3-D. (JSmol is an open-source HTML5 viewer for chemical structures, available at http://wiki.jmol.org.)

Links to featured PDB structures allow users to access specific RCSB PDB pages containing detailed information about the structure. In addition, links to abridged versions of the structure pages provide access to short descriptions of its biological function, an interactive JSmol 3-D interactive viewer, and sequences of protein and nucleic acid polymers of the molecules. Related PDB structures are also provided as starting points for further explorations, giving non-experts access to a curated subset of PDB structures to increase familiarity with visualization and help them take advantage of more complex resources, including more powerful visualization tools, available from the RCSB PDB. Users can access the features by topic (such as "Health and Disease" or "Biotechnology and Nanotechnology") or an alphabetical index of categories.

MotM features are widely used by diverse audiences (Box 3). Examples of educational programs, projects, and resources developed based on or impacted by our efforts are described in the following sections.

Box 3. Usage of the Molecule of the Month

Access and download statistics of individual monthly features over a period of time (e.g., approximately 18,000 unique visits for the feature on hemoglobin during the first nine months of 2014) document their popularity. Interestingly, MotM features are not only
accessed immediately upon publication, but continue to be used many years later. An online survey of PDB-101 users, conducted in 2014, revealed that 74% of the approximately 700 respondents used one or more MotM features, with 49% downloading images for their own use, 46% using the interactive JSmol viewer, 40% viewing related structures, and 33% using our suggestions for exploration of related topics, provided at the end of each feature. While this survey is not exhaustive, it provides self-reported measures of which feature components are being used. Complementary metrics available from Google Analytics (http://www.google.com/analytics) documented over 350,000 sessions that accessed MotM pages between August 1, 2014, and January 18, 2015. These sessions originated from various search engines, links from scientific databases worldwide, and resources such as Foldit (http://fold.it/portal). MotM content is also reproduced with formal written permission by textbook authors, editors, and journalists (25 requests were processed in 2014).

Molecule of the Month In Student Research

Two projects initiated at the Center for BioMolecular Modeling at the Milwaukee School of Engineering (CBM MSOE) illustrate how student research projects incorporate the MotM. The Connecting Researchers, Educators and Students (CREST) program (http://cbm.msoe.edu/stupro/crest) pairs local university researchers and educators with undergraduate students to work together on a research topic. MotM features have been assigned as foundational reading material for several of the CREST projects and have been incorporated into their posters, videos, tutorials, and scientific articles for use in undergraduate education.

In another CBM program, teachers from high schools nationwide learn about biological molecules by visualizing protein structures using MotM features. The teachers then lead a team of students to work with a local scientist on a research topic in a program called Students Modeling a Research Topic (SMART) teams. The teams model biomolecular structures based on available literature under the scientist’s guidance, often using MotM features for background research. Physical models created by these SMART teams provide useful “tools” for discussing the topic and designing new experiments [9,10]. The teams often present their work and models at local, regional, or national meetings organized by professional societies (e.g., ASBMB), giving students in these teams a taste of research experience.

Informing and Challenging the Next Generation of Citizens and Scientists

The high school student protein modeling event in the Science Olympiad (http://soinc.org) [11], a large, nationwide scientific challenge, was originally developed by CBM MSOE and is now being operated jointly by RCSB PDB and CBM MSOE (Fig 3). To participate, students are required to read about a specific theme, build 3-D molecular models of proteins, and answer exam questions about the biology of the modeled molecular structures [12]. Links to specially prepared or extant MotM features provide the necessary background information for participants and serve as the source for many of the exam questions. Annually, 5,000–7,000 teams (with two to three students per team) participate throughout the US. Competitions are conducted at regional and state levels in late fall through early spring, leading up to the national-level competitions in May. Preparation for the competitions motivates students, teachers, coaches, and judges to read MotM features relevant to the competition.
Other competitions and extracurricular scientific events use MotM features as foundational materials, as well. For example, the RCSB PDB video challenge for high school students, started in 2014, challenged students to create two-minute videos highlighting aspects of the structural biology of HIV [13]. Challenge contributors with diverse interests cited the MotM features as resources used for creating their videos. Another example is noted in RCSB PDB’s "Education Corner" newsletter: the Protein Modeling Challenge at Stony Brook University, which pointed students to MotM features for background research on Ras proteins as molecular switches [14].

**Molecule of the Month In the Classroom**

We have received many anecdotal reports from educators using MotM features in manifold and creative ways. Some have assigned different topics to each student for presentation in their classrooms, using relevant MotM features as starting points for further study. One educator combined art with science, creating original paintings of molecules using proteins highlighted in MotM features as sources of inspiration (Fig 4). This project is particularly exciting, as the students are drawing directly from primary scientific data. Professional curriculum developers include links to MotM features (see, for instance, [15] and http://www.hhmi.org/biointeractive/how-do-fibers-form). As part of its outreach and educational mission, the RCSB PDB has developed short lesson plans on topics, such as double helical DNA, green fluorescent protein, and the HIV capsid, in which relevant MotM features are combined with hands-on activities (Fig 5). In addition to engaging students, these lesson plans are used to train teachers at local and national workshops. Scientists at all levels (undergraduate, graduate, postdoctoral, and faculty) also use these lesson plans and materials in specialist courses, such as in the Applied Bioinformatics Course at Peking University (since 2000) and the annual Interdisciplinary Quantitative Biology Boot Camp at Rutgers University (since 2014).

Our recent survey of PDB-101 users documented that, today, a majority of our users come from colleges, universities, and research institutes. Only 8% of respondents identified
themselves as being from K–12 institutions. We have, therefore, taken concerted steps to strengthen our outreach, with particular focus on high school science educators and students. We plan to augment K–12 usage by introducing focused curricula, teacher training, and competitions such as the video challenge. In 2014, RCSB PDB coordinated development of HIV/AIDS curricula for general biology and AP biology students, in collaboration with scientists,

Fig 4. Molecular machines at High Tech High. Parag Chowdhury and his students at High Tech High in North County San Diego, California, combined art and science to explore molecular machines. Using the RCSB PDB MotM features as starting points, they created original paintings of biomolecules and wrote one-page descriptions of the role played by each biomolecule in the cell. They then combined all of this creative work into a book, Molecular Machines: How Are We Assembled?, currently available through the on-demand publisher Blurb. The painting of a bacterial virus or bacteriophage [28], shown above, was used on the cover of the book, and was kindly provided by Parag Chowdhury with permission.

doi:10.1371/journal.pbio.1002140.g004
curriculum design experts, educators, clinicians, and local high school teachers. These curricula use authentic (primary scientific) data, conform to current science teaching standards, and include hands-on teaching materials, individual and group activities, and suggestions for student assessment and curriculum evaluation. Initial activities in the curricula explain the structure and function of biological molecules playing key roles in human immunity. A separate module focuses on the targeted nature of HIV infection affecting helper T-cells and both current and emerging strategies for treating HIV infection (Fig 6). Version 1.0 of the general biology and AP biology curricula are available from PDB-101 at http://education.rcsb.org/curriculum.

Future Plans

Since 2013, RCSB PDB educational and outreach efforts have been integrating basic science and medicine by focusing on an important public health-related topic. Our current focus is on HIV/AIDS. We use customized educational materials that showcase various HIV proteins from the PDB archive and explain at the molecular level in atomic detail how the US Food and Drug Administration (FDA)-approved combination anti-retroviral therapy (cART) regimens target selected HIV proteins. Relevant MotM features are presented using multiple modalities, including a video challenge and newly developed hands-on educational materials. Beginning in 2016, we will focus on type 2 diabetes mellitus. Going beyond the MotM feature on insulin, new educational material will be developed on the insulin receptor, insulin signaling, glucagon and its receptor, glycated hemoglobin A1c, and the mechanisms of action of glucose-lowering drugs approved by the US FDA.

Historically, some of the topics discussed in the MotM features have been selected in response to community requests. Features on catalase and carbonic anhydrase, commonly used in high school and college laboratory exercises, were both developed following educator suggestions, and are currently the third and fourth most frequently used features. Most recently,
MotM features have been inspired by current events, such as the Ebola crisis in west Africa, which inspired us to produce a video on the molecular anatomy of the virus (http://www.youtube.com/watch?v=9SptF9bCyd0). The RCSB PDB welcomes requests and suggestions for new MotM features. Our overall goal in all these efforts is to bring a molecular structural view of biology and medicine to a large and diverse community.
Acknowledgments

We thank our users and our current and past website developers, particularly Boki Beran, Chunxiao Bi, Wolfgang Bluhm, Cole Christie, Andreas Prlić, Greg Quinn, Alex Ramos, Peter Rose, and Jesse Woo, for their support in organizing and presenting this project.

References

1. Barber NC, Stark LA. Engaging with molecular form to understand function. CBE Life Sci Educ 2014; 13: 21–24. doi: 10.1187/cbe.13-12-0247 PMID: 24591499
2. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. Nucleic Acids Res 2000; 28: 235–242. PMID: 10592235
3. Martz E. Protein Explorer: easy yet powerful macromolecular visualization. Trends Biochem Sci 2002; 27: 107–109. PMID: 11852249
4. Velankar S, Alhroub Y, Best C, Caboche S, Conroy MJ, Dana JM, et al. PDBe: Protein Data Bank in Europe. Nucleic Acids Res 2012; 40: D445–452. doi: 10.1093/nar/gkr998 PMID: 22110033
5. Hodis E, Prilusky J, Martz E, Silman I, Moult J, Sussman JL. Proteopedia—a scientific ‘wiki’ bridging the rift between three-dimensional structure and function of biomacromolecules. Genome Biol 2008; 9: R121. doi: 10.1186/gb-2008-9-8-r121 PMID: 18673581
6. Berman H, Henrick K, Nakamura H. Announcing the worldwide Protein Data Bank. Nature structural biology 2003; 10: 980. PMID: 14634627
7. Rose PW, Prlic A, Bi C, Bluhm WF, Christie CH, Dutta S, et al. The RCSB Protein Data Bank: views of structural biology for basic and applied research and education. Nucleic Acids Res 2015; 43: D345–356. doi: 10.1093/nar/gku1214 PMID: 25428375
8. Rose PW, Bi C, Bluhm WF, Christie CH, Dimitropoulos D, Dutta S, et al. The RCSB Protein Data Bank: new resources for research and education. Nucleic Acids Res 2013; 41: D475–482. doi: 10.1093/nar/gks1200 PMID: 23193259
9. Herman T. Teaching Science Through Protein Modeling. ASBMB Today, September 2010. http://www.asmbb.org/asmbbtoday/asmbbtoday_article.aspx?id=9158
10. Hata TS. PDB Education Corner. RCSB PDB Newsletter 2004; 21. http://www.pdb.org/pdb/static.do?p=general_information/news_publications/newsletters/2004q1/hata.html
11. Putz G. J. K. Science Olympiad and the ASBMB: Inspiring the next generation of scientists. ASBMB Today, February 2013. http://www.asmbb.org/asmbbtoday/asmbbtoday_article.aspx?id=32460
12. Dutta S, Zardecki C, Goodsell DS, Berman HM. Promoting a structural view of biology for varied audiences: an overview of RCSB PDB resources and experiences. J Appl Crystallogr 2010; 43: 1224–1229. PMID: 20877496
13. Bhattacharya M, Jaworsky C, Lucci K. Education Corner: Mentoring High School Students for the RCSB PDB HIV Video Challenge. RCSB PDB Newsletter 2015; 64. http://www.pdb.org/pdb/general_information/news_publications/newsletters/2015q1/education_corner.html
14. Kiely J. Education Corner: Protein Challenge. RCSB PDB Newsletter 2010; 47. http://www.pdb.org/pdb/general_information/news_publications/newsletters/2010q3/education_corner.html
15. Chowning JT, Kovarik D., Griswold J. Modeling Protein Structure & Function: Pencil Transferase. The American Biology Teacher 2012; 74: 581–582.
16. Beese LS, Derbyshire V, Stelitz TA. Structure of DNA polymerase I klenow fragment bound to duplex DNA. Science 1993; 260: 352–355. PMID: 8469987
17. Klein DJ, Schmeing TM, Moore PB, Stelitz TA. The kink-turn: a new RNA secondary structure motif. The EMBO journal 2001; 20: 4214–4221. PMID: 11483524
18. Wimberly BT, Brodersen DE, Clemons WM Jr., Morgan-Warren R, Carter AP, Vonrhein C, et al. Structure of the 30S ribosomal subunit. Nature 2000; 407: 327–339. PMID: 11014182
19. Fermi G, Perutz MF, Shaanan B, Fourme R. The crystal structure of human deoxyhaemoglobin at 1.74 Å resolution. J Mol Biol 1984; 175: 159–174. PMID: 6726807
20. Rastogi VK, Girvin ME. Structural changes linked to proton translocation by subunit c of the ATP synthase. Nature 1999; 402: 263–268. PMID: 10580496
21. Gibbons C, Montgomery MG, Leslie AG, Walker JE. The structure of the central stalk in bovine F(1)-ATPase at 2.4 Å resolution. Nat Struct Biol 2000; 7: 1055–1061. PMID: 11062563
22. Rasmussen SG, DeVree BT, Zhou Y, Kruse AC, Chung KY, Kobila TS, et al. Crystal structure of the beta2 adrenergic receptor-Gs protein complex. Nature 2011; 477: 549–555. doi: 10.1038/nature10361 PMID: 21772288
23. Kramer R, Bella J, Mayville P, Brodsky B, Berman HM. Sequence dependent conformational variations of collagen triple-helical structure. Nat Struct Biol 1999; 6: 454–457. PMID: 10331873

24. Schmeing TM, Voorhees RM, Kelley AC, Gao YG, Murphy FVT, Weir JR, et al. The crystal structure of the ribosome bound to EF-Tu and aminoacyl-tRNA. Science 2009; 326: 688–694. doi: 10.1126/science.1179700 PMID: 19833920

25. Voorhees RM, Weixlbaumer A, Loakes D, Kelley AC, Ramakrishnan V. Insights into substrate stabilization from snapshots of the peptidyl transferase center of the intact 70S ribosome. Nat Struct Mol Biol 2009; 16: 528–533. doi: 10.1038/nsmb.1577 PMID: 19363482

26. Gao YG, Selmer M, Dunham CM, Weixlbaumer A, Kelley AC, Ramakrishnan V. The structure of the ribosome with elongation factor G trapped in the posttranslational state. Science 2009; 326: 694–699. doi: 10.1126/science.1179709 PMID: 19833919

27. Fremont DH, Matsumura M, Stura EA, Peterson PA, Wilson IA. Crystal structures of two viral peptides in complex with murine MHC class I H-2Kb. Science 1992; 257: 919–927. PMID: 1323877

28. Dokland T, Bernal RA, Burch A, Pletnev S, Fane BA, Rossmann MG. The role of scaffolding proteins in the assembly of the small, single-stranded DNA virus phiX174. J Mol Biol 1999; 288: 595–608. PMID: 10329166

29. Yang F, Moss LG, Phillips GN Jr. The molecular structure of green fluorescent protein. Nat Biotechnol 1996; 14: 1246–1251. PMID: 9631087

30. Zhao G, Perilla JR, Yufenyuy EL, Meng X, Chen B, Ning J, et al. Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. Nature 2013; 497: 643–646. doi: 10.1038/nature12162 PMID: 23719463