

RCSB Protein Data Bank Advisory Committee

Report of April 26, 2021 Annual Meeting

Teleconference

Chair: Paul Adams

Membership:

Present: Peter Andolfatto, Judith Blake, Andy Byrd, Wah Chiu, Kirk Clark, Paul Craig, Robert B. Darnell, Roland Dunbrack, Paul Falkowski, Thomas Ferrin, Mandë Holford, Cathy Peishoff, Sue Rhee, Torsten Schwede, Jill Trehwella

Absent: Bridget Carragher

RCSB PDB AC E-mail Addresses:

PDAdams@LBL.gov, pa2543@columbia.edu, judith.blake@jax.org, bcarr@nysbc.org, byrdra@mail.nih.gov, wahc@stanford.edu, kirk.clark@novartis.com, paul.craig@rit.edu, Robert.Darnell@Rockefeller.edu, roland.dunbrack@fcc.edu, falko@marine.rutgers.edu, tef@cgl.ucsf.edu, mholford@hunter.cuny.edu, peishoffc@gmail.com, srhee@carnegiescience.edu, torsten.schwede@unibas.ch, jill.trehwella@sydney.edu.au

RCSB PDB Leadership: Stephen Burley (Director), Helen Berman (Director Emerita), Andrej Sali (UCSF Site Head)

RCSB PDB Leadership E-mail Addresses:

sburley@proteomics.rutgers.edu, berman@rcsb.rutgers.edu, sali@salilab.org

Executive Summary

The Advisory Committee (AC) to the Research Collaboratory for Structural Bioinformatics (RCSB) held a virtual meeting on April 26th, 2021 to review recent progress and provide feedback on specific questions.

Agenda items included

- Welcome and Introductions
- Overview and Response to COVID-19
- Deposition/Biocuration
- PDB Archive Status
- New and Improved RCSB.org
- PDB Data Delivery Status
- Outreach/Education and PDB50
- Operations and Funding
- Discussion

The meeting was opened by Dr. Stephen Burley. Other RCSB PDB participants were Helen M. Berman, Robert Lowe, John Westbrook, Jasmine Young, Christine Zardecki (Rutgers); Andrej Sali (UCSF); and Jose Duarte (UCSD), Yana Rose (UCSD). Also in attendance for parts of the meeting were representatives of funding agencies: Steven Ellis (NSF), Ramana Madupu (DOE), and Amy Swain (DOE).

Appendix 1 provides a summary of the RCSB responses to the 2020 Advisory Panel meeting recommendations. Appendix 2 provides a summary of global PDB deposition and data access statistics in 2020.

Overall Comments from the Advisory Panel

The team is congratulated on their outstanding work over the last 12 months under challenging circumstances. Structural biology has played a critical role in fighting the pandemic, and the RCSB has stepped up in educating the community and curating the deposition of many SARS-CoV-2 related structures, including bulk depositions. The team gave an excellent set of focused presentations, and their collective leadership of the RCSB and their individual projects over the last year is applauded. The committee recognizes the continued efforts of the RCSB Director, Stephen Burley, and his colleagues to seek additional funding for the RCSB activities, which have paid off over the last year.

Recommendations for future meetings

- Keep up the great work!

Detailed Advisory Panel Comments and Feedback

Response to COVID-19

The COVID-19 pandemic has been challenging for everyone in the world but has also presented some opportunities for the RCSB. They prioritized the biocuration of SARS-CoV-2 structures (more than 1000 by March 2021). They also seized the opportunity to provide educational material about the virus to the community, with information about the virus, viral proteins, and therapeutics. There was a noticeable increase in traffic to PDB-101 as a result. PDB data also formed an important part of several other COVID-19 related databases created in the last year. RCSB staff also participated in education activities such as a summer bootcamp on COVID-19. The committee also heard how working remotely was embraced by the RCSB team, and that hybrid work was a likely model for the future. The committee recognizes that there may be further opportunities to combine education and outreach activities around COVID-19 with fund raising activities, especially with the PDB50 celebrations this year.

Recommendations

- Beyond the recent publication in PROTEINS about the structural biology of the SARS-CoV-2 virus, the RCSB should consider whether there are other venues to highlight the role of structure, and the PDB, in the pandemic response. Longer term there may be opportunities to study the long-term impact that structural biology and the availability of structures from the PDB has had on the fight against the virus.

Deposition/Biocuration

The committee heard from Jasmine Young about deposition and biocuration activities. We continue to be very impressed with Jasmine's leadership in this area, and the report on last year's activities was very positive. The load balancing across the wwPDB locations appears to be working well, and new features such as GroupDep are making it easier for some groups to make their depositions, especially for the SARS-CoV-2 ligand screening studies. There were a record number of depositions and biocurations by the RCSB team. The continued growth in the number of cryo-EM structure depositions was also noted. The committee was very pleased to hear that the efficiency of biocuration has continued to improve, with the median time for large structures decreasing from 3-4 days to less than 2 days. Finally, the committee was happy to hear that the new charter, which formally incorporates the EMDB, has been signed. We were asked if we had any concerns about the Deposition/Biocuration

work underway with the wwPDB partners. The committee again congratulates Jasmine for her leadership of OneDep and creating a strong collaboration across the wwPDB. However, there were some concerns from the committee about the level of effort available from partners during the pandemic.

Recommendations

- The improvements in deposition efficiency appear to have come in part from the flexibility afforded staff with remote work. The RCSB is encouraged to think about how to continue this model in the future.
- The committee welcomes the new validation features for cryo-EM depositions. However, the RCSB (in collaboration with the broader wwPDB community) is encouraged to implement new accepted metrics in a timely manner as they become available.
- Continue to monitor the growth of cryo-EM depositions and be prepared to prioritize the implementation of deposition standards and tools to help respond to the increased load.

PDB Archive Status

The committee heard from John Westbrook on the status of the archive, and its continued growth. Of note has been the deployment of new storage for the core archive, which will accommodate projected growth through 2022. The committee was also very pleased to hear that the carbohydrate remediation project has been completed, and that the chemical reference data has been extended to include additional definitions. Finally, the securing of several new funding sources will clearly help extend the reach of PDB data into new research areas.

Recommendations

- There are several other opportunities to improve the archive through remediation campaigns. The committee encourages the RCSB to prioritize targets such as lipids/detergents, metalloproteins, and assemblies.
- The remediation of antibody structures is an outstanding problem.
- With the recent technical developments in the field of cryo-electron tomography it seems likely that there will be increased demand for archiving tomography data as part of depositions. The committee wonders whether the current EMD/EMPIAR system will be sustainable and suggests the RCSB look into how complementary services might be established in the US.

New and Improved RCSB.org

The committee heard from John Westbrook on the ongoing efforts to improve the infrastructure for the rcsb.org website and associated backends. The new search functionality and structure visualization tools were also presented. The committee was impressed by all these developments. However, there were some concerns about the complexity of the search system for many users. The committee feels that the option of a simplified search and visualization interfaces would benefit many of the RCSB users. The committee thought the development of tools for visualization of protein features in 3D was an important step, and likely to be of interest to many users. The committee recognizes that it may be challenging for the team to adopt the mindset of the non-structural biology users of the site when thinking about improving interfaces (many of the committee members face the same problem). However, we think that engagement locally with the Rutgers user experience (UX) program and with non-structural communities could provide resources to help.

Recommendations

- Reach out to the Rutgers UX program to leverage their expertise, and if possible, raise awareness of UX thinking across the team.
- Create a 1 year plan for engagement with non-structural biology user communities and updates of user interfaces as a result, including identifying the top 3 communities to target and build relationships with.
- Consider the development of a simplified search and visualization interface for the more novice or casual user.
- Consider seeking input from the community as a potentially productive route for thoughts about how to improve the site. Committee members suggested looking at BioMolViz.org and engaging with undergraduate life science education researchers.

PDB Data Delivery Status

The committee heard about the tracking of RCSB data access. It is a testament to the stability of the system the RCSB has implemented that it can seamlessly support so many users, sessions, and downloads. The example of an errant script producing a large data spike prompted discussion about the potential need for data throttling. The consensus was this shouldn't be implemented without community input and analysis of historical data. The committee also felt that the NCBI model for throttling was one that could be adopted if necessary, and that tokens or similar impediments to site access should be avoided.

Outreach/Education and PDB50

Christine Zardecki presented many of the great efforts in outreach and education over the last year. Unsurprisingly, much of the outreach and education was focused on COVID-19. It was very encouraging to see that this led to PDB-101 supporting substantially more users in 2020 (30% over 2019). The committee also heard about several non-COVID-19 activities, including some interesting health-related videos. The 2022-23 focus on cancer will likely be very popular. The community outreach by the team is excellent and clearly very important for educating the research community about the RCSB and structural biology. The committee members expressed their support of, and interest in participating in, the proposed RCSB PDB User Community series. The PDB50 celebration also provides a great opportunity to promote the RCSB widely and emphasize the impact of structural biology on science, technology and medicine.

Recommendations

- Leverage the PDB50 celebrations for outreach where possible.
- In response to the request for the committee to comment on whether users should be surveyed, and if so, on what topic/s, we saw several potential opportunities. Receiving input on integrated/hybrid structures and their potential uses would be valuable in these early stages of that field. It was also felt that obtaining community input on the RCSB.org website features would likely be beneficial. Reaching out to educators for their input is one way to help make sure that the site meets the needs of that important group, and by extension many students. The committee suggests gathering focus groups of undergraduate instructors who use the PDB in teaching at national conferences like ASBMB, ASCB, ABRCMS and SACNAS to learn more about how they use the PDB and to even teach them how to use the newer tools that are being developed. We encourage the PDB to host

focus groups at national conferences that target specific areas, like COVID-19, cancer research or drug development.

- Given the likely reduction in the number of conferences in the coming years it would be advantageous to make webinars, with presentations and Q&A sessions, a regular RCSB offering.

Operations and Funding

The committee heard about several new arrivals in the team, and some departures. The RCSB is applauded for looking broadly in their recruitment efforts, including outreach to underrepresented minorities and the LGBTQ+ community. The committee was also very excited to hear about new successful proposals for funding to NSF, NIH and DOE. The team is commended for their efforts in this area, which are clearly paying off. The efforts to migrate to cloud computing and expand the capabilities of the PDB-Dev system are both strongly supported by the committee. Finally, we heard that Andy Byrd will be stepping down from the committee after 12 years of service. We all thank Andy for his efforts and leadership.

Recommendations

- Identify a replacement for Andy Byrd from the NMR community in a timely manner.

Next Advisory Committee Meeting: scheduling, location

The committee supports the plan for an in-person meeting in Washington to enable engagement with federal program managers. Spring 2022 is good timing, with the meeting timed to not conflict with other major meetings.

Appendix 1: Responses to 2020 RCSB PDB AC Recommendations

RCSB PDB thanks the Advisory Committee for their participation and thoughtful meeting report. Our responses to recommendations bulleted in the report follow (from September 2020) with updates noted as of March 2021.

Executive Summary

A brief report in the next 2 months on the impact of the COVID-19 pandemic on the RCSB operations and PDB depositions would be very helpful for the committee.

RCSB PDB Response:

The majority of RCSB PDB staff have been working remotely since March with no discernible impact on ongoing Operations. All meetings take place regularly via Zoom. The [Summer 2020 RCSB PDB Newsletter "Message from the RCSB PDB" article](#) described on-going activities during the pandemic. The team has worked hard and very well together despite the challenges of remote working, child care issues, home schooling, etc.

The biggest change from our usual sequence of events annually has been on undergraduate research support. Originally, two students were scheduled to perform research on-campus with the RCSB PDB at Rutgers. When programming moved online, we quickly planned a one-week Boot Camp that hosted 31 students followed by a five-week virtual research experience for 12. The results of this research project are currently in preparation for submission to a journal.

As of September 16, 390 SARS-CoV-2 structures have been released in the PDB archive. Each entry has been validated, expeditiously reviewed, and annotated by wwPDB biocurators following these "guiding principles":

- Biocuration of COVID-19 structures is prioritized over that of other structures, including post release revisions such as citation updates
- Authors are encouraged to release their structures immediately
- Consistent taxonomy name and ID ([Severe acute respiratory syndrome coronavirus 2; 2697049](#)) are applied to all COVID-19 structures
- Consistent UniProt referencing is incorporated: [P0DTD1](#), [P0DTC1](#), [P0DTC2](#), [P0DTC9](#) All released SARS-CoV-2 structures and related resources are highlighted at <http://RCSB.org/covid19>.

Update: RCSB PDB has continued to work remotely, and we anticipate this to continue at least until Fall 2021. Summer undergraduate experiences will be virtual.

All RCSB PDB services have maintained a high level of availability and/or productivity. The main challenges have been adapting activities that work best in-person (e.g., wwPDB OneDep Summit; educational activities; professional society meetings).

Detailed Advisory Panel Comments and Feedback

COVID-19: Are there other projects in this area we could develop to support research and education?

Beyond their COVID-19 current efforts, the RCSB should consider additional activities. The committee identified a number of opportunities, including:

- Creating PDB101s on viral infection processes, immunity, virus-mediated acute respiratory syndrome.
- Extending the current site to provide structural information and links to other material (experiments, recent news, etc) about each protein from the viral genome. Enabling or more

directly supporting the collection of revisions of structures from the community, which could eventually lead to new version uploads by the original authors.

- Reaching out to the local community to provide information about the basic structure of COVID proteins, viral RNAs, interacting cellular proteins, virus and pathogen in relationship to human diseases through TV news stations, school districts or public health departments.

The committee also recognizes that there may be opportunities to combine education and outreach activities around COVID-19 with fund raising activities, especially with the PDB50 celebrations next year.

Recommendations

Develop an action plan for expanding the RCSB role in educating the community about COVID-19 and other related pandemics, and the role of structural biology. These plans would ideally be integrated with current and future fundraising activities.

Track access to the RCSB maintained COVID-19 materials; this would be very helpful for future efforts to highlight the impact of the resource.

RCSB PDB Response:

Since the AC meeting, new SARS-CoV-2 materials at PDB-101 have included

- Molecule of the Month: [SARS-CoV-2 Spike](#) and [RNA-dependent RNA Polymerase](#)
[New series: Resources to Fight the COVID-19 Pandemic](#)
- [Coronavirus Life Cycle painting](#)
- [Coronavirus Background for virtual meetings](#)
- Curricula: [COVID-19 in Molecular Detail](#) and [COVID-19 Evolution and Structural Biology](#)

Educational materials are added to the [Coronavirus Browse feature at PDB-101](#). We plan to continue developing resources throughout the pandemic.

As of June 30, coronavirus-related content accounted for 13% of 1.3 million page views at PDB-101, including

- MOTM main protease: 69,239 views
- MOTM Spike protein: 11,817 views
- Coronavirus images, video, *etc.*: 99,247 views

In addition, the coronavirus hand-washing video has been viewed >400K times directly on YouTube. This video was also translated into Japanese and distributed by Protein Data Bank Japan.

At RCSB.org, the URL <http://rcsb.org/covid19> links to all SARS-CoV-2 structures and related resources; this page has been accessed >75K since March 25. Review of RCSB.org traffic shows frequent activity on SARS-CoV-2 pages. The Structure Summary page for the first structure (PDB ID 6lu7) has been accessed >100K times.

We will continue to monitor coronavirus traffic to PDB-101 and RCSB.org, and plan to use the coronavirus story to develop materials for fundraising.

Related SARS-CoV-2 activities have included

- [Image contest held in May](#)
- [Virtual Boot camp](#) focused on the SARS-CoV-2 Nsp5 main protease held June 22-26 (31 students)

- Boot camp described in “COVID-19 Evolution and Structural Biology” (2020) *Biochemistry and Molecular Biology Education*, doi: [10.1002/bmb.21428](https://doi.org/10.1002/bmb.21428).

- Undergraduate research experience exploring the full virus June 29-July 30 (12 students) ○ 4 students presented posters at the American Crystallographic Association Meeting

- [1 student won the MiTeGen-Society of Physics Students Undergraduate Poster Prize](#)

Update: Additional PDB-101 COVID-19-related content has included

Molecule of the Month:

[SARS-CoV-2 Spike](#) | [SARS-CoV-2 Spike and Antibodies](#)

Resources to Fight the Pandemic:

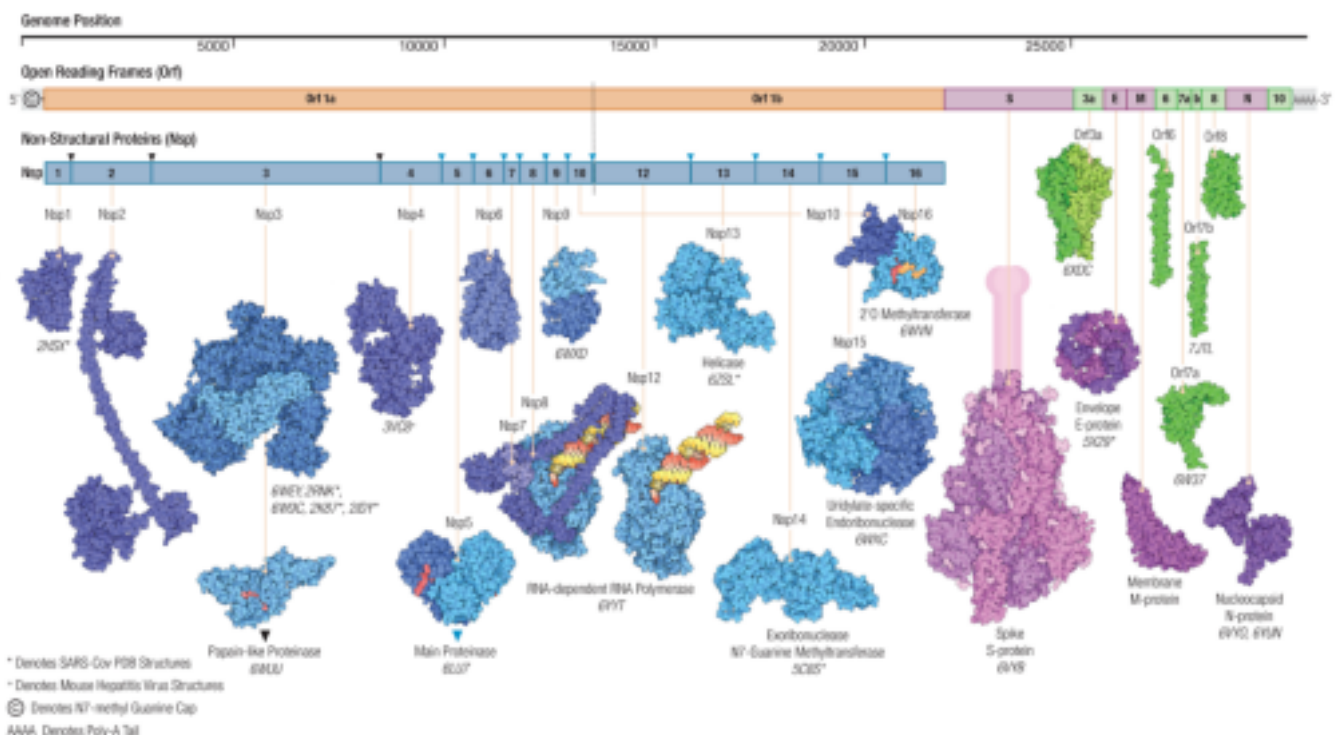
[Dexamethasone and Cytokine Storms](#) | [SARS-CoV-2 mRNA Vaccine](#)

Paintings:

[Respiratory Droplet](#) | [SARS-CoV-2 and Neutralizing Antibodies](#) | [SARS-CoV-2 Fusion](#) | [SARS-CoV-2 mRNA Vaccine](#)

Publications:

- COVID-19 Evolution and Structural Biology (2020) *BAMBed* 48:511–513 doi: [10.1002/bmb.21428](https://doi.org/10.1002/bmb.21428).
- Architecture of the SARS-CoV-2 genome and proteome from *bioRxiv* 2020.12.01.406637; doi: [10.1101/2020.12.01.406637](https://doi.org/10.1101/2020.12.01.406637)



At RCSB.org, a dedicated SARS-CoV-2 resources page was updated weekly at <http://rcsb.org/covid19>. In 2020, this page was visited by ~152,000 visitors more than 227,000 times to access the latest SARS CoV-2 structures and related educational resources.

At PDB101.rcsb.org, COVID-19-related content received 294,311 page views (~11% of total traffic calendar 2020). 180,303 views were in the first 6 months The most popular feature (Coronavirus

Proteases) was accessed >83,000 times; the spike protein (June 2020) had 20,358 views; and RNA dependent RNA Polymerase (September 2020), 17,860 views.

For comparison, the most popular article released in 2019 was Measles Virus Proteins, which was released in March and received 14,721 views that year.

Deposition/Biocuration: Any concerns about the Deposition/Biocuration work underway with our wwPDB partners?

It is also clear that the wwPDB will need to accommodate significant growth in the deposition of atomic resolution models from cryo-EM in the next 5 years. At the same time new XFEL approaches are gaining in popularity. It also currently looks unlikely that there will be dramatic reduction in the number of crystallographic structures deposited each year. This increased volume of structures will need to be processed without a backlog developing. The committee was very pleased to see that an analysis had been performed to provide projections of depositions from 2020 to 2024 for all of the experimental techniques. However, there was a concern that the projections for cryo-EM might be underestimated and not reflective of the current exponential growth.

Recommendations

- Continue to monitor the growth of cryo-EM depositions, and be prepared to prioritize the implementation of deposition standards and tools to help respond to the increased load.
- Continue to track the time taken for depositions, to both measure the load on annotators, and to provide metrics about how process improvements are increasing deposition throughput. This information will be helpful for funding justifications in the future.

RCSB PDB Response:

We shall continue to monitor the growth in cryo-EM depositions closely with the goal of improving our ability to forecast. In addition, we shall continue to evaluate the performance of the wwPDB OneDep System and the efficiency of wwPDB Biocurators for processing cryo-EM depositions and make improvements to the software and our standard operating procedures where indicated.

Update: We have improved metadata capturing by providing a controlled vocabulary. This includes generation of enumeration lists for microscope, detectors, and plunger. In addition, EM map analysis and 3D views of map overlay with models have been added to the validation PDF reports with the goal to reduce depositor communication and reprocessing workload on biocurators. Overall biocuration efficiency has been increased by the implementation of sequence autoprocessing based on the previous annotation of the same protein in an existing entry.

New and Improved RCSB.org: Additional Site and Search Functionality requests?

The committee heard from John Westbrook on the ongoing efforts to improve the infrastructure for the rcsb.org website and associated backends. We were impressed how quickly this has been implemented without any substantial interruptions to providing services. Demonstrations of the new search functionality highlighted useful new features. However, there were some concerns about the complexity of the search system for many users, and the loss of important features (such as refining a search to provide a non redundant set of results). One suggestion was the creation of question-driven functionalities and workflows for popular search activities. The committee recognizes that the RCSB has undertaken community outreach to get user feedback, but these efforts might need to be extended. The new Mol* 3D visualization system was also presented, and clearly shows great potential for interactive display of molecules and maps. However, the committee feels that further development, and in some cases simplification, of the interface would benefit many of the RCSB users.

Recommendations

- Seek further community input, maybe through the creation of focus groups or targeted outreach, to refine the search functionality and the Mol* visualization services.

RCSB PDB Response:

Mol*: Since the AC meeting, the user interface has been greatly improved based upon feedback from the community and [user documentation was published at RCSB.org](#). Mol* was used extensively during the summer with students and their feedback was positive. After boot camp, which included many students new to molecular visualization, 65% said Mol* was “easy to use.”

RCSB.org: Since April 2020, we have released many additional features (some new and some improved versions of previous offerings), such as returning non-redundant search results, to our searching and reporting services. Due to the nature of the database/software architecture, we were unable to release all services at the same time.

We have started a project to improve documentation to help support users and plan to initiate a user survey at the end of 2020. We are monitoring access to the website to see how it is being used, and are developing tools to better analyze which searches are being performed. These metrics will guide further enhancements of usability/documentation and help prioritize development and release of new features.

Update: The plan to increase documentation [has been published under a new menu at the top of every RCSB.org page](#).

Traffic at the help desk at info@rcsb.org regularly monitored. The spike in traffic that occurred when the new search functionality was released has decreased as users have become more familiar with the new features.

Outreach/Education: Suggestions for new materials and virtual venues for celebrating PDB50 throughout 2021?

The committee had several suggestions for ways to engage the community in the current circumstances and leverage the importance of structure in the COVID-19 response. One idea was the creation of virtual reality resources, perhaps using ChimeraX. This might provide a platform to propose something similar to Folding@Home where people would be in VR or AR space and attempting to design drugs for COVID-19 proteins. Another suggestion was a competition for creating protein structures from found objects around the home.

The PDB50 celebration in 2021 provides a great opportunity to promote the RCSB widely and emphasize the impact of structural biology. The committee suggested addressing this in multiple forums. Large conferences provide an opportunity for outreach, and in some cases these may be well organized as virtual conferences - the Intelligent Systems for Molecular Biology in 2021 was one example. Museums and other public facing organizations may also provide a great opportunity for engaging a broader audience. The American Museum of Natural History was put forward as an organization looking for online content. Ultimately, the committee feels that there is an opportunity for either local or national recognition through mainstream media, such as NPR and network TV channels. Science Friday at NPR would be a great target for a PDB50 piece, as would a NOVA documentary.

Recommendations

- Create a plan for online outreach and communications for the next 12 months, which incorporates some PDB50 celebration activities
- Develop a PDB50 media communications strategy targeted both locally and nationally.

RCSB PDB Response:

Select PDB50 activities and materials are being developed in collaboration with our wwPDB partners as listed at <https://foundation.wwpdb.org/pdb50.html>. The wwPDB is also collaborating on a themed

calendar for 2021. Other projects are also being discussed.

In addition, we are developing pitch materials to provide to committee members who have volunteered to help us target museums and media about the PDB using the coronavirus story.

Update: PDB50 celebrations and related materials are being updated at <http://wwpdb.org/pdb50> and <http://rcsb.org/pdb50>. Meetings are being coordinated by the wwPDB. Current registration numbers for the inaugural May meeting indicate >700 registrants and 280 posters.

Next Advisory Committee Meeting location?

While the committee looks forward to the next meeting in person, it seems unlikely that it will come to pass in the first half of 2021. We expect that the next committee meeting will be held virtually. If an in-person meeting is possible, the option of coordinating with the PDB50 event at the ACA in Baltimore seems reasonable.

Recommendations

- Plan for a virtual meeting in the first half of 2021. The duration of the meeting will be increased to 4 hours (per AC Chair feedback).

RCSB PDB Response:

A doodle poll for available dates in April 2021 will be circulated. Given the AC Chair's other duties, two meeting dates will be confirmed during that month to provide a reserve date in the event that the Chair becomes unavailable on short notice.

Discussion: What advice do you have regarding PDB Legacy Format sunseting?

The committee agrees that a timeline for a permanent transition needs to be established, communicated, and enacted. The details of that timeline will require some careful consideration. Firstly, depositions for all structure types will need to move to mmCIF/PDBx before the legacy format can be dropped. Currently, there are plans for cryo-EM submissions to move to mmCIF/PDBx later this year. The committee is still uncertain as to the plans for NMR-based structure depositions. Provided that the milestone of technique-wide mmCIF/PDBx deposition is either reached or scheduled, it seems reasonable to set a firm deadline for dropping support for serving the legacy format files to the RCSB user community. Clearly, extensive outreach and communication should be performed prior to this.

A general transitional approach to the legacy format was also raised by the committee, where the legacy format versions of structures currently available are frozen, and they are not changed any further. This would send the message that mmCIF is the only format capable of handling complex structures but wouldn't remove access to all structures on a specific date, which could present problems for some. This approach also minimizes the work necessary for the RCSB to that of creating a static repository of legacy format structures. The committee did see that there are some additional remaining issues. For example, biological assemblies are currently not available in mmCIF format from RCSB, although they are available from PDBe. If the legacy format is going to be dropped, these files need to be provided sooner rather than later. The mmCIF format is much better than the legacy format for assemblies, which use the MODEL-ENDMDL format to show multiple copies of the ASU in larger assemblies. There is also a good case for the RCSB to generate additional material to help educators know how to best use mmCIF format files. The PDB 101 has a site about beginners using mmCIF/PDBx, which is very helpful. However, it would be even better to have one or two videos for educators that explain how to teach with structures in this file format.

Recommendations

- Develop a timeline for sunseting the PDB format, which can be widely socialized for feedback. The timeline will need to be consistent with the termination of legacy format model submission for all experimental methods. Consideration could be given to providing an unsupported set of

legacy version files to ease the transition for some researchers.

- Make it a priority to address cases where mmCIF files are currently not available, such as assemblies.

RCSB PDB Response:

At the wwPDB AC Meeting in October 2020, the wwPDB PIs will ask the advisors to concur with work with the user community to understand how many rely on “best efforts” PDB Legacy format files and how they use these files in their day-to-day research and teaching activities.

The wwPDB will analyze the results of the survey and develop plans for deprecation of the legacy PDB file format that minimizes disruption to users and promotes the FAIR principles of findability, accessibility, interoperability, and reusability. The wwPDB will also determine how best to help our diverse user community transition to the PDBx/mmCIF format regime with webinars, conversion software, on-line tutorials, etc.

Update: We have begun the work to provide better logs and tracking that will enable statistics analysis on file download for different file types which will inform us about the usage of “best efforts” PDB format files.

Funding and Operations

It was also noted that building stronger links to other research communities traditionally outside of structural biology could pay dividends in the future. Areas identified included bioengineering and synthetic biology, which are relevant to all three federal funding agencies. Developing stronger plans for funding to support education activities could be another productive route, which might leverage the scientific response to COVID-19.

Recommendations

- Reassess the approaches to obtaining new funding given the current global financial situation.
- Consider alternative routes to funding to support RCSB educational activities.
- Develop a strategy for broader community engagement outside structural biology to develop new routes to joint funding.

RCSB PDB Response:

We will continue to identify funding opportunities focusing on federal funders, and will emphasize the impact PDB data and RCSB PDB services have had on the COVID-19 pandemic.

Appendix 2: PDB 2020 Metrics

In aggregate, 15,436 depositions were received and processed between January 1st and December 31st, 2020, with an average turnaround of two weeks by the wwPDB. This represents an increase from the 13377 entries deposited in 2019. Based upon the number of entries deposited this year to date, it is estimated that PDB will receive a number of depositions in-between 2019 and 2020 totals.

Breakdown of depositions by discipline was as follows:

X-ray: 12242 (79% of entries deposited, up from 10952 in 2019)

NMR: 390 (2.5%, down from 403)

EM: 2780 (18%, up from 1996)

Other: 24 (.2%, down from 26)

Breakdown of depositions by wwPDB processing site was as follows:

RCSB PDB: 7190 (47%)

PDBj: 3438 (22%)

PDBe-EBI: 4808 (31%)

Breakdown of depositors by location was as follows:

North America 33.3%

Europe 40.9%

Asia 22.0%

South America 1.0%

Oceania 2.5%

Africa <1%

During 2020, RCSB PDB's website at <http://rcsb.org> was visited by millions of unique visitors.

In 2020, data files from the PDB archive were accessed >575 million times from RCSB PDB-hosted FTP and websites (up from 547 million in 2019). Additional data were downloaded from wwPDB partners PDBe and PDBj.