Celebrating
PROTEIN
DATA BANK
Years

RCSB PDB Advisory Committee

Virtual Meeting

Monday, April 26, 2021

Background Information Slides

Slides with a blue title are provided as background information and can be presented and discussed at the meeting by request.

Main Slides Appear with Green Titles

Agenda

11:30am PDT	2:30pm EDT	Executive Session	Paul Adams/Breakout Room *via* Ken Dalenberg
11:45	2:45	Welcome and Introductions	Stephen K. Burley
11:55am	2:55	Overview; Response to COVID-19	Stephen K. Burley
12:10pm	3:10	Deposition/Biocuration Status	Jasmine Young
12:25	3:25	PDB Archive Status	John Westbrook
12:35	3:35	New/Improved RCSB.org	Jose Duarte/Yana Rose
12:50	3:50	PDB Data Delivery Status	Robert Lowe
1:00	4:00	Outreach/Education; PDB50	Christine Zardecki
1:15	4:15	Operations and Funding	Stephen K. Burley
1:30	4:30	Break	
1:45	4:45	Discussion	
2:15	5:15	Executive Session	Breakout Room *via* Ken Dalenberg
3:00	6:00	Feedback from AC	Stephen K. Burley, Andrej Sali, Christine Zardecki
3:30	6:30	End	

Introductions to Participants

Advisory Committee (full Roster)

- Paul Adams (Chair), Peter Andolfatto, Judy Blake, Andy Byrd, Bridget Carragher (absent), Wah Chiu, Kirk Clark, Paul Craig, Robert B. Darnell, Roland Dunbrack, Paul Falkowski, Thomas Ferrin, Mandë Holford, Cathy Peishoff, Sue Rhee, Torsten Schwede, Jill Trewhella

RCSB PDB

- Stephen K. Burley, Helen M. Berman
- Rutgers: Robert Lowe, John Westbrook, Jasmine Young, Christine Zardecki
  - IT Support: Ken Dalenberg
- UCSD: Jose Duarte, Yana Rose
- UCSF: Andrej Sali

Funders

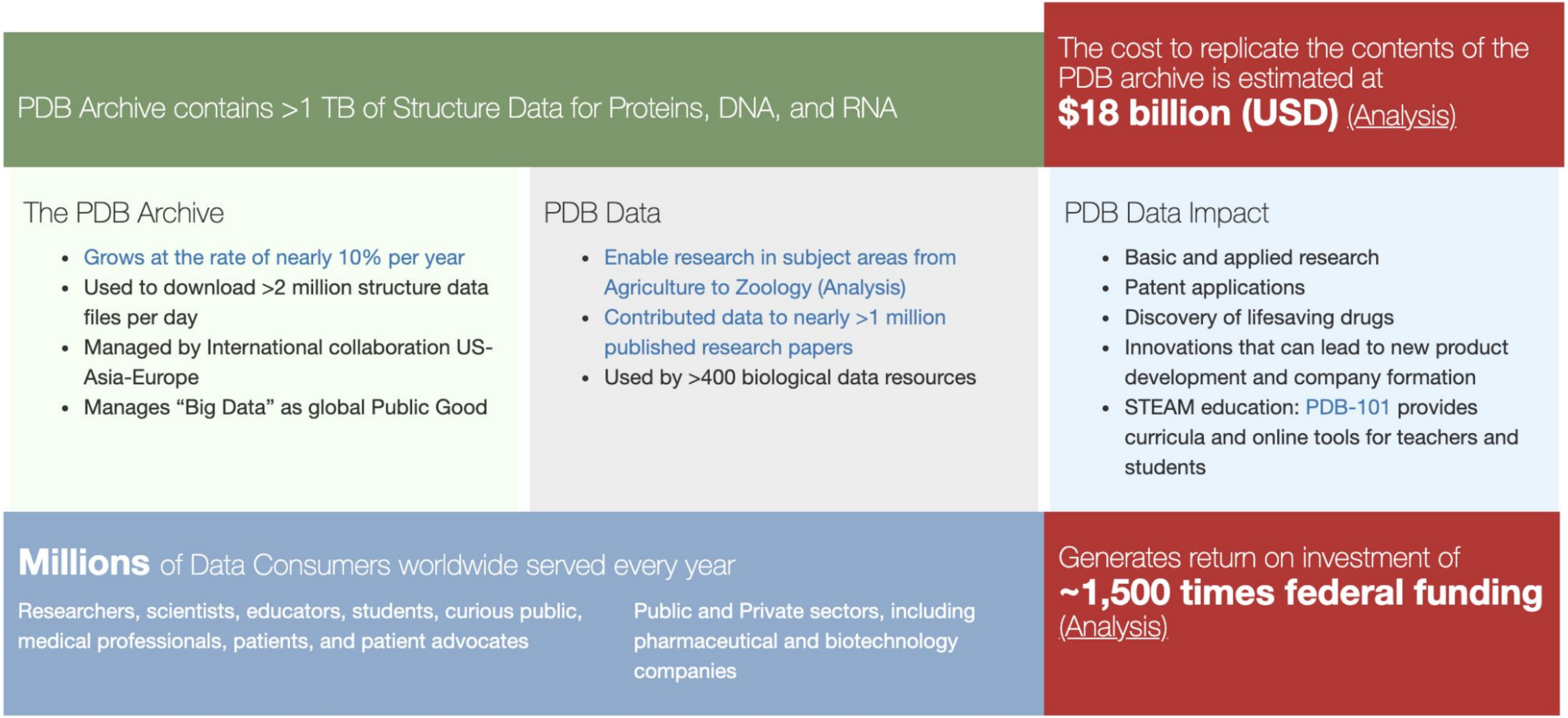
- Steven Ellis (NSF), Ramana Madupu (DOE), Amy Swain (DOE), Susan Gregurick (NIH OD, absent) Jerry Li (NCI, absent), Paula Flicker (NIGMS)

Overview

Stephen K. Burley

2020 Covers by RCSB PDB

Publicly Available PDB Data



PDB Archive contains >1 TB of Structure Data for Proteins, DNA, and RNA

The cost to replicate the contents of the PDB archive is estimated at **$18 billion (USD)** (Analysis)

The PDB Archive

- Grows at the rate of nearly 10% per year
- Used to download >2 million structure data files per day
- Managed by International collaboration US-Asia-Europe
- Manages "Big Data" as global Public Good

PDB Data

- Enable research in subject areas from Agriculture to Zoology (Analysis)
- Contributed data to nearly >1 million published research papers
- Used by >400 biological data resources

PDB Data Impact

- Basic and applied research
- Patent applications
- Discovery of lifesaving drugs
- Innovations that can lead to new product development and company formation
- STEAM education: PDB-101 provides curricula and online tools for teachers and students

Millions of Data Consumers worldwide served every year

Researchers, scientists, educators, students, curious public, medical professionals, patients, and patient advocates

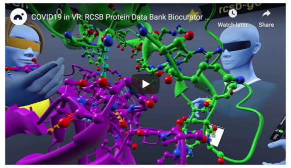
Public and Private sectors, including pharmaceutical and biotechnology companies

Generates return on investment of **~1,500 times federal funding** (Analysis)

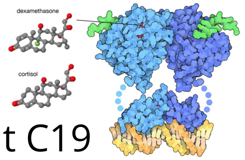
Background Information: https://www.rcsb.org/pages/about-us/index

2020 Highlights

~15K depositions (~800 SARS-CoV-2)
End of Classic Architecture

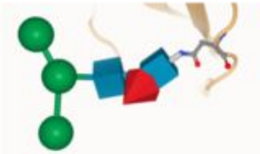


Gregg enters VR



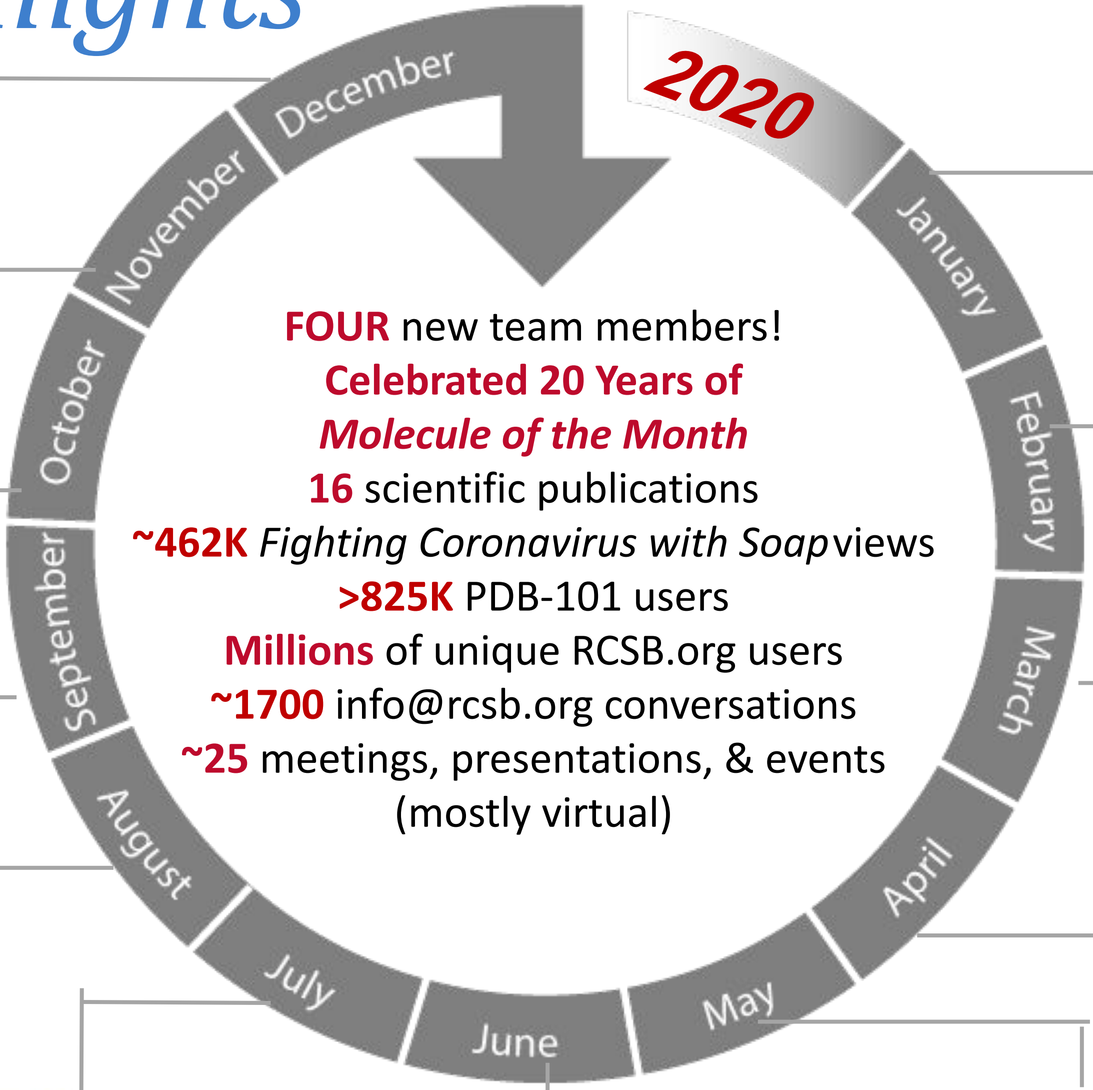
Resources to Fight C19

Data.rcsb.org
API tutorials

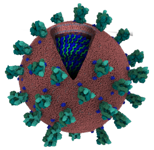


Carbohydrate remediation

virtual ISMB 2020
iSCB



FOUR new team members!
Celebrated 20 Years of *Molecule of the Month*
16 scientific publications
~462K *Fighting Coronavirus with Soap* views
>825K PDB-101 users
Millions of unique RCSB.org users
~1700 info@rcsb.org conversations
~25 meetings, presentations, & events (mostly virtual)



RISE at Rutgers
Research Intensive Summer Experience

IQB Boot Camp

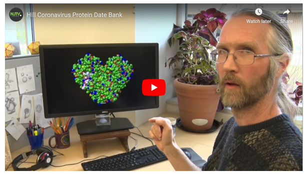
HS Video Challenge

IQB Boot Camp
Cloud Technologies

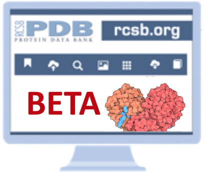
The New York Structural Biology Discussion Group Winter Meeting

RU: Undergraduate Course
Drugs and the Brain

BioXFEL



NJTV

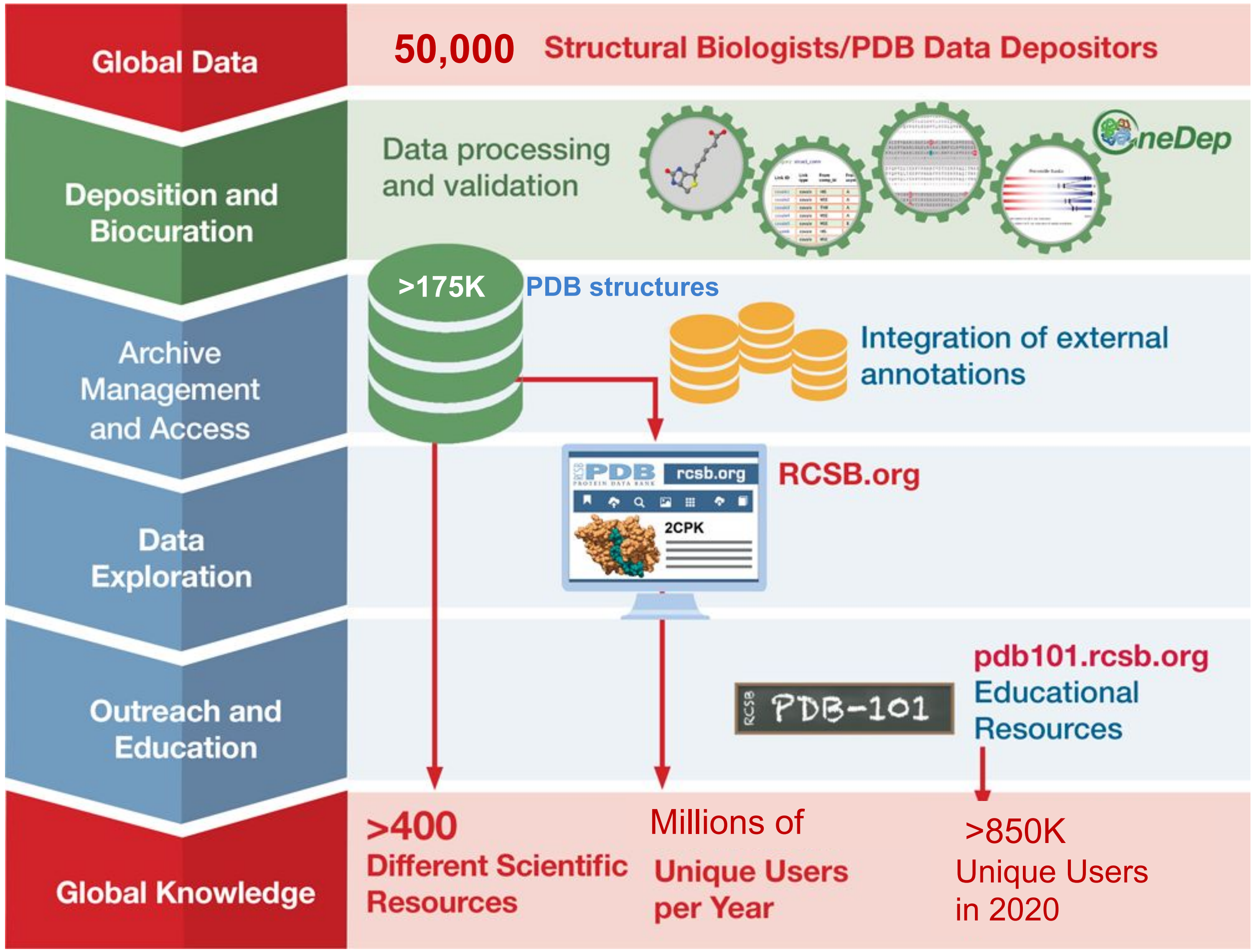


104 SARS-CoV-2 PDBs

New Search Released

RCSB PDB
Advisory Committee

RCSB PDB: Driving Breakthroughs in Research and Education



Global Data

50,000 Structural Biologists/PDB Data Depositors

Deposition and Biocuration

Data processing and validation

OneDep

Archive Management and Access

>175K PDB structures

Integration of external annotations

Data Exploration

RCSB.org

Outreach and Education

pdb101.rcsb.org Educational Resources

Global Knowledge

>400 Different Scientific Resources

Millions of Unique Users per Year

>850K Unique Users in 2020

Deposition, Validation, Biocuration, Remediation

Data Standards, Integration, Storage, Access

Web Portal, Search/Analysis, 3D Visualization

Website and In-person Outreach

Protein Structure/Central Dogma Now Part of 2020 Culture

PROTEIN DATA BANK


How to Make Your Own Coronavirus Piñata (and Then Smash It)

Join the craft renaissance and cope at the same time.

nytimes.com

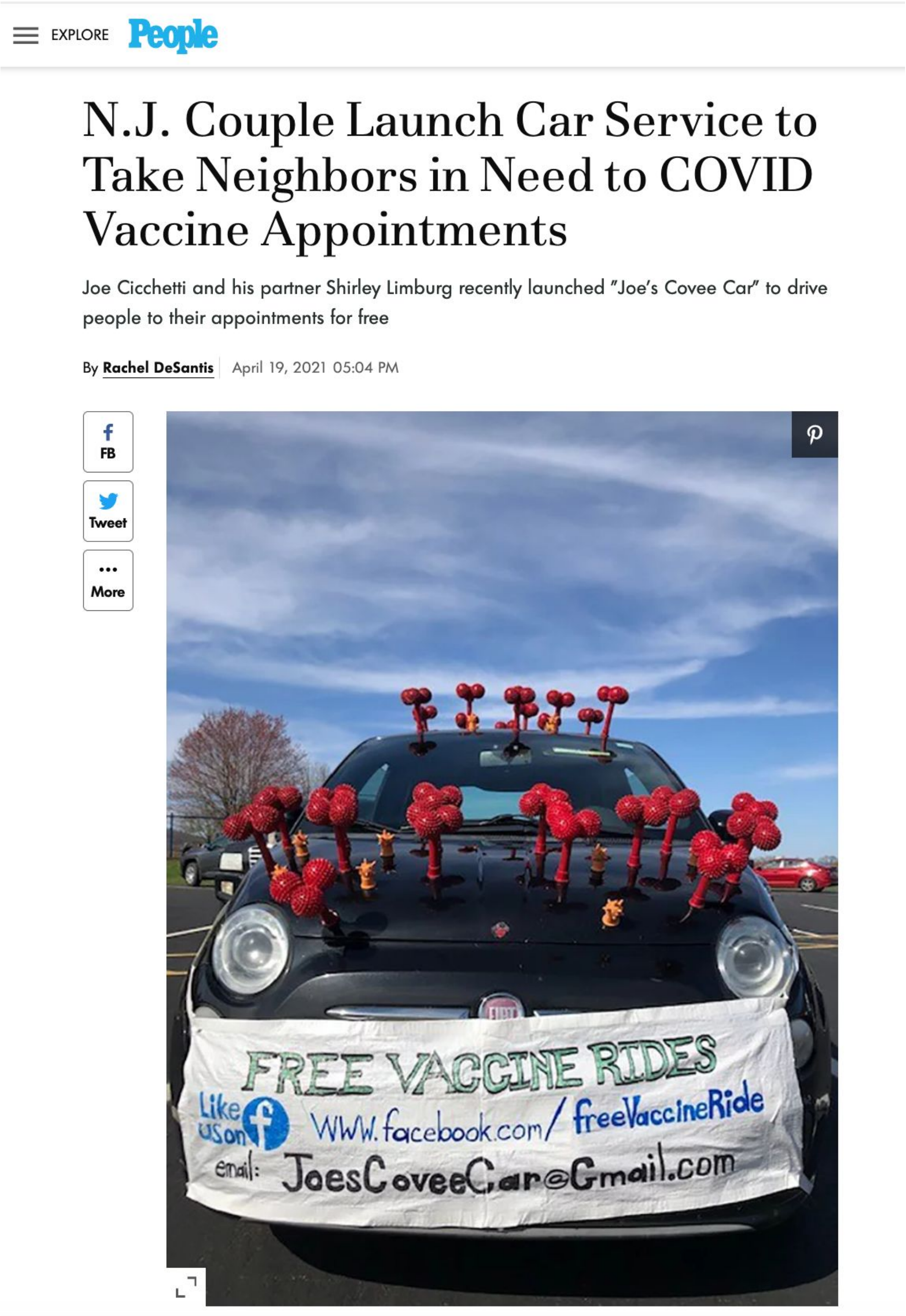

The Coronavirus Is Unveiled Down to Its Blobs and Pores

By CARL ZIMMER

Learning from History: Fauci Donates Model to Smithsonian's COVID-19 Collection

Posted on April 15th, 2021 by Dr. Francis Collins

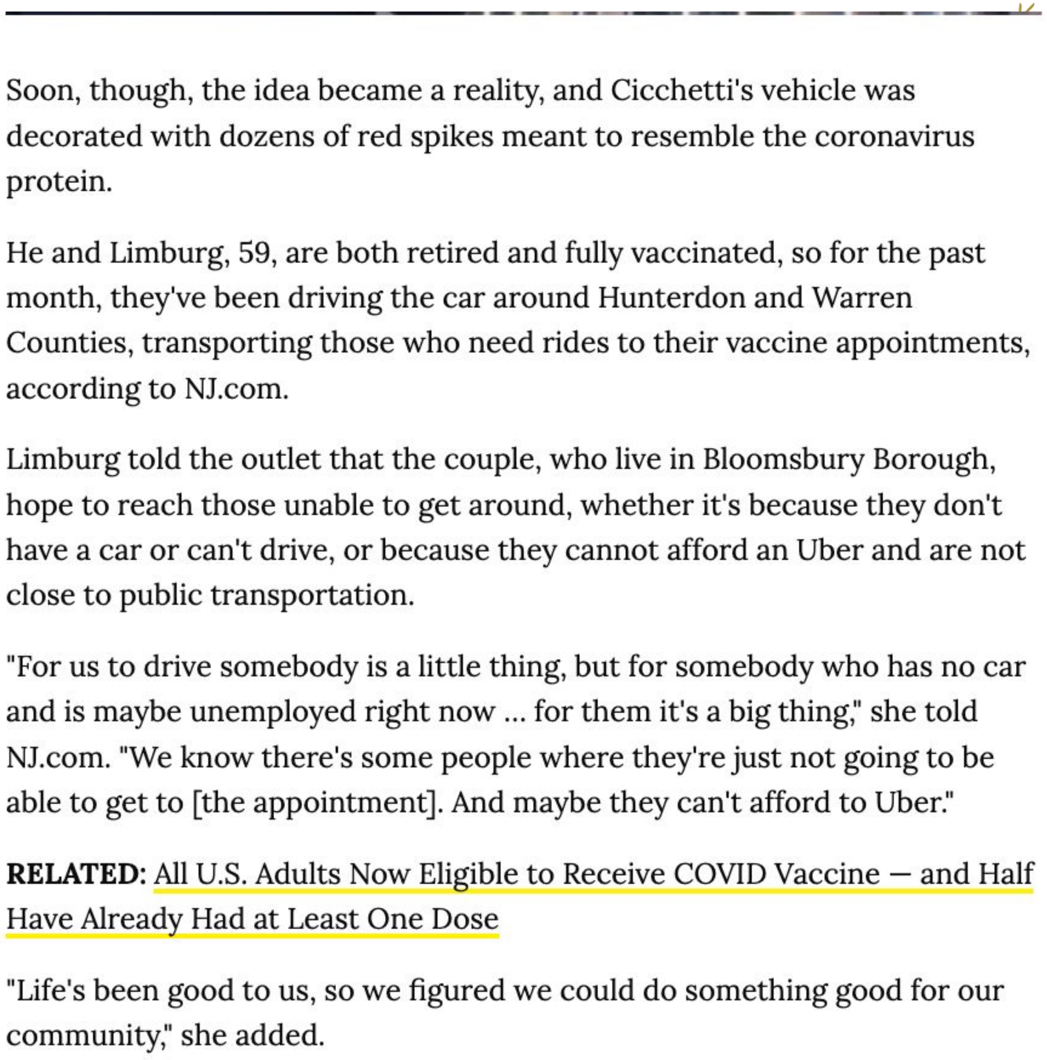

Smithsonian's Great Americans Medal | Dr. Anthony S. Fauci Explains ...



N.J. Couple Launch Car Service to Take Neighbors in Need to COVID Vaccine Appointments

Joe Cicchetti and his partner Shirley Limburg recently launched "Joe's Covee Car" to drive people to their appointments for free

By **Rachel DeSantis** | April 19, 2021 05:04 PM



Soon, though, the idea became a reality, and Cicchetti's vehicle was decorated with dozens of red spikes meant to resemble the coronavirus protein.

He and Limburg, 59, are both retired and fully vaccinated, so for the past month, they've been driving the car around Hunterdon and Warren Counties, transporting those who need rides to their vaccine appointments, according to NJ.com.

Limburg told the outlet that the couple, who live in Bloomsbury Borough, hope to reach those unable to get around, whether it's because they don't have a car or can't drive, or because they cannot afford an Uber and are not close to public transportation.

"For us to drive somebody is a little thing, but for somebody who has no car and is maybe unemployed right now ... for them it's a big thing," she told NJ.com. "We know there's some people where they're just not going to be able to get to [the appointment]. And maybe they can't afford to Uber."

RELATED: All U.S. Adults Now Eligible to Receive COVID Vaccine — and Half Have Already Had at Least One Dose

"Life's been good to us, so we figured we could do something good for our community," she added.

https://people.com/human-interest/nj-couple-launch-car-service-to-take-neighbors-in-need-to-covid-vaccine-appointments/

First *Coronavirus* painting
~55K views since Feb 6, 2020

Response to COVID-19

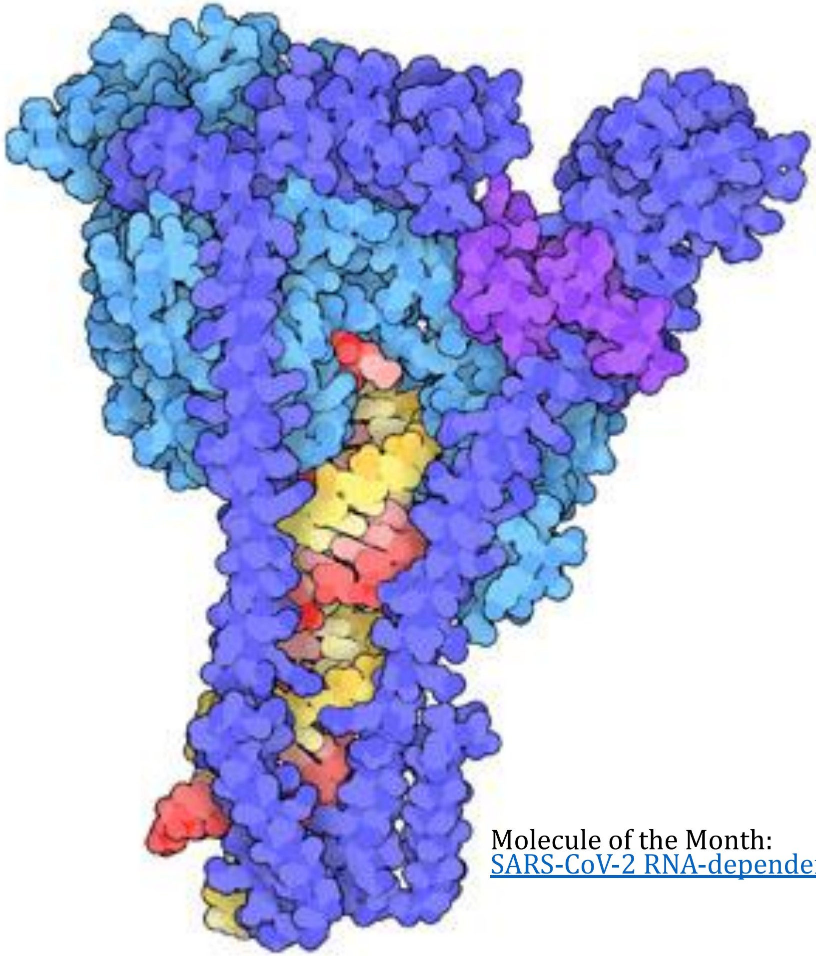
Stephen K. Burley

Impact of COVID 19 on Operations

- Remote working since March 2020
  - Regular All Hands and social hours in addition to ongoing meetings
- Biocuration of SARS-CoV-2 structures prioritized
  - >1,000 structures released as of March 2, 2021 (most immediately)
  - Covid-19 accounted ~7% of 2020 depositions
- Pandemic-related content on PDB-101 increased traffic
- Related resources updated weekly at RCSB.org/covid19

Covid-19 Depositions in 2020

- ~1000 deposited/795 released
  - 416 entries from PanDDA method submitted *via* GroupDep
- 173 PIs
- 17 Countries
- Median 34 days to release
- 6 RDRP structures (Remdesivir)
- 263 protease co-crystal structures
  - 210 with bound ligand
- 266 spike protein structures
- Numerous 3DEM contributions



Molecule of the Month:
SARS-CoV-2 RNA-dependent RNA Polymerase

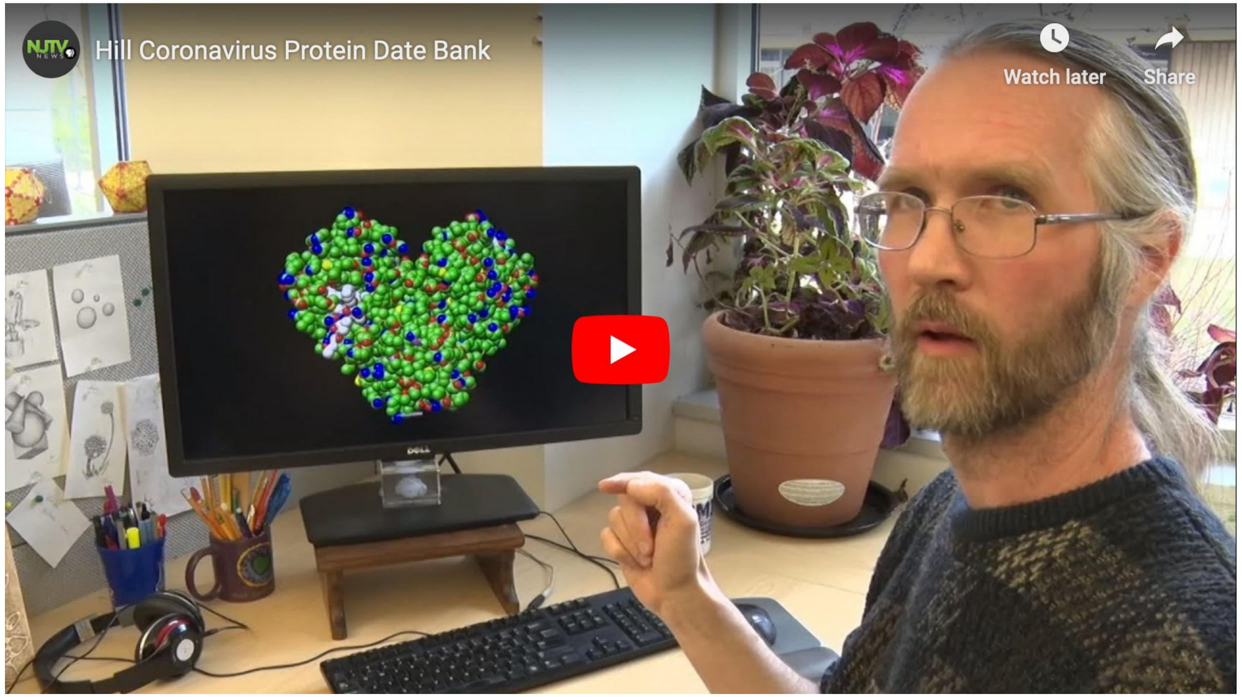
Covid-19 Data Utilization

First Structure (PDB ID 6lu7)

- >260K website views
- Downloads ~100K
- Interest in 6lu7 peaked March/April

Therapeutic Interventions

- Pfizer has put two 6lu7 inhibitors into clinical trials (IV prodrug and Oral)
- Second generation vaccine design
- Monoclonal antibody design
- *In silico* screening efforts



NJTV News spoke with Rutgers University researchers Stephen Burley and Brian Hudson at the Protein Data Bank about finding the key to fighting COVID-19. Shown: PDB structure 6lu7

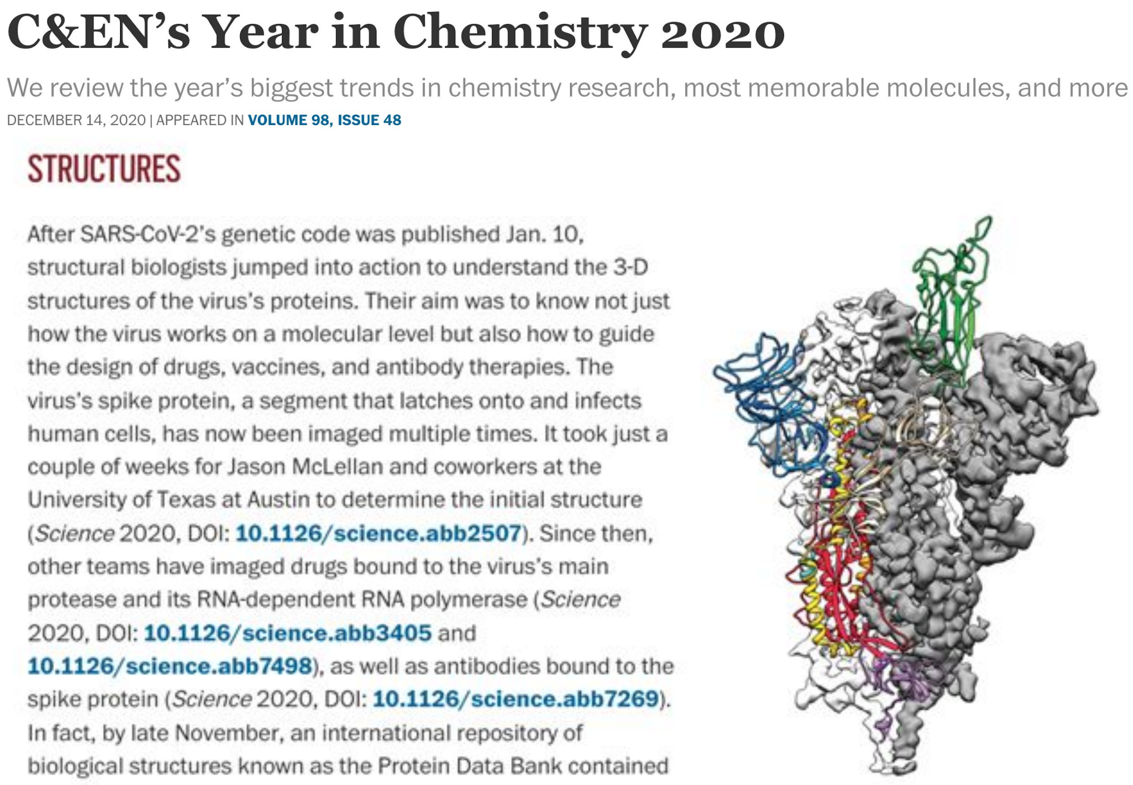
PDB Data Driving Research and Education

New SARS-CoV-2 databases utilizing PDB data (NAR 2021 Database Issue)

- CoV3D - Experimental coronavirus protein structures
- DockCoV2 - In silico drug docking against SARS-CoV2 targets
- LitCovid - Curated COVID-19 literature
- ViruSurf - Portal to SARS-CoV2 sequences and variants

Resources Providing Additional Curation

- Coronavirus Structural Task Force
- Covid-19.bioreproducibility.org



C&EN's Year in Chemistry 2020

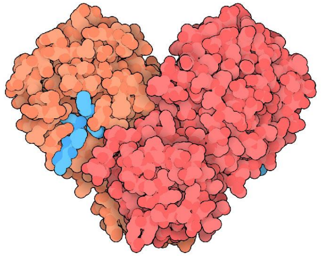
We review the year's biggest trends in chemistry research, most memorable molecules, and more

DECEMBER 14, 2020 | APPEARED IN **VOLUME 98, ISSUE 48**

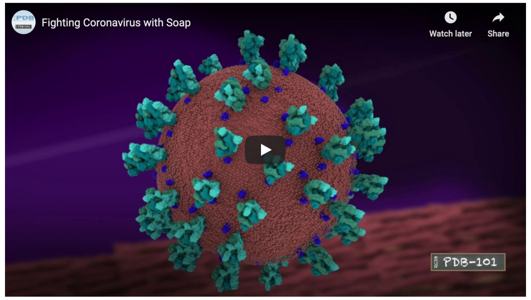
STRUCTURES

After SARS-CoV-2's genetic code was published Jan. 10, structural biologists jumped into action to understand the 3-D structures of the virus's proteins. Their aim was to know not just how the virus works on a molecular level but also how to guide the design of drugs, vaccines, and antibody therapies. The virus's spike protein, a segment that latches onto and infects human cells, has now been imaged multiple times. It took just a couple of weeks for Jason McLellan and coworkers at the University of Texas at Austin to determine the initial structure (*Science* 2020, DOI: **10.1126/science.abb2507**). Since then, other teams have imaged drugs bound to the virus's main protease and its RNA-dependent RNA polymerase (*Science* 2020, DOI: **10.1126/science.abb3405** and **10.1126/science.abb7498**), as well as antibodies bound to the spike protein (*Science* 2020, DOI: **10.1126/science.abb7269**). In fact, by late November, an international repository of biological structures known as the Protein Data Bank contained

CV19: 11% of 2020 PDB-101 Views

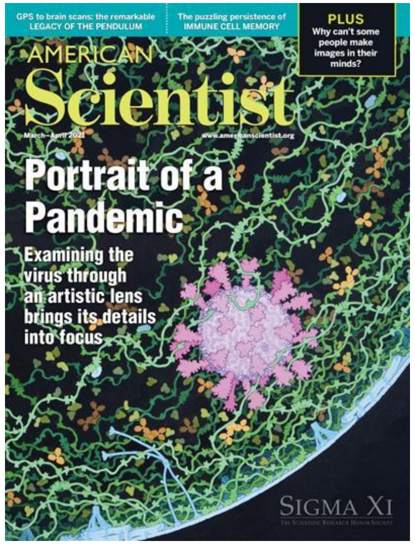


February Molecule of the Month: Coronavirus Proteases; >85K views



Video: >493K views
2020 American Public Health Association Film Festival
Official Selection

APHA Public Health FILM FESTIVAL

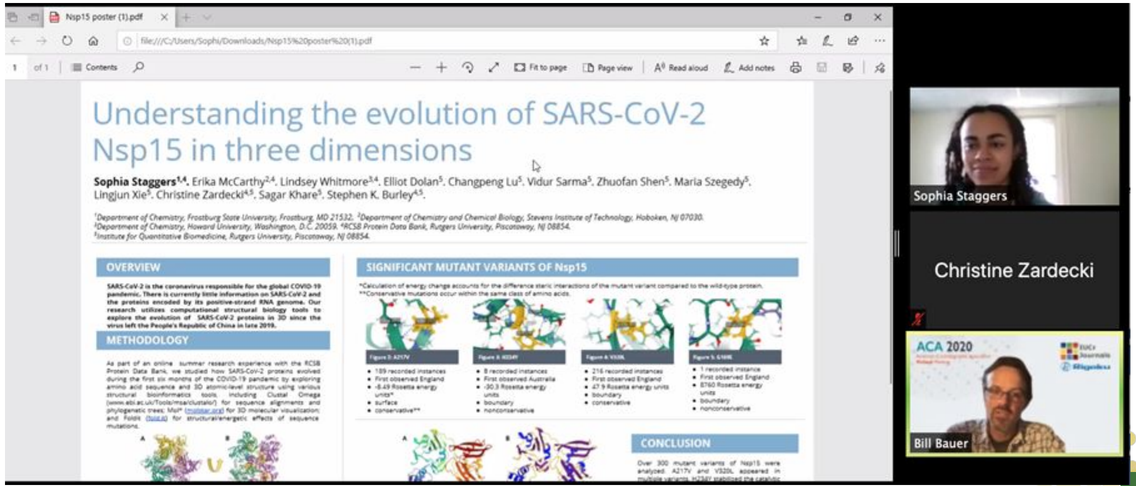


6 Molecular Landscapes by David Goodsell
73K views

Feature	Number of Page Views
MOTM Main Protease	85,645
MOTM Spike	20,047
MOTM RDRP	17,945
Soap Video	36,752
Series: Fighting the Pandemic (4 articles)	5,832
Coloring Book page	24,824
Goodsell Landscapes (6)	73,493
Images (landing pages)	2,141
Browse landing page	10,400
Covid-19 Teach	17,232
Total	**294,311**
(~11% of total traffic calendar 2020) **180,303 views in first 6 months**	

Summer of the "Coronaverse"

- One-week Boot Camp: 35 student researchers studied sequence-structure evolution of the main protease
    - Co-mentored by BASIL participants
    - Materials: https://covid19-bootcamp.rcsb.org/
    - Description: COVID-19 Evolution and Structural Biology (2020) *BAMBed* 48:511–513 doi: [10.1002/bmb.21428](https://doi.org/10.1002/bmb.21428)
- 12 continued for 5 weeks to explore the other structures
    - 3 supported by NSF REU
    - 4 presented at American Crystallographic Association
    - Sophia awarded *MiTeGen-Society of Physics Students Undergraduate Poster Prize*
- Publications
    - Architecture of the SARS-CoV-2 genome and proteome *bioRxiv* 2020 doi: [10.1101/2020.12.01.406637](https://doi.org/10.1101/2020.12.01.406637)
    - Under revision for *PROTEINS*: *Structure, Function*, and *Bioinformatics*



Understanding the evolution of SARS-CoV-2 Nsp15 in three dimensions

Sophia Staggers[1,4], Erika McCarthy[2,4], Lindsey Whitmore[3,4], Elliot Dolan[5], Changpeng Lu[5], Vidur Sarma[5], Zhuofan Shen[5], Maria Szegedy[5], Lingjun Xie[5], Christine Zardecki[4,5], Sagar Khare[5], Stephen K. Burley[4,5].

[1]Department of Chemistry, Frostburg State University, Frostburg, MD 21532. [2]Department of Chemistry and Chemical Biology, Stevens Institute of Technology, Hoboken, NJ 07030. [3]Department of Chemistry, Howard University, Washington, D.C. 20059. [4]RCSB Protein Data Bank, Rutgers University, Piscataway, NJ 08854. [5]Institute for Quantitative Biomedicine, Rutgers University, Piscataway, NJ 08854.

OVERVIEW

SARS-CoV-2 is the coronavirus responsible for the global COVID-19 pandemic. There is currently little information on SARS-CoV-2 and the proteins encoded by its positive-strand RNA genome. Our research utilizes computational structural biology tools to explore the evolution of SARS-CoV-2 proteins in 3D since the virus left the People's Republic of China in late 2019.

METHODOLOGY

As part of an online summer research experience with the RCSB Protein Data Bank, we studied how SARS-CoV-2 proteins evolved during the first six months of the COVID-19 pandemic by exploring amino acid sequence and 3D atomic-level structure using various structural bioinformatics tools, including Clustal Omega (www.ebi.ac.uk/Tools/msa/clustalo/) for sequence alignments and phylogenetic trees; Mol* (molstar.org) for 3D molecular visualization; and Foldit (fold.it) for structural/energetic effects of sequence mutations.

SIGNIFICANT MUTANT VARIANTS OF Nsp15

*Calculation of energy change accounts for the difference steric interactions of the mutant variant compared to the wild-type protein.
**Conservative mutations occur within the same class of amino acids.

CONCLUSION

Boot Camp Acknowledgements

- IQB Faculty and Graduate Students: Stephen K. Burley, Shuchismita Dutta, Brian Hudson, Jennifer Jiang, Sagar Khare, Joseph Lubin, Vidur Sarma, Lingjun Xie, Christine Zardecki
- IQB's Ken Dalenberg for technical oversight
- BASIL: Biochemistry Authentic Scientific Inquiry Laboratory) biochemistry consortium faculty: Paul Craig (Rochester Institute of Technology), Bonnie Hall (Grand View University), Julia Koeppe (SUNY Oswego), Stephen Mills (Xavier University), Mike Pikaart (Hope College), Rebecca Roberts (Ursinus College)
- RISE at Rutgers: nationally acclaimed summer research program for outstanding undergraduates from diverse backgrounds
- Rutgers Faculty Lectures: Yana Bromberg (SEBS), Siobain Duffy (SEBS), Jay Tischfield (SAS/RUCDR)

Protein Structure/Central Dogma Now Part of 2020 Culture

No Service Disruptions

Managed More Depositions

Released More Structures

Enabled More Downloads

Published More Papers

Submitted More Grants

Secured More Funding

Supported More Outreach

Hosted More URM Interns

Ran Two Boot Camps

Spoke Virtually at More Meetings

Planned wwPDB PDB50 ASBMB

Planned RCSB PDB PDB50 ACA

Planned RCSB PDB PDB50 ACS

Questions and Comments?

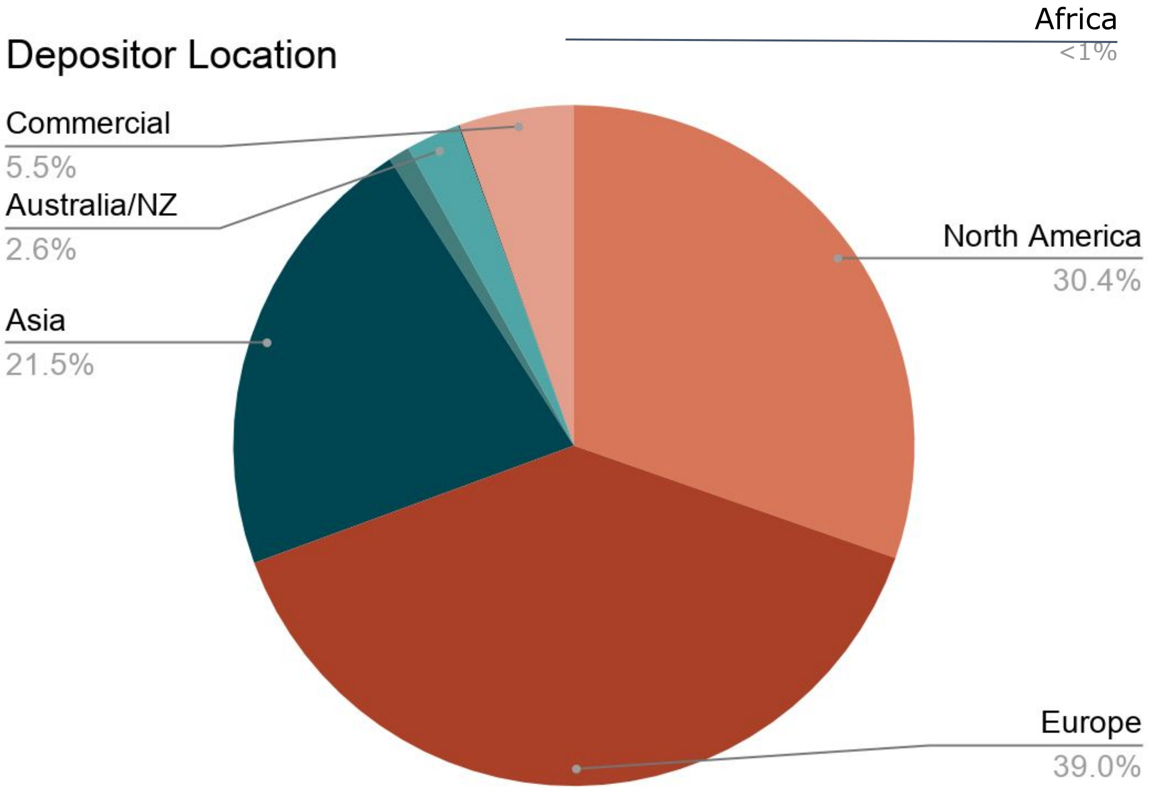
Deposition/Biocuration

Jasmine Young

2020 wwPDB Deposition/Biocuration Statistics

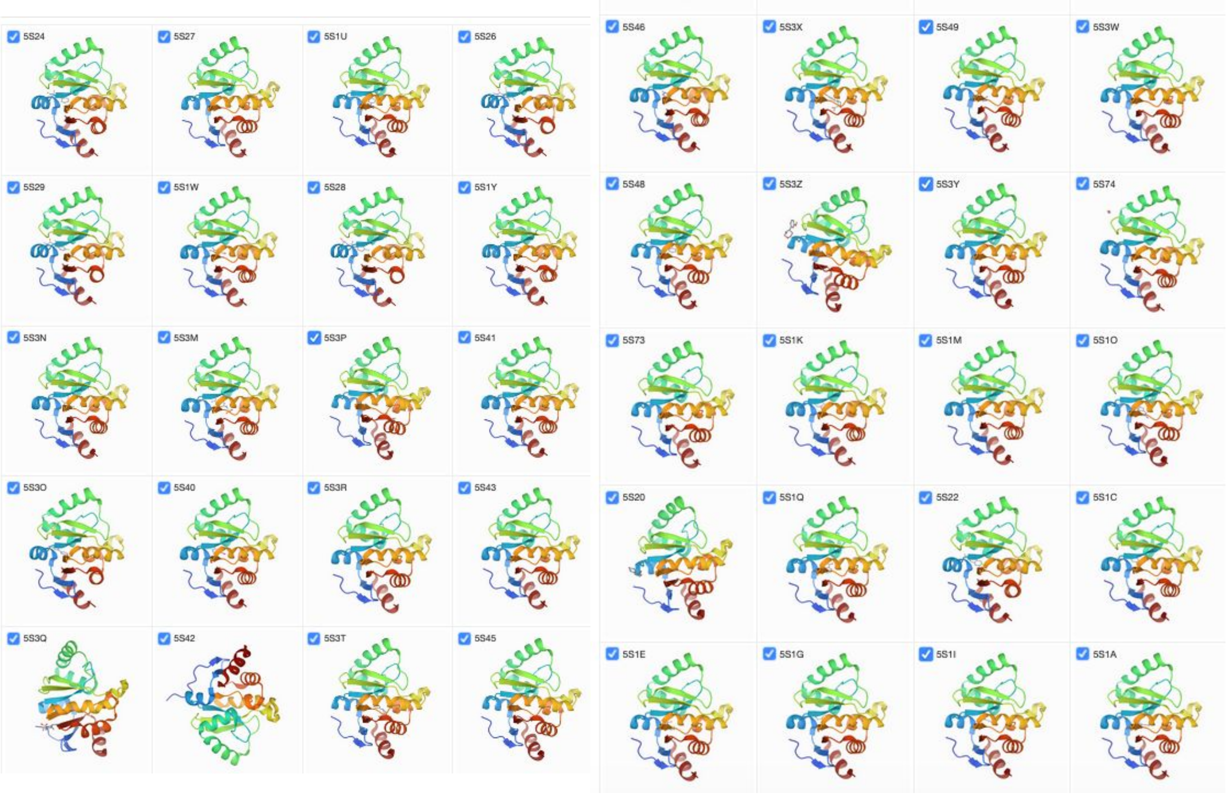
- Record 15,436 new depositions (15.4% increase over 2019)
- Record 7,190 RCSB PDB-biocurated (29.9% increase over 2019)
- Biocuration workload balanced geographically
    - 47%* Americas, Oceania
    - 31% Europe, Africa
    - 22% Asia

*Global Group Depositions included



Efficient Group Deposition/Biocuration

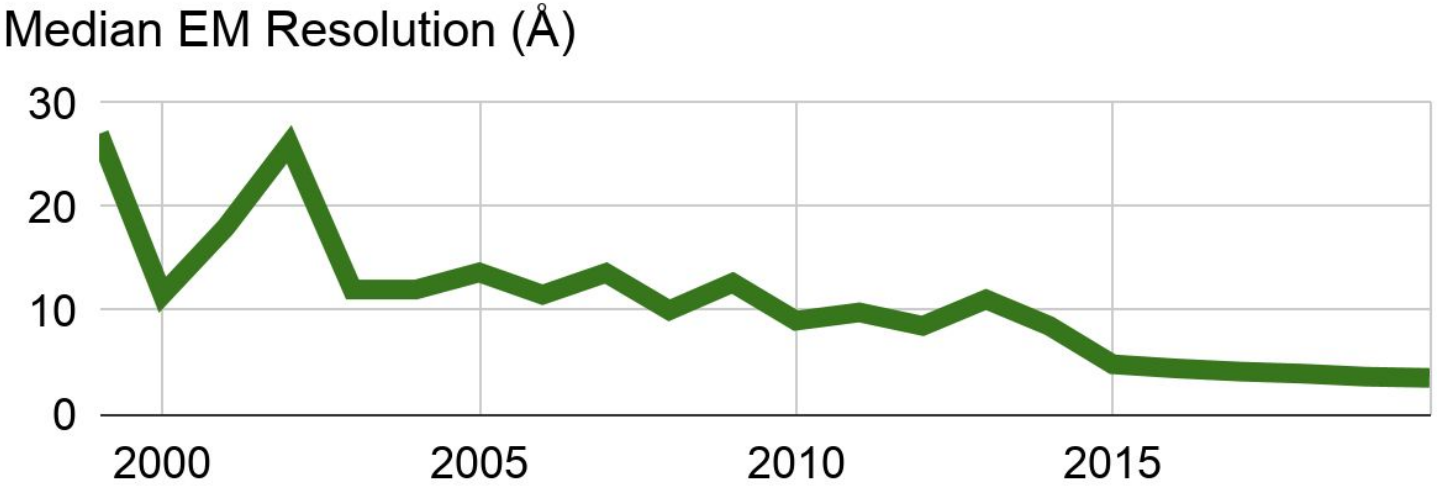
- Depositors can submit multiple, similar structures *via* GroupDep
- Total of 416 SARS-CoV-2 structures with 172 novel ligands from PanDDA fragment screening
- All processed and released within two weeks from date of deposition

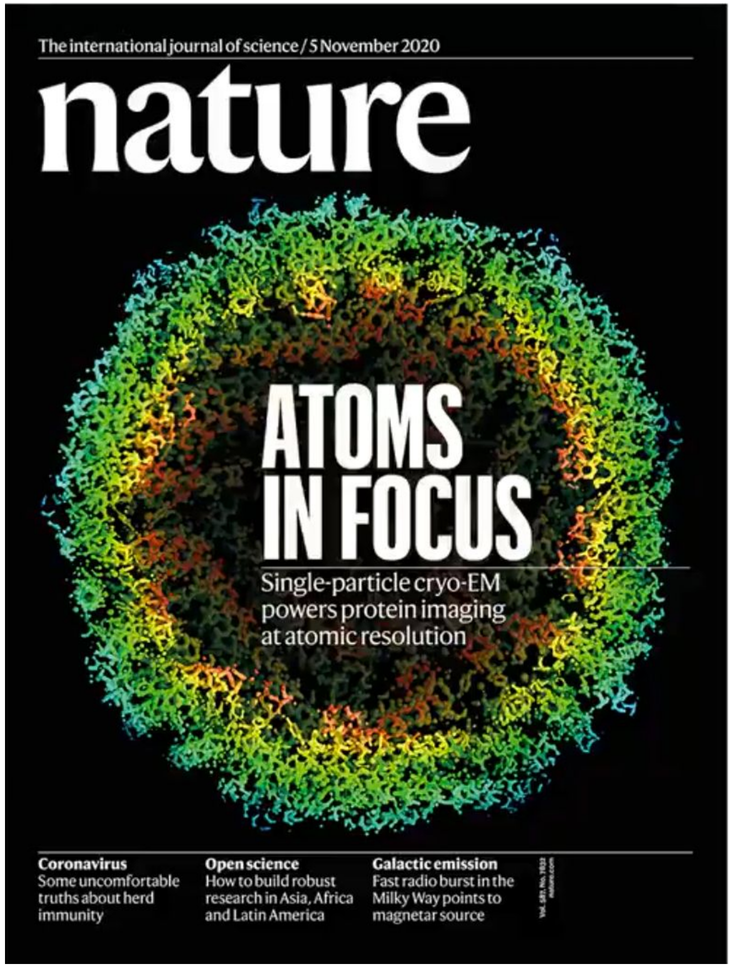


Nsp3 Macrodomain

3DEM: Fast Evolving

- Record 2,780 new 3DEM structures in 2020
  - 40% increase from 2019
- Map volume analysis part of validation
- Reaching very high resolution limits (~1Å)




The international journal of science / 5 November 2020

nature

ATOMS IN FOCUS

Single-particle cryo-EM powers protein imaging at atomic resolution

Coronavirus
Some uncomfortable truths about herd immunity

Open science
How to build robust research in Asia, Africa and Latin America

Galactic emission
Fast radio burst in the Milky Way points to magnetar source

Cathy Lawson serves as EMDB/EMDR Liaison on OneDep Project Team

Deposition/Biocuration in 2020

- Better automation for biocurating incoming depositions
  - Based on existing structure annotations
  - Median processing time/entry reduced for large and/or complex entries (from 3-4 days to <2 days)
- Streamlined NMR data deposition in NEF and NMR-STAR formats
- Full coordinate replacement enabled (131 entries versioned)
- Carbohydrate remediation project completed in July
- Enhanced Validation: carbohydrates, EM map volume analysis, NMR restraints assessments
- Wholesale validation report recalculation and 3DEM map analysis added to Archive

Deposition/Biocuration Goals for 2021

Goal	Impact/Gain
Provide validation reports in mmCIF format	Improve data quality Enable interoperability
Set up PDB China OneDep instance at PDBj ● Provide more automated OneDep software installation ● Document software packaging and maintenance ● Provide Biocuration training/quality assessment ● Configure OneDep software to include PDBc site ● Improve CCD code distribution	Increase wwPDB sustainability Enhance wwPDB collaboration Reduce Biocuration workload
Upgrade on Python 3 and Centos 8	Improve system maintenance
Biocuration Automation	Increase Biocuration efficiency
Assembly Annotation by Depositor	Increase Biocuration efficiency Improve data quality
Project planning and community communication on the extension of CCD code and PDB accession code	Better data management
Improve EM deposition	Better data validation

Background Information

Questions and Comments?

PDB Archive Status

John Westbrook

Archive Growth in 2020

Year-end holdings > 173K structures

- Record 14,044 new structures released (22% increase over 2019)
  - 8.8 % growth in the archive

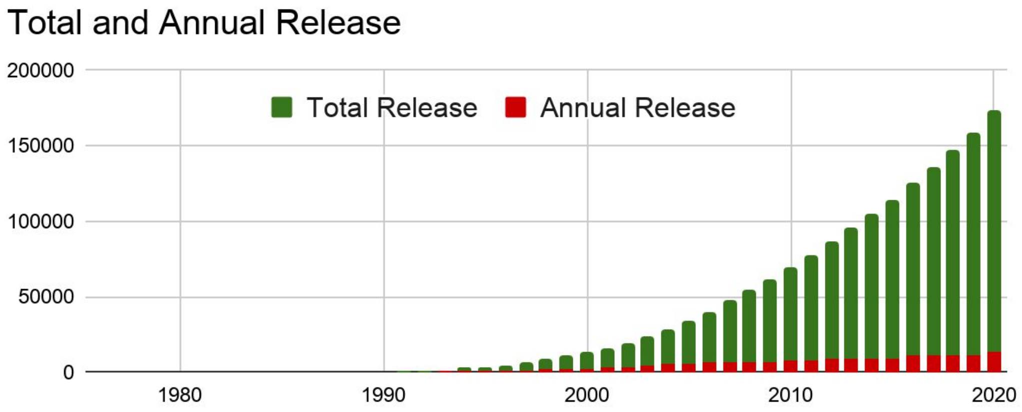
Core Archive Storage

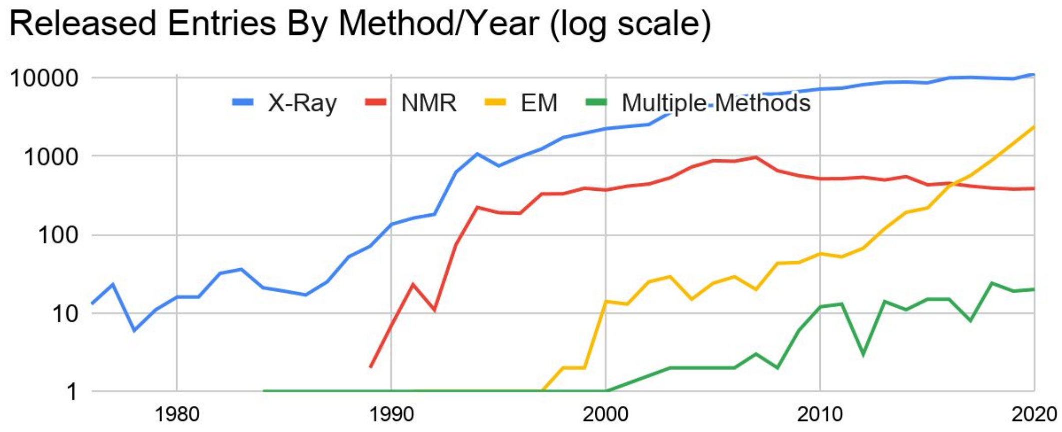
- OneDep Sessions: 26 TB
- ftp (legacy + versioned): 1.5TB
- EMDB ftp: 4 TB

Storage for Deployed Core Archive Services: 132 TB

- Redundant copies of all ftp servers, map and model services, images, and annual snapshots
- Based on total storage footprint, projected growth can be supported through 2022 for current RCSB.org services with current hardware

CoreTrustSeal certification renewed through April 2024 (CoreTrustSeal.org)


Total and Annual Release


Released Entries By Method/Year (log scale)

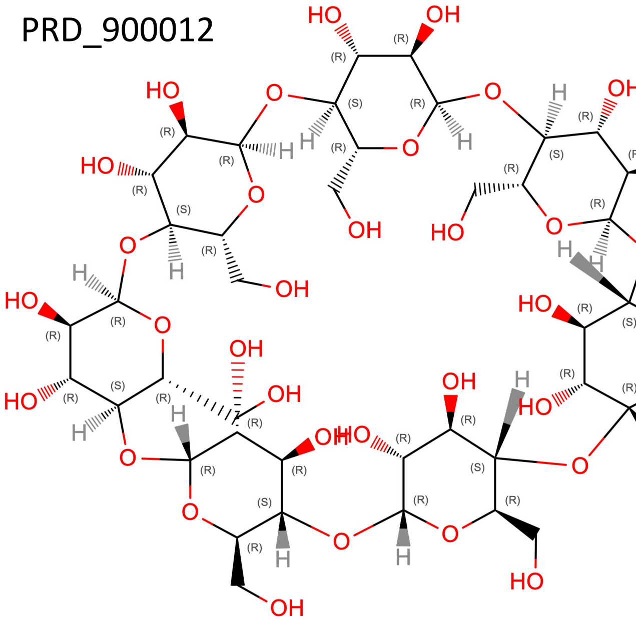
Chemical Reference Data

~34K Chemical Component Definitions
- 3118 new in 2020 (Up 29% over 2019)
- 6598 updated (Up 1.6x over 2019)

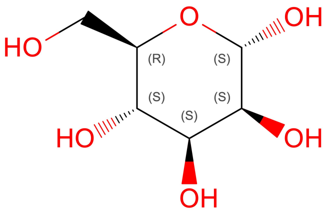
~1K Biologically Interesting molecule Reference Dictionary (BIRD) Definitions
- 174 new in 2020 (Up 9x over 2019)
- 130 updated (Up 60x over 2019)

Overlap with External Resources
- PubChem – 31603 (92% of CCD)
- ChEMBL – 12902 (38%)
- ChEBI – 4616 (14%)
- DrugBank – 6022 (18%)
- Pharos – 4277 (13%)

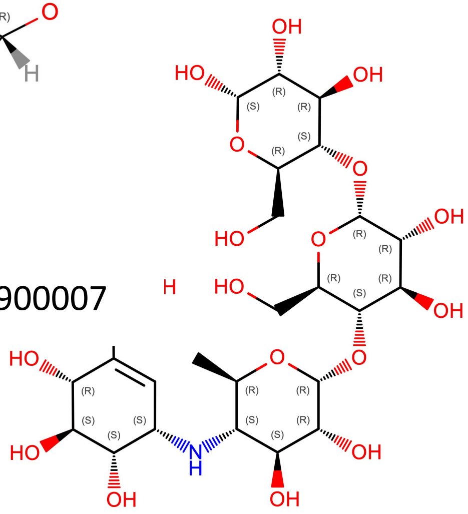
PRD_900012



MAN



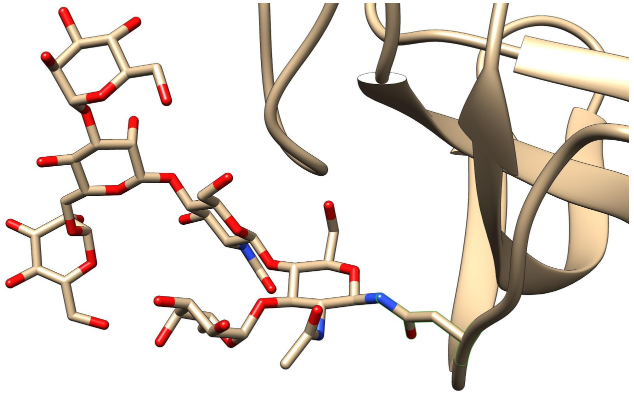
PRD_900007



July 2020: Carbohydrates Released

Scope

- 862 monosaccharides in the Chemical Component Dictionary
- 134 BIRD molecules
- ~10% of PDB structures that contain carbohydrates remediated
  - ~14,000 PDB structures
  - ~13,500 oligosaccharides



Process

Communicated with glycoscience, PDB, and software developer communities *via* conferences/virtual meetings

Standardized nomenclature following IUPAC/IUBMB

Adopted glycoscience community software

Provided uniform representation for oligosaccharides with community linear descriptor(s)

Annotated glycosylation with enumerated types

NIGMS grant U01 CA221216 with U Georgia (Jasmine Young served as Rutgers PI)

Additional Archive Projects In 2020

- NIH Common Fund Supplement (1 year)
  - Extends PDB data integration to NIH Common Fund data resources
    - Pharos: Illuminating the Druggable Genome (IDG; 49.8K PDB entries)
    - Genotype-Tissue Expression (GTEx) (48.7K PDB entries)
    - International Mouse Phenotyping Consortium (IMPC) (2.3K PDB entries)
- DOE KnowledgeBase (KBase) PDB Integration Project (2 years)
  - Reciprocal data exchange APIs with the KBase platform
- Joint NSF/BBSRC Project with PDBe (3 years)
  - Developing APIs for automated data deposition
  - Next generation archive with enriched/updated derived and integrated data products

Questions and Comments?

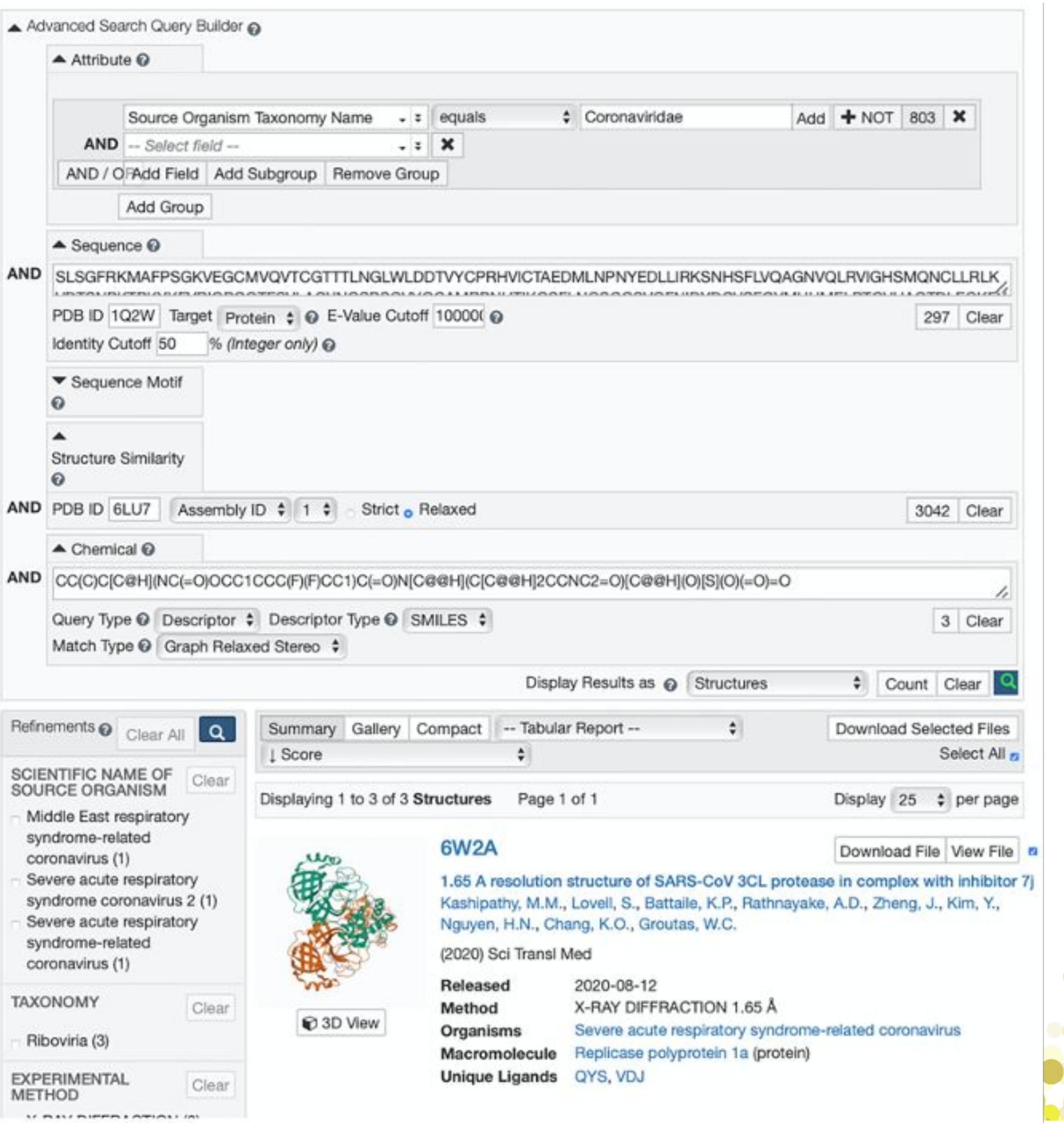
New and Improved RCSB.org

Jose Duarte and Yana Rose

RCSB.org visualization on NOVA

New RCSB.org Released April 2020

- Combined structure, sequence, chemistry, and annotation searches
- Search results now shown at Entry, Assembly, and Entity levels
- Left-hand menu for quick refinement of search results
- Improved tabular report
- Bookmarkable searches
- Improved MyPDB access to stored searches
- New APIs support data access and searching



Source Organism Taxonomy Name equals Coronaviridae 803

AND -- Select field --

AND SLSGFRKMAFPSGKVEGCMVQVTCGTTTLNGLWLDDTVYCPRHVICTAEDMLNPNYEDLLIRKSNHSFLVQAGNVQLRVIGHSMQNCLLRLK

PDB ID 1Q2W Target Protein E-Value Cutoff 10000 297

Identity Cutoff 50 % *(Integer only)*

AND PDB ID 6LU7 Assembly ID 1 Strict ● Relaxed 3042

AND CC(C)C[C@H](NC(=O)OCC1CCC(F)(F)CC1)C(=O)N[C@@H](C[C@@H]2CCNC2=O)[C@@H](O)[S](O)(=O)=O

Query Type Descriptor Descriptor Type SMILES 3

Match Type Graph Relaxed Stereo

Display Results as Structures

↓ Score

SCIENTIFIC NAME OF SOURCE ORGANISM
- Middle East respiratory syndrome-related coronavirus (1)
- Severe acute respiratory syndrome coronavirus 2 (1)
- Severe acute respiratory syndrome-related coronavirus (1)

TAXONOMY
- Riboviria (3)

EXPERIMENTAL METHOD

Displaying 1 to 3 of 3 **Structures** Page 1 of 1 Display 25 per page

6W2A

1.65 A resolution structure of SARS-CoV 3CL protease in complex with inhibitor 7j

Kashipathy, M.M., Lovell, S., Battaile, K.P., Rathnayake, A.D., Zheng, J., Kim, Y., Nguyen, H.N., Chang, K.O., Groutas, W.C.

(2020) Sci Transl Med

Released	2020-08-12
Method	X-RAY DIFFRACTION 1.65 Å
Organisms	Severe acute respiratory syndrome-related coronavirus
Macromolecule	Replicase polyprotein 1a (protein)
Unique Ligands	QYS, VDJ

New Features May-Dec 2020

- New images (Mol*-generated) for entries, polymer entities, and assemblies
- New Protein Feature View
- Sequence Alignments in Search Results
- Access and visualization of updated carbohydrate data
- Genome View: graphical summary of PDB entity sequences and reference genomes
- Documentation for all new features

Streamlined Delivery

- Legacy APIs retired

6VXX

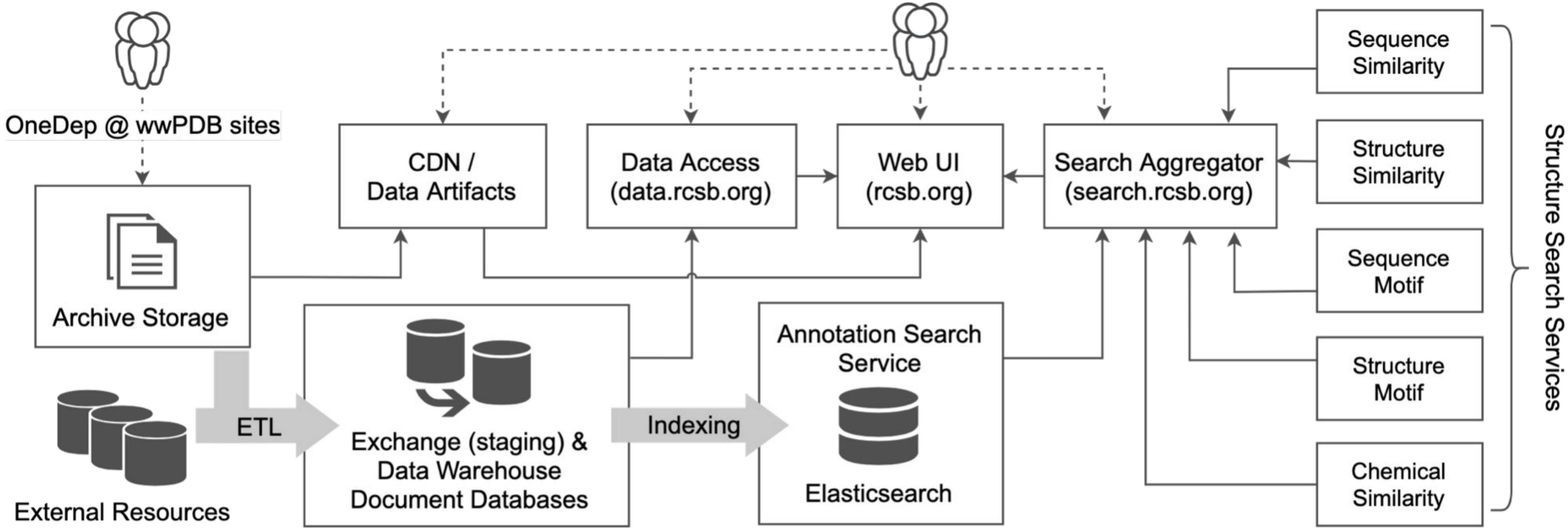
Structure of the SARS-CoV-2 spike glycoprotein (closed state)

(Mouse scroll to zoom / Hold left click to move)

6VXX.A

SARS-CoV-2 spike glycoprotein - Severe acute respiratory syndrome coronavirus 2

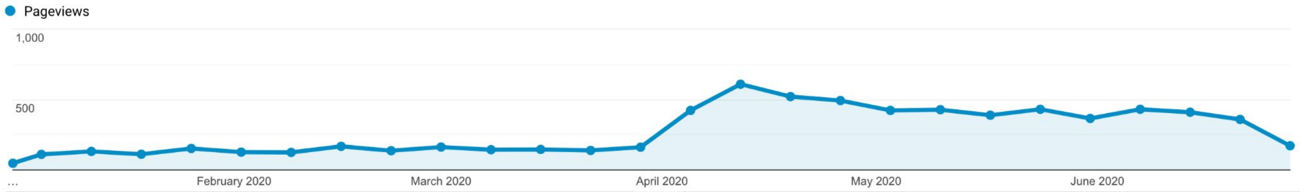
Search Architecture Driving RCSB.org



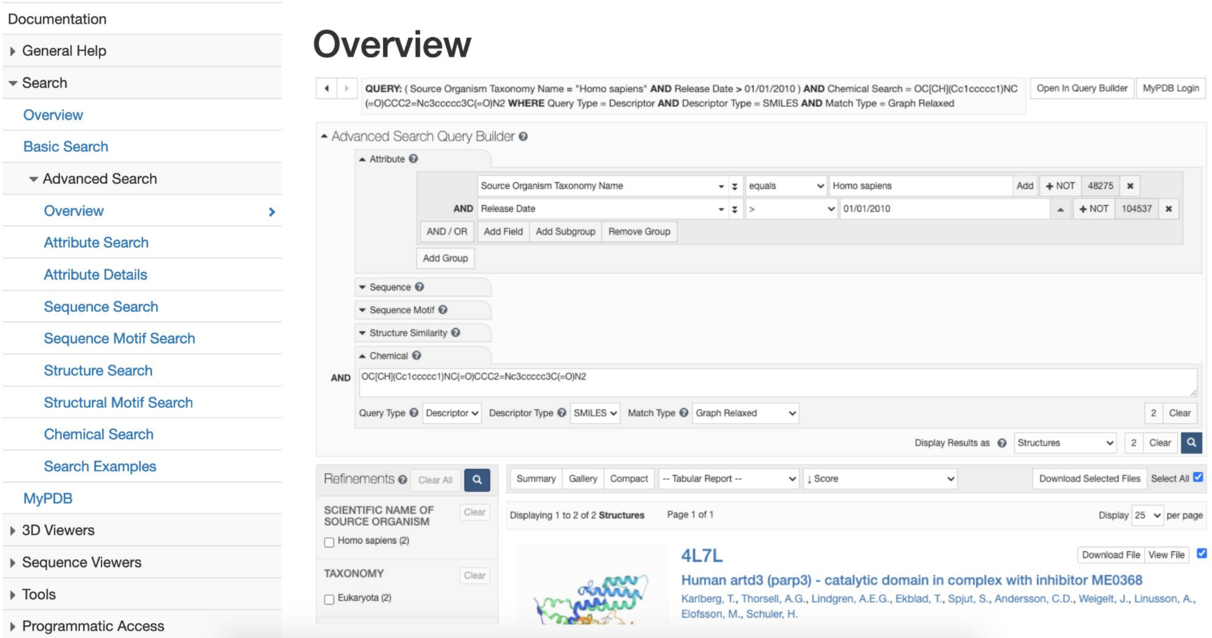
Background Information; *RCSB Protein Data Bank: Architectural Advances Towards Integrated Searching and Efficient Access to Macromolecular Structure Data from the PDB Archive,* https://doi.org/10.1016/j.jmb.2020.11.003

Monitoring Customer Service

- Customer Service enquiries increased after initial release of new features
  - 2020 Q2: 643 total questions/comments
  - 2019 Q2: 288 total
- Users
  - Provided positive feedback
  - Wanted more guidance on usage
  - Identified bugs and requested new features
- Feedback returned to baseline Fall 2020
- Detailed documentation project initiated



Q1-Q2 Visits to https://www.rcsb.org/pages/contactus



Overview

QUERY: (Source Organism Taxonomy Name = "Homo sapiens" **AND** Release Date > 01/01/2010) **AND** Chemical Search = OC[CH](Cc1ccccc1)NC(=O)CCC2=Nc3ccccc3C(=O)N2 **WHERE** Query Type = Descriptor **AND** Descriptor Type = SMILES **AND** Match Type = Graph Relaxed

Source Organism Taxonomy Name | equals | Homo sapiens | 48275

AND Release Date | > | 01/01/2010 | 104537

AND OC[CH](Cc1ccccc1)NC(=O)CCC2=Nc3ccccc3C(=O)N2

Query Type: Descriptor; Descriptor Type: SMILES; Match Type: Graph Relaxed; 2

Display Results as: Structures; 2

SCIENTIFIC NAME OF SOURCE ORGANISM

☐ Homo sapiens (2)

TAXONOMY

☐ Eukaryota (2)

Displaying 1 to 2 of 2 **Structures** Page 1 of 1 Display 25 per page

4L7L

Human artd3 (parp3) - catalytic domain in complex with inhibitor ME0368

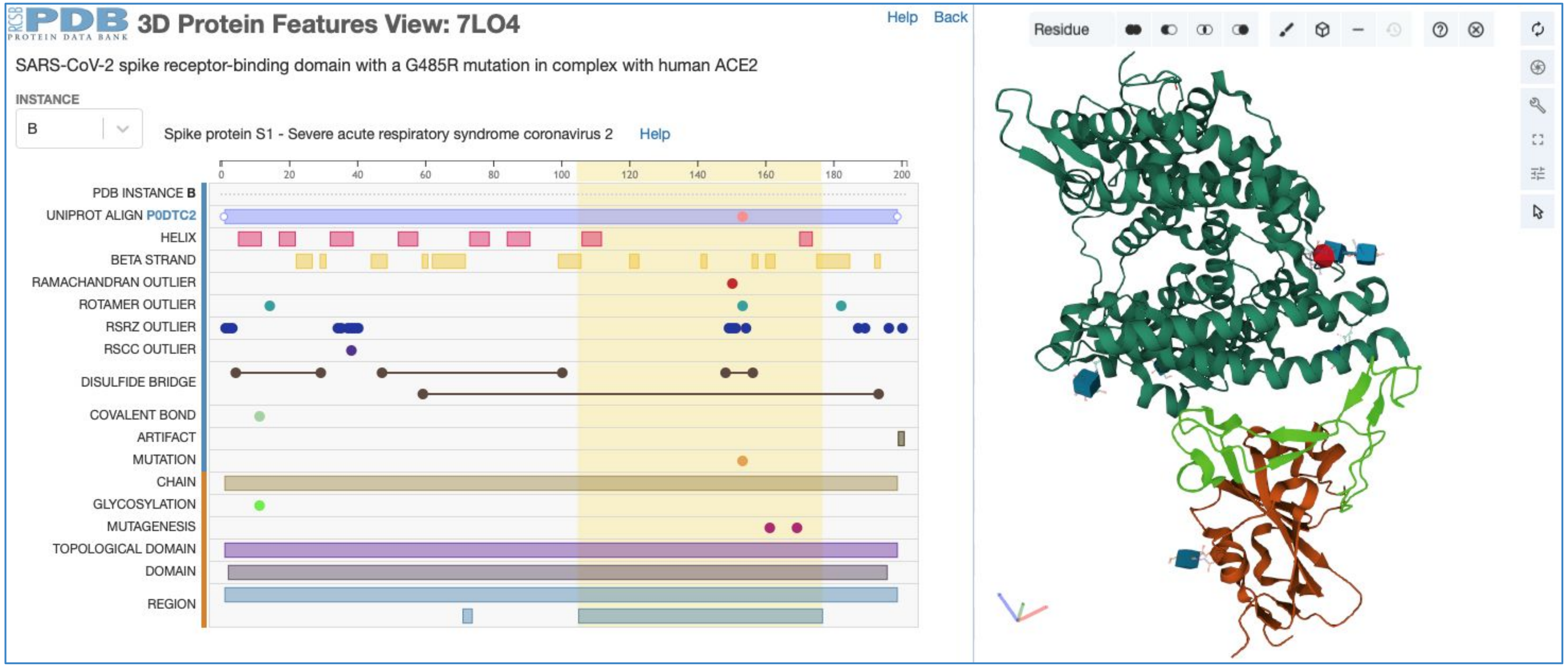
Karlberg, T., Thorsell, A.G., Lindgren, A.E.G., Ekblad, T., Spjut, S., Andersson, C.D., Weigelt, J., Linusson, A., Elofsson, M., Schuler, H.

New RCSB.org documentation (released early 2021)

Visualizing Protein Features in 3D

3D Protein Feature View maps protein sequence features (annotations from RCSB PDB and external resources) on 3D proteins structures

- Navigate between 1D sequence features and 3D structure
- Select regions of interest
- Highlight interactions of specific residues
- Explore an environment of an ion/ligand



3D Protein Features View: 7LO4

SARS-CoV-2 spike receptor-binding domain with a G485R mutation in complex with human ACE2

INSTANCE: B

Spike protein S1 - Severe acute respiratory syndrome coronavirus 2

https://www.rcsb.org/3d-sequence/7LO4

Questions and Comments?

PDB Data Delivery

Rob Lowe

RCSB PDB Usage in 2020

RCSB.org

- ~5.4 million unique users (~4 m in 2019)
    - 11 million sessions
    - 44 million page views
    - ~79 TB of data accessed

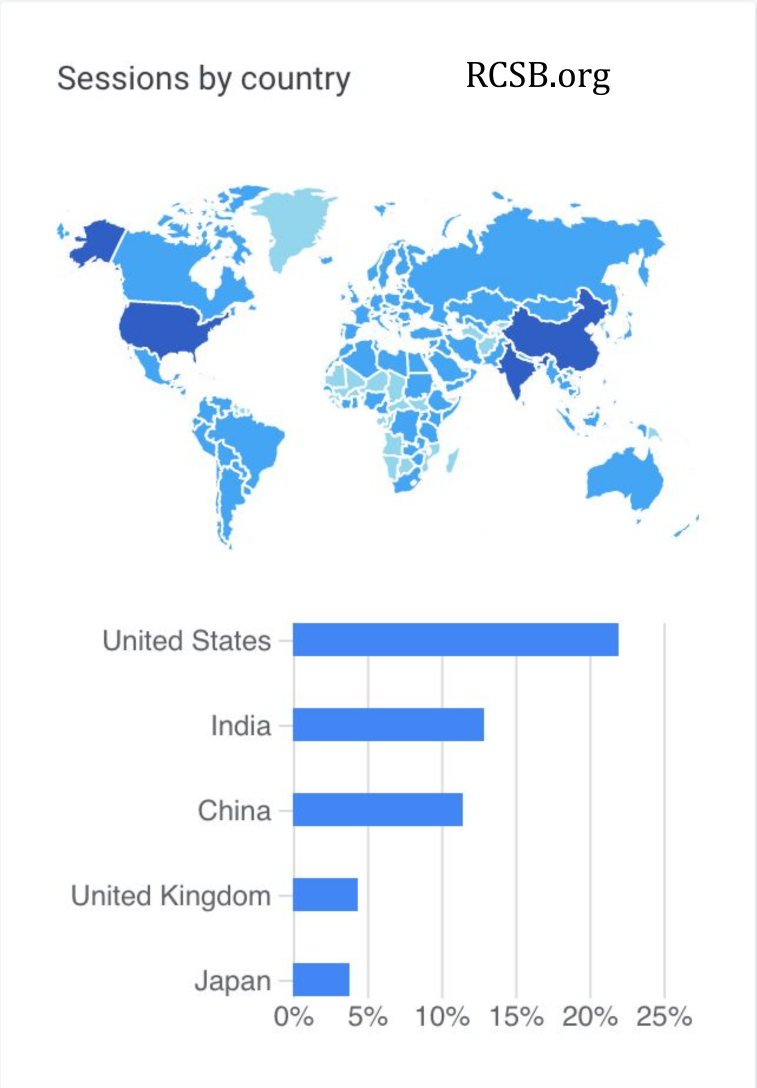
Data File Downloads

- >575 million times from RCSB PDB-hosted FTP and websites (~547 million in 2019)
- Additional downloads from PDBe and PDBj

External Utilization

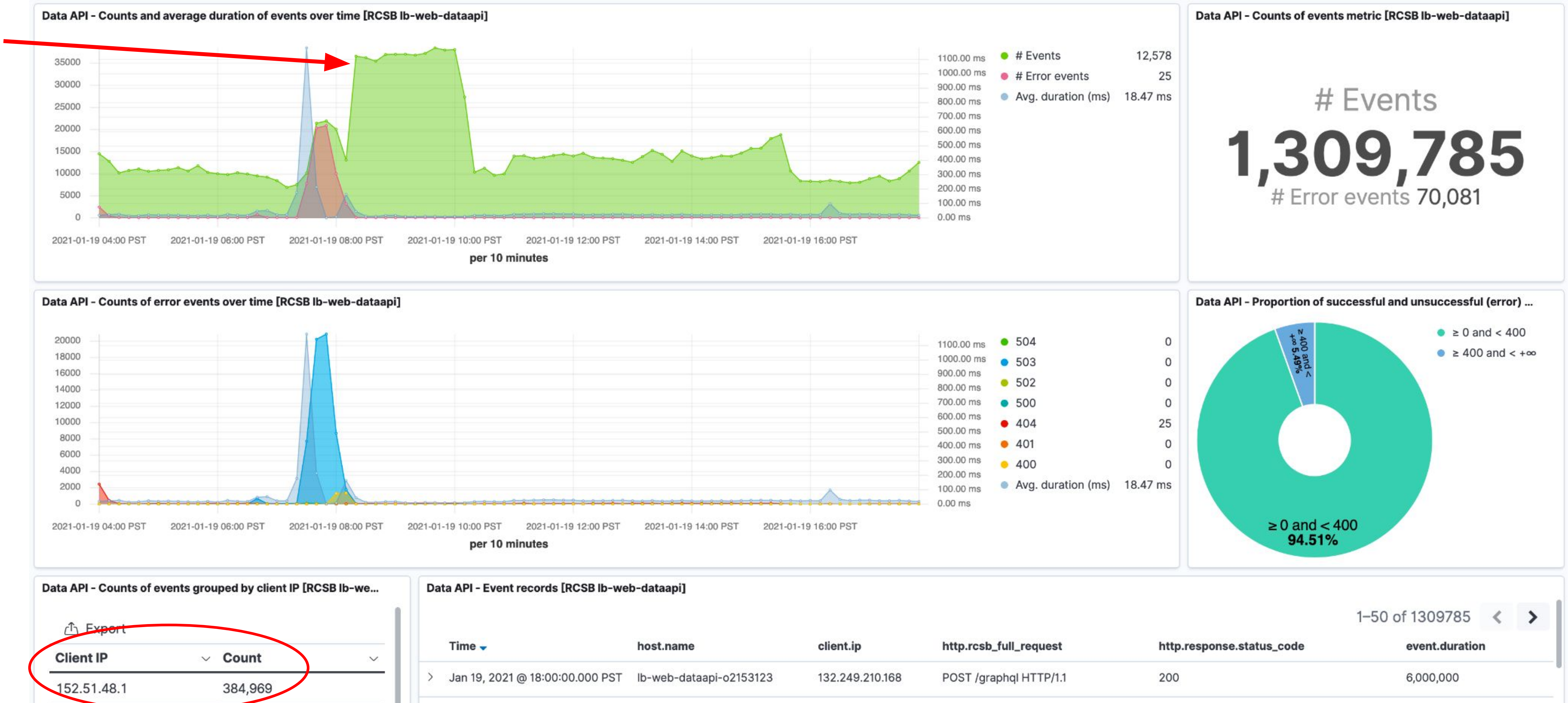
- >430 external resources repackage and distribute data
- Biopharm/biotech companies use PDB data internally

Tracked using Google Analytics, AWstats, internal logging services



Website Traffic Monitoring

Jan 19, 2021: Data API queries spike produced by a single scripting user from GSK in Pennsylvania



Data API - Counts and average duration of events over time [RCSB lb-web-dataapi]

Data API - Counts of events metric [RCSB lb-web-dataapi]

\# Events

1,309,785

\# Error events 70,081

Data API - Counts of error events over time [RCSB lb-web-dataapi]

Data API - Proportion of successful and unsuccessful (error) ...

Data API - Counts of events grouped by client IP [RCSB lb-we...

Client IP	Count
152.51.48.1	384,969

Data API - Event records [RCSB lb-web-dataapi]

	Time	host.name	client.ip	http.rcsb_full_request	http.response.status_code	event.duration
›	Jan 19, 2021 @ 18:00:00.000 PST	lb-web-dataapi-o2153123	132.249.210.168	POST /graphql HTTP/1.1	200	6,000,000

Questions and Comments?

Outreach/Education

Christine Zardecki

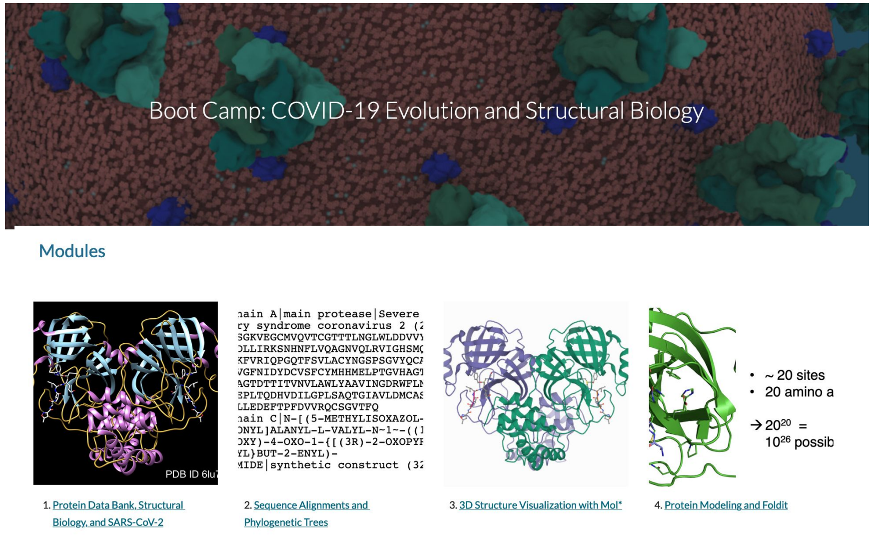
2020 Focus: COVID-19


The New York Times

CALIFORNIA TODAY

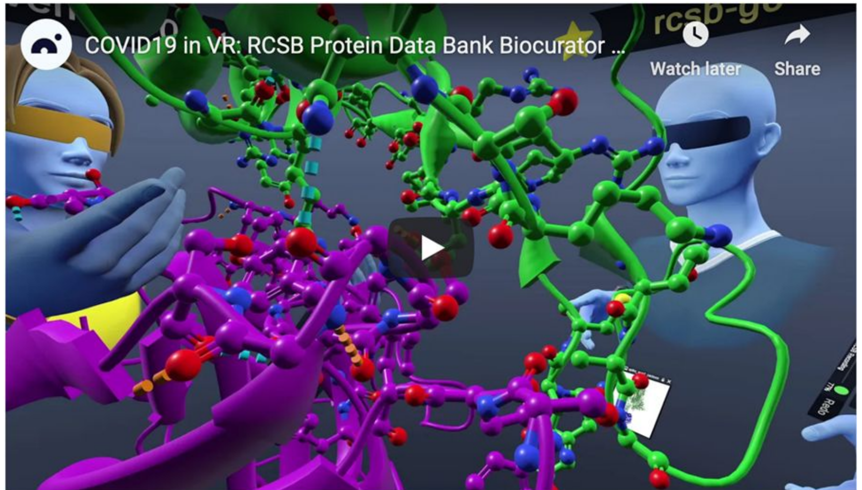
Why Are California and New York Different in the Virus Crisis?

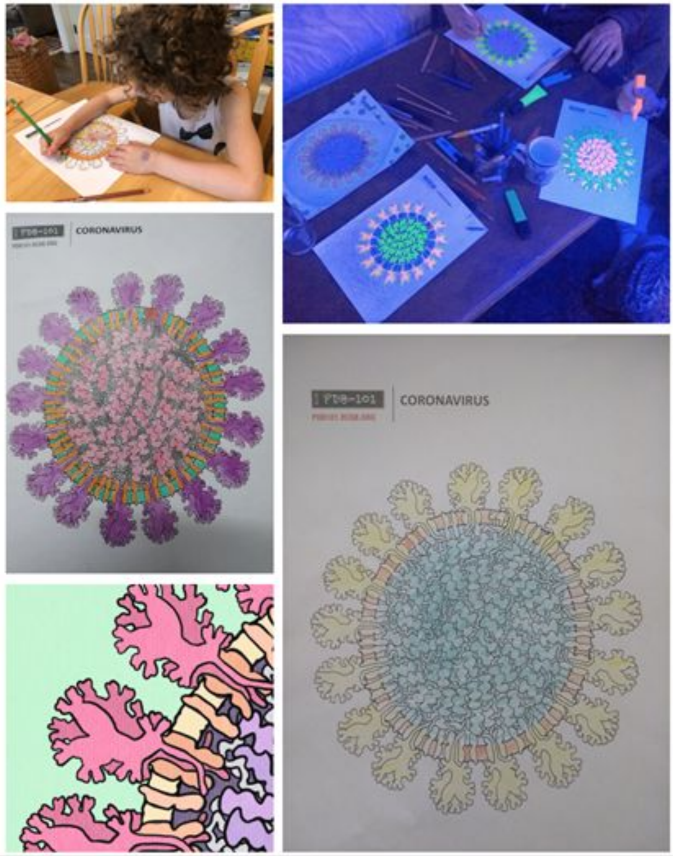
Monday: It's too early to say for sure, but some experts say early action by California leaders has helped. Also: Art during the pandemic.


Boot Camp: COVID-19 Evolution and Structural Biology

Modules

1. Protein Data Bank, Structural Biology, and SARS-CoV-2
2. Sequence Alignments and Phylogenetic Trees
3. 3D Structure Visualization with Mol*
4. Protein Modeling and Foldit


COVID19 in VR: RCSB Protein Data Bank Biocurator ...



PDB-101 Supported More Users

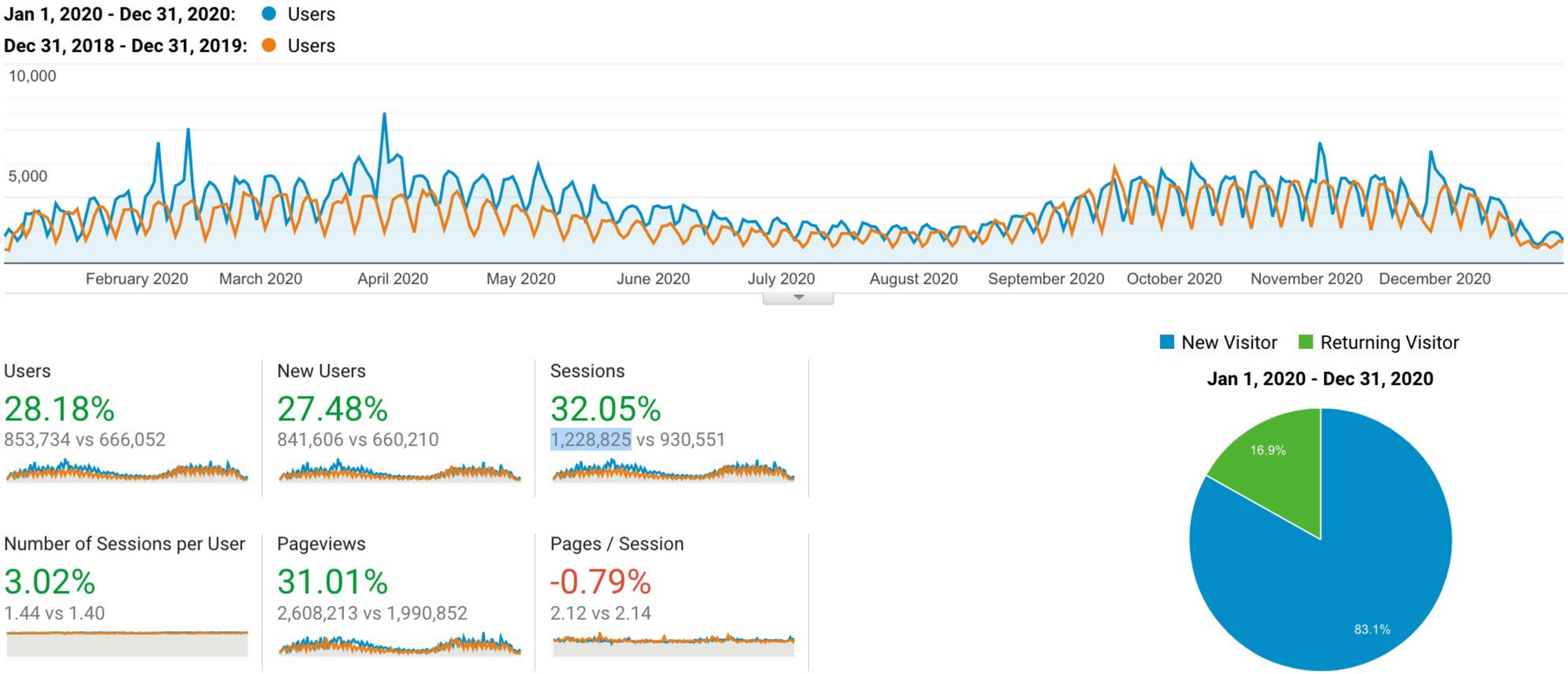
Users

2020: 853,734

2019: 665,958

2018: 594,073

2017: 620,784



Jan 1, 2020 - Dec 31, 2020: ● Users

Dec 31, 2018 - Dec 31, 2019: ● Users

Users
28.18%
853,734 vs 666,052

New Users
27.48%
841,606 vs 660,210

Sessions
32.05%
1,228,825 vs 930,551

Number of Sessions per User
3.02%
1.44 vs 1.40

Pageviews
31.01%
2,608,213 vs 1,990,852

Pages / Session
-0.79%
2.12 vs 2.14

🟦 New Visitor 🟩 Returning Visitor

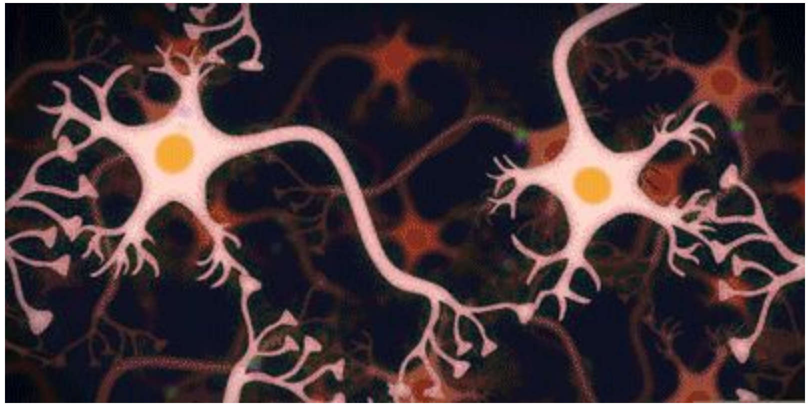
Jan 1, 2020 - Dec 31, 2020

PDB-101 in 2020: Beyond Covid

- 2021 calendar (with wwPDB partners)
- 9 Molecule of the Month articles
- Molecular Origami paper models: Green and Red Fluorescent Proteins
- Understanding Integrative Modeling
- Video: Neuronal Signaling and Sodium-Potassium Pump
- Virtual Meeting Backgrounds
- Health Focus: Drugs & Brain
  - HS Video Challenge
  - Undergraduate MAP course

2022-23 Health Focus: *Cancer*

PDF template for dsRED in color. PDF template for GFP in color.



Full video at PDB-101. >16K views on YouTube.

Inaugural PDB50 Event (Virtual)

PDB50

A special symposium celebrating the 50th anniversary of the Protein Data Bank

May 4–5, 2021

PROTEIN DATA BANK

Years

March 22:
Deadline for abstracts and early registration

asbmb.org/meetings-events/pdb50

May 4 | 11 a.m. – 4:30 p.m. ET

Helen M. Berman, *Rutgers, The State University of New Jersey & Univ. of Southern California*
The evolution of the Protein Data Bank as a community resource

Zihe Rao, *ShanghaiTech University and Tsinghua University*
Structural insight into SARS-CoV-2 replication and transcription complex (RTC)

Stephen K. Burley, *Rutgers, The State University of New Jersey & UC San Diego*
Impact of structural biologists and fifty years of Protein Data Bank operations on drug discovery and development

Alexandre M. J. J. Bonvin, *Utrecht University*
Solving 3D puzzles by integrative modelling using PDB

Eddy Arnold, *Rutgers, The State University of New Jersey*
Using HIV-1 reverse transcriptase structures to guide anti-AIDS drug discovery

Hao Wu, *Harvard Medical School, Boston Children's Hospital*
"Speck"tacular inflammasomes: structures of supramolecular complexes in innate immunity

Jennifer L. Martin, *University of Wollongong*
Science, crystallography, reflections: A journey with the PDB over 35 years

May 5 | 11 a.m. – 4:30 p.m. ET

Johann Deisenhofer, *University of Texas Southwestern Medical Center*
50 years of PDB — from crazy idea to treasure

Juli Feigon, *University of California, Los Angeles*
Structural biology of telomerase

Thomas L. Blundell, *University of Cambridge*
A personal history of five decades of structural biology and the PDB: From the X-ray structure of 2-Zinc insulin hexamer in 1970 to Cryo-EM structures of DNA-PK from DNA repair in 2020

Stephen L. Mayo, *California Institute of Technology*
Antibody small molecule conjugates with computationally designed target binding synergy

Wah Chiu, *Stanford University*
Cryo-EM of biomolecules at Ångström resolutions

Angela Gronenborn, *University of Pittsburgh*
Integrated BioNMR — getting by with a little help from my friends

WORLDWIDE PDB PROTEIN DATA BANK
Organized by the wwPDB Foundation

ASBMB
Hosted by the American Society for Biochemistry and Molecular Biology

Registration so far

- >800 registered
- 275 posters
  - ~175 students and postdocs
  - 16 wwPDB
- Let Christine know if you want to register as our guest

wwPDB Foundation Industrial Sponsors (to date)

- Platinum: ThermoFisher, Novartis
- Silver: Astex Pharmaceuticals, Bristol Myers Squibb, Cambridge Isotope, Dectris, Genentech, Discngine, National Biomedical Research Foundation, OpenEye, Silicon Therapeutics
- Bronze: Anton Paar, Cell Press, Rigaku

Other PDB50 Celebrations

- ACA Transactions Symposium 2021: Function Follows Form: Celebrating the 50th Anniversary of the Protein Data Bank (Jul 30-31, virtual)
- ACS Understanding Enzyme Function in 3D (Aug ~25, Virtual)
- Bringing Molecular Structure to Life: 50 Years of the PDB
  (Oct 20-22, virtual EMBL Conference)
- Celebrating Protein Data Bank Contributions from Rutgers University Structural Biologists (Nov 2, virtual)
- Biophysical Society of Japan
  (Nov 24, TBD)

Collection of JBC Reviews shows synergism between the PDB and explosive progress witnessed in many scientific areas.

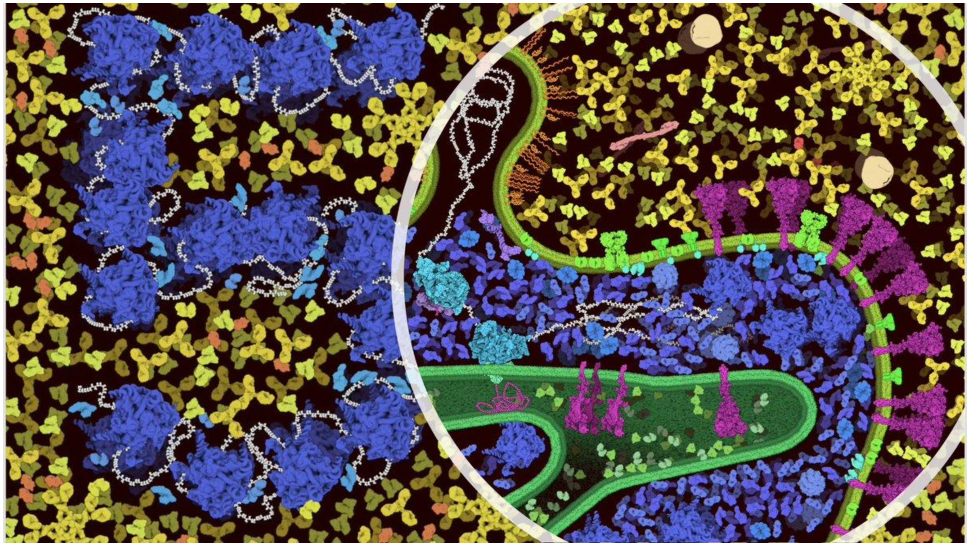
Coordinating Editors: Lila Gierasch and Helen Berman (guest)

Highlights a few of the seminal structures published in JBC since 1971, including classics across a variety of fields

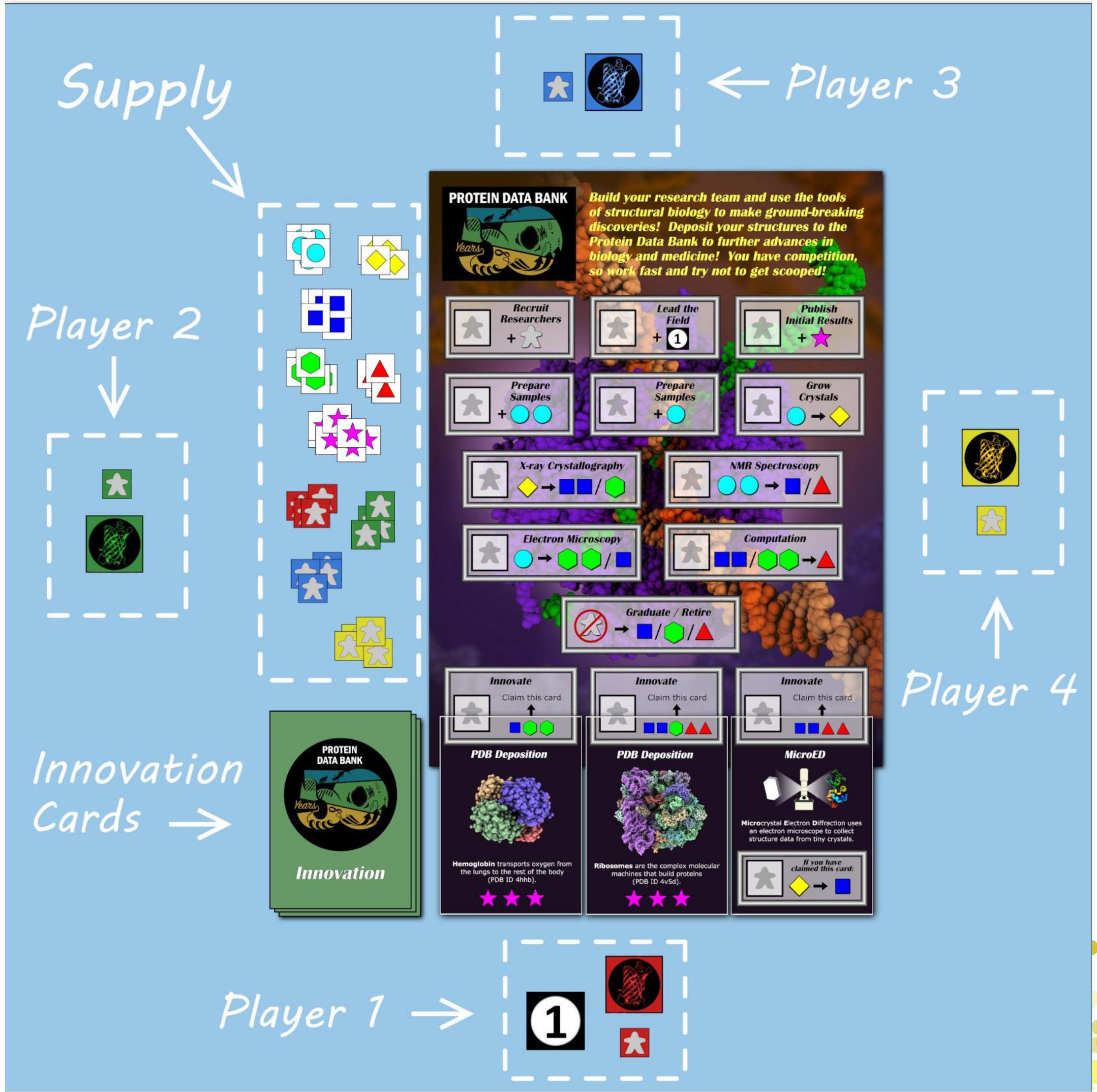
Assembled by Lila Gierasch and Helen Berman

Select Works in Progress

- October Molecule of the Month
- [CellPAINT image contest]



[Printable game]



Build your research team and use the tools of structural biology to make ground-breaking discoveries! Deposit your structures to the Protein Data Bank to further advances in biology and medicine! You have competition, so work fast and try not to get scooped!

Hemoglobin transports oxygen from the lungs to the rest of the body (PDB ID 4hhb).

Ribosomes are the complex molecular machines that build proteins (PDB ID 4v5d).

Microcrystal **E**lectron **D**iffraction uses an electron microscope to collect structure data from tiny crystals.

Questions and Comments?

Operations and Funding

Stephen K. Burley

Team Transitions

Rutgers

- New Hires: Maryam Fayazi (Developer), Justin Flatt (Biocurator), Harmeet Kaur (Developer), Dennis Piehl (Developer),
Shamara Whetstone (Admin)
- Departure: Cindy Kleiman (Admin, to retirement)

UCSD

- New Hire: Jeremy Henry (DevOps)
- Promoted: Sebastian Bittrich (Developer)
- Departure: Alex Rose (Developer, to Industry)

Dr. Sebastian Bittrich

Scientific Software Developer

Maryam Fayazi

Scientific Application/Web Developer

Dr. Justin Flatt

Biocurator

Jeremy Henry

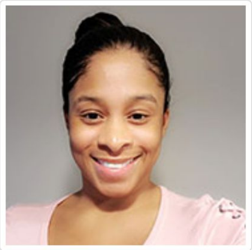
DevOps Lead

Dr. Harmeet Kaur

Scientific Application/Web Developer

Dr. Dennis Piehl

Scientific Application/Web Developer



Shamara Whetstone

Administrative Coordinator/Executive Assistant to Dr. Burley

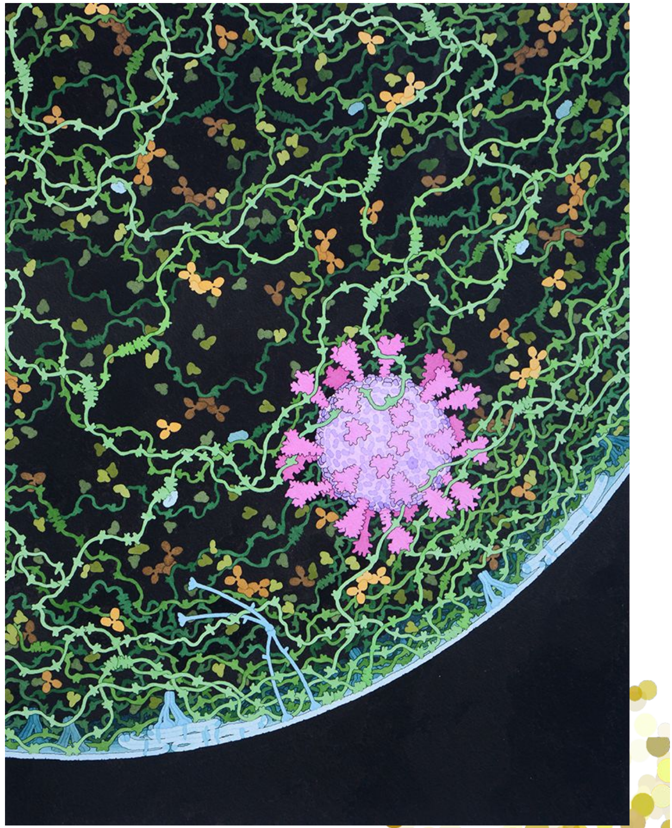
Recruitment in 2020

- Biophysical Society Job Fair
- Professional societies: American Crystallographic Association (ACA), American Society for Biochemistry and Molecular Biology (ASBMB), BioXFEL, International Society for Computational Biology (ISCB), Nature, Society for Advancement of Chicanos/Hispanics and Native Americans in Science (SACNAS)
- Informal postings: Black Women in Computational Biology Network, 500 Queer Scientists, Facebook, Twitter, LinkedIn, CCP4, pdbl

CAREER OPPORTUNITIES for SCIENTIFIC SOFTWARE DEVELOPERS

PROTEIN DATA BANK 50 Years

Join the multidisciplinary RCSB Protein Data Bank Team at Rutgers, The State University of New Jersey and the University of California San Diego



Develop innovative analysis, integration, query, and visualization tools for 3D biomolecular structures at **RCSB.org** to help accelerate research and training in biology, medicine, and related disciplines.

Design, develop, and deploy modern web and data applications and complex interactive graphical user interfaces.

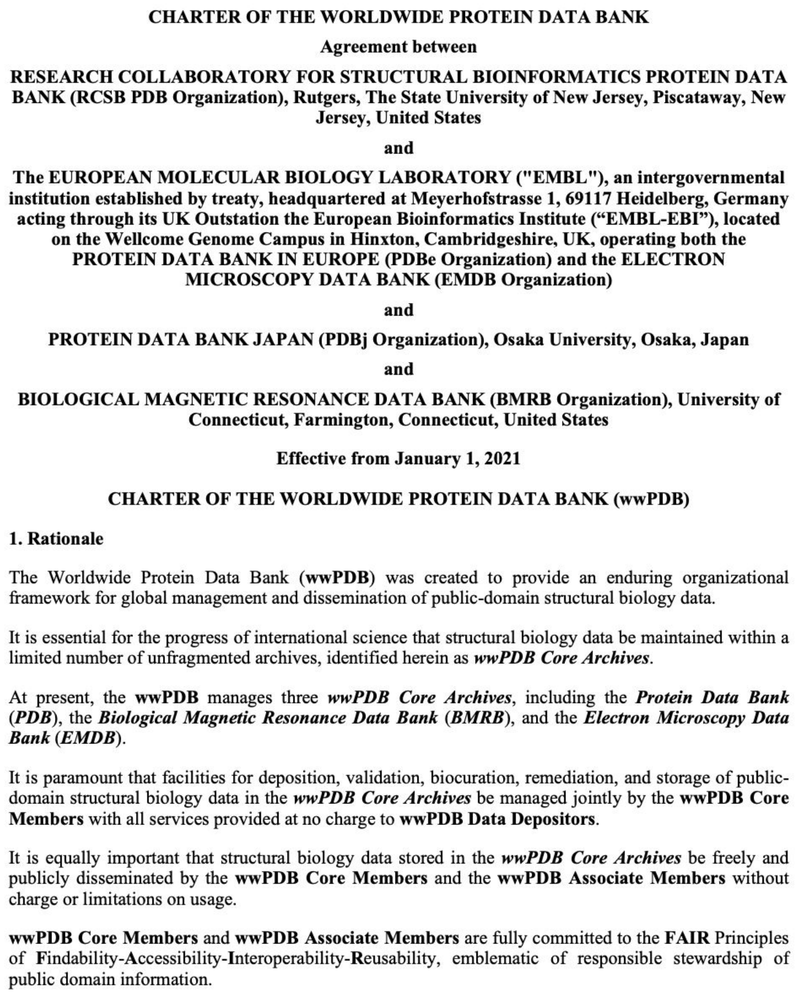
Open Positions: Scientific Software Developer (Rutgers)
Scientific Software Developer (UCSD)
Postdoctoral Researcher (UCSD)

Details: **www.rcsb.org/pages/jobs**

RUTGERS UC San Diego SDSC SAN DIEGO SUPERCOMPUTER CENTER UCSF University of California San Francisco

New wwPDB Charter

- Effective Date January 1 2021
- Admitted Electron Microscopy Data Bank (EM Map Archive)
- Framework for admission of PDB China and PDB India as Associate wwPDB Members (with roadmap for transition to Full Membership)
- Adopted Creative Commons CC0 License



CHARTER OF THE WORLDWIDE PROTEIN DATA BANK

Agreement between

RESEARCH COLLABORATORY FOR STRUCTURAL BIOINFORMATICS PROTEIN DATA BANK (RCSB PDB Organization), Rutgers, The State University of New Jersey, Piscataway, New Jersey, United States

and

The EUROPEAN MOLECULAR BIOLOGY LABORATORY ("EMBL"), an intergovernmental institution established by treaty, headquartered at Meyerhofstrasse 1, 69117 Heidelberg, Germany acting through its UK Outstation the European Bioinformatics Institute ("EMBL-EBI"), located on the Wellcome Genome Campus in Hinxton, Cambridgeshire, UK, operating both the PROTEIN DATA BANK IN EUROPE (PDBe Organization) and the ELECTRON MICROSCOPY DATA BANK (EMDB Organization)

and

PROTEIN DATA BANK JAPAN (PDBj Organization), Osaka University, Osaka, Japan

and

BIOLOGICAL MAGNETIC RESONANCE DATA BANK (BMRB Organization), University of Connecticut, Farmington, Connecticut, United States

Effective from January 1, 2021

CHARTER OF THE WORLDWIDE PROTEIN DATA BANK (wwPDB)

1. Rationale

The Worldwide Protein Data Bank (**wwPDB**) was created to provide an enduring organizational framework for global management and dissemination of public-domain structural biology data.

It is essential for the progress of international science that structural biology data be maintained within a limited number of unfragmented archives, identified herein as ***wwPDB Core Archives***.

At present, the **wwPDB** manages three ***wwPDB Core Archives***, including the ***Protein Data Bank*** (***PDB***), the ***Biological Magnetic Resonance Data Bank*** (***BMRB***), and the ***Electron Microscopy Data Bank*** (***EMDB***).

It is paramount that facilities for deposition, validation, biocuration, remediation, and storage of public-domain structural biology data in the ***wwPDB Core Archives*** be managed jointly by the **wwPDB Core Members** with all services provided at no charge to **wwPDB Data Depositors**.

It is equally important that structural biology data stored in the ***wwPDB Core Archives*** be freely and publicly disseminated by the **wwPDB Core Members** and the **wwPDB Associate Members** without charge or limitations on usage.

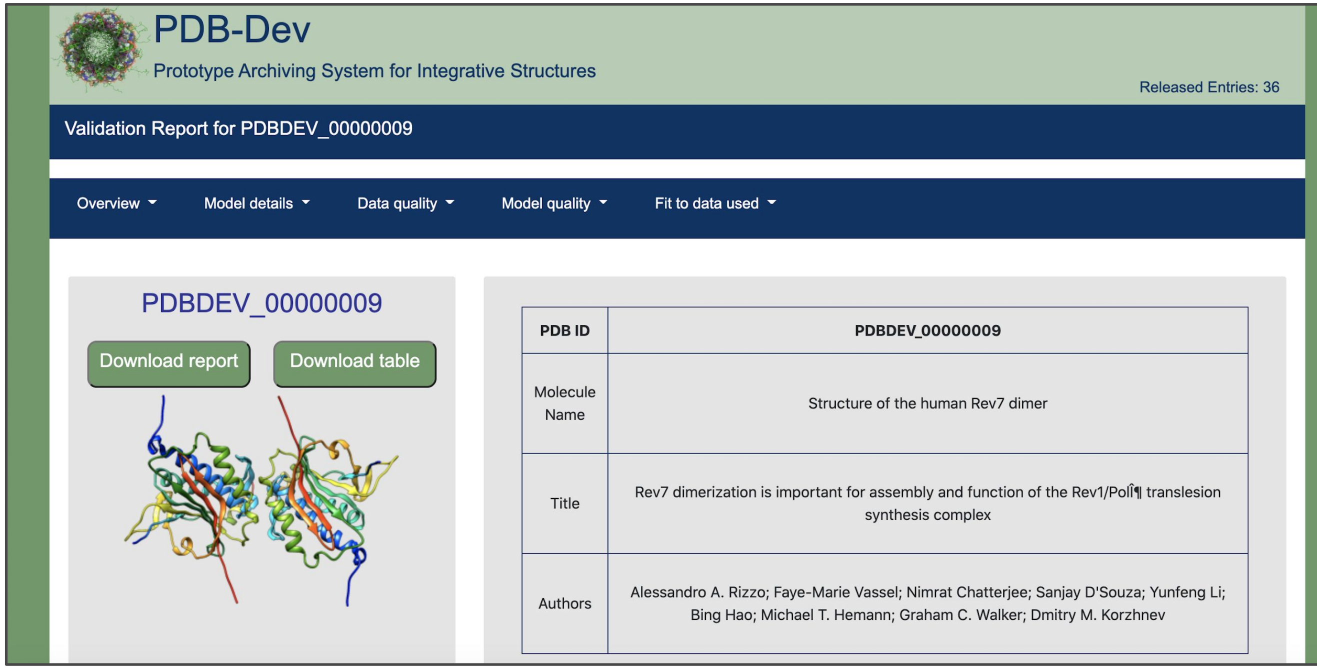
wwPDB Core Members and **wwPDB Associate Members** are fully committed to the **FAIR** Principles of **F**indability-**A**ccessibility-**I**nteroperability-**R**eusability, emblematic of responsible stewardship of public domain information.

Strategic Initiatives

Stephen K. Burley
Andrej Sali

PDB-Dev: Validation of Models

- I/HM validation report will be available for all ~80 entries in PDB-Dev by August 2021
- First version of the report will include model quality assessment and SAS validation (HTML and PDF)
- Summary table will be available for download
- Well-tested and code-reviewed IHM-validation package available on GitHub
- Next: Validation based on CX-MS data

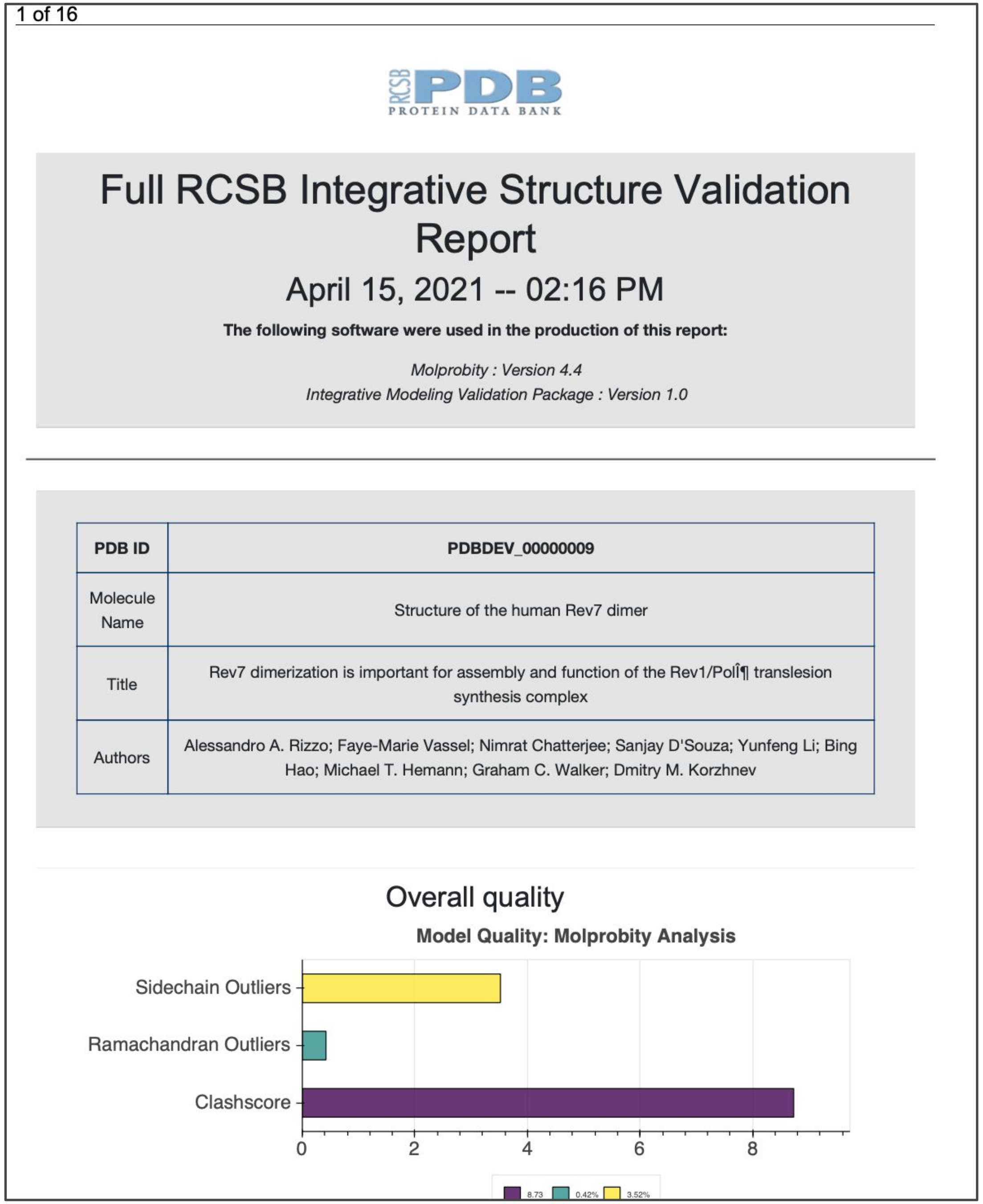


Validation Report for PDBDEV_00000009

PDBDEV_00000009

PDB ID	PDBDEV_00000009
Molecule Name	Structure of the human Rev7 dimer
Title	Rev7 dimerization is important for assembly and function of the Rev1/PolÎ¶ translesion synthesis complex
Authors	Alessandro A. Rizzo; Faye-Marie Vassel; Nimrat Chatterjee; Sanjay D'Souza; Yunfeng Li; Bing Hao; Michael T. Hemann; Graham C. Walker; Dmitry M. Korzhnev

I/HM Validation project funded by RCSB PDB.
Development of the I/HM dictionary and the PDB-Dev prototype system funded by NSF awards DBI-1519158, DBI-1756248, and DBI-1756250.



1 of 16

RCSB PDB PROTEIN DATA BANK

Full RCSB Integrative Structure Validation Report

April 15, 2021 -- 02:16 PM

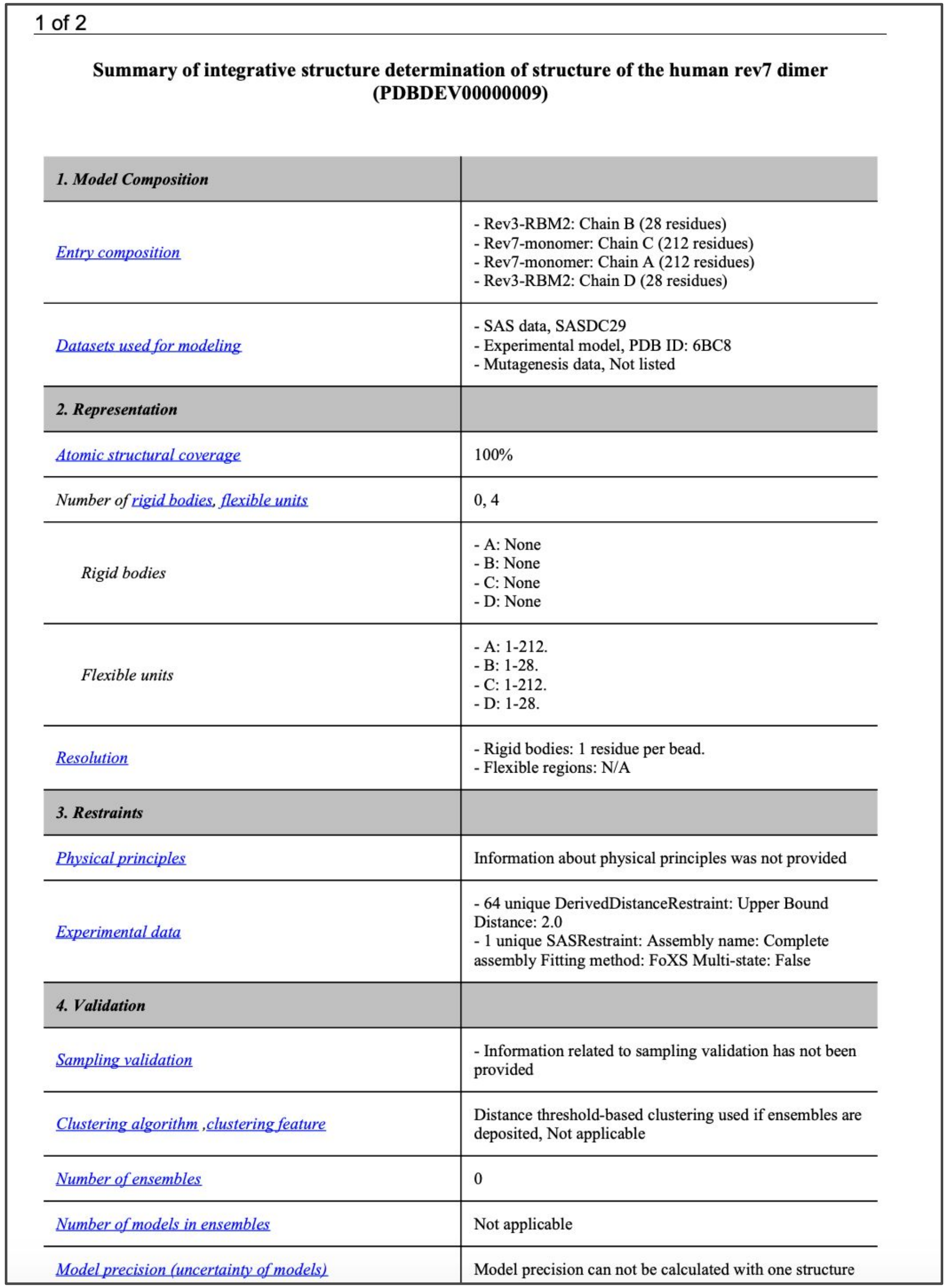
The following software were used in the production of this report:

Molprobity : Version 4.4
Integrative Modeling Validation Package : Version 1.0

PDB ID	**PDBDEV_00000009**
Molecule Name	Structure of the human Rev7 dimer
Title	Rev7 dimerization is important for assembly and function of the Rev1/PolÎ¶ translesion synthesis complex
Authors	Alessandro A. Rizzo; Faye-Marie Vassel; Nimrat Chatterjee; Sanjay D'Souza; Yunfeng Li; Bing Hao; Michael T. Hemann; Graham C. Walker; Dmitry M. Korzhnev

Overall quality

Model Quality: Molprobity Analysis



1 of 2

Summary of integrative structure determination of structure of the human rev7 dimer (PDBDEV00000009)

1. Model Composition	
Entry composition	- Rev3-RBM2: Chain B (28 residues) - Rev7-monomer: Chain C (212 residues) - Rev7-monomer: Chain A (212 residues) - Rev3-RBM2: Chain D (28 residues)
Datasets used for modeling	- SAS data, SASDC29 - Experimental model, PDB ID: 6BC8 - Mutagenesis data, Not listed
2. Representation	
Atomic structural coverage	100%
Number of rigid bodies, flexible units	0, 4
Rigid bodies	- A: None - B: None - C: None - D: None
Flexible units	- A: 1-212. - B: 1-28. - C: 1-212. - D: 1-28.
Resolution	- Rigid bodies: 1 residue per bead. - Flexible regions: N/A
3. Restraints	
Physical principles	Information about physical principles was not provided
Experimental data	- 64 unique DerivedDistanceRestraint: Upper Bound Distance: 2.0 - 1 unique SASRestraint: Assembly name: Complete assembly Fitting method: FoXS Multi-state: False
4. Validation	
Sampling validation	- Information related to sampling validation has not been provided
Clustering algorithm ,clustering feature	Distance threshold-based clustering used if ensembles are deposited, Not applicable
Number of ensembles	0
Number of models in ensembles	Not applicable
Model precision (uncertainty of models)	Model precision can not be calculated with one structure

Next Steps Funded by NSF

Funded by the NSF:

Collaborative Research: Creating mechanisms to make integrative structures of large macromolecular assemblies available from the Protein Data Bank

Principal Investigators

- Carl Kesselman (USC)
- Andrej Sali (UCSF)
- Brinda Vallat (Rutgers)

Questions and Comments?

Many thanks to Andy Byrd!

RCSB PDB AC member 2009-2021 and wwPDB AC member (2004-2021) and chair (2015-2018)

2009: Toronto, Canada

2019: Berkeley, CA

Acknowledgements

RCSB PDB PROTEIN DATA BANK

RCSB.ORG

info@rcsb.org

Funding

National Science Foundation (DBI-1832184),
National Institute of General Medical Sciences,
National Institute of Allergy and Infectious Disease, and
National Cancer Institute, (R01GM133198), and the
US Department of Energy (DE-SC0019749)

Management

RUTGERS

UC San Diego

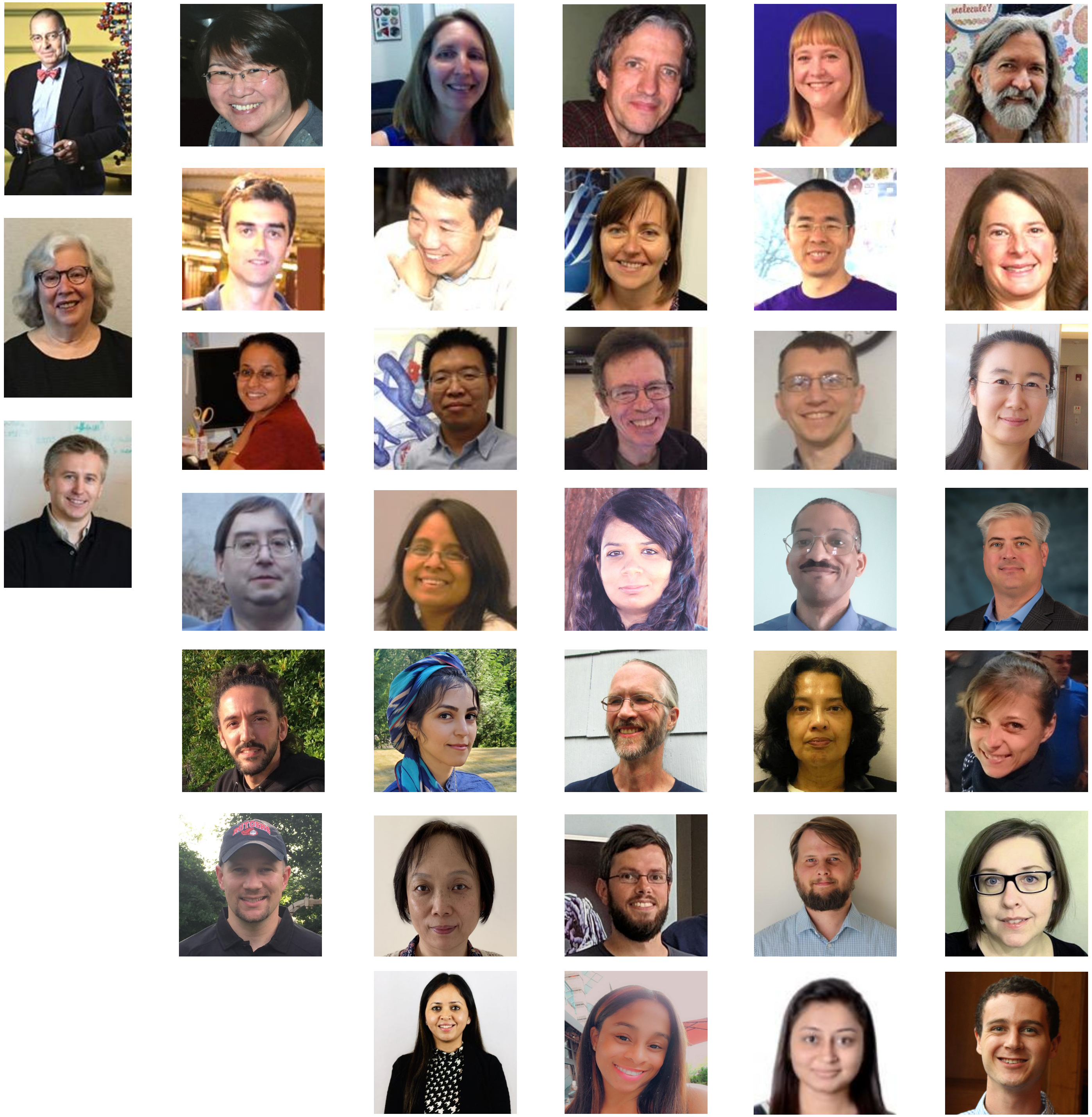
SDSC SAN DIEGO SUPERCOMPUTER CENTER

UCSF University of California San Francisco

WORLDWIDE PDB PROTEIN DATA BANK

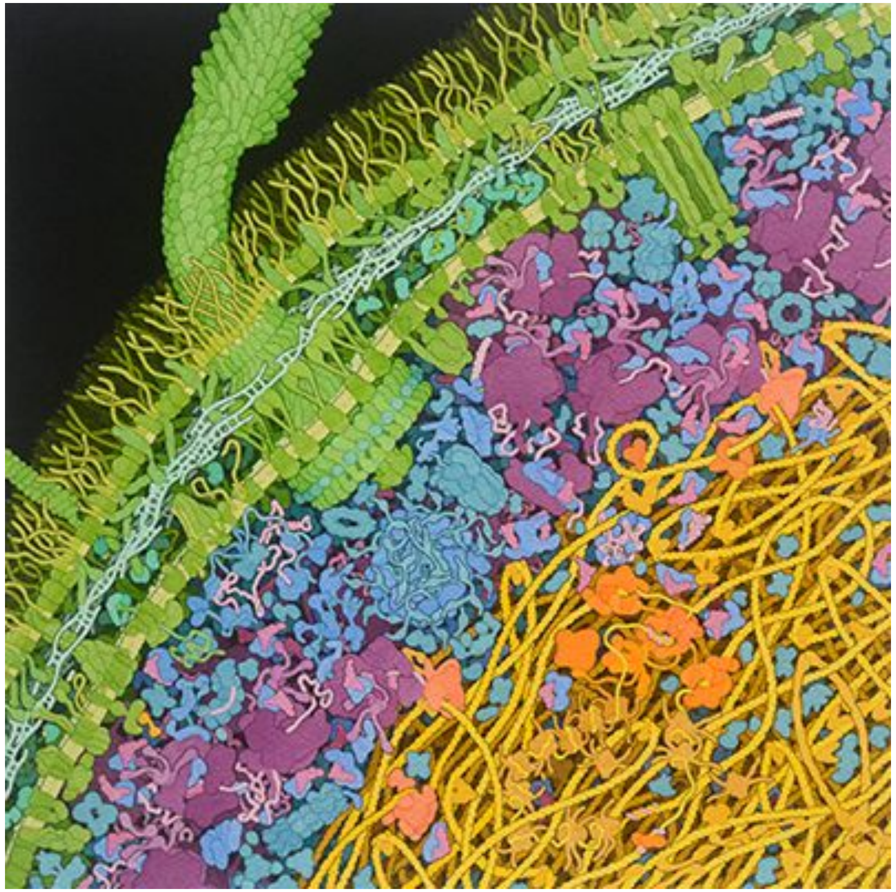
Member of the
Worldwide Protein Data Bank
(wwPDB; **wwpdb.org**)

Follow us



Group Screenshot Photo

Thank you



Molecular Landscapes: *E. coli*