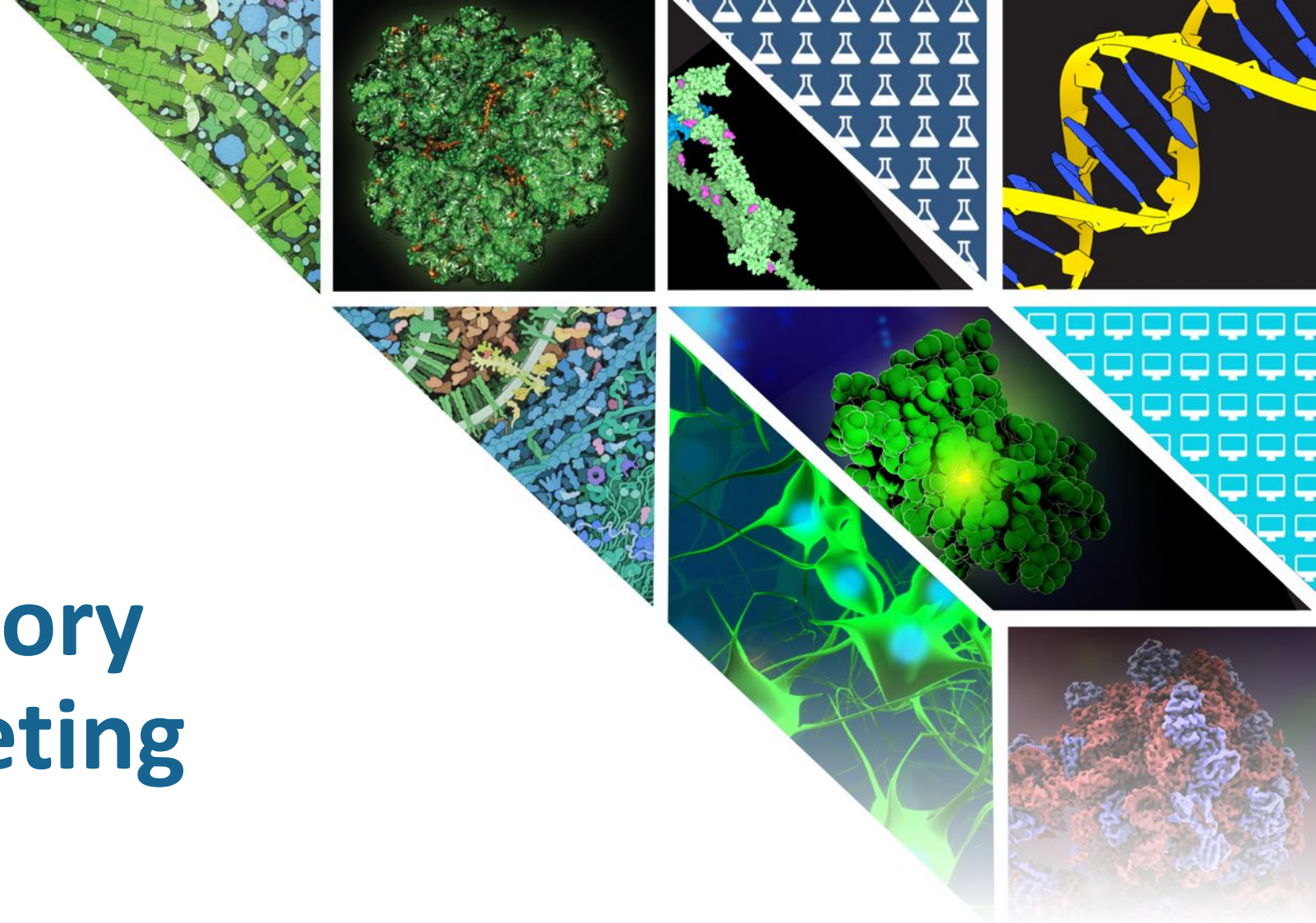**rcsb.org**

RCSB PDB Advisory Committee Meeting

April 12, 2023

Today's Participants: Welcome and Introductions

- **Advisory Committee**
  - *Present*: Paul Adams, Peter Andolfatto, Bridget Carragher, Wah Chiu, Kirk Clark, Roland Dunbrack, Paul Falkowski, Thomas Ferrin, Mandë Holford, Cathy Peishoff, Sue Rhee, Torsten Schwede, Lance Stewart, Kevin H. Gardner, Takita F. Sumter
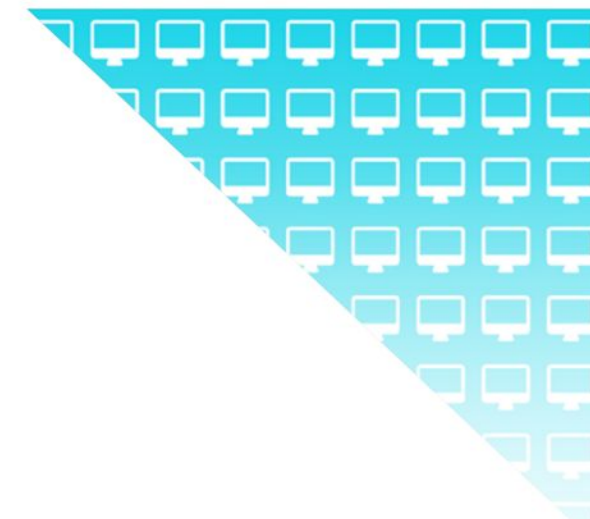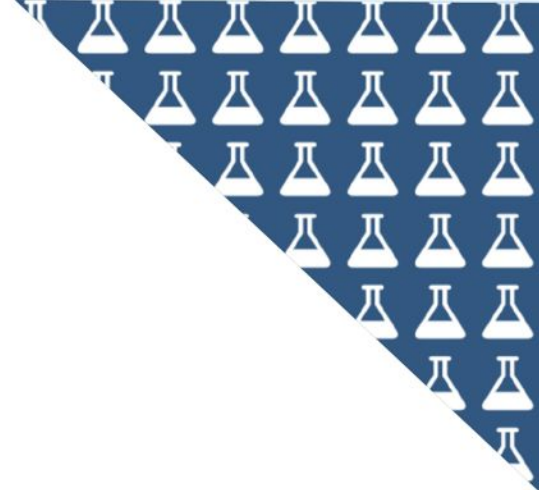 - *Absent*: Robert B. Darnell

- **RCSB PDB Participants**
  - *Leadership*: Stephen K. Burley (Director/PI), Andrej Sali (UCSF Site Head), Helen M. Berman (Director Emerita)
  - *Operations Team Representatives*: Jose Duarte (Scientific Software Lead and UCSD Manager), Henry Chao (DevOps Lead East), Zukang Feng (Principal Scientific Software Developer), Yana Rose (S3 Lead; Scientific Software Developer & Data Architect), Jasmine Young (S1-2 Lead; RCSB PDB Biocuration Team Lead & wwPDB Global Project Lead), Christine Zardecki (Deputy Director, S4 Lead)

Proposed Agenda: 2022 & Site Visit Planning

- Welcome, Introductions, and Group Photo
- 2022 Overview and Brief Updates
- Non-grant related questions for AC and Executive Session
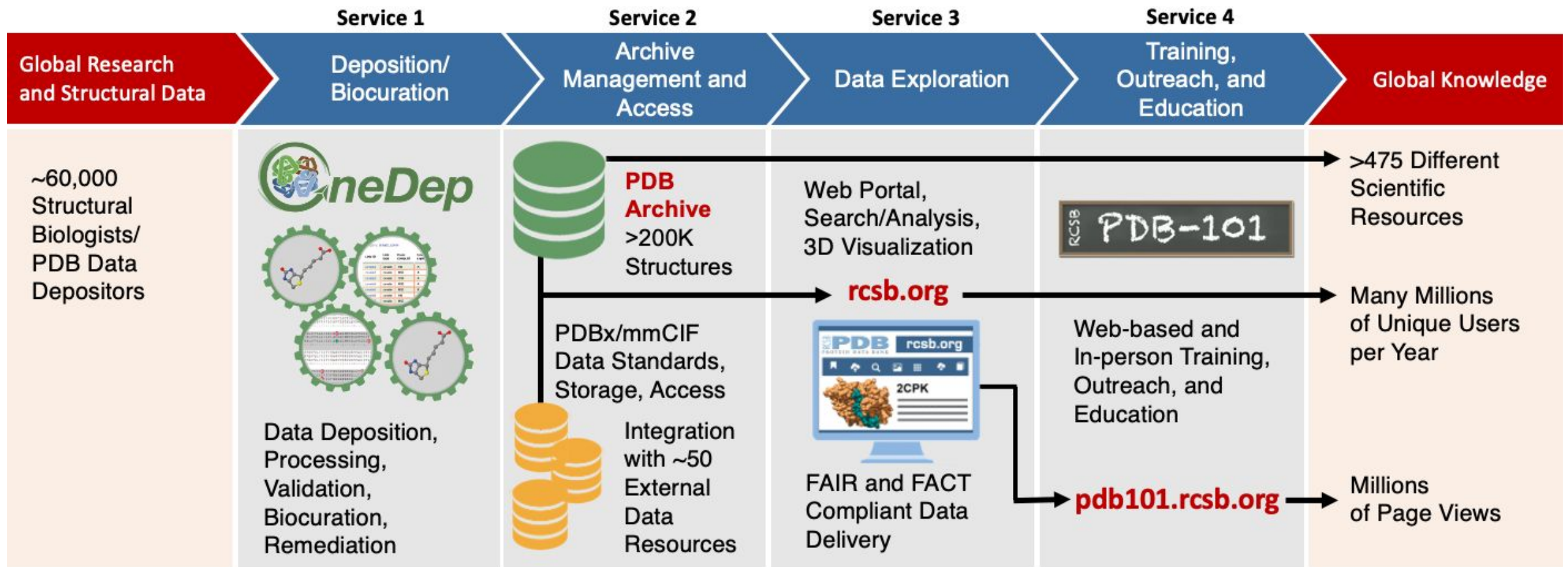- Grant preparations

Background Information Slides

- Slides with an *italicized, light blue title* are provided as background information and can be presented and discussed at the meeting by request.

- Main slides appear with non-italicized blue titles

- Underlined text indicates an active link

2022 Highlights

Stephen K. Burley

Combined Operations Across Four Services Drive Research and Innovation

Global Research and Structural Data	Service 1 Deposition/ Biocuration	Service 2 Archive Management and Access	Service 3 Data Exploration	Service 4 Training, Outreach, and Education	Global Knowledge
~60,000 Structural Biologists/ PDB Data Depositors	OneDep Data Deposition, Processing, Validation, Biocuration, Remediation	**PDB Archive** >200K Structures PDBx/mmCIF Data Standards, Storage, Access Integration with ~50 External Data Resources	Web Portal, Search/Analysis, 3D Visualization rcsb.org FAIR and FACT Compliant Data Delivery	RCSB PDB-101 Web-based and In-person Training, Outreach, and Education pdb101.rcsb.org	>475 Different Scientific Resources Many Millions of Unique Users per Year Millions of Page Views

2022 By The Numbers: What a Year!

Scientific Support and User Engagement

- Maintained 99.9% availability of RCSB.org and APIs in 2022

- RCSB PDB help desk supported ~800 conversations with users
  - Additional OneDep and structure-related questions transferred to wwPDB Help Desk

S1: Deposition/Biocuration

- Record 16,344 structures deposited and processed–PDB record (up from 14,571 in 2021)
  - 1,151 SARS-CoV-2 structures released (>3,000 available)

- 3327 new ligands and 45 new BIRD dictionary items

S2: Archive Management and Access

- Record 14,286 new PDB structures released
  - 7% of the year-end total holdings of 199,755 entries

- PDB surpassed the 200,000 structures on January 10, 2023

- Record ~3 billion data file downloads across the wwPDB

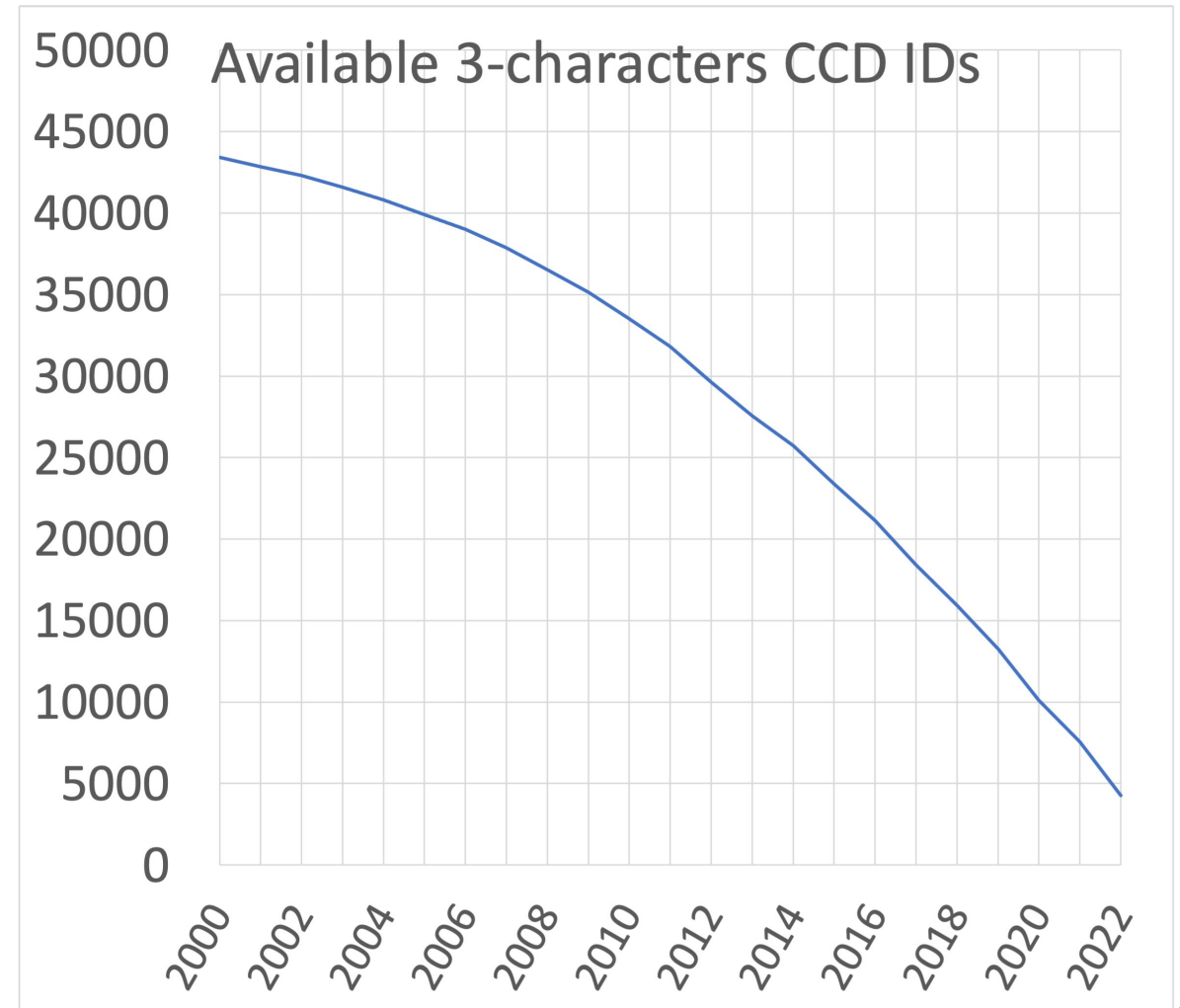- PDB Certified as a Global Core Biodata Resource

S3: Data Exploration

- Record ~7.2 million unique clients (unique IP addresses, up from ~6.8 million in 2021)

- 61 million sessions (~37 million estimated to be human users)

- ~2.8 billion requests/interactions (e.g., data downloads, service usage, web page views)

S4: Training, Outreach, Education

- ~663,000 PDB-101 users (down from ~692,000 in 2021)

- >3 million page views (up from ~2.3 million in 2021)

- 945,718 YouTube Channel views

- Virtual "crash courses"
  - CSMs and RCSB.org: ~180 attendees
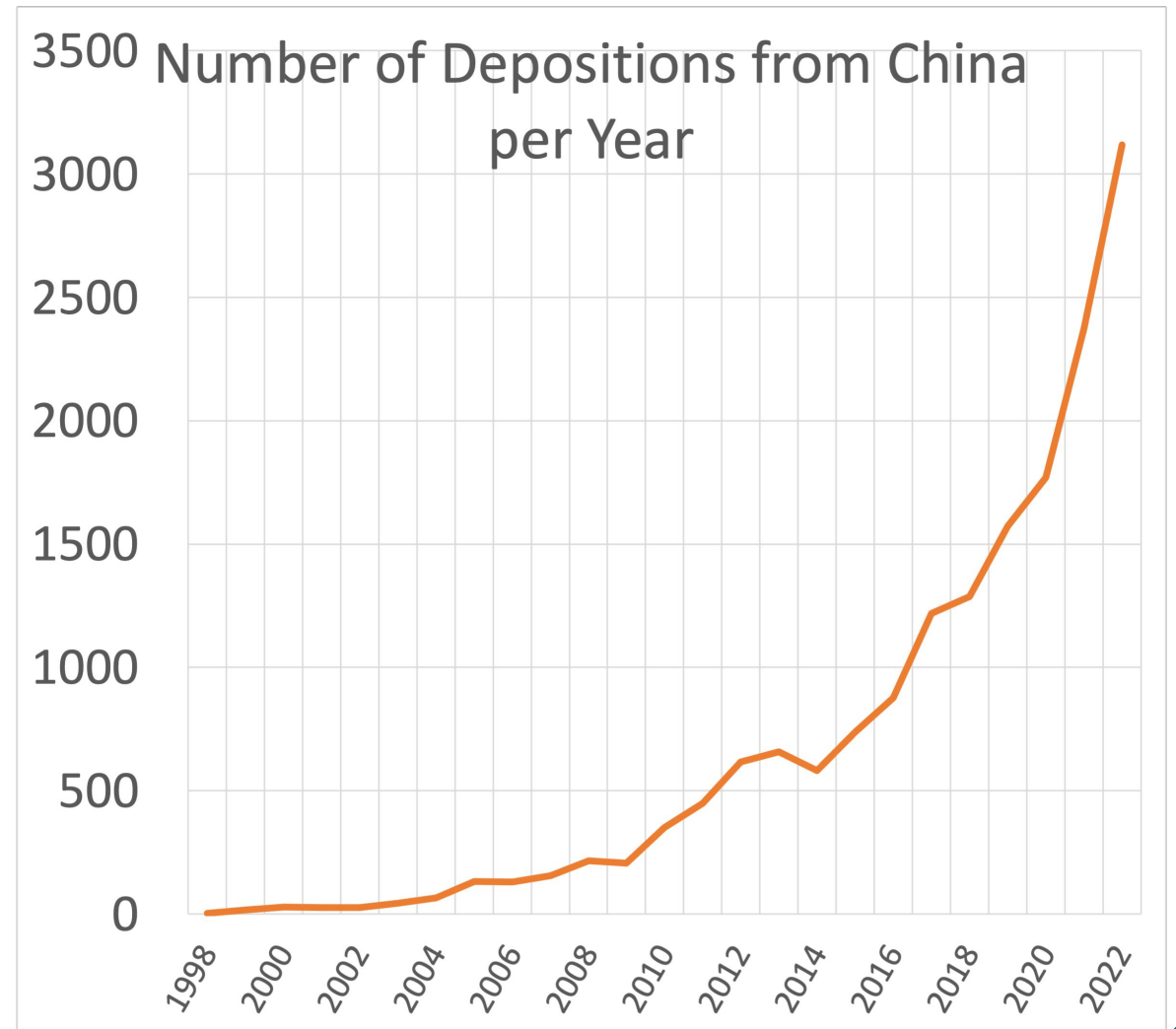 - KBase and RCSB PDB: ~125 participants

S1: Deposition and Biocuration Highlights

- Remediation
  - ~600 entries: funding agency data
  - ~400 entries: additional mutation annotation
  - ~200 corrections to PDB and EMDB ID mapping
  - 70 entries: translated from non-crystal to crystal frame

- OneDep System for Deposition, Validation, and Biocuration: 12 software upgrade releases
  - Mol* 3D assembly views for depositors
  - Mandatory half-map depositions for non-tomogram EM submissions
  - Enhanced validation reports with Q-score for 3DEM entries for depositors
  - Improvements in Biocuration efficiency

- Established plan to launch and support extended CCD IDs in 2023

- PDB China (PDBc) now wwPDB Associate Member; processing depositions

Available 3-characters CCD IDs

PDB China (PDBc)

- Associate member of wwPDB

- PDBc OneDep instance operated by PDBj is now in production

- Biocuration training provided onsite by PDBj and remotely by RCSB PDB/PDBe

- ~300 entries processed by PDBc in Q4 2022 during training (~10% of 3118 depositions from PRC in 2022)

- Geographic re-direction will be in place when the PDBc OneDep instance at Shanghai is in production
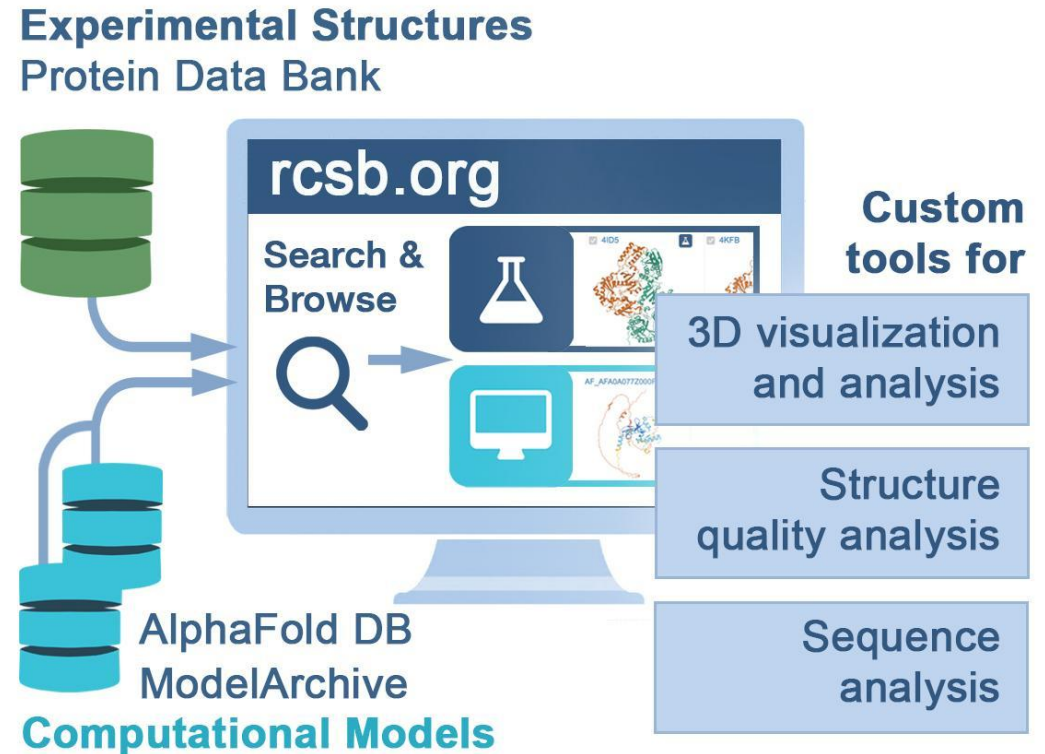
Number of Depositions from China per Year

Background Information

S2: Archive Management and Access Highlights

- Distribution of PDBx/mmCIF formatted assembly files

- Regeneration of all 3DEM validation reports to include Q-scores

- Inclusion of extended PDB IDs and DOIs in PDBx/mmCIF files

- Community engagement on extended PDB IDs and CCD IDs

- Extended ModelCIF to support Computed Structure Models (CSMs)

- Incorporated >1,000,000 CSMs from AlphaFold DB and the ModelArchive, requiring adaptation of RCSB PDB APIs to scale data delivery volume by a factor of 10

- Developed RCSB PDB Usage Analytics system (based on Elasticsearch) to monitor API access and performance, and real-time PDB Data Consumer experience

- Initiated data access scaling to support future incorporation of ~200 million additional CSMs (covering nearly every protein sequence in UniProt)
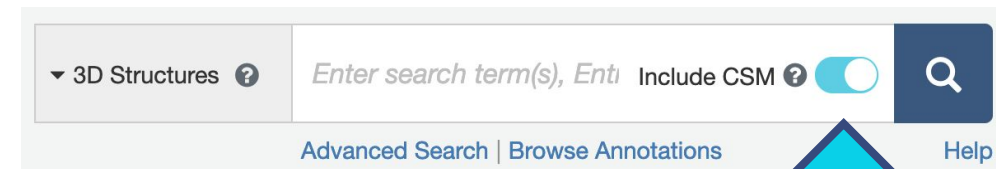
S3: Data Exploration Highlights

- Extended RCSB.org tools to support parallel delivery of >1M CSMs from AlphaFold DB and the ModelArchive alongside all PDB structures

- In collaboration with Rutgers graduate students, performed a User Experience Design (UXD) review of basic RCSB.org and Mol* services; Initial recommendations have been implemented

- Enabled delivery of data for groups of structures based on UniProt ID, sequence identity, or GroupDep structures as search results and in Group Summary Pages

- Launched a visualization tool for protein and nucleic acid sequences which (together with Mol*) enabled new 1D-3D views for single and groups of PDB and CSM structures
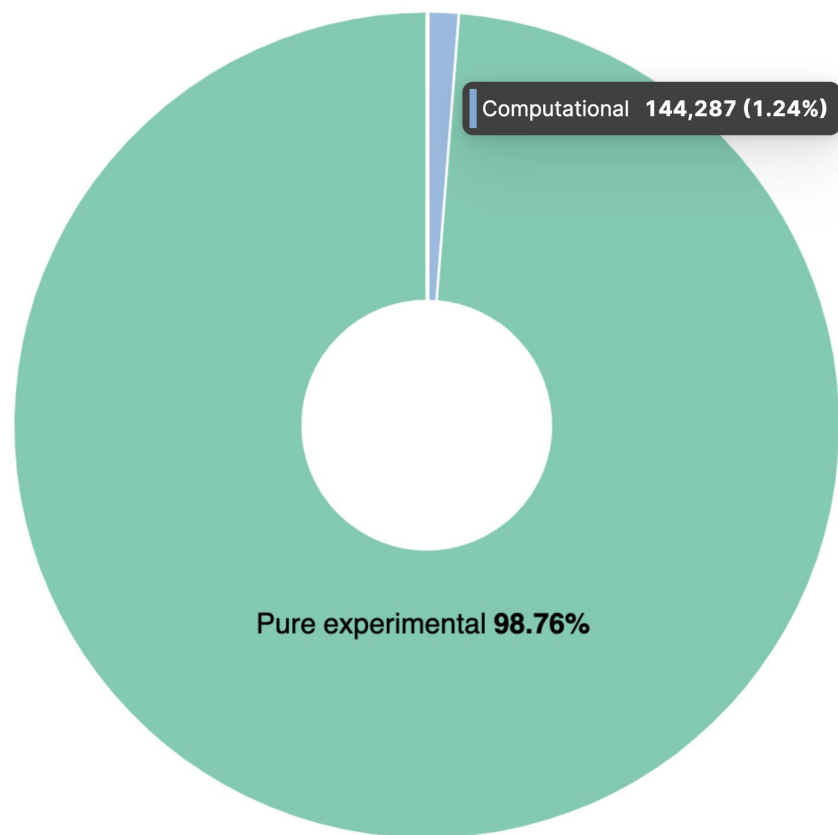

RCSB.org is one-stop shop for studying 3D structures of biomolecules
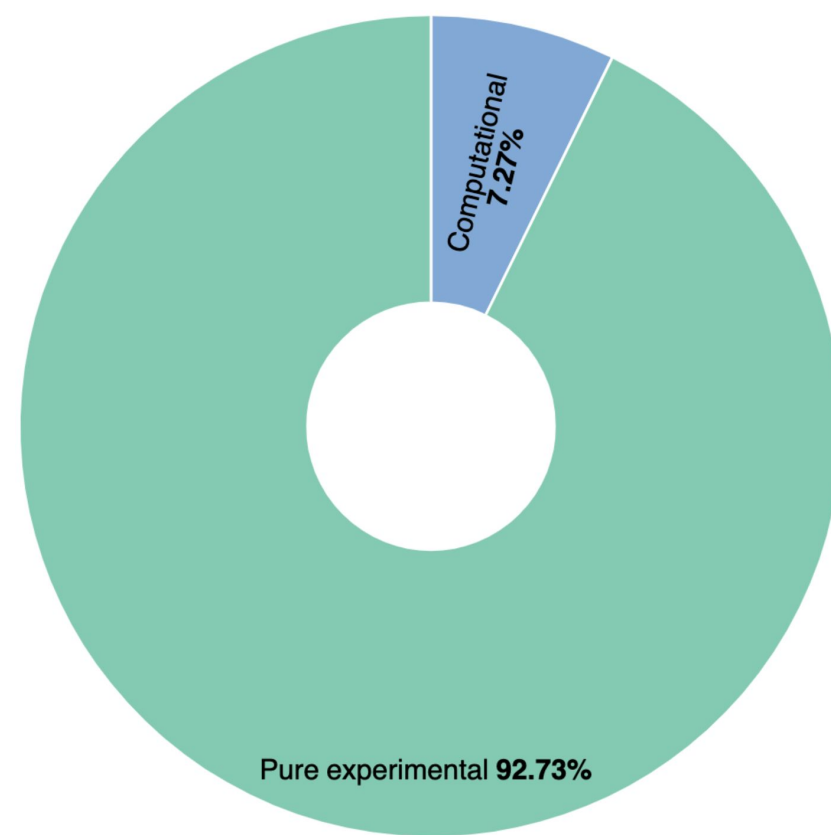
CSM "Opt-in" Utilization

Ratio of searches including CSMs vs experimental-only

Human searches

Computational 144,287 (1.24%)

Pure experimental **98.76%**

Programmatic searches

Computational 7.27%

Pure experimental **92.73%**

(Data from Oct. 7th, 2022–Mar. 31st, 2023)

Rutgers User Experience Design (UXD) Review

Spring Semester 2022: Two Master of Business and Science program students
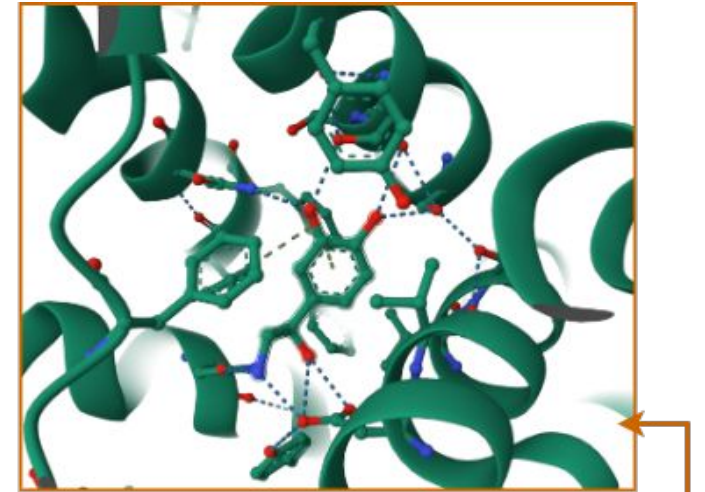
Process
- Design study to identify how undergraduate professors use RCSB.org in their classroom
- Collect and analyze user needs and pain points
- Deliver recommendations on how to improve the usability and experience of the tools they frequently use
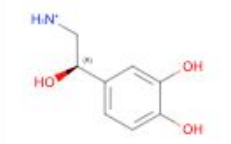
2022 Activities
- Search and Mol*

Summer 2023 Activities
- Structure Summary Pages

E5E (Subject of Investigation/LOI)
Query on E5E

Download Ideal Coordinates CCD File
Download Instance Coordinates

D [auth A]

Noradrenaline
$C_8 H_{12} N O_3$
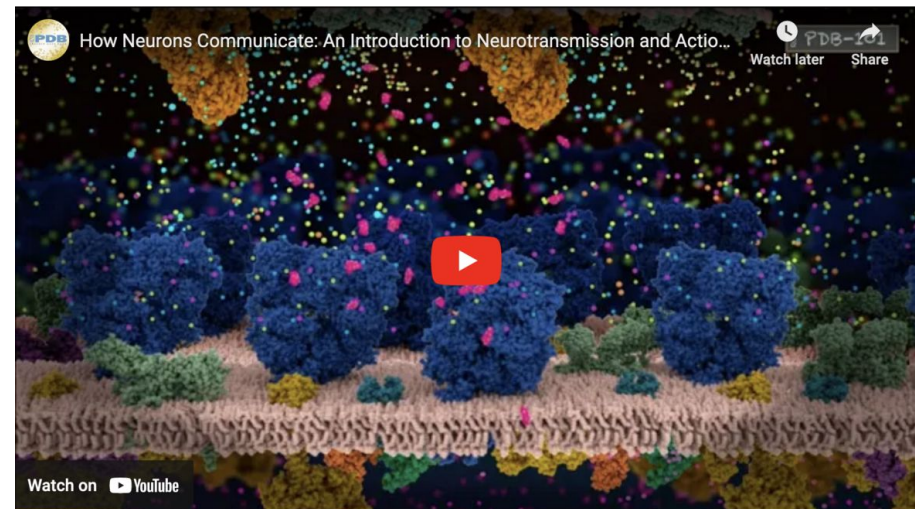SFLSHLFXELFNJZ-QMMMGPOBSA-O

Ligand Interaction

https://www.rcsb.org/3d-view/7BU6?preset=ligandInteraction&label_asym_id=D

Reports and Current Status of 2022 Reviews

- Spring 2022: Overall RCSB.org Usability
  - Initial recommendations have been implemented
    - RCSB PDB AC extremely enthusiastic
  - Jira tickets created for additional improvements/implementations
    - Modifications specified for implementation in 2023-2024

- Summer 2022: Initial Mol* Usability
  - Translation of report recommendations → Jira still outstanding
    - Modifications need to be specified for implementation in 2023-2024

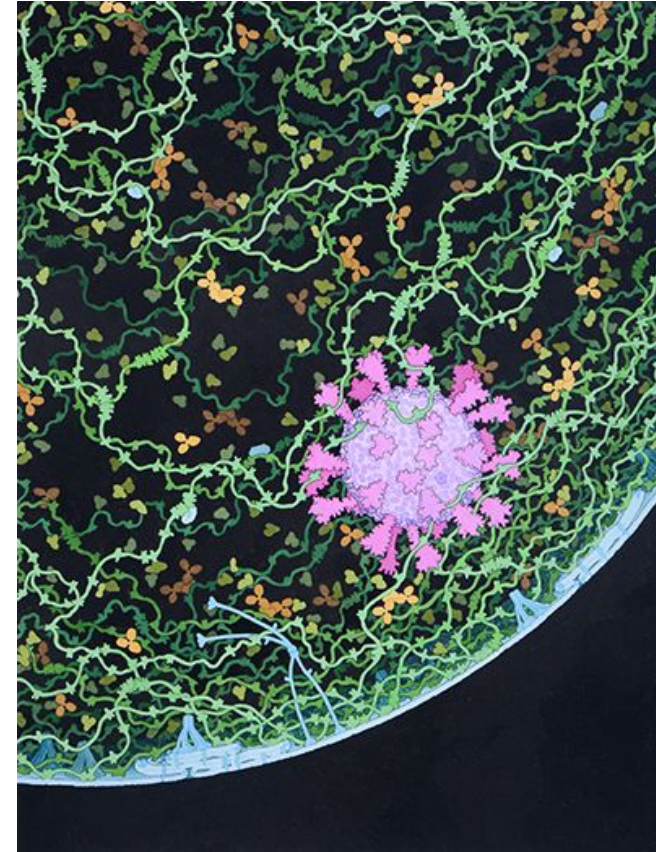- Summer 2023: Review of Structure Summary Page navigation

Service 4: PDB-101 Activities

- Website redesigned to highlight Training materials
  - Includes Crash Courses on KBase and CSMs
- 2022 IQB Winter Boot Camp with undergraduate and graduate students generated Molecule of the Month articles on Cancer Drug Targets
- New Features
  - Videos: Intro to Neurotransmission and Action Potential; Photosynthesis
  - SciArt: new Molecular Landscapes
  - Guide to Computed Structure Models
  - Exploring the Structural Biology of Cancer

Award-winning Outreach and Education

- David S. Goodsell, Carl Brändén Award from The Protein Society and the I. Fankuchen Award from the American Crystallographic Association

- Shuchismita Dutta named a 2022 Fellow of the American Association for the Advancement of Science in the section of Education

- PDB-101 images selected as runners-up in international Wiki Science Competition (WSC) 2021: Maria Voigt's *Molecular view of glutamatergic synapse* and David Goodsell's *Respiratory Droplet*

Respiratory Droplet

Training, Outreach, and Education (TOE) Working Group Established

- Members, facilitated by former RCSB PDB Advisor Paul Craig (Rochester Institute of Technology), provide feedback on RCSB PDB resources and collaborate on promoting PDB-101 training

- Tested RCSB.org CSM incorporation

- Provided feedback on funding renewal application

- Will review TOE-related Reverse Site Visit presentation

Select TOE Working Group Members

Full Membership: https://pdb101.rcsb.org/more/training-outreach-education-working-group

Questions for Advisor Feedback

- Any concern on our responses to last year's AC report?

- Any concerns regarding 2022 wwPDB AC activities?

- Any suggestions on venues to [communicate extended PDB/CCD IDs](#)?

- Thoughts on approaches to CSM integration?
  - Are your lab members "opting in" to perform CSM searches?
  - Should we be advertising CSM features more (and suggestions for where to disseminate)?

- What is your preference for CSM searching at RCSB.org: opt-in to include models, or opt-out to exclude?

- What would be best approach for making new RCSB PDB features better known and more accessible?
  - Website Quick Guides, How-to videos, Home page redesign, … Other suggestions?

- Topics for future crash courses/training?
  - Advanced Search, APIs, CSMs, Mol*, … Other suggestions?