Welcome

Leveraging RCSB PDB APIs for Bioinformatics Analyses and Machine Learning

Oct 12th, 2023

Jose Duarte
jose.duarte@rcsb.org
Course split into 2 parts

- Part 1 (today): introduction to APIs
- Part 2: hands-on exercise
  - Oct 19: starting 16:00 UTC
  - Oct 27: starting 00:00 UTC
Crash course recording and presentations

- Zoom recording will be posted on RCSB PDB’s YouTube Channel
- Presentations will be available to participants via Exit Survey
- Q&A Summary: go.rutgers.edu/86q3uya7
- Part 2 (hands-on) registration will be shared with today’s participants
Tutors

Jose Duarte
Brinda Vallat
Yana Rose
Dennis Piehl
Sebastian Bittrich
Joan Segura
## Part 1 Agenda

<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
<th>Presenter</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>9:00 - 9:10 AM PDT</strong></td>
<td>Welcome</td>
<td></td>
</tr>
<tr>
<td>12:00 - 12:10 PM EDT</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>9:10 - 9:40 AM PDT</strong></td>
<td>Introduction to RCSB PDB APIs and data schemas</td>
<td>Brinda Vallat, PhD</td>
</tr>
<tr>
<td>12:10 - 12:40 PM EDT</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>9:40 - 10:10 AM PDT</strong></td>
<td>Data API</td>
<td>Jose Duarte, PhD</td>
</tr>
<tr>
<td>12:40 - 1:10 PM EDT</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>10:10 - 10:20 AM PDT</strong></td>
<td>Break</td>
<td></td>
</tr>
<tr>
<td>1:10 - 1:20 PM EDT</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>10:20 - 10:50 AM PDT</strong></td>
<td>Search API</td>
<td>Yana Rose, PhD</td>
</tr>
<tr>
<td>1:20 - 1:50 PM EDT</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>10:50 - 11:00 AM PDT</strong></td>
<td>Search and Data API hands-on teaser</td>
<td>Dennis Piehl, PhD</td>
</tr>
<tr>
<td>1:50 - 2:00 PM EDT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>Activity</td>
<td>Speaker(s)</td>
</tr>
<tr>
<td>-----------------</td>
<td>--------------------------------------------------------------------------</td>
<td>---------------------------------------</td>
</tr>
<tr>
<td>9:00 - 9:10 AM PDT</td>
<td>Introduction</td>
<td>Dennis Piehl, PhD</td>
</tr>
<tr>
<td>12:00 - 12:10 PM EDT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9:10 - 9:40 AM PDT</td>
<td>Walk through a worked example in Colab</td>
<td>Joan Segura, PhD Sebastian Bittrich, PhD</td>
</tr>
<tr>
<td>12:10 - 12:40 PM EDT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9:40 - 10:30 AM PDT</td>
<td>Time for participants to work on their own problem</td>
<td>Tutors available for questions and breakout discussions as needed</td>
</tr>
<tr>
<td>12:40 - 1:30 PM EDT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10:30 - 11:20 AM PDT</td>
<td>Present results (report out). More time for Q&amp;A</td>
<td></td>
</tr>
<tr>
<td>1:30 - 2:20 PM EDT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11:20 - 11:30 AM PDT</td>
<td>Closing remarks</td>
<td></td>
</tr>
<tr>
<td>2:20 - 2:30 PM EDT</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Introduction to RCSB PDB APIs and Data Schema

Leveraging RCSB PDB APIs for Bioinformatics Analyses and Machine Learning

October 12th, 2023

Brinda Vallat
brinda.vallat@rcsb.org
The RCSB PDB Web Portal (RCSB.org)

- **RCSB.org**: Tools for searching, accessing, visualizing, analyzing, and downloading the contents of the PDB Archive
- Open access to >210,000 experimental structures of macromolecules
- >1 million Computed Structure Models (CSMs) predicted using AI/ML methods
- Living data resource integrated with annotations from ~50 external biodata resources (UniProt, SCOPe, CATH, …)
- **PDB-101 (pdb101.rcsb.org)**: Educational resources and training

3D structural data from around the world
Data in the PDB Archive

- Experimentally determined structures of macromolecules and their complexes
  - Atomic coordinates
  - Molecular descriptions, references
  - Source organisms, details about sample, experiment
  - Citation, software, authors
- Depositor provided and software generated
- Well-curated data ensures data standardization and completeness
- Validation metrics for assessment of structure quality

https://pdb101.rcsb.org/motm/253
PDB Data Organization: Molecular Hierarchy

- Entry: 4hhb
- Entity: 1 (alpha subunit)
- Entity Instance: C (chain C)
- Assembly: (quaternary structure)
- Chemical Component: HEM (HEME molecule)

- e.g. title of the entry, list of depositors
- e.g. protein, DNA, RNA, membrane lineage
- e.g. structural connectivity, secondary structure
- e.g. transformations required to generate the biological assembly
- e.g. chemical descriptors (SMILES & InChI), chemical formula

https://data.rcsb.org/index.html#data-organization
RCSB.org: Integration of External Annotations
### RCSB.org: Information Integrated from External Resources

<table>
<thead>
<tr>
<th>Data Content Type</th>
<th>Resource Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protein domain classification</td>
<td>SCOP2, CATH</td>
</tr>
<tr>
<td>Evolution related</td>
<td>ECOD</td>
</tr>
<tr>
<td>Antibiotic resistance</td>
<td>CARD</td>
</tr>
<tr>
<td>Immunology related</td>
<td>IMGT, SAbDab</td>
</tr>
<tr>
<td>Small molecule related</td>
<td>CSD, COD, ChEBI, ChEMBL, NCBI PubChem</td>
</tr>
<tr>
<td>Glycan correspondences</td>
<td>Glygen, GlyToucan, GlyCosmos</td>
</tr>
<tr>
<td>Membrane protein related</td>
<td>MemProtMD, PDBTM, OPM, mpstruc</td>
</tr>
<tr>
<td>Diffraction images</td>
<td>SBGrid, proteindiffraction.org</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Data Content Type</th>
<th>Resource Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene and Taxonomy</td>
<td>NCBI Gene and Taxonomy, Gencode</td>
</tr>
<tr>
<td>Molecular function, Cellular component, Biological processes</td>
<td>GO</td>
</tr>
<tr>
<td>Drug target related</td>
<td>Pharos, Drugbank</td>
</tr>
<tr>
<td>Enzyme nomenclature</td>
<td>ExploreEnZ</td>
</tr>
<tr>
<td>Binding affinities</td>
<td>BindingDB, Binding MOAD</td>
</tr>
<tr>
<td>Protein families</td>
<td>Pfam, InterPro</td>
</tr>
<tr>
<td>Citation and literature</td>
<td>Pubmed, PMC</td>
</tr>
<tr>
<td>Sequence related</td>
<td>UniProt</td>
</tr>
</tbody>
</table>
RCSB.org: Powered by APIs

PDB Archive

RCSB Data API

```
{
  "id": "4HHB",
  "title": "The crystal structure of human deoxyhaemoglobin at 1.74 Å resolution"
}
```

Custom Code

ExternalResource.com

THE CRYSTAL STRUCTURE OF HUMAN DEOXYHAEMOGLOBIN AT 1.74 ÅNGSTOMS RESOLUTION

PDB Identifier 4hhh

RCSB.org APIs

Application Programming Interface or APIs provide programmatic access to data available from RCSB.org

Two main APIs that power the RCSB.org website are:

- **Data API** serves to retrieve data when the PDB identifiers are known
- **Search API** serves to find out what identifiers match specific search conditions
RCSB.org: Other APIs

- **ModelServer** is a service for accessing subsets of macromolecular model data. It delivers atomic coordinates together with annotations in a compressed BinaryCIF encoding (BCIF).
- **VolumeServer** is a service for accessing subsets of volumetric data and provides near-instant access to large data sets.
- **1D Coordinate Server** compiles alignments between structural and sequence databases and integrates protein positional features from multiple resources. Alignments are available for Refseq, UniProt, and PDB sequences.
Search Tools at RCSB.org

Basic search

Advanced search
RCSB.org: Search Results Page

- **Structure attribute search**

- **Search Results**
  - Search API retrieves the identifiers matching the search criteria
  - Data API fetches data related to the identifiers (Title, Authors, Publication, Release date, Experimental method, Macromolecule name)
Data API provides information seen on the “Structure Summary” page.
RCSB.org: Experiment Details

Data API provides information seen on the “Experiment” tab
Data API provides information seen on the “Annotations” tab.
RCSB.org APIs provide information seen on the 1D-3D “Protein Feature View”
RCSB.org Data Schema

Supports Search and Data APIs
What is Schema?

• Schema is metadata: Provides information about how data is structured

• Schema in the context of databases (data schema or data model)
  • Schema provides organization, structure, and architecture for a database
  • Describes the shape of the database and how different data tables or objects relate to one another
  • Applicable to both relational and document-oriented databases

• Schema in the context of APIs
  • Like database schemas but for APIs
  • Describes API operations and how to interact with APIs
RCSB.org Data Schema

- Follows the PDBx/mmCIF data dictionary
  - Data standard followed by the PDB archive
  - Definitions for macromolecular structures and associated metadata
  - Dictionary definition language supports specifications for data types, controlled vocabularies, mandatory attributes
  - Designed to be extensible and has software support
  - Virtual crash course: https://pdb101.rcsb.org/train/training-events/mmcif

- RCSB extension with additional definitions
  - Definitions specific for RCSB.org data delivery
  - Definitions for annotations integrated from external resources
  - Prefixed with “rcsb_”

- Data organization based on molecular hierarchy
  - Core objects: Entry, Entity, Entity Instance, Assembly, Chemical Component

- Powered by JSON schema language

https://data.rcsb.org/index.html#data-schema
PDBx/mmCIF Data in JSON Format

Source data from PDBx/mmCIF

```json
{  
  "key1": "value1",
  "key2": "value2",
}
```

```json
{  
  "id": "4HHB",
  "title": "THE CRYSTAL STRUCTURE OF HUMAN DEOXYHAEMOGLOBIN AT 1.74 ANGSTROMS RESOLUTION",
  "method": "X-RAY DIFFRACTION",
}
```

Keys (objects and attributes)  Values

RCSB Extension: Common Data Objects

• Positional features
  • Entities and instances: rcsb_<core_object>_feature
  • Example: CATH, SCOP, ECOD, mutations, model quality metrics, validation metrics, ligand binding sites, accessible surface area

• Positional feature summaries
  • Entities and instances: rcsb_<core_object>_feature_summary
  • Feature statistics: count, coverage, minimum value, maximum value

• Annotations
  • Entities and instances: rcsb_<core_object>_annotation
  • Example: GO, InterPro, Pfam

• Provenance information: PDB or other data source
Examples

Data fetched from the Data API

Data in PDBx/mmCIF

Annotations integrated from external resources

Data fetched from the Data API

Data from PDBx/mmCIF

Positional features from the PDB
Summary

- RCSB.org is powered by APIs
  - Search API provides programmatic access to all functionalities supported by RCSB.org basic and advanced search
  - Data API provides programmatic access all static data delivered on RCSB.org
  - Available to all users

- RCSB.org APIs are supported by the underlying data schema
  - Information from the PDB archive in PDBx/mmCIF
  - Information integrated from external resources
  - Data organized and mapped into molecular hierarchy
  - Powered by JSON schema language
RCSB PDB Data API

Leveraging RCSB PDB APIs for Bioinformatics Analyses and Machine Learning

October 12th, 2023

Jose Duarte
jose.duarte@rcsb.org
Data API: What Is It

Retrieve data once you know **identifiers**

Identifiers may come from external API/resource, references in some publication …

If identifiers not known: use Search API first. Typical workflow:

1. Use **Search API** to find identifiers that match a specific set of conditions
2. For each identifier, use **Data API** to retrieve data related to the identifier
Interfaces for Data API

REST
- Endpoints per granularity
- Get ALL data for given object
- Note some endpoints are offered only in REST (e.g. holdings)

GraphQL
- Single entry point that can traverse the entire schema
- Get only the data you need for your use-case
- JSON-based query language

The output is in JSON format
GraphQL Basics

Graph Query Language

Why GraphQL: prevent under and over-fetching

Despite its name: it is NOT a query language that can be used to express a search condition

Error handling in GraphQL: not your standard http response codes

- All queries return 200 response code
- Errors come in the response payload in json format
GraphQL Tooling

Very large ecosystem of tools in many languages: exploration tools, IDEs, visualization tools, editors, validators, automatic type generation

Most importantly: plenty of client implementations
The GraphQL Tool

Interactive query builder (IDE-like):

- Autocomplete
- Documentation as you type
- Navigating to schema

https://data.rcsb.org/graphql/index.html
Submitting a Query

Both GET and POST available for query submission

GET example:

https://data.rcsb.org/graphql?query={entry(entry_id:"4HHB"){exptl{method}}}  
URL encoded:

https://data.rcsb.org/graphql?query=%7Bentry%28entry_id%3A%224HHB%22%7D%7Bexptl{method%7D%7D

We recommend using a client: it will take care of everything. Especially important for error handling.
Query by Example

The Data API button in rcsb.org. Present in:

• Structure Summary Pages and tabs
• Ligand Summary Pages

Gets the data that is needed to render the page
Reminder: PDB Data Hierarchy (Granularities)

Identifiers
"rcsb_id"

<table>
<thead>
<tr>
<th>4HHB</th>
<th>4HHB_1</th>
<th>4HHB.C</th>
<th>4HHB-1</th>
<th>HEM</th>
</tr>
</thead>
</table>

**Data API core object hierarchy**

- **Entry**: 4hhb
- **Entity**: 1 (alpha subunit)
- **Entity Instance**: C (chain C)
- **Assembly**: (quaternary structure)
- **Chemical Component**: HEM (HEME molecule)

Each level of hierarchy has its own set of attributes in the Data API:
- e.g. title of the entry, list of depositors
- e.g. protein, DNA, RNA, membrane lineage
- e.g. structural connectivity, secondary structure
- e.g. transformations required to generate the biological assembly
- e.g. chemical descriptors (SMILES & InChI), chemical formula
Connections Between Granularity Levels

- Top level accessors for each granularity in 2 flavours: single and list
  
  entry(entry_id: "4HHB")
  
  entries(entry_ids: ["4HHB", "3HBX"])
  
  assembly(entry_id: "4HHB", assembly_id: "1")
  
  assemblies(assembly_ids: ["4HHB-1", "3HBX-1"])

- Connection accessors from other granularity levels
Some Examples

Title, experimental method and resolution and Rfree for some entries

Organisms and cluster membership of polymeric entities

Annotations at chain level (e.g. CATH or SCOP)

Data associated to a Computed Structure Model (e.g. pLDDT)

Interface properties for a certain assembly
FAQ

Q: How do I find what field has the data I want?
   A: Query by example, GraphiQL schema browser and contextual help

Q: Does the request allow for filtering?
   A: No. Filtering must be done by consumer

Q: How do I get data for the whole archive?
   A: Holdings REST endpoint and GraphQL queries by batches
Resources

Tutorial and many examples at: https://data.rcsb.org

The full list of data attributes:
https://data.rcsb.org/data-attributes.html

The schema browser in GraphiQL:
https://data.rcsb.org/graphql/index.html ("Docs" link on top right)

Reference for REST endpoints: https://data.rcsb.org/redoc/index.html
Questions?
RCSB PDB Search API

Leveraging RCSB PDB APIs for Bioinformatics Analyses and Machine Learning

October 12th, 2023

Yana Rose

yana.rose@rcsb.org
In this session, we'll dive into the RCSB PDB Search API's capabilities. You’ll learn how to utilize advanced query options tailored to the needs of structural bioinformaticians.

Search API is a powerful tool that allows you to programmatically query the RCSB PDB data.

- REST over HTTP using JSON
- Search API defines a language for writing complex queries that can be used to retrieve a list of the PDB IDs that match these criteria.
Overview of Available Search Options

- **Text Search**
  Structures with metadata that matches specific keywords or values, e.g. release date, resolution, experimental details, taxonomy

- **Structure**
  Assemblies and chains resembling a target, in terms of the global volumetric shape

- **Sequence**
  "Homologous" nucleotide or protein sequences, statistically significant similarity that reflects common ancestry

- **Sequence Motif**
  Related sequences that share small conserved regions (motifs) that have biological meaning

- **Structure Motif**
  Assemblies with similar patterns of local structure associated with function, e.g. catalytic sites

- **Chemical**
  Small molecules which are similar to the query chemical structure, in terms of calculated molecular descriptors or "fingerprints"

**_exptl.method_**

**X-RAY DIFFRACTION**

**HUMAN**

**CHIMP**

**MOUSE**

**RAT**

**COW**

**AND**

**OR**

**1BFB, 1KHT, 2J9Z, 2LBG, 3ZOE, 4BIK**
Search API: Query Language

Search API provides a full DSL (Domain Specific Language) based on JSON to define queries. Query context consists of two types of clauses:

1. **Terminal Query Clauses**: individual search criteria, e.g. match a particular value in a particular field or run a sequence search

2. **Group Query Clauses**: wrap other terminal or group queries and are used to combine multiple queries in a logical fashion (AND, OR)

[https://search.rcsb.org/#query-language](https://search.rcsb.org/#query-language)
Search API: Return Type

Underlying macromolecular structure hierarchy progresses from atoms through amino acids and chains to assemblies of interacting macromolecules and ligands. Search API can return identifiers for the following levels:

- Entry
- Entity
- Entity Instance
- Assembly
- Chemical Component

RCSB PDB identifiers: 4HHB, 4HHB_1, 4HHB.C, 4HHB-1, 4HHB_3

Instance identifier corresponds to the _label_asym_id from the mmCIF schema (assigned by the PDB). It can differ from _auth_asym_id (selected by the author at the time of deposition).

https://search.rcsb.org/#return-type
Search API: Request Options

Request Options context determines what is included in a search response:

Include Computed Structures:

```json
{
  "request_options": {
    "results_content_type": ["experimental", "computational"],
    "paginate": {
      "start": 0,
      "rows": 2
    }
  }
}
```

Return All Hits:

```json
{
  "request_options": {
    "results_verbosity": "compact",
    "return_all_hits": true
  }
}
```

Redundancy Filter:

```json
{
  "request_options": {
    "group_by": {
      "aggregation_method": "sequence_identity",
      "similarity_cutoff": 30
    },
    "group_by_return_type": "representatives"
  }
}
```

Sorting:

```json
{
  "request_options": {
    "sort": [
      {"sort_by": "rcsb_accession_info.initial_release_date",
       "direction": "desc"}
    ]
  }
}
```
Search API is documented with the OpenAPI Specification ([https://swagger.io/specification](https://swagger.io/specification)). This reference documentation describes how to use the endpoints Search API is exposing:

```json
{
  "openapi": "3.0.1",
  "info": {
    "title": "UCSF Search API",
    "description": "Provides programmatic access to RCSB search API."
  },
  "servers": [{
    "url": "/search/v2/
  }]
}
```

Download: [https://search.rcsb.org/openapi.json](https://search.rcsb.org/openapi.json)
UI: [https://search.rcsb.org/redoc/index.html](https://search.rcsb.org/redoc/index.html)
Attributes Available for Search

API endpoint:
https://search.rcsb.org/rcsbsearch/v2/metadata/schema

Comprehensive documentation:
https://search.rcsb.org/structure-search-attributes.html

RCSB.org UI:
https://www.rcsb.org/search/advanced
Search API: Query by Example

[Diagram showing search options for sequence, structure, and text search with example JSON query and URLs]

https://www.rcsb.org/structure/4HHB

https://search.rcsb.org/#examples

https://search.rcsb.org/query-editor.html
Useful Resources

Learn more about Search API:

• Reference Documentation: https://search.rcsb.org/redoc/index.html
• User Guide: https://search.rcsb.org/#search-api
• Tutorial: https://pdb101.rcsb.org/learn/guide-to-understanding-pdb-data/introduction-to-rcsb-pdb-apis
• Help Desk: info@rcsb.org
Search API In Real-world Application

Search API powers the search features provided on the RCSB.org
Questions?
Search and Data API Hands-on Teaser

Leveraging RCSB PDB APIs for Bioinformatics Analyses and Machine Learning

October 12th, 2023

Dennis Piehl, Ph.D.
dennis.piehl@rcsb.org
Goal: Get the citation information for all structures of insulin.

Strategy:

1. Search for all structures of the protein, “insulin”

   - XYZ1
   - XYZ2
   - XYZ3
   - XYZ4
   - XYZ5
   - ...

2. Get citation data for each structure

   - Citation 1
   - Citation 2
   - Citation 3
   - Citation 4
   - Citation 5
   - ...

RCSB.org
Search and Data API: Example Pipeline

[Live Demo]
Goal: Get the citation information for all structures of insulin.

Strategy:

1. Search for all structures of the protein, “insulin”

   Search API query: https://tinyurl.com/4d6hbptj

2. Get citation data for each structure

   Data API query: https://tinyurl.com/mr28xyan
Additional Tools and Modes of API Access

- Command-line tools, such as `curl`:
  ```bash
curl -X GET "https://search.rcsb.org/rcsbsearch/v2/query?json=..."
  ```

- Programming libraries, such as Python, e.g.:
  - Python’s `requests` module:
    ```python
    requests.get("https://search.rcsb.org/rcsbsearch/v2/query?json=...")
    ```
  - Python’s GraphQL module: [https://pypi.org/project/python-graphql-client/](https://pypi.org/project/python-graphql-client/)
    ```python
    client = GraphqlClient(endpoint="https://data.rcsb.org/graphql")
    ```

- [NEW] RCSB PDB Search API Python package: [rcsbsearchapi.readthedocs.io](https://rcsbsearchapi.readthedocs.io)
RCSB PDB Search API Python Package

- [NEW] RCSB PDB Search API Python package: rcsbsearchapi.readthedocs.io
  - Access search API via Python interface
  - Install from PyPI or GitHub (https://github.com/rcsb/py-rcsbsearchapi)

- QuickStart tutorial: rcsbsearchapi.readthedocs.io/en/latest/quickstart.html

```python
from rcsbsearchapi.search import TextQuery
from rcsbsearchapi import rcsb_attributes as attrs

# Create terminals for each query
q1 = TextQuery("heat-shock transcription factor")
q2 = attrs.rcsb_struct_symmetry.symbol == "C2"
q3 = attrs.rcsb_struct_symmetry.kind == "Global Symmetry"
q4 = attrs.rcsb_entry_info.polymer_entity_count_DNA >= 1

# combined using bitwise operators (&, |, ~, etc)
query = q1 & (q2 & q3 & q4)

# Call the query to execute it
for assemblyid in query("assembly"): print(assemblyid)
```
Register for Part 2: Hands-on APIs

Offered at two different times:

- **October 19** 16:00 - 18:30 UTC
  (12:00 - 2:30 PM EDT / 9:00 - 11:30 AM PDT)
- **October 27** 00:00 - 02:30 UTC
  (**October 26** 8:00 - 10:30 PM EDT / 5:00-7:30 PM PDT)

Requirements for participation:

- Registration (form will be sent to today’s participants): *Fill out exit survey!*
- Familiarity with Python basics
- Google account (for accessing a Google Colab notebook)
- Questions and real use cases that you wish to investigate

Space is limited; Zoom link will be provided to accepted participants
Part 2 Teaser: Hands-on ML/AI use case

• Learn how to create a dataset to use for training ML/AI models

• Use case will focus on predicting protein-protein binding sites:
  ○ Search for hetero-dimer complexes
  ○ Use biological features/annotations from Data API to remove redundancy and/or split between training and testing sets

• Explore how to map positional features onto structures:
  ○ Secondary structure
  ○ Binding site residues
RCSB PDB Team

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

Management

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

Follow us

John D. Westbrook
In memoriam
1957-2021
OPPORTUNITIES for
SCIENTIFIC SOFTWARE DEVELOPER
Undergraduates and Graduates

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.

Visit www.rcsb.org/pages/jobs for more information

- DevOps Developer (Rutgers)
- Database Programmer (Rutgers)
- Postdoctoral Researchers (Rutgers, UCSD)
- Gap Year Opportunities (Rutgers)
- Undergraduate Summer Research (Rutgers)
Thank You for Joining Us Today