



KBase

PREDICTIVE BIOLOGY

DOE Systems Biology Knowledgebase

Review of Crash Course Objectives and Arabidopsis Case Study

Christopher S. Henry

November 10th, 2022

RUTGERS



U.S. DEPARTMENT OF
ENERGY

Office of
Science

Office of Biological and Environmental Research



INTEGRATION and
MODELING *for*
PREDICTIVE BIOLOGY

Course Objectives

1. Teach those who are new to PDB or KBase a little more about these platforms
2. Highlight new tools that have been added to KBase to enhance connections to PDB
3. Demonstrate how these tools can be applied to discover new functions for genes in plant and microbial genomes
4. Generally improve audience knowledge in key tools in KBase (modeling, genomics, annotation) and PDB (mol*, structure query, structure alignment)
5. Obtain your feedback on the tools we have built so far and your suggestions on what we should build next



U.S. DEPARTMENT OF
ENERGY

Office of Science



KBase
PREDICTIVE BIOLOGY

DOE Systems Biology Knowledgebase

Course Outline

1. Burley - Introduction
2. Henry - Crash Course Objectives and Arabidopsis Case Study
3. Edirisinghe - Microbial Case Study Involving Pyridine
4. Piehl - Studying Experimental Structures in PDB - Arabidopsis resistosome
5. Vallat - Exploring tools and data in PDB to aid in function discovery
6. Dutta - Detailed tutorial on visualizing structure data in PDB
7. Zhang - Overview of all structure tools currently in KBase
8. Henry - Exploring other information to be gained from PDB tools in KBase and discussing future plans
9. Stephen - Future plans and concluding remarks



U.S. DEPARTMENT OF
ENERGY

Office of Science

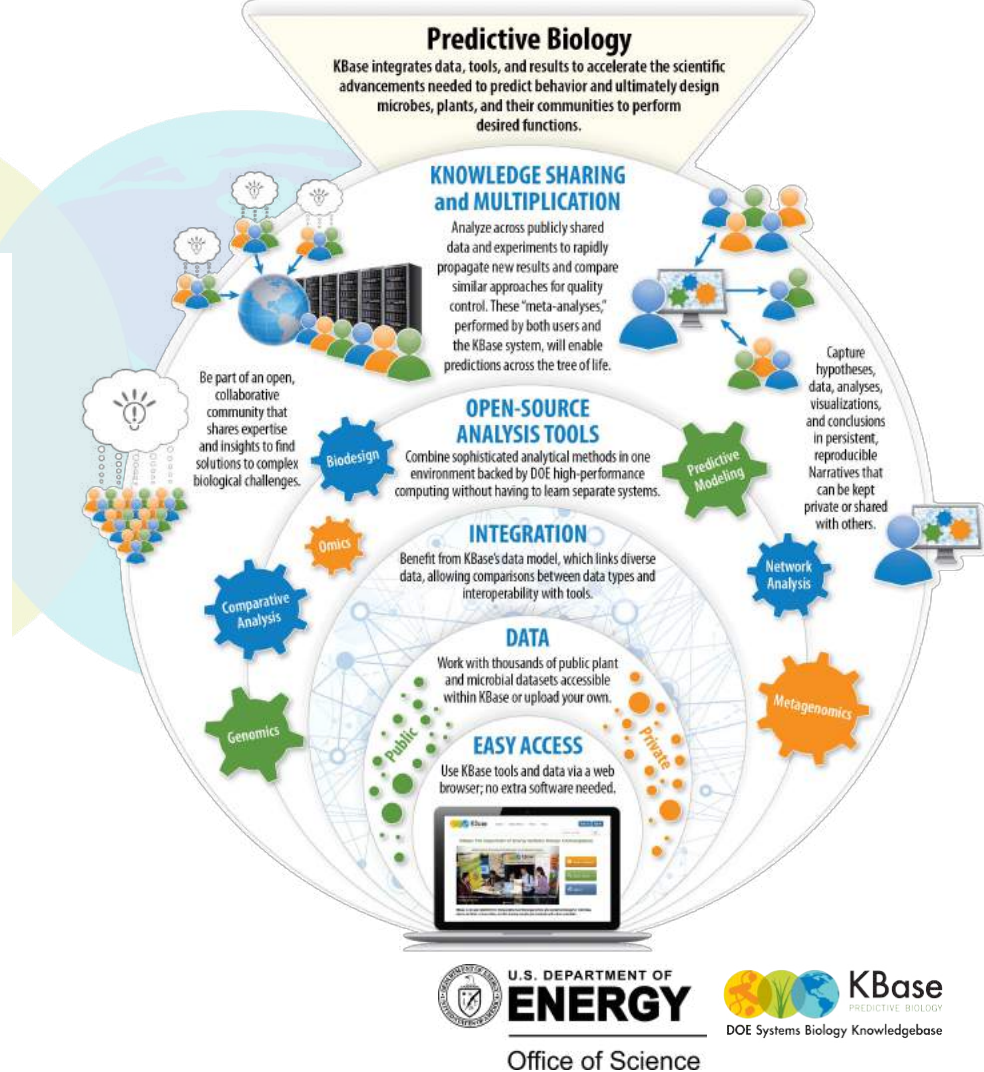


KBase
PREDICTIVE BIOLOGY

DOE Systems Biology Knowledgebase

What is KBase?

KBase is an integrated platform for aggregating and sharing tools and data in order to collaboratively solve scientific problems



What is KBase?

Data Management

- Unlimited data storage
- Import data by
 - Drag & drop
 - Globus
 - FTP, HTTP, Gdrive, Box/Dropbox
- File types
 - FASTQ, FASTA, SRA, GenBank, gff, expression matrix, media, phenotype, FBA models

The screenshot displays the KBase web interface in a browser window. The URL is <https://narrative.kbase.us/narrative/ws.27609.obj.1>. The page title is "Untitled" and it was created by "Paramvir Dehal (psdehal)". The interface has a top navigation bar with tabs: "Analyze", "Narratives", "My Data", "Shared With Me", "Public", "Example", "Import", and "Staging (Beta)". The "My Data" tab is currently selected.

Under the "DATA" section, it states "This Narrative has no data yet." with a red "Add Data" button. To the right, there is a large dashed box for uploading data, with instructions: "Drag and drop data files (or click) in this box to upload them to your staging area." and "(Still in development -- more upload types will be added soon.) Click the ? below for help."

Below the upload area, it says "Or upload to this staging area by using Globus Online" with a link to "psdehal". A search bar is present.

A table lists data files in the staging area:

Name	Size	Modified	Import As...
12042.2.235334.CCAGTGT-AACACTG.fastq.gz	688.56 MB	21 hours ago	Select a format
.12042.2.235334.CCAGTGT-AACACTG.fastq....	12.48 KB	21 hours ago	Select a format
12074.5.235993.ATCTCAG-TATCCTC.filter-IS...	3.58 KB	14 days ago	Select a format
.12074.5.235993.ATCTCAG-TATCCTC.filter-IS...	10.02 KB	14 days ago	Select a format
12073.3.235898.CAGAGTG-ACACTCT.filter-IS...	357.34 MB	26 days ago	Select a format
.12073.3.235898.CAGAGTG-ACACTCT.filter-IS...	10.09 KB	26 days ago	Select a format
83908.assembled.fna	203.18 KB	28 days ago	Select a format
.83908.assembled.fna.jgi	5.19 KB	28 days ago	Select a format

At the bottom, there is an "APPS" section with a list of categories and their counts:

- My Favorites: 2
- Annotation: 7
- Assembly: 26
- Communities: 9
- Comparative Genomics: 34
- Expression: 33
- Metabolic Modeling: 15
- Reads: 12
- Sequence: 14
- Uncategorized: 2
- Upload: 5
- Util: 13

What is KBase?

Data Analysis

• App Catalog

- ~280 apps
- Read processing
- Genome assembly
- Genome annotation
- Sequence alignment
- Comparative genomics
- Metabolic modeling
- Expression
- Microbial communities

The screenshot displays the KBase App Catalog interface. On the left is a sidebar with navigation icons: a hamburger menu, a magnifying glass, and icons for 'Navigator', 'Orgs', 'Catalog', 'Search', 'Jobs', 'Account', and 'Feeds'. The main content area is titled 'Genome Annotation' and features a grid of application cards. Each card includes an icon, the app name, a brief description, the version number, the developer's name, a star rating, and a download count. The apps listed include Prokka, RAST, DRAM, vConTACT2, Domains, and OrthoFinder, among others.

App Icon	App Name	Description	Version	Developer	Stars	Downloads
Prokka	Annotate Assembly and Re-annotate Genomes with Prokka	ProkkaAnnotation	v1.14.5	cherry, olson	61	15026
RAST	Annotate Microbial Assembly with RASTik	RAST_SDK	v1.073	cherry, olson	54	56349
RAST	Annotate Microbial Genome with RASTik	RAST_SDK	v1.073	cherry, olson	25	8077
DRAM	Annotate and Distill Assemblies with DRAM	kb_DRAM		michael, shaffer	16	
Classify	Classify Microbes with GTDB-Tk	kb_gtdbtk	v1.7.0	donovan, parks, aaronmussig, +3 more	15	
vConTACT2	vConTACT2	vConTACT2	0.9.19	bybitolovic	14	113
Domains	Annotate Domains in a Genome	DomainAnnotation		jmc, panovchikov, +1 more	13	1932
Prokka	Annotate Metagenome Assembly with Prokka	ProkkaAnnotation	v1.14.5	by sebras	13	275
RAST	Annotate Genome/Assembly with RASTik	RAST_SDK	v1.073	by cherry, olson, +1 more	11	5
RAST	Annotate Multiple Microbial Genomes with RASTik	RAST_SDK	v1.073	by cherry, olson	11	932
RAST	Annotate Multiple Microbial Assemblies with RASTik	RAST_SDK	v1.073	by landini	9	1202
Domains	Annotate Domains in a GenomeSet	kb_phylogenomics		by dylan, jmc, +2 more	8	859
KB	Build AssemblySet	kb_SetUtilities	v1.0.1	by dylan	6	323
OrthoFinder	Annotate Plant Enzymes with OrthoFinder	kb_orthofinder		by seaver	5	95
DATA	Batch Import Genome from Staging Area	kb_uploadmethods		by tgu2	5	191
RAST	Annotate Metagenome Assembly and Re-annotate Metagenome with RASTik	RAST_SDK	v1.073	by sebras	4	275
KB	Batch Create Assembly Set	kb_SetUtilities	v1.2.0	by dylan	4	108
Compare	Compare Metabolic Annotations	MergeMetabolicAnnotations		by kimbrel, cherry, +1 more	4	62
Import	Import Annotations from Staging	MergeMetabolicAnnotations		by kimbrel, cherry, +1 more	4	165
DRAM	Annotate and Distill Genomes with DRAM	kb_DRAM		by michael, shaffer	3	
DRAM	Annotate and Distill Viral Assemblies with DRAM-v	kb_DRAM		by michael, shaffer	3	
Plant	Annotate Plant Transcripts with Metabolic Functions	kb_plant_rast		by seaver	3	98
RAST	Bulk Annotate Genomes/Assemblies with RASTik	RAST_SDK	v1.073	by sebras	2	7
Import	Bulk Import Annotations from Staging	MergeMetabolicAnnotations		by kimbrel, cherry, +1 more	2	9

What is KBase?

Collaboration

- As users conduct their work within notebooks called “narratives”, they can share those narratives with other users
- Users can also gather their “narratives” into “Organizations”, or groups of users operating with a common mission in KBase

The screenshot displays the KBase 'Organizations' page. On the left is a sidebar with navigation icons for Navigator, Orgs, Catalog, Search, Jobs, Account, and Feeds. The main area features a search bar and a list of organizations. Each organization entry includes a logo, name, description, owner, creation date, and statistics for members, documents, and recent activity.

Organization	Owner	Created	Members	Documents	Recent Activity
ENIGMA Ecosystems and Networks Integrated with Genes and Molecular Assemblies SFA <i>Microbial Ecology, Metagenomics, Systems Biology, Microbiology, OMICS, genetics, modeling</i>	Adam Paul Arkin	Jan 7, 2019	100	77	27 days ago
KBase Educators Using KBase in the classroom and lab	Zach Crockett	Mar 16, 2020	164	59	Sep 7, 2022
NC State Elementary Genetics Laboratory Fall 2022 <i>sourdough metagenomics, osmotic stress</i>	Claire L. Gordy	Sep 26, 2022	34	47	2 days ago
Published MAG/SAG genome database Open Data. Microbial Ecology, Genomics and Metagenomics.	Sean Jungbluth	Apr 5, 2019	8	43	Mar 18, 2022
Boca Raton High School <i>Metagenomics</i>	Jonathan Benskin	May 28, 2020	25	34	Jan 13, 2021

URL: <https://docs.kbase.us/getting-started/narrative/orgs>

What is KBase?

Narrative

- Object-oriented list of data entities
- Point and click app panel
- Notebook with analyses and markdown cells
- Buttons to share with users, make public, publish to HTML, write code cells

KBase KBase PDB workshop, November 10th 2022
Created by: Janaka Edirisinghe (janakabase)

Analyze Narratives Outline

DATA

- DraftModel_MLuteus.pyridine.gf v2
FBAModel
1 day ago by janakabase
- DraftModel_MLuteus.v3
FBAModel
1 day ago by janakabase
- MLuteus_ATCC_49442_RAST_PROKKA... v2
Genome: MLuteus
1 day ago by janakabase
- PyridineMinimalMedia v2
Media
2 days ago by janakabase
- MLuteus_ATCC_49442_RAST_PROKKA v3
Genome: MLuteus
2 days ago by janakabase
- Pyridine_degradationEscherMap v2
EscherMap
2 days ago by janakabase
- FBA_Mluteus_Pyridine_Degradation.v1
FBA
2 days ago by janakabase
- DraftModel_MLuteus.pyridine.v7
FBAModel

APPS

- My Favorites
- Comparative Genomics
- Expression
- Genome Annotation
- Genome Assembly
- Host
- Metabolic Modeling
- Microbial Communities
- Read Processing
- Sequence Analysis
- Uncategorized
- Upload
- Utilities
- Virus

Annotation and constructing metabolic models

Sequencing Annotations Metabolic Models

Organisms / Microbiomes

Documentation and a tutorial Narrative on Intro to Metabolic Models
<https://docs.kbase.us/apps/analysis/metabolic-modeling>
<https://narrative.kbase.us/narrative/18302>

1a. Explore the Annotated Genome

Below, the *M. luteus* genome is shown in a genome viewer. This viewer provides a concise, text-based overview of the genome as well as its contigs and genes.

In the Contigs and Features tabs, each entry is clickable, opening either a browser for the contig or another tab with expanded information about the gene. You can sort these entries by clicking on a column header to sort by that field (e.g., Length). Clicking the same column header again will reverse the sort order.

The *M. luteus* genome has two contigs. Click on one to see neighboring genes and potential operons in this species.

To further explore this genome, click on "Browse Features" tab, where you can search for gene annotations/functions by name (e.g. pyruvate synthase, EC numbers etc.), extract DNA or protein sequences, explore the neighboring genes/gene clusters

MLuteus_ATCC_49442
v2 - KBaseGenomes.Genome-10.0

Annotate and Distill Genomes with DRAM
Annotate MAGs with DRAM and distill resulting annotations to create an interactive functional summary per genome. For

Annotate Assembly and Re-annotate Genomes with Prokka - v1.14.5
Annotate Assembly and Re-annotate Genomes with Prokka annotation pipeline.

Output from Annotate Assembly and Re-annotate Genomes with Prokka - v1.14.5

What is KBase?

Documentation

KBase Tutorial Narratives

These Narrative tutorials provide step-by-step examples that show how to use KBase tools and data to perform useful analyses. You can copy these Narratives and rerun the steps, modify the analyses, or try on your own data.

Multi-Omics Modeling Of Biochemical Pathways

EMSL Summer School 2021

Learn how to use visualization tools, analysis, and modeling, of multi-omics data for understanding biochemical pathways in this virtual course from EMSL.

[Course Page →](#)

Searching For Features

Genome Analysis in KBase Series focuses on how to use various tools to analyze genomic sequencing data in KBase. This second Narrative demonstrates a workflow for searching and identifying features within...

[View Tutorial →](#)

Drafting Isolate Genomes

Genome Analysis in KBase Series focuses on how to use various tools to analyze genomic sequencing data in KBase. This first Narrative shows how to draft isolate genomes in...

[View Tutorial →](#)

Multiscale Microbial Dynamics Modeling

EMSL Summer School 2020

Learn how to incorporate microbial metagenomic and environmental metabolite data from watershed ecosystems into metabolic and community modeling using computational frameworks in this virtual course.

KBase For Educators

Made by Ellen Dow

Join the community of educators using KBase to teach bioinformatics and computational biology!

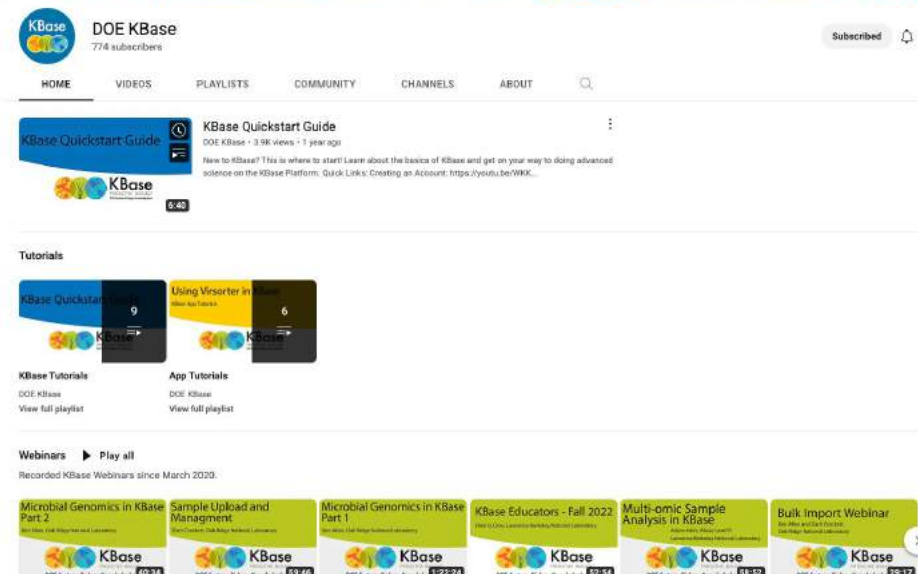
Why KBase?

While KBase is routinely used for advanced analysis by scientists around

Assembly And Annotate Prokaryotic Genome Tutorial

Made by the KBase Documentation team

KBase provides multiple Apps for *de novo* assembly of prokaryotic Next-Generation Sequencing (NGS) reads from various sequencing platforms. These assemblies can then be annotated with RAST or Prokka, enabling you to...



Many workflow specific tutorials and hundreds of videos on KBase youtube channel



U.S. DEPARTMENT OF
ENERGY

Office of Science



DOE Systems Biology Knowledgebase

Why KBase and PDB?

1. There is enormous and rapidly growing synergy between KBase and PDB
2. KBase contains extensive genomics data, particular non-reference user data
3. KBase has numerous workflows to integrate multi-omics data and comparative genomics approaches to predict and understand protein function
4. Advances in protein folding now make it possible to easily obtain a predicted structure for proteins of interest discovered in KBase, but what next?
5. Here PDB steps in offering a large database of experimental structures to aid in contextualizing new predicted structures generated for genes in KBase
6. PDB has a growing body of structure-related tools and data to enable users to actually gain functional insights from structure data



U.S. DEPARTMENT OF
ENERGY



DOE Systems Biology Knowledgebase

Office of Science

Case study in Arabidopsis...



U.S. DEPARTMENT OF
ENERGY

Office of Science



KBase
PREDICTIVE BIOLOGY

DOE Systems Biology Knowledgebase

Acknowledgements



Stephen Burley



John Westbrook



Ed O'loughlin



Qizhi Zhang



Kelly Skinner



Jack Gilbert



Sam Seaver



Claudia Lerma-Ortiz



Nidhi Gupta

KBase team



We thanks DOE BER for funding this work!

Special thanks to all those who attended and contributed to our KBase-PDB design workshops!



KBbase

PREDICTIVE BIOLOGY

DOE Systems Biology Knowledgebase

Protein Candidates from Function Queries in KBbase

Janaka N. Edirisinghe, Ph.D

November 10th, 2022

RUTGERS



U.S. DEPARTMENT OF
ENERGY

Office of
Science

Office of Biological and Environmental Research



INTEGRATION and
MODELING *for*
PREDICTIVE BIOLOGY

Identifying a Novel degradation Pathway with KBase Discovery Pipeline and PDB tools

Scientific Problem

- There are many microbial transformations yet to be characterized. How should we computationally predict potential gene candidates for novel enzymatic transformations?
- How can we effectively use RCSB data in aid in identifying predicted high confidence gene candidates?



U.S. DEPARTMENT OF
ENERGY

Office of Science



DOE Systems Biology Knowledgebase

Science example: Discover potential novel degradation pathway genes for heteroaromatic compound Pyridine in *Micrococcus letus* using KBase Functional Discovery Pipeline

- Phenotypic data such as biology/growth experiments often show that microbes can degrade certain nutrients, but the degradation pathway is either poorly characterized or unexplained.
- Scientists at ANL identified a *Micrococcus letus* strain that could degrade pyridine, but the pathway was unknown. The degradation pathway wasn't explained until very recently
- We will apply KBase Functional Discovery Pipeline and knowledge that derives from RCSB data to identify potential pyridine degradation enzymes in *M.letus*



U.S. DEPARTMENT OF
ENERGY

Office of Science



KBase
PREDICTIVE BIOLOGY

DOE Systems Biology Knowledgebase

MaSuRCA MaSuRCA Assembler - v3.2.9
★ kb_MaSuRCA v1.1.3

RAST Annotate Genome/Assembly with RASTtk - v1.073
★ RAST_SDK v1.9.4

Model SEED Build Metabolic Model
★ fba_tools v2.0.0

PickAxe - Generate novel compounds from reaction rules
★ kb_pickaxe v1.3.1
7320 novel reactions

Model SEED Gapfill Metabolic Model Fill gaps with novel and known compounds
★ fba_tools v2.0.0

Escher Pathway Viewer View key reactions and pathways
★ kb_escher v0.0.1

DESeq2 Create Differential Expression Matrix using DESeq2 - v1.20.0
★ kb_deseq v1.1.2
> 20 gene candidates (~15)

PDB - Import PDB Metadata into KBase Genome
★ KBaseAnnotationApps v0.0.1
Query candidate proteins from RSCB few genes

DATA Query RCSB databases for protein structures
★ ProteinStructureUtils v0.0.2

4719 genes

4719 genes

1127 genes
1135 reactions

7320 novel reactions

Fill gaps with novel and known compounds

View key reactions and pathways

> 20 gene candidates (~15)

Query candidate proteins from RSCB few genes

Link: [Base Narrative Workflow](https://narrative.kbase.us/narrative/127880)

(<https://narrative.kbase.us/narrative/127880>)

Identification of strains member that have a biosynthesis/degradation pathway of interest

Succinate-semialdehyde dehydrogenase (NAD(P)+) (EC 1.2.1.16)
L-lactate dehydrogenase (EC 1.1.2.3)
Phosphoenolpyruvate carboxylase (EC 4.1.1.31)
Shikimate kinase I (EC 2.7.1.71)
Biotin: 1,2-dimethylglyoxal beta-lyase

Annotation

Models

$H_2 + \text{Acetoacetate} \rightarrow CO_2$

$H_2O + \text{Methanol} \rightarrow H_2O + \text{Formaldehyde}$

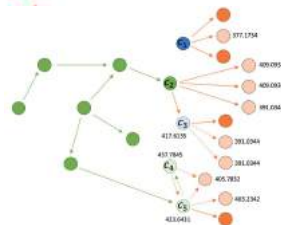
$NADH + H^+ + \text{Butanol} \rightarrow NAD + H^+$



Metabolic pathways

Cheminformatics

Expansion from known biochemistry into novel compounds



Metabolomics



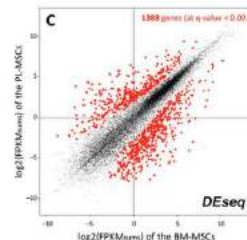
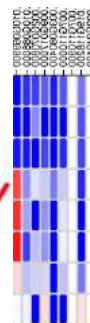
Potential genes candidates

Structure Analysis



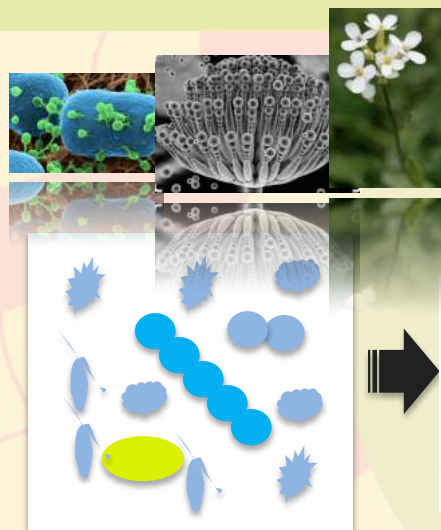
Experimental verification

KBase Functional Discovery Pipeline



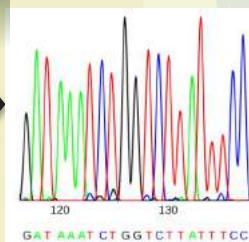
Transcriptomics

Annotation of genomes and constructing metabolic models



Organisms / Microbiomes

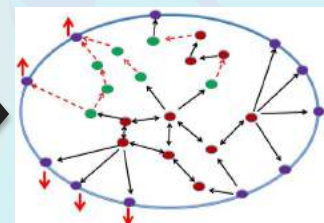
Sequencing



Annotations



Metabolic Models



U.S. DEPARTMENT OF
ENERGY

Office of Science



KBase
PREDICTIVE BIOLOGY

DOE Systems Biology Knowledgebase

Model SEED Build Metabolic Model
Generate a draft metabolic model

Model SEED Build Metagenome Metabolic Model
Generate a draft metabolic model based on an annotat

Fungal Model Build Fungal Model
Allows users to build a Fungal metabolic model.

RAST Rapid Annotation using
Subsystem Technology

Documentation and a tutorial Narrative on Intro to Metabolic Models

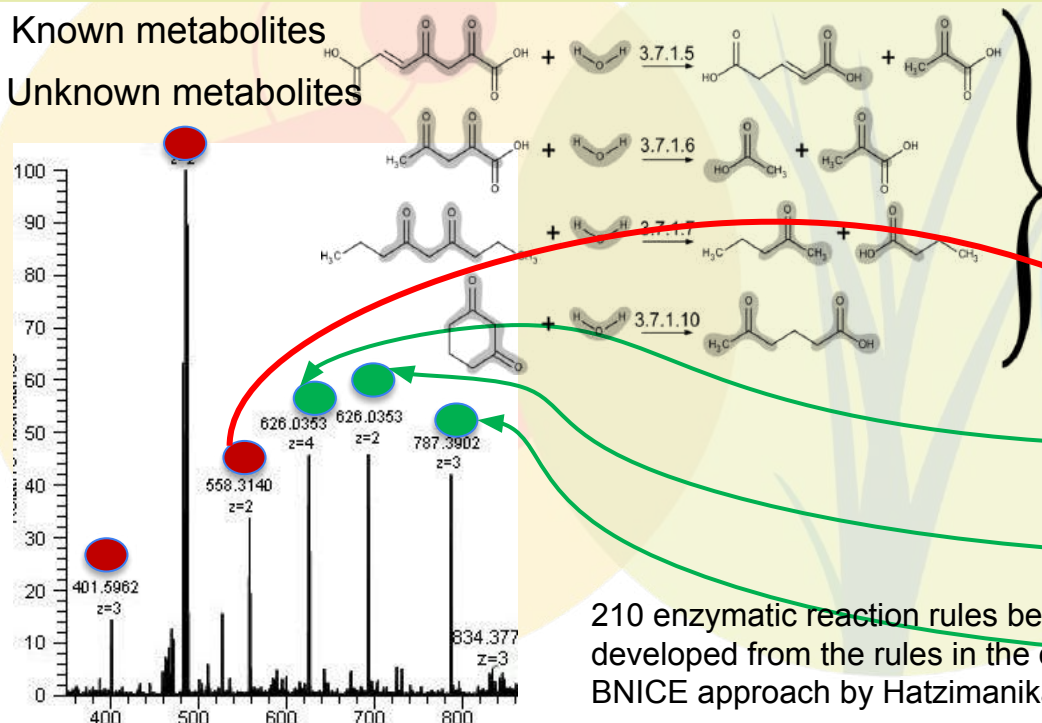
<https://docs.kbase.us/apps/analysis/metabolic-modeling>

<https://narrative.kbase.us/narrative/18302>

Introducing new biochemistry in filling metabolic gaps

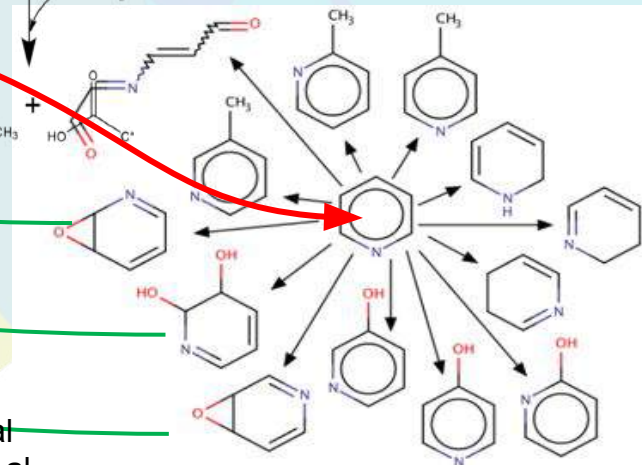
PickAxe app – Based on biochemical rules can be derived from known biochemistry is used to predict novel or promiscuous enzymatic reactions

- Known metabolites
- Unknown metabolites



210 enzymatic reaction rules been developed from the rules in the original BNICE approach by Hatzimanikatis et al

Biochemistry Databases



PickAxe - Generate novel compounds from reaction rules

Generate novel compounds based enzymatic and spontaneous reaction rules



U.S. DEPARTMENT OF
ENERGY

Office of Science



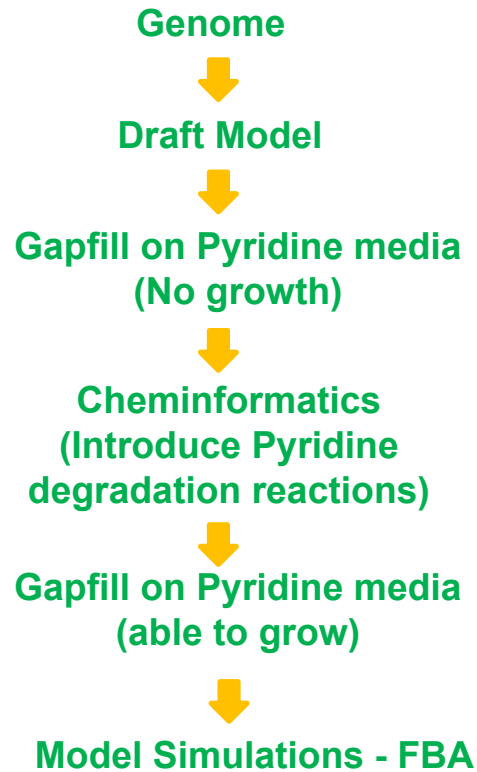
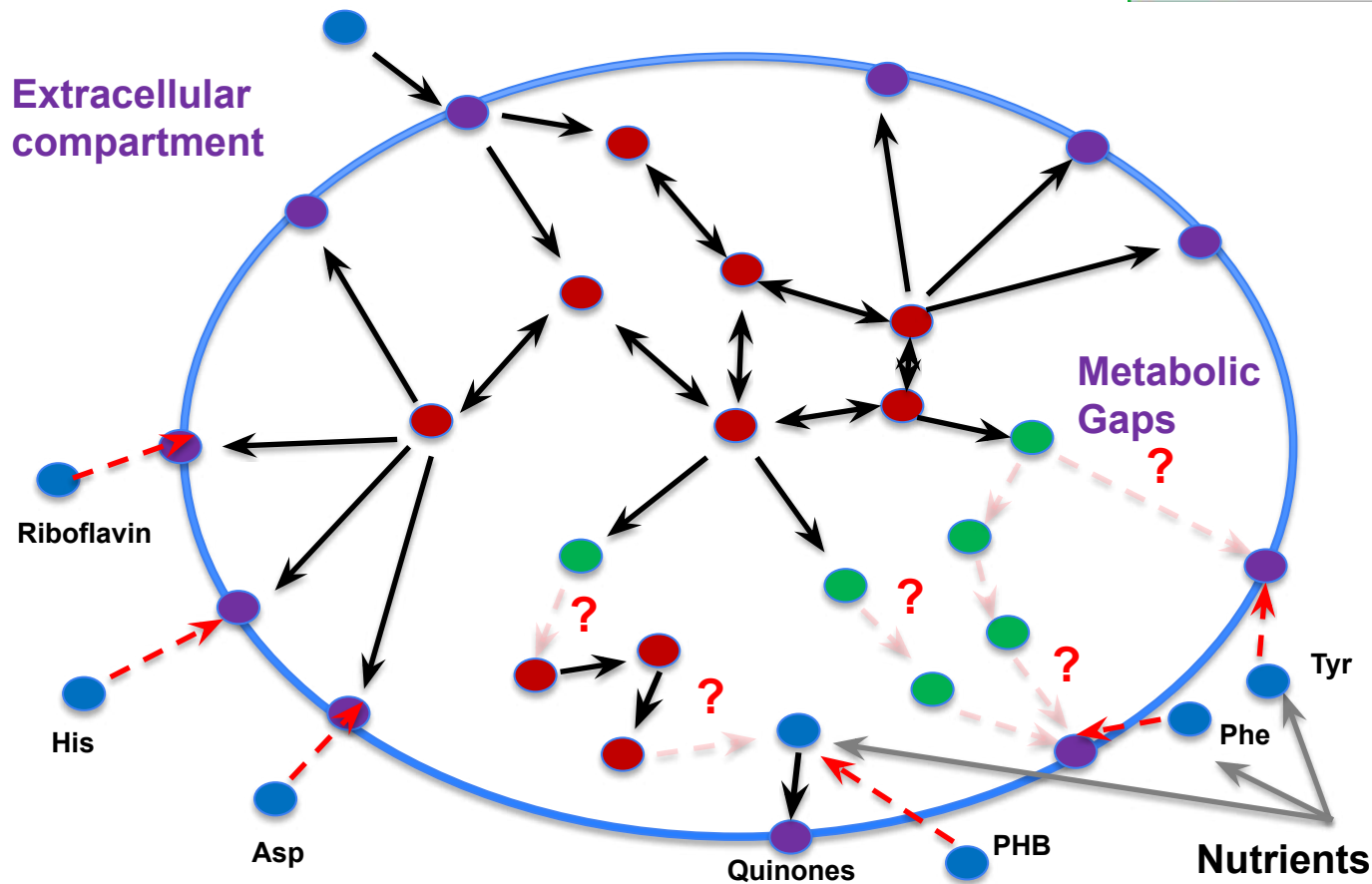
KBase
PREDICTIVE BIOLOGY

DOE Systems Biology Knowledgebase

Filling gaps in a metabolic network



Gapfill Metabolic Model
Identify the minimal set of biochem



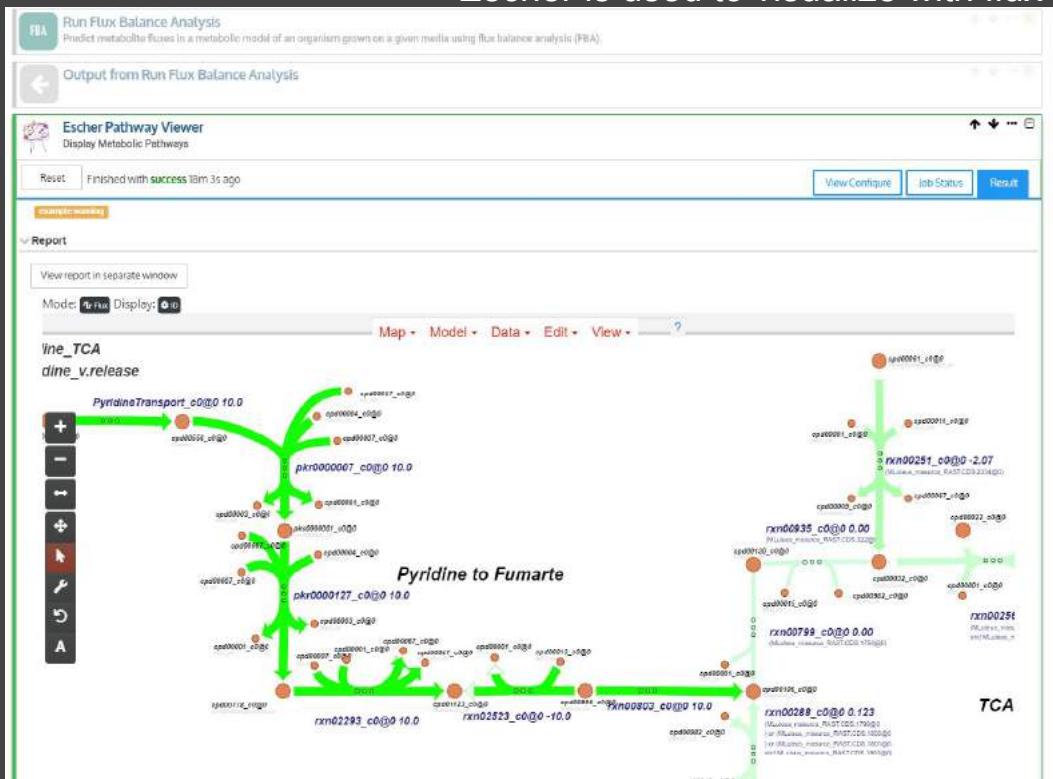
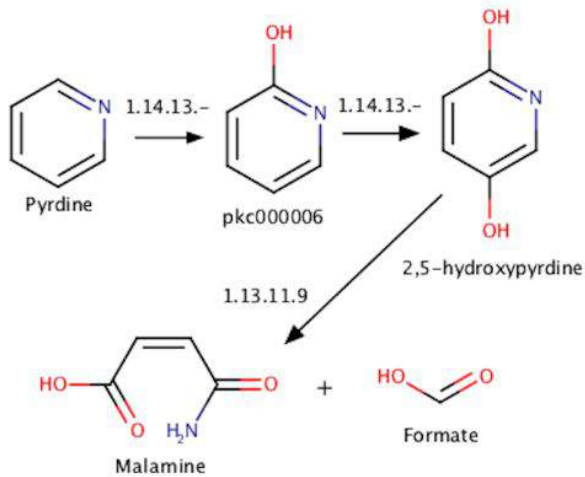
Discovering genes involved in pyridine degradation

Using cheminformatics to predict a novel pathway

Metabolic modeling workflow in KBase is used to predict the most likely pathway, then Escher is used to visualize with flux

PickAxe - Generate novel compounds from reaction rules
Generate novel compounds based enzymatic and spontaneous reaction rules

Reaction	Name	Equation	Genes	Gapfilling
pkc000000[0][c0]	pkc000000[0][c0]	Pyridine[c0] + H[c0] + NADH[c0] => NAD[c0] + pkc000000[0][c0]		
pkc0000002[c0]	pkc0000002[c0]	NADH[c0] + H[c0] + O2[c0] + Pyridine[c0] => pkc00000006[c0] + H2O[c0] + NAD[c0]		
pkc0000003[c0]	pkc0000003[c0]	O2[c0] + Pyridine[c0] => pkc00000008[c0]		
pkc0000004[c0]	pkc0000004[c0]	O2[c0] + H[c0] + NADH[c0] + Pyridine[c0] => pkc00000005[c0] + NAD[c0]		
pkc0000005[c0]	pkc0000005[c0]	S-Adenosyl-L-methionine[c0] + Pyridine[c0] => S-Adenosyl-homocysteine[c0] + H[c0] + pkc00000010[c0]		



Linking cheminformatics to genes in the genome



PickAxe - Generate novel compounds from reaction rules

Generate novel compounds based enzymatic and spontaneous reaction rules

PyridineNovelReactions

v2 - KBaseFBA.FBAModel-14.0

Overview

Reactions

Compounds

Genes

Compartments

Biomass

Gapfilling

Pathways

enzc274_c0 ✕

Show 10 entries

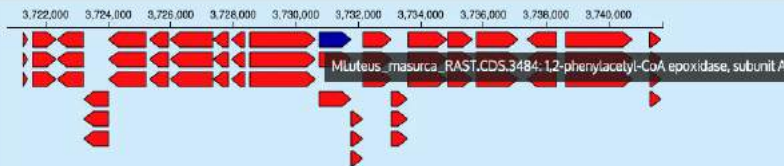
Reaction	Name	Equation
enzr106[c0]	1.14.13_10_14	2-Pyridone[c0] + O2[c0] + NADH[c0] + H+[c0] => dihydroxypyridine[c0] + H2O[c0] + NAD[c0]
enzr10[c0]	2.1.1_02_3	Pyridine[c0] + 5-Adenosyl-L-methionine[c0] => e 5-Adenosyl-homocysteine[c0] + H+[c0]
enzr1148[c0]	1.14.13_10_136	2,5-Dihydroxypyridine[c0] + O2[c0] + NADH[c0] => enz15800[c0] + H2O[c0] + NAD[c0]
enzr118[c0]	1.3.1_03;1.3.1_05_1	enzc3[c0] + NAD[c0] => 2,3-dihydroxypyridine[c0] + NADH[c0] + H+[c0]
enzr11[c0]	1.14.13_08_1	Pyridine[c0] + O2[c0] + NADH[c0] + H+[c0] => e H2O[c0] + NAD[c0]
enzr12[c0]	1.14.13_08_2	Pyridine[c0] + O2[c0] + NADH[c0] + H+[c0] => e H2O[c0] + NAD[c0]
enzr13[c0]	1.14.13_10_3	Pyridine[c0] + O2[c0] + NADH[c0] + H+[c0] => enz7[c0] + H2O[c0] + NAD[c0]
enzr14[c0]	1.3.1_05.rev_3	Pyridine[c0] + NADH[c0] + H+[c0] <=> enz10[c0] + NAD[c0]

Feature ID ▾	Type	Function	Ontology
MLuteus_masurca_RAST.CDS.3484	gene	1,2-phenylacetyl-CoA epoxidase, subunit A	ec:1.14.13.149- Phenylacetyl-CoA 1,2-epoxidase.

Product Function	Aldehyde dehydrogenase (EC1.2.1.3), PaaZ
Function Descriptions	
Ontology Terms	None
Location	Contig: scf7180000000004 3,728,527 - 3,730,639 (+ Strand)
Feature Context	



Ontology Terms	ec:1.14.13.149- Phenylacetyl-CoA 1,2-epoxidase.
Location	Contig: scf7180000000004 3,730,746 - 3,731,772 (+ Strand)
Feature Context	



U.S. DEPARTMENT OF
ENERGY

Office of Science

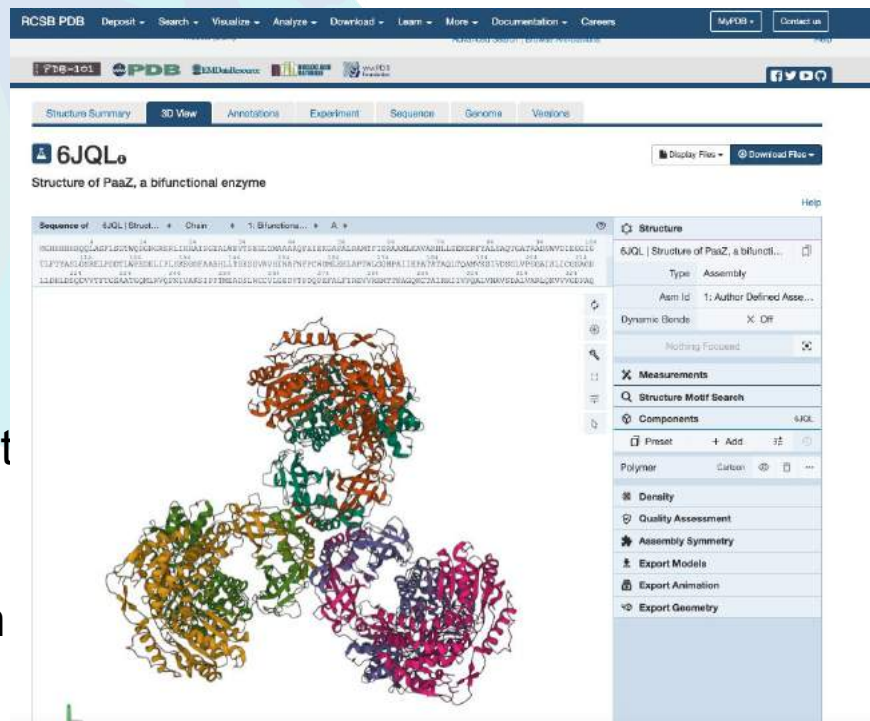


KBase
PREDICTIVE BIOLOGY

DOE Systems Biology Knowledgebase

Structure prediction and analysis on selected gene candidates by the PDB team – example gene: CDS. 3483

- Query experimentally resolved structures that are corresponding to the potential gene candidates
- Query and learn from co-crystallized structures with the docking of the substrate
- Align experimental and computational structures to aid binding site identification and functional characterization



U.S. DEPARTMENT OF
ENERGY

Office of Science



Acknowledgements



Stephen Burley



John Westbrook



Ed O'loughlin



Qizhi Zhang



Kelly Skinner



Jack Gilbert



Sam Seaver



Claudia Lerma-Ortiz



Nidhi Gupta

KBase team

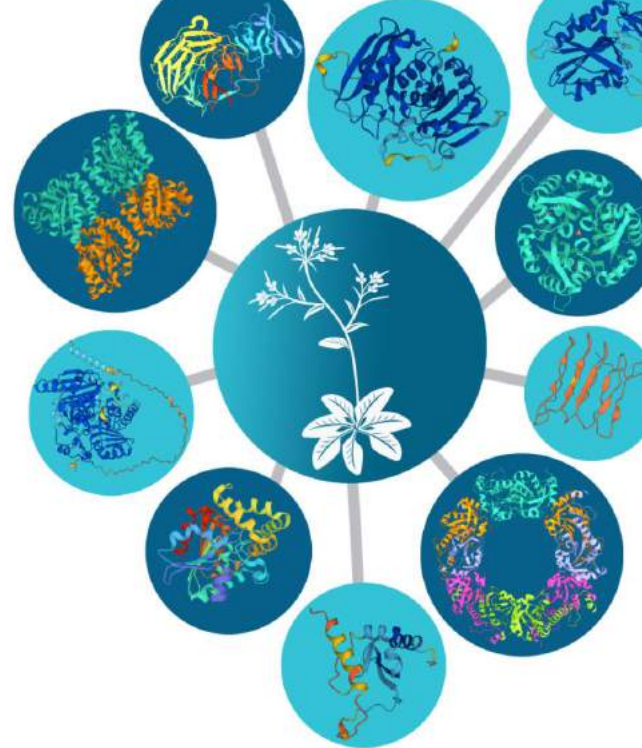


We thanks DOE BER for funding this work!

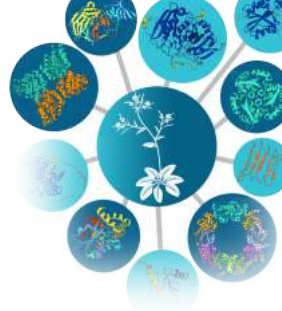
Special thanks to all those who attended and contributed to our KBase-PDB design workshops!

Accessing Experimental Structures from the PDB

Dennis Piehl, Ph.D., RCSB PDB/Rutgers



Outline

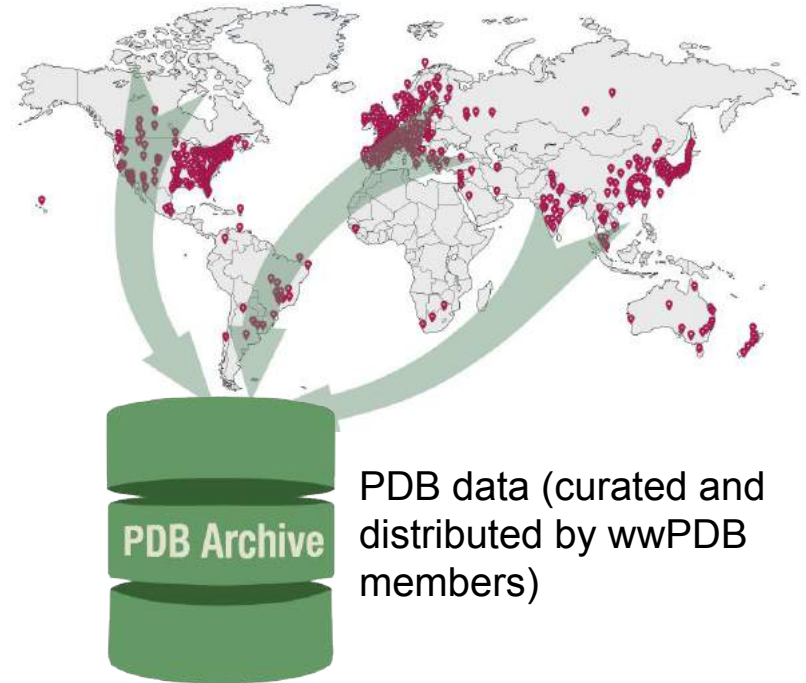


- Introduction to the PDB Archive and RCSB PDB
- Overview of search tools
- Overview of download services and structure data format
- Tutorial: Accessing data from RCSB.org
 - Case study: *Arabidopsis thaliana* resistosome
 - Searching and browsing
 - Structure summary pages
 - Downloading data
 - Programmatic access (searching and downloading)
- Additional resources and documentation

History of the Protein Data Bank (PDB)

- Established in 1971 as the first open-access digital data resource in biology, with just seven protein structures
- Now hosts >190,000 experimental 3D structures of biomolecules, deposited by researchers worldwide
- The PDB Core Archive is jointly managed by the Worldwide PDB (wwPDB), comprised of the RCSB PDB (U.S.), PDBe (Europe), PDBj (Japan), BMRB (U.S./Japan), and EMDB (Europe)
- Committed to making PDB data “FAIR” (Findable, Accessible, Interoperable, & Reusable); all data is made freely available to the public

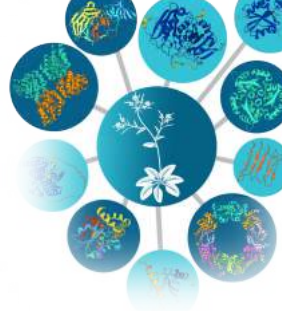
3D structural data from around the world



Protein Data Bank (1971) *Nature New Biology* **233**, 223.

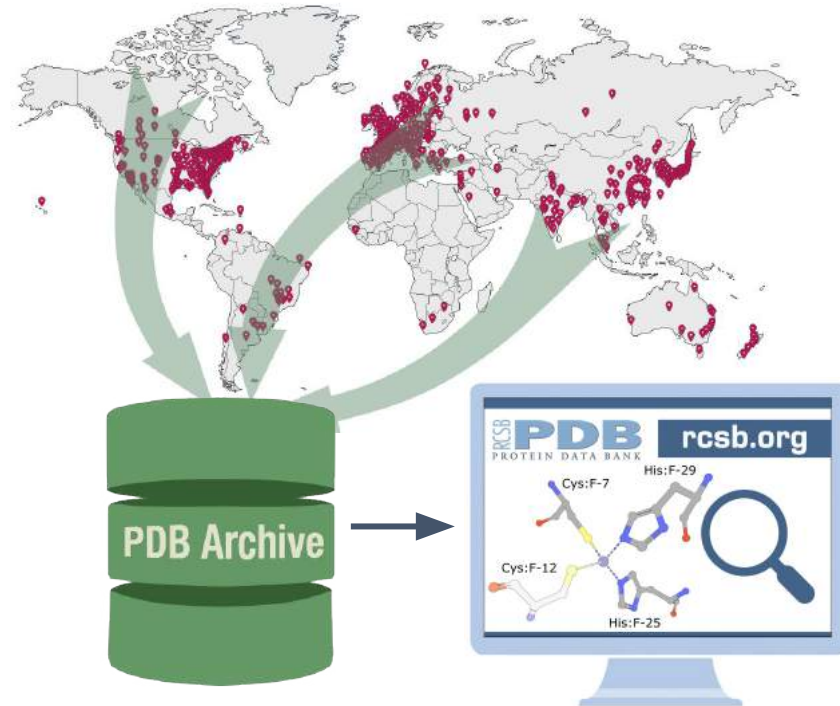
Worldwide Protein Data Bank (2019) *Nucleic Acids Research* **47**, D520–D528.

The RCSB PDB Web Portal (RCSB.org)

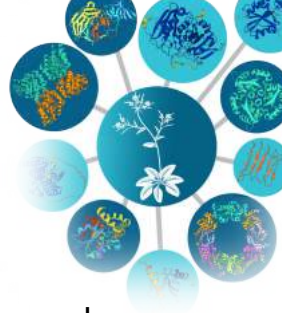


- **RCSB.org:** Tools for searching, visualizing, analyzing and downloading the contents of the PDB Archive
- Integration of PDB Archive data with annotations from ~50 external data resources (UniProt, SCOPe, CATH, ...)
- Recently introduced tools and features:
 - Protein 1D-3D Feature View
 - Groups 1D-3D Alignment View
 - Integration of Computed Structure Models (CSMs)
- **PDB-101 (pdb101.rcsb.org):** Educational resources and training

3D structural data from around the world



Search Tools at RCSB.org



RCSB PDB Deposit Search Visualize Analyze Download Learn More Documentation Careers MyPDB Contact us

RCSB PDB PROTEIN DATA BANK
197,848 Structures from the PDB
1,000,361 Computed Structure Models (CSM)

3D Structures Enter search term(s), Entry ID(s), or sequence Include CSM [toggle] [Search]

Advanced Search | Browse Annotations Help

PDB-101 PDB EMDatabank NUCLEIC ACID DATABASE wwPDB Foundation

f t y d

← Basic search

Search Query History Browse Annotations MyPDB

QUERY: Scientific Name of the Source Organism HAS EXACT PHRASE "Arabidopsis thaliana" JSON MyPDB Login

Advanced Search Query Builder Help

Full Text

Structure Attributes Help

Scientific Name of the Source Organism	x	has exact phrase	Arabidopsis thaliana	+ NOT	Count	x
Add Attribute Add Subquery				Remove Subquery		
Add Subquery						

Chemical Attributes

Sequence Similarity

Sequence Motif

Structure Similarity

Structure Motif

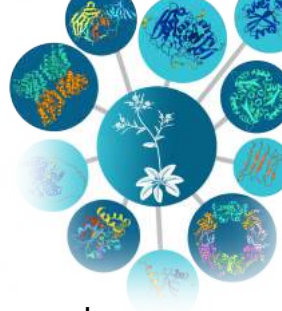
Chemical Similarity

Return Structures grouped by No Grouping

Include Computed Structure Models (CSM) [toggle] Count Clear Search

Advanced search

Search Tools at RCSB.org



RCSB PDB Deposit Search Visualize Analyze Download Learn More Documentation Careers MyPDB Contact us

RCSB PDB PROTEIN DATA BANK 197,848 Structures from the PDB 1,000,361 Computed Structure Models (CSM)

3D Structures Enter search term(s), Entry ID(s), or sequence Include CSM [toggle] [Search]

Advanced Search | Browse Annotations Help

PDB-101 PDB EMDatabank NUCLEIC ACID DATABASE wwPDB Foundation

f t y d

← Basic search

Search Query History Browse Annotations MyPDB

QUERY: Scientific Name of the Source Organism HAS EXACT PHRASE "Arabidopsis thaliana" [JSON] [MyPDB Login]

Advanced Search Query Builder [Help]

Full Text [input]

Structure Attributes [Help]

Scientific Name of the Source Organism	x	has exact phrase	Arabidopsis thaliana	+ NOT	Count	x
Add Attribute Add Subquery				Remove Subquery		
Add Subquery						

Chemical Attributes [input]

Sequence Similarity [input]

Sequence Motif [input]

Structure Similarity [input]

Structure Motif [input]

Chemical Similarity [input]

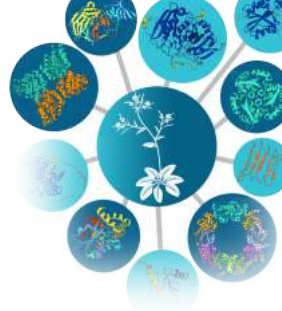
Return Structures [input] grouped by No Grouping [input]

Include Computed Structure Models (CSM) [toggle] Count Clear [Search]

Search API
(search.rcsb.org)

Advanced search

Advanced Search Tools at RCSB.org



▲ Advanced Search Query Builder ?

▼ Full Text ?

▼ Structure Attributes ?

▼ Chemical Attributes ?

▼ Sequence Similarity ?

▼ Sequence Motif ?

▼ Structure Similarity ?

▼ Structure Motif ?

▼ Chemical Similarity ?

- Full-text “Google-like” search
- Structure or entry property (author, method, size, ...)
- Chemical property (name, weight, ...)
- Sequence identity search (MMseqs2)
- Amino acid regular expression (e.g., ST*G)
- Global structure similarity (in-house algorithm)
- Local structural motif (e.g., binding site residues)
- Chemical structure similarity (SMILES, InChI, sketch)

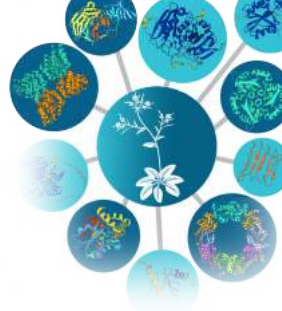
Return ▼ ? grouped by No Grouping ▼ ?

▼ Structures

- Polymer Entities
- Assemblies
- Non-polymer Entities
- Molecular Definitions

→ Group structures by various criteria (sequence, deposit id, ...)

Search API at RCSB.org



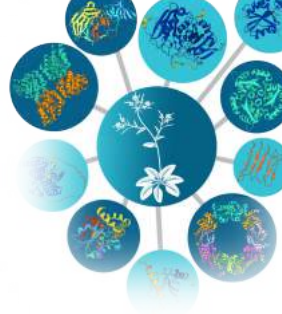
```
← → ↺ search.rcsb.org/query-editor.html?json=%7B"query"%3A%7B"type"%3A"terminal"%2C"label"%3A"text"%2C"service"%3A"text"%2C"parameters"... ☆

RCSB PDB: Search API Query Editor [Icons] [Refresh] [Copy] [Run]

1- {
2-   "query": {
3-     "type": "terminal",
4-     "label": "text",
5-     "service": "text",
6-     "parameters": {
7-       "attribute": "rcsb_entity_source_organism.ncbi_scientific_name",
8-       "operator": "contains_phrase",
9-       "negation": false,
10-      "value": "Arabidopsis thaliana"
11-    }
12-  },
13-  "return_type": "entry",
14-  "request_options": {
15-    "paginate": {
16-      "start": 0,
17-      "rows": 25
18-    },
19-    "results_content_type": [
20-      "experimental"
21-    ],
22-    "sort": [
23-      {
24-        "sort_by": "score",
25-        "direction": "desc"
26-      }
27-    ],
28-    "scoring_strategy": "combined"
29-  }
30- }
```

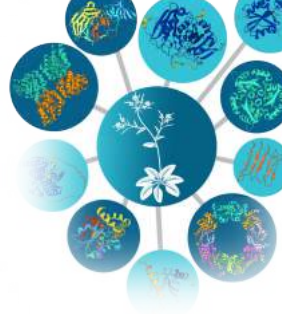
```
{
  "query_id": "330fcd0f-62c0-4b9e-a710-fbe40c9c93d7",
  "result_type": "entry",
  "total_count": 1879,
  "result_set": [
    {
      "identifier": "7AR8",
      "score": 1
    },
    {
      "identifier": "7ARB",
      "score": 1
    },
    {
      "identifier": "6XYW",
      "score": 0.93261934140986
    },
    {
      "identifier": "7AR7",
      "score": 0.843479411762288
    },
    {
      "identifier": "7WG5",
      "score": 0.7614635080689911
    },
    {
      "identifier": "7OUI",
      "score": 0.7074923938322015
    },
    {
      "identifier": "7AQQ",

```



Download Services

- While browsing RCSB.org:
 - Structure summary pages (i.e., structure entry pages)
 - Search results page
 - Bulk download page (<https://www.rcsb.org/downloads>)
- PDB Archive:
 - Over HTTP (via direct navigation on browser or programmatic access):
 - **Structure data under the directory: /pub/pdb/data/structures/divided/mmCIF/*
 - <https://s3.rcsb.org>
 - <https://files.rcsb.org> (supports recursive directory retrieval)
 - Can use command-line/scripting tools (e.g., curl, wget, or Python)
 - Using rsync: `rsync://rsync.rcsb.org`
 - Using FTP: `ftp://ftp.wwpdb.org`



Structure Data Format: PDBx/mmCIF

- PDBx/mmCIF Data Standard (<https://mmcif.wwpdb.org/>)
 - Became the standard file format for data in the PDB Archive in 2014
 - Files are given the extension “.cif” (or “.cif.gz”)
 - Supersedes the now legacy “PDB” (“.pdb”) file format, which is limited by fixed-column width restrictions and is no longer extended to support new types of metadata

Display Files ▾

Download Files ▾

FASTA Sequence

PDBx/mmCIF Format

PDBx/mmCIF Format (gz)

PDB Format

PDB Format (gz)

PDBML/XML Format (gz)

data_6J5T

#

_entry.id 6J5T

#

_audit_conform.dict_name mmcif_pdbx.dic

_audit_conform.dict_version 5.303

_audit_conform.dict_location http://mmcif.pdb.org/dictionaries/ascii/mmcif_pdbx.dic

#

loop_

_database_2.database_id

_database_2.database_code

PDB 6J5T

WWPDB D_1300010525

EMDB EMD-0680

#

_pdbx_database_related.db_name EMDB

_pdbx_database_related.details 'Reconstitution and structure of a plant NLR resistosome conferring immunity'

_pdbx_database_related.db_id EMD-0680

_pdbx_database_related.content_type 'associated EM volume'

#

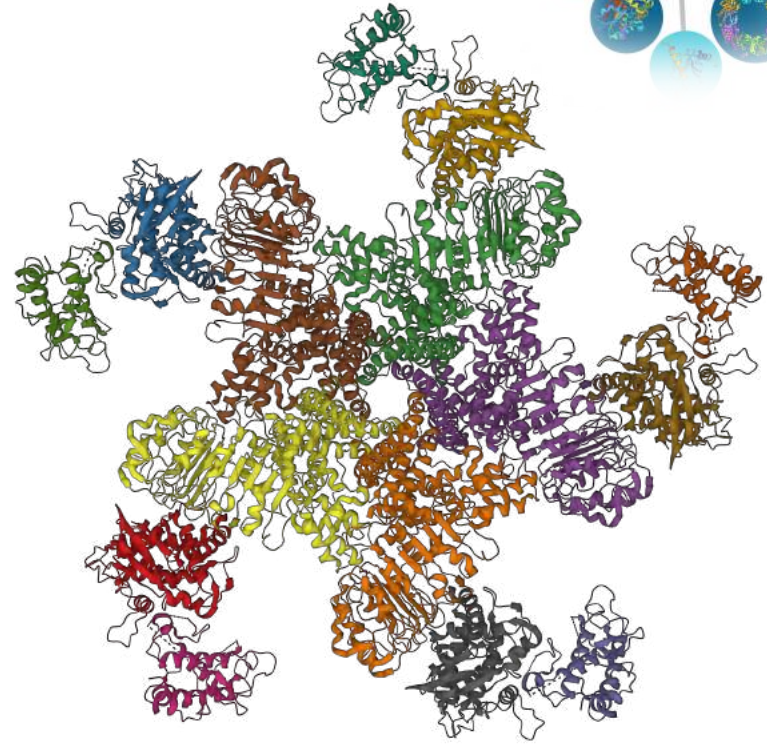
...

→

← Legacy

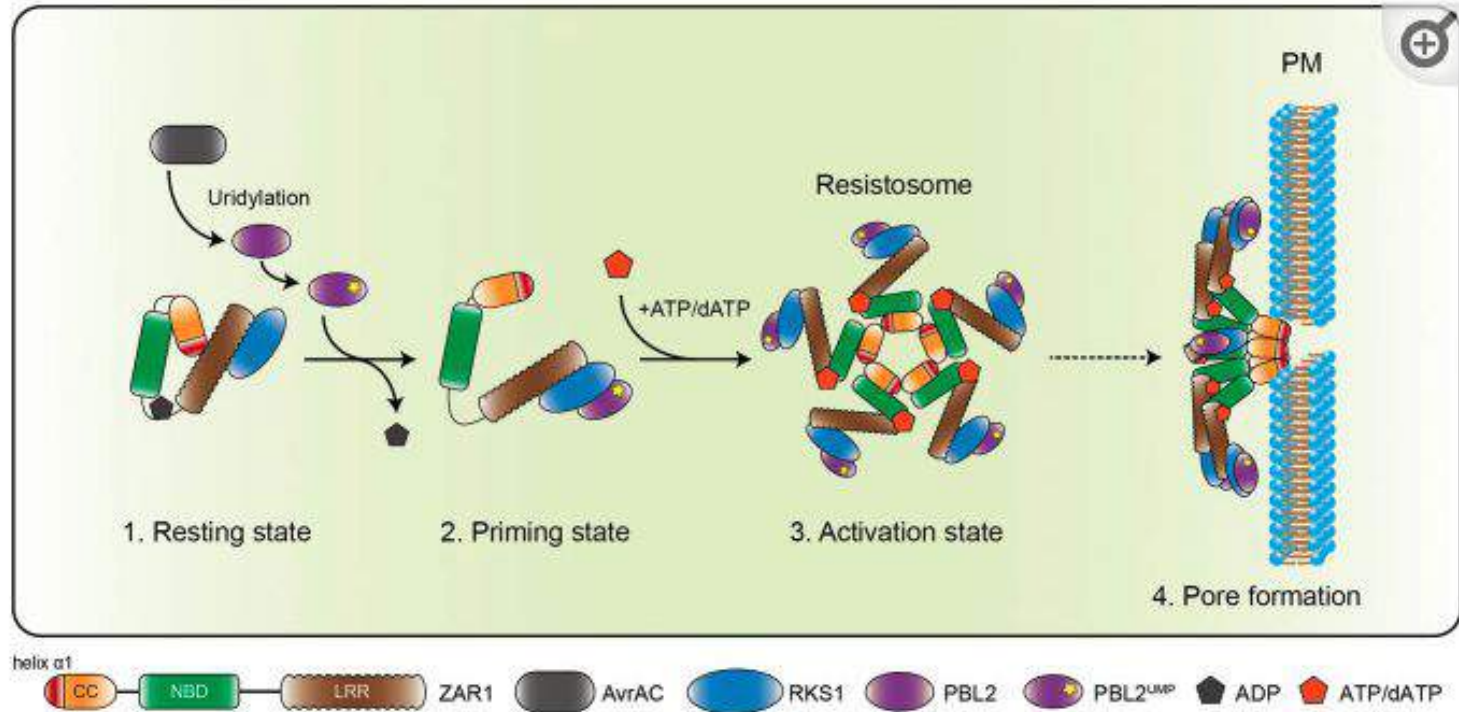
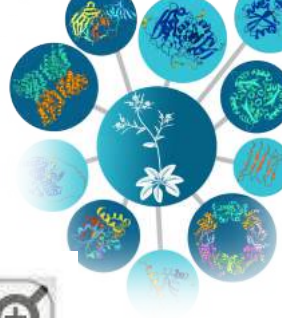
Case study: *Arabidopsis thaliana* resistosome

- *Arabidopsis thaliana*: the gold standard model organism in plant molecular biology
- In plants, the innate immune response is triggered via the sensing of pathogens by nucleotide-binding, leucine-rich repeat receptors (NLRs)
- Upon pathogen sensing, NLRs oligomerize to form the activated “resistosome”
- Ultimately, this process culminates in programmed cell death

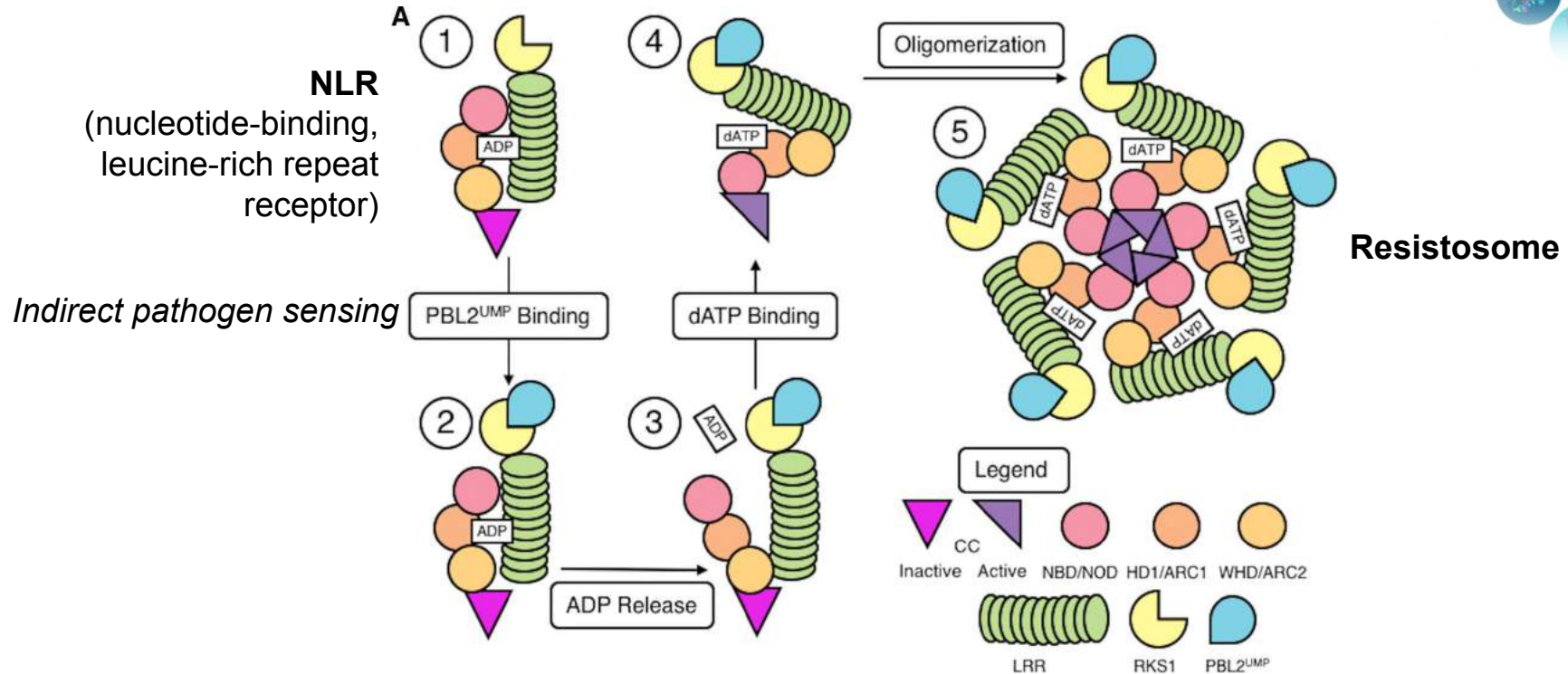
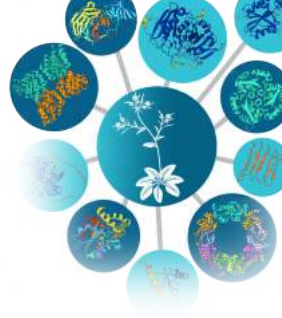


***A. thaliana* resistosome**
PDB 6J5T (Wang 2019)

Case study: *Arabidopsis thaliana* resistosome



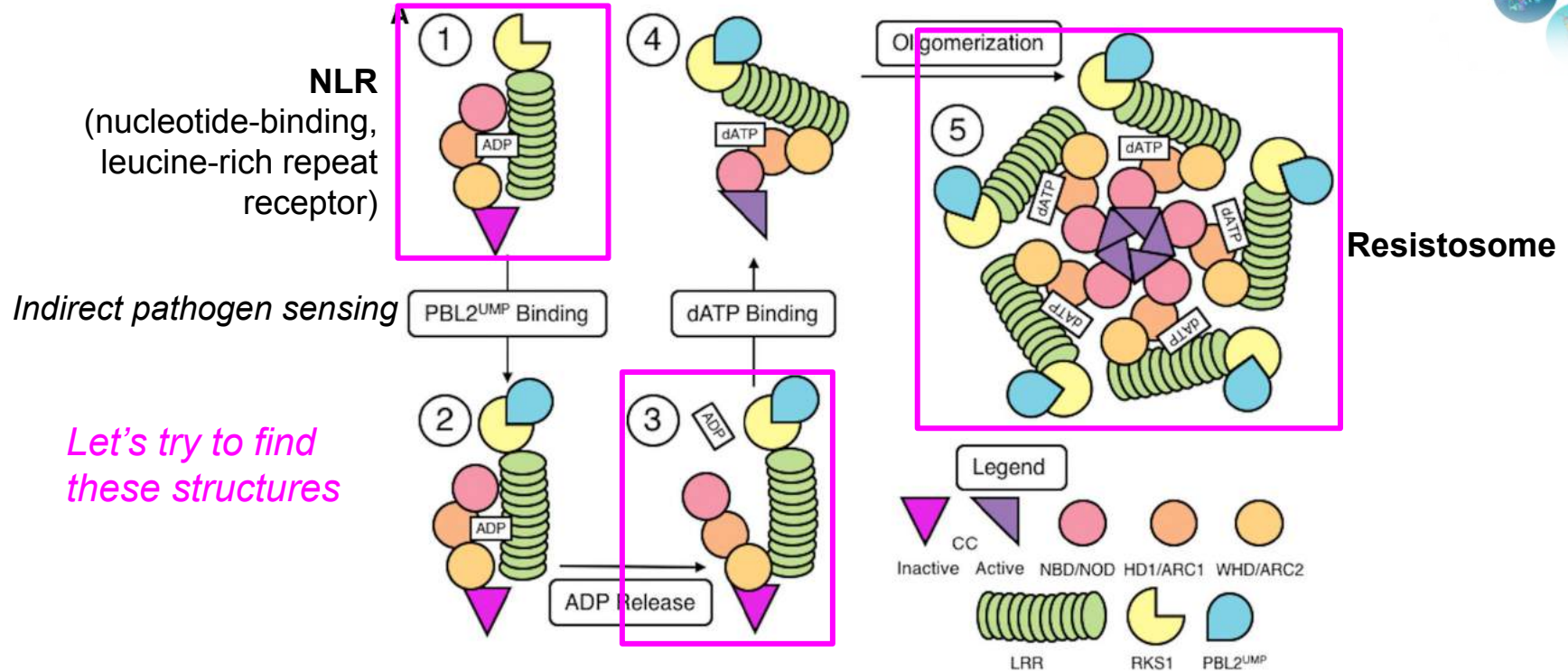
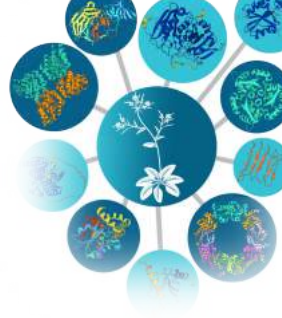
Case study: *Arabidopsis thaliana* resistosome



Burdett, H., et al. (2019) *Cell Host & Microbe* **26**(2), 193–201.

Wang, J., et al. (2019) *Science* **364**(6435), eaav5870.

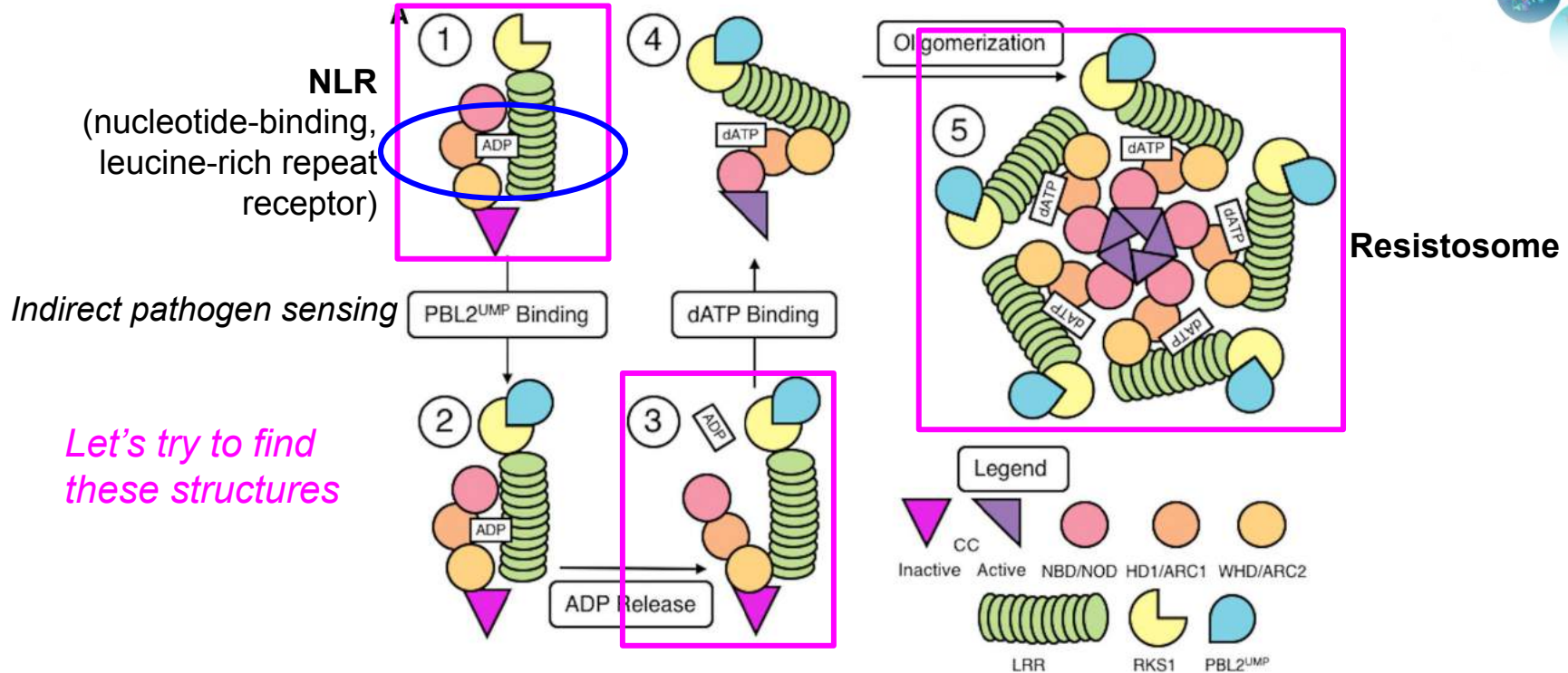
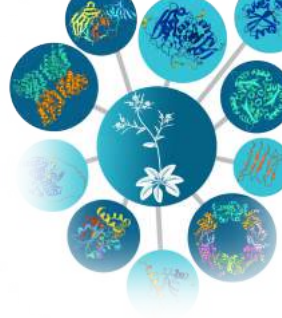
Tutorial: Accessing structure data in the PDB



Burdett, H., et al. (2019) *Cell Host & Microbe* **26**(2), 193–201.

Wang, J., et al. (2019) *Science* **364**(6435), eaav5870.

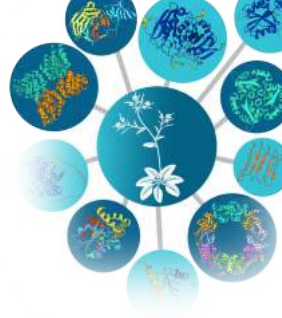
Tutorial: Accessing structure data in the PDB



Burdett, H., et al. (2019) *Cell Host & Microbe* **26**(2), 193–201.

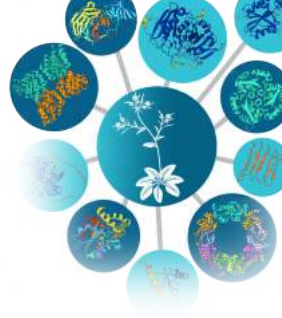
Wang, J., et al. (2019) *Science* **364**(6435), eaav5870.

Tutorial: Accessing structure data in the PDB

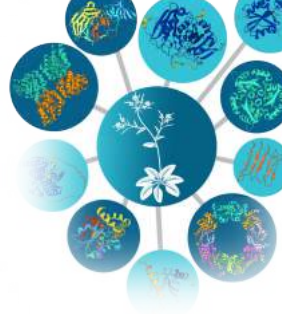


Tutorial time

Summary: Accessing structure data in the PDB



- Interactive access (i.e., while browsing RCSB.org):
 - Search data with “Basic” and “Advanced” search tools (<https://www.rcsb.org/search/advanced>)
 - Download data via:
 - Structure summary pages
 - Search results page
 - Direct navigation of the PDB Archive at <https://s3.rcsb.org> or <https://files.rcsb.org>
 - Structure data under the directory: /pub/pdb/data/structures/divided/mmCIF/
- Programmatic access:
 - Search data with our Search API (try the query editor: <https://search.rcsb.org/query-editor.html>)
 - Download data via command-line/scripted access of the PDB Archive
 - Over HTTP (e.g., using curl, wget, or Python):
 - <https://s3.rcsb.org>
 - <https://files.rcsb.org> (supports recursive directory retrieval)
 - Using rsync: `rsync://rsync.rcsb.org`
 - Using FTP: `ftp://ftp.wwpdb.org`



Resources and Documentation

- Basic and advanced searching:
 - Infographic: <https://cdn.rcsb.org/rcsb-pdb/search/SearchnBrowse2go.pdf>
 - <https://www.rcsb.org/docs/search-and-browse/overview-search-and-browse>
- Programmatic access:
 - Search API: <https://search.rcsb.org/index.html#search-api>
 - Downloading structures:
 - <https://www.rcsb.org/docs/programmatic-access/file-download-services>
 - <https://www.wwpdb.org/ftp/pdb-ftp-sites>
 - AWS S3: <https://www.rcsb.org/news/6266e0379c3931864b072861>
 - <https://www.rcsb.org/docs/programmatic-access/batch-downloads-with-shell-script>

Questions?

RCSB PDB Team



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

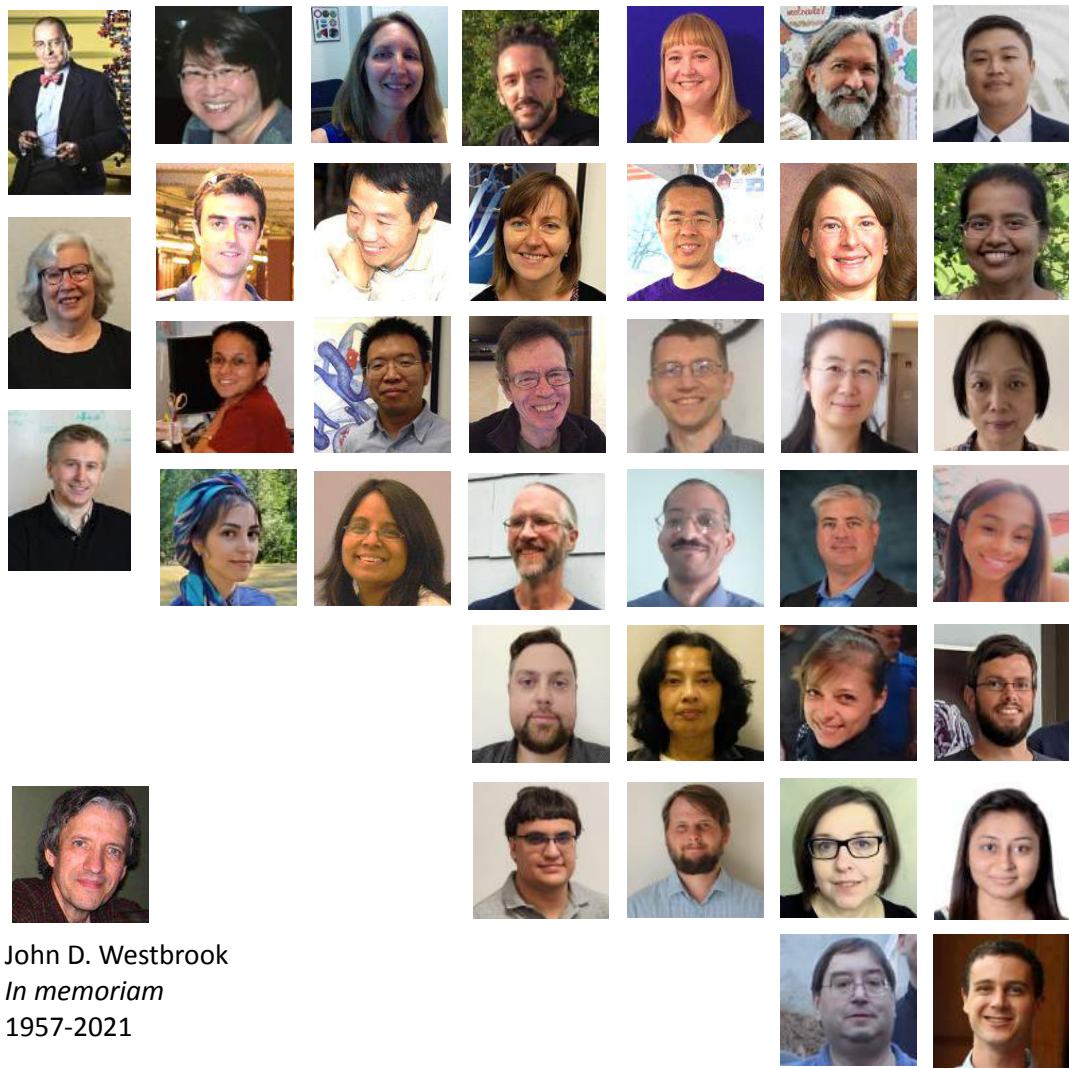


University of California
San Francisco



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

Follow us



John D. Westbrook
In memoriam
1957-2021

Accessing Computed Structure Models Generated Using AlphaFold2 and RoseTTAFold



DOE KBASE/RCSB PDB VIRTUAL CRASH COURSE
NOVEMBER 10, 2022

Brinda Vallat, Ph.D., RCSB PDB/Rutgers
brinda.vallat@rcsb.org

Outline



- What are computed structure models (CSMs)?
 - Protein structure prediction
 - AlphaFold2 and RoseTTAFold
 - AlphaFoldDB and ModelArchive
 - Model quality metrics
 - ModelCIF data standard
- CSMs in RCSB.org
- Accessing CSMs on RCSB.org
 - Live demo: RCSB.org tools and functionalities
 - Case studies: Disease resistance RPP13-like protein 4 from *Arabidopsis thaliana* resistosome and *Micrococcus luteus* aldehyde dehydrogenase

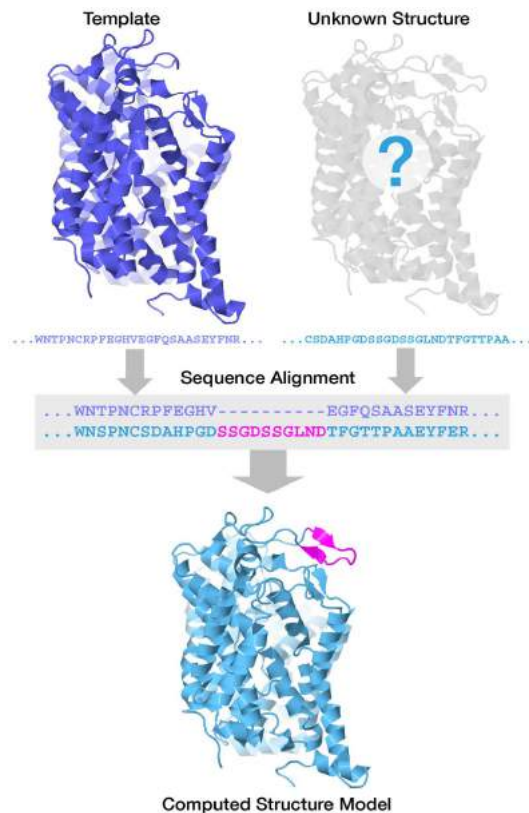


What are computed structure models?

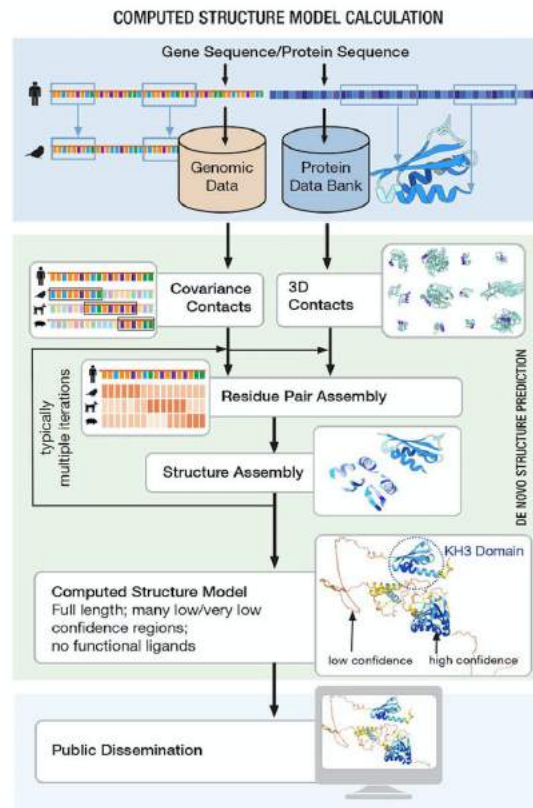
Protein structure prediction



Template-based
modeling



Template-free
modeling



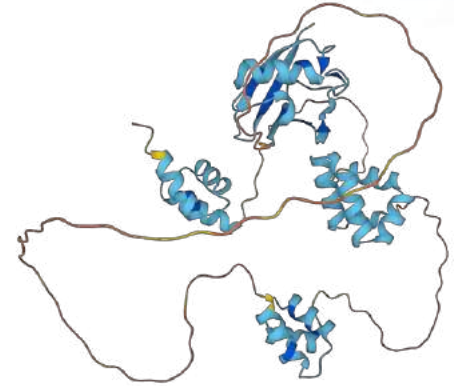


**Computed structure models (CSMs)
are predicted models obtained using
template-based or template-free
modeling methods**

AlphaFold2 and RoseTTAFold



- AlphaFold2 and RoseTTAFold are artificial intelligence / machine learning (AI/ML) methods that predict a protein's 3D structure from its amino acid sequence
- AlphaFold2 is developed by Google Deepmind
 - Achieves accuracy comparable to low resolution experimental structures
- RoseTTAFold is developed by HHMI investigator David Baker and his team at the University of Washington
 - Prediction accuracies approach that of AlphaFold2
- These prediction methods were trained using thousands of sequences and structures of known proteins from the PDB
- **PDB data is the primary input for training these AI/ML prediction methods**



Jumper, J., et al. Nature. 2021. 596, 583–589
Baek, M., et al. Science. 2021. 373, 871-876
Humpheys, I.R., et al. Science. 2021. 374, eabm4805
Shao, C. et al. Structure. 2022. 30, 1385-1394



AlphaFold Protein Structure Database and ModelArchive

- AlphaFold Protein Structure Database (<https://alphafold.ebi.ac.uk>) provides open access to AlphaFold2 predictions
 - Hosted at the European Bioinformatics Institute (EBI)
 - Includes >200 million protein structure predictions, providing a broad structural coverage of sequences in UniProt
- ModelArchive (<https://www.modelarchive.org>) is a repository for CSMs referenced in publications
 - Hosted at the Swiss Institute of Bioinformatics (SIB)
 - Includes predictions of core eukaryotic heteromeric protein complexes modeled using a combination of RoseTTAFold and AlphaFold2

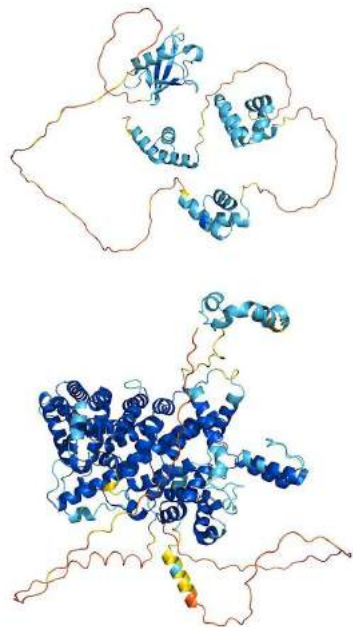
Varadi, M., et al. Nucleic Acids Res. 2022. 50, D439–D444
UniProt Consortium. Nucleic Acids Res. 2021. 49, D480–D489
Humpheys, I.R., et al. Science. 2021. 374, eabm4805

Model quality metrics

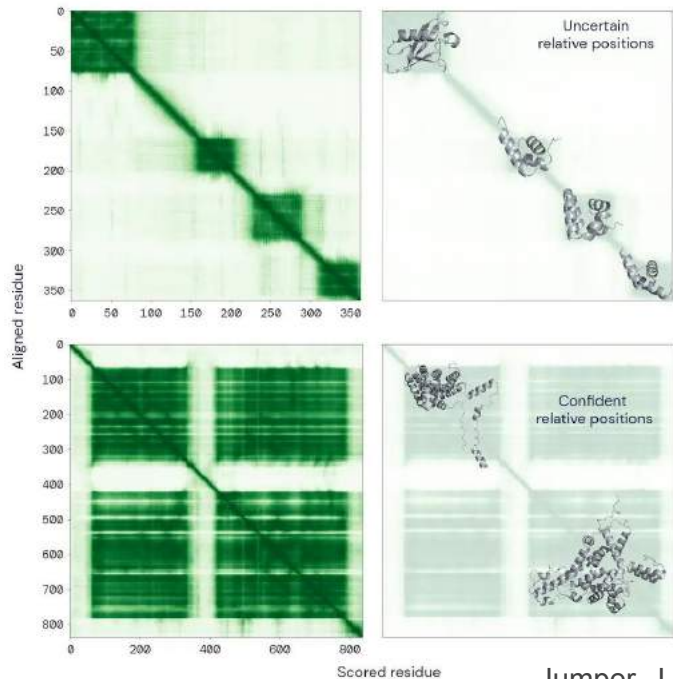


Per residue confidence

■ High (> 90)	■ Low (50 - 70)
■ Medium (70 - 90)	■ Very low (< 50)



Predicted Aligned Error



Two types of intrinsic model accuracy estimates:

- pLDDT: per-residue measure of local confidence on a scale from 0 - 100
- PAE: expected position error at residue x, when the predicted and true structures are aligned on residue y

Jumper, J., et al. Nature. 2021. 596, 583–589

Tunyasuvunakool, K., et al. Nature. 2021. 596, 590-6

<https://www.deepmind.com/publications/enabling-high-accuracy-protein-structure-prediction-at-the-proteome-scale>



ModelCIF Data Standard

- Extension of PDBx/mmCIF for CSMs
- Definitions retained from PDBx/mmCIF
 - Representation of small molecules, polymeric macromolecules, complexes, atomic coordinates and relevant metadata (e.g., authors, citations, software)
- New definitions in ModelCIF
 - Targets, templates, alignments, coevolution data, predicted contacts, model quality metrics
- Publicly available via GitHub: github.com/ihmwg/ModelCIF
- Managed by the wwPDB ModelCIF working group:
wwpdb.org/task/modelcif
- Supported by AlphaFoldDB and the ModelArchive



CSMs in RCSB.org

CSMs in RCSB.org



- RCSB.org now delivers >1,000,000 CSMs together with ~200,000 experimental structures
- Made possible because ModelCIF extension of PDBx/mmCIF facilitates interoperation with PDB data
- Motivation in integrating CSMs:
 - Ameliorates the paucity of experimental 3D structures in the PDB
 - Improves structural coverage of proteomes of interest
 - Extensive set of tools and services provided by RCSB.org can be used to search, analyze, and visualize CSMs alongside experimentally-determined PDB structures

<https://www.rcsb.org/docs/general-help/computed-structure-models-and-rcsborg>
<https://pdb101.rcsb.org/learn/guide-to-understanding-pdb-data/computed-structure-models>

CSMs in RCSB.org



- CSMs are consistently distinguished from experimental structures through source-specific icons and unique coloring schemes
- Provenance information about the CSMs are clearly provided



Experimental
Structures



CSMs

<https://www.rcsb.org/docs/general-help/computed-structure-models-and-rcsborg>
<https://pdb101.rcsb.org/learn/guide-to-understanding-pdb-data/computed-structure-models>

CSMs in RCSB.org: AlphaFold2 and RoseTTAFold



- Pre-packaged collection of 999,255 CSMs from AlphaFoldDB (v3)
 - Model organism proteomes: Proteomes from 48 different model organisms e.g., Arabidopsis, *E. coli*, fruit fly, human, soybean, and zebrafish
 - Global health proteomes: Proteomes of disease-causing organisms, e.g., *H. pylori*, *K. pneumoniae*, *M. tuberculosis*, and *P. falciparum*
 - Swiss-Prot sequences
 - MANE (Matched Annotation from NCBI and EMBL-EBI) Select sequences
- RoseTTAFold CSMs from the ModelArchive
 - Computed structures of core eukaryotic protein complexes produced by the Baker lab computed using a combination of RoseTTAFold and AlphaFold2
 - Set of 1,106 heteromeric complexes archived in the ModelArchive (<https://modelarchive.org/doi/10.5452/ma-bak-cepc>)

Humpheys, I.R., et al. Science. 2021. 374, eabm4805

Baek, M., et al. Science. 2021. 373, 871-876

<https://www.modelarchive.org/>

Jumper, J., et al. Nature. 2021. 596, 583–589

Varadi, M., et al. Nucleic Acids Research. 2022. 50, D439–D444



Accessing CSMs on RCSB.org

Live Demo

Summary



- RCSB.org now provides access to >1,000,000 CSMs from AlphaFold2 and RoseTTAFold alongside ~200,000 experimental structures of macromolecules
- Various search, analysis and visualization tools on RCSB.org facilitate the study of CSMs together with the experimental structures to obtain additional structural and functional insights
- Depending upon the scientific question being investigated and the data available to address the question, different search and analysis strategies can be devised

RCSB PDB Team



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

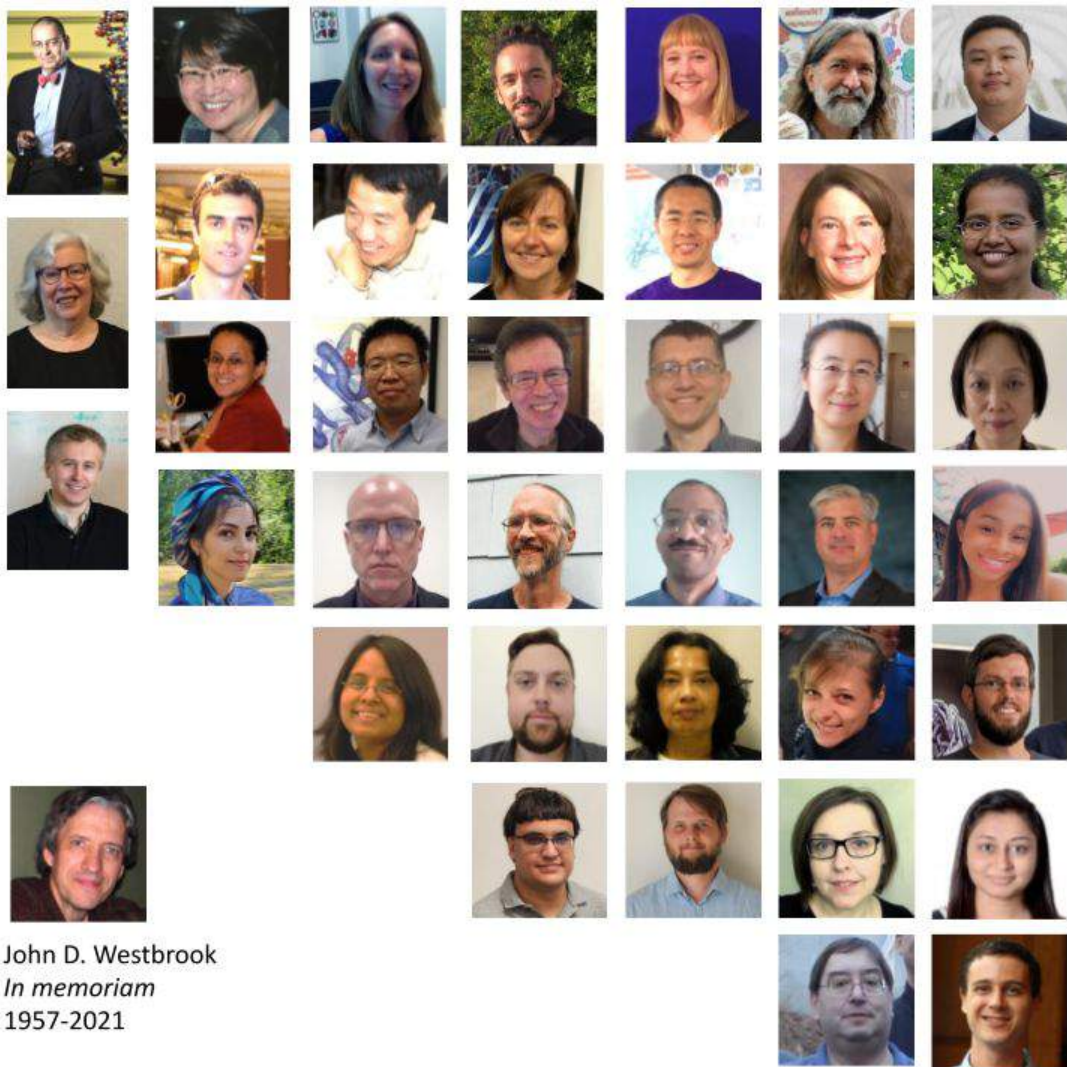


University of California
San Francisco



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

Follow us



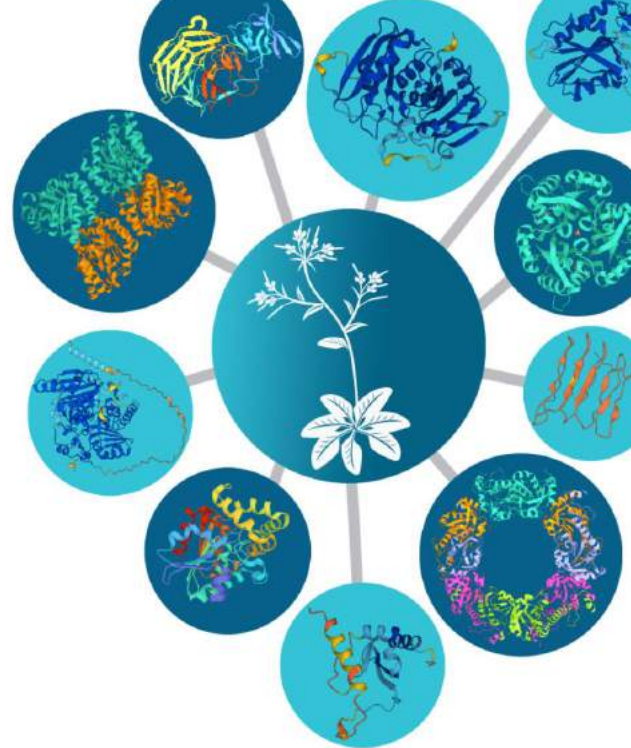
John D. Westbrook
In memoriam
1957-2021



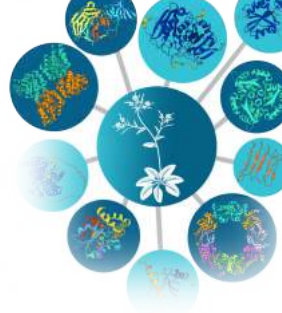
Thank you!

Introduction to the Mol* Molecular Graphics System

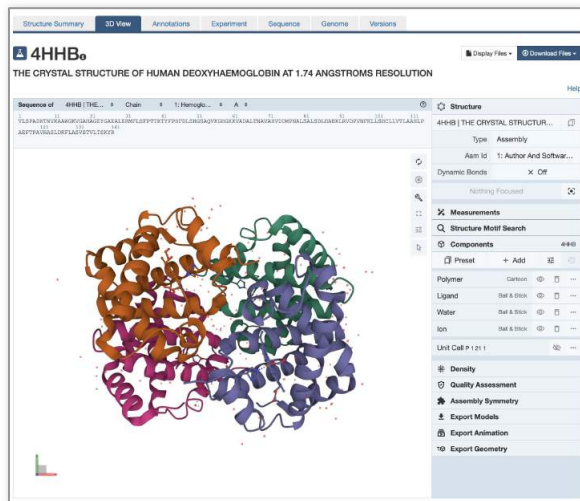
Shuchismita Dutta, Ph.D., RCSB PDB/Rutgers
shuchi.dutta@rcsb.org



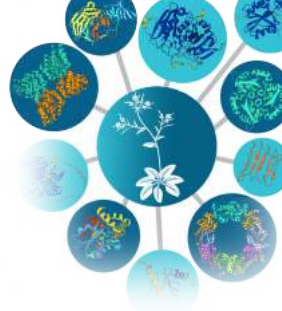
Overview



- Exploring a 3D structure from RCSB.org using Mol*
 - Overall structure
 - Focus vs Select
 - Display and Color
 - Measure
 - Explore Annotations (1D-3D viewer)
- Standalone Mol*
- Exploring and Comparing multiple 3D structures using Mol*



Overall Structure (Access from Structure Summary Page → 3D View)



Sequence Panel

Structure Summary 3D View Annotations Experiment Sequence Genome Versions

4HHB
THE CRYSTAL STRUCTURE OF HUMAN DEOXYHAEMOGLOBIN AT 1.74 ANGSTROMS RESOLUTION

Sequence of 4HHB | THE... Chain 1: Hemoglo...
STWVKAAGKRGDAAGEYGALEKLEWPLSFPTTKTFFPHFOLSHQAQVKGKGAADALYNAYAVYDMPNHALSLEHFAKRLAVDFNPKELEKCLLPLLAHLP
ASFTFAVRLSLQFLADYGVTVLTSKTB

Structure
4HHB | THE CRYSTAL STRUCTUR...
Type Assembly
Asm Id 1: Author And Softwar...
Dynamic Bonds X Off
Nothing Focused
Measurements
Structure Motif Search
Components 4HHB
Preset + Add
Polymer Cartoon
Ligand Ball & Stick
Water Ball & Stick
Ion Ball & Stick
Unit Cell P 121 1
Density
Quality Assessment
Assembly Symmetry
Export Models
Export Animation
Export Geometry

Controls Panel

Components

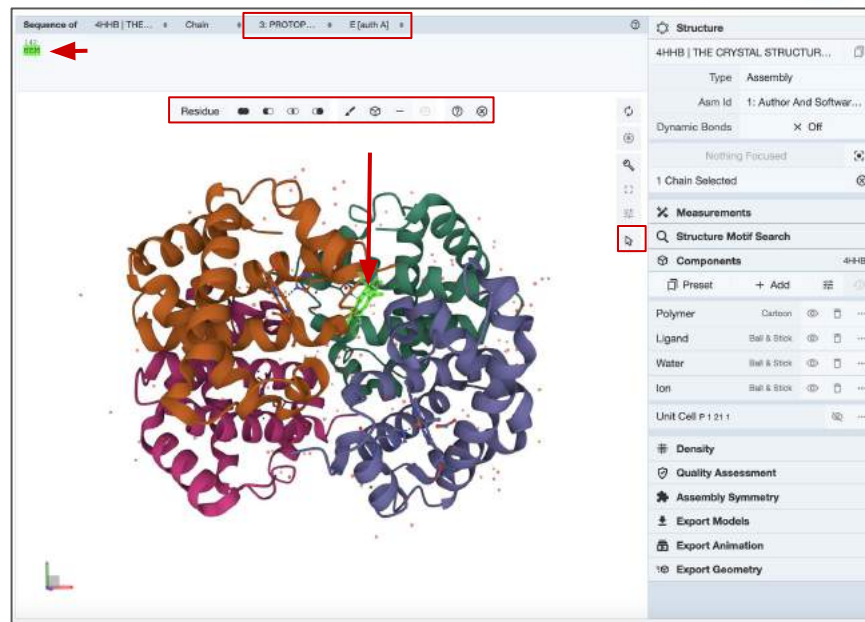
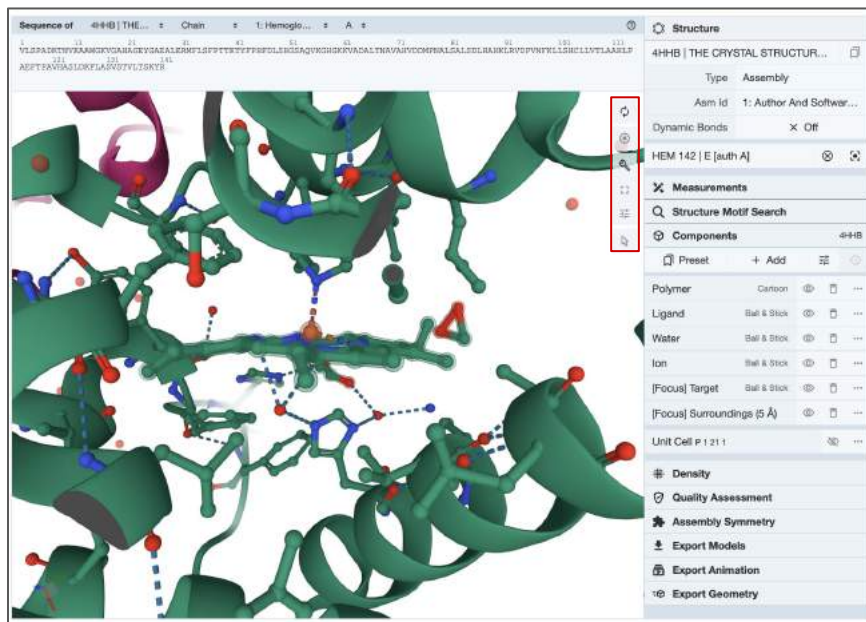
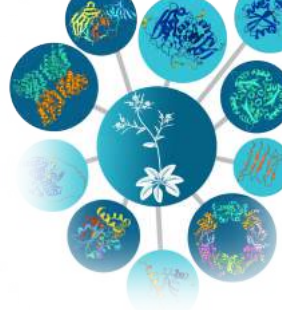
3D-Canvas

Quality Assessment

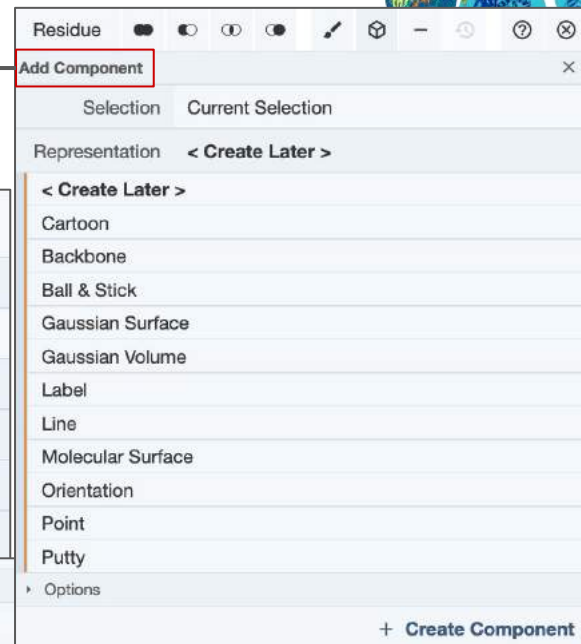
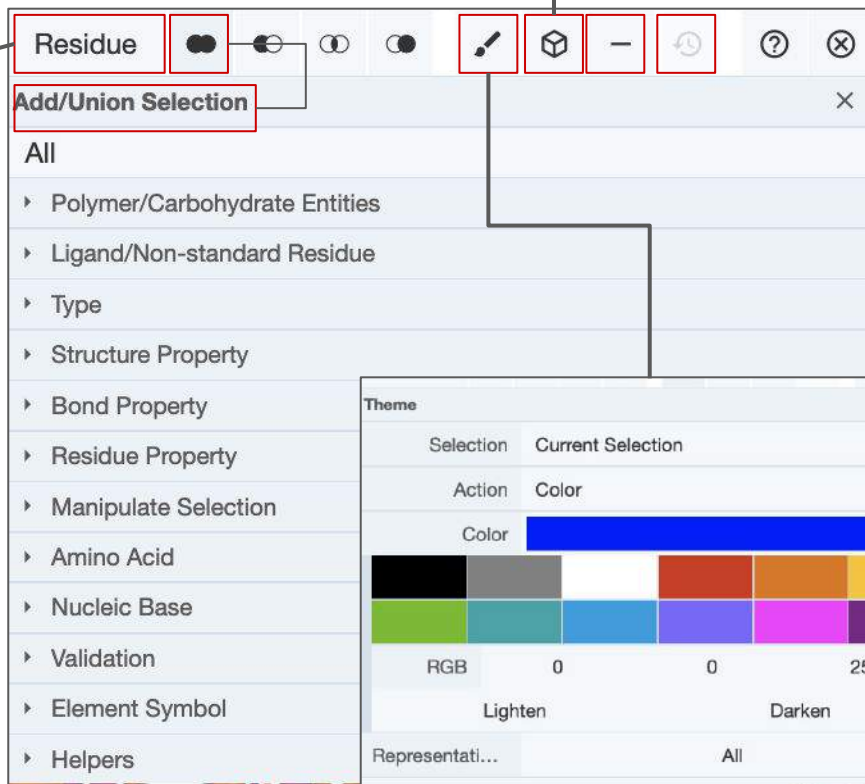
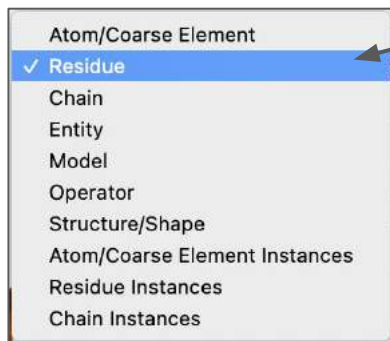
Assembly

Export Options

Focus vs Select

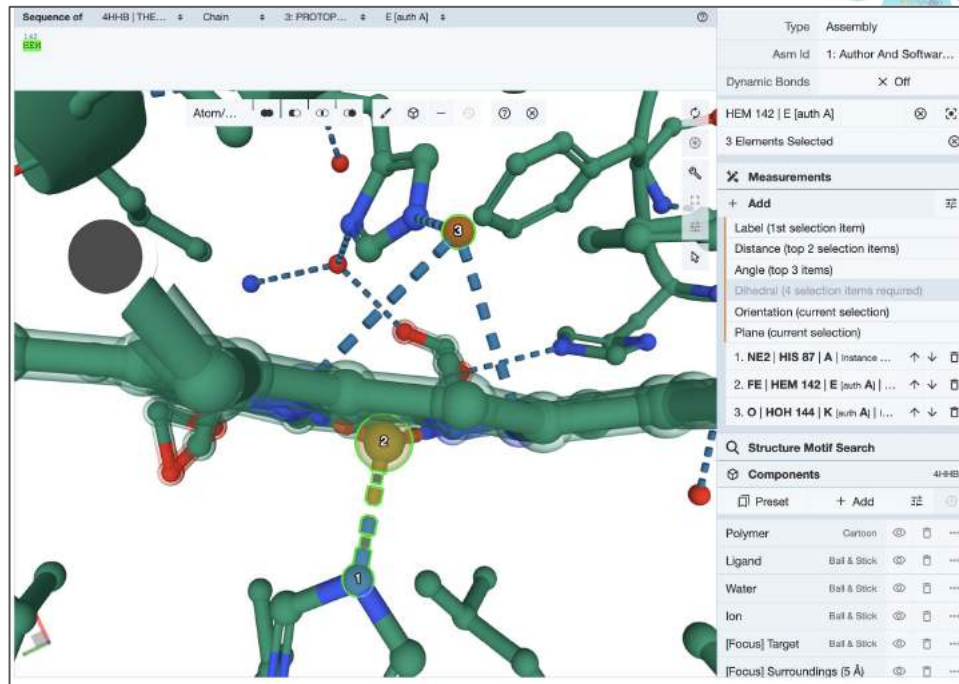


Select to Display and Color



Measure

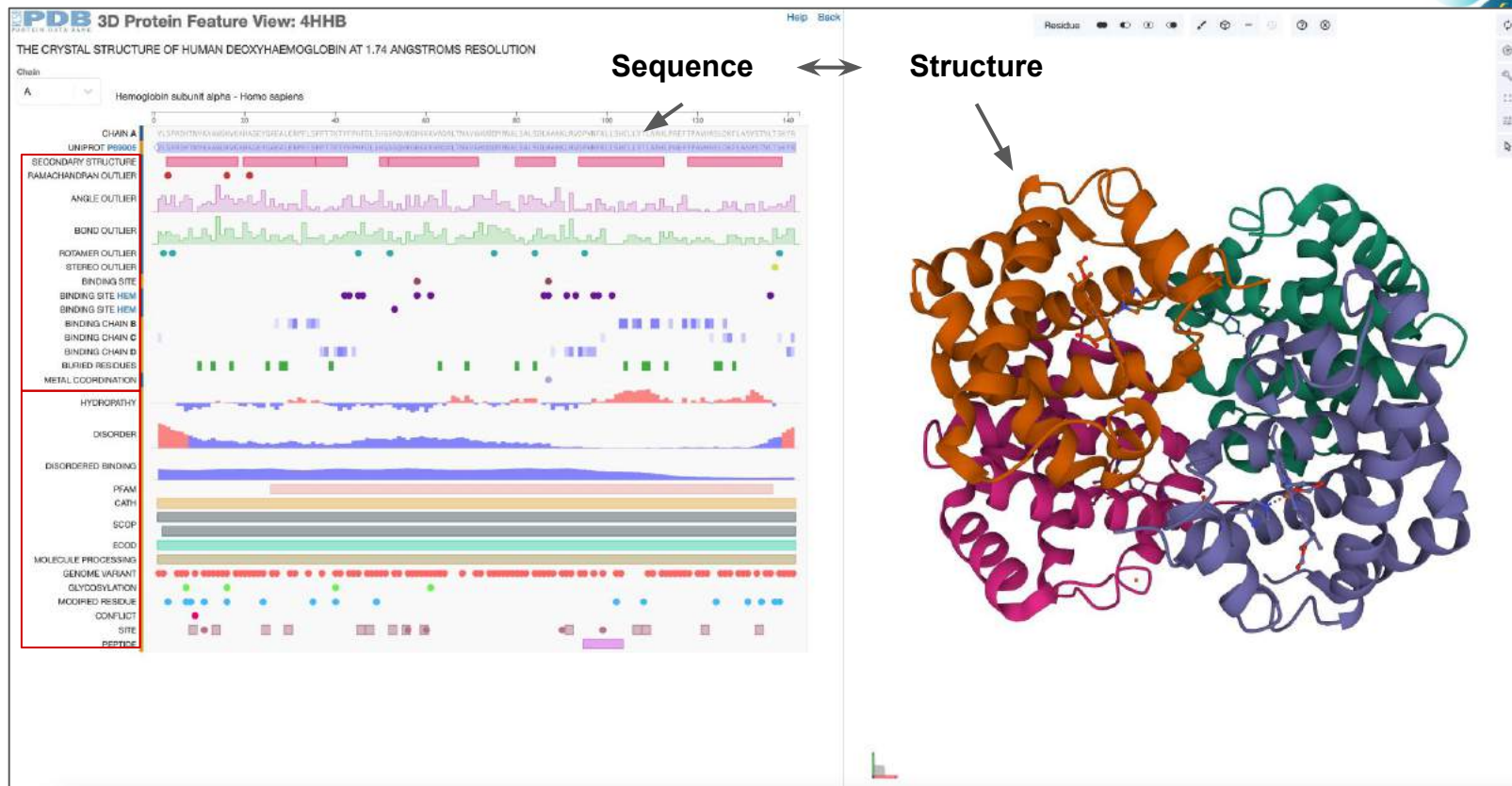
1. Select atom(s) on 3D Canvas
2. Measurements → +Add
3. Select options to measure distance, angle, torsion



Explore Annotations (1D-3D View)

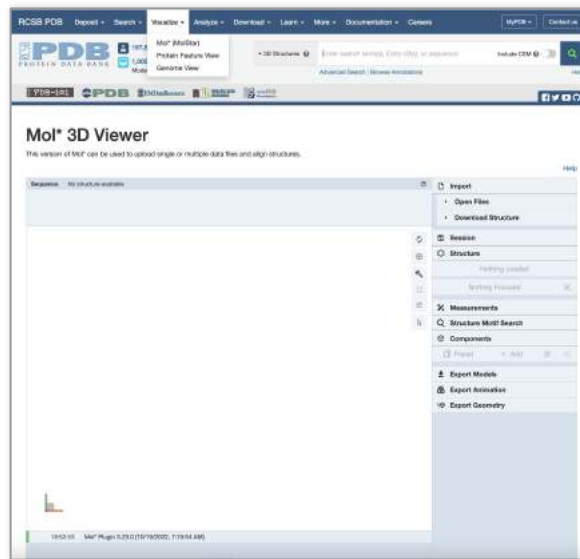
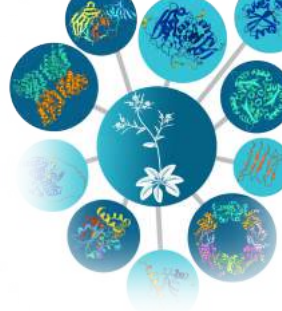
Structure
features

Annotations



Overview


- Exploring a 3D structure from RCSB.org using Mol*
- Standalone Mol* (<https://www.rcsb.org/3d-view>)
 - Upload your coordinates
 - Compare structures
- Exploring and Comparing multiple 3D structures using Mol*



Upload Coordinates

Sequence of 4HHB | THE... Chain 1: Hemoglo... A

1 31 51 71 91 101 111
VLSFACKTVKAAWKGYSAGSSTGAEALRPLSFPTIKTYPPHOLFSGSAQVKGKRVADALTNVAHVVDMPHLSALSOLHARKLVDPVNFKLSNCLLVTAARLP
321 337 347
AEFTPAVHAELEKFLASVSTVLTSEYR



19:25:01 Created Ball & Stick in 6ms.
19:25:01 Created Ball & Stick in 4ms.
19:25:01 Updated Structure Focus Representation in 2ms.

Import

Open Files

Select files...

Format Auto

Visuals ☒ On

Apply

Download Structure

Source PDB

PDB Id(s) 4hhb

Options

Apply

Session

Structure

4HHB | THE CRYSTAL STRUCTUR...

Type Assembly

Asm Id 1: Author And Softwar...

Dynamic Bonds ☒ Off

Nothing Focused

Measurements

Structure Motif Search

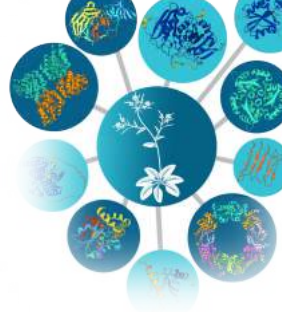
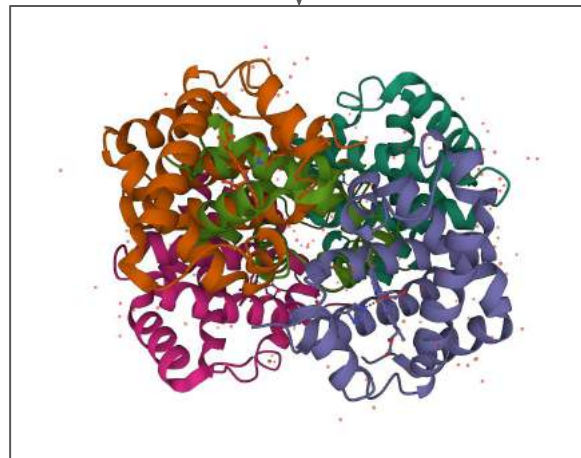
Components 4HHB

Preset + Add

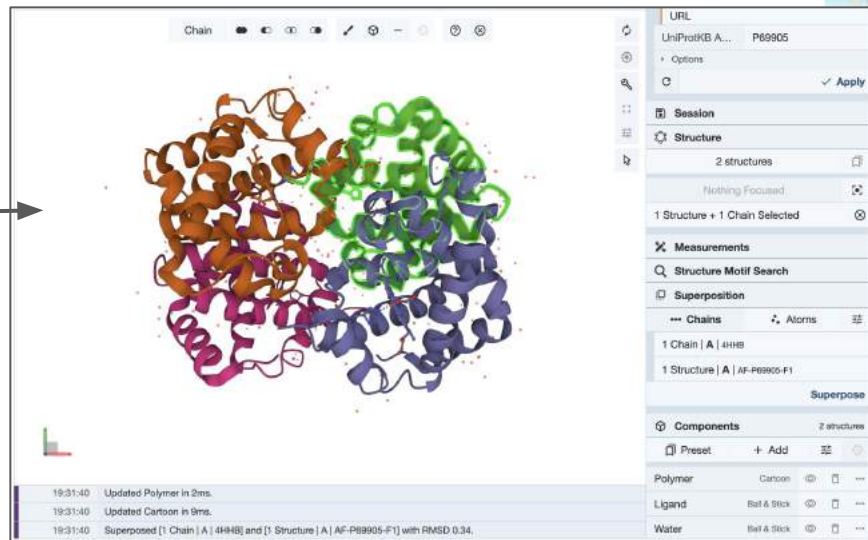
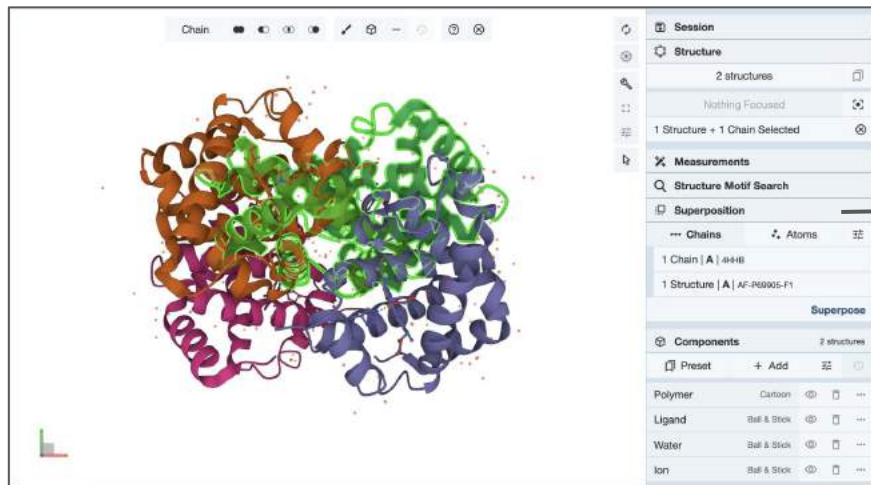
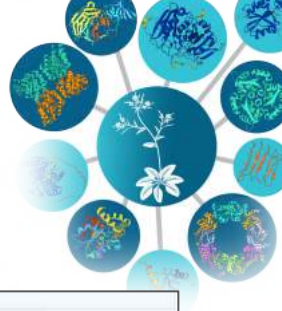
Polymer Cartoon ☒ ☐ ☐

PDB	
PDBDEV	
SWISS-MODEL	
AlphaFold DB	
Model Archive	
PubChem	
URL	
UniProtKB A...	P69905

Download Structure	
Source	AlphaFold DB
UniProtKB A...	P69905
Options	
Apply	

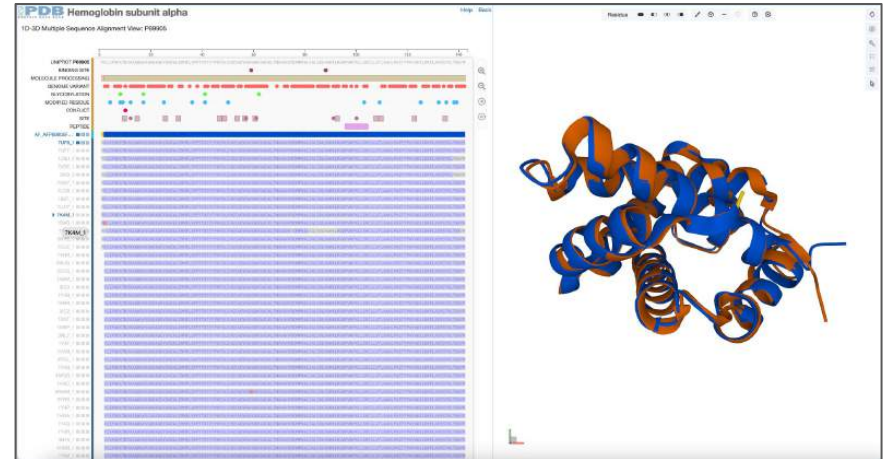
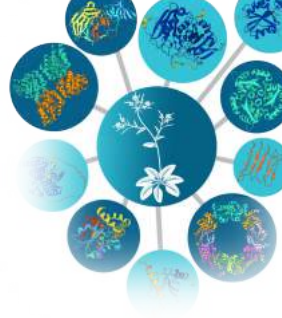


Compare Structures

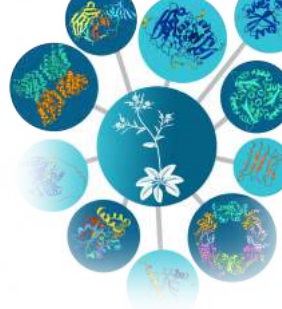


Overview

- Exploring a 3D structure from RCSB.org using Mol*
- Standalone Mol*
- Exploring and Comparing multiple 3D structures using Mol*
 - View 1D-3D Alignments



View 1D-3D Alignments



Group Summary

Group Sequence

Member 1 of 311

Hemoglobin subunit alpha

UniProtKB accession: P09905
Grouped By: Matching UniProtKB accession
Group Content: Polymer Entrez matching query 311
Go to UniProtKB: [P09905](#)
UniProtKB description: Involved in oxygen transport from the lung to the various peripheral tissues.

Group Members:

Hemoglobin subunit alpha

Release Date:

3D View: 1D-3D Alignments

1A00_1 - Summary | Structure

Name: Hemoglobin subunit alpha

Organism: Homo sapiens

Experimental Method: X-ray

Resolution: 2 Å

Molecular Weight: 15.15 kDa

Group Summary

Group Sequence

Hemoglobin subunit alpha

UniProtKB accession: P09905
Grouped By: Matching UniProtKB accession
Group Content: Polymer Entrez matching query 311

[See 1D-3D Alignments](#)

Warning: Some alignment gaps might be an artifact produced by displaying a subset of the whole group for which the multiple sequence alignment was calculated (link).

ALIGNMENTS | STRUCTURAL FEATURES | BINDING SITES

ALL (311)

UniProtKB accessions

SEQUENCE SITE

MOLECULE PROCESSING

GENOME VARIANT

GLYCOSYLATION

MODIFIED RESIDUE

CONFLICT

SITE

PEPTIDE

1A00_1

1A00_2

1A00_3

1A00_4

1A00_5

1A00_6

1A00_7

1A00_8

1A00_9

1A00_10

1A00_11

1A00_12

1A00_13

1A00_14

1A00_15

1A00_16

1A00_17

1A00_18

1A00_19

1A00_20

1A00_21

1A00_22

1A00_23

1A00_24

1A00_25

1A00_26

1A00_27

1A00_28

1A00_29

1A00_30

1A00_31

1A00_32

1A00_33

1A00_34

1A00_35

1A00_36

1A00_37

1A00_38

1A00_39

1A00_40

1A00_41

1A00_42

1A00_43

1A00_44

1A00_45

1A00_46

1A00_47

1A00_48

1A00_49

1A00_50

1A00_51

1A00_52

1A00_53

1A00_54

1A00_55

1A00_56

1A00_57

1A00_58

1A00_59

1A00_60

1A00_61

1A00_62

1A00_63

1A00_64

1A00_65

1A00_66

1A00_67

1A00_68

1A00_69

1A00_70

1A00_71

1A00_72

1A00_73

1A00_74

1A00_75

1A00_76

1A00_77

1A00_78

1A00_79

1A00_80

1A00_81

1A00_82

1A00_83

1A00_84

1A00_85

1A00_86

1A00_87

1A00_88

1A00_89

1A00_90

1A00_91

1A00_92

1A00_93

1A00_94

1A00_95

1A00_96

1A00_97

1A00_98

1A00_99

1A00_100

1A00_101

1A00_102

1A00_103

1A00_104

1A00_105

1A00_106

1A00_107

1A00_108

1A00_109

1A00_110

1A00_111

1A00_112

1A00_113

1A00_114

1A00_115

1A00_116

1A00_117

1A00_118

1A00_119

1A00_120

1A00_121

1A00_122

1A00_123

1A00_124

1A00_125

1A00_126

1A00_127

1A00_128

1A00_129

1A00_130

1A00_131

1A00_132

1A00_133

1A00_134

1A00_135

1A00_136

1A00_137

1A00_138

1A00_139

1A00_140

1A00_141

1A00_142

1A00_143

1A00_144

1A00_145

1A00_146

1A00_147

1A00_148

1A00_149

1A00_150

1A00_151

1A00_152

1A00_153

1A00_154

1A00_155

1A00_156

1A00_157

1A00_158

1A00_159

1A00_160

1A00_161

1A00_162

1A00_163

1A00_164

1A00_165

1A00_166

1A00_167

1A00_168

1A00_169

1A00_170

1A00_171

1A00_172

1A00_173

1A00_174

1A00_175

1A00_176

1A00_177

1A00_178

1A00_179

1A00_180

1A00_181

1A00_182

1A00_183

1A00_184

1A00_185

1A00_186

1A00_187

1A00_188

1A00_189

1A00_190

1A00_191

1A00_192

1A00_193

1A00_194

1A00_195

1A00_196

1A00_197

1A00_198

1A00_199

1A00_200

1A00_201

1A00_202

1A00_203

1A00_204

1A00_205

1A00_206

1A00_207

1A00_208

1A00_209

1A00_210

1A00_211

1A00_212

1A00_213

1A00_214

1A00_215

1A00_216

1A00_217

1A00_218

1A00_219

1A00_220

1A00_221

1A00_222

1A00_223

1A00_224

1A00_225

1A00_226

1A00_227

1A00_228

1A00_229

1A00_230

1A00_231

1A00_232

1A00_233

1A00_234

1A00_235

1A00_236

1A00_237

1A00_238

1A00_239

1A00_240

1A00_241

1A00_242

1A00_243

1A00_244

1A00_245

1A00_246

1A00_247

1A00_248

1A00_249

1A00_250

1A00_251

1A00_252

1A00_253

1A00_254

1A00_255

1A00_256

1A00_257

1A00_258

1A00_259

1A00_260

1A00_261

1A00_262

1A00_263

1A00_264

1A00_265

1A00_266

1A00_267

1A00_268

1A00_269

1A00_270

1A00_271

1A00_272

1A00_273

1A00_274

1A00_275

1A00_276

1A00_277

1A00_278

1A00_279

1A00_280

1A00_281

1A00_282

1A00_283

1A00_284

1A00_285

1A00_286

1A00_287

1A00_288

1A00_289

1A00_290

1A00_291

1A00_292

1A00_293

1A00_294

1A00_295

1A00_296

1A00_297

1A00_298

1A00_299

1A00_300

1A00_301

1A00_302

1A00_303

1A00_304

1A00_305

1A00_306

1A00_307

1A00_308

1A00_309

1A00_310

1A00_311

1A00_312

1A00_313

1A00_314

1A00_315

1A00_316

1A00_317

1A00_318

1A00_319

1A00_320

1A00_321

1A00_322

1A00_323

1A00_324

1A00_325

1A00_326

1A00_327

1A00_328

1A00_329

1A00_330

1A00_331

1A00_332

1A00_333

1A00_334

1A00_335

1A00_336

1A00_337

1A00_338

1A00_339

1A00_340

1A00_341

1A00_342

1A00_343

1A00_344

1A00_345

1A00_346

1A00_347

1A00_348

1A00_349

1A00_350

1A00_351

1A00_352

1A00_353

1A00_354

1A00_355

1A00_356

1A00_357

1A00_358

1A00_359

1A00_360

1A00_361

1A00_362

1A00_363

1A00_364

1A00_365

1A00_366

1A00_367

1A00_368

1A00_369

1A00_370

1A00_371

1A00_372

1A00_373

1A00_374

1A00_375

1A00_376

1A00_377

1A00_378

1A00_379

1A00_380

1A00_381

1A00_382

1A00_383

1A00_384

1A00_385

1A00_386

1A00_387

1A00_388

1A00_389

1A00_390

1A00_391

1A00_392

1A00_393

1A00_394

1A00_395

1A00_396

1A00_397

1A00_398

1A00_399

1A00_400

1A00_401

1A00_402

1A00_403

1A00_404

1A00_405

1A00_406

1A00_407

1A00_408

1A00_409

1A00_410

1A00_411

1A00_412

1A00_413

1A00_414

1A00_415

1A00_416

1A00_417

1A00_418

1A00_419

1A00_420

1A00_421

1A00_422

1A00_423

1A00_424

1A00_425

1A00_426

1A00_427

1A00_428

1A00_429

1A00_430

1A00_431

1A00_432

1A00_433

1A00_434

1A00_435

1A00_436

1A00_437

1A00_438

1A00_439

1A00_440

1A00_441

1A00_442

1A00_443

1A00_444

1A00_445

1A00_446

1A00_447

1A00_448

1A00_449

1A00_450

1A00_451

1A00_452

1A00_453

1A00_454

1A00_455

1A00_456

1A00_457

1A00_458

1A00_459

1A00_460

1A00_461

1A00_462

1A00_463

1A00_464

1A00_465

1A00_466

1A00_467

1A00_468

1A00_469

1A00_470

1A00_471

1A00_472

1A00_473

1A00_474

1A00_475

1A00_476

1A00_477

1A00_478

1A00_479

1A00_480

1A00_481

1A00_482

1A00_483

1A00_484

1A00_485

1A00_486

1A00_487

1A00_488

1A00_489

1A00_490

1A00_491

1A00_492

1A00_493

1A00_494

1A00_495

1A00_496

1A00_497

1A00_498

1A00_499

1A00_500

1A00_501

1A00_502

1A00_503

1A00_504

1A00_505

1A00_506

1A00_507

1A00_508

1A00_509

1A00_510

1A00_511

1A00_512

1A00_513

1A00_514

1A00_515

1A00_516

1A00_517

1A00_518

1A00_519

1A00_520

1A00_521

1A00_522

1A00_523

1A00_524

1A00_525

1A00_526

1A00_527

1A00_528

1A00_529

1A00_530

1A00_531

1A00_532

1A00_533

1A00_534

1A00_535

1A00_536

1A00_537

1A00_538

1A00_539

1A00_540

1A00_541

1A00_542

1A00_543

1A00_544

1A00_545

1A00_546

1A00_547

1A00_548

1A00_549

1A00_550

1A00_551

1A00_552

1A00_553

1A00_554

1A00_555

1A00_556

1A00_557

1A00_558

1A00_559

1A00_560

1A00_561

1A00_562

1A00_563

1A00_564

1A00_565

1A00_566

1A00_567

1A00_568

1A00_569

1A00_570

1A00_571

1A00_572

1A00_573

1A00_574

1A00_575

1A00_576

1A00_577

1A00_578

1A00_579

1A00_580

1A00_581

1A00_582

1A00_583

1A00_584

1A00_585

1A00_586

1A00_587

1A00_588

1A00_589

1A00_590

1A00_591

1A00_592

1A00_593

1A00_594

1A00_595

1A00_596

1A00_597

1A00_598

1A00_599

1A00_600

1A00_601

1A00_602

1A00_603

1A00_604

1A00_605

1A00_606

1A00_607

1A00_608

1A00_609

1A00_610

1A00_611

1A00_612

1A00_613

1A00_614

1A00_615

1A00_616

1A00_617

1A00_618

1A00_619

1A00_620

1A00_621

1A00_622

1A00_623

1A00_624

1A00_625

1A00_626

1A00_627

1A00_628

1A00_629

1A00_630

1A00_631

1A00_632

1A00_633

1A00_634

1A00_635

1A00_636

1A00_637

1A00_638

1A00_639

1A00_640

1A00_641

1A00_642

1A00_643

1A00_644

1A00_645

1A00_646

1A00_647

1A00_648

1A00_649

1A00_650

1A00_651

1A00_652

1A00_653

1A00_654

1A00_655

1A00_656

1A00_657

1A00_658

1A00_659

1A00_660

1A00_661

1A00_662

1A00_663

1A00_664

1A00_665

1A00_666

1A00_667

1A00_668

1A00_669

1A00_670

1A00_671

1A00_672

1A00_673

1A00_674

1A00_675

1A00_676

1A00_677

1A00_678

1A00_679

1A00_680

1A00_681

1A00_682

1A00_683

1A00_684

1A00_685

1A00_686

1A00_687

1A00_688

1A00_689

1A00_690

1A00_691

1A00_692

1A00_693

1A00_694

1A00_695

1A00_696

1A00_697

1A00_698

1A00_699

1A00_700

1A00_701

1A00_702

1A00_703

1A00_704

1A00_705

1A00_706

1A00_707

1A00_708

1A00_709

1A00_710

1A00_711

1A00_712

1A00_713

1A00_714

1A00_715

1A00_716

1A00_717

1A00_718

1A00_719

1A00_720

1A00_721

1A00_722

1A00_723

1A00_724

1A00_725

1A00_726

1A00_727

1A00_728

1A00_729

1A00_730

1A00_731

1A00_732

1A00_733

1A00_734

1A00_735

1A00_736

1A00_737

1A00_738

1A00_739

1A00_740

1A00_741

1A00_742

1A00_743

1A00_744

1A00_745

1A00_746

1A00_747

1A00_748

1A00_749

1A00_750

1A00_751

1A00_752

1A00_753

1A00_754

1A00_755

1A00_756

1A00_757

1A00_758

1A00_759

1A00_760

1A00_761

1A00_762

1A00_763

1A00_764

1A00_765

1A00_766

1A00_767

1A00_768

1A00_769

1A00_770

1A00_771

1A00_772

1A00_773

1A00_774

1A00_775

1A00_776

1A00_777

1A00_778

1A00_779

1A00_780

1A00_781

1A00_782

1A00_783

1A00_784

1A00_785

1A00_786

1A00_787

1A00_788

1A00_789

1A00_790

1A00_791

1A00_792

1A00_793

1A00_794

1A00_795

1A00_796

1A00_797

1A00_798

1A00_799

1A00_800

1A00_801

1A00_802

1A00_803

1A00_804

1A00_805

1A00_806

1A00_807

1A00_808

1A00_809

1A00_810

1A00_811

1A00_812

1A00_813

1A00_814

1A00_815

1A00_816

1A00_817

1A00_818

1A00_819

1A00_820

1A00_821

1A00_822

1A00_823

1A00_824

1A00_825

1A00_826

1A00_827

1A00_828

1A00_829

1A00_830

1A00_831

1A00_832

1A00_833

1A00_834

1A00_835

1A00_836

1A00_837

1A00_838

1A00_839

1A00_840

1A00_841

1A00_842

1A00_843

1A00_844

1A00_845

1A00_846

1A00_847

1A00_848

1A00_849

1A00_850

1A00_851

1A00_852

1A00_853

1A00_854

1A00_855

1A00_856

1A00_857

1A00_858

1A00_859

1A00_860

1A00_861

1A00_862

1A00_863

1A00_864

1A00_865

1A00_866

1A00_867

1A00_868

1A00_869

1A00_870

1A00_871

1A00_872

1A00_873

1A00_874

1A00_875

1A00_876

1A00_877

1A00_878

1A00_879

1A00_880

1A00_881

1A00_882

1A00_883

1A00_884

1A00_885

1A00_886

1A00_887

1A00_888

1A00_889

1A00_890

1A00_891

1A00_892

1A00_893

1A00_894

1A00_895

1A00_896

1A00_897

1A00_898

1A00_899

1A00_900

1A00_901

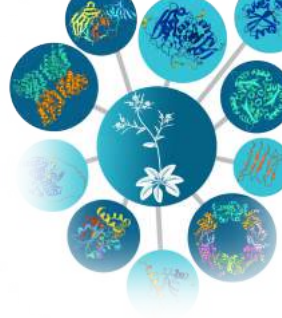
1A00_902

1A00_903

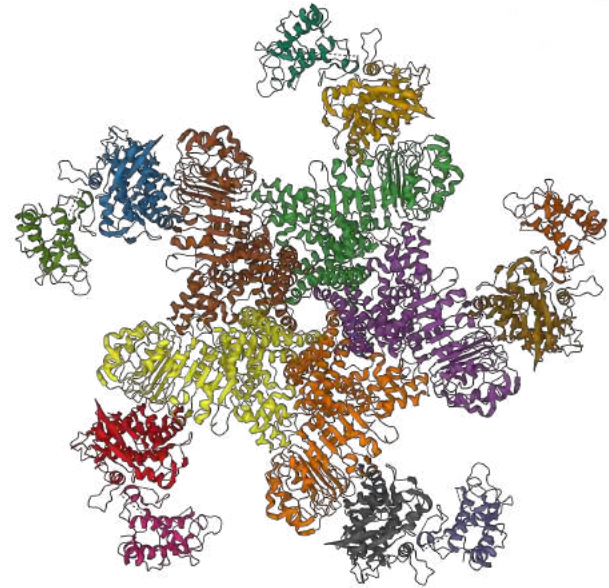
1A00_904

1

Case Study: *Arabidopsis thaliana* Resistosome

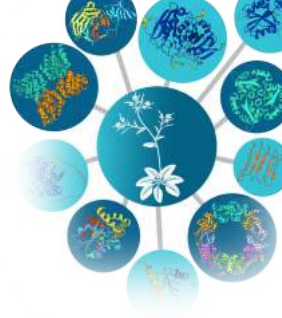


- Which is the Disease resistance RPP13-like protein 4?
- What proteins does it interact with?
- What ligands does it interact with?
- Mutations @ 297 vs 359?
- How do other (experimental/CSM) structures of this protein compare with this structure

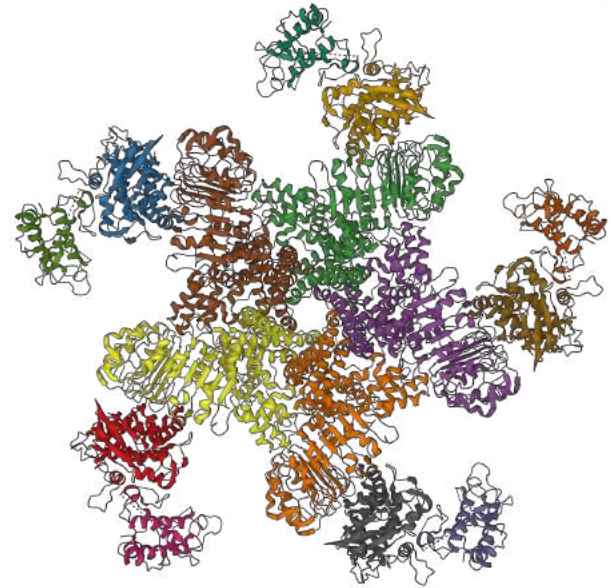


PDB 6J5T (Wang 2019)

Exploring: *Arabidopsis thaliana* Resistosome



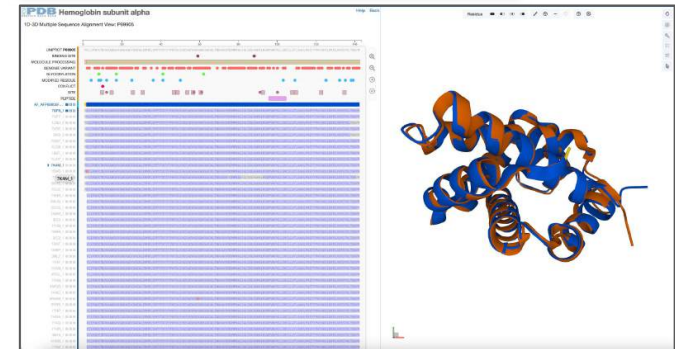
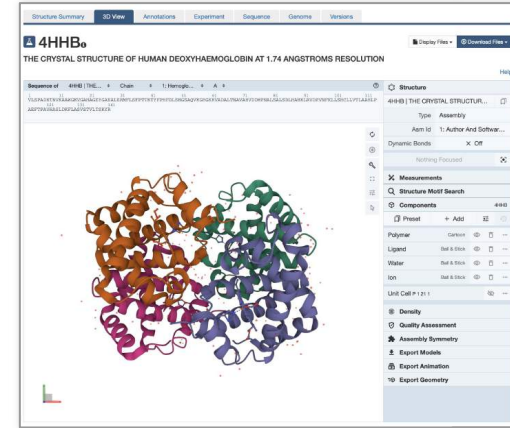
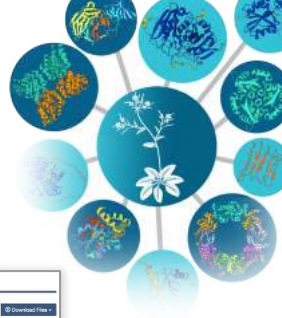
- Which is the Disease resistance RPP13-like protein 4?
 - [Structure Summary Page](#)
- What proteins does it interact with?
 - [1D-3D viewer](#)
- What ligands does it interact with?
 - [1D-3D viewer, Mol*](#)
- Mutations @ 297 vs 359?
 - [1D-3D viewer](#)
- How do other (experimental/CSM) structures of this protein compare with this structure
 - [View 1D-3D Alignments](#)
 - [Pairwise alignment tool](#)



PDB 6J5T (Wang 2019)

Summary

- Exploring a 3D structure from RCSB.org using Mol*
 - Overall structure
 - Focus vs Select
 - Display and Color
 - Measure
 - Explore Annotations (1D-3D viewer)
- Standalone Mol*
 - Upload your coordinates
 - Compare structures
- Exploring and Comparing multiple 3D structures using Mol*
 - View 1D-3D Alignments



RCSB PDB Team



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

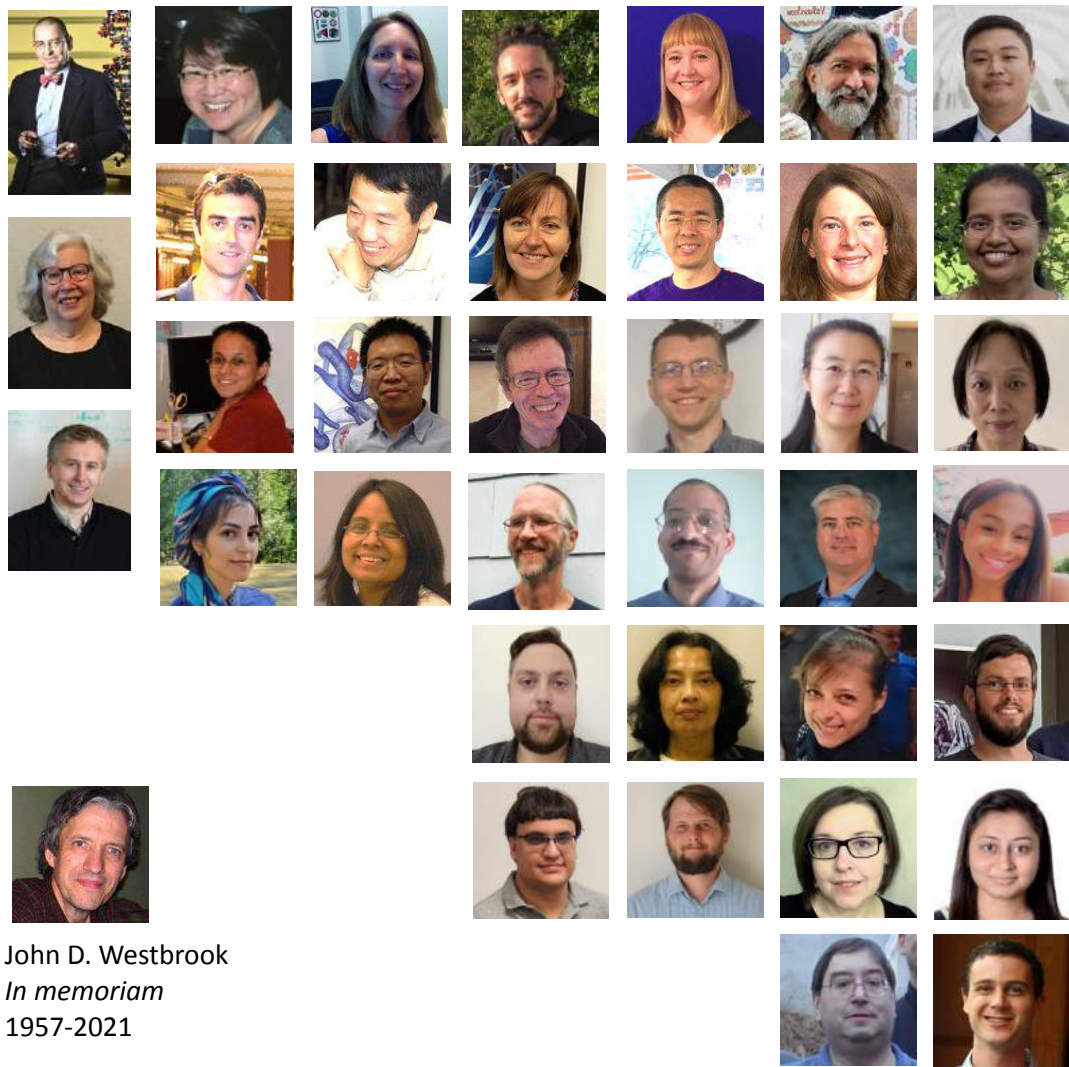


University of California
San Francisco



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

Follow us



John D. Westbrook
In memoriam
1957-2021



KBbase

PREDICTIVE BIOLOGY

DOE Systems Biology Knowledgebase

KBbase Apps for Protein Structure Data Communication and Integration with RCSB PDB

(DOE/KBbase/RCSB PDB) Virtual Crash Course, Nov.10, 2022

Qizhi Zhang, PhD, KBbase/Argonne National Laboratory



U.S. DEPARTMENT OF
ENERGY

Office of
Science

Office of Biological and Environmental Research

INTEGRATION and
MODELING *for*
PREDICTIVE BIOLOGY

Introducing 4 KBase Apps interfacing with RCSB PDB

- Import ProteinStructures from a Metadata File in Staging Area
- Query RCSB databases for protein structures
- **PDB - Import PDB Metadata into KBase Genome**
- Import RCSB Structures (to associate with KBase genomic data objects)

Demo Narrative: <https://narrative.kbase.us/narrative/130799>



Import ProteinStructures from a Metadata File in Staging Area

DATA **Import ProteinStructures from a Metadata File in Staging Area**
Import a set of PDBs from your staging area into your narrative into a ProteinStructures

Run

Parameters
Metadata File Path

Output Objects
ProteinStructures name

Purpose: With your own protein structure files, you can link them with KBase genome data that is in your narratives or shared with you. Given certain sequence similarity threshold, annotation data of structures and genes that satisfy the threshold will be uploaded (saved) to KBase as an object of the type of **ProteinStructures**.

Narrative ID	Object name (Genome AMA feature set)	Feature ID	PDB filename	Is model
107138	MLuteus_ATCC_49442	MLuteus_masurca_RAST.CDS.3483	MLuteus_AlphaFold_3483.pdb	yes
107138	MLuteus_ATCC_49442	MLuteus_masurca_RAST.CDS.3484	MLuteus_Alphafold_3484.pdb	yes
107138	MLuteus_ATCC_49442	MLuteus_masurca_RAST.CDS.133	pdb_extract_MLuteus_133.cif	yes
107138	MLuteus_ATCC_49442	MLuteus_masurca_RAST.CDS.133	ma-bak-cepc-0245.cif	yes
107138	MLuteus_ATCC_49442	MLuteus_masurca_RAST.CDS.133	vinaoutNtarget_133.pdb	yes



Query RCSB databases for protein structures

Purpose: By specifying query filters, you can search the RCSB database to find structure hits that meet the criteria you give.

DATA Query RCSB databases for protein structures
Given a json format query constraint, query RCSB databases for a list of protein structures

Run Configure Info Job Status Result

Parameters

Protein Sequences ×

Uniprot IDs ×

EC Numbers ×

InChI codes ×


SMILES codes ×

Logical AND Operator ☐

Evalue Cutoff

Identity Cutoff 0 ≤ 0.75 ≤1

PDB - Import PDB Metadata into KBase Genome

**PDB - Import PDB Metadata into KBase Genome**
Queries PDB API with genome proteins and annotates proteins with associated PDB metadata

Reset

Finished with **success** at Nov 8, 2022 at 5:41pm

Input Objects

Genome

Athaliana_TAIR10_2012

Parameters

Suffix for annotated genomes

.pdb

Similarity threshold type

E-value

Similarity threshold

Import RCSB Structures

Purpose: Of the RCSB query structure hits, after examining the result metadata, the user can choose a subset of those structures (in RCSB IDs) and associate them with KBase genome data by using this app.

The screenshot shows the 'Import RCSB Structures' web application interface. At the top, there is a 'DATA' tab and the title 'Import RCSB Structures'. Below the title, a subtitle reads: 'Given a list of given RCSB IDs and KBase Genome data, upload their protein structure annotation info'. A status bar indicates 'Finished with success at Nov 7, 2022 at 2:15pm'. There are buttons for 'Reset', 'View Configure', 'Info', 'Job Status', and 'Result'. The main section is titled 'Parameters' and contains a form for importing RCSB structures. The form is divided into two sections: 'Info for RCSBs to be Imported' and a section for individual structure details. The first section includes fields for 'RCSB ID' (6TUK), 'RCSB file extension' (pdb), 'Narrative ID' (107138), 'Genome name' (MLuteus_ATCC_49442), 'Feature ID' (MLuteus_masurca_RAST.CDS.133), and 'Is Model' (checked). The second section includes fields for 'RCSB ID' (3LXD) and 'RCSB file extension' (pdb).

Parameters	
Info for RCSBs to be Imported	
RCSB ID	6TUK
RCSB file extension	pdb
Narrative ID	107138
Genome name	MLuteus_ATCC_49442
Feature ID	MLuteus_masurca_RAST.CDS.133
Is Model	<input checked="" type="checkbox"/>
RCSB ID	3LXD
RCSB file extension	pdb

To the demo...

Demo in a KBase Narrative:

<https://narrative.kbase.us/narrative/130799>



Todo next...

1. Add viewer for the KBase ProteinStructures datatype
2. More RCSB PDB query features and tools (e.g., Docking/Alignment)
3. Better differentiation of experimental from computational structures (source-specific icon, provenance info, RCSB ids)
4. Search results into sets of KBase data objects
5. Other Suggestions?





KBase

PREDICTIVE BIOLOGY

DOE Systems Biology Knowledgebase

Making the Best use of Protein Structure Data in KBase and PDB

Christopher S. Henry

November 10th, 2022

RUTGERS



U.S. DEPARTMENT OF
ENERGY

Office of
Science

Office of Biological and Environmental Research



INTEGRATION and
MODELING *for*
PREDICTIVE BIOLOGY

Review of PDB-KBase Workflows

1. Using gene function discovery pipelines to identify candidate genes for functions of interest
 - Candidates could be proposed from gapfilling, pathway prediction, annotation apps, alignments, blast and many other algorithms in KBase
2. Pulling closest experimental structures associated with candidate genes from PDB
3. Exporting candidate genes as FASTA, generating structures in google collab, and importing computational structures back into KBase or query and analyze structures in PDB
4. Using PDB query and import app to also pull structures with desired functions or substrates cocrystalized
5. Aligning computational and experimental structures in the PDB site
6. Comparing PDB complexes and cocrystalized ligands with complexes and reaction substrates in models
7. Applying Import PDB Metadata app to determine extent of structure coverage of genome, identify gaps in structure coverage, test for existing and completeness of protein complexes, import PDB/uniprot data on annotation and co-crystalization, and identify structures of interest for deeper analysis.

Future plans for KBase PDB apps

1. Mol* viewer in KBase
 - Currently only available on upload - adding a widget to view any time
 - Enabling painting of other data in KBase on structure views (e.g. sequence alignments, variation, domains, docking, annotations, fold quality)
 - Integrating more mol* functionality into KBase
2. PDB Query app
 - Enabling query by feature ID in KBase objects (genome, feature set, metagenome) rather than copying and pasting sequence
 - Saving query output in KBase with annotations as protein sequence set
 - Enabling query by structure alignment
3. Structure import
 - Expanding import to include computational structures from PDB and other databases
4. Docking
 - Linking KBase autodock-VINA app to new structure datatype and enabling saving of docked poses for future viewing

Future plans for New PDB apps

1. Links to modeling
 - a. Comparing all model complexes to complex data in structures
 - b. Comparing reaction substrate to cocrystallized ligands and annotated models with co-enzyme data (e.g. PLP)
2. Structural alignment
 - a. Mitchell and Sedova team at ORNL are adding tools for structure alignment
3. Experimental/computational structures
 - a. Enabling automated linking of imported computational structures to closest experimental structures and facilitating rapid alignment of these pairs
4. Adding one or more computational structure prediction apps in KBase
5. Other ideas?

Acknowledgements



Stephen Burley



John Westbrook



Ed O'loughlin



Qizhi Zhang



Kelly Skinner



Jack Gilbert



Sam Seaver



Claudia Lerma-Ortiz



Nidhi Gupta

KBase team



We thanks DOE BER for funding this work!

Special thanks to all those who attended and contributed to our KBase-PDB design workshops!