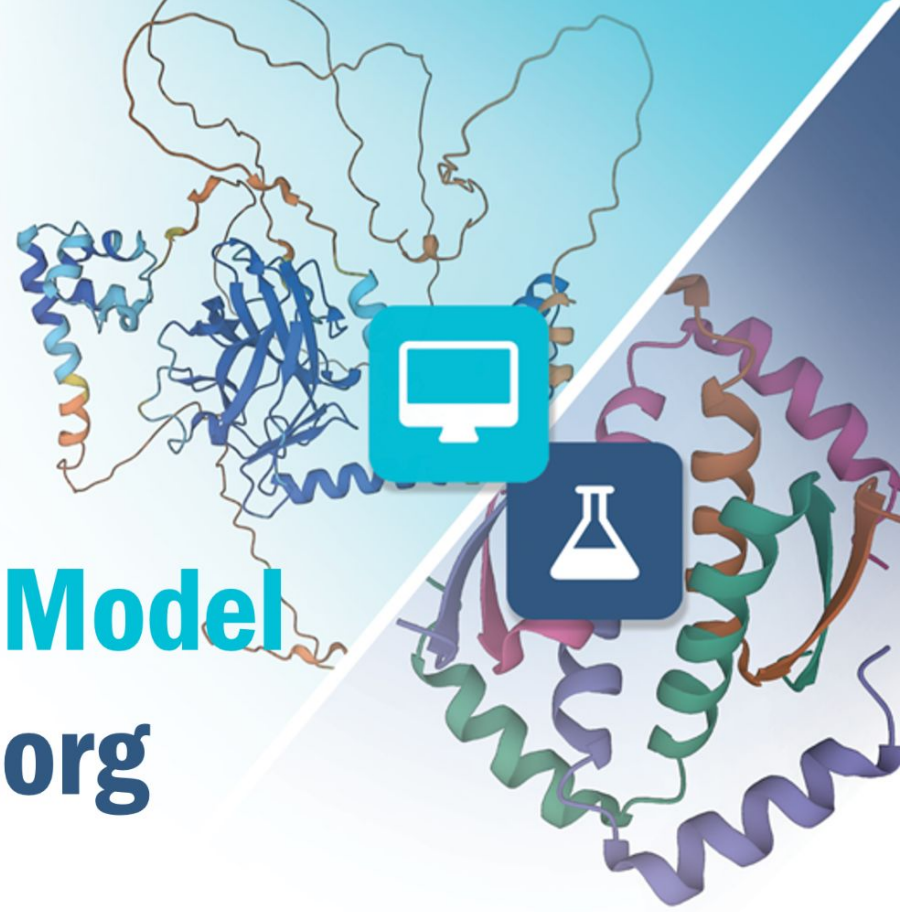
RCSB PDB
PROTEIN DATA BANK

IQB



VIRTUAL WEBINAR

A Deep Dive into Computed Structure Model Exploration at RCSB.org

Tuesday April 30th 2024

9-10am Pacific | 12-1pm Eastern

RCSB PDB
PROTEIN DATA BANK

rcsb.org

Welcome

Stephen K. Burley, M.D., D.Phil.

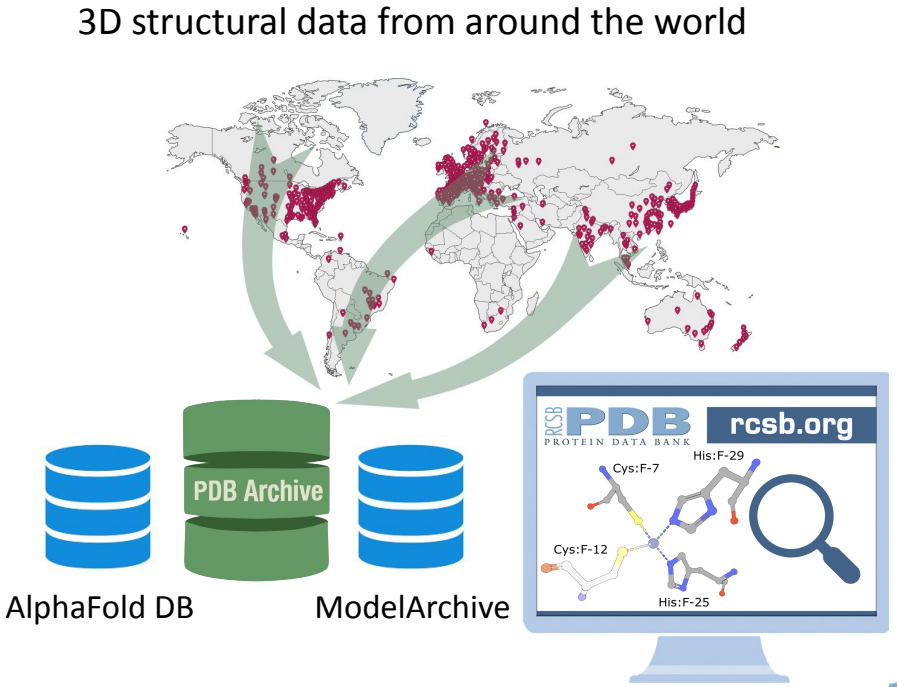
stephen.burley@rcsb.org

Agenda

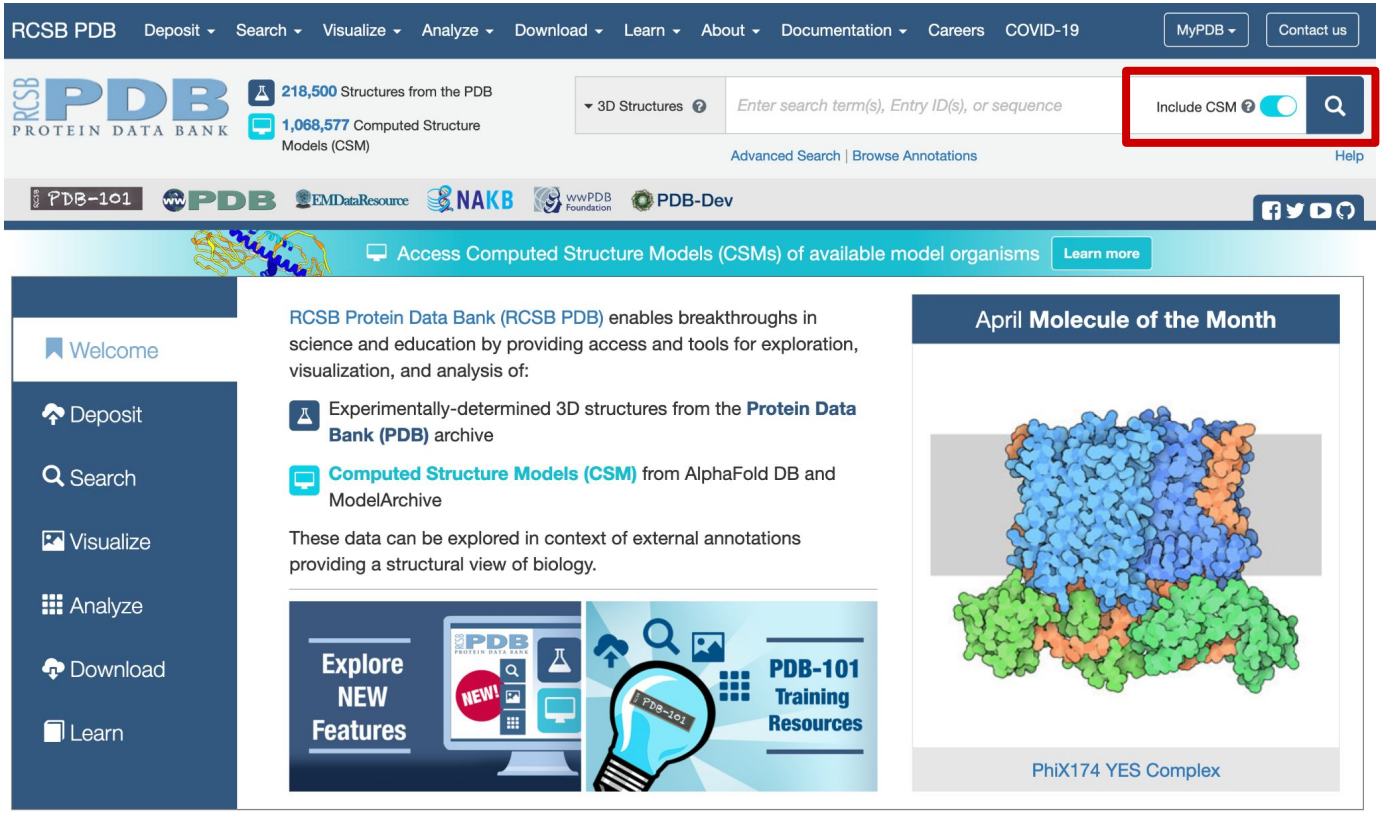
9:00 - 9:05 AM PDT 12:00 - 12:05 PM EDT	Welcome and Introduction	Stephen K. Burley, Director
9:05 - 9:25 AM PDT 12:05- 12:25 PM EDT	A case study of Low-density Lipoprotein Receptor Adapter Protein 1 (LDLRAP1)	Yana Rose, Scientific Software Developer & Data Architect
9:25 - 09:45 AM PDT 12:25 - 12:45 PM EDT	Case study of Class II aminoacyl-tRNA synthetases	Sebastian Bittrich, Scientific Software Developer
09:45 - 10:00 AM PDT 12:45 - 1:00 PM EDT	Q&A	Jose Duarte, Brinda Vallat, Yana Rose, Sebastian Bittrich, Joan Segura

RCSB.org: One-Stop-Shop for Public 3D Biostructure Data

- **RCSB.org:** Tools for searching, accessing, visualizing, analyzing, and downloading data
  - Open access to ~220,000 experimental structures of macromolecules
  - >1 million Computed Structure Models (CSMs) predicted using AI/ML methods
- Provenance/reliability of both data types clearly identified



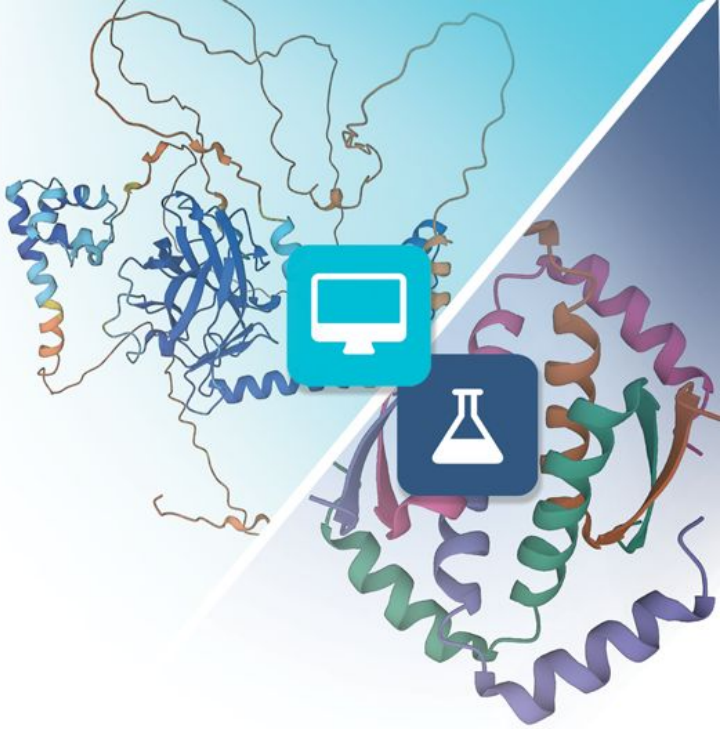
RCSB.org: Opt-in to include CSMs



RCSB PDB
PROTEIN DATA BANK

rcsb.org

Case Study of Low-density Lipoprotein Receptor Adapter Protein 1 (LDLRAP1)

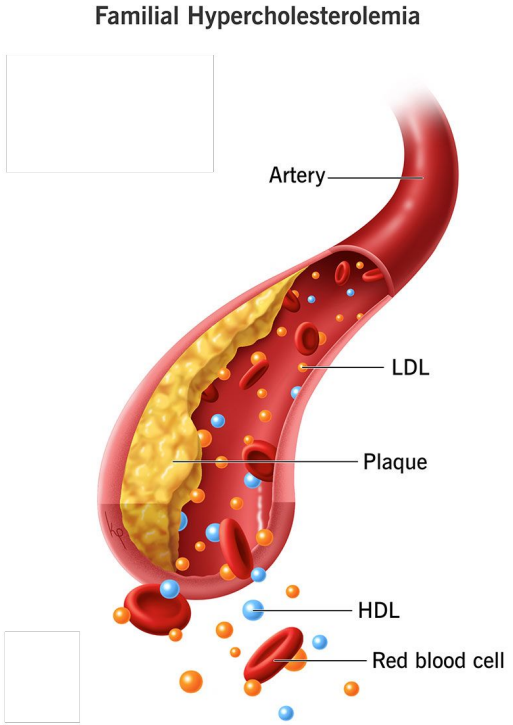


Yana Rose, PhD

yana.rose@rcsb.org

Low-density lipoprotein receptor adapter protein 1 (LDLRAP1)

- LDLRAP1 protein helps remove low-density lipoprotein (LDL) “bad” cholesterol from the bloodstream
- Mutations in the *LDLRAP1* gene has been shown to cause a form of familial hypercholesterolemia due to high levels of LDR
- Familial hypercholesterolemia can cause heart attacks at an early age



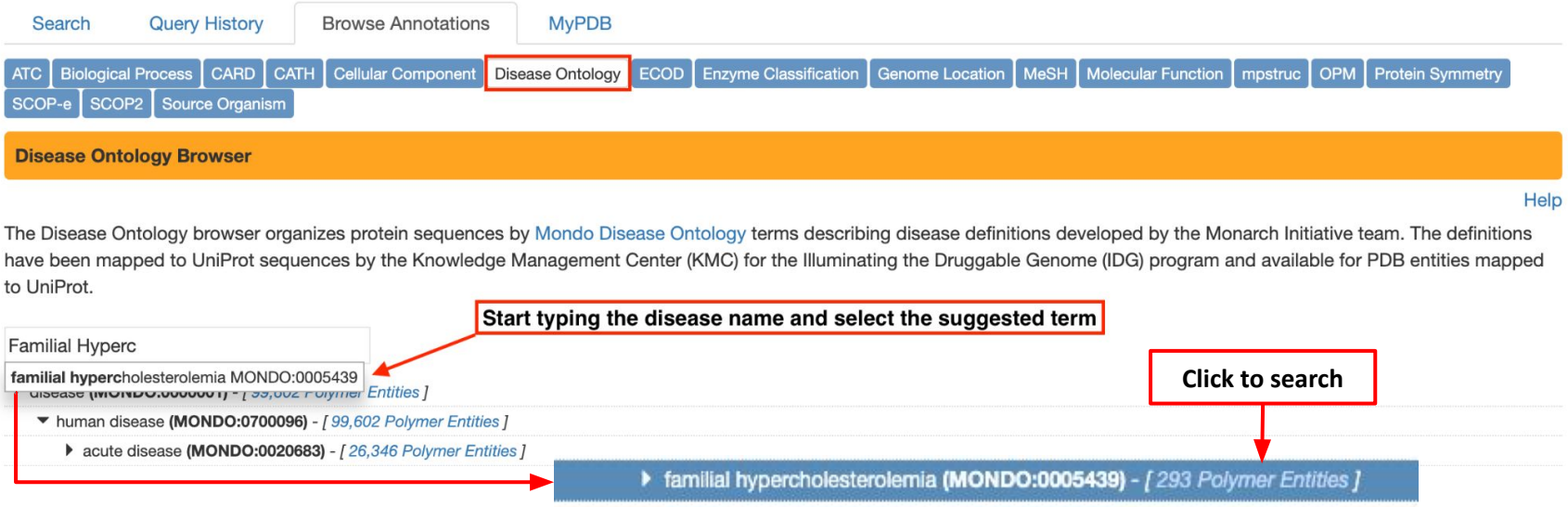
The structural coverage of the top 5 proteins causing familial hypercholesterolemia

Gene	UniProt ID	Protein	Protein length (residues, n)	Experimental coverage (residues, n)	Experimental coverage (residues, %)	AlphaFold (pLDDT ≥ 70) coverage (residues, n)	AlphaFold (pLDDT ≥ 70) coverage (residues, %)
LDLRAP1	Q5SW96	Low density lipoprotein receptor adapter protein 1	308	16	**5.2**	159	**51.6**
LDLR	P01130	Low-density lipoprotein receptor	860	705	82.0	643	74.8
APOB	P04114	Apolipoprotein B-100	4563	N/A	N/A	-	-
APOE	P02649	Apolipoprotein E	317	298	94.0	218	68.8
PCSK9	Q8NBP7	Proprotein convertase subtilisin/kexin type 9	692	642	92.8	563	81.4

Live demo

RCSB.org

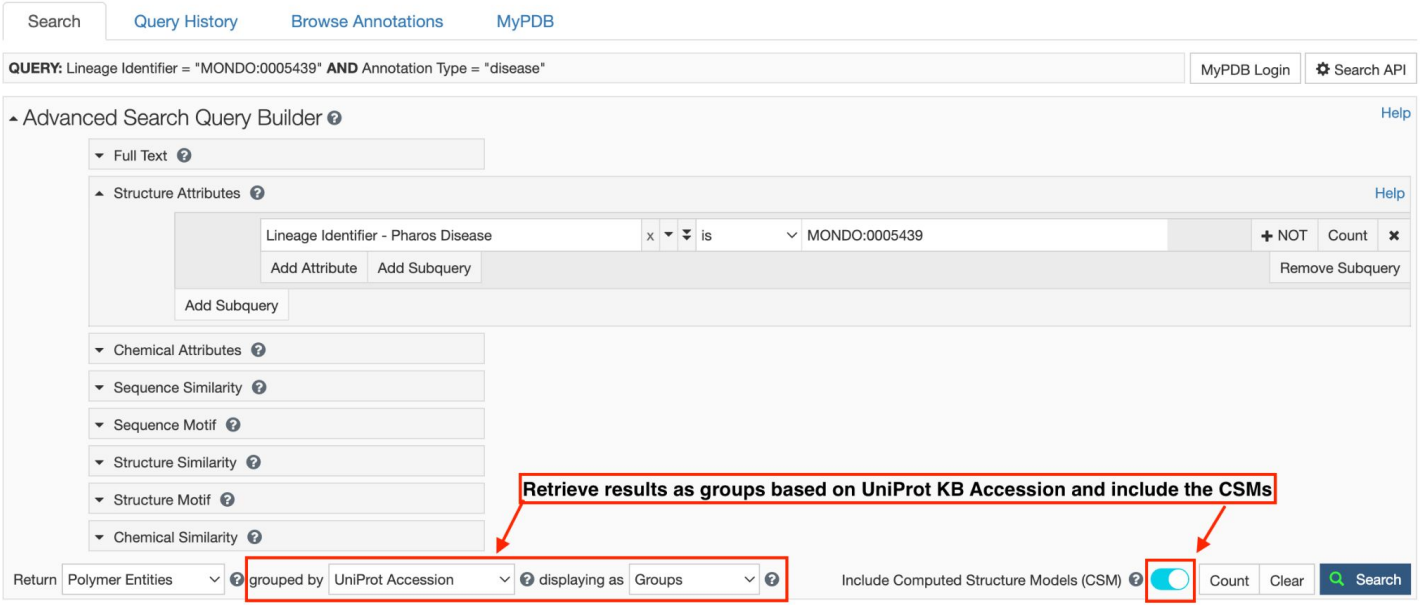
Browse disease annotations on RCSB.org



Browse Disease Annotations

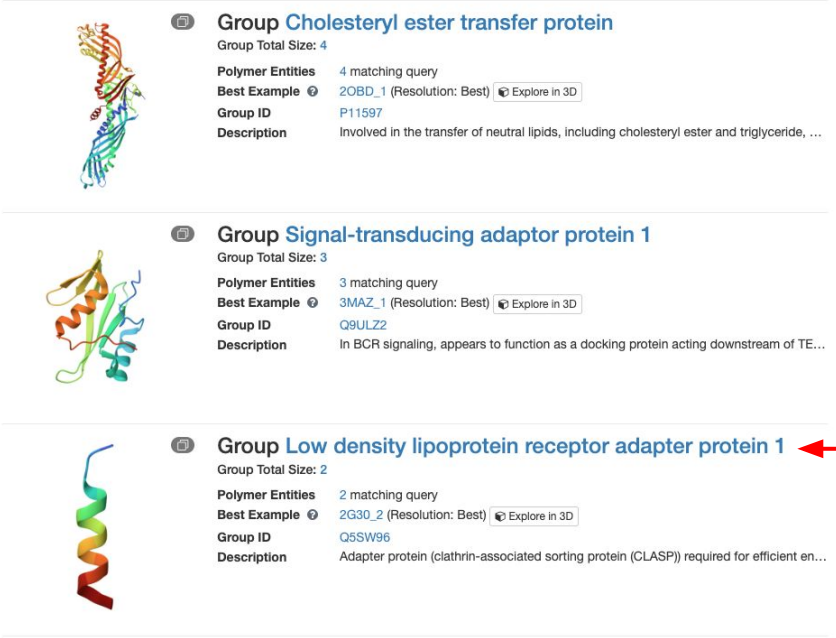
Learn more about Mondo DO

Use Advanced Search to group structures based on the reference sequence



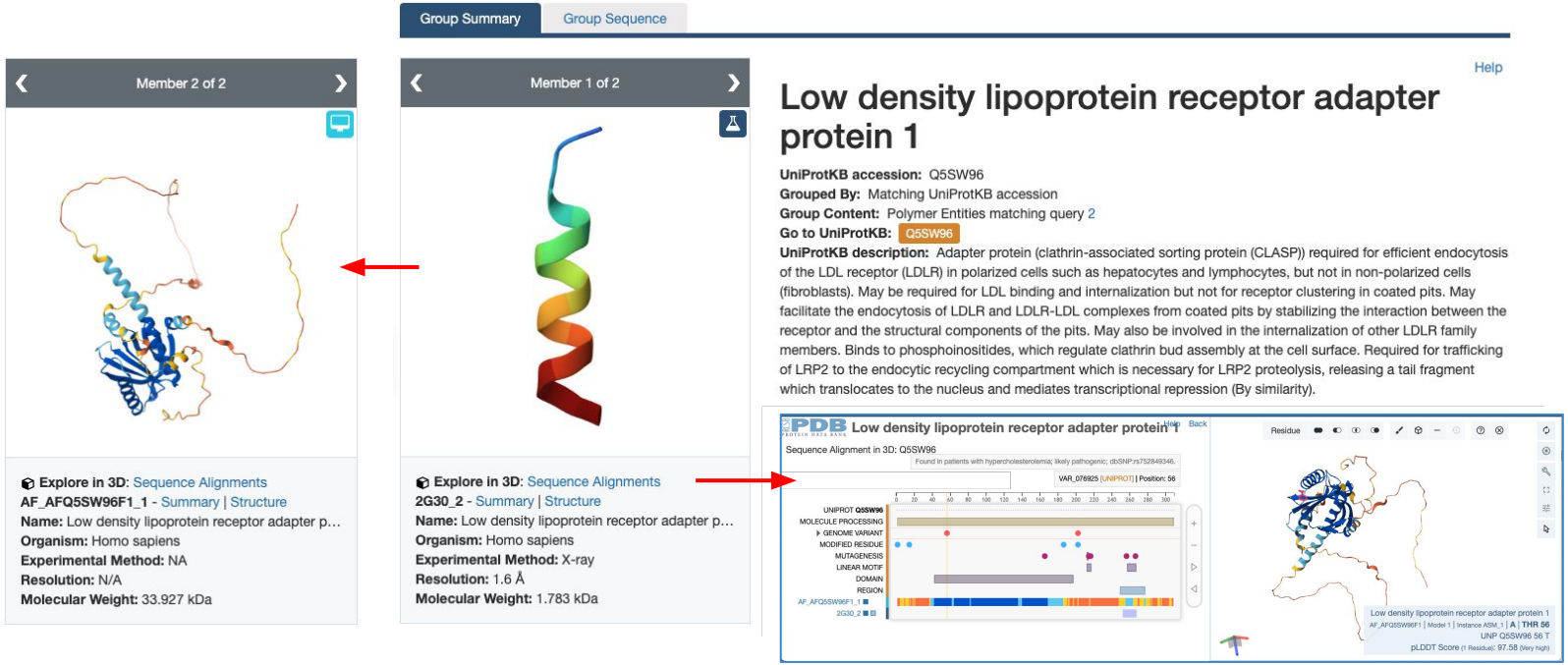
[Use Advanced Search tool to explore individual proteins]

Explore search results and navigate to the GSP



Click to navigate to the Group Summary page

Explore LDLRAP1 protein group



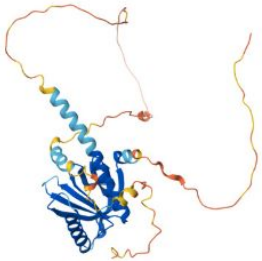
[Explore Group Summary]

[Explore Sequence Alignment in 3D]

Explore AlphaFold model of LDLRAP1 protein

Structure Summary | Structure | Annotations | Sequence | Genome

Assembly



Explore in 3D: Structure | Sequence Annotations

Global Symmetry: Asymmetric - C1
Global Stoichiometry: Monomer - A1

Find Similar Assemblies

Display Files | Download Files | Data API

AF_AFQ5SW96F1

COMPUTED STRUCTURE MODEL OF LOW DENSITY LIPOPROTEIN RECEPTOR ADAPTER PROTEIN 1

AlphaFold DB: AF-Q5SW96-F1

Released in AlphaFold DB: 2021-07-01
Last Modified in AlphaFold DB: 2022-09-30

Organism(s): Homo sapiens
UniProtKB: Q5SW96

There are no experimental data to verify the accuracy of this computed structure model. See Model Confidence metrics below for all regions of the polypeptide chain.

Click to search for proteins with similar 3D structure

Macromolecules

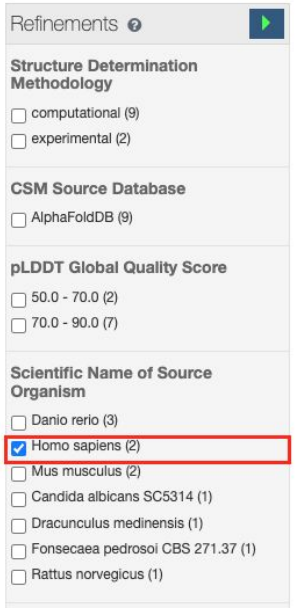
Find similar proteins by: Sequence (by identity cutoff) | 3D Structure

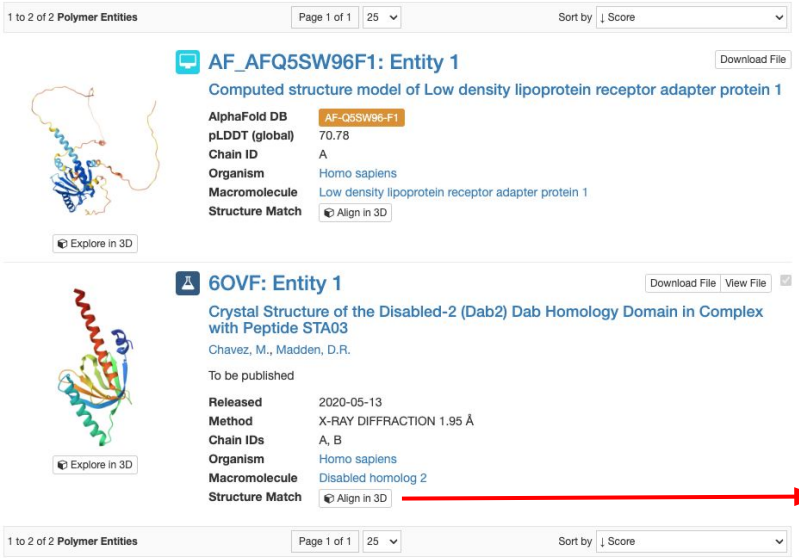
Entity ID: 1

Molecule	Chains	Sequence Length	Organism	Details	Image
Low density lipoprotein receptor adapter protein 1	A	308	Homo sapiens	Mutation(s): 0 Gene Names: LDLRAP1	

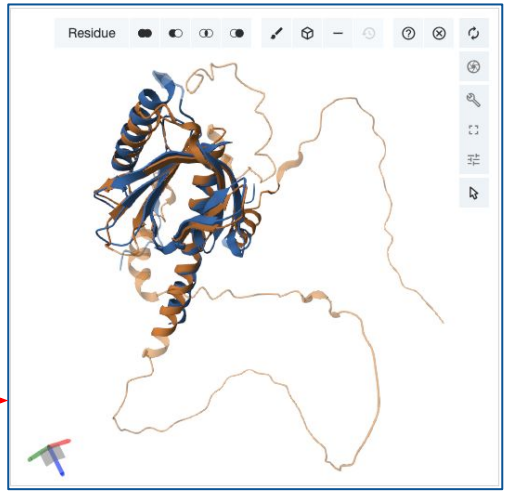
Explore Structure Summary

Combine Structure Similarity and text annotation searches





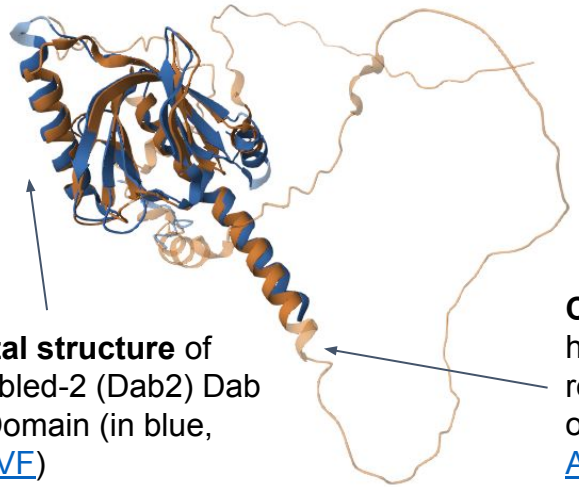
[Navigate to Pairwise Structure Alignment tool]



[Search for human proteins with similar shape]

Disabled homolog 2 (DAB2) protein

Entry	Chain	RMSD	TM-score	Identity	Equivalent Residues	Sequence Length	Modelled Residues
AF_AFQ5SW96F1	A	-	-	-	-	308	308
6OVF	A	2.43	0.43	18%	148	161	154



Experimental structure of human Disabled-2 (Dab2) Dab Homology Domain (in blue, PDB ID: 6OVF)

Computed structure model of human low density lipoprotein receptor adapter protein 1 (in orange, AlphaFoldDB ID: AF-Q5SW96-F1)

Disabled-2 (Dab2) is a signal-transduction protein that contains phosphotyrosine binding (PTB) domain and, like LDLRAP1, binds to and receives signals from members of the low-density lipoprotein receptor (LDLR) family

Summary

- RCSB.org offers seamless access to a suite of powerful tools designed to explore structure data
- Computed Structure Model (CSM) data provided alongside experimental structures helps to bridge gaps in structural data exploration for proteins with low experimental coverage

RCSB PDB
PROTEIN DATA BANK

rcsb.org

Case study of Class II aminoacyl-tRNA synthetases


Florian Kaiser


Sarah Krautwurst


Christoph Leberecht


Dirk Labudde


Sebastian Salentin


V. Joachim Haupt

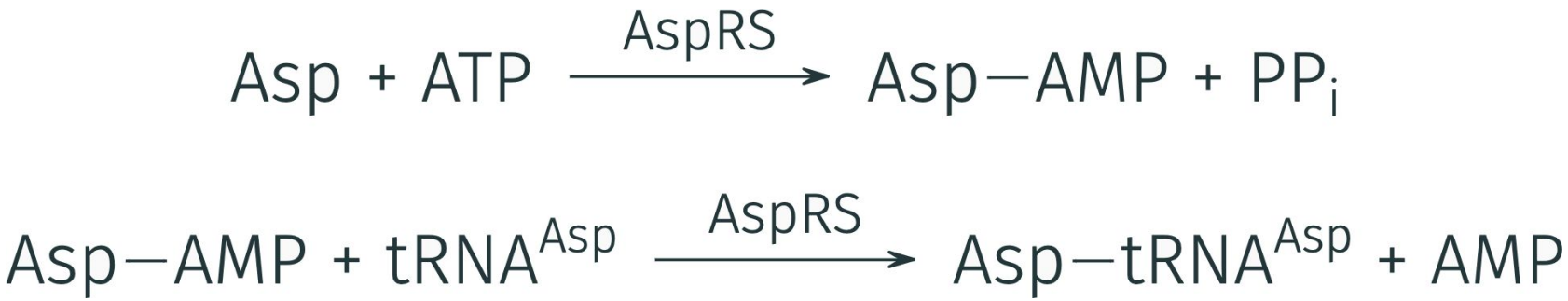

Michael Schroeder

Sebastian Bittrich, PhD

sebastian.bittrich@rcsb.org

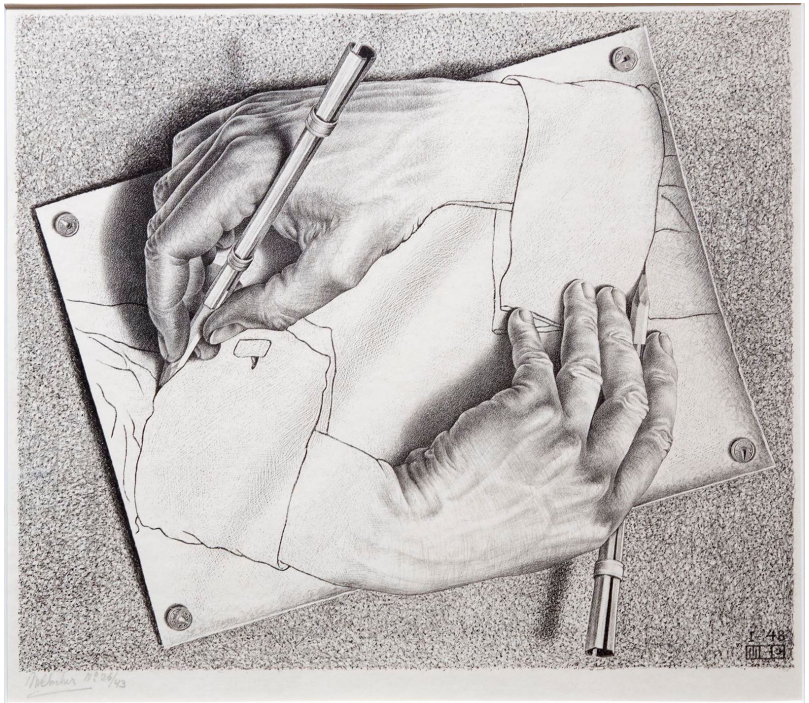
Aminoacyl-tRNA Synthetases (aaRS)

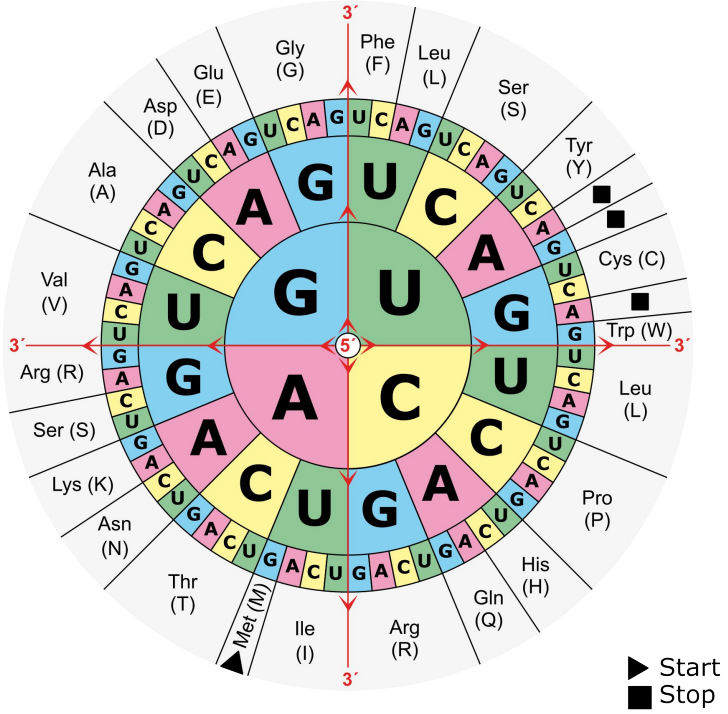
- Present in all organisms and all cells
- One implementation per amino acid
- 2-step reaction that uses ATP as energy source



Kaiser 2018, PLOS CB; Bittrich 2018, PLOS One; Kaiser 2020, Sci. Rep.

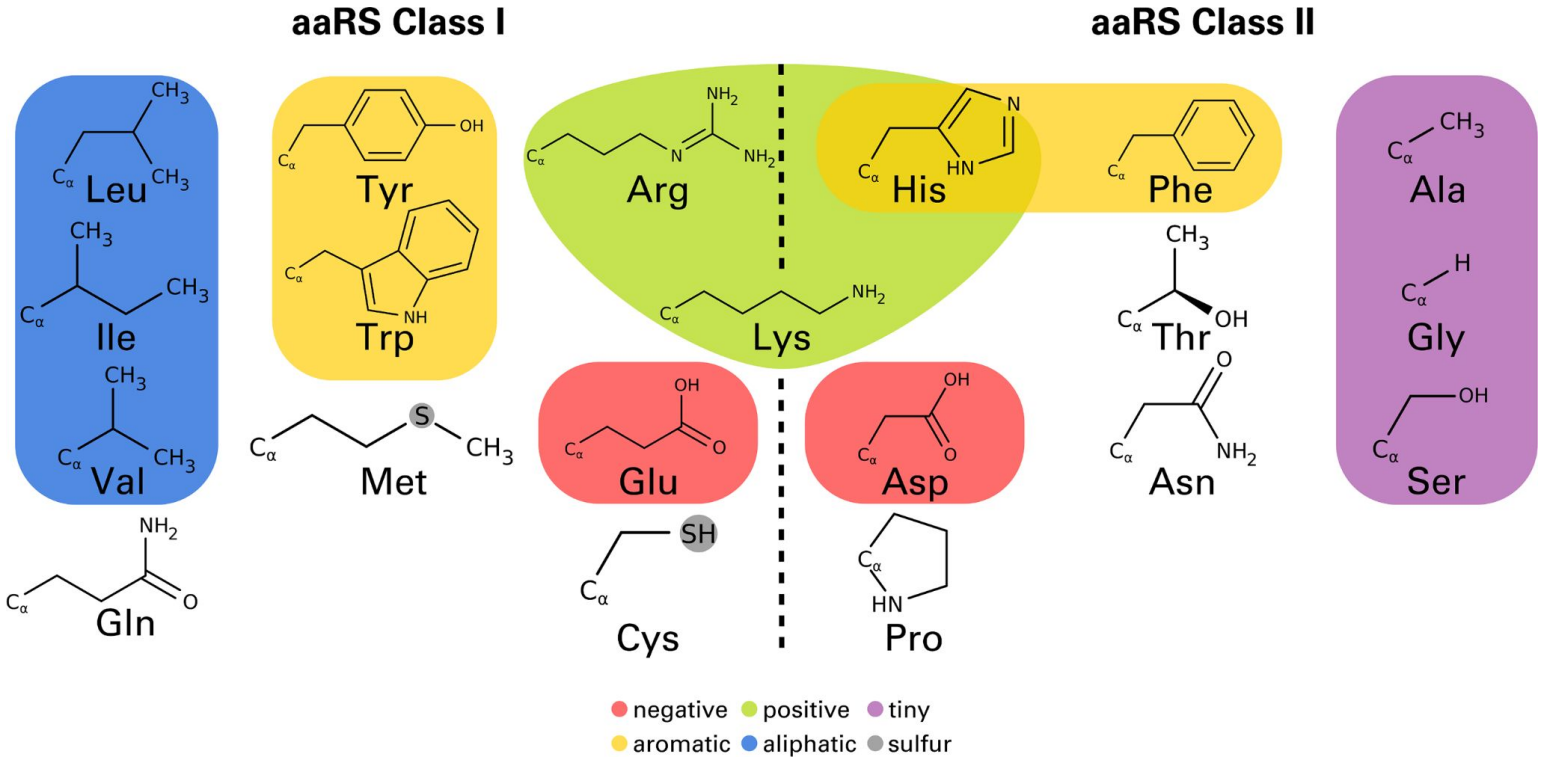
A Self-Referential System





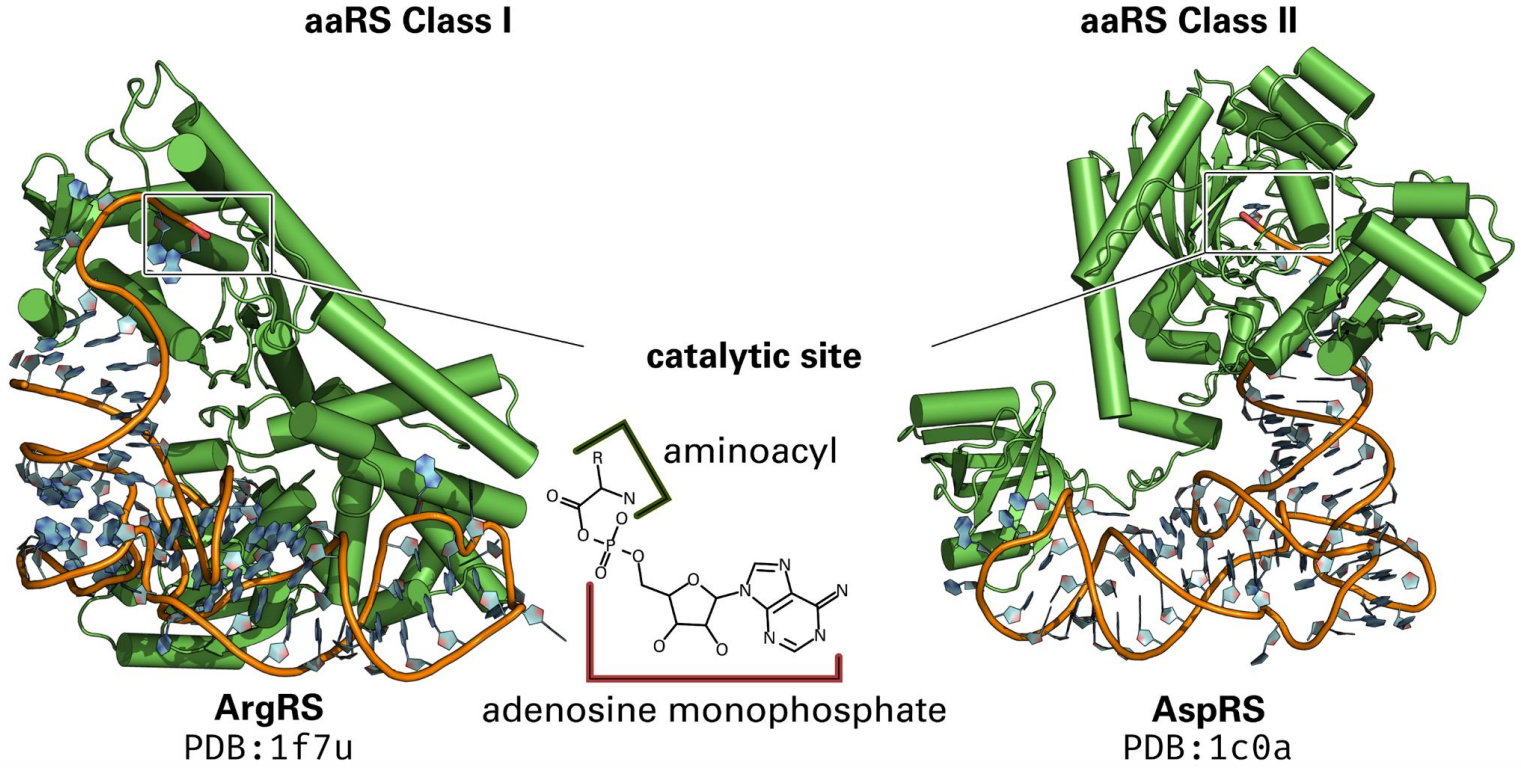
M. C. Escher “Drawing Hands” 1948; de.wikipedia.org/wiki/Code-Sonne

2 Classes of Aminoacyl-tRNA Synthetases



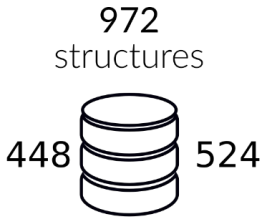
Kaiser 2018, PLOS CB

2 Classes of Aminoacyl-tRNA Synthetases



Kaiser 2018, PLOS CB

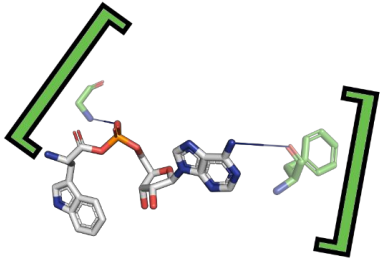
2018: First Archive-Wide Structural Analysis



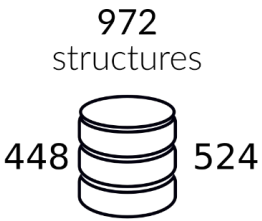
Kaiser 2018, PLOS CB

aaRS Class Signatures

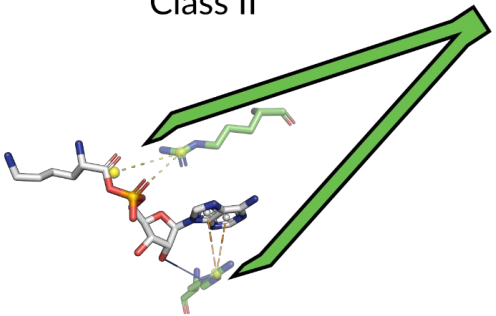
Class I



Backbone Brackets



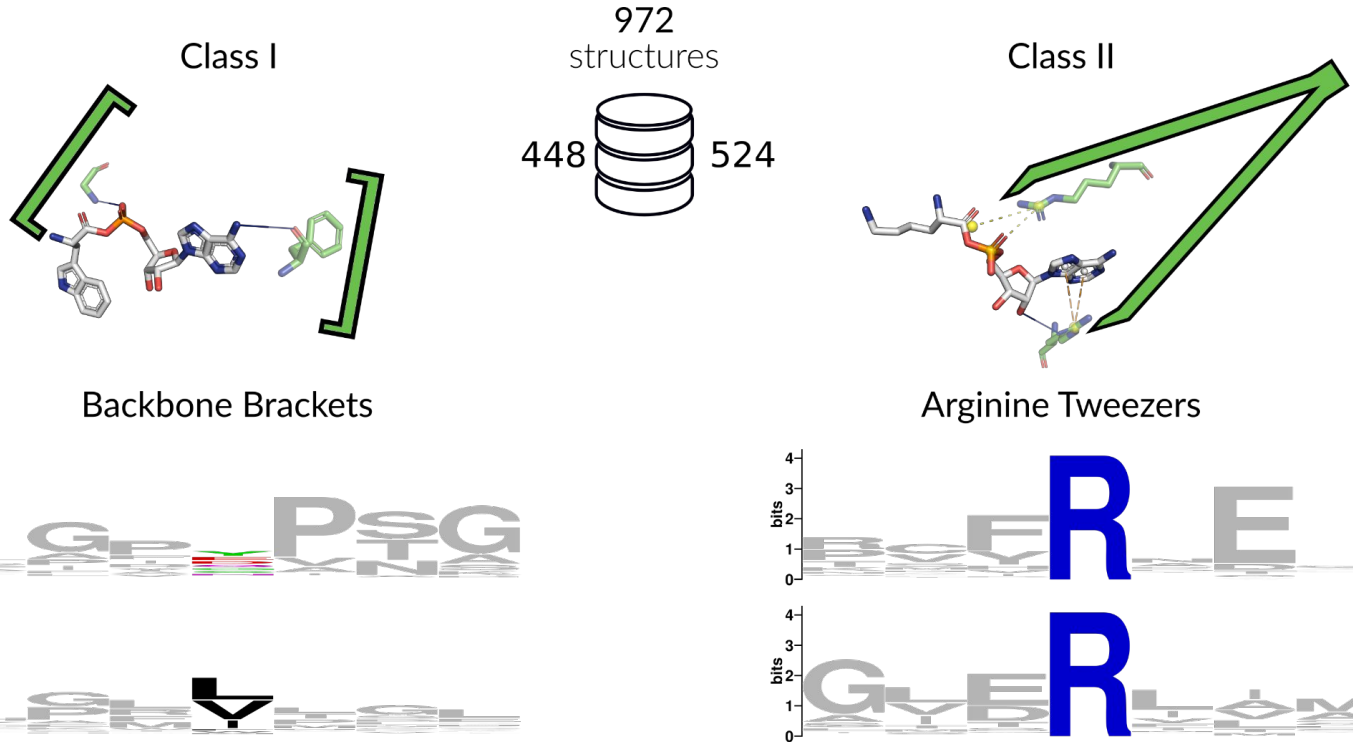
Class II



Arginine Tweezers

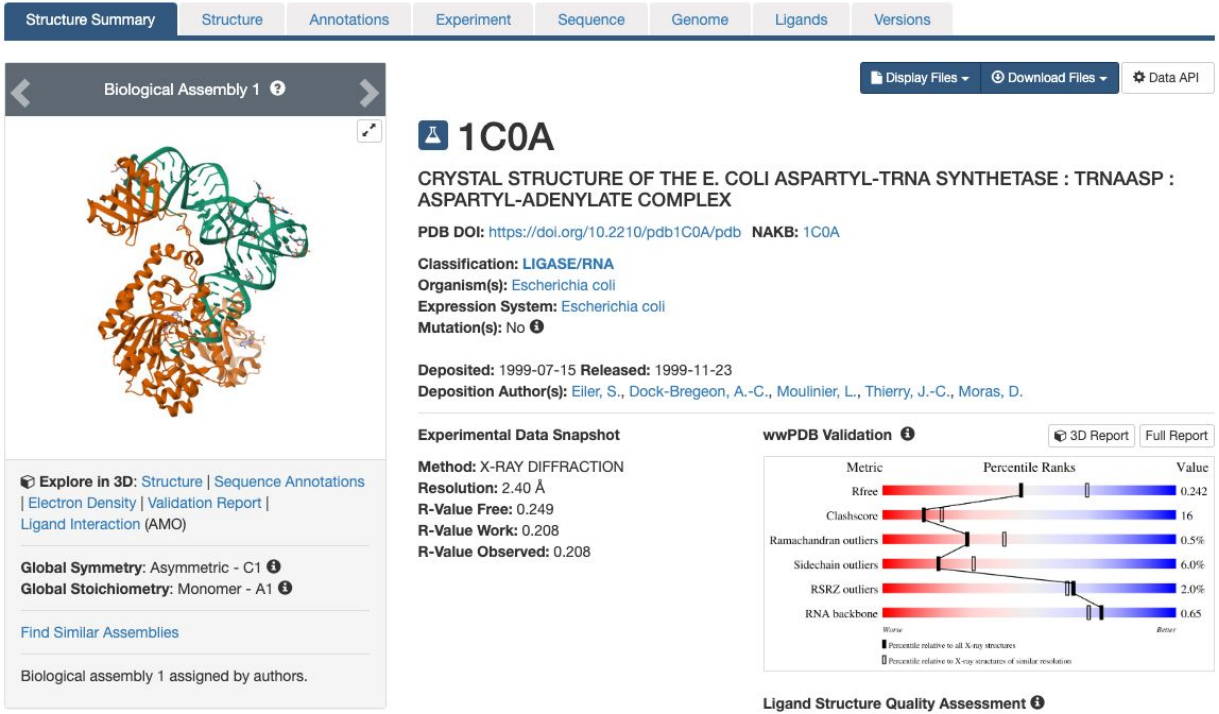
Kaiser 2018, PLOS CB

aaRS Class Signatures



Kaiser 2018, PLOS CB

Using 1c0a as Reference for Class II aaRS



RCSB.org

PDB-ID: 1c0a

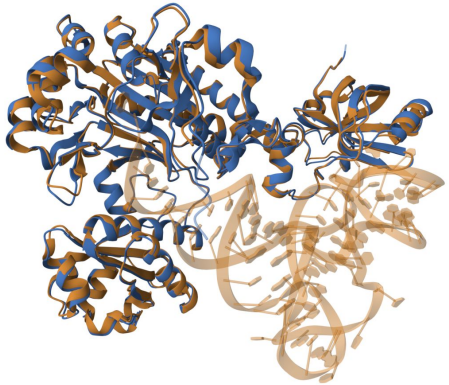
Revisiting Class II – 6 Years and 1M CSMs Later

Sequence Search

Example Queries
20 experimental
7 new since 2018

796 CSMs

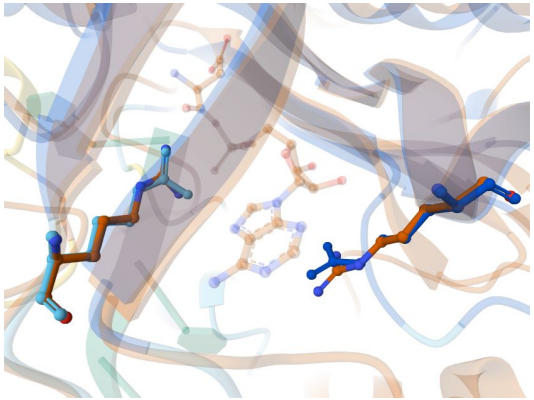
Global Structure Search



Example Queries
18 experimental
6 new since 2018

715 CSMs

Structure Motif Search



Example Queries
1,604 experimental
956 new since 2018

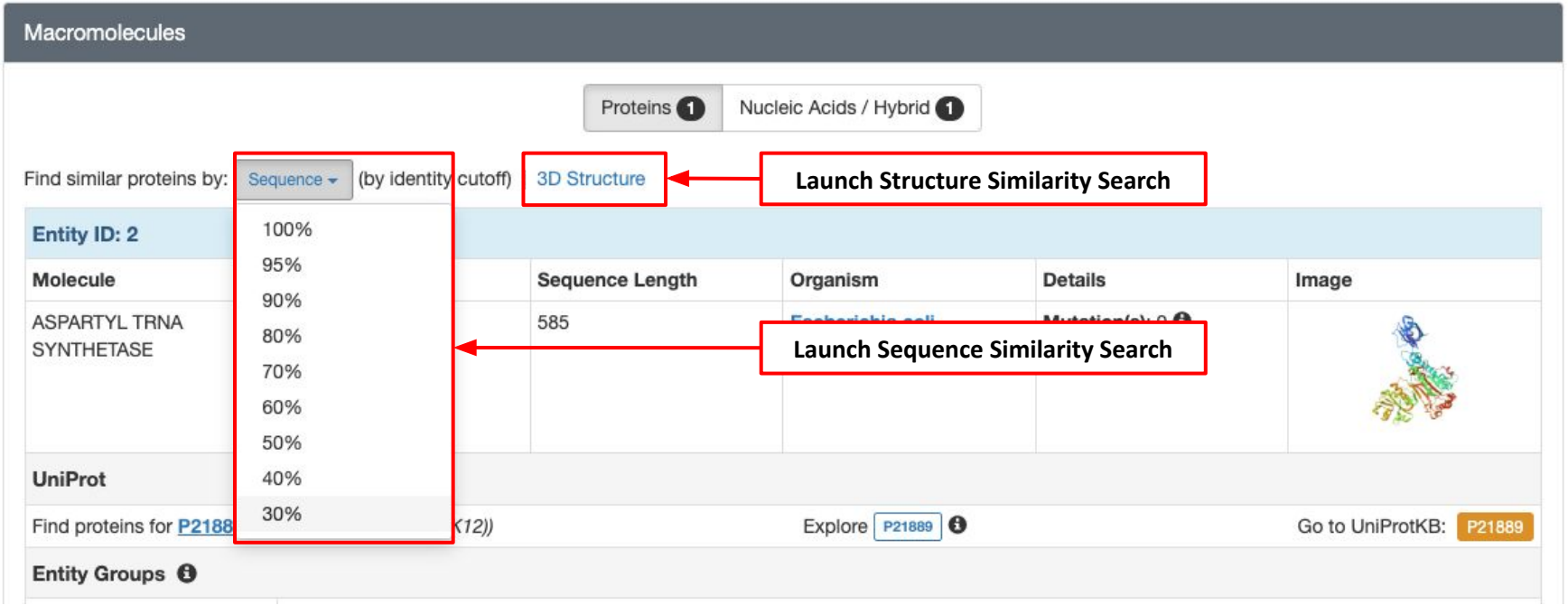
6,073 CSMs

Steinegger 2017, Nat. Biotechnol.; Guzenko 2020, PLOS CB; Bittrich 2020, PLOS CB

Live demo

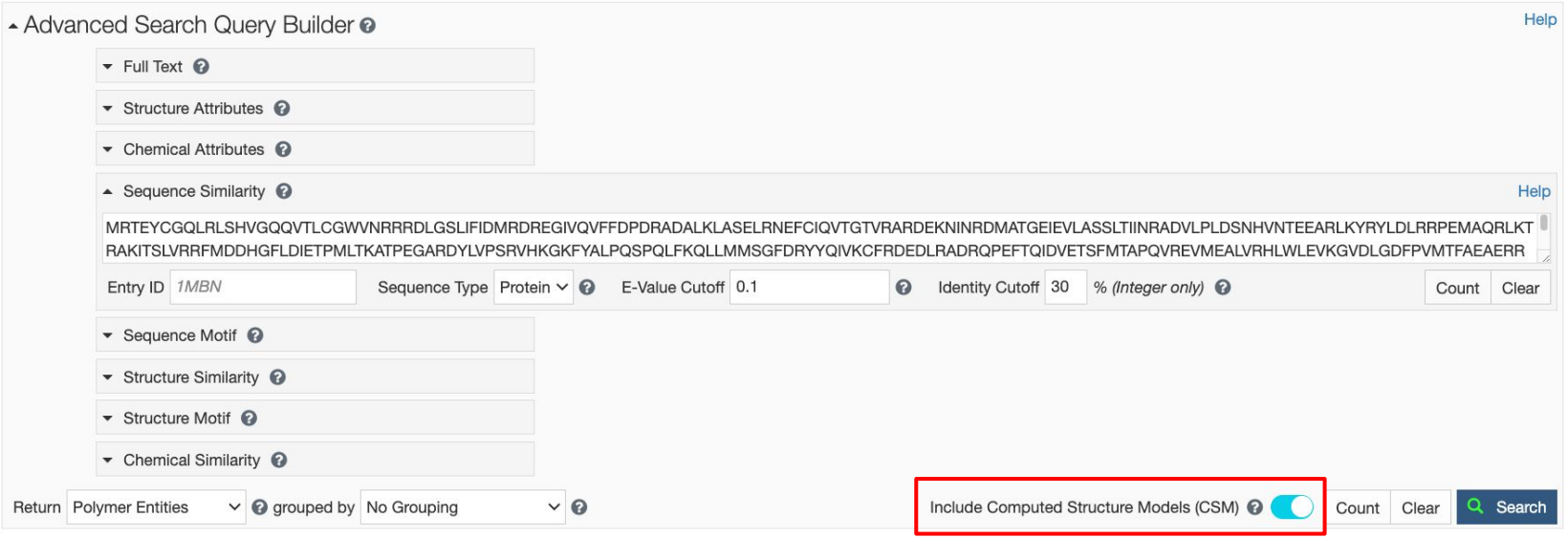
RCSB.org

Structure Summary Page: Similarity Queries



PDB-ID: 1c0a

Remember to Toggle CSMs if Relevant



CSMs aren't included by default when launching queries from Structure Summary Pages of experimentally-determined entries

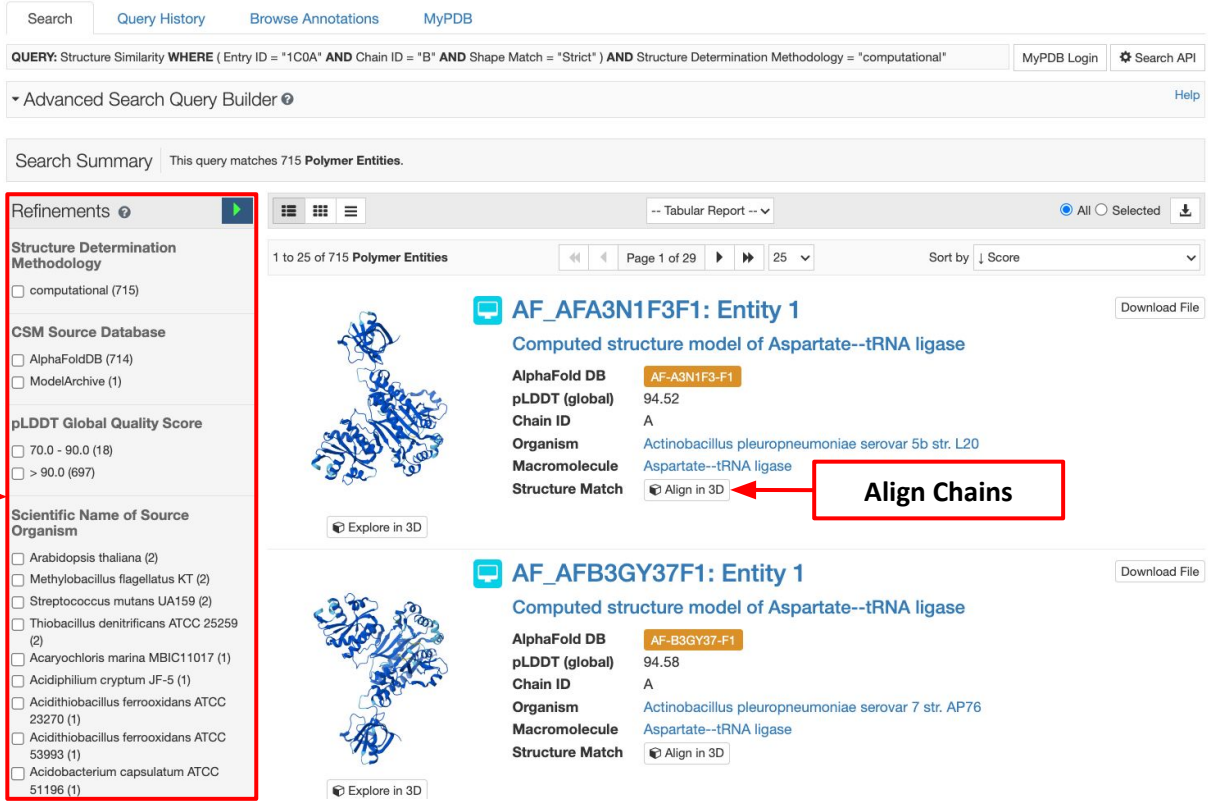
Toggle to Include CSMs

PDB-ID: 1c0a

Sequence Similarity Search Results

Query: Sequence Similarity Search on 1c0a_2

Structure Similarity Search Results

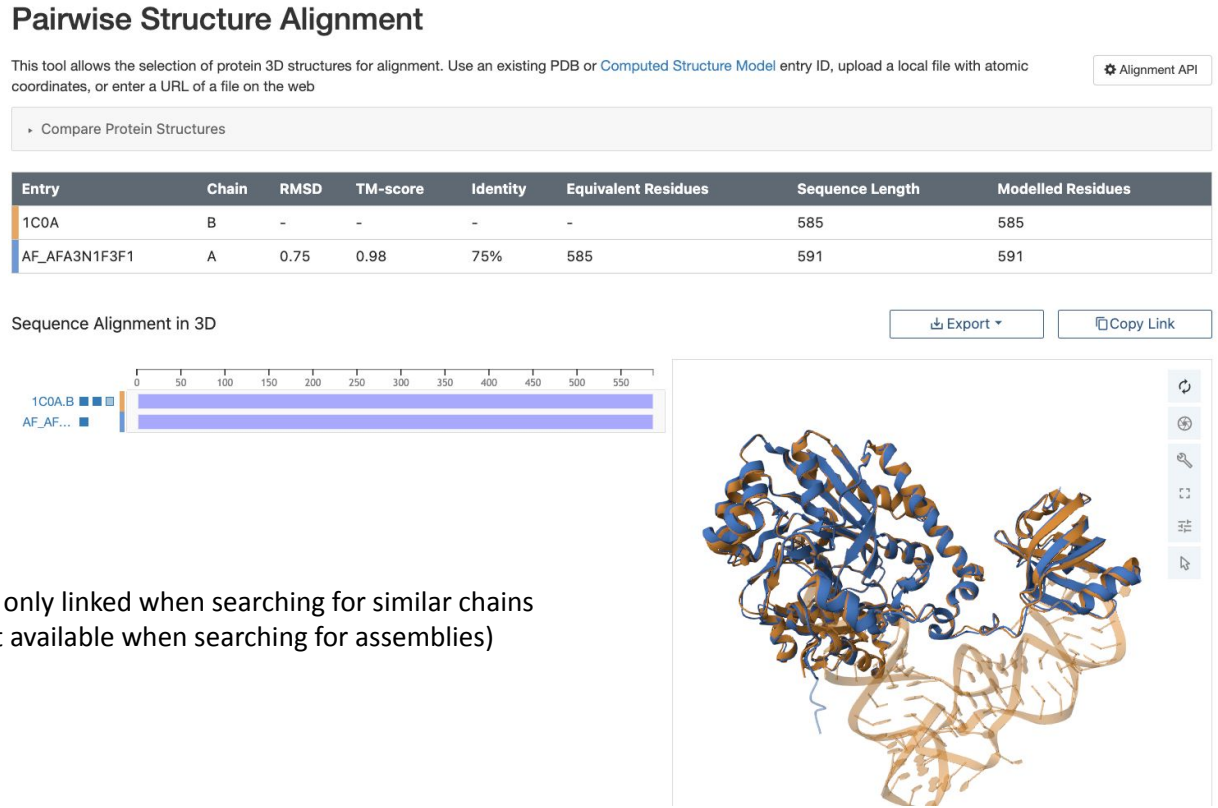


Refine Search

Align Chains

[Query: Structure Similarity Search on 1c0a.B]

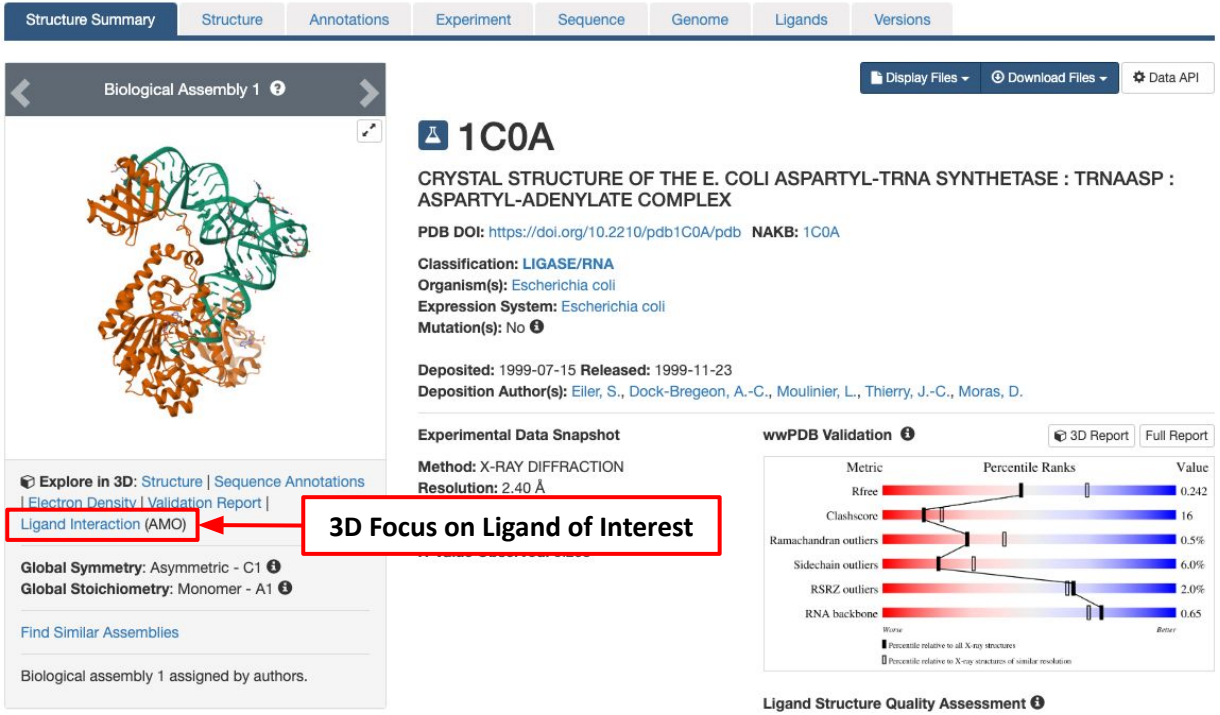
Alignment Visualization



This view is only linked when searching for similar chains (and not available when searching for assemblies)

Alignment Between 1c0a.B and AF_AFA3N1F3F1.A

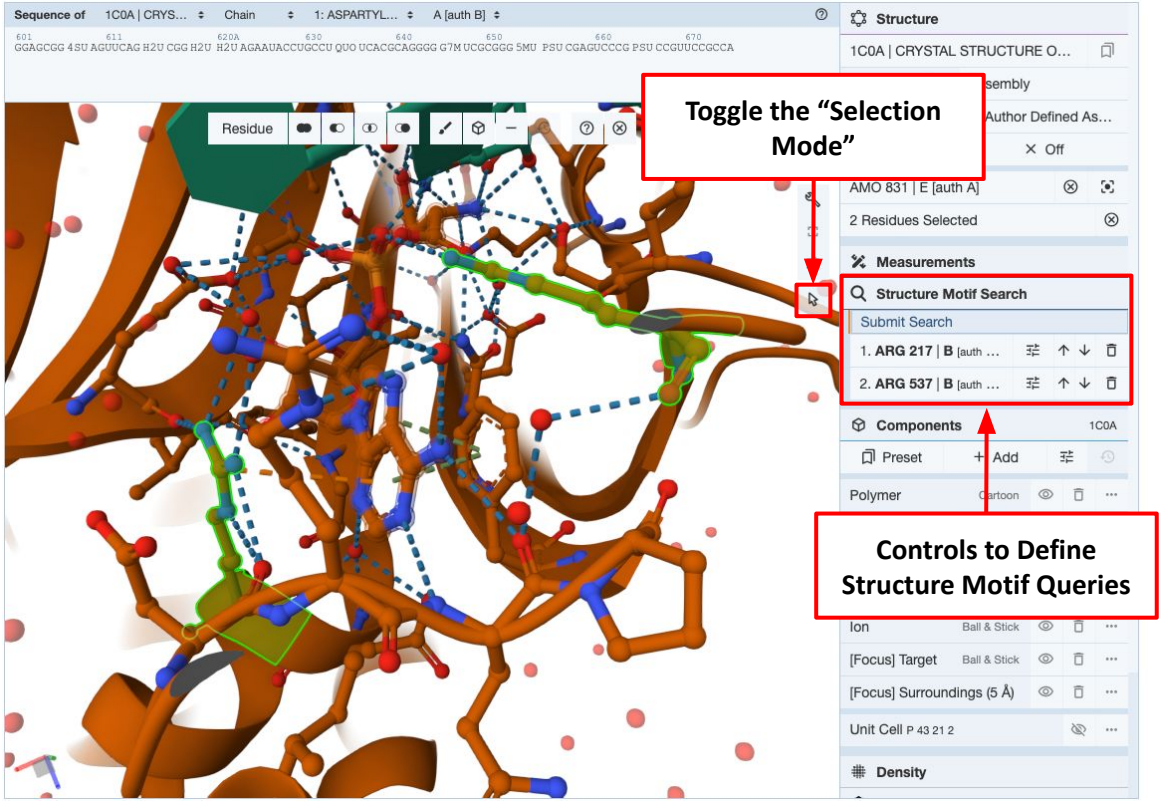
Structure Summary Page: Ligand Focus View



PDB-ID: 1c0a

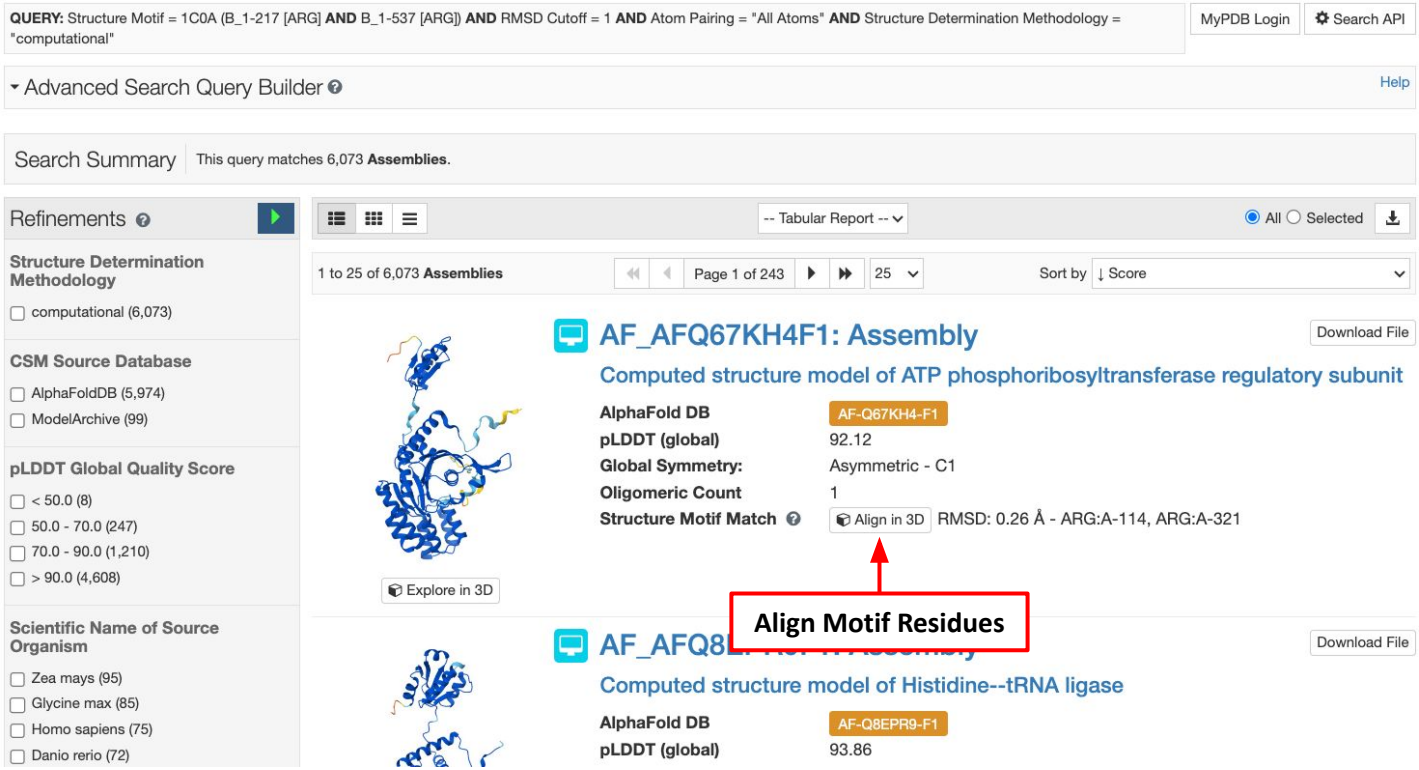
Define Structure Motifs using Mol*

Select residues in 3D to include in your search. Make sure to active the “Selection Mode”.



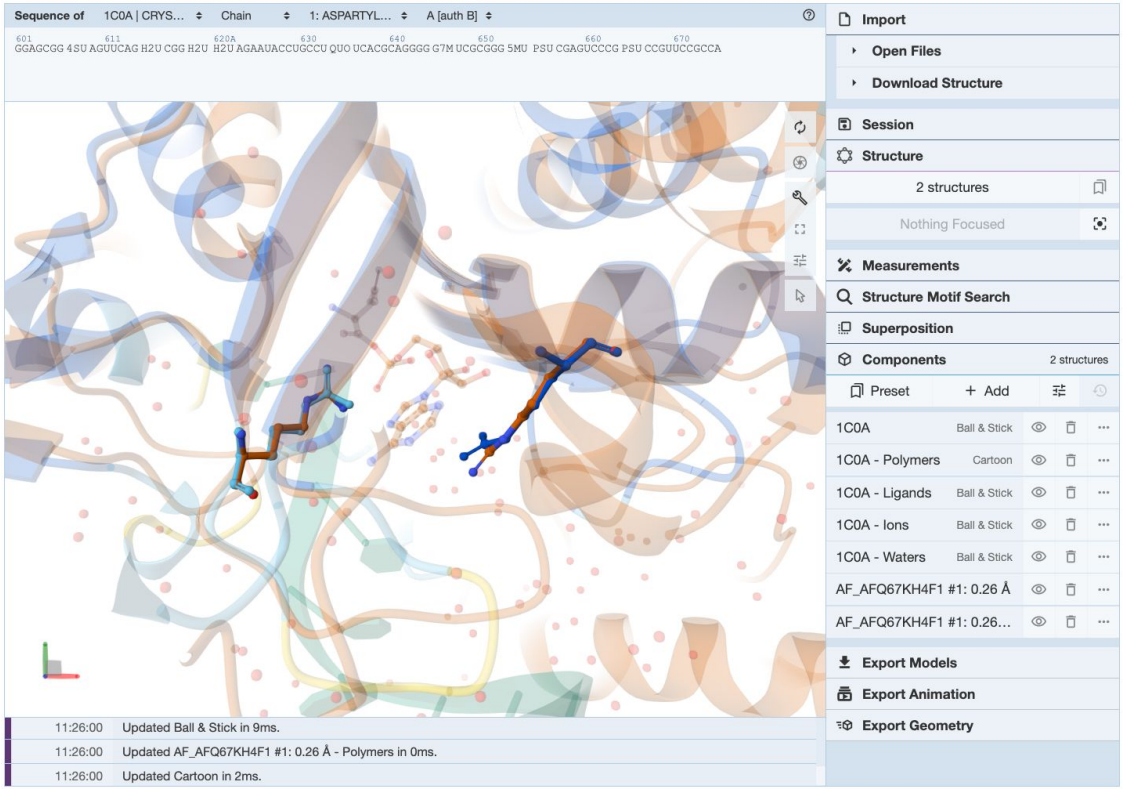
Ligand Focus View on AMO in 1c0a

Structure Motif Search Results



Query: Structure Motif Search on Arg-217 and Arg-537 from 1c0a.B

Structure Motif Alignment View



Motif Alignment Between 1c0a and AF_AFQ67KH4F1; Global Structure Alignment

Summary: RCSB.org is a powerful tool for CSM exploration

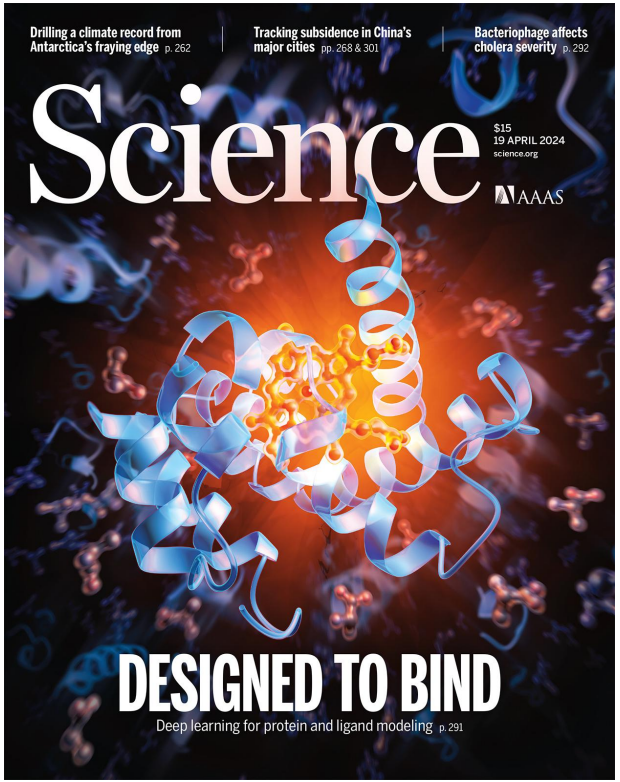
- Different ways to search for relevant Class II aaRS entries
  - Sequence search **primarily** finds aspartyl-RS
  - Global structure search **primarily** finds aspartyl-RS
  - Structure motif search finds **all Class II aaRS**, even for other amino acids plus ATP-binding paralogs
  - In general: Different types of searching may be relevant and RCSB.org is the **one-stop-shop to run different search types**

- Relevant CSMs identified from AlphaFold DB and ModelArchive
  - Numerous aaRS predictions available from model organisms and genomes with global health implications from AlphaFold DB
  - NB: ModelArchive contains i.a. predictions relevant for climate change

Varadi 2024, Nucleic Acids Res; Schwede 2009, Structure

Related Resources

- **AlphaFold**: Highly accurate protein structure prediction with AlphaFold. J. Jumper et al., Nature 596, 583-589 (2021).
  - **AlphaFold DB**: alphafold.ebi.ac.uk
- **RoseTTAfold**: Accurate prediction of protein structures and interactions using a three-track neural network. M. Baek et al., Science 373, 871-876 (2021).
- **ModelArchive:** Computational Structural Biology group at the SIB - Swiss Institute of Bioinformatics and the Biozentrum University of Basel
  - **ModelArchive**: modelarchive.org
- **ModelCIF**: An Extension of PDBx/mmCIF Data Representation for Computed Structure Models. B. Vallat et al., J Mol Biol, 168021 (2023).
- Open-access data: A cornerstone for artificial intelligence approaches to protein structure prediction. S. K. Burley, H. M. Berman, Structure 29, 515-520 (2021).



RCSB PDB Team

RCSB PDB PROTEIN DATA BANK

RCSB.ORG

info@rcsb.org

Core Operations Funding

US National Science Foundation (DBI-2321666), 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

RUTGERS THE STATE UNIVERSITY OF NEW JERSEY

UC San Diego

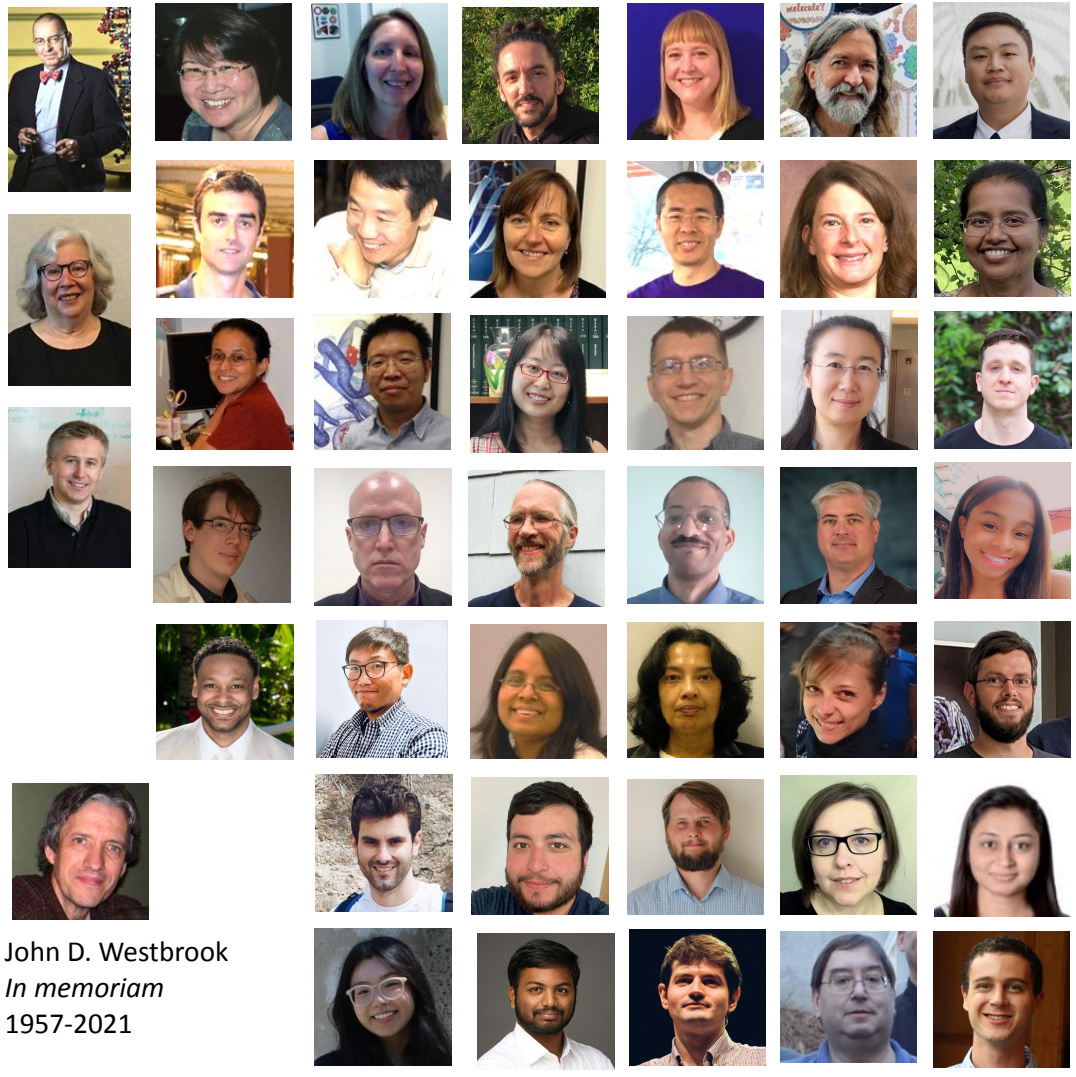
SDSC SAN DIEGO SUPERCOMPUTER CENTER

UCSF University of California San Francisco

WORLDWIDE PDB PROTEIN DATA BANK

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

Follow us



John D. Westbrook
In memoriam
1957-2021