VIRTUAL WEBINAR
A Deep Dive into
Computed Structure Model Exploration at RCSB.org
Tuesday April 30th 2024
9-10am Pacific | 12-1pm Eastern
Welcome

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

stephen.burley@rcsb.org
<table>
<thead>
<tr>
<th>Time</th>
<th>Session Title</th>
<th>Speaker(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>9:00 - 9:05 AM PDT</td>
<td>Welcome and Introduction</td>
<td>Stephen K. Burley, Director</td>
</tr>
<tr>
<td>9:05 - 9:25 AM PDT</td>
<td>A case study of Low-density Lipoprotein Receptor Adapter Protein 1 (LDLRAP1)</td>
<td>Yana Rose, Scientific Software Developer &amp; Data Architect</td>
</tr>
<tr>
<td>9:25 - 9:45 AM PDT</td>
<td>Case study of Class II aminoacyl-tRNA synthetases</td>
<td>Sebastian Bittrich, Scientific Software Developer</td>
</tr>
<tr>
<td>09:45 - 10:00 AM PDT</td>
<td>Q&amp;A</td>
<td>Jose Duarte, Brinda Vallat, Yana Rose, Sebastian Bittrich, Joan Segura</td>
</tr>
</tbody>
</table>
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
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

<table>
<thead>
<tr>
<th>Gene</th>
<th>UniProt ID</th>
<th>Protein</th>
<th>Protein length (residues, n)</th>
<th>Experimental coverage (residues, n)</th>
<th>Experimental coverage (residues, %)</th>
<th>AlphaFold (pLDDT ≥ 70) coverage (residues, n)</th>
<th>AlphaFold (pLDDT ≥ 70) coverage (residues, %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LDLRAP1</td>
<td>Q5SW96</td>
<td>Low density lipoprotein receptor adapter protein 1</td>
<td>308</td>
<td>16</td>
<td>5.2</td>
<td>159</td>
<td>51.6</td>
</tr>
<tr>
<td>LDLR</td>
<td>P01130</td>
<td>Low-density lipoprotein receptor</td>
<td>860</td>
<td>705</td>
<td>82.0</td>
<td>643</td>
<td>74.8</td>
</tr>
<tr>
<td>APOB</td>
<td>P04114</td>
<td>Apolipoprotein B-100</td>
<td>4563</td>
<td>N/A</td>
<td>N/A</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>APOE</td>
<td>P02649</td>
<td>Apolipoprotein E</td>
<td>317</td>
<td>298</td>
<td>94.0</td>
<td>218</td>
<td>68.8</td>
</tr>
<tr>
<td>PCSK9</td>
<td>Q8NB7</td>
<td>Proprotein convertase subtilisin/kexin type 9</td>
<td>692</td>
<td>642</td>
<td>92.8</td>
<td>563</td>
<td>81.4</td>
</tr>
</tbody>
</table>
Live demo

RCSB.org
Browse disease annotations on RCSB.org

The Disease Ontology browser organizes protein sequences by Mondo Disease Ontology terms describing disease definitions developed by the Monarch initiative team. The definitions have been mapped to UniProt sequences by the Knowledge Management Center (KMC) for the Illuminating the Druggable Genome (IDG) program and available for PDB entities mapped to UniProt.

Start typing the disease name and select the suggested term
Click to search

Browse Disease Annotations
Learn more about Mondo DO
Use Advanced Search to group structures based on the reference sequence

Retrieve results as groups based on UniProt KB Accession and include the CSMs

Use Advanced Search tool to explore individual proteins
Explore search results and navigate to the GSP

- **Group Cholesteryl ester transfer protein**
  - Group Total Size: 4
  - Polymer Entities: 4 matching query
  - Best Example: 2ORO_1 (Resolution: Best) Explore in 3D
  - Group ID: P11897
  - Description: Involved in the transfer of neutral lipids, including cholesteryl ester and triglyceride, ...

- **Group Signal-transducing adaptor protein 1**
  - Group Total Size: 3
  - Polymer Entities: 3 matching query
  - Best Example: 3MAZ_1 (Resolution: Best) Explore in 3D
  - Group ID: G9ULZ2
  - Description: In BCR signaling, appears to function as a docking protein acting downstream of TE...

- **Group Low density lipoprotein receptor adapter protein 1**
  - Group Total Size: 2
  - Polymer Entities: 2 matching query
  - Best Example: 2G30_2 (Resolution: Best) Explore in 3D
  - Group ID: Q65W96
  - Description: Adapter protein (clathrin-associated sorting protein (CLASPI) required for efficient en...
Explore LDLRAP1 protein group

Explore Group Summary

Low density lipoprotein receptor adapter protein 1

UniProtKB accession: Q5SWW6
Grouped By: Matching UniProtKB accession
Group Content: Polymer Entities matching query 2
Go to UniProtKB: Q5SWW6

UniProtKB description: Adapter protein (clathrin-associated sorting protein (CLASPI) required for efficient endocytosis of the LDL receptor (LDLR) in polarized cells such as hepatocytes and lymphocytes, but not in non-polarized cells (fibroblasts). May be required for LDL binding and internalization but not for receptor clustering in coated pits. May facilitate the endocytosis of LDL and LDL-LDL complex from coated pits by stabilizing the interaction between the receptor and the structural components of the pits. May also be involved in the internalization of other LDLR family members. Binds to phosphoinositides, which regulate clathrin bud assembly at the cell surface. Required for trafficking of LRP2 to the endocytic recycling compartment which is necessary for LRP2 proteolysis, releasing a tail fragment which translocates to the nucleus and mediates transcriptional repression (by similarity).

Explore Sequence Alignment in 3D
Explore AlphaFold model of LDLRAP1 protein

Click to search for proteins with similar 3D structure
Combine Structure Similarity and text annotation searches

Search for human proteins with similar shape

Navigate to Pairwise Structure Alignment tool
Disabled homolog 2 (DAB2) protein

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.

<table>
<thead>
<tr>
<th>Entry</th>
<th>Chain</th>
<th>RMSD</th>
<th>TM-score</th>
<th>Identity</th>
<th>Equivalent Residues</th>
<th>Sequence Length</th>
<th>Modelled Residues</th>
</tr>
</thead>
<tbody>
<tr>
<td>AF_AFQ65SW96F1</td>
<td>A</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>308</td>
<td>308</td>
</tr>
<tr>
<td>6OVF</td>
<td>A</td>
<td>2.43</td>
<td>0.43</td>
<td>18%</td>
<td>148</td>
<td>161</td>
<td>154</td>
</tr>
</tbody>
</table>

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)
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
Case study of Class II aminoacyl-tRNA synthetases
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

\[
\text{Asp} + \text{ATP} \xrightarrow{\text{AspRS}} \text{Asp-AMP} + \text{PP}_i
\]

\[
\text{Asp-AMP} + \text{tRNA}^{\text{Asp}} \xrightarrow{\text{AspRS}} \text{Asp-tRNA}^{\text{Asp}} + \text{AMP}
\]

A Self-Referential System

M. C. Escher “Drawing Hands” 1948; de.wikipedia.org/wiki/Code-Sonne
2 Classes of Aminoacyl-tRNA Synthetases

aaRS Class I
- Leu
- Tyr
- Trp
- Arg
- His
- Lys
- Met
- Glu
- Asp
- Cys
- Pro

aaRS Class II
- Ile
- Val
- Phe
- Ala
- Gly
- Ser
- Thr
- Asn

Legend:
- negative
- positive
- tiny
- aromatic
- aliphatic
- sulfur
2 Classes of Aminoacyl-tRNA Synthetases

aaRS Class I

ArgRS
PDB:1f7u

adenosine monophosphate

catalytic site

aminoacyl

aaRS Class II

AspRS
PDB:1c0a

Kaiser 2018, PLOS CB
2018: First Archive-Wide Structural Analysis

Class I

972 structures

448

Class II

524

Kaiser 2018, PLOS CB
aaRS Class Signatures

Class I
Backbone Brackets

Class II
Arginine Tweezers

972 structures
448
524

Kaiser 2018, PLOS CB
aaRS Class Signatures

Class I

972 structures

Class II

448 524

Backbone Brackets

Arginine Tweezers

Kaiser 2018, PLOS CB
Using 1c0a as Reference for Class II aaRS
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

Live demo

RCSB.org
Structure Summary Page: Similarity Queries

Launch Structure Similarity Search

Launch Sequence Similarity Search

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.

PDB-ID: 1c0a
Sequence Similarity Search Results

**Search Summary**

**Manipulate Search**

**Result Summary**

**Sequence Alignment**

**Query:** Sequence Similarity Search on 1c0a_2
Structure Similarity Search Results

Refine Search

Align Chains

Query: Structure Similarity Search on 1c0a.B
Alignment Visualization

Pairwise Structure Alignment

This tool allows the selection of protein 3D structures for alignment. Use an existing PDB or Computed Structure Model entry ID, upload a local file with atomic coordinates, or enter a URL of a file on the web.

- Compare Protein Structures

<table>
<thead>
<tr>
<th>Entry</th>
<th>Chain</th>
<th>RMSD</th>
<th>TM-score</th>
<th>Identity</th>
<th>Equivalent Residues</th>
<th>Sequence Length</th>
<th>Modelled Residues</th>
</tr>
</thead>
<tbody>
<tr>
<td>1C0A.B</td>
<td>B</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>585</td>
<td>585</td>
</tr>
<tr>
<td>AF_AFA3N1F3F1.A</td>
<td>A</td>
<td>0.75</td>
<td>0.98</td>
<td>75%</td>
<td>585</td>
<td>591</td>
<td>591</td>
</tr>
</tbody>
</table>

Sequence Alignment in 3D

⚠️ 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](https://alphafold.再来) and [ModelArchive](https://modelarchive.再来)
  - 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 DB: alphafold.ebi.ac.uk
- **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).
RCSB PDB Team

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

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

Follow us

In memoriam
John D. Westbrook
1957-2021