

National Institute of Allergy and Infectious Diseases

Rocky Mountain Laboratories Training

# Computational Structural Biology

September 2019

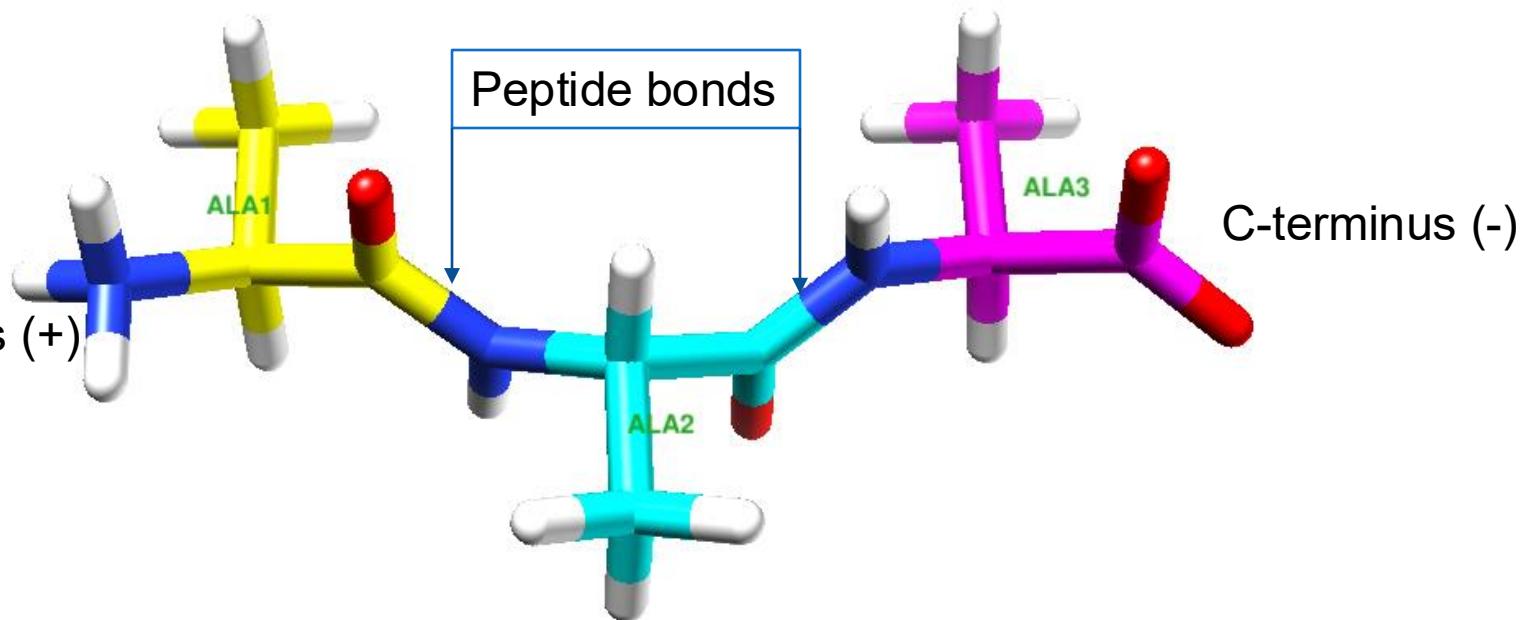
NIAID



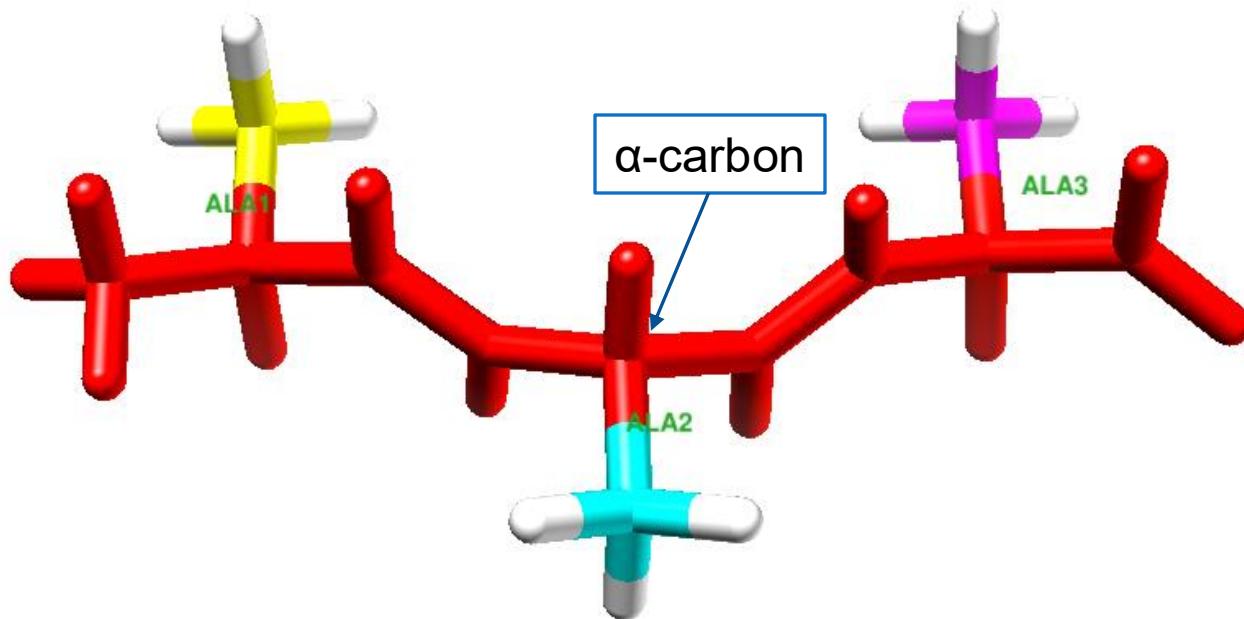
National Institute of  
Allergy and  
Infectious Diseases

**Phillip Cruz, Ph.D.**  
Computational Structural Biologist  
BCBB

# Quick review of protein structure

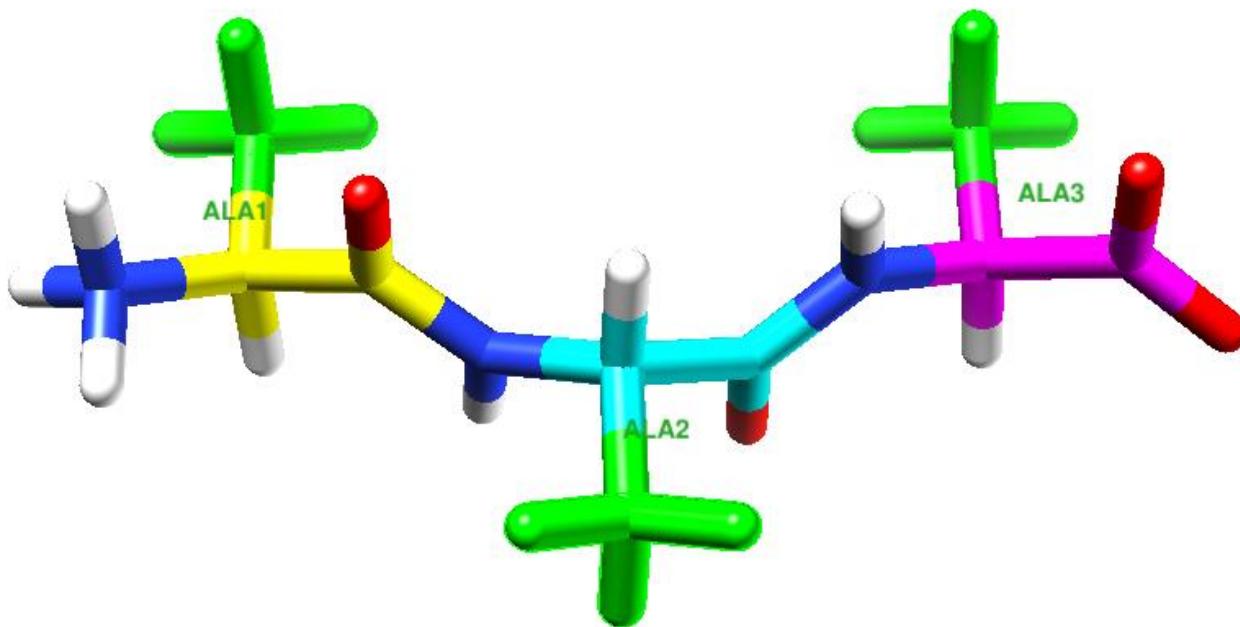


# Backbone



National Institute of  
Allergy and  
Infectious Diseases

# Sidechains



National Institute of  
Allergy and  
Infectious Diseases

# Intro to the PDB (Protein Data Bank)

- PDB refers to two things
  - Online repository of experimental biomacromolecule structures- <https://www.rcsb.org>
    - All entries represented by 4 character accession code
  - The file format(s) used to represent structures on the PDB web site- .pdb (or .cif, .xml)
    - .pdb is by far the most universal

# PDB Home page

RCSB PDB Deposit Search Visualize Analyze Download Learn More MyPDB

**PDB** 155169 Biological Macromolecular Structures Enabling Breakthroughs in Research and Education

Search by PDB ID, author, macromolecule, sequence, or ligands Go Advanced Search | Browse by Annotations

PDB-101 SPDB Worldwide Data Resources by RSCB Protein Data Bank Foundation

Welcome

Deposit Search Visualize Analyze Download Learn

A Structural View of Biology

This resource is powered by the Protein Data Bank archive-information about the 3D shapes of proteins, nucleic acids, and complex assemblies that helps students and researchers understand all aspects of biomedicine and agriculture, from protein synthesis to health and disease.

As a member of the wwPDB, the RCSB PDB curates and annotates PDB data. The RCSB PDB builds upon the data by creating tools and resources for research and education in molecular biology, structural biology, computational biology, and beyond.

Job Opportunities for Biocurators and Developers

JOIN OUR TEAM

August Molecule of the Month

Cyclin and Cyclin-dependent Kinase

Content List

Latest Entries As of Tuesday Aug 20 2019

6RHL Booy temperature data of Galectin-3C in Nitrocellulose

PDB Entry

Features & Highlights

Mandatory PDBx/mmCIF format files submission for MX depositions

Submission of PDBx/mmCIF format files for crystallographic depositions to the PDB will be mandatory from July 1<sup>st</sup> 2019 onward. PDB format files will no longer be accepted for deposition of structures solved by MX techniques.

Join Our Team as a Biocurator

Curate, validate, and standardize macromolecular structures from the PDB community at Rutgers, The State University of New Jersey.

News Publications

Poster Prize Awarded at ACA

Congratulations to Shaobo Dai for *Structural basis of SETD3 as an active histidine methytransferase SETD3* 08/20/2019

Join Our Biocuration Team 08/09/2019

Improve your previously released coordinates AND keep your original PDB ID with OneDep 07/31/2019

Education Corner: How Does Life Work? 07/23/2019



National Institute of  
Allergy and  
Infectious Diseases

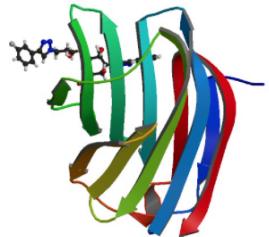
# PDB Home page

The screenshot shows the RCSB PDB homepage. At the top, there is a navigation bar with links for RCSB PDB, Deposit, Search, Visualize, Analyze, Download, Learn, and More. A search bar is located in the top right corner, with the placeholder text "Search by PDB ID, author, macromolecule, sequence, or ligands" and a "Go" button. Below the search bar is a map of the world. To the right of the map, the text "Search Field" is written in red. The main content area features a "Welcome" sidebar on the left with links for Deposit, Search, Visualize, Analyze, Download, and Learn. The main content area includes sections for "A Structural View of Biology", "Job Opportunities for Biocurators and Developers" (with a "JOIN OUR TEAM" button), "August Molecule of the Month" (showing a 3D model of Cyclin and Cyclin-dependent Kinase), "Latest Entries" (listing 6RHL and Booy temperature data of Galectin-3C), "Features & Highlights" (mentioning mandatory PDBx/mmCIF format files submission for MX depositions), "News" (listing a poster prize at ACA for Shaobo Dai), and "Publications" (listing a publication about SETD3). A "Contact Us" link is also present.

# PDB Structure page

Structure Summary    3D View    Annotations    Sequence    Sequence Similarity    Structure Similarity    Experiment

Biological Assembly 1



6RHL  
Room temperature data of Galectin-3C in complex with a pair of enantiomeric ligands: R enantiomer  
DOI: 10.2210/pdb6RHL/pdb

Classification: SUGAR BINDING PROTEIN  
Organism(s): Homo sapiens  
Expression System: Escherichia coli BL21(DE3)

Deposited: 2019-04-22 Released: 2019-08-21  
Deposition Author(s): Kumar, R., Verteramo, M.L., Nilsson, U.J., Logan, D.T.  
Funding Organization(s): Knut and Alice Wallenberg Foundation

Experimental Data Snapshot

Method: X-RAY DIFFRACTION  
Resolution: 1.299 Å  
R-Value Free: 0.170  
R-Value Work: 0.136

wwPDB Validation

Metric	Percentile Ranks	Value
Rfree	0	0.170
Clashscore	0	0
Ramachandran outliers	0	0
Sidetchain outliers	0.8%	0.8%
RSRZ outliers	2.2%	2.2%

This is version 1.0 of the entry. See complete history.

Literature

Are crystallographic B-factors suitable for calculating protein conformational entropy?

Caldararu, O., Kumar, R., Oksanen, E., Logan, D.T., Ryde, U.  
(2019) Phys Chem Chem Phys --, --  
PubMed: 31389436 Search on PubMed

Macromolecule Content

- Total Structure Weight: 16261.62



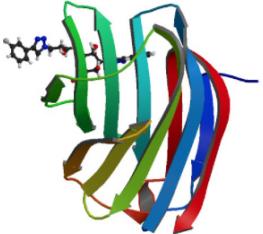
National Institute of  
Allergy and  
Infectious Diseases

NIAID

# PDB Structure page

Structure Summary    3D View    Annotations    Sequence    Sequence Similarity    Structure Similarity    Experiment

Biological Assembly 1



6RHL  
Room temperature data of Galectin-3C in complex with a pair of enantiomeric ligands: R enantiomer  
DOI: 10.2210/pdb6RHL/pdb

Classification: SUGAR BINDING PROTEIN  
Organism(s): Homo sapiens  
Expression System: Escherichia coli BL21(DE3)

Deposited: 2019-04-22 Released: 2019-08-21  
Deposition Author(s): Kumar, R., Verteramo, M.L., Nilsson, U.J., Logan, D.T.  
Funding Organization(s): Knut and Alice Wallenberg Foundation

Experimental Data Snapshot

Method: X-RAY DIFFRACTION  
Resolution: 1.299 Å  
R-Value Free: 0.170  
R-Value Work: 0.136

wwPDB Validation

Metric	Percentile Ranks	Value
Rfree	0	0.170
Clashscore	0	0
Ramachandran outliers	0	0
Sidetchain outliers	0.8%	0.8%
RSRZ outliers	2.2%	2.2%

This is version 1.0 of the entry. See complete history.

Literature

Are crystallographic B-factors suitable for calculating protein conformational entropy?  
Caldararu, O., Kumar, R., Oksanen, E., Logan, D.T., Ryde, U.  
(2019) Phys Chem Chem Phys --, --  
PubMed: 31389436 Search on PubMed



Download File



National Institute of  
Allergy and  
Infectious Diseases

NIAID

# PDB Structure page- 2

Macromolecules

Find similar proteins by: Sequence | Structure

Entity ID: 1

Molecule	Chains	Sequence Length	Organism	Details
Galectin-3	A	138	Homo sapiens	Mutation(s): 0 ⓘ Gene Names: LGALS3 (MAC2)

Find proteins for P17931 (Homo sapiens) Go to Gene View: [LGALS3](#) Go to UniProtKB: [P17931](#) Full Protein Feature View for [P17931](#)

Protein Feature View

P17931 - LEG3 HUMAN - Galectin-3  
Molec. Processing Motif: Galactin-3 6 X 9 AA tandem repeats of Y-P-G-X(3)-P-G-A  
UP Sites Secstruc PDB Validation 6RHLA  
Uniprot

Small Molecules

Ligands 1 Unique

ID	Chains	Name / Formula / InChI Key	2D Diagram & Interactions	3D Interactions
J0T <a href="#">Query on J0T</a>	A	(2-[S],3-[R],4-[S],5-[R],6-[R])-4-[4-(3-fluorophenyl)-1,2,3-triazol-1-yl]-2-[2-(R)]-3-[4-(3-fluorophenyl)-1,2,3-triazol-1-yl]-2-oxidanyl-propyl[sulfanyl-6-hydroxymethyl]oxane-3,5-diol C <sub>25</sub> H <sub>26</sub> F <sub>2</sub> N <sub>6</sub> O <sub>5</sub> S QSKQHLBUOMOHKU-LAMOWQQJSA-N		<a href="#">Ligand Interaction</a>

Links to UniProt



National Institute of  
Allergy and  
Infectious Diseases

NIAID

# Related UniProt Entry

UniProtKB - P17931 (LEG3\_HUMAN)

Display

Entry

Protein | Galectin-3  
Gene | LGALS3  
Organism | Homo sapiens (Human)  
Status | Reviewed - Annotation score: 5/5 - Experimental evidence at protein level<sup>i</sup>

Function<sup>i</sup>

None

Function<sup>i</sup>

Galactose-specific lectin which binds IgE. May mediate with the alpha-3, beta-1 integrin the stimulation by CSPG4 of endothelial cells migration. Together with DMBT1, required for terminal differentiation of columnar epithelial cells during early embryogenesis (By similarity). In the nucleus: acts as a pre-mRNA splicing factor. Involved in acute inflammatory responses including neutrophil activation and adhesion, chemoattraction of monocytes/macrophages, opsonization of apoptotic neutrophils, and activation of mast cells. Together with TRIM16, coordinates the recognition of membrane damage with mobilization of the core autophagy regulators ATG16L1 and BECN1 in response to damaged endomembranes. [By similarity](#) [4 Publications](#)

GO - Molecular function<sup>i</sup>

- IgE binding [Source: BHF-UCL](#)
- RNA binding [Source: UniProtKB](#)
- carbohydrate binding [Source: ProtInc](#)
- chemoattractant activity [Source: BHF-UCL](#)
- laminin binding [Source: BHF-UCL](#)
- oligosaccharide binding [Source: UniProtKB](#)
- protein phosphatase binding [Source: ARUK-UCL](#)
- protein phosphatase inhibitor activity [Source: ARUK-UCL](#)

Complete GO annotation on QuickGO ...

GO - Biological process<sup>i</sup>

- RNA splicing [Source: UniProtKB-KW](#)
- antimicrobial humoral immune response mediated by antimicrobial peptide [Source: UniProtKB](#)
- eosinophil chemotaxis [Source: BHF-UCL](#)
- epithelial cell differentiation [Source: UniProtKB](#)

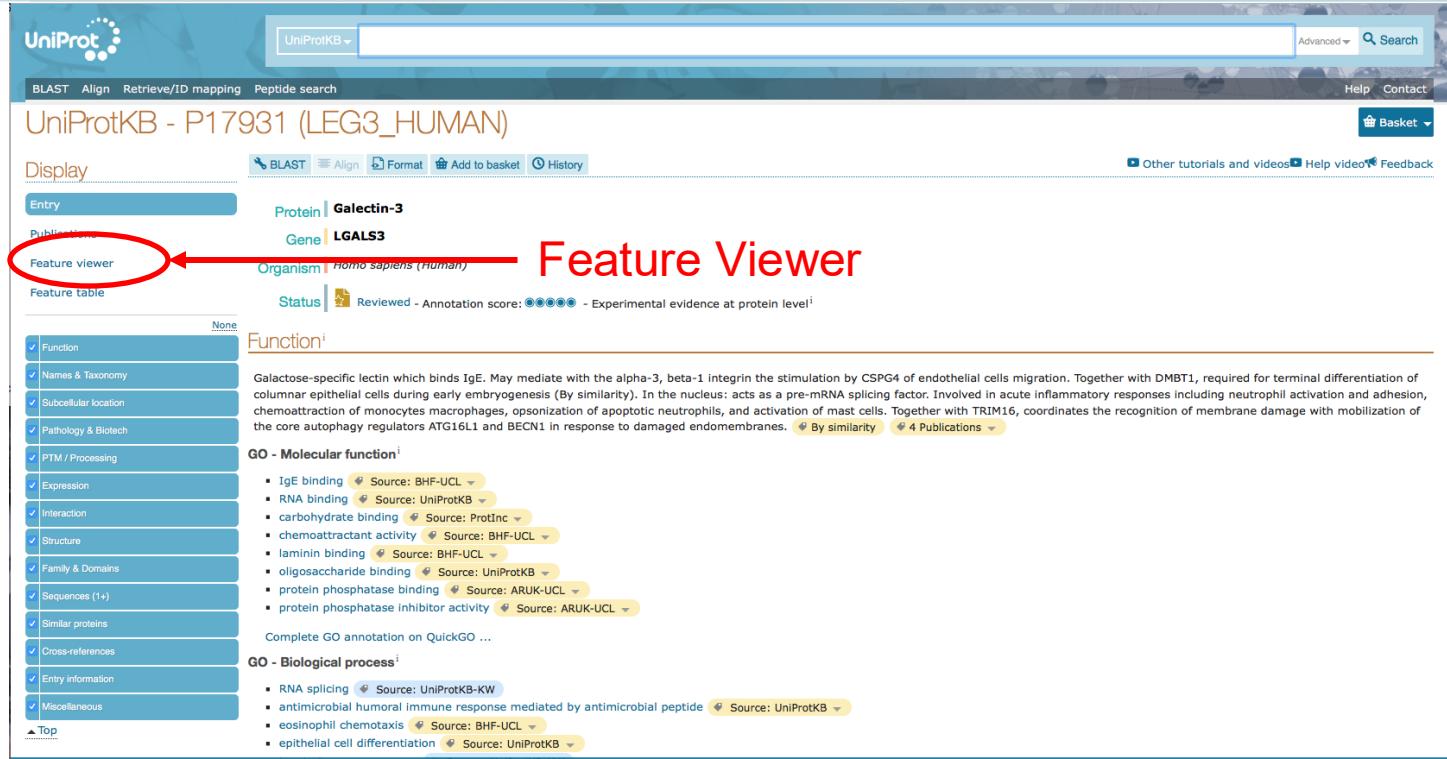
Cross-references

Entry information

Miscellaneous

Top

# Related UniProt Entry



UniProtKB - P17931 (LEG3\_HUMAN)

Display

Entry

Feature viewer

Feature table

Protein | Galectin-3  
Gene | LGALS3  
Organism | *Homo sapiens (Human)*

Status | Reviewed - Annotation score: 5/5 - Experimental evidence at protein level!

Function

None

- Function
- Names & Taxonomy
- Subcellular location
- Pathology & Biotech
- PTM / Processing
- Expression
- Interaction
- Structure
- Family & Domains
- Sequences (1+)
- Similar proteins
- Cross-references
- Entry information
- Miscellaneous

GO - Molecular function<sup>i</sup>

- IgE binding
- RNA binding
- carbohydrate binding
- chemoattractant activity
- laminin binding
- oligosaccharide binding
- protein phosphatase binding
- protein phosphatase inhibitor activity

Complete GO annotation on QuickGO ...

GO - Biological process<sup>i</sup>

- RNA splicing
- antimicrobial humoral immune response mediated by antimicrobial peptide
- eosinophil chemotaxis
- epithelial cell differentiation

Top

BLAST Align Retrieve/ID mapping Peptide search Advanced Search Help Contact Basket Other tutorials and videos Help video Feedback

Feature Viewer

# UniProt Feature Viewer

The screenshot shows the UniProt Feature Viewer interface for the protein P17931 (LEG3\_HUMAN). The top navigation bar includes links for BLAST, Align, Retrieve/ID mapping, Peptide search, Advanced search, Help, and Contact. A sidebar on the left provides options for Display (Entry, Publications, Feature viewer, Feature table), Domains & sites, Molecule processing, PTM, Sequence information, Structural features, Proteomics, Antigenic sequences, and Variants. The main content area displays a protein structure with a yellow highlighted residue at position 203. A tooltip labeled "LEU A 203 [TAK3]" is shown above the residue. Above the structure, several tracks show domain distribution, PTM sites, and sequence features across the protein's length (1-250). Below the structure is a table of related PDB entries:

PDB Entry	Method	Resolution	Chain	Positions	Links
1A3K	X-ray	2.10 Å	A	114-250	PDBe RCSB ... PDBj PDBsu...
1KJL	X-ray	1.40 Å	A	105-250	PDBe RCSB ... PDBj PDBsu...
1KJR	X-ray	1.55 Å	A	105-250	PDBe RCSB ... PDBj PDBsu...
2NMN	X-ray	2.45 Å	A	113-250	PDBe RCSB ... PDBj PDBsu...
2NMO	X-ray	1.35 Å	A	113-250	PDBe RCSB ... PDBj

At the bottom, a banner informs users about the update of the Privacy Notice to comply with GDPR, with a "Do not show this banner again" link.

List of related protein structures, with links to PDB.

Highlight residue with mouse (yellow) and get feedback label.



National Institute of  
Allergy and  
Infectious Diseases

NIAID

# UniProt Feature Viewer- Variants

The screenshot shows the UniProt Feature Viewer interface for the protein P17931 (LEG3\_HUMAN). The left sidebar has a red circle highlighting the 'Variants' link under the 'Feature viewer' section. The main panel displays a protein structure with a yellow arrow pointing to a specific residue. Below the structure is a table of PDB entries:

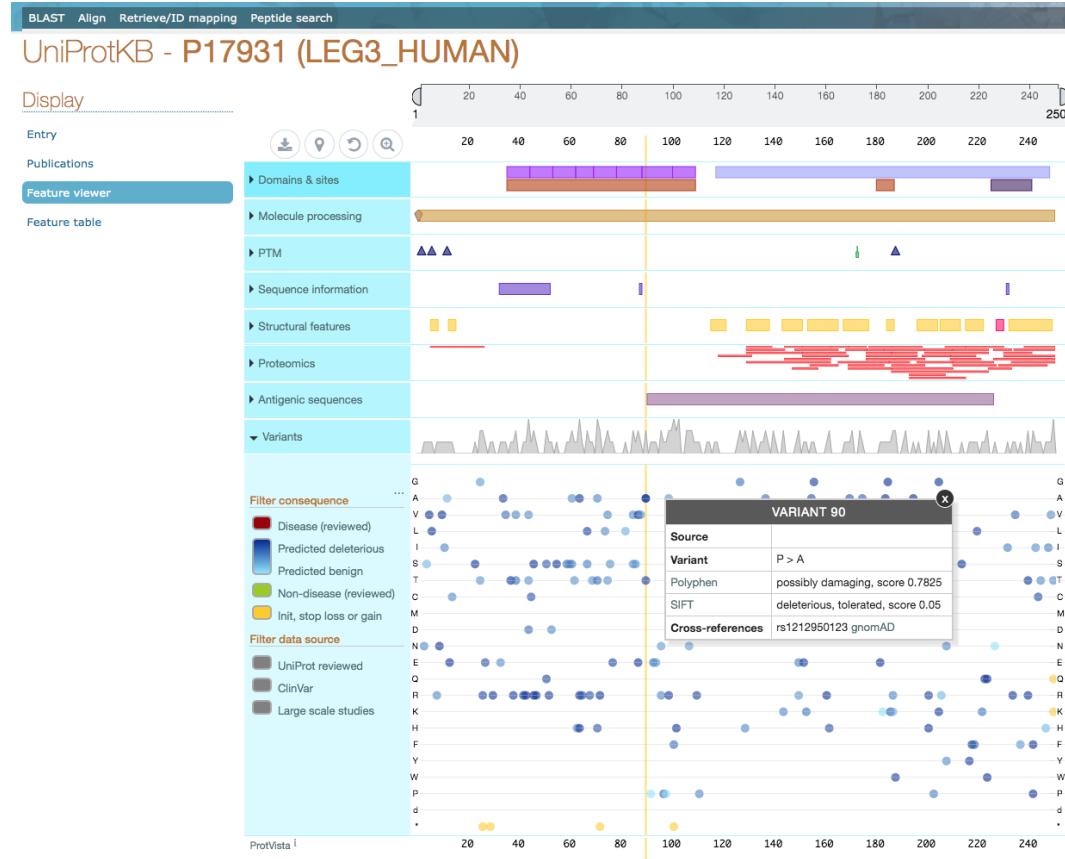
PDB Entry	Method	Resolution	Chain	Positions	Links
1A3K	X-ray	2.10 Å	A	114-250	PDBe RCSB ... PDBj PDBsu...
1KJL	X-ray	1.40 Å	A	105-250	PDBe RCSB ... PDBj PDBsu...
1KJR	X-ray	1.55 Å	A	105-250	PDBe RCSB ... PDBj PDBsu...
2NMN	X-ray	2.45 Å	A	113-250	PDBe RCSB ... PDBj PDBsu...
2NMO	X-ray	1.35 Å	A	113-250	PDBe RCSB ... PDBj



National Institute of  
Allergy and  
Infectious Diseases

NIAID

# UniProt Feature Viewer- Variants



National Institute of  
Allergy and  
Infectious Diseases

NIAID

# Common Structure File Formats

- Proteins
  - PDB (Protein Data Bank)
  - mol2 (SYBYL)
- Ligands (small molecules)
  - SDF (structure-data file)
  - mol2
  - SMILES (2D information only)
- Proteins and Ligands
  - PDB (but issues with ligands)
  - mol2

# A Cautionary Tale: Hydrogen atoms and PDB files

- Most structures in the PDB are from x-ray crystallography
  - Typically no hydrogen atoms are present
    - Water molecules consist of a single oxygen atom
- PDB files don't indicate "bond order", that is, whether a given bond is single, double, triple, etc.

Problem: Without hydrogen atoms and bond orders, the structure of ligands can be ambiguous.

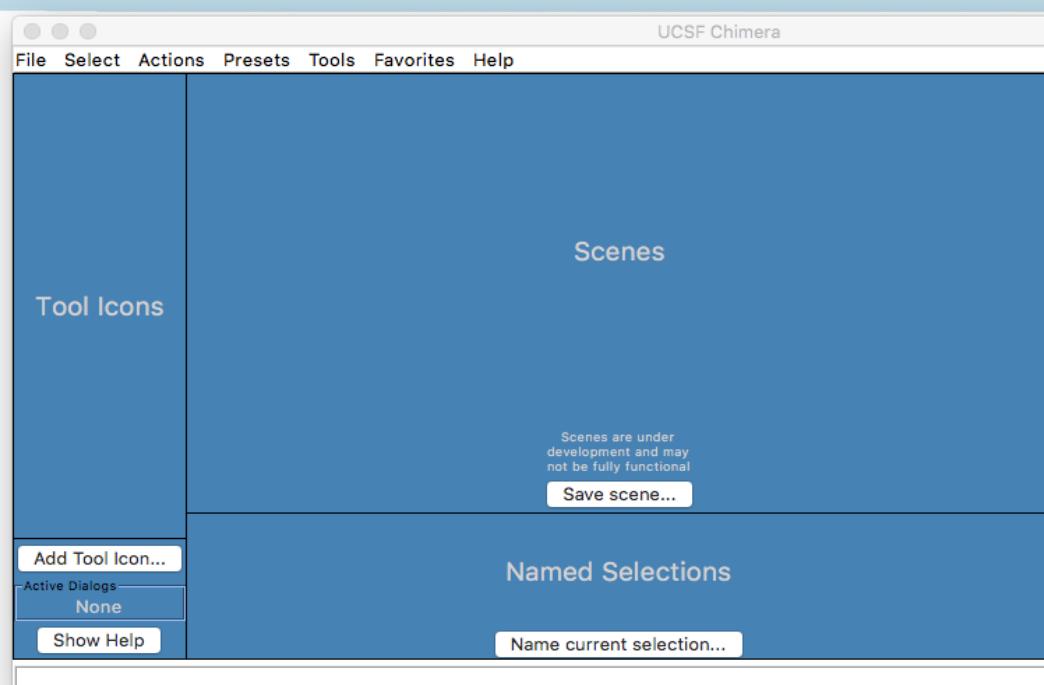
Molecular modeling programs make educated guesses, but can be wrong.

# UCSF Chimera- Molecular Modeling

- Download from:  
<https://www.cgl.ucsf.edu/chimera/download.html>
- Multiplatform:
  - Windows
  - Mac
  - Linux
- Graphical interface and command line

# Chimera- Exercise 1

## Create “review of protein structure” scenes

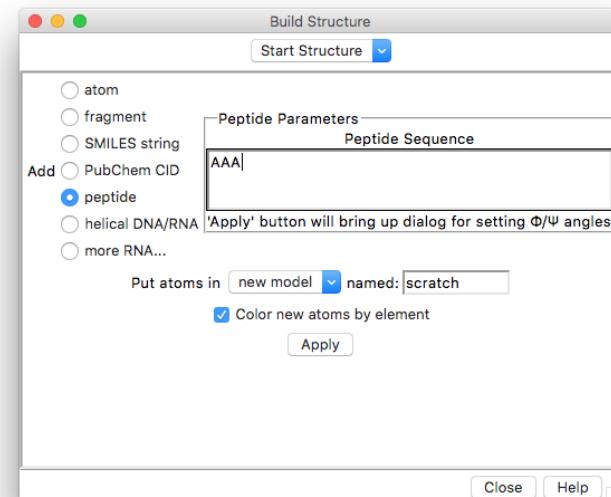


Build an AlaAlaAla peptide:

Tools>>Structure Editing>>Build Structure

Choose peptide

Enter "AAA" in the Peptide Sequence field  
Click "Apply"

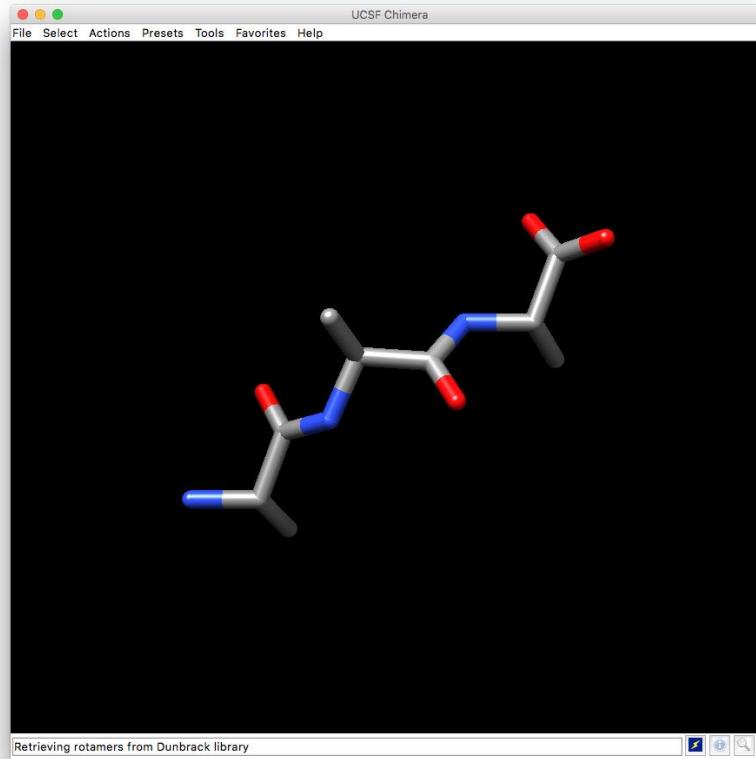


National Institute of  
Allergy and  
Infectious Diseases

NIH

# Chimera

## Create “review of protein structure” scenes

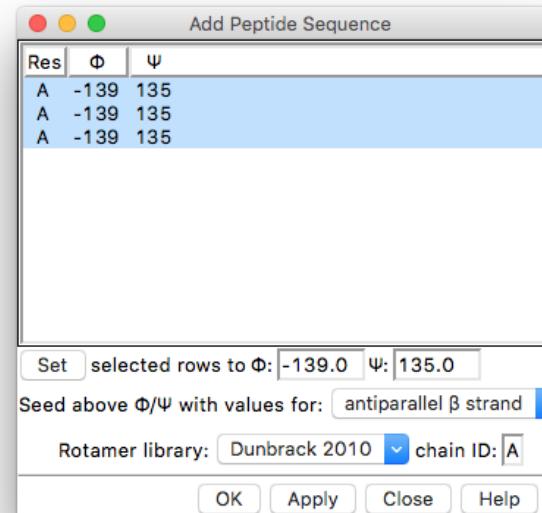


Build an AlaAlaAla peptide (continued):

Select all three rows in the "Add Peptide Sequence" dialog box

Seed above Phi/Psi with values for "antiparallel beta strand"

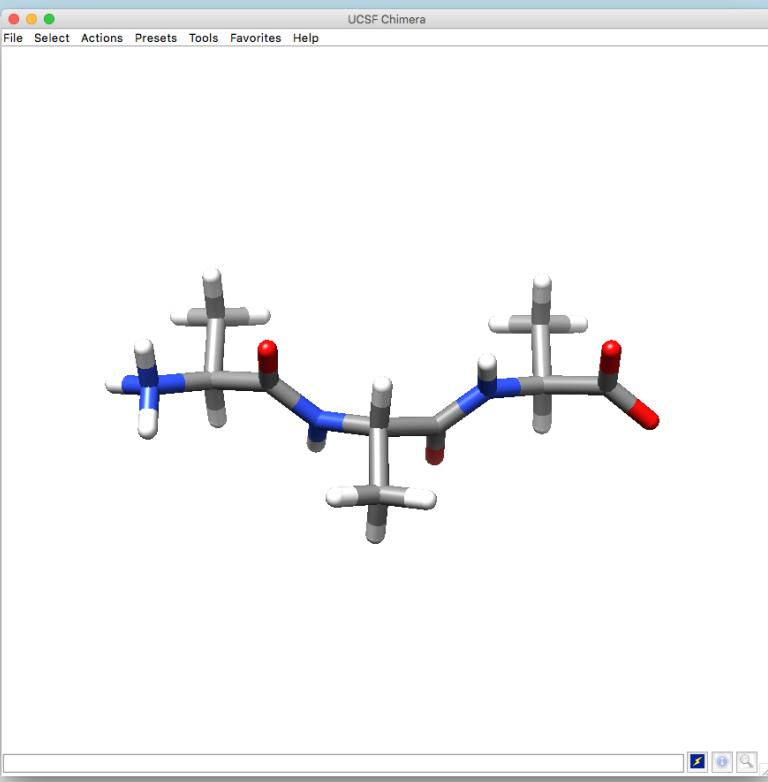
Click "Set", then OK



National Institute of  
Allergy and  
Infectious Diseases

# Chimera

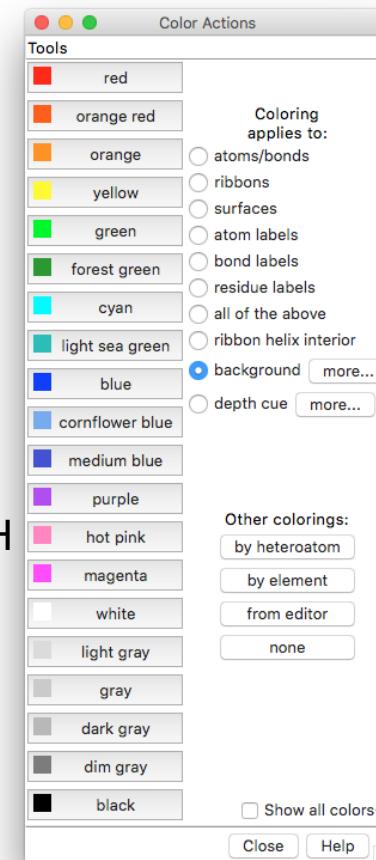
## Make background white; add hydrogen



Actions>>Color>all options...>>  
Click “background” >> white

Click “atoms/bonds”  
Click “Close”

Tools>>Structure Editing>>AddH  
OK

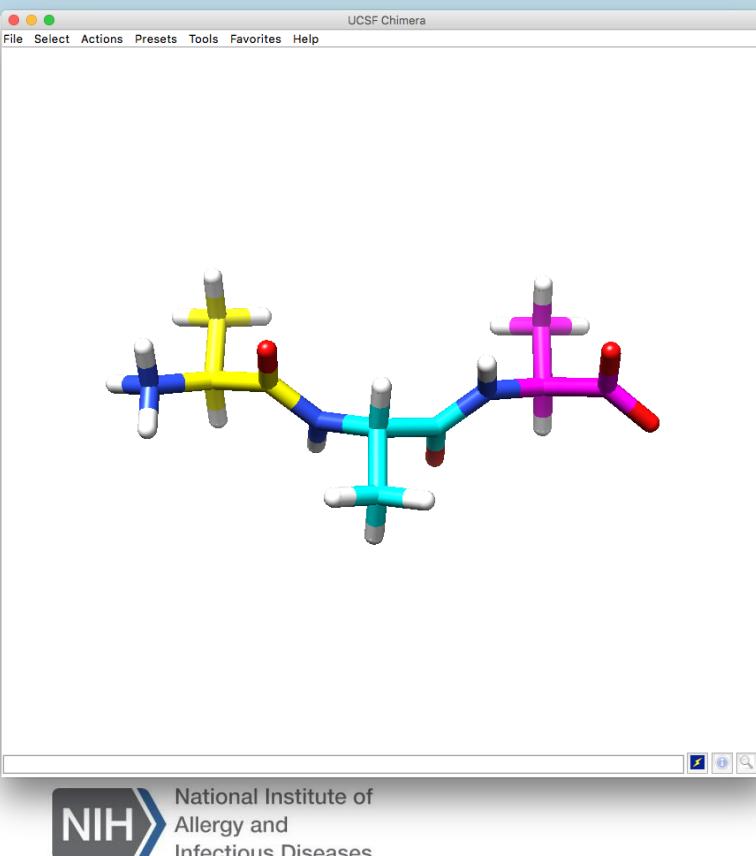


National Institute of  
Allergy and  
Infectious Diseases

NIH

# Chimera

## Color the residues



Control-click on the left-most alpha-carbon to select it.  
(highlighted in green outline)

Press the up arrow to select the entire amino acid.  
Actions>>Color>>yellow

Control-click on the middle alpha-carbon to select it.  
Press the up arrow to select the entire amino acid.

Actions>>Color>>cyan

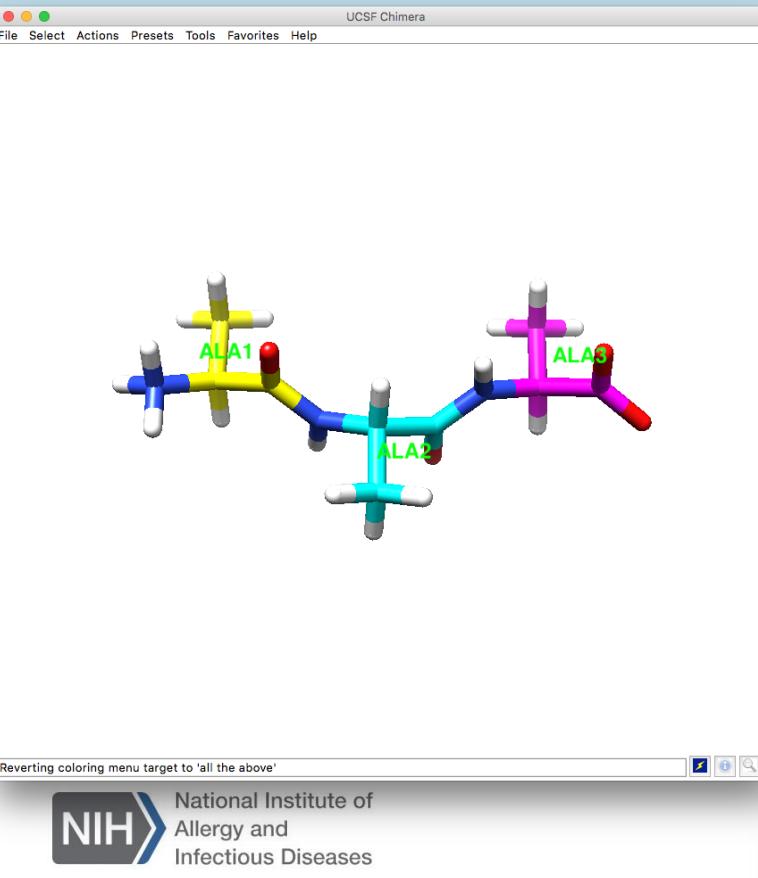
Control-click on the right-most alpha-carbon to select it.  
Press the up arrow to select the entire amino acid.

Actions>>Color>>magenta

Select>>Clear Selection. (or control-click away from any atoms)  
Actions>>Color>>by heteroatom

# Chimera

## Add labels



Actions>>Label>>residue>>Custom

Clear the "Residue label" field

Click "name" and "number"

Click OK

Actions>>Label>>Options

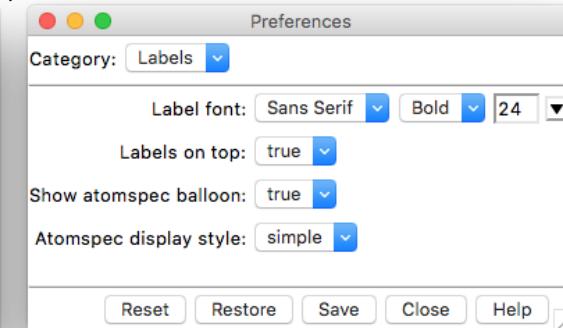
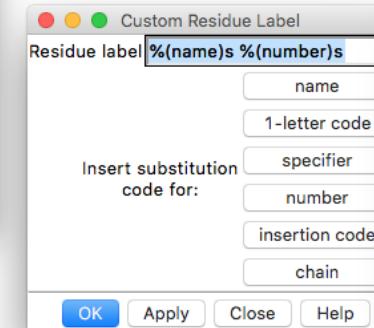
Set Label font to "Bold" and Size "24"

Click Close

Actions>>Color>>all options...>

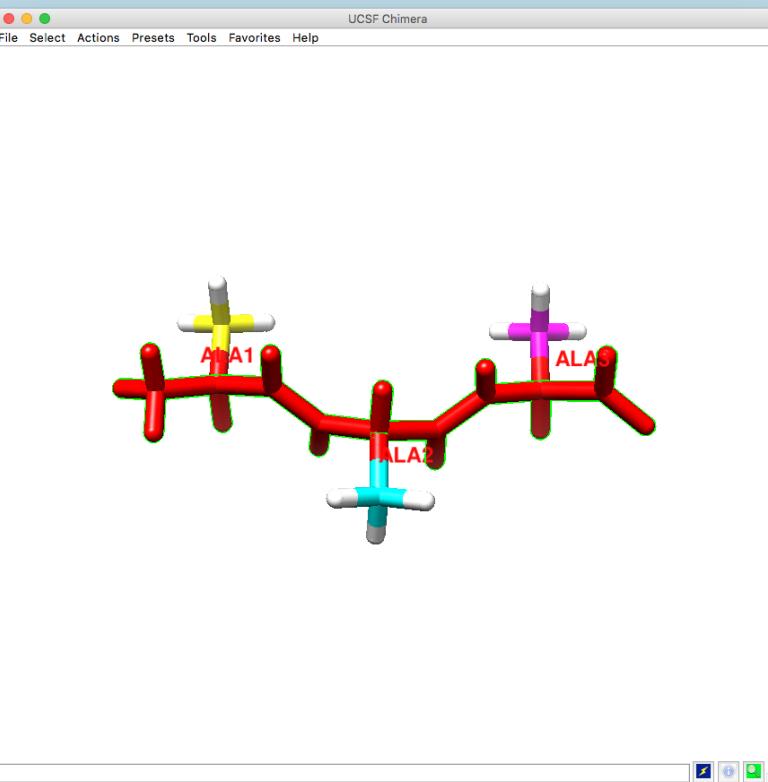
Choose "residue labels" and click "green"

Choose "atoms/bonds", then Close



# Chimera

## Color backbone red (and side chains green)



Select>>Structure>>backbone>>full  
Actions>>Color>>Red

Select>>Structure>>side chain/base>>with CA/C1'  
Actions>>Color>>Green

See if you can figure out how to get the backbone colors back.

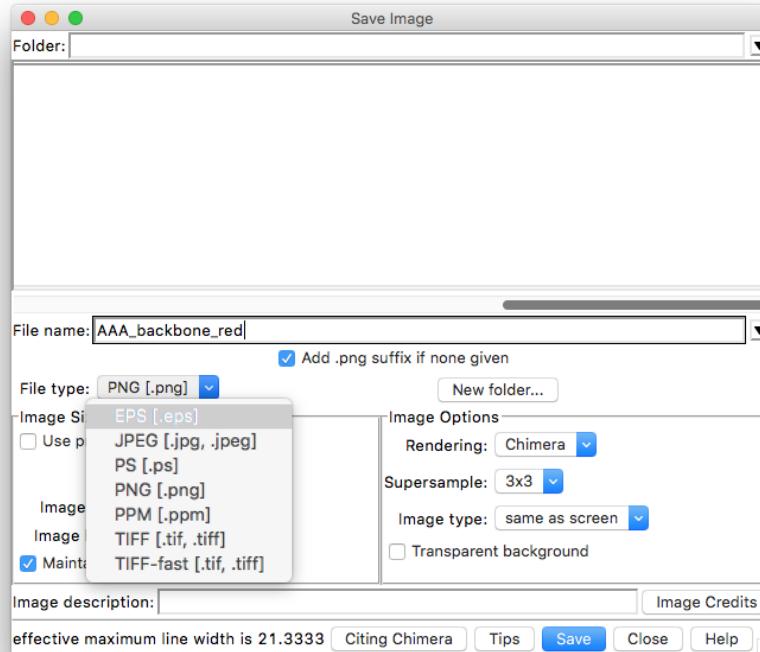


National Institute of  
Allergy and  
Infectious Diseases

NIAID

# Chimera

## Save the current scene as an image file



File>>Save Image...

Choose a folder to save the image into.

Enter a File name, and click "Save"

Choose a "File type".

Note: Different file types will have different "Image Options".



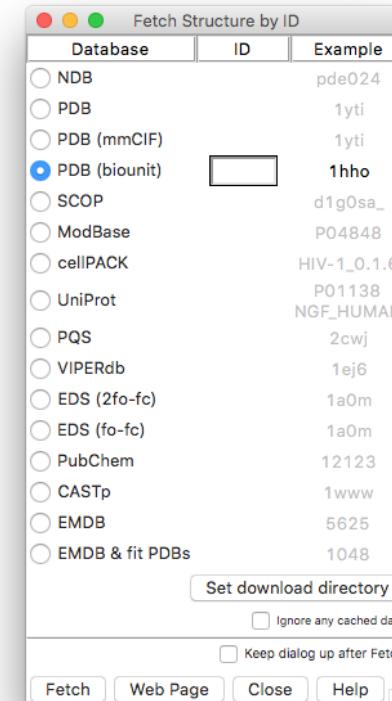
National Institute of  
Allergy and  
Infectious Diseases

# Chimera- Exercise 2

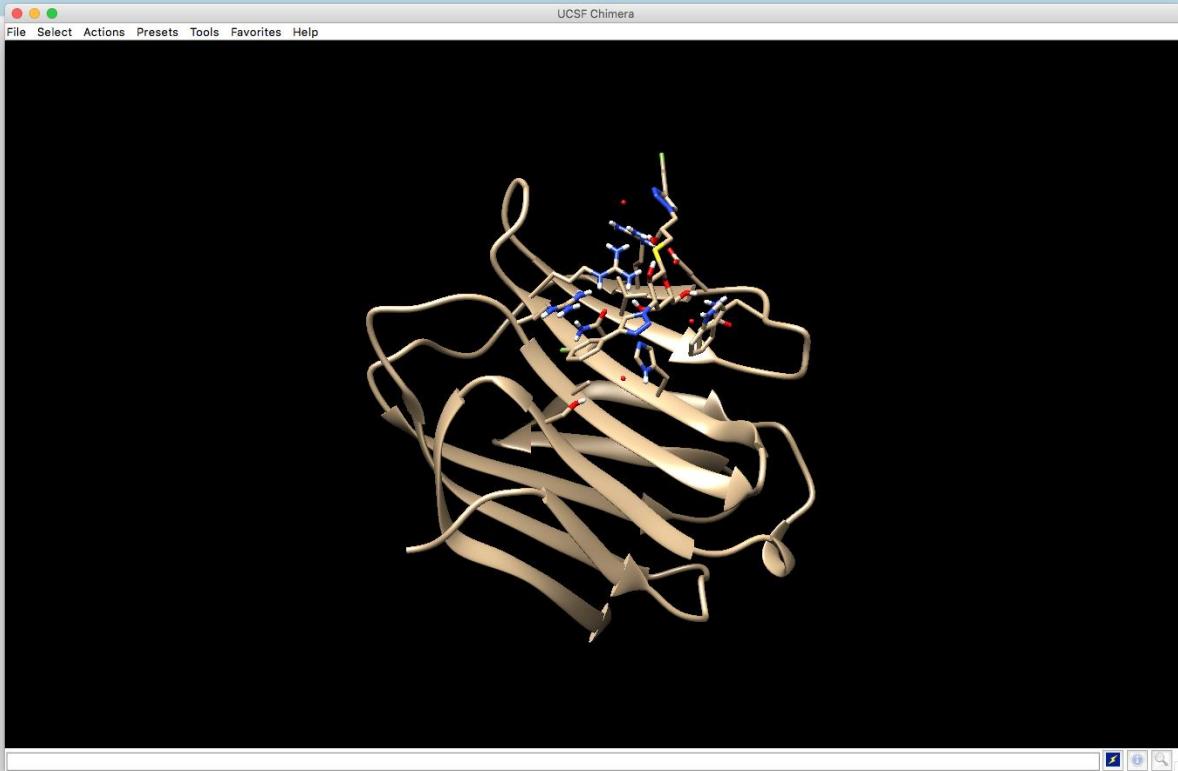
## “Fetch” structures from the PDB; common operations



File>>Fetch by ID>>6RHL



# 6RHL- Galectin-3C



## Mouse controls

Left (near screen center)- xy rotation

Left (near screen edge)- z rotation

Middle- translation

Wheel- Resize

Note ribbon structure  
consists of sheets, coil,  
And one small helix



National Institute of  
Allergy and  
Infectious Diseases

NIH

# Chimera's Standard Representation

- Protein backbone is shown as a ribbon
- Ligands are shown as “sticks”
- Protein sidechains near ligands are shown as sticks
- Water molecules near ligands are shown as a red dot
- Hovering mouse cursor over part of model gives a popup description

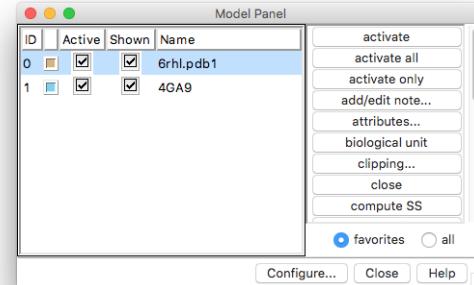
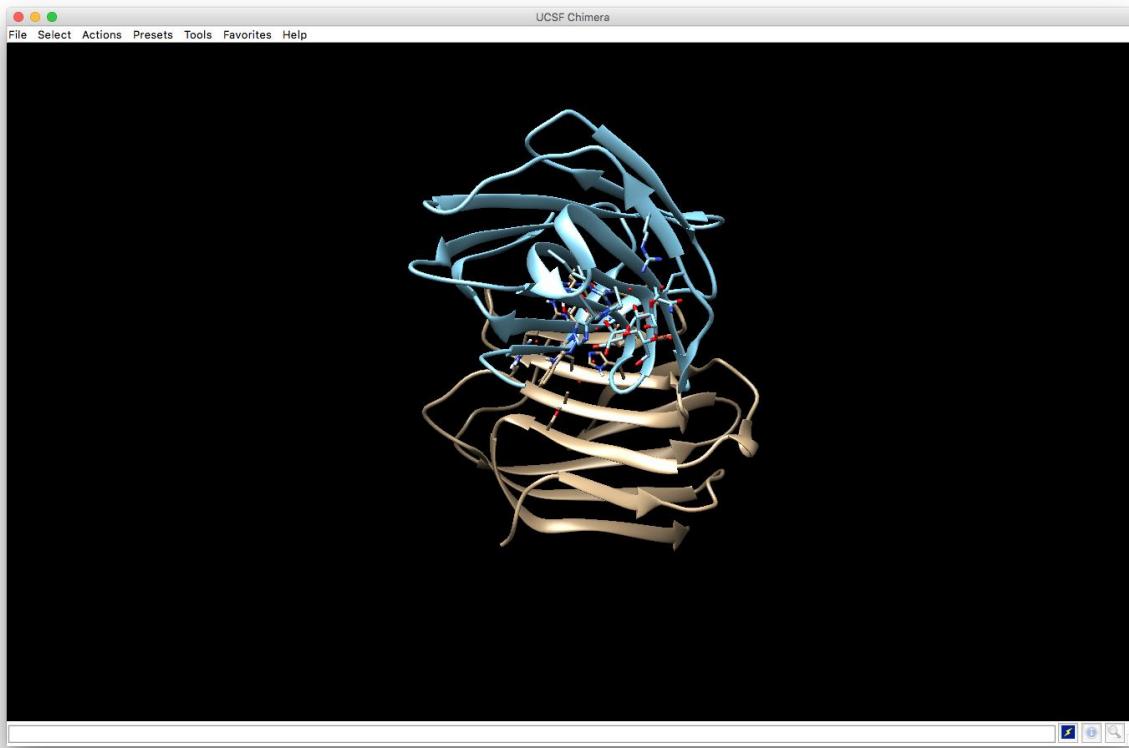
# Fetch 4GA9 (Rat Galectin-1)



Chimera Paradigm:  
Select *then* Actions (on selection)

File>>Fetch by ID>>PDB>>4GA9  
Select>>Chain B  
Actions>>Atoms/Bonds>>Delete

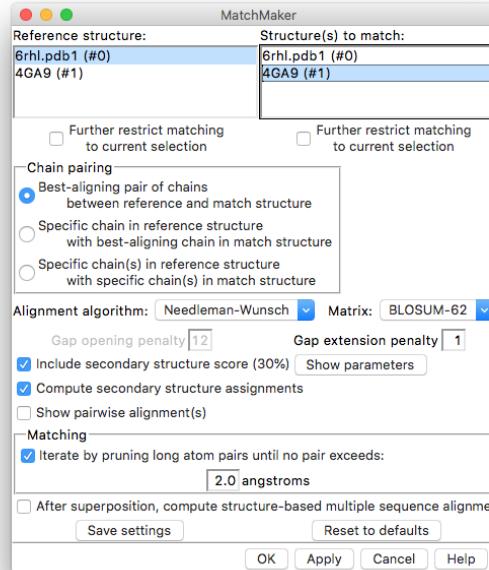
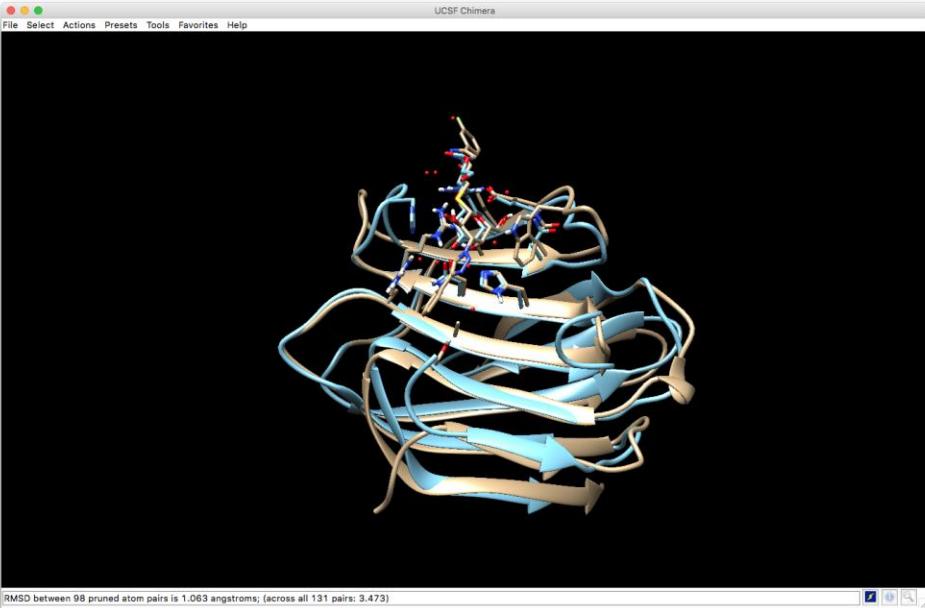
# Model Panel (Favorites menu)



Check boxes in “Shown” column turn models on and off

# Align models

Tools>>Structure Comparison>>Matchmaker

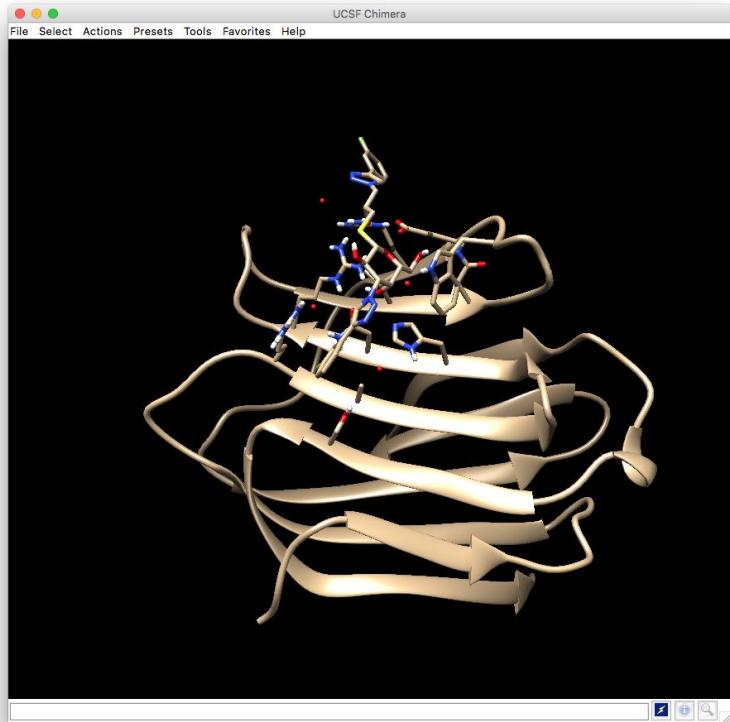


Select structures as shown and press “OK”.  
Examine similarities and differences.  
When done, turn off 4GA9 in Model Panel.



National Institute of  
Allergy and  
Infectious Diseases

# Sequence and Structure



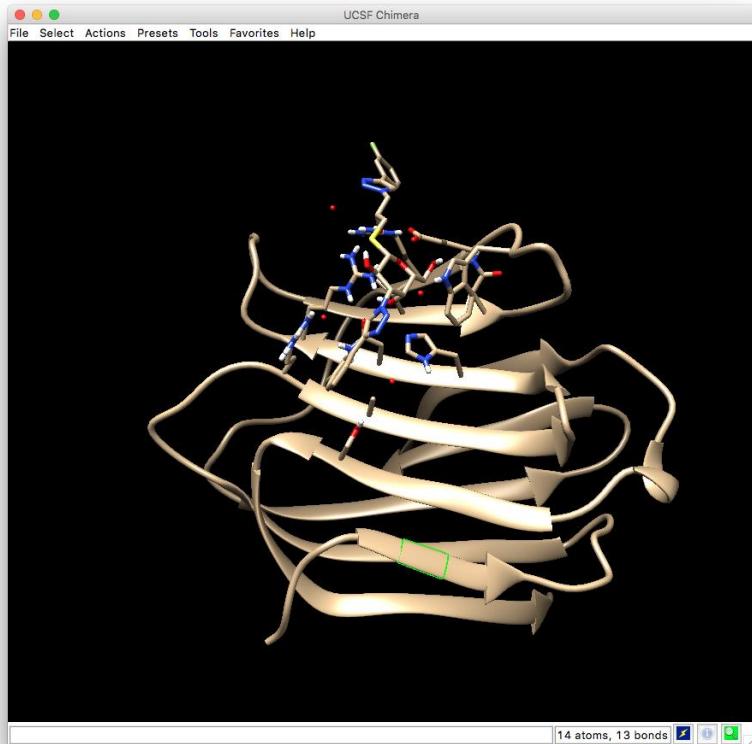
Tools>>Sequence>>Sequence  
Choose 6rhl and click “Show”

6rhl.pdb1 (#0) chain A

	File	Edit	Structure	Headers	Numberings	Tree	Info	Preferences																																										
6rhl.pdb1 (#0) chain A	113	P	L	I	V	P	Y	N	L	P	L	P	G	G	V	V	P	R	M	L	I	T	I	L	G	T	V	K	P	N	A	R	I	A	L	D	F	Q	R	G	N	D	V	A	F	H	F	N	P	R
6rhl.pdb1 (#0) chain A	163	F	N	E	N	N	R	R	V	I	V	C	N	T	K	L	D	N	N	W	G	R	E	R	Q	S	V	F	F	E	S	G	K	P	F	K	I	Q	V	L	V	E	P	D	H	F	K	V	A	
6rhl.pdb1 (#0) chain A	213	V	N	D	A	H	L	L	Q	Y	N	H	R	V	K	K	L	N	E	I	S	K	L	G	I	S	G	D	I	D	L	T	S	A	S	Y	T	M	I											

Quit Hide Help

# Selecting with mouse



Control-click at a point in the ribbon  
Selection is highlighted in green  
Selection also is bright green in sequence



Additional selection operations:

- Control-shift-click adds to current selection
- Typing up arrow in graphics window extends selection
- Left-mouse-button drag in sequence window selects in both sequence and graphics windows.

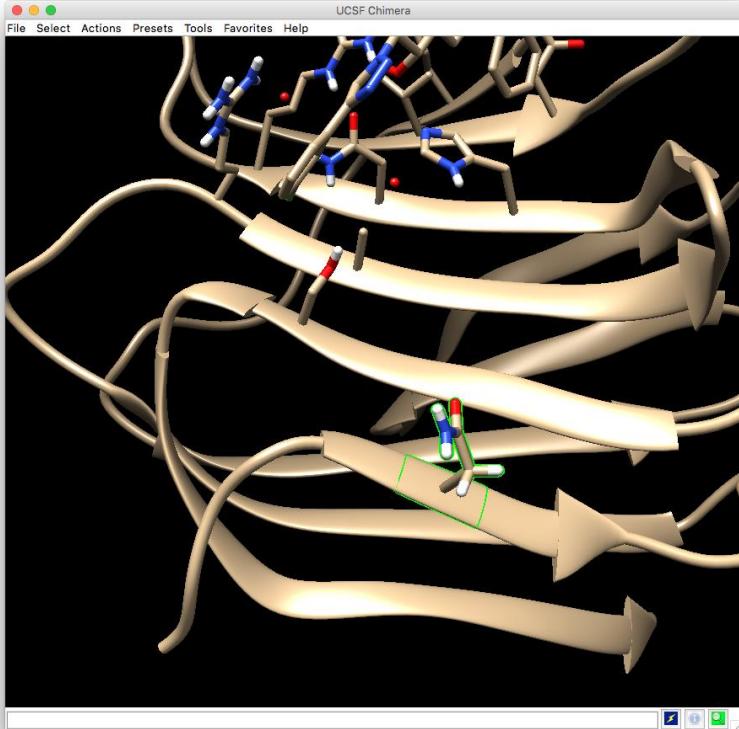


National Institute of  
Allergy and  
Infectious Diseases

# Visualization Options- On Actions menu based on current selection

- Types of molecular visualization
  - Atoms/bonds
    - Stick
    - Ball and Stick
    - Sphere
    - Wire
  - Ribbons
  - Surfaces
- Color
- Labels

# Atoms/Bonds visualization

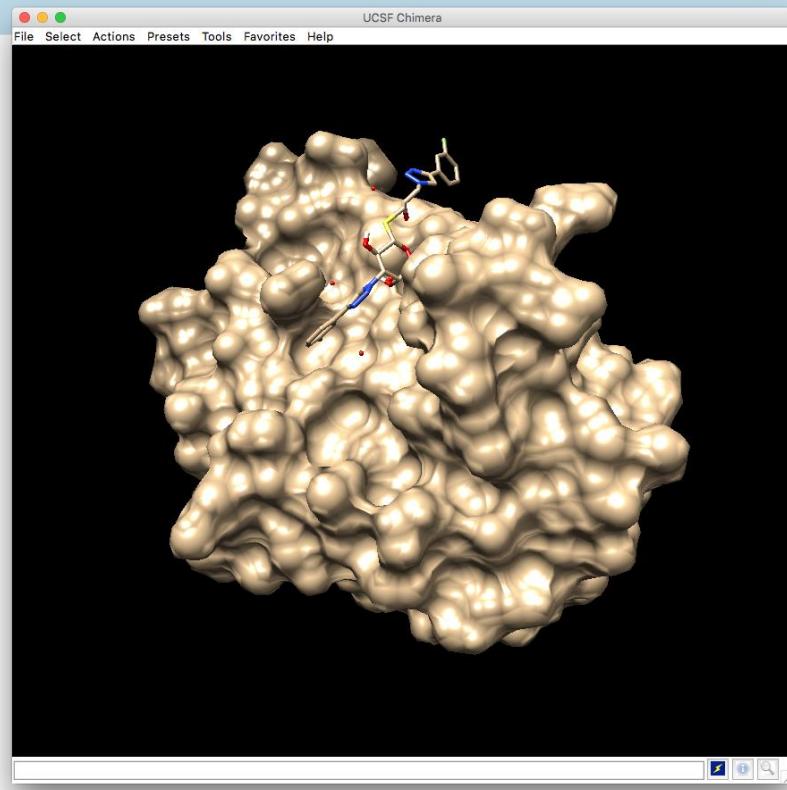


Control-click a residue in a sheet to select it  
Actions>>Atoms-bonds>>Show

Side chain atoms are now shown.

Try the different choices for viewing atoms  
Actions>>Atoms/Bonds>>  
Ball and Stick  
Sphere  
Wire

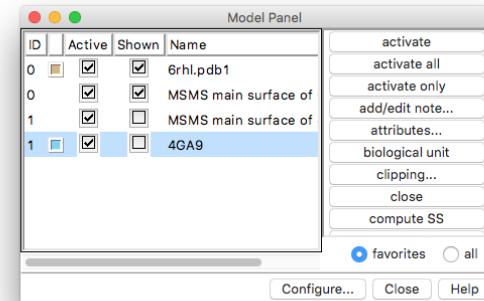
# Surfaces



Control-click in background to clear selection  
Actions>>Surface>>Show

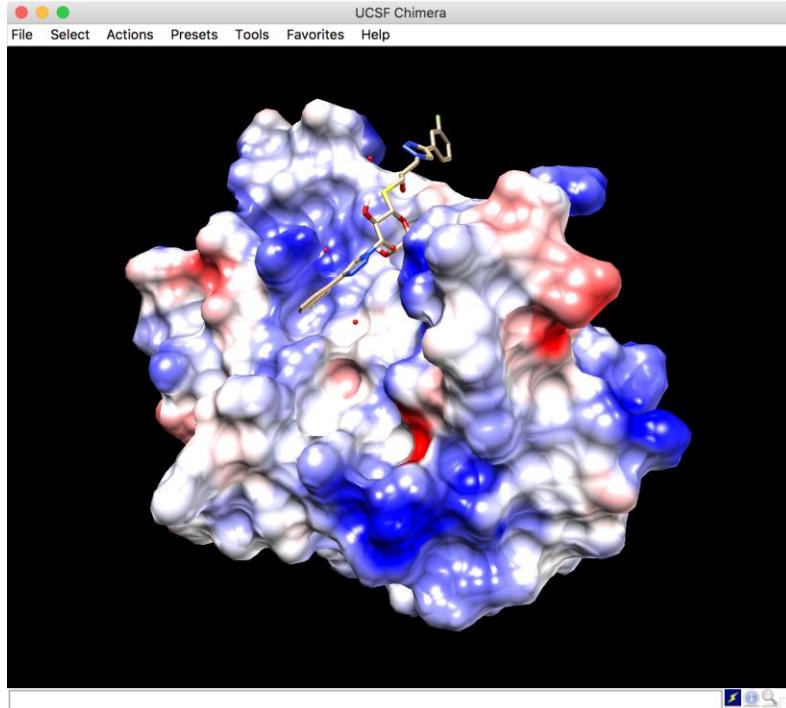
You will see a surface representation with bound ligand

Look at the Model Panel. There are additional entries “MSMS main surface...”  
Note that a surface was also created for 4GA9.  
You can turn on and off the various components of the scene with the “Shown” check boxes.



National Institute of  
Allergy and  
Infectious Diseases

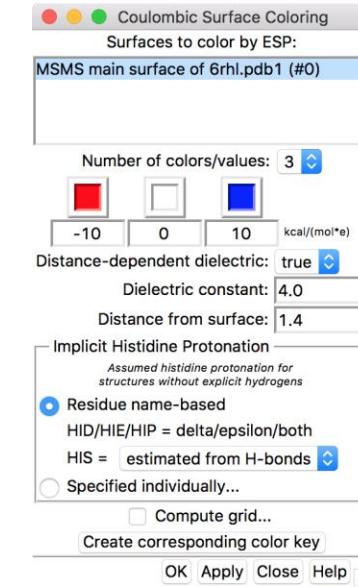
# Surfaces- Coulombic Coloring



Tools>>Surface/Binding Analysis>>Coulombic Surface Coloring

In the “Coulombic Surface Coloring” dialog box,  
Click OK.

When done, click off the  
surface display in the Model  
Panel



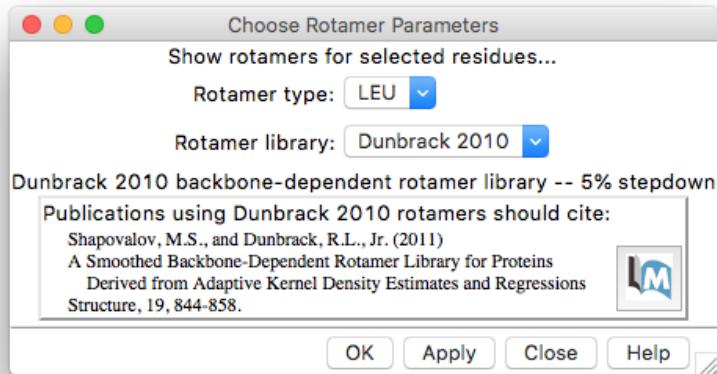
National Institute of  
Allergy and  
Infectious Diseases

# Chimera- Exercise 3

## Mutating a monomer

- In Chimera, **mutations** are performed based on the concept of a **rotamer**.
- Rotamer- Any of a number of isomers of a molecule that can be interconverted by rotation of part of the molecule around a particular (usually) single-bond.
- In the context of protein structure, refers to the combination of side chain torsion angles for a specific amino acid
- Different “libraries” of rotamers are created by statistical analysis of PDB structures. We will use “Dunbrack 2010”.

# Select the amino acid to mutate



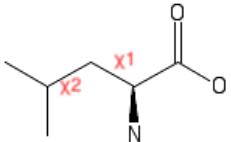
On the 6RHL ribbon, control-click on the ASN119 residue to select it.

Tools>>Structure Editing>>Rotamers

Now you can choose the amino acid type to mutate into. In this example, “LEU”:

Set “Rotamer type:” to “LEU”  
Click OK

# Example rotamer library entry



Leucine

Lists the  $\chi_1$  and  $\chi_2$  combinations for LEU in decreasing order of probabilities, based on analysis of known structures.

#0 ASN 119.A Side-Chain...

Select Columns

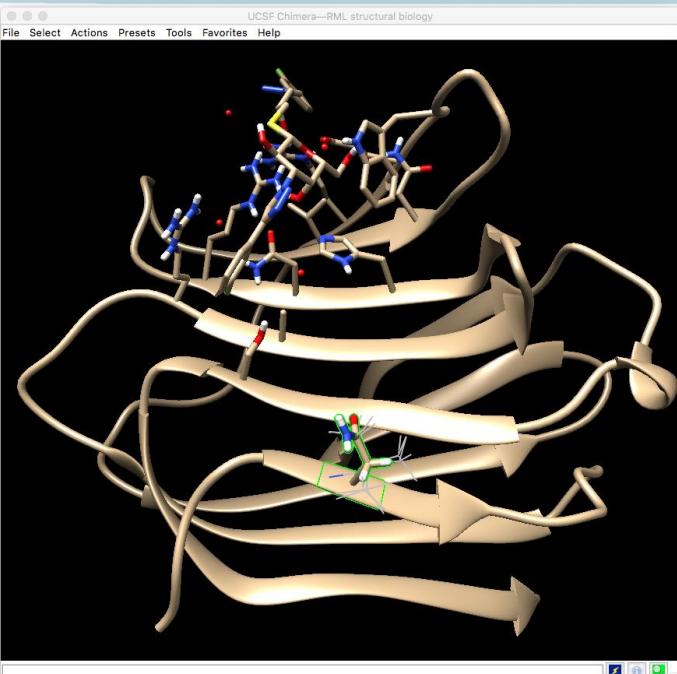
Dunbrack 2010 LEU rotamers

Chi 1	Chi 2	Probability
177.4	64.0	0.545940
-59.2	176.5	0.356426
-76.3	69.0	0.043495
-171.4	157.9	0.040667
-171.9	-69.9	0.008550
-74.6	-59.9	0.004913
61.5	81.7	0.000007
69.2	163.5	0.000003
70.3	-63.0	0.000000

Existing side chain(s):

OK Apply Close Help

# Visualizing rotamers



#0 ASN 119.A Side-Chain...

Select Columns

Dunbrack 2010 LEU rotamers

Chi 1	Chi 2	Probability
177.4	64.0	0.545940
-59.2	176.5	0.356426
-76.3	69.0	0.043495
-171.4	157.9	0.040667
-171.9	-69.9	0.008550
-74.6	-59.9	0.004913
61.5	81.7	0.000007
69.2	163.5	0.000003
70.3	-63.0	0.000000

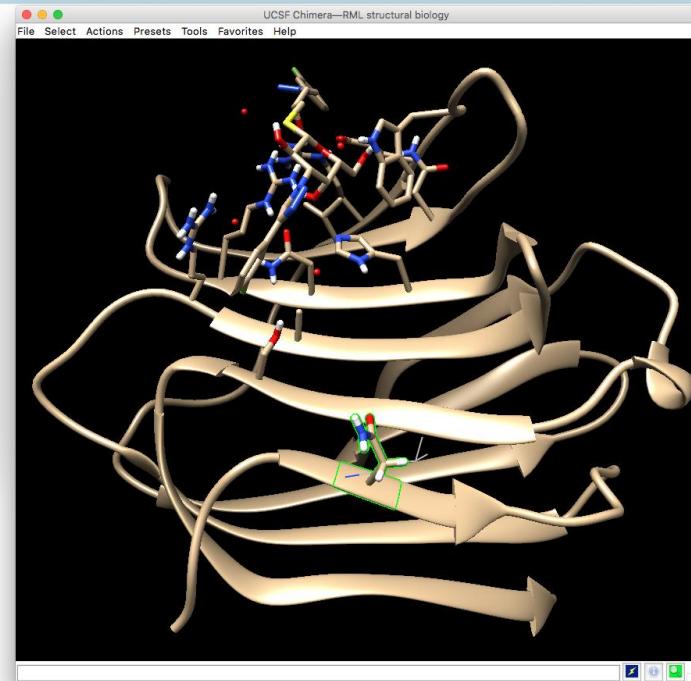
Existing side chain(s): replace

OK Apply Close Help

Click on the top row to show that rotamer.

Use up and down arrows to scroll through list.

When satisfied, click “Apply”.



National Institute of  
Allergy and  
Infectious Diseases

NIH

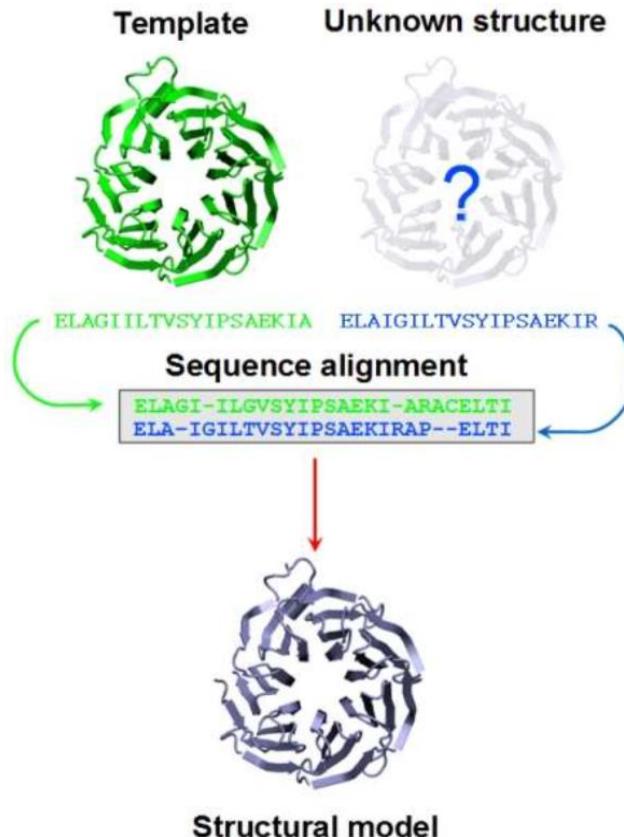
# Rules of thumb for picking rotamers

1. Avoid “bumps” with rest of protein.
2. Choose a rotamer that mimics the “wild-type” amino acid as much as possible.
3. Choose a higher probability rotamer.

# Protein homology (comparative) modeling

- constructing an atomic-resolution **model** of the "target" protein from its amino acid sequence and an experimental three-dimensional structure of a related **homologous** protein (the "template").
- There are other types of protein modeling
  - Secondary Structure Prediction (local structure only)
  - Ab initio protein folding (VERY computationally difficult)

# High-level Concept



National Institute of  
Allergy and  
Infectious Diseases

# Homology Modeling with I-TASSER

## Iterative Threading ASSEmbley Refinement (I-TASSER)

- on-line platform for protein structure and function predictions (although it can be downloaded)
- a hierarchical approach
  - structural templates first identified from the PDB by multiple threading approach LOMETS
  - full-length atomic models are then constructed by iterative template fragment assembly simulations
  - function insights of the target are derived by threading the 3D models through protein function database BioLiP
- Consistently ranked at or near the top in the Community-wide Assessment for Structure Prediction
  - I-TASSER was ranked as the No 1 server for protein structure prediction in CASP7, CASP8, CASP9, CASP10, CASP11, CASP12



National  
Allergy and  
Infectious

NIAID

# I-TASSER web server

**Zhang Lab**

UNIVERSITY OF MICHIGAN

**I-TASSER**  
Protein Structure & Function Predictions

(The server completed predictions for 499484 proteins submitted by 107129 users from 143 countries!  
(The template library was updated on 2019/02/24)

I-TASSER (Iterative Threading ASSEMBly Refinement) is a hierarchical approach to protein structure and function prediction. It first identifies structural templates from the PDB by multiple threading approach (LOMETS), with full-length atomic models constructed by iterative template-based fragment assembly simulations. Function insights of the target are then derived by threading the 3D models through protein function database BioLit. I-TASSER was ranked as the No 1 server for protein structure prediction in recent community-wide CASP7, CASP8, CASP9, CASP10, CASP11, CASP12, and CASP13 experiments. It was also ranked as the best for function prediction in CASP9. The server is in active development with the goal to provide the most accurate structural and function predictions using state-of-the-art algorithms. Please report problems and questions at I-TASSER message board and our developers will study and answer the questions accordingly. (See More about the server...)

[Queue](#) | [Forum](#) | [Download](#) | [Search](#) | [Registration](#) | [Statistics](#) | [Remove](#) | [Potential](#) | [Decoy](#) | [News](#) | [Annotation](#) | [About](#) | [FAQ](#)

**I-TASSER On-line Server** [View an example of I-TASSER output](#):

Copy and paste your sequence below (10, 1500) residues in FASTA format! [Click here for a sample input](#)

Or upload the sequence from your local computer:  
[Choose File](#)

Email: (mandatory, where results will be sent to)

Password: (mandatory, please click [here](#) if you do not have a password)

ID: (optional, your given name of the protein)

► [Option I: Assign additional restraints & templates to guide I-TASSER modeling.](#)

► [Option II: Exclude some templates from I-TASSER template library.](#)

► [Option III: Specify secondary structure for specific residues.](#)

Keep my results public (uncheck this box if you want to keep your job private. A key will be assigned for you to access the results)

[Run I-TASSER](#) [Clear form](#)  
(Please submit a new job only after your old job is completed)

**I-TASSER Suite:**  
[Download I-TASSER Standalone Package \(Version 5.1\)](#)

**I-TASSER Resource:**

TM-align

<https://zhanglab.ccmb.med.umich.edu/I-TASSER/>



National Institute of  
Allergy and  
Infectious Diseases

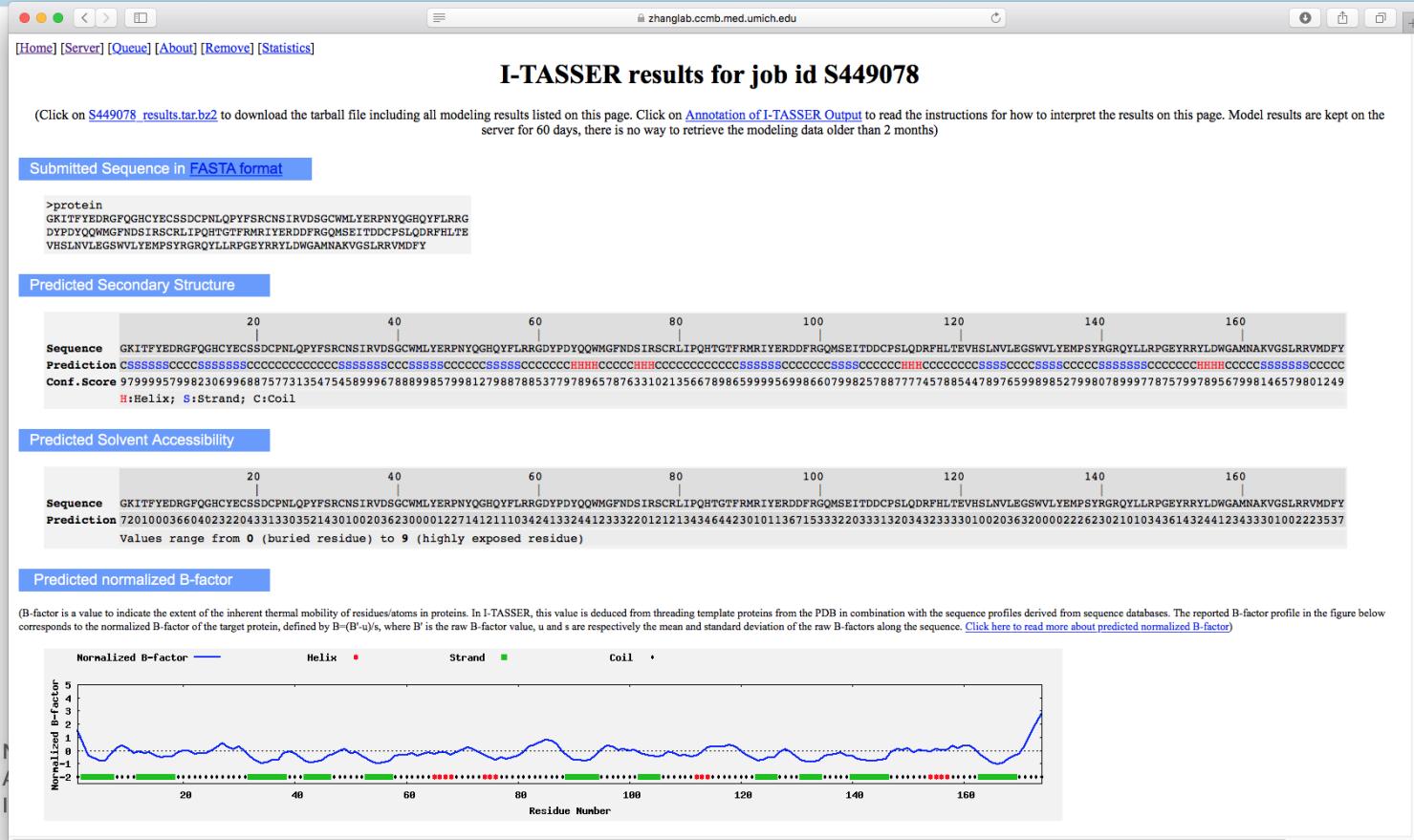
NIH

# Submitting a job on I-TASSER server

- Register for an account
- Enter sequence (or upload file) in FASTA format
- Enter the email address you used to register, and your password.
- Enter an ID that describes the protein for your reference to the modeling run.
- Click “Run I-TASSER”

Major Limitation: Only one concurrent job per IP address

# Results from an I-TASSER run



# Results from an I-TASSER run



National Institute of  
Allergy and  
Infectious Diseases

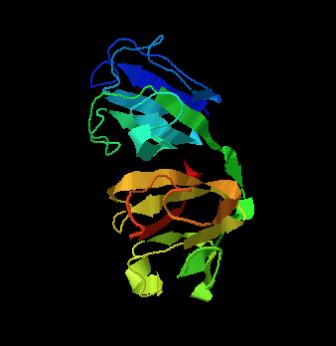
# Results from an I-TASSER run

Top 5 final models predicted by I-TASSER

(For each target, I-TASSER simulations generate a large ensemble of structural conformations, called decoys. To select the final models, I-TASSER uses the SPICKER program to cluster all the decoys based on the pair-wise structure similarity, and reports up to five models which corresponds to the five largest structure clusters. The confidence of each model is quantitatively measured by C-score that is calculated based on the significance of threading template alignments and the convergence parameters of the structure assembly simulations. C-score is typically in the range of [-5, 2], where a C-score of a higher value signifies a model with a higher confidence and vice-versa. TM-score and RMSD are estimated based on C-score and protein length following the correlation observed between these qualities. Since the top 5 models are ranked by the cluster size, it is possible that the lower-rank models have a higher C-score in rare cases. Although the first model has a better quality in most cases, it is also possible that the lower-rank models have a better quality than the higher-rank models as seen in our benchmark tests. If the I-TASSER simulations converge, it is possible to have less than 5 clusters generated; this is usually an indication that the models have a good quality because of the converged simulations.)

- [More about C-score](#)
- [Local structure accuracy profile of the top five models](#)

(By right-click on the images, you can export image file or change the configurations, e.g. modifying the background color or stopping the spin of your models)



Reset to initial orientation  Spin On/Off

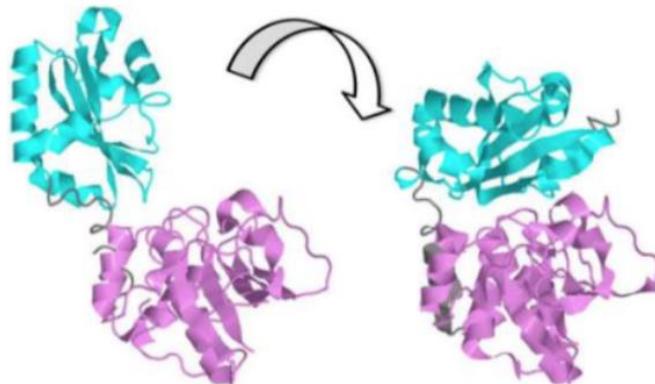
- [Download Model 1](#)
- C-score=1.60 ([Read more about C-score](#))
- Estimated TM-score = 0.94±0.05
- Estimated RMSD = 2.0±1.6Å

# Scoring the Results

**C-score** is a confidence score for estimating the quality of models.

- calculated based on the significance of threading template alignments and the convergence parameters of the structure assembly simulations
- C-score is typically in the range of [-5 to 2], where a C-score of **higher** value signifies a model with a high confidence and vice-versa.

**Tm-score** - solves the problem of local error when calculating RMSD



# Factors Determining Model Quality

- % sequence identity to templates

% ID	Confidence?
> 30	good to great
25 - 30	low to maybe?
< 25	low

- Coverage
- Steric or electrostatic clashes
- Agreement with bench data
- Agreement with general protein structure knowledge
- Scoring (RMSD, C-score, Tm-score, others...)

# Is my model good enough?

- **The answer depends on your purpose.**
- Good enough for drug design? – probably if the sequence identity is very high (>50%)
- Sometimes good enough if far lower sequence id but accurate around site of interest.
- High confidence but low seq i.d. still very likely correct fold, useful for a range of tasks.

# Other I-TASSER capabilities

- I-TASSER accepts several types of user-specified restraints and filters
  - Inter-residue contact and distance restraints
  - Specified template structures and template-target alignments
  - Secondary structure assignment
- Special algorithm for GPCR modeling

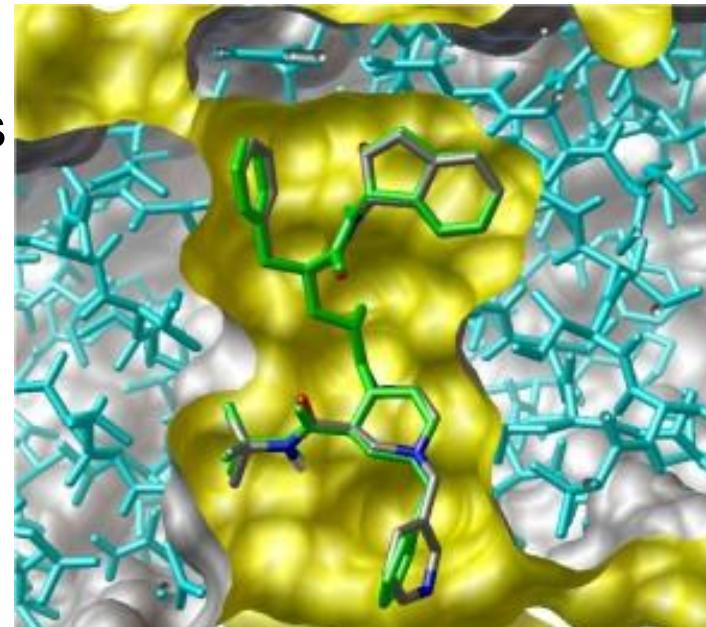
# Mutations and Homology Modeling

- The STRUCTURAL effects of point mutations on the structure will **NOT** be modelled accurately by homology modeling programs.
- Need to use other modeling methods based on energy
  - Force Field methods model a molecule as charged spheres, with weights on springs
    - **very fast in a computer**
    - Topic for another day

# Chimera- Exercise 4

## Docking with AutoDock Vina using Chimera

- AutoDock Vina
  - A suite of automated docking tools
  - Free
  - Cross-platform
  - Open source
  - Online Server
  - Available on Biowulf
  - Docks ligands up to 2048 atoms
- From Laboratory of Dr. Oleg Trott, Scripps



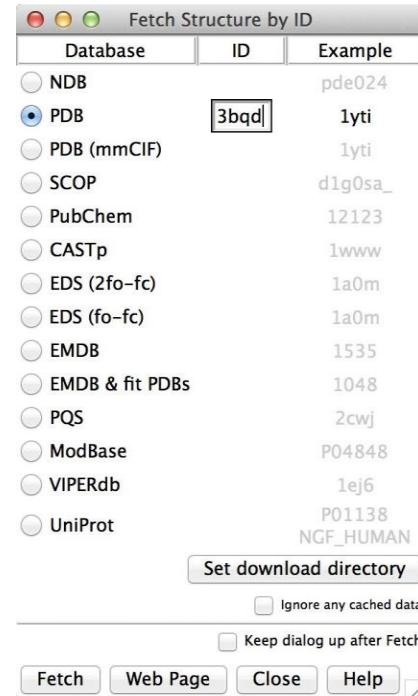
# Generalized steps for docking setup

- Prepare protein
  - Add hydrogens (or not!)
  - Treat terminal groups
  - Set side chain torsions
  - Remove ligand
- Prepare ligand(s)
  - Add hydrogens
  - Adjust ionization state (pH)
  - Choose stereoisomers/tautomers
- Identify active site on protein

# Open Chimera and load the protein

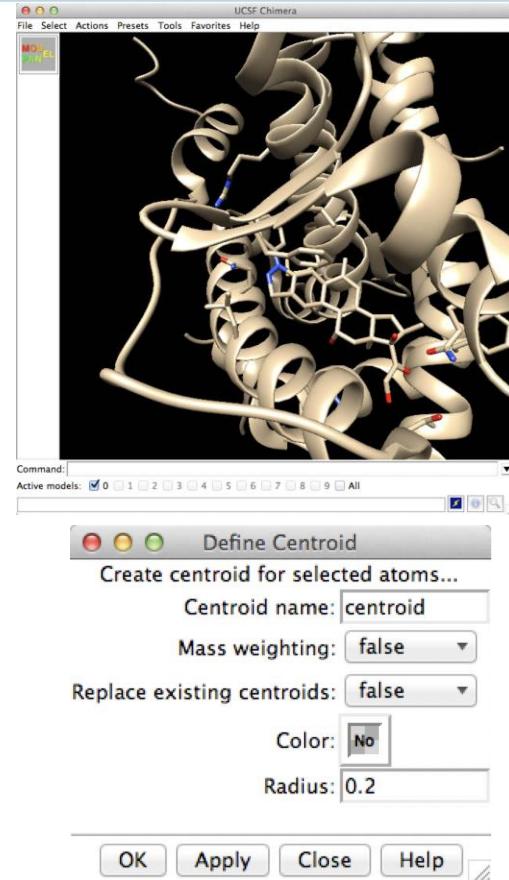
- Open Chimera and load the protein:
- File>>Fetch by ID...
- Choose PDB and type “3bqd” into the ID box.
- Click “Fetch”.

A crystal structure of the Glucocorticoid Receptor will appear in the graphics window.



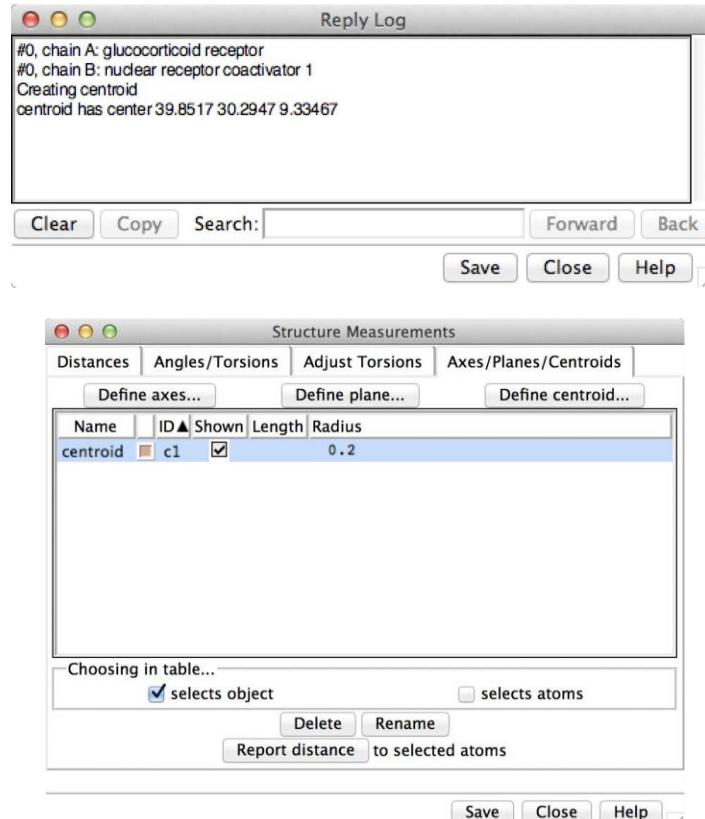
# Find the Ligand Centroid- 1

- Tools>>Utilities>>Reply log
- Select>>Residue>>DAY
  - This selects the ligand
- Tools>>Structure Analysis>>Axes/Planes/Centroids
- Click “Define centroid...”
- Click “OK” in the Defined Centroid dialog box



# Find the Ligand Centroid- 2

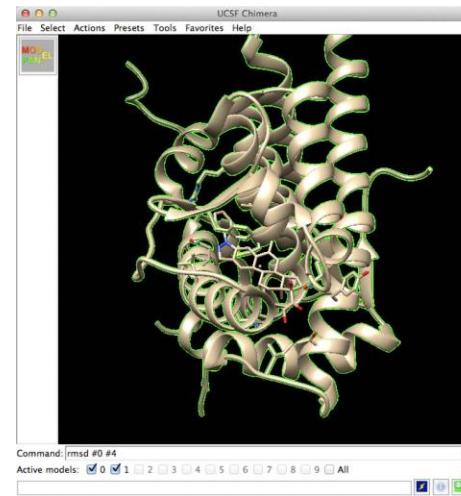
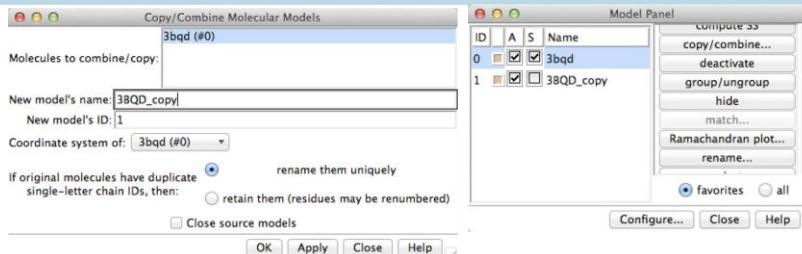
- Note coordinates of centroid in Reply Log.
- Select the centroid “c1” in the Structure Measurements dialog box and click delete.
- Close the dialog box.



National Institute of  
Allergy and  
Infectious Diseases

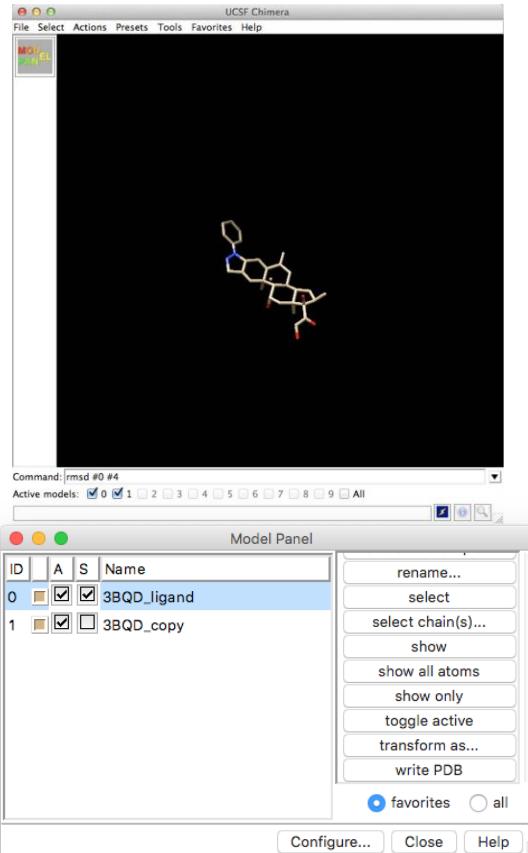
# Separate the ligand from the protein- 1

- In the Model Panel, scroll menu and choose “copy/combine...”
- Select #0
- In the field “New model’s name” enter 3BQD\_copy
- Click OK
- In Model Panel, uncheck 3BQD\_copy in “S” column
- Verify that the ligand is still selected, then choose Select>>Invert (selected models)->.
  - The protein will now be selected.



# Separate the ligand from the protein- 2

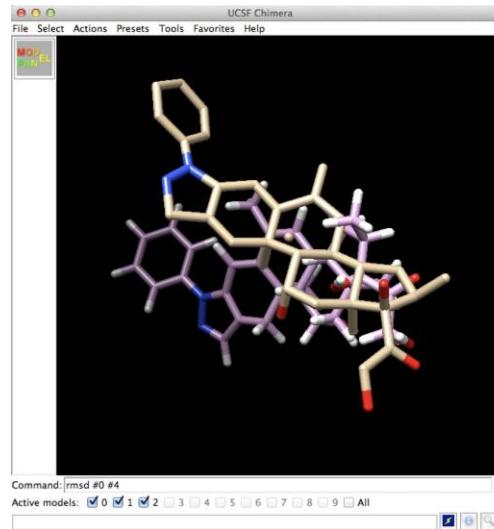
- Choose the menu item:  
Actions>>Atoms/Bonds>>delete.
- Only the ligand should remain visible.
- In the Model Panel, select the model with ID #0 (3BQD.pdb), and scroll to choose “rename...” in the Model Panel menu.
- In the dialog box, enter “3BQD\_ligand” for the new name and click OK.



# Load a version of the ligand with arbitrary pose, that will be docked into the protein

- File>>Open and choose the file:

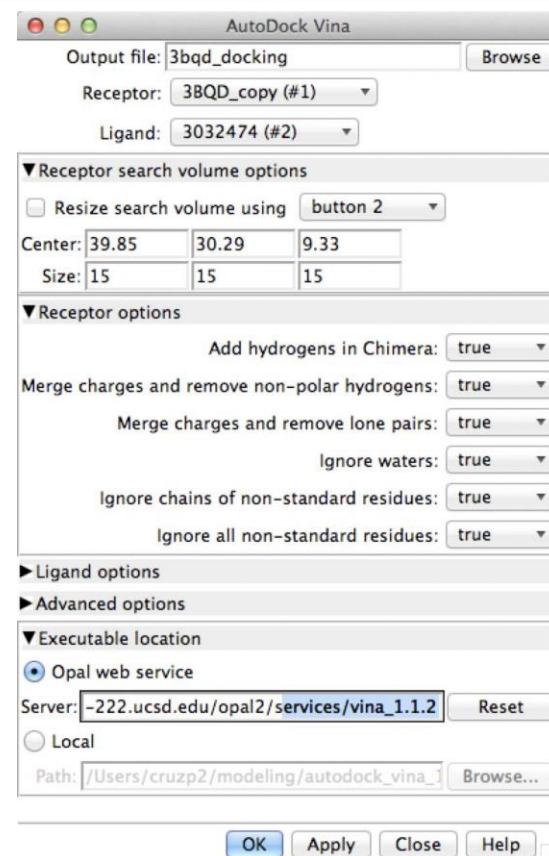
dac\_3d\_renum\_vina.mol2



National Institute of  
Allergy and  
Infectious Diseases

# Submit the AutoDock Vina docking calculation

- Tools>>Surface/Binding Analysis>>AutoDock Vina
- Click “Browse”, navigate to current directory, enter an output filename and “Set Output Location”
- Specify Receptor: 3BQD\_copy (#1)
- Specify Ligand: 3032474 (#2)
- Enter Receptor search volume options
  - Center: centroid coords from Reply Log  
(2 decimal places)
  - Size: 15 for x, y, z (graphics box will appear)
- Enter Receptor options
  - Ignore all non-standard residues: true
- Executable location>>Opal web service
- Click OK to start job

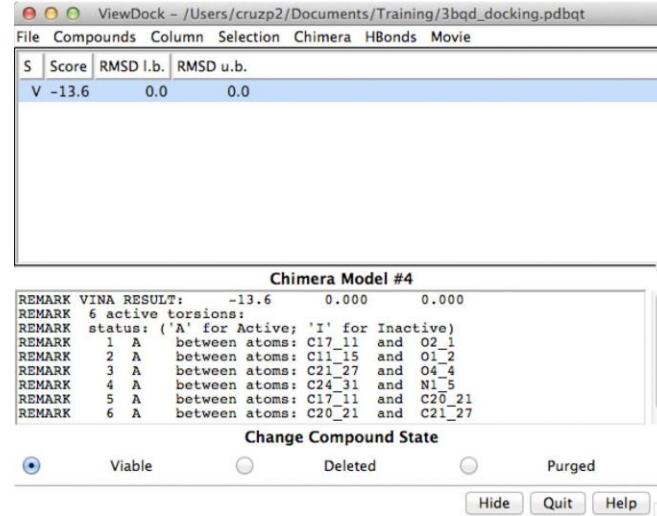


National Institute of  
Allergy and  
Infectious Diseases

NIH

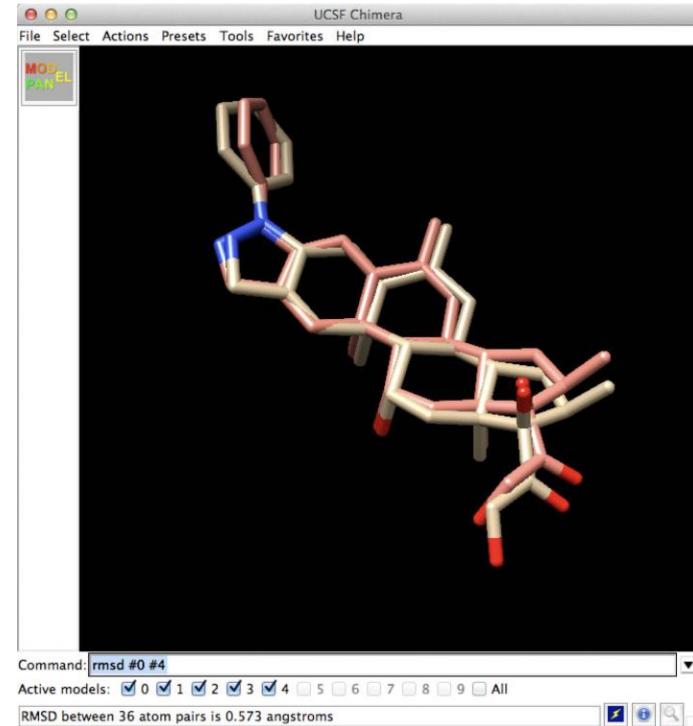
# Examine results

- When the job concludes, a new model, Docked 3032474, will appear in the graphics window with ID 4 in the Model Panel.
- The ViewDock dialog box, which shows the docking score, will also appear.
- Make sure only the new model #4 and #0 are selected in the “S” column of the Model Panel, to visually examine how well the program performed.



# Do an RMSD calculation to quantify how well the program did

- Delete hydrogen atoms on the docked molecule (ID 4)
  - a. Note that there are 3 H atoms, colored white, attached to oxygen atoms.
  - b. Control-click one of them to select
  - c. Actions>>Atoms/Bonds/delete
  - d. Repeat steps b. and c. for the other 2 atoms
- Favorites>>Command Line
- At the Command prompt type:  
rmsd #0 #4
- The RMSD will appear in the Reply Log, and in the message window at the bottom of the main graphics window.



# Sources of Ligands

- Zinc- <https://zinc15.docking.org/>
  - Over 230 million purchasable compounds in ready-to-dock, 3D formats
- Pubchem- <https://pubchem.ncbi.nlm.nih.gov/>
  - Nearly 100 million compounds, most with 3D formats
- ChEMBL- <https://www.ebi.ac.uk/chembl/>
  - Nearly 2 million bioactive molecules with drug-like properties
  - Can search by target

# VR with ChimeraX

- Sign up for a VR demo with ChimeraX if you haven't already done so.

# Exercise: Where is my mutation?

- PDB ID: 5DGO
- \*Mutation P321T
  - A form of Meier-Gorlin syndrome, a syndrome characterized by bilateral microtia, aplasia/hypoplasia of the patellae, and severe intrauterine and postnatal growth retardation with short stature and poor weight gain.
- Find the location of the mutation, highlight it either with color or depiction, and then perform the mutation.

\*From UniProt VAR\_080870