



[www.esaunggul.ac.id](http://www.esaunggul.ac.id)

# PENGANTAR BIOINFORMATIKA

## IBT 431

*By Seprianto S.Pi, M.Si*

Meeting 12

# Akses Struktur Protein Data Base

# Sasaran Perkuliahan

- Mahasiswa mampu mengakses Struktur database (PDB, Swiss – model, Pyre2, PyMol, EXPASY, )
- Mahasiswa Mampu mempelajari cara dan mendapatkan data struktur 3D protein dari GenBank

# 1. Protein Data Bank (PDB)

The screenshot shows the PDB website interface. On the left, the RCSB PDB logo is displayed with the text '139357 Biological Macromolecular Structures Enabling Breakthroughs in Research and Education'. Below the logo are several partner logos: PDB-101, Worldwide PDB, EMDatabank, Nucleic Acid Database, and Worldwide Protein Data Bank Foundation. In the center, there is a search bar with the placeholder text 'Search by PDB ID, author, macromolecule, sequence, or ligands' and a 'Go' button. Below the search bar are links for 'Advanced Search' and 'Browse by Annotations'. On the right side, there are social media icons for Facebook, Twitter, YouTube, and LinkedIn.

- Protein Data Bank: maintained by the **R**esearch **C**ollaboratory for **S**tructural **B**ioinformatics (**RCSB**)
- Also contains structures of other bio-macromolecules: DNA, carbohydrates and protein-DNA complexes.

➤ They are big databases too: **Penelusuran database protein**

**Swiss-Prot** (very high level of annotation)

[https://web.expasy.org/docs/swiss-prot\\_guideline.html](https://web.expasy.org/docs/swiss-prot_guideline.html)

**UniProt** (protein identification resource) the world's most comprehensive catalog of information on proteins

<http://www.uniprot.org/>

➤ Translated databases:

**TREMBL** (translated EMBL): includes entries that have not been annotated yet into Swiss-Prot.

<http://www.ebi.ac.uk/trembl/access.html>

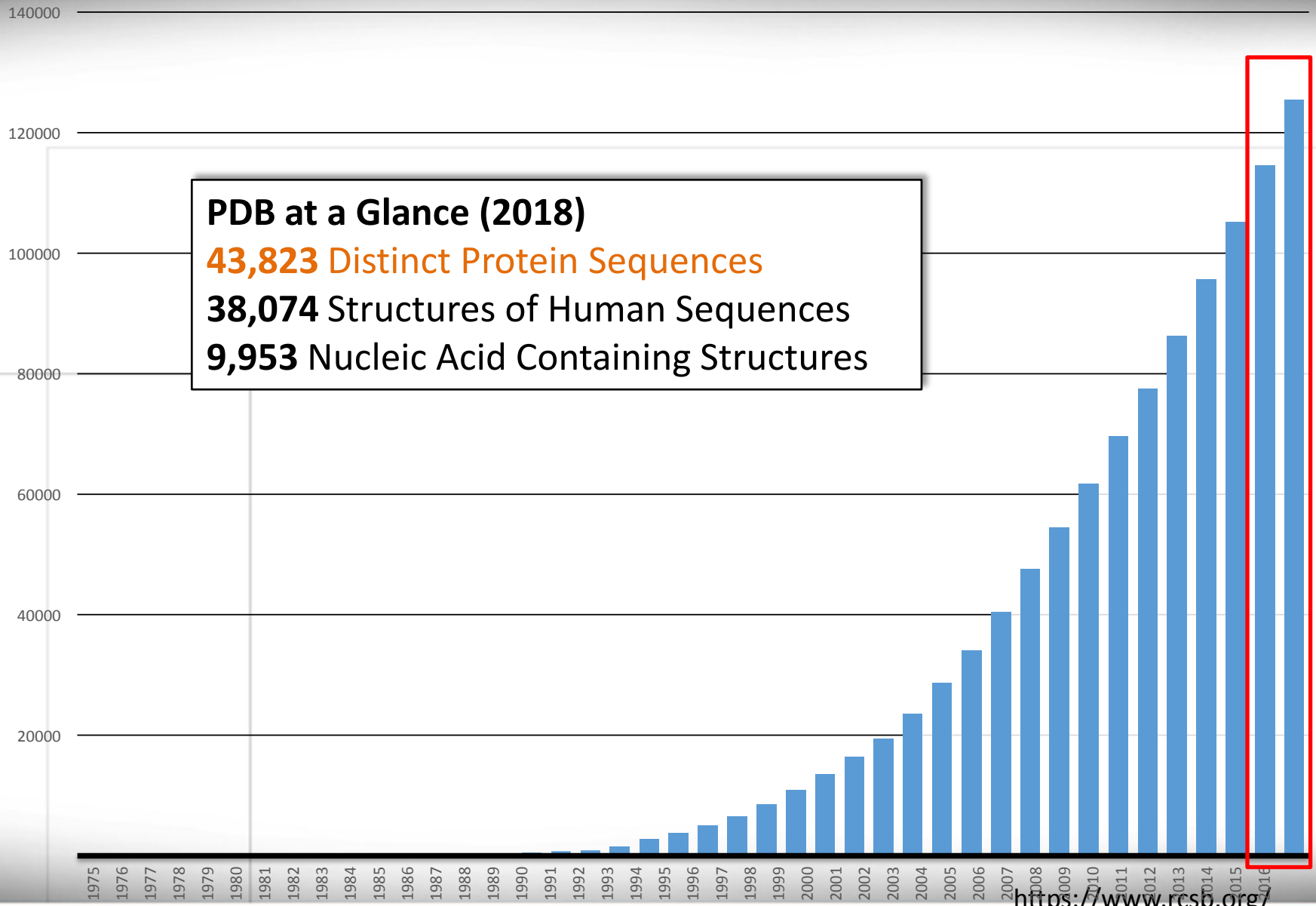
**GenPept** (translation of coding regions in GenBank)

**PDB** (sequences derived from the 3D structure Brookhaven PDB) <http://www.rcsb.org/pdb/>

# Protein Data Bank: 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.

# Protein structure submission in PDB (2018)



## PDB at a Glance (2018)

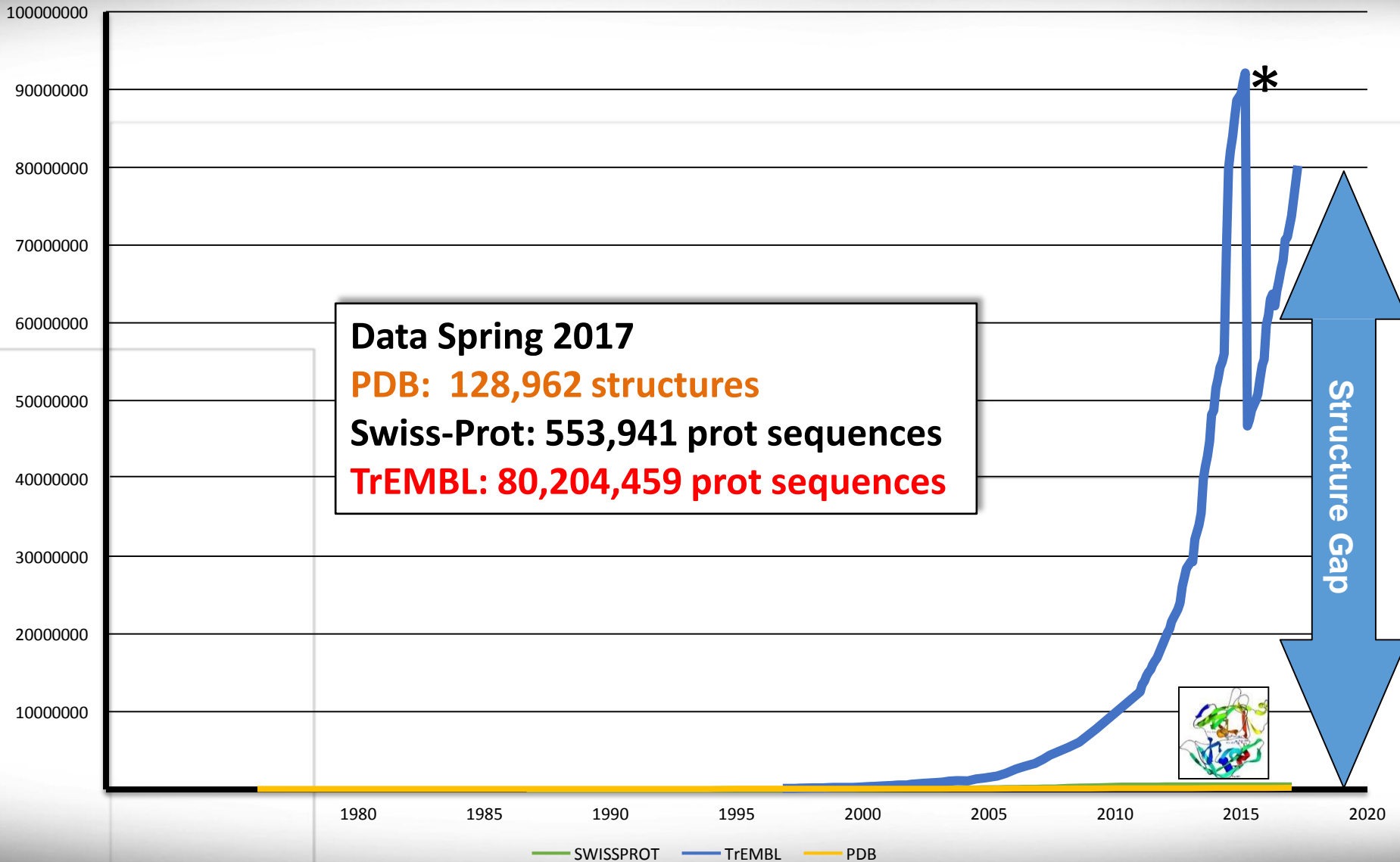
**43,823** Distinct Protein Sequences

**38,074** Structures of Human Sequences

**9,953** Nucleic Acid Containing Structures

# Structure gap in databases: a real problem

Redundant data removed

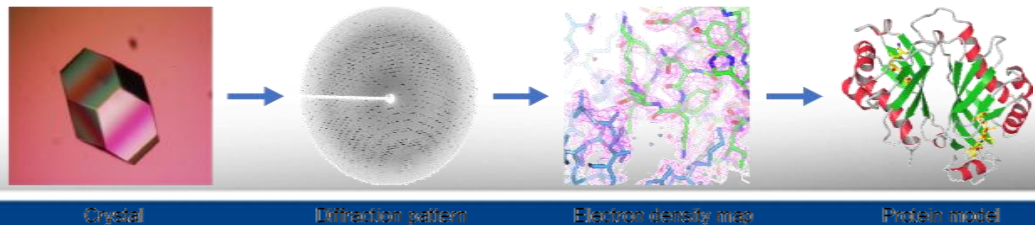
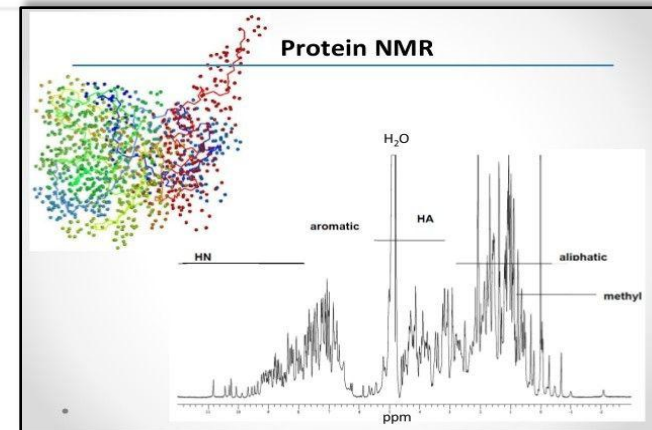




Experimental protein structure solution (eg. by NMR or X-Ray crystallography) is **labor** intensive and **expensive**.

For the majority of proteins in any given proteome, experimental structures are not available.

1. Is it possible to **predict** 3-dimensional protein structures **computationally**?
2. Which computational methods are **feasible** and applicable in a life science research context?



# Definitions of the components

## Uses of structural databases

- 1 Protein Structural Alignment:** The geometry of two given protein structures can be compared by means of available software tools that analyse their three dimensional similarity to each other.
- 2 Protein Structure Prediction:** The prospective secondary structures of peptides or proteins can be predicted from a given stretch of amino acid residues by using machine learning algorithms.
- 3 Functional Annotation:** For novel proteins that are yet to be characterized, the potential functions can be predicted by techniques such as Homology Modelling which provide an initial insight into the protein's properties.

# Homology modeling

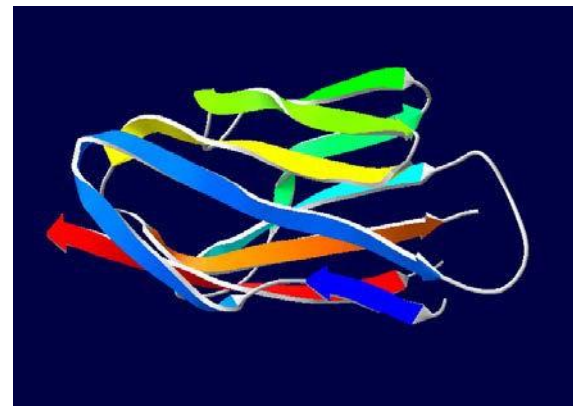
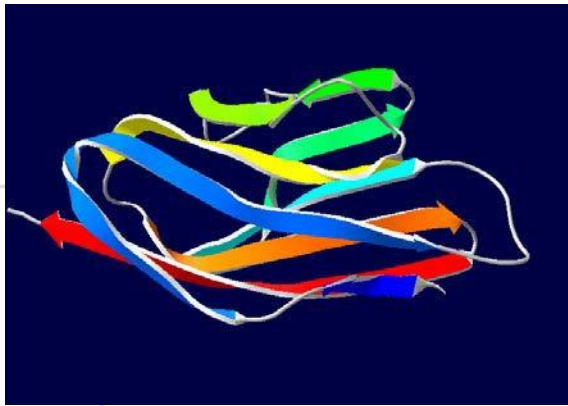
## “The biology perspective”

Homologous proteins have **evolved** by molecular evolution from a common ancestor over millions of years. If we can establish **homology to a known protein**, we can **predict aspects of structure and function** of a protein by similarity - **Charles Darwin**

# Darwin's evolution of protein structures

**Protein structure is better conserved than sequence**

**Similar Sequence = Similar Structure**



Homology modeling = Comparative protein modeling

**Idea:** Using experimental 3D-structures of related family members (templates) to calculate a model for a new sequence (target).

## SWISS-MODEL-Hhpred

Server for homology detection and structure prediction by **HMM-HMM comparison**.

## I-Tasser

I-TASSER is a server for protein structure and function predictions. 3D models are built based on multiple-threading alignments by **LOMETS** and **iterative TASSER assembly simulations**.

## QUARK

QUARK is a computer algorithm for **ab initio protein structure** prediction and protein peptide folding, which aims to construct the correct protein 3D model from amino acid sequence only. QUARK models are built from small fragments (1-20 residues long) by replica-exchange Monte Carlo simulation under the guide of **an atomic-level knowledge-based force field**.

# Protein modeling tools

## M4T

Comparative Modelling using a combination of **multiple templates** and **iterative optimization** of alternative alignments.

## Modeller

Software for homology or comparative modeling of protein three-dimensional structures. MODELLER implements comparative protein structure modeling by **satisfaction** of **spatial restraints**.

## ModWeb

A web server for automated comparative modeling that relies on **PSI-BLAST**, **IMPALA** and **MODELLER**.

## Phyre2

A fold recognition server for predicting the structure and/or function of your protein sequence.

# SWISS-MODEL

- **SWISS-MODEL** is a **web-based integrated service** dedicated to protein structure homology modelling.
- Building a homology model comprises **four main steps**: (1) identification of structural **template(s)**, (2) **alignment** of target sequence and template structure(s), (3) **model-building**, and (4) **model quality evaluation**.
- Modelling modes:
  - ✓ **Automated** - requires the amino acid sequence or the UniProtKB accession code
  - ✓ **Alignment** - if the template protein is known
  - ✓ **Project** - visual inspection and manual manipulation



# Chose your modelling mode

- You can either **paste the protein sequence** or **provide the UniprotKB** of your target sequence in the input form.
- To search for available template structures, click on the **“Search for Templates”** button

### Start a New Modelling Project

Target Target `MVVKAVCVINGDAKGTVFFEQESSGTPVKVSGEVCGLAKGLHGFHVHEFGDNTNGCMSSGPHFNFPYGKEHGAPVDENRHL` 80

Sequence(s):  
(Format must be  
FASTA, Clustal,  
plain string, or a valid  
UniProtKB AC)

Target `GDLGNIEATGDCPTKVNITDSKITLFGADSIIGRTVVVHADADDLGQGGHELKSTGNAGARIGCGVIGIAKV` 153

Project Title:

By using the SWISS-MODEL server, you agree to comply with the following [terms of use](#) and to cite the corresponding [articles](#).

### Supported Inputs

- Sequence(s)
- Target-Template Alignment
- User Template
- DeepView Project



# Your template results

- Build model by **selecting your template(s)**
- Chose the **best sequence similarity** (above 30% preferred)

## Template Results

Templates		Quaternary Structure	Sequence Similarity	Alignment of Selected Templates	More	
Name	Title	Coverage	Identity	Method	Oligo State	Ligands
<input checked="" type="checkbox"/> <input checked="" type="checkbox"/>	319e.1.A	Superoxide dismutase [Cu-Zn]	<div style="width: 68.63%;"></div>	68.63	X-ray, 2.0Å	homo-dimer ✓ 2 x ZN <sup>CS</sup>
<input checked="" type="checkbox"/>	319y.1.A	Superoxide dismutase [Cu-Zn]	<div style="width: 67.97%;"></div>	67.97	X-ray, 1.8Å	homo-dimer ✓ 2 x ZN <sup>CS</sup> , 2 x CU <sup>CS</sup>
<input checked="" type="checkbox"/>	319e.1.A	Superoxide dismutase [Cu-Zn]	<div style="width: 67.97%;"></div>	67.97	X-ray, 2.0Å	homo-dimer ✓ 2 x ZN <sup>CS</sup>
<input checked="" type="checkbox"/>	319y.1.A	Superoxide dismutase [Cu-Zn]	<div style="width: 67.32%;"></div>	67.32	X-ray, 1.8Å	homo-dimer ✓ 2 x ZN <sup>CS</sup> , 2 x CU <sup>CS</sup>
<input type="checkbox"/>	2zky.1.A	Superoxide dismutase [Cu-Zn]	<div style="width: 61.84%;"></div>	61.84	X-ray, 2.4Å	homo-dimer ✓ 2 x ZN <sup>CS</sup>
<input type="checkbox"/>	4b3e.1.A	SUPEROXIDE DISMUTASE [CU-ZN]	<div style="width: 61.84%;"></div>	61.84	X-ray, 2.1Å	homo-dimer ✓ 2 x ZN <sup>CS</sup> , 2 x CU <sup>CS</sup>
<input type="checkbox"/>	1n18.1.A	Superoxide dismutase [Cu-Zn]	<div style="width: 61.84%;"></div>	61.84	X-ray, 2.0Å	homo-dimer ✓ 2 x ZN <sup>CS</sup> , 2 x CU1 <sup>CS</sup>
<input type="checkbox"/>	2zky.1.A	Superoxide dismutase [Cu-Zn]	<div style="width: 61.18%;"></div>	61.18	X-ray, 2.4Å	homo-dimer ✓ 2 x ZN <sup>CS</sup>
<input type="checkbox"/>	3ltv.1.A	Superoxide dismutase [Cu-Zn], Superoxide dismutase [Cu-Zn]	<div style="width: 61.33%;"></div>	61.33	X-ray, 2.5Å	homo-dimer ✓ 2 x ZN <sup>CS</sup>
<input type="checkbox"/>	4b3e.1.A	SUPEROXIDE DISMUTASE [CU-ZN]	<div style="width: 61.18%;"></div>	61.18	X-ray, 2.1Å	homo-dimer ✓ 2 x ZN <sup>CS</sup> , 2 x CU <sup>CS</sup>
<input type="checkbox"/>	1n19.1.A	Superoxide Dismutase [Cu-Zn]	<div style="width: 61.18%;"></div>	61.18	X-ray, 1.9Å	homo-dimer ✓ 2 x ZN <sup>CS</sup> , 2 x CU1 <sup>CS</sup>
<input type="checkbox"/>	2zky.1.B	Superoxide dismutase [Cu-Zn]	<div style="width: 61.18%;"></div>	61.18	X-ray, 2.7Å	homo-dimer ✓ 2 x ZN <sup>CS</sup> , 2 x CU1 <sup>CS</sup>
<input type="checkbox"/>	2zkw.1.A	Superoxide dismutase [Cu-Zn]	<div style="width: 61.18%;"></div>	61.18	X-ray, 1.9Å	homo-dimer ✓ 2 x ZN <sup>CS</sup> , 2 x CU1 <sup>CS</sup>
<input type="checkbox"/>	3gtt.1.A	Superoxide dismutase [Cu-Zn]	<div style="width: 60.67%;"></div>	60.67	X-ray, 2.4Å	homo-dimer ✓ 2 x ZN <sup>CS</sup>
<input type="checkbox"/>	3ltv.1.A	Superoxide dismutase [Cu-Zn], Superoxide dismutase [Cu-Zn]	<div style="width: 60.67%;"></div>	60.67	X-ray, 2.5Å	homo-dimer ✓ 2 x ZN <sup>CS</sup>

Build Models 4

Clear Selection

PV
Cartoon
📷
▶
▲
🔄

319e.1.A	<input type="text"/>	✕
319y.1.A	<input type="text"/>	✕
319e.1.A	<input type="text"/>	✕
319y.1.A	<input type="text"/>	✕

# Model results & evaluation

- **GMQE (Global Model Quality Estimation)** combines properties from the target–template alignment and the template search method. GMQE score is expressed as a **number between 0 and 1**, reflecting the **expected accuracy of a model built** with that alignment and template and the coverage of the target.

## Template Results

Templates    Quaternary Structure    Sequence Similarity    Alignment of Selected Templates    More ▾

Name	Title	Coverage	Identity	Method	Oligo State	Ligands
319e.1.A	Superoxide dismutase [Cu-Zn]		68.63	X-ray, 2.0Å	homo-dimer ✓	2 x ZN <sup>CS</sup>
319y.1.A	Superoxide dismutase [Cu-Zn]		67.97	X-ray, 1.8Å	homo-dimer ✓	2 x ZN <sup>CS</sup> , 2 x CU <sup>CS</sup>

**Method** X-RAY DIFFRACTION 1.80 Å

**Found By** BLAST

**GMQE** 0.84

**Seq Similarity** 0.51

**QSQE** 0.72

Biounit Oligo State homo-dimer

Target Prediction It is possible to build a homo-dimer. The target model is also **predicted** to be a homo-dimer.  
Build a homo-dimer  monomer

Target M V V K A V C V I N G D A K G T V F F E Q E S S G P V P V K V S G E V C G L A K G L H G F H V H E F G D N I N G C M S S G P H F N P Y G K E H G A P V D 75  
319y.1.A M V V K A V C V L R G D V S G T V F F L Q Q E K S P V V S G E V G L I K G K H G F H V H E F G D N I N G C I S A G A H F N P E K Q D H G P P S 75

Target E N R H L G D L G N I E A T G D C - P T K V N I T D S K I T L F G A D S I I G R T V V V H A D A D D L G Q G G H E L S K S T I G N A G A R I G C G V I G 149  
319y.1.A A V R H V G D L G N I E A T E D A G V T K V S I Q D S Q I S L H G P N S I I G R T L V V H A D P D D L G L G G N E L S K T I G N A G R I A C G V I G 150

Target I A K V 153  
319y.1.A A K 154

Build Models 4

Clear Selection

Build Model

PV ↑    Cartoon ↑    [Camera]    [Play]    [Up]    [Refresh]

# Model results & evaluation

- **QMEAN**, a **composite estimator** based on different geometrical properties and provides both **global** (i.e. for the entire structure) and **local** (i.e. per residue) absolute quality estimates on the basis of **one single model**.
- QMEAN Z-scores **around zero is good**, but of **-4.0 or below** are an indication of models with **low quality**

## Model Results

Order by: GMQE

GMQE	QMEAN
0.88	0.66



Model 01

**Oligo-State**  
Homo-dimer (matching prediction)

**Ligands**  
2 x CU<sup>2+</sup>, 2 x ZN<sup>2+</sup>  
2 x COPPER (II) ION  
2 x ZINC ION

Ligand 2 in contact with: Chain A : H45, H47, H62, V117, H119  
Ligand 4 in contact with: Chain B : H45, H47, H62, V117, H119

Ligand 1 in contact with: Chain A : H62, H70, H79, D82  
Ligand 3 in contact with: Chain B : H62, H70, H79, D82

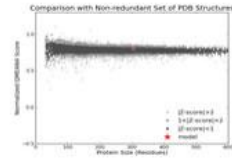
**Global Quality Estimate**

QMEAN	0.66
C $\beta$	0.29
All Atom	-1.23
Solvation	-1.25
Torsion	0.94

**Local Quality Estimate**



**Comparison**



**Template** 319y.1.A  
**Seq Identity** 67.97%  
**Coverage**

**Description**  
Superoxide dismutase [Cu-Zn]

**Model-Template Alignment**

Model_01:A	MVVKAVCVINGDAKGTVPFPEQESSGTPVKVSGEVCGLAKGLHGPHVHEFGDNTNGCMSGGPHFNPYGKEHGAPVDENRHL	80
Model_01:B	MVVKAVCVINGDAKGTVPFPEQESSGTPVKVSGEVCGLAKGLHGPHVHEFGDNTNGCMSGGPHFNPYGKEHGAPVDENRHL	80
319y.1.A	MKAVCVLRDQSGIVFFDDEKSPVVVSGEYGLTKGKHGFHVDFEGDNTNGCCTSGADFNPEKQDHGGPSSAVRHV	80
Model_01:A	GDLGNIEATGDCPTKVNITDSKITLFGADSIIGRTVVVHADADDLGGGGHLSKSTGNAGARIGCGVIGIAKV	153
Model_01:B	GDLGNIEATGDCPTKVNITDSKITLFGADSIIGRTVVVHADADDLGGGGHLSKSTGNAGARIGCGVIGIAKV	153
319y.1.A	GDLGNIEADIEDAGVTKVSIKDSQILDHGNISIIIGRTLVVHADDDDLGGNELSKITGNAGARICGVIGIAKI	154

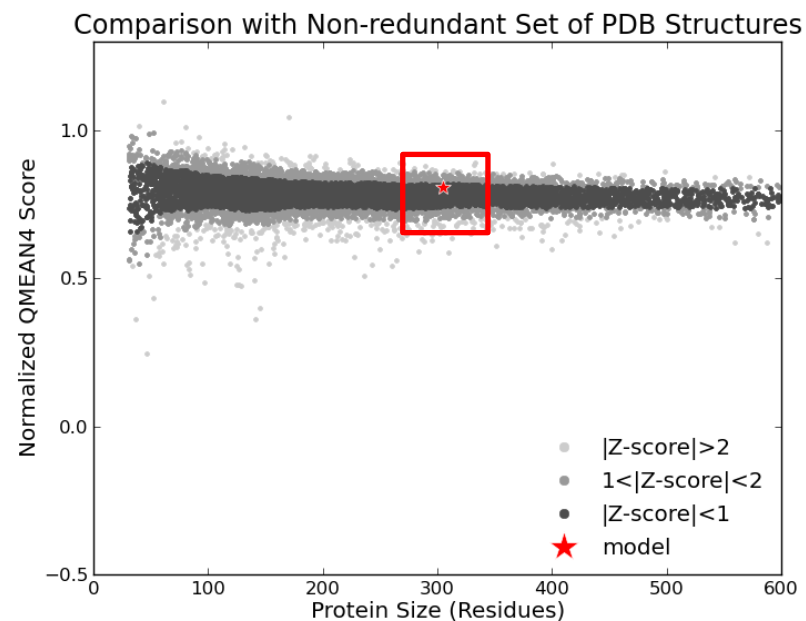
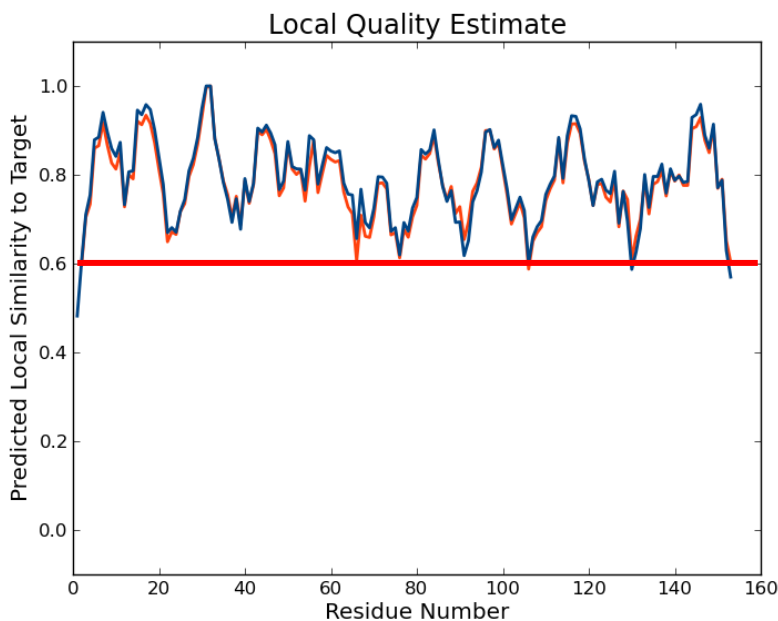


PV Cartoon

1 153

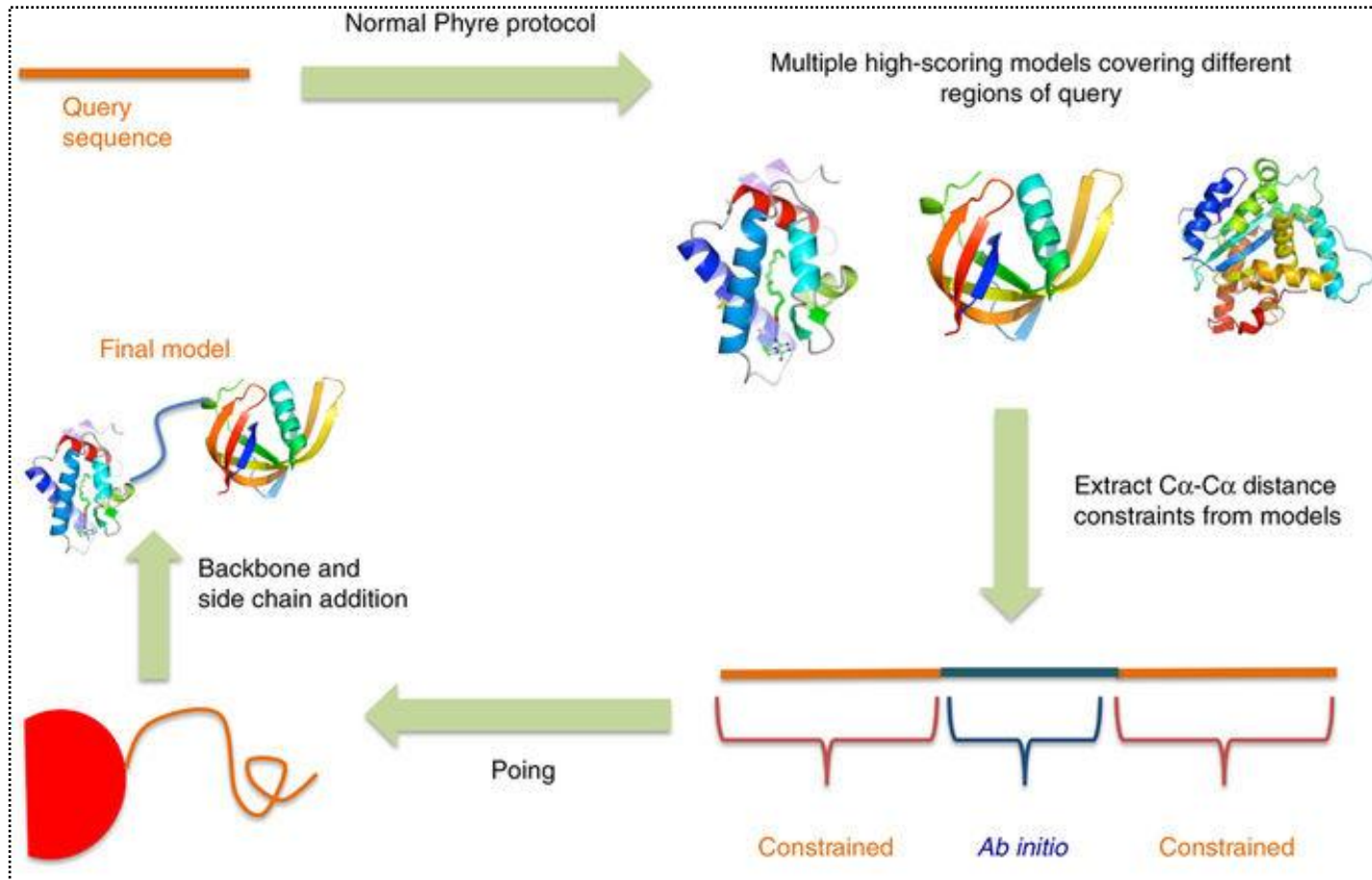
# Model results & evaluation

- The “**Local Quality**” estimates, for each residue of the model (reported on the x-axis), the expected similarity to the native structure (y-axis). Typically, residues showing **a score below 0.6 are expected to be of low quality**.
- The “**Comparison**” plot models quality scores of **individual models** related to scores obtained for **experimental structures** of similar size. **Query inside normal distribution** of existing model is **great**.



# Phyre2

- **Protein Homology/analogY Recognition Engine V 2.0** (Phyre2) provides an intensive mode to create a complete full-length model of a sequence through a **combination of multiple template modeling** and simplified **ab initio folding simulationists** with a simple and intuitive interface.



Eliminate **constraints** from several models and create final model using **Poing** (protein-folding simulator)

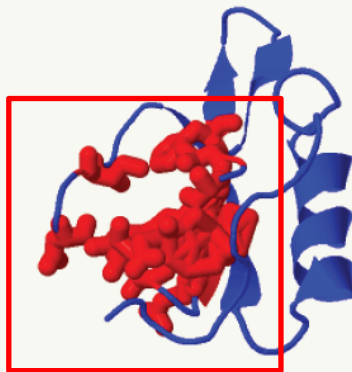
## A. Identification of pocket using Phyre2 investigators

### Pocket detection

Large pockets are frequently found to be the location of active sites. The largest pocket as detected by the [fpocket2](#) program are shown in wireframe mode, coloured red.

[Download raw data](#)

Largest pocket  
Pocket



### Analyses

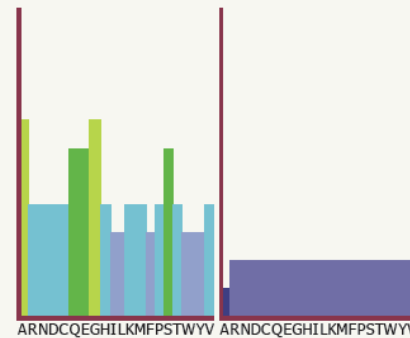
Residue: ALA 2

#### Quality

- Conservation
- Pocket detection
- Mutational sensitivity

#### Sequence profile

#### Mutations

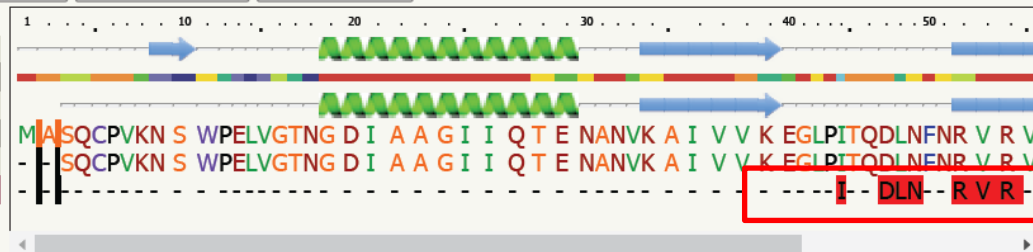


JSmol

Take JMol snapshot Show All analyses Hide All analyses Clear Selection

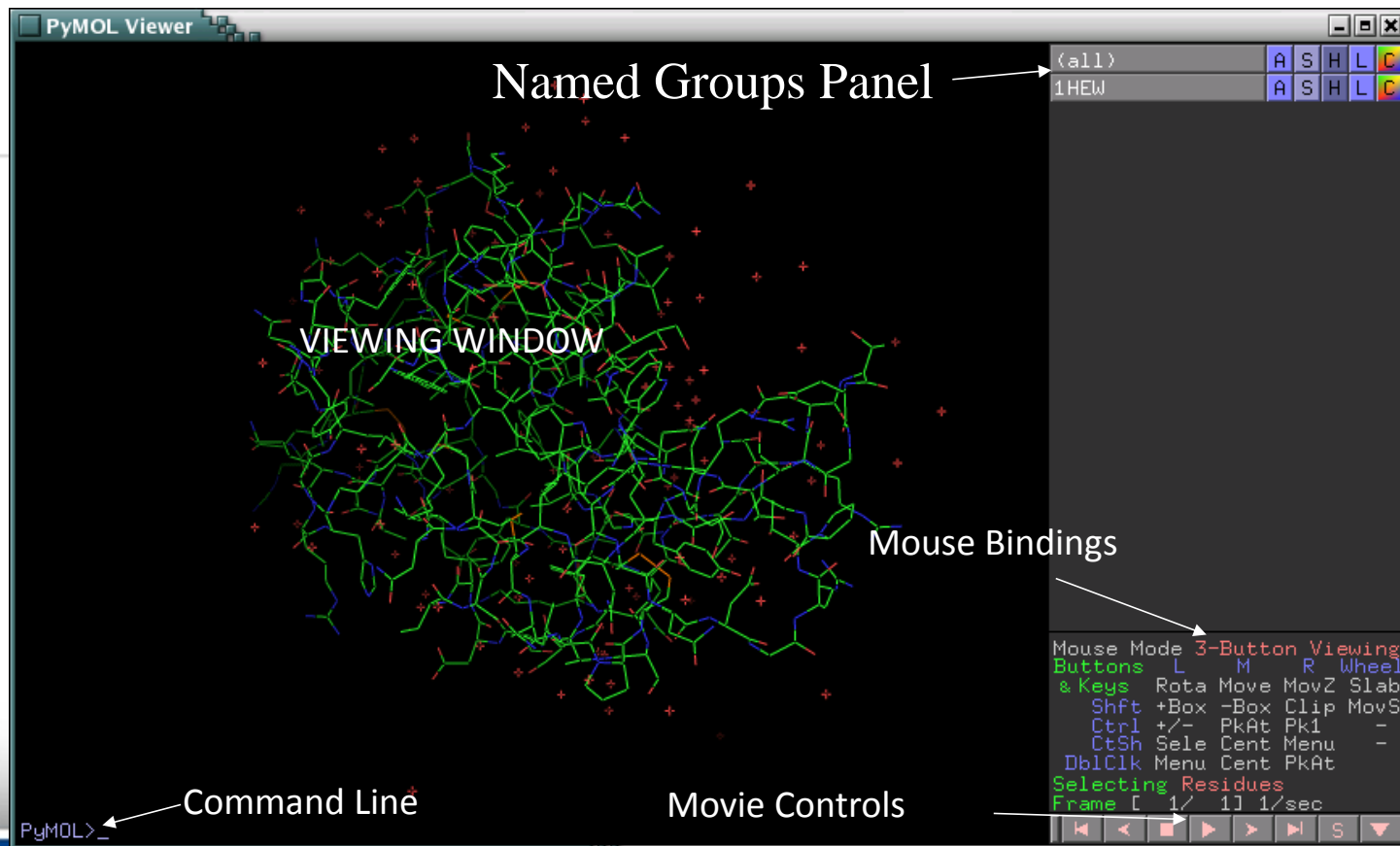
Pocket was identified at the C-terminal sites

Predicted Secondary structure  
SS Confidence  
Model Secondary structure  
Query Sequence  
Modelled Residues  
Pocket detection

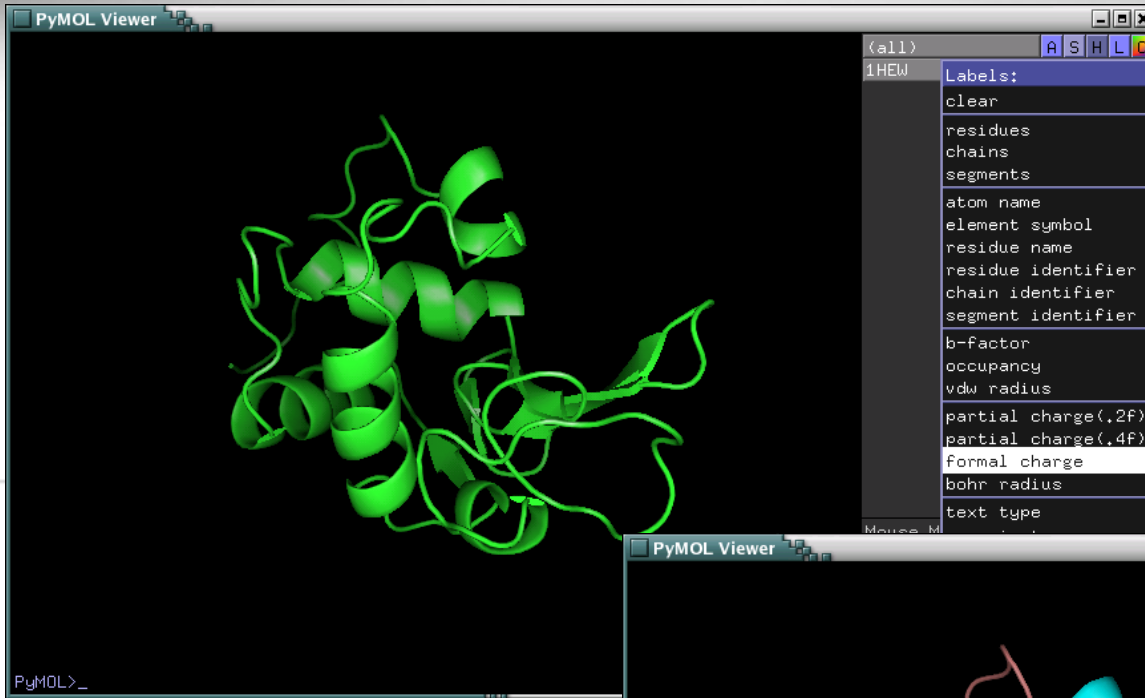


# Visualization of protein model: PyMOL

- **PyMOL** is a user-sponsored molecular **visualization system** on an open-source foundation, maintained and distributed by Schrödinger.

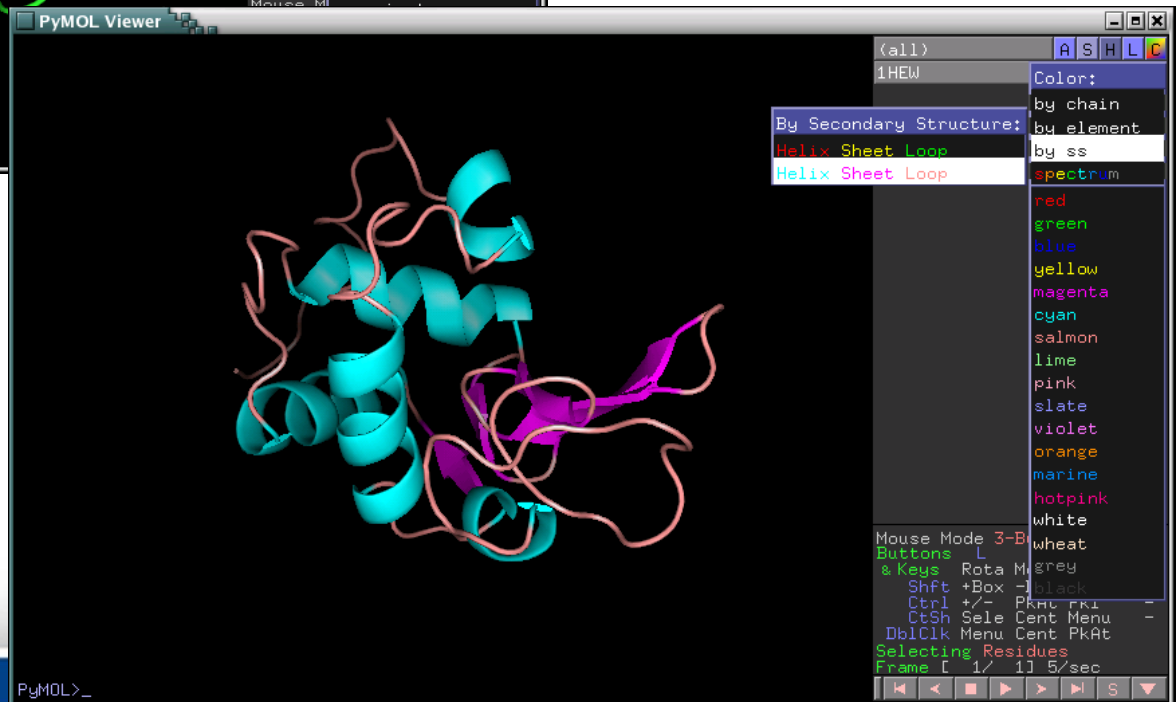


# Play with PyMOL



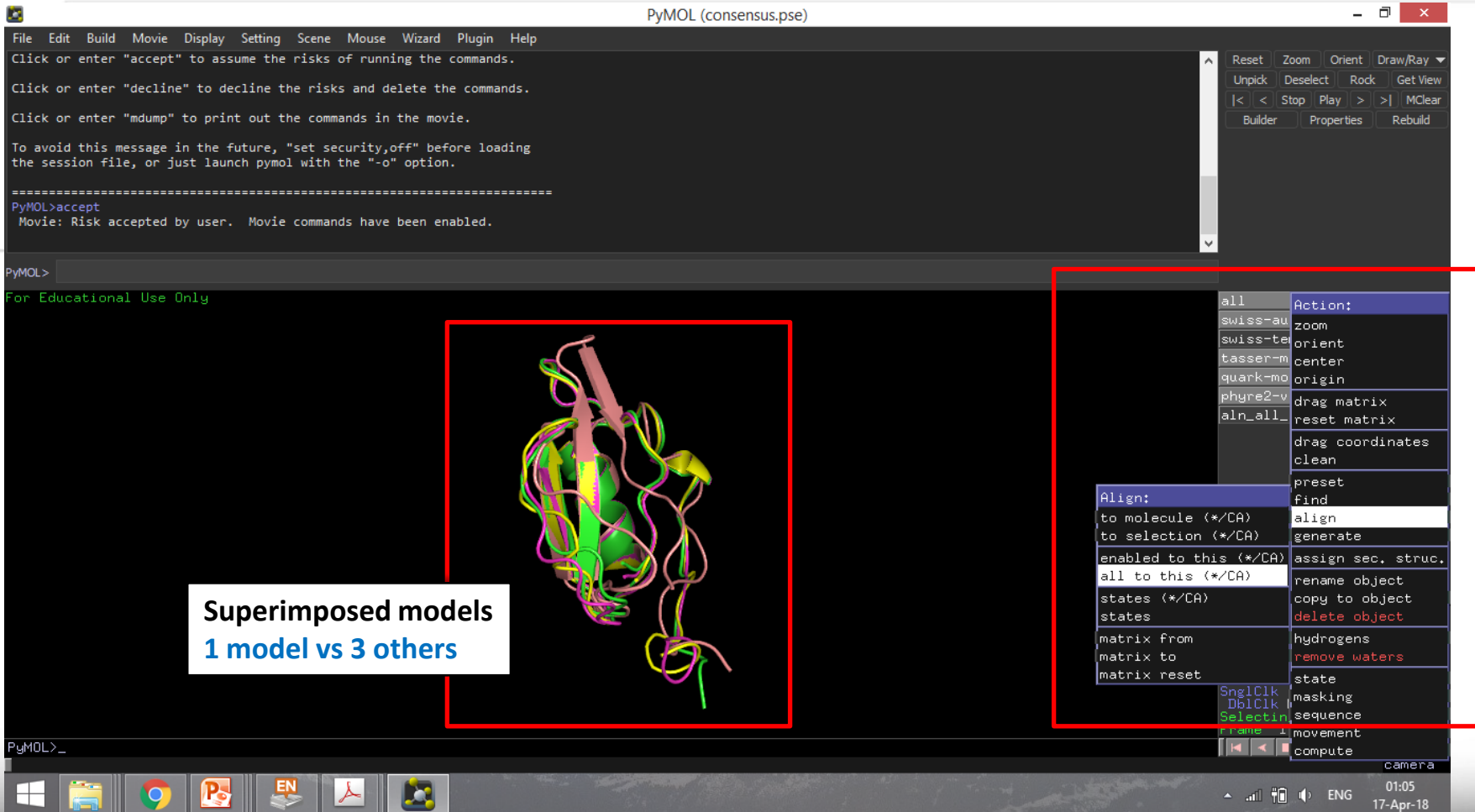
**Label panel:**  
residues, chains,  
segments

**Color panel:** by chain,  
by element, by  
secondary structure,  
by spectrum





- Take on protein model and **align with molecule or all models** viewed in PyMOL



PyMOL (consensus.pse)

File Edit Build Movie Display Setting Scene Mouse Wizard Plugin Help

Click or enter "accept" to assume the risks of running the commands.

Click or enter "decline" to decline the risks and delete the commands.

Click or enter "mdump" to print out the commands in the movie.

To avoid this message in the future, "set security,off" before loading the session file, or just launch pymol with the "-o" option.

=====  
PyMOL>accept  
Movie: Risk accepted by user. Movie commands have been enabled.

PyMOL>  
For Educational Use Only

all Action:  
swiss-au zoom  
swiss-te orient  
tasser-m center  
quark-mo origin  
phyre2-v drag matrix  
aln\_all\_ reset matrix  
drag coordinates  
clean  
preset  
find  
Align:  
to molecule (\*CA) align  
to selection (\*CA) generate  
enabled to this (\*CA) assign sec. struc.  
all to this (\*CA) rename object  
states (\*CA) copy to object  
states delete object  
matrix from hydrogens  
matrix to remove waters  
matrix reset state  
Sng1C1k masking  
Db1C1k sequence  
Selectin movement  
frame compute  
camera

PyMOL>\_

01:05  
17-Apr-18

**Superimposed models  
1 model vs 3 others**

## TUGAS PRAKTIKUM ---- KUMPULKAN MINGGU DEPAN

1. Buat lah struktur 3D protein dengan menggunakan analisis **Swiss-model** dan **PyMol** dari gen sebelumnya yang menjadi tugas saudara
2. Laporan dikumpul minggu depan

\*gen penyandi protein tugas sebelumnya.

THANK  
YOU



607132.wordpress.com

Noviani's Blog

