Universitas Esa Unggul

Smart, Creative and Entrepreneurial

www.esaunggul.ac.id

PENGANTAR BIOINFORMATIKA
IBT 431

By Seprianto S.Pi, M.Si

Meeting 12

Akses Struktur Protein Data Base

Universitas Esa Unggul

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)

- 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.

Penelusuran database protein

- **They are big databases too:**

Swiss-Prot (very high level of annotation)
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)

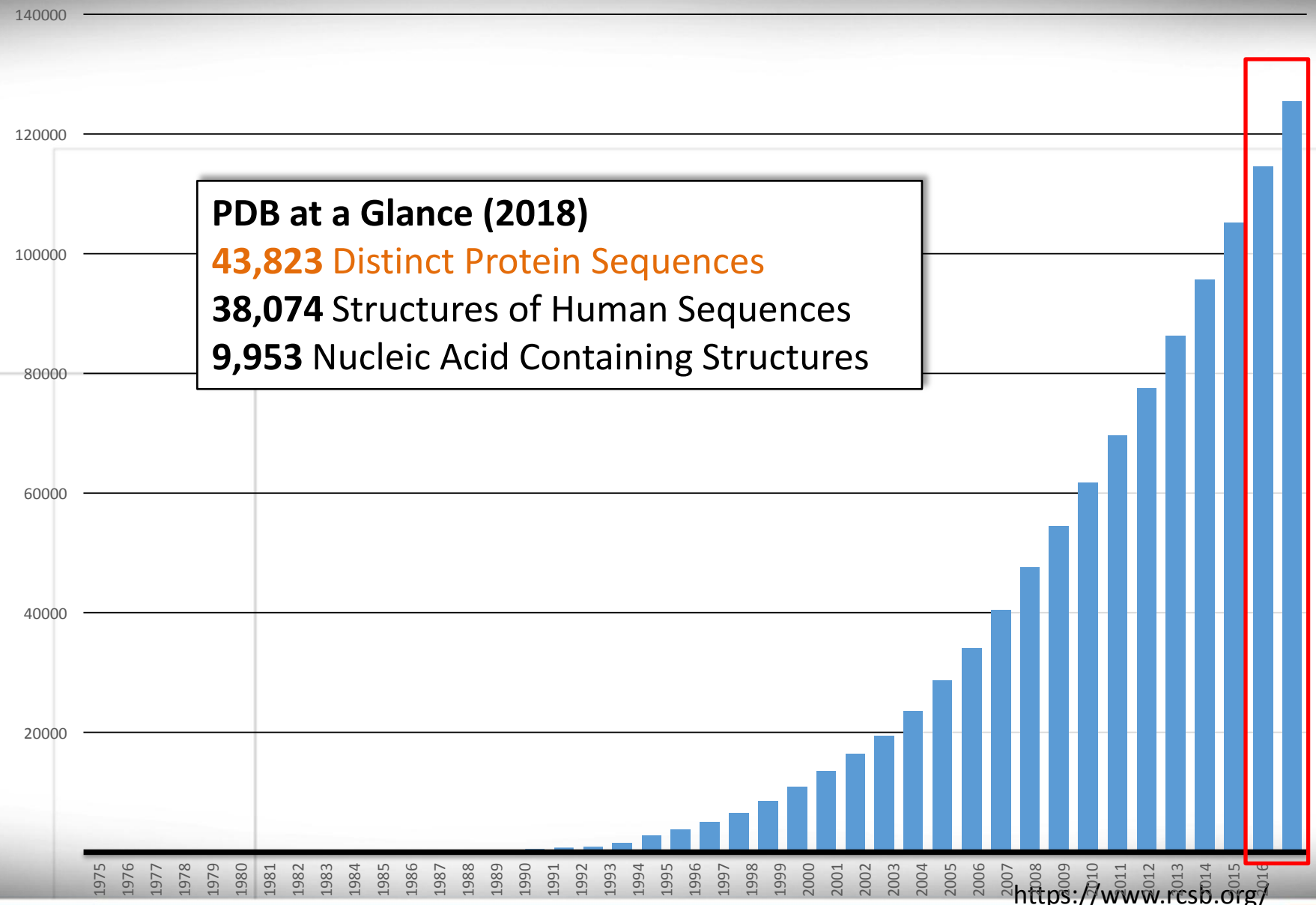
PDB (sequences derived from the 3D structure Brookhaven PDB) http://www.rcsb.org/pdb/

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.

https://www.rcsb.org/

Protein structure submission in PDB (2018)



PDB at a Glance (2018)

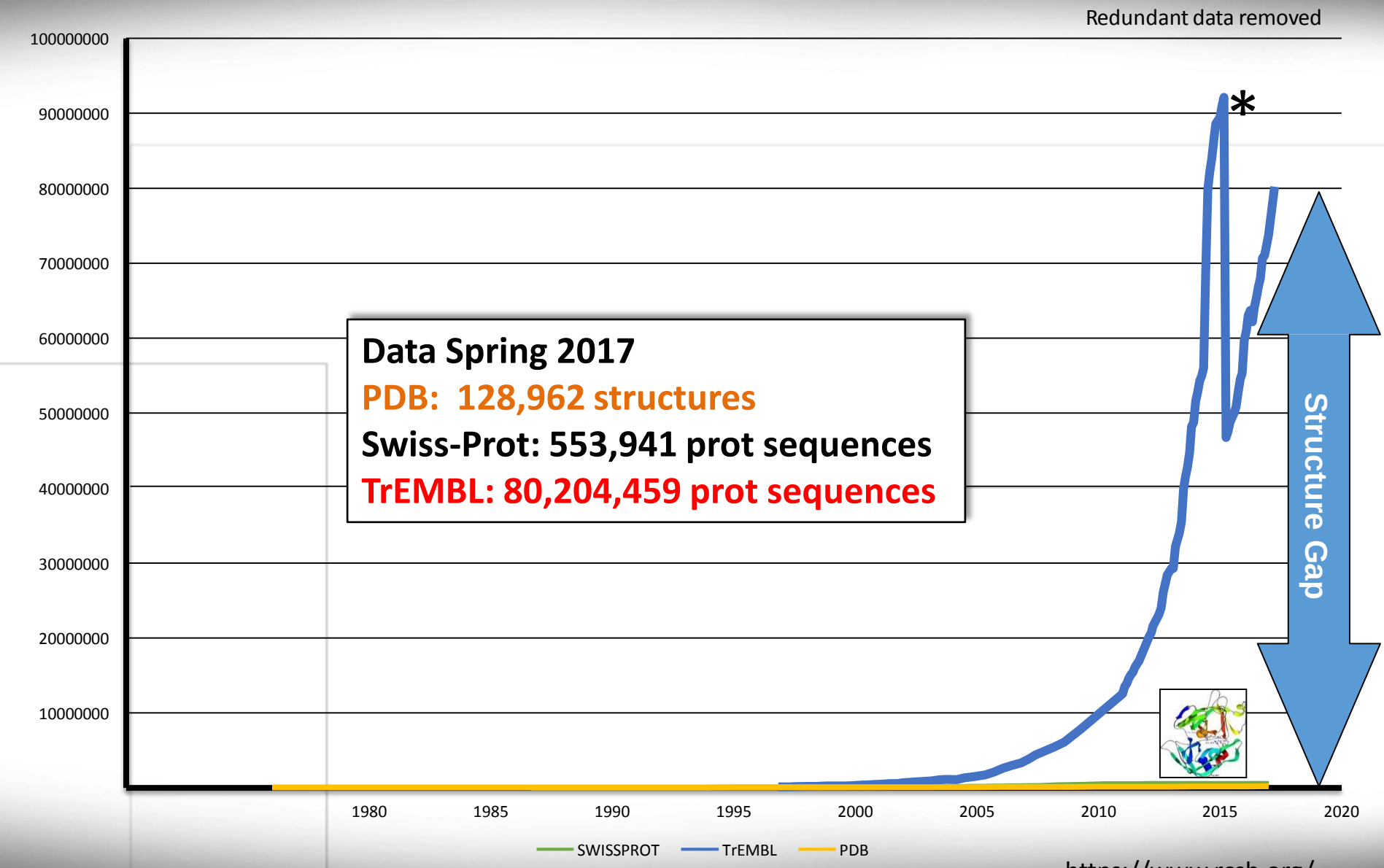
43,823 Distinct Protein Sequences

38,074 Structures of Human Sequences

9,953 Nucleic Acid Containing Structures

https://www.rcsb.org/

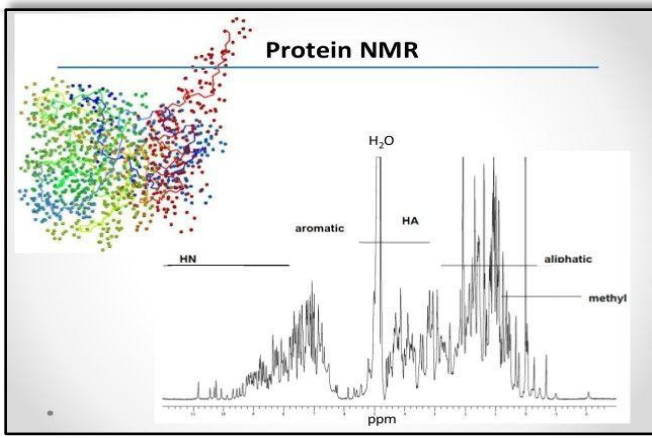
Structure gap in databases: a real problem



https://www.rcsb.org/

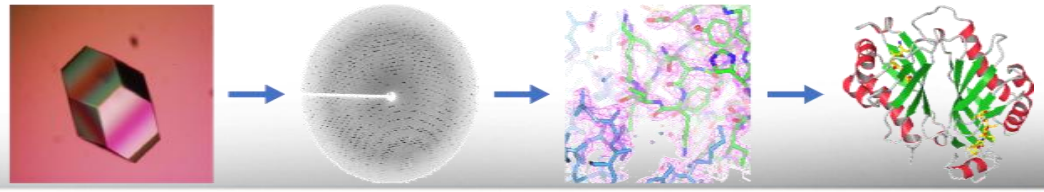
Protein structure gap

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

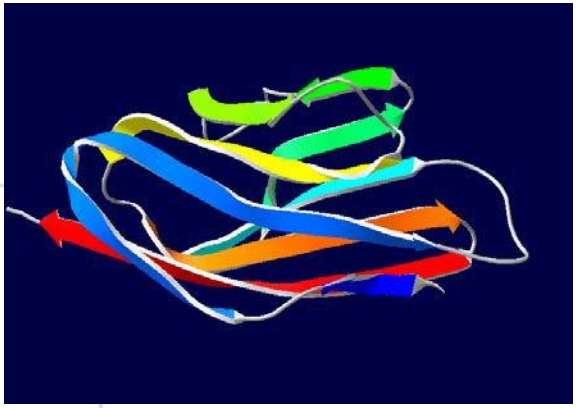
"The biology perspective"

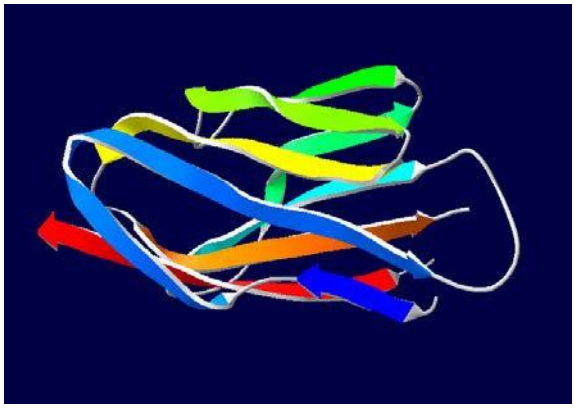
Homologous proteins have **evolved** by molecular evolution from a common ancestor over millions of years. If we can establish **homology to a known protein**, we can **predict aspects of structure and function** of a protein by similarity - **Charles Darwin**

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).**

Protein modeling tools

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.

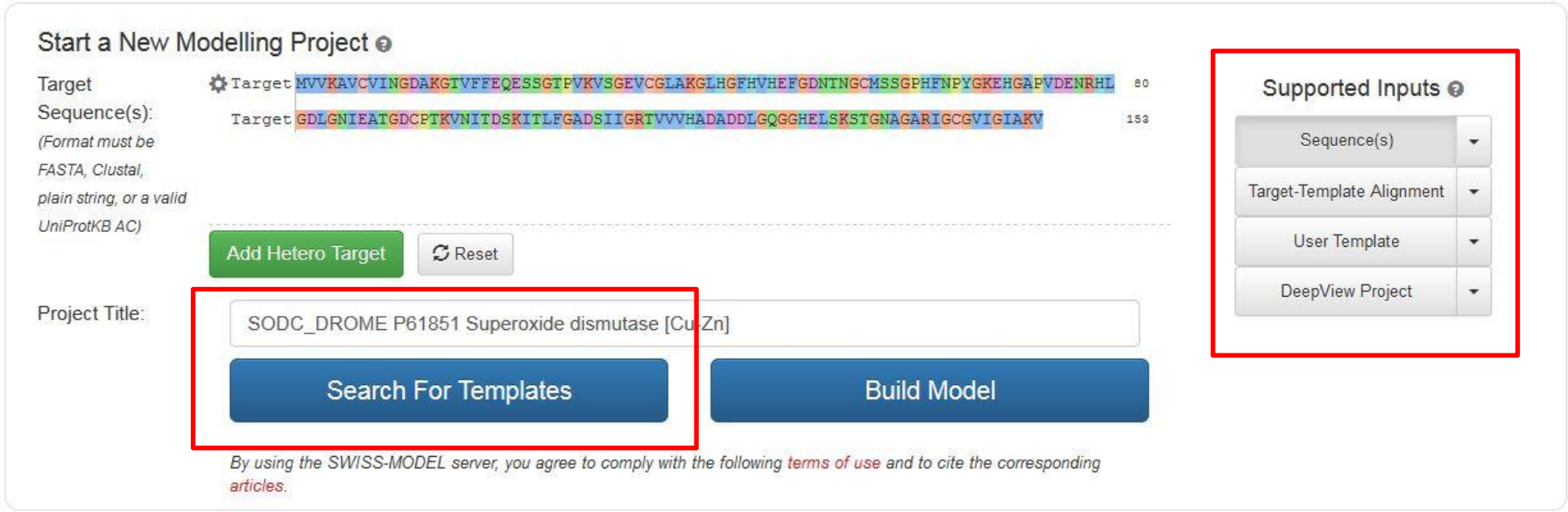
SWISS-MODEL

- **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

SIB
Swiss Institute of Bioinformatics

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



Your template results

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

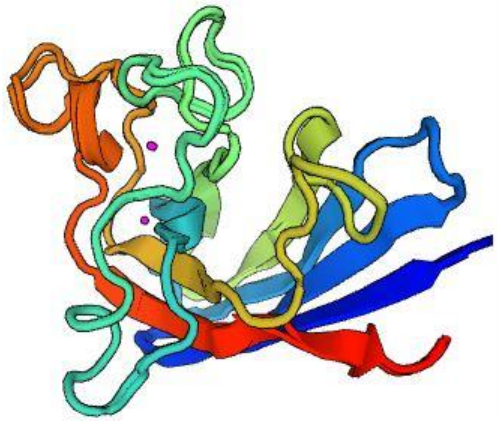
Template Results

Templates | Quaternary Structure | Sequence Similarity | Alignment of Selected Templates | More

	Name	Title	Coverage	Identity	Method	Oligo State	Ligands
☑ ✓	3l9e.1.A	Superoxide dismutase [Cu-Zn]		68.63	X-ray, 2.0Å	homo-dimer ✓	2 x ZN
☑	3l9y.1.A	Superoxide dismutase [Cu-Zn]		67.97	X-ray, 1.8Å	homo-dimer ✓	2 x ZN, 2 x CU
☑	3l9e.1.A	Superoxide dismutase [Cu-Zn]		67.97	X-ray, 2.0Å	homo-dimer ✓	2 x ZN
☑	3l9y.1.A	Superoxide dismutase [Cu-Zn]		67.32	X-ray, 1.8Å	homo-dimer ✓	2 x ZN, 2 x CU
☐	2zky.1.A	Superoxide dismutase [Cu-Zn]		61.84	X-ray, 2.4Å	homo-dimer ✓	2 x ZN
☐	4b3e.1.A	SUPEROXIDE DISMUTASE [CU-ZN]		61.84	X-ray, 2.1Å	homo-dimer ✓	2 x ZN, 2 x CU
☐	1n18.1.A	Superoxide dismutase [Cu-Zn]		61.84	X-ray, 2.0Å	homo-dimer ✓	2 x ZN, 2 x CU1
☐	2zky.1.A	Superoxide dismutase [Cu-Zn]		61.18	X-ray, 2.4Å	homo-dimer ✓	2 x ZN
☐	3ltv.1.A	Superoxide dismutase [Cu-Zn],Superoxide dismutase [Cu-Zn]		61.33	X-ray, 2.5Å	homo-dimer ✓	2 x ZN
☐	4b3e.1.A	SUPEROXIDE DISMUTASE [CU-ZN]		61.18	X-ray, 2.1Å	homo-dimer ✓	2 x ZN, 2 x CU
☐	1n19.1.A	Superoxide Dismutase [Cu-Zn]		61.18	X-ray, 1.9Å	homo-dimer ✓	2 x ZN, 2 x CU1
☐	2zkx.1.B	Superoxide dismutase [Cu-Zn]		61.18	X-ray, 2.7Å	homo-dimer ✓	2 x ZN, 2 x CU1
☐	2zkw.1.A	Superoxide dismutase [Cu-Zn]		61.18	X-ray, 1.9Å	homo-dimer ✓	2 x ZN, 2 x CU1
☐	3gtt.1.A	Superoxide dismutase [Cu-Zn]		60.67	X-ray, 2.4Å	homo-dimer ✓	2 x ZN
☐	3ltv.1.A	Superoxide dismutase [Cu-Zn],Superoxide dismutase [Cu-Zn]		60.67	X-ray, 2.5Å	homo-dimer ✓	2 x ZN

Build Models 4

Clear Selection



PV | Cartoon

3l9e.1.A

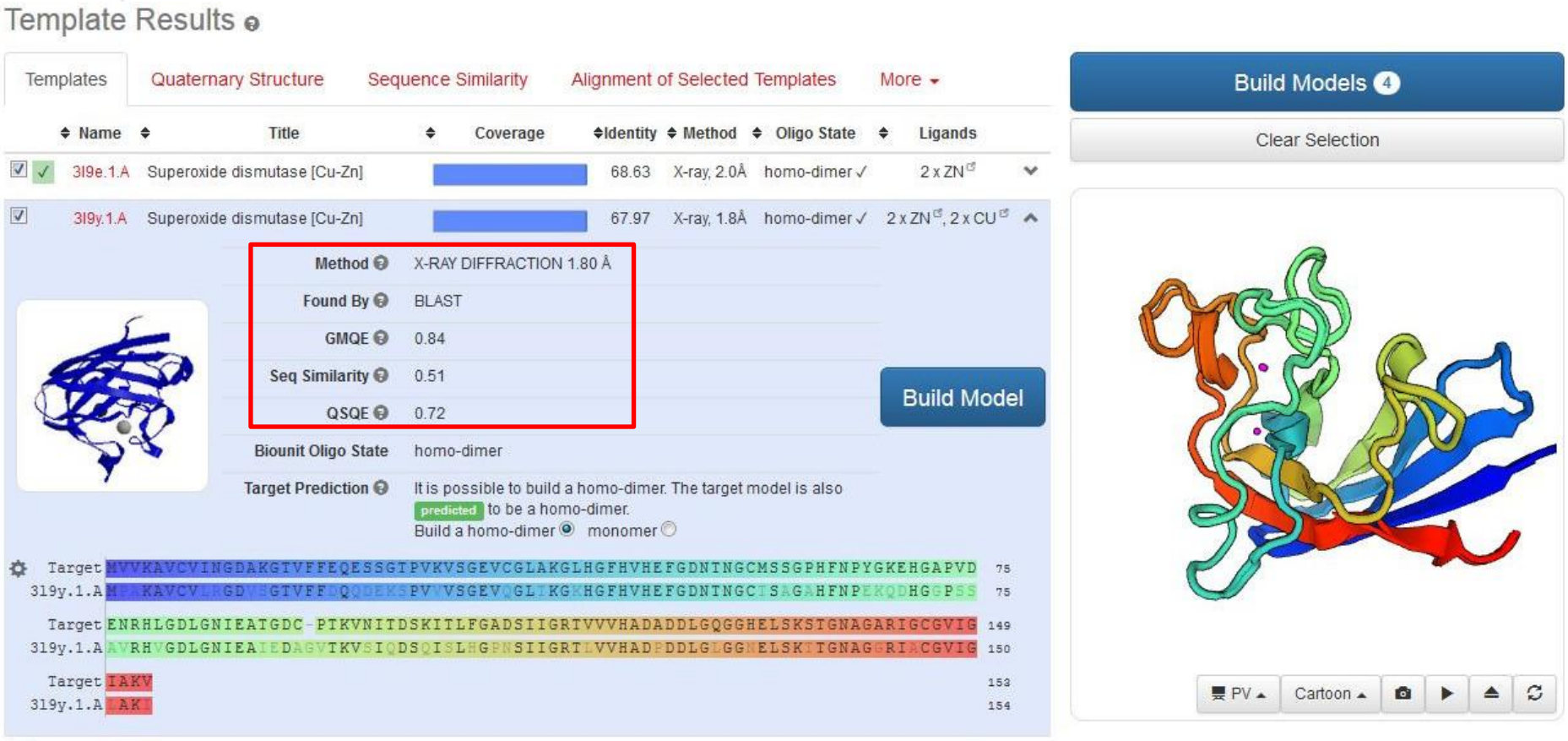
3l9y.1.A

3l9e.1.A

3l9y.1.A

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.

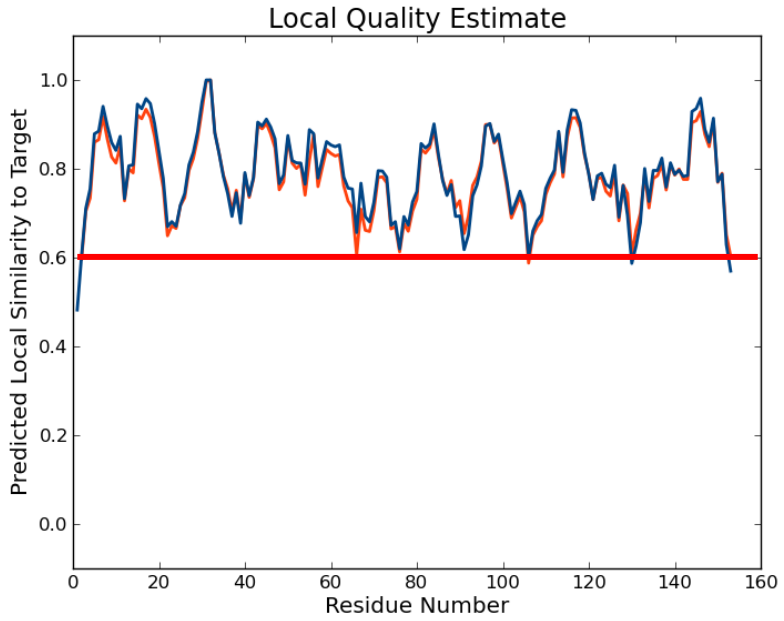


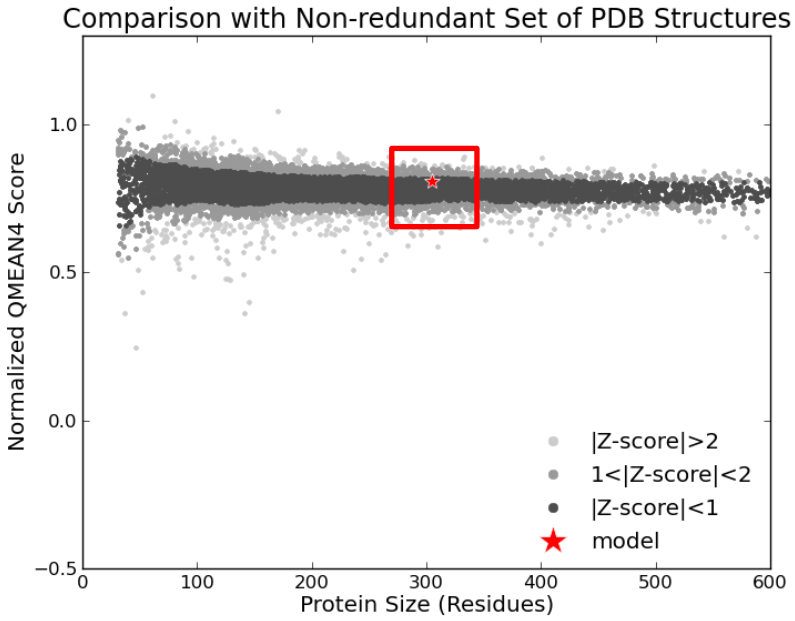
Model results & evaluation

- **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 & evaluation

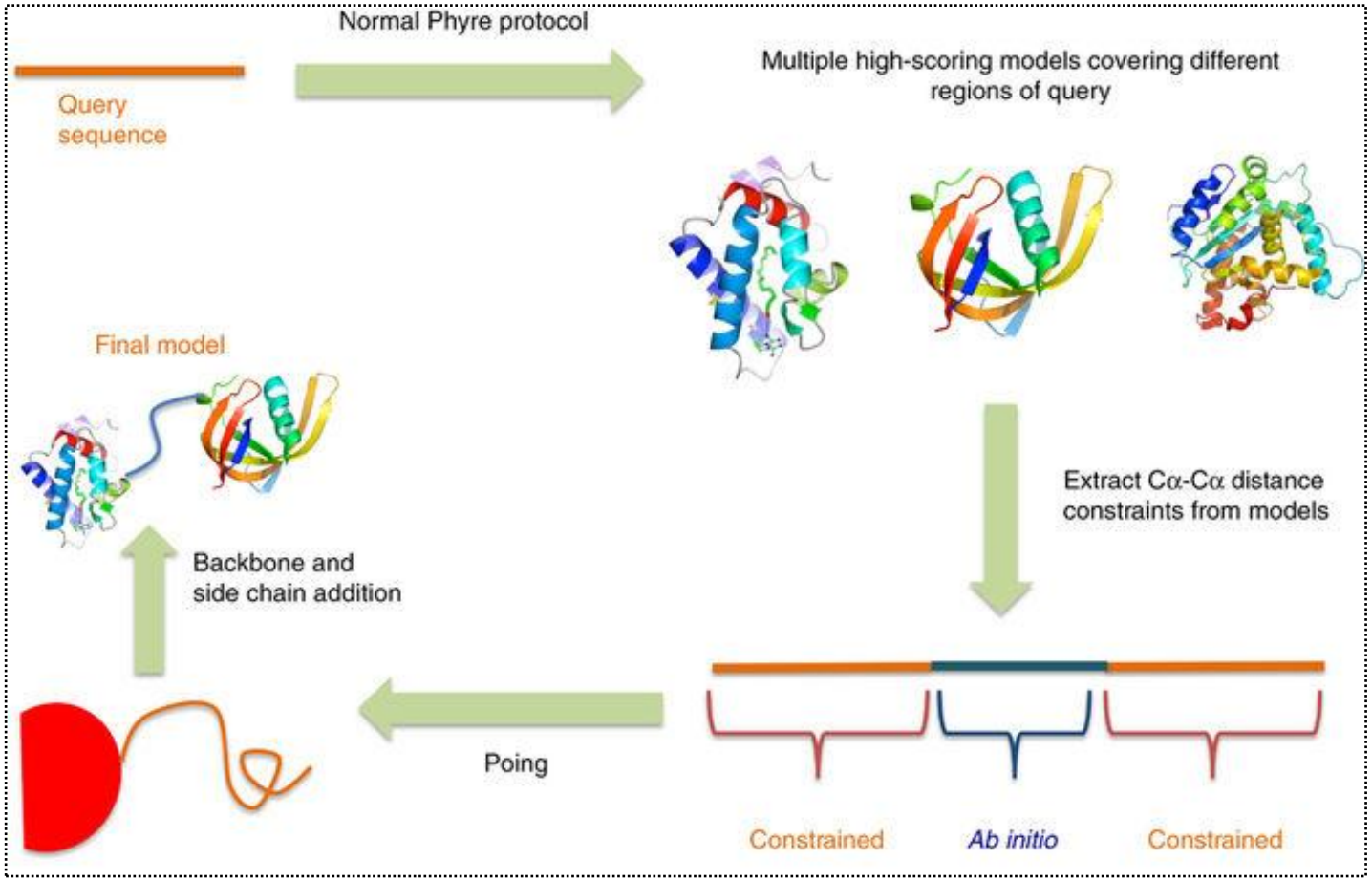
- The "**Local Quality**" estimates, for each residue of the model (reported on the x-axis), the expected similarity to the native structure (y-axis). Typically, residues showing **a score below 0.6 are expected to be of low quality**.
- The "**Comparison**" plot models quality scores of **individual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





Phyre2

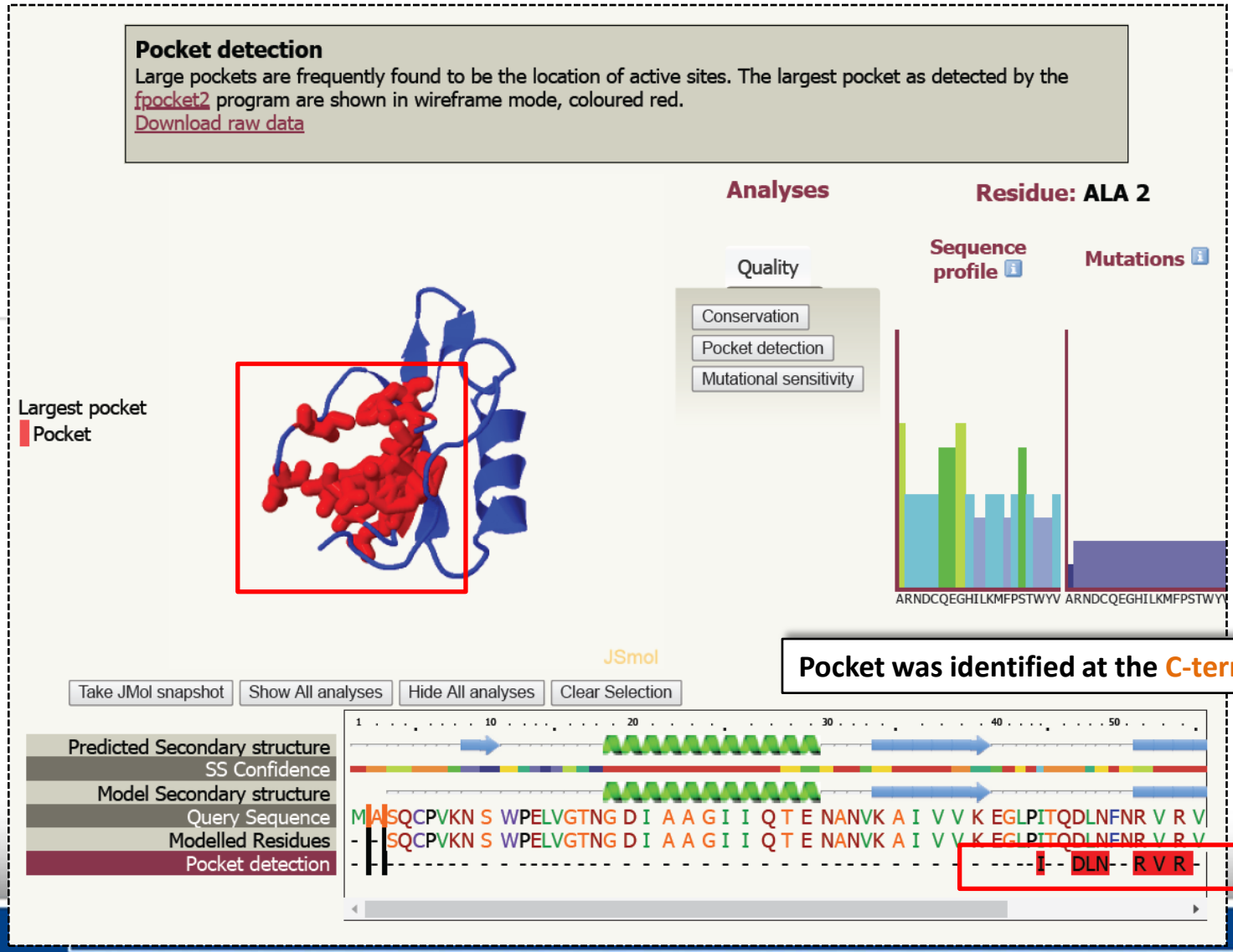
- **Protein Homology/analogY Recognition Engine V 2.0** (Phyre2) provides an intensive mode to create a complete full-length model of a sequence through a **combination of multiple template modeling** and simplified **ab initio folding simulationists** with a simple and intuitive interface.



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

Phyre2 results

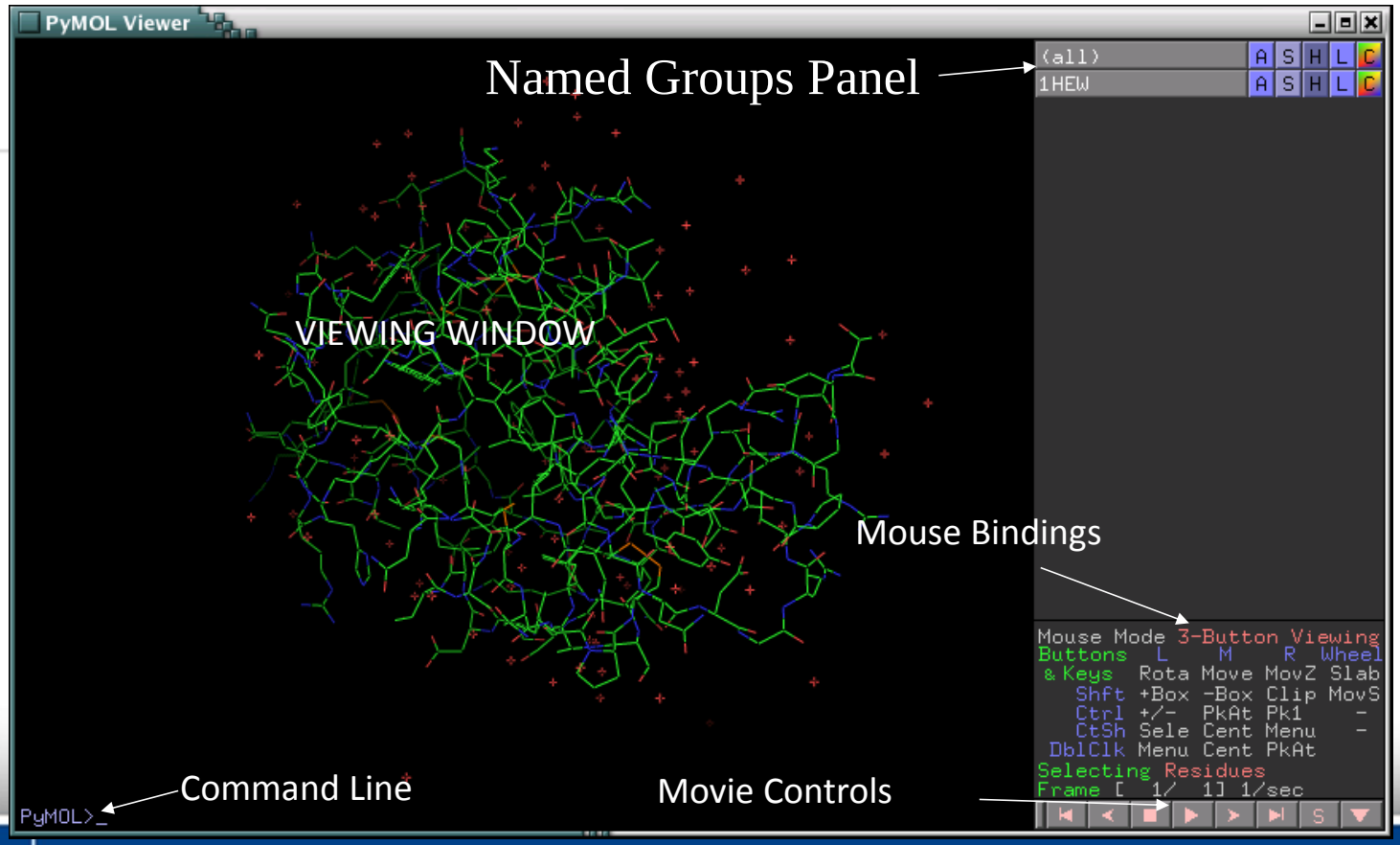
A. Identification of pocket using Phyre2 investigators



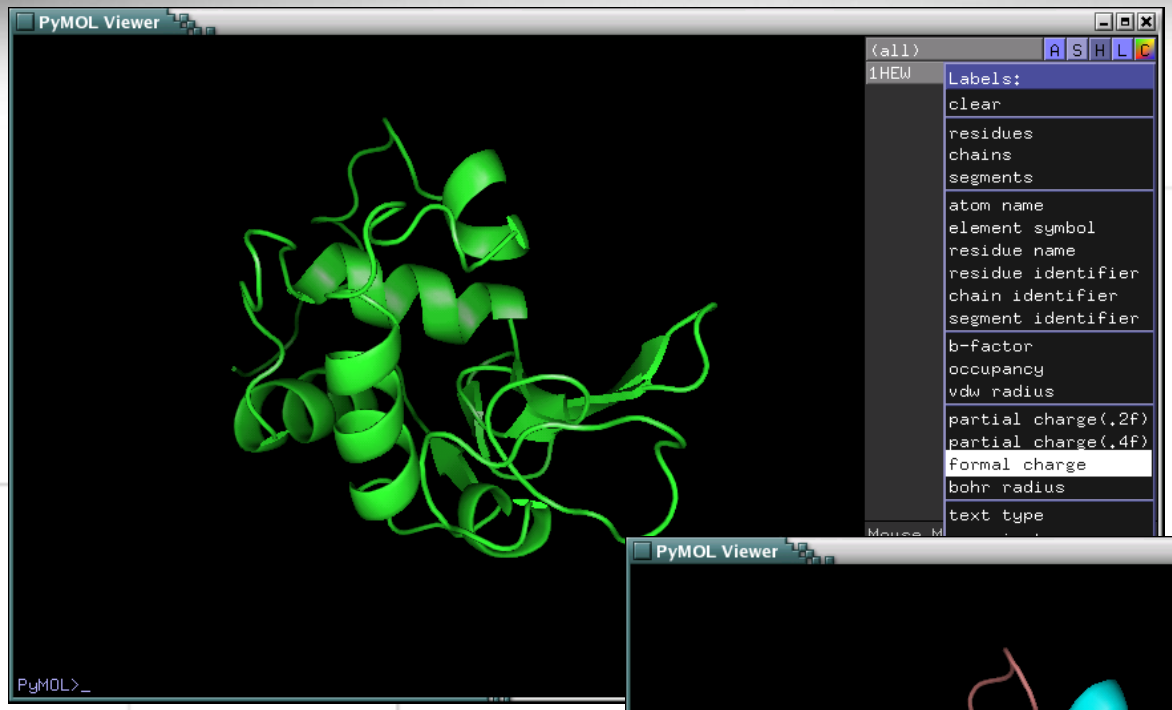
Pocket was identified at the C-terminal sites

Visualization of protein model: PyMOL

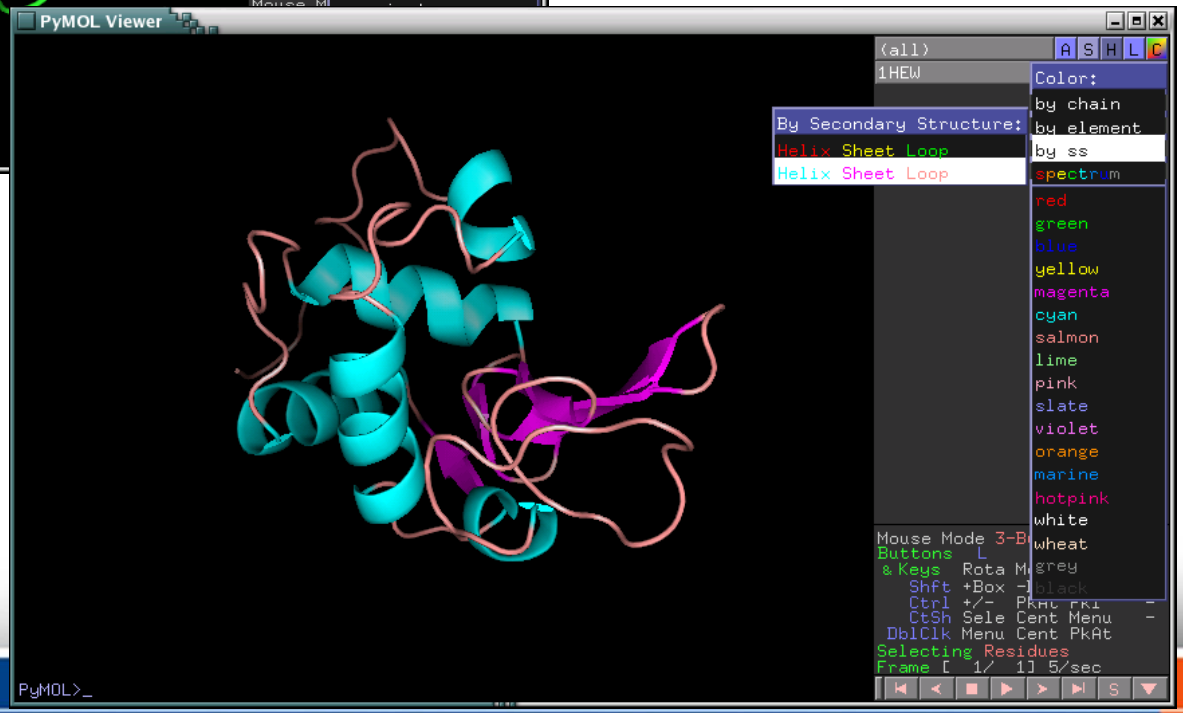
- **PyMOL** is a user-sponsored molecular **visualization system** on an open-source foundation, maintained and distributed by Schrödinger.



Play with PyMOL



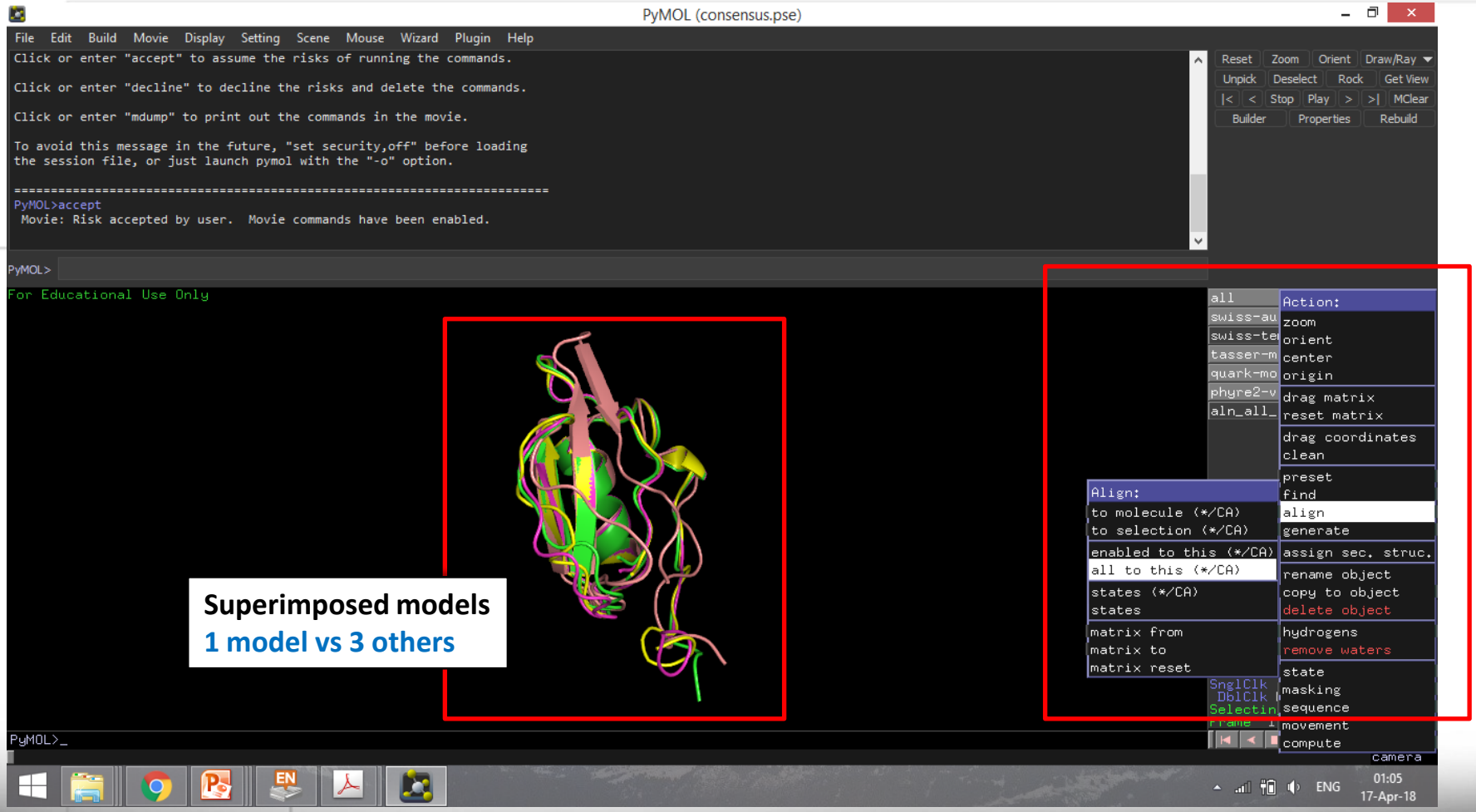
Label panel: residues, chains, segments



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

Super[imposition] with PyMOL

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



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
Novi Jan's Blog