

LAPORAN
WORKSHOP BIOINFORMATIKA
“BIOINFORMATIKA UNTUK PROTEIN MODELLING”

Bogor, 17 – 18 April 2018
Pusat Penelitian Bioteknologi dan Bioindustri Indonesia (PPBBI)

Oleh
SEPRIANTO, S.Pi, M.Si



UNIVERSITAS ESA UNGGUL

2018

PENDAHULUAN

Bioinformatika didefinisikan sebagai cabang keilmuan komputasi dari biologi molekuler dan berbagai keilmuan sains terapan lainnya yang merupakan teknologi pengumpulan, penyimpanan, analisa, interpretasi, penyebaran, dan aplikasi dari informasi biologi. Bioinformatika menggunakan program komputer maupun website untuk analisa data biologi dan penyimpanan sejumlah data biologi yang dihasilkan oleh proyek genom. Bioinformatika banyak berhubungan dengan sekuen nukleotida dan protein termasuk desain primer, struktur, fungsi, perbandingan seluruh genom dan gen, struktur tiga dimensi protein (protein modelling), dan manajemen data. Melalui bioinformatika kita juga dapat melakukan berbagai desain eksperimen untuk mengetahui penyakit manusia, drug discovery dan pembuatan peta genom.

Secara kimia/fisika, bentuk struktur protein diungkap dengan kristalografi sinar- X ataupun spektroskopi NMR, namun kedua metode tersebut sangat memakan waktu dan relatif mahal. Sementara itu, metode sekuensing protein relatif lebih mudah mengungkapkan sekuens asam amino protein. Prediksi struktur protein berusaha meramalkan struktur tiga dimensi protein berdasarkan sekuens asam aminonya (dengan kata lain, meramalkan struktur tersier dan struktur sekunder berdasarkan struktur primer protein). Secara umum, metode prediksi struktur protein yang ada saat ini dapat dikategorikan ke dalam dua kelompok, yaitu metode pemodelan protein komparatif (comparative protein modelling) dan metode pemodelan de novo.

TUJUAN

1. Workshop ini bertujuan untuk memudahkan para peneliti atau dosen dalam membuat permodelan suatu struktur protein
2. Peserta dapat Hands On (Langsung praktek) dalam membuat permodelan struktur protein
3. Memberikan pemahaman tentang cara membuat Anotasi Sekuen dalam sebuah Genom DNA
4. Peserta dapat memanfaatkan software – software yang open akses yang terhubung dengan internet dalam pembuatan protein modelling

PELAKSANAAN

Waktu dan tempat

Workshop dilaksanakan pada:

- Hari dan tanggal : Selasa - Rabu, 17 - 18 April 2018
- Waktu : 08.00 WIB – selesai
- Lokasi : Pusat Penelitian Bioteknologi dan Bioindustri Indonesia,
- Jl. Taman Kencana No.1 Bogor

Peserta Workshop

Dosen pengampu matakuliah Bioinformatika dan Peneliti dalam bidang Bioteknologi

Rincian kegiatan

Kegiatan 1. Pengantar Bioinformatika

Materi ini berkaitan dengan pengenalan tentang ilmu bioinformatika dengan ilmu terkait, serta perkembangan pemanfaatan bioinformatika dalam dunia sains terutama dalam dunia industri, kesehatan dan obat

Narasumber: Dr. Riza Arief Putranto

Kegiatan 2. Pengenalan Database Protein PDB

Materi ini mencakup tentang database protein terutama pada PDB (Protein Data Bank) berisi semua publisitas yang ada secara eksperimen telah dideterminasi (oleh x-ray crystallography dan NMR) sebagai model structural proteins dan asam nukleat.

Narasumber: Dr. Riza Arief Putranto

Kegiatan 3. Pengantar Protein Modelling

Materi ini berkaitan dengan permodelan stuktur protein dengan menggunakan perangkat bioinformatik (software yang terintegrasi internet). Metode ini, struktur suatu protein ditentukan berdasarkan struktur protein lain (protein templat) yang sudah diketahui dan memiliki kemiripan sekuens dengan protein target.

Narasumber : Dr. Rer. Nat Arli Aditya Parikensit

Kegiatan 4. Anotasi Sekuens protein dan Komparasi sekuen protein

Materi ini berkaitan tentang, menentukan letak gen (Lokus) dalam suatu sekuens genom. Setiap kali genom disekuensing maka dibutuhkan anotasi agar dapat digunakan sebagai informasi yang jelas.

Narasumber : irfan Martiansyah, Ssi

Kegiatan1. Pengenalan BIG DATA

Materi terkait dengan pengenalan Database secara keseluruhan termasuk data yang digunakan dalam analisis bioinformatika. kumpulan dari set data yang besar yang tidak dapat diolah dengan menggunakan teknik komputasi tradisional. Data besar bukan hanya data, melainkan telah menjadi subjek yang lengkap, melibatkan berbagai alat, teknik dan kerangka kerja.

Narasumber : Ir. Beno Kunto Pradekso

Rundown acara

Selasa, 17 April 2018

Waktu

08.00 – 08.30

08.30 – 08.40

08.40 – 09.00

09.00 – 09.30

09.30 – 11.00

11.00 – 12.30

12.30 – 13.30

13.30 – 15.00

15.00 – 16.30

Rabu, 18 April 2018

08.30 – 10.00

10.00 – 10.30

10.30 – 12.00

12.00 – 13.00

15.30 – 16.00

Kegiatan

Registrasi

Pembukaan oleh panitia

Sambutan dan pembukaan

Oleh Direktur PPBBI

Coffee break

Materi 1: Pengantar

Bioinformatika

Materi II. Pengenalan

Database Protein

Isoma

Materi III. Pengantar Protein

Modelling

Anotasi Sekuens Protein dan

Komparasi Sekuen protein

Kegiatan

Materi VI. Pengenalan BIG

DATA

Coffe break

Praktikum : Hands-on Protein

Modelling


Praktikum : Hands-on Protein

Modelling (Lanjutan)

Diskusi & Penutupan

Pendaftaran

Pendaftaran dapat dilakukan melalui formulir online:
<http://iribb.org/pelatihan/bioinformasi/> dengan Biaya pendaftaran sebesar Rp 2. 500.000,-.
Untuk kategori umum dan mahasiswa S3 sedangkan kategori mahasiswa S1 dan S2 Rp. 1.500.000, Peserta akan mendapatkan konsumsi, seminar kit, dan sertifikat.




Panitia pelaksana

- Pusat Penelitian Bioteknologi dan Bioindustri Indonesia.

Output dan Outcome

Output




Peserta diharapkan setelah mengikuti pelatihan ini dapat secara mandiri mempraktekan dalam pembuatan protein modelling serta dapat memberikan pengayaan dalam pengembangan matakuliah Bioinformatika di prodi masing – masing, sedangkan untuk peneliti agar dapat memberikan kemudahan dalam memprediksi struktur 3 dimensi protein.

Outcome



Dapat memberikan practice dan hands on dalam Permodelan protein struktur 3 dimensi protein

Penutup



Demikian Term of Reference ini dibuat sebagai panduan kegiatan. Semoga dapat memberikan informasi bagi pihak-pihak yang terkait.

Ringkasan Materi:

1.1. Pengertian Bioinformatika

Bioinformatika didefinisikan sebagai cabang komputasi dari biologi molekuler yang merupakan teknologi pengumpulan, penyimpanan, analisa, interpretasi, penyebaran, dan aplikasi dari informasi biologi. Bioinformatika menggunakan program komputer maupun website untuk analisa data biologi dan penyimpanan sejumlah data biologiyang dihasilkan oleh proyek genom. Bioinformatika banyak berhubungan dengan sekuen nukleotida termasuk desain primer, struktur, fungsi, perbandingan seluruh genom dan gen, struktur tiga dimensi protein, dan manajemen data. Melalui bioinformatik kita juga dapat melakukan berbagai desain eksperimen untuk mengetahui penyakit manusia dan pembuatan peta genom.

1.2. Perkembangan Bioinformatika Dalam Dunia Sains

Perkembangan Penetrasi Teknologi Informasi (TI) dalam berbagai disiplin ilmu telah melipatgandakan perkembangan ilmu bersangkutan. Berbagai kajian baru bermunculan, sejalan dengan perkembangan TI itu sendiri dan disiplin ilmu yang didukungnya. Aplikasi TI dalam bidang biologi molekuler telah melahirkan bidang Bioinformatika. Kajian ini semakin penting, sebab perkembangannya telah mendorong kemajuan bioteknologi di satu sisi, dan pada sisi lain memberi efek domino pada bidang kedokteran, farmasi, lingkungan dan lainnya.

Kajian baru Bioinformatika ini tak lepas dari perkembangan biologi molekuler modern yang ditandai dengan kemampuan manusia untuk memahami genom, yaitu cetak biru informasi genetik yang menentukan sifat setiap makhluk hidup yang disandi dalam bentuk pita molekul DNA (asam deoksiribonukleat). Kemampuan untuk memahami dan memanipulasi kode genetik DNA ini sangat didukung oleh TI melalui perangkat perangkat keras maupun lunak. Hal ini bisa dilihat pada upaya Celera Genomics, perusahaan bioteknologi Amerika Serikat yang melakukan pembacaan sekuen genom manusia yang secara maksimal memanfaatkan TI sehingga bisa melakukan pekerjaannya dalam waktu yang singkat (hanya beberapa tahun), dibanding usaha konsorsium lembaga riset publik AS, Eropa, dan lain-lain, yang memakan waktu lebih dari 10 tahun.

Kelahiran Bioinformatika modern tak lepas dari perkembangan bioteknologi di era tahun 70-an, dimana seorang ilmuwan AS melakukan inovasi dalam mengembangkan teknologi DNA rekombinan. Berkat penemuan ini lahirlah perusahaan bioteknologi pertama di dunia, yaitu Genentech di AS, yang kemudian memproduksi protein hormon insulin dalam

bakteri, yang dibutuhkan penderita diabetes. Selama ini insulin hanya bisa didapatkan dalam jumlah sangat terbatas dari organ pankreas sapi. Bioteknologi modern ditandai dengan kemampuan pada manipulasi DNA. Rantai/sekuen DNA yang mengkode protein disebut gen. Gen ditranskripsikan menjadi mRNA, kemudian mRNA ditranslasikan menjadi protein. Protein sebagai produk akhir bertugas menunjang seluruh proses kehidupan, antara lain sebagai katalis reaksi biokimia dalam tubuh (disebut enzim), berperan serta dalam sistem pertahanan tubuh melawan virus, parasit dan lain-lain (disebut antibodi), menyusun struktur tubuh dari ujung kaki (otot terbentuk dari protein actin, myosin, dan sebagainya) sampai ujung rambut (rambut tersusun dari protein keratin), dan lain-lain. Arus informasi, DNA -> RNA -> Protein, inilah yang disebut sentral dogma dalam biologi molekuler

1.3 Pengenalan DATABASE

NCBI merupakan server yang memuat data base tentang informasi kesehatan dan bioteknologi. Data base terus menerus di update sesuai dengan penemuan-penemuan terkini yang menyangkut DNA, Protein, Senyawa aktif dan taksonomi. Disamping data base, ncbi juga menyediakan

berbagai macam software untuk analisis DNA, protein 3D, pencarian primer, pencarian conserve domain dan lain sebagainya. NCBI merupakan salah satu bank data gen, protein dan literature khususnya dibidang kesehatan yang terlengkap dan diacu oleh para peneliti di dunia. NCBI (<http://www.ncbi.nlm.nih.gov>) memiliki database dan software (analysis tools) yang sering digunakan untuk analisis adalah sebagai berikut:

DNA-RNA TOOLS:

GenBank

Bank Gen ini bagian dari database nukleotida internasional yang bekerja sama dengan DataBank of Japan (DDBJ), the European Molecular Biology Laboratory (EMBL), and GenBank at NCBI.

BioSystems

Database yang berisi tentang korelasi biologi dari gen, protein dan small molecule berdasarkan literature. Kita dapat menggunakan untuk menganalisis fungsi protein dan interaksi protein yang kita teliti pada level sel. System ini terhubung dengan beberapa database biosystem seperti reactome dan KEGG. Misalnya kita ingin mengetahui keberadaan/fungsi protein p53 di dalam sel, maka tinggal memasukkan p53 kedalam keyword

box, maka akan diperoleh hasil seperti dibawah yang menunjukkan p53 berfungsi sebagai DNA-damage response yang terdapat di database reactome

Database of Expressed Sequence Tags (dbEST)

- GenBank yang berisi short single-pass reads of cDNA (transcript) sequences/EST.

Database of Genome Survey Sequences (dbGSS)

- GenBank yang berisi short single-pass reads of genomic DNA/GSS.

BLAST (*Basic Local Alignment Search Tool*)

Software yang dapat digunakan untuk menentukan homologi suatu urutan DNA atau asam amino dengan data yang ada di NCBI. BLAST memiliki beberapa pilihan menu sesuai dengan analisis yang akan dikerjakan seperti pada table di bawah ini

Program	Program Description
nucleotide blast	Search a nucleotide database using a nucleotide query <i>Algorithms:</i> blastn, megablast, discontinuous megablast
protein blast	Search protein database using a protein query <i>Algorithms:</i> blastp, psi-blast, phi-blast
blastx	Search protein database using a translated nucleotide query
tblastn	Search translated nucleotide database using a protein query
tblastx	Search translated nucleotide database using a translated nucleotide query

Allignment analysis

Sekuen yang diperoleh dari hasil penelitian di laboratorium dapat dianalisis dengan data serupa yang telah dipublikasikan sebelumnya di gen bank. Salah satu bentuk analisis yang dapat dilakukan misalnya adalah analisis penyejajaran. Analisis penyejajaran dapat digunakan untuk membandingkan dua sekuen atau lebih. Program yang digunakan untuk analisis penyejajaran yaitu program BLAST (Basic Local Allignment Search Tools). Program ini dapat diakses melalui website National Center for Biotechnology Information at The National

Library of Medicine in Washington, DC (<http://www.ncbi.nlm.nih.gov/BLAST>)

Pangkalan data primer untuk sekuens asam nukleat saat ini adalah

- [GenBank](#) (Amerika Serikat),
- [EMBL](#) (the European Molecular Biology Laboratory, Eropa), dan
- [DDBJ](#) (DNA Data Bank of Japan, [Jepang](#)).

Ketiga pangkalan data tersebut bekerja sama dan bertukar data secara harian untuk menjaga keluasan cakupan masing-masing pangkalan data.

Sumber utama data sekuens asam nukleat adalah submisi (pengumpulan) langsung dari

- peneliti individual,
- proyek sekuensing [genom](#), dan
- pendaftaran [paten](#).
- Selain berisi sekuens asam nukleat, entri dalam pangkalan data sekuens asam nukleat pada umumnya mengandung informasi tentang
 - jenis asam nukleat ([DNA](#) atau [RNA](#)),
 - nama [organisme](#) sumber asam nukleat tersebut, dan segala sesuatu yang berkaitan dengan sekuens asam nukleat tersebut

1. OMIM, (*Online Mendelian Inheritance in Man—woman*), adalah insiklopedia gen-gen manusia dan penyakit genetik, merupakan penghubung untuk *entry* gen pada GenBank dan literatur ilmiah pada PubMed, berisi informasi berbagai gen manusia komplis dan paling baru.

2. PDB (*Protein Data Bank*) berisi semua publisitas yang ada secara eksperimen telah dideterminasi (oleh *x-ray crystallography* dan *NMR*) sebagai model structural proteins dan asam nukleat. Tidak berisi model homologi atau tipe model teoritis lainnya. **PDB** (*Protein Data Bank*, Bank Data Protein) ialah pangkalan data tunggal yang menyimpan model struktur tiga dimensi [protein](#) dan [asam nukleat](#) hasil penentuan eksperimental (dengan [kristalografi sinar-X](#), [spektroskopi NMR](#), dan [mikroskopi elektron](#)). PDB menyimpan data struktur sebagai [koordinat tiga dimensi](#) yang menggambarkan posisi [atom](#)-atom dalam protein atau pun asam nukleat.

3. PubMed adalah diskripsi pada Wikipedia sebagai “suatu kebebasan mengakses sititasi database MEDLINE dan abstrak artikel riset biomedik.

- Subjek utama adalah riset di bidang kedokteran, dan PubMed juga mempublikasi bidang yang terkait dengan bidang kedokteran, seperti kebidanan dan disipiin kesehatan lainnya.
- Hal ini secara menyeluruh meliputi keilmuan yang berhubungan dengan ilmu seperti biokemia dan biologi sel.

Situs ini ditawarkan oleh *the United States National Library of Medicine* di *the National Institutes of Health* sebagai bagian dari *the Entrez information retrieval system*.”

4. UniProt Knowledgebase (Swiss-Prot and TrEMBL),

dioperasikan oleh **SIB** (*Swiss Institute of Bioinformatics*) dan **EBI** (*European Bioinformatics Institute*), berisi sebagian besar publikasi yang ada berupa **sekuens protein** (bukan DNA atau RNA).

- Sekuens dalam *Swiss-Prot* dijelaskan secara manual dan menyediakan atau menghubungkan pengguna dengan semua informasi publisitas yang berisi sekuens tersebut.
- Sequences pada TrEMBL dikoleksi dan dijabarkan secara otomatis dari sekuens database, dan akan membuat jalannya menuju Swiss-Prot, tetapi tidak hanya setelah mereka secara manual menjabarkan *Swiss-Prot standards*.

Beberapa situs informasi data base DNA, RNA dan protein

- NCBI: www.ncbi.nlm.nih.gov
- EMBL: www.ebi.ac.uk
- DDBJ: www.ddbj.nig.ac.jp
- SWISS-PROT: www.expasy.ch/sprot/sprot_details.html
- ENSEMBL: www.ensembl.org
- Univeristy California Santa Cruz: genome.cse.ucsc.edu MGD the Jackson Lab: www.informatics.jax.org

Protein Structure Modelling

- 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 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?

Homology Protein

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**

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**.

[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

✓ SMTL **aggregates information** of experimental structures from the **Protein Data Bank** and augments it with derived information. When a new structure is released by the PDB, the coordinates and accompanying information are processed and imported into the template library.

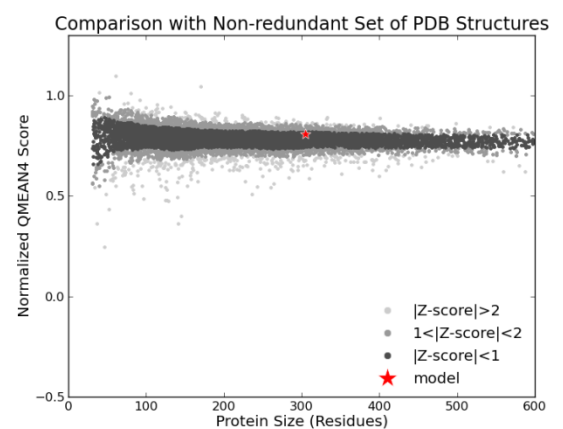
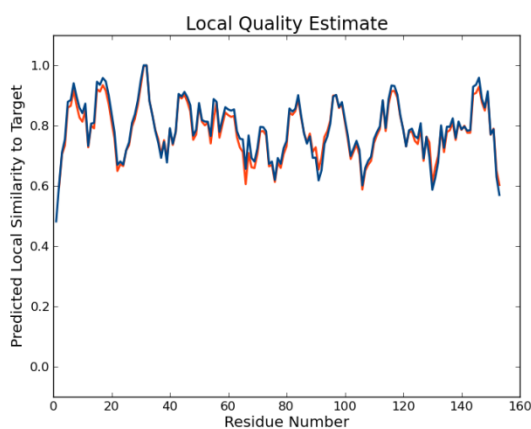
✓ The current SMTL contains:

✓ **548.438 chains**

✓ **94.303 unique SEQRES sequences** (primary sequence of the polymeric molecules present) 223.434 biounits

Model results & evaluation

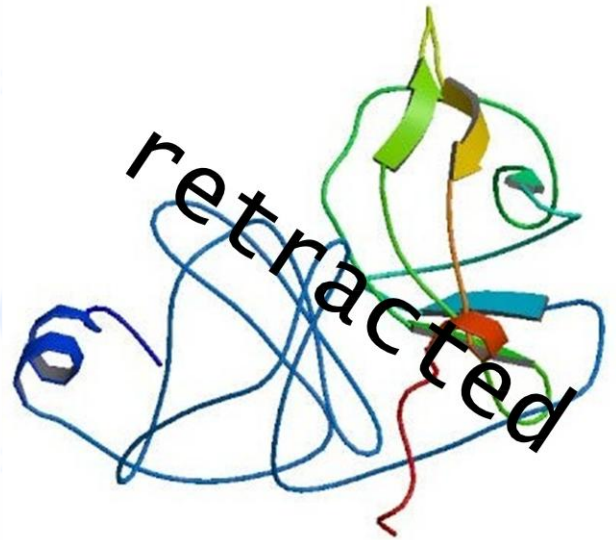
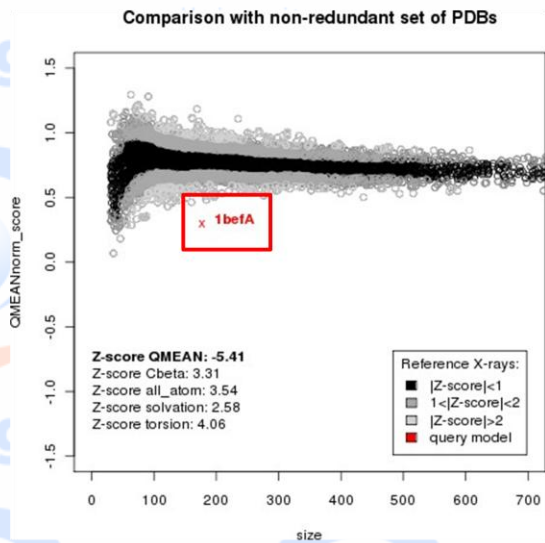
- **GMQE (Global Model Quality Estimation)** combines properties from the target–template alignment and the template search method. GMQE score is expressed as a **number between 0 and 1**, reflecting the **expected accuracy of a model built** with that alignment and template and the coverage of the target.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



Sometimes in modelling, you can **fail**. It is real!

If your model is **so far away from normal distribution of references**.

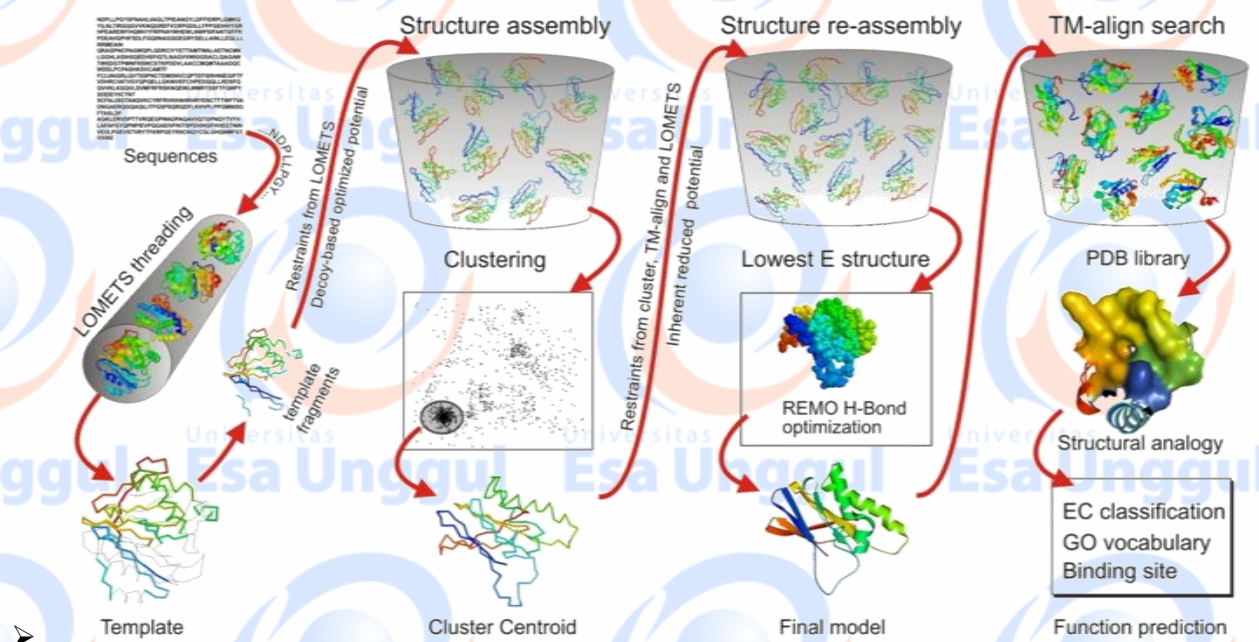
REDO your modelling, compare it with other methods



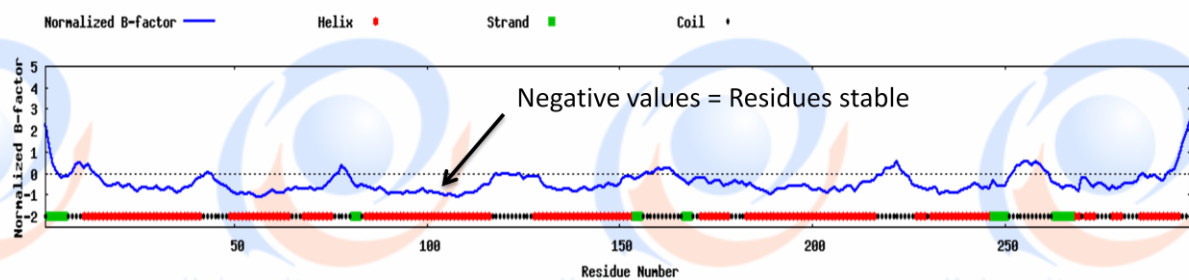
The case of Dengue Virus NS3 Serine Protease

I-TASSER

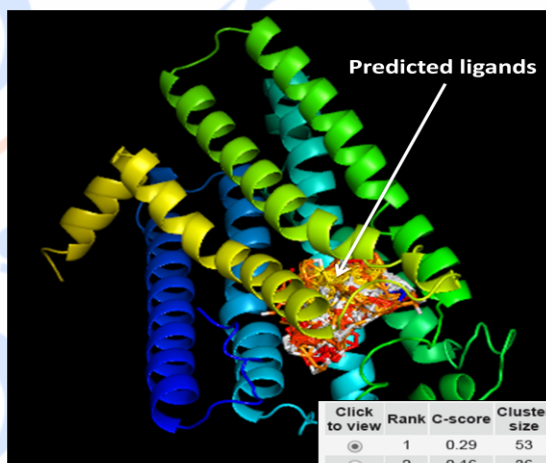
- **I-TASSER (Iterative Threading ASSEMBLY Refinement)**, a hierarchical approach to protein structure and function prediction.
- It forms structural templates from the PDB by multiple threading approach LOMETS, with full-length atomic models constructed by iterative template fragment assembly simulations.



- **C-score** is typically **in the range of [-5,2]**, where a C-score of higher value signifies a model with a high confidence and vice-versa.
- **TM-score >0.5** indicates a model of correct topology and **TM-score <0.17** means a random similarity.
- B-factor, a value to indicate the extent of the inherent thermal mobility of residues/atoms in proteins
- The normalized B-factor (**B-factor profile, BFP**), predicted using a combination of both **template-based assignment** and **profile-based prediction**.



- C-score is the confidence score of the prediction. C-score ranges [0-1], where a higher score indicates a more reliable prediction.
- Cluster size is the total number of templates in a cluster.
- Lig Name is name of possible binding ligand. Click the name to view its information in the BioLiP database.
- Rep is a single complex structure with the most representative ligand in the cluster, i.e., the one listed in the Lig Name column.
- Mult is the complex structures with all potential binding ligands in the cluster.



Click to view	Rank	C-score	Cluster size	PDB Hit	Lig Name	Download Complex	Ligand Binding Site Residues
<input checked="" type="radio"/>	1	0.29	53	4amiA	G90	Rep. Mult	90,91,94,95,98,180,181,182,192,196,242,243
<input type="radio"/>	2	0.16	36	2y04A	Y01	Rep. Mult	51,55,58,96,128,132,135,139
<input type="radio"/>	3	0.08	24	4ea3A	QNN	Rep. Mult	71,91,94,95,98,99,192,196,242,246,249,264
<input type="radio"/>	4	0.05	12	3dqba	PEPTIDE	Rep. Mult	48,112,115,116,118,211,215,218,221,284,285
<input type="radio"/>	5	0.04	18	2ycwA	2CY	Rep. Mult	208,211,212,225,228,229,235

- CscoreGO is a combined measure for evaluating global and local similarity between query and template protein. It's range is [0-1] and higher values indicate more confident predictions.
- TM-score is a measure of global structural similarity between query and template protein.

Consensus prediction of GO terms

Molecular Function	GO:0003796	GO:0004940	GO:0004941	GO:0004995	
GO-Score	0.58	0.48	0.35	0.33	
Biological Process	GO:0071875	GO:0019835	GO:0009253	GO:0042742	GO:0016998
GO-Score	0.66	0.58	0.58	0.58	0.58

GO:0003796   [JSON](#)

lysozyme activity

Molecular Function

Definition ([GO:0003796 GONUTS page](#))

Catalysis of the hydrolysis of the beta-(1->4) linkages between N-acetylmuramic acid and N-acetyl-D-glucosamine residues in a peptidoglycan. PMID:22748813

GO:0071875   [JSON](#)

adrenergic receptor signaling pathway

Biological Process

Definition ([GO:0071875 GONUTS page](#))

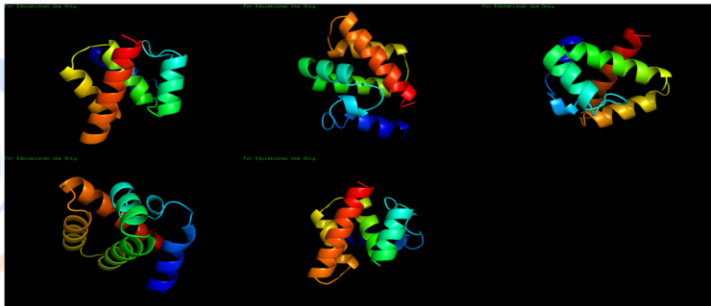
A series of molecular signals generated as a consequence of an adrenergic receptor binding to one of its physiological ligands.

QUARK

- **QUARK**, 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, 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.
- QUARK was ranked as the No 1 server in Free-modeling (FM) in [CASP9](#) and [CASP10](#) experiments.

QUARK results

A. Prediction of five ab initio proteins from query



Confidential score 1-9, close to 10 perfect

B. Predicted secondary structure

	20	40	60	80
Sequence	TTSQKHRDFVAEPGEKPVGSLAGIGEVLGKLEERGFDKAYVVLGQFLVLKKDEDLFREWLKDTCGANAKQSRDCFGCLRWCDAFL			
Prediction	CCCHHHHHHCCCCCCCCCCCCCHHHHHHHHHCCCHHHHHHHHHHHHHCCCHHHHHHHHHHHHHCCCCCHHHHHHHHHHHHHHHHC			
Conf. Score	988899999879999987447898899999999979659999999999588899999999996889999999999999859			
	H:Helix; S:Strand; C:Coil			

C. Predicted solvent accessibility

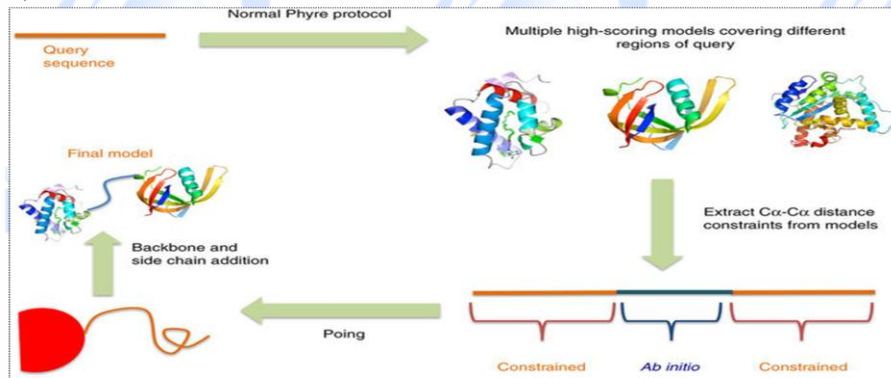
	20	40	60	80
Sequence	TTSQKHRDFVAEPGEKPVGSLAGIGEVLGKLEERGFDKAYVVLGQFLVLKKDEDLFREWLKDTCGANAKQSRDCFGCLRWCDAFL			
Prediction	553330221123223321120110032002102421132002000200113232310220022102031310310010022003324			
	Values range from 0 (buried residue) to 9 (highly exposed residue)			

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.

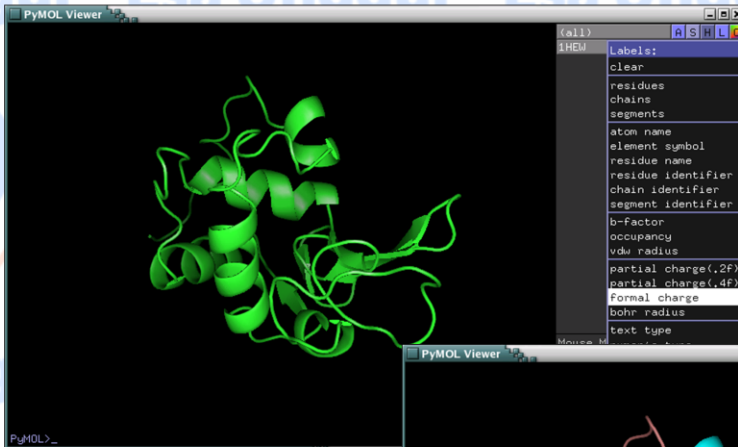
Phyre2

- **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.



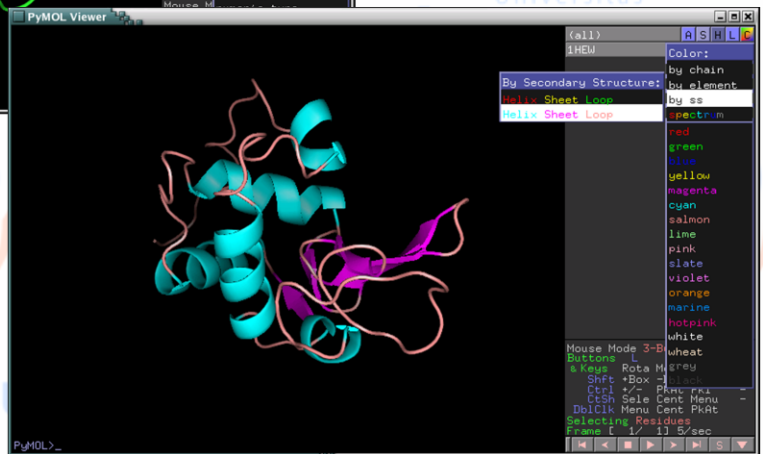
Eliminate **constraints** from several models and create final model using **Poing** (protein-folding simulator)

Play with PyMOL



Label panel: residues, chains, segments

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



TAHAPAN SWISS MODEL (sequence)>> automated modelling

Model	Template	Seq Identity	Coverage	Description
Model 01	1mit.1.A	47.83%		TRYPSIN INHIBITOR V
Model 02	1dwm.1.A	50.72%		LINUM USITATISSIMUM TRYPSIN INHIBITOR

Template ada di bagian putih

Kalau ada yg kosong di kanan / kiri itu menentukan urutan basa yang dicoverage

Quality :

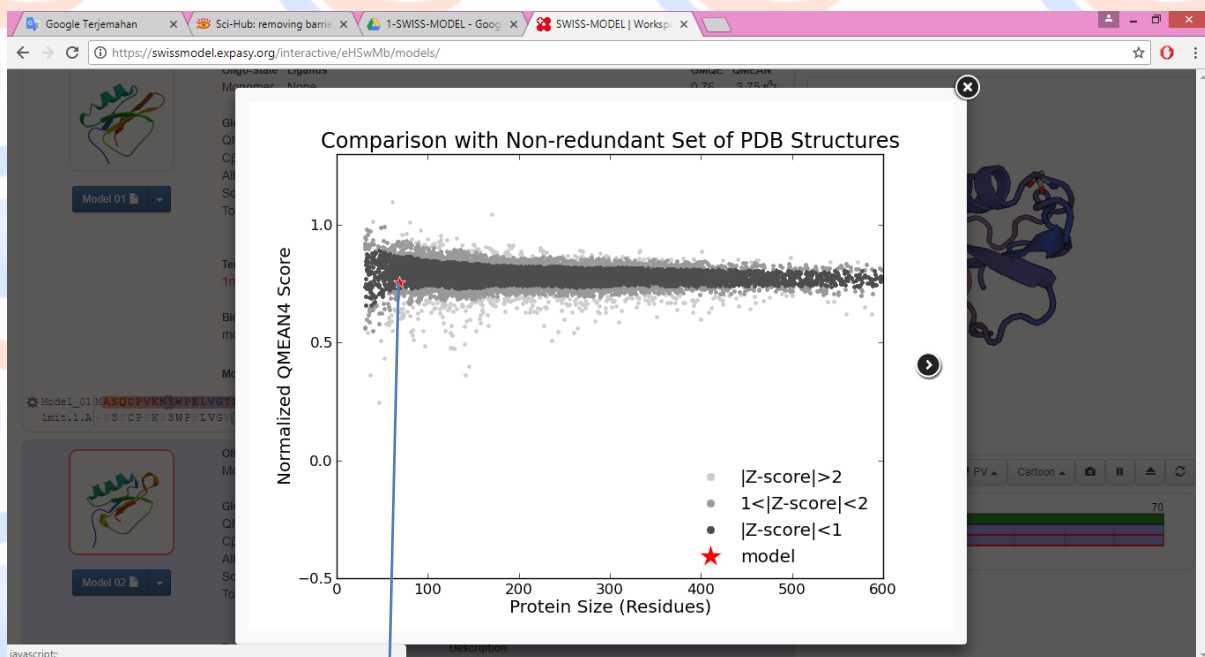
GMQE > mendekati 1 artinya mendekati modelnya

QMEAN > nilai harus diatas -4, lebih baik jika mendekati 0



residu

Batas standar

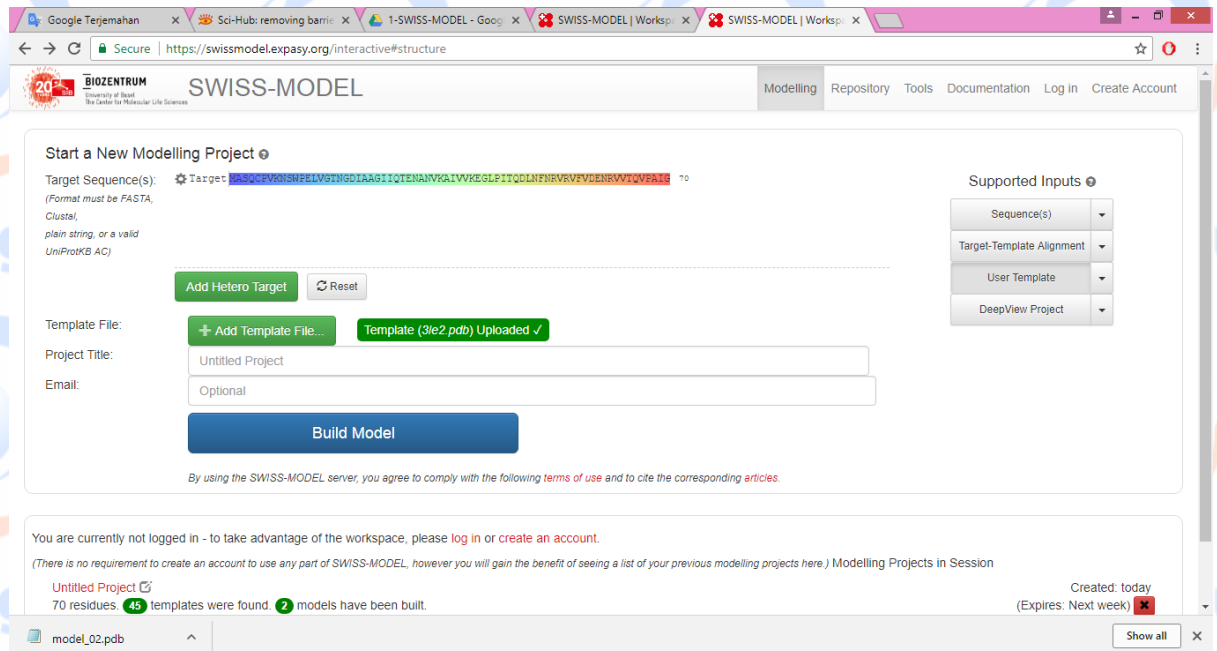


Model ada disekitar sebaran

Model yg pertama diluar sebaran, jadi tidak diambil

TAHAPAN SWISS MODEL (user template)>> bersifat manual, berfungsi pendekatan untuk memastikan modelling kita

- Masukin sequence
- Add template file.pdb
- Build model



Coverage nya lebih dikit

Ganti background alfahelix > didisplay phymol

TAHAPAN I-TASSER

Masuk ke <https://zhanglab.cmb.med.umich.edu/I-TASSER/>

Copy sequence/choose file

Masukin email berinstitusi (tidak boleh yang gratisan email)

Password send by I-TASSER

Namain

Run (keep my result public check or not)

Modelnya di sent ke email

Bisa mempredikasikan apa saja yang mampu menempel di sisi aktif

Melakukan 3x pengulangan untuk membuat model

Manfaat i-tasser :

b-factor

c score >2

ligan

enzyme

ontology funcion



Mengcompare struktur sekunder

Prediction 8764514763315503640541274147303010332675614515343211347232330231044317146631010102432246433034015634442100000132412233648
 Values range from 0 (buried residue) to 9 (highly exposed residue)

Predicted normalized B-factor

(B-factor is a value to indicate the extent of the inherent thermal mobility of residues atoms in proteins. In I-TASSER, this value is deduced from threading template proteins from the PDB in combination with the sequence profiles derived from sequence databases. The reported B-factor profile in the figure below corresponds to the normalized B-factor of the target protein, defined by $B = (B-u)^2/s^2$, where B' is the raw B-factor value, u and s are respectively the mean and standard deviation of the raw B-factors along the sequence. [Click here to read more about predicted normalized B-factor](#))

Top 10 threading templates used by I-TASSER

(I-TASSER modeling starts from the structure templates identified by LOMETS from the PDB library. LOMETS is a meta-server threading approach containing multiple threading programs, where each threading program can generate tens of thousands of template alignments. I-TASSER only uses the templates of the highest significance in the threading alignments, the significance of which are measured by the Z-score, i.e. the difference between the raw and average scores in the unit of standard deviation. The templates in this section are the 10 best templates selected from the LOMETS threading programs. Usually, one template of the highest Z-score is selected from each threading program, where the threading programs are sorted by the average performance in the large-scale benchmark test experiments.)

Rank	PDB Hit	I den1	I den2	Cov	Norm. Z-score	Download Align.
1	1ac5B	0.60	0.57	0.96	3.47	Download
2	3h94A	0.53	0.51	0.95	2.81	Download

Model secara statistika baik ketika nilai B-factor dibawah nol

Reset to initial orientation Spin On/Off

- [Download Model 1](#)
- C-score=1.03 ([Read more about C-score](#))
- Estimated TM-score = 0.85±0.08
- Estimated RMSD = 1.4±1.3Å

Reset to initial orientation Spin On/Off

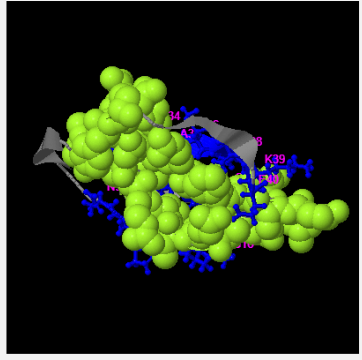
- [Download Model 2](#)
- C-score = -5

C-score : -5 sd 2 (high confidence, mendekati 2 semakin baik)

Secure | https://zhanglab.cmb.med.umich.edu/I-TASSER/output/S392253/

(This section reports biological annotations of the target protein by COFACTOR and COACH based on the I-TASSER structure prediction. While COFACTOR deduces protein functions (ligand-binding sites, EC and GO) using structure comparison and protein-protein networks, COACH is a meta-server approach that combines multiple function annotation results (on ligand-binding sites) from the COFACTOR, TM-SITE and S-SITE programs.)

Ligand binding sites



Click to view	Rank	C-score	Cluster size	PDB Hit	Lig Name	Download Complex	Ligand Binding Site Residues
<input type="radio"/>	1	0.14	3	1c9A	PEPTIDE	Rep , Multi	9,10,11,12,14,15,16,17,18,19,22,26,30,33,34,35,36,37,38,39,40
<input type="radio"/>	2	0.13	2	1Y3DB	1Y3DB00	Rep , Multi	36,44,45,52,54
<input type="radio"/>	3	0.12	3	3ze7B	FCO	Rep , Multi	14,19,24,25,26,67
<input type="radio"/>	4	0.08	2	1e3dB	F3S	Rep , Multi	34
<input type="radio"/>	5	0.08	1	1VBWA	1VBWA00	Rep , Multi	2,49,50

[Download](#) the residue-specific ligand binding probability, which is estimated by SVM.
[Download](#) the all possible binding ligands and detailed prediction summary.
[Download](#) the templates clustering results.

(a) C-score is the confidence score of the prediction. C-score ranges [0-1], where a higher score indicates a more reliable prediction.
 (b) Cluster size is the total number of templates in a cluster.
 (c) Lig Name is name of possible binding ligand. Click the name to view its information in [the BioLiP database](#).
 (d) Rep is a single complex structure with the most representative ligand in the cluster, i.e., the one listed in the Lig Name column. Multi is the complex structures with all potential binding ligands in the cluster.

Spin On/Off

Enzyme Commission (EC) numbers and active

[model1_3rdzC \(1\).pdb](#) [model1_3rdzC.pdb](#) [model1.pdb](#)

Download multi untuk melihat ligan yang bisa menempel

TAHAPAN PHYRE2

Gungsi : Mutasi, pocket dan ramachandran plot

- Masuk kesini

Not secure | www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index

Nominate your protein to be experimentally solved! New CASP13 experiment upcoming. [MORE INFO](#)

[FBI 2017 Workshop](#) | [Older Workshops](#) | [Phyre2 paper](#)

E-mail Address:

Optional Job description:

Amino Acid Sequence:

Or try the [sequence finder](#)

Modelling Mode: Normal Intensive

2492451 submissions since Feb 14 2011

Phyre is for non-commercial use only
Commercial users please contact [Michael Sternberg](#)

Nominate your protein to be experimentally solved! New CASP13 experiment upcoming. [MORE INFO](#)

[EBI 2017 Workshop](#) | [Older Workshops](#) | [Phyre2 paper](#)

E-mail Address:

Optional Job description:

Amino Acid Sequence:

```
>vibranium
NASQCPKNSHPFLVGTNGDIAAGIIQTENANNKAIVVKEGLPITQDLNFRVRV
FVDENRVVTQVPAIG
```

Or try the [sequence finder](#)

Modelling Mode: Normal Intensive

2492451 submissions since Feb 14 2011

Phyre is for **non-commercial use only**

Commercial users please contact [Michael Sternberg](#)

Please cite: The Phyre2 web portal for protein modeling, prediction and analysis
Kelley LA et al. *Nature Protocols* 10, 845-858 (2015) [[paper](#)] [[Citation link](#)]

Result

Untuk ab initio

Summary

Top model

Model (left) based on template [c1tnA](#)

Top template information

PDB header: serine protease inhibitor
Chain: A: **PDB Molecule:** trypsin inhibitor v;
PDBTitle: three-dimensional structure in solution of cucurbita maxima trypsin2 inhibitor-v determined by nmr spectroscopy

Confidence and coverage

Confidence: **100.0%** Coverage: **97%**

68 residues (97% of your sequence) have been modelled with 100.0% confidence by the single highest scoring template.

3D viewing

[Interactive 3D view in JSmol](#)

For other options to view your downloaded structure offline see the [FAQ](#)

Image coloured by rainbow N → C terminus

Model dimensions (Å): X:39.265 Y:28.128 Z:23.340

- mirip dengan swiss model
- tersimpan hanya 30 hari, jadi segeralah screenshot
- jangan lupa save pdf dan link http://www.sbg.bio.ic.ac.uk/phyre2/phyre2_output/47b7fc950cec76dc/summary.html
- perhatikan : confidence, identity
- run investigator the best protein model

Phyre2
Investigator

Rank 1 model (template c1tinA_)
Job Description: Vibranium_

Beta Testing (What is this?)
Email [Lawrence Kelley](#) with problems or suggestions

Phyre Investigator allows you to interactively examine many features of your protein model including:

- Model quality predictions (clashes, rotamers etc)
- Conservation analysis
- Pocket detection and interface prediction
- Predicted effects of mutations using [SuSPect](#)

Processing typically takes 5-10 minutes

Warning: You should be using an up-to-date HTML5 compliant browser to view Phyre Investigator results

Analyses :

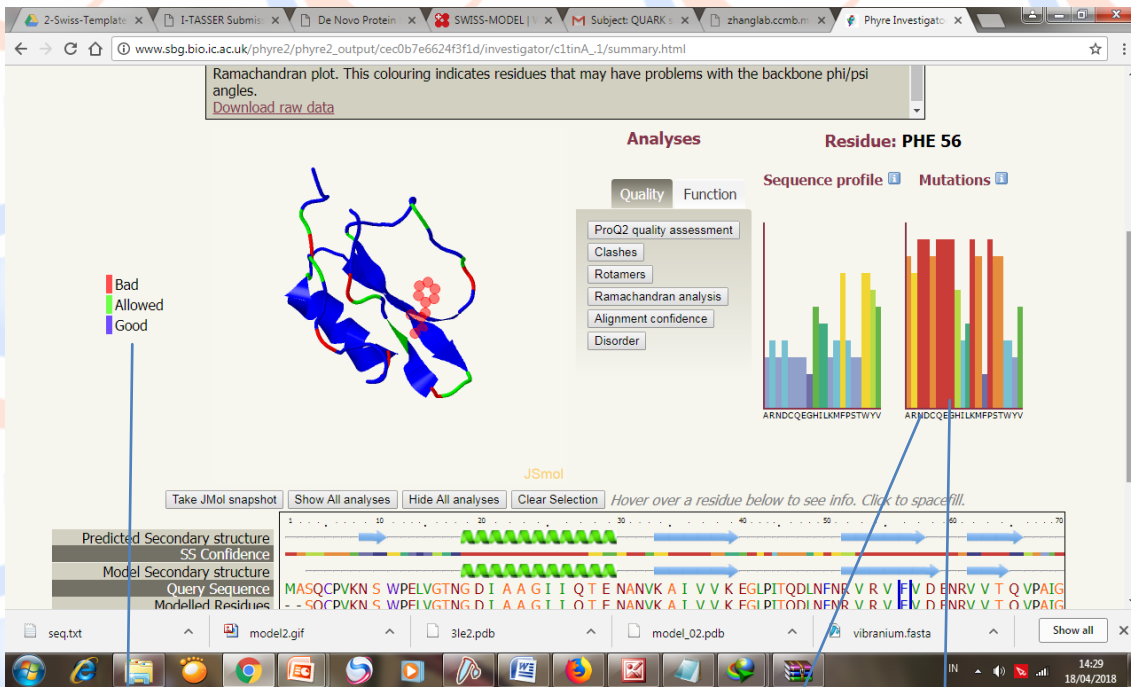
Klik : untuk melihat data, mutasi, pocket (sisi aktif)

Rampage > usaha untuk mengatur suatu asam amino akan memancarkan sudut kimia dan fisika nya

Jsmol > warna gelap, tingkat mutasi rendah

Jika warna terang maka mutasi tinggi

Jadi tidak boleh mutasi di daerah pocket karena daerah pocket merupakan daerah sisi aktif yang mempengaruhi kerja enzim



Interprestasi data

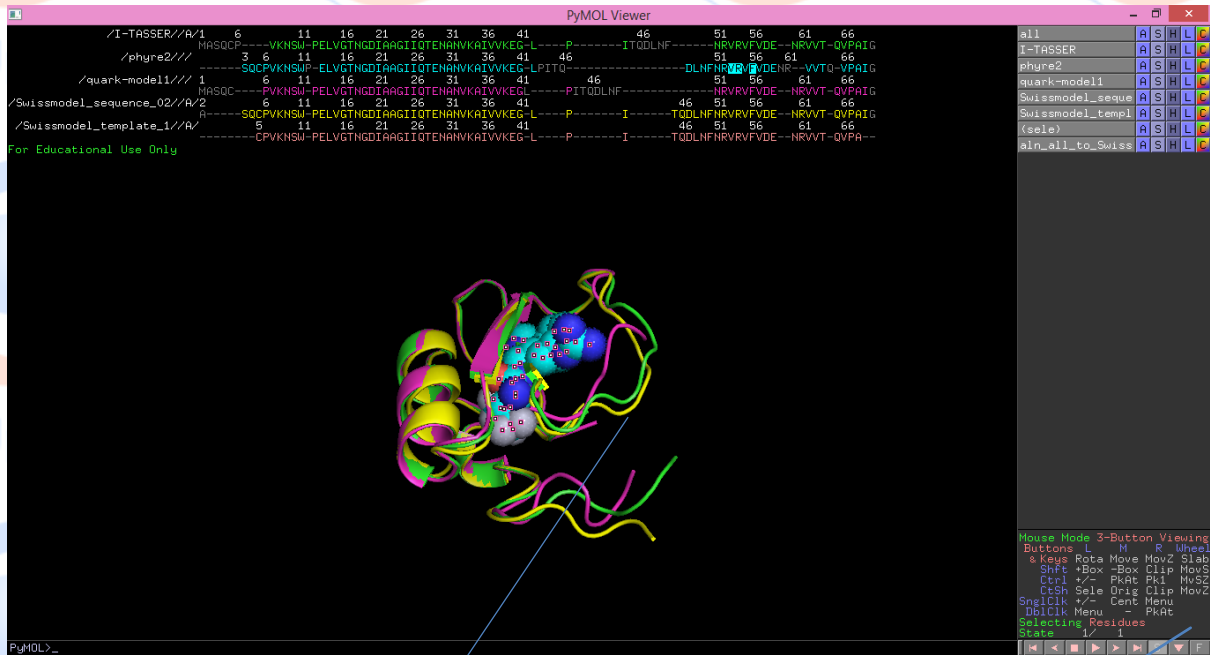
Semakin berwarna merah, semakin tinggi peluang untuk menggantikan posisi asam amino valine

Bagian dibawah sekuens berwarna adalah protein yang akan menggantikan dia

Cara pakai phymol :

- klik insert data
- klik all
- klik salah satu teru di align> all to this
- sele : untuk menentukan sisi aktif

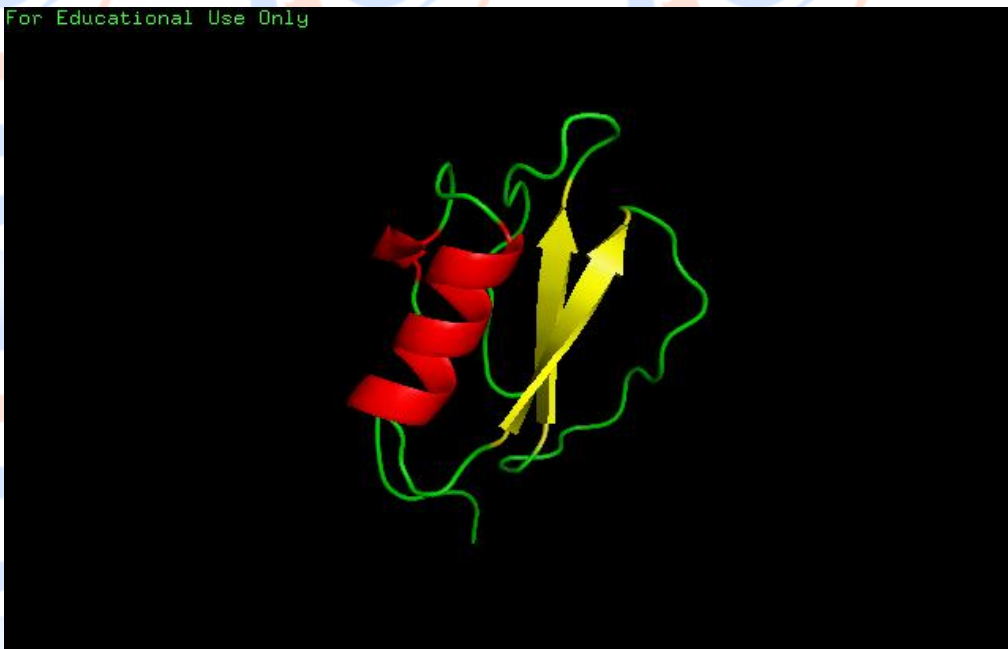
show :



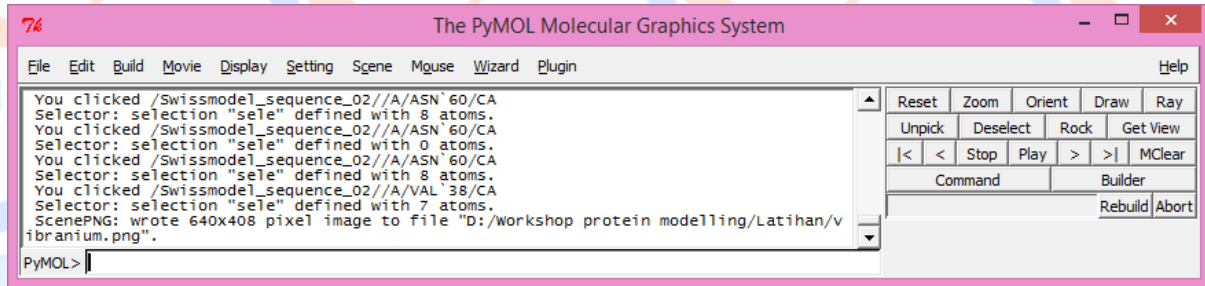
Sisi aktif :
Klik urutan asam amino 53,54,56
Klik sele > show > spheres

Memunculkan sekuens

Final result by swiss model



How to play movie :



Window pymol 1

Klik movie > program > camera loop> nutate> terserah

Dokumentasi Acara



Diselenggarakan oleh :



Pusat Penelitian Bioteknologi dan Bioindustri Indonesia
PT. Riset Perkebunan Nusantara



**PELATIHAN
BIOINFORMATIKA UNTUK PROTEIN MODELLING
BOGOR, 17 - 18 APRIL 2018**

Universitas
Esa U

Universitas
Esa U

Universitas
Esa U

Universitas
Esa U

Universitas
Esa U

Universitas
Esa Unggul

Universitas
Esa Unggul

Universitas
Esa Unggul

Universitas
Esa Unggul

gul

gul

gul

gul

gul