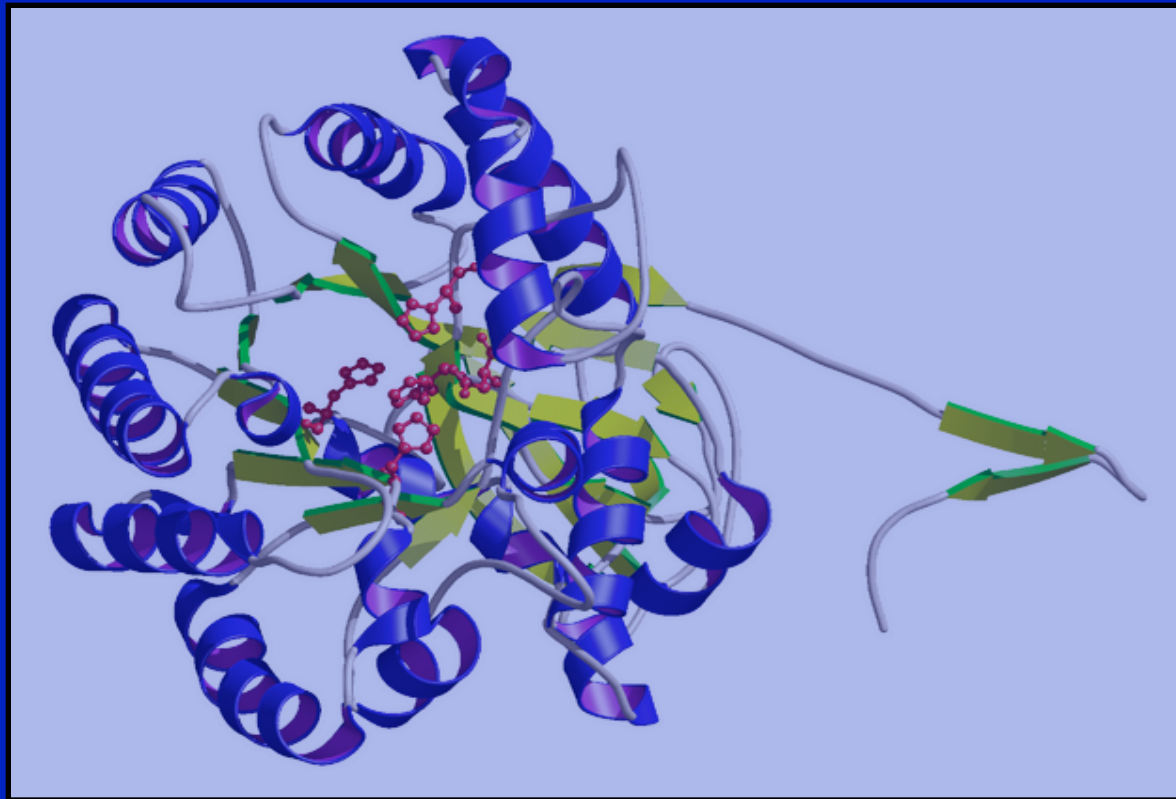


HOMOLOGY MODELLING



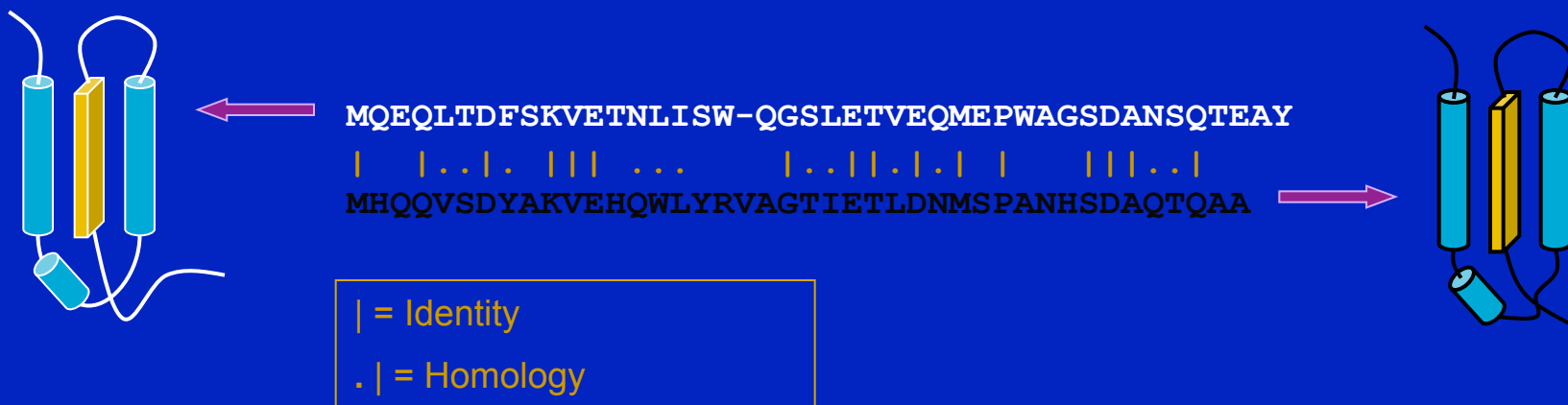
Chris Wilton

Homology Modelling

- What is it and why do we need it?
 - principles of modelling, applications available
- Using Swiss-Model
 - preparation, workspace tools, automated-mode
 - refining with manually-adjusted alignments
- Model Assessment
 - using SwissModel tools to check model quality
 - what to look for, what we expect...

What is Homology Modelling?

- Given an unknown protein, make an informed guess on its 3D structure based on its sequence:
 - Search structure databases for *homologous* sequences
 - Transfer coordinates of known protein onto unknown



Why Modelling is Necessary

- Current structure determination methods:
 - NMR – conc. protein solution; limited size + resolution
 - XRay Crystallography – protein crystals, high resolution
 - Expensive, slow, difficult (especially membrane proteins!)
- Can't keep up with growth rate of sequence databases:
 - PDB: **40132** structures (14/11/2006)
 - pairsdb: **4000000+** sequences (11/2006)

Current (Free) Servers & Software

- **SWISSMODEL** (www)
<http://swissmodel.expasy.org/SWISS-MODEL.html>
- **3D-Jigsaw** (www)
<http://www.bmm.icnet.uk/servers/3djigsaw/>
- **ESyPred3D** (www)
<http://www.fundp.ac.be/sciences/biologie/urbm/bioinfo/esypred/>
- **WHATIF** (www)
<http://swift.cmbi.kun.nl/WIWWWI>
- **CPH Models 2.0** (www)
<http://www.cbs.dtu.dk/services/CPHmodels>
- **MODELLER 8v2** (standalone for windows, mac, linux; also web submission)
<http://salilab.org/modeller>
<http://alto.compbio.ucsf.edu/modweb-cgi/main.cgi>

Assumptions & Principles

- Increase in sequence identity correlates with increase in structural similarity
- RMSD of core α -carbon coordinates for two homologous proteins sharing 50% identity expected at $\sim 1\text{\AA}$ (GLY $3.5\text{\AA}$, helix $5\text{\AA}$ diameter)
- Theoretical models are low resolution, and depend on quality of input alignment!

The SwissModel Workspace

The screenshot shows the Swiss Model Server workspace interface in a web browser. The browser window is titled "Swiss Model Server - Opera" and the address bar shows the URL: http://swissmodel.expasy.org/workspace/index.php?func=workspace_admin&mode=deleteno.... The page title is "SWISS MODEL WORKSPACE".

Navigation links include: [Workspace] [Modelling] [Tools] [Repository] [General Info] [Links] [Help] [Settings] [logout]

Workspace ?

Workunit	Type	Title	Status
P000014	Template Identification	APYRHO	finished
P000015	Template Identification	PORYSA	finished
P000016	Modelling - Automated Mode	APYRHO_automated_five	finished
P000017	Modelling - Automated Mode	APYRHO_automated_best	finished
P000018	Modelling - Automated Mode	APYRHO_automated_2bufA	finished
P000019	Modelling - Automated Mode	PORYSA_automated_five	finished
P000020	Modelling - Automated Mode	PORYSA_automated_best	finished
P000024	Structure Assessment	APYRHO_automated_five	finished
P000025	Structure Assessment	APYRHO_automated_2bufA	finished
P000026	Structure Assessment	PORYSA_automated_five	finished
P000027	Structure Assessment	PORYSA_automated_2akoA	finished
P000028	Modelling - Alignment Mode	APYRHO_alignment_2bnfA	finished
P000029	Modelling - Alignment Mode	PORYSA_alignment_1e19A	finished
P000030	Structure Assessment	APYRHO_alignment_2bnfA	finished
P000031	Structure Assessment	PORYSA_alignment_1e19A	finished

Symbols:

- submission not finished
- queued
- running
- failed/stopped
- finished
- 7days left
- 1days left
- will be deleted
- keep 7 days longer
- delete workunit

swissmodel.expasy.org/workspace [SWISS-MODEL Team]

Template Search

Swiss Model Server - Opera

File Edit View Bookmarks Widgets Tools Help

PS JS

Google

HOW ... Mail ... Gmail ... FMI ... BBC N... BBC S... Conte... Chris ... Sprog... The La... Univer... Genes... The B... Amaz... Exerci... Swiss ...

http://swissmodel.expasy.org/workspace/index.php?func=tools_targetidentification1&userid=chris.wilton@helsinki.fi&tol=

100%

Transfers
History
Widgets
Search
Info

SWISS MODEL WORKSPACE

[Workspace] [Modelling] [Tools] [Repository] [General Info] [Links] [Help] [Settings] [Logout]

Template Identification ?

Email:

Project Title:

Please provide a protein sequence or a UniProt AC Code: ?

MRWVSMRVIKVGGSVLENLEKVKPEIFRDAVVHGGSRVYDELAKKLGLNVEKLTSPSG
VTFRRTRTRKVLVDVYIAAVMKANRELVSFLRROGIEATGVSGKEVII GRRKKLIKAVING
KIMAIRDDYSGIIEKVNVMIRNMYKVGVPVIAP IAYDPVENVPLNVDGDKVAYHVALAL
KSRELYFLSDTAFLINGEVIDKIPRNRIEYFPYSSGGMRKKLLMAKKAIESGVERVIE
GLNGRTVIS

☒ **InterPro Domain Scan:** ?

☒ HMMProfam ☒ HMMTigr ☒ ProfileScan ☒ SuperFamily
☒ BlastProDom ☐ FPrintScan ☐ HMMPIR ☐ HMMSmart
☐ ScanRegExp

☒ **Gapped Blast Query:** ?

E-value cutoff: Matrix:

Descriptions: Alignments:

☒ **Iterative Profile Blast:** ?

Stage 1: PSI-BLAST Profile generation vs. NR

Iterations: Value: (threshold for multipass inclusions)

Matrix: SEG Filter:

Stage 2: PSI-BLAST Profile Search

Search Database:

E-value cutoff: Matrix:

Descriptions: Alignments:

Swiss Model Server - Opera

File Edit View Bookmarks Widgets Tools Help

PS JS

Google

HOW... Mail... Gmail... FMI... BBC... Cont... Chris... Spro... The L... Jao Unive... Jao Gene... Blank... The B... Amaz... Hom... Swiss...

http://swissmodel.expasy.org/workspace/index.php?func=workspace_templateidentification&userid=chris.wilton@helsir

100%

Transfers

History

Widgets

Search

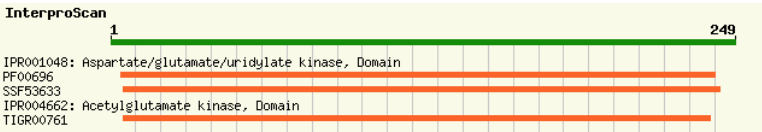
Info

SWISS MODEL WORKSPACE

[Workspace] [Modelling] [Tools] [Repository] [General Info] [Links] [Help] [Settings] [Logout]

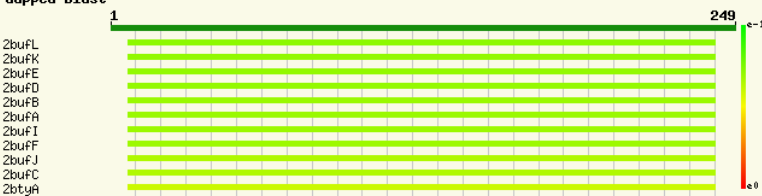
Workunit: P000002
Title: APYRHO_first_approach
Status: running

InterproScan



IPR001048: Aspartate/glutamate/uridylylate kinase, Domain
 PF00696
 SSF53633
 IPR004662: Acetylglutamate kinase, Domain
 TIGR00761

Gapped Blast



ZbuFL
 ZbuFK
 ZbuFE
 ZbuFD
 ZbuFB
 ZbuFA
 ZbuFI
 ZbuFF
 ZbuFJ
 ZbuFC
 ZbtYA

Interpro: top

Interpro Scan has finished. Here are the results:

IPR001048: Aspartate/glutamate/uridylylate kinase, Domain
 PF00696: 5 - 241 AA_kinase

IPR001048: Aspartate/glutamate/uridylylate kinase, Domain
 SSF53633: 6 - 243 Aa_kinase

IPR004662: Acetylglutamate kinase, Domain
 TIGR00761: 6 - 239 argB

Gapped Blast: top

BLASTP 2.2.9 [May-01-2004]

Reference: Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schaffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", Nucleic Acids Res. 25:3389-3402.

Preparation

- Automated mode:
 - Select a template structure with high % identity and average resolution, or let SwissModel choose best one.
- Alignment mode:
 - Obtain multiple alignment between possible templates and your query sequence.
 - Check alignment – are key motifs correctly aligned? Manually adjust as necessary, using Jalview.
 - Upload alignment, select template, and submit.

Submit a Model Request

Swiss Model Server - Opera

File Edit View Bookmarks Widgets Tools Help

PS JS

Google

ssmodel.expasy.org/workspace/index.php?func=modelling_sim

100%

SWISS MODEL WORKSPACE

[Workspace] [Modelling] [Tools] [Repository] [General Info] [Links] [Help] [Settings] [Logout]

SwissModel Automatic Modelling Mode

Email:

Project Title:

Provide a protein sequence or a UniProt AC Code:

MRWVSMRVIKVGGSVLENLEKVKPEIFRDAVVHGGSRVDELAKKLGLNVEKLTSPSG
VTFRRTTRKVLVDYIAAVMKANRELVSFLRRQIEAIGVSGVKEVII GRRKKLIKAVING
KIMAIRDDYSGIIEVNVEMIRNYMKVGPVIAPIAYDPVENVPLNVDGDKVAYHVALAL
KSRELYFLSDTAFLINGEVIDKIPNRRIEYFPYSSGGMRRKLLMAKKAIESGVERVII
GLNGRTVIS

Options:

BLAST E-value limit to select

Use a specific template:

swissmodel.expasy.org/workspace [SWISS-MODEL Team]

Submit a Model Request

The screenshot shows the Swiss Model Server web interface in an Opera browser. The browser's address bar displays the URL: `http://swissmodel.expasy.org/workspace/index.php?func=modelling_align6&user=chris.wilt...`. The page title is "SWISS MODEL WORKSPACE".

The interface includes a sidebar on the left with navigation links: Transfers, History, Widgets, Search, and Info. The main content area displays the following information:

- Workunit:** P000032
- Title:**
- Status:** submission
- [back](#)

Below this, a message states: "Before submission, please ensure that the alignment has been interpreted correctly:". This is followed by the text "CLUSTAL W(1.81) multiple sequence alignment".

The alignment data is presented in a table-like format, comparing a Target sequence (1-331) with a Template sequence (1-331). The sequences are aligned, with gaps indicated by dashes. The alignment is as follows:

Target/1-331	Template/1-331
MRVSMRVIKVGGSVLEN-----LEKVKPE--IFRDAVVHGGSRVY	---GKRVIIALGGNALQQRGQKGSYEEMDNVRKTARQIAEIIAR-GYEVVITHGNPQV
---	..**:*:*:*:
DELAKKLGLNVEKLTSPSG--VTFRRTTRKVLVDVYIAAVMKANRELVSFLRROG--IEA	G--SLLLHM-DAGQATYGIPAPMDVAGAMSQGWIGYMIQQ--ALKNELRKRGMKQV
---	..* : : : * : : : : * : : : * : ** : *
IGVSGVKEVIIGRR-----KKLIKAVIN-----GKIMAIRDDYSG--	TIITQ---TIVDKNDPAFQNP TKPVGP FYDEETAKRLAREKGWIKEDSGRGMRRVVPSP
---	.. : * : : . : : . : .
-IIKEVNVEMIRNYMKV--GVPVIAPIAYDPVE-----NVPLNVDGDKVAYHVALALK	DPKGHVEAETIKKLVE-RGVIVIASGGGVPILEDGEIKGVEAVIDKDLAGEKLAEEVN
---	..* : * : : : : : : : * : * : * : : * : :
SRELYFLSDTAFLNGEVIDKIPNRRIEEYFPYS-----SGGMRKLLMAKK	ADIFMILTDVNGAALYYGTEKEQ--WLREVKEELRKYEEGHFKAG--SMGPKVLAIR
---	.. : * : * : : * : : * : . : * : * : * :
AIESGVERVIEG-----	FIEWGGERAIIAHLEKAVEALEGKTGTQVLP
---	** * * *

At the bottom of the main content area, there is a button labeled "submit alignment".

The footer of the page displays the URL `swissmodel.expasy.org/workspace` and the text "[SWISS-MODEL Team]".

Hidden Modelling Steps

- Automated mode, multiple templates:
 - Superimposition and optimisation of corresponding α -carbon pairs for all template structures
 - Alignment of all residues of template structures: acceptable alpha-carbon atoms located within 3Å radius of mean
- All modes:
 - Framework construction of query protein structure, using coordinates of template structure, or mean coordinate positions if multiple templates.
 - Building unconserved loop regions / insertions, closing gaps in backbone - uses penta-peptide fragment library derived from all PDB entries of resolution 2Å or better (note: this can fail badly!)

Output

- Within a few minutes of submission, your results are returned to the Workspace.
- Output Page:
 1. Your model in pdb format (plus simple viewer)
 2. Query to template alignment
 3. Simple assessment graphs
 4. Logging data of the modelling process
- Save the model in DeepView Project format, then open in DeepView, and colour by B-factor.

917:856 - Swiss Model Server - Opera

File Edit View Bookmarks Widgets Tools Help

HOW 2006: Protein An... Mail :: INBOX JalView - Download Page RCSB PDB : Structure E... Swiss Model Server HEL LOOKS | Exhibition

http://swissmodel.expasy.org/workspace/index.php?userid=chris.wilton@helsinki.fi&token=2d98a2e42

History Transfers Search Widgets Info

Template Selection Log: [\[top\]](#)

Template Selection

Searching ExPDB library for suitable templates ...

ExPDB Version: 4.0.20060920

Search Query : userseq (249 residues)
Putative Hits: 14

Hit	ID	Evalue	Resolution	Link
Hit 1:	2bufL	4.42e-11	2.95	[Template]
Hit 2:	2bufK	5.92e-11	2.95	[Template]
Hit 3:	2bufE	7.06e-11	2.95	[Template]
Hit 4:	2bufD	7.67e-11	2.95	[Template]
Hit 5:	2bufB	1.11e-10	2.95	[Template]
Hit 6:	2bufA	1.14e-10	2.95	[Template]
Hit 7:	2bufI	1.31e-10	2.95	[Template]
Hit 8:	2bufF	1.39e-10	2.95	[Template]
Hit 9:	2bufG	4.69e-10	2.95	[Template]
Hit 10:	2bufJ	4.69e-10	2.95	[Template]
Hit 11:	2bufH	4.81e-10	2.95	[Template]
Hit 12:	2bufC	5.32e-10	2.95	[Template]
Hit 13:	2btyA	3.07e-09	2.75	[Template]
Hit 14:	2btyB	3.07e-09	2.75	[Template]
Hit 15:	2btyC	3.07e-09	2.75	[Template]

Scoring Hits by Evalue, Resolution, Anolea and Gromos energy

Graphical Template Match Overview

```

=====
Target      1 ++++++ 249
2bufL      *****
2bufE      *****
2bufD      *****
2bufB      *****
2bufA      *****
2bufI      *****
2bufF      *****
2bufG      *****
2bufH      *****
2bufC      *****
2bufJ      *****
2btyA      *****
2btyB      *****
2btyC      *****
2bufK      *****
  
```

Clustering hits into Batches to be modeled seperately.

Batch : 1
Residue Range : 8 - 241
Templates : 2bufL.pdb 2bufE.pdb 2bufD.pdb 2bufB.pdb 2bufA.pdb

Template selection end.

Multiple Sequence Alignment

- More information than pairwise alignment
- Shows conserved *regions* in protein family
 - Critical secondary structure elements
- Shows strongly conserved *residues* and *motifs*
 - Structural importance
 - Functional importance

Alignment Adjustment

- Look at alignment used to select template
 - Check alignment against 3D structure (Jalview)
 - Secondary structure elements preserved?
 - Motifs and key residues aligned?
 - Are gaps in acceptable locations (ie: loops)?
- Try to improve alignment manually ...
- Or use different alignment application

The SwissModel Repository

- Contains theoretical models (996876 on 05/09/2006)
- Searchable by UniProt / Trembl accession number
- **Fully automated procedure - models may be bad!**
- It can be found here:

<http://swissmodel.expasy.org/repository>

- **ModBase** also holds threshold-validated theoretical models [4064704 structures currently]:

<http://alto.compbio.ucsf.edu/modbase-cgi/index.cgi>

Model Assessment

Swiss Model Server - Opera

File Edit View Bookmarks Widgets Tools Help

PS JS http://swissmodel.expasy.org/workspace/index.php?func=tools_structureassessment2&useri... Google

http://swissmodel.expasy.org/workspace/index.php?func=tools_str 100%

SWISS MODEL WORKSPACE

[Workspace] [Modelling] [Tools] [Repository] [General Info] [Links] [Help]

[Settings] [Logout]

Protein Structure & Model Assessment Tools ?

Email:

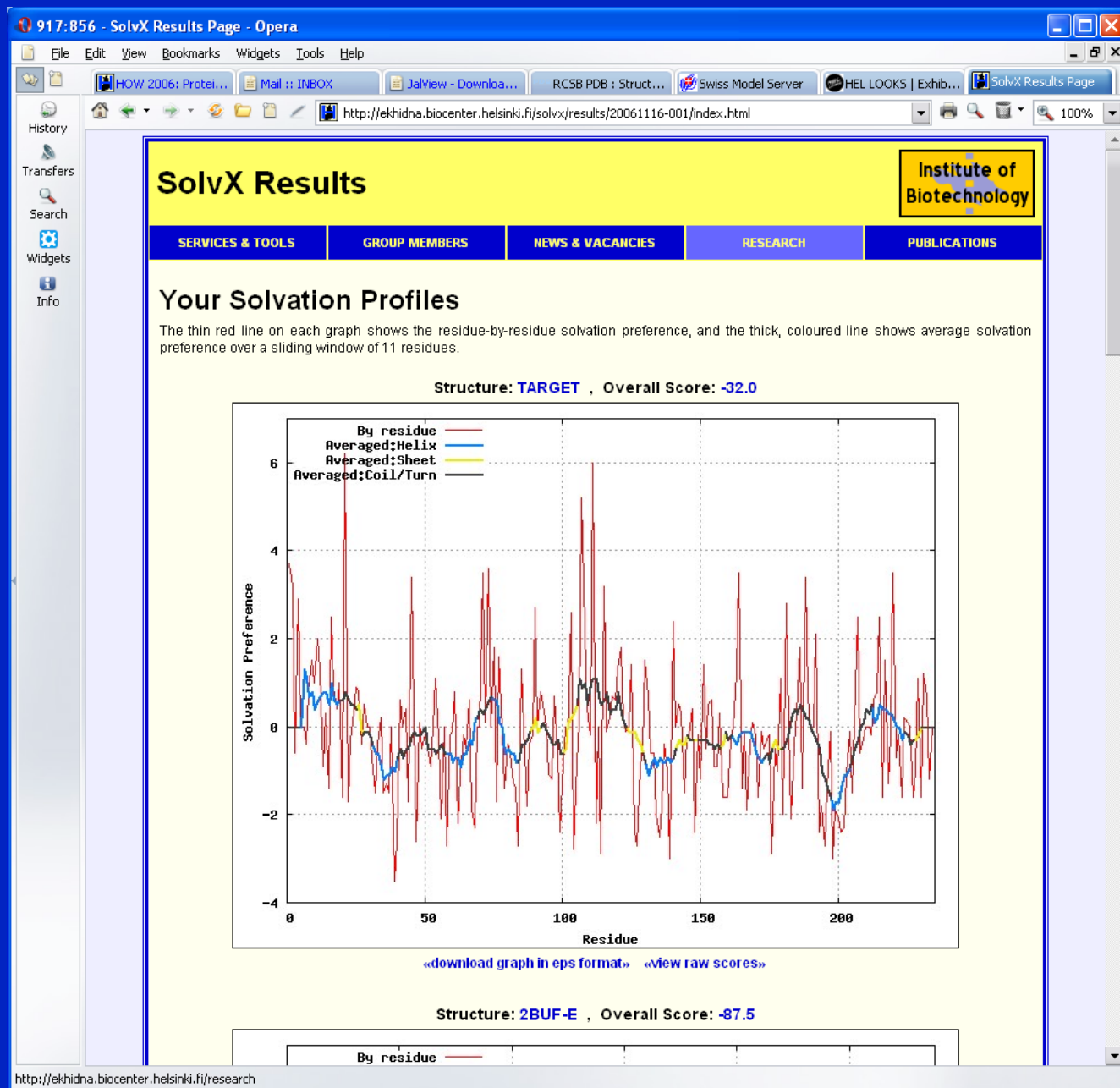
Project Title:

The following tools are provided to assess the quality and structural features of protein models and template structures. Please upload a model or template structure in **PDB format**. ?

Protein Structure Assessment*:

☒ Anolea ?	Anolea atomic mean force potential
☒ Verify3D ?	Assessment of protein models with 3D profiles
☐ Gromos ?	Empirical force field
☒ Whatcheck ?	Protein structure verification
☒ Procheck ?	Stereochemical quality check; min. Resolution: 2.2 Å
☐ DSSP ?	Secondary Structure, geometrical features, and solvent exposure assignment
☐ Promotif ?	Analysis of protein structure motifs

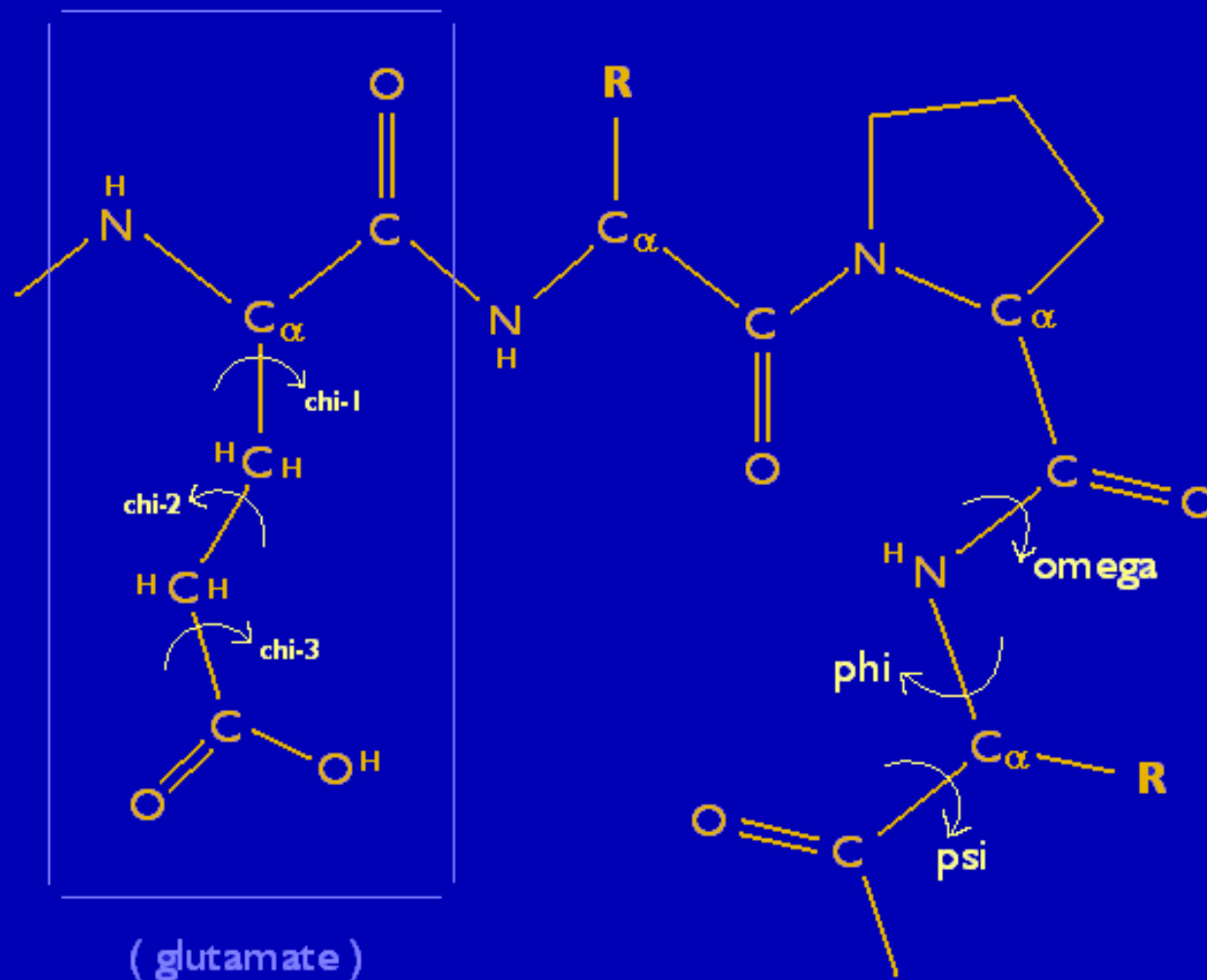
*Disclaimer and user agreement: The use of SWISS-MODEL Server and SWISS-MODEL Workspace is free for both academic and commercial users. However, some of the tools provided on this page may require a valid license for some users. If you are in doubt, please check carefully the licensing conditions for each tool. Links to the original sites are provided in the help file. By using SWISS-MODEL workspace, you confirm that you are legally entitled to use these tools. You also agree to cite SWISS-MODEL Workspace AND the corresponding reference of these tools in any of your publications reporting results obtained from this service.



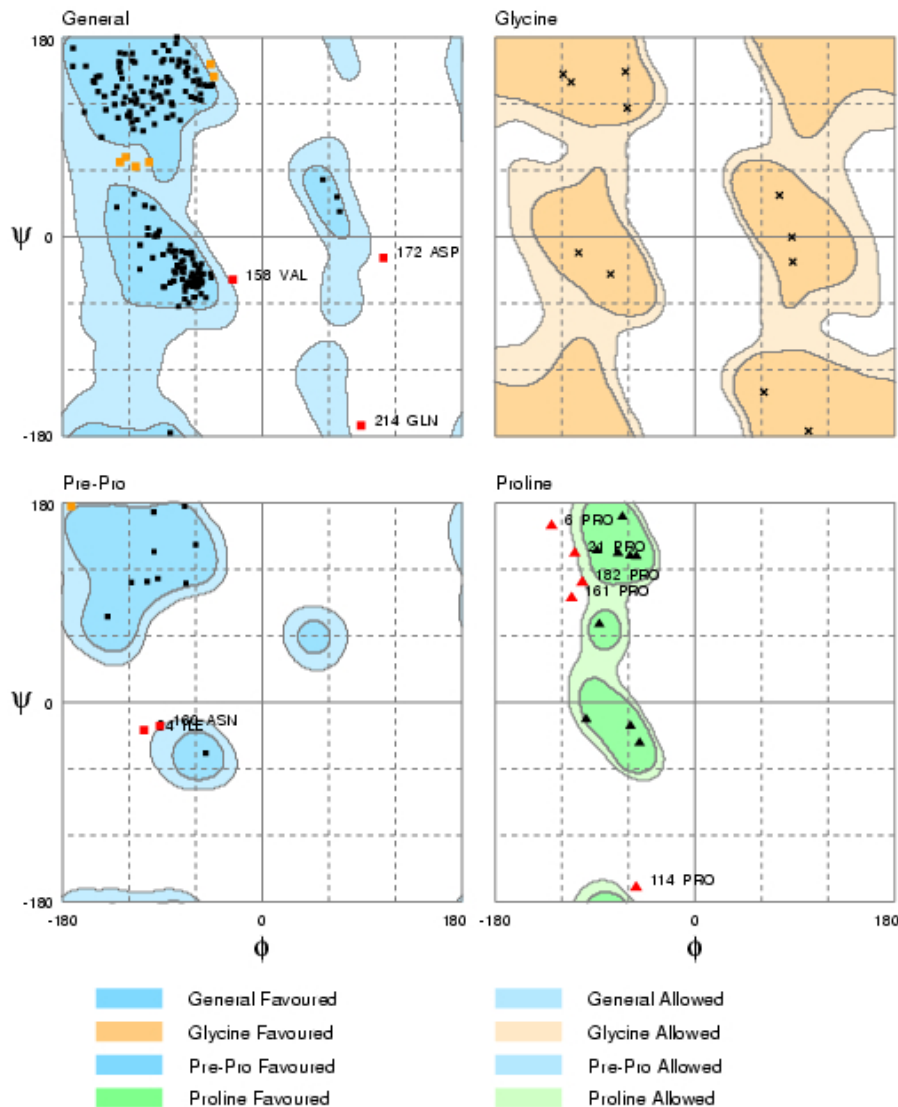
Model Evaluation

- Look at model in DeepView: is it globular and compact, are there odd unstructured loops?
- Colour by B-factor and note bad areas – in loops, or in important secondary elements?
- WhatCheck - always lots of errors! Look for packing, inside-outside, torsion angle errors.
- Check *Solvation Profile* – how well packed is the model? Are there regions which are too loose, and therefore solvent-exposed?

Torsion Angles



Ramachandran Plot



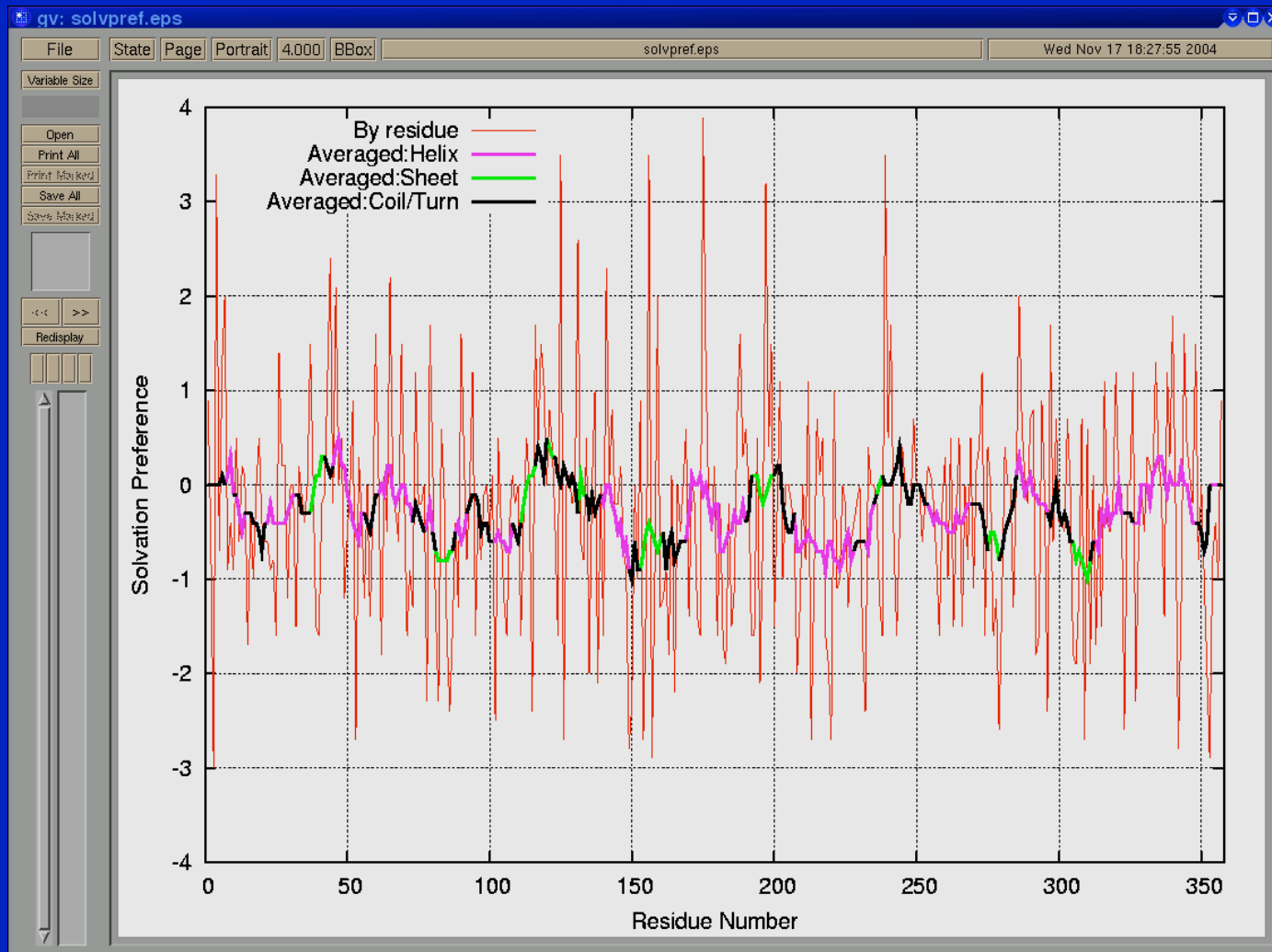
Experimentally Resolved Model
(NMR, XRay):

- Rare for residues other than Glycine & Alanine to be in disallowed regions

Homology Model:

- Bad residues common
- May have to remodel small sections and loops
- Large residues in bad regions worth further investigation ...

Solvation Profile



- Long stretches above zero probably loops
- Most-negative regions well buried in core
- Functional residues above zero!
- Majority of residues must be below zero
- **Compare your model to pdb template!**

Further details

- DeepView tutorial
 - <http://au.expasy.org/spdbv/text/tutorial.htm>
- Advanced tutorial on SwissModel & DeepView
 - <http://www.usm.maine.edu/~rhodes/SPVTut/text/HGHomMod.html>

Suggested Reading

- Swiss-Model's advice:
 - http://www.expasy.org/swissmod/SM_ModelAccuracy.html
- A good interactive tutorial on Protein Modelling:
 - <http://www.ncbi.nlm.nih.gov/Class/minicourses/x1a.html>

References

Schwede T, Kopp J, Guex N, and Peitsch MC (2003) SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Research* 31:3381-3385.

Guex, N. and Peitsch, M. C. (1997) SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modelling. *Electrophoresis* 18:2714-2723.

Peitsch, M. C. (1995) Protein modeling by Email. *Bio/Technology* 13:658-660.

R.W.W.Hooft, G.Vriend, C.Sander and E.E.Abola, Errors in protein structures. *Nature* 381, 272 (1996).

Ramachandran GN, Ramakrishnan C, Sasisekharan V (1963). *J Mol Biol* 7:95-99.

Demonstrations

- Uniprot entry **Q11CR8_MESSB** is described as a "Fructose-bisphosphate aldolase class 1, from *Mezorhizobium* sp. (strain BNC1)"
- It belongs to Pfam [PF00274](#), the Fructose-1,6-Bisphosphate Aldolases (Class 1); this family contains 24 PDB structures.
- When we run BlastP on Q11CR8_MESSB against the PDB, we get hits to all these structures; %ids range from 48 to 54%, which is reassuringly high.
- Let's try SwissModel in Automated mode:
- <http://swissmodel.expasy.org/workspace>

Demonstrations

- Uniprot entry **Q07159** is another Class I FBP Aldolase from *Staphylococcus carnosus*. It is unusual, however – most bacterial FBP Aldolases belong to Class II ([PF00596](#))
- When we run a Blast against the PDB, again we get hits to all twenty-four PDB structures, but identities range from 25 to 32% – not good.
- Try SwissModel in Automated mode...
- Oh dear, only part of our sequence has been modelled – automated mode has failed this time. Pfam has a [family alignment](#), so let's use it as input to Alignment mode.

Demonstrations

Now let's assess our two complete models:

- Go to [Tools] >> Structure Assessment, and upload each model in turn:
 - Add Whatcheck and Procheck to the selection, and add the resolution (check result file for automated mode, or check PDB website for alignment mode)
- Also, go to the [SolvX Server](#) and input our models. It should give us (at least) two graphs per input – the first will be our model, the other(s) our template(s).

Demonstrations

- What validation checks are the important ones?
 - Whatcheck: Accessibility-related (2.6), 3D-Database (2.8), Geometric (2.5.4 2.5.12 2.5.13 2.5.14) Checks. Make a note of bad residues and regions.
 - Procheck: Residues Report (G-factors) and Ramachandran Plot. Again, note bad residues and regions.
 - SolvX: Compare your model with the template(s). Where are the worst (most solvent accessible) regions? Note them down.
- How does your alignment model for ALF1_STACA compare with your neighbour's model?