Homology modeling Tutorial

Sequence to structure

Dr. Ratna Singh

Homology modeling

Principle

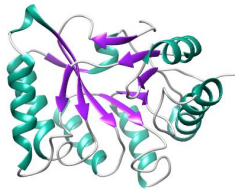
Evolutionary related protein tend to have similar structure

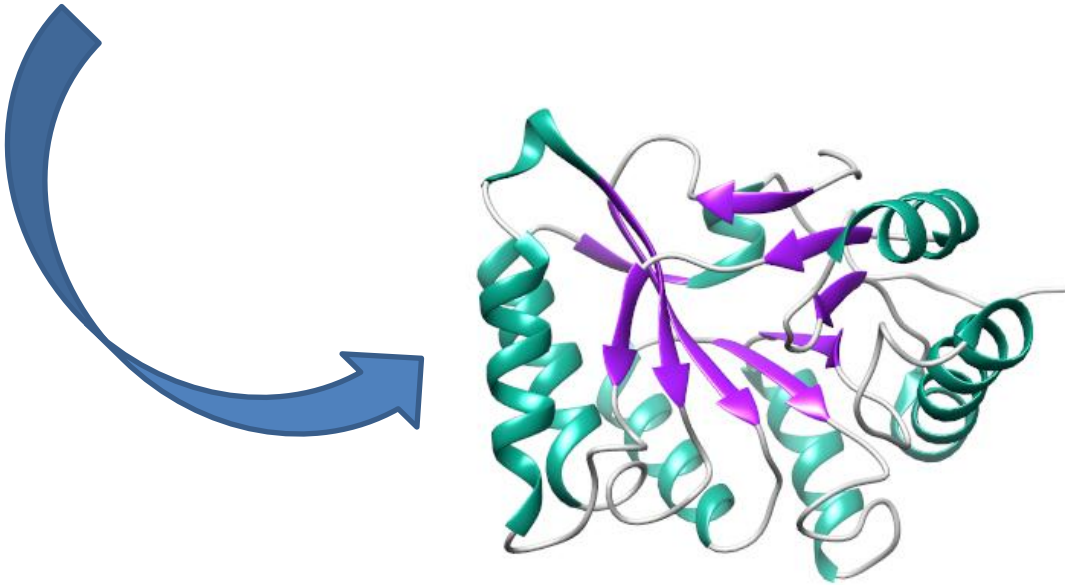
Homology modeling

Predicting structure from the sequence

```
 89 SKSISFGGCLTQMYFMIALGNTDSYILAAMAYDRAVAIS 127
 68 -FCAACHGCLFIACFVLVLTQSSIFSLLAIAIDRYIAIR 105

128 RPLHYTTIMSPRSCIWLIAGSWVIGNANALPHTLL-TAV 165
106 IPLRYNGLVTGTRAKGIIAICWVLSFAIGLTP-MLGWNA 143
```

3d structure of template





Homology modeling/Comparative modeling

TEMPLATES ⇢ **TARGET**

→ Predicts the three-dimensional structure of a given protein sequence (TARGET) based on an alignment to one or more known protein structures (TEMPLATES)

→ If similarity between the TARGET sequence and the TEMPLATE sequence is detected, structural similarity can be assumed.

Comparative Protein Models will be increasingly utilized to help solve biological problem

Homology modeling

Basic steps in Homology modeling :

i) Model building based on template

ii) Refinement

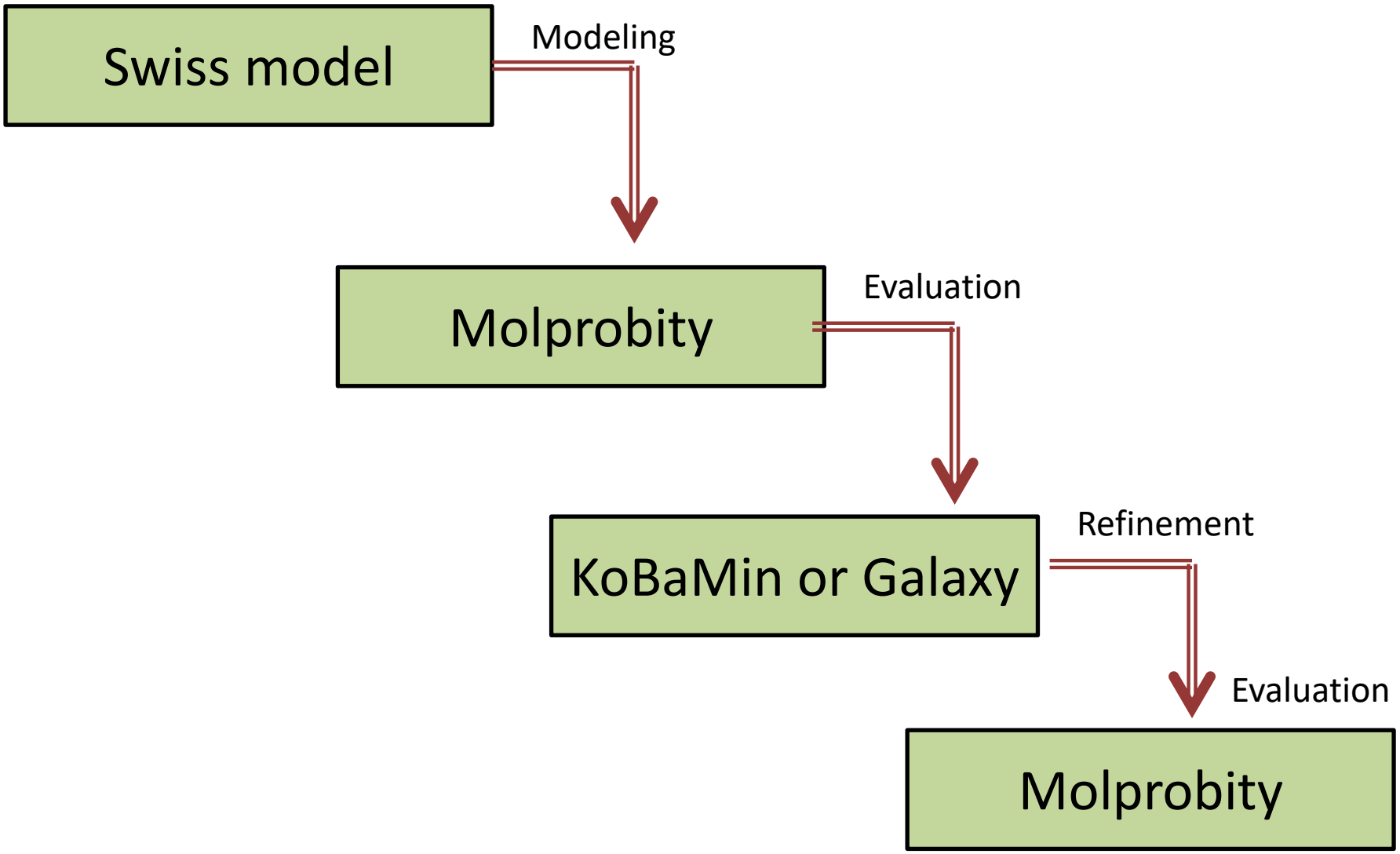
ii) Evaluation

Building the model

Software for homology molecular modeling

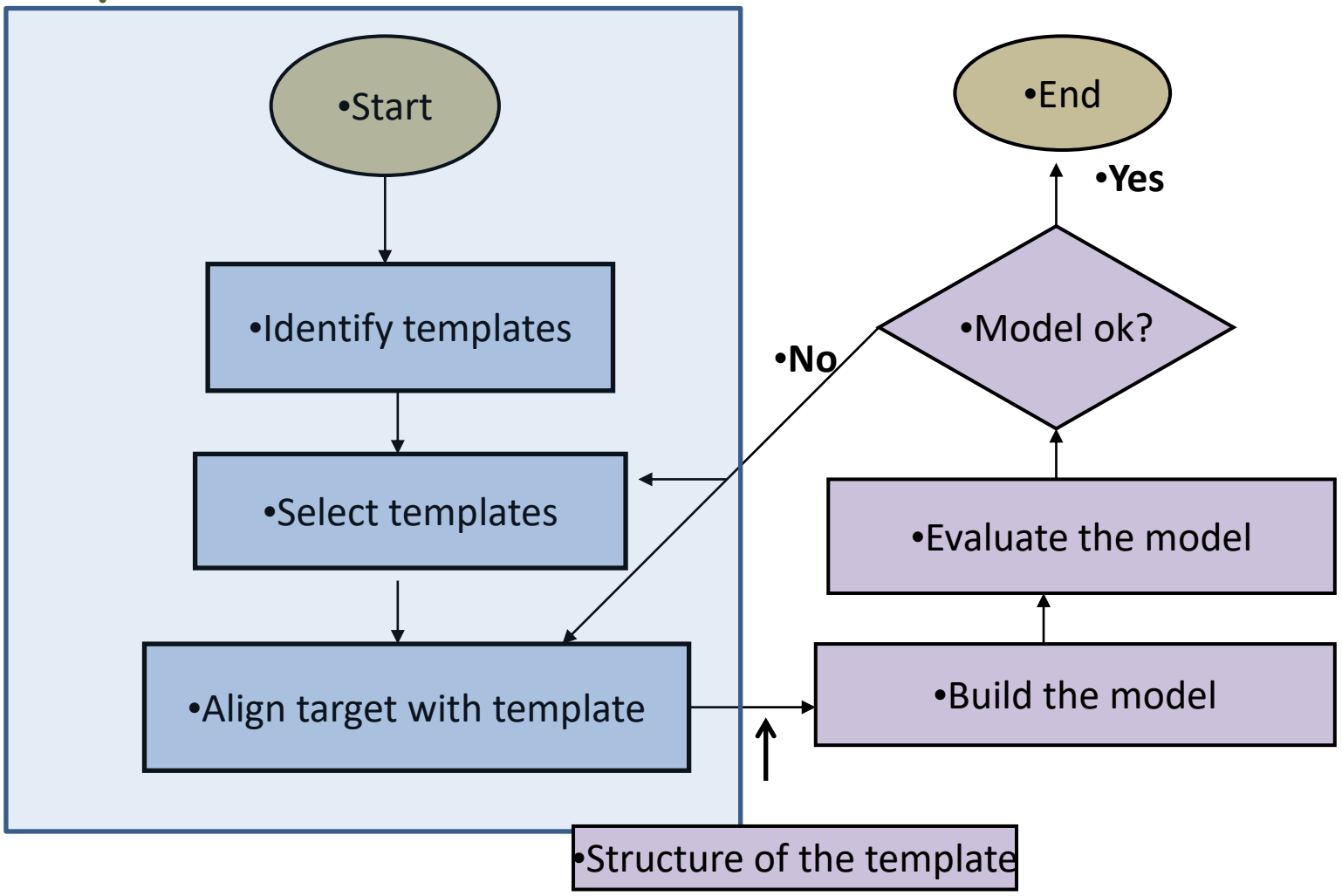
- Freeware: available for all OS
  - Downloadable
    - Modeller (Sali, 1998)
    - DeepView (SwissPDB viewer)
  - Web based:
    - **I-TASSER** (http://zhanglab.ccmb.med.umich.edu/I-TASSER/SDSC1)
    - **SWISS MODEL** server (www.expasy.org/swissmod/SWISS-MODEL.html)
    - **Hhpred server** (https://toolkit.tuebingen.mpg.de/tools/hhpred)
    - **PSIPRED** (http://bioinf.cs.ucl.ac.uk/psipred/)

Summary steps for modeling



Basic Comparative Protein Modeling Procedures

Step1

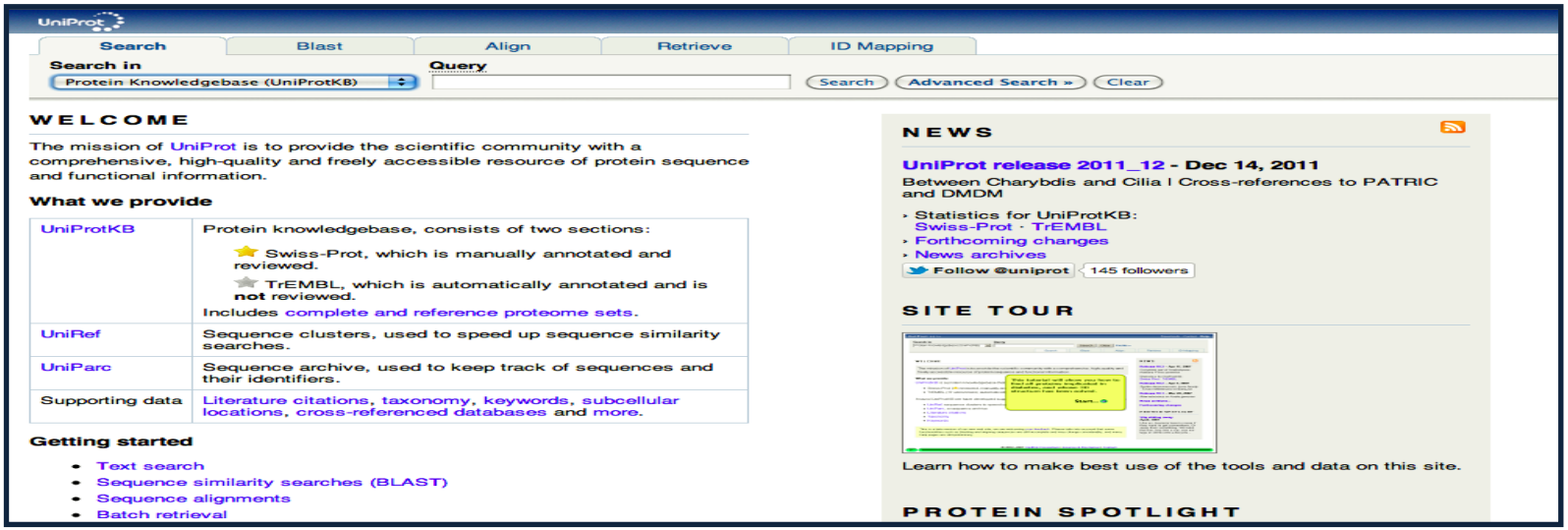


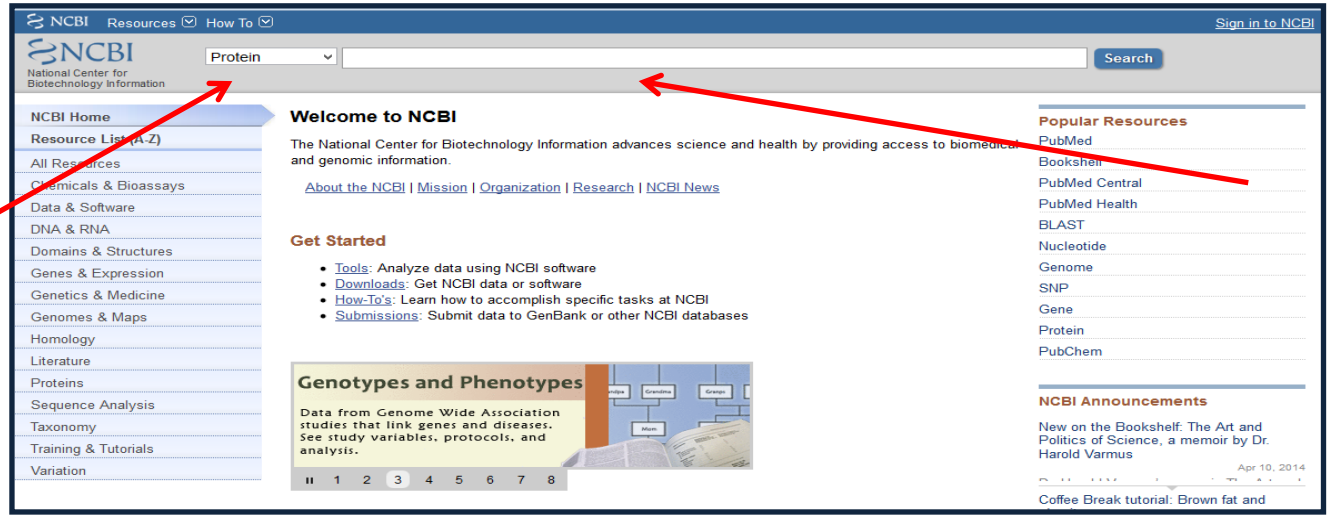
Homology modeling/Comparative modeling

Input for the modeling

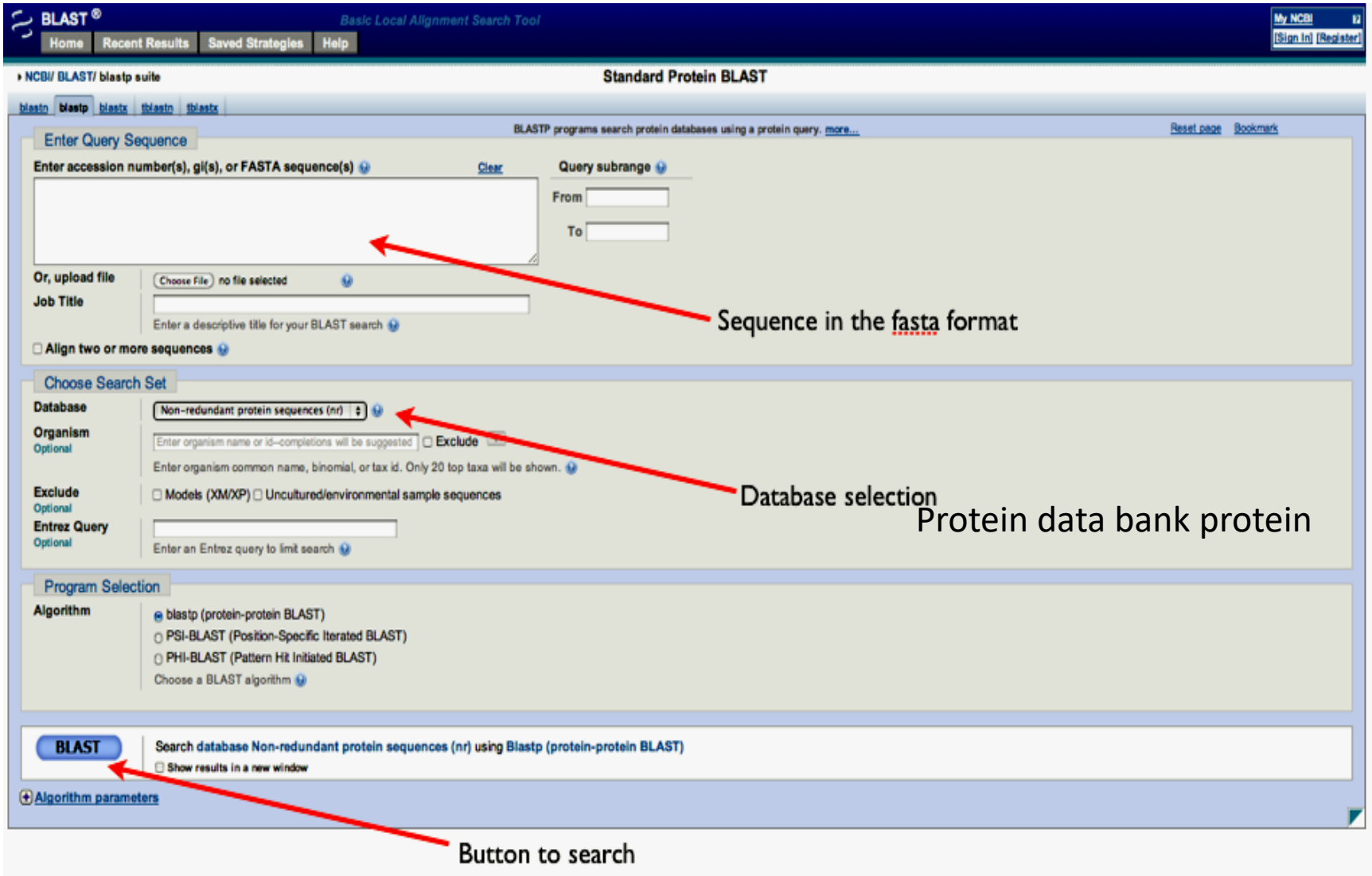
- Target sequence- Sequence of the structure with unknown 3D structure
- 3D template- selected based on the highest sequence identity with the target sequence
- Alignment between the target and template

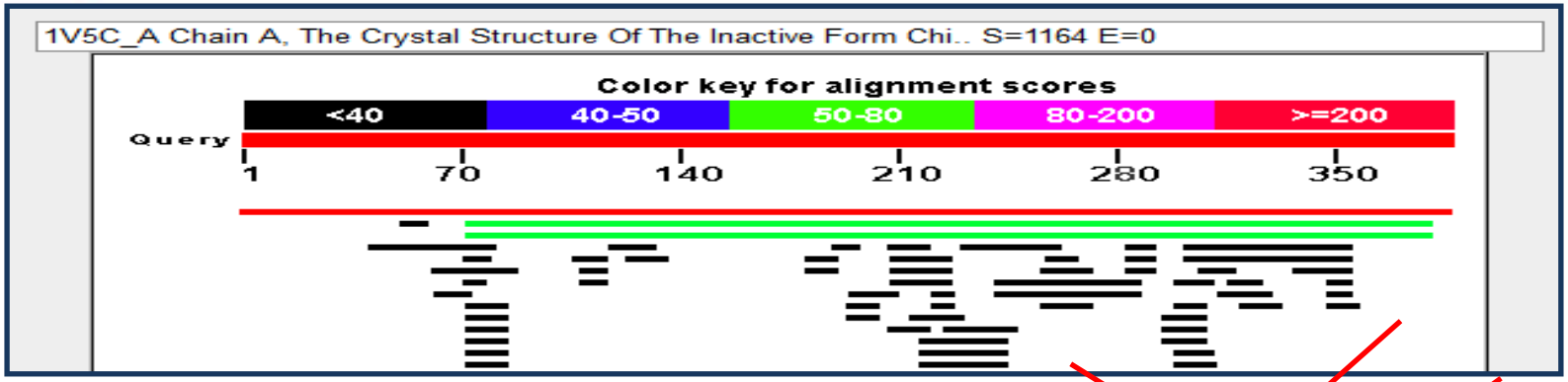
Target selection





Template selection





	Description	Max score	Total score	Query cover	E value	Ident	Accession	Select for PSI blast	Used to build PSSM
☐	Chain A, The Crystal Structure Of The Inactive Form Chitosanase From Bacillus Sp. K17 At Ph3.7 >gi|58176989|pdb|1V5D|A Chair	1164	1164	100%	0.0	97%	gi|58176988|1V5C_A	☑	
☐	Chain A, Endoglucanase A From Clostridium Thermocellum At Atomic Resolution >gi|157830584|pdb|1CEM|A Chain A, Endoglucı	65.1	65.1	79%	1e-10	30%	gi|23200137|1IS9_A	☑	
☐	Chain A, Atomic Resolution Structure Of An Inverting Glycosidase In Complex With Substrate	57.1	57.1	79%	4e-08	29%	gi|20151090|1KWF_A	☑	

Chain A, The Crystal Structure Of The Inactive Form Chitosanase From Bacillus Sp. K17 At Ph3.7

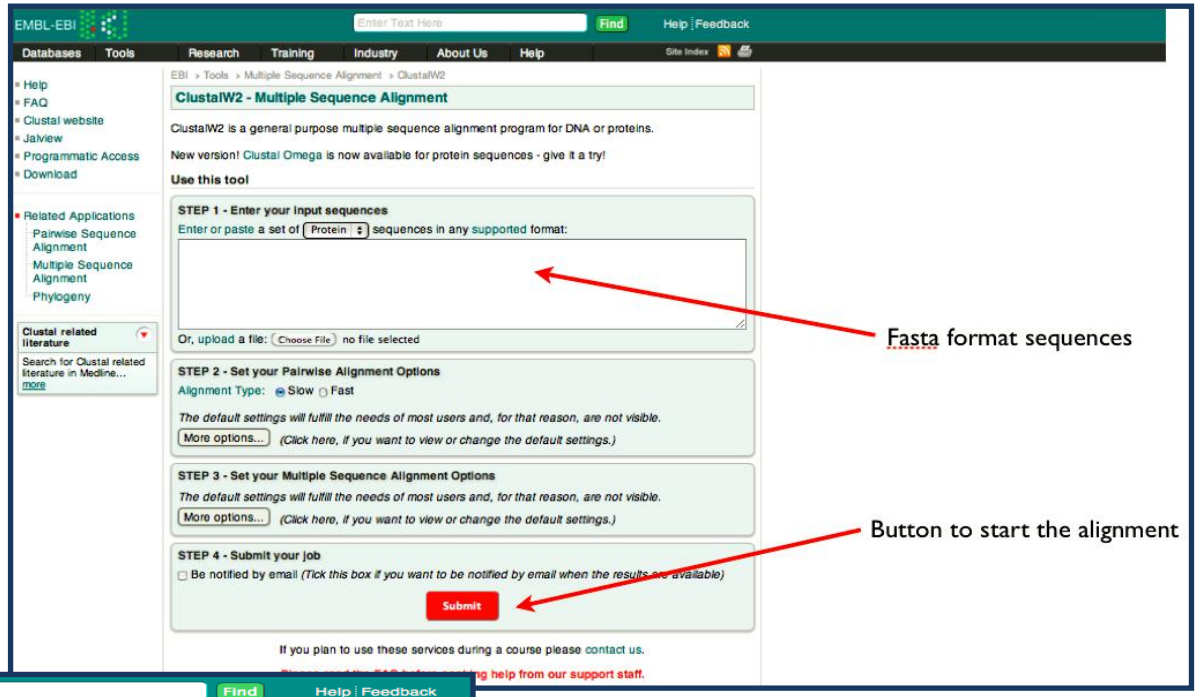
Sequence ID: gi|58176988|pdb|1V5C|A Length: 386 Number of Matches: 1

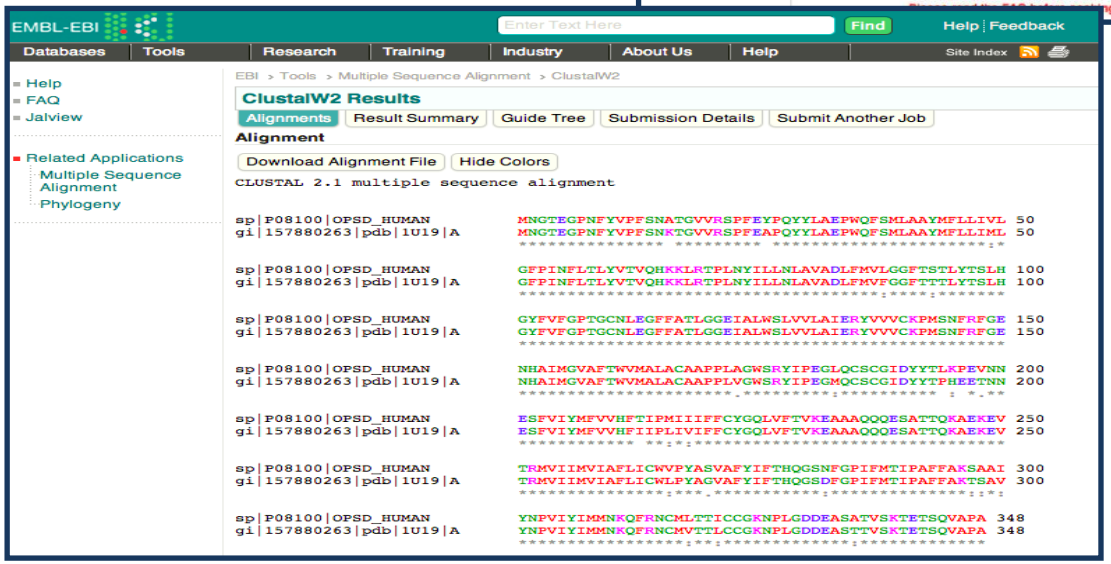
► See 2 more title(s)

Range 1: 1 to 386 GenPept Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
1164 bits(2737)	0.0	Composition-based stats.	376/386(97%)	378/386(97%)	0/386(0%)

```
Query  1    AKEMKPFPQQVNYAGVIKPNHVTQESLNASVRSYYDNWKKKYLKNDLSSLPGGYYVKGEI  60
            AKEMKPFPQQVNYAGVIKPNHVTQESLNASVRSYYDNWKKKYLKNDLSSLPGGYYVKGEI
Sbjct  1    AKEMKPFPQQVNYAGVIKPNHVTQESLNASVRSYYDNWKKKYLKNDLSSLPGGYYVKGEI  60

Query  61   TGDADGFKPLGTSEGQGYGMIITVLMAGYDSNAQKIYDGLFKTARTFKSSQNPNLMGWVV  120
            TGDADGFKPLGISEGQGYGMIITVLMAGYDSNAQKIYDGLFKTARTFKSSQNPNLMGWVV
Sbjct  61   TGDADGFKPLGISEGQGYGMIITVLMAGYDSNAQKIYDGLFKTARTFKSSQNPNLMGWVV  120

Query  121  ADSKKAQGHFDSATDGDLDIAYSLLLAHKQWGSNGAVNCLKEAQDMITKGIKASNVTNNS  180
            ADSKKAQGHFDSATDGDLDIAYSLLLAHKQWGSNG VN LKEAQDMITKGIKASNVTNN
Sbjct  121  ADSKKAQGHFDSATDGDLDIAYSLLLAHKQWGSNGTVNYYLKEAQDMITKGIKASNVTNNN  180

Query  181  RLNLGDWDSKSSLDTRPSDWMMSHLRAFYEFTGDKTWLTVINNLYDVYTQFSNKYSPNTG  240
             LNLGDWDSKSSLDTRPSDWMMSHLRAFYEFTGDKTWLTVINNLYDVYTQFSNKYSPNTG
Sbjct  181  QLNLGDWDSKSSLDTRPSDWMMSHLRAFYEFTGDKTWLTVINNLYDVYTQFSNKYSPNTG  240

Query  241  LISDFVVKNPPQPAPKGFLNESEYTNAYYYNASRVPLRIVMDYAMYGEKRSKVISDKVSS  300
            LISDFVVKNPPQPAPK FL+ESEYTNAYYYNASRVPLRIVMDYAMYGEKRSKVISDKVSS
Sbjct  241  LISDFVVKNPPQPAPKDFLDESEYTNAYYYNASRVPLRIVMDYAMYGEKRSKVISDKVSS  300

Query  301  WIQNKTNGNPSKIVDGYQLNGSNIGNYPTAVFVSPFIAASITNSNNQKWVNSGWDWMKNK  360
            WIQNKTNGNPSKIVDGYQLNGSNIG YPTAVFVSPFIAASIT SNNQKWVNSGWDWMKNK
Sbjct  301  WIQNKTNGNPSKIVDGYQLNGSNIGSYPTAVFVSPFIAASITSSNNQKWVNSGWDWMKNK  360

Query  361  RESYFSDSYNLLTMLFITGNWWKPIP  386
            RE YFSDSYNLLTMLFITGNWWKP+P
Sbjct  361  RERYFSDSYNLLTMLFITGNWWKPVP  386
```





Alignment

Dr. Ratna Singh

Rules for Template selection

- The quality of the homology model is dependent on the quality of the sequence alignment and template structure

- Model quality declines with decreasing sequence identity.

- The approach can be complicated by the presence of alignment gaps that indicate a structural region present in the target but not in the template

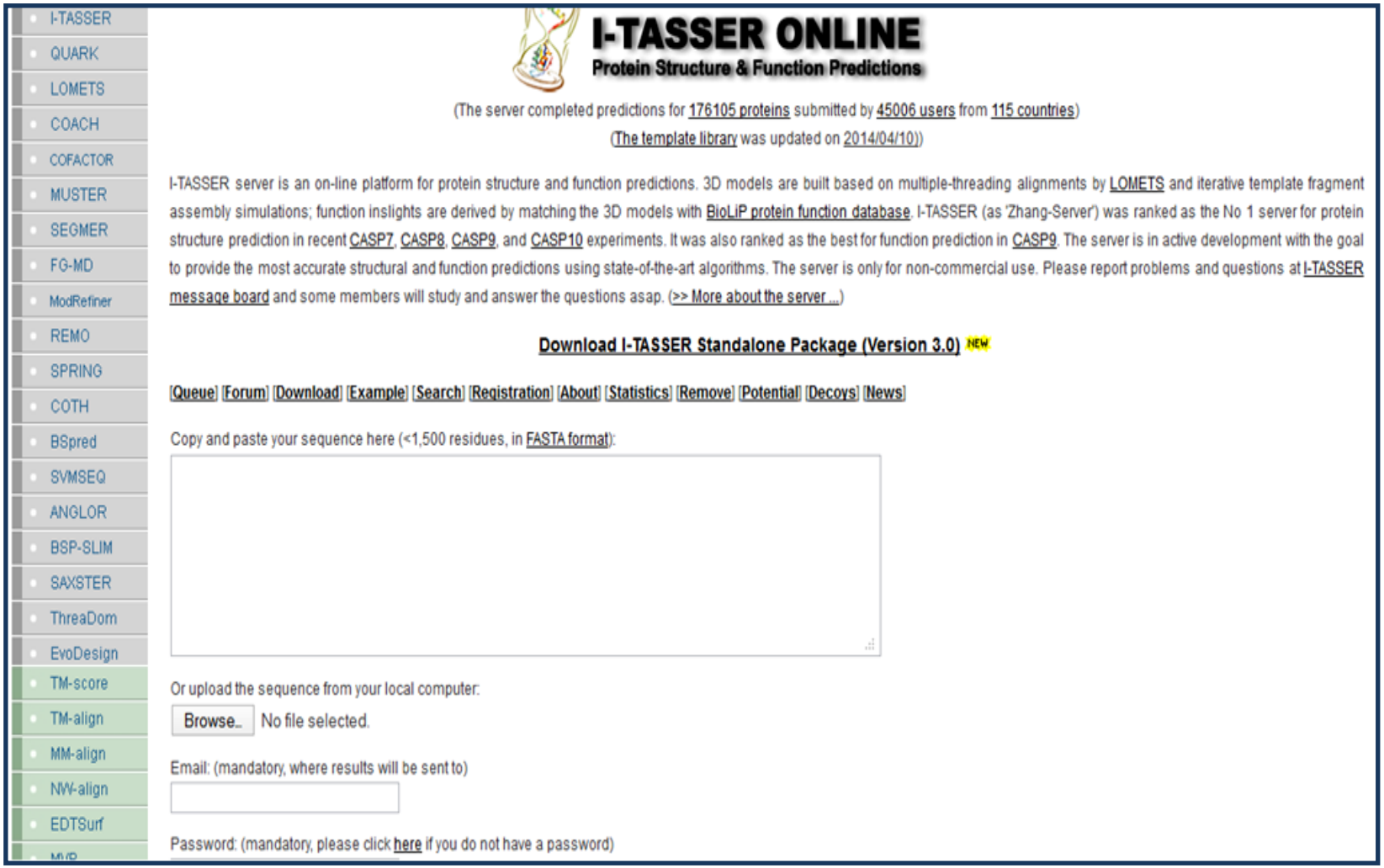
- Is an X-ray crystallography structure (NMR structure are usually much less resolved)?

- All the atoms are resolved in the selected structure?

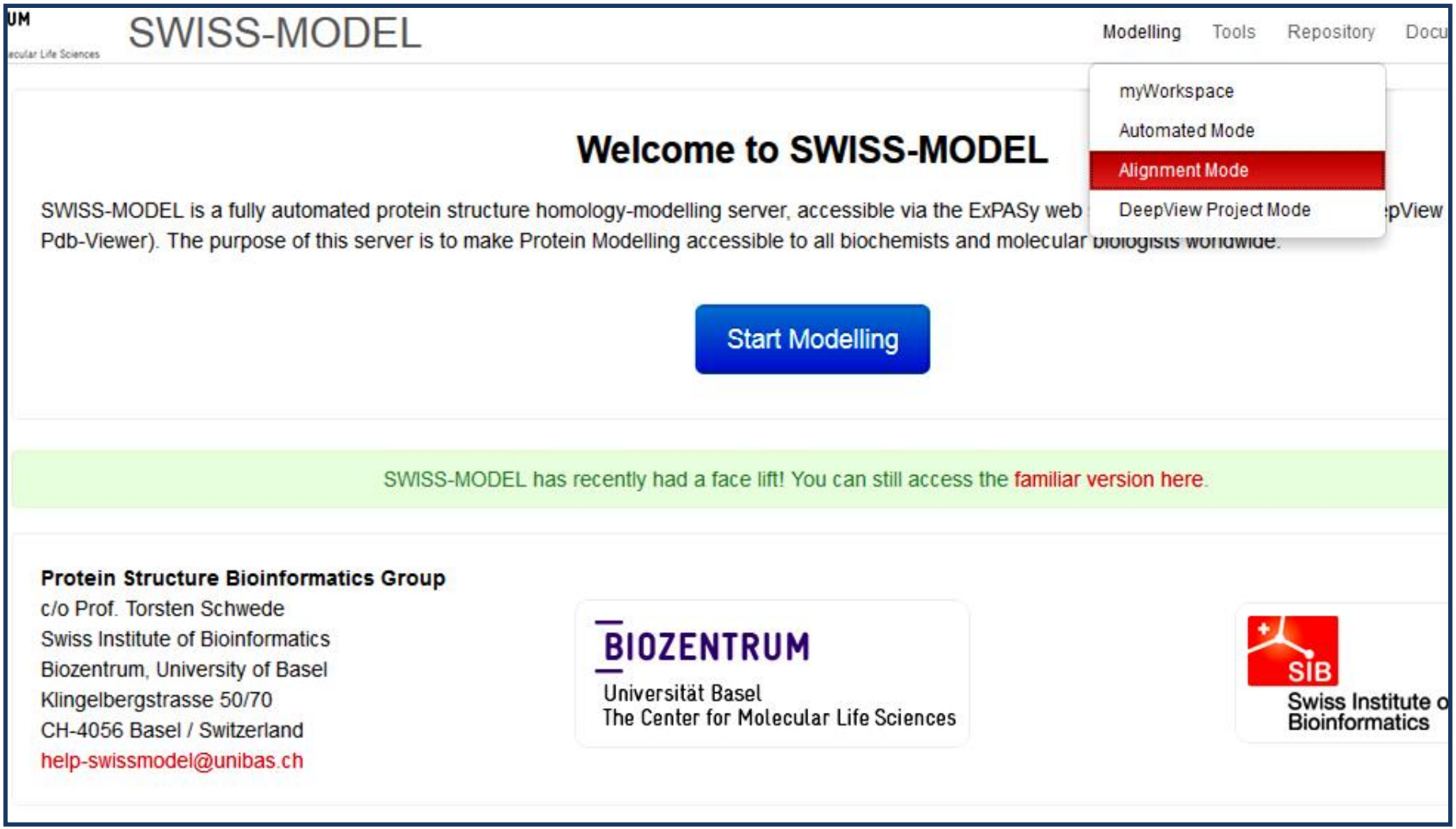
- Is the chosen structure the best resolved one (typical good resolution is smaller than 2Å)

As a rule of thumb we should never use template with an E value larger than 0.1 .

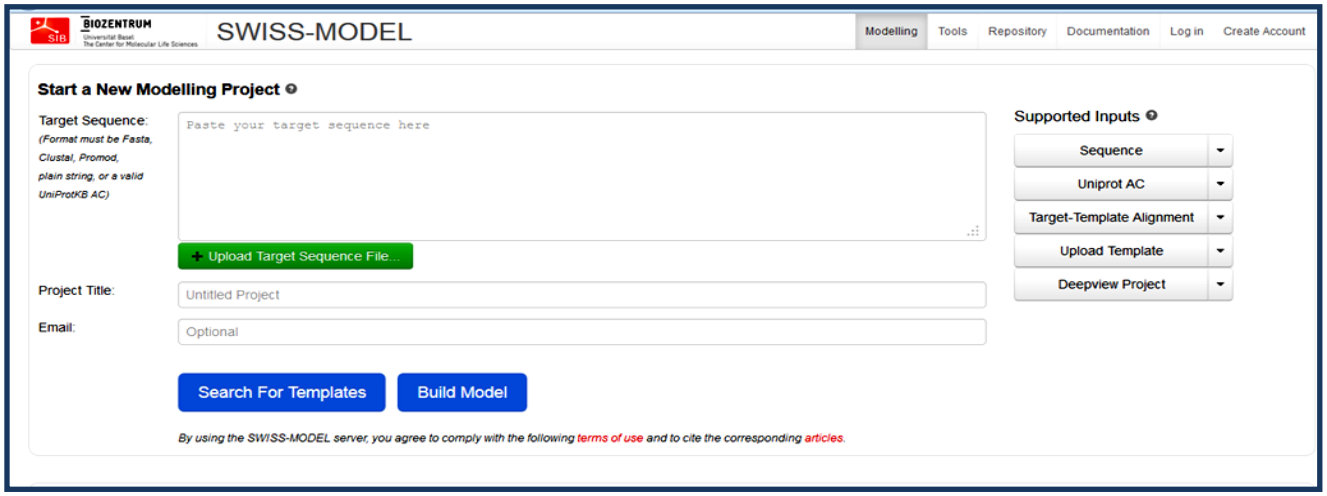
Building the model



Building the model

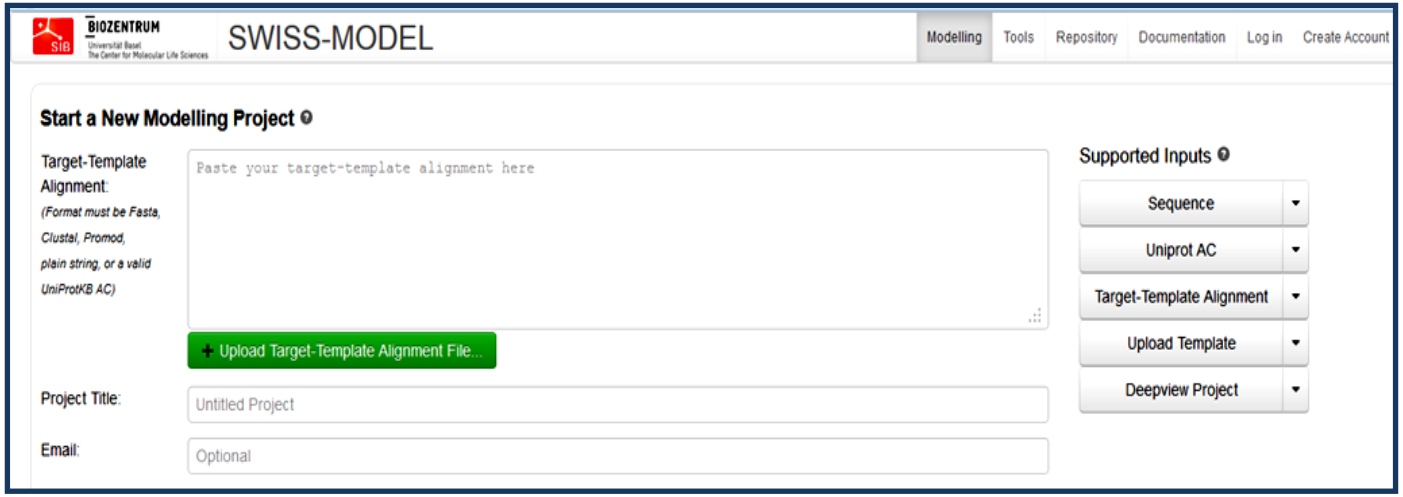


Building the model



Automated mode

Alignment mode



Dr. Ratna Singh

Building the model

Summary | Templates 38 | Models 0

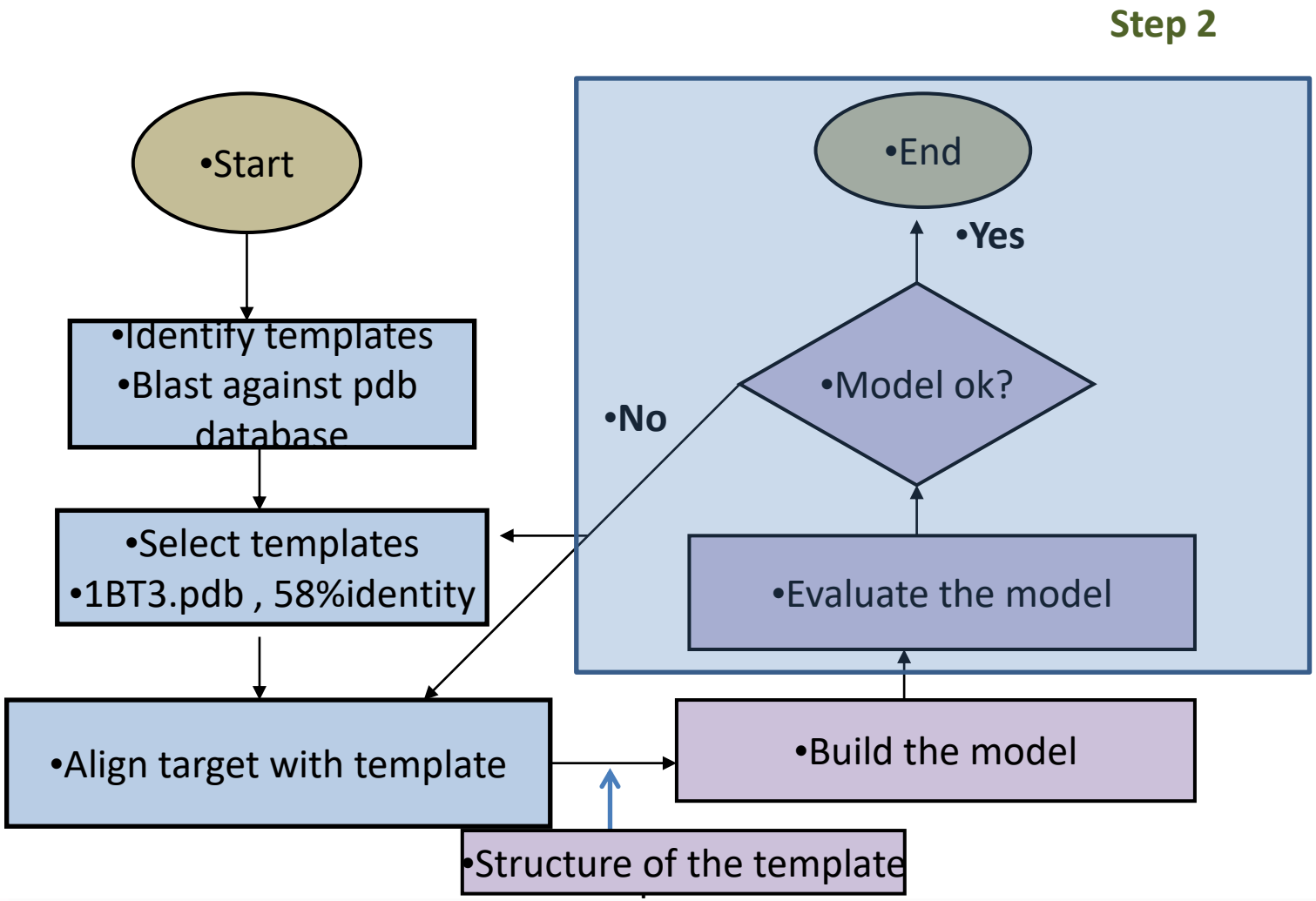
Template Results

Templates | Sequence Similarity | Alignment of Selected Templates | More

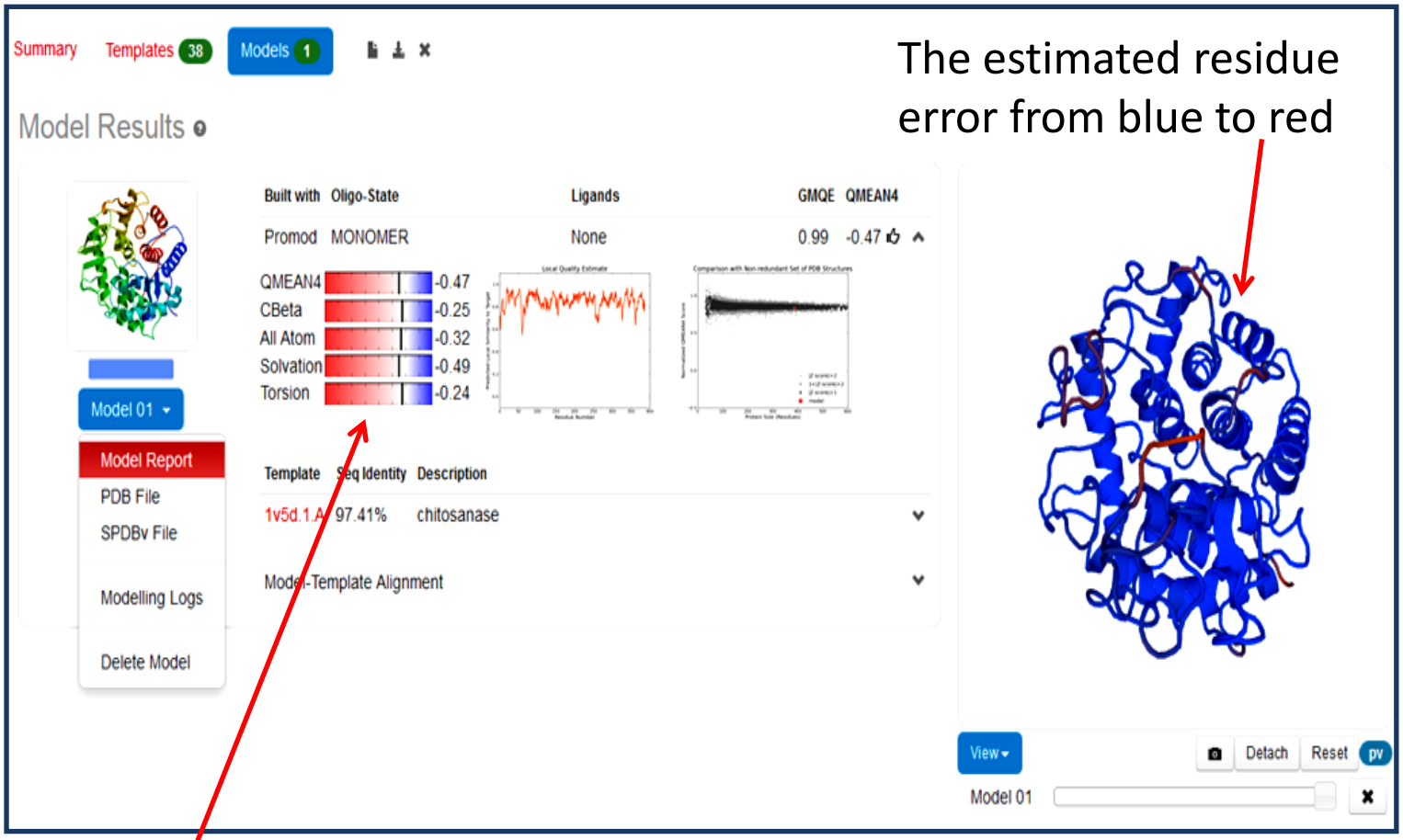
Build Models 1 | Clear Selection

Name	Title	Coverage	Identity	Method	Oligo State	Ligands
1v5d.1.A	chitosanase		97.41	X-ray, 1.5Å	monomer	None
1cem.1.A	CELLULASE CELA (1,4-BETA-D-GLUCAN-GLUCANOHYDROLASE)		30.86	X-ray, 1.6Å	monomer	None
1kwf.1.A	Endoglucanase A		30.29	X-ray, 0.9Å	monomer	None
1cem.1.A	CELLULASE CELA (1,4-BETA-D-GLUCAN-GLUCANOHYDROLASE)		33.63	X-ray, 1.6Å	monomer	None
1kwf.1.A	Endoglucanase A		33.03	X-ray, 0.9Å	monomer	None
1h13.1.A	ENDO-1,4-BETA-XYLANASE		19.54	X-ray, 1.3Å	monomer	None
1xwt.1.A	endo-1,4-beta-xylanase		19.54	X-ray, 1.3Å	monomer	None
1xw2.1.A	Endo-1,4-beta-Xylanase		19.25	X-ray, 1.8Å	monomer	None
1h14.1.A	ENDO-1,4-BETA-XYLANASE		18.97	X-ray, 1.5Å	monomer	None
2a8z.1.A	endo-1,4-beta-xylanase		18.97	X-ray, 3.2Å	monomer	None
2drr.1.A	Xylanase Y		17.20	X-ray, 1.6Å	monomer	1 x NI
1wu6.1.A	xylanase Y		16.62	X-ray, 1.4Å	monomer	1 x XYP-XYP 1 x NI

Basic Comparative Protein Modeling Procedures



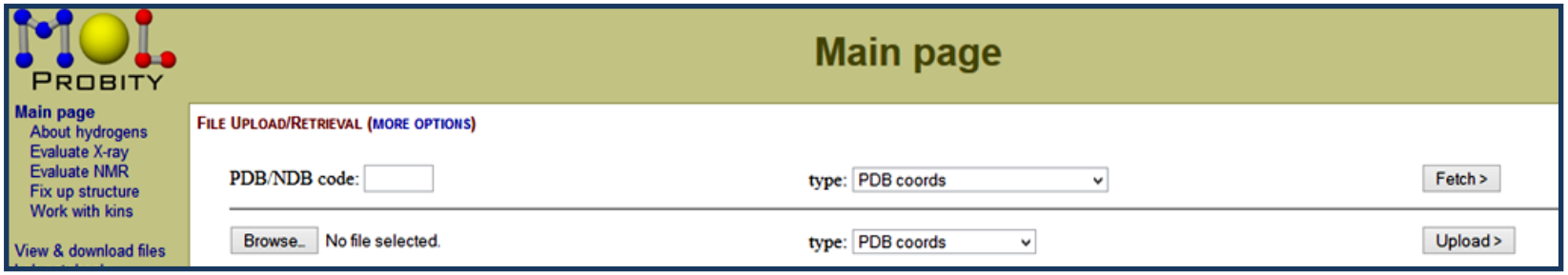
Model Evaluation



reflecting the predicted model reliability

Model Evaluation

http://molprobity.biochem.duke.edu/

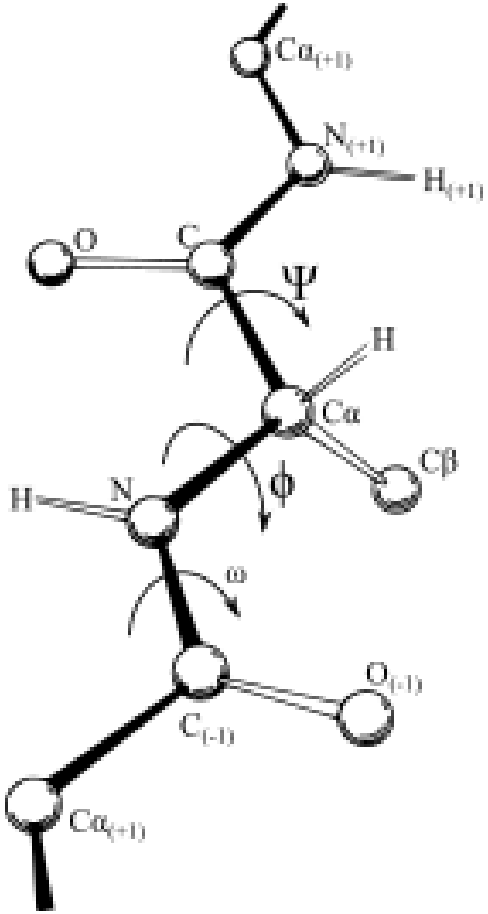


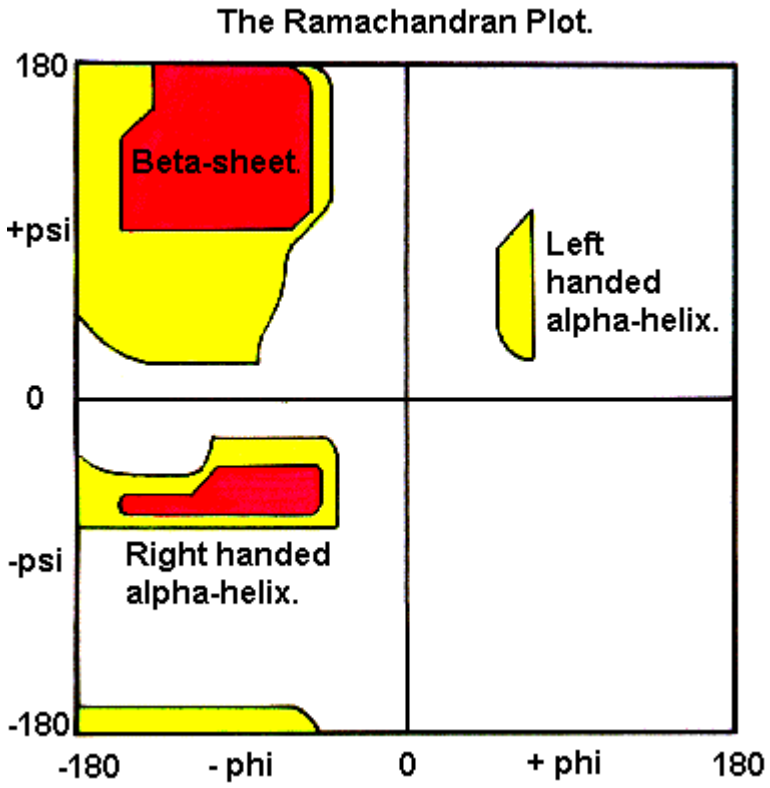
Protein Geometry	Poor rotamers	2	0.60%	Goal: <1%
	Ramachandran outliers	0	0.00%	Goal: <0.05%
	Ramachandran favored	374	97.40%	Goal: >98%
	Cβ deviations >0.25Å	0	0.00%	Goal: 0
	Bad backbone bonds:	25 / 3149	0.79%	Goal: 0%
	Bad backbone angles:	29 / 4275	0.68%	Goal: <0.1%

In the two column results, the left column gives the raw count, right column gives the percentage.

#	Alt	Res	High B	Ramachandran	Rotamer	Cβ deviation	Bond lengths	Bond angles
			Avg: 0.85	Outliers: 0 of 384	Poor rotamers: 2 of 334	Outliers: 0 of 357	Outliers: 17 of 386	Outliers: 17 of 386
A 1		ALA	0.71	-	-	0.052Å	-	-
A 2		LYS	0.6	Favored (18.81%) General / 61.3,29.8	96.1% (*mttt*) chi angles: 298.4,185.6,177.2,185.8	0.036Å	-	-
A 3		GLU	0.72	Favored (58.09%) General / -60.4,137.7	25.4% (*mp0*) chi angles: 286.3,67.3,34	0.075Å	-	-
A 4		MET	0.75	Favored (17.3%) General / -99.5,-18.6	77.8% (*mtm*) chi angles: 293.2,180.9,285.3	0.014Å	-	-
A 5		LYS	0.76	Favored (11.06%) Pre-proline / -127.7,53.7	45.7% (*mmtm*) chi angles: 290,293.6,190.6,297.9	0.071Å	-	-
A 6		PRO	0.83	Favored (81.68%) Trans-proline / -63.6,141.8	98.7% (*Cg_endo*) chi angles: 30.5	0.029Å	-	-
A 7		PHE	0.85	Favored (96.43%) Pre-proline / -65.7,137.8	54% (*t80*) chi angles: 181.9,273.7	0.069Å	-	-
A 8		PRO	0.88	Favored (3.55%) Cis-proline / -97.9,137.2	33.5% (*Cg_endo*) chi angles: 37.1	0.06Å	-	-
A 9		GLN	0.76	Favored (43.56%) General / -79.0,-28.3	26.6% (*mm100*) chi angles: 287.4,299.5,108.8	0.051Å	-	-

Ramachandran PLOT





Homology Modeling Results

Many places can go wrong:

→ Bad template - it doesn't have the same structure as the target

→ Bad alignment (a very common problem)

→ Bad loop construction

→ Bad side chain positioning

Model Refinement

→ Side chain optimization

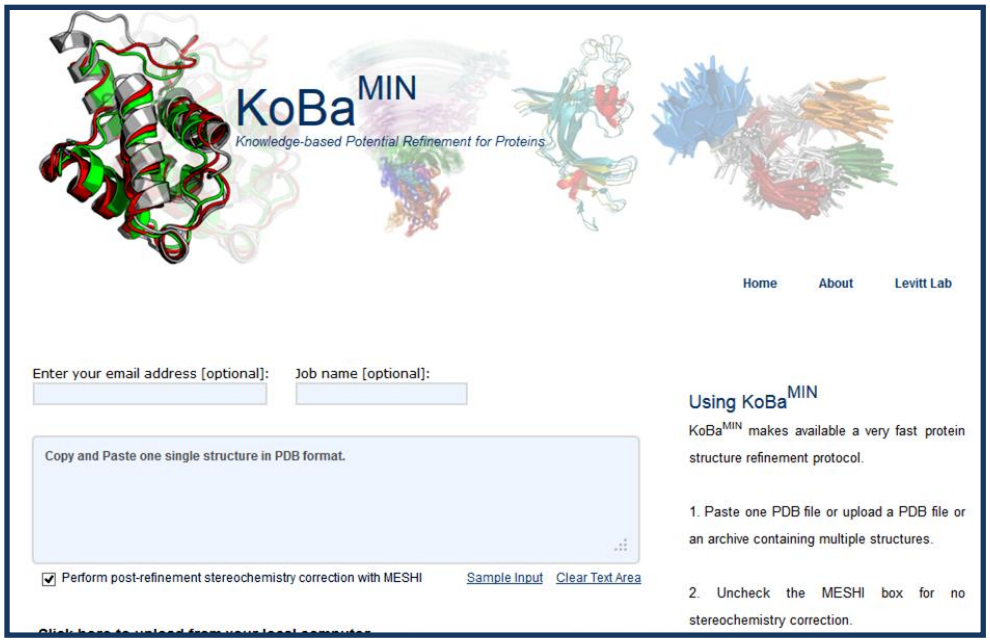
KOBAMIN server, GALAXY websever

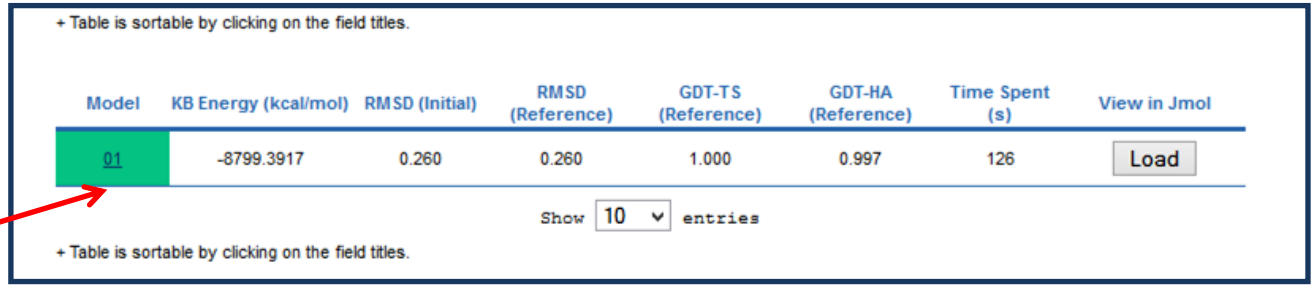
→ Loop modeling

MODLOOP

Model optimization

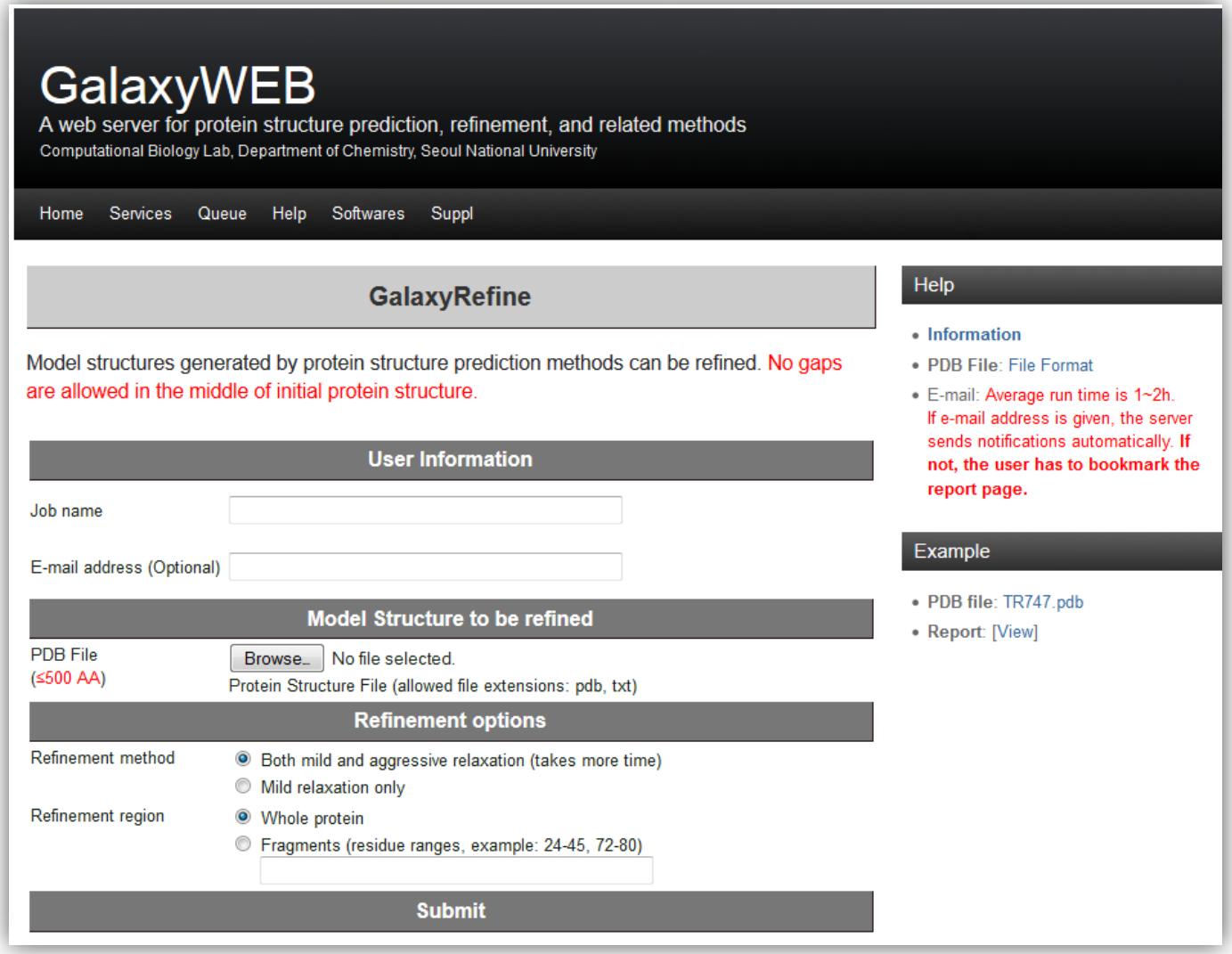
http://csb.stanford.edu/kobamin/



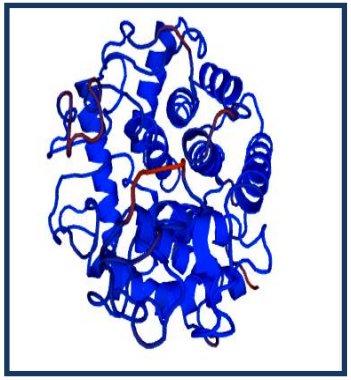


Model optimization

http://galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE



Model optimization



ModLoop

• Sali Lab Home • ModWeb • ModBase • ModEval • PCSS • FoXS • IMP • MultiFit • ModPipe •

Login · ModLoop Home · Current ModLoop queue · Help · Contact

ModLoop: Modeling of Loops in Protein Structures

ModLoop is a web server for automated modeling of loops in protein structures. The server relies on the loop modeling routine in MODELLER that predicts the loop conformations by satisfaction of spatial restraints, without relying on a database of known protein structures.

Developer:

Andras Fiser

Acknowledgements:

Ben Webb
Ursula Pieper

Andrej Sali

Version r194

General information

Email address (optional)

Modeller license key

Upload coordinate file — Browse... No file selected.

Enter loop segments

Name your model — loop

Process Reset

Key

Residues for model

Excercise

Generate 3D model using SWISS-PDB and evaluate your model