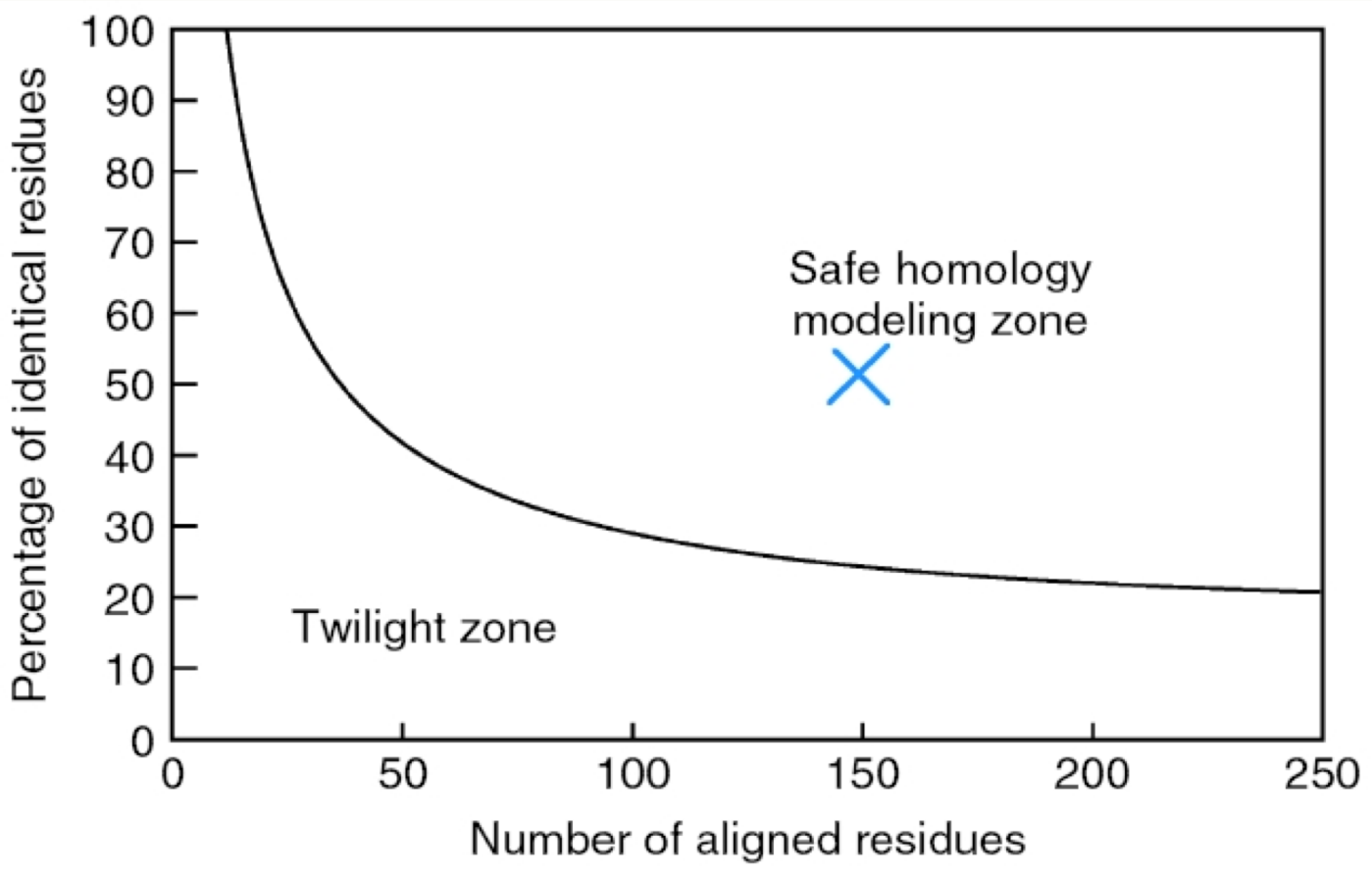
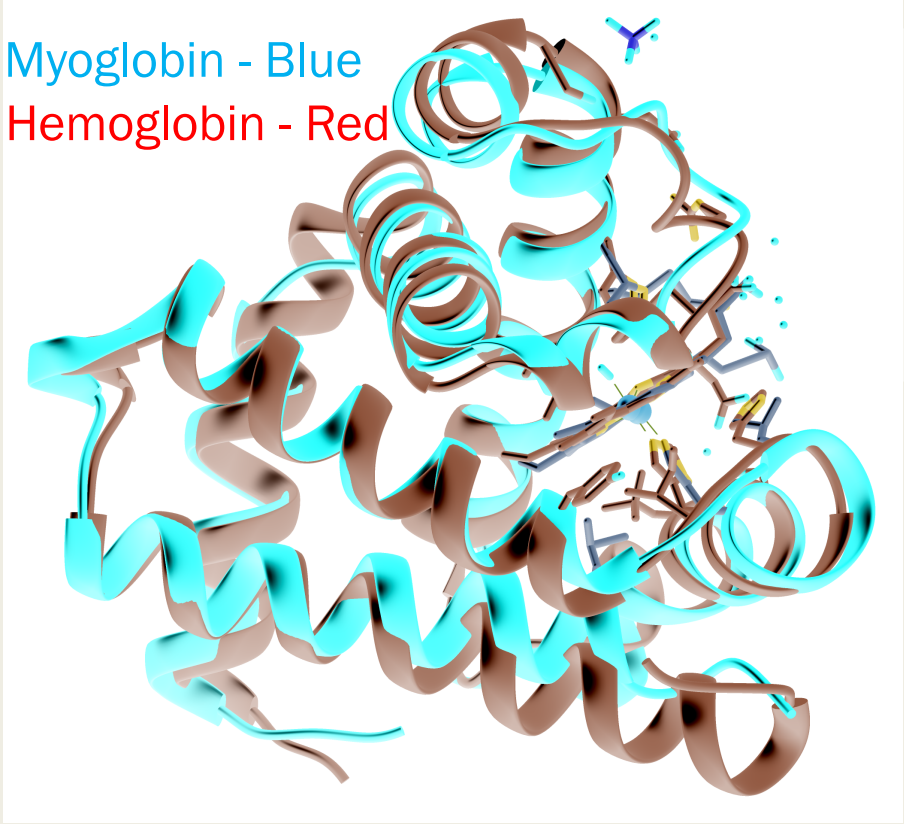
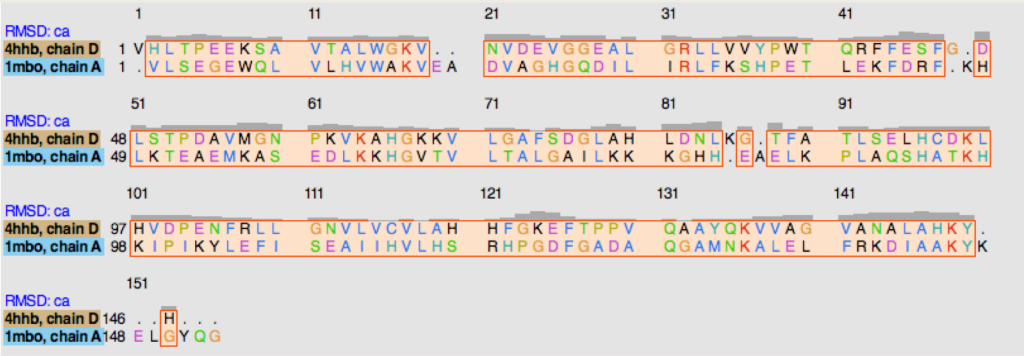
Homology modelling best practices



There are exceptions!



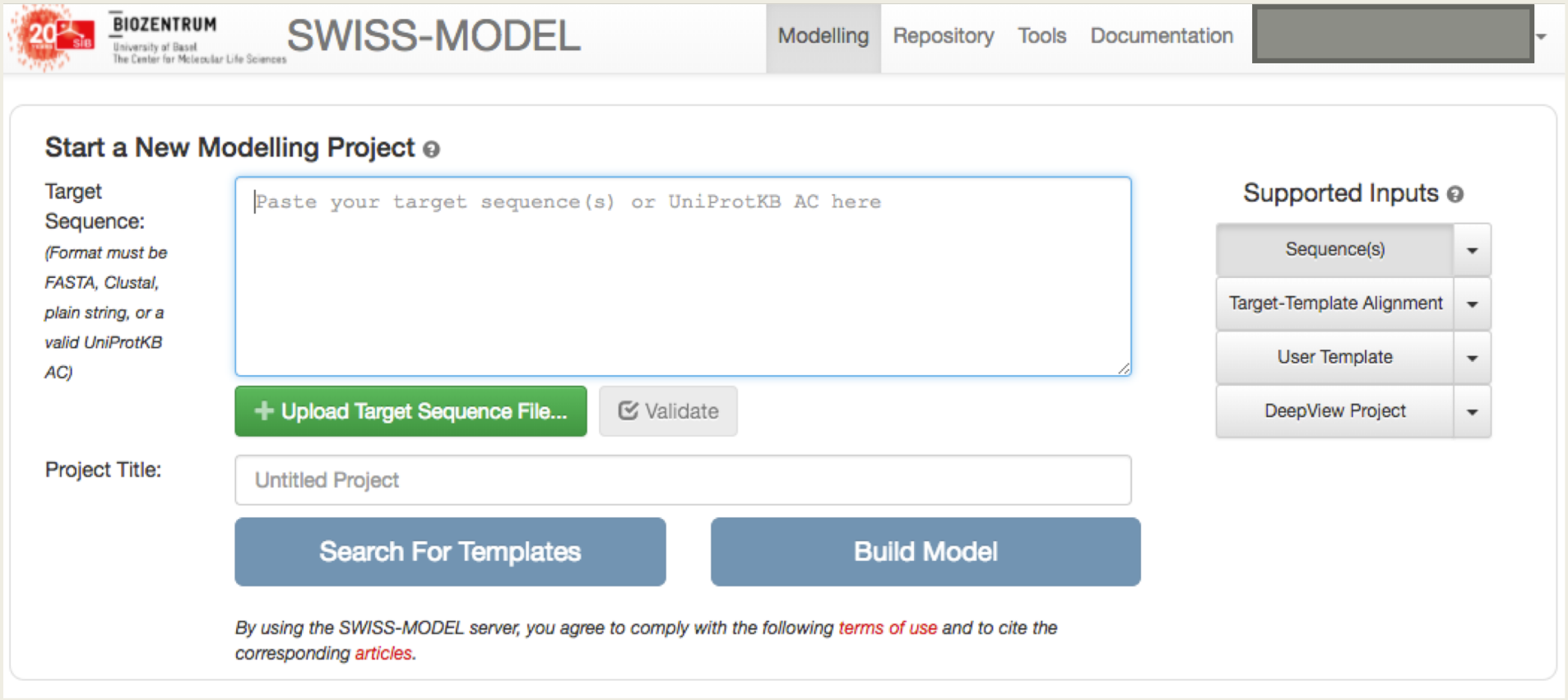


Sequence similarity is <25%
But identity is high.

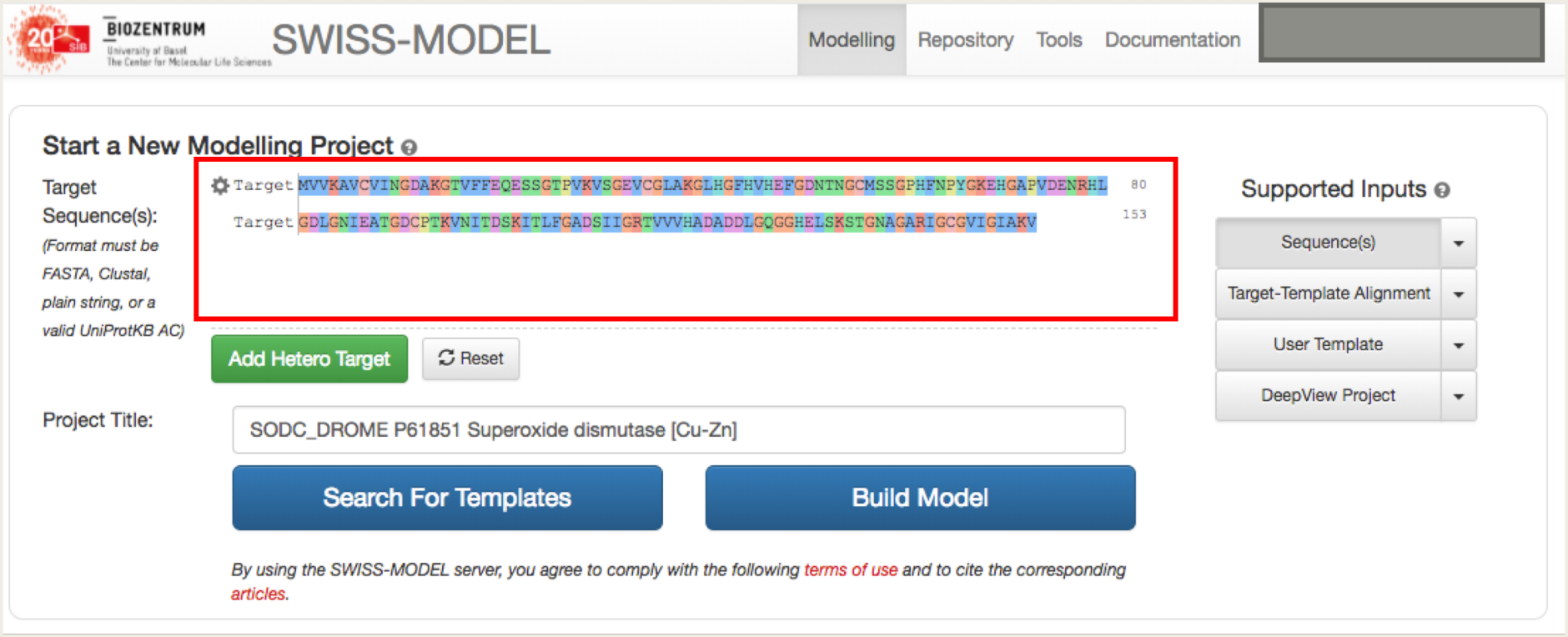
Online modelling software (mostly automated)

- https://swissmodel.expasy.org/ -We will be using this today.
- https://zhanglab.ccmb.med.umich.edu/I-TASSER/ - One of the highest ranked for accuracy of produced models.
- http://robetta.bakerlab.org/ - Online version of ROSETTA protein design suite (still in Beta, long wait times).
- https://salilab.org/modeller/ - Very popular and a lot of webservers run this.
- https://modbase.compbio.ucsf.edu/modweb/ - webserver version of modeller.
- http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index – works by recognition of conserved folds and also attempts to predict function.

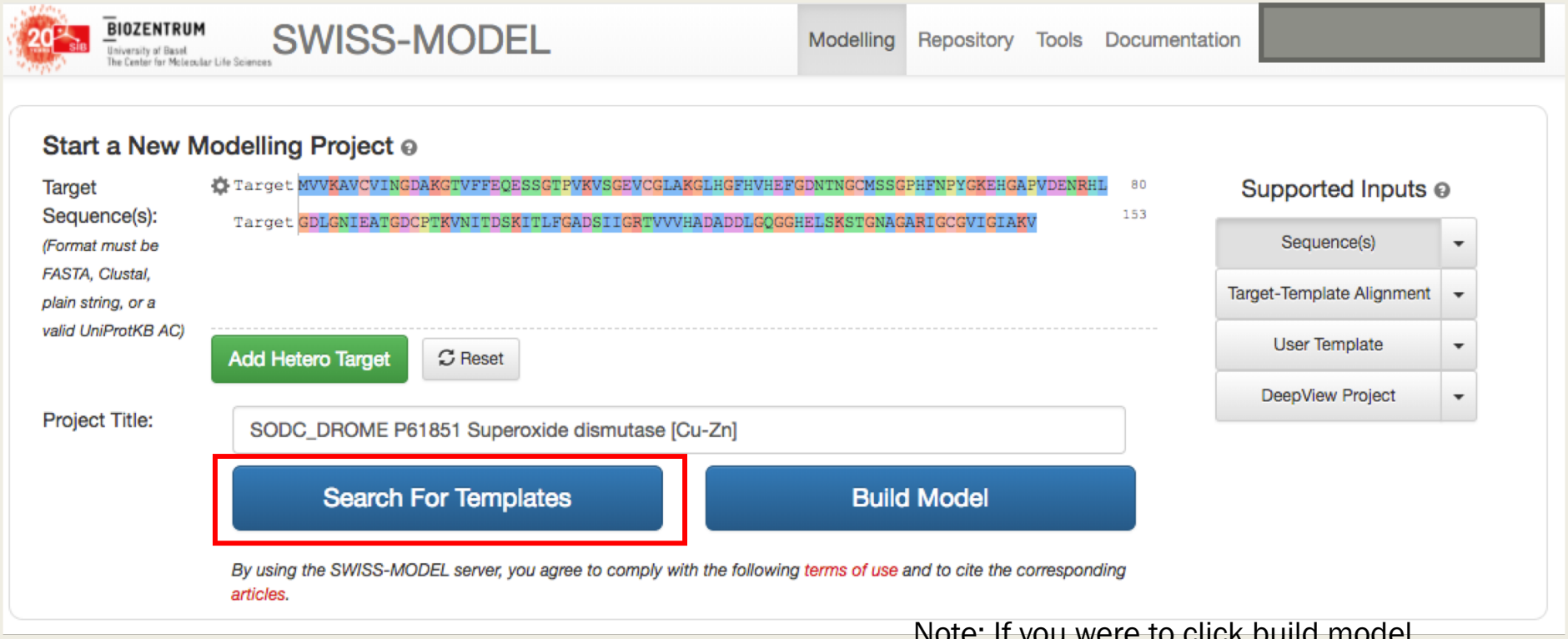
An example using SwissModel (automated mode)



An example using SwissModel (automated mode)



An example using SwissModel (automated mode)



Note: If you were to click build model SwissModel would automatically select the template

An example using SwissModel (automated mode)

SODC_DROME P61851 Superoxide dismutase [Cu-Zn] Created: today at 02:16

Summary | Templates 50 | Models

Template Results

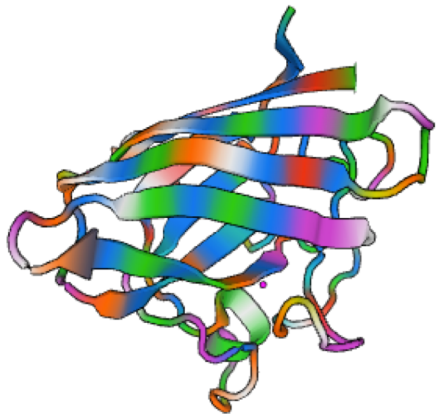
Templates | Quaternary Structure | Sequence Similarity | Alignment of Selected Templates

More

Build Models 1

Clear Selection

Sort	Name	Title	Coverage	GMQE	QSQE	Identity	Method	Oligo State	Ligands
☐	3gtv.1.A	Superoxide dismutase [Cu-Zn]		0.80	0.94	60.26	X-ray, 2.2Å	homo-dimer ✓	2 x ZN
☑	3gtv.1.A	Superoxide dismutase [Cu-Zn]		0.82	0.93	61.33	X-ray, 2.2Å	homo-dimer ✓	2 x ZN
☐	1n19.1.A	Superoxide Dismutase [Cu-Zn]		0.81	0.92	61.18	X-ray, 1.9Å	homo-dimer ✓	2 x ZN, 2 x CU1
☐	4b3e.1.A	SUPEROXIDE DISMUTASE [CU-ZN]		0.79	0.91	61.18	X-ray, 2.1Å	homo-dimer ✓	2 x ZN, 2 x CU
☐	4b3e.1.A	SUPEROXIDE DISMUTASE [CU-ZN]		0.81	0.91	61.84	X-ray, 2.1Å	homo-dimer ✓	2 x ZN, 2 x CU



An example using SwissModel (automated mode)

SODC_DROME P61851 Superoxide dismutase [Cu-Zn] Created: today at 02:16

Summary | Templates 50 | Models

GMQE – General Model Quality Estimate (close to 1 is better)
QSQE – Quaternary Structure Quality Estimate (only relevant if your model has multiple subunits)

Template Results

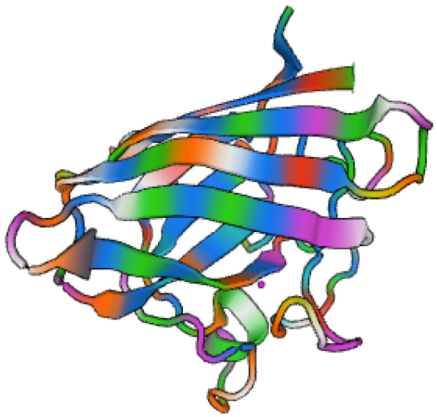
Templates | Quaternary Structure | Sequence Similarity | Alignment of Selected Templates

More

Build Models 1

Clear Selection

Sort	Name	Title	Coverage	GMQE	QSQE	Identity	Method	Oligo State	Ligands
☐	3gtv.1.A	Superoxide dismutase [Cu-Zn]		0.80	0.94	60.26	X-ray, 2.2Å	homo-dimer ✓	2 x ZN
☑	3gtv.1.A	Superoxide dismutase [Cu-Zn]		0.82	0.93	61.33	X-ray, 2.2Å	homo-dimer ✓	2 x ZN
☐	1n19.1.A	Superoxide Dismutase [Cu-Zn]		0.81	0.92	61.18	X-ray, 1.9Å	homo-dimer ✓	2 x ZN, 2 x CU1
☐	4b3e.1.A	SUPEROXIDE DISMUTASE [CU-ZN]		0.79	0.91	61.18	X-ray, 2.1Å	homo-dimer ✓	2 x ZN, 2 x CU
☐	4b3e.1.A	SUPEROXIDE DISMUTASE [CU-ZN]		0.81	0.91	61.84	X-ray, 2.1Å	homo-dimer ✓	2 x ZN, 2 x CU



An example using SwissModel (automated mode)

SODC_DROME P61851 Superoxide dismutase [Cu-Zn] Created: today at 02:16

Summary | Templates 50 | Models

Template Results

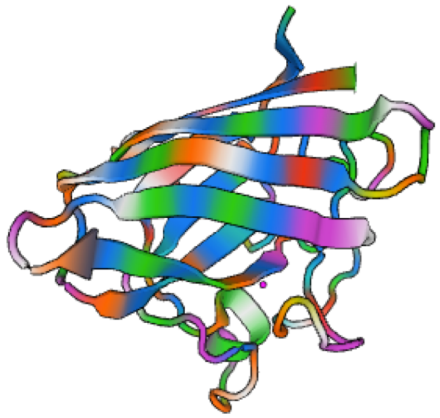
Templates | Quaternary Structure | Sequence Similarity | Alignment of Selected Templates

More

Build Models 1

Clear Selection

Sort	Name	Title	Coverage	GMQE	QSQE	Identity	Method	Oligo State	Ligands
☐	3gtv.1.A	Superoxide dismutase [Cu-Zn]		0.80	0.94	60.26	X-ray, 2.2Å	homo-dimer ✓	2 x ZN
☑	3gtv.1.A	Superoxide dismutase [Cu-Zn]		0.82	0.93	61.33	X-ray, 2.2Å	homo-dimer ✓	2 x ZN
☐	1n19.1.A	Superoxide Dismutase [Cu-Zn]		0.81	0.92	61.18	X-ray, 1.9Å	homo-dimer ✓	2 x ZN, 2 x CU1
☐	4b3e.1.A	SUPEROXIDE DISMUTASE [CU-ZN]		0.79	0.91	61.18	X-ray, 2.1Å	homo-dimer ✓	2 x ZN, 2 x CU
☐	4b3e.1.A	SUPEROXIDE DISMUTASE [CU-ZN]		0.81	0.91	61.84	X-ray, 2.1Å	homo-dimer ✓	2 x ZN, 2 x CU



A side note on template selection

- Your final model is only ever as trustworthy as your template.
- Fortunately SwissModel provides you with a PDB code for the template.
- The PDB has its own set of validation metrics that we can use to asses the model we wish to use.

An example using SwissModel (automated mode)

SODC_DROME P61851 Superoxide dismutase [Cu-Zn] Created: today at 02:16

Summary | Templates 50 | Models

Template Results

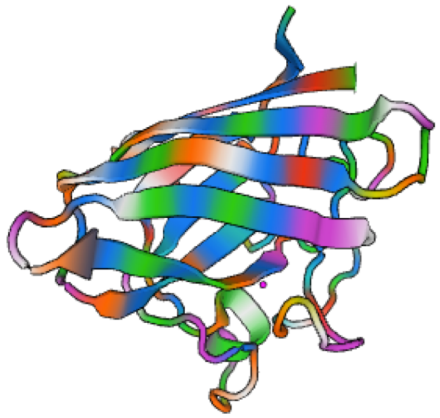
Templates | Quaternary Structure | Sequence Similarity | Alignment of Selected Templates

More

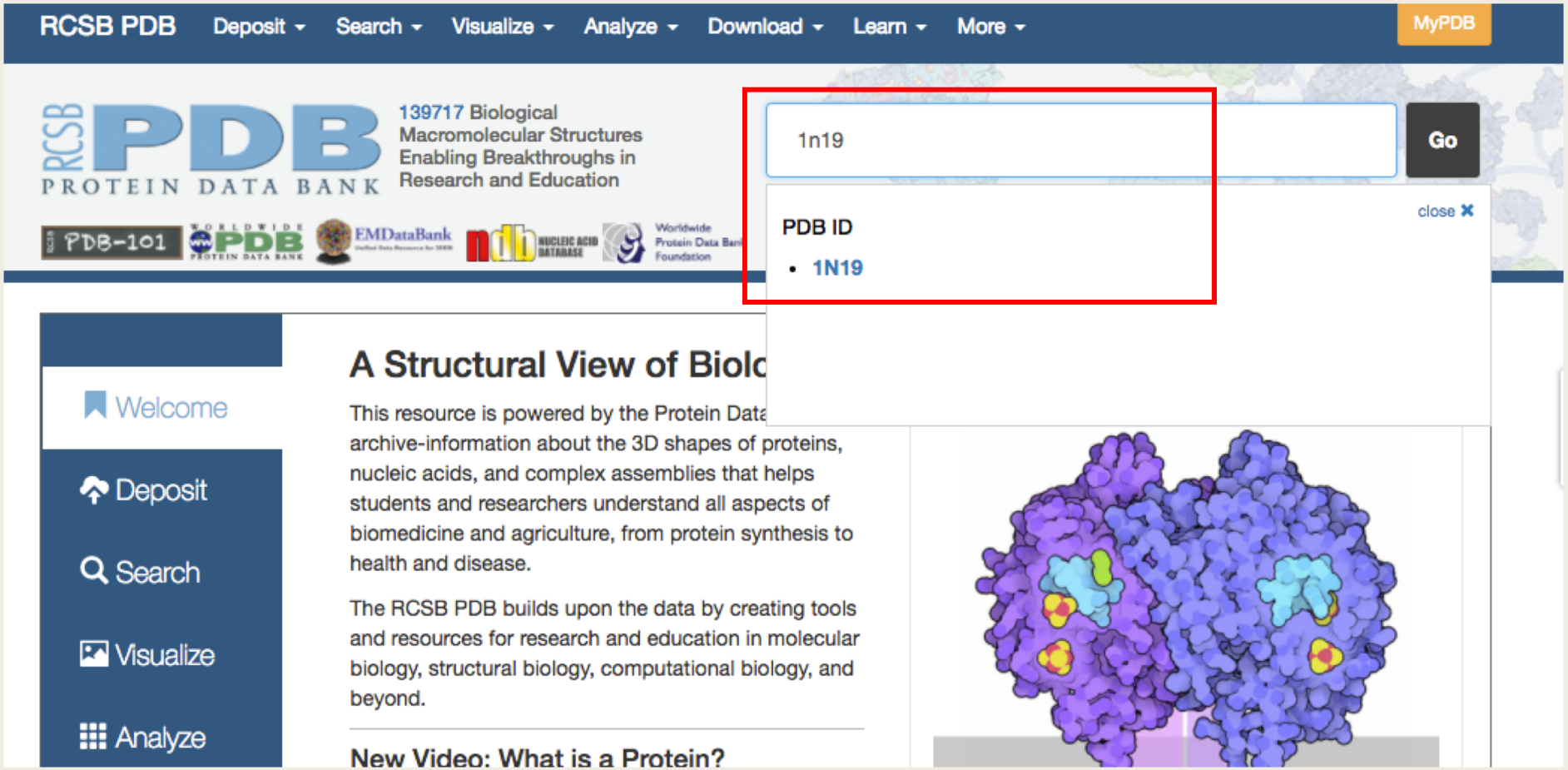
Build Models 1

Clear Selection

Sort	Name	Title	Coverage	GMQE	QSQE	Identity	Method	Oligo State	Ligands
☐	3gtv.1.A	Superoxide dismutase [Cu-Zn]		0.80	0.94	60.26	X-ray, 2.2Å	homo-dimer ✓	2 x ZN
☑	3gtv.1.A	Superoxide dismutase [Cu-Zn]		0.82	0.93	61.33	X-ray, 2.2Å	homo-dimer ✓	2 x ZN
☐	1n19.1.A	Superoxide Dismutase [Cu-Zn]		0.81	0.92	61.18	X-ray, 1.9Å	homo-dimer ✓	2 x ZN, 2 x CU1
☐	4b3e.1.A	SUPEROXIDE DISMUTASE [CU-ZN]		0.79	0.91	61.18	X-ray, 2.1Å	homo-dimer ✓	2 x ZN, 2 x CU
☐	4b3e.1.A	SUPEROXIDE DISMUTASE [CU-ZN]		0.81	0.91	61.84	X-ray, 2.1Å	homo-dimer ✓	2 x ZN, 2 x CU



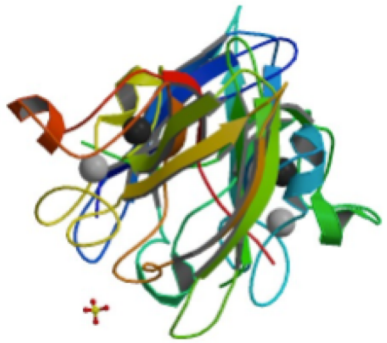
Checking the structure in the PDB



Checking the structure in the PDB

Structure Summary | 3D View | Annotations | Sequence | Sequence Similarity | Structure Similarity | Experiment

Biological Assembly 1



3D View: Structure | Electron Density | Ligand Interaction

Standalone Viewers
Protein Workshop | Ligand Explorer

Global Symmetry: Cyclic - C2 (3D View)
Global Stoichiometry: Homo 2-mer - A2

Display Files | Download Files

1N19

Structure of the HSOD A4V mutant

DOI: 10.2210/pdb1N19/pdb

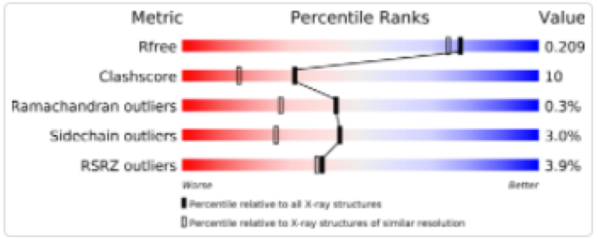
Classification: OXIDOREDUCTASE
Organism(s): Homo sapiens
Expression System: Escherichia coli
Mutation(s): 3

Deposited: 2002-10-16 **Released:** 2002-11-27
Deposition Author(s): Cardoso, R.M.F., Thayer, M.M., DiDonato, M., Lo, T.P., Bruns, C.K., Getzoff, E.D., Tainer, J.A.

Experimental Data Snapshot

Method: X-RAY DIFFRACTION
Resolution: 1.86 Å
R-Value Free: 0.260
R-Value Work: 0.207

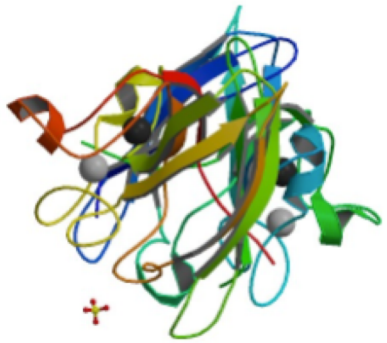
wwPDB Validation 3D Report | Full Report



Checking the structure in the PDB

Structure Summary | 3D View | Annotations | Sequence | Sequence Similarity | Structure Similarity | Experiment

Biological Assembly 1



3D View: Structure | Electron Density | Ligand Interaction

Standalone Viewers
Protein Workshop | Ligand Explorer

Global Symmetry: Cyclic - C2 (3D View)
Global Stoichiometry: Homo 2-mer - A2

Display Files | Download Files

1N19

Structure of the HSOD A4V mutant

DOI: 10.2210/pdb1N19/pdb

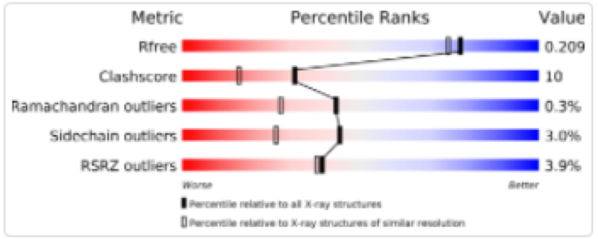
Classification: OXIDOREDUCTASE
Organism(s): Homo sapiens
Expression System: Escherichia coli
Mutation(s): 3

Deposited: 2002-10-16 **Released:** 2002-11-27
Deposition Author(s): Cardoso, R.M.F., Thayer, M.M., DiDonato, M., Lo, T.P., Bruns, C.K., Getzoff, E.D., Tainer, J.A.

Experimental Data Snapshot

Method: X-RAY DIFFRACTION
Resolution: 1.86 Å
R-Value Free: 0.260
R-Value Work: 0.207

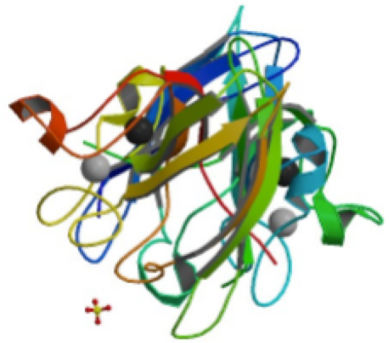
wwPDB Validation 3D Report | Full Report



Checking the structure in the PDB

Structure Summary | 3D View | Annotations | Sequence | Sequence Similarity | Structure Similarity | Experiment

Biological Assembly 1



3D View: Structure | Electron Density | Ligand Interaction

Standalone Viewers
Protein Workshop | Ligand Explorer

Global Symmetry: Cyclic - C2 (3D View)
Global Stoichiometry: Homo 2-mer - A2

Display Files | Download Files

1N19

Structure of the HSOD A4V mutant

DOI: 10.2210/pdb1N19/pdb

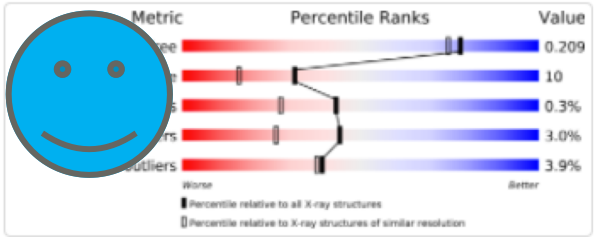
Classification: OXIDOREDUCTASE
Organism(s): Homo sapiens
Expression System: Escherichia coli
Mutation(s): 3

Deposited: 2002-10-16 **Released:** 2002-11-27
Deposition Author(s): Cardoso, R.M.F., Thayer, M.M., DiDonato, M., Lo, T.P., Bruns, C.K., Getzoff, E.D., Tainer, J.A.

Experimental Data Snapshot

Method: X-RAY DIFFRACTION
Resolution: 1.86 Å
R-Value Free: 0.260
R-Value Work: 0.207

wwPDB Validation 3D Report | Full Report



An example using SwissModel (automated mode)

SODC_DROME P61851 Superoxide dismutase [Cu-Zn] Created: today at 02:16

Summary | Templates 50 | Models

Template Results

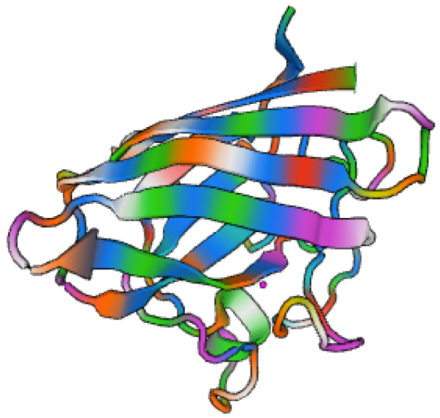
Templates | Quaternary Structure | Sequence Similarity | Alignment of Selected Templates

More

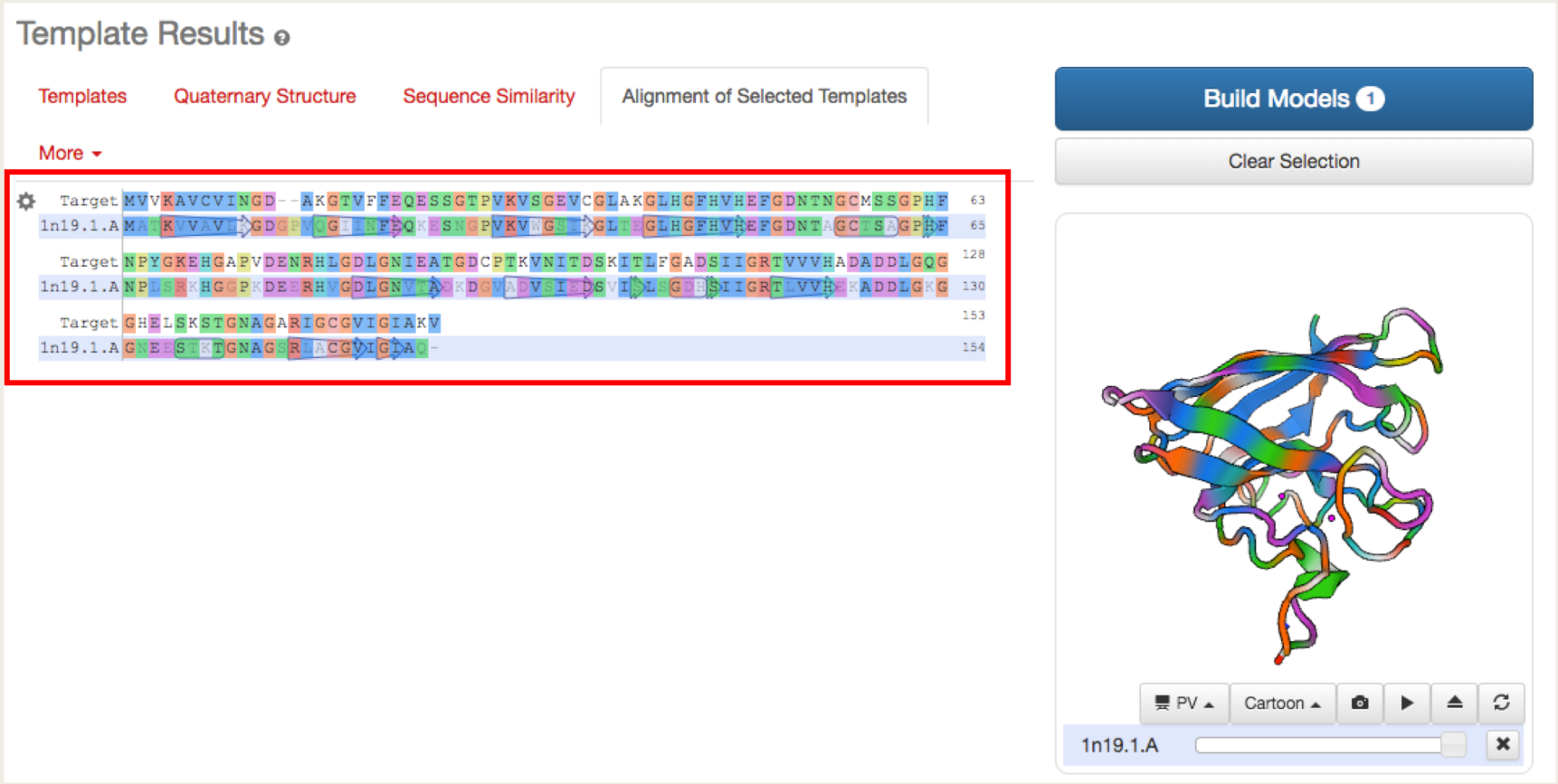
Build Models 1

Clear Selection

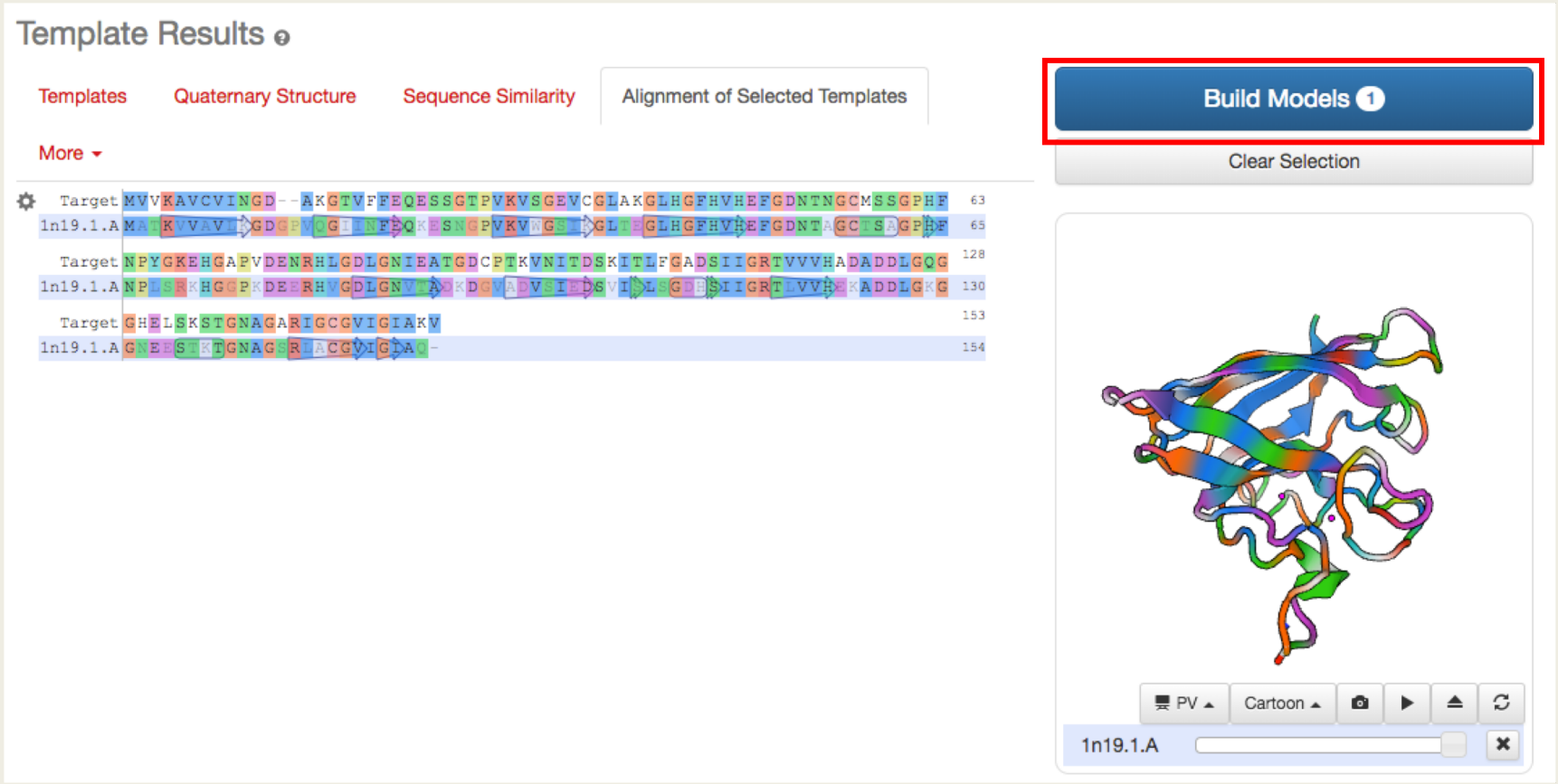
Sort	Name	Title	Coverage	GMQE	QSQE	Identity	Method	Oligo State	Ligands
☐	3gtv.1.A	Superoxide dismutase [Cu-Zn]		0.80	0.94	60.26	X-ray, 2.2Å	homo-dimer ✓	2 x ZN
☑	3gtv.1.A	Superoxide dismutase [Cu-Zn]		0.82	0.93	61.33	X-ray, 2.2Å	homo-dimer ✓	2 x ZN
☐	1n19.1.A	Superoxide Dismutase [Cu-Zn]		0.81	0.92	61.18	X-ray, 1.9Å	homo-dimer ✓	2 x ZN, 2 x CU1
☐	4b3e.1.A	SUPEROXIDE DISMUTASE [CU-ZN]		0.79	0.91	61.18	X-ray, 2.1Å	homo-dimer ✓	2 x ZN, 2 x CU
☐	4b3e.1.A	SUPEROXIDE DISMUTASE [CU-ZN]		0.81	0.91	61.84	X-ray, 2.1Å	homo-dimer ✓	2 x ZN, 2 x CU



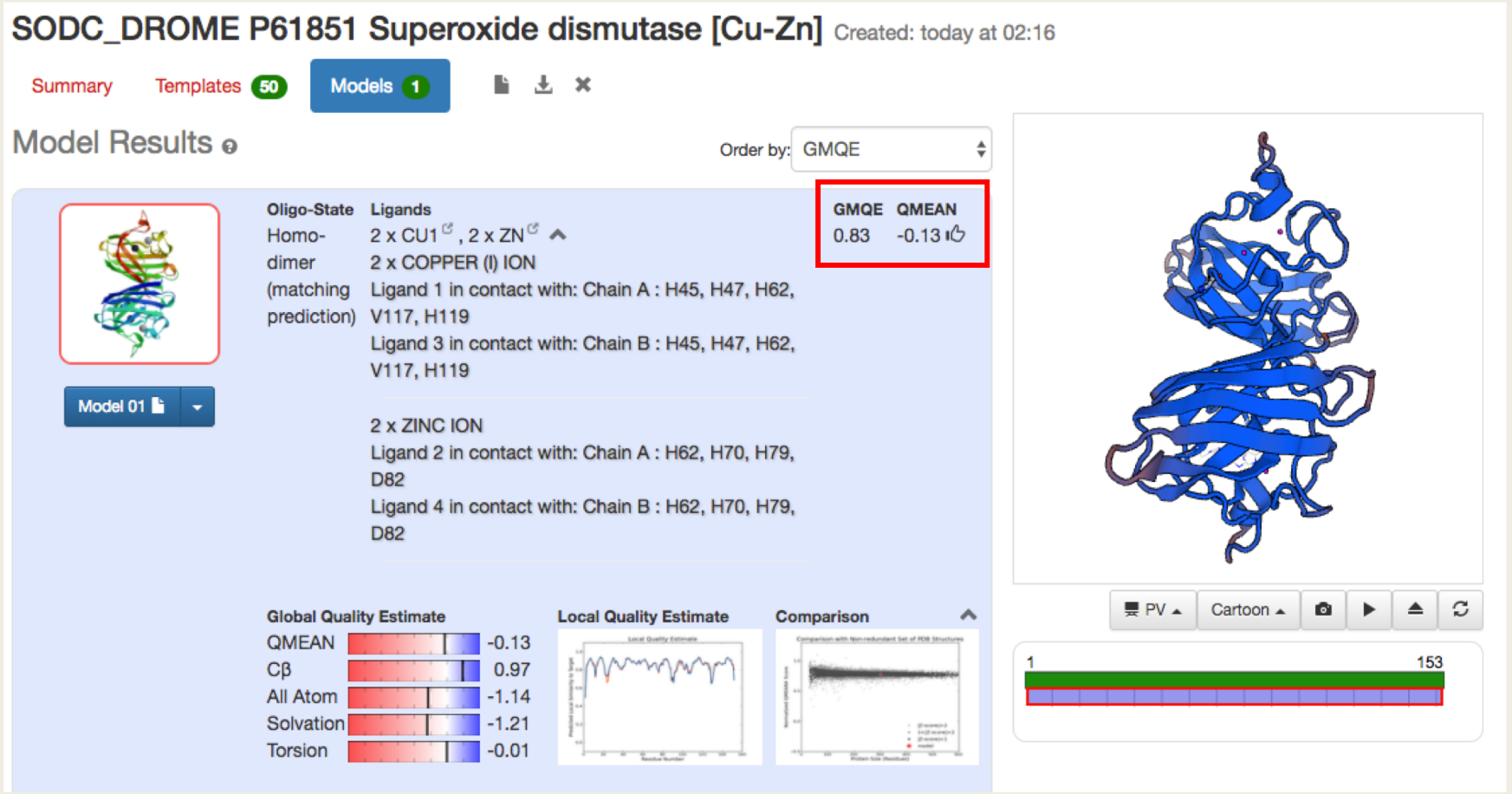
An example using SwissModel (automated mode)



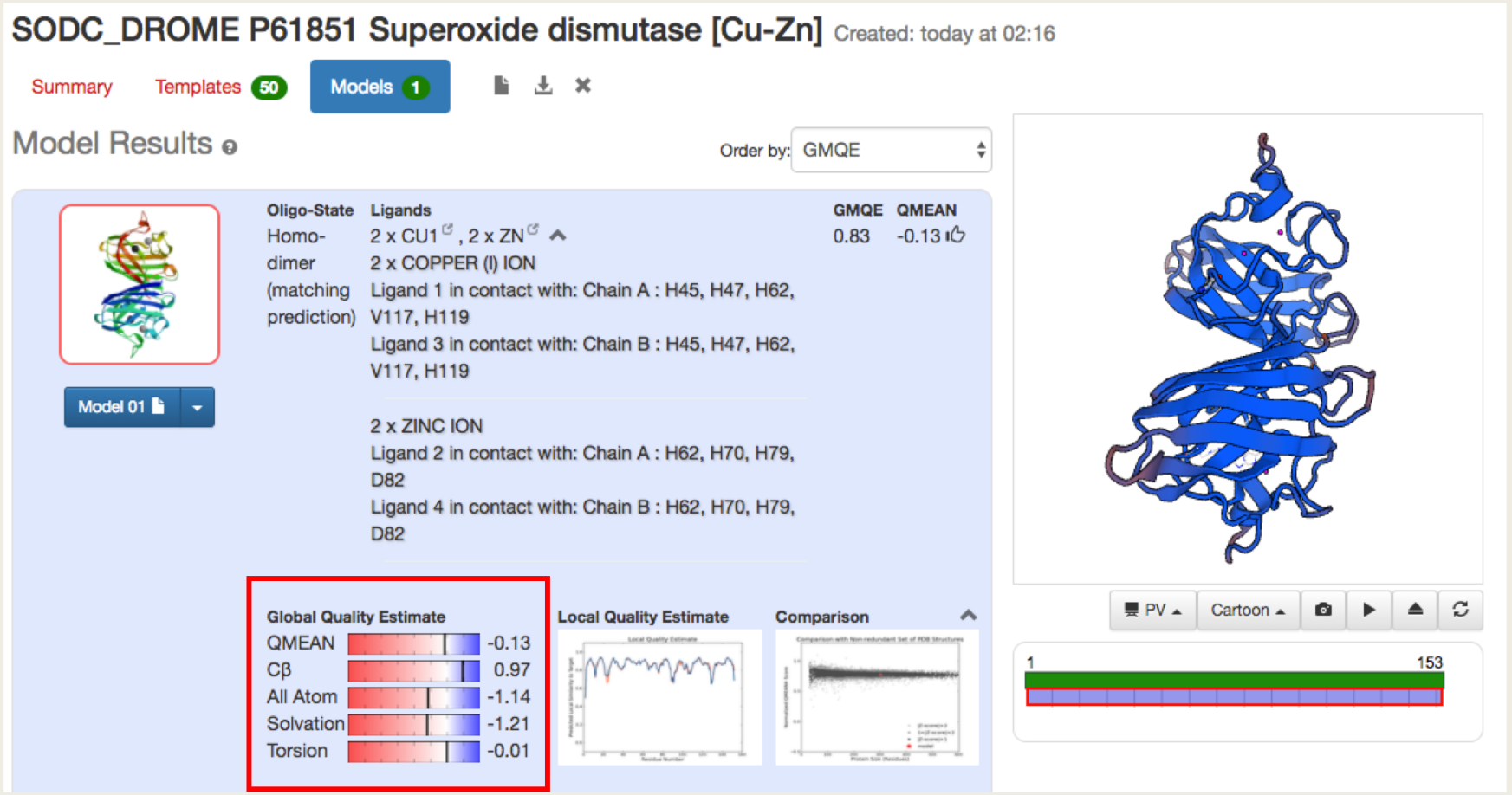
An example using SwissModel (automated mode)



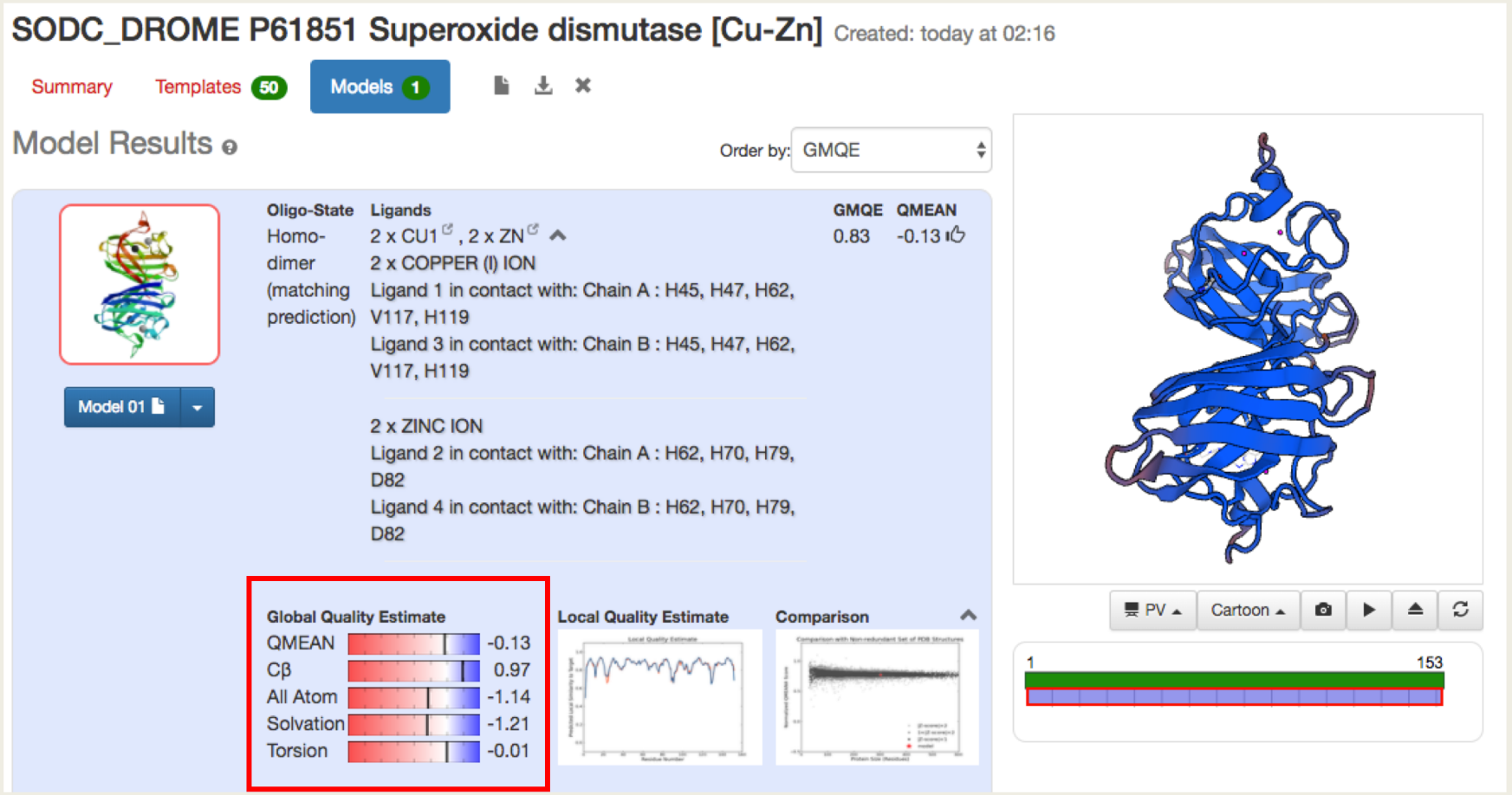
An example using SwissModel (automated mode)



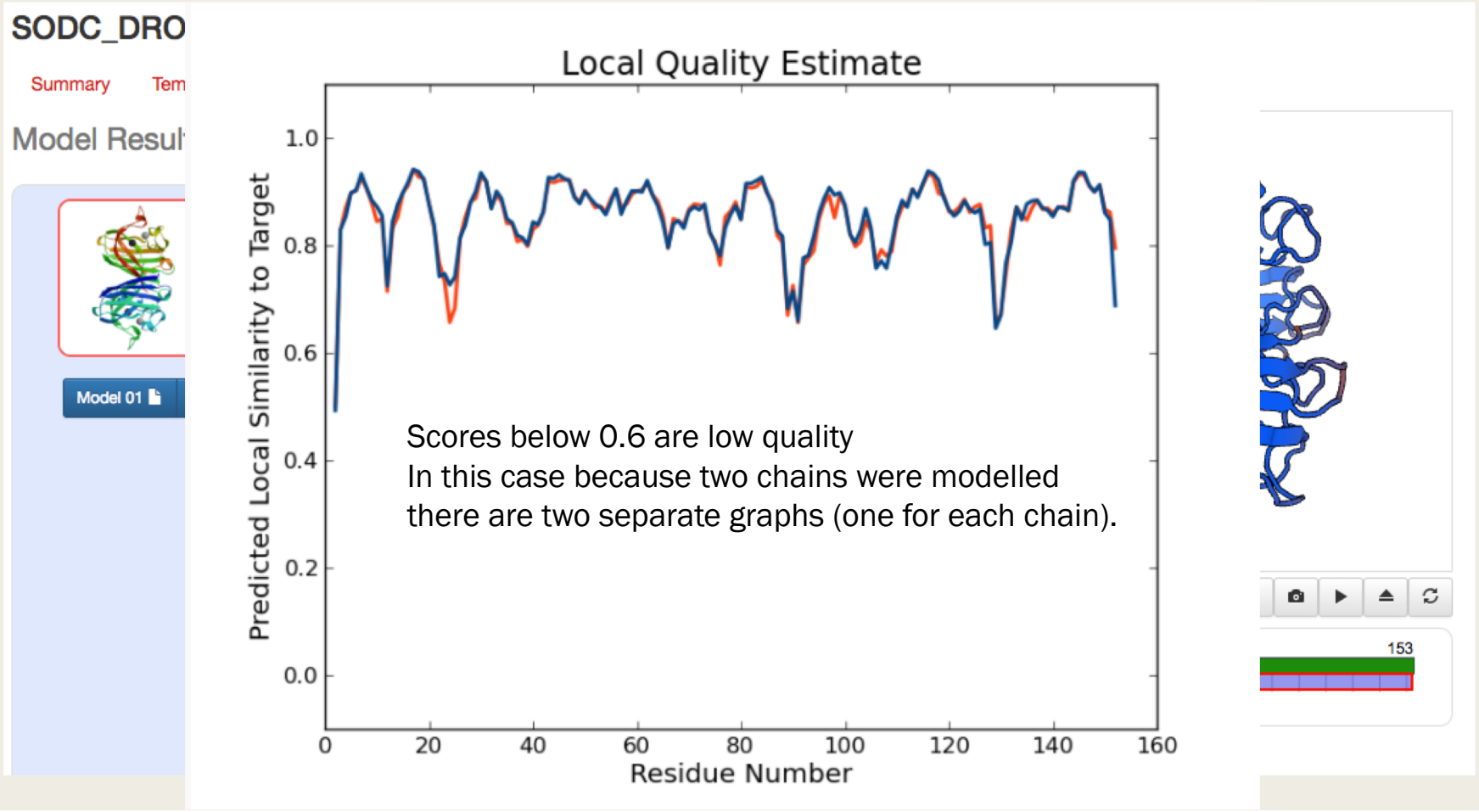
An example using SwissModel (automated mode)



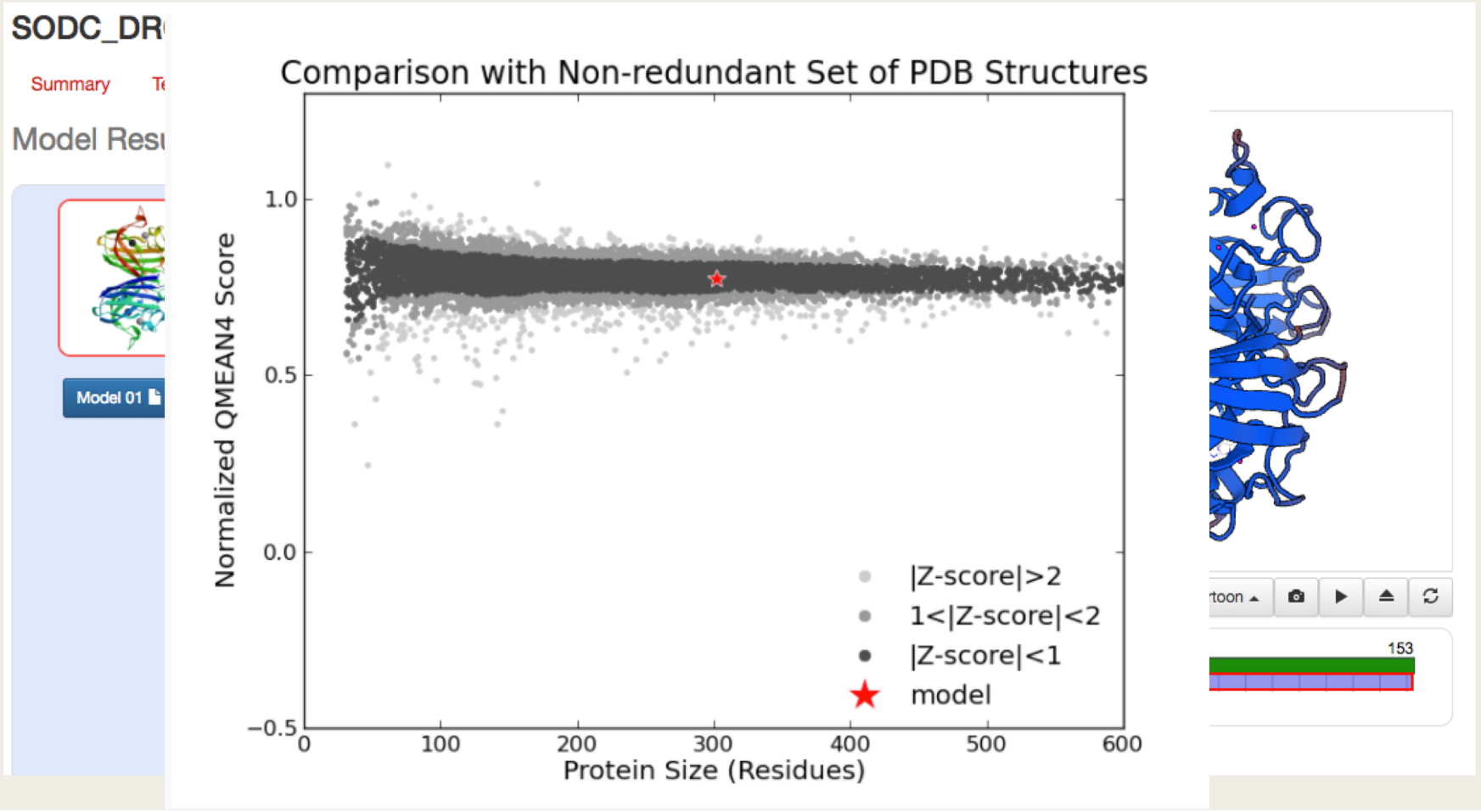
An example using SwissModel (automated mode)



An example using SwissModel (automated mode)



An example using SwissModel (automated mode)



Cross validate with several other templates if possible

Template Results

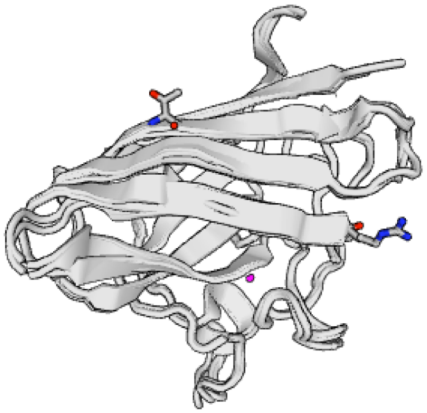
Templates | Quaternary Structure | Sequence Similarity | Alignment of Selected Templates

More

Build Models 5

Clear Selection

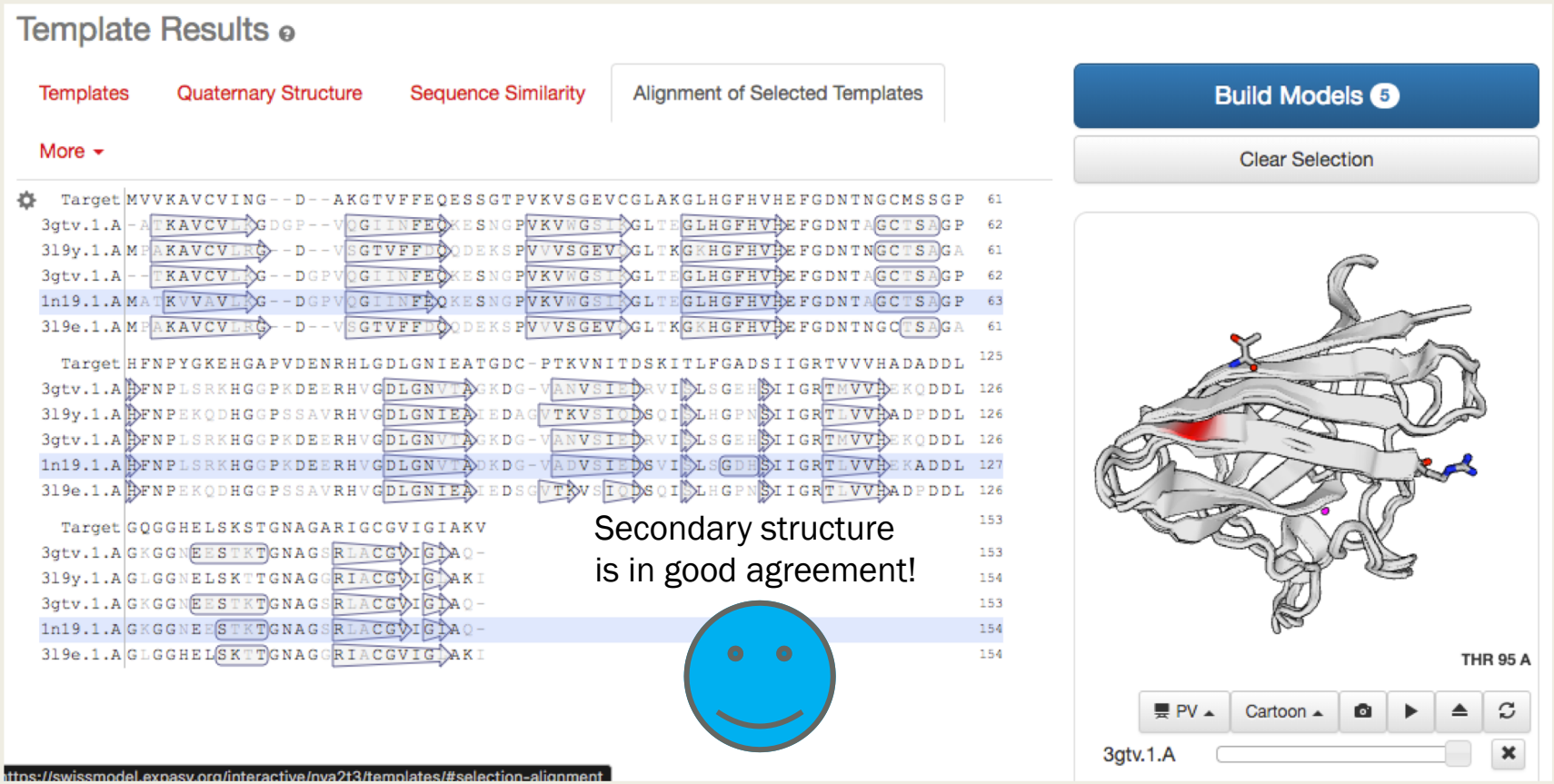
Sort	Name	Title	Coverage	GMQE	QSQE	Identity	Method	Oligo State	Ligands
☑	3gtv.1.A	Superoxide dismutase [Cu-Zn]		0.82	0.93	61.33	X-ray, 2.2Å	homo-dimer ✓	2 x ZN
☑	3l9y.1.A	Superoxide dismutase [Cu-Zn]		0.86	0.88	67.97	X-ray, 1.8Å	homo-dimer ✓	2 x ZN, 2 x CU
☑	3gtv.1.A	Superoxide dismutase [Cu-Zn]		0.80	0.94	60.26	X-ray, 2.2Å	homo-dimer ✓	2 x ZN
☑ ✓	1n19.1.A	Superoxide Dismutase [Cu-Zn]		0.81	0.92	61.18	X-ray, 1.9Å	homo-dimer ✓	2 x ZN, 2 x CU1
☑	3l9e.1.A	Superoxide dismutase [Cu-Zn]		0.85	0.87	68.63	X-ray, 2.0Å	homo-dimer ✓	2 x ZN
☐	1n18.1.A	Superoxide dismutase [Cu-Zn]		0.81	0.90	61.84	X-ray, 2.0Å	homo-dimer ✓	2 x ZN, 2 x CU1
☐	4b3e.1.A	SUPEROXIDE		0.81	0.91	61.84	X-ray, 2.1Å	homo-dimer	2 x ZN,



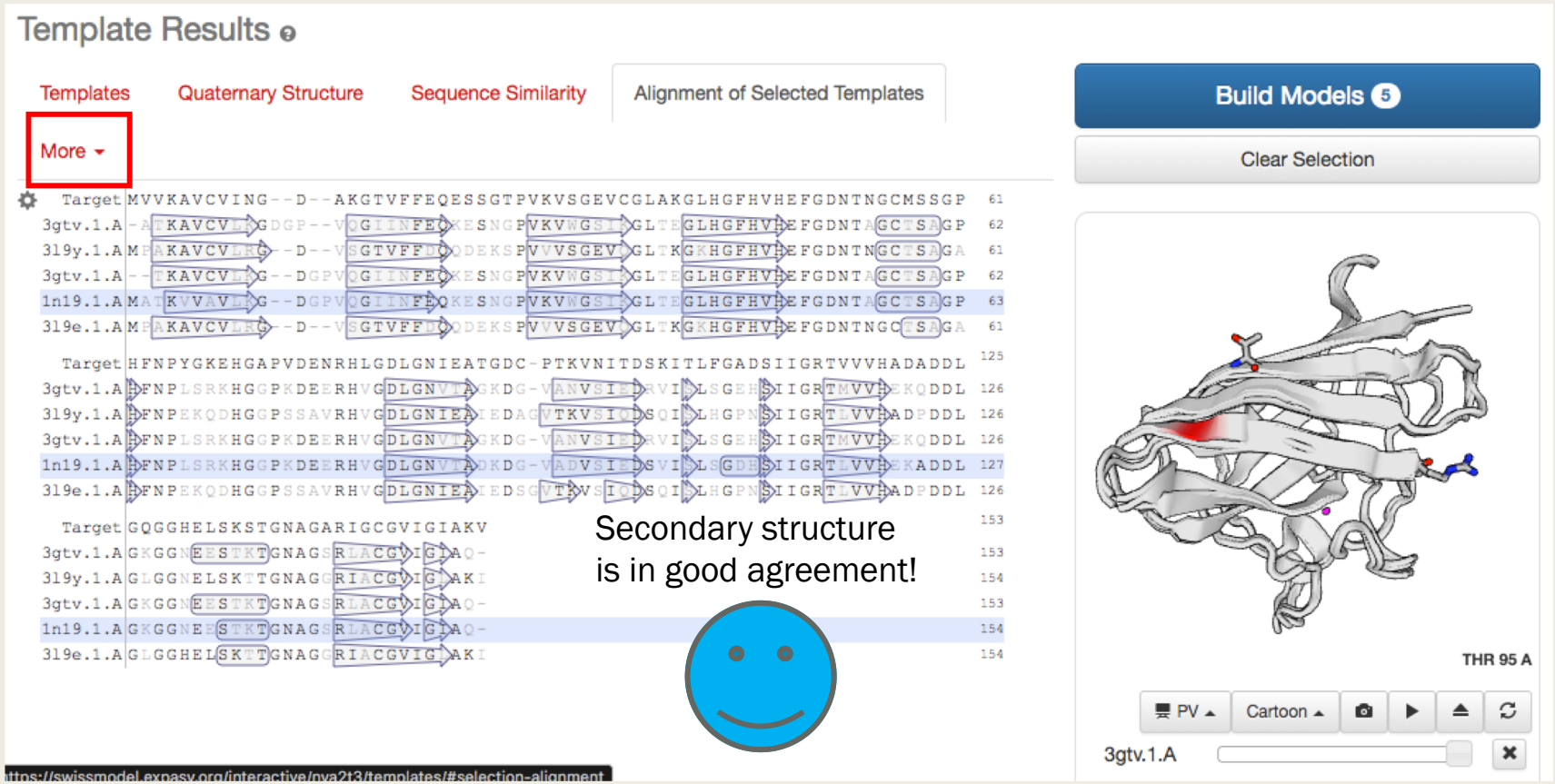
PV | Cartoon

3gtv.1.A

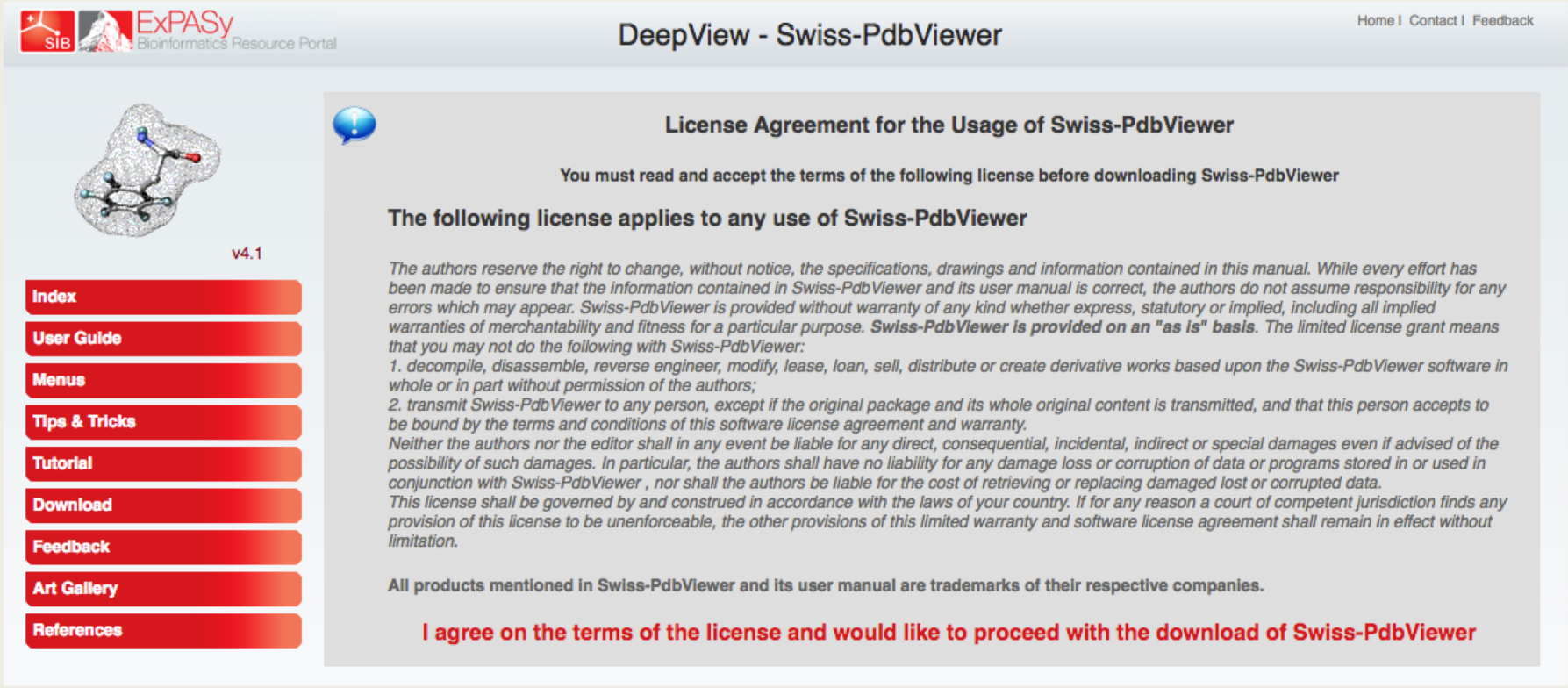
Cross validate with several other templates if possible



Adjusting alignments with DeepView

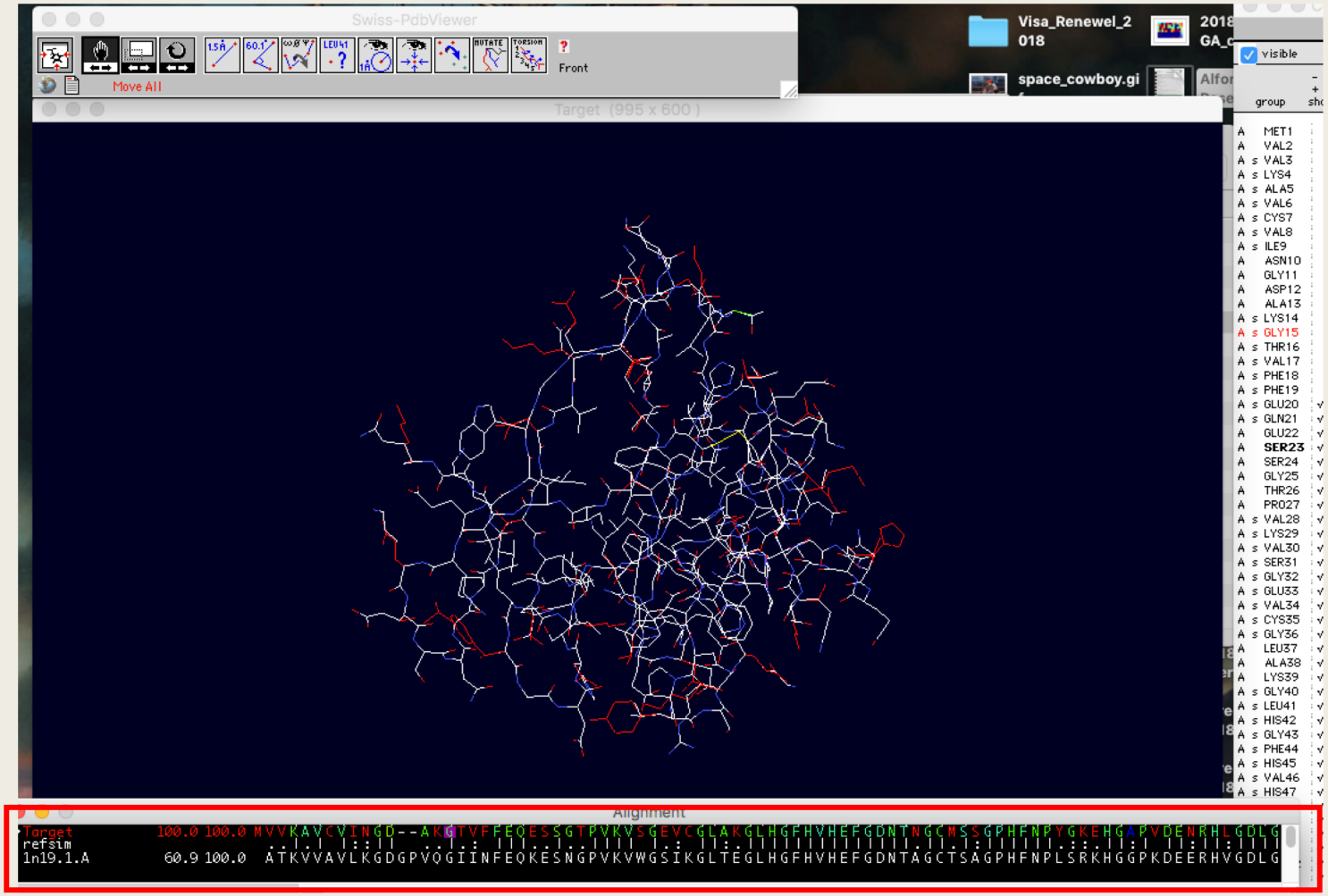


Adjusting alignments with DeepView

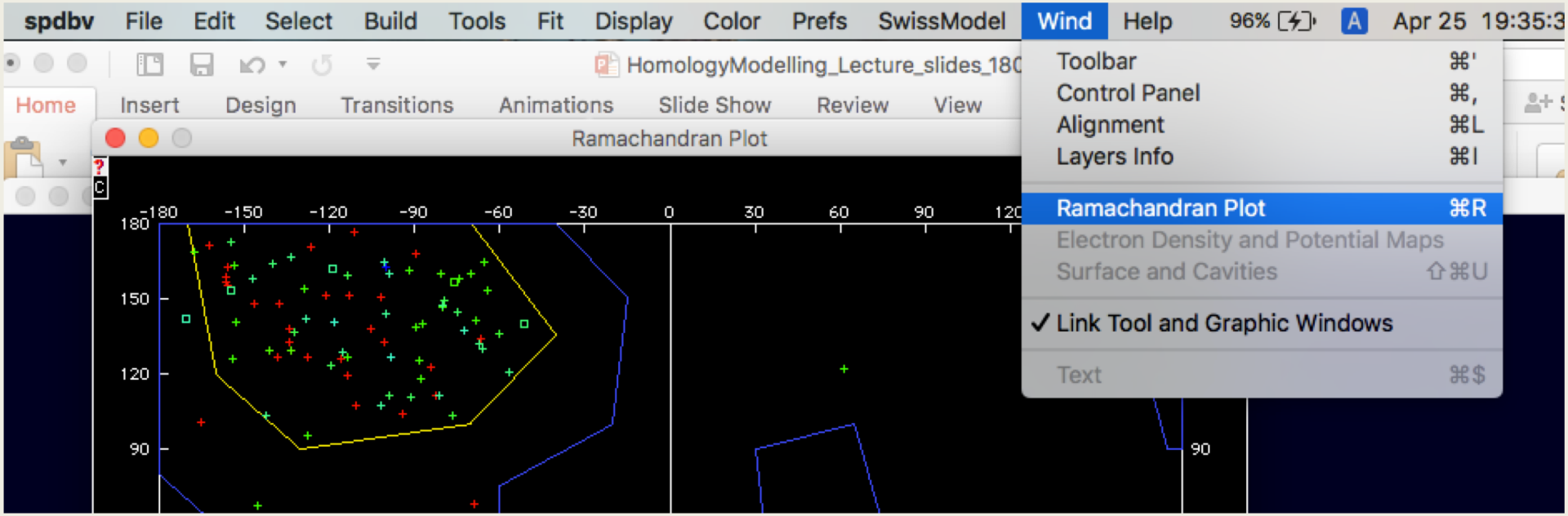


https://spdbv.vital-it.ch/

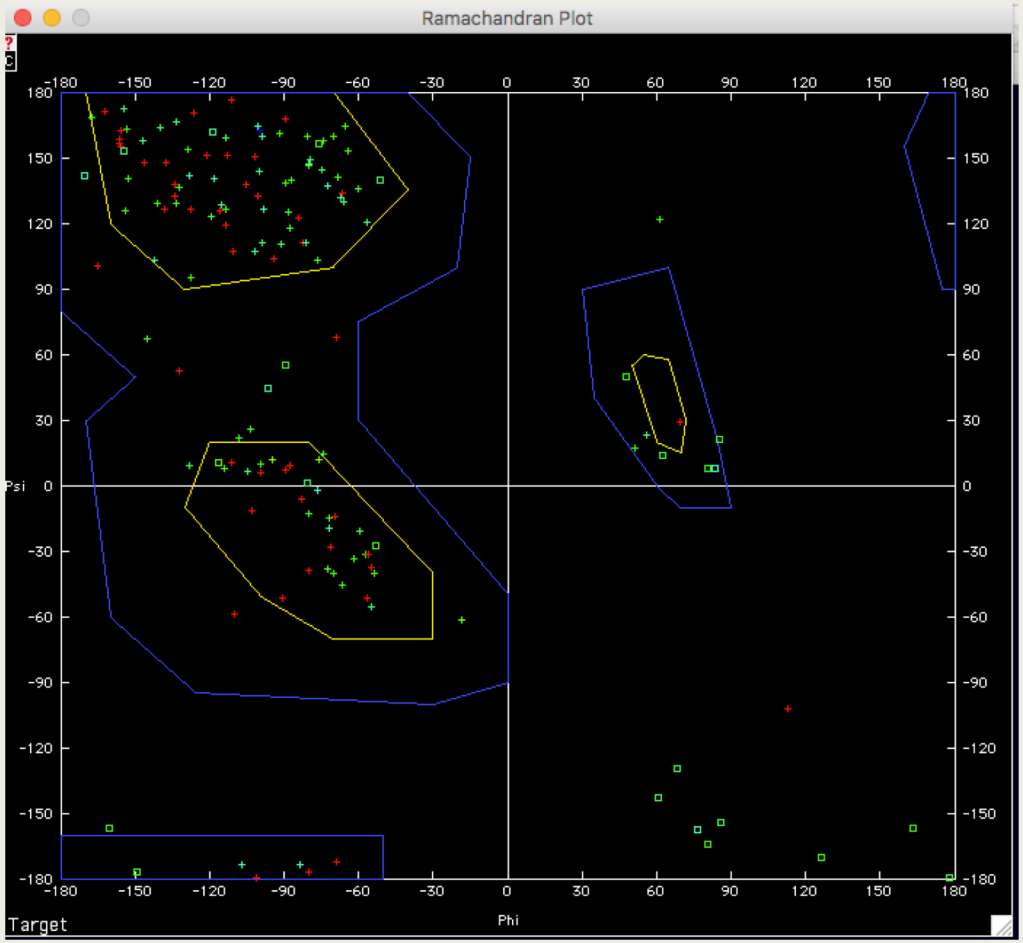
Adjusting alignments with DeepView



Viewing Ramachandran plot in deepview



Viewing Ramachandran plot in deepview



- Green crosses indicates residues with ***favored*** *dihedral angles*
- Red cross indicated residues with ***unfavored*** dihedral angles
- If most of these are in loop regions that is acceptable

Validating your model - SAVES

The Structure Analysis and Verification Server version 4

[SAVES] | XdVal | MTZdump | Ramachandran Plot | pdbU | pdbSNAFU (Check for ADIT compliance) | PROCHECK | Verify3D | ERRAT

This metaserver runs 6 programs for checking and validating protein structures during and after model refinement.

This server processed 3,677.3 jobs per month in the last 12 months. See more usage statistics here

Monthly totals plot

PROCHECK	Checks the stereochemical quality of a protein structure by analyzing residue-by-residue geometry and overall structure geometry. [Reference]
WHAT_CHECK	Derived from a subset of protein verification tools from the WHATIF program (Vriend, 1990), this does extensive checking of many sterochemical parameters of the residues in the model. [Reference]
ERRAT	Analyzes the statistics of non-bonded interactions between different atom types and plots the value of the error function versus position of a 9-residue sliding window, calculated by a comparison with statistics from highly refined structures. [Reference]
VERIFY_3D	Determines the compatibility of an atomic model (3D) with its own amino acid sequence (1D) by assigning a structural class based on its location and environment (alpha, beta, loop, polar, nonpolar etc) and comparing the results to good structures. References: [Bowie *et al.*, 1991; Luethy *et al.*, 1992].
PROVE	Calculates the volumes of atoms in macromolecules using an algorithm which treats the atoms like hard spheres and calculates a statistical Z-score deviation for the model from highly resolved (2.0 Å or better) and refined (R-factor of 0.2 or better) PDB-deposited structures. [PUBMED Reference].
CRYST1 record matches	We take the CRYST1 record and search the entire PDB for matches and report these as possibly similar structures.
Ramachandran Plot	We produce an interactive Ramachandran plot. Also a standalone server linked above.

http://services.mbi.ucla.edu/SAVES/

Summary

- Final model is only as good as the input template.
- Identity is usually the best benchmark for final quality but don’t ignore SS prediction.
- Always remember to cross check, more models means more confidence (for the most part).
- Try different model tools (my preference are I-TASSER and robetta but they are slow)
- You may need to try several different alignments when homology is low – Alignment is the place most likely to introduce errors.
- Always validate! Does your model make chemical sense?

Further reading

- For more on definitions of protein structure:
  - *Ramachandran GN, Ramakrishnan C, Sasisekharan V (1963). JMol Biol 7:95-99.*
  - *Richardson. (1981). The Anatomy and Taxonomy of Protein Structure, Advances in protein chemistry; 34. 167 -339 (New York: Academic Press).*
  - *Kuhlman, B., and Baker, D. (2000). Native protein sequences are close to optimal for their structures. Proceedings of the National Academy of Sciences 97, 10383–10388.*
  - *Zhang, Y. (2009). Protein structure prediction: when is it useful? Current Opinion in Structural Biology 19, 145–155.*

Further reading

- For more on folding and algorithms:
  - *Li, Z., and Scheraga, H.A. (1987). Monte Carlo-minimization approach to the multiple-minima problem in protein folding. Proceedings of the National Academy of Sciences 84, 6611–6615.*
  - *Simons, K.T., Kooperberg, C., Huang, E., and Baker, D. (1997). Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and bayesian scoring functions1. Journal of Molecular Biology 268, 209–225.*
  - *Jones, D.T. (1999). Protein secondary structure prediction based on position-specific scoring matrices1. Journal of Molecular Biology 292, 195–202.*
  - *Song, Y., DiMaio, F., Wang, R.Y.-R., Kim, D., Miles, C., Brunette, T., Thompson, J., and Baker, D. (2013). High-Resolution Comparative Modeling with RosettaCM. Structure 21, 1735–1742.*

Further reading

- For more on Swiss model in particular:
  - *Benkert, P., Biasini, M., and Schwede, T. (2011). Toward the estimation of the absolute quality of individual protein structure models. Bioinformatics 27, 343–350.*
  - *Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Cassarino, T.G., Bertoni, M., Bordoli, L., et al. (2014). SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Research 42, W252–W258.*
  - *Bienert, S., Waterhouse, A., de Beer, T.A.P., Tauriello, G., Studer, G., Bordoli, L., and Schwede, T. (2017). The SWISS-MODEL Repository—new features and functionality. Nucleic Acids Research 45, D313–D319.*