Homology modelling best practices

![Graph showing the relationship between the percentage of identical residues and the number of aligned residues. The graph highlights a 'Twilight zone' and a 'Safe homology modeling zone.'](image-url)
There are exceptions!

Myoglobin - Blue
Hemoglobin - Red

Sequence similarity is <25%
But identity is high.
Online modelling software (mostly automated)

- [https://swissmodel.expasy.org/](https://swissmodel.expasy.org/) - We will be using this today.
- [https://zhanglab.ccmb.med.umich.edu/I-TASSER/](https://zhanglab.ccmb.med.umich.edu/I-TASSER/) - One of the highest ranked for accuracy of produced models.
- [https://salilab.org/modeller/](https://salilab.org/modeller/) - Very popular and a lot of webservers run this.
- [https://modbase.compbio.ucsf.edu/modweb/](https://modbase.compbio.ucsf.edu/modweb/) - Webservice version of modeller.
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)
An example using SwissModel (automated mode)

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

```
<table>
<thead>
<tr>
<th>ID</th>
<th>Name</th>
<th>Title</th>
<th>Coverage</th>
<th>GMQE</th>
<th>QSQE</th>
<th>Identity</th>
<th>Method</th>
<th>Oligo State</th>
<th>Ligands</th>
</tr>
</thead>
<tbody>
<tr>
<td>3gtv.1.A</td>
<td>Superoxide dismutase [Cu-Zn]</td>
<td>Superoxide dismutase [Cu-Zn]</td>
<td>0.80</td>
<td>0.94</td>
<td></td>
<td>60.26</td>
<td>X-ray, 2.2Å</td>
<td>homo-dimer</td>
<td>2 x ZN</td>
</tr>
<tr>
<td>1n19.1.A</td>
<td>Superoxide Dismutase [Cu-Zn]</td>
<td>Superoxide Dismutase [Cu-Zn]</td>
<td>0.81</td>
<td>0.92</td>
<td></td>
<td>61.18</td>
<td>X-ray, 1.9Å</td>
<td>homo-dimer</td>
<td>2 x ZN</td>
</tr>
<tr>
<td>4b3e.1.A</td>
<td>SUPEROXIDE DISMUTASE [Cu-Zn]</td>
<td>SUPEROXIDE DISMUTASE [Cu-Zn]</td>
<td>0.79</td>
<td>0.91</td>
<td></td>
<td>61.18</td>
<td>X-ray, 2.1Å</td>
<td>homo-dimer</td>
<td>2 x ZN</td>
</tr>
</tbody>
</table>
```
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 assess the model we wish to use.
An example using SwissModel (automated mode)

<table>
<thead>
<tr>
<th>ID</th>
<th>Name</th>
<th>Title</th>
<th>Coverage</th>
<th>GMQE</th>
<th>GSQE</th>
<th>Identity</th>
<th>Method</th>
<th>Oligo State</th>
<th>Ligands</th>
</tr>
</thead>
<tbody>
<tr>
<td>3gvt.1.A</td>
<td>Superoxide dismutase [Cu-Zn]</td>
<td></td>
<td>0.80</td>
<td>0.94</td>
<td>60.26</td>
<td>X-ray, 2.2Å</td>
<td>homo-dimer</td>
<td>✓</td>
<td>2 x ZN$^{\text{c}}$</td>
</tr>
<tr>
<td>1n19.1.A</td>
<td>Superoxide dismutase [Cu-Zn]</td>
<td></td>
<td>0.81</td>
<td>0.92</td>
<td>61.18</td>
<td>X-ray, 1.9Å</td>
<td>homo-dimer</td>
<td>✓</td>
<td>2 x ZN$^{\text{c}}$, 2 x CU$^{\text{c}}$</td>
</tr>
</tbody>
</table>
Checking the structure in the PDB
Checking the structure in the PDB

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
Checking the structure in the PDB

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
Checking the structure in the PDB

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

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

Scores below 0.6 are low quality. In this case because two chains were modelled, there are two separate graphs (one for each chain).
An example using SwissModel (automated mode)
Cross validate with several other templates if possible
Cross validate with several other templates if possible

Secondary structure is in good agreement!
Adjusting alignments with DeepView

Secondary structure is in good agreement!
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 | pdbSNAPU (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

<table>
<thead>
<tr>
<th>Module</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PROCHECK</td>
<td>Checks the stereochemical quality of a protein structure by analyzing residue-by-residue geometry and overall structure geometry. [Reference]</td>
</tr>
<tr>
<td>WHAT_CHECK</td>
<td>Derived from a subset of protein verification tools from the WHATIF program (Vriend, 1990), this does extensive checking of many stereochemical parameters of the residues in the model. [Reference]</td>
</tr>
<tr>
<td>ERRAT</td>
<td>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]</td>
</tr>
<tr>
<td>VERIFY_3D</td>
<td>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].</td>
</tr>
<tr>
<td>PROVE</td>
<td>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].</td>
</tr>
<tr>
<td>CRIST1 record matches</td>
<td>We take the CRIST1 record and search the entire PDB for matches and report these as possibly similar structures.</td>
</tr>
<tr>
<td>Ramachandran Plot</td>
<td>We produce an interactive Ramachandran plot. Also a standalone server linked above.</td>
</tr>
</tbody>
</table>

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:
Further reading

For more on folding and algorithms:


Further reading