BIOZENTRUM
University of Basel
The Center for Molecular Life Sciences

SWISS-MODEL Homology Modelling Report

Model Building Report

This document lists the results for the homology modelling project "Untitled Project" submitted to SWISS-MODEL workspace on Dec. 24, 2020, 2:09 p.m..The submitted primary amino acid sequence is given in Table T1.

If you use any results in your research, please cite the relevant publications:

- Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L., Lepore, R., Schwede, T. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. 46(W1), W296-W303 (2018). doi
- Guex, N., Peitsch, M.C., Schwede, T. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective. Electrophoresis 30, S162-S173 (2009). doi
- Bienert, S., Waterhouse, A., de Beer, T.A.P., Tauriello, G., Studer, G., Bordoli, L., Schwede, T. The SWISS-MODEL Repository - new features and functionality. Nucleic Acids Res. 45, D313-D319 (2017). doi
- Studer, G., Rempfer, C., Waterhouse, A.M., Gumienny, G., Haas, J., Schwede, T. QMEANDisCo - distance constraints applied on model quality estimation. Bioinformatics 36, 1765-1771 (2020). doi
- Bertoni, M., Kiefer, F., Biasini, M., Bordoli, L., Schwede, T. Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology. Scientific Reports 7 (2017). doi

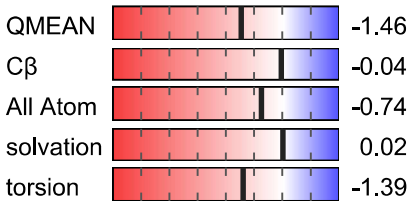
Results

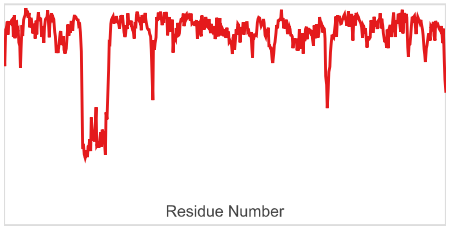
The user uploaded a template structure to use for the modelling process.

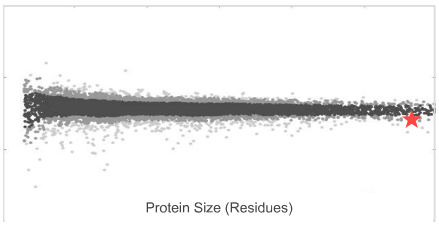
Models

The following model was built (see Materials and Methods "Model Building"):

Model #01	File	Built with	Oligo-State	Ligands	GMQE	QMEAN
	PDB	ProMod3 3.2.0	monomer	1 x 760: 760; 1 x SO4: SO4;	0.57	-1.46







Template	Seq Identity	Oligo-state	QSQE	Found by	Method	Resolution	Seq Similarity	Range	Coverage	Description
template_upload.1.A	91.42	monomer	0.00	BLAST	Unknown	-	0.60	142 - 709	0.72	Polypeptide

Included Ligands

Ligand	Description
1 x 760	760
1 x SO4	SO4

Excluded ligands

Ligand Name.Number	Reason for Exclusion	Description
BMA.7	Binding site not conserved.	BMA
MAN.6	Not in contact with model.	MAN
NAG.2	Not in contact with model.	NAG
NAG.4	Clashing with protein.	NAG
NAG.5	Not in contact with model.	NAG

```
Target                  MARFITSSSLSPSTTPSFASATASPAIASSLPRSKSTKATKARKTILSANSFKTS
template_upload.1.A     ------------------------------------------------------

Target                  SLSSLSSSASSPSLSSSPSFFKSSSSSKSRYSSSHLFNAFASLIPSSLAFSSLSF
template_upload.1.A     ------------------------------------------------------

Target                  SSLTSTNDLYRGFFATLVILLRMSSVAYGITDRLIVQTTSGPVRGRAVTVQGREV
template_upload.1.A     ------------------------------DRLVVQTSSGPVRGRSVTVQGREV

Target                  HVFTGIPYAKPPVDDLRFRKPVPAEPWHGVLDATRLPATCVQERYEYFPGFSGEE
template_upload.1.A     HVYTGIPYAKPPVEDLRFRKPVPAEPWHGVLDATGLSATCVQERYEYFPGFSGEE

Target                  IWNPNTNVSEDCLYMNIWAPSKARLRHGRGANGVEHAAKTDPDHLIHSATPQNTT
template_upload.1.A     IWNPNTNVSEDCLYINVWAPA--------------------------------TT

Target                  NGLPILIWIYGGGFMTGSATLDIYNADIMSAVGNVIVASFQYRVGAFGFLHLSPV
template_upload.1.A     NGLPILIWIYGGGFMTGSATLDIYNADIMAAVGNVIVASFQYRVGAFGFLHLAPE

Target                  MPG-FEEEAPGNVGLWDQALALRWLKENARAFGGNPEWMTLFGESAGSSSVNAQL
template_upload.1.A     MPSEFAEEAPGNVGLWDQALAIRWLKDNAHAFGGNPEWMTLFGESAGSSSVNAQL

Target                  VSPVTRGLVKRGMMQSGTMNAPWSHMTSEKAVEIGKALINDCNCNASLLSENPQS
template_upload.1.A     MSPVTRGLVKRGMMQSGTMNAPWSHMTSEKAVEIGKALINDCNCNASMLKTNPAH

Target                  VMACMRSVDAKTISVQQWNSYSGILSFPSAPTIDGAFFTADPMTLMKTADMTGYD
template_upload.1.A     VMSCMRSVDAKTISVQQWNSYSGILSFPSAPTIDGAFLPADPMTLMKTADLKDYD

Target                  IMIGNVKDEGTYFLLYDFIDYFDKDEATSLPRDKYLEIMNNIFNKATQAEREAII
template_upload.1.A     ILMGNVRDEGTYFLLYDFIDYFDKDDATALPRDKYLEIMNNIFGKATQAEREAII

Target                  FQYTSWEGNPGYQNQQQIGRAVGDHFFTCPTNEYAQALAERGAQVHYYYFTHRTS
template_upload.1.A     FQYTSWEGNPGYQNQQQIGRAVGDHFFTCPTNEYAQALAERGASVHYYYFTHRTS

Target                  TSLWGEWMGVLHGDEIEYFFGQPLNTSLQYRPVERELGKRMLNSVIEFAKTGNPA
template_upload.1.A     TSLWGEWMGVLHGDEIEYFFGQPLNNSLQYRPVERELGKRMLSAVIEFAKTGNPA

Target                  VDGEEWPNFSKEDPVYYVFSTDEKTEKLQRGPLAKRCSFWNDYLPKVRSWVGSEC
template_upload.1.A     QDGEEWPNFSKEDPVYYIFSTDDKIEKLARGPLAARCSFWNDYLPKVRSW-----

Target                  ENNNAASAAVNNINGQQYLLKWVIMLTIMVTCIFQ
template_upload.1.A     ----------------------------------
```

Materials and Methods

User Template Alignment

The user entered their own target sequence together with an uploaded a template structure file in PDB format.

Model Building

Models are built based on the target-template alignment using ProMod3. Coordinates which are conserved between the target and the template are copied from the template to the model. Insertions and deletions are remodelled using a fragment library. Side chains are then rebuilt. Finally, the geometry of the resulting model is regularized by using a force field. In case loop modelling with ProMod3 fails, an alternative model is built with PROMOD-II (Guex et al.).

Model Quality Estimation

The global and per-residue model quality has been assessed using the QMEAN scoring function (Studer et al.).

Ligand Modelling

Ligands present in the template structure are transferred by homology to the model when the following criteria are met: (a) The ligands are annotated as biologically relevant in the template library, (b) the ligand is in contact with the model, (c) the ligand is not clashing with the protein, (d) the residues in contact with the ligand are conserved between the target and the template. If any of these four criteria is not satisfied, a certain ligand will not be included in the model. The model summary includes information on why and which ligand has not been included.

Oligomeric State Conservation

The quaternary structure annotation of the template is used to model the target sequence in its oligomeric form. The method (Bertoni et al.) is based on a supervised machine learning algorithm, Support Vector Machines (SVM), which combines interface conservation, structural clustering, and other template features to provide a quaternary structure quality estimate (QSQE). The QSQE score is a number between 0 and 1, reflecting the expected accuracy of the interchain contacts for a model built based a given alignment and template. Higher numbers indicate higher reliability. This complements the GMQE score which estimates the accuracy of the tertiary structure of the resulting model.

References

- **BLAST**
  Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L. BLAST+: architecture and applications. BMC Bioinformatics 10, 421-430 (2009). doi>
- **HHblits**
  Steinegger, M., Meier, M., Mirdita, M., Vöhringer, H., Haunsberger, S. J., Söding, J. HH-suite3 for fast remote homology detection and deep protein annotation. BMC Bioinformatics 20, 473 (2019). doi>

Table T1:

Primary amino acid sequence for which templates were searched and models were built.

```
MARFITSSSLSPSTTPSFASATASPAIASSLPRSKSTKATKARKTILSANSFKTSSLSSLSSSASSPSLSSSPSFFKSSSSSKSRYSSSHLFNAFASLIP
SSLAFSSLSFSSLTSTNDLYRGFFATLVILLRMSSVAYGITDRLIVQTTSGPVRGRAVTVQGREVHVFTGIPYAKPPVDDLRFRKPVPAEPWHGVLDATR
LPATCVQERYEYFPGFSGEEIWNPNTNVSEDCLYMNIWAPSKARLRHGRGANGVEHAAKTDPDHLIHSATPQNTTNGLPILIWIYGGGFMTGSATLDIYN
ADIMSAVGNVIVASFQYRVGAFGFLHLSPVMPGFEEEAPGNVGLWDQALALRWLKENARAFGGNPEWMTLFGESAGSSSVNAQLVSPVTRGLVKRGMMQS
GTMNAPWSHMTSEKAVEIGKALINDCNCNASLLSENPQSVMACMRSVDAKTISVQQWNSYSGILSFPSAPTIDGAFFTADPMTLMKTADMTGYDIMIGNV
KDEGTYFLLYDFIDYFDKDEATSLPRDKYLEIMNNIFNKATQAEREAIIFQYTSWEGNPGYQNQQQIGRAVGDHFFTCPTNEYAQALAERGAQVHYYYFT
HRTSTSLWGEWMGVLHGDEIEYFFGQPLNTSLQYRPVERELGKRMLNSVIEFAKTGNPAVDGEEWPNFSKEDPVYYVFSTDEKTEKLQRGPLAKRCSFWN
DYLPKVRSWVGSECENNNAASAAVNNINGQQYLLKWVIMLTIMVTCIFQ
```

Table T2:

Template	Seq Identity	Oligo-state	QSQE	Found by	Method	Resolution	Seq Similarity	Coverage	Description
template_upload.1.A	91.42	monomer	-	BLAST	Unknown	NA	0.60	0.72	Polypeptide
template_upload.1.A	90.84	monomer	-	HHblits	Unknown	NA	0.60	0.71	Polypeptide