

October 23, 2016

Molecular Modeling of the p92 RdRp

Logan Donaldson / Dep't Biology / York University / logand@yorku.ca

This directory contains all sequences and protocols used to model the p92 RdRp.

An initial set of 16 homologous sequences to p92 was identified and automatically aligned to p92 using HH-PRED (<https://toolkit.tuebingen.mpg.de/hhpred>). A preliminary molecular model was made using the Rosetta comparative modeling implementation as described in:

http://www.rosettacommons.org/docs/latest/application_documentation/structure_prediction/RosettaCM

This preliminary model was compared with all entries in the PDB using the EBI SSM server (<http://www.ebi.ac.uk/msd-srv/ssm>). The HCV NS5B polymerase (2XYM) was selected as a single template for detailed modeling with 17% sequence identity to p92 in the finger and palm domains.

Stage 1

All template PDBs were prepared for Rosetta with the *clean_pdb.py* script.

Stage 2

Among the files listed in this directory, the prefix *kawf* refers to the p92 sequence from amino acids 458-818. The relevant files going forward in this stage were *kawf_2xym.grishin*, an alignment of the p92 to HCV NS5B RdRp, and *kawf-on-2xym.pdb*, an initial model of p92 that was created using the *partial_thread* module of Rosetta.

Stage 3

Three and nine amino acid fragment files of the p92 sequence (*full_03/09.frag*s) were calculated by the Robetta server at <http://robeta.bakerlab.org>. Hybridization of all templates and fragments was achieved with a Rosetta script that performed three rounds of relaxation with different parameter weightings. At this stage, it was essential to iteratively examine and adjust the p92-2XYM alignment to improve packing and eliminate unbalanced, buried charges in the model.

Stage 4

The purpose of this stage was to add protons to the hybridized model and improve geometry. This was achieved by running the *relax* module of Rosetta with default parameters. The accompanying script *pdbconvert.pl* was used to convert the Rosetta atom nomenclature to a standard format and renumber the sequence.

The final model of p92 consists of amino acids 458-686 and is provided as *p92-458-686.pdb*.