**Rosetta comparative homology modeling (RosettaCM), command lines and code**

**Data and code availability:**

All analysis scripts and Nsp1-ribosome homology models are available at [https://github.com/glasgowlab/Interface\_Energy\_nsp1](https://github.com/glasgowlab/Interface_Energy_nsp1/tree/master). Their use is documented in a README file in the same repository.

**Rosetta version:**

2020.11.post.dev+202.master.0840a913640 0840a913640fb154357316971a8f4b0903f2659d

**RosettaCM reference:**

“High-resolution comparative modeling with RosettaCM.” Song Y, DiMaio F, Wang RY, Kim D, Miles C, Brunette T, Thompson J, Baker D., *Structure*. 2013 Oct 8;21(10):1735-42. doi:10.1016/j.str.2013.08.005. Epub 2013 Sep 12. [1]

**Preparing input structures with Rosetta Relax:**

Relaxing prepares high-resolution experimentally-solved ribosome-Nsp1 structures as templates for comparative homology modeling in Rosetta [2]. We relaxed eight structures with backbone constraints. Input structure PDB codes: 6ZLW, 6ZM7, 6ZME, 6ZMI, 6ZMO, 6ZMT, 6ZN5, 6ZOJ. [3]

Example command line to relax the whole structure, PDB ID 6zlw:

/Users/anumglasgow/Rosetta/source/bin/relax.macosclangrelease -in:file:s 6zlw\_chain-i.pdb -database /Users/anumglasgow/Rosetta/database -relax:constrain\_relax\_to\_start\_coords -out:suffix \_relax -beta\_nov16 -corrections::beta\_nov16

Example command line to relax the ribosome-Nsp1 complex in a 150 Å radius around the Nsp1 C-terminal fragment, PDB ID 6zlw:

/Users/anumglasgow/Rosetta/source/bin/relax.macosclangrelease -in:file:s 6zlw\_small.pdb -database /Users/anumglasgow/Rosetta/database -relax:constrain\_relax\_to\_start\_coords -out:suffix \_relax -beta\_nov16 -corrections::beta\_nov16

**Partial threading:**

To guide the folding of the five evolutionarily-related Nsp1 C-terminal domain sequences, we threaded them on the eight ribosome-Nsp1 templates using the partial thread application in Rosetta. We also performed this step on the SARS-CoV-2 Nsp1 sequence. As input files for each model, we provided fasta files, alignment files in the Grishin format (one for every template), and the relaxed PDB template. Partial threading generates one model PDB file for each template, for each target sequence.

Example command line to build one threaded model:

/Users/anumglasgow/Rosetta/source/bin/partial\_thread.macosclangrelease -in:file:fasta ../fasta\_files/CoV1\_Nsp1.fasta -in:file:alignment ../grishin\_files/Cov1\_grishin/Cov1\_6zlw\_thread.grishin -in:file:template\_pdb ../minimize\_small\_pdbs/6zlw\_small\_relax.pdb

Fasta files:

>SARS-CoV-1\_Nsp1

ELGTDPIEDYEQNWNTKHGSGALRELTRELNGG

>SARS-CoV-2\_Nsp1

ELGTDPYEDFQENWNTKHSSGVTRELMRELNGG

>HKU1\_Nsp1

DAYAEVHAEPKGKYSQKAYALLRQYRG

>MERS\_Nsp1

EWMDDFEADPKGKYAQNLLKKLIGG

>MHV\_Nsp1.fasta

DACEEVHLNPKGKYSCKAYALLKGYRG

>OC43\_Nsp1

DAYDQVHDEPKGKFSKKAYALIRGYRG

Example alignment files for SARS-CoV-1 Nsp1 in the Grishin format:

## CoV1\_Nsp1 6zlw\_small\_thread

#

scores\_from\_program: 0

0 ELGTDPIEDYEQNWNTKHGSGALRELTRELNGG

0 ELGTDPYEDFQENWNTKHSSGVTRELMRELNGG

--

## CoV1\_Nsp1 6zm7\_small\_thread

#

scores\_from\_program: 0

0 ELGTDPIEDYEQNWNTKHGSGALRELTRELNGG

0 ----DPYEDFQENWNTKHSSGVTRELMRELNGG

--

## CoV1\_Nsp1 6zme\_small\_thread

#

scores\_from\_program: 0

0 ELGTDPIEDYEQNWNTKHGSGALRELTRELNGG

0 ----DPYEDFQENWNTKHSSGVTRELMRELNGG

--

## CoV1\_Nsp1 6zmi\_small\_thread

#

scores\_from\_program: 0

0 ELGTDPIEDYEQNWNTKHGSGALRELTRELNGG

0 ELGTDPYEDFQENWNTKHSSGVTRELMRELNGG

--

## CoV1\_Nsp1 6zmo\_small\_thread

#

scores\_from\_program: 0

0 ELGTDPIEDYEQNWNTKHGSGALRELTRELNGG

0 ELGTDPYEDFQENWNTKHSSGVTRELMRELNGG

--

## CoV1\_Nsp1 6zmt\_small\_thread

#

scores\_from\_program: 0

0 ELGTDPIEDYEQNWNTKHGSGALRELTRELNGG

0 ----DPYEDFQENWNTKHSSGVTRELMRELNGG

--

## CoV1\_Nsp1 6zn5\_small\_thread

#

scores\_from\_program: 0

0 ELGTDPIEDYEQNWNTKHGSGALRELTRELNGG

0 ----DPYEDFQENWNTKHSSGVTRELMRELNGG

--

## CoV1\_Nsp1 6zoj\_small\_thread

#

scores\_from\_program: 0

0 ELGTDPIEDYEQNWNTKHGSGALRELTRELNGG

0 ELGTDPYEDFQENWNTKHSSGVTRELMRELNGG

--

**Hybridization:**

In this step, we wrote a RosettaScript to build a homology model for each Nsp1 C-terminal domain sequence as it may look when bound to the human ribosome. The inputs files were: the RosettaScript XML files, the Nsp1 fasta files, and the threaded template PDB models from the previous step. We generated 1500-2000 models for each Nsp1 sequence from the threaded structures.

Example command line for hybridization to build the SARS-CoV1 Nsp1 models:

/Users/anumglasgow/Rosetta/source/bin/rosetta\_scripts.macosclangrelease -database /Users/anumglasgow/Rosetta/database -in:file:fasta CoV1\_Nsp1.fasta -parser:protocol CoV1\_hybridize.xml -nstruct 10000 -relax:jump\_move true -relax:dualspace -out:suffix \_hyb -beta -default\_max\_cycles 2000

Example RosettaScript [4] for hybridization to build the SARS-CoV1 Nsp1 models:

<ROSETTASCRIPTS>

<TASKOPERATIONS>

</TASKOPERATIONS>

<SCOREFXNS>

<ScoreFunction name="stage1"

weights="score3" symmetric="1">

<Reweight scoretype="atom\_pair\_constraint" weight="0.1"/>

</ScoreFunction>

<ScoreFunction name="stage2"

weights="score4\_smooth\_cart" symmetric="1">

<Reweight scoretype="atom\_pair\_constraint" weight="0.1"/>

</ScoreFunction>

<ScoreFunction name="fullatom"

weights="beta\_cart" symmetric="1">

<Reweight scoretype="atom\_pair\_constraint" weight="0.1"/>

</ScoreFunction>

</SCOREFXNS>

<FILTERS>

</FILTERS>

<MOVERS>

<Hybridize name="hybridize"

stage1\_scorefxn="stage1" stage2\_scorefxn="stage2"

fa\_scorefxn="fullatom" batch="1">

<Template pdb="6zlw\_small\_thread.pdb"

weight="1.0" cst\_file="AUTO"/>

<Template pdb="6zm7\_small\_thread.pdb"

weight="1.0" cst\_file="AUTO"/>

<Template pdb="6zme\_small\_thread.pdb"

weight="1.0" cst\_file="AUTO"/>

<Template pdb="6zmi\_small\_thread.pdb"

weight="1.0" cst\_file="AUTO"/>

<Template pdb="6zmo\_small\_thread.pdb"

weight="1.0" cst\_file="AUTO"/>

<Template pdb="6zmt\_small\_thread.pdb"

weight="1.0" cst\_file="AUTO"/>

<Template pdb="6zn5\_small\_thread.pdb"

weight="1.0" cst\_file="AUTO"/>

<Template pdb="6zoj\_small\_thread.pdb"

weight="1.0" cst\_file="AUTO"/>

</Hybridize>

</MOVERS>

<PROTOCOLS>

<Add mover="hybridize"/>

</PROTOCOLS>

</ROSETTASCRIPTS>

**Modeling Nsp1-ribosome complexes:**

The 1500-2000 models for each Nsp1 were ranked according to total score using the Rosetta energy function. We selected the four models with the lowest total energies for each sequence. These Nsp1 models were then aligned to the experimentally-solved Nsp1 C-terminal domain fragment in the SARS-CoV-2 high-resolution PDB structure 6LZW using the *align* command in the protein modeling software PyMOL (Schrödinger). We then relaxed and scored the ribosome-Nsp1 complexes in Rosetta with backbone constraints, as described in the previous section, for a total of four models for each Nsp1-ribosome complex.

**Analyzing Nsp1-ribosome complex models:**

We wrote a custom script (InterfaceAnalysis\_Run.py, available at https://github.com/glasgowlab/Interface\_Energy\_nsp1/blob/master/InterfaceAnalysis\_Run.py) for scoring using PyRosetta (version 2022.16: 2022.16+release.839b00f59d92ff5270c37073376bd37d27064c7e). The following sections describe how the calculations were performed.

Example command line for scoring Nsp1-ribosome complex and interface:

python InterfaceAnalaysis\_Run.py --pdb\_path inputs/COV1\_01.pdb --output\_dir output --target\_chain A

Scoring the Nsp1-ribosome complexes.

Each nsp1-ribosome complex model was scored using the Rosetta energy function with the *rna\_res\_level\_energy4* weights settings. We divided the scoring process into per residue energies total score, interface energy, and score type of the following rna terms: *fa\_elec\_rna\_phos\_phos, rna\_torsion, fa\_stac rna\_sugar\_close, hbond\_sr\_bb\_sc.* Additionally, we also scored the experimentally-solved SARS-CoV-2 Nsp1 structure 6ZLW.

Characterizing the Nsp1 interaction.

To evaluate the models and explore the basis for binding interactions, we calculated the intermolecular interface energies and individual score terms for Nsp1, ribosomal proteins, and ribosomal RNA. First, we scored the complexes using a score function optimized for RNA-protein interactions [5]. Residues within 6 Å of the C-terminal Nsp1 protein were considered to be in contact with the protein. We then calculated the interface score using the Rosetta Interface Energy application [6], which calculates all inter-residue pairwise energies over the two faces of an interface. The residues in contact with Nsp1 were enumerated based on the PDB chain index and used as faces for interface scoring. Residues in the same chain belonged to the same ribosomal protein or RNA across the Nsp1-ribosome complex models. We used the SARS-CoV-2 experimentally solved Nsp1 C-terminal domain-ribosome (PDB ID 6ZLW [3]) complex structure as a point of comparison for our models. Our models demonstrated similar predicted total energies across SARS-CoV-01, HKU, MERS, and OC43 Nsp1-ribosome complexes while the interface scores varied by model, suggesting that the individual Nsp1 interactions with the ribosome each involve different pairwise interactions of different strengths.

Determining which Nsp1 residues contribute *most strongly* to the interaction with the ribosome.

To further investigate the energetic basis for the predicted stability of the Nsp1-ribosome complex models, we checked individual per-residue energies by score terms related to RNA-protein interactions. Specifically, we focused on the following score terms: fa\_atr, fa\_elec, rna\_phos\_phos, rna\_torsio, fa\_stack, rna\_sugar\_close, hbond\_sc, and hbond\_sr\_bb\_sc [5, 7]. The fa\_stack scores suggest a well-conserved interaction across all Nsp1-ribosome complex models between aromatic amino acids and RNA nucleotides. This score term reports on stacking interactions of pi bonds. Additionally, we observed energetically favorable polar interactions among the Nsp1 C-terminal domain and ribosomal proteins close to the binding site. As an example, in the case of the SARS-CoV-1 Nsp1-ribosome model, we found Nsp1 residues H18, E25, and N31 to make significant interactions with the ribosome. Residue H18 in the SARS-CoV-1 Nsp1-ribosome model interacts with RNA bases 607U and 630U through pi stacking. Residues E25 and N31 hydrogen bond with backbone residues of ribosomal proteins.

Determining which Nsp1 residues make *similar or conserved* ribosomal interactions.

We then considered analogous amino acids across all Nsp1-ribosome complex models to determine similar/conserved interactions. The pi stacking predicted interactions of the C-terminal Nsp1 protein with ribosomal RNA were highly conserved across the models. Specifically, residue H18 in SARS-CoV-1 Nsp1, 14Y in HKU Nsp1 and MERS Nsp1, and F14 in OC43 Nsp1 all include an identical hydrophobic interaction with RNA bases U607U and U630. The experimentally solved structure of SARS-CoV-2 also shows the same interaction with residue H18. This well-conserved interaction suggests that pi stacking plays a strong role in maintaining the Nsp1-ribosome complex.

References:

1. Yifan Song *et al.* (2013), High-Resolution Comparative Modeling with RosettaCM, *Structure*, 21(10): 1735-1742.
2. Nivón LG, Moretti R, Baker D (2013) A Pareto-Optimal Refinement Method for Protein Design Scaffolds. *PLoS ONE* 8(4): e59004.
3. Matthias Thoms *et al.* (2020), Structural basis for translational shutdown and immune evasion by the Nsp1 protein of SARS-CoV-2. *Science* 369,1249-1255.
4. Fleishman SJ *et al.* (2011) RosettaScripts: A Scripting Language Interface to the Rosetta Macromolecular Modeling Suite. *PLOS ONE* 6(6): e20161.
5. A. M. Watkins, R. Rangan, R. Das (2020), FARFAR2: Improved De Novo Rosetta Prediction of  
   Complex Global RNA Folds*.* *Structure* 28, 963-976.e6
6. Bazzoli A *et al.* (2017) Using homology modeling to interrogate binding affinity in neutralization of ricin toxin by a family of single domain antibodies. *Proteins* 85(11):1994-2008.
7. Hahnbeom Park *et al*. (2016) Optimization of Biomolecular Energy Functions on Features from Small Molecules and Macromolecules. *Journal of Chemical Theory and Computation* 12(12): 6201-6212.