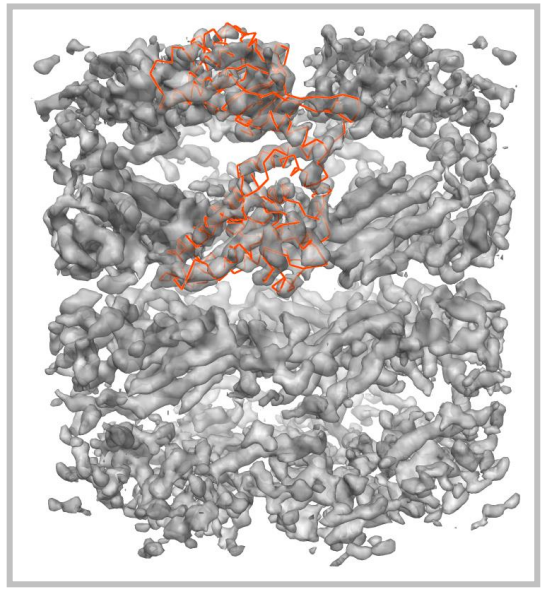
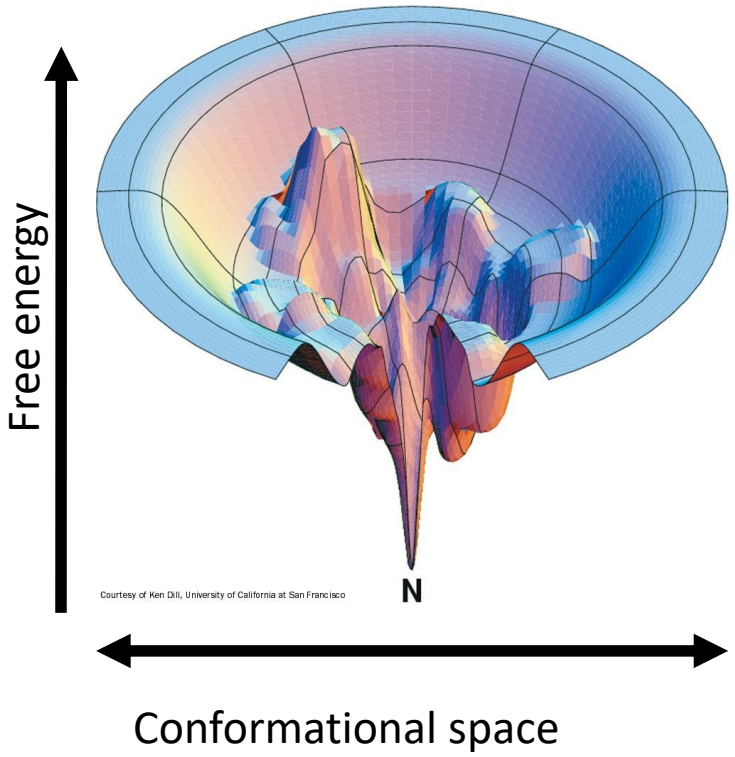


Introduction to Rosetta for density-guided model building

Frank DiMaio

Institute for Protein Design
University of Washington
dimaio@uw.edu

Protein structure prediction and protein design with Rosetta

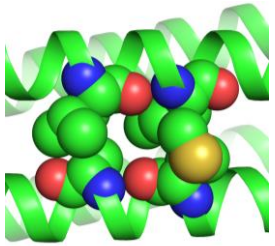


- Represent the various enthalpic and entropic effects governing folding with *parameterized equations*
  - vdW interactions
  - electrostatic interactions
  - solvent entropy
  - etc.
- **Structure prediction**: Identify the *lowest-energy* conformation of the protein
- **Protein design**: identify a sequence whose *lowest energy state* corresponds to the desired conformation

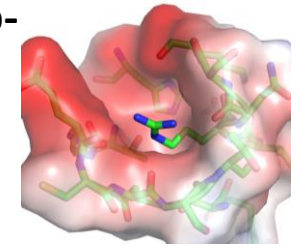
The “core” of Rosetta

Energy function

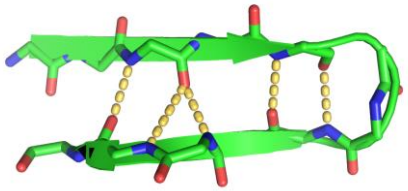
Core packing



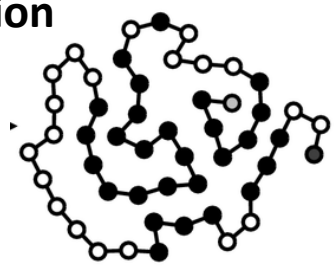
Electro-statics



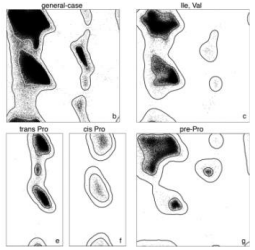
Hydrogen bonding



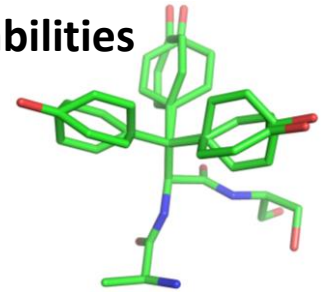
Implicit solvation



Torsional probabilities

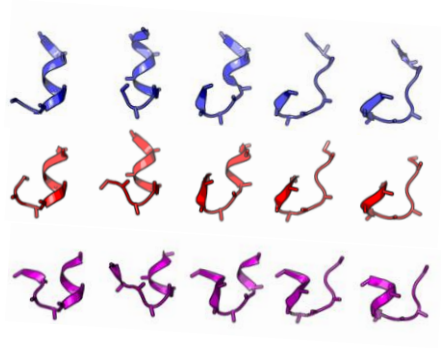


Rotamer probabilities



Sampling

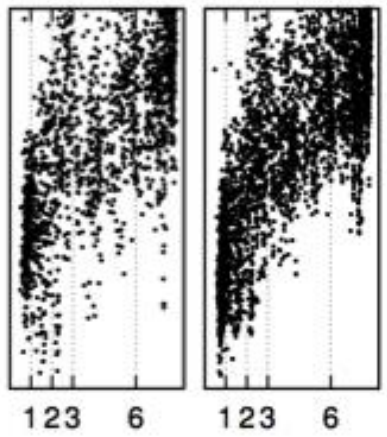
…VTKPLVARAKL…



The Rosetta energy function

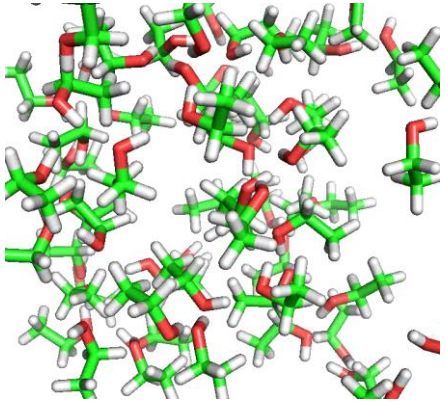
Rosetta

- Scorefunction contains a mixture of physically and statistically derived terms
- Weights/parameters fit to recapitulate high-resolution crystal structures



Molecular mechanics

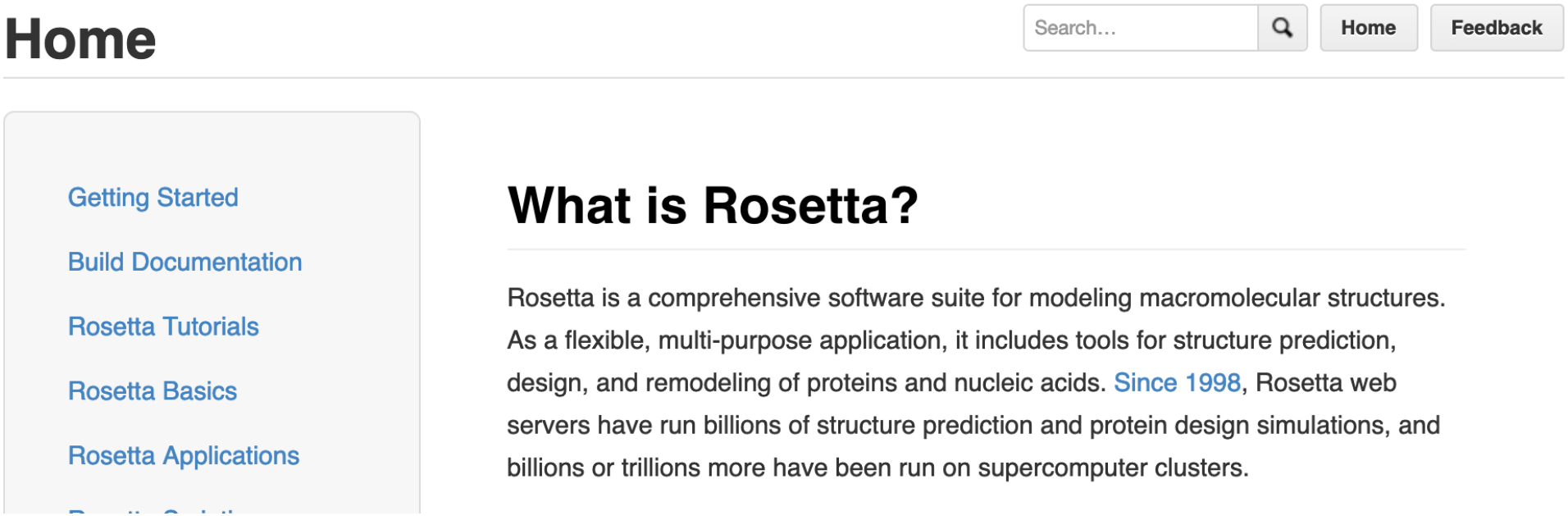
- Scorefunction contains only physically derived terms
- Weights/parameters (largely) fit to recapulate small molecule data



Getting Rosetta

Rosetta can be downloaded from
http://www.rosettacommons.org/software/index.html

Documentation wiki: https://www.rosettacommons.org/docs/latest/Home



Rosetta + EM tutorials

https://dimaiolab.ipd.uw.edu/software/

Tutorials

Using Rosetta with cryoEM density (*last updated August 2018*)
This tutorial gives an overview of all the EM-related tools available within Rosetta.
[instructions (pdf)] [all files (tgz)]

The main tutorial, covering all Rosetta EM tools

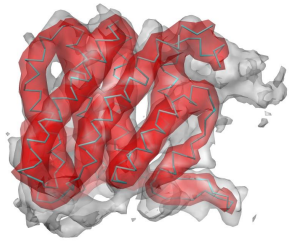
You may also download sections individually:

- **Part 1:** Introduction to Rosetta for cryoEM [pdf][tgz]
- **Part 2:** Fragment-based structure refinement [pdf][tgz]
- **Part 3:** Model rebuilding using RosettaCM [pdf][tgz]
- **Part 4:** De novo EM-guided structure determination [pdf][tgz]
- **Part 5:** Rebuilding long protein segments with RosettaES [pdf][tgz]

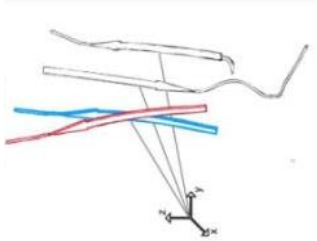
A tutorial given at EMBL in August 2022
This tutorial features new examples.
[instructions (pdf)] [all files (tgz)]

This tutorial

Today's lecture



1) Introduction to the Rosetta relax

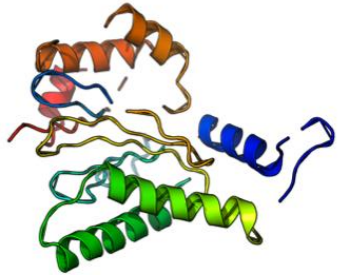


Completing a model with RosettaCM

2) RosettaCM to fix an Alphafold model

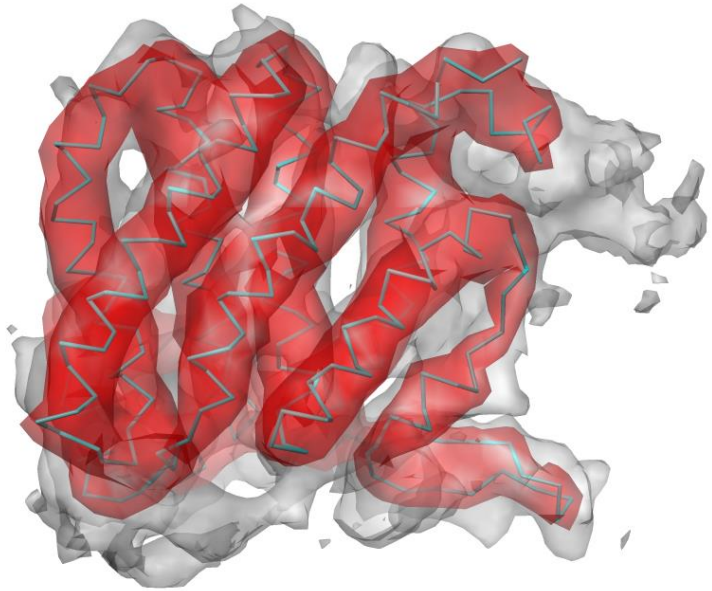
3) Refining a symmetric complex

4) Advanced RosettaCM: calculating energetics of alternate threadings



5) Rebuilding large segments with RosettaES

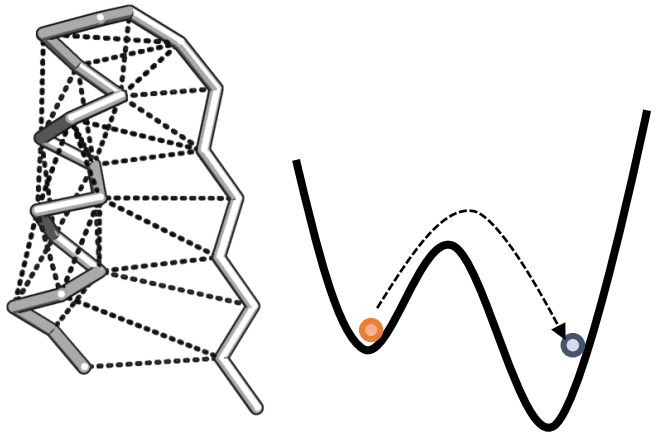
Rosetta density basics: Density scoring functions

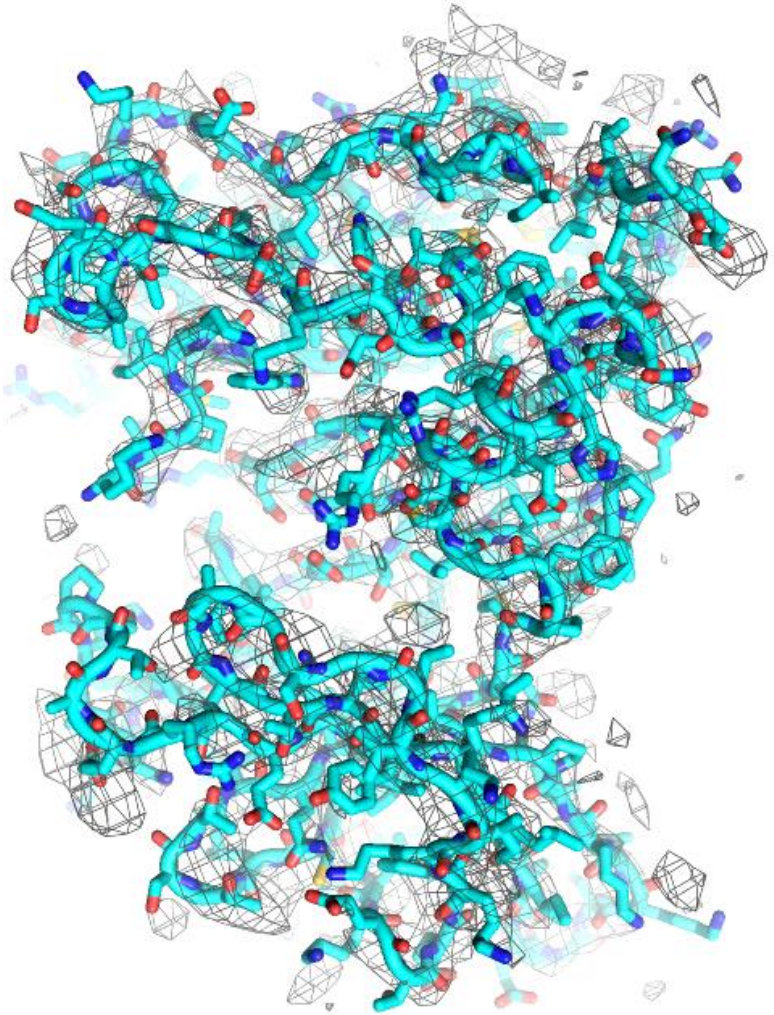


elec_dens_fast

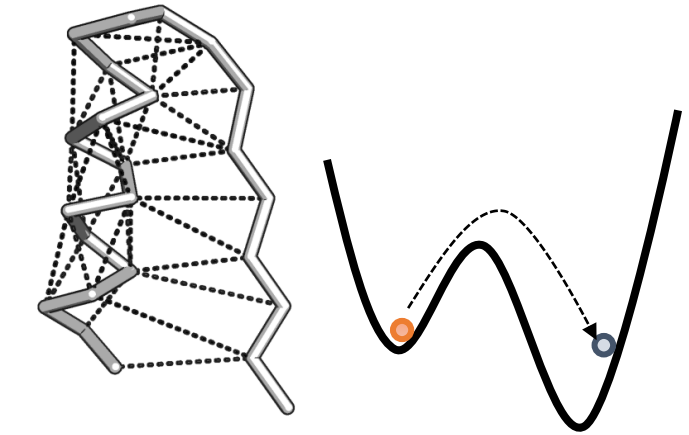
- Adds a **scoring term** to Rosetta assessing agreement of model and map
- Uses real-space correlation (RSCC) between the density given the model and experimental data
- **'_fast'** variant precomputes much of scoring making the density score not exactly a correlation
- Significantly faster (~10x) then a naïve correlation

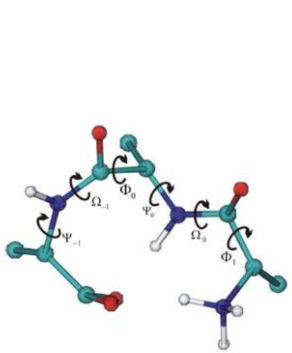
FastRelax: allatom refinement in Rosetta

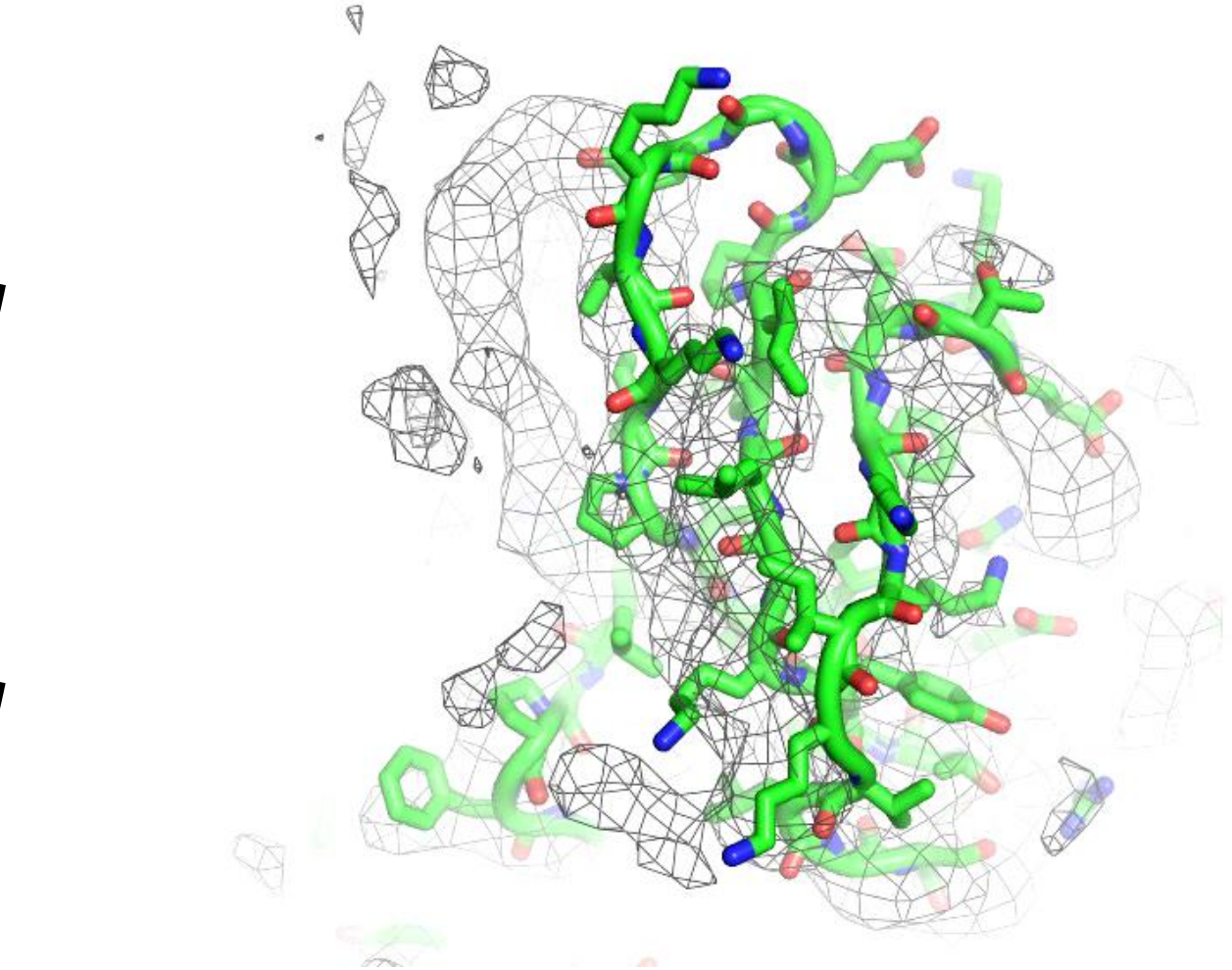




FastRelax: allatom refinement in Rosetta

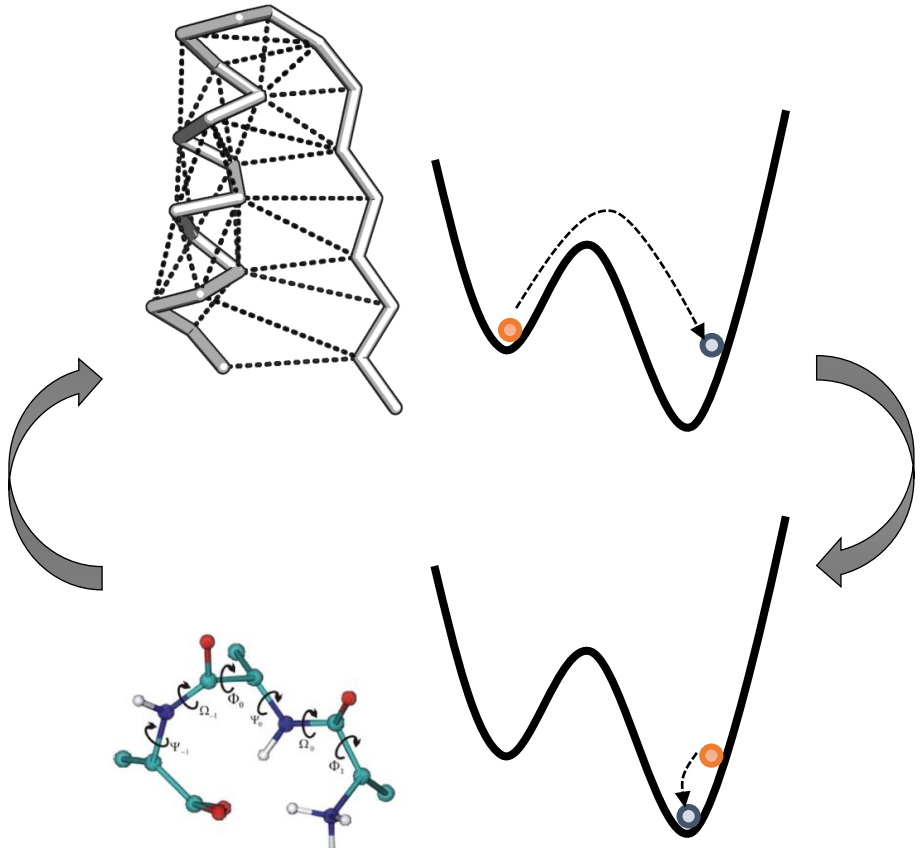




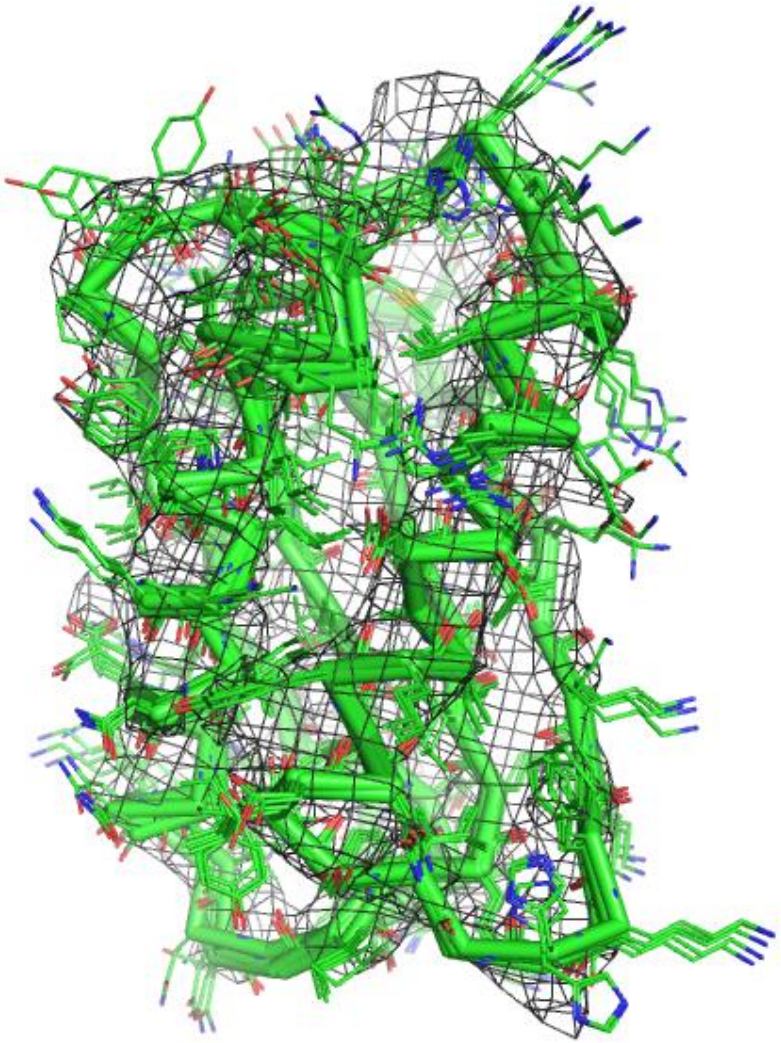


FastRelax: allatom refinement in Rosetta

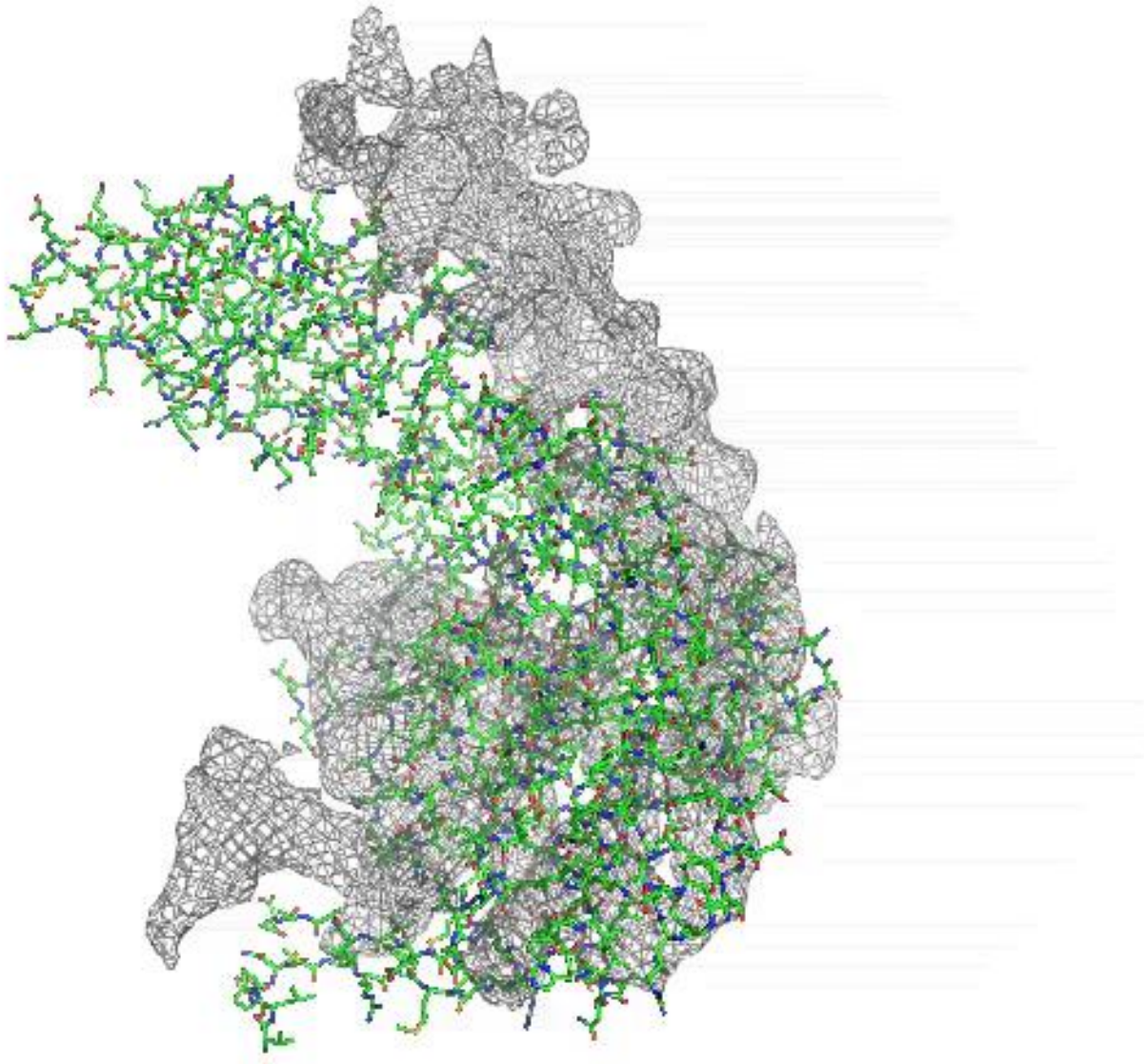
Discrete sidechain rotamer optimization



Continuous torsion-space minimization



FastRelax can capture large movements



Demo #1A

Rosetta output: scorefiles

Outputs a scorefile '*score.sc*'
(may be *default.sc* for some protocols)

```
SCORE:     score     fa_atr … elec_dens_fast … description
SCORE:  -354.953   -279.331   …   -344.503   …  2JEL_P_0001
SCORE:  -312.988   -279.933   …   -344.185   …  2JEL_P_0002
SCORE:  -546.073   -291.698   …   -360.408   …  2JEL_P_0003
SCORE:  -552.000   -288.205   …   -361.356   …  2JEL_P_0004
SCORE:  -549.890   -282.653   …   -366.543   …  2JEL_P_0005
```

Density energy alone

Rosetta energy + density energy

Rosetta density basics: Introduction to RosettaScripts XML interface

More complicated protocols make use of RosettaScripts

Declare score functions

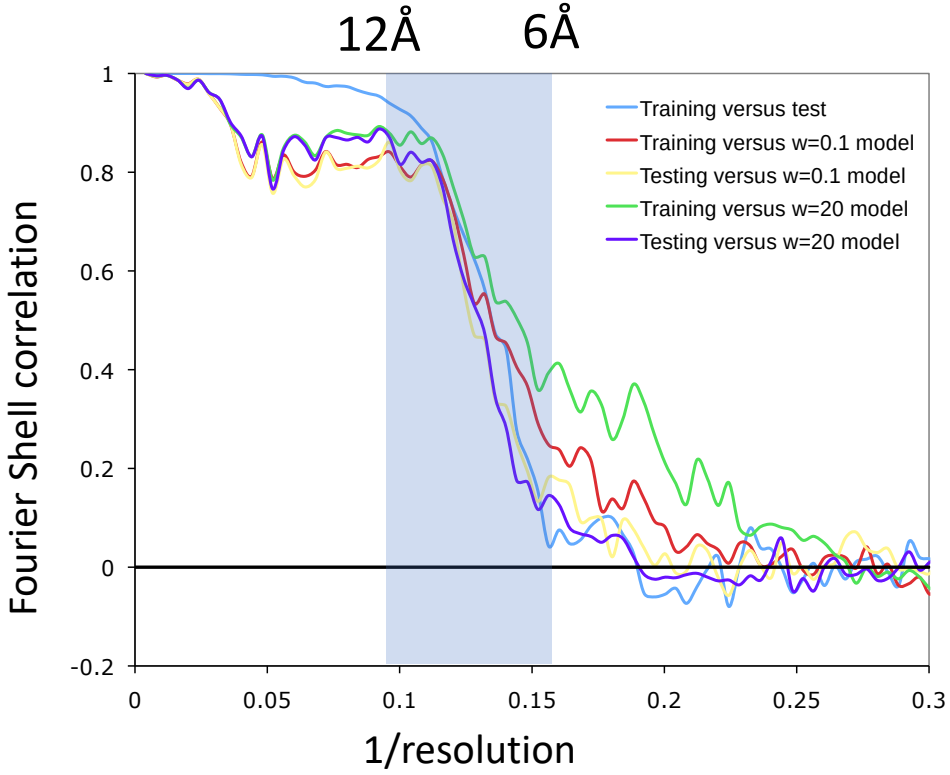
Declare movers (atomic conformation operations)

Protocol is a sequence of movers

```
<ROSETTASCRIPTS>
  <SCOREFXNS>
     <ScoreFunction name="dens" weights="ref2015">
        <Reweight scoretype="elec_dens_fast" weight="35.0"/>
     </ScoreFunction>
  </SCOREFXNS>

  <MOVERS>
     <SetupForDensityScoring name="setupdens"/>
     <LoadDensityMap name="loaddens" mapfile="1issA_6A.mrc"/>
     <FastRelax name="relax" scorefxn="dens"/>
  </MOVERS>

  <PROTOCOLS>
     <Add mover="setupdens"/>
     <Add mover="loaddens"/>
     <Add mover="relax"/>
   </PROTOCOLS>
   <OUTPUT scorefxn="dens"/>
</ROSETTASCRIPTS>
```

Demo #1C

Extra: Report model/map agreement

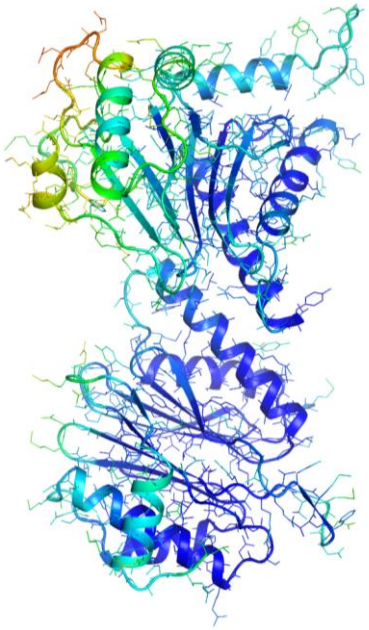



```
<ReportFSC name="report" res_high="12.0" res_low="6.0"/>
```

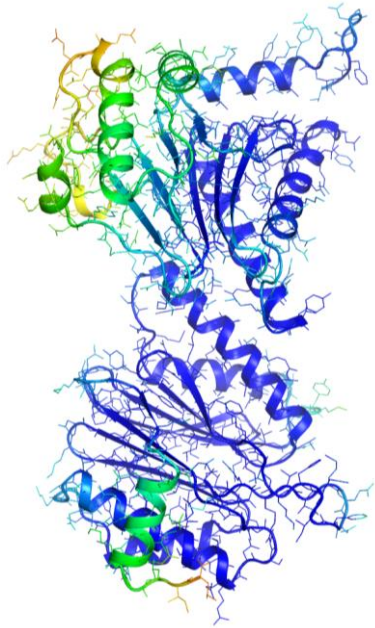
REMARK 1 FSC(12:6) = 0.587

Extra: Atomic B factor fitting

Deposited crystal structure (1pma)



20S proteasome at 3.3Å resolution, real-space B factors



```
<BfactorFitting name="fit_bs"
   max_iter="100" wt_adp="0.0005" init="1" exact="1"/>
```

Exercise

```
<ReportFSC name="report" res_high="12.0" res_low="6.0"/>
```

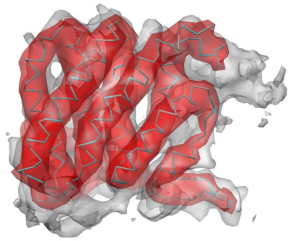
```
<BfactorFitting name="fit_bs" max_iter="100" wt_adp="0.0005"
   init="1" exact="1"/>
```

Write an XML for a refinement protocol that will:

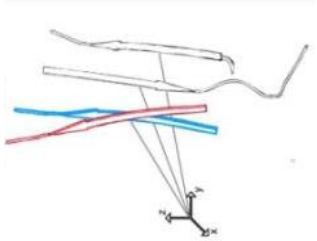
- Load 1isr and report integrated FSC between
- Relax 1isr into density map for 1 cycle in torsion space of 1iss and report integrated FSC
- Relax again in Cartesian space for 1 cycle and report integrated FSC
- Fit B factors and report integrated FSC

Run this protocol. What FSCs do you get?

Today's lecture



1) Introduction to the Rosetta relax

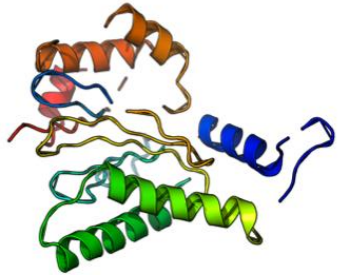


Completing a model with RosettaCM

2) RosettaCM to fix an Alphafold model

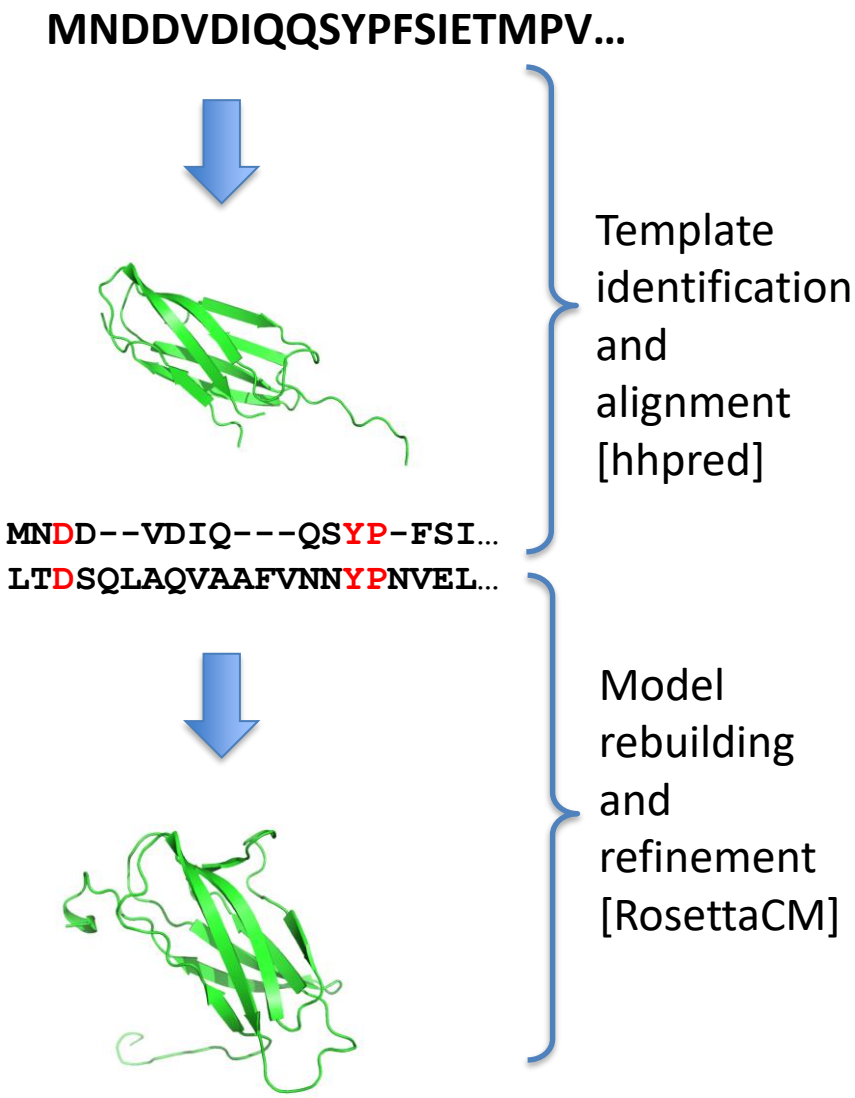
3) Refining a symmetric complex

4) Advanced RosettaCM: calculating energetics of alternate threadings



5) Rebuilding large segments with RosettaES

Model completion with RosettaCM

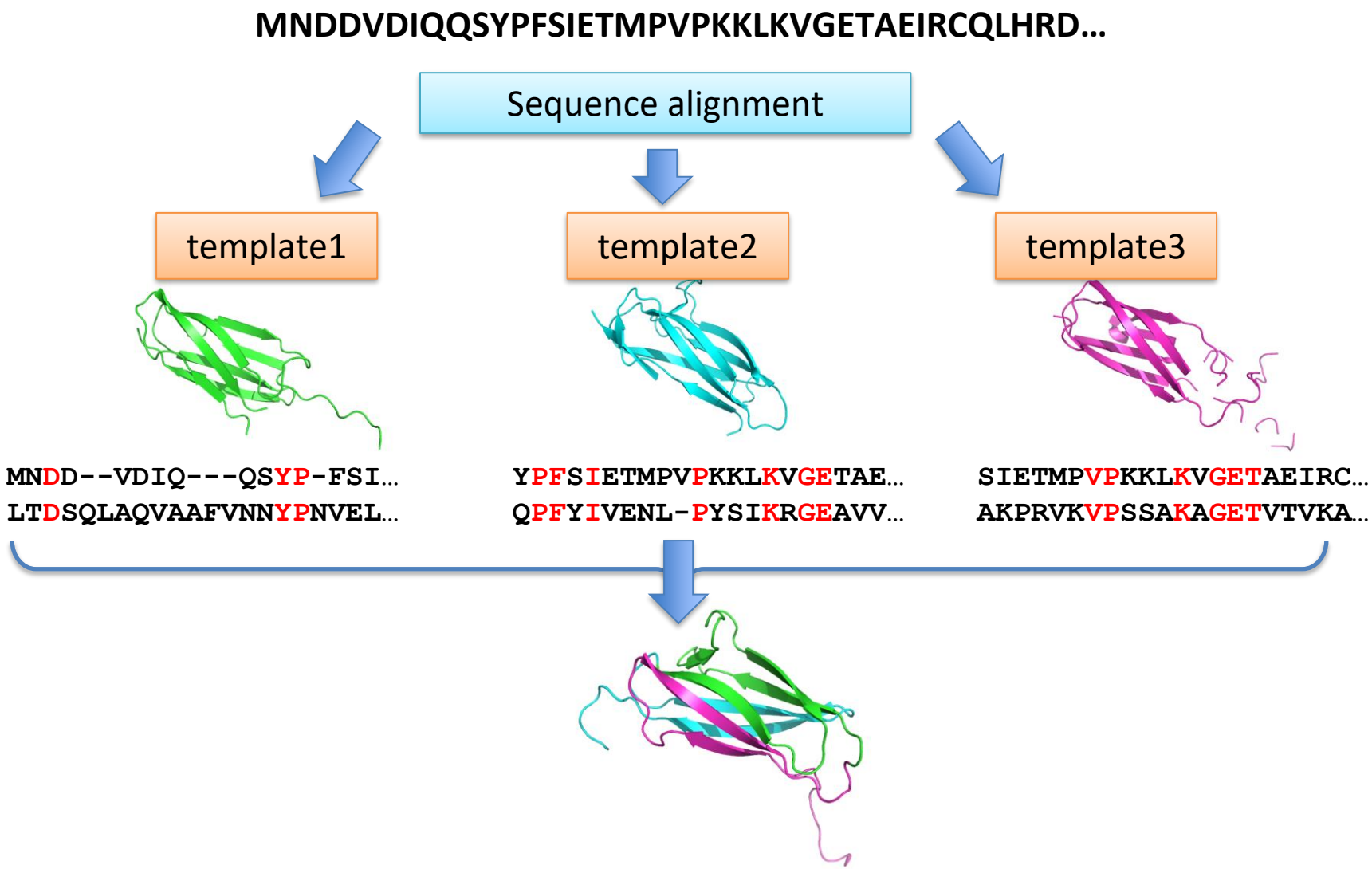


- RosettaCM (Comparative Modeling) is a tool originally developed for modelling structures from homologues
  - Modelling changes nearby mutations
  - Conformational sampling of backbone near **insertions** and **deletions**

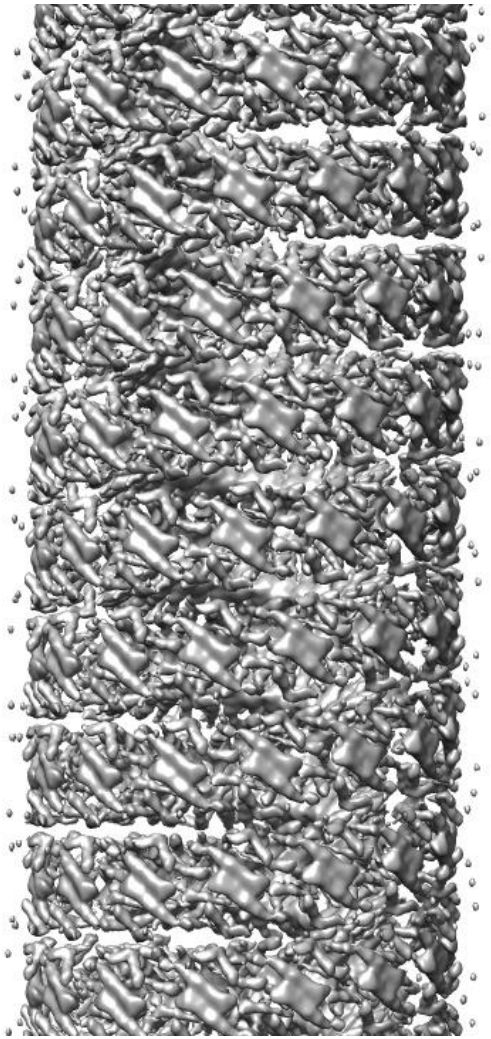
- When combined with density scoring, it is a very powerful general tool for modelling

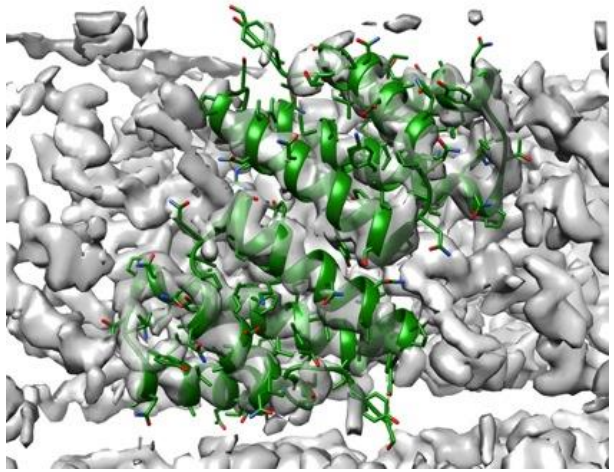
Song *et al.* (2013) *Structure*.

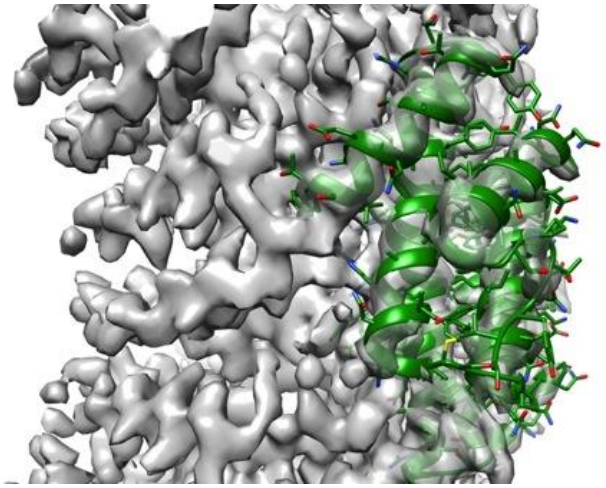
Model rebuilding with RosettaCM



RosettaCM
3.8Å cryo-EM reconstruction of SIRV2

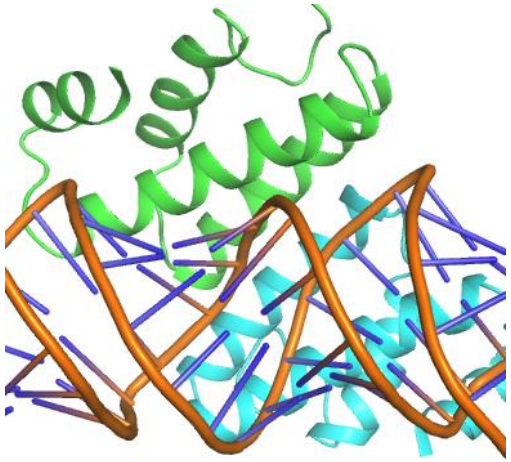




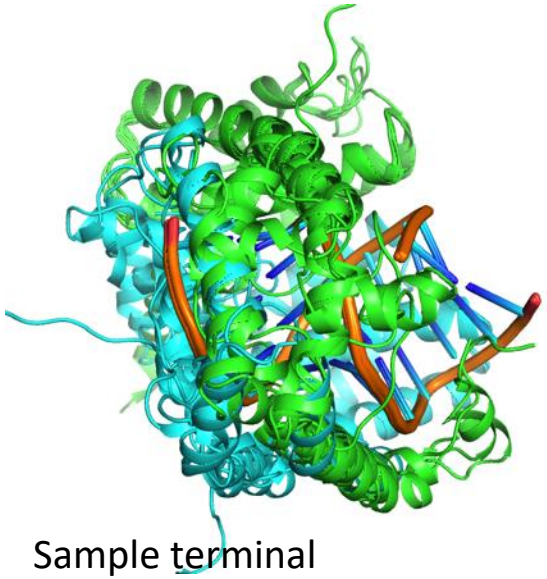


DiMaio *et al Science* (2015)

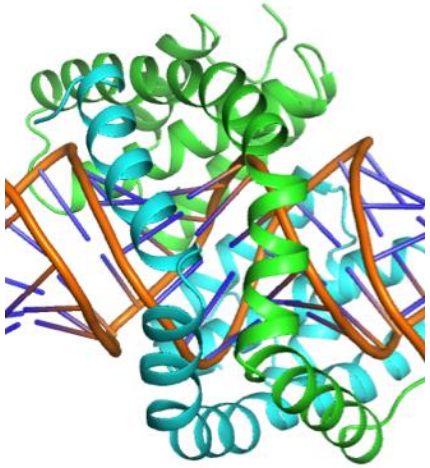
Rebuilding missing N-terminus



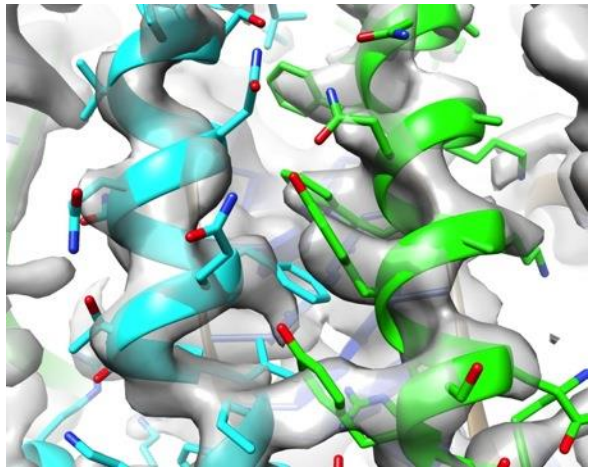
Initial model from docked crystal structure + DNA

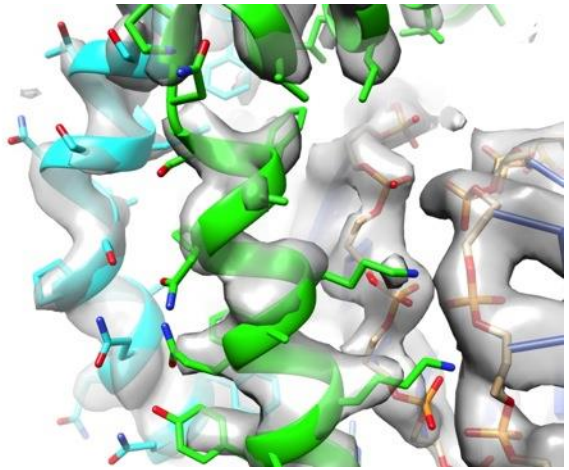


Sample terminal conformations with Rosetta guided by density



Further refinement via iterative local rebuilding





Model rebuilding into density with RosettaCM

```
 …
<Hybridize name="hybridize" stage1_scorefxn="stage1" stage2_scorefxn="stage2"
   fa_scorefxn="fullatom"  batch="1" stage1_increase_cycles="1.0"
   stage2_increase_cycles="1.0">
   <Template pdb="7s9d_trim_forCM.pdb" weight="1.0" cst_file="AUTO"/>
</Hybridize>
 …
```

Demo #2

Extra: Adding user-defined constraints

User-specified constraints:

```
AtomPair CA 107 CA 143 HARMONIC 5.0 1.0
```

```
…
<Hybridize name="hybridize" stage1_scorefxn="stage1"
    stage2_scorefxn="stage2" fa_scorefxn="fullatom"  batch="1"
    stage1_increase_cycles="1.0" stage2_increase_cycles="1.0">
   <Template pdb="7s9d_trim_forCM.pdb" weight="1.0" cst_file="custom.cst"/>
</Hybridize>
…
```

Extra: A full constraint language

```
Constraint_Type Constraint_Def
```

Basic types

- AtomPair <name1> <res1> <name2> <res2>
- Angle <name1> <res1> <name2> <res2> <name3> <res3>
- Dihedral <name1> <res1> <name2> <res2> <name3> <res3> …

```
AmbiguousConstraint
Constraint_Type1 Constraint_Def1
[Constraint_Type2 Constraint_Def2
[...]]
END
```

- lowest energy of all constraints

```
KofNConstraint k
Constraint_Type1 Constraint_Def1
[Constraint_Type2 Constraint_Def2
[...]]
END
```

- sum of lowest energy k of all constraints

https://www.rosettacommons.org/docs/latest/rosetta_basics/file_types/constraint-file

Extra: A full constraint language

```
Constraint_Type Constraint_Def
```

Basic types

- Harmonic x0 stdev
- CircularHarmonic x0 stdev
- Bounded xmin xmax stdev tag
- Sigmoid x0 m

- ConstantFunc x
- SumFunc func1_type func1_def func2_type func2_def
- ScalarWeightedFunc weight func_type func_def

https://www.rosettacommons.org/docs/latest/rosetta_basics/file_types/constraint-file

Exercise

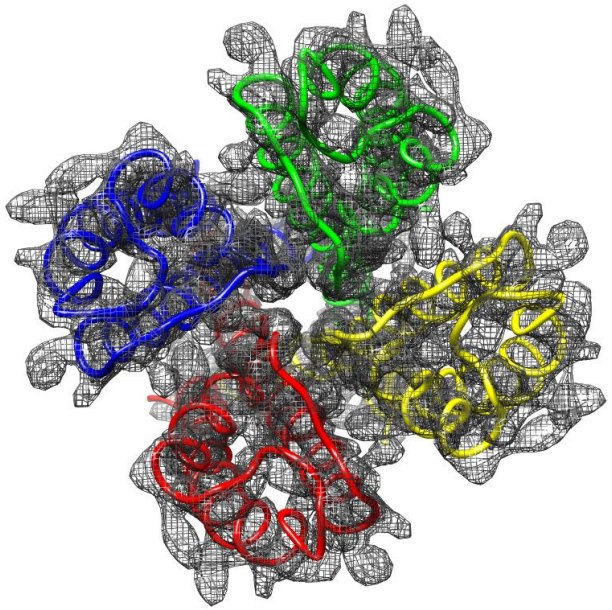
```
AtomPair CA 107 CA 143 HARMONIC 5.0 1.0
```

```
…
<Hybridize name="hybridize" stage1_scorefxn="stage1"
    stage2_scorefxn="stage2" fa_scorefxn="fullatom"  batch="1"
    stage1_increase_cycles="1.0" stage2_increase_cycles="1.0">
    <Template pdb="7s9d_trim_forCM.pdb" weight="1.0" cst_file="custom.cst"/>
</Hybridize>
…
```

Add a set of constraints that force a hydrogen bond to be formed between T621 –NH and G668 –C=O

Note: use the script scripts/renumberPDB.pl to get proper numbering in the input sequence (83 and 130, respectively)

Symmetric Modeling

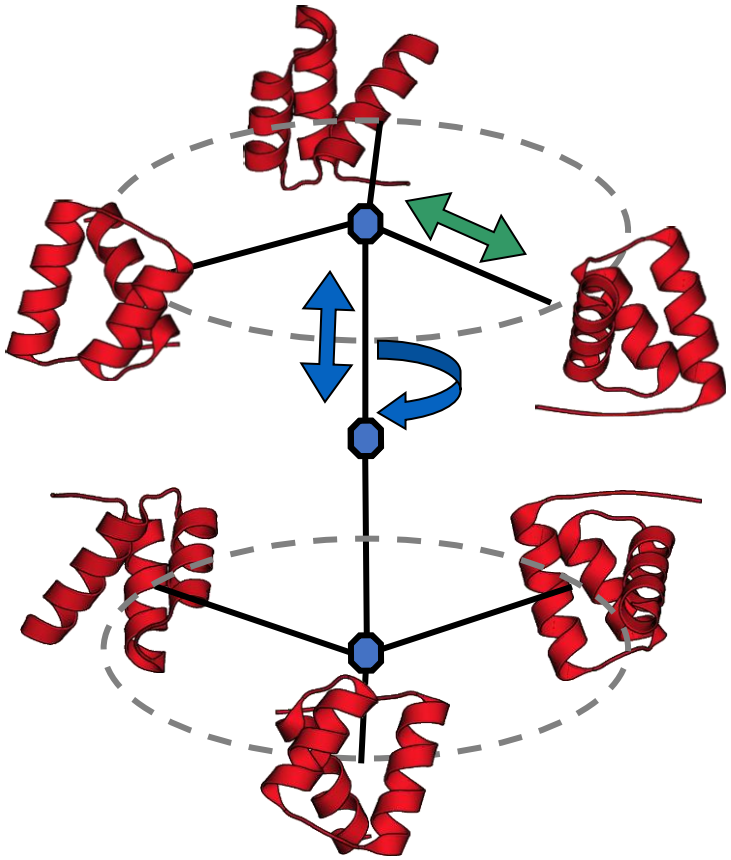


- Rosetta includes support for modeling and refinement of symmetric systems
- Symmetry is *implicit*: only the symmetric degrees of freedom are allowed to move
- Symmetry is controlled in Rosetta through a *symmetry definition file*

Symmetric Modeling

Symmetry definition file tells Rosetta…

- How to construct symmetric system from monomer
- Symmetric degrees of freedom
- How to score the symmetric system



The make_symmdef_file script

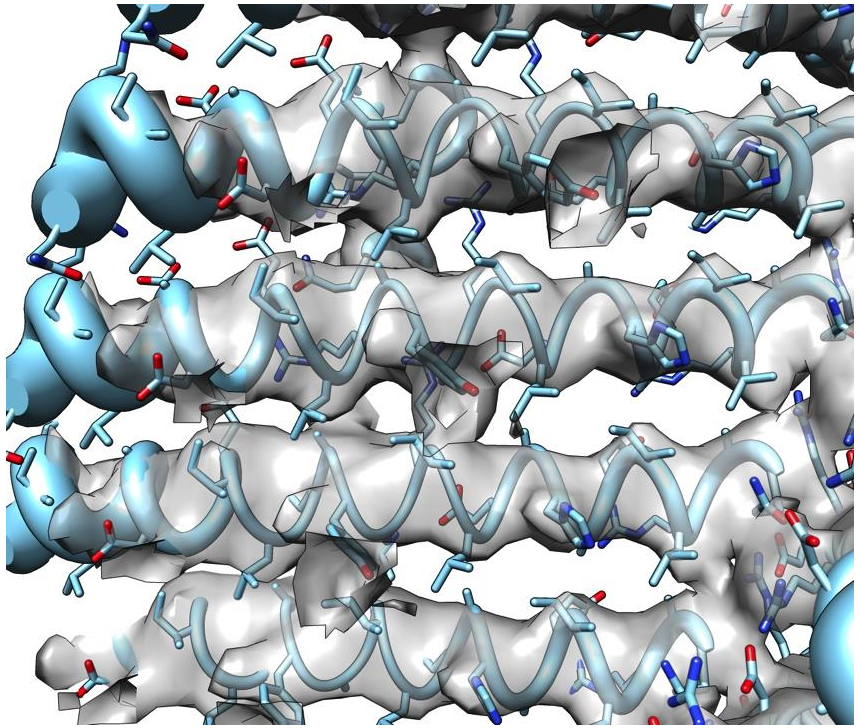
```
usage: $ROSETTA3/source/src/apps/public/symmetry/make_symmdef_file.pl [options]

common options:
  -m (NCS|CRYST|HELIX|PSEUDO) : [default NCS] symmetric mode to run
      NCS: generate noncrystallographic (point) symmetries
             from multiple chains in a PDB file
      CRYST: generate crystallographic symmetry (fixed unit cell)
             from the CRYST1 line in a PDB file
      HELIX: generate helical/fiber symmetry from multiple chains
             in a PDB file
      PSEUDO: (EXPERIMENTAL) generate pseudo-symmetric system
  -p <string> : Input PDB file
  -r <real>   : [default 10.0] the max CA-CA distance between
                two interacting chains
  -f          : [default false] enable fast distance checking
                (recommended for large systems)
```

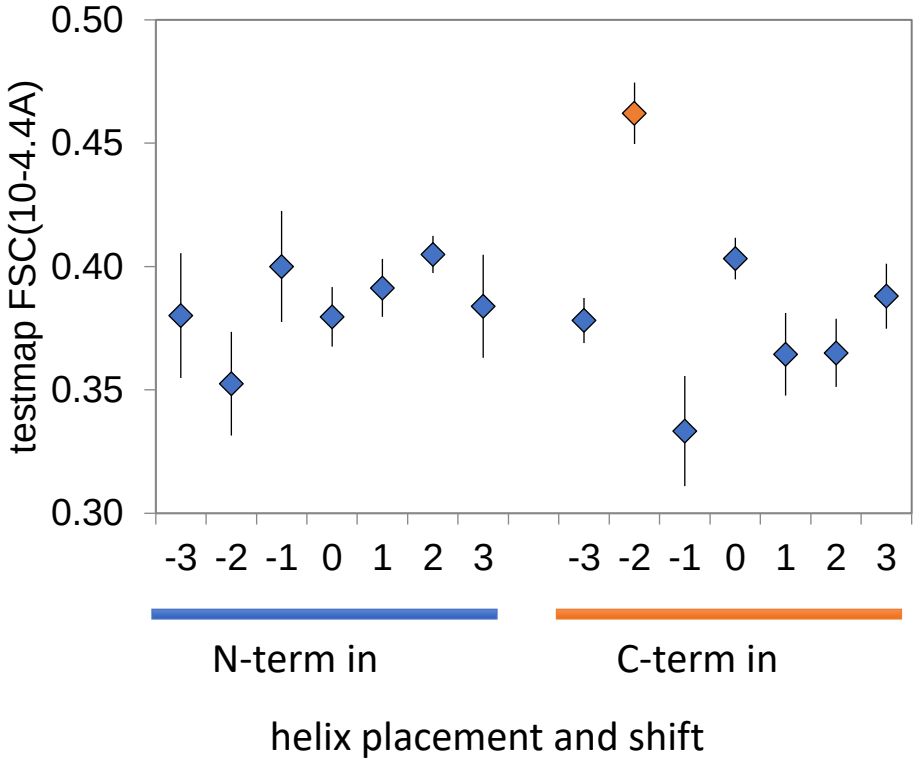
Using symmetry in RosettaCM

```
<Hybridize name="hybridize" stage1_scorefxn="stage1"
   stage2_scorefxn="stage2"
   fa_scorefxn="fullatom" batch="1”>
   <Template pdb="3uoiV_201_aln.pdb" weight="1.0"
      cst_file="AUTO" symmdef="ferritin.symm"/>
   <Template pdb="3gvyC_202_aln.pdb" weight="1.0"
      cst_file="AUTO" symmdef="ferritin.symm"/>
</Hybridize>
```

Demo #3

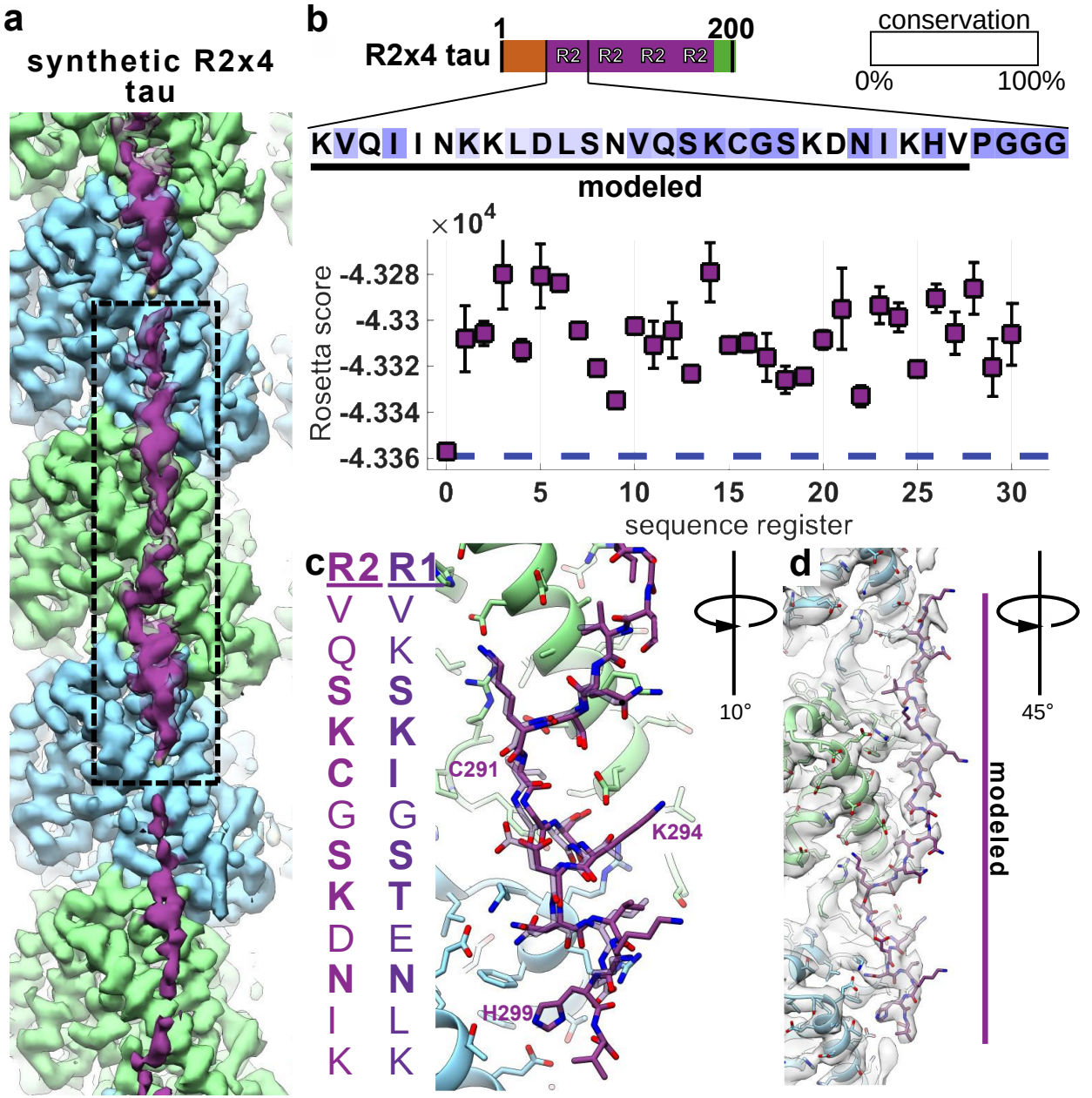
Refinement identifies helix registration of a designed assembly





E Egelman, et al. (2015). *Structure*.

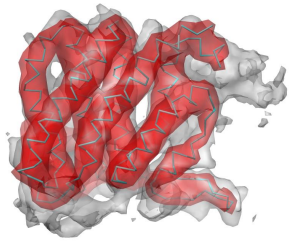
Tau/Tubulin complex



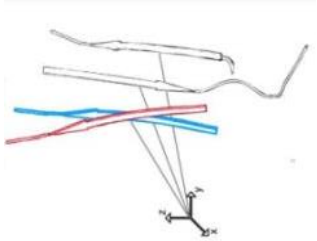
Kellogg *et al*. *Science* 2018

Demo #4

Today's lecture



1) Introduction to the Rosetta relax

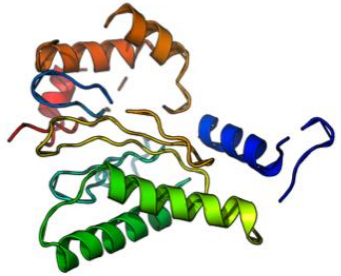


Completing a model with RosettaCM

2) RosettaCM to fix an Alphafold model

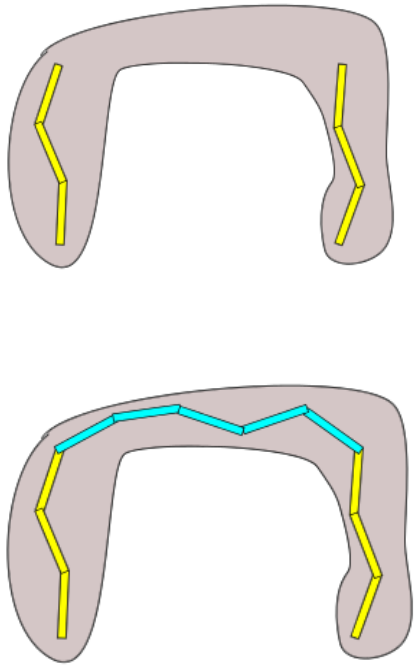
3) Refining a symmetric complex

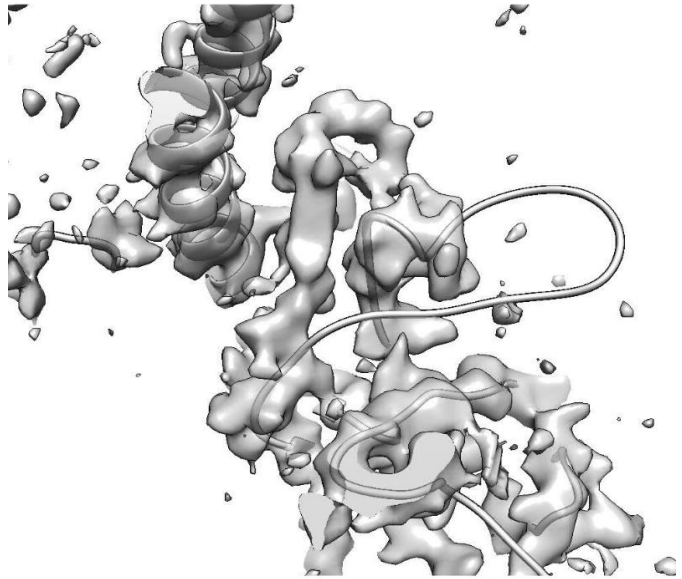
4) Advanced RosettaCM: calculating energetics of alternate threadings



5) Rebuilding large segments with RosettaES

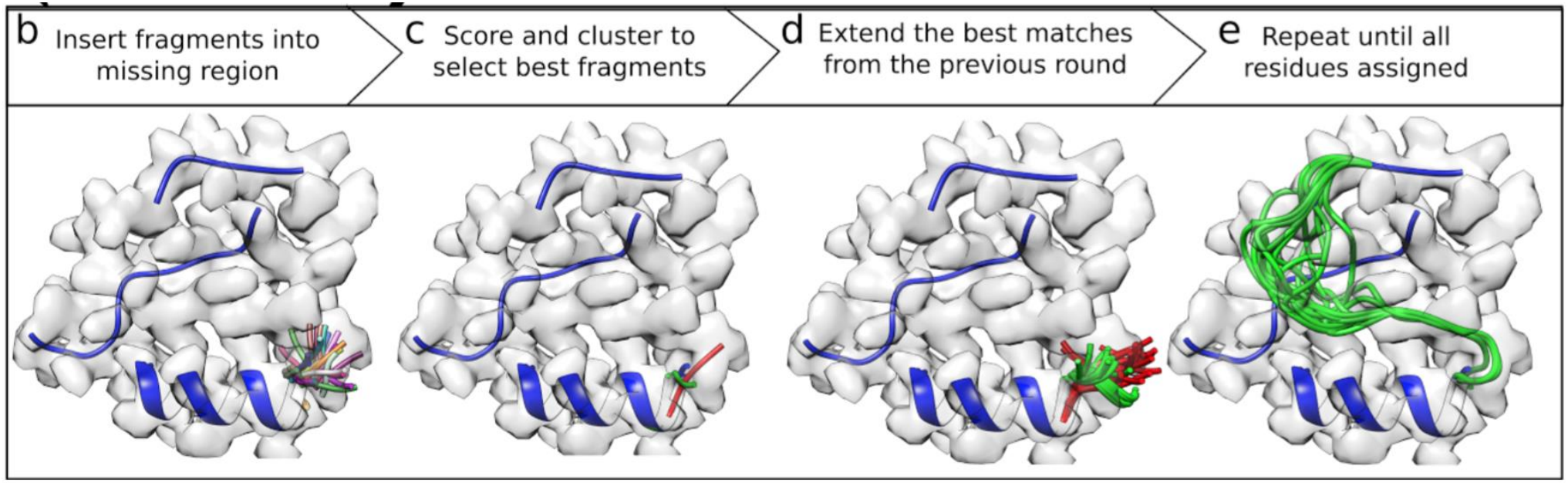
Model completion limits applicability of *de novo* model-building





- Entire missing segment is sampled simultaneously (RosettaCM)
- Much time is spent sampling conformations outside of density

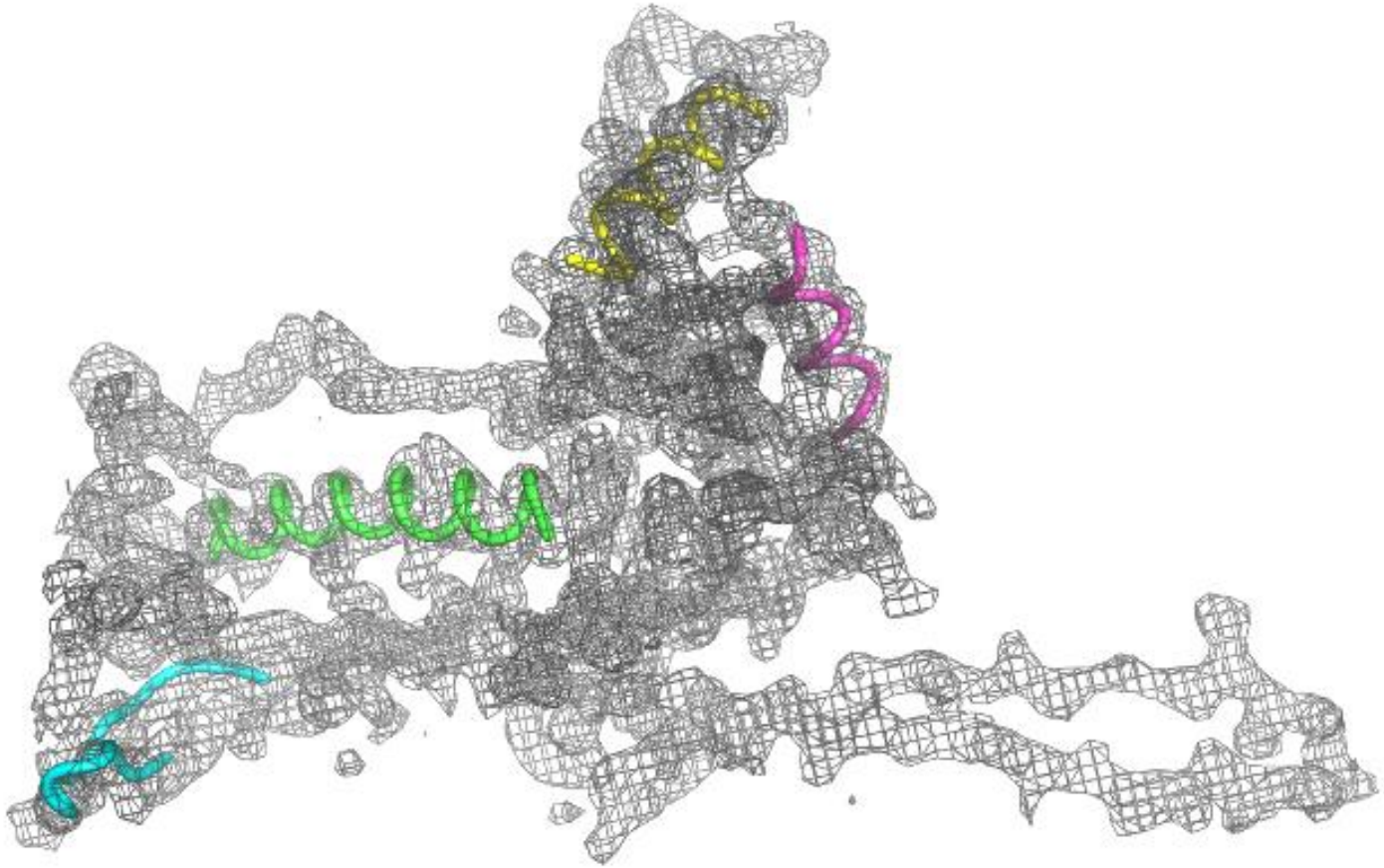
An exhaustive enumeration strategy



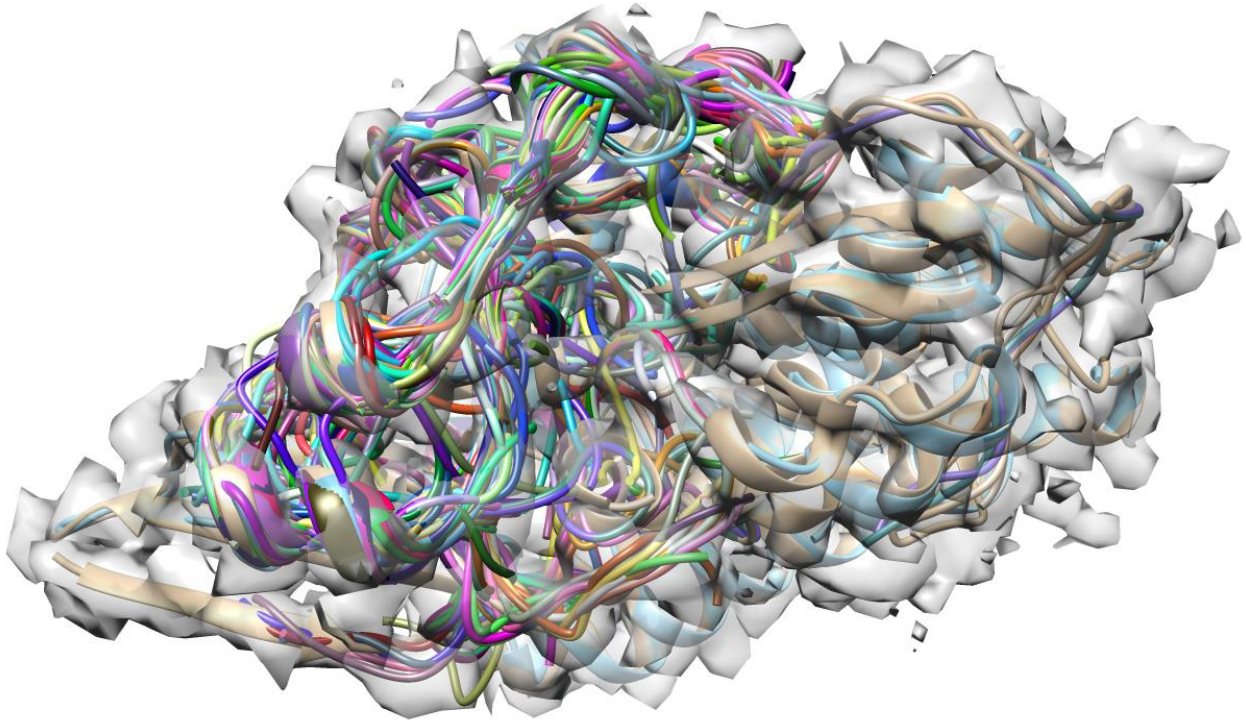
- Maintain a conformational ensemble of structures consistent with the data and local sequence information

Frenz *et al.* *Nature Methods* 2017

An exhaustive enumeration strategy

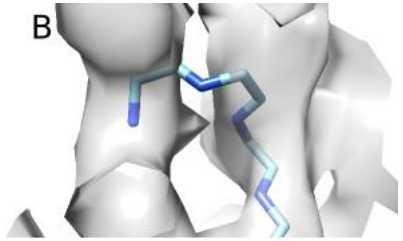


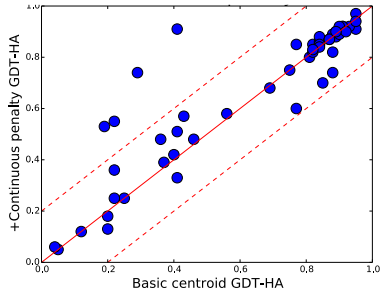
There may be many conformations consistent with the density



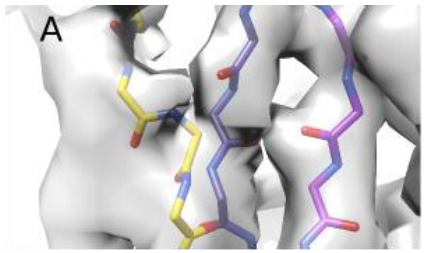
Reduction of search space leads to improvements

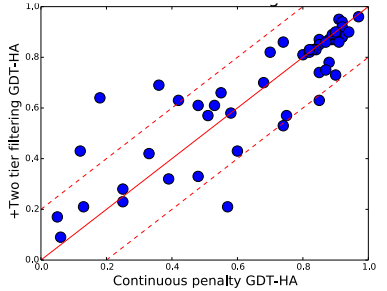
Penalizing density discontinuities



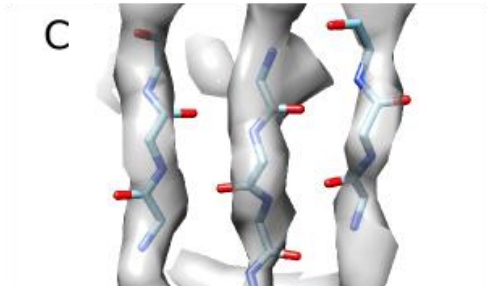


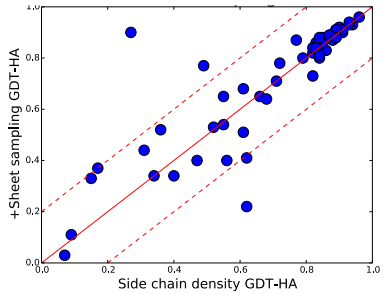
Balancing exploration versus exploitation



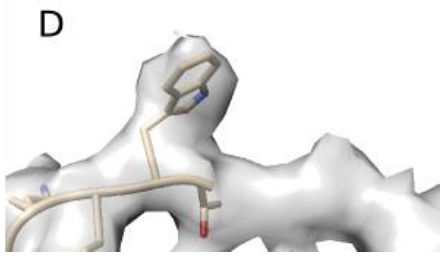


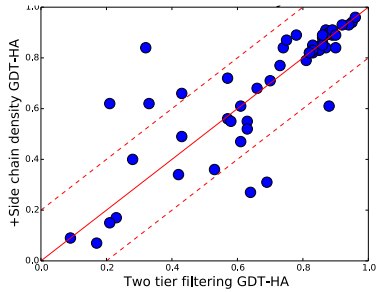
Explicit modeling of beta sheets



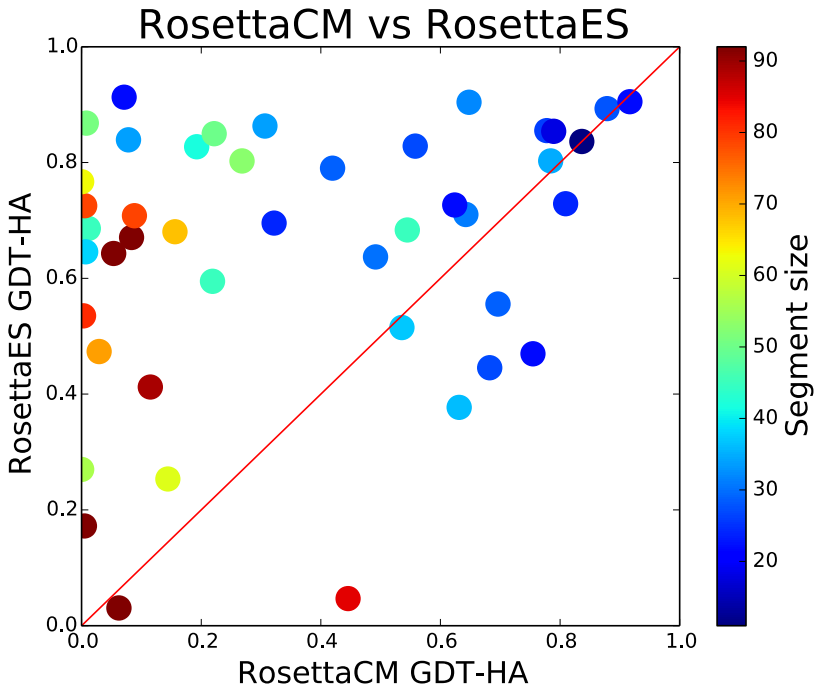


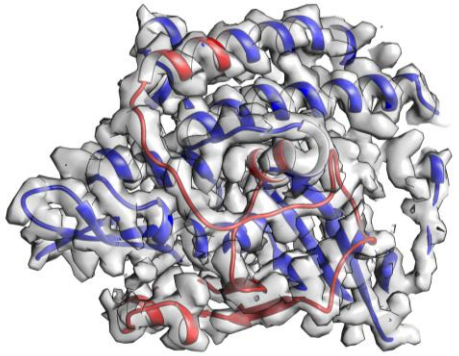
"Rotamer-like" density matching



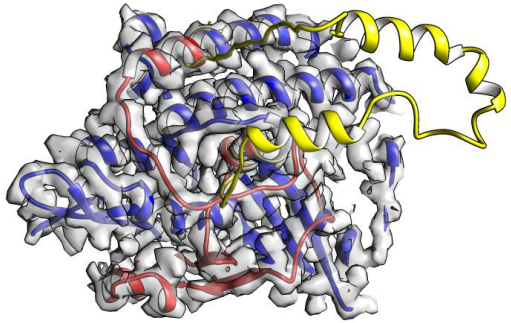


Enumerative backbone sampling

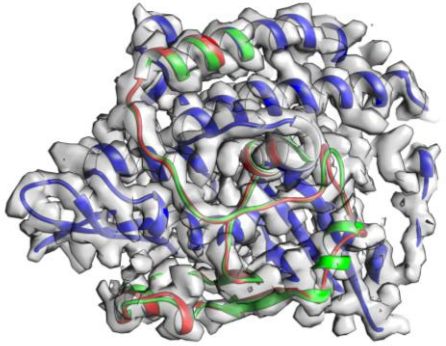




Native structure

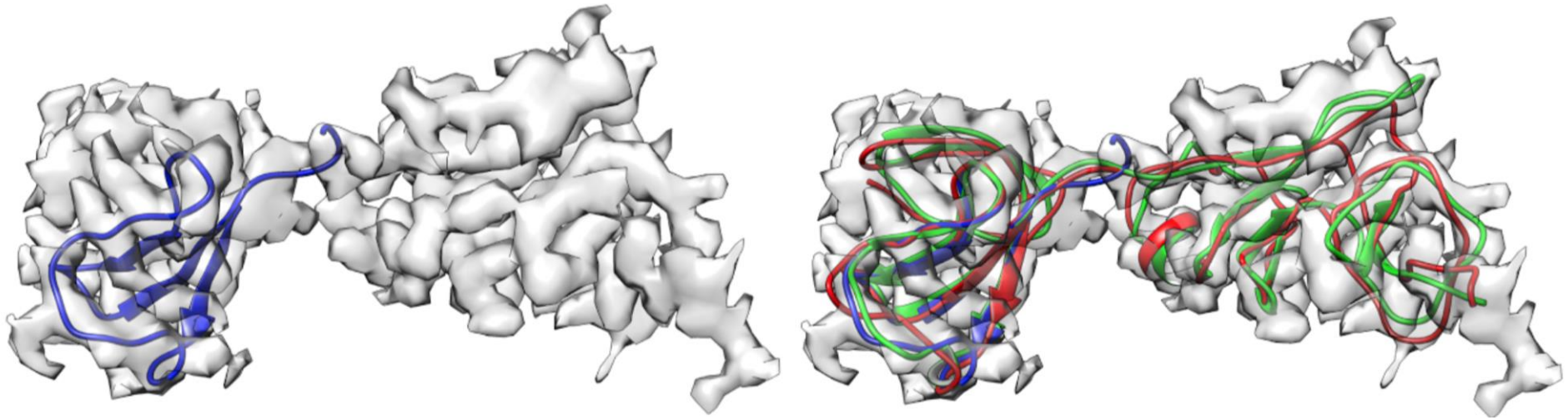


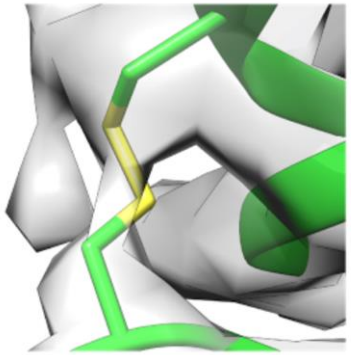
RosettaCM

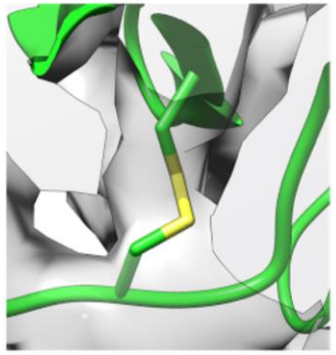


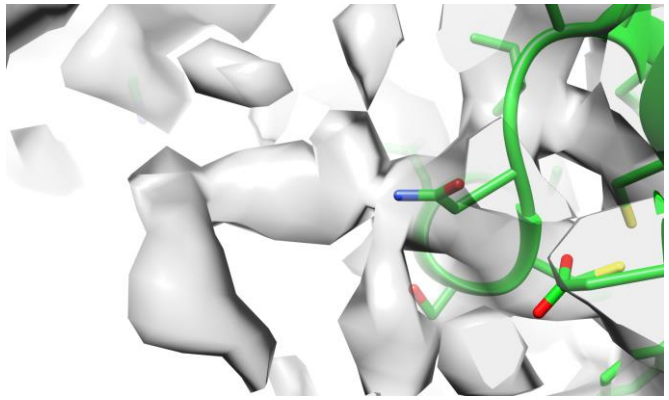
Enumerative Sampling

Modeling the mouse hepatitis virus spike



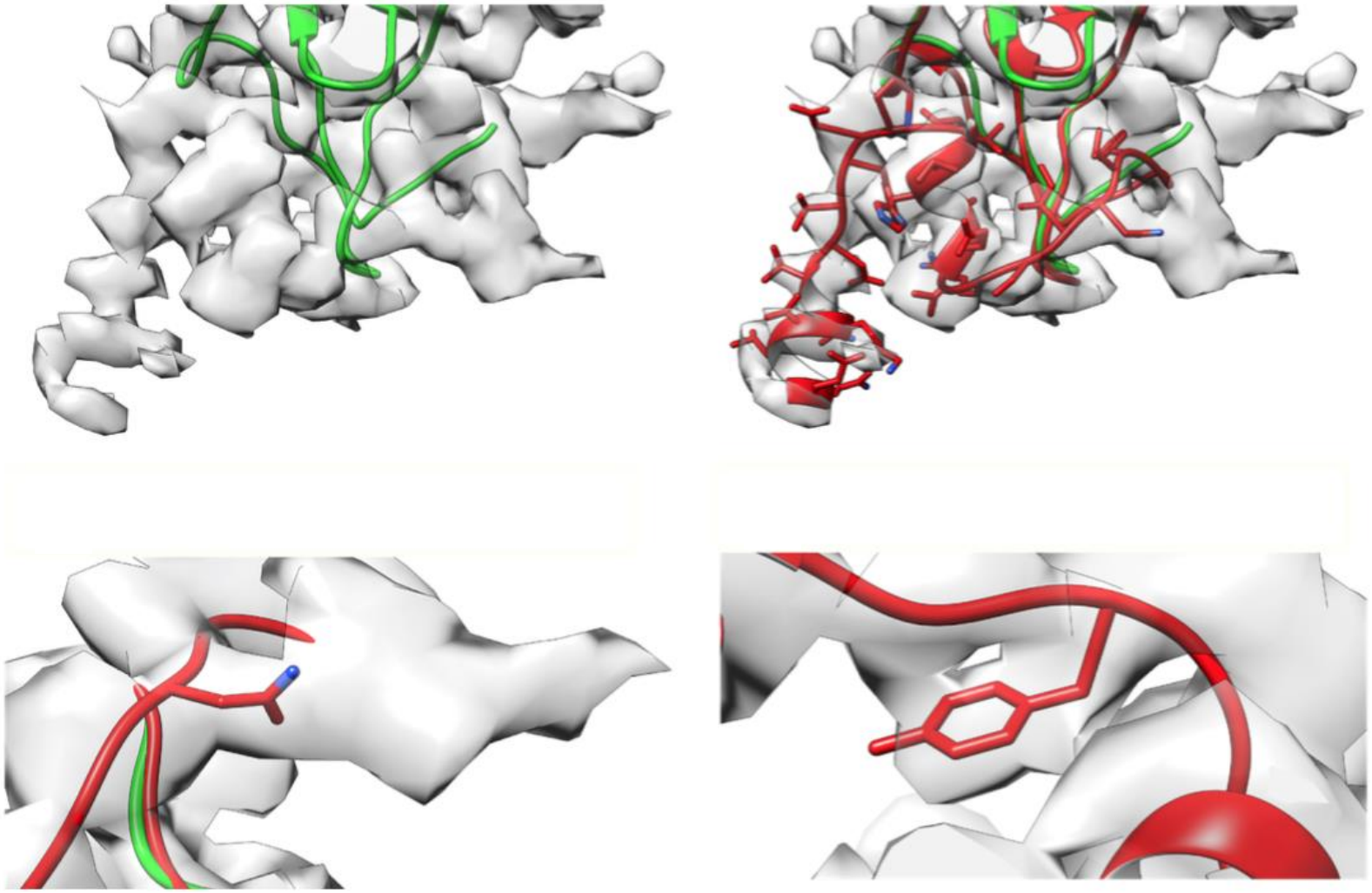






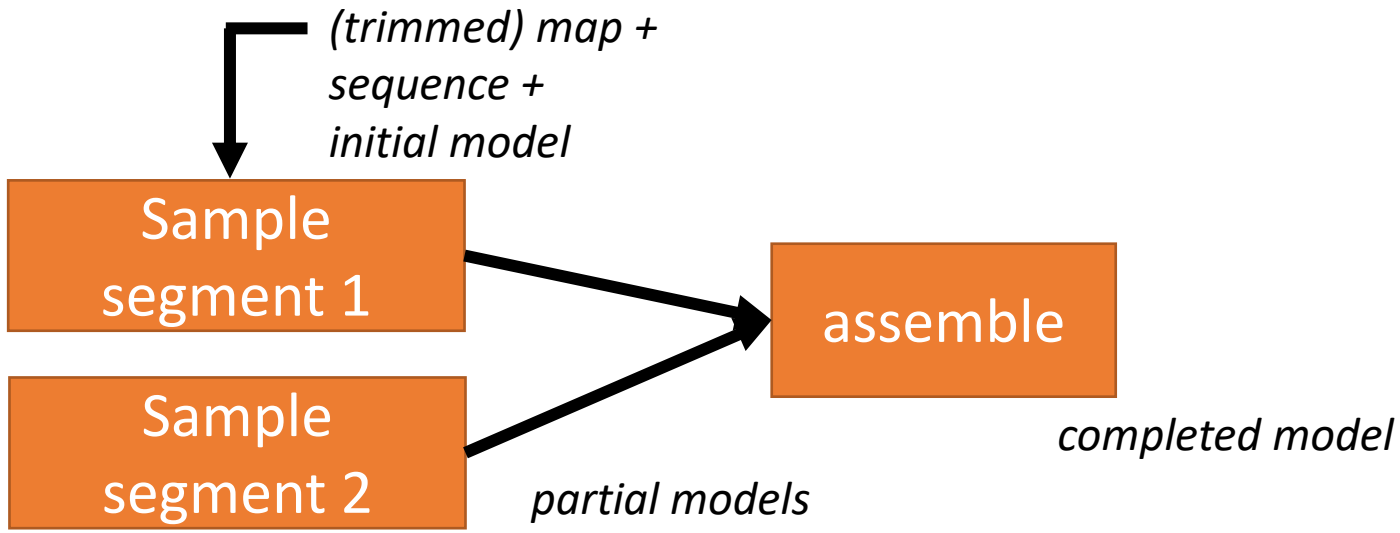
Walls *et al. Nature* (2016)

HCoV-NL63 spike glycoprotein



Walls *et al. NSMB* (2016)

RosettaES segment rebuilding




```
$ROSETTA3/source/scripts/python/public/EnumerativeSampling/RunRosettaES.py \
    -rs runES.sh \
    -x RosettaES.xml \
    -f t2OsA.fasta \
    -p input.pdb \
    -d T2OS_48A_alpha_chainA.mrc \
    -l 1\
    -c 16\
    -n loop_1
```

Demo #5