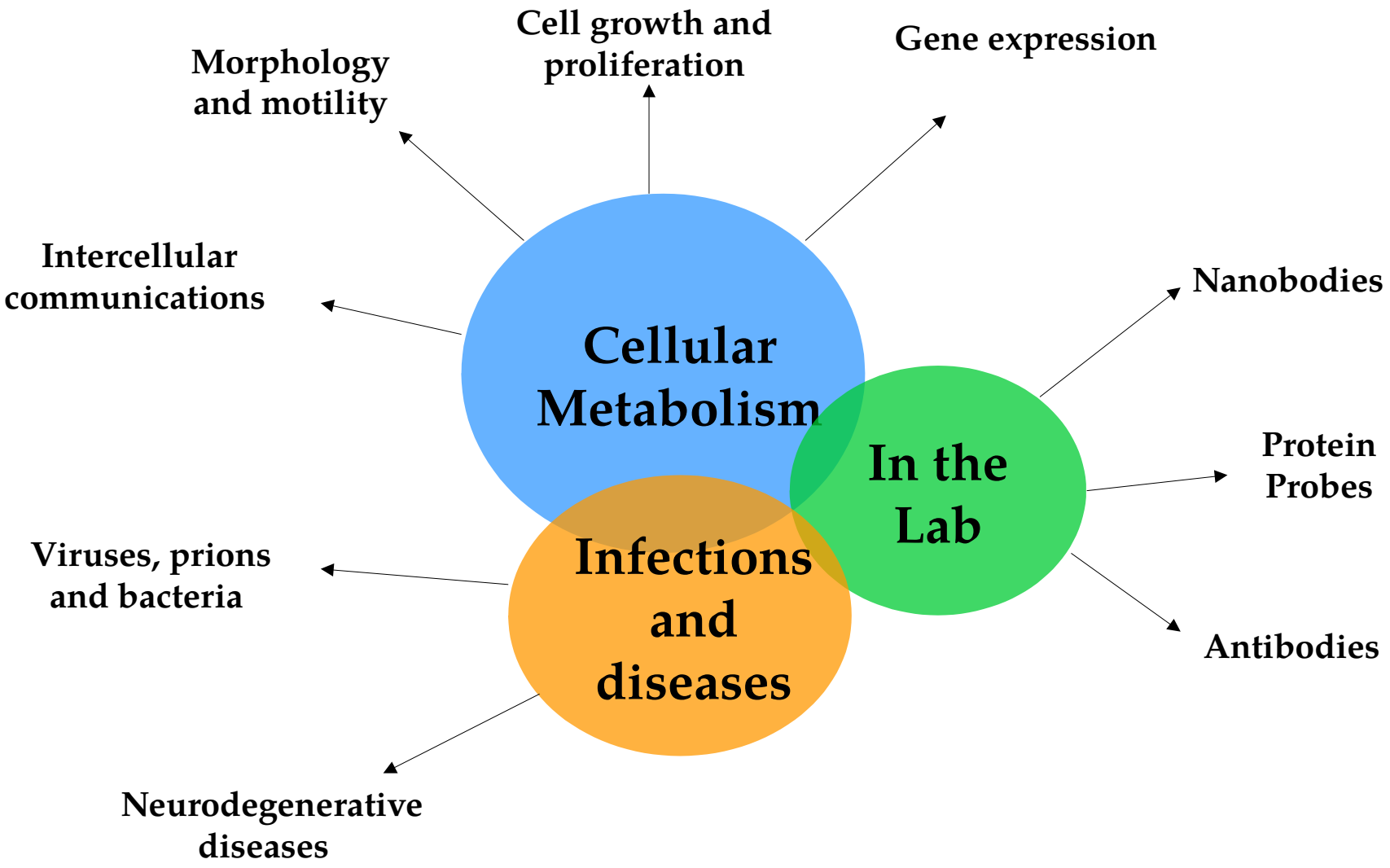
Protein-Protein Docking

VANDERBILT
UNIVERSITY

Cristina Elisa Martina
Rosetta Workshop
Meiler Lab

Protein-Protein Interactions



Protein-Protein Interactions are essential

Protein-Protein Docking in Rosetta

Global Protein-Protein Docking
Short overview in this presentation

Local Protein-Protein Docking
Overview and tutorial

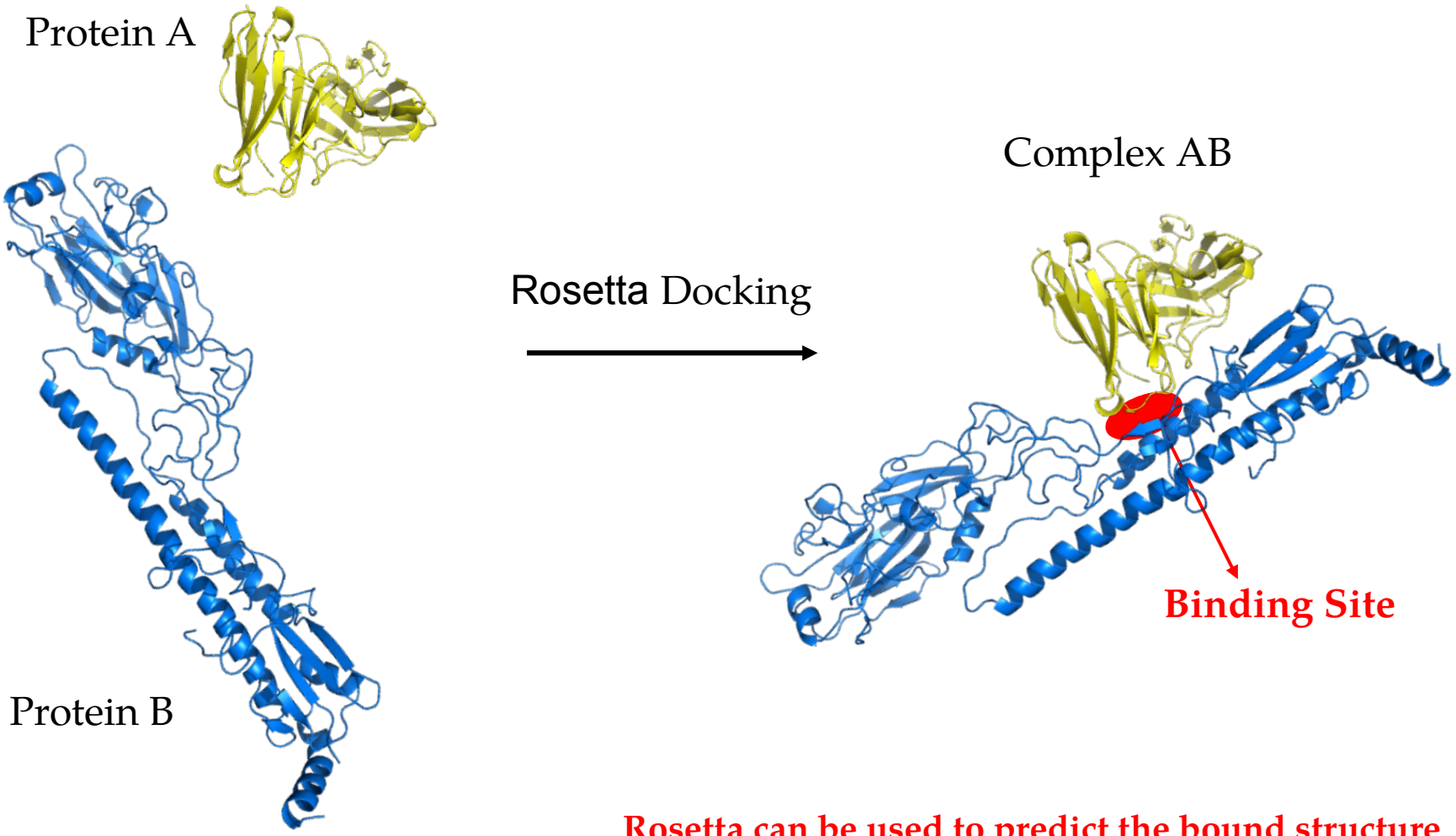
FlexPepDock
For docking of short peptides with increased flexibility

References:
- Alam N., et al., "High-resolution global peptide-protein docking using fragments-based PIPER-FlexPepDock ", 2017, PLOS Computational Biology

SnugDock
For antibody/nanobody docking with increased flexibility at the CDR regions

References:
- Jeliazkov JR., et al., "Robustification of RosettaAntibody and Rosetta SnugDock", 2021, PLOS One

Protein-Protein Docking in Rosetta



Rosetta can be used to predict the bound structure of two proteins starting from unbound structures.

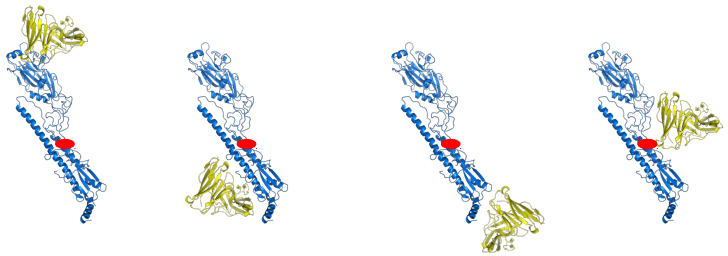
Protein-Protein Docking in Rosetta

Global Docking

Global Docking is used when there is no information about the binding between two proteins, or when the binding is known but not the exact position on the proteins.

Advantages:
- No need of prior info about the proteins

Limitations:
- Only two partners are accepted
- Less accurate than the Local Docking
- Works best for small complexes (<450 aa)

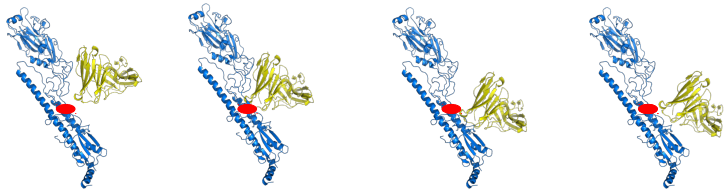


Local Docking

Local Docking is used when the interaction and the position of the two (or more) proteins is known.

Advantages:
- More accurate than the Global Docking
- Multiple partners accepted
- Can integrate ulterior experimental data

Limitations:
- Requires prior information about the binding site



Rosetta can perform global or local docking. In the tutorial we will do a local docking.

Rosetta Protein Docking - Protocol

Local Docking Protocol

Rosetta is very flexible

The general protocol can be adapted to different needs:

Global Docking Protocol
Stage 0 - Randomization of the initial positions
Stage 1 - Low Resolution
Stage 2 - High Resolution

Low Resolution Protocol
Stage 1 only - Low resolution

Local Refinement Protocol
Stage 2 only - High Resolution

Rosetta performs first a low resolution docking and then a high resolution docking.

Local Protein Docking - Protocol

Inputs Preparation

Stage 1
Low Resolution

Stage 2
High Resolution

Outputs Analysis

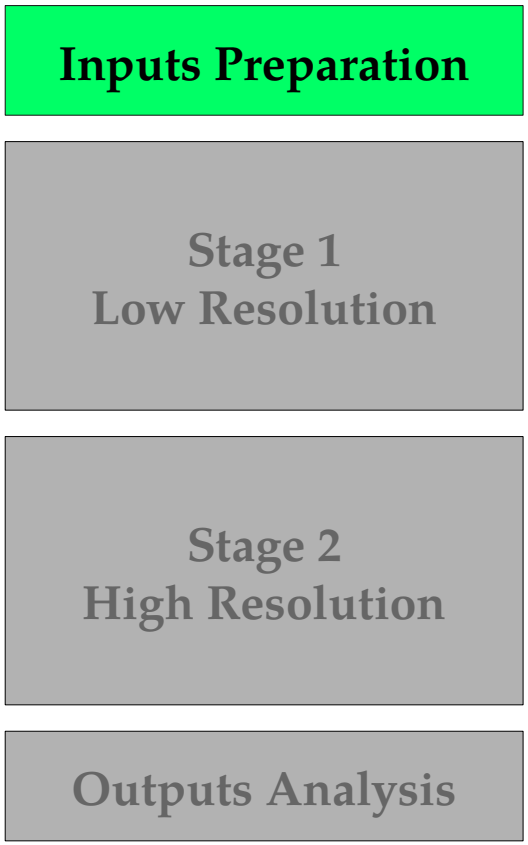
1- The PDB structure:
The two (or more) partners has to be in the same pdb file, with different chain names, and within 10 Å distance at the binding pocket site.

The starting structure must be prepacked in order to lower the energy of the side-chains outside of the docking interface...

Extra steps might be required to prepare the partners:
- reducing size to reduce the calculation time
- closing breakchains / modeling loops
- preparing ensembles of conformers

The docking protocol requires a single pdb with all partners in a close distance.

Local Protein Docking - Protocol



2- The XML file:
The Rosetta protocol instruction file. We will see it more in detail in the next slides.

3- Other files:
Other instructions (rather than the protocol itself) can be included to tune the Rosetta docking run. These might include the option file and the constraints file, usually used when experimental data have to be included in the run.

Experimental information can be included in the docking run.

Local Protein Docking - Protocol

Input Preparation

Stage 1
Low Resolution

Stage 2
High Resolution

Outputs Analysis

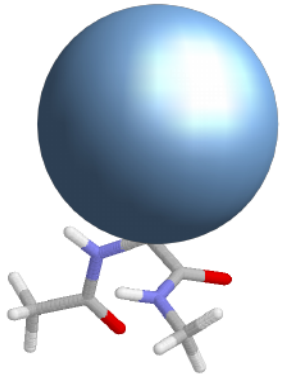
Low Resolution Docking:
Centroid-mode / Coarse-grain

Side-chains are represented by spheres with similar properties (charge, size, ...).

In this stage, Rosetta attempts to find the rough orientation of the docking partners.

Advantages:
- Faster calculation time

Limitations:
- Lower accuracy



The low-resolution stage define the initial orientation of the docking partners.

Local Protein Docking - Protocol

Input Preparation

Stage 1
Low Resolution

Stage 2
High Resolution

Outputs Analysis

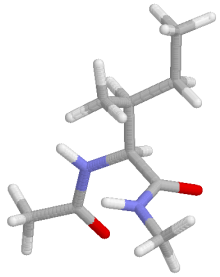
High Resolution Docking:
All-Atom / Full-Atom

Centroid residues are replaced with the side-chains atoms in unbound conformation.

Rotamers are tested and, if accepted, the complex is minimized and repacked.

Advantages:
- More accurate than centroid-mode

Limitations:
- Requires higher calculation time



The high resolution stage includes minimization and repacking.

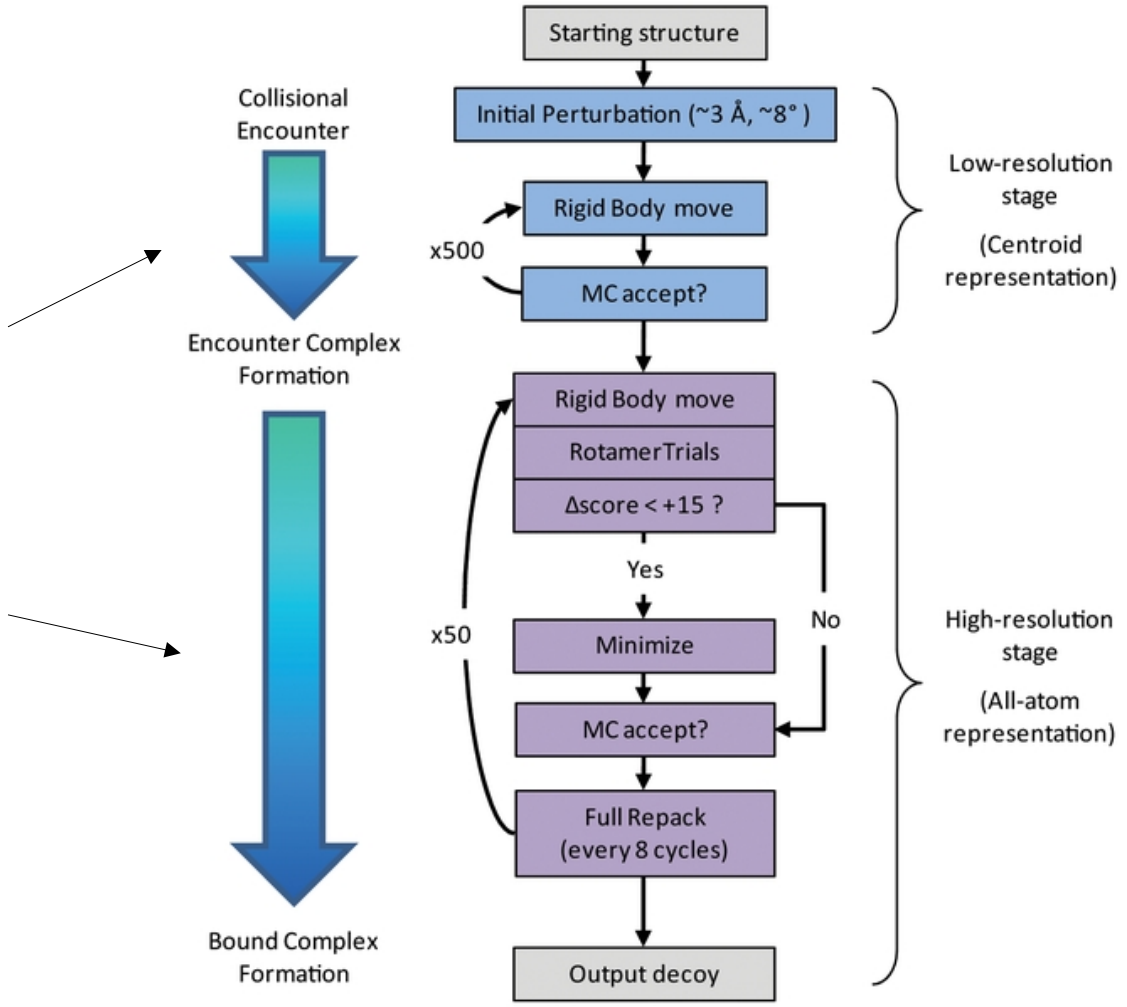
Local Protein Docking - Protocol

Input Preparation

Stage 1
Low Resolution

Stage 2
High Resolution

Outputs Analysis



Rosetta performs multiple cycles for both the low- and the high-resolution stages.

Local Protein Docking - Protocol

General Protocol:

At least 500 output structures should be requested for a local docking protocol (1000 recommended).

Not all the requested outputs will necessarily be generated: if one of the run fails at any stage of the protocol, the output will not be created and Rosetta will restart with the next run.

Local Protein Docking - Protocol

Input Preparation

Stage 1
Low Resolution

Stage 2
High Resolution

Outputs Analysis

Interface Analyzer:

The Rosetta InterfaceAnalyzer mover can give you many information about the complex structure, including:
- the binding energy of the two partners
- the residues involved in the interaction
- the RMSD between the model and the native structure.

The Interface Analyzer mover is used after completation of the docking protocol, when all the outputs have been generated.

The requested inputs are:
- a list with all the output pdb (from docking)
- an XML file with the Interface Analyzer protocol
- the native strucutre for calculation of the RMSD
- other files (i.e. option file)

Interface Analyzer gives information about the binding interface.

Local Protein Docking - Protocol

Input Preparation

Stage 1
Low Resolution

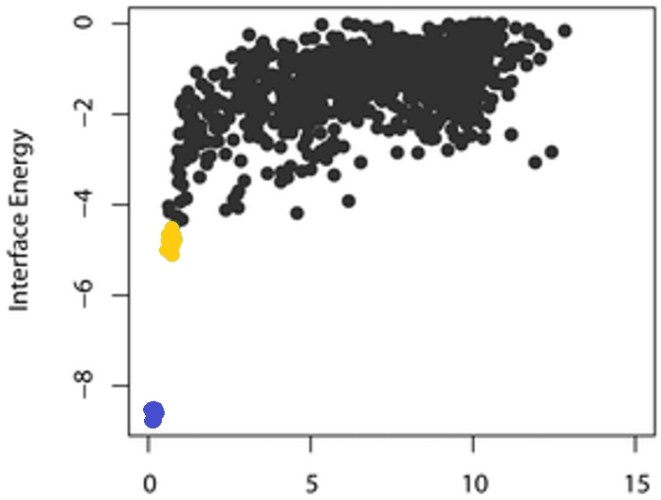
Stage 2
High Resolution

Outputs Analysis

From the interface analyzer we can extract and plot:

1- RMSD vs Total Energy
2- RMSD vs Binding Energy

For both we expect a funnel-like plot, in which the lower scoring models should be close to the native sturcture in terms of energy and conformation.



All models
Top models
Native complex

RMSD vs Total Energy and RMSD vs Binding Energy should give a funnel-plot.

Local Protein Docking - XML file

```
<ROSETTASCRIPTS>
        <TASKOPERATIONS>
                <InitializeFromCommandline name="ifcl"/>
                <RestrictToRepacking name="rtr" />
                Restrict to residues within a distance and vector cutoff of the protein-protein interface
                <RestrictToInterfaceVector name="rtiv" chain1_num="1,2" chain2_num="3,4" CB_dist_cutoff="10.0" nearby_atom_cutoff="5.5" vector_angle_cutoff="75" vector_dist_cutoff="9.0" />
                Fix residues known experimentally to be critical in interaction
                <PreventResiduesFromRepacking name="prfrp" residues="11,41,345" />
        </TASKOPERATIONS>
        <MOVERS>
                MINIMIZATION MOVERS
                Single cycle of FastRelax to minimize backbone of docking partners
                <FastRelax name="minimize_interface" scorefxn="REF2015" repeats="1" task_operations="ifcl,rtr,rtiv,prfrp" />
                DOCKING MOVERS
                <Docking name="dock_low" score_low="score_docking_low" score_high="REF2015" fullatom="0" local_refine="0" optimize_fold_tree="1" conserve_foldtree="0" ignore_default_docking_task="0" design="0" task_operations="ifcl,prfrp" jumps="1"/>
                <Docking name="dock_high" score_low="score_docking_low" score_high="REF2015" fullatom="1" local_refine="1" optimize_fold_tree="1" conserve_foldtree="0" design="0" task_operations="ifcl,prfrp" jumps="1"/>

                <SaveAndRetrieveSidechains name="srsc" allsc="0" /> Speeds the move from centroid to full atom mode
        </MOVERS>
        <PROTOCOLS>
                Run docking protocol
                <Add mover="dock_low"/>
                <Add mover="srsc" />
                <Add mover="dock_high" />
                Minimize interface
                <Add mover="minimize_interface" />
        </PROTOCOLS>
        <OUTPUT scorefxn="REF2015" />
</ROSETTASCRIPTS>
```

Blue	= protocol name (fix)
Green	= options (fix)
"Yellow"	= values (edit)
White	= comments (edit)

Local Protein Docking - XML file

```
<ROSETTASCRIPTS>
        <TASKOPERATIONS>
                <InitializeFromCommandline name="ifcl"/>
                <RestrictToRepacking name="rtr" />
                Restrict to repacking within a certain distance cutoff of the protein-protein interface
                <RestrictInterfaceVector name="rtiv" chain1_num="1" chain2_num="3,4" distance_cutoff="10.0"
 nearby_atom_cutoff="5.5" vector_angle_cutoff="75" vector_dist_cutoff="9.0" />
                Fix residues known experimentally to be critical in interaction
                <PreventResiduesFromRepacking name="prfrp" residues="11,41,345" />
        </TASKOPERATIONS>
        <MOVERS>
                MINIMIZATION MOVERS
                Single cycle of FastRelax to minimize backbone of docking partners
                <FastRelax name="minimize_interface" scorefxn="REF2015" repeats="1" task_operations="ifcl,rtr,
rtiv,prfrp" />
                DOCKING MOVERS
                <Docking name="dock_low" score_low="score_docking_low" score_high="REF2015" fullatom="0" local
_refine="0" optimize_fold_tree="1" conserve_foldtree="0" score_filter="default_docking_task="0" design="0" task_opera
tions="ifcl,prfrp" jumps="1"/>
                <Docking name="dock_high" score_low="score_docking_low" score_high="REF2015" fullatom="1" loca
l_refine="1" optimize_fold_tree="1" conserve_foldtree="0" design="0" task_operations="ifcl,prfrp" jumps="1"/>

                <SaveAndRetrieveSidechains name="srsc" allsc="0" /> Speeds the move from centroid to full atom
 mode
        </MOVERS>
        <PROTOCOLS>
                Run docking protocol
                <Add mover="dock_low"/>
                <Add mover="srsc" />
                <Add mover="dock_high" />
                Minimize interface
                <Add mover="minimize_interface" />
        </PROTOCOLS>
        <OUTPUT scorefxn="REF2015" />
</ROSETTASCRIPTS>
```

3- TASK OPERATIONS

2- MOVERS

1- PROTOCOL

XML file - Protocol

```
28          <PROTOCOLS>
29                  Run docking protocol
30                  <Add mover="dock_low"/>
31                  <Add mover="srsc" />
32                  <Add mover="dock_high" />
33
34                  Minimize interface
35                  <Add mover="minimize_interface" />
36          </PROTOCOLS>
```

Rosetta will perform the task contained in the protocol section in order:
1- **dock_low** (low resolution docking)
2- **srsc** (to speed up the step from centroid-mode to full-atom)
3- **dock_high** (high resolution docking)
4- **minimize_interface** (minimization of the interface residues only)

The information relative to each step can be found in the “movers” section.

XML file - Movers

```
14        <MOVERS>
15                MINIMIZATION MOVERS
16                Single cycle of FastRelax to minimize backbone of docking partners
17                <FastRelax name="minimize_interface" scorefxn="REF2015" repeats="1" task_operations="ifcl,rtr,rtiv,prfrp" />
18
19                DOCKING MOVERS
20                <Docking name="dock_low" score_low="score_docking_low" score_high="REF2015" fullatom="0" local_refine="0" optimiz
  e_fold_tree="1" conserve_foldtree="0" ignore_default_docking_task="0" design="0" task_operations="ifcl,prfrp" jumps="1"/>
21                <Docking name="dock_high" score_low="score_docking_low" score_high="REF2015" fullatom="1" local_refine="1" optimi
  ze_fold_tree="1" conserve_foldtree="0" design="0" task_operations="ifcl,prfrp" jumps="1"/>
22
23                <SaveAndRetrieveSidechains name="srsc" allsc="0" /> Speeds the move from centroid to full atom mode
24
25        </MOVERS>
```

In the "Movers" section, the order of the tasks is not taken in account. Here we found details regarding the different steps, including for example the scoring functions to be used (score_docking_low for the low resolution step, REF2015 for the high resolution step and the interface minimization).

Three out of the four movers recall the "task_operation" that are defined in the last section of the .xml file.

More information about the available movers can be found online:
https://www.rosettacommons.org/docs/latest/scripting_documentation/RosettaScripts/Movers/Movers-RosettaScripts

XML file - Task Operation

```
4         <TASKOPERATIONS>
5                 <InitializeFromCommandline name="ifcl"/>
6                 <RestrictToRepacking name="rtr" />
7                 Restrict to residues within a distance and vector cutoff of the protein-protein interface
8                 <RestrictToInterfaceVector name="rtiv" chain1_num="1,2" chain2_num="3,4" CB_dist_cutoff="10.0" nearby_atom_cutoff
  ="5.5" vector_angle_cutoff="75" vector_dist_cutoff="9.0" />
9                 Fix residues known experimentally to be critical in interaction
10                <PreventResiduesFromRepacking name="prfrp" residues="11,41,345" />
11        </TASKOPERATIONS>
```

The task operation section defines other useful commands:

InitializeFromCommandline -> accept option from the commandline
RestrictToRepacking -> necessary to avoid re-design of interface residues
RestrictToInterfaceVector -> restrict to residues at the interface, based on distance and vector cut-off.
PreventResiduesFromRepacking -> experimentally relevant residues will not repack

More information about different Task Operation can be found online:
https://www.rosettacommons.org/docs/latest/scripting_documentation/RosettaScripts/TaskOperations/TaskOperations-RosettaScripts

Local Protein Docking - Option file

```
-docking                                    # the docking option group
        -partners AB_HL                     # set rigid body docking partners
        -dock_pert 3 8                      # set coarse perturbation parameters (degrees and angstroms)
        -dock_mcm_trans_magnitude 0.1       # refinement translational perturbation
        -dock_mcm_rot_magnitude 5.0         # refinement rotational perturbation
-s 3gbm_HA_3gbn_Ab.pdb                      # input model
-run:max_retry_job 10                       # if the mover fails, retry 50 times
-use_input_sc                               # add the side chains from the input pdb to the rotamer library
-ex1                                        # increase rotamer bins to include mean +- 1 standard deviation
-ex2                                        # increase rotamer bins to include mean +- 2 standard deviations
-out                                        # out option group
        -file                               # out:file option group
                -scorefile docking.fasc     # the name of the model score file
-score:weights ref2015.wts                  # Set ref2015 as default score function
```

In the option file we can find importan information such as:
- the binding partners (`-partners AB_HL`)
- the input structure (`-s 3gbm_HA_3gbn_Ab.pdb`)
- the rotation and translation values (`-dock_pert 3 8`)
- the output file (`-scorefile docking.fasc`)

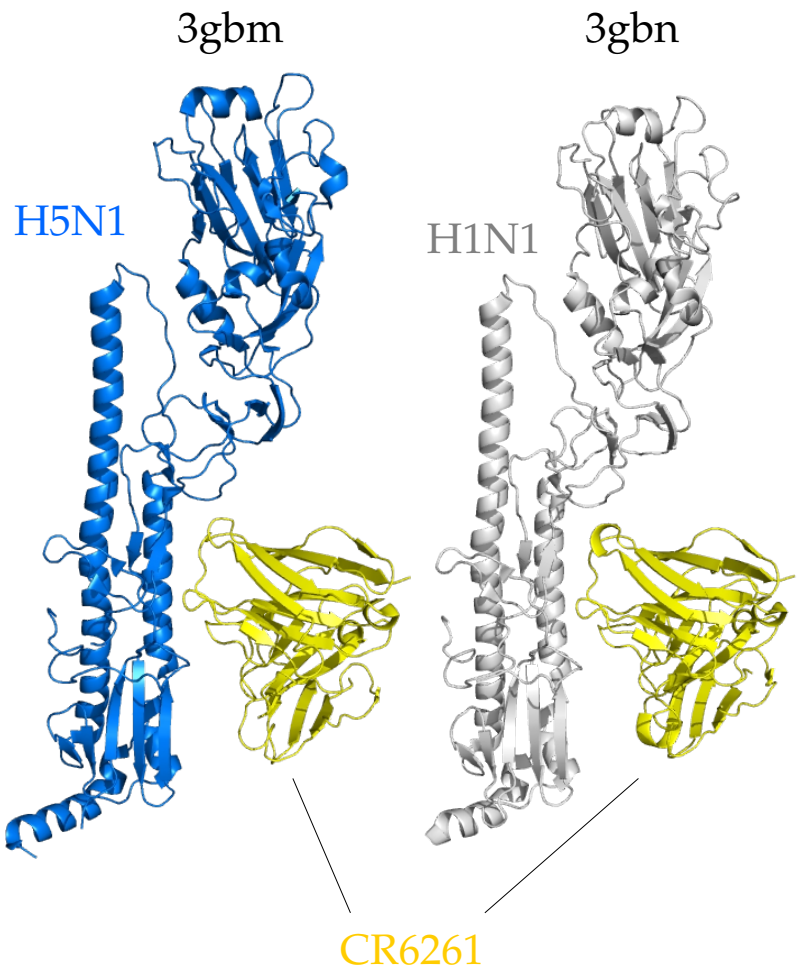
More information about different options can be found online:
https://www.rosettacommons.org/docs/latest/full-options-list

The protein-protein docking tutorial:

The antibody CR6261 binds to multiple sub-types of influenza antigen hemagglutinin (HA). In the crystal structure 3gbm, the antibody is bound to the sub-type H5N1, while in the structure 3gbm it is bound to H1N1.

The sequence of the antibody is the same in the two structures, but the conformations are slightly different.

Here we will perform a cross-docking experiment, in which we will dock the CR6261 protein from 3gbn to the H5N1 structure of 3gbm.



The protein-protein docking tutorial:

Input Preparation:
- Download the pdbs
- Clean the pdbs
- Close a chain-break
- Repack the structures
- Orient the structures

Rosetta Docking:
- Generate docking models
- Minimize the native structure for comparison

Analysis of the Outputs:
- Perform Interface Analyzer
- Plot RMSD vs Total Energy
- Plot RMSD vs Binding Energy

Bibliography - Docking:

- Gray, J. J.; Moughon, S.; Wang, C.; Schueler-Furman, O.; Kuhlman, B.; Rohl, C. A.; Baker, D., **Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations**. *Journal of molecular biology* **2003,** *331* (1), 281-99.
- Daily, M. D.; Masica, D.; Sivasubramanian, A.; Somarouthu, S.; Gray, J. J., **CAPRI rounds 3-5 reveal promising successes and future challenges for RosettaDock**. *Proteins* **2005,** *60* (2), 181-6.
- Chaudhury, S.; Sircar, A.; Sivasubramanian, A.; Berrondo, M.; Gray, J. J., **Incorporating biochemical information and backbone flexibility in RosettaDock for CAPRI rounds 6-12**. *Proteins* **2007,** *69* (4), 793-800.
- Sivasubramanian, A.; Sircar, A.; Chaudhury, S.; Gray, J. J., **Toward high-resolution homology modeling of antibody Fv regions and application to antibody-antigen docking**. *Proteins* **2009,** *74* (2), 497-514.
- Sircar, A.; Chaudhury, S.; Kilambi, K. P.; Berrondo, M.; Gray, J. J., **A generalized approach to sampling backbone conformations with RosettaDock for CAPRI rounds 13-19**. *Proteins* **2010,** *78* (15), 3115-23.
- Sircar, A.; Gray, J. J., **SnugDock: paratope structural optimization during antibody-antigen docking compensates for errors in antibody homology models.** *PLoS computational biology* **2010,** *6* (1), e1000644.
- Chaudhury, S., Berrondo, M., Weitzner, B. D., Muthu, P., Bergman, H., Gray, J. J.; **Benchmarking and analysis of protein docking performance in RosettaDock** v3.2. *PLoS One* **2011**.
- Marze, N. A., Jeliazkov, J. R., Roy Burman, S. S., Boyken, S. E., DiMaio, F., Gray, J. J.; **Modeling oblong proteins and water-mediated interfaces with RosettaDock in CAPRI rounds 28-35**. *Proteins* **2016**, 85(3):479-486.