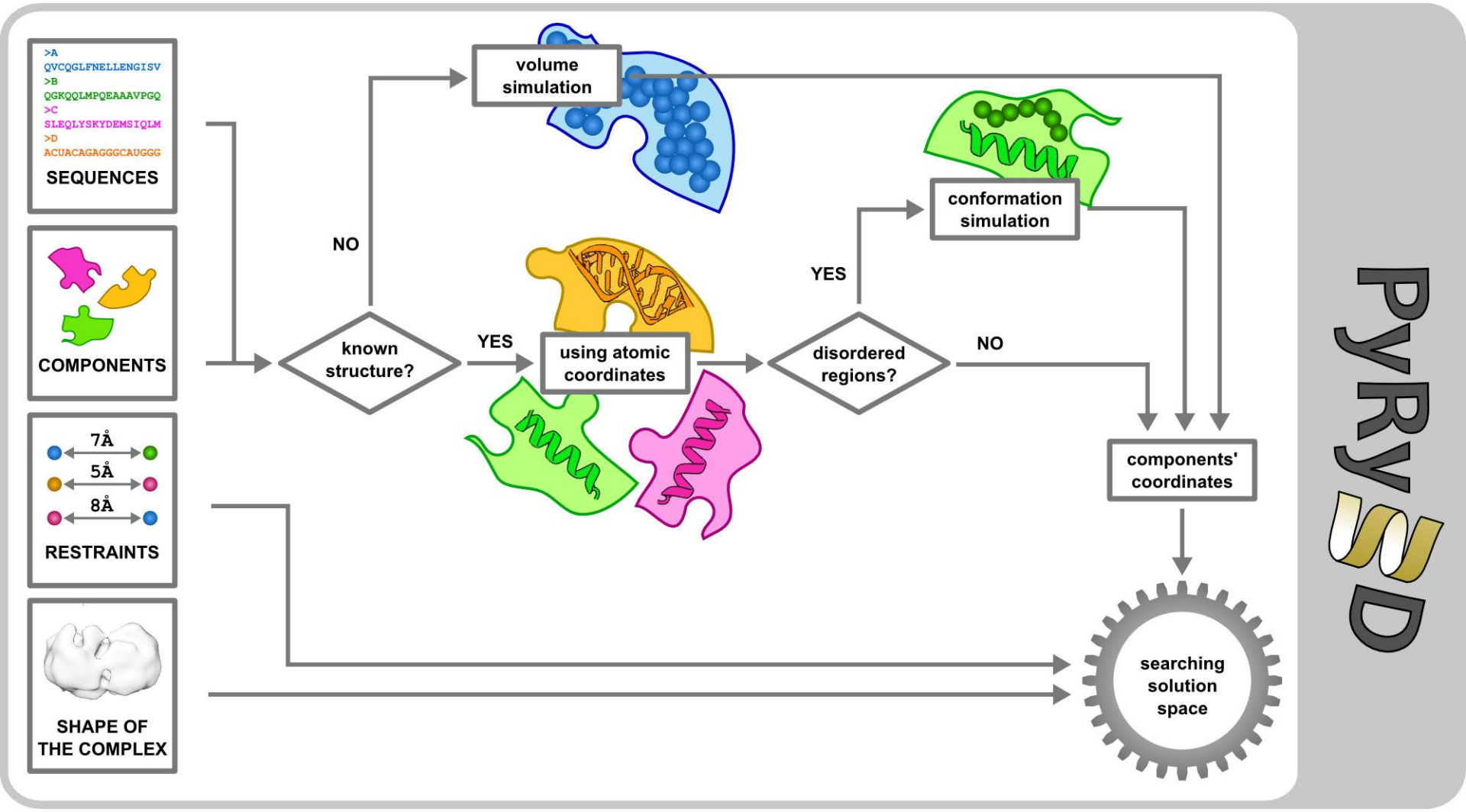
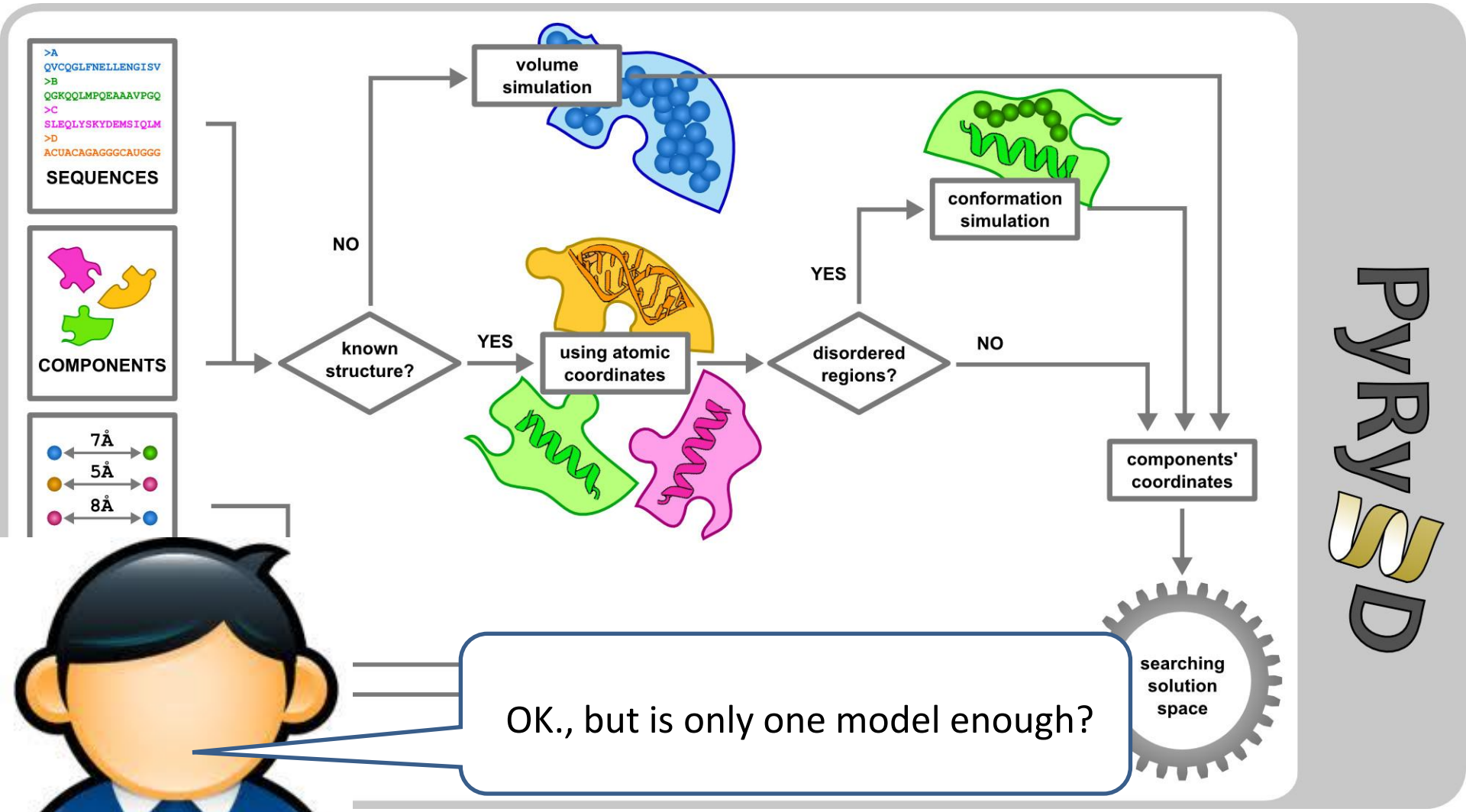
integrative modeling with PyRy3D step by step

PyRy3D motivation

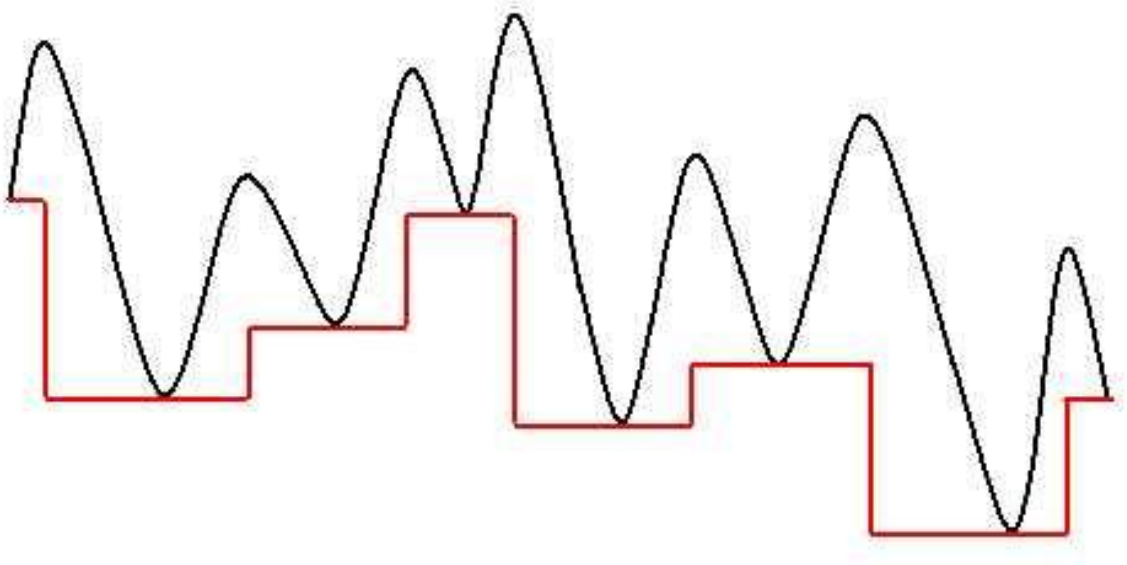


PyRy3D motivation



OK., but is only one model enough?

Monte Carlo



Monte Carlo simulations:

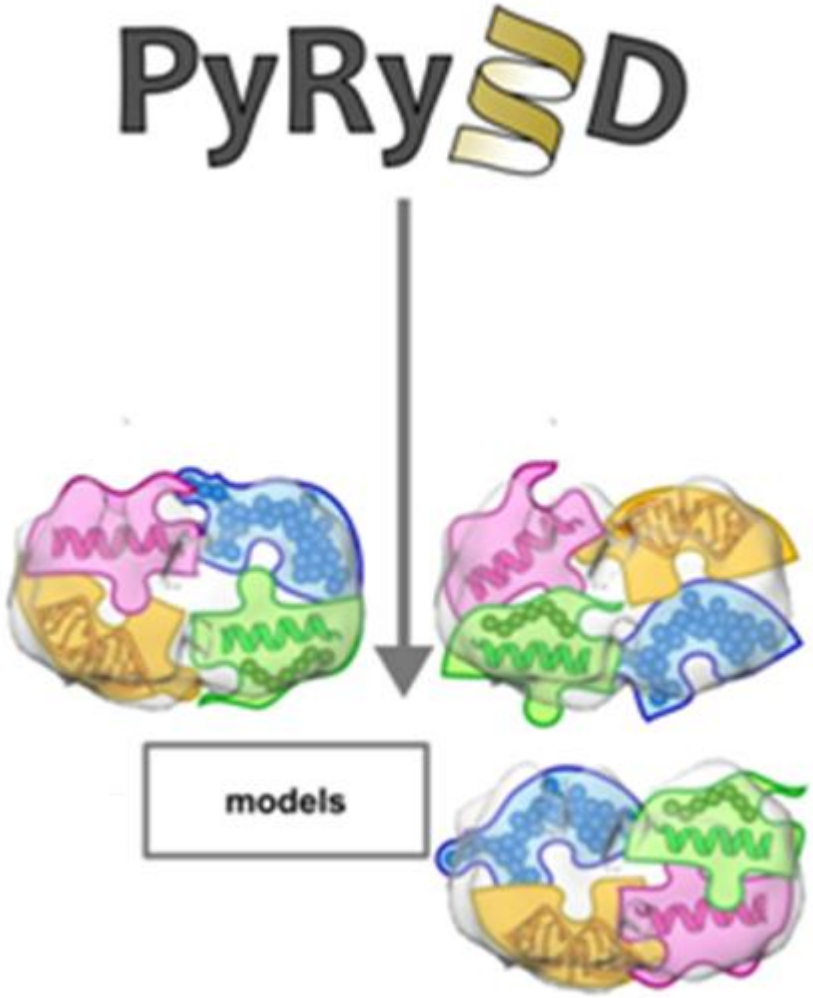
- It is very important where you start
- Simulation should be repeated many times to get statistically significant solutions
- Scoring function should be able to distinguish between similar solutions
- Methods should be equipped with protocols to avoid getting stuck in local energy minima(e.g., Metropolis criteria to accept some worse solutions) or advanced sampling protocols (Replica Exchange, genetic algorithm)

Modeling process

•A number of different configurations may be consistent with the input restraints.

•The aim is to obtain as many structures as possible that satisfy all input restraints.

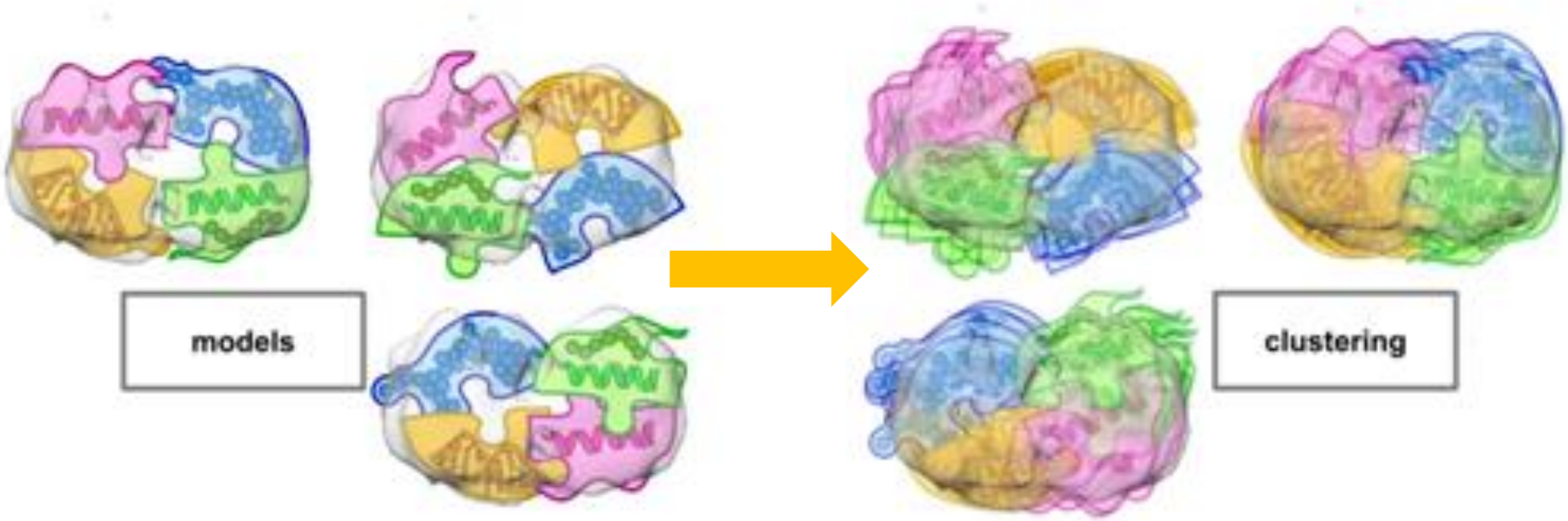
•To comprehensively sample such structural solutions consistent with the data, independent optimizations of randomly generated initial configurations need to be performed until an ensemble of structures satisfying the input restraints is obtained.



Clustering models

To check similarity between obtained models

clustering procedures can be applied



Possible outcomes

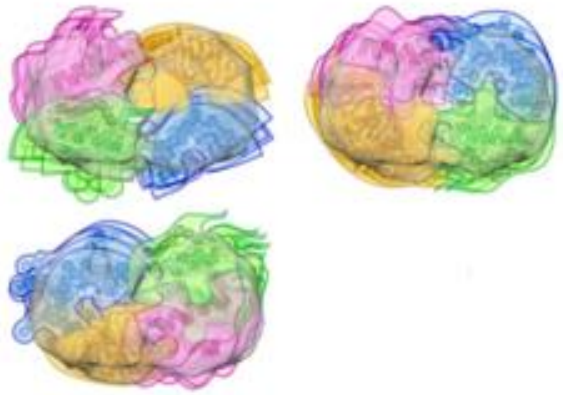
One single model

Sufficient data for prediction of native structure

Alternative models

Insufficient data for prediction of single native structure or there are multiple native structures

Additional experiments to narrow down the possible solutions



NO models

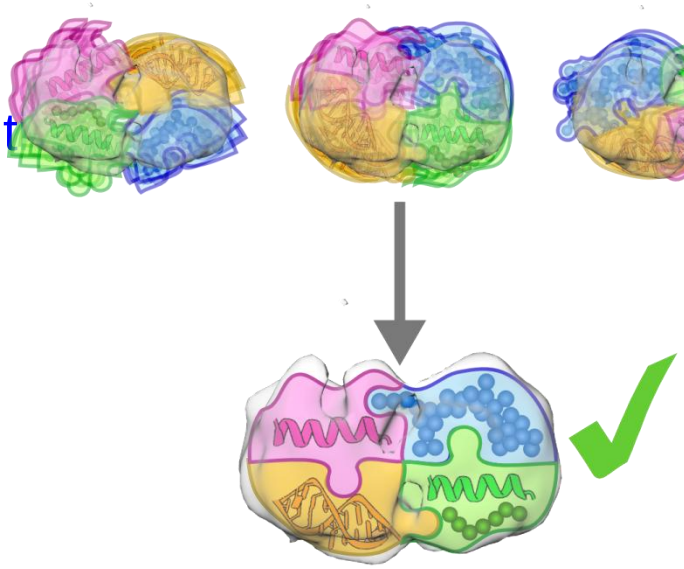
Data or their interpretation in terms of restraints are incorrect

Predicting accuracy of models

1. Self consistency of independent experimental data
2. Structural similarity among all models
3. Agreement with restraints in simulations where a native structure is assumed
4. Confirmatory experimental data that were not used in the model calculation
5. Patterns emerging from a mapping of independent and unused data on the structure that are unlikely to occur by chance

Models refinement

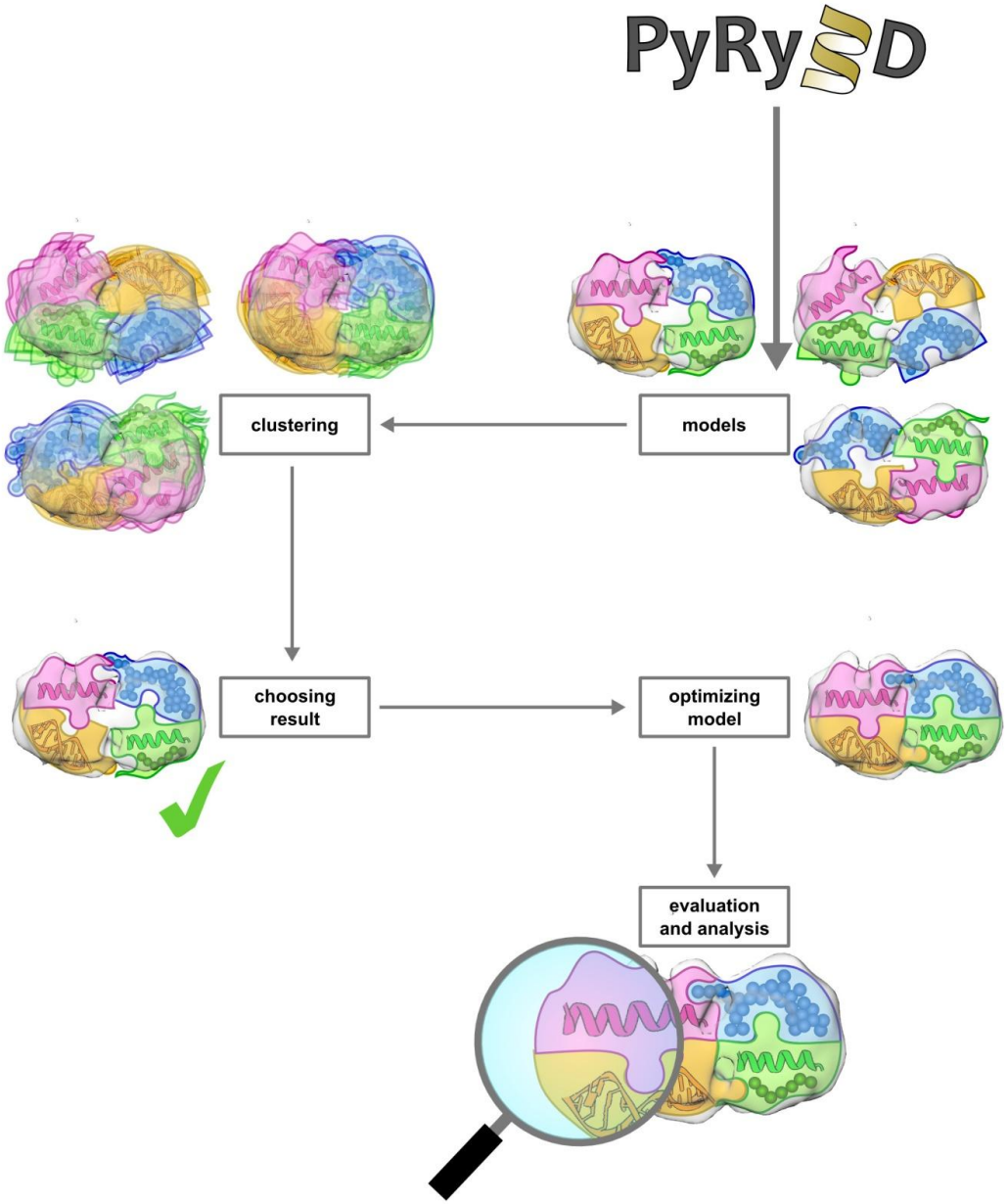
Models generated with PyRy3D are low-resolut



The refinement includes:

- Local fit into density map optimization (colores, DEFINER)
- Removing clashes between side chains (SWISS-MODEL, Modeller, NMD, Zephyr)
- Rebuilding full-atom representation for flexible/disordered refions (Refiner, ROSSETA, Mod-EM, Moulder-EM, qplasty)

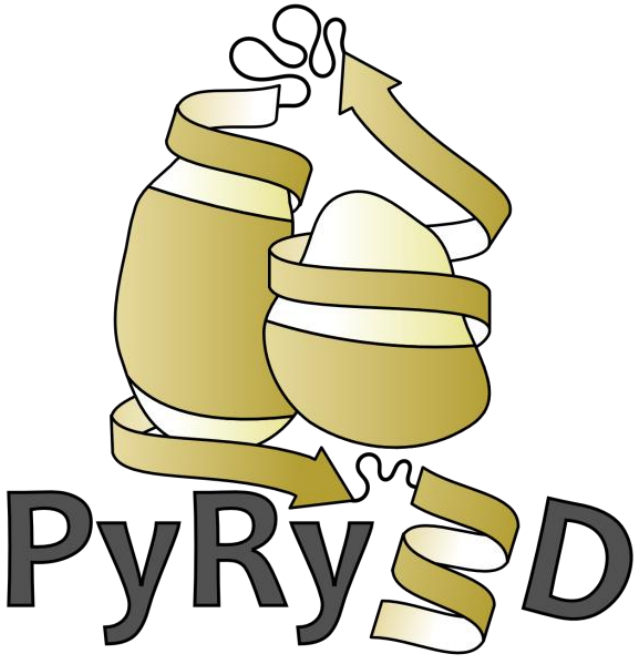
Modeling process



Advantages of integrative approach

1. Allows use of data from various sources
2. Increases accuracy, precision, completeness and efficiency in comparison to what can be achieved by individual methods
3. Achieves a synergy among the available data
4. Can produce all models that satisfy the input data, not only one
5. Minimizes the drawback of incomplete, inaccurate and/or imprecise datasets

PyRy3D in practice



How to prepare input files?

DATA TYPE	FORMAT	COMMENTS
STRUCTURES	PDB	• EACH COMPONENT AS ONE CHAIN; • IF DISORDERED REGIONS OCCUR – NUMBERING OF RESIDUES MATTERS
SEQUENCES	MULTI FASTA	• REQUIRED ONLY WHEN DISORDERED REGIONS OCCUR
DENSITY MAP AB INITIO MODEL SAXS CURVE	MRC PDB – DAMMIN/IF .DAT	
RESTRAINTS	FILTREST3D	• POSSIBLE TO ASSIGN WEIGHTS TO RESTRAINTS AND TO COMBINE THEM BY BOOLEAN OPERATORS
CONFIGURATION FILE	TEXT FILE	• ALL PARAMETERS CAN BE MODIFIED HERE

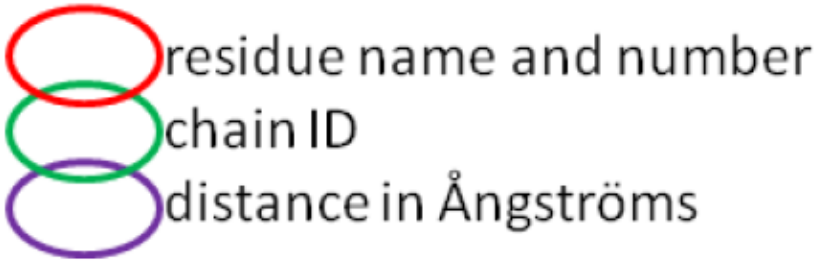
* required

Configuration file

```
#   SIMULATION PARAMETERS FILE
## ----   General simulation parameters   ----------
SIMMETHOD SimulatedAnnealing #"Genetic" or "SimulatedAnnealing" for simulated annealing (default) or "ReplicaExchange" for replica exchange
STEPS   50                   #default 100; how many simulation steps to perform?
ANNTEMP 10                    #from range X to Y
REHEAT false 1
MAXROT 5                    #default is 5
MAXTRANS    5 5 5            #default is [5, 5, 5]

## ---    Scoring function parameters      ----------
OUTBOX 1 1                   #default 1 1; can be in range from 0 to 10 (float numbers)
MAP_FREESPACE  1 1           #default 1 1; can be in range from 0 to 10 (float numbers); also used for SAXS shapes
CLASHES   0 0                #default 1 1; can be in range from 0 to 10 (float numbers)
RESTRAINTS  1 1              #default 1 1; can be in range from 0 to 10 (float numbers)
DENSITY     0 0              #default 1 1; can be in range from 0 to 10 (float numbers)

### ----  Mutation frequencies -----------
ROTATION_FREQ 0.4            #from range 0 to 1, default is 0.25, sum of FREQ params must be equal to 1
ROTATION_COV_FREQ 0.0        #from range 0 to 1, default is 0.25, sum of FREQ params must be equal to 1
TRANSLATION_FREQ 0.3         #from range 0 to 1, default is 0.25, sum of FREQ params must be equal to 1
EXCHANGE_FREQ 0.3            #from range 0 to 1, default is 0.25, sum of FREQ params must be equal to 1
SIMUL_DD_FREQ 0.0            #from range 0 to 1, default is 0.25, sum of FREQ params must be equal to 1
TRANSLATION_ALL_FREQ 0.0     #from range 0 to 1, default is 0.25, sum of FREQ params must be equal to 1; mutation were all components are translated together
ROTATION_ALL_FREQ 0.0        #from range 0 to 1, default is 0.25, sum of FREQ params must be equal to 1; mutation were all components are translated together

## -----  Input data descriptors   --------
THRESHOLD 1.6                #float value existing in map file, default is 0; alternative parameter: KVOL
SIMBOX 1.4                   #default simulation box diameter
GRIDRADIUS  2.0              #default is 1.0
GRIDTYPE    cubic               #cubic or diamond
COMPONENT_REPRESENTATION ca  #CA - only calfas/c4' (default); cacb - coarse grain, 3p - 3points, fa - full atom

### ----- Simulation process control -------
SCALEPARAMS on               #on/off
PARAMSCALINGRANGES 0 25 50   #default 0 25 50; at what point of simulation should parameter scaling ranges kick in; % zakresu symulacji (postep)
PARAMSCALINGR1 50 100        #default 50 100 in %
PARAMSCALINGR2 25 50         #default 25 50 in %
PARAMSCALINGR3 0 25          #default 0 25 in %

## ------ Output parameters ---------
WRITE_N_ITER 1               #default 0, minimum 1 max=STRUCT_NR
```

Restraints file – extented Filtrest3D format

```
dist (
    (G8) "C" - (U9) "D" (<=1.50)
    (U11) "D" - (C12) "A" (<=1.50)
    (G59) "A" - (A60) "D" (<=1.50)
    (C62) "D" - (C63) "B" (<=1.50)
    (G83) "B" - (U84) "D" (<=1.50)
    (G88) "D" - (C89) "C" (<=1.50)
    )
```



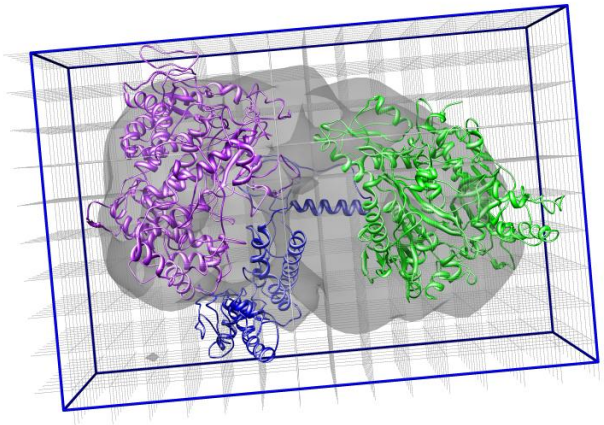
- Boolean operators: AND, OR
- Other types of restraints: PointDistance, SurfaceAccess, Symmetry, Relation

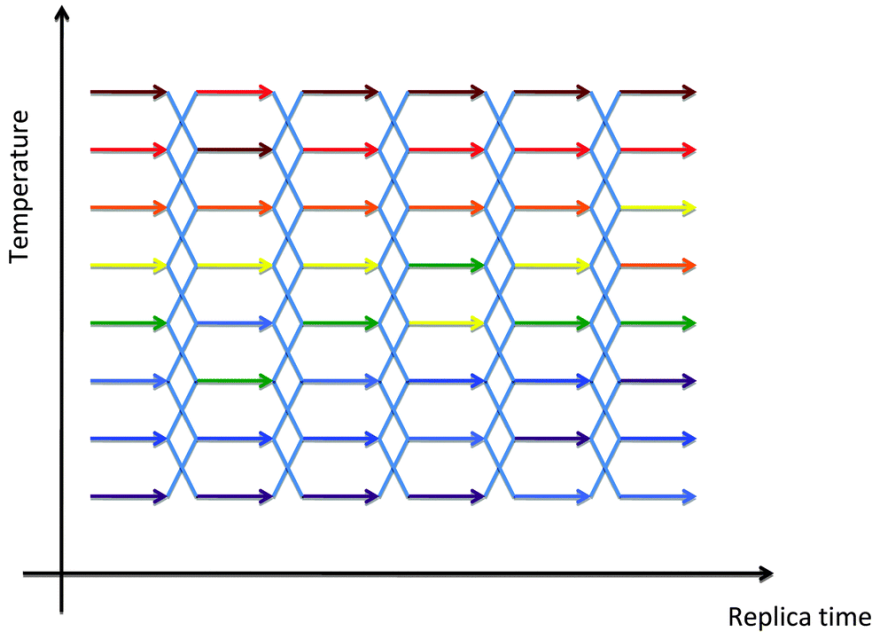
How to set the first modeling?

SIMMETHOD SimulatedAnnealing
ANNTEMP 10

STEPS 100
WRITE_N_ITER 10

SIMBOX 1.2
GRIDRADIUS 1.0





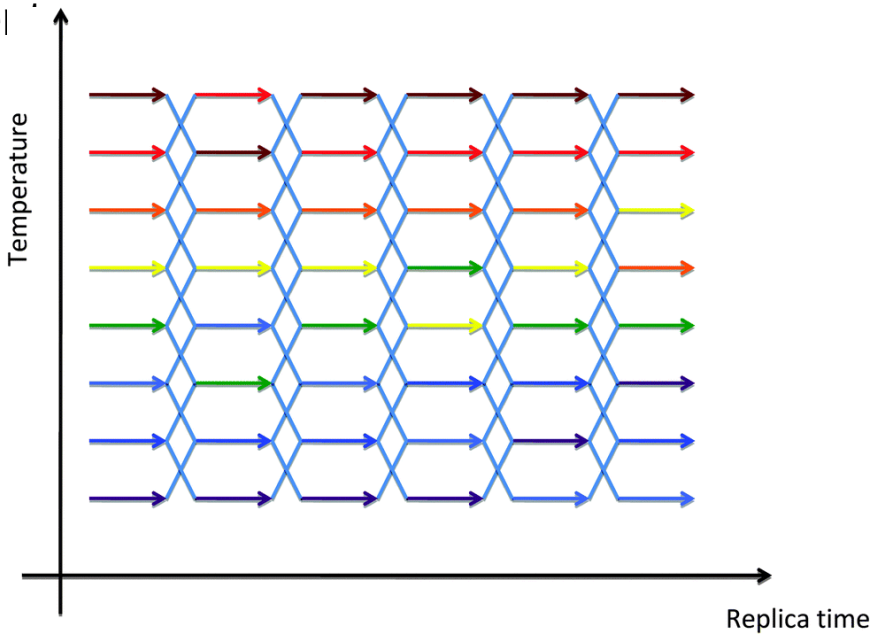
My simulation stuck in local minima..

Play with sampling protocol:
SIMMETHOD ReplicaExchange
REPLICAEXCHANGE_FREQ 2 #after each 2 steps replicas will be exchanged;
REPLICATEMPERATURES 200 100 50 0 #list of temperatures for all replicas

SIMMETHOD Genetic
MAXPOOLSIZE 4
REDUCTMETHOD roulette #cutoff, tournamer

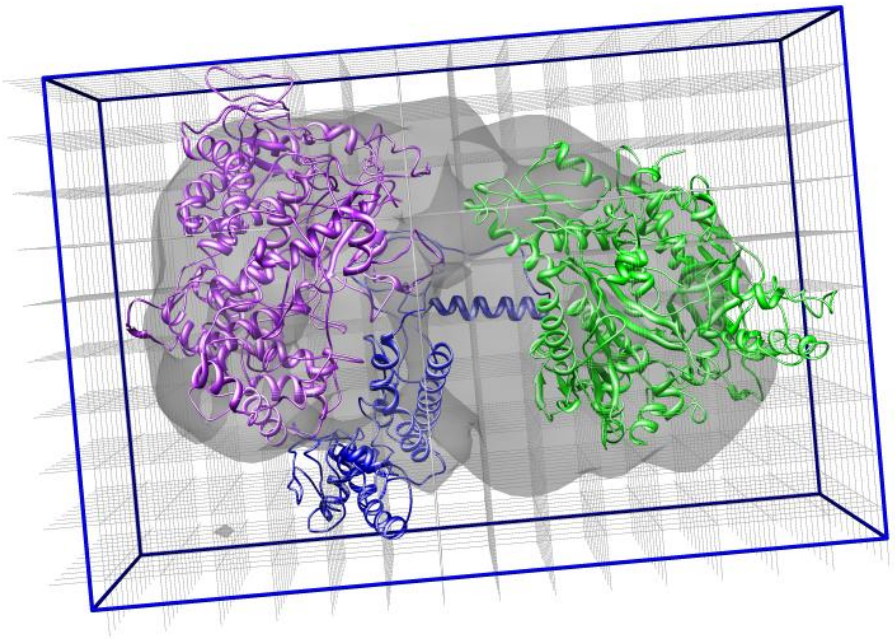
Play with annealing temperature:
REHEAT True 1000

ANNTEMP 1000



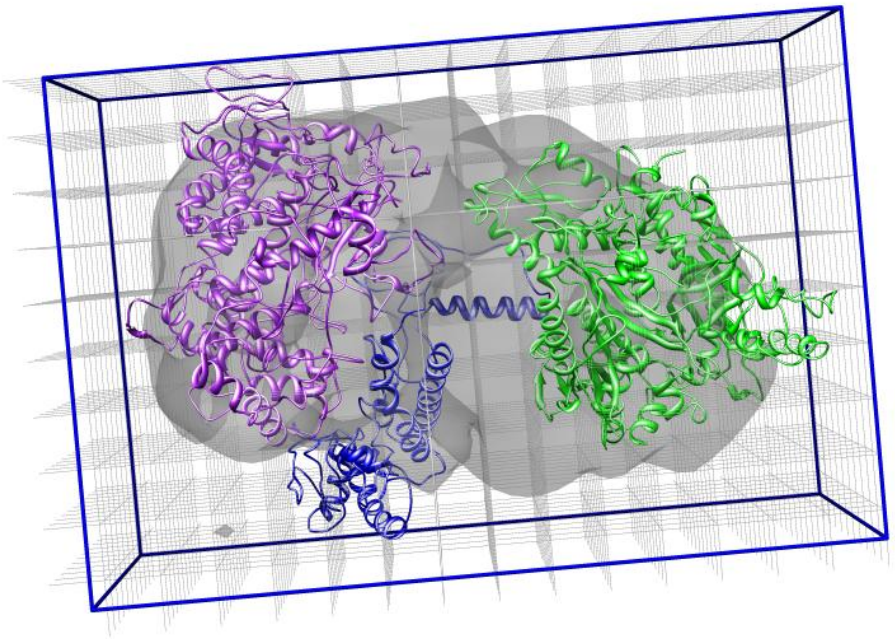
How to speed up the analysis?

- Increase grid size (**GRIDRADIUS**)
- Use reduced models (**COMPONENT_REPRESENTATION**)
- Decrease the simulation area (**SIMBOX**)
- Change the density map (**THRESHOLD / KVOL**)



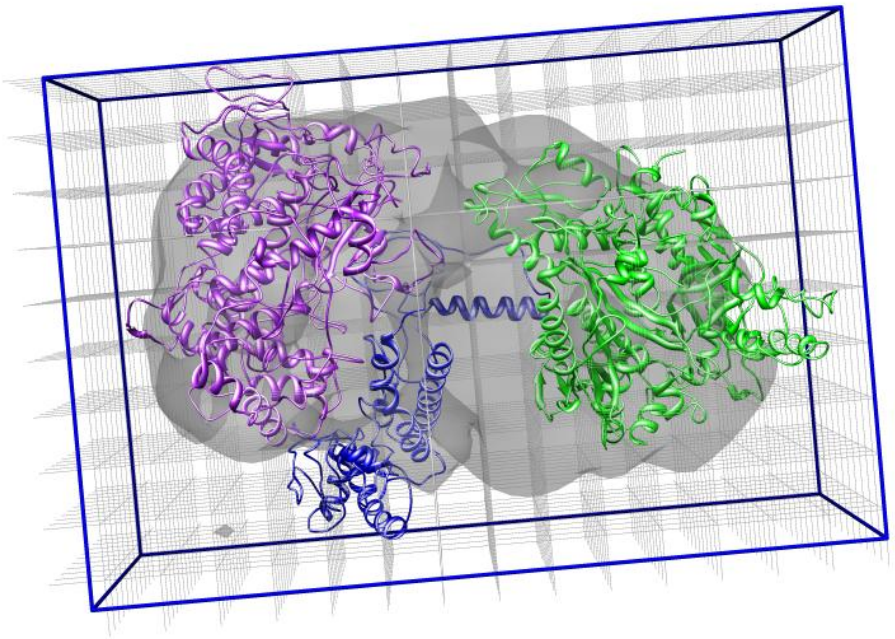
How to reconstruct full-atom model?

- Use command-line version
- Run PyRy3D with BOTH options:
  - -f → to build fullatom model
  - -v → to save history of simulation movements



How to define complex shape?

- Download a map in MRC format and check the details provided by its authors such as: contour level, resolution, volume/mass
- Define grid size (**GRIDRADIUS**)
- Limit the simulation area (**SIMBOX**)
- Change the density map size (**THRESHOLD / KVOL**)



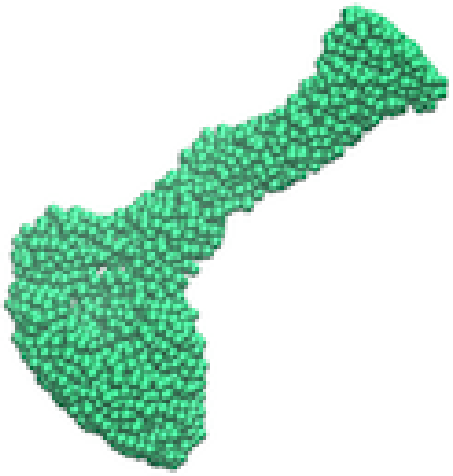
How to work with SAXS data?

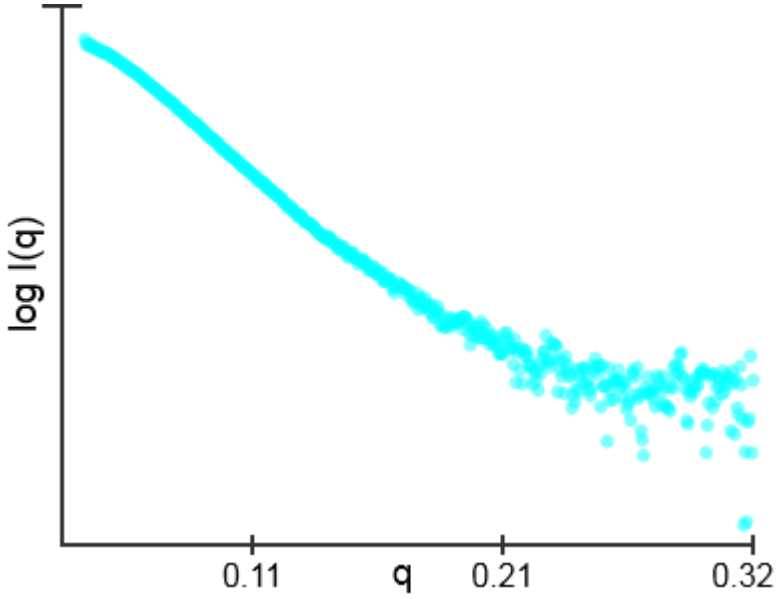
There are two ways to use the results from SAXS experiment:

- *ab initio* models from Dammin/dammif programs (option -x)
- raw data from SAXS – experimental curve (option -y)

In the configuration file, remember to define:

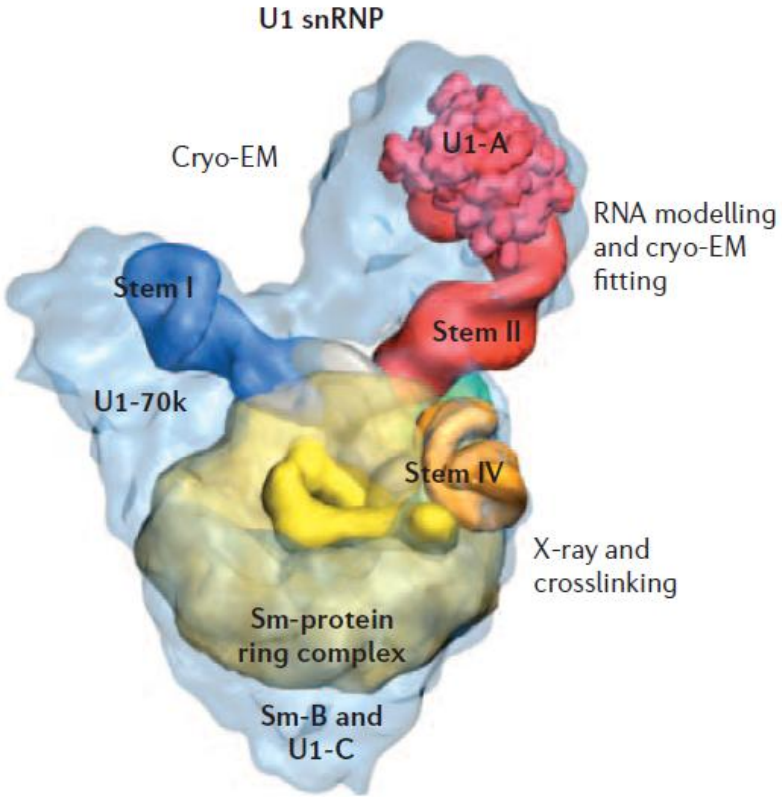
- CHI2, RG_VA
- SAXSRADIUS
- CRYSOL_PA





How to optimize model with PyRy3D?

- Start with predefined orientation (**START_ORIENTATION**)
- Limit component movements (**MOVE_STATE**)
- Use small moves (like **MAXTRANS** and **MAXROT** 1)
- Decrease temperature parameter value (e.g. **ANNTEMP** 1)



How to limit movements of component?

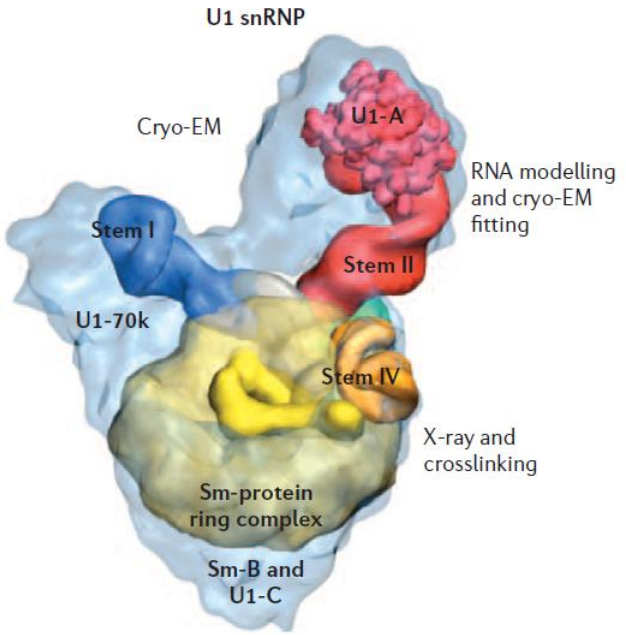
- **MOVE_STATE** A fixed
- **MOVE_STATE** A 5 5 5 10 10 10 60 60 60 50 50 50 5 60

5 5 5 – maximal rotation around X, Y, Z axes respectively in a SINGLE move

60 60 60 - maximal rotation around X, Y, Z axes respectively during entire simulation

10 10 10 maximal translation along X, Y, Z axes respectively in a SINGLE move

50 50 50 maximal translation along X, Y, Z axes respectively during the entire simulatio…



5 60 refers to rotation around a covalent bond only; first refers to the limitation in a single move, second refers to the entire simulation

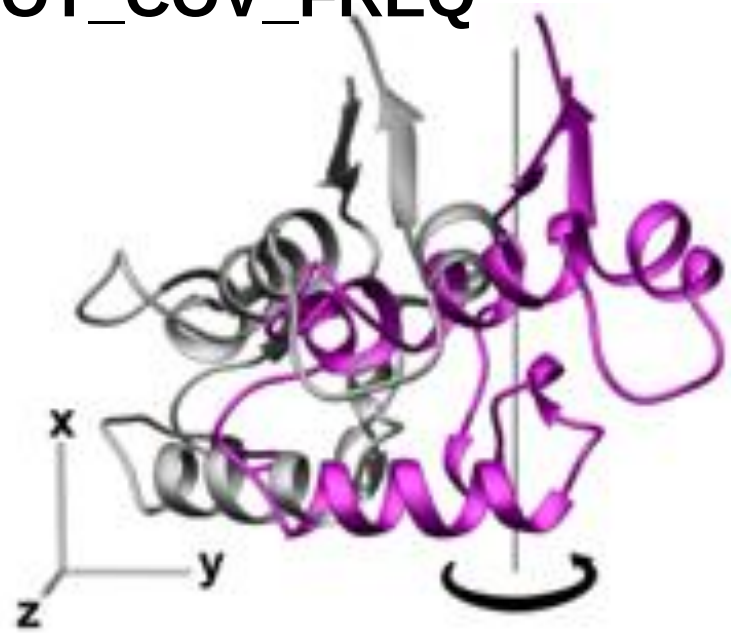
What if some components should be moved together?

- put them into single PDB file and treat as a single component

OR:

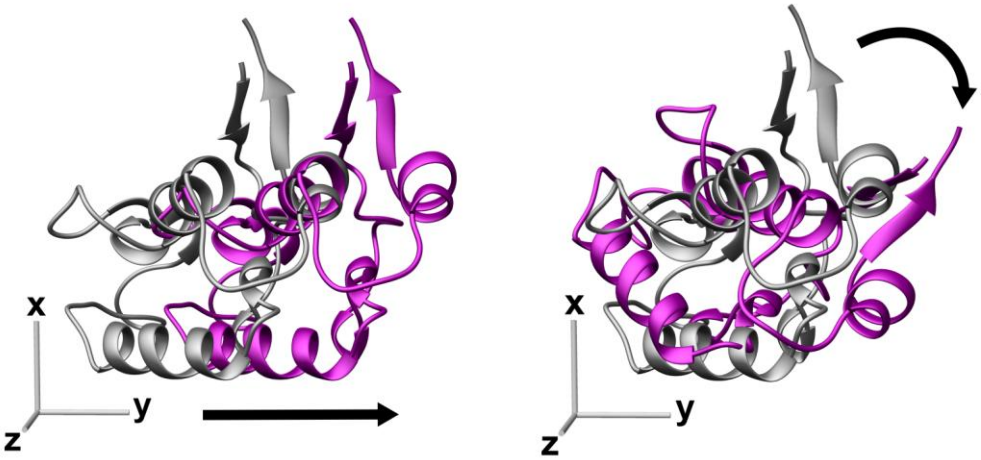
- define **COVALENT_BOND A [„B”,”C”] [10,11]**

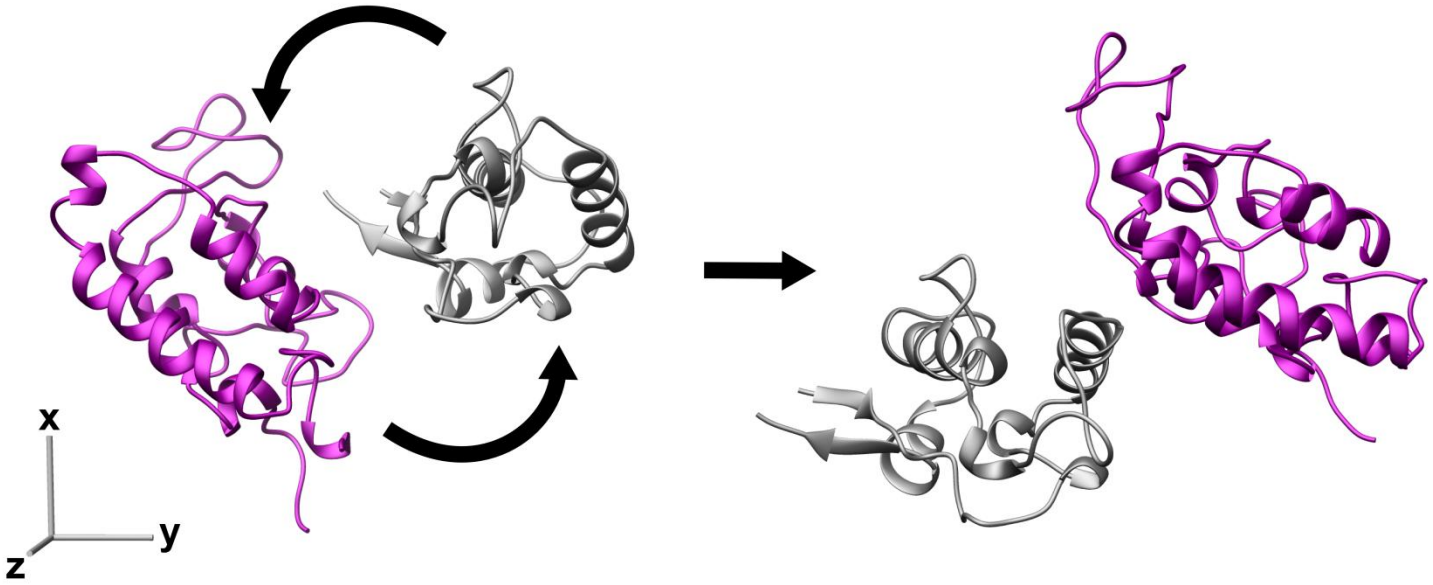
- remember you can apply rotation around the covalent bond to such components: **ROT_COV_FREQ**



How to choose allowed moves

ROTATION_FREQ	0.3
ROTATION_COV_FREQ	0.0
TRANSLATION_FREQ	0.3
EXCHANGE_FREQ	0.4
EXCHANGESAMPLE_FREQ	0.0
SIMUL_DD_FREQ	0.0
TRANSLATION_ALL_FREQ	0.0
ROTATION_WHOLE_FREQ	0.0

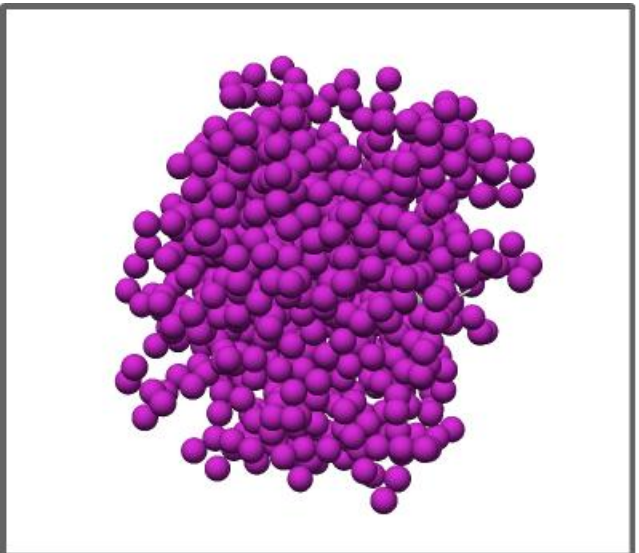
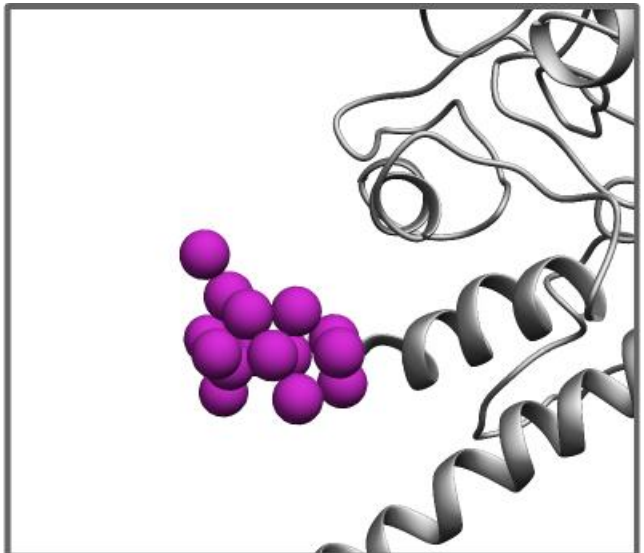
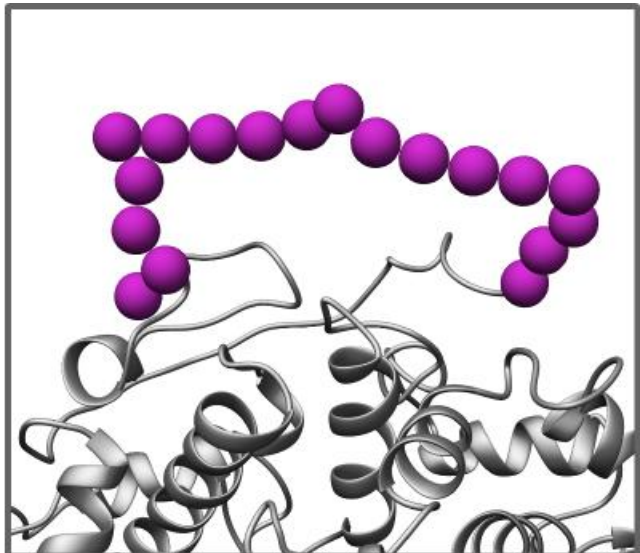




How to set components flexibility?

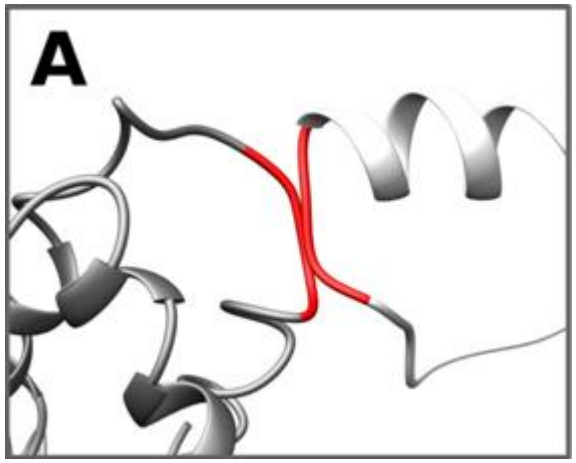
SIMUL_DD_FREQ 0.5
IDENTIFY_DISORDERS True

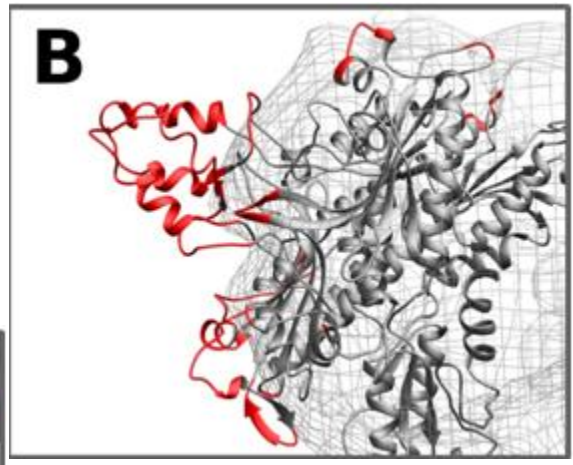
FASTA file
>X_protein
PVTLAGMLEMGVSYLPVNQNLPRSALPRAVIRHPDYDEEGLYGAILPQVVTAGTITRR
AVEPTWLTASNARPDRVGSELKAMVQAPPGYTLVGADVDSQELWIAAV
>Z
FAYKLGLNDLPQSVAFFSAVDIDRCLRKEVTMDCKTPSNPTGMERRYGIPQGEALDIY
QIIELTKGSLEKRSQPGPhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

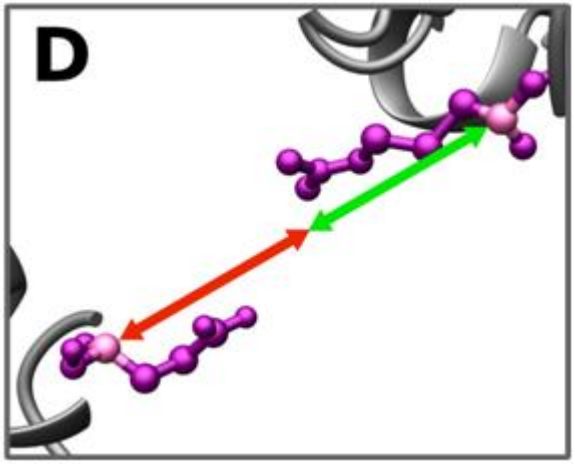
How to set scoring function weights

CLASHES 10 10
CLASHES_ALLATOMS 1 1





OUTBOX 10 10
MAP_FREESPACE 1 5
DENSITY 0 0



RESTRAINTS 5 1
SYMMETRY 0 0

CHI2 1 1
RG 1 1

SCALE_PARAMS ON

Happy modeling!!