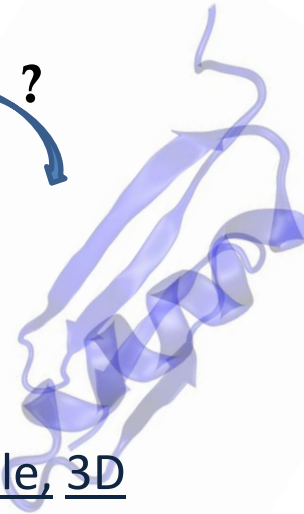


Designing highly specific probes with tunable affinity

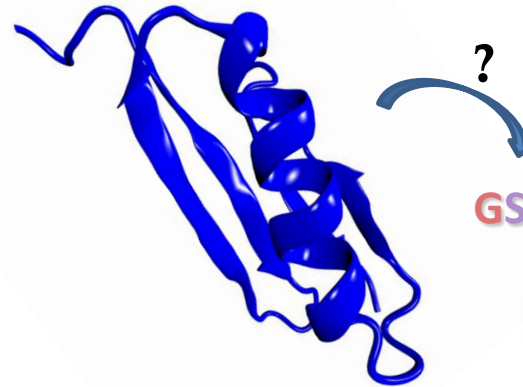
GSHT...TCPKC



FOLDING



Find a specific, stable, 3D structure given a fixed amino acid sequence



GSHT...TCPKC

DESIGN

Find the sequence of amino acids which will fold in a given target structure



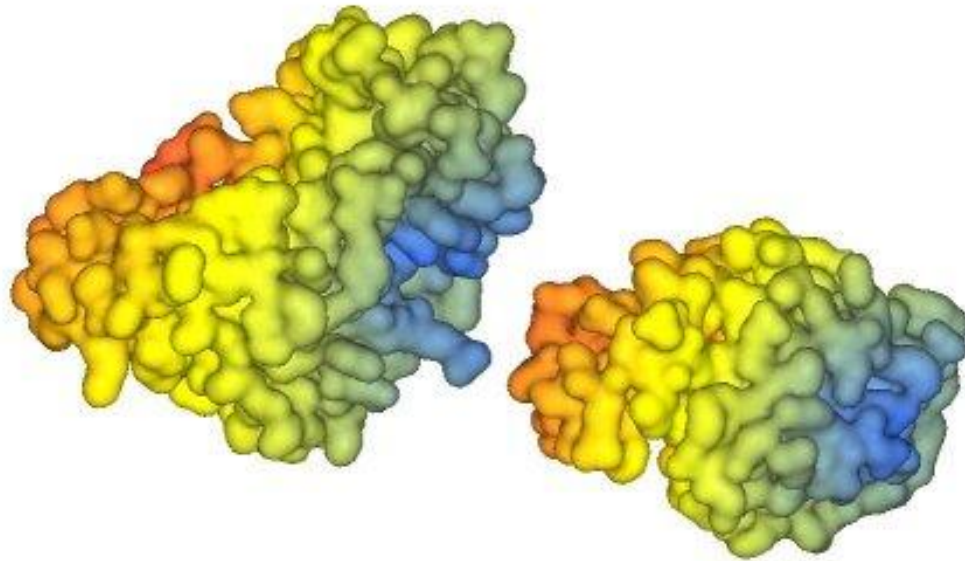
universität
wien

Francesca Nerattini



Computational Physics Group, University of Vienna

Protein-ligand binding



Goal:

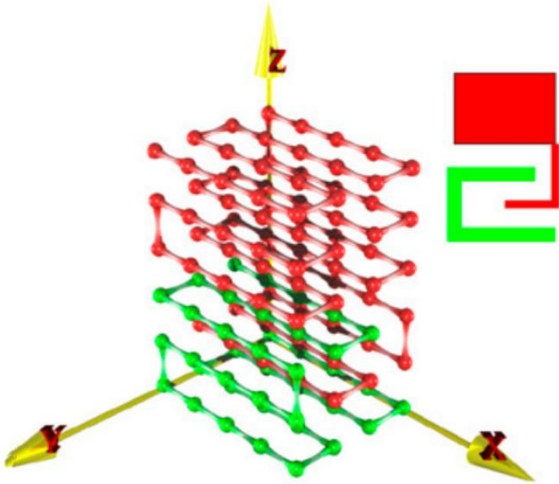
Design protein-ligand systems characterized by high specificity and tunable affinity

Protein-ligand binding ON LATTICE

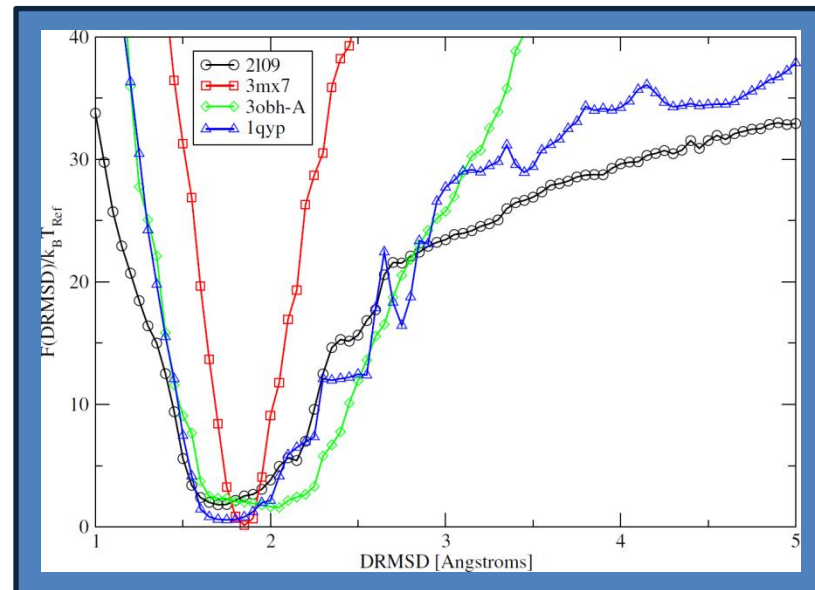
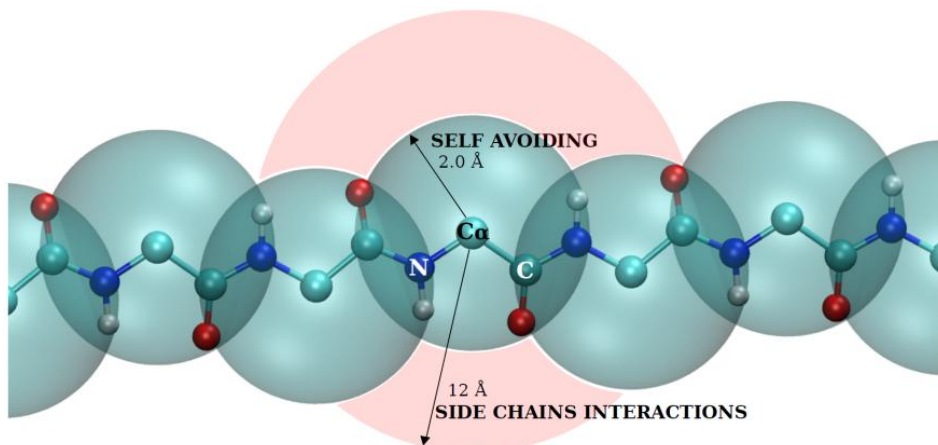
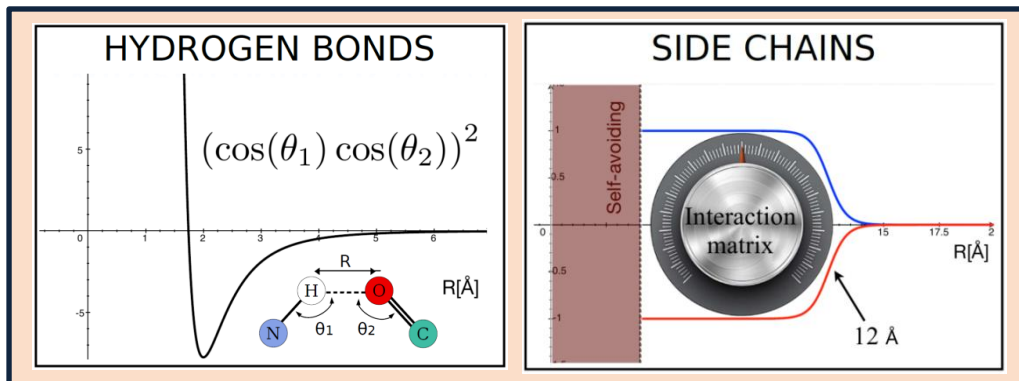
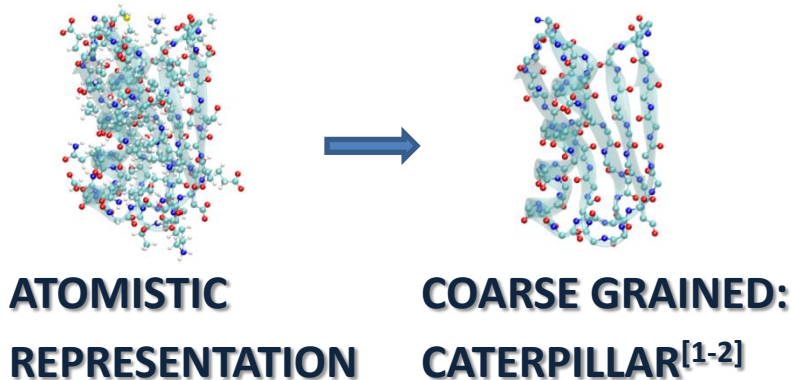
- Design (MC,CATERPILLAR) sequences of both protein and substrate

- Introduce different percentages of “randomness”

- Evaluate the Binding Affinity (folding,MC,CATERPILLAR)
$$\Delta f \equiv -k_B T \ln(Q_b/Q_f)$$



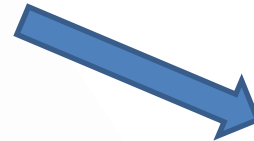
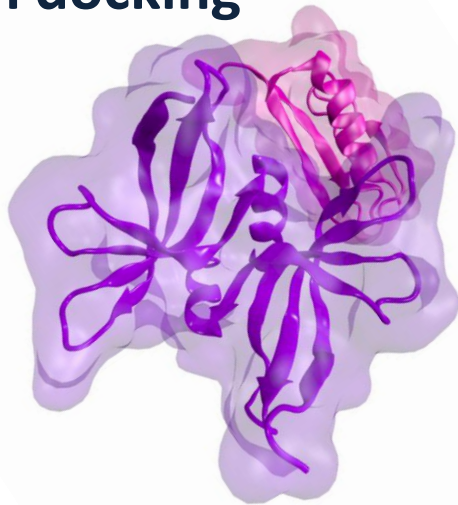
The CATERPILLAR model



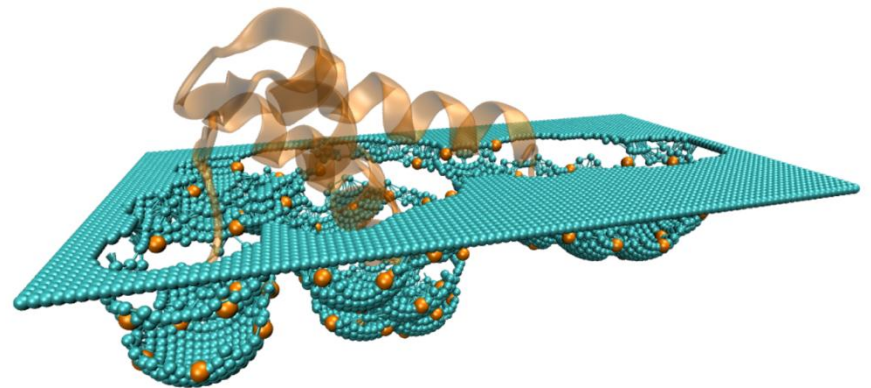
- [1] Coluzza, Ivan. "A coarse-grained approach to protein design: learning from design to understand folding." PloS one, 6, e20853 (2011)
- [2] Coluzza, Ivan. "Transferable coarse-grained potential for de novo protein folding and design." PloS one, 9, e112852 (2014)

Protein-ligand binding OFF LATTICE

**REAL SYSTEM:
protein docking**

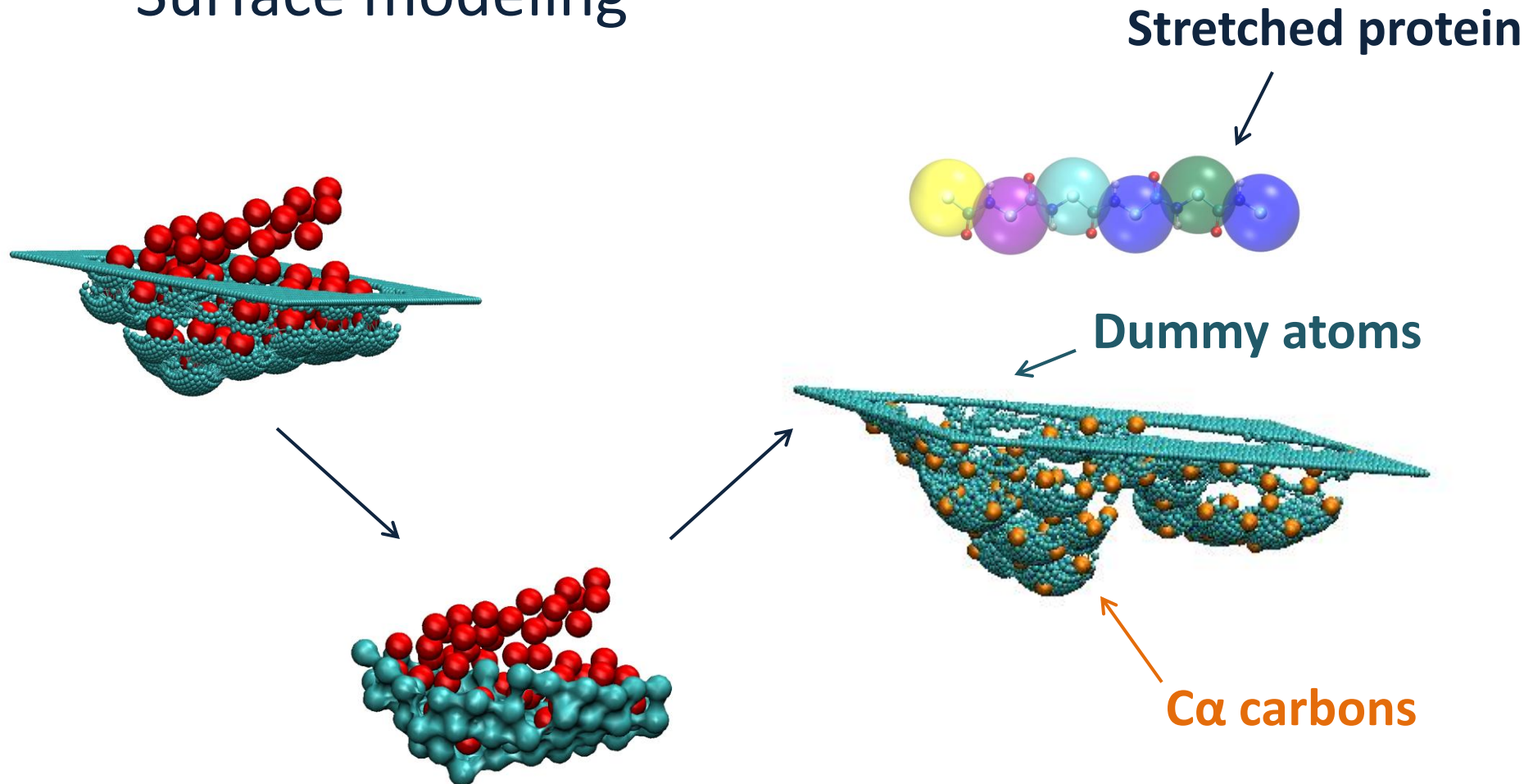


**TEST SYSTEM:
protein-like surface**



Protein-ligand binding OFF LATTICE

- Surface modeling



Protein-ligand binding OFF LATTICE

WORK IN PROGRESS



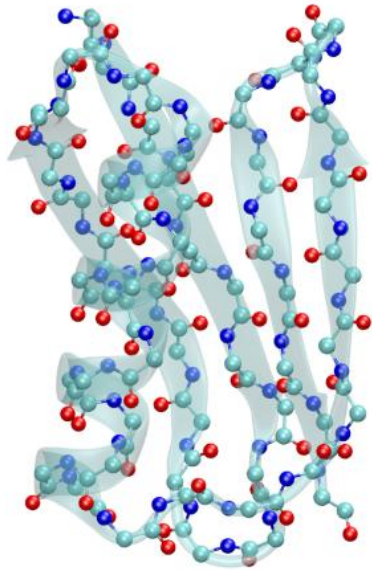
- Design of protein-ligand sequences
- Evaluation of the binding affinity

FUTURE DIRECTION

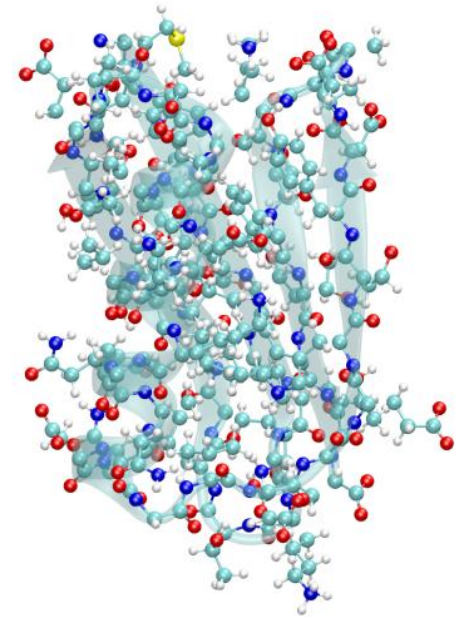


- Protein-protein docking

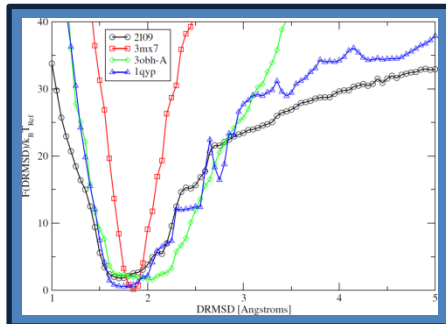
Test designed sequences stability



- Replace the native sequence with the designed one
- Side chain rotamer analysis (**SCWRL4**)
- Set up simulation box
- System equilibration



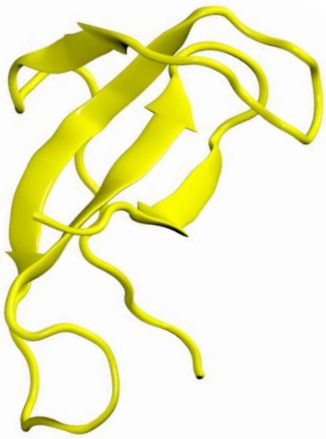
**COARSE GRAINED:
CATERPILLAR**



ATOMISTIC REPRESENTATION

GROMACS

- amber99sb
- TIP3P water
- T = 298 K ; P = 1 atm



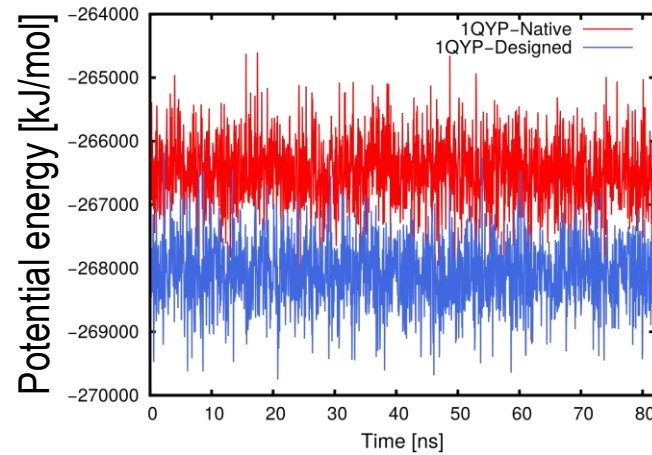
Full atomistic 1QYP

GROMACS

- amber99sb
- TIP3P water
- T = 298 K ; P = 1 atm

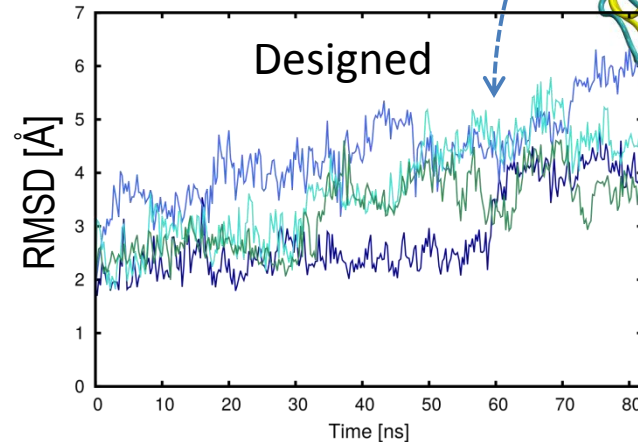
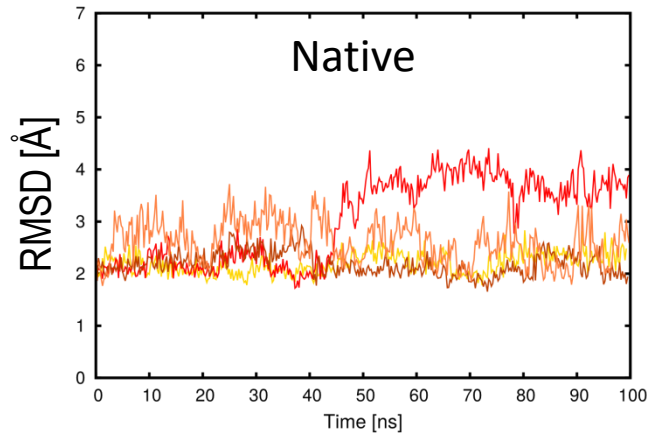
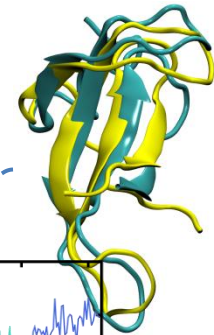


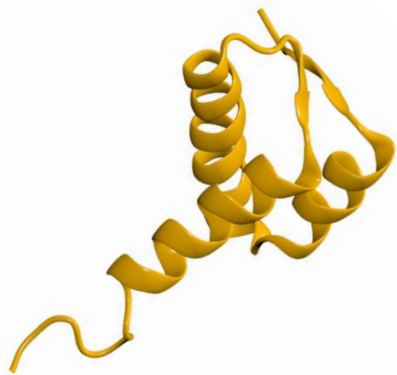
Reference Structure
for RMSD from RCSB



Native seq.

Designed seq.





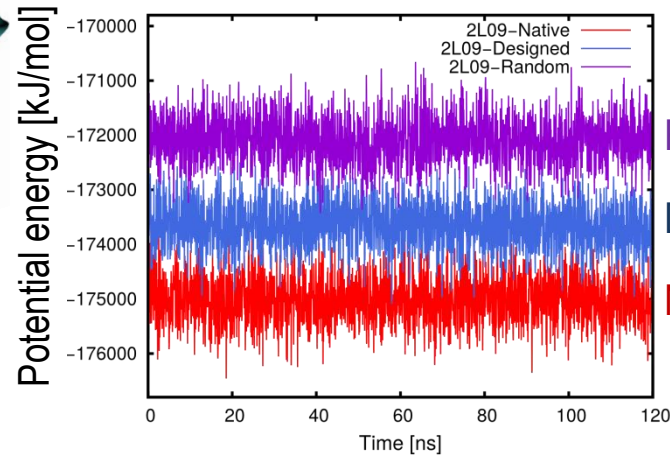
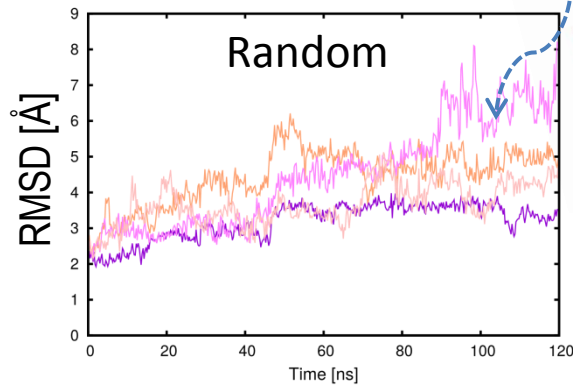
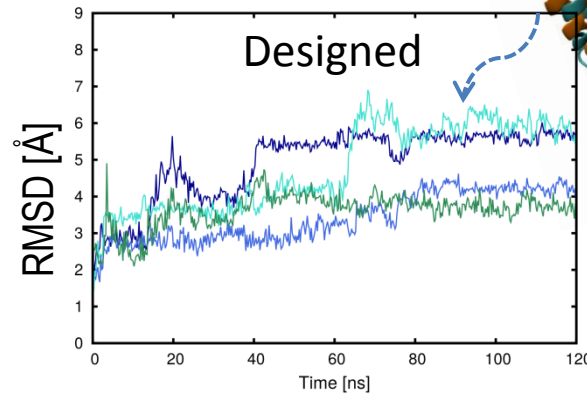
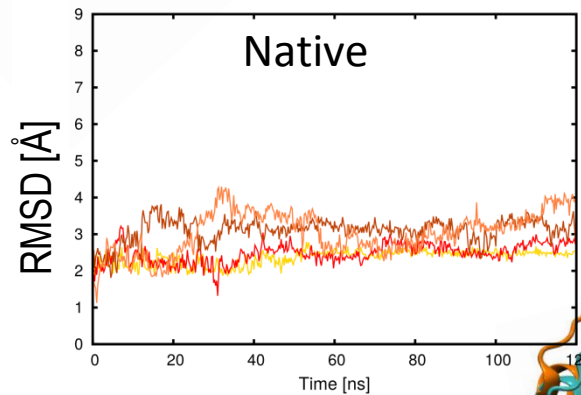
Full atomistic 2L09

GROMACS

- amber99sb
- TIP3P water
- T = 298 K ; P = 1 atm



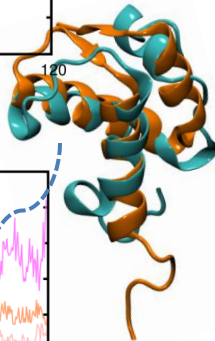
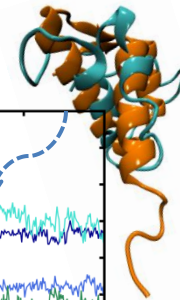
Reference Structure
for RMSD from RCSB



Random seq.

Designed seq.

Native seq.

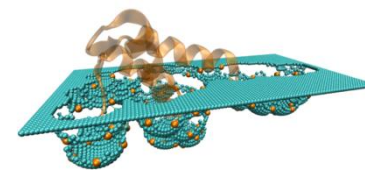




From here...

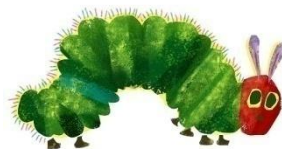
➤ Protein-ligand binding:

- Evaluate binding free energy and affinity of the test system.
- Protein-protein docking.



➤ Test of designed sequences:

- Take advantage of MD feedback in order to improve the CATERPILLAR MODEL.
- Set up BIAS-EXCHANGE METADYNAMICS with PLUMED.



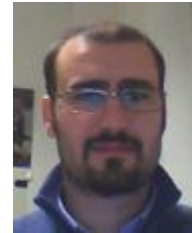
Acknowledgements

- Prof. Christoph Dellago and Dr. Ivan Coluzza



- Automated Bio Marker (ABM) - In silico automated tumor targeting

- Dr. Luca Tubiana



- Theory and simulations of designable modular bionic proteins

- Dr. Valentino Bianco



- M.Sc. Chiara Cardelli

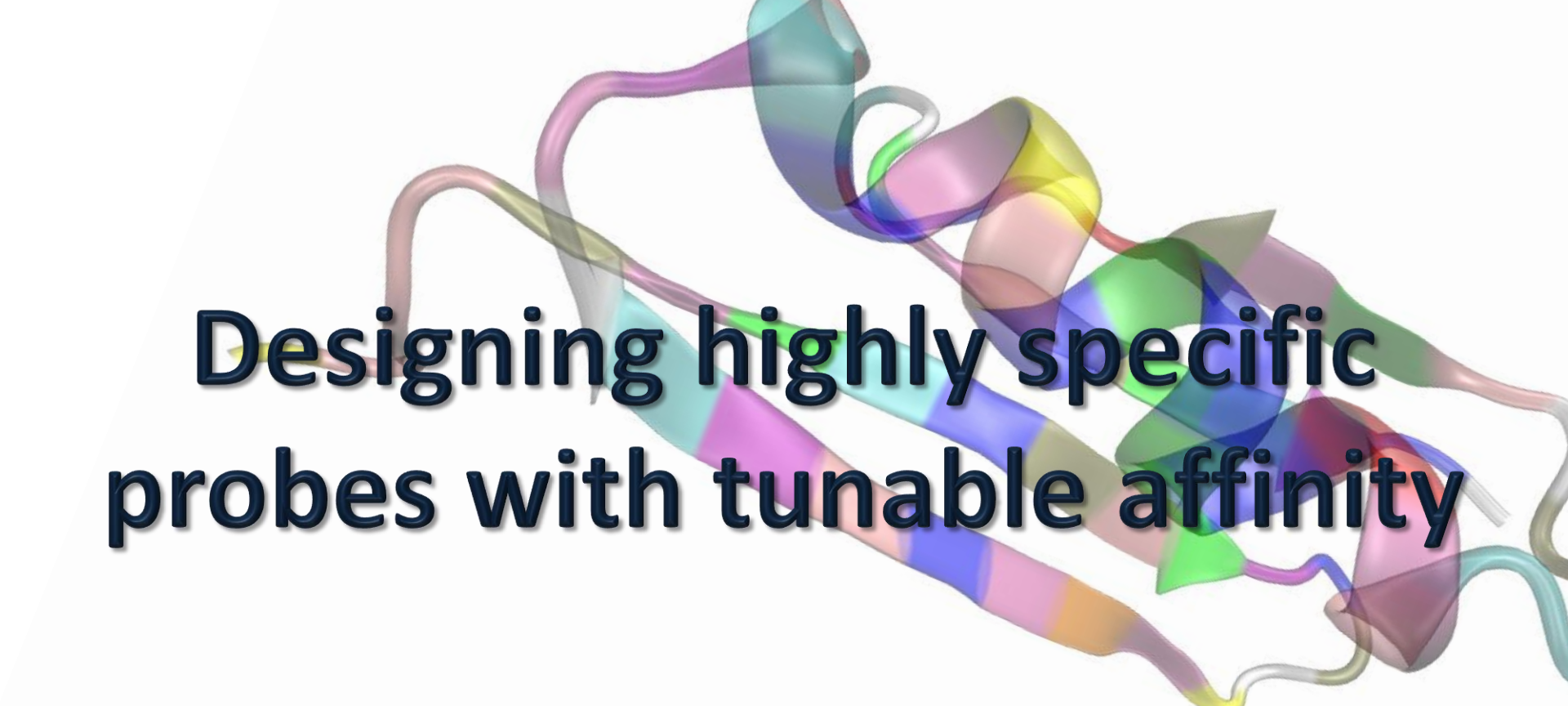


Thank you for your attention



universität
wien





Designing highly specific probes with tunable affinity

Francesca Nerattini

Computational Physics Group, University of Vienna



universität
wien

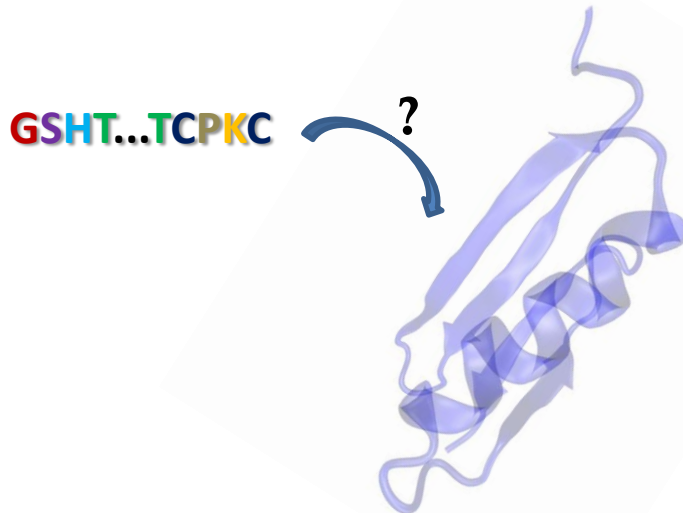


Working with proteins..

- Protein folding and design are major biophysical problems, the solution of which would lead to important applications

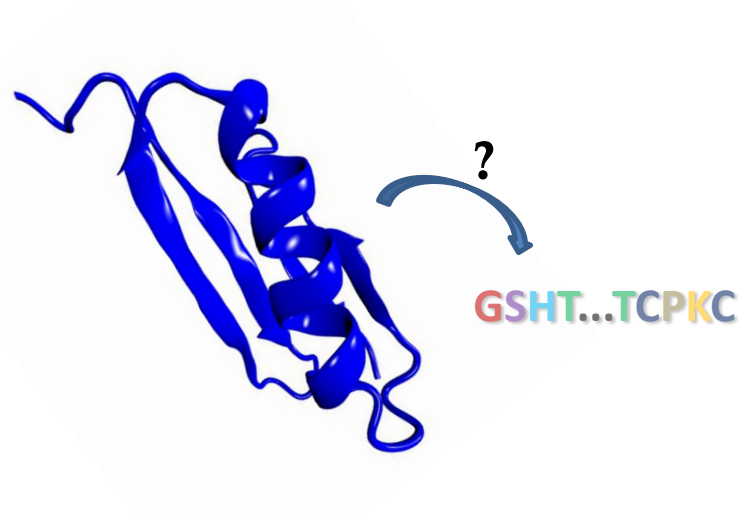
FOLDING

Find a specific, stable,
3D structure given a
fixed amino acid
sequence



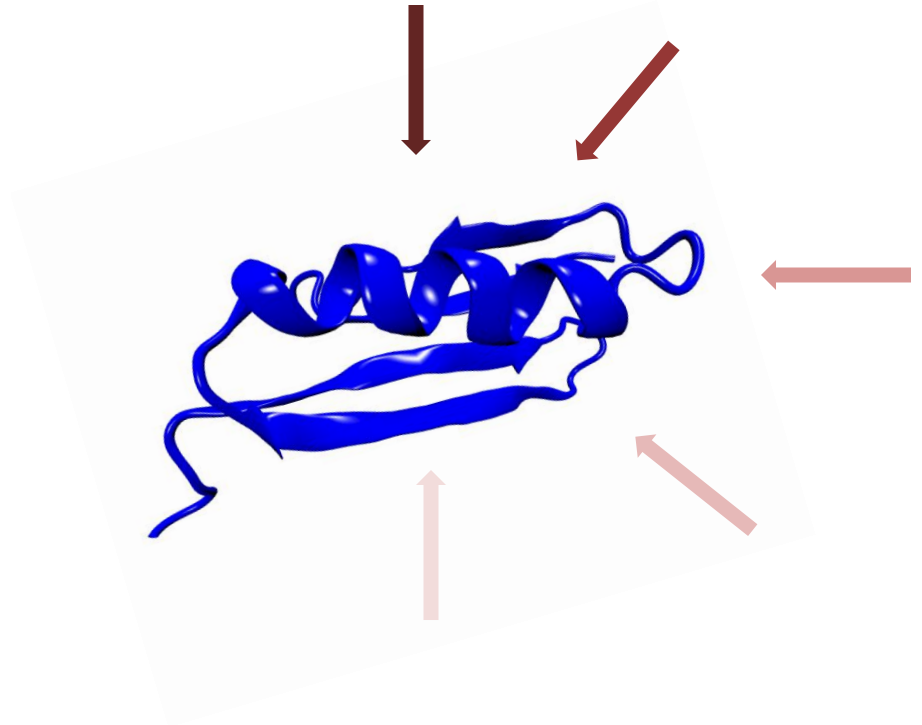
DESIGN

Find the sequence of
amino acids which will
fold in a given target
structure



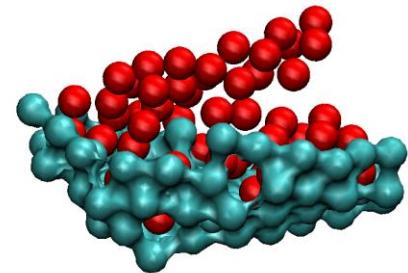
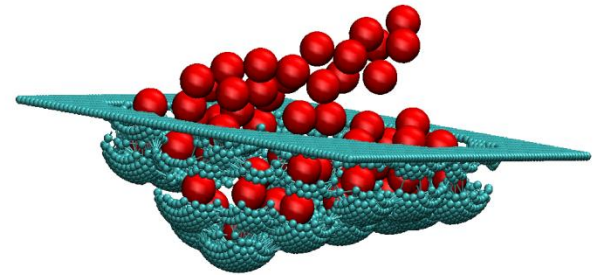
Protein-ligand binding OFF LATTICE

- Identification of the pocket
 - Active site
 - Rotational analysis



Protein-ligand binding OFF LATTICE

- Identification of the pocket
- Surface modeling
 - define parameters
 - construct a 2D grid
 - move grid point until the minimum distance from protein
 - smooth the surface
 - reduce the number of points

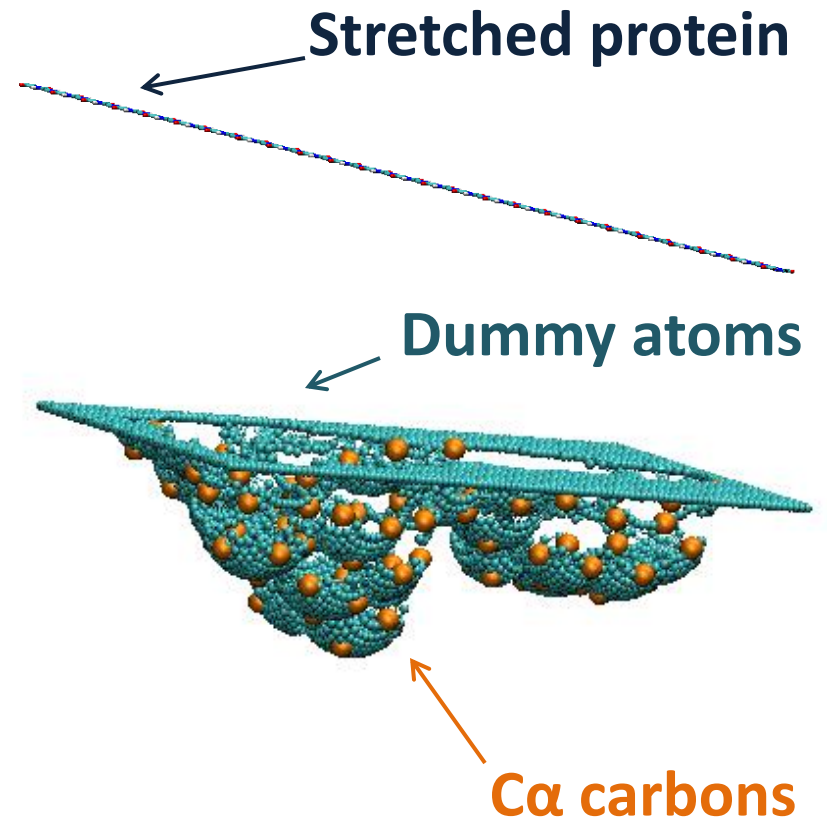
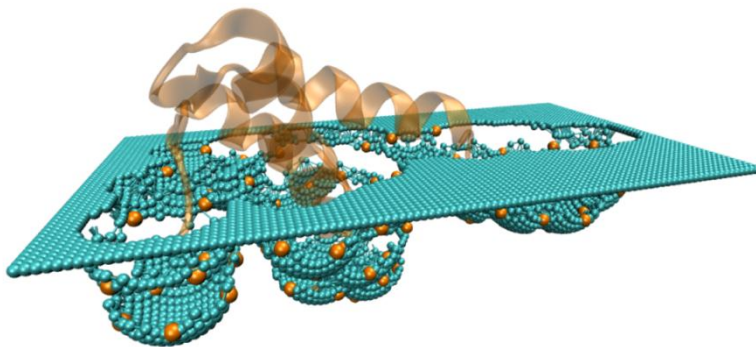


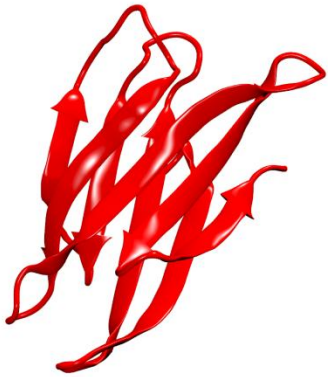
Protein-ligand binding OFF LATTICE

DESIGN



FOLDING





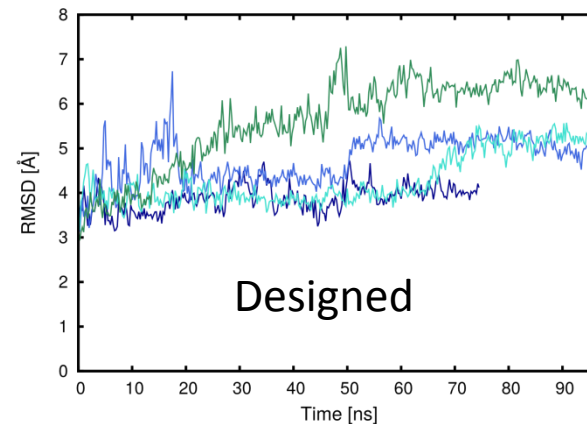
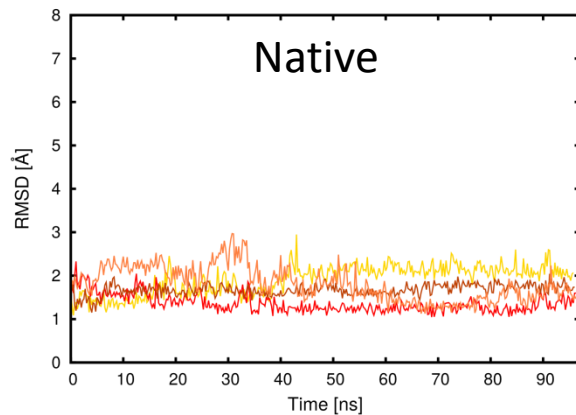
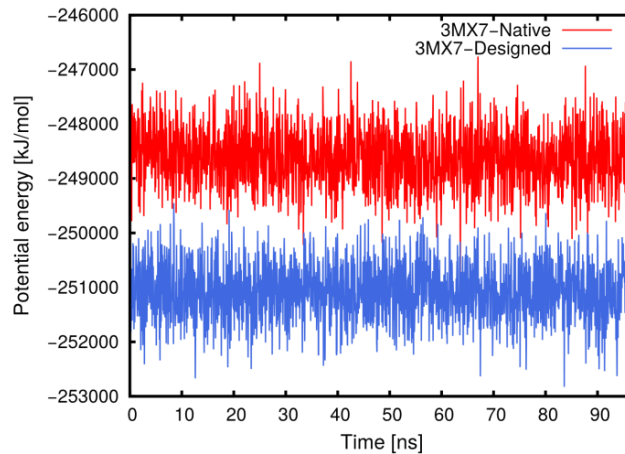
Full atomistic 3MX7

GROMACS

- amber99sb
- TIP3P water
- T = 298 K ; P = 1 atm

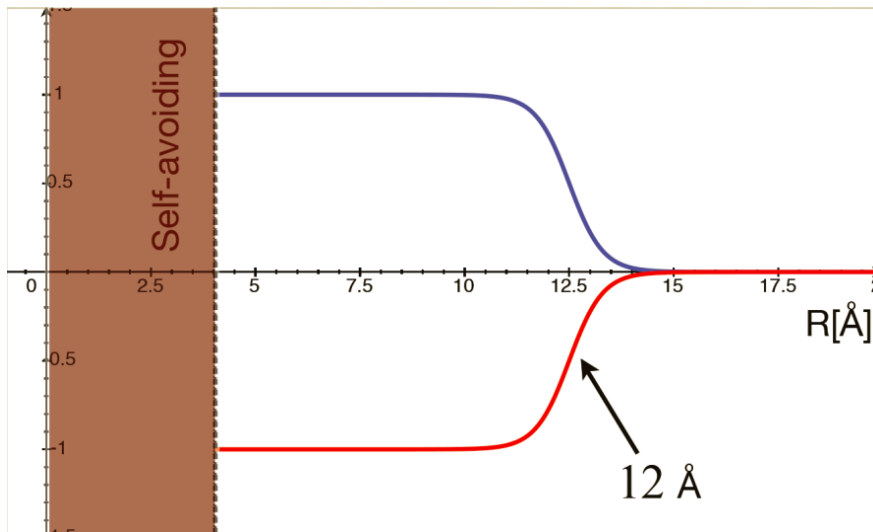


Reference Structure
for RMSD from RCSB

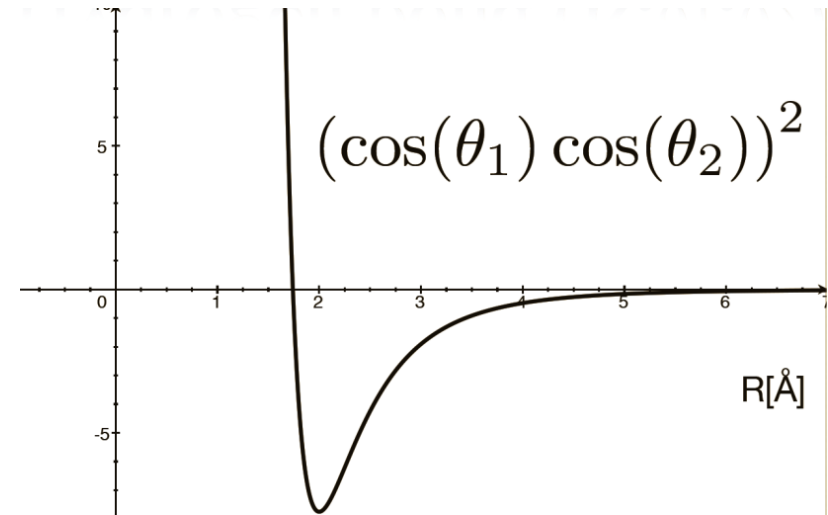


Model

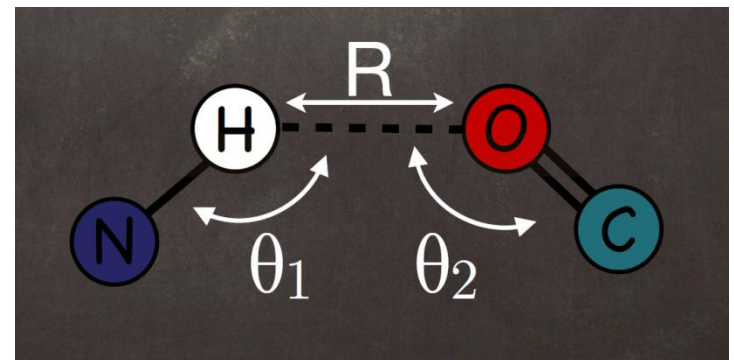
Side chains interactions (R)



Hydrogen bonds (R, θ_1, θ_2)



The depth depends on the interaction matrix, i.e. on the different nature of the amino acids



Model

Implicit solvent model

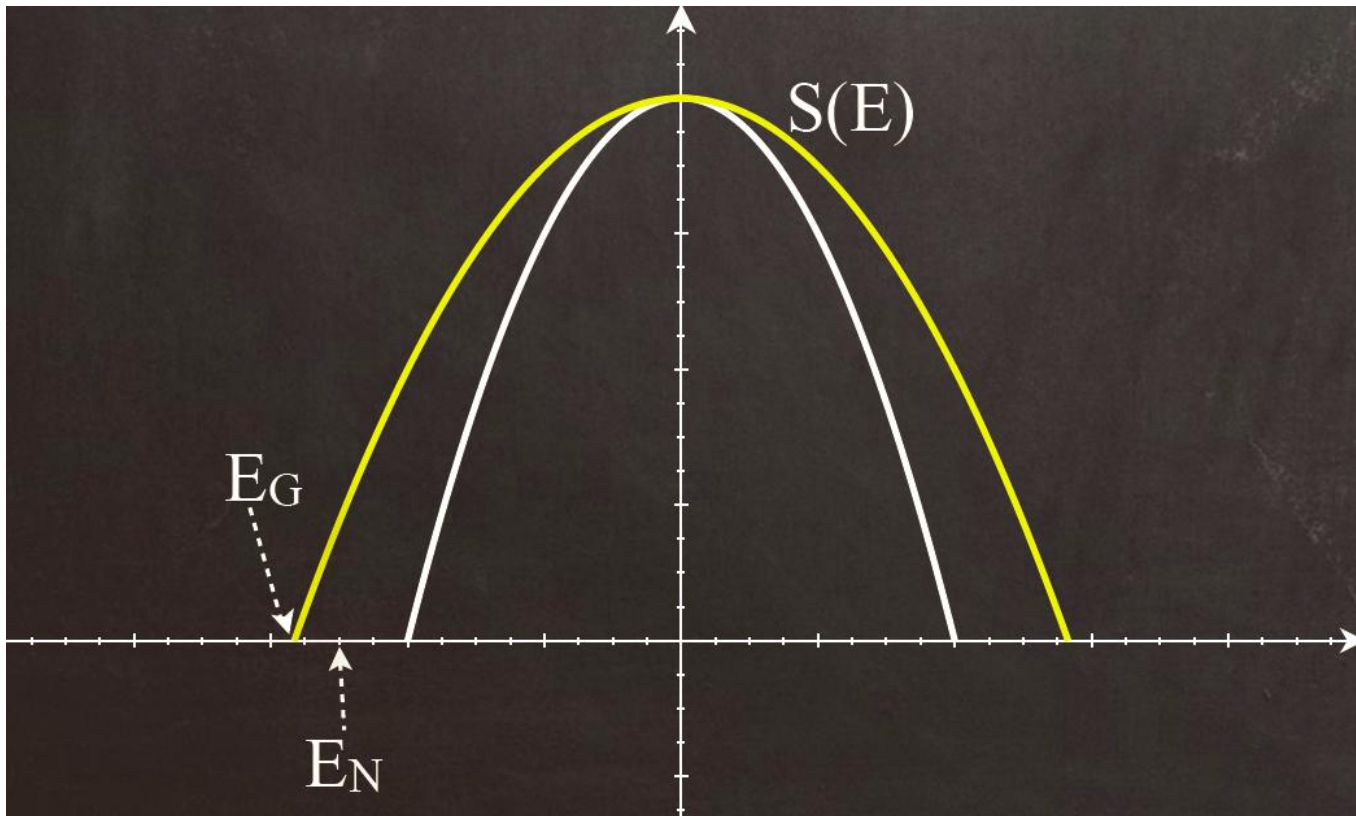
$$E_{\text{Sol}}^i = \begin{cases} \epsilon_{\text{Sol}}^i [\Omega - \Omega^i] & \Omega^i \leq \Omega \\ 0 & \Omega^i > \Omega \end{cases}$$
$$\Omega^i = \sum_j \left(1 - \frac{1}{1 + e^{2.5(r_{\text{max}} - r_{ij})}} \right)$$

Ω = threshold for the number of contacts above which the amino acid is considered fully buried

ϵ_{Sol}^i = (rescaled) Dolittle hydrophobicity index
> 0 hydrophobic aa
< 0 hydrophilic aa

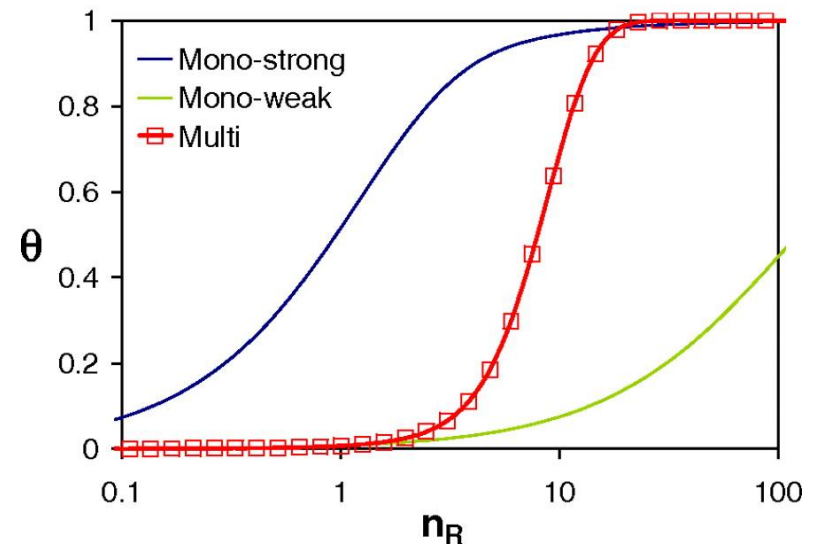
Model

Mean field theory Random Energy Model

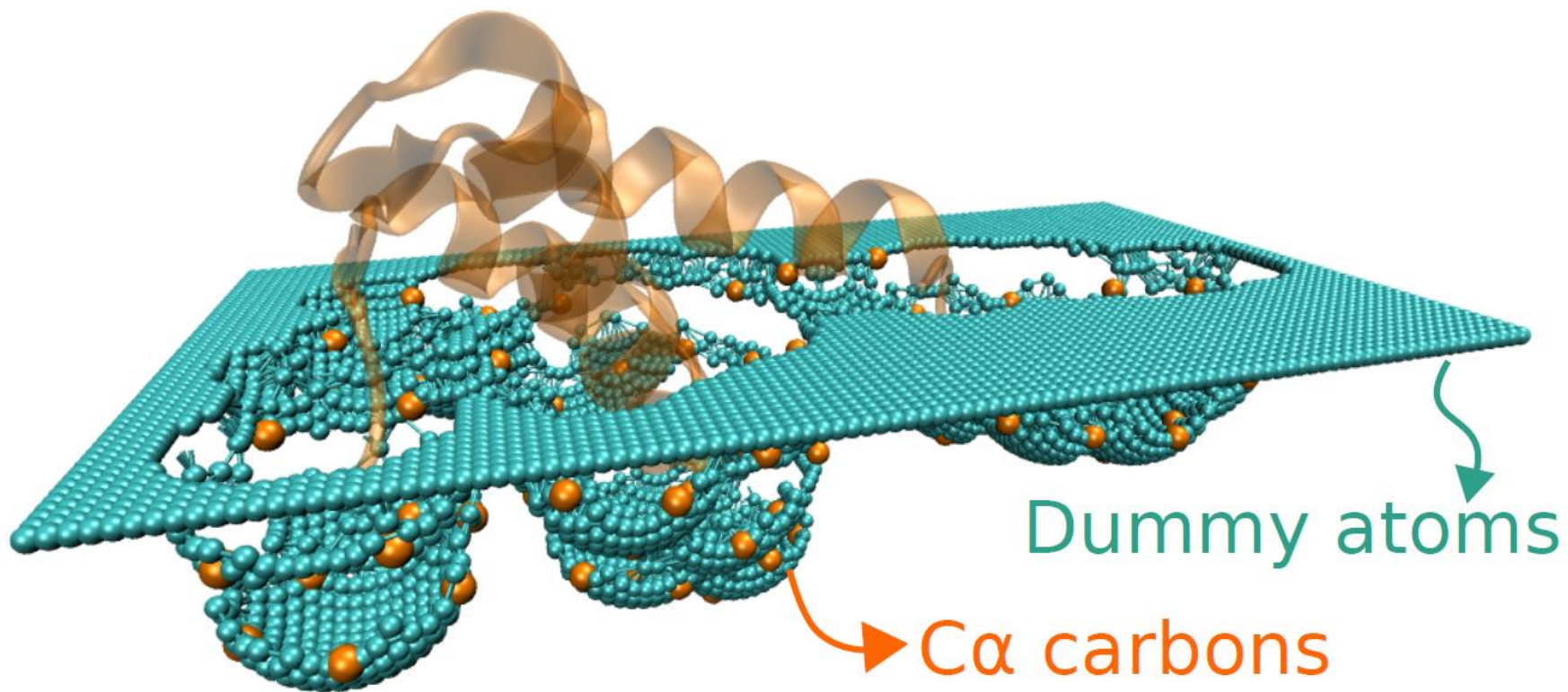


Bio Velcro

- Automated Bio Marker (ABM) - In silico automated tumor targeting
- Selective tumor cell targeting using low-affinity, multivalent interactions



- CD47-signal regulatory protein alpha

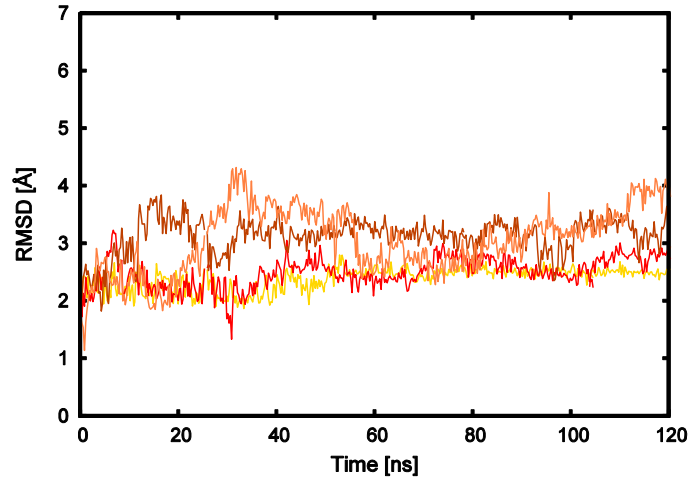


Full atomistic 2L09

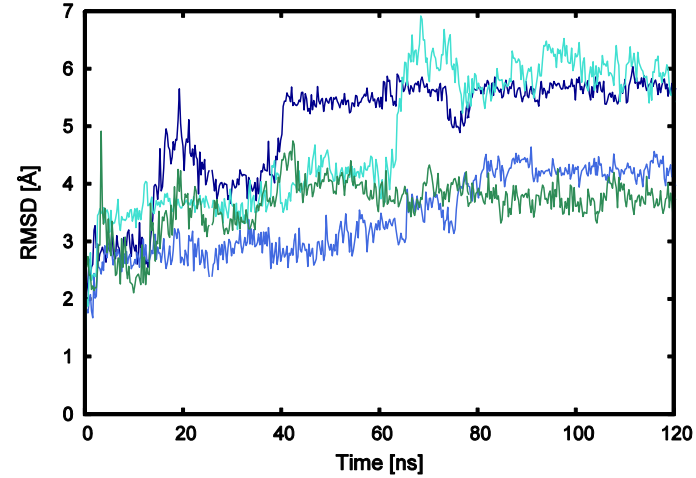
GROMACS

- amber99sb
- TIP3P water
- T = 298 K ; P = 1 atm

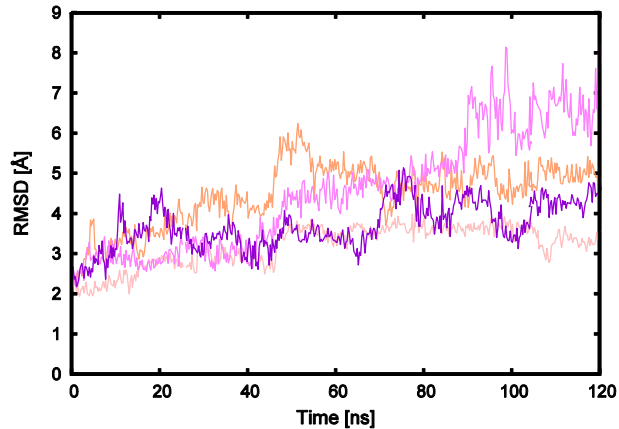
Native



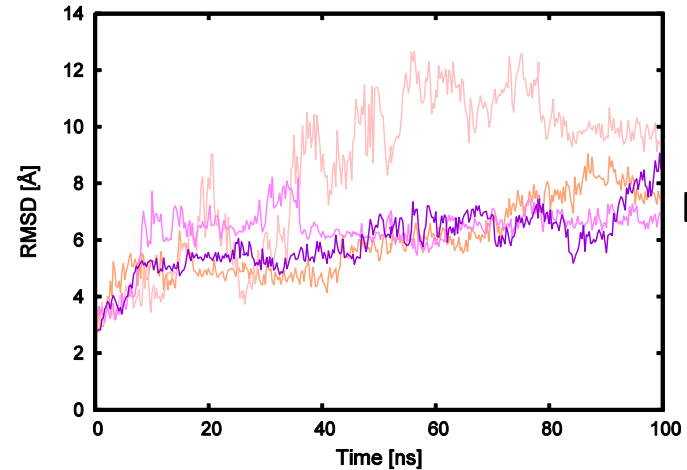
Designed



Random



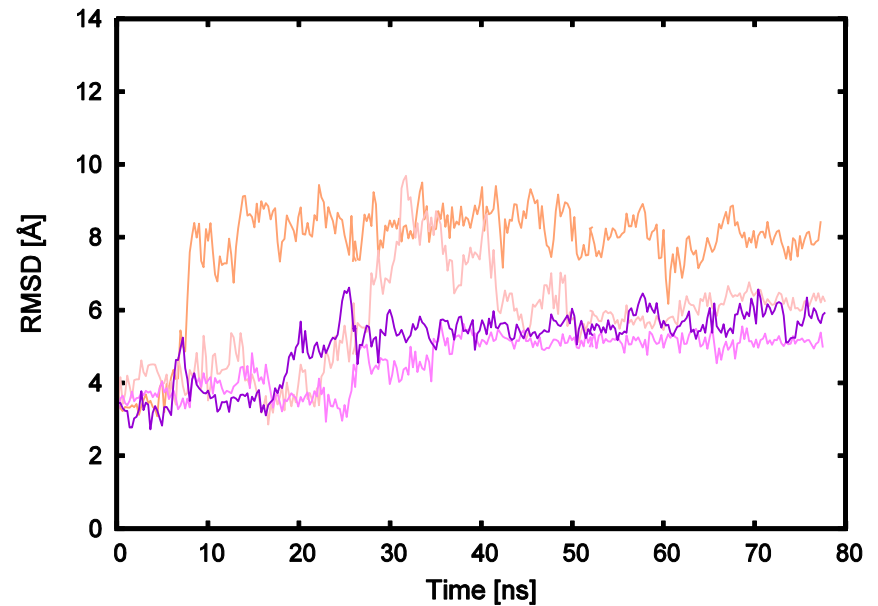
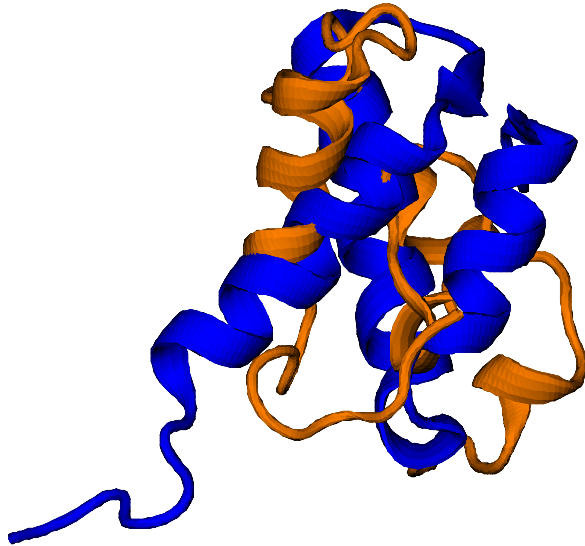
Designed Repeat



Full atomistic 2L09

GROMACS

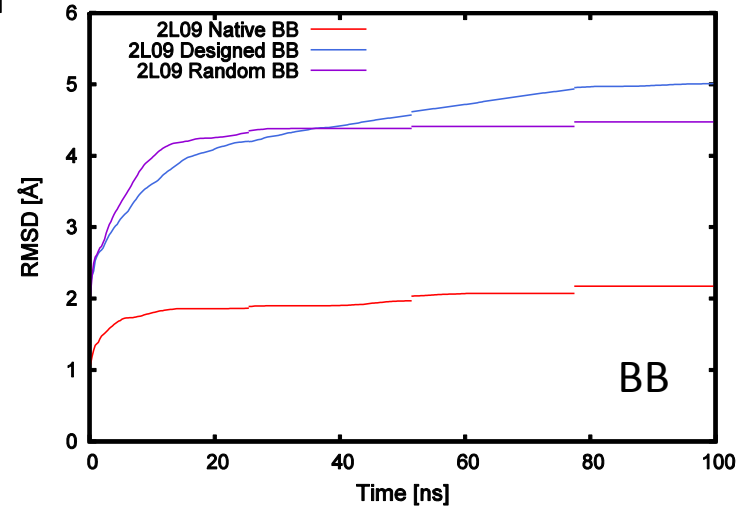
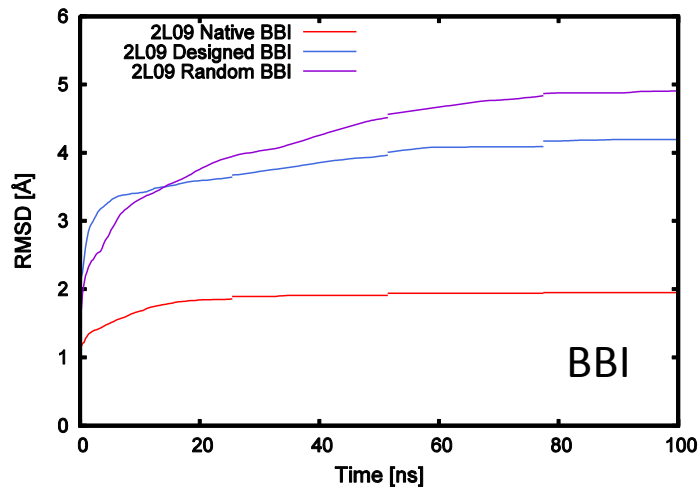
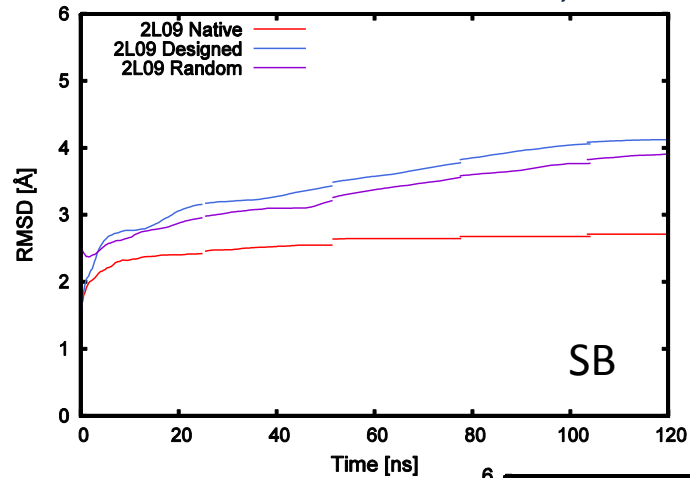
- amber99sb
- TIP3P water
- T = 358 K ; P = 1 atm



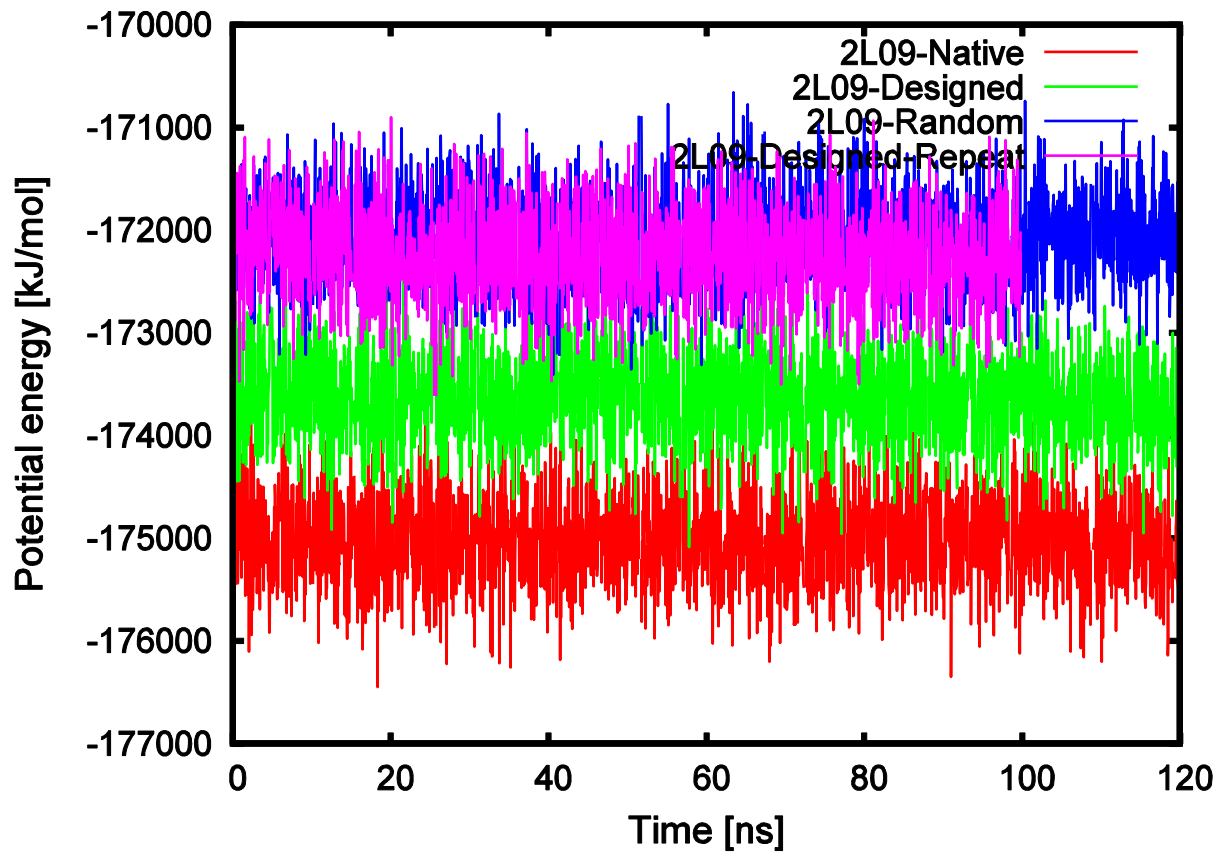
Full atomistic 2L09

GROMACS

- amber99sb
- TIP3P water
- T = 358 K ; P = 1 atm



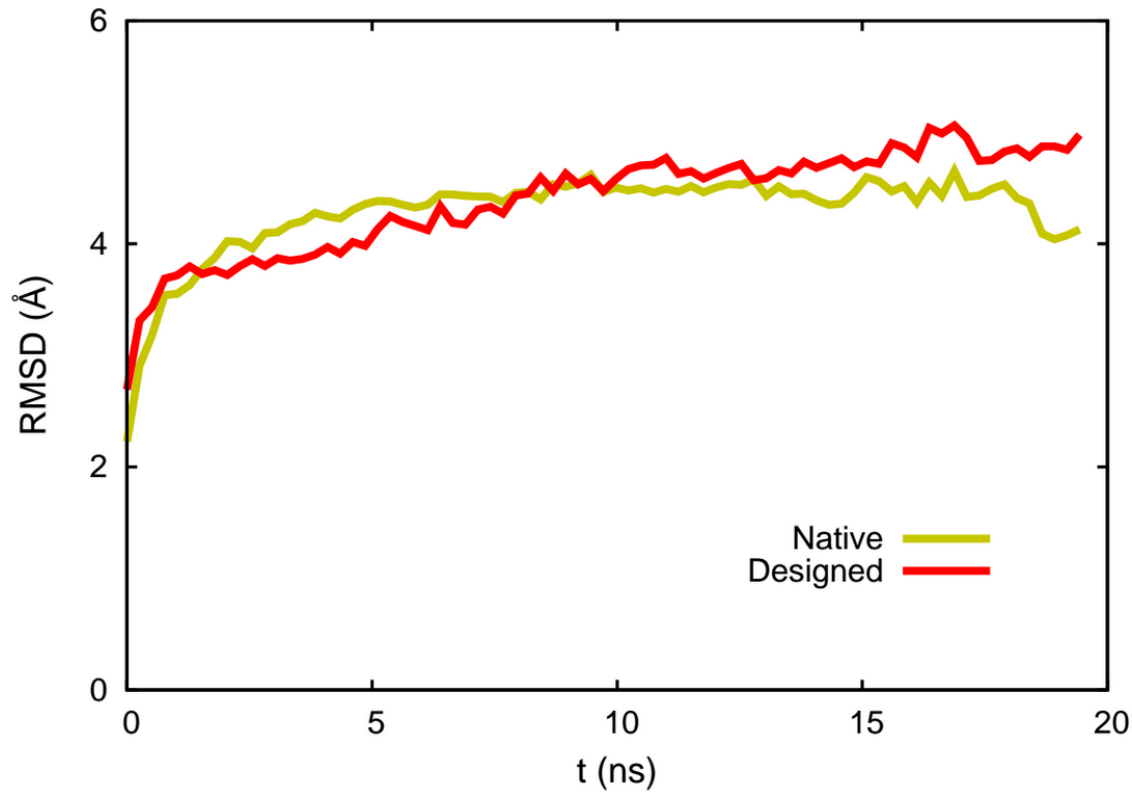
Full atomistic 2L09



Full atomistic 2L09

ORAC

- amber99sb
- TIP3P water
- T = 298 K ; P = 1 atm



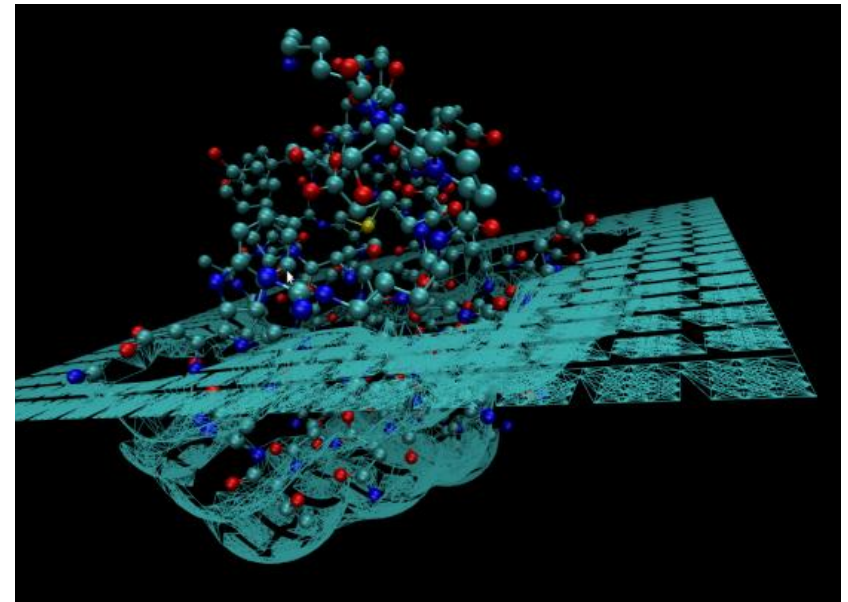
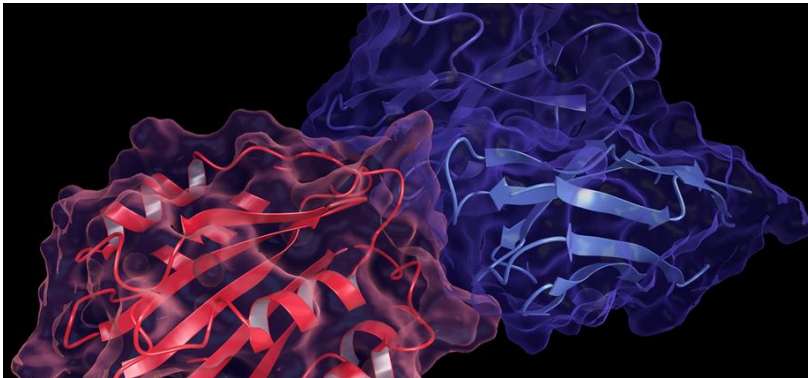
Reference Structure: SOLUTION NMR from RCSB

Protein-ligand binding OFF LATTICE

REAL SYSTEM:
protein docking

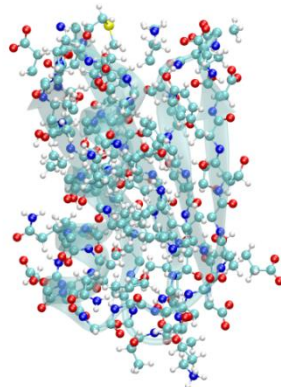


TEST SYSTEM:
protein-like surface

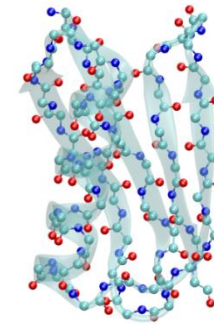


The CATERPILLAR model

Constraints are the key to
understand proteins



**ATOMISTIC
REPRESENTATION**

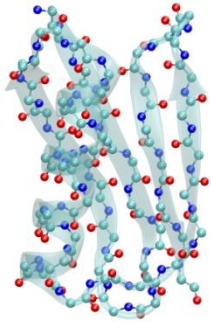


**COARSE GRAINED:
CATERPILLAR^[1-2]**

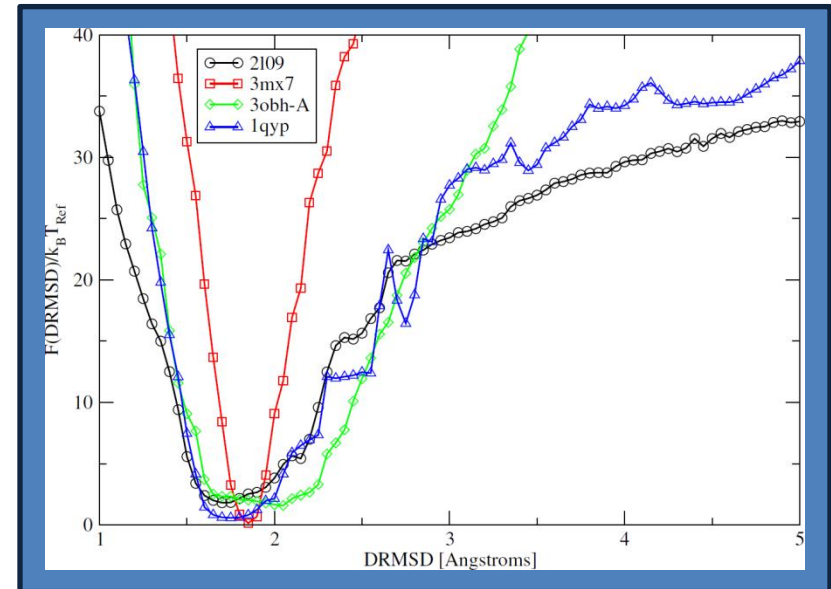
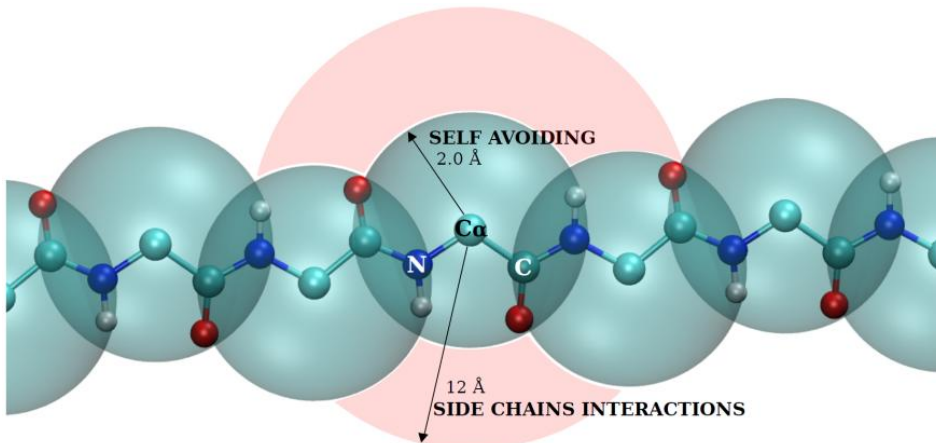
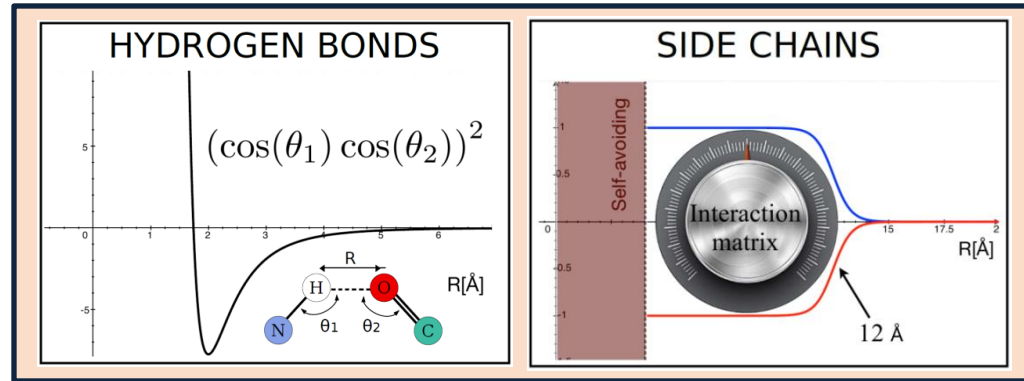
[1] Coluzza, Ivan. "A coarse-grained approach to protein design: learning from design to understand folding." PloS one, 6, e20853 (2011)

[2] Coluzza, Ivan. "Transferable coarse-grained potential for de novo protein folding and design." PloS one, 9, e112852 (2014)

The CATERPILLAR model



**COARSE GRAINED:
CATERPILLAR^[1-2]**



- [1] Coluzza, Ivan. "A coarse-grained approach to protein design: learning from design to understand folding." PloS one, 6, e20853 (2011)
 [2] Coluzza, Ivan. "Transferable coarse-grained potential for de novo protein folding and design." PloS one, 9, e112852 (2014)