The Abdus Salam
International Centre for Theoretical Physics

United Nations
Educational, Scientific and
Cultural Organization

IAEA
International Atomic Energy Agency

2268-6

Conference on Nanotechnology for Biological and Biomedical Applications (Nano-Bio-Med)

10 - 14 October 2011

Designing short peptides with high binding affinity for organic molecules

Alessandro LAIO

SISSA, International School for Advanced Studies
via Bonomea 265, 34136 Trieste
Italy

Strada Costiera 11, 34151 Trieste, Italy - Tel.+39 040 2240 111; Fax +39 040 224 163 - sci_info@ictp.it

Designing short peptides with high binding affinity for organic molecules.

Alessandro Laio

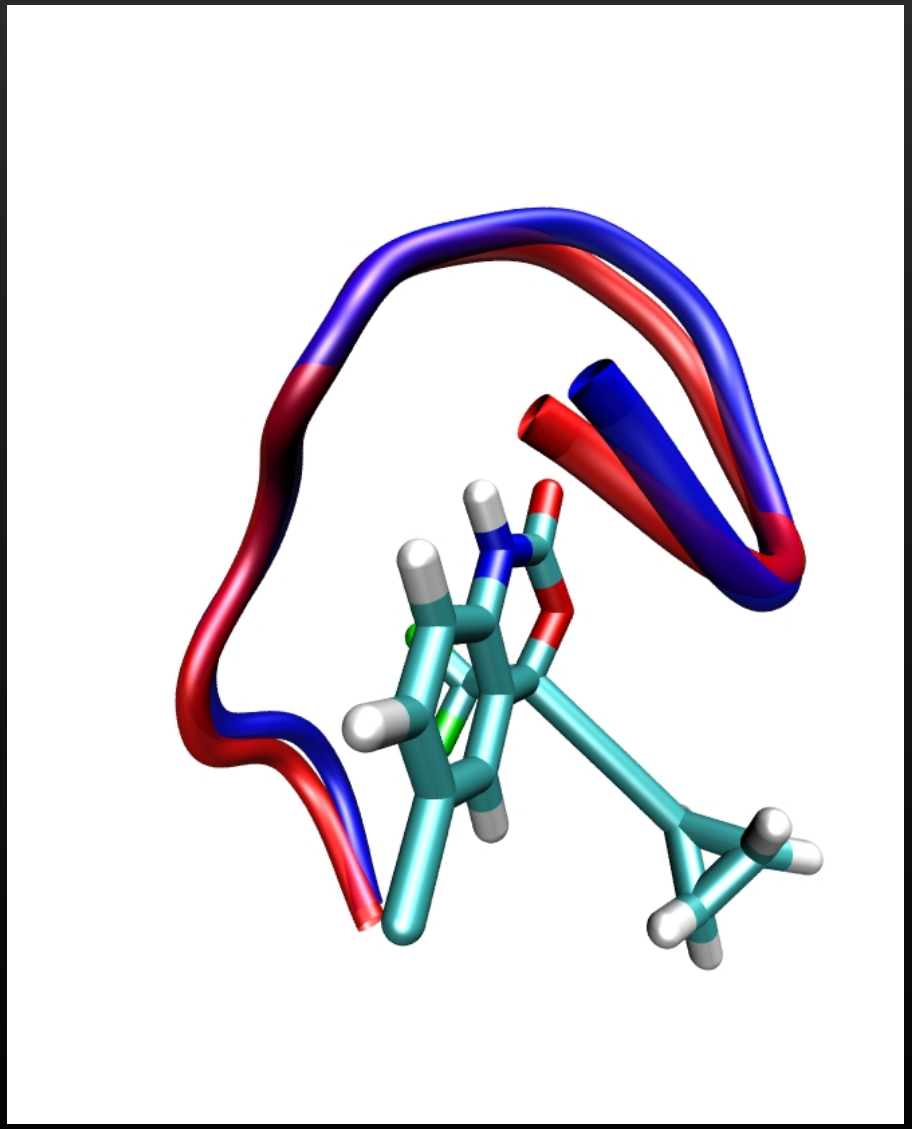
SISSA, Trieste

Rolando Hong, Federico Berti Pilar Cossio,
Flavio Seno and Antonio Trovato

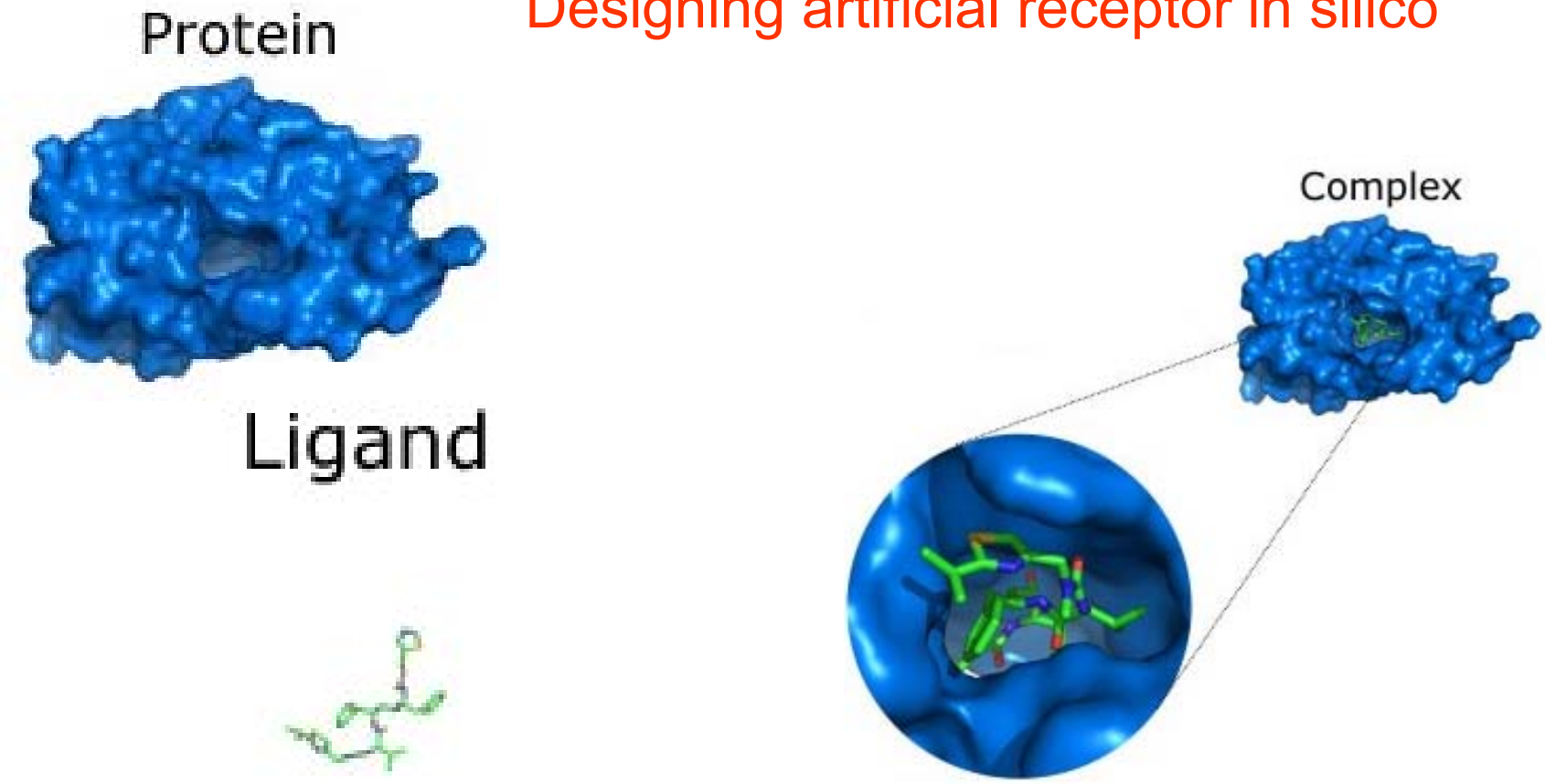
The goal:

Optimizing the primary sequence of a short peptide in order to bind strongly and selectively to:

- **drugs/metabolites/small molecules**
- **proteins**



Designing artificial receptor in silico



EMBEDDING THESE PEPTIDE IN NANOSCALE SENSOR WILL ALLOW:

- Detecting the presence of markers
- Monitoring the concentrations of drugs

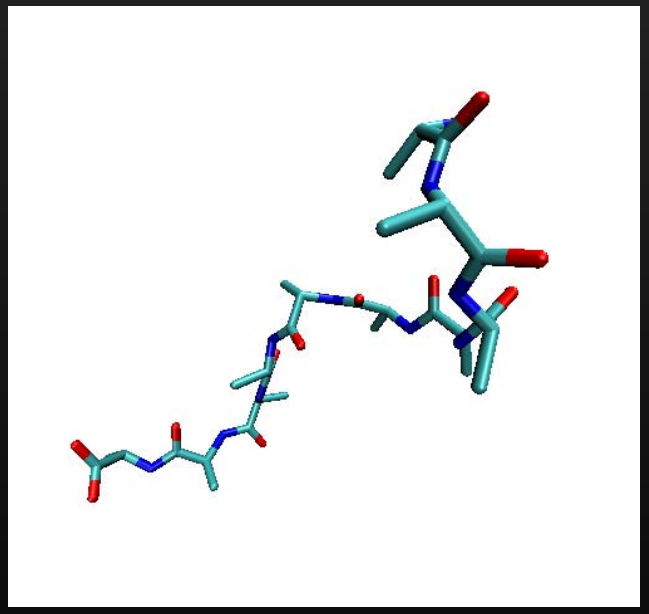
The challenge:

Sequence space

Number of sequences = 20^N

For N= 10
~ 10 000 billions !!!

Conformational space

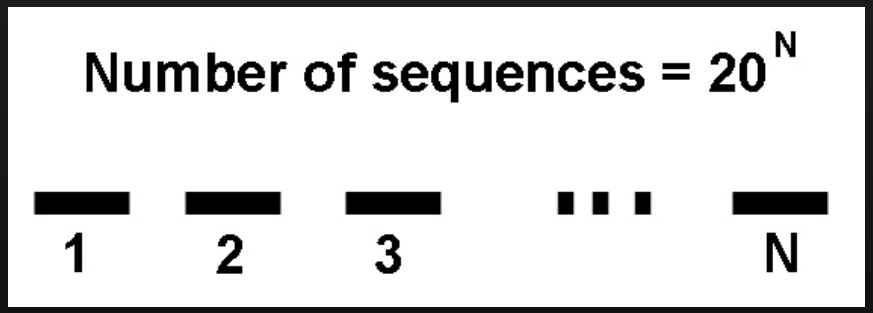


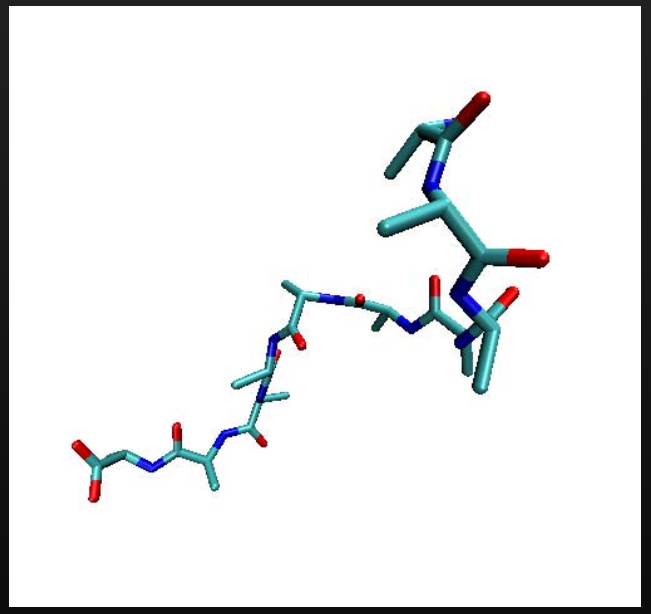
For N= 10
~ 20 relevant dihedrals !!!

The challenge:

Sequence space

Conformational space

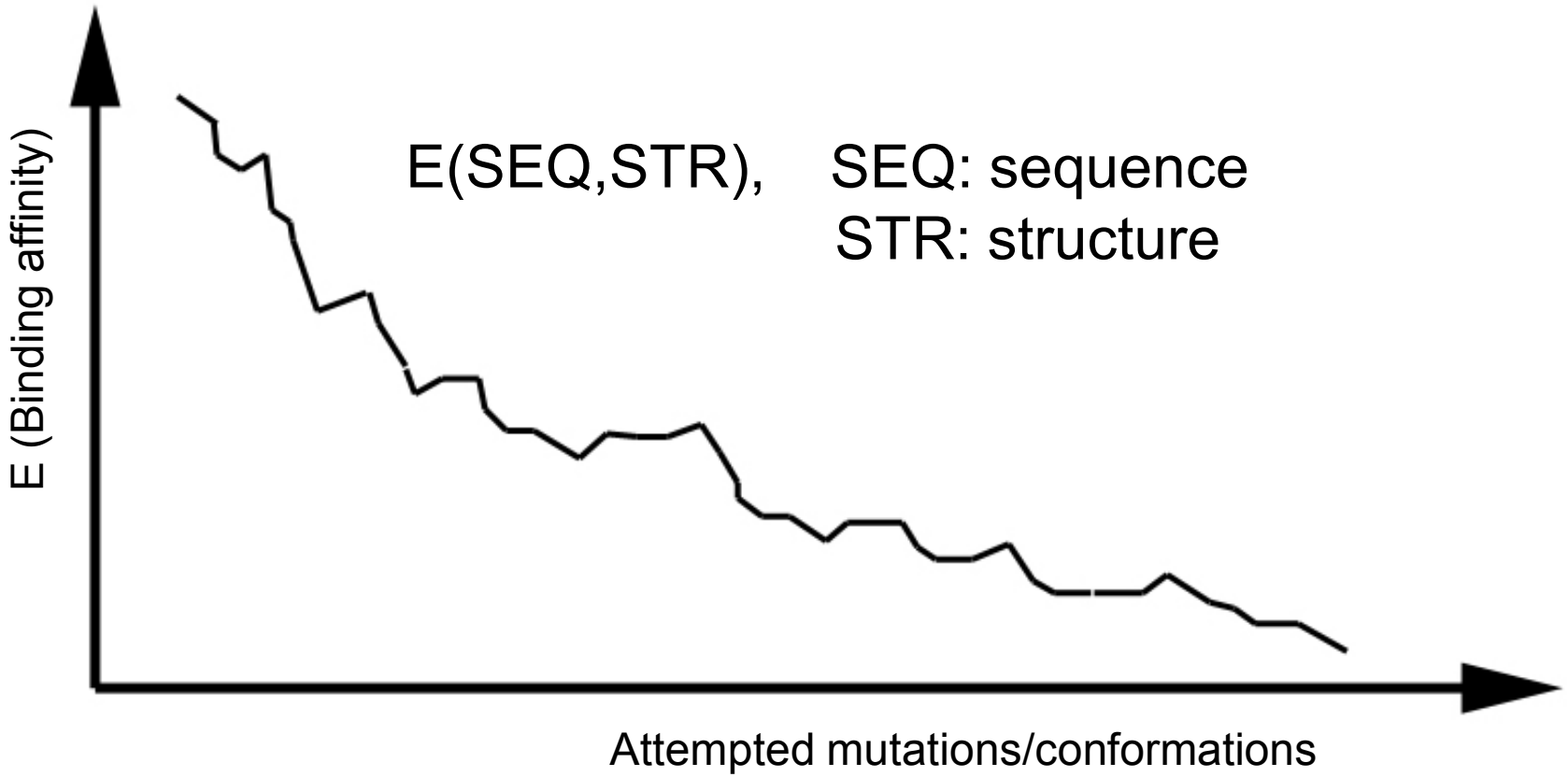




Discrete space

Continuous space

Optimization



What are we optimizing?

E(SEQ,STR)

Is the rigid docking scoring function [1] towards the target molecule of a peptide of sequence SEQ and structure STR.

[1] Trott O and Olson AJ. *Autodock vina: improving the speed and accuracy of docking with a new scoring function and mul tithreading.* **J. Comput. Chem.**, 31:450-461, 2010.

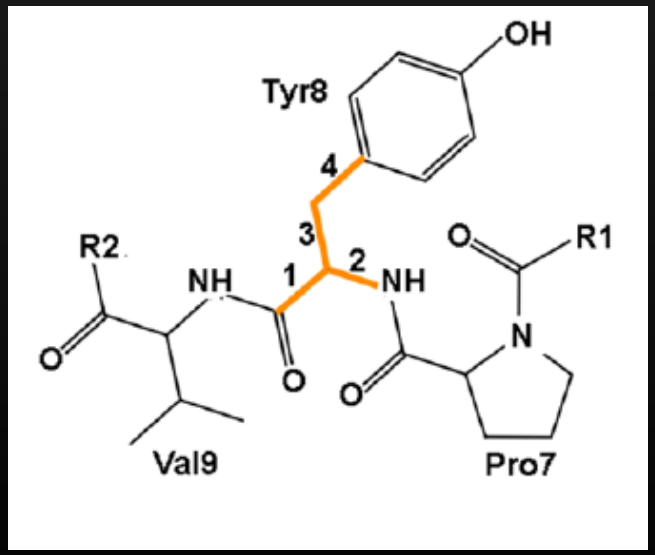
Monte Carlo Optimization

- Initial state: SEQ,STR
- Randomly change SEQ in SEQ'
- Find the structure (STR') maximizing the binding energy of SEQ'
- Accept the "move" (SEQ,STR) → (SEQ',STR') according to a Metropolis criterion:

$$P = \min\left(1, \exp\left(E(\mathrm{SEQ}, \mathrm{STR}) - E(\mathrm{SEQ}', \mathrm{STR}')\right)/T\right)$$

THE BOTTLENECK:
Find the structure (STR') maximizing the binding energy of SEQ'

- **Selectively explore only a few rotamers,** the ones generated rotating the backbone and sidechain dihedrals of the mutated residue

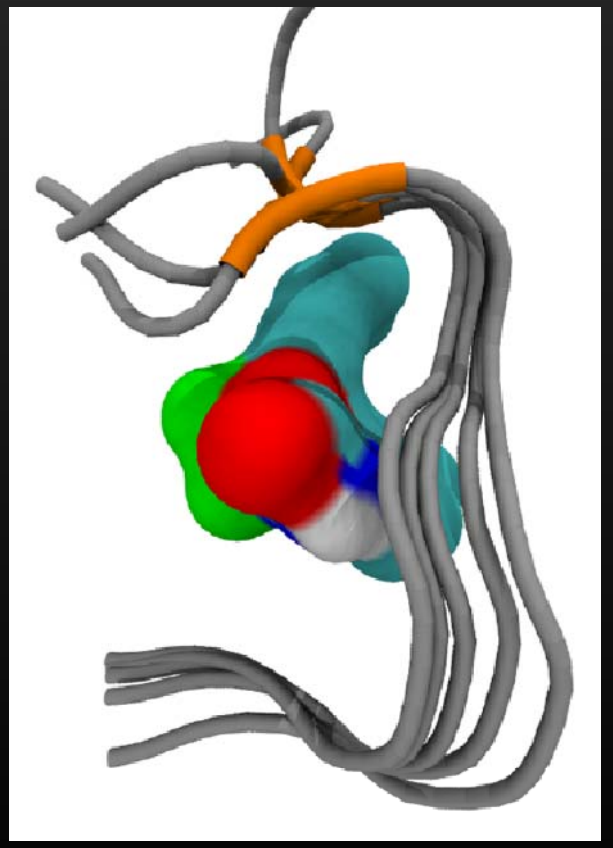


THE BOTTLENECK:
Find the structure (STR') maximizing the binding energy of SEQ'

- Each rotamer is relaxed by finite temperature molecular dynamics around the target

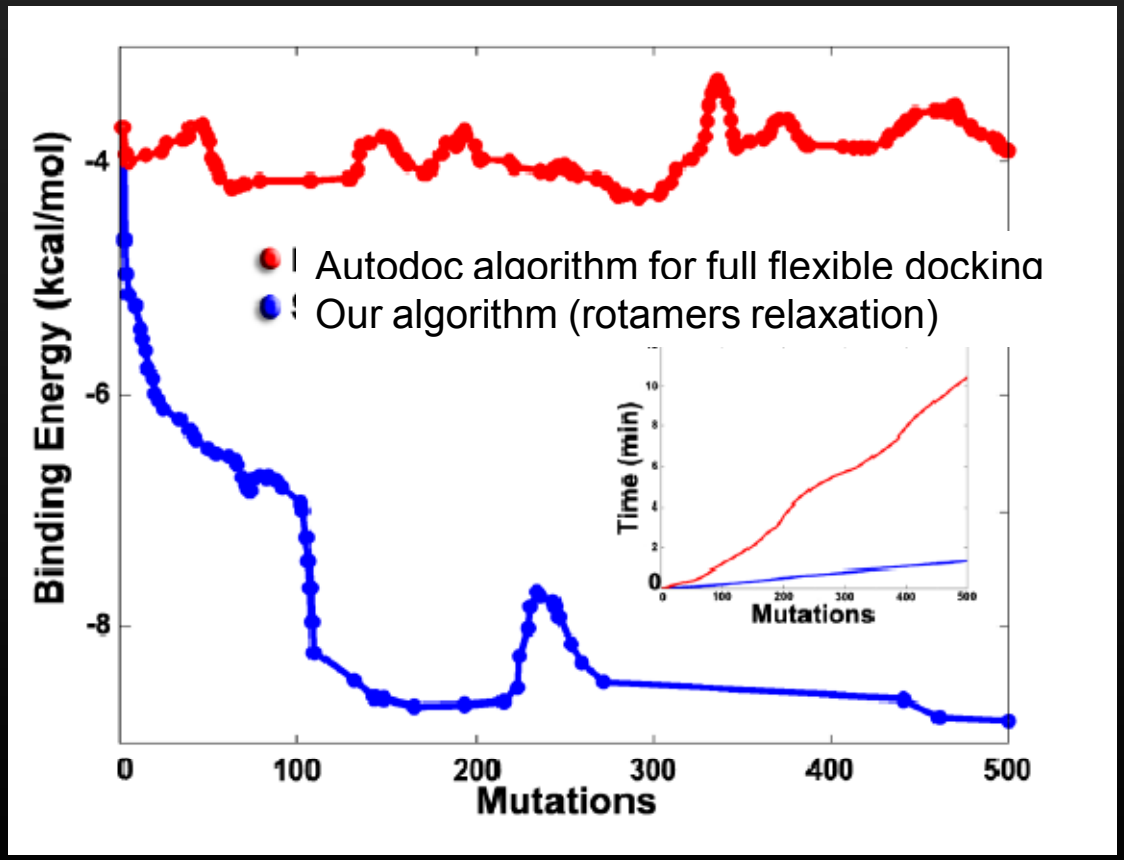
- For each final structure one computes the rigid docking energy

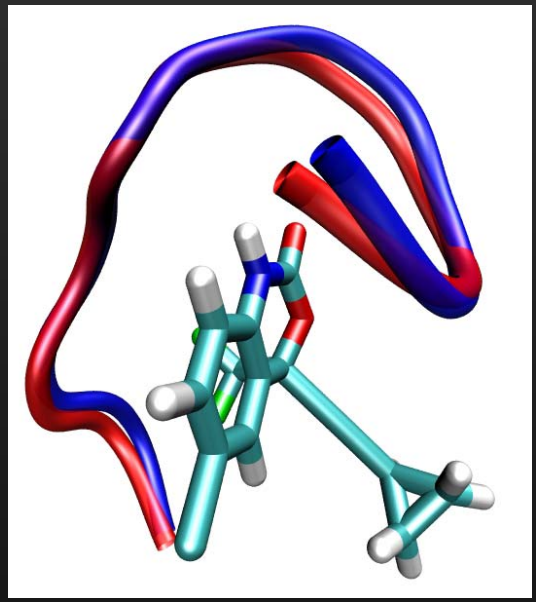
- We select the best binding pose. E(SEQ',STR') is the corresponding binding energy



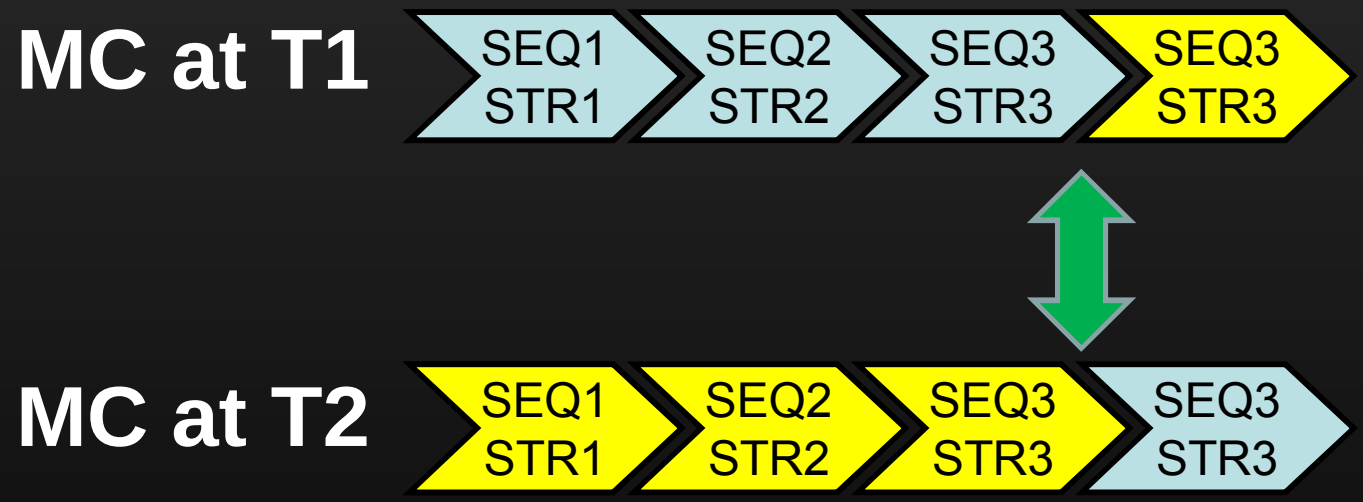
Monte Carlo Optimization

Target: Efavirenz



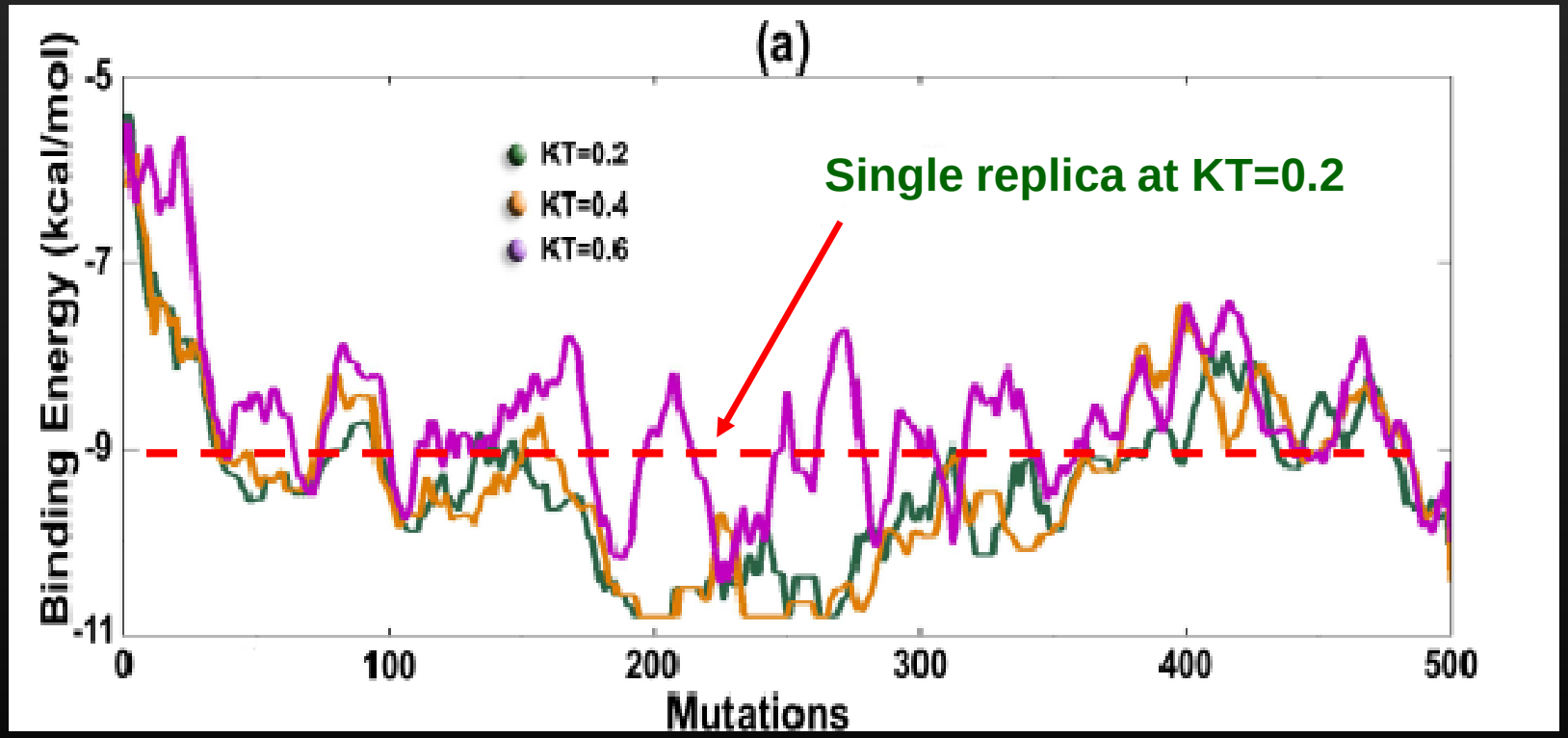


Avoiding traps: Replica Exchange MC

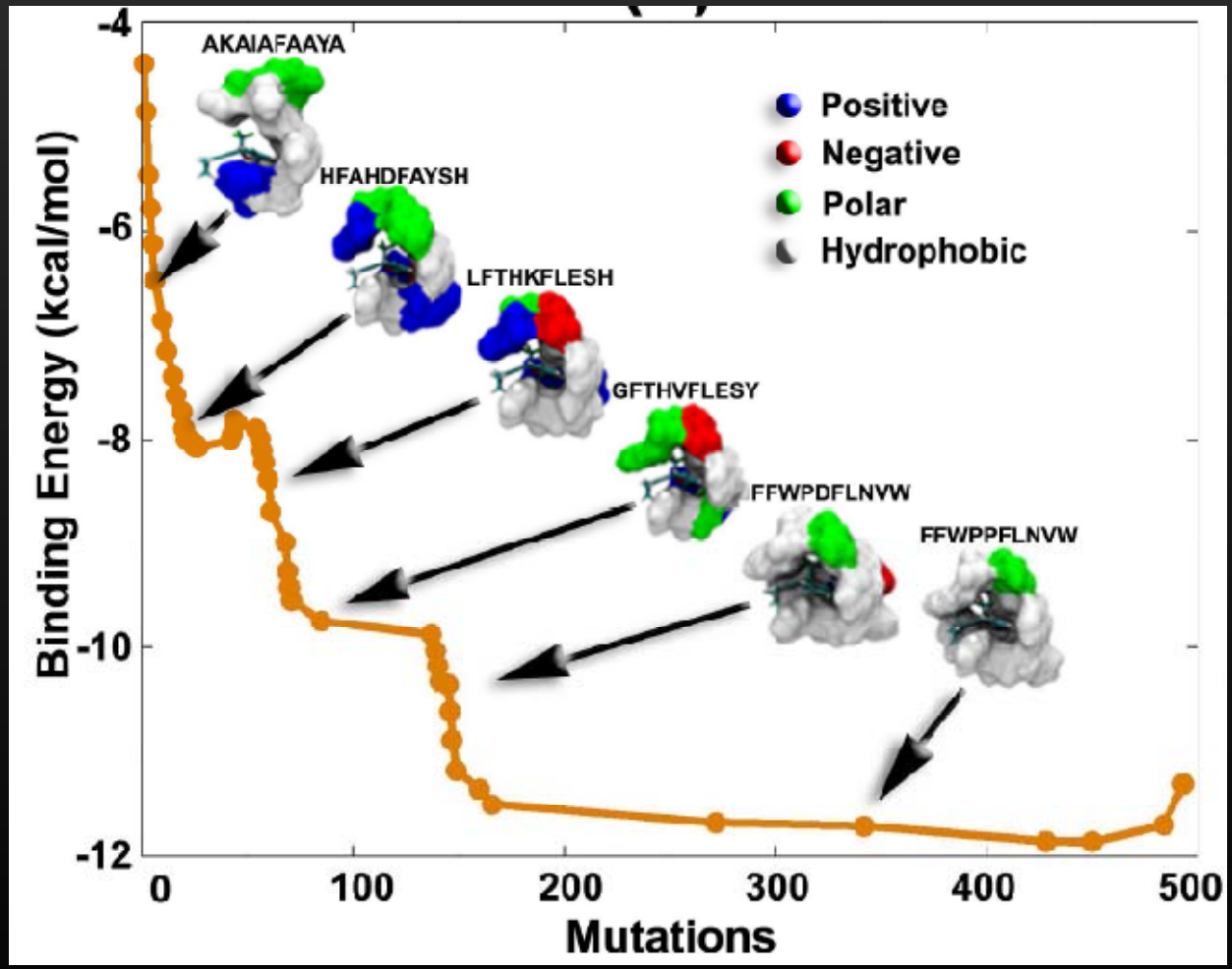


$$P = \min\left(1, \exp\left(E(\mathrm{SEQ},\mathrm{STR}) - E(\mathrm{SEQ}',\mathrm{STR}')\right) / \left(\frac{1}{T1} - \frac{1}{T2}\right)\right)$$

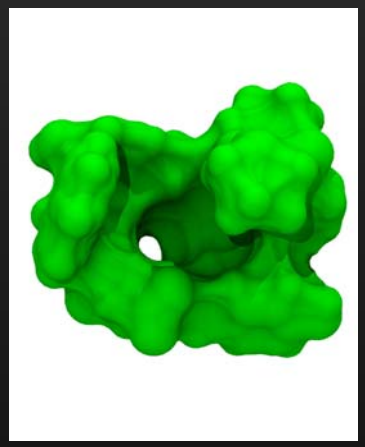
Avoiding traps: Replica Exchange MC

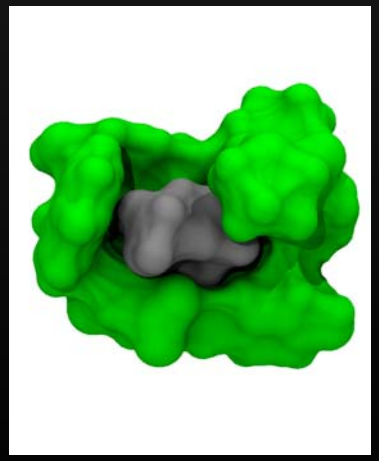


Wrapping the target molecule

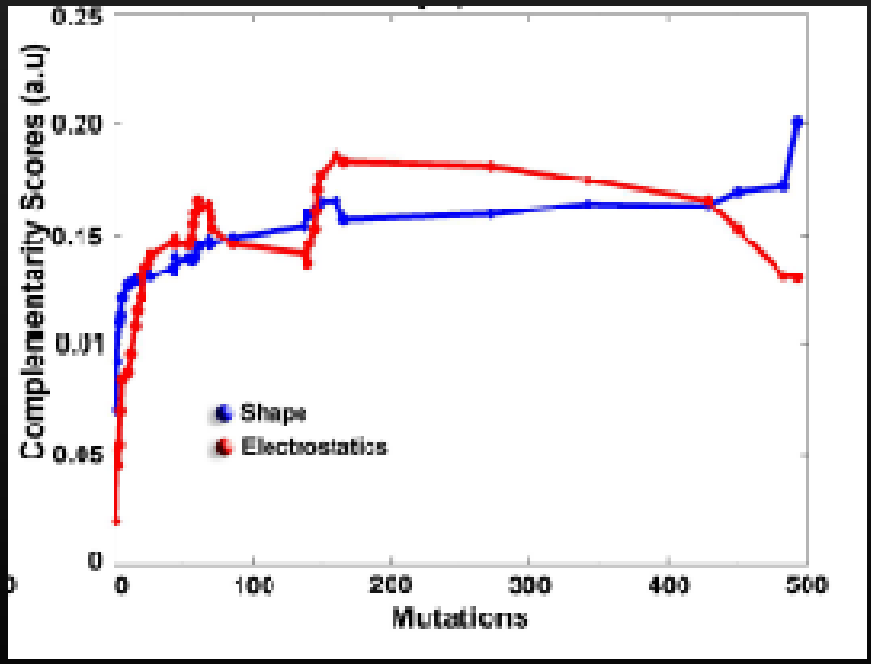


Wrapping the target molecule: shape complementarity

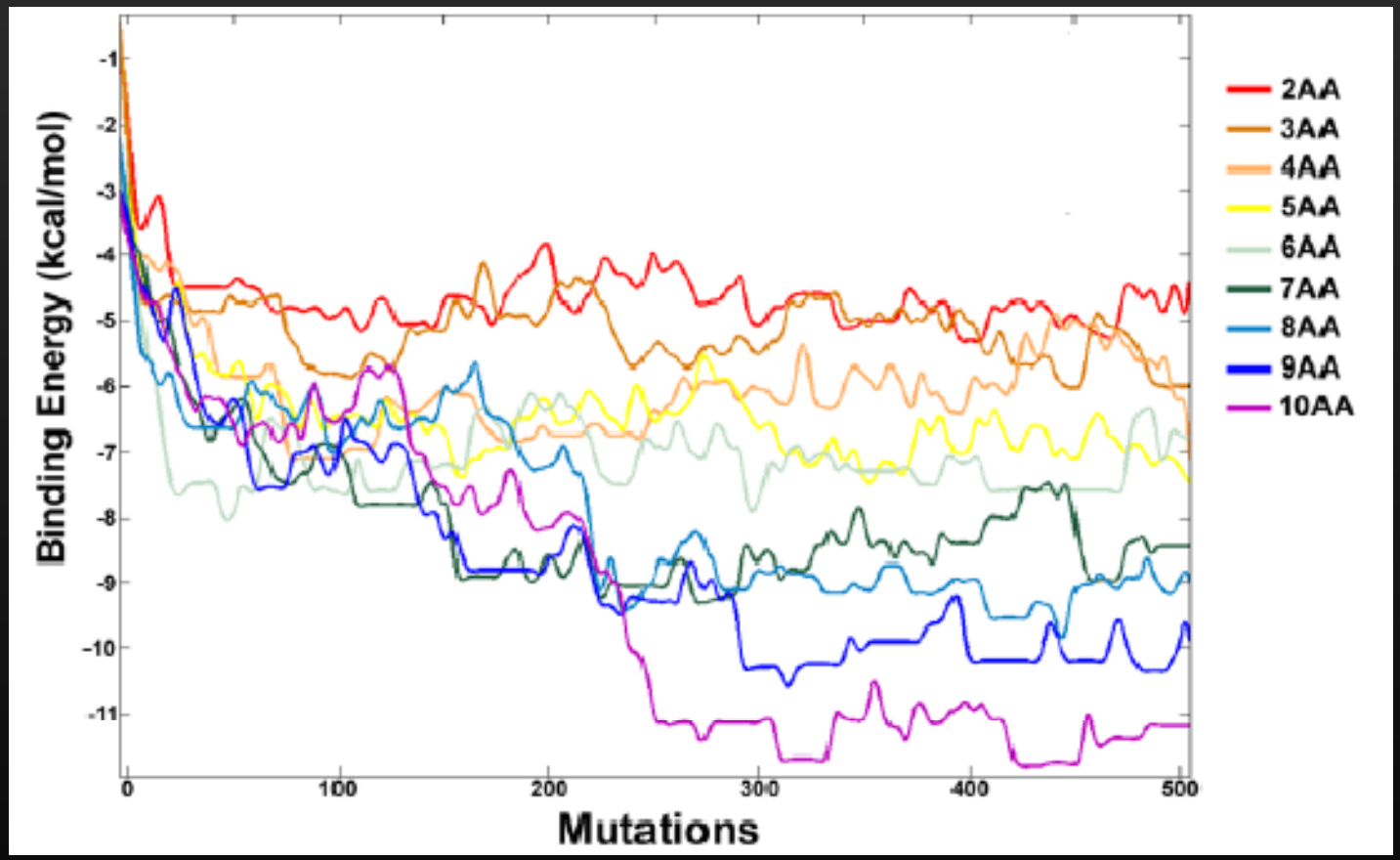




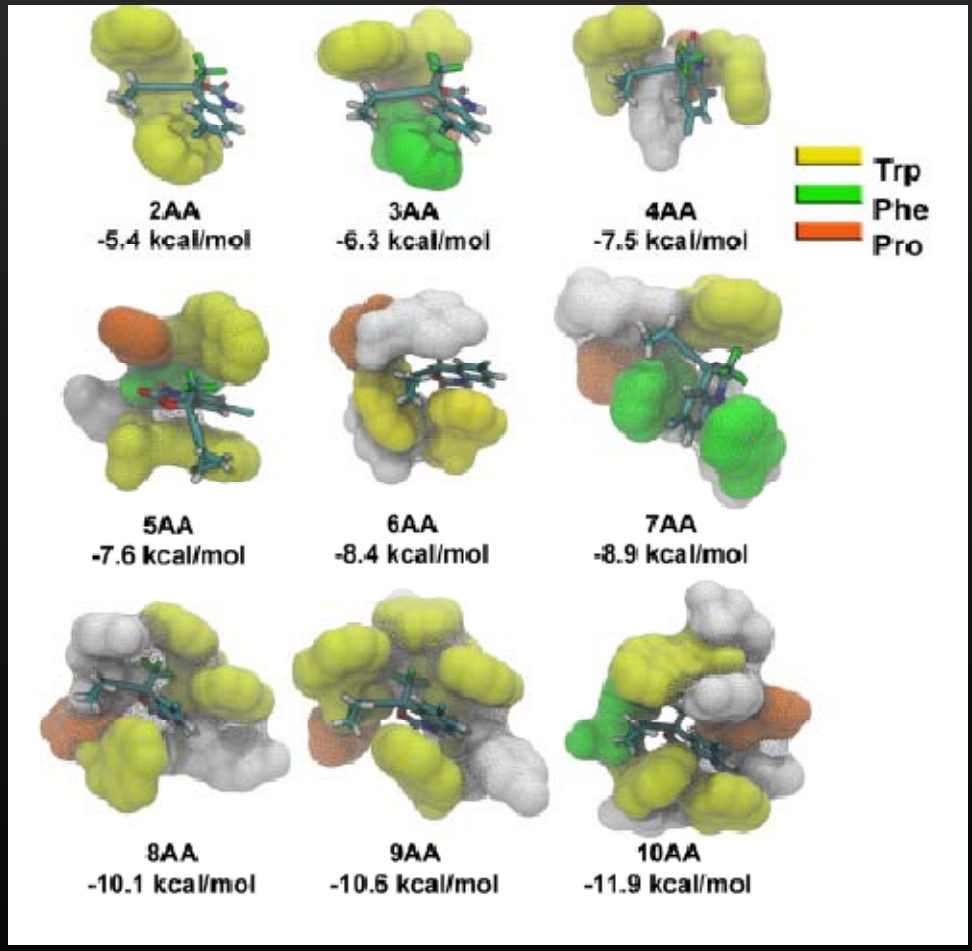
We count the number of vertices of the two surfaces that are in contact



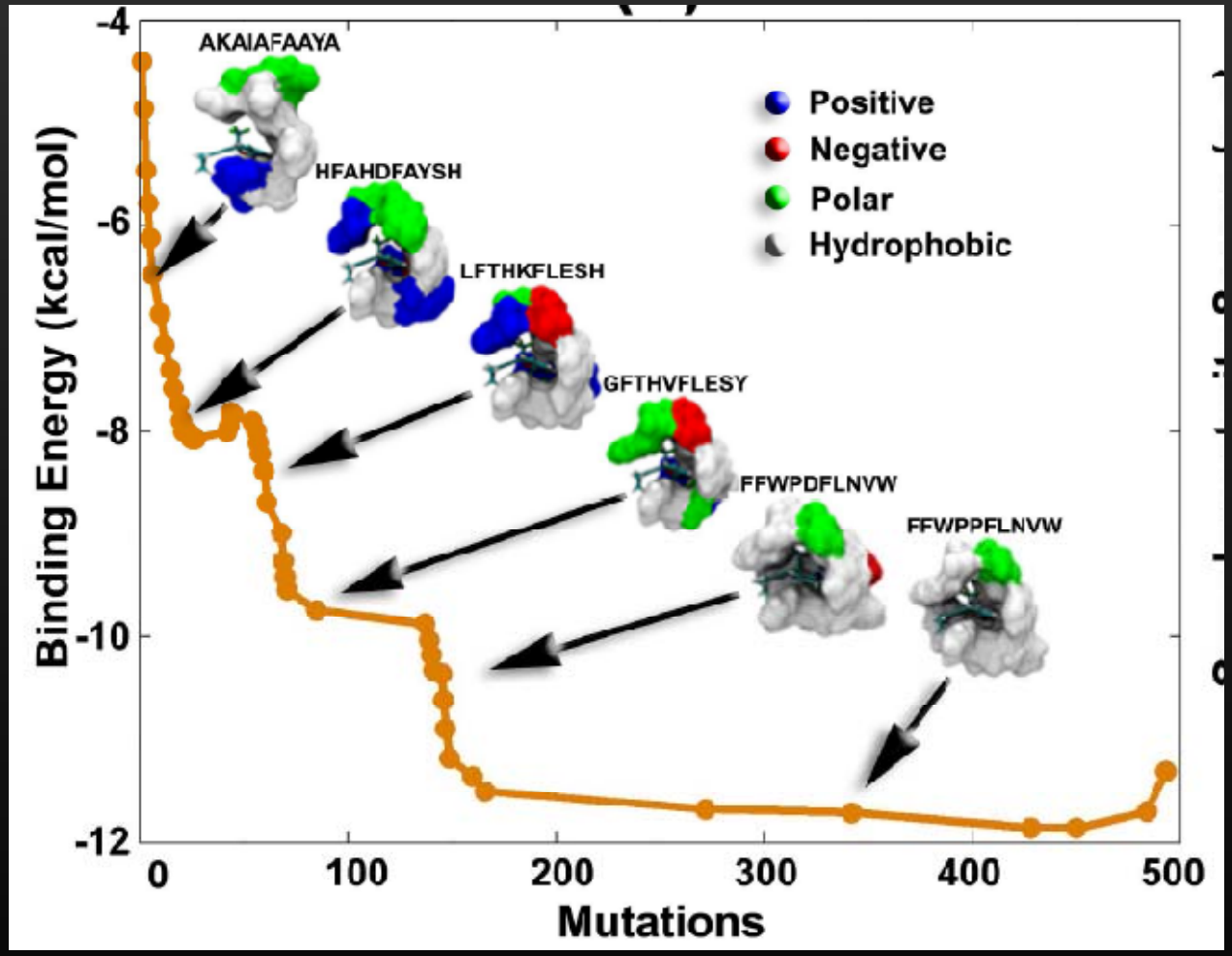
Optimal length of the peptide?



Optimal length of the peptide?



Results: experimental validation

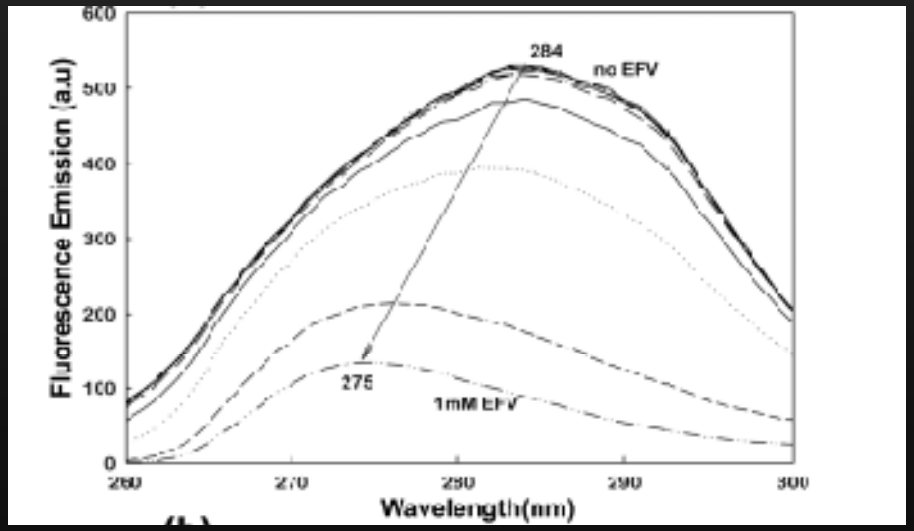


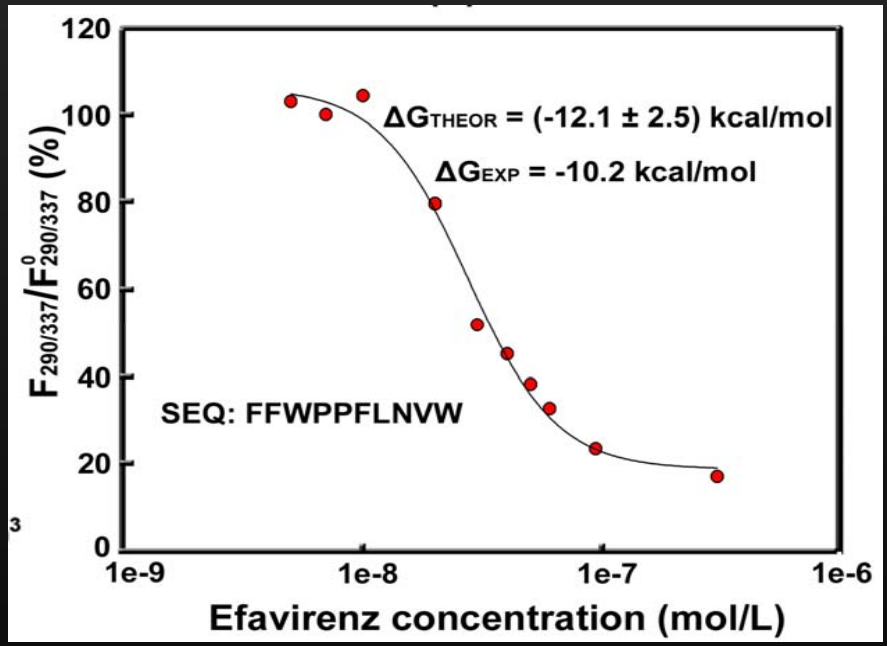
FFWPPFLNVW

Binding energy~-12 kcal/mol

Results: experimental validation

Fluorescence quenching of 10 nM solution of FFWPPFLNVW in 50 mM phosphate, 20% acetonitrile by EFV

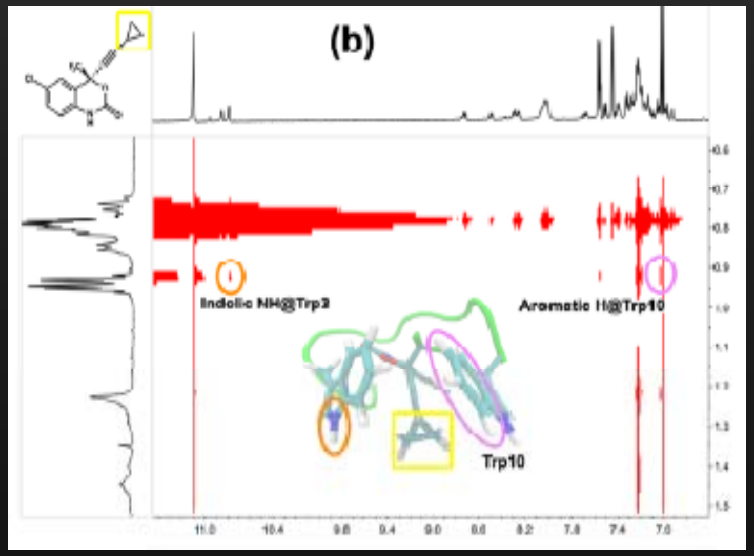


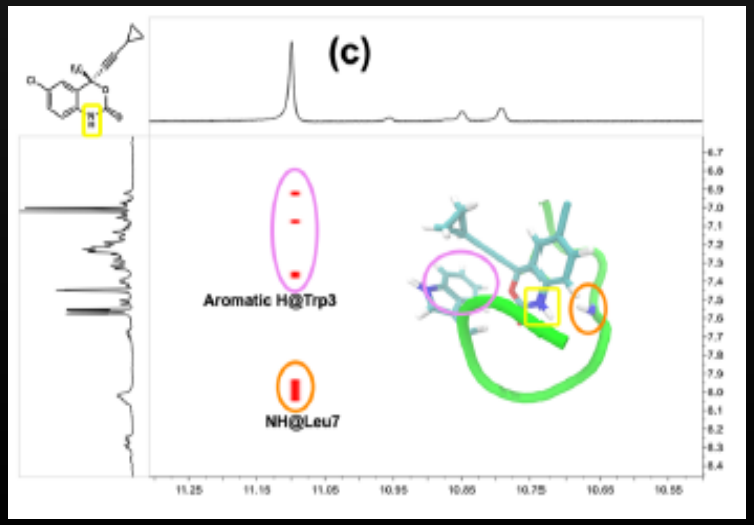


Fluorescence spectroscopy confirms a significant binding affinity

Is the structure of the complex predicted by the algorithm correct?

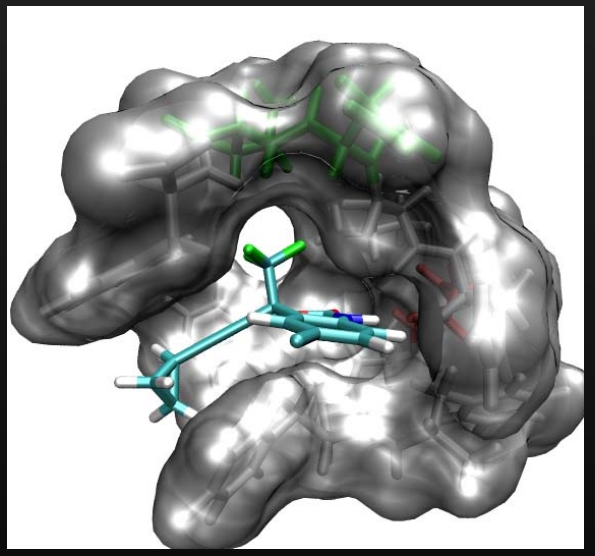
2D-NOESY crosspeaks between EFV and FFWPPFLNVW





Cycloprop. Group ←→ Arom. Trp10

Cycloprop. Group ←→ Indolic NH @Trp3



NH Group ←→ NH @Leu7

NH Group ←→ Arom. Trp10: NOT PRESENT

Result: our designed peptide actually binds to EFV.

We also correctly predict the structure of the complex

Have we been lucky?

We made two more attempts: in one case we also found a good binding affinity. In another the peptide is not soluble

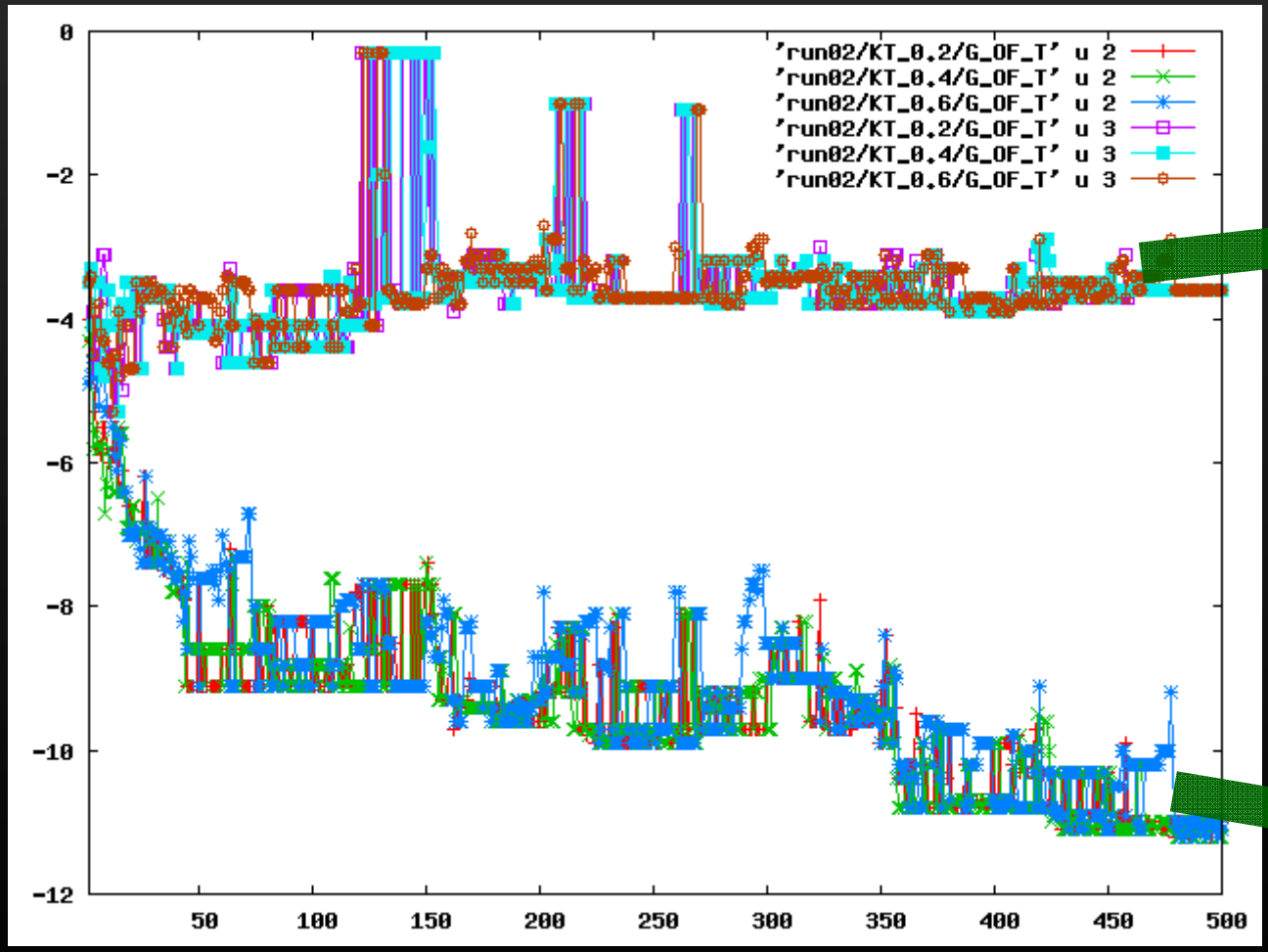
Ongoing:

1) Impose solubility and low aggregation propensity

2) Impose biocompatibility

3) SELECTIVITY

Selectivity: simultaneously optimize the binding to a target, and discourage the binding towards another target



Tenofovir

Efavirenz

Ongoing:

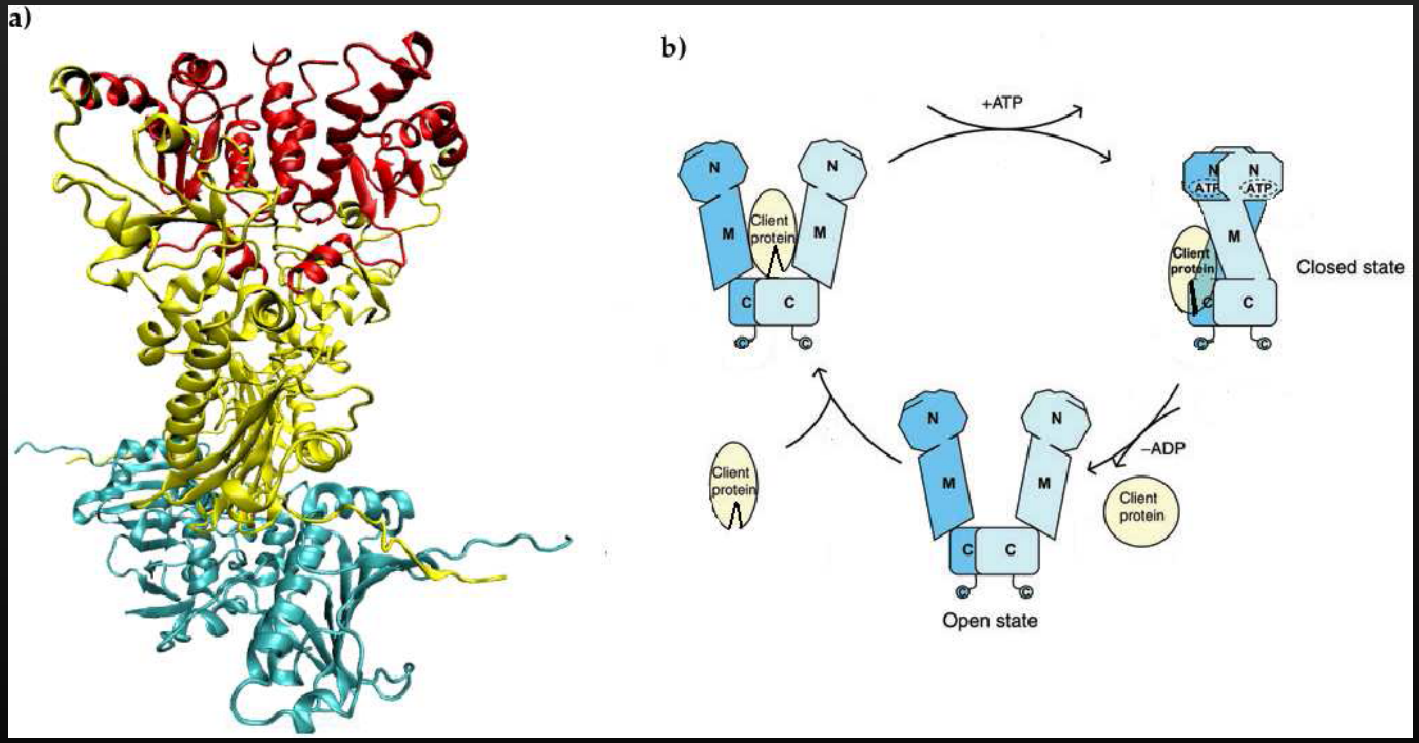
1) Impose solubility and low aggregation propensity

2) Impose biocompatibility

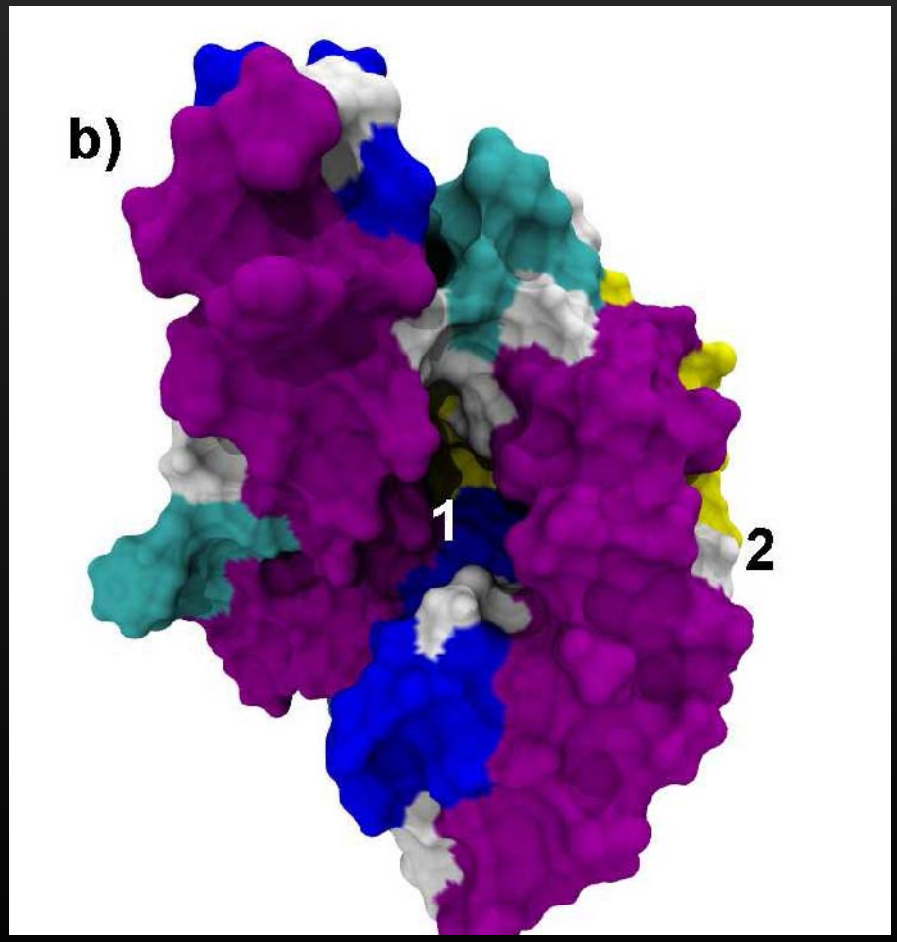
3) SELECTIVITY

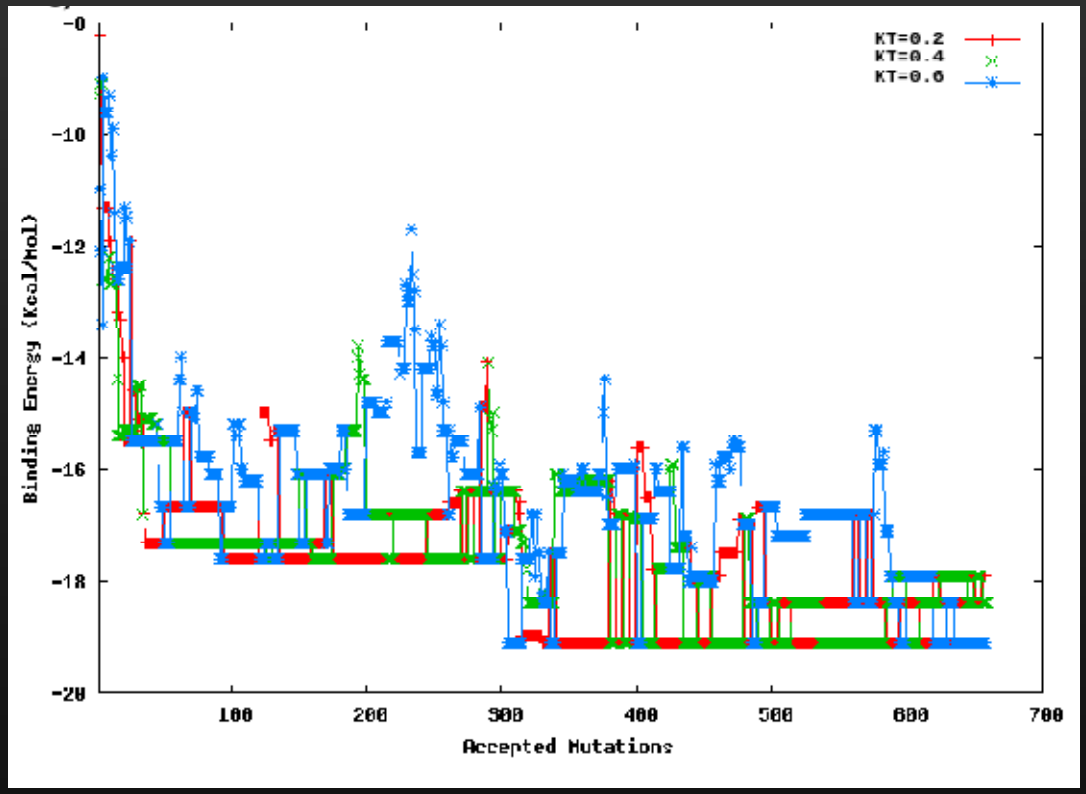
4) Designing binders for proteins

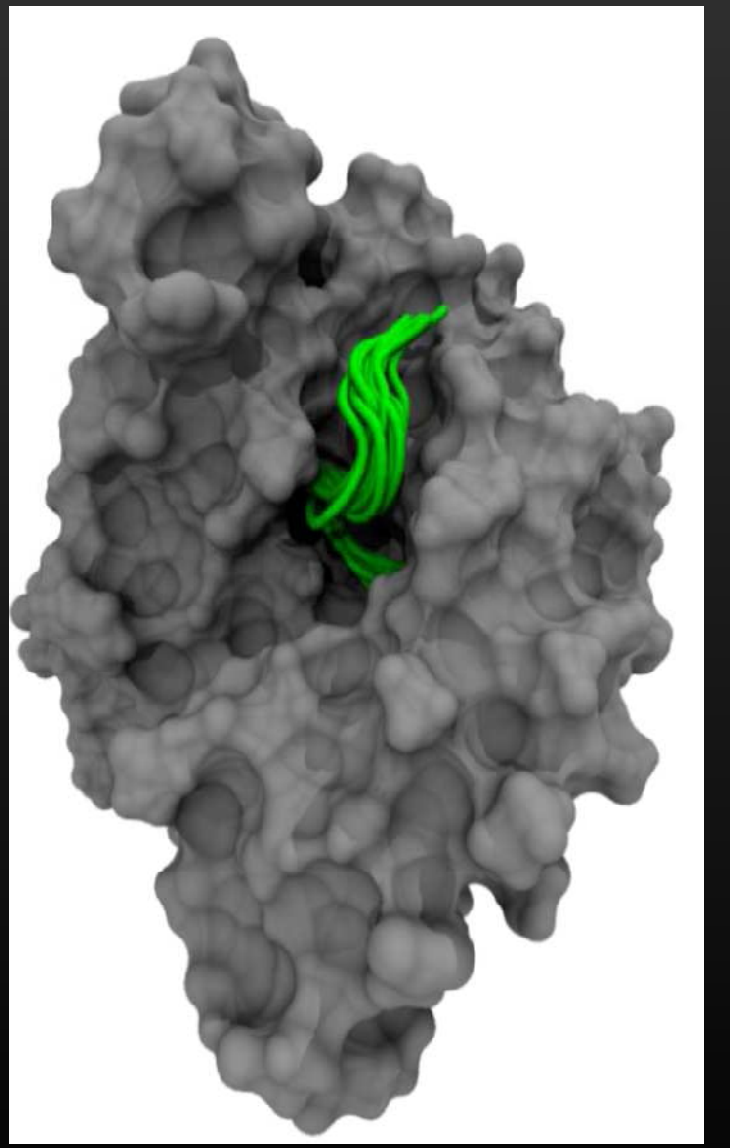
optimizing a peptide with a high binding affinity for the ATP binding pocket of Hsp90



optimizing a peptides with a high binding affinity for the ATP binding pocket of Hsp90







Ala-Trp-Arg-Trp-Ala-Trp-Gly-Gln:
Binding affinity of ~-19 kcal/mol

Ongoing:

1) Impose solubility and low aggregation propensity

2) Impose biocompatibility

3) SELECTIVITY

4) Designing binders for proteins

5) New and more reliable scoring functions

Classes of scoring function to estimate the strength of the binding

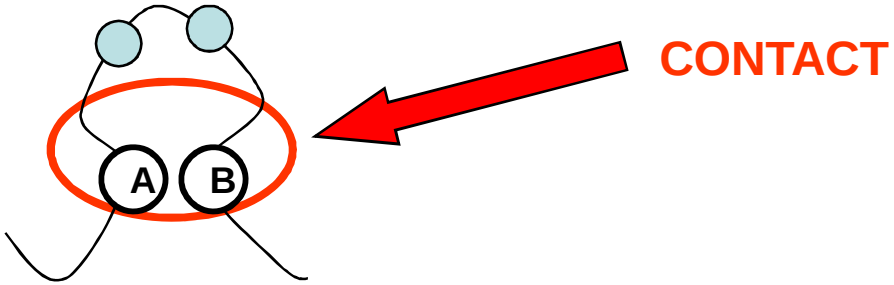
- Broadly speaking, scoring functions can be divided into the following classes:

  - Physics-based energy functions
    - Derived from a fundamental analysis of the forces between the particles. Example: molecular mechanics forcefields (Amber, Charmm, etc)
    - Greater computational cost
  - Knowledge-based potentials
    - Derived by a statistical analysis of known structures (Rosetta, QMEAN)
    - They are more robust and easier to compute

Knowledge-based potentials

Parameters estimated from probabilities observed in a database of experimentally determined proteins

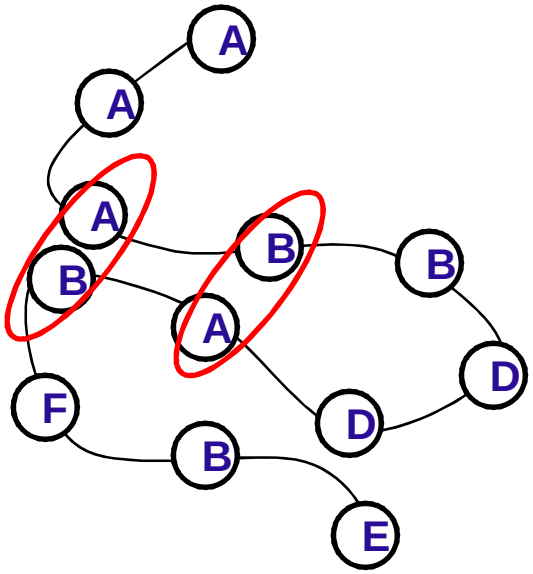
SIMPLEST EXAMPLE: PRESENCE OF A CONTACT BETWEEN TWO AMINOACIDS



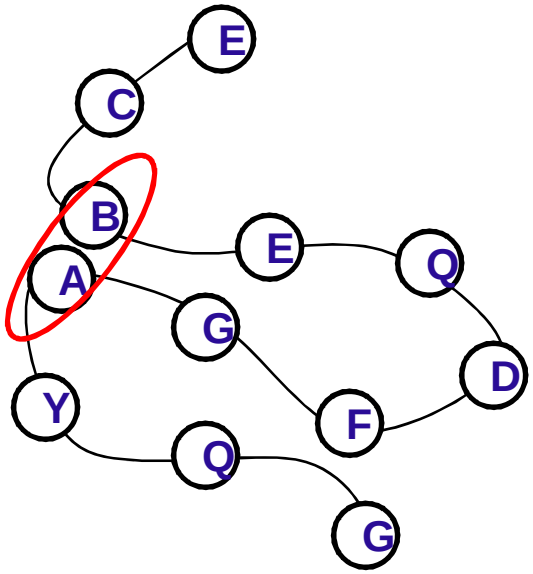
IF (A) AND (B) LIKE EACH OTHER THEY ARE OBSERVED VERY OFTEN IN CONTACT IN THE DATABASE → FAVOURABLE INTERACTION ENERGY!

What does “often” mean?

PROTEIN 1



PROTEIN 2



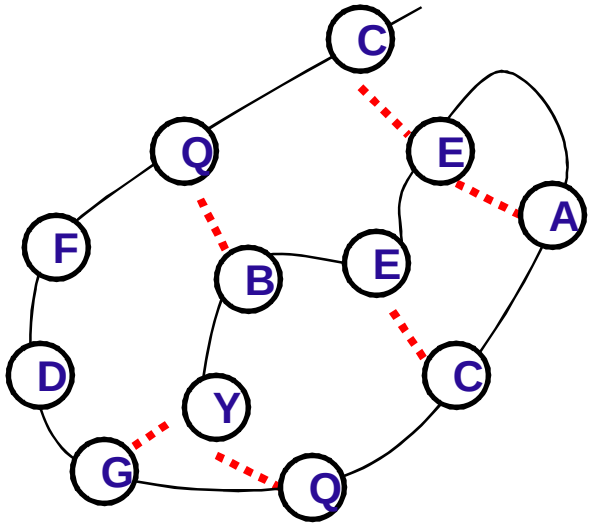
N_{AB}: number of contacts between A and B

N_A and N_B: number of A and number of B

$$E_{AB} = -\sum_{i \in \text{structures}} \log\left(\frac{N_{AB}^{i}}{N_{A}^{i} N_{B}^{i}}\right)$$

Scoring a structure

$$\mathrm{SCORE} = \sum_{i,j \in A,B,C,....} N_{ij} E_{ij}$$



$$N_{AE} = 1$$

$$N_{QB} = 1$$

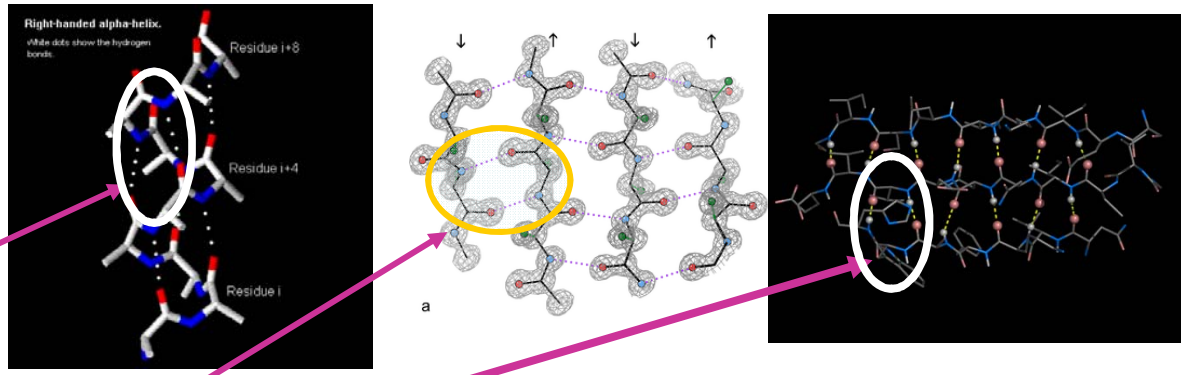
$$N_{CE} = 2$$

$$N_{YQ} = 1$$

$$N_{YG} = 1$$

$$\mathrm{SCORE} = E_{AE} + E_{QB} + 2E_{CE} + E_{YQ} + E_{YG}$$

What do we mean with “contacts”



A given residue pair can

- Form a α-bridge
- Form a anti-parallel β-bridge
- Form a parallel β-bridge
- Form a side-chain side-chain contact (heavy atoms within 4.5 Å)
- Do not interact (include residues non crystallized)

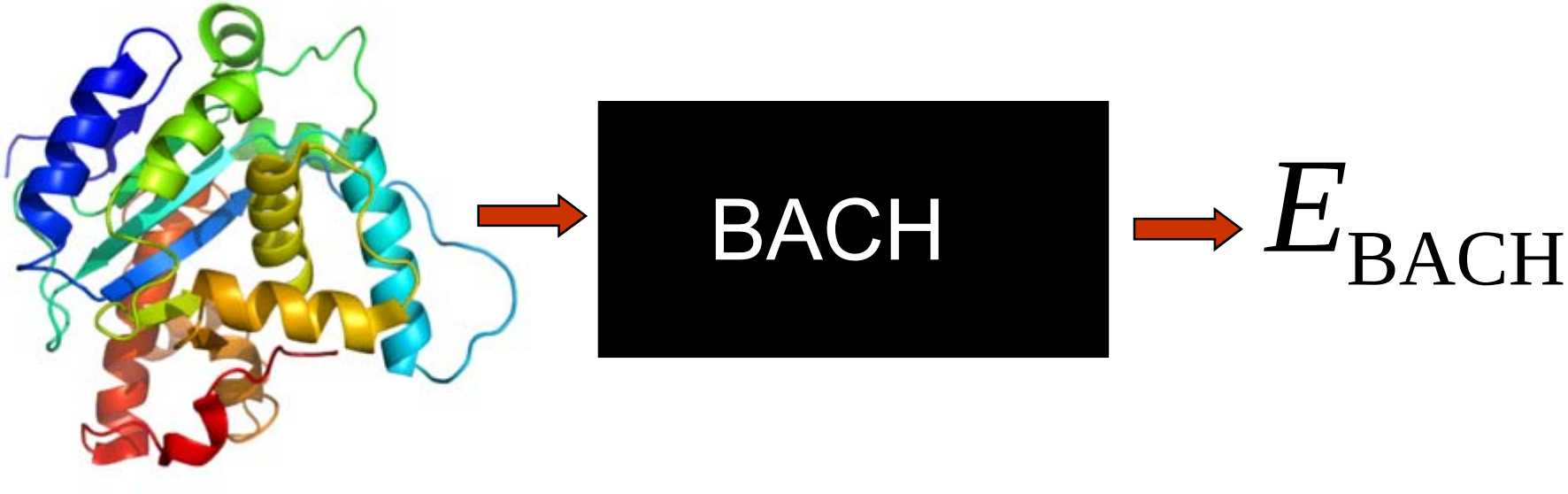
5 mutually exclusive classes → *x=1,2,3,4,5*

BACH ENERGY FUNCTION

5*(20)*(20+1)/2 parameters 40 parameters =1091

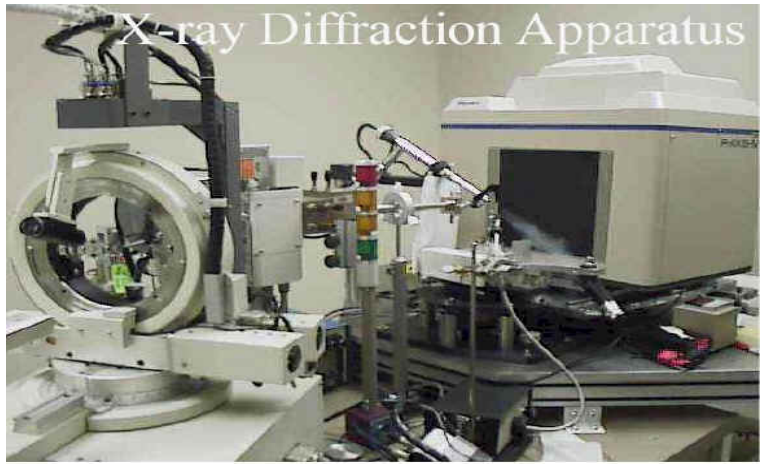
$$E_{\mathrm{BACH}} = E_{\mathrm{contact}} + p * E_{\mathrm{solv}}$$

Exposed/Buried

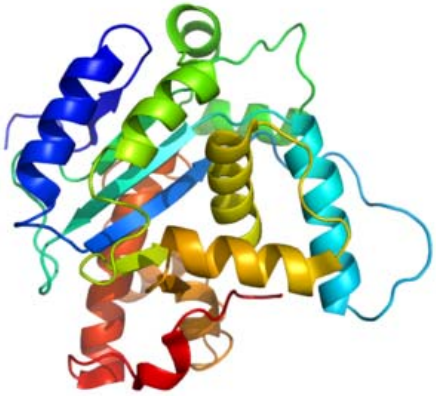


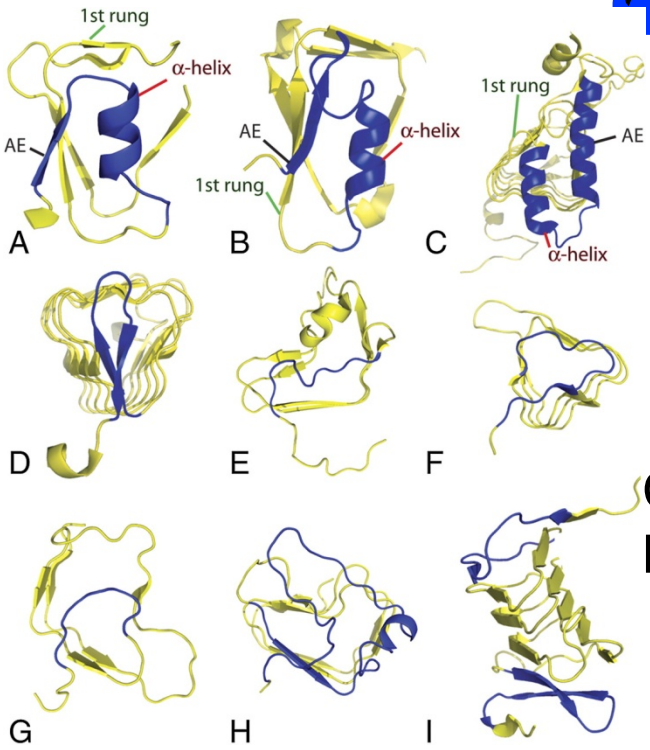
Critical Assessment of *Protein Structure Prediction (CASP)*

PROTEIN SEQUENCE

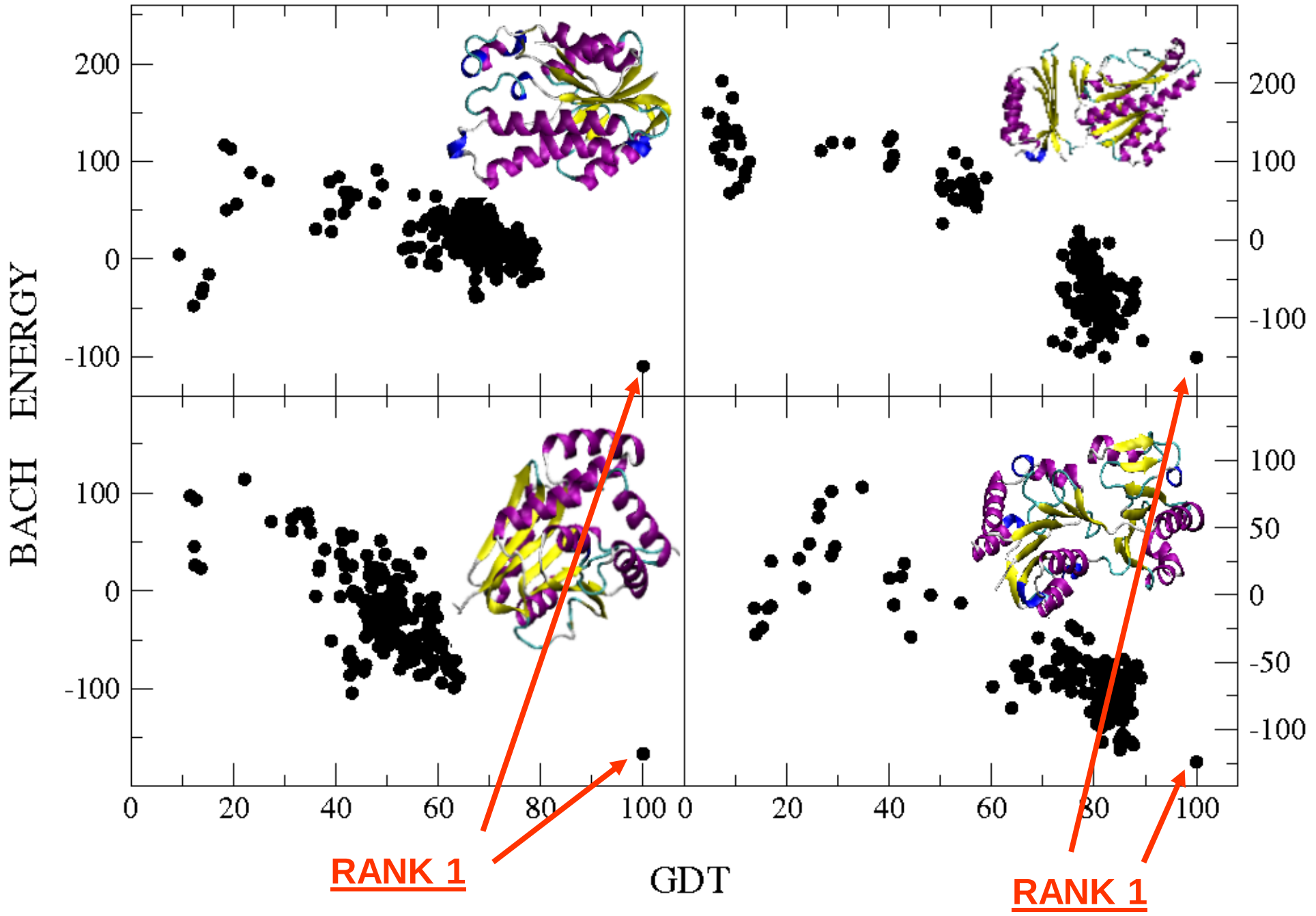


EXPERIMENTAL TARGET





COMPUTATIONAL MODELS


BACH ENERGY
GDT
RANK 1
RANK 1
200
100
0
-100
200
100
0
-100
100
0
-100
100
50
0
-50
-100
0
20
40
60
80
100
0
20
40
60
80
100

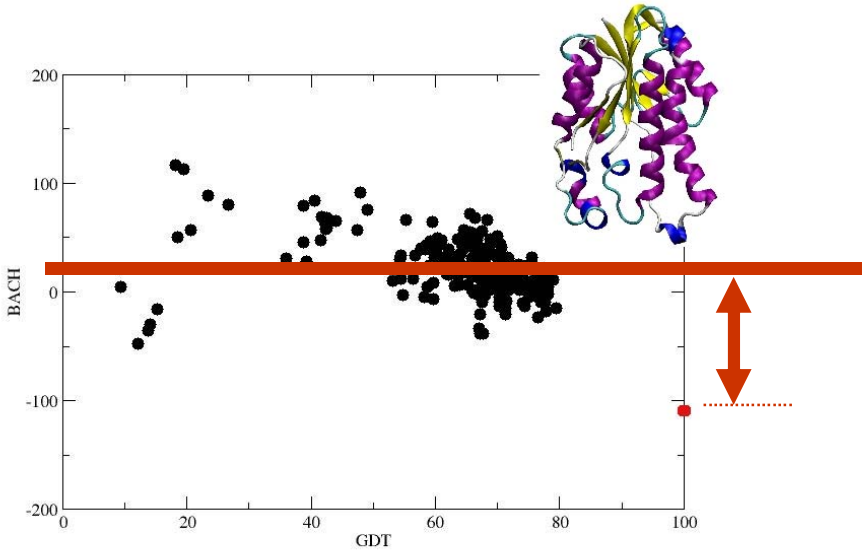
CASP8 and CASP9 TARGETS
32 SETS

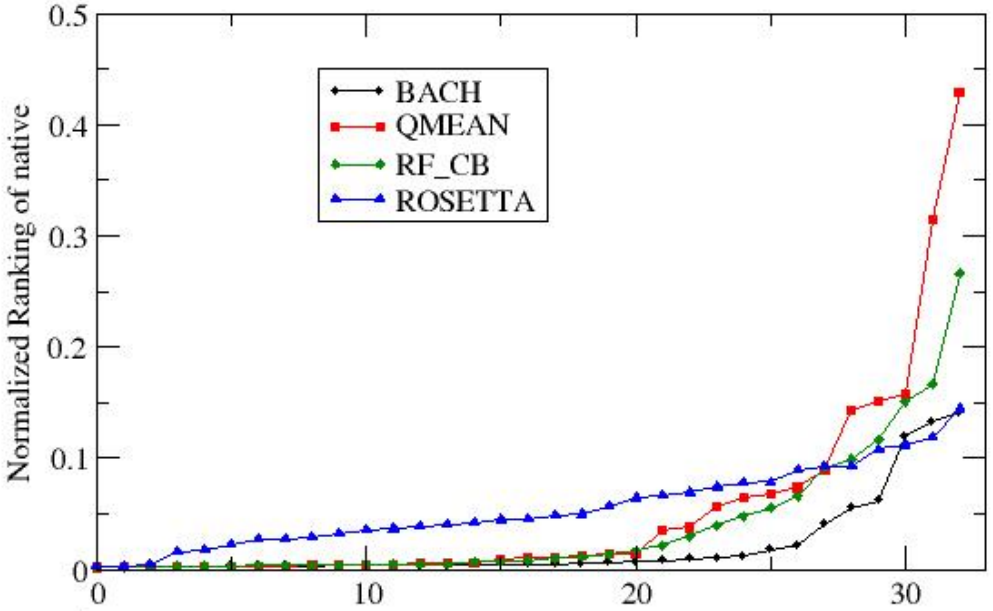
Normalized Ranking:

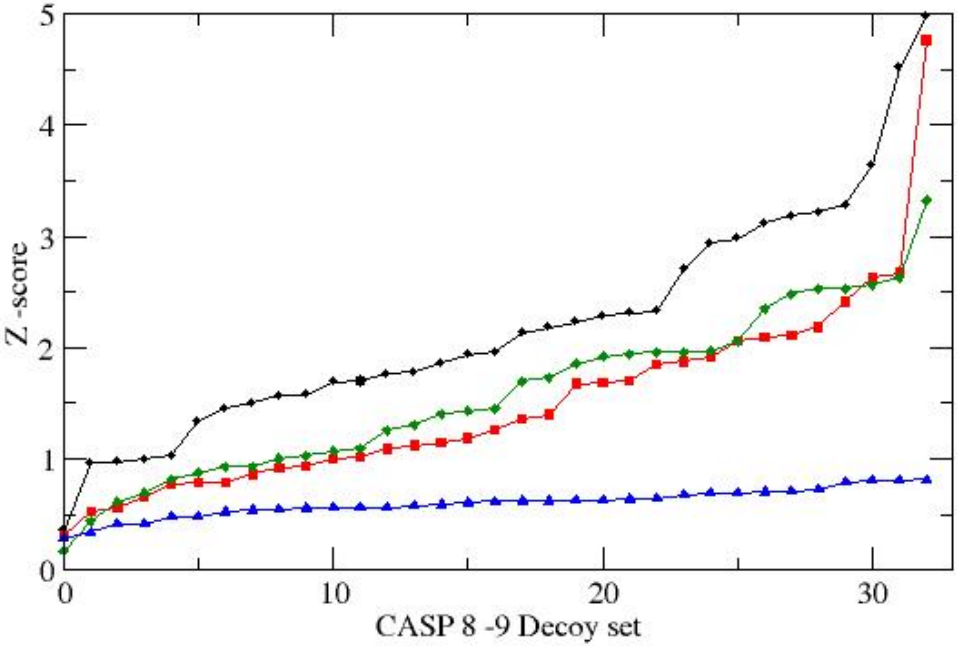
$$\frac{RANK}{Number\ of\ proteins} \rightarrow 0$$

$$\rightarrow 1$$

Z- score:

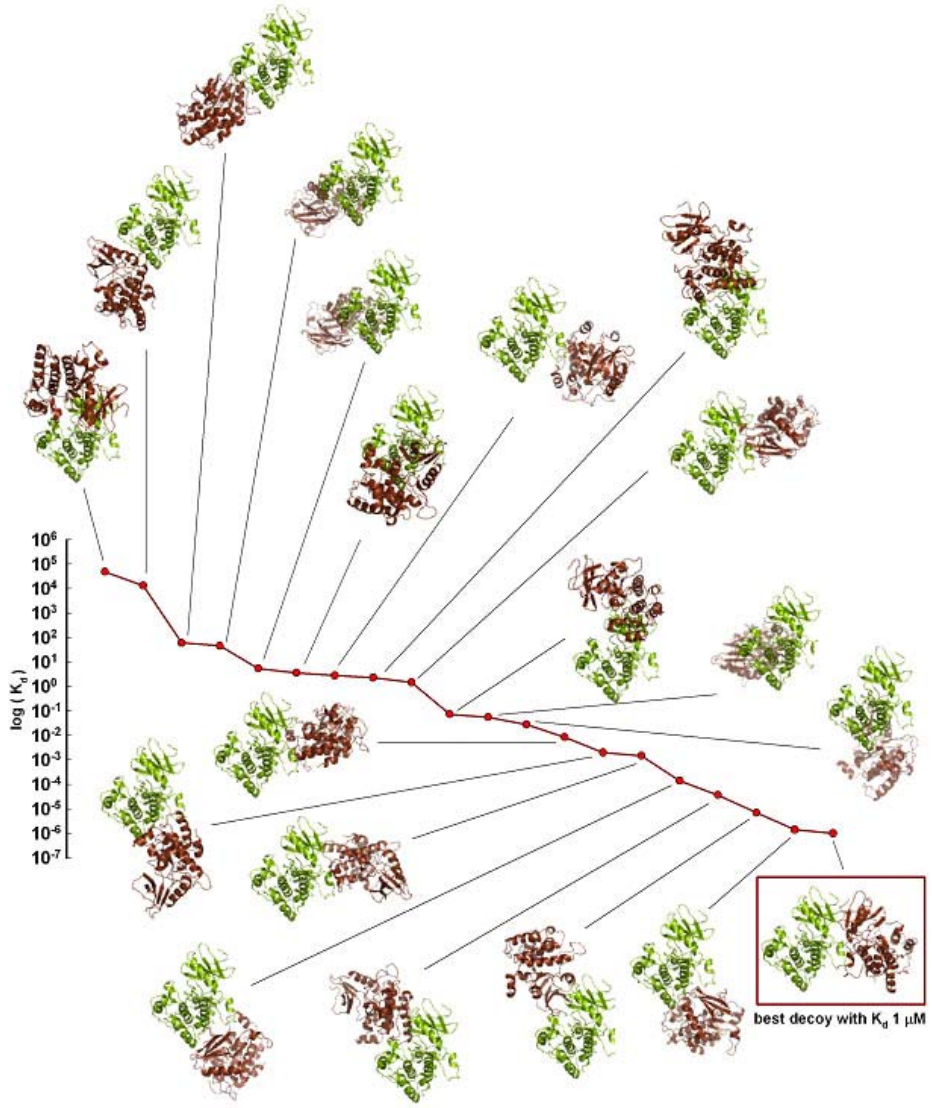


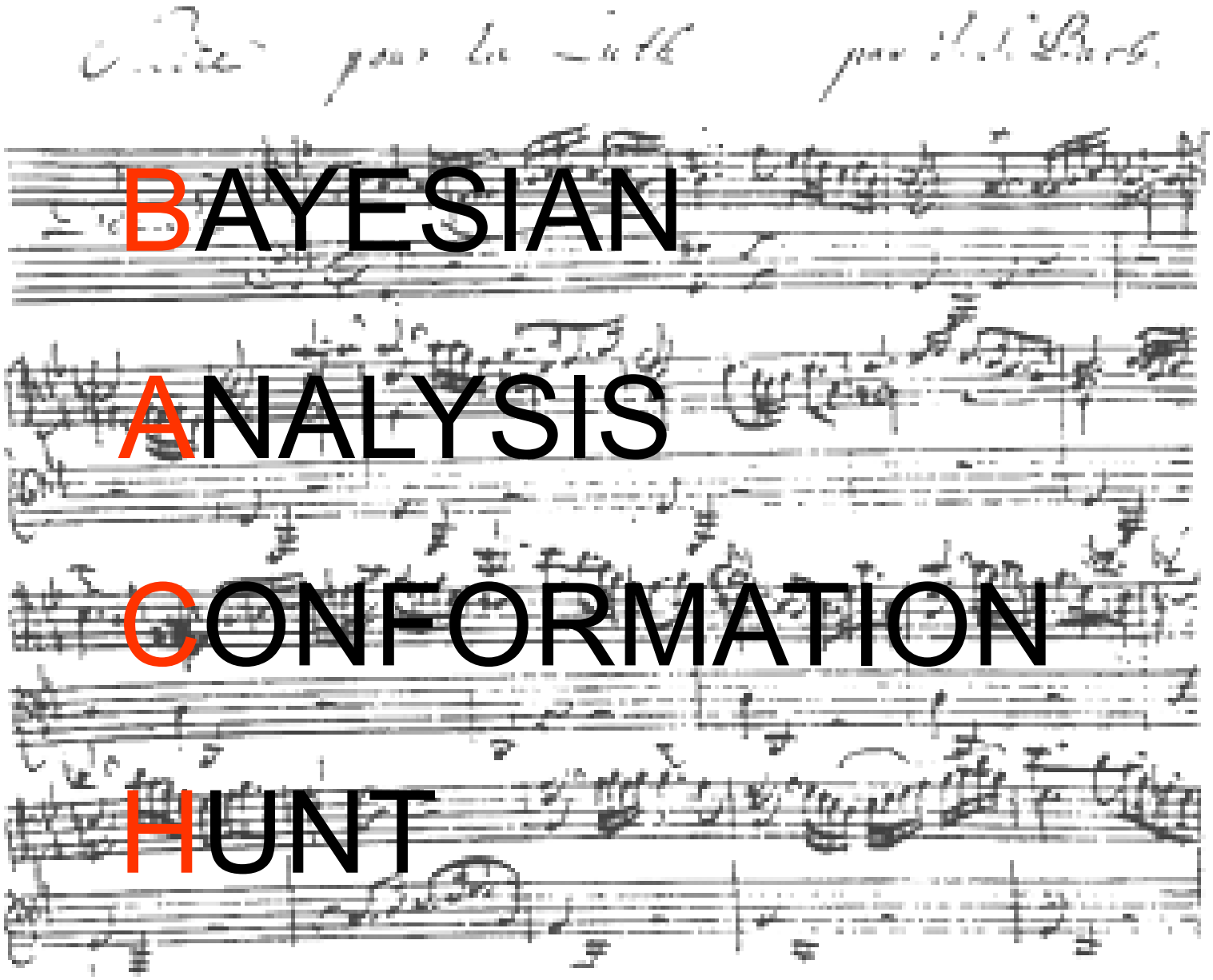




PROTEIN-PROTEIN COMPLEXES

Edoardo Sarti




BAYESIAN
ANALYSIS
CONFORMATION
HUNT

- BACH outperforms state of the art scoring functions in recognizing protein native states
- It is simple, robust and based on few parameters: easy to export
- Probably it will have to be adjusted for describing interactions between a protein and a peptide with variable primary sequence

Result:

- We developed a method that allows designing short peptides capable of binding to drugs or proteins.
- First experimental validations
- Critical points: selectivity and reliability of the scoring function
- We want to develop a knowledge-based potential specific for design.
- We developed BACH, a function with excellent capability of discriminating the folded state.

Acknowledgements

Rolando Hong

Emmanuel Njumbe

Federico Berti

Pilar Cossio, Antonio Trovato & Flavio Seno

Giacinto Scoles