Designing highly specific probes with tunable affinity

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

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

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)
  \[ \Delta f \equiv -k_B T \ln \left( \frac{Q_b}{Q_f} \right) \]

Protein-ligand binding OFF LATTICE

REAL SYSTEM:
protein docking

TEST SYSTEM:
protein-like surface
Protein-ligand binding OFF LATTICE

- Surface modeling

Stretched protein

Dummy atoms

Cα carbons
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

GROMACS
- amber99sb
- TIP3P water
- $T = 298 \text{ K} \; ; \; P = 1 \text{ atm}$
Full atomistic 1QYP

GROMACS

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

Reference Structure for RMSD from RCSB

Potential energy [kJ/mol]

Native seq.

Designed seq.

RMSD [Å]

Native

Designed
Full atomistic 2L09

GROMACS

- amber99sb
- TIP3P water
- $T = 298 \text{ K} ; P = 1 \text{ atm}$

Reference Structure for RMSD from RCSB

**native seq.**
**designed seq.**
**random seq.**

**Potential energy [kJ/mol]**

**Native**

**Designed**

**Random**
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
  o Dr. Luca Tubiana

• Theory and simulations of designable modular bionic proteins
  o Dr. Valentino Bianco
  o M.Sc. Chiara Cardelli
Thank you for your attention
Designing highly specific probes with tunable affinity

Francesca Nerattini

Computational Physics Group, University of Vienna
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
  o Active site
  o 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

Stretched protein

Dummy atoms

Cα carbons
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 \((R, \theta_1, \theta_2)\)

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

\[ E^i_{\text{Sol}} = \begin{cases} \varepsilon^i_{\text{Sol}} [\Omega - \Omega^i] & \Omega^i \leq \Omega \\ 0 & \Omega^i > \Omega \end{cases} \]

\[ \Omega^i = \sum_j \left( 1 - \frac{1}{1 + e^{2.5(r_{\text{max}} - r_{ij})}} \right) \]

\( \Omega = \) threshold for the number of contacts above which the amino acid is considered fully buried

\( \varepsilon^i_{\text{Sol}} = \) (rescaled) Dolittle hydrophobicity index

\( > 0 \) hydrophobic aa

\( < 0 \) hydrophilic aa
Model

Mean field theory Random Energy Model

![Graph depicting the S(E) function with points E_G and E_N.]
• Automated Bio Marker (ABM) - In silico automated tumor targeting
  o Selective tumor cell targeting using low-affinity, multivalent interactions

  o CD47-signal regulatory protein alpha

F. J. Martinez-Veracoechea, and D. Frenkel PNAS 2011;108:10963-10968
Full atomistic 2L09

GROMACS

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

Native

Random

Designed

Designed Repeat
Full atomistic 2L09

GROMACS

- amber99sb
- TIP3P water
- $T = 358$ K ; $P = 1$ atm
Full atomistic 2L09

GROMACS

- amber99sb
- TIP3P water
- $T = 358 \text{ K} ; P = 1 \text{ atm}$
Full atomistic 2L09

![Graph showing potential energy over time for different conditions]
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

The CATERPILLAR model

COARSE GRAINED: CATERPILLAR\(^{[1-2]}\)