Modeling Protein- RNA complexes

AlgoSB_2019

Isaure Chauvot de Beauchêne

Protein – RNA interactions

Many RNA functions require interactions with proteins:

- Genes regulation
- Signal transduction
- Protein synthesis
- Replication of RNA viruses

A 3D representation of the interface allows to:

- Understand protein/RNA function
- Predict mutations effect
- Compute binding energy
- Design drugs to target the interface
- Design a protein to taget an RNA
- Design an RNA to target a protein („*aptamer*“)

Protein – RNA interfaces : Experimental data
Main features

The docking problem: Sampling
Evaluation of docking models
Scoring

Flexibility: Flexible docking
Fragment-based docking

Data-driven docking: Contact/interface -driven
Fitting in 3D shape

Other docking paradigms

Modeling pipeline

Experimental structures

Protein-RNA Interactions in an Icosahedral Virus at 3.0 Å Resolution

ZHONGGUO CHEN, CYNTHIA STAUFFACHER, YUNGE LI, TIM SCHMIDT, WU BOMU, GREG KAMER, MICHAEL SHANKS, GEORGE LOMONOSSOFF, JOHN E. JOHNSON*

Science, Jul 1989

Experimental structures

X-ray crystallography

Very high resolution (~ 1 - 3 Å)
Most heavy atoms visible

One static view, no dynamics
Most flexible parts not visible
Artefacts from crystallisation
(incl. false interfaces !)

1 Å = 10 nm

Experimental structures

Nuclear magnetic resonance

Multiple conformations
=> info on dynamics

Limited to small systems:
< 50 kDa (~150 nucleotides)

weak oscillating magnetic field

↓

Nucleus in strong
static magnetic field

↓

Electromagnetic signal

Frequency = f(atomic environment)

Protein – RNA interfaces : Experimental data

Main features

The docking problem: Sampling

Evaluation of docking models

Scoring

Flexibility: Flexible docking

Fragment-based docking

Data-driven docking: Contact/interface -driven

Fitting in 3D shape

Other docking paradigms

Modeling pipeline

Protein – RNA interfaces

Main characteristics of macro-molecular interfaces

- Shape complementarity
  - Large
  - Compact

  } Desolvation effect
  } Favourable contacts

Protein – RNA interfaces

Main characteristics of macro-molecular interfaces

- Shape complementarity
  - Large
  - Compact
  - → Desolvation effect
  - → Favourable contacts
- Electrostatic complementarity
  - Ionic bridges
  - hydrogen bonds
  - Water bridges

Protein – RNA interfaces

Main characteristics of macro-molecular interfaces

- Shape complementarity
  - Large
  - Compact
  - → Desolvation effect
  - → Favourable contacts
- Electrostatic complementarity
  - Ionic bridges
  - hydrogen bonds
  - Water bridges

Examples of protein – RNA interfaces

- sequence (non-)specificity run: *pymol toxin_antitoxin.pse*
- electrostatic-driven (*Phosphate groups*)
- stacking *pymol RRM_polyU.pse*

Protein – RNA interfaces

Dec. 2018:

137,000 protein structures

1,400 RNA structures

2,400 protein-RNA structures ← *Much lower than what has been experimentally proved to exist in vivo*

The structure of a complex is more difficult to solve experimentally than a single protein or RNA

=> We often need to model the complex from the structures of the single molecules.

obligatory vs non-obligatory complex

Experimental data on protein – RNA interfaces

The docking problem: Sampling
Evaluation of docking models
Scoring

Flexibility: Flexible docking
Fragment-based docking

Data-driven docking: Contact/interface -driven
Fitting in 3D shape

Other docking paradigms

Modeling pipeline

The Docking Problem

Predicting *in silico* the most probable 3D structures of a complex at the thermodynamic equlibrium, using the 3D structures of its individual components.

! It is assumed that the 2 molecules do bind !

The thermodynamically most stable conformation is called **native conformation**. It is assumed to be the one observed experimentally (and carrying the biological function).

The docking is mainly based on geometrical and chemical properties.

The rigid approximation:

- The 2 molecules are kept (mostly) rigid
- One molecule being fixed, the other has 6 degrees of freedom
- We neglect intra-molecular energies
- We minimize the intermolecular interaction energy

The Docking Problem

Sampling → Model 1 | Model 2 | Model 3

Scoring → E = - 40 kcal/mol | E = - 35 kcal/mol | E = - 3 kcal/mol

The rigid approximation:

- The 2 molecules are kept (mostly) rigid
- One molecule being fixed, the other has 6 degrees of freedom
- We neglect intra-molecular energies
- We minimize the intermolecular interaction energy

Sampling of docking models

ex. ATTRACT : Gradient-based minimisation

Sampling of docking models

ex. ATTRACT : Gradient-based minimisation

Sampling of docking models

Place the ligand at random position

Compute the inter-molecular energy

Assumption: pairwise additive interactions

$$V_{(lig\,,\,rec)} = \sum_{(A \in Rec\,;\,B \in Lig)} V_{(A\,,\,B)}$$

Van der Waals *coulomb* *

$$V_{(A,B)} = \left(\frac{R_{AB}}{r_{ab}}\right)^{12} - \left(\frac{R_{AB}}{r_{ab}}\right)^{6} + \frac{q_a q_b}{\varepsilon r_{ab}}$$

q = electrostatic charge
R = mean atomic radius
r = inter-atomic distance
ε = dielectric constant

Compute the derivative (gradient)

$$\frac{\delta V}{dx\ dy\ dz\ d\theta\ d\varphi\ d\omega}$$

Displace and rotate the ligand in that direction from a small distance δθ

* Van der Waals in r ^ -12 = short range effect
coulomb in r ^ -1 = long range (esp. RNA negative charges)

Sampling of docking models

How can we increase the probability to end-up in the global minimum?

Sampling of docking models

How can we increase the probability to end-up in the global minimum?

- Allow some increases of energy : Monte-Carlo procedure
  ☹ Computationally demanding, Stochastic results

pos2 = pos1 + random(δθ)

Accept move with proba P

$$P = e^{-\frac{E2 - E1}{k_b T}}$$

Sampling of docking models

How can we increase the probability to end-up in the global minimum?

- Allow some increases of energy : Monte-Carlo procedure
  ☹ Computationally demanding, Stochastic results
- Start from many multiple starting positions

pos2 = pos1 + random(δθ)

Accept move with proba P

$$P = e^{-\frac{E2 - E1}{k_b T}}$$

Sampling of docking models

How can we increase the probability to end-up in the global minimum?

- Allow some increases of energy : Monte-Carlo procedure
  ☹ Computationally demanding, Stochastic results
- Start from many multiple starting positions

pos2 = pos1 + random(δθ)

Accept move with proba P

$$P = e^{-\frac{E2 - E1}{k_b T}}$$

Sampling of docking models

How can we increase the probability to end-up in the global minimum?

- Allow some increases of energy : Monte-Carlo procedure
  ☹ Computationally demanding, Stochastic results
- Start from many multiple starting positions
- Reduce the number of local minima
  by smoothing the energy landscape

pos2 = pos1 + random(δθ)

Accept move with proba P

$$P = e^{-\frac{E2 - E1}{k_b T}}$$

Sampling of docking models

Coarse-grained model : replace few atoms by one pseudo-atom

→ Faster pairwise computations: $O(N^2)$ => 36x faster

→ Smoother energy landscape => less local minima

→ Account for some inaccuracies in the structures

Sampling of docking models

Parametrization

2 parameters x 31 amino-acid beads x 17 RNA beads

attraction $$U_{ij}^{\mathrm{attr}}(r) = \epsilon_{ij}\left(\frac{\sigma_{ij}}{r^8} - \frac{\sigma_{ij}}{r^6}\right)$$

repulsion $$U_{ij}^{\mathrm{rep}}(r) = \begin{cases} U_{ij}^{\mathrm{attr}}(r) + 2U_{ij}^{m} & \text{for } r \le r_{ij}^{m} \\ -U_{ij}^{\mathrm{attr}}(r) & \text{for } r > r_{ij}^{m} \end{cases}$$

Adjust parameters based on known structure, so as to give better score to models close to known structures

Experimental data on protein – RNA interfaces

The docking problem: Sampling
Evaluation of docking models
Scoring

Flexibility: Flexible docking
Fragment-based docking

Data-driven docking: Contact/interface -driven
Fitting in 3D shape

Other docking paradigms

Modeling pipeline

Evaluation of docking models

We put ourselves in the **artificial case** were the real structure of the complex is known

To assess the quality of a solution, we compute the **Root mean squared deviation** (RMSD): Average deviation of the atoms in the **docking solution** vs in the **experimental structure**.

$$RMSD = \sqrt{\frac{\sum_{i=1}^{N}\left(\left(x_i^a - x_i^b\right)^2 + \left(y_i^a - y_i^b\right)^2 + \left(z_i^a - z_i^b\right)^2\right)}{N}}$$

Ligand-RMSD (L-RMSD)
RMSD of docked vs bound ligand, after fitting the model on the bound receptor

Evaluation of docking models

We put ourselves in the **artificial case** were the real structure of the complex is known

To assess the quality of a solution, we compute the **Root mean squared deviation** (RMSD): Average deviation of the atoms in the **docking solution** vs in the **experimental structure**.

$$RMSD = \sqrt{\frac{\sum_{i=1}^{N}\left(\left(x_i^a - x_i^b\right)^2 + \left(y_i^a - y_i^b\right)^2 + \left(z_i^a - z_i^b\right)^2\right)}{N}}$$

Ligand-RMSD (L-RMSD)

RMSD of docked vs bound ligand, after fitting the model on the bound receptor

interface-RMSD (i-RMSD)

RMSD of only the interface after fitting on the bound interface the equivalent atoms of the model.

Evaluation of docking models

We put ourselves in the **artificial case** were the real structure of the complex is known

To assess the quality of a solution, we compute the **Root mean squared deviation** (RMSD): Average deviation of the atoms in the **docking solution** vs in the **experimental structure**.

$$RMSD = \sqrt{\frac{\sum_{i=1}^{N}\left(\left(x_i^a - x_i^b\right)^2 + \left(y_i^a - y_i^b\right)^2 + \left(z_i^a - z_i^b\right)^2\right)}{N}}$$

Ligand-RMSD (L-RMSD)

RMSD of docked vs bound ligand, after fitting the model on the bound receptor

interface-RMSD (i-RMSD)

RMSD of only the interface after fitting on the bound interface the equivalent atoms of the model.

Fraction of native contacts (Fnat)

A contact is a pair of RNA-protein atoms at <5 A from each other. Fnat is the % of contacts recovered by the model.

Evaluation of docking models

What should we aim for?

iRMSD: 0.5 Å
***

iRMSD: 1.2 Å
**

iRMSD: 3.5 Å
*

Experimental data on protein – RNA interfaces

The docking problem: Sampling
Evaluation of docking models
Scoring

Flexibility: Flexible docking
Fragment-based docking

Data-driven docking: Contact/interface -driven
Fitting in 3D shape

Other docking paradigms

Modeling pipeline

Scoring

Sampling : Producing models as close as possible from the native structure
Scoring : Discriminating good („near-native“) solutions from wrong solutions („decoys“)

Solutions are scored and **ranked** from the smallest to highest score.

Scores are only an approximation of the energy => ranking is approximative !

Scoring

RMSD	Rank
0.22	1
0.53	2
0.21	3
1.10	4
...	

Test on an **artificial case**

Sampling : Producing models as close as possible from the native structure
Scoring : Discriminating good („near-native“) solutions from wrong solutions („decoys“)

Solutions are scored and **ranked** from the smallest to highest score.

Scores are only an approximation of the energy => ranking is approximative !

Scoring

Test on another **artificial case**

The first good model is typically in the ~10-100 top-ranked model

=> You have to consider as many models to not throw away all good ones!

Scoring

How to check if any “predicted good structure” is really good ?

reference structure unknown
=> no RMSD available

real case

The first good model is typically in the ~10-100 top-ranked model

=> You have to consider as many models to not throw away all good ones!

=> What about a **real case**, where we don't have access to the RMSD ?
(there is no reference structure)

Scoring

How to check if any “predicted good structure” is really good ?

> Visualization: *large interface*
good shape complementarity (no holes at interface)
clustering at same area

real case

Wrong model

Scoring

How to check if any "predicted good structure" is really good ?

> Visualization: *large interface*
good shape complementarity (no holes at interface)
clustering at same area

> Correspondence to biological function

G protein-coupled receptor kinase 2 (GRK2). However, during chronic heart failure GRK2 is upregulated and believed to contribute to disease progression. We have determined crystallographic structures of GRK2 bound to an RNA aptamer that potently and selectively inhibits kinase activity. Key to the mechanism of inhibition is the positioning of an adenine nucleotide into the ATP-binding pocket and interactions with the basic αF-αG loop region of the GRK2 kinase domain.

Tesmer, Lennarz, Mayer, Tesmer (2012) Structure 20(8):1300-9.

pymol : GPCR.pse

Scoring

How to check if any “predicted good structure” is really good ?

> Visualization: *large interface*
good shape complementarity (no holes at interface)
clustering at same area

> Correspondence to biological function

> Mutagenesis on amino-acids at the interface

→ should change the binding affinity

→ should not disrupt the individual structures !!

Experimental data on protein – RNA interfaces

The docking problem: Sampling
Evaluation of docking models
Scoring

Flexibility: Flexible docking
Fragment-based docking

Data-driven docking: Contact/interface -driven
Fitting in 3D shape

Other docking paradigms

Modeling pipeline

Flexibility

The rigid approximation:

- The molecules are kept rigid
- receptor is fixed, ligand has 6 DOF
- We neglect intra-molecular energies
- We minimize the interaction energy

Flexibility

Known protein – RNA interaction

Rigid docking

The rigid approximation:

- The molecules are kept rigid
- receptor is fixed, ligand has 6 DOF
- We neglect intra-molecular energies
- We minimize the interaction energy

Flexibility

Known protein – RNA interaction

Experimental unbound structures

RNA

Protein

Flexible docking

models

E = -40 kcal/mol

E = -35 kcal/mol

Flexible docking

Flexibility

Conformational changes upon binding occurs by (a combination of) 2 ways :

Flexibility

Conformational changes upon binding occurs by (a combination of) 2 ways :

Flexibility

Proteins

Side chain (protein surface)
Backbone in loop regions
Secondary structures
Domains

amplitude

RNA

Base flipping
Backbone in loop regions
Helices bend / twist
Unstructured regions (single-strand)

Flexibility

Proteins

Side chain (protein surface)
Backbone in loop regions
Secondary structures
Domains

amplitude

RNA

Base flipping
Backbone in loop regions
Helices bend / twist
Unstructured regions (single-strand)

Flexibility

Proteins

Side chain (protein surface)
Backbone in loop regions
Secondary structures
Domains

amplitude

RNA

Base flipping
Backbone in loop regions
Helices bend / twist
Unstructured regions (single-strand)

unbound
bound

Flexibility

Proteins

Side chain (protein surface)
Backbone in loop regions
Secondary structures
Domains

amplitude

RNA

Base flipping
Backbone in loop regions
Helices bend / twist
Unstructured regions (single-strand)

13 Å RMSD

Flexibility

Proteins

Side chain (protein surface)
Backbone in loop regions
Secondary structures
Domains

amplitude

RNA

Base flipping
Backbone in loop regions
Helices bend / twist
Unstructured regions (single-strand)

Flexibility

Proteins

Side chain (protein surface)
Backbone in loop regions
Secondary structures
Domains

amplitude

RNA

Base flipping
Backbone in loop regions
Helices bend / twist
Unstructured regions (single-strand)

Flexibility

Proteins

Side chain (protein surface)
Backbone in loop regions
Secondary structures
Domains

amplitude

RNA

Base flipping
Backbone in loop regions
Helices bend / twist
Unstructured regions (single-stranded)

Flexibility

Flexibility

Proteins

Side chain (protein surface)
Backbone in loop regions
Secondary structures
Domains

amplitude

RNA

Base flipping
Backbone in loop regions
Helices bend / twist
Unstructured regions (single-stranded)

Undefined in solution

Flexibility

The perfect superpositions of unbound structures on the bound complex gives the upper limit of model quality that can be obtained by rigid body modeling.

If the perfect fit brings atomic clashes, it will not be found by docking.

Best fit : 3.2 Å

Experimental data on protein – RNA interfaces

The docking problem: Sampling
Evaluation of docking models
Scoring

Flexibility: Flexible docking
Fragment-based docking

Data-driven docking: Contact/interface -driven
Fitting in 3D shape

Other docking paradigms

Modeling pipeline

Flexibility

Known protein – RNA interaction

Experimental unbound structures

RNA

Protein

Flexible docking

models

E = -40 kcal/mol

E = -35 kcal/mol

Flexible docking

Harmonic modes

Energetically favorable directions of deformation, orthogonal to each-other.

- Computed based on the structure of each molecule and its inter-atomic forces.
- Added as additional DOF along the minimization.

Hopefully one or few modes will bring the unbound RNA closer to the bound form (but it is not guaranteed !).

ATTRACT: harmonic modes

Conformational ensemble

Ensemble of unbound conformations docked in parallel.

Can be obtained from Molecular Dynamics, NMR …

Increase the probability to be close enough to the bound form (but it is not guaranteed !),

Flexible docking

Refinement

After docking, use MD with all-atom force-field to refine models

=> Only small ajustements. Needs to have already a quite correct solution

!! time-consuming

Experimental data on protein – RNA interfaces

The docking problem: Sampling
Evaluation of docking models
Scoring

Flexibility: Flexible docking
Fragment-based docking

Data-driven docking: Contact/interface -driven
Fitting in 3D shape

Other docking paradigms

Modeling pipeline

Fragment – based docking

Flexible docking

Known RNA sequence

~~Flexible docking~~

Fragment – based docking

- combinatorial explosion
- Protein-induced conformations

Approximation : discrete **local** conformations
=> structural fragment library

Known RNA sequence

ssRNA: Disordered

?

| {conformations} | ~ ∞

Flexible docking

Fragment – based docking

RNA sequence

A U G G U

A U G

U G G

G G U

Fragment library

~ 3000 per sequence

Protein structure

Sampling

Combinatorial assembly

Chains of poses

Search for paths (RNA chains) with low total energy

Fragment – based docking

~ $10^{5\text{-}6}$ poses / fragment

~ 5-8 fragmt / RNA

=> $10^{25\text{-}48}$ chains

Sampling

Combinatorial assembly

Chains of poses

Search for paths (RNA chains) with low total energy

Main difficulties :

- All frag must be correctly sampled
- Scoring problem => lot of false positive

Poses compatibility

Poses compatibility

Chain sampling with mean-rank constrain

- Starting backward, consider the best-case meanrank at each pose

Ex : ($\prod rank_i$) threshold = 5000

Chain sampling with mean-rank constrain

- Starting backward, consider the best-case meanrank at each pose
- For next pose, eliminate connections that cannot lead to correct meanrank

Ex : ($\prod rank_i$) threshold = 5000

Chain sampling with mean-rank constrain

- Starting backward, consider the best-case meanrank at each pose
- For next pose, eliminate connections that cannot lead to correct meanrank

Ex : ($\prod$rank$_i$) threshold = 5000

Chain sampling with mean-rank constrain

- Starting backward, consider the best-case meanrank at each pose
- For next pose, eliminate connections that cannot lead to correct meanrank

Ex : ($\prod rank_i$) threshold = 5000

Chain sampling with mean-rank constrain

- Starting backward, consider the best-case meanrank at each pose
- For next pose, eliminate connections that cannot lead to correct meanrank
- Eliminate non-connected poses if needed

Ex : ($\prod rank_i$) threshold = 5000

Chain sampling with mean-rank constrain

- Starting backward, consider the best-case meanrank at each pose
- For next pose, eliminate connections that cannot lead to correct meanrank
- Eliminate non-connected poses if needed
- Retain at each new pose the best-case meanrank

Ex : ($\prod rank_i$) threshold = 5000

Chain sampling with mean-rank constrain

- Starting backward, consider the best-case meanrank at each pose
- For next pose, eliminate connections that cannot lead to correct meanrank
- Eliminate non-connected poses if needed
- Retain at each new pose the best-case meanrank

Ex : ($\prod rank_i$) threshold = 5000

Chain sampling with mean-rank constrain

- Starting backward, consider the best-case meanrank at each pose
- For next pose, eliminate connections that cannot lead to correct meanrank
- Eliminate non-connected poses if needed
- Retain at each new pose the best-case meanrank

Ex : ($\prod rank_i$) threshold = 5000

Chain sampling with mean-rank constrain

- Starting backward, consider the best-case meanrank at each pose
- For next pose, eliminate connections that cannot lead to correct meanrank
- Eliminate non-connected poses if needed
- Retain at each new pose the best-case meanrank

Ex : ($\prod rank_i$) threshold = 5000

Chain sampling with mean-rank constrain

- Starting backward, consider the best-case meanrank at each pose
- For next pose, eliminate connections that cannot lead to correct meanrank
- Eliminate non-connected poses if needed
- Retain at each new pose the best-case meanrank

Ex : ($\prod rank_i$) threshold = 5000

Chain sampling with mean-rank constrain

- Starting backward, consider the best-case meanrank at each pose
- For next pose, eliminate connections that cannot lead to correct meanrank
- Eliminate non-connected poses if needed
- Retain at each new pose the best-case meanrank

Walk the filtered tree w. dead-end elimination

Ex : ($\prod rank_i$) threshold = 5000

Current $\prod rank_i$ = **1**

Best possible tot $\prod rank_i$

Chain sampling with mean-rank constrain

- Starting backward, consider the best-case meanrank at each pose
- For next pose, eliminate connections that cannot lead to correct meanrank
- Eliminate non-connected poses if needed
- Retain at each new pose the best-case meanrank

Walk the filtered tree w. dead-end elimination

Ex : ($\prod rank_i$) threshold = 5000

Current $\prod rank_i$ = **5**

Best possible tot $\prod rank_i$

Chain sampling with mean-rank constrain

- Starting backward, consider the best-case meanrank at each pose
- For next pose, eliminate connections that cannot lead to correct meanrank
- Eliminate non-connected poses if needed
- Retain at each new pose the best-case meanrank

Walk the filtered tree w. dead-end elimination

Ex : ($\prod rank_i$) threshold = 5000

Current $\prod rank_i$ = **10**

Best possible tot $\prod rank_i$

Chain sampling with mean-rank constrain

- Starting backward, consider the best-case meanrank at each pose
- For next pose, eliminate connections that cannot lead to correct meanrank
- Eliminate non-connected poses if needed
- Retain at each new pose the best-case meanrank

Walk the filtered tree w. dead-end elimination

Ex : ($\prod rank_i$) threshold = 5000

Current $\prod rank_i$ = **10**

Best possible tot $\prod rank_i$

Chain sampling with mean-rank constrain

- Starting backward, consider the best-case meanrank at each pose
- For next pose, eliminate connections that cannot lead to correct meanrank
- Eliminate non-connected poses if needed
- Retain at each new pose the best-case meanrank

Walk the filtered tree w. dead-end elimination

Ex : ($\prod rank_i$) threshold = 5000

Current $\prod rank_i$ = **1000**

Best possible tot $\prod rank_i$

chain ok

Chain sampling with mean-rank constrain

- Starting backward, consider the best-case meanrank at each pose
- For next pose, eliminate connections that cannot lead to correct meanrank
- Eliminate non-connected poses if needed
- Retain at each new pose the best-case meanrank

Walk the filtered tree w. dead-end elimination

Ex : ($\prod rank_i$) threshold = 5000

Current $\prod rank_i$ = **250**

Best possible tot $\prod rank_i$

Chain sampling with mean-rank constrain

- Starting backward, consider the best-case meanrank at each pose
- For next pose, eliminate connections that cannot lead to correct meanrank
- Eliminate non-connected poses if needed
- Retain at each new pose the best-case meanrank

Walk the filtered tree w. dead-end elimination

Ex : ($\prod rank_i$) threshold = 5000

Current $\prod rank_i$ = **100**

Best possible tot $\prod rank_i$

Chain sampling with mean-rank constrain

- Starting backward, consider the best-case meanrank at each pose
- For next pose, eliminate connections that cannot lead to correct meanrank
- Eliminate non-connected poses if needed
- Retain at each new pose the best-case meanrank

Walk the filtered tree w. dead-end elimination

Ex : ($\prod rank_i$) threshold = 5000

Current $\prod rank_i$ = **500**

Best possible tot $\prod rank_i$

Chain sampling with mean-rank constrain

- Starting backward, consider the best-case meanrank at each pose
- For next pose, eliminate connections that cannot lead to correct meanrank
- Eliminate non-connected poses if needed
- Retain at each new pose the best-case meanrank

Walk the filtered tree w. dead-end elimination

Ex : ($\prod rank_i$) threshold = 5000

Current $\prod rank_i$ = **500**

Best possible tot $\prod rank_i$

Chain sampling with mean-rank constrain

- Starting backward, consider the best-case meanrank at each pose
- For next pose, eliminate connections that cannot lead to correct meanrank
- Eliminate non-connected poses if needed
- Retain at each new pose the best-case meanrank

Walk the filtered tree w. dead-end elimination

Ex : ($\prod rank_i$) threshold = 5000

Current $\prod rank_i$ = **2000**

Best possible tot $\prod rank_i$

Chain sampling with mean-rank constrain

- Starting backward, consider the best-case meanrank at each pose
- For next pose, eliminate connections that cannot lead to correct meanrank
- Eliminate non-connected poses if needed
- Retain at each new pose the best-case meanrank

Walk the filtered tree w. dead-end elimination

Ex : ($\prod rank_i$) threshold = 5000

Current $\prod rank_i$ = **1000**

Best possible tot $\prod rank_i$

chain ok

Experimental data on protein – RNA interfaces

The docking problem: Sampling
Evaluation of docking models
Scoring

Flexibility: Flexible docking
Fragment-based docking

Data-driven docking: Contact/interface -driven
Fitting in 3D shape

Other docking paradigms

Modeling pipeline

Data-driven docking

Contact / interface data

Experimental: NMR
Mutagenesis
FRET
…

Biological: Existing mutation
Concerved contact in homologs
...

Ambiguous interaction restraints

Experimental structures

Cryo-Electron Microscopy

For very large systems
e.g. Ribosomes
Multiple conformations

Not for small systems
Resolution often > 4 A

Shape data => Fitting in 3D map

SAXS, Cryo-EM ...

Cryo-electron microscopy generates electron density maps.
The resolution varies a lot!

=> Fitting in 3D map

Shape data => Fitting in 3D map

SAXS, Cryo-EM ...

Sampling: Voxel-based atom density mask

Electron density map contour

Protein voxels
(inside electron density contour)

Allowed atom density:
0.824 Da/Å^3 * 1.5 * voxelsize$^{2/3}$

Solvent voxels
(outside electron density contour)

Allowed atom density: 0

ATTRACT-EM

Experimental data on protein – RNA interfaces

The docking problem: Sampling
Evaluation of docking models
Scoring

Flexibility: Flexible docking
Fragment-based docking

Data-driven docking: Contact/interface -driven
Fitting in 3D shape

Other docking paradigms

Modeling pipeline

Docking paradigms

Minimisation-driven sampling

Gradient-descent — HADDOCK, ATTRACT

Monte Carlo — ATTRACT, RosettaDock

Swarm optimization

Docking paradigms

Systematic sampling

FFT — ClusPro, ZDOCK, FroDock, GRAMM, FTDock, DOT ...

Minimisation-driven sampling

Gradient-descent — HADDOCK, ATTRACT

Monte Carlo — ATTRACT, RosettaDock

Docking paradigms

Systematic sampling

FFT — ClusPro, ZDOCK, FroDock, GRAMM, FTDock, DOT ...

Minimisation-driven sampling

Gradient-descent — HADDOCK, ATTRACT

Monte Carlo — ATTRACT, RosettaDock

IFFT

speed-up ~10^7

Docking paradigms

Systematic sampling

FFT — ClusPro, ZDOCK, FroDock, GRAMM, FTDock, DOT ...

Spherical FFT — HEX

Minimisation-driven sampling

Gradient-descent — HADDOCK, ATTRACT

Monte Carlo — ATTRACT, RosettaDock

Swarm optimization — SwarmDock

Docking paradigms

Systematic sampling

FFT	ClusPro, ZDOCK, FroDock, GRAMM, FTDock, DOT ...
Spherical FFT	HEX
Geometric hashing	PatchDock

Minimisation-driven sampling

Gradient-descent — HADDOCK, ATTRACT

Monte Carlo — ATTRACT, RosettaDock

Swarm optimization — SwarmDock

Experimental data on protein – RNA interfaces

The docking problem: Sampling
Evaluation of docking models
Scoring

Flexibility: Flexible docking
Fragment-based docking

Data-driven docking: Contact/interface -driven
Fitting in 3D shape

Other docking paradigms

Modeling pipeline

Protein and RNA bind ?
dunno
Protein – RNA interaction prediction
minutes
yes
No sure => test
Biol. or exp. data
weeks
Structures available ?
no
yes
Structure prediction
Min / hours
Docking
restraints
minutes
weeks
Experimental validation
select best models
Clustering, to keep most representative
filter
Energy computation
days
Keep top 100-1000 models
Multiple MD simulations
hours
Convert to all-atom
Visual check
Minimisation