Free energy methods to study conformational transitions and ligand-binding to flexible targets

Francesco Luigi Gervasio

Computational Biophysics Group,
Spanish National Cancer Research Centre, Madrid, Spain
flgervasio@cnio.es
Flexible targets in cancer therapy

Kinase inhibitors are effective and promising in cancer therapy

However structure-based design is difficult due to:

Conformational plasticity [1]

High degree of homology in “rigid” regions[2]

Role of solvent in inducible hydrogen bonds[2]

Problems with classical docking

Fast docking methods:
1. simplified Hamiltonian,
2. rigid cavity/local flexibility
3. efficient search methods
4. clustering

Advantages:
Fast, Correct docking geometry

Problems:
1. Accurate scoring is hard (error~3kcal/mol)
2. Highly ordered water difficult to treat(Cox2, biotin/streptavidin [1])
3. Flexible targets/induced docking (CDK2 [2])

A possible alternative: Molecular dynamics

A computer simulation, wherein atoms and molecules interact under the laws of physics.

\[
M_I \frac{d^2 R_I}{dt^2} = -\nabla_I V_{\text{eff}} (\{R\})
\]

\( V_{\text{eff}} \) fit on experiment

\( V_{\text{eff}} \) from first principle cal.

First macromolecular MD simulation:
Bovine Pancreatic Trypsine Inhibitor (Nature 1977, 500 atoms, 9.2 ps=0.0092 ns) Protein motion as essential in function.
A fulfilled promise?

Albeit the many successful applications of MD and FPMD, two main problems hinder its full acceptance:

**Accuracy of FF and DFT Functionals**

Some progress: Amber99SB, ParmBSC0, Charmm22-cmap, DESRES FF,…

**Sampling of phase space (time scale problem)**

Hinders direct comparison with experiment
The time scale problem

In principle an atomistic simulation could describe many biologically relevant event.

Direct simulation allows only short runs (~10M steps ~10 ns)

while most interesting events take place on a longer time scale.
Possible solutions:

- Purpose built machines (Anton, IBM)
- Use of GPUs (NAMD, ACEMD)
- Methods to sample rare events:
  - Temperature enhanced sampling (histogram reweighting, parallel tempering, ...)
  - “flattening” the surface (Conformational flooding, local elevation, hyperdynamics, puddle-skimming, self-healing umbrella sampling, Wang-Landau. “Well tempered Metadynamics”)
- Trajectory-based schemes (transition path sampling, Lagrangean action minimization, nudged elastic band, String method...)

Best results by combining several approaches!
Metadynamics/Self healing US

1. Set-up a MD
2. Choose collective variables (S) approx. RC
3. The algorithm to explore $F(s)$:
   Each N time steps add a Gaussian

\[ V_G(S(x), t) = \sum_{t' = \tau_G, 2\tau_G, \ldots} \exp\left( -\frac{(S(x) - s(t'))^2}{2\delta s^2} \right) \]

\[ w = \omega e^{-[V(s,t)/\Delta T] \tau_G} \]

5. For large times:

\[ \tilde{F}(s, t) = -\frac{T + \Delta T}{\Delta T} V(s, t) \]

Advantages:

• Efficient (Well tempered MetaD only explores only relevant configurations)
• Knowledge of final state not required*
• Full FES reconstructed

Disadvantages:

• Choice of CVs not always obvious (PTMetaD)

Selected Biological Applications:
Docking benzamidine to β-trypsin

- All-atom description
- Solvated with 2072 TIP3 waters
- Counter-ions
- Amber force field
- ORAC MD

Choice of CVs

1. Distance from the cavity or CM
2. Angle between the principle axis of inertia and that distance

*In the case of large flexible ligands deeply buried in the cavity more CVs are needed
Starting with the inhibitor in solution
The problem of hidden degrees of freedom.

1. Metadynamics fails to reconstruct the FES.
2. Often rare events are still sampled.
3. Lack of convergence is diagnostic.
A difficult problem for CV methods: large scale motions in biomolecules.

- Complex transition
- CVs: distances, dihedrals, native contacts, salt-bridges?

Trajectory of Kinase activation. From closed to open and back
Path-like CVs

Path-like CVs

Where the square distance can be defined in different spaces as MSD between 2 aligned structures or in contact map space as:

\[
\mathbf{s(x)} = \lim_{\lambda \to \infty} \frac{\int_0^1 t e^{-\lambda \|S(x) - S(t)\|^2} \, dt}{\int_0^1 e^{-\lambda \|S(x) - S(t)\|^2} \, dt},
\]

\[
\mathbf{z(x)} = \lim_{\lambda \to \infty} -\frac{1}{\lambda} \ln \int_0^1 e^{-\|S(x) - S(t)\|^2} \, dt
\]

Unique advantage: non local exploration

Potential (a.u.)

Free Energy (a.u.)

FES(s,z)
Cyclin-dependent kinases control the cell division. Cdk deregulation in tumors.

2 sub-units: kinase and cyclin.

Interaction with cyclin and phosphorillation:
- Conformational transition
  - Rotation of the $\alpha$C helix
  - Rearrangement of the T-loop (~ 25 Å)
Malfunctioning of Cdk5 is associated with some neurodegenerative diseases, Alzheimer, Parkinson etc.

45000 atoms
OPLS force field

The case of Cdk5
1. Modeling the closed state using the homology with CDK2
2. Interpolating the 2 structures with bio-informatic tools (MOLMOV: molmovdb.mbb.yale.edu/)
3. Optimizing the path with PCV (NAMD 2.6 modified + OPLS FF).
4. Using Metadynamics to reconstruct FES and explore alternative paths (>400 ns dynamics = several weeks on 128 cores Cray XT4)
Mechanism
The path finds a metastable intermediate

Finding an intermediate opens the avenue to the design of selective inhibitors!
Cdk5: the reactive trajectory

Induced fit?: Adenosine Kinase

AK converts adenosine to AMP

Two classes of inhibitors

Distinct binding modes: AK small domain is rotated by 30° relative to large domain.

Metadynamics + PCV

Metric = MSD
Code = NAMD / ACEMD (GPU) + PLUMED*
FF = Amber99SB

1. Find and optimize a path

2. Metadynamics run with S, distance and angle

Masetti, Branduardi and Gervasio, Manuscript in preparation

*http://merlino.mi.infn.it/~plumed/PLUMED/Home.html
Ile293-Asn296 located in $\alpha11$ are involved in a cofactor dependent conformational transition that defines the oxyanaion hole (highly conserved in Ribokinase-like enzymes: DTXGAGD motif)
Differential binding to CDK2

2-anilino-4(hetero)aryl-pyrimidine derivatives

Bind differently to the active and inactive form of CDK2

Experimental inhibition constants correlate only with interaction energies calculated for the active form

2 compounds are predicted to have higher affinity to the inactive form

Kontopidis et al.; *Chem. Biol.* 13, 201-211, 2006
Binding FES

Code: NAMD and ACEMD with PLUMED
FF: Amber 99SB

Undocking path obtained by metadynamics (large hills)

Path build on metadynamics surfaces
And optimized

FES(S,Z) obtained for docking and undocking to CDK2A

FES in 2 days on 3GPUs

FL Gervasio et al.; Manuscript in preparation
Calculated binding FE

<table>
<thead>
<tr>
<th>Compound</th>
<th>ΔΔG (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exp.</td>
<td>Calc.</td>
</tr>
<tr>
<td>-1.5</td>
<td>-1.0</td>
</tr>
<tr>
<td>-2</td>
<td>-2.5</td>
</tr>
<tr>
<td>-4</td>
<td>-6</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>-2</td>
<td>-2.5</td>
</tr>
</tbody>
</table>
Conclusions

Biased MD come of age?

Path CV + biased MD + faster MD (GPU/ Anton?) are able to
1. Model very complex conformational transitions
2. Reconstruct FES and crucial metastable intermediates
3. Discern between induced fit and conformational selection
4. Accurately score binding energies with full complexity

Open issues:

• Computationally expensive
• Systematic errors in the FF could invalidate results
• More experimental validation needed!
Acknowledgments and Collaborations

Dr. Davide Branduardi
Dr. Matteo Masetti
Massimiliano Bonomi
Anna Berteotti

Prof. Michele Parrinello

Dr. Marc Bianciotto

Prof. Scapozza
Andrea Cristiani

Dr. Ludovico Sutto
Dr. Jarek Juraszek

Dr. Alfonso DeSimone
Thank you for your attention

1 Staff Scientist position available!
(>2 years of post-doc required)

Send CV to flgervasio@cnio.es or apply at www.cnio.es
Supporting Material
Well Tempered MetaD

For large times, $V(s; t)$ varies so slowly that one can assume that the $q$'s reach equilibrium, the probability distribution becomes

$$P(s, t)ds \propto \exp\left(-\frac{F(s) - V(s, t)}{T}\right)ds$$

$$\dot{V}(s, t) = \omega e^{-[V(s, t)/\Delta T]}P(s, t)$$

$$= \omega e^{-[V(s, t)/\Delta T]} \frac{e^{-[F(s) + V(s, t)]/T}}{\int ds e^{-[F(s) + V(s, t)]/T}}.$$ 

$$V(s, t \to \infty) = -\frac{\Delta T}{\Delta T + T} F(s)$$

$$\tilde{F}(s, t) = -\frac{T + \Delta T}{\Delta T} V(s, t)$$


FIG. 1 (color). Panels (a–c) Green dots represent 6 ns long trajectories in the $(\Phi, \Psi)$ space for different choices of $\Delta T$ [600 K (a), 1800 K (b), and 4200 K (c)]. The underlying color map (kcal mol$^{-1}$) shows the reference free-energy landscape. Panel (d) Estimate of the free-energy difference between the two metastable minima $C_{mix}(70, -70)$ and $C_{eq}(-83, 74)$ as a function of the simulation time, as obtained from the same trajectories.
Optimizing the path
Optimizing the guess and Finding new paths

Correspondence between minimum free energy paths in Ramachandran plot representation (panel a) and s, z space (panel b). The isoline separation is 1.0 kcal mol$^{-1}$. The yellow path is the reference path. The cyan and red paths represent alternative low-free energy paths found by PCV.