Sampling Conformational Changes with Metadynamics and PT-MetaD

Francesco L. Gervasio

Chair of Bio-molecular Modeling

Department of Chemistry & Institute of Structural and Molecular Biology

University College London
Dynamics links structure to function.
The synergy between structure and dynamics (plasticity) is essential to the function of proteins.

M. Perutz and J. C. Kendrew
Nobel Prize 1962

V. Ramakrishnan, Thomas A. Steitz, AE Yonath; Nobel Price 2009

CDK5 activation [Gervasio et al. JACS 2008]
Conformational changes in Protein kinases

Signal transduction, cellular growth/development/death...
Protein kinases are tightly regulated

Disease
Most frequently contributing to cancer

Chronic Myeloid Leukemia (Bcr/Abl fusion 95%)
Kinase activation is accompanied by conformational changes

Epidermal Growth Factor Receptor activation
Sutto & Gervasio Proc Natl Acad Sci, in press, doi: 10.1073/pnas.1221953110

2. The flip of the DFG-motif (c-Abl)
Lovera,…, Gervasio*, JACS 134, 2496 (2012)
Understanding target dynamics in great details is challenging.

Conformational dynamics, allosterism, solvation play a crucial role.

**Molecular Dynamics** is useful.

\[ M_t \frac{d^2 R}{dt^2} = -\nabla_l V_{\text{eff}}(\{R\}) \]

Crystallography: EGFR – lung cancer
Wild type (yellow) and mutant
Small differences!

[Yun et al. Cancer Cell, 11, 217, 2007]
Limiting factors in Molecular dynamics simulations

Accuracy of potential

The time-scale problem!

AMBER 99SB ILDN∗

\[ M_I \frac{d^2 R_I}{dt^2} = -\nabla I V_{\text{eff}}(\{R\}) \]
MANY DIFFERENT SOLUTIONS PROPOSED

• Thermodynamic integration.

• “flattening” the surface (umbrella sampling, hyperdynamics, etc.).

• Trajectory-based schemes (nudged elastic band, string method, transition path sampling, Lagrangean action minimization, …).

• Finding the saddle points (eigenvalue following, dimer method, hessian based methods,…).

• Temperature enhanced sampling (histogram reweighting, parallel tempering, …)

• …
Well-tempered Metadynamics

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

\[ V(S, t) = \int_0^t dt' \omega \exp \left( - \sum_{i=1}^d \frac{(S_i(R) - S_i(R(t'))^2}{2\sigma_i^2} \right) \]

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

4. For large times:

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

Laio and Parrinello, PNAS, 2002;

PLUMED gpl plug-in for most MD codes
http://www.plumed-code.org/
\[ V(S, t) = \int_0^t \, dt' \omega \exp \left( - \sum_{i=1}^d \frac{(S_i(R) - S_i(R(t'))^2}{2\sigma_i^2} \right) \]

**What do we need to choose?**

1. **Collective variables**
   \[ S = (S_1(R), \ldots, S_d(R)) \]

2. **Energy rate**
   \[ \omega = \frac{W}{\tau_G} \]

3. **Gaussian width**
Example.

HILLS W_stride 1000 Height 0.4
TORSION LIST 5 7 9 15 SIGMA 0.35
ENDMETA

- The main output is the HILLS file

<table>
<thead>
<tr>
<th>time (ps)</th>
<th>CV (rad)</th>
<th>Gaussian sigma (rad)</th>
<th>Gaussian height (kJ/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.000</td>
<td>-2.617548716</td>
<td>0.3500000000</td>
<td>0.4000000000</td>
</tr>
<tr>
<td>2.000</td>
<td>-2.718742869</td>
<td>0.3500000000</td>
<td>0.4000000000</td>
</tr>
<tr>
<td>3.000</td>
<td>-2.662313657</td>
<td>0.3500000000</td>
<td>0.4000000000</td>
</tr>
<tr>
<td>4.000</td>
<td>-2.378730722</td>
<td>0.3500000000</td>
<td>0.4000000000</td>
</tr>
<tr>
<td>5.000</td>
<td>-2.120031391</td>
<td>0.3500000000</td>
<td>0.4000000000</td>
</tr>
</tbody>
</table>
Well-tempered Metadynamics

\[ W = W_0 e^{-V_B(s,t)/(f-1)T} \]

- \( f \) an input parameter, called **bias factor**, (\( f = +\infty \) -> regular metadynamics)
  \[ f = \frac{(T + \Delta T)}{T} \]

- The height of the added Gaussian **decreases** depending on the underlying bias
  \( \rightarrow \) improves the convergence of the free energy

- Tuning the parameter to avoid exploration of too high energy states, limiting the **exploration of the CV space to the relevant regions**

Example.

HILLS W_STRIDE 1000 HEIGHT 0.4
WELLTEMPERED SIMTEMP 300 BIASFACTOR 15
TORSION LIST 5 7 9 15 SIGMA 0.35
ENDMETA

- The **HILLS** file looks like this:

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1.000</td>
<td>-2.617548716</td>
<td>0.350000000</td>
<td>0.428571429</td>
<td>15.000</td>
<td></td>
</tr>
<tr>
<td>2.000</td>
<td>-2.718742869</td>
<td>0.350000000</td>
<td>0.423889087</td>
<td>15.000</td>
<td></td>
</tr>
<tr>
<td>3.000</td>
<td>-2.662247736</td>
<td>0.350000000</td>
<td>0.419017741</td>
<td>15.000</td>
<td></td>
</tr>
<tr>
<td>4.000</td>
<td>-2.380845469</td>
<td>0.350000000</td>
<td>0.418270876</td>
<td>15.000</td>
<td></td>
</tr>
<tr>
<td>5.000</td>
<td>-2.119639700</td>
<td>0.350000000</td>
<td>0.420668977</td>
<td>15.000</td>
<td></td>
</tr>
</tbody>
</table>
• Free-energy reconstruction with SUM_HILLS

\[ V(S, t \to \infty) = -F(S) + C \]

• First we have to compile it!

Example.
sum_hills -file HILLS -out fes.dat -ndim 3 -ndw 1 2 -kt 0.6 -ngrid 100 100 100

- file input file with the list of Gaussians
- out output file with the FES
- ndim number of collective variables
- ndw ID of the variables for FES in output
- ngrid grid mesh dimension
- dp grid bin size
- kt \( kT \) in the energy units
- stride how often the FES is written
- 2pi ID of the variables with \([0; 2\pi]\) periodicity
- pi ID of the variables with \([-\pi; \pi]\) periodicity
Advantages

• accelerates sampling by pushing the system away from explored regions
• no a-priori knowledge of the FE required
• quantitative a-posteriori reconstruction of the FES

Limitations

• the dynamics is altered
  no kinetic information (but... arXiv:1309.5323, Tiwari & Parrinello)

Critical points: choice of the CVs

• CVs must:
  – describe the transition of interest
  – include all slow varying degrees of freedom
  – be a small number (typically max 3)
How to choose the CV...

- use your chemical/physical intuition of the system and the transition
- look for similar problems in the literature
- trial & error!
Induced fit and conformational selection
Imatinib and the DFG flip

Bcr/Abl

Capdeville et al. Nat Rev Drug Discov 1, 493, 2002
Imatinib binding requires conformational changes
The DFG-flip

The flip of the DFG-motif (c-Abl)
Lovera, ... Gervasio*, JACS 134, 2496 (2012)

Despite a 47% sequence identity the cSRC affinity >2000x lower (Seeliger, Structure 2007)
Is DFG-out favored in Abl vs cSrc?

Y Shan et al. PNAS 106, 139 (2009)
c-Abl and c-Src simulations

FF: Amber99SB-ILDN* TIP3P water
MD Code: ACEMD [Harvey et al. JCTC 5,1632, 2009]
3μs long simulation of c-Abl at 340K (2 months run) only 2 DFG flips. The WT cSrc DFG does not flip.
Local differences

Abl CV3: distance PHE – ILE
SRC CV3: distance PHE – LEU
Choice of CV and hidden degrees of freedom

1. Free energy methods based on CV fail to reconstruct the correct FES.
2. The advantage of Metadynamics
   Non diffusive behavior of CVs is diagnostic

**Solutions:**

1. Systematic definition of CVs by PCA
   [Sutto, D'Abramo and FL Gervasio, JCTC 2010]

2. Path-like CV
   [D Branduardi, FL Gervasio et al., J Chem Phys 2007]

3. Parallel Tempering + Metadynamics
   [G Bussi, FL Gervasio et al., JACS 2006]
Parallel Tempering + Metadynamics with 3CVs + 1

- N different biases (same CVs)
- Exchange probability is

\[
\min\{1, \exp[(\beta_j - \beta_i)(U(q_j) - U(q_i)) + \beta_iV_i(s(q_j)) - V_i(s(q_i)) + \beta_jV_j(s(q_i)) - V_j(s(q_j))]\}
\]

only change in plumed.dat: add “PTMETAD”

Prepare 6 input files (gromppX.mdp with temperatures)
300 380 460 540 620 700

Prepare 6 topology files (topolX.tpr)

only change in plumed.dat: add “PTMETAD”

mpirun -np 6 mdrun -multi 6 -replex 100 -plumed plumed.dat

swap every 100 steps
swap every 6 replicas steps
6 processors (you can use N*6 here)
PT+ Metadynamics* 4CVs

28 replicas $T=[308;400]$

4 CVs:

1. Dist Lys45 Asp154
2. Dist Phe155 Leu/Ile
3. Beta Sheet
4. Dihedral combination DFG

GROMACS 4.5 + PLUMED 1.2.1 plug-in on 28x10 cores

CV1: distance ASP – LYS [nm]

CV2: distance PHE – ILE [nm]

CV2: distance PHE – LEU [nm]

Abl

DFG-Asp in

DFG-Asp out

7 kcal/mol

5 kcal/mol

Src

TS

7 kcal/mol

ΔΔG_{exp} = 2.7 kcal/mol

Estimating binding affinities with optimal collective variables

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:

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

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

\[
    \sum_{j > i} (C_{i,j}(x) - C_{i,j}(l))^2
\]

Non local exploration!

Potential (a.u.)

Free Energy (a.u.)

FES(s,z)
PathCV in practice

1. Select the subspace in which your path will be defined (RMSD, DRMSD, CMAP)

2. Define a reference path that connect state A to state B

3. Refine the initial path (e.g. by using steered dynamics)

4. Run Metadynamics on s(x) and z(x).

# the two path variables #
S_PATH TYPE RMSD FRAMESET frame_ NFRAMES 12 LAMBDA 10300 SIGMA 0.3
Z_PATH TYPE RMSD FRAMESET frame_ NFRAMES 12 LAMBDA 10300 SIGMA 0.002
#
Do we get the correct finding free energy?

\[
\Delta G = -8.5 \text{ kcal/mol}
\]

\[
\Delta G^\ddagger = 8.0 \text{ kcal/mol}
\]

\[
\Delta G^\ddagger = 11.0 \text{ kcal/mol}
\]

\[
\Delta G^\ddagger = 2.5 \text{ kcal/mol}
\]

Exp: -7 kcal/mol
Effects of pathogenic mutations on EGFR
Activating mutations in EGFR are found in many cancers.
Changes in the inactive to inactive equilibrium (above) linked to oncogenic potential?
What can we learn on EGFR from very long MD?

Anton, the most powerful computer for MD:

25 µs (10^{10} integration steps) = 1 event

Shan et al. PNAS 110, 7270, 2013
Which CVs would you use?
PT-MetaD – well tempered ensemble

32 (6) replicas $T = [300; 370]$  

3 CVs:

1. Dist from Active Contact Map
2. Dist from Inactive CM
3. Diff of Salt Bridge distances

GROMACS 4.6 + PLUMED$^1$ 1.3 plug-in on 6 x 128 cores

[Sutto & Gervasio PNAS, 110, 10616, 2013]
PT-metaD with well-tempered ensemble

without WTE

25-30 Replicas

WTE = Enhanced fluctuations in U
5 instead of 30 replicas!

[Sutto & Gervasio PNAS, 110, 10616, 2013]
[Sutto & Gervasio PNAS, 110, 10616, 2013]
[Sutto & Gervasio PNAS, 110, 10616, 2013]
Allosteric inhibition of Fibroblast Growth Factor Receptor

FGFs regulate a plethora of developmental processes, as brain patterning and limb development.

Angiogenesis, tissue repair

Over-expressed in cancer
Folkman et al. Am J Pathol, 130, 1988
Fibroblast Growth Factor Receptor 2

Ig-like D1

D2

heparin

FGF

FGF binding site

D3

Kinase domain

Experimental evidence

By screening of 100,000 compounds we found “SSR”

- Binds to ECD FGF inhibits endocytosis

Isothermal titration calorimetry, SPR

- Drug docks to D3, FGF still binds

FTIR measurements upon SSR binding

- Helical structure increase

• No binding pocket is visible in the X-ray structure

• NMR of the D3 domain fails due to excessive flexibility
Can we find the cavity?
Metadynamics with first PCA vectors of D2 D3

A new cavity opens, where SSR can bind, but...

When we calculate the binding free energy, the minimum is too shallow.
Collective variables for Well-Tempered PT-MetaD

CV 1: distance CM - ligand

CV 2: orientation

CV 3: $\alpha$-helical h-bonds

CV 4: $\beta$-sheet h-bonds

CV 5: PCA vectors

CV 6: dihedral combinations

CV 7: ligand hydration

GROMACS 4.5 + PLUMED 1.2.1 plug-in  AMBER 99SB-ILDN* on 64x6 cores

http://www.plumed-code.org/
2. Native helix stabilization through unspecific binding around the FGF docking site
Conformational changes and helix growth

N

H1

H2

Cavity opens

Unfolding

U
A superimposition of the D3 domains of solved FGF–FGFR complex structures. The variation in the conformation of the βC’–βE loop between the structures is evident. The plasticity of this loop is a major determinant of FGF–FGFR binding specificity.
State H2

- Tyr328 goes in and forms contact with the hydrophobic core
- Beta sheet grows
- Helix grows to 2 turns and points up
- A cavity is formed
- Beta sheet half broken

H2 vs N
Comparison of backbone B-factor (temperature-factor) distribution of the FGFR2 D2D3/FGF1 crystal structures in the absence (left) and the presence (right) of SR128545 at 4Å resolution.
Estimating affinities

Undocking profile – H2

undocking profile – NMA site

free-energy [kcal/mol]

11 kcal/mol

free-energy [kcal/mol]

5 kcal/mol
Table 1. Structure Activity Relationship of SSR Analogs

<table>
<thead>
<tr>
<th>Compound</th>
<th>R₁</th>
<th>R₂</th>
<th>R’</th>
<th>R”</th>
<th>pERK Inhibition (IC₅₀ Range, nM)</th>
<th>ΔG Calculation</th>
<th>Glide XP Score</th>
<th>NMR Binding Event (fSTD/fWL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>OMe</td>
<td>Me</td>
<td>COOH</td>
<td>NH₂</td>
<td>&lt;100</td>
<td>-11</td>
<td>-10.3</td>
<td>2.0/3.8</td>
</tr>
<tr>
<td>(SSR)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>OMe</td>
<td>cPr</td>
<td>COOH</td>
<td>NH₂</td>
<td>&lt;100</td>
<td>-12</td>
<td>-11.3</td>
<td>4.9/15.0</td>
</tr>
<tr>
<td>3</td>
<td>H</td>
<td>Me</td>
<td>COOH</td>
<td>NH₂</td>
<td>&lt;100</td>
<td>-10⁺</td>
<td>-11.3</td>
<td>4.4/12.6</td>
</tr>
<tr>
<td>Weak</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>OMe</td>
<td>Me</td>
<td>NH₂</td>
<td>COOH</td>
<td>&lt;1,000</td>
<td>-8</td>
<td>-8.2</td>
<td>1.5/3.2</td>
</tr>
<tr>
<td>6</td>
<td>OMe</td>
<td>Me</td>
<td>COOH</td>
<td>NHMe</td>
<td>&lt;1,000</td>
<td>-9</td>
<td>-9.8</td>
<td>3.6/11.6</td>
</tr>
<tr>
<td>7</td>
<td>OMe</td>
<td>Me</td>
<td>CONH₂</td>
<td>NH₂</td>
<td>&lt;1,000</td>
<td>-9</td>
<td>-9.1</td>
<td>ND</td>
</tr>
<tr>
<td>9</td>
<td>Et</td>
<td>Me</td>
<td>COOH</td>
<td>NH₂</td>
<td>&lt;1,000</td>
<td>-8</td>
<td>-12</td>
<td>ND</td>
</tr>
<tr>
<td>Inactive</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>OMe</td>
<td>Me</td>
<td>COOMe</td>
<td>NH₂</td>
<td>&gt;1,000</td>
<td>-8</td>
<td>-10.3</td>
<td>ND</td>
</tr>
<tr>
<td>14</td>
<td>OMe</td>
<td>Me</td>
<td>COOH</td>
<td>OH</td>
<td>&gt;1,000</td>
<td>-5</td>
<td>-11.9</td>
<td>ND</td>
</tr>
<tr>
<td>17</td>
<td>3-Methoxybenzamide</td>
<td>Me</td>
<td>COOH</td>
<td>H</td>
<td>&gt;1,000</td>
<td>-4</td>
<td>-4.7</td>
<td>0.7/1.9</td>
</tr>
<tr>
<td>18</td>
<td>3-Methoxybenzene-sulfonamide</td>
<td>Me</td>
<td>COOH</td>
<td>NH₂</td>
<td>&gt;1,000</td>
<td>ND</td>
<td>-4.2</td>
<td>Not detectable</td>
</tr>
</tbody>
</table>

Compound 1 is SSR. ND, not determined; fSTD, STD amplification factor; fWL, quantitative waterLOGSY effect.
Mode of action

Acknowledgments

Nicola D’Amelio
Giorgio Saladino
Kristen Marino
Ludovico Sutto

P. Carmeliet / Leuven
Tom L. Blundell / Cambridge

SANOFI

GOBIERNO DE ESPAÑA
MINISTERIO DE CIENCIA E INNOVACIÓN
IMI
MARIE CURIE ACTIONS
PRACE