Applications of molecular modeling

Understanding, verification and prediction of macromolecular structure

- Homology modeling (determine the structure of a protein from the knowledge of the structure of a related protein)
- Protein folding (determine the structure from the knowledge of its sequence)
- Protein-protein, protein-DNA, protein-drug interaction (HIV protease and reverse transcriptase inhibitors)

Analysis of macromolecular dynamics and mechanism of action

- Protein stability and conformational changes (RNA polymerase, ribosome, helicase, etc.)

Complications

- Disparate time scales:
  - atomic motions: $10^{-14}$ s
  - surface sidechain rotations: $10^{-11}$ s
  - DNA bending: $10^{-10} - 10^{-7}$ s
  - domain motions: $10^{-4} - 1$ s
  - protein folding: $10^{-5} - 10$ s
- Thermal fluctuations

Statistical Mechanics

The probability density for configuration (or microstate) $x$ is

$$p(x) = \frac{e^{-E(x)/kT}}{Z}$$

where $k = 1.38 \times 10^{-23} \text{ m}^2 \text{kg} \text{s}^{-2} \text{K}^{-1}$ is the Boltzmann constant, $T$ is the temperature in Kelvins and $Z$ is the partition function

$$Z = \int e^{-E(x)/kT} \, dx$$

Deep local minima of $E$ correspond to observable macrostates

Methods of molecular modeling

Energy minimization – local or global minima

Monte Carlo sampling – occupancy probabilities for macrostates, distribution of microstates

Molecular, Langevin, Brownian dynamics – dynamics of transitions between macrostates
(a) Molecular Dynamics
TBP
TATA / TBP snapshots

(b) Brownian Dynamics
3000 bp supercoiled DNA

(c) Monte Carlo Simulation
repressor protein R_{G/E} map

(d) Poisson Boltzmann
acetylcholine esterase

(e) Targeted Molecular Dynamics
polymerase β enzyme motion

(f) Minimization
DNA/aminofluorene adduct
<table>
<thead>
<tr>
<th>Method</th>
<th>Pros</th>
<th>Cons</th>
<th>CPU</th>
</tr>
</thead>
<tbody>
<tr>
<td>* Molecular Dynamics (MD)</td>
<td>continuous motion, experimental bridge between structure and macroscopic kinetic data</td>
<td>expensive; short timespan</td>
<td>high</td>
</tr>
<tr>
<td>[37, 38, 834]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>* Targeted MD (TMD) [835]</td>
<td>connection between two states; useful for ruling out steric clashes and suggesting high barriers</td>
<td>not necessarily physical</td>
<td>moderate</td>
</tr>
<tr>
<td>* Stochastic Path Approach [836]</td>
<td>high-frequency motion filtering; approximate long-time trajectories</td>
<td>expensive (global optimization of entire trajectory)</td>
<td>high</td>
</tr>
<tr>
<td>* Continuum Solvation [703, 712, 704, 711, 701, 702]</td>
<td>mean-force potential approximates environment and reduces model's cost; useful information on ionic atmosphere and intermolecular associations</td>
<td>approximate</td>
<td>high (if repeated in time)</td>
</tr>
<tr>
<td>* Brownian Dynamics (BD) [703, 269]</td>
<td>large-scale and long-time motion</td>
<td>approximate hydrodynamics; limited to systems with small inertia</td>
<td>moderate</td>
</tr>
<tr>
<td>* Monte Carlo (MC) [833]</td>
<td>large-scale sampling; useful statistics</td>
<td>move definitions are difficult; unphysical paths</td>
<td>low</td>
</tr>
<tr>
<td>* Minimization [758]</td>
<td>valuable equilibria information; experimental constraints can be incorporated</td>
<td>no dynamic information</td>
<td>low</td>
</tr>
</tbody>
</table>

Energy Minimization

Unconstrained minimization problem: Given function \( E(x) \) find \( x_{\text{min}} \in \mathbb{D} \) such that

\[
E(x_{\text{min}}) = \min_{x \in \mathbb{D}} \{ E(x) \}
\]

Minimum: \( x_{\text{min}} \) is a strict (weak) local minimum of \( E(x) \) iff it has an open neighborhood \( \mathbb{D} \) such that \( E(x_{\text{min}}) < E(x) \) \( (E(x_{\text{min}}) \leq E(x)) \) for all \( x \in \mathbb{D}, x \neq x_{\text{min}} \)

Necessary conditions: If \( E(x) \) is smooth and \( x_{\text{min}} \) is a strict local minimum then the gradient \( g \) ( \( g_i = \partial E / \partial x_i \) ) and Hessian \( H (H_{ij} = \partial^2 E / \partial x_i \partial x_j) \) of \( E \) obey

\[
\begin{align*}
g(x_{\text{min}}) &= 0 \\
H(x_{\text{min}}) &\text{ is positive definite}
\end{align*}
\]

Algorithms

Basic algorithm

\[
x_{k+1} = x_k + \lambda_k p_k
\]

where \( p_k \) is appropriately chosen descend direction and \( \lambda_k \) is obtained by 1D line search

Variants: steepest descend \( p_k = -g_k \) where \( g_k = g(x_k) \)

Newton’s method \( p_k = -H_k^{-1}g_k \) where \( H_k = H(x_k) \)

Strategy: Accept \( \lambda_k \) if the following conditions (*) are satisfied

\[
\begin{align*}
E(x_{k+1}) &\leq E(x_k) + \alpha \lambda_k g_k^T p_k \\
g_k^T p_k &\geq \beta g_k^T p_k
\end{align*}
\]

with \( 0 < \alpha < \beta < 1 \) (e.g., \( \alpha = 10^{-4}, \beta = 0.9 \))

THEOREM: If \( E(x) \) is bounded below, continuously differentiable, \( g(x) \) is Lipschitz continuous, \( p_k \) are such that \( -g_k^T p_k / \|g_k\| \|p_k\| \geq \delta > 0 \) and \( \lambda_k \) obey (*) then

\[
\lim_{k \to \infty} \|g_k\| = 0.
\]

Convergence criteria check the smallness of \( \|g_k\|, |E(x_k) - E(x_{k-1})|, \|x_k - x_{k-1}\| \) e.g.,

\[
\|g_k\| \leq \varepsilon(1 + E(x_k))
\]
\textit{Quasi-Newton methods}

- Avoid computation of Hessian by using approximations $B_k$ of $H$ that are updated at each step so as to obey the quasi-Newton condition:
  \[ B_{k+1}s_k = y_k \]
  where $s_k = x_{k+1} - x_k$ and $y_k = g(x_{k+1}) - g(x_k)$

\textbf{BFGS method}

- Updates the inverse $\hat{B}$ of $B$ rather than $B$
  \[ \hat{B}_{k+1} = \left(I - \frac{s_ky_k^T}{y_k^Ts_k}\right)\hat{B}_k \left(I - \frac{y_k^Ts_k}{y_k^Ty_k}\right) + \frac{s_ky_k^T}{y_k^Ts_k} \]
  \[ x_{k+1} = x_k - \hat{B}_k g(x_k) \]

Nocedal’s improvements save memory

\textit{Nonlinear conjugate gradient}

The descend direction $p_k$ obeys
  \[ p_k = -g_k + \beta_k p_{k-1} \]
with $p_0 = -g_0$ and $\beta_k$ is chosen so that if $E(x)$ was convex quadratic, linear CG process would result. Examples:
  \[ \beta_k^{FR} = g_k^Tg_k / g_{k-1}^Tg_{k-1} \quad \text{Fletcher-Reeves} \]
  \[ \beta_k^{PR} = g_k^Ty_{k-1} / g_{k-1}^Ty_{k-1} \quad \text{Polak-Ribiere} \]
  \[ \beta_k^{PR} = g_k^Ty_{k-1} / p_{k-1}^Ty_{k-1} \quad \text{Hestenes-Stiefel} \]

\textit{Truncated Newton method}

- Far away from the minimum instead of exact descent vector $p_k = -H(x_k)^{-1}g(x_k)$ use a cheaply evaluated approximation. Increase accuracy when nearing solution, e.g., by requiring
  \[ \|H_k p_k + g_k\| \leq \eta_k \|g_k\| \]
  where
\[ \eta_k = \min \left\{ \frac{c}{k}, \|g_k\| \right\} \quad 0 < c \leq 1 \]

Uses nonlinear conjugate gradient to find an appropriate \( p_k \) iteratively

**Difficult problem**
- Finding global minima (number of local minima \( \sim a^N \))

**Practical solutions**
- Use many starting points
- Compare results of different algorithms
- Compare results from different force fields
- Check eigenvalues
- Beware of artificial minima introduced by cutoffs