Molecular dynamics at constant pH: theory and application of the stochastic titration method

António M. Baptista

Molecular Simulation Lab
Instituto de Tecnologia Química e Biológica
Universidade Nova de Lisboa
Portugal
Motivation: lack of thermodynamic parameters in MD simulations

- External thermodynamic parameters:
  - In standard molecular mechanics/dynamics (MM/MD):
    - Temperature
    - Pressure
  - Relevant for (redox) proteins and other biomolecules:
    - Temperature
    - Pressure
    - pH

- Thus, we need an MD method that includes the effect of the pH of the solution.
Addressing protonation with standard methods
Classical Hamiltonians for protein electrostatics

- Molecular mechanics (MM) Hamiltonians.
  - Atomic-level model of solute.
  - Atomic-level model of solvent.
  - Used to sample solute and solvent configurations by molecular dynamics (MD) simulations.

- Electrostatics-oriented simplified Hamiltonians.
  - Atomic-level model of solute.
  - Simplified model of solvent:
    - Poisson–Boltzmann (PB)
    - generalized Born
    - grid of dipoles (PDLD)
  - Used to sample protonation states by Monte Carlo simulations or other approximate methods (total or partial mean field, etc)

- These two approaches address protonation differently.
Molecular mechanics/molecular dynamics (MM/MD) approach

- A single “typical” protonation state is chosen for the solute.

- Energetics of that protonation state is described with a MM model:

  \[ E_{\text{pot}} = \sum_{\text{bonds}} K_b (b - b_0)^2 + \sum_{\text{angles}} K_a (\theta - \theta_0)^2 + \sum_{\text{torsions}} K_t (1 \pm \cos n\phi) \]

  \[ + \sum_{\text{pairs}} \frac{q_i q_j}{r_{ij}} + \sum_{\text{pairs}} \left( \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) + \cdots \]

- Configurations of solute and solvent are sampled using molecular dynamics (MD) simulations:

  \[ \cdots \rightarrow \quad \rightarrow \quad \rightarrow \cdots \]

  \( \bullet/\circ = \text{site protonated/deprotonated} \)
• The MM/MD stochastic operator (matrix) $\mathcal{M}$ conserves $\rho(q, p, \bar{q}, \bar{p}|n)$:

$$\mathcal{M}\rho(q, p, \bar{q}, \bar{p}|n) = \rho(q, p, \bar{q}, \bar{p}|n)$$

where

- **solute** configuration: $(q, p)$
- **solvent** configuration: $(\bar{q}, \bar{p})$
- protonation state: $n = (n_1, n_2, \ldots), n_i = 0$ or $1$ (deprot. or prot.)

• Protonation/deprotonation processes can be treated with MM/MD free energy methods, but these are slow and may have problems with the treatment of long-range electrostatics.
Poisson–Boltzmann/Monte Carlo (PB/MC) approach

- A single “typical” conformation is chosen for the solute.
- Energetics of that conformation is described with a PB model:

\[
\Delta G(n) = -2.3RT \sum_i n_i p K^\text{int}_i + \sum_i \sum_{j<i} (n_i n_j + n_i z_j^o + n_j z_i^o) W_{ij}
\]

- Protonation states of solute are sampled using MC simulations:

Dielectric treatment assumes automatically relaxed solvent.
The PB/MC stochastic operator (matrix) $C$ conserves $\rho(n|q)$:

$$C\rho(n|q) = \rho(n|q)$$

For (large) solutes with little conformational freedom, the PB/MC and similar methods (GB, PDLD, etc) gives the best $pK_\text{a}$ predictions.

The solute dielectric constant ($\varepsilon$) has a clear physical meaning for a rigid solute: $\varepsilon \approx 2$, due to induced dipoles.

A higher value of $\varepsilon$ (4–20) can be used to model a flexible solute, but $\varepsilon$ becomes an empirical parameter, without a clear physical meaning.
Complementarity of MM/MD and PB/MC

- MM/MD samples all degrees of freedom except protonation states:

  \[ \cdots \rightarrow \text{MM/MD conformation} \rightarrow \text{PB/MC conformation} \rightarrow \cdots \]

- PB/MC samples (explicitly) only protonation states:

  \[ \cdots \rightarrow \text{PB/MC protonation states} \rightarrow \cdots \]

- Instead of trying to converge both methods (e.g., using conformer states in the PB model), we explored that complementarity:
  - Constant-pH MD using implicit titration
  - Constant-pH MD using stochastic titration
Constant-pH MD using stochastic titration
Hybrid Monte Carlo of protonation and solvent states

- We borrow the concept of hybrid Monte Carlo (HMC).

- The HMC move consists of the following two-step process:

  ![Diagram of HMC move](image)

  - If we devise a proper way to compute the acceptance/rejection of the HMC moves, we get a correct sampling of protonation and solvent states, at fixed solute conformation.

  ![Diagram of HMC move](image)

  - However, the original HMC algorithm holds for changes of momenta, being invalid for changes of protonation state.
• But the probability of the complete HMC move can be estimated by a PB/MC move or run, since the physical process is the same:

![Diagram showing protonation change and solvent relaxation](image)

• PB-estimated HMC move:

1. Solute protonation state is changed by PB/MC.
2. Solvent configuration is relaxed by MM/MD.

• PB-estimated HMC moves sample protonation and solvent states, at fixed solute conformation.

• The PB-estimated HMC stochastic operator (matrix) $\mathcal{H}$ is assumed to conserve $\rho(\bar{q}, \bar{p}, n|q)$:

$$\mathcal{H}\rho(\bar{q}, \bar{p}, n|q) = \rho(\bar{q}, \bar{p}, n|q)$$
Stochastic titration (ST) algorithm for constant-pH MD


• Consists of alternated MM/MD and PB-estimated HMC moves:

\[
\text{... MM/MD} \rightarrow \text{PB-estimated HMC} \rightarrow \text{MM/MD} \rightarrow \text{PB-estimated HMC} \rightarrow \text{...}
\]

• The ST method samples simultaneously:
  
  ○ Protonation states, like PB/MC.
  
  ○ Solute and solvent configurations, like MM/MD.
  
  ○ Reflects temperature, pressure and pH.

• The ST method captures the coupling between protonation and conformation.
HMC

PB-estimated

MM/MD of Solute and Solvent

Changed Solute and Solvent configurations

PB/MC of Solute

Changed Solute protonation state

MM/MD of Solvent

Changed Solvent configuration
Markov chain of stochastic titration is ergodic

- System at constant pH: semi-grand canonical (SGC) ensemble.
- The MM/MD stochastic operator $\mathcal{M}$ conserves the SGC distribution $\rho(q, p, \bar{q}, \bar{p}, n)$:
  \[
  \mathcal{M}\rho(q, p, \bar{q}, \bar{p}, n) = \mathcal{M}\rho(q, p, \bar{q}, \bar{p}|n)\rho(n) = \rho(n)\mathcal{M}\rho(q, p, \bar{q}, \bar{p}|n)
  \]
  \[
  = \rho(n)\rho(q, p, \bar{q}, \bar{p}|n) = \rho(q, p, \bar{q}, \bar{p}, n)
  \]
  In short, $\mathcal{M}\rho = \rho$.
- The PB-estimated HMC stochastic operator $\mathcal{H}$ conserves the SGC distribution $\rho(q, p, \bar{q}, \bar{p}, n)$:
  \[
  \mathcal{H}\rho(q, p, \bar{q}, \bar{p}, n) = \mathcal{H}\rho(q, \bar{q}, \bar{p}, n)\rho(p) = \mathcal{H}\rho(\bar{q}, \bar{p}, n|q)\rho(q)\rho(p)
  \]
  \[
  = \rho(q)\rho(p)\mathcal{H}\rho(\bar{q}, \bar{p}, n|q) = \rho(q)\rho(p)\rho(\bar{q}, \bar{p}, n|q)
  \]
  \[
  = \rho(q, p, \bar{q}, \bar{p}, n)
  \]
  In short, $\mathcal{H}\rho = \rho$. 
A stochastic operator $\mathcal{P}$ that conserves the SGC distribution can be obtained by performing PB-estimated HMC moves:

- **Randomly**, with probability $\lambda$:
  \[
  \mathcal{P}\rho = [(1 - \lambda)\mathcal{M} + \lambda\mathcal{H}]\rho = (1 - \lambda)\mathcal{M}\rho + \lambda\mathcal{H}\rho \\
  = (1 - \lambda)\rho + \lambda\rho = \rho
  \]

- **Periodically**, once every $k$ MM/MD steps:
  \[
  \mathcal{P}\rho = (\mathcal{H}\mathcal{M}^k)\rho = \mathcal{H}\mathcal{M}^{k-1}\mathcal{M}\rho = \mathcal{H}\mathcal{M}^{k-1}\rho \\
  = \cdots = \mathcal{H}\mathcal{M}\rho = \mathcal{H}\rho = \rho
  \]

Both approaches are valid.

- The other conditions for ergodicity (e.g., irreducibility) can also be proved or shown to be reasonable.

- Therefore, the ST method samples from the SGC ensemble, being suitable for constant-pH MD.
Some applications of the constant-pH MD method
Study of the helix–coil transition of decalysine


- Decalysine displays a pH-induced helix–coil transition:

(coil at low pH) \rightleftharpoons \text{ (helix at high pH) }

- Constant-pH MD simulations correctly predict the protonation transition midpoint and the conformational transition midpoint:
**Study of hen egg white lysozyme (HEWL)**


- **HEWL** is a standard benchmark system for $pK_a$ calculations: several groups with *shifted* $pK_a$s, both up and down.

- **Very good $pK_a$ predictions in the acidic pH range:**

<table>
<thead>
<tr>
<th>Residue</th>
<th>Experimental “$pK_a$”</th>
<th>Calculated $pK_{a\text{mid}}$</th>
<th>GRF</th>
<th>PME</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Range</td>
<td>Avg.</td>
<td>$pK_a$</td>
<td>Error</td>
</tr>
<tr>
<td>Glu-7</td>
<td>2.60–3.10</td>
<td>2.85</td>
<td>4.04</td>
<td>1.19</td>
</tr>
<tr>
<td>His-15</td>
<td>5.29–5.43</td>
<td>5.36</td>
<td>4.22</td>
<td>1.14</td>
</tr>
<tr>
<td>Asp-18</td>
<td>2.58–2.74</td>
<td>2.66</td>
<td>4.11</td>
<td>1.45</td>
</tr>
<tr>
<td>Glu-35</td>
<td>6.1–6.3</td>
<td>6.20</td>
<td>5.50</td>
<td>0.70</td>
</tr>
<tr>
<td>Asp-48</td>
<td>1.2–2.0</td>
<td>1.60</td>
<td>2.56</td>
<td>0.96</td>
</tr>
<tr>
<td>Asp-52</td>
<td>3.60–3.76</td>
<td>3.68</td>
<td>3.96</td>
<td>0.28</td>
</tr>
<tr>
<td>Asp-66</td>
<td>0.4–1.4</td>
<td>0.90</td>
<td>1.50</td>
<td>0.60</td>
</tr>
<tr>
<td>Asp-87</td>
<td>1.92–2.22</td>
<td>2.07</td>
<td>2.68</td>
<td>0.61</td>
</tr>
<tr>
<td>Asp-101</td>
<td>4.02–4.16</td>
<td>4.09</td>
<td>3.77</td>
<td>0.32</td>
</tr>
<tr>
<td>Asp-119</td>
<td>3.11–3.29</td>
<td>3.20</td>
<td>2.97</td>
<td>0.23</td>
</tr>
<tr>
<td>Cter-129</td>
<td>2.63–2.97</td>
<td>2.75</td>
<td>3.23</td>
<td>0.48</td>
</tr>
<tr>
<td><strong>RMSD</strong></td>
<td></td>
<td></td>
<td><strong>0.82</strong></td>
<td></td>
</tr>
</tbody>
</table>

(\textcolor{red}{\textbf{Grayed cells}}: error > 1)

- **Excellent prediction**: errors comparable to experimental uncertainty.
pH-induced misfolding of the human prion


- **What is known:**
  - Involved in amyloid diseases: Creutzfeldt-Jakob, scrapie, ...
  - The prion or prionic protein (PrP) is probably the only infectious agent.
  - Known structure for the native or cellular (PrPC) form, but not for the misfolded or scrapie (PrPSc) form.
  - Contradictory experimental evidences about the misfolded region: N-terminal versus C-terminal.
  - It was suggested that the endosome low pH induces prion misfolding:
    \[ \text{PrPC} \xrightarrow{\text{low pH}} \text{PrPSc} \xrightarrow{\text{aggregation}} \text{amyloid fibrils} \]

- To study this we did constant-pH MD of PrP at acidic pH.
The simulations indicate that low pH induces:

- Loss of helical structure:

- Gain of $\beta$ structure:

This is in agreement with experimental results.
- Localization of structural changes:
  - pH has a clear effect on the prion average structure and fluctuations:
    [chain width = fluctuations]
  - Example of formation of persistent $\beta$ structure on C-terminal region:
Conclusions

- Low pH:
  - loss of helical structure
  - gain of \( \beta \) structure

- Structural changes:
  - globular part
  - C-terminal region

- Persistent \( \beta \)-rich structure:
  - \( \text{PrP}^{\text{Sc}} \)?
  - intermediate to \( \text{PrP}^{\text{Sc}} \)?

- Low pH may be enough to induce misfolding, without other factors (protein X, etc).
Concluding remarks
Current state of the stochastic titration method

• Extensions to:
  ◦ Inclusion of proton isomerism [Proteins 2008, 72:289]
  ◦ MD at constant pH and reduction potential [JACS 2009, 131:12586]
  ◦ Titration of lipidic membranes [in progress]
  ◦ Membrane electrochemical gradients [in progress]

• Other applications from our Lab:
  ◦ Cytochrome c3 [JACS 2009, 131:12586]
  ◦ PrP reversibility [submitted]
  ◦ Peptide dendrimers [in preparation]

• Currently the best approach to treat protonation–conformation coupling.

• But other constant-pH MD methods remain to be explored: implicit titration, etc.
Acknowledgments

- People involved:
  - Molecular Simulation Laboratory at ITQB-UNL:
    - Miguel Machuqueiro (implementation, decalysine, HEWL, prion)
      Now Researcher at Faculty of Sciences, University of Lisbon
    - Sara Campos (prion, implementation)
    - Pedro Magalhães (kyotorphin, implementation)
    - Luís Filipe (peptide dendrimers, implementation)
  - Protein Modelling Laboratory at ITQB-UNL
    - Cláudio Soares (initial implementation, discussions)

- Funding:
  Fundação para a Ciência e a Tecnologia
  MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR