BT5420: Computer Simulations of Biomolecular Systems

Take Home Exam : Submitted by Sahana G (BE17B038)

I understand that it is a take home exam, not an assignment, and I'm aware that disciplinary action would be taken if I'm found to be involved in any form of malpractice.

Sahana

Question 1.

The three site TIP3P model (in which point charges are centred on each of the three atoms) is commonly used in AMBER simulations. TIP3P reproduces the key features of bulk water at $25°C$, 1 atm (ie, density of $0.997 g/cm^3$ & $\Delta H_{vap} = 10.53 kcal/mol$) but it underestimates the height of the second (tetrahedral) peak in the O-O radial distribution function & and overestimates the diffusion constant.

Both the four site TIP4P and five site TIP5P models gives better fit for experimental O-O radial distribution functions and is also more accurate than the TIP3P model in different conditions. TIP4P has a dummy atom near O along the bisector of H-O-H angle, with a negative charge. The TIP5P model places negative charges on dummy atoms (representing the lone pair of oxygen atom) and therefore has a tetrahedral-like geometry, thus better reproducing the experimental radial distribution functions from neutron diffraction and the temperature of maximal density of water.

However, as we move from TIP3P to TIP4P to TIP5P by adding additional extra point charges, increases the computational cost of higher versions. TIP4P is ~1.5 times and TIP5P is ~2.5 times as complex (computationally) & expensive as TIP3P in terms of simulation time.

Variants of these water models :

> TIP3P-PME and TIP3P-PME/LRC are generated ~~using~~ by changing a few parameters in the original TIP3P model. They show improved performance in reproducing bulk properties of liquid water (dielectric constant, diffusion coefficient, etc.) relative to TIP3P using short non-bonded cutoffs.

> TIP4P-Ew model is similar to TIP4P, but used with Ewald Summation method.

> TIP4P/Ice is used for simulation of solid water ice.

> TIP4P/2005 is obtained by a general parameterization for simulating the entire phase diagram of condensed water.

> TIP5P-E model is a reparameterization of TIP5P for use of Ewalds Summation Method.

Question 2:

The *.itp file is a portable topology file. There can be different specifications for the
*.itp file. The position restraint file (posre.itp) contains the entries for all the heavy
atoms (hence posing no restraints on the protons in the system) from the original PDB file
and specifies the force field applied to it.

While specifying topology, the *.itp file only lists the information of one particular
molecule and allows you to re-use the [moleculetype] directive in multiple systems
with revoking pdb2gmx command/manually copy-pasting. Unlike *.tpr files, atom IDs
and residue IDs in *.itp files are indexed from 1. Therefore a TPR file created from an
ITP file will have different numbering during MD analysis, than while using ITP file.
The urea.itp file consists of the following directives — molecule types, atoms, bonds,
dihedrals. Example:

```
[moleculetype]
; name    nrexcl
URE        3          residue number
                          residue
nr  [atoms] type       atom    cgnr    charge          mass
    1   C   1   URE   C    1   0.880229    12.01000  ; amber C type
    2   O   1   URE   O    2   -0.613359   16.00000  ; amber O type
    3   N   1   URE   N1   3   -0.923545   14.01000  ; amber N type
    4   H   1   URE   H11  4   0.395055    1.00800   ; amber H type
    5   H   1   URE   H12  5   0.395055    1.00800   ; amber H type
    6   N   1   URE   N2   6   -0.923545   14.01000  ; amber N type
    7   H   1   URE   N21  7   0.395055    1.00800   ; amber H type
    8   H   1   URE   N22  8   0.395055    1.00800   ; amber H type

[bonds]
    1   2  ⎫
    1   3  ⎬ → meaning that atom 1 is connected to atoms 2,3,5.
    1   5  ⎭

    3   4
    3   5
    6   7
    6   8

[dihedrals]
; ai aj ak al function definition
..... → details about all possible dihedral angles will be present
```

Difference with *.top file -

An *.itp file corresponds to a single molecule type, while a *.top describes an entire system. Hence we would expect details for all the residues/small molecules with the *.top file. If we have *.itp file present for individual residues/molecules, we should use #include statement in the *.top file. This allows for a modular representation of residues in the *.top file.

$$\boxed{\text{\#include "urea.itp"}}$$

Question 3

mass: m_1, m_2. positions: r_1, r_2. velocities: $\dot{r}_1, \dot{r}_2$ and $U \propto |r_1 - r_2|$

{magnitude since U is a scalar quantity}

(i) Lagrangian of the system:

$$\mathcal{L}(r_1, r_2, \dot{r}_1, \dot{r}_2) = KE(\dot{r}_1, \dot{r}_2) - PE(r_1, r_2)$$

$$\boxed{\mathcal{L} = \frac{1}{2} m_1 \dot{r}_1^2 + \frac{1}{2} m_2 \dot{r}_2^2 - V(|r_1 - r_2|)} \quad — ①$$

(ii) $R = \dfrac{m_1 r_1 + m_2 r_2}{M}$, $M = m_1 + m_2$, $r = r_1 - r_2$, $\mu = \dfrac{m_1 m_2}{M}$

$\downarrow$ multiply both sides with $m_2 \Rightarrow m_2 r_1 - m_2 r_2 = m_2 r$

$\Rightarrow m_1 r_1 + m_2 r_2 = MR \quad — ②$ $\rightharpoondown ③$

> Adding ② and ③ $\Rightarrow (m_1 + m_2) r_1 = m_2 r + MR$

$$\Rightarrow r_1 = \frac{m_2 r + RM}{M} \quad — ④$$

$\rightharpoondown ⑤$

> Substituting ④ in ③ $\Rightarrow r_2 = r_1 - r = \dfrac{m_2 r + RM - m_1 r - m_2 r}{M} \Rightarrow r_2 = \dfrac{RM - m_1 r}{M}$
> (divide by m_2)

> Substituting ④ & ⑤ in ①, we get:

$$\mathcal{L} = \frac{1}{2} m_1 \left[\frac{m_2 \dot{r}}{M}\right]^2 + \frac{1}{2} m_1 \left[\frac{R\dot{M}}{M}\right]^2 + \frac{1}{2} m_2 \left[\frac{-m_1 \dot{r}}{M}\right]^2 + \frac{1}{2} m_2 \left[\frac{R\dot{M}}{M}\right]^2 - V(r)$$

$$\mathcal{L} = \frac{1}{2} \frac{m_1 m_2}{M^2} \left[m_1 + m_2\right] \dot{r}^2 + \frac{1}{2}(m_1 + m_2) \dot{R}^2 - V(r)$$

$$\mathcal{L} = \frac{1}{2} M \dot{R}^2 + \frac{1}{2} \frac{m_1 m_2}{M} \dot{r}^2 - V(r)$$

$$\boxed{\mathcal{L} = \frac{1}{2} M \dot{R}^2 + \frac{1}{2} \mu \dot{r}^2 - V(r)} \quad — ⑥$$

$\underbrace{\phantom{\frac{1}{2}M\dot{R}^2}}_{\text{Centre of mass}}$ $\underbrace{\phantom{\frac{1}{2}\mu\dot{r}^2}}_{\text{relative motion term}}$

(iii) The second set of variables with the Centre of mass term, reduced mass and relative motion is best suited for this problem. The cartesian variables have more complicated math and solving is more computationally extensive. Also with the use of second set of variables (R, M, μ, r), a potential (diatomic) two-body problem is reduced to a single-body problem. Here, we also decouple the mechanics as centre-of-mass related term and the relative motion term.

Question 4

TPT: $\dfrac{PV}{KT} = \ln \Xi = \dfrac{VG}{\lambda^3} = \langle N \rangle$

Proof: $\Xi = \displaystyle\sum_{N=0}^{\infty} G^N Q(N,V,T) \longleftarrow$ Grand canonical partition function in terms of fugacity and canonical partition function.
$\qquad\qquad\qquad\qquad \underset{\llcorner \text{①}}{}$

Let us first derive canonical partition function:

$$Q = \dfrac{1}{N! h^{3N}} \int_{-\infty}^{\infty} dp_1 \cdots \int_{-\infty}^{\infty} dp_N \int_{-\infty}^{\infty} dq_1 \cdots \int_{-\infty}^{\infty} dq_N \, e^{-\beta H \{p\}\{q\}}$$

Splitting the kinetic and potential energy parts separately, we get:

$$K = \dfrac{1}{h^{3N}} \int_{-\infty}^{\infty} dp_1 \cdots \int_{-\infty}^{\infty} dp_N \, e^{-\beta p^2/2m} = \dfrac{1}{\lambda^{3N}} \qquad \text{②}$$

In ideal gas, the particles don't interact and there is no attraction/repulsion. Therefore there is no overall change in their position.

$$Z = \dfrac{1}{N!} \int_{-\infty}^{\infty} dq_1 \cdots \int_{-\infty}^{\infty} dq_N \, e^{\frac{-\beta \{q\}}{KT}} \qquad \text{Here}: U\{q\}=0. \Rightarrow e^0 = 1.$$

Also, $\displaystyle\int_{-\infty}^{\infty} dq_i \Rightarrow$ position of particle from $-\infty$ to ∞. Since We can find particle anywhere within the volume V $\Rightarrow \displaystyle\int_{-\infty}^{\infty} dq_i = V \qquad \Rightarrow Z = \dfrac{1}{N!} V^N \qquad \text{③}$

$$\Rightarrow Q = \dfrac{V^N}{N! \, \lambda^{3N}} = \left(\dfrac{V}{\lambda^3}\right)^N \dfrac{1}{N!} \qquad \text{④} \longrightarrow \text{Canonical partition function.}$$

Substitute ④ in ①, we get:

$$\Xi = \sum_{N=0}^{\infty} G^N \dfrac{V^N}{\lambda^{3N} N!} \qquad\qquad \text{Also,} \quad G = e^{\mu \beta N} \Rightarrow e^{\mu N/KT}$$

$$\Rightarrow \Xi(T,V,\mu) = \sum_{N=0}^{\infty} \frac{V^N}{\lambda^{3N} N!} e^{\mu N/kT} = \sum_{N=0}^{\infty} \left[\frac{Ve^{\mu/kT}}{\lambda^3}\right]^N \frac{1}{N!}$$

Using $\sum_{n=0}^{\infty} \frac{x^n}{n!} = e^x$, we get:

$$\Xi = e^{\frac{Ve^{\mu/kT}}{\lambda^3}} \longrightarrow \text{Taking } \ln \text{ on both sides} \Rightarrow \ln\Xi = \frac{Ve^{\mu/kT}}{\lambda^3} \quad ⑤$$

$$\Rightarrow \boxed{\ln\Xi = \frac{VG}{\lambda^3}} \quad ⑤$$

We already know that : $PV = kT\ln\Xi \underset{⊛}{} \Rightarrow PV = kT\left[\frac{V}{\lambda^3} e^{\mu/kT}\right] \leftarrow \text{Using} ⑤$

$$\Rightarrow P = \frac{kT}{\lambda^3} e^{\mu/kT} \quad ⑦$$

To get the ideal gas equation / In the ideal gas scenario:

$$\langle N\rangle = \left[\frac{\partial (PV)}{\partial \mu}\right]_{T,V} = kT\left[\frac{\partial\left[\frac{Ve^{\mu/kT}}{\lambda^3}\right]}{\partial\mu}\right]_{T,V} \quad \{\because PV = kT\ln\Xi\}$$

$$\Rightarrow \langle N\rangle = \left[\frac{kTV}{\lambda^3}\right]\cdot\left[\frac{1}{kT}\right]\cdot e^{\mu/kT} = \frac{V}{\lambda^3}e^{\mu/kT} = \ln\Xi \quad \{\text{from eq } ⑤\} \underset{⑧}{}$$

Comparing ⑥, ⑦, ⑧ we get : $\boxed{\frac{PV}{kT} = \ln\Xi = \langle N\rangle = \frac{VG}{\lambda^3}}$ // Hence proved.

Question 5

(i) Site-Site radial distribution function

This is a measure of the probability of finding a particle at a distance r away from a given reference particle, relative to that for an ideal gas.

In a molecular system, tracking only the distance between two molecules does not suffice. Two molecules can be at the same distance away from the reference molecule, but they can be at very different orientations. To avoid this confusion, site-site radial distribution function, is hence used to track the distance between specific atoms (sites).

Hence if we are looking at radial distribution function for water molecules, the different site-site interactions that can be tracked are g_{OO} (commonly measured), g_{OH} (corresponds to the distribution function of all H atoms with a specific O as the central atom) and g_{HO}.

Therefore, $g_{ab}(r_{ab})$ is a set of site-site distribution function, can be calculated in the same way as the academic atomic $g(r)$ for each type of site between molecules.

Here $g(r) = \frac{V}{N^2}\left\langle \sum_i \sum_{j\neq i} \delta(r-r_{ij})\right\rangle$

ii) Hydrogen bond life time analysis based on autocorrelation function.

The time autocorrelation function is calculated for all the selected hydrogen bonds that are passed to it. This set of hydrogen bonds, starting from $t = t_0$ are measured based on geometric criteria and their lifetime is monitored over time

$$C_x(t) = \left\langle \frac{h_{ij}(t_0) h_{ij}(t_0 + t)}{h_{ij}(t_0)^2} \right\rangle$$

Multiple passes through the same trajectory is recorded to obtain the average behaviour. Here the subscript x denotes whether we're defining the lifetime of the hydrogen bonds in continuum time or intermittent time. The continuum definition measures the time that a hydrogen bond remains continuously attached, whereas the intermittent definition also includes the same H bond if it gets cleaved and re-attached during the measurement. The relevant lifetime can be measured as follows:

$$\tau_x = \int_0^\infty C_x(t) dt'$$

while performing MD simulations, this can be measured using the command:

• MD Analysis. analysis. hbonds. hbond_autocorrel

with input as hydrogen bond donors and acceptors.

output as → results (raw result of autocorrelation function)
→ time (time axis is ps)
→ fit ──┐ → continuous lifetime : (A_1, A, τ_1, τ_2) and
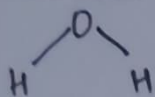→ τ ├─   intermittent lifetime : $(A_1, A_2, \tau_1, \tau_2, \tau_3)$
→ estimate ─┘
                    ↓

$$C_x(t) = A_1 \exp(-t/\tau_1) + A_2 \exp(-t/\tau_2) \left[+ A_3 \exp(-t/\tau_3) \right]$$
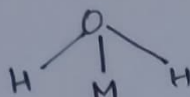
Here & $\sum_{i=1}^{n} A_i = 1 \implies A_n = 1 - \sum_{i=1}^{n-1} A_i$

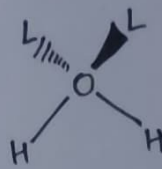## Supplementary information for question 1.

Water models:

2-site   4-site   5-site.

Properties compared across TIP3P, TIP4P, TIP5P.

Property	TIP3P	TIP4P	TIP5P	Experimental
distance (OH) A°	0.9572	0.9572	0.9572	0.9584
$\angle$ HOH, degree	104.52	104.52	104.52	104.45°
A, 10^3 kcal A^{012}/mol	582.0	600.0	544.5	—
B, kcal A^{06}/mol	595.6	610.0	590.3	—
charge: q(X)	X:O, q=-0.834	X:M, q=-1.04	X:L; q=-0.241	—
charge: q(H)	+0.417	+0.52	+0.241	—
dipole moment	2.35	2.18	2.29	2.95
dielectric constant	82	53	81.5	78.4
self diffusion, 10^5 cm/s	5.19	3.29	2.62	2.30
density max @T, ℃	-91	-25	+4	+3.984
expansion coefficient (10^{-4}℃$^{-1}$)	9.2	4.4	6.3	2.53

Using the Leonard Jones potential:

$$E_{ab} = \sum_i^a \sum_j^b \frac{K_c q_i q_j}{r_{ij}} + \frac{A}{r_{00}^{12}} \neq \frac{B}{r_{00}^6}$$

→ K_c: electrostatic constant
      = 332.1 A° kcal/mol e²

→ q_i, q_j are the partial charges relative to charge on electron.

→ r_{ij}: distance b/w charged atoms

→ A,B: LJ parameters.