Computational Biochemistry

Lecture 4
Molecular Dynamics (MD) Simulation

Why dynamic simulation?

- **Biomolecules are dynamic – no single structure is a complete model.**

- **MD simulations remove investigator bias.**

Limitations of energy minimization

- Minimization of the potential energy function (force field) leads to a single structure.

- This structure may represent a reasonable "average" conformation if the molecule is essentially rigid.

- When a molecule is flexible, at room temperature the conformations may interconvert, and so no single structure is very representative.

Average structure or average properties?

- For a flexible molecule, its average structure is not physically meaningful!

- If a snake wiggles equally right and left, what is its average shape?
- Does that average ever look like the snake?
- But, the snake has an average length, so the average properties can be used to define the snake.

Stochastic vs deterministic modeling of motion

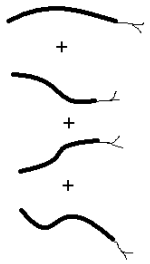
- The "structure" of a flexible molecule can be thought of as an ensemble of individual conformations (also called snapshots), which give rise to characteristic average properties.

- These average properties can be generated by random sampling of the potential conformations (**stochastic sampling**).

- By modeling the time-dependency of the conformations (**molecular dynamics simulation**).

The Ergodic Hypothesis

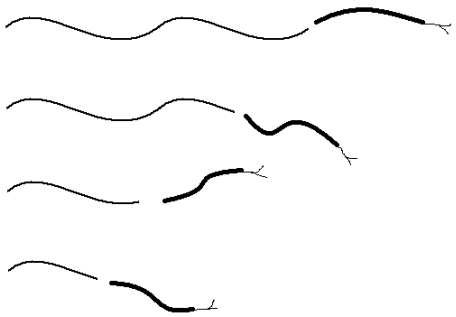
Multiple shapes of a snake – determined at from random sampling (stochastic)



If enough shapes are observed, both stochastic and deterministic methods will result in the same average properties.

This is called the “Ergodic” principle.

Shapes of a snake observed as the snake moves under its own force (deterministic)



The difference is that deterministic methods explain how the shapes occur.

That is, deterministic methods explain the path of the process, where as stochastic methods do not.

The conclusion:

"A snake is mostly elongated, but wiggles from side to side”: This would be reached from either average, but the time-dependent approach (molecular dynamics, MD) lets you predict where the snake will go next, and understand how it got to where it is now.

This is very important for understanding biochemical processes.

Molecular dynamics (MD) simulation

- All molecules are in motion except at absolute zero. The motion can be associated with overall **translation**, **rotation** or **vibration**, which are relevant to temperature and pressure of a system.

- However, more importantly, some of the motions are internal. These internal motions can drastically alter the conformation of the molecule, particularly through torsion angle rotation.

- In reality, one conformational state changes to another in a smooth time-dependent manner.

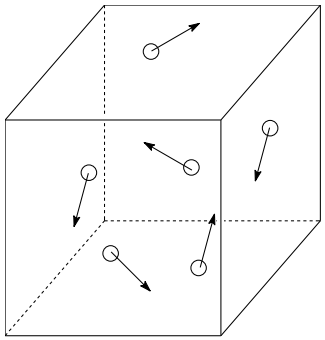
- In MD simulations the motion is divided into very small time steps (Δt = 1-2 fs, 1fs = 1×10^{-15}s).

Potential and kinetic energy, temperature, and pressure

$$PE = V_{tot} = \sum_{bonds} V_r + \sum_{angles} V_q + \sum_{torsions} V_t + \sum_{atoms} V_{vanderWaals} + \sum_{atoms} V_{electrostatics}$$

$$KE = \frac{1}{2} m \langle v^2 \rangle \qquad KE = \frac{3}{2} kT$$

For an ideal gas, the temperature and pressure are functions of mean squared atomic velocity ($< v^2 >$), atomic mass (m), the Boltzmann constant (k) and number of particles in a unit volume (N):

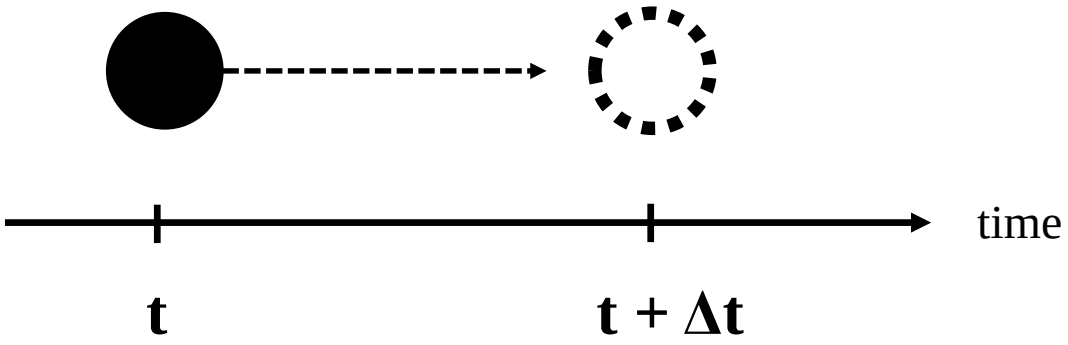


$$T = \frac{m}{3k} \langle v^2 \rangle \qquad P = \frac{1}{3} Nm \langle v^2 \rangle$$

Predicting Position

Question 1:

How to predict the position of a particle for the next step (t+Δt) ?



Predicting Position

Given the position (x) of a particle at time t, its new position after time Δt may be obtained by the familiar Taylor expansion:

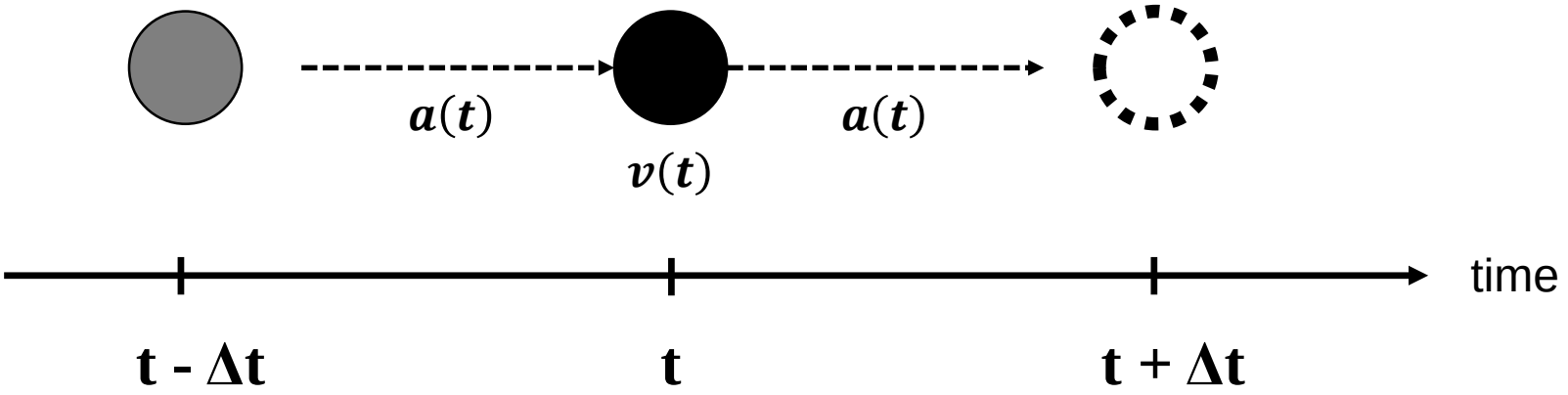
$$x(t+\Delta t) = x(t) + v(t)\Delta t + \frac{1}{2}a(t)\Delta t^2 + \ldots$$

Example:

- You are at home (x = ***home***) and walk **5 km/hr** (**v = 5 kph**) for **1 hr** (**Δt = 1 hr**), where are you now (***x(t+Δt)*** = *future position*)?
- This is a trick question – the answer depends on the direction in which you are walking.
- Velocity (v) and acceleration (a) have directions – that is, they are vector properties.
- So – if you new the velocity of an atom or molecule, as well as any accelerations on the particles, you could predict where it would be a short time later.
- This is the same principle employed when a GPS predicts your arrival time.

Verlet Algorithm

$$x(t-\Delta t) = x(t) - v(t)\Delta t + \frac{1}{2}a(t)\Delta t^2 \qquad x(t+\Delta t) = x(t) + v(t)\Delta t + \frac{1}{2}a(t)\Delta t^2$$



$$x(t+\Delta t) = \underset{1}{2x(t)} - \underset{2}{x(t-\Delta t)} + \underset{4}{a(t)}\underset{3}{\Delta t^2}$$

1. Current position
2. Previous position
3. Time step
4. Acceleration

Force and acceleration

Remember Newton's Second Law?

$$F_i = m_i a_i$$

The force on particle *i* equals its mass (*m*) multiplied by its acceleration (*a*)

So if we knew the force on an atom, we could determine its acceleration (we know the atomic masses):

$$a_i = \frac{F_i}{m_i}$$

Force and potential energy

Now, recall that the derivative of the potential energy of a particle (*V*) with respect to its position (*x*) is force (*F*). In energy minimization we used this to find the minimum or maximum of the potential energy surface.

Given the potential energy (*V*) of a group of atoms (force field):

$$V_{tot} = \sum_{bonds} V_r + \sum_{angles} V_\theta + \sum_{torsions} V_\tau + \sum_{atoms} V_{vanderWaals} + \sum_{atoms} V_{electrostatics}$$

If the atomic positions change, the force on the atoms can be directly determined:

$$F_i = -\frac{\partial V}{\partial x_i} \approx -\frac{\Delta V}{\Delta x_i}$$

Force = gradient of the potential energy

= slope of energy with respect to position

Force, potential energy, and acceleration

So putting it all together…

1) Given a force field, we can compute the way a molecule's energy depends on atomic positions.
2) From that dependency, we can compute the forces on each atom.
3) From the forces we can compute the atomic accelerations, and therefore the new positions.
4) From the new positions we can compute new forces.
5) Repeat this cycle until enough motion has been observed.

$$F_i = -\frac{\partial V}{\partial x_i} \approx -\frac{\Delta V}{\Delta x_i} = -\frac{V(t+\Delta t) - V(t)}{x_i(t+\Delta t) - x_i(t)} \approx m_i a_i$$

Summary of MD simulation

Given the position (x) of a particle at time t, its new position after time Δt is described by the familiar Taylor expansion:

$$x(t+\Delta t) = x(t) + v(t)\Delta t + \frac{1}{2}a(t)\Delta t^2 + \ldots$$

or:

$$x(t+\Delta t) = x(t) - x(t-\Delta t) + a(t)\Delta t^2$$

From the atomic forces (F_i) we can compute the atomic accelerations (a_i) from Newton's second law:

$$F_i = -\frac{\partial V}{\partial x_i} = m_i a_i$$

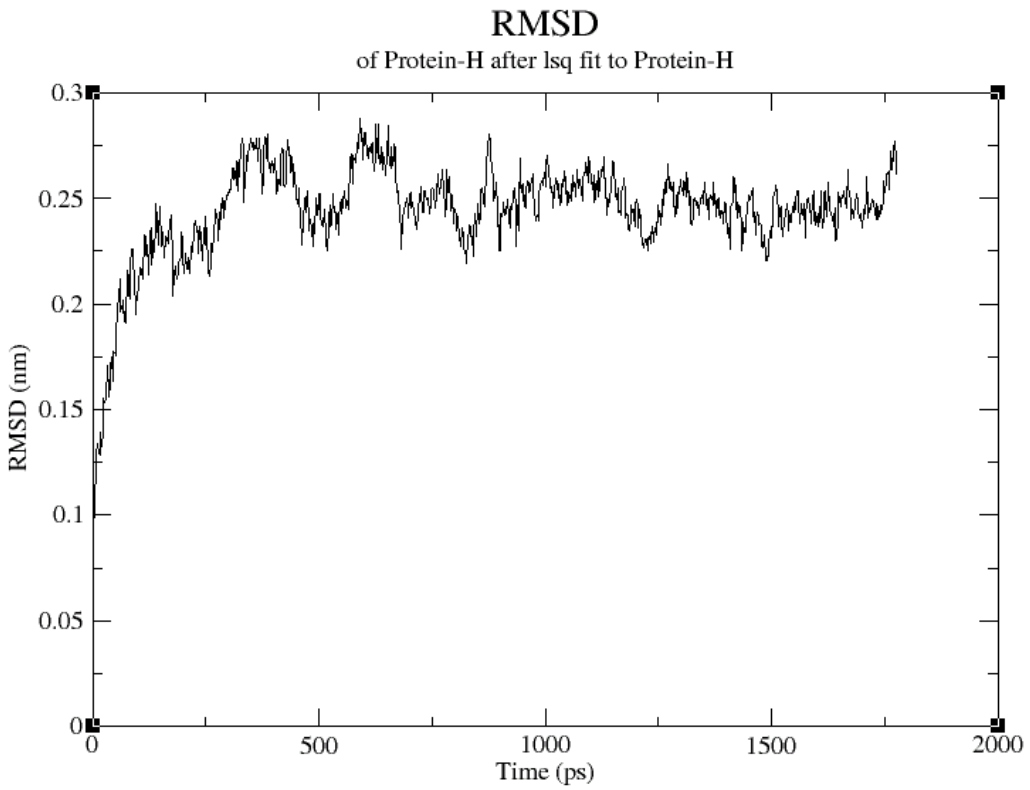
For which we need the potential energy (V), as defined by the force field:

$$V_{total} = \sum_{bonds} K_r (r - r_{eq})^2 + \sum_{angles} K_\theta (\theta - \theta_{eq})^2 + \sum_{dihedrals} \sum_n \frac{V_n}{2} \left[1 + \cos(n\phi - \gamma_n)\right]$$

$$+ \sum_{\substack{non-bonded \\ i<j}} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{C_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\varepsilon R_{ij}} \right]$$

How long do we have to run the simulation? When is enough, enough?

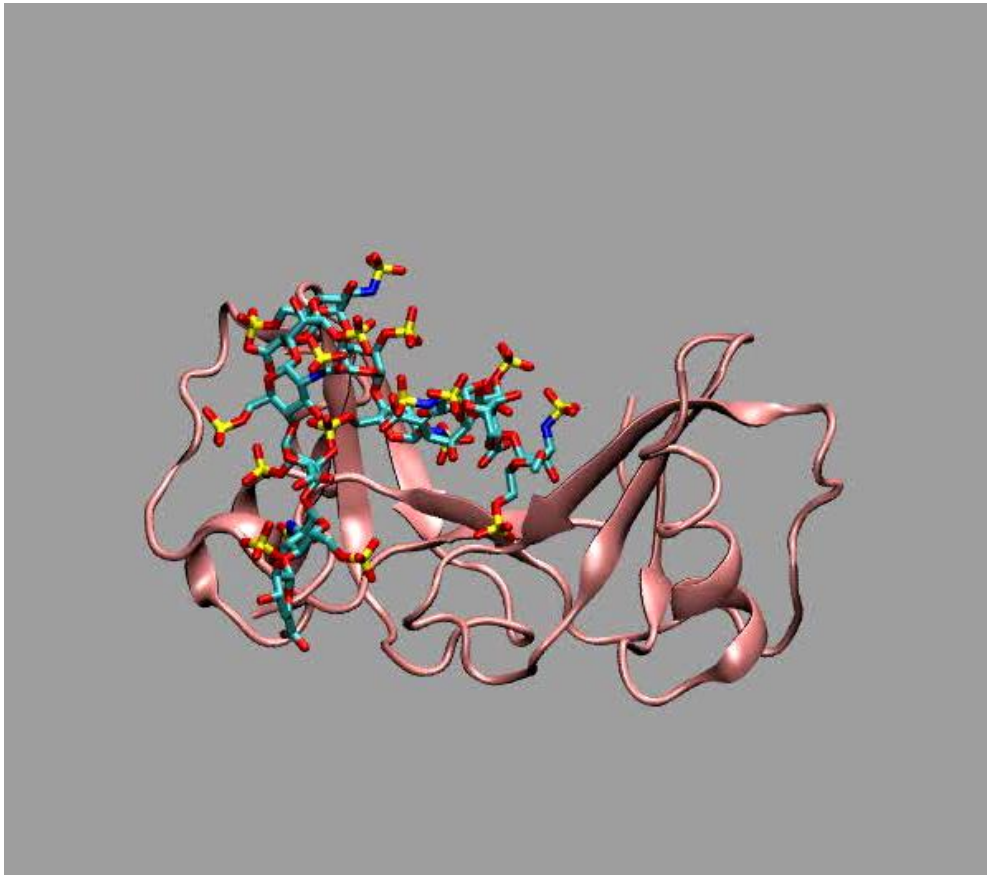
The MD simulation is continued until the average property that we are monitoring no longer changes. That is, the simulation is continued until convergence is reached.



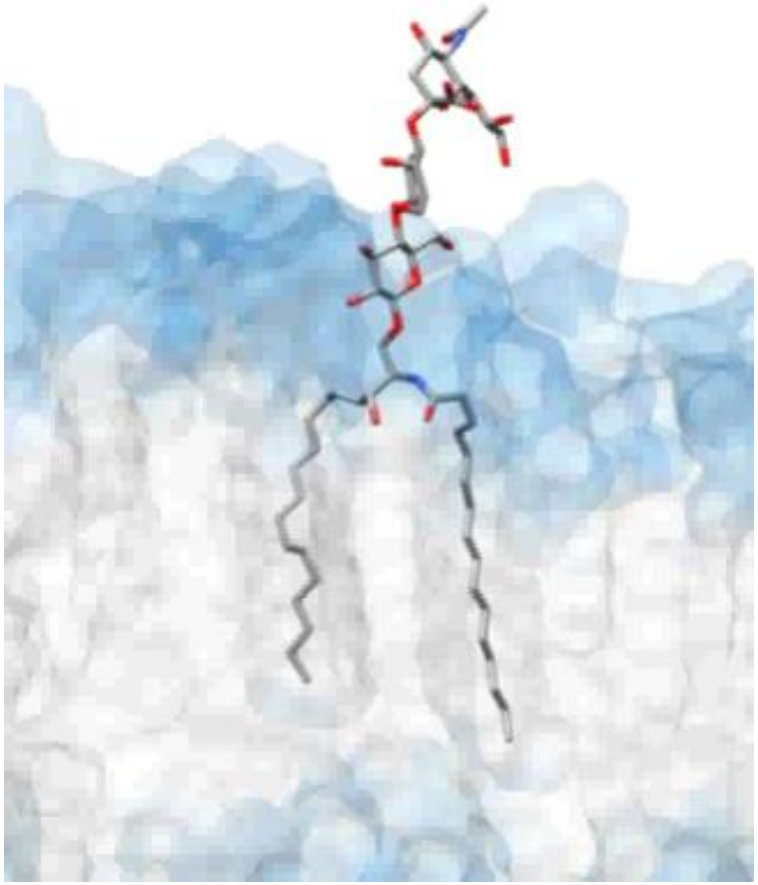
Root-mean-squared-deviation (RMSD) is a common way to monitor the change in atomic positions x_i with respect to a reference state x_o:

$$RMSD = \sqrt{\frac{\sum_{i=1}^{N} (x_i - x_o)^2}{N}}$$

Complex systems can be simulated

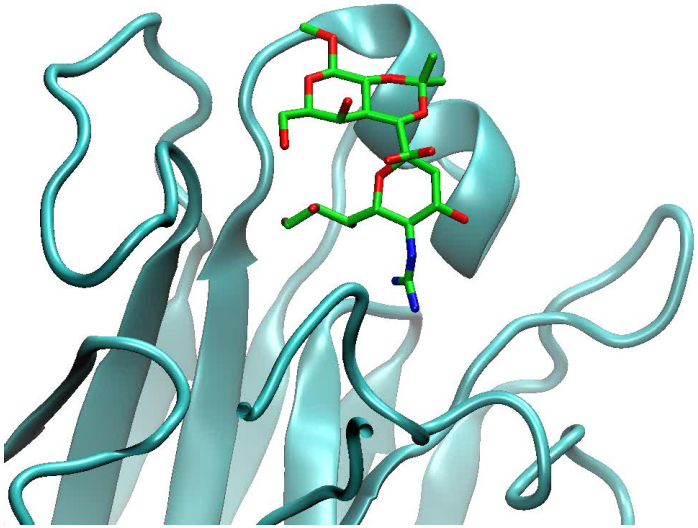


Protein-heparin complex

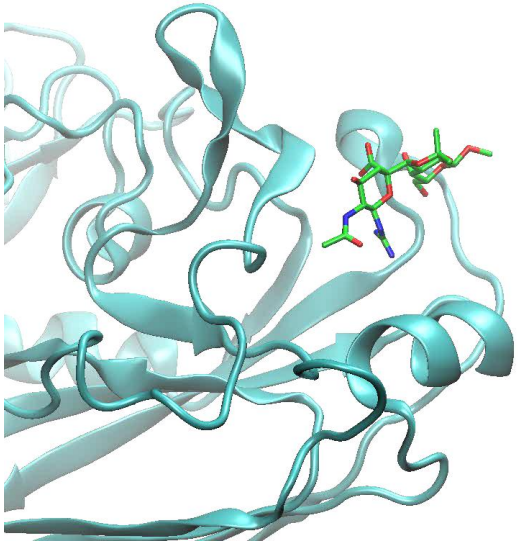


Glycolipid-Membrane Complex

Complex systems can be simulated



Putative Inhibitor (5-GAN) Complex 1

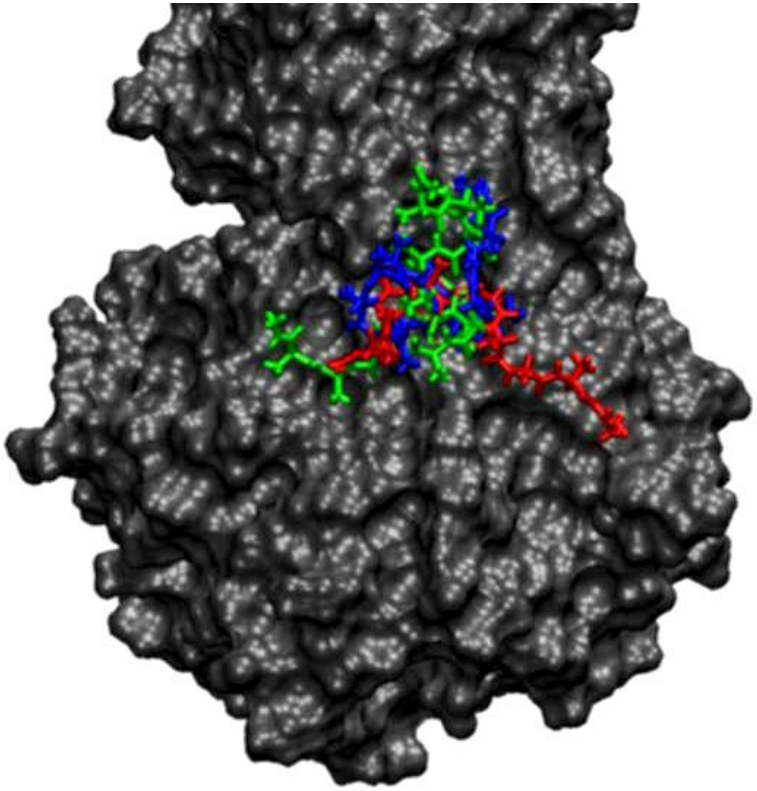


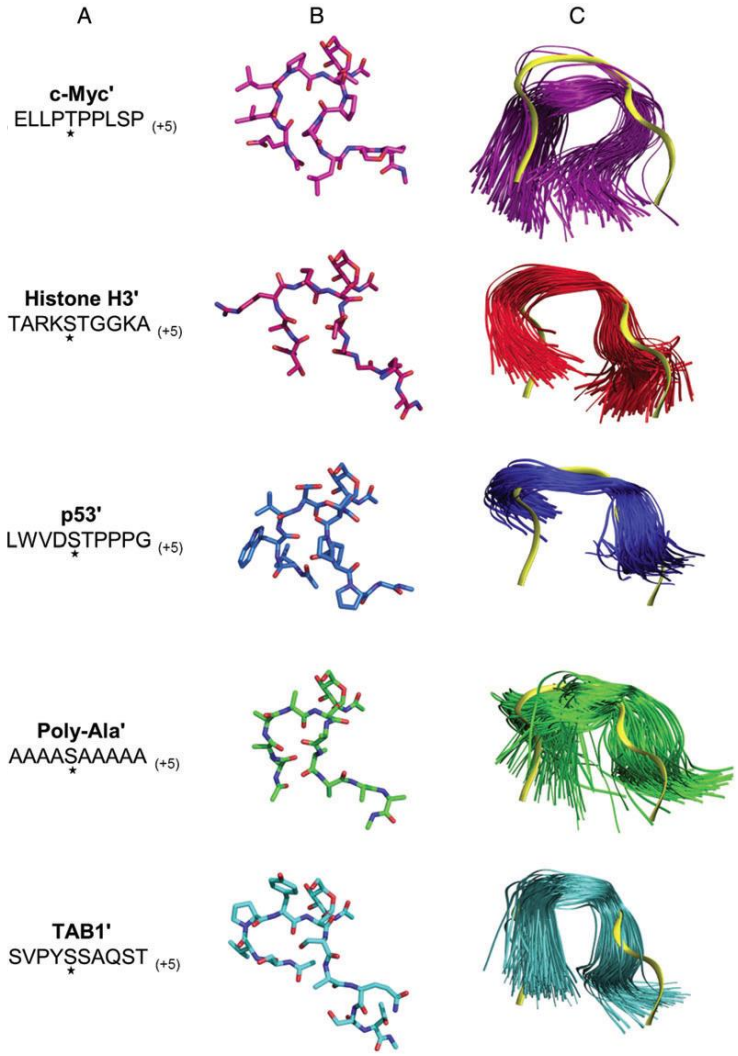
Putative Inhibitor (6-GAN) Complex 2

Representing Molecular Motion

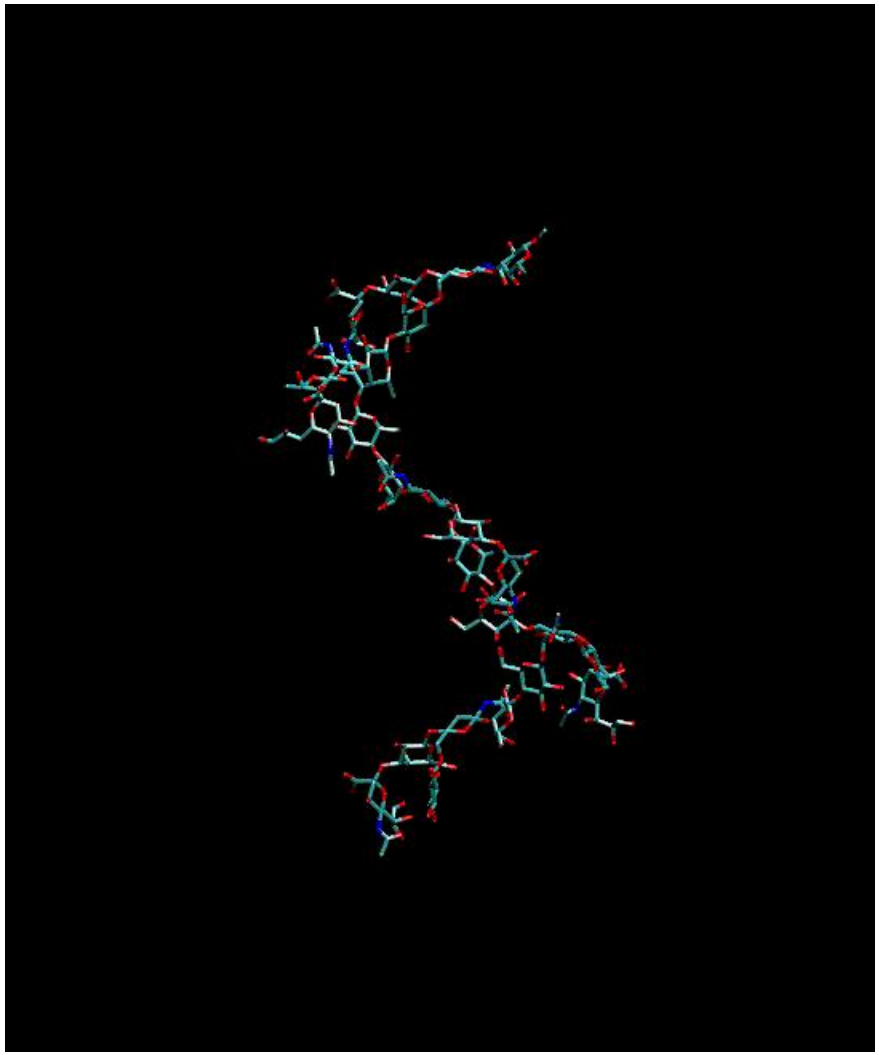
Glycobiology vol. 24 no. 1 pp. 85–96, 2014
doi:10.1093/glycob/cwt094
Advance Access publication on October 16, 2013

Defining the structural origin of the substrate sequence independence of O-GlcNAcase using a combination of molecular docking and dynamics simulation



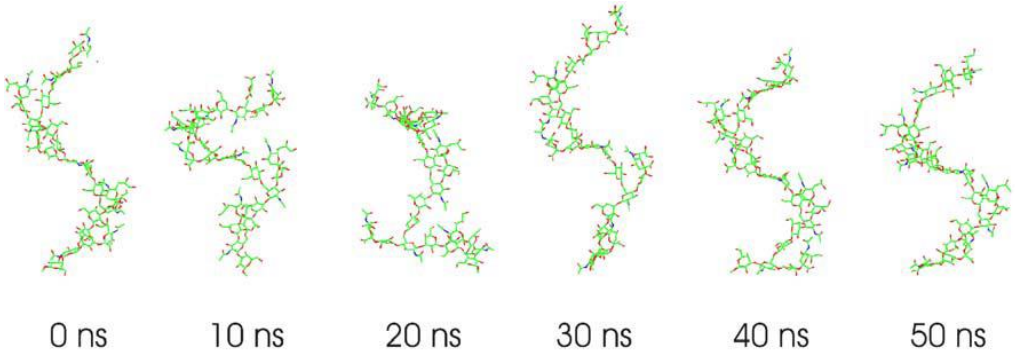


Visualize trajectory and quantify properties



What does the motion look like?

How can we describe it in a publication?



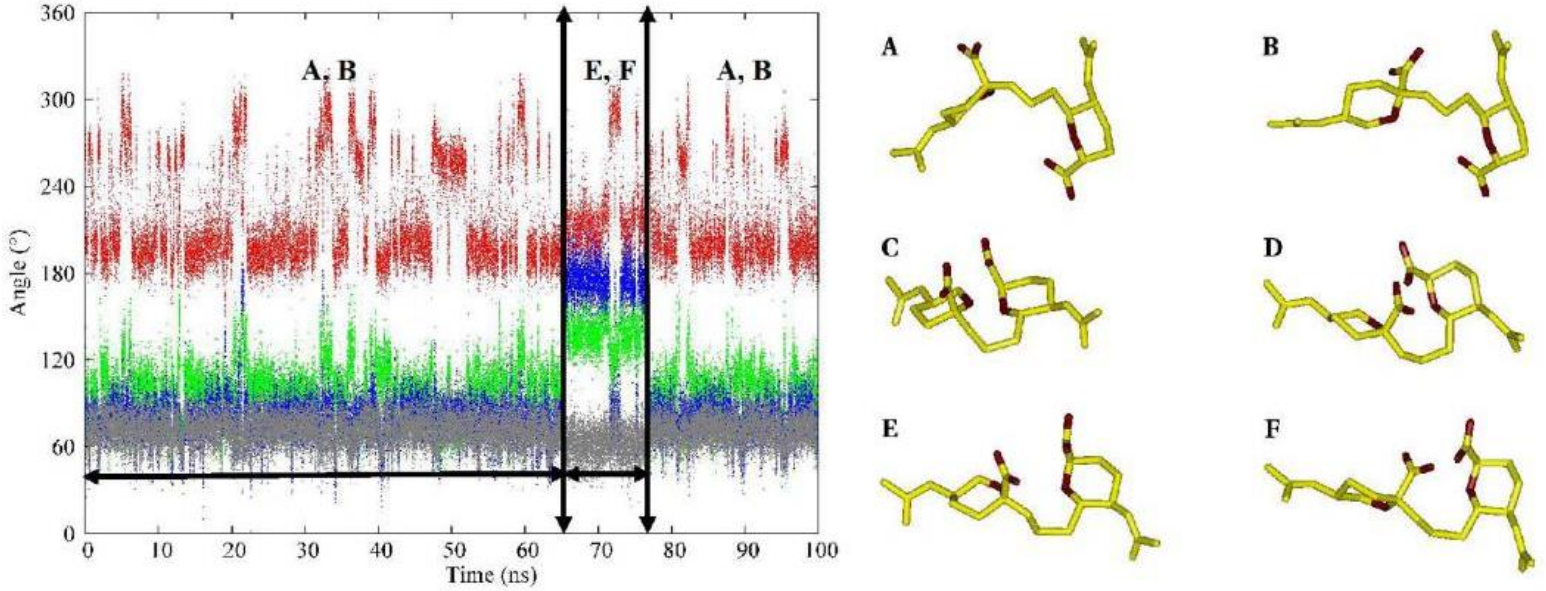
How can we confirm that it is real? Compute experimentally-observable properties, such as NMR coupling constants

ELSEVIER Carbohydrate Research 340 (2005) 1007–1018

Structural elucidation of type III group B *Streptococcus* capsular polysaccharide using molecular dynamics simulations: the role of sialic acid

How long do we have to run the simulation? When is enough, enough?

Monitoring torsion angles versus simulation time is useful for identifying conformational states, and for judging when conformational sampling has converged.



Biochemistry. 2008 November 25; 47(47): 12493–12514. doi:10.1021/bi800431c.

The conformational properties of methyl α-(2,8)-di/trisialosides and their *N*-acyl analogs: Implications for anti-*Neisseria meningitidis* B vaccine design

General ” Rules” for running MD simulations

- Include *Positive Controls* – to demonstrate the suitability of the force field/protocol, and to detect any methodological errors or instabilities or *PEBCAK** errors.
- Use a force field that has been shown to perform well for relevant systems.
- Use explicit water.
- Run for as long as necessary to ensure adequate sampling of the properties of interest.
- Look for trends rather than absolutes, hence run multiple simulations and/or simulations of multiple related systems.
- Look for sources of error in the simulation and in the experiment.
  - *Everyone trusts an experimental result, except the person who performed the experiment*
  - *No one trusts a computed result, except the person who did the calculation*

* PEBCAK: problem exists between chair and keyboard

Solvation

- The term “solvation” refers to the interaction of a solute with the solvent molecules surrounding it.

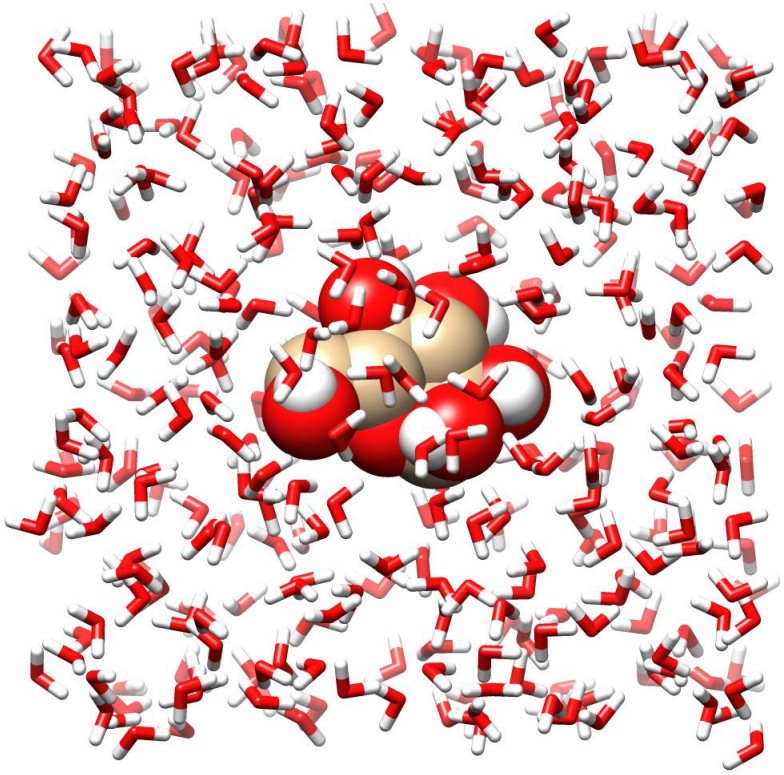
- There are several approaches to modeling solvent. Questions to consider:

  - How detailed does the solvent (water) model need to be?
    - Explicit or Implicit
    - Rigid or Flexible
    - Polarizable or Non-polarizable
    - Classical or Quantum Mechanical

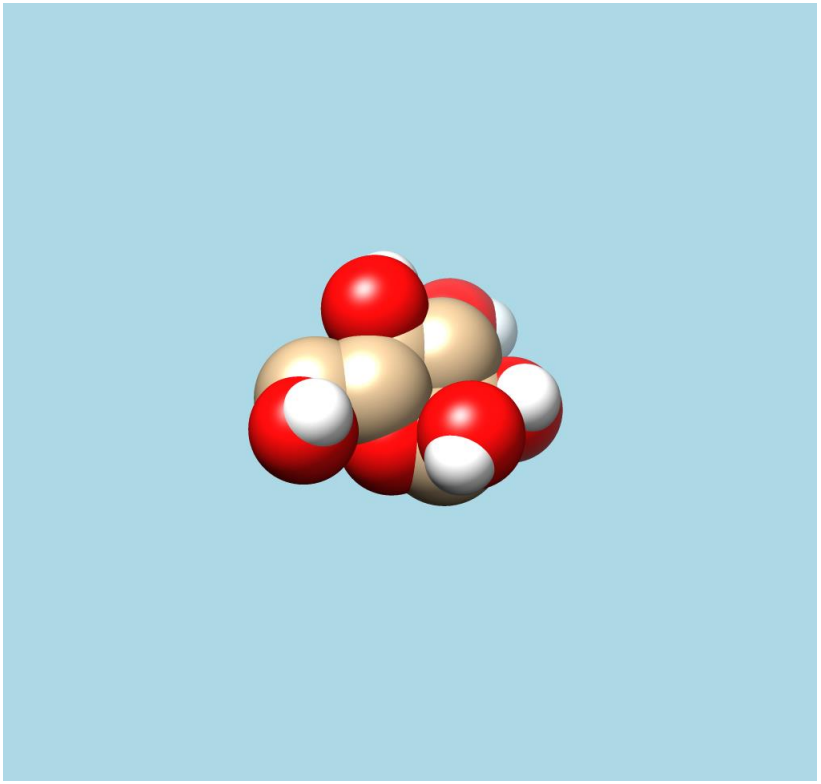
  - Is the solvent model compatible with the solute model? What does it mean to be “compatible”?

  - Is dynamics important?

Explicit versus implicit solvation

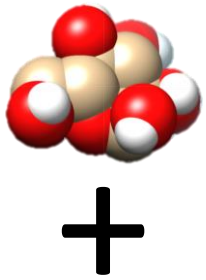


Every solvent (water) molecule is included explicitly, and all interactions between molecules are computed from the force field



Solvent is approximated as a uniform field around the solute (dielectric constant), and its influence on the solute is calculated from an implicit solvent model (generalized Born or Poisson-Boltzmann)

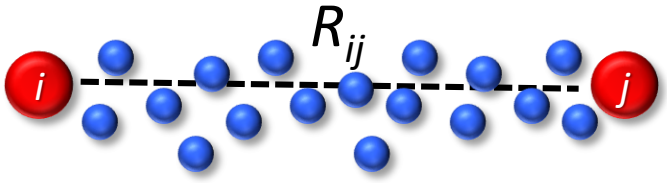
Thermodynamics of Solvation



$$\Delta G_{Solvation} = \Delta G_{Cavity} + \Delta G_{Ele} + \Delta G_{vdW}$$

Electrostatics and solvation

When other molecules exist between the charges, they may weaken the electrostatic interaction depending on the extent of their polarity. The more polar the intervening molecules the more they screen the two charges from each other.



This effect is very important for water because water is very polar.

Electrostatics and solvation

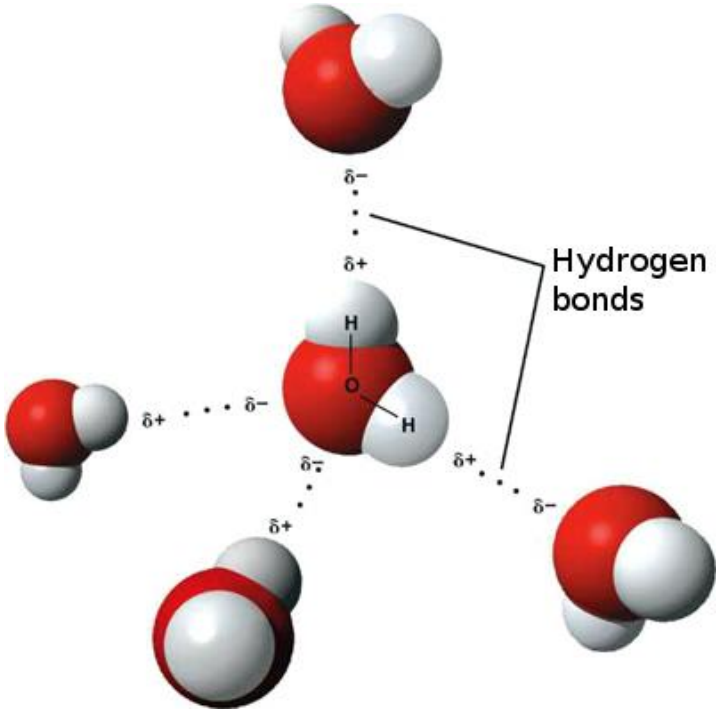
Charge screening can be explicitly modeled if individual waters are present, or it can be approximated implicitly using the dielectric constant of the solvent.

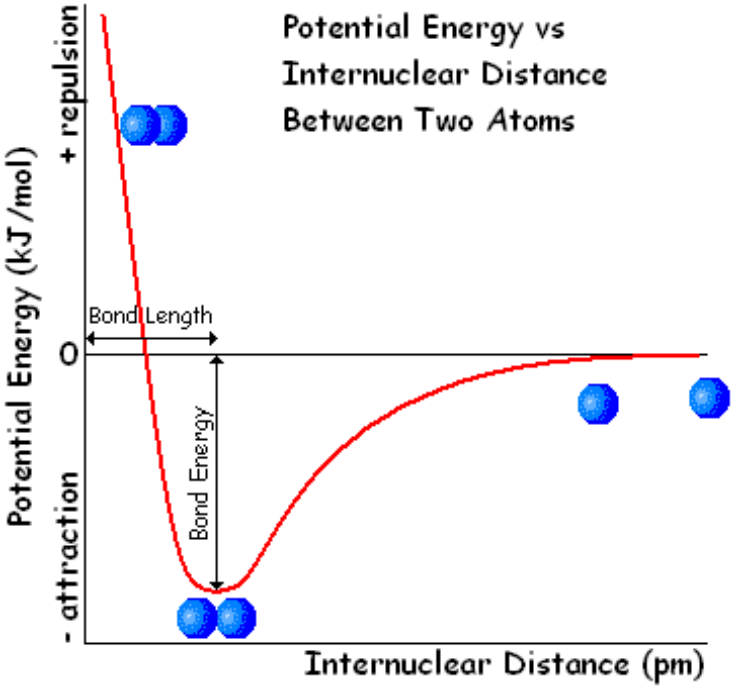
$$V_{Electrostatic} = \frac{q_i q_j}{\varepsilon_r R_{ij}}$$

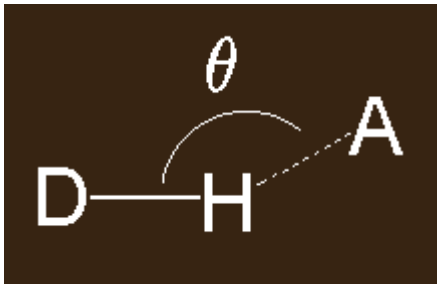
Here ε_r is the relative electric permitivity also known as the "dielectric constant".

For example, *water* (very polar, dipole = 1.85D) has a *dielectric constant* of 80 while n-butanol (less polar, dipole = 1.63D) has a *dielectric constant* of 18, at 20 °C

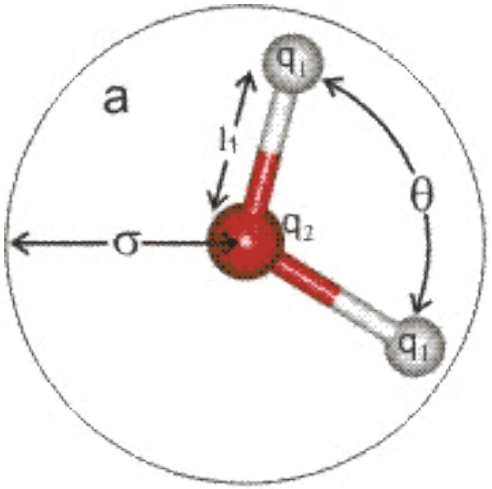
Water structure – directional hydrogen bonds





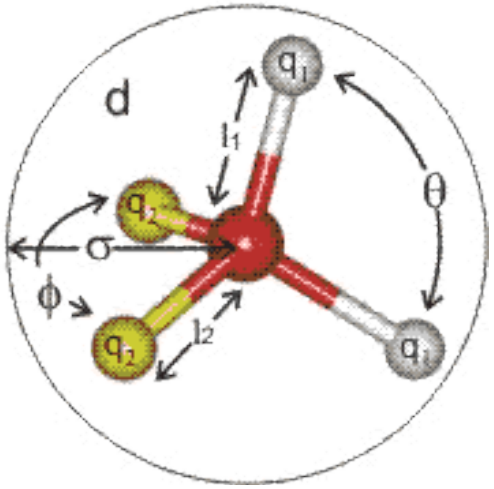


Common classical water models

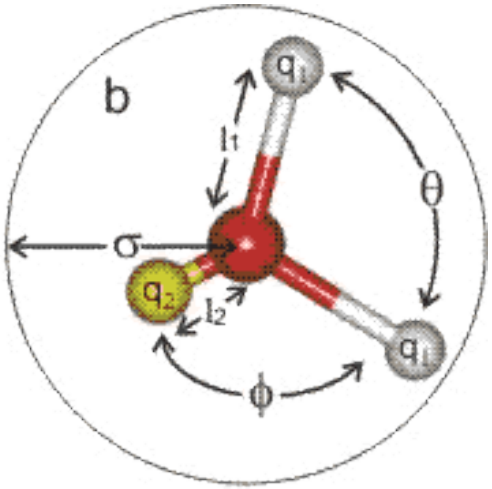


TIP3P (Jorgensen et al., 1983)
SPC (Berendsen et al., 1981)

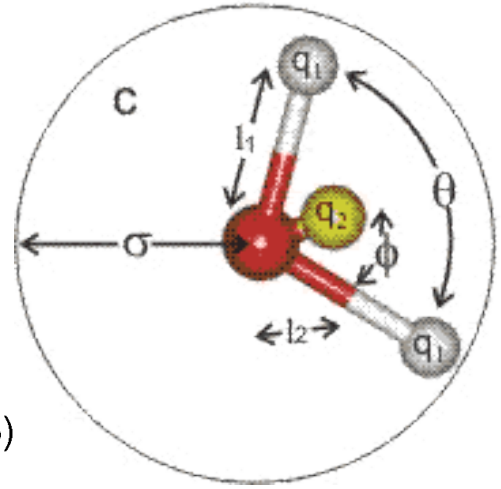
Each model performs differently at reproducing the properties of water



ST2 (Stillinger & Rahman, 1972)
TIP5P (Jorgensen et al., 2000)



BF (Bernal & Fowler, 1933)
TIP4P (Jorgensen et al., 1983)



Water model performance

Model	Type	σ Å [6]	ε kJ mol^{-1} [6]	l_1 Å	l_2 Å	q_1	q_2	θ °	ϕ °
SSD [511]	- [8]	3.016	15.319	-	-	-	-	109.47	109.47
SPC [94]	a	3.166	0.650	1.000(	-	+0.410	-0.8200	109.47	-
SPC/E [3]	a	3.166	0.650	1.000(	-	+0.4238	-0.8476	109.47	-
SPC/HW (D_2O) [220]	a	3.166	0.650	1.000(	-	+0.4350	-0.8700	109.47	-
TIP3P [180]	a	3.15061	0.6364	0.9572	-	+0.4170	-0.8340	104.52	-
PPC [1, 2] [3]	b	3.23400	0.6000	0.943(	0.06	+0.5170	-1.0340	106.0(	127.0(
TIP4P [180] [10]	c	3.15365	0.6480	0.9572	0.15	+0.5200	-1.0400	104.52	52.26
TIP4P-FQ [197]	c	3.15365	0.6480	0.9572	0.15	+0.63 [1]	-1.26 [1]	104.52	52.26
SWFLEX-AI [2] [201]	c	four terms used		0.968 [1]	0.14 [1,3]	+0.6213	-1.2459	102.7 [1]	51.35 [1]
COS/G3 [704] [11]	c	3.17459	0.9445	1.000(	0.15	+0.450672	-0.901344	109.47	-
TIP5P [180]	d	3.12000 [9]	0.6694 [9]	0.9572	0.70	+0.2410	-0.2410	104.52	109.47
POL5/TZ [2] [256]	d	2.9837 [4]	[4]	0.9572	0.5	varies [5]	-0.42188	104.52	109.47
Six-site [491]	c/d [7]	3.115_{OO} 0.673_{HH}	0.715_{OO} 0.115_{HH}	0.980	0.8892_{L} 0.230_{M}	+0.477	-0.044_{L} -0.866_{M}	108.0(	111.0(

Properties of water that may be parameterized

- Simple averages:
  - density (ρ)
  - enthalpy of vaporization (ΔHvap)
  - structure (O-O RDF)
  - **dipole moment**
- Higher-order averages:
  - heat capacity (C_p)
  - dielectric constant (ε)
  - isothermal compressibility (κ)
  - coefficient of thermal expansion (α)
- Kinetic observables:
  - diffusion constant (D)
  - rotational autocorrelation (τ_{rot})
  - hydrogen bond lifetimes (τ_{HB})

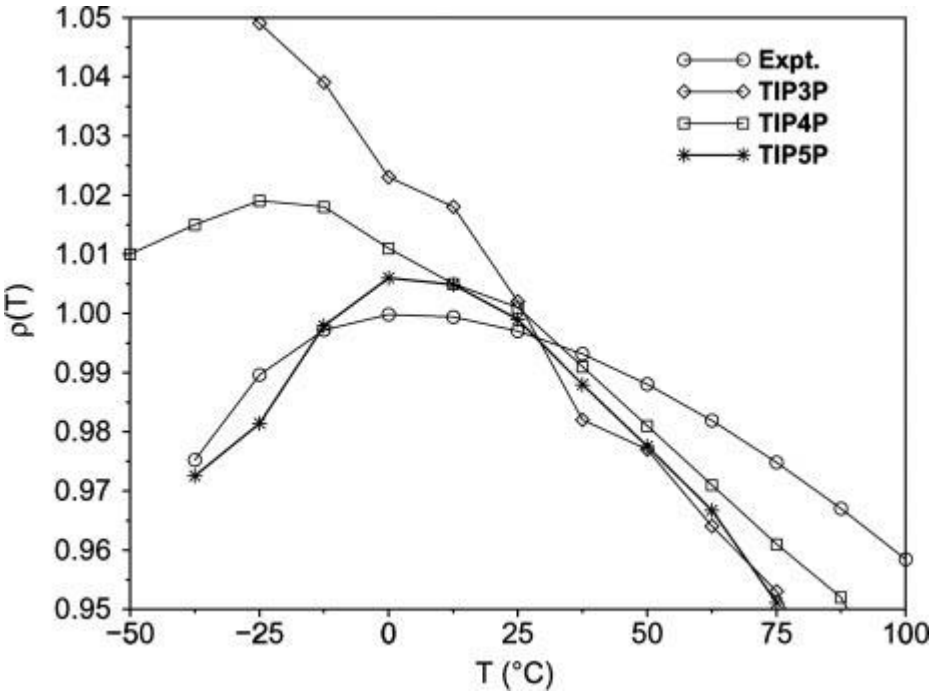
A particular challenge is getting the temperature dependence of these right, e.g., the density maximum of water around 4 C.

Properties of water that may be parameterized

Some of the calculated physical properties of the water models.

Model	Dipole moment	Dielectric constant	Self diffusion, 10^{-5} cm^2/s	Average configurational energy, kJ mol^{-1}	Density maximum, °C	Expansion coefficient, 10^{-4} $°C^{-1}$
SSD	2.35 [511]	72 [511]	2.13 [511]	-40.2 [511]	-13 [511]	-
SPC	2.27 [181]	65 [185]	3.85 [182]	-41.0 [185]	-	7.3 [704] **
SPC/E	2.35 [3]	71 [3]	2.49 [182]	-41.5 [3]	-38 [183]	-
PPC	2.52 [3]	77 [3]	2.6 [3]	-43.2 [3]	+4 [184]	-
TIP3P	2.35 [180]	82 [3]	5.19 [182]	-41.1 [180]	-13 [180]	9.2 [180]
TIP4P	2.18 [3,180]	53[a] [3]	3.29 [182]	-41.8 [180]	-25 [180]	4.4 [180]
TIP4P-FQ	2.64 [197]	79 [197]	1.93 [197]	-41.4 [201]	+7 [197]	-
SWFLEX-AI	2.69 [201]	116 [201]	3.66 [201]	-41.7 [201]	-	-
COS/G3 **	2.57 [704]	88 [704]	2.6 [704]	-41.1 [704]	-	7.0 [704]
TIP5P	2.29 [180]	81.5 [180]	2.62 [182]	-41.3 [180]	+4 [180]	6.3 [180]
POL5/TZ	2.712 [256]	98 [256]	1.81 [256]	-41.5 [256]	+25 [256]	-
Six-site*	1.89 [491]	33 [491]	-	-	+14 [491]	2.4 [491]
Expt.	2.65, 3.0	78.4	2.30	-41.5 [180]	+3.984	2.53

Water density versus temperature



Computed and experimental results for the density (g/cm^3) of liquid water vs. temperature at 1 atm.

All models display the correct density at room temperature, but only TIP5P shows correct density behavior over a range of temperatures.

Does this matter? It depends on the conditions that you want to simulate

Water density versus temperature

Example Publications:

Effects of water models on Biomolecular Structure

1) **Explicit Water Models Affect the Specific Solvation and Dynamics of Unfolded Peptides While the Conformational Behavior and Flexibility of Folded Peptides Remain Intact**

   Petra Florov, Petr Sklenovsk, Pavel Ban, and Michal Otyepka

   J. Chem. Theory Comput., **2010**, *6*, 3569–3579

2) **On the role of water models in quantifying the standard binding free energy of highly conserved water molecules in proteins: the case of Concanavalin A.**

   Elisa Fadda and Robert J. Woods

   J. Chem. Theory Comput., **2011**, *7*, 3391–3398.