BIOC404 Biochemical Methods

Introduction to Biomolecular Simulations

Dr. Erich Kuechler

erich.kuechler@msl.ubc.ca

My Office Doesn't Matter Anymore

Introduction to biomolecular simulation

Lecture 1:

Molecular Dynamics & Ensembles

Lecture 2:

Force Fields & Solvation Models

Lecture 3:

Enhanced Sampling Methods

Assignment:

Analyze simulation output and relate the findings to concepts introduced by the course.

Introduction to biomolecular simulation: Lecture 2



Part 1: Introduction to force fields

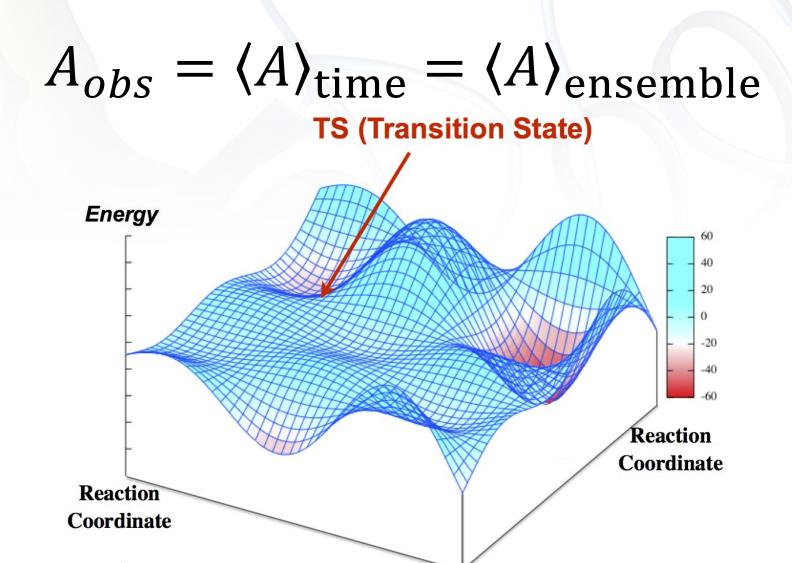
Part 2: Common force field terms

Part 3: Explicit and implicit solvation

Introduction to biomolecular simulation

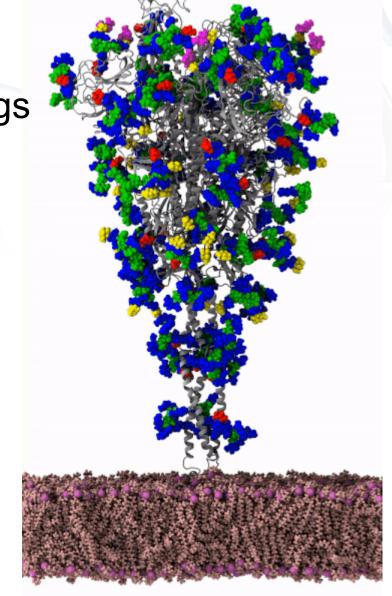
Lecture 1:

Molecular Dynamics & Ensembles



"everything that living things do can be understood in terms of the jigglings and wigglings of atoms." – Richard P. Feynman

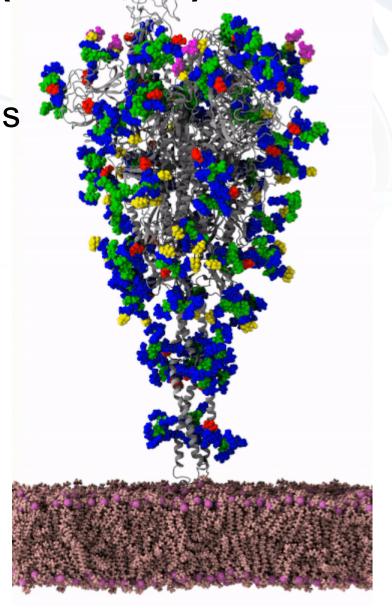
$$\boldsymbol{F}_i = \dot{\boldsymbol{p}}_i - \frac{\partial U(\boldsymbol{r}^N)}{\partial \boldsymbol{r}_i}$$



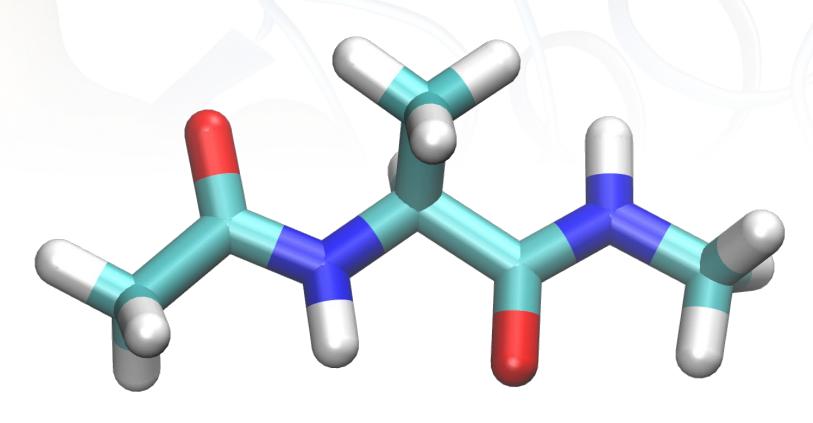
"everything that living things do can be understood in terms of the jigglings and wigglings of atoms." – Richard P. Feynman

$$\boldsymbol{F}_i = \dot{\boldsymbol{p}}_i - \frac{\partial U(\boldsymbol{r}^N)}{\partial \boldsymbol{r}_i}$$

Force fields are a mathematical expression to define the potential energy surface in a simulation.



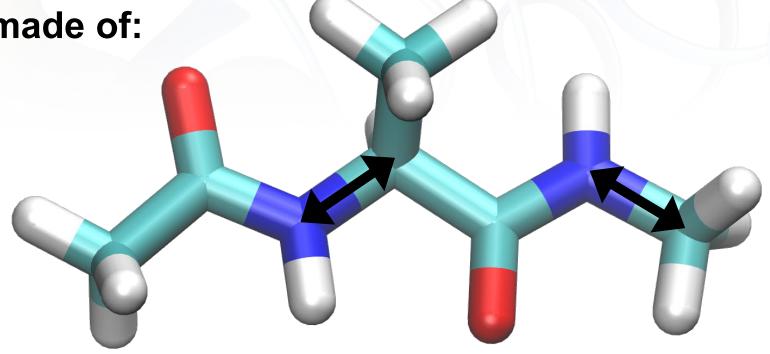
Force fields are a mathematical expression to define the potential energy surface in a simulation.



Force fields are a mathematical expression to define the potential energy surface in a simulation.

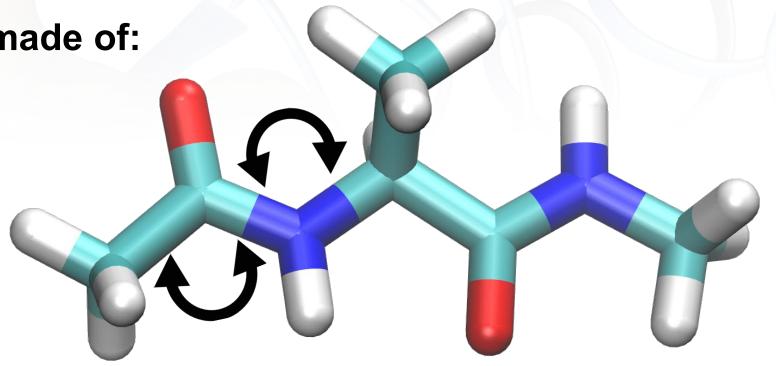
In MD, FFs are generally made of:

Bond



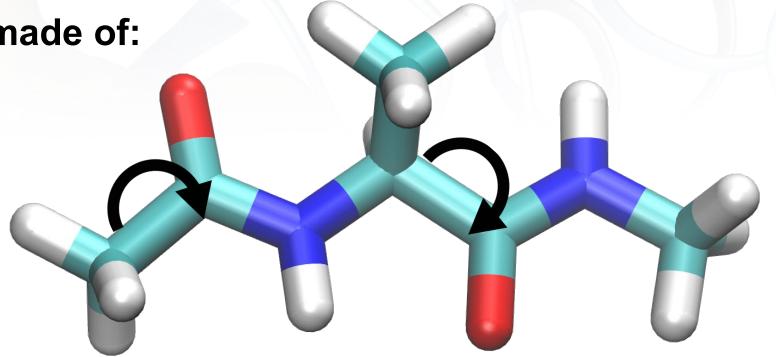
Force fields are a mathematical expression to define the potential energy surface in a simulation.

- Bond
- Angle



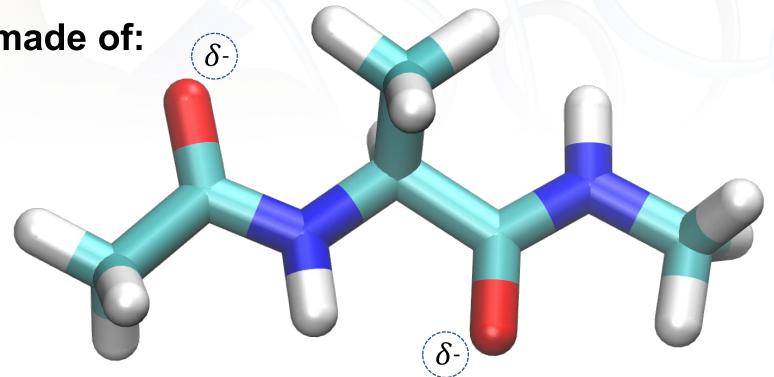
Force fields are a mathematical expression to define the potential energy surface in a simulation.

- Bond
- Angle
- Torsion



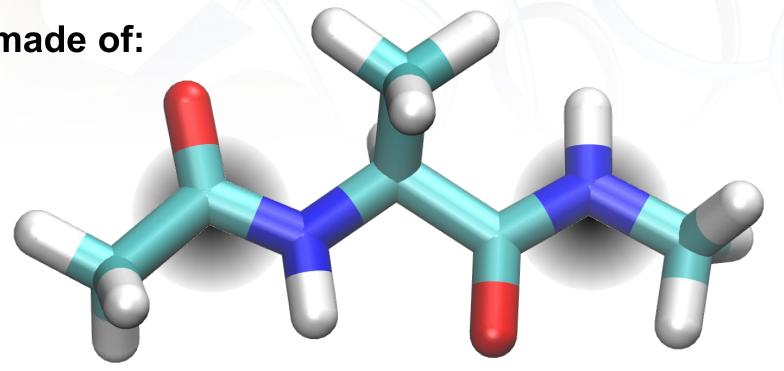
Force fields are a mathematical expression to define the potential energy surface in a simulation.

- Bond
- Angle
- Torsion
- Charge



Force fields are a mathematical expression to define the potential energy surface in a simulation.

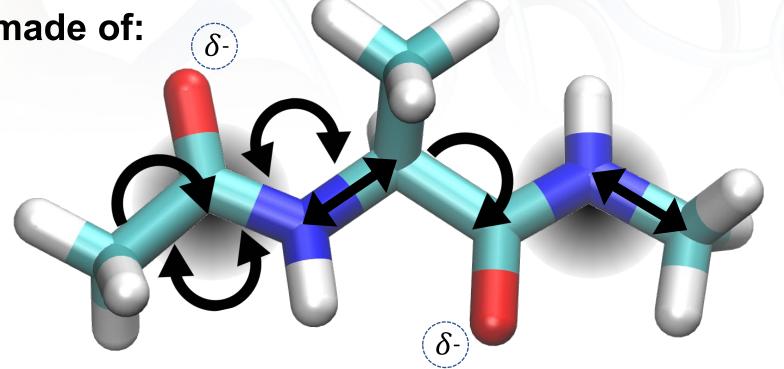
- Bond
- Angle
- Torsion
- Charge
- van der Waals



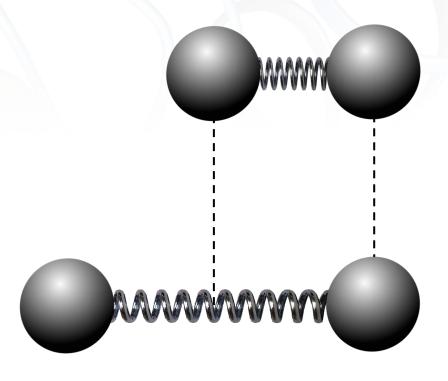
Force fields are a mathematical expression to define the potential energy surface in a simulation.

In MD, FFs are generally made of:

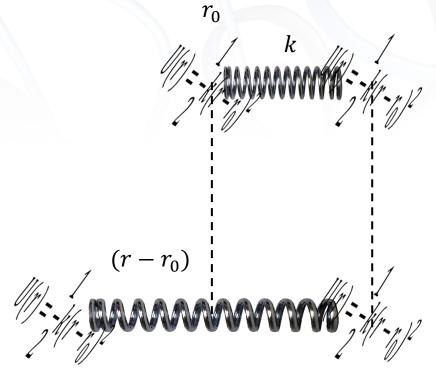
- Bond
- Angle
- Torsion
- Charge
- van der Waals



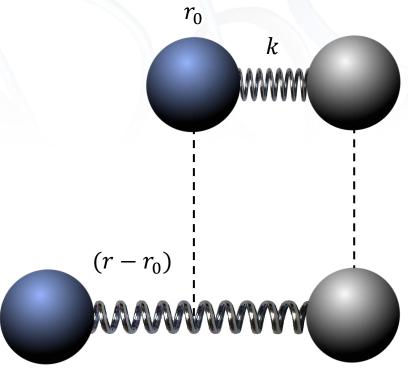
Force field parameters depend on local electronic environment and might be tuned for specific properties. Choice of force field can be extremely important!

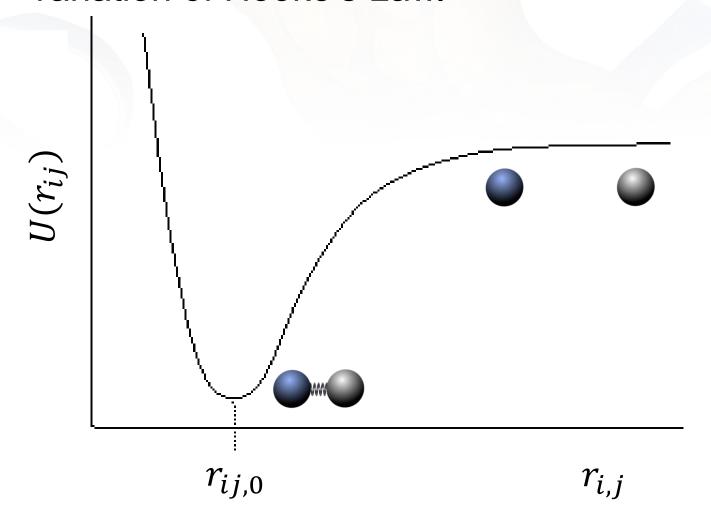


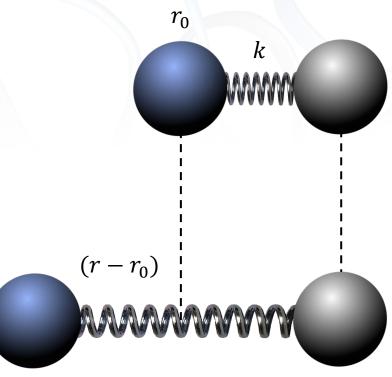
$$U(r) = \frac{1}{2}k(r - r_0)^2$$

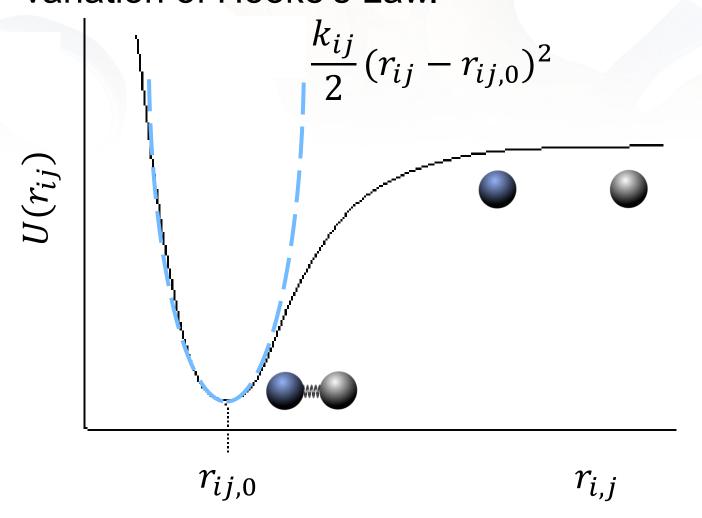


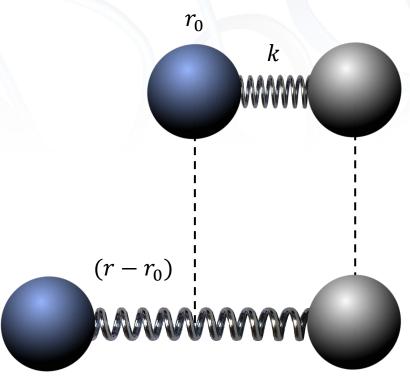
$$U(r_{ij}) = \frac{1}{2} k_{ij} (r_{ij} - r_{ij,0})^2$$

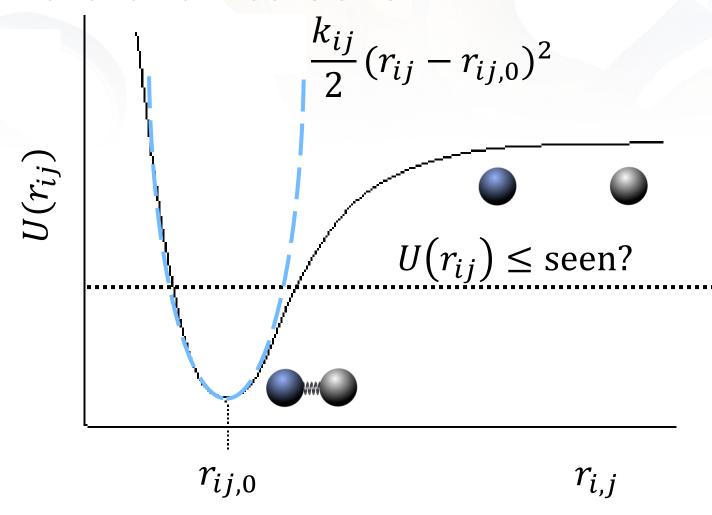


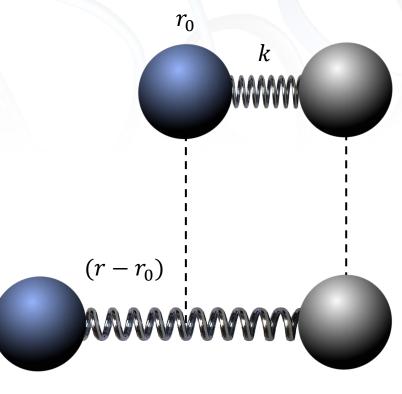




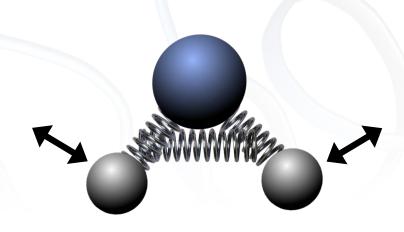






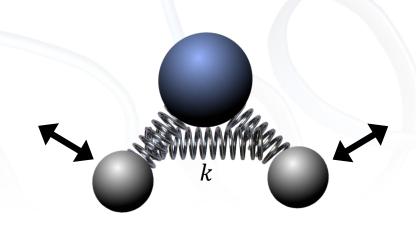


Bond angles are also expressed similarly.



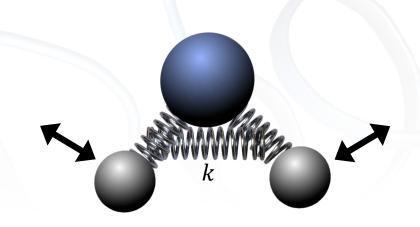
Bond angles are also expressed similarly.

$$U(r) = \frac{1}{2}k(\theta - \theta_0)^2$$



Bond angles are also expressed similarly.

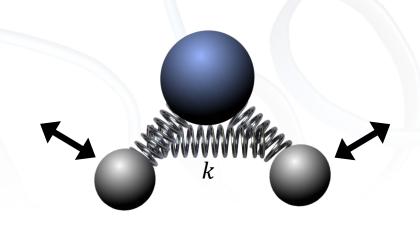
$$U(r) = \frac{1}{2}k(\theta - \theta_0)^2$$



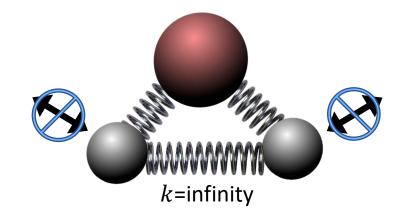
Bond distances and angles are sometimes called 'hard' degrees of freedom as compared to the remaining terms.

Bond angles are also expressed similarly.

$$U(r) = \frac{1}{2}k(\theta - \theta_0)^2$$

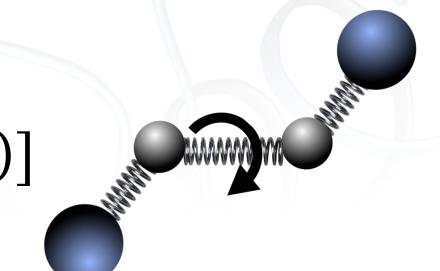


Special case: water



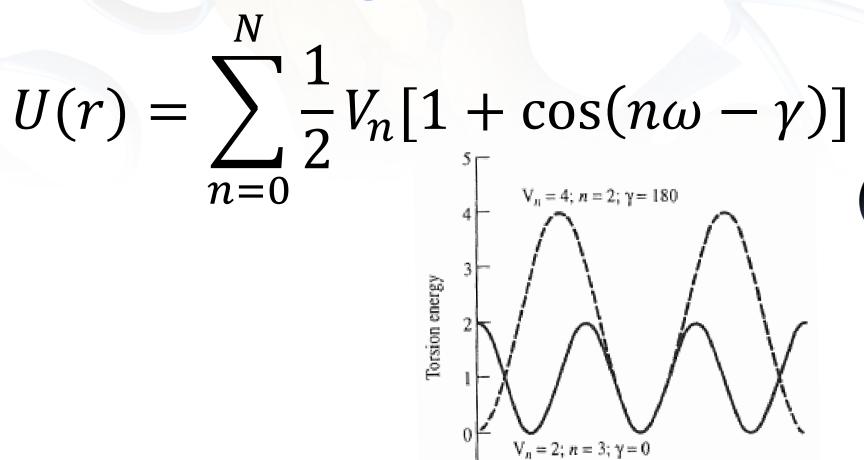
So are dihedral angles

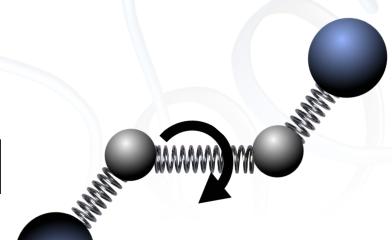
$$U(r) = \sum_{n=0}^{N} \frac{1}{2} V_n [1 + \cos(n\omega - \gamma)]$$



Torsion angle

So are dihedral angles



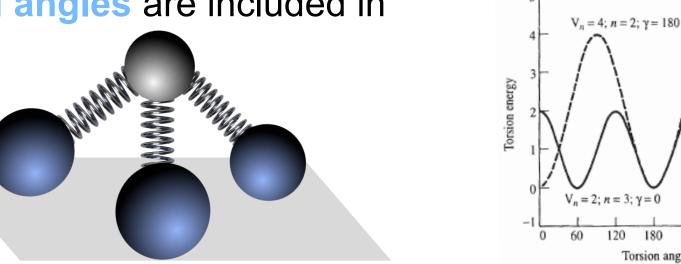


So are dihedral angles

$$U(r) = \sum_{n=0}^{N} \frac{1}{2} V_n [1 + \cos(n\omega - \gamma)]$$

Also, improper dihedral angles are included in

force fields

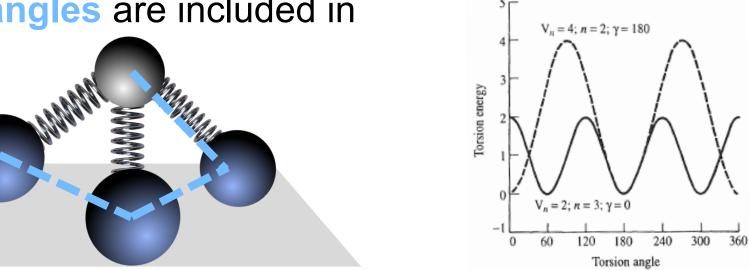


So are dihedral angles

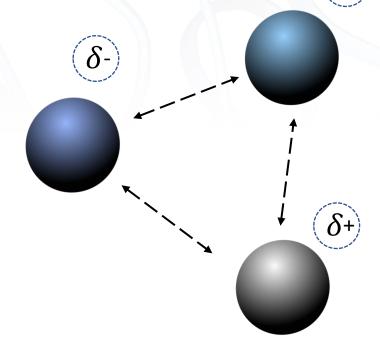
$$U(r) = \sum_{n=0}^{N} \frac{1}{2} V_n [1 + \cos(n\omega - \gamma)]$$

Also, improper dihedral angles are included in

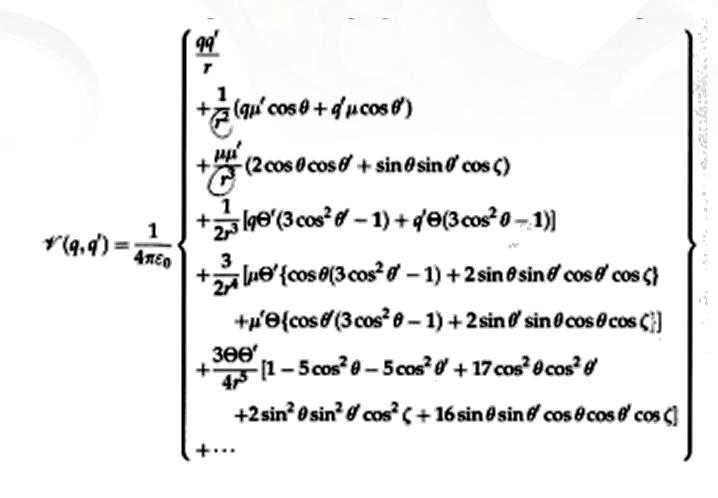
force fields

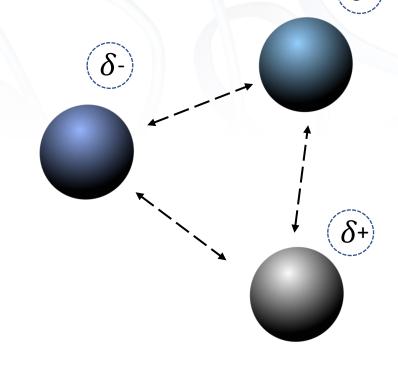


Electrostatic interactions are *most commonly* expressed through some variation of Coulomb's Law. δ



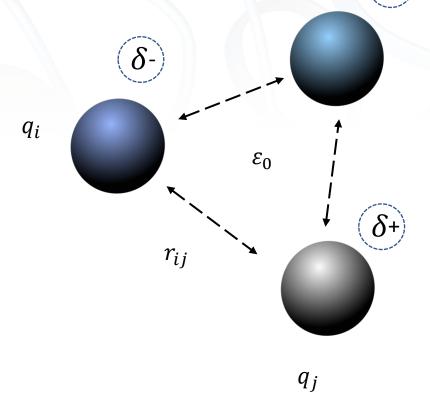
Electrostatic interactions are *most commonly* expressed through some variation of Coulomb's Law. δ





Electrostatic interactions are *most commonly* expressed through some variation of Coulomb's Law. δ

$$U(r_{ij}) = \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}$$



Electrostatic interactions are most commonly expressed through some variation of Coulomb's Law. δ

$$U(r_{ij}) = \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}$$

Note: Physics would demand that force fields include many-bodied interactions

$$\delta$$
-
 δ -
 δ +

$$U_{\text{tot}}(r_{(i,j,k)}) = U(r_{ij}) + U(r_{ik}) + U(r_{jk}) + U(r_{ijk})$$

Electrostatic interactions are *most commonly* expressed through some variation of Coulomb's Law. δ

$$U(r_{ij}) = \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}$$

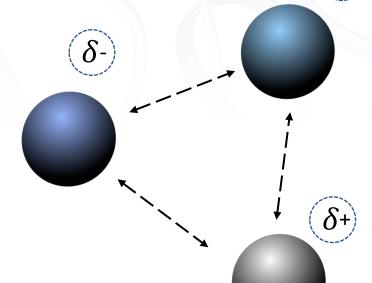
Assuming that many-bodied interactions are small and can be subsumed makes an **effective** pairwise-potential of the potential energy surface.

$$U_{\text{tot}}(r_{(i,j,k)}) \approx U(r_{ij}) + U(r_{ik}) + U(r_{jk})$$

 δ +)

Electrostatic interactions are most commonly expressed through some variation of Coulomb's Law. δ

$$U(r_{ij}) = \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}$$



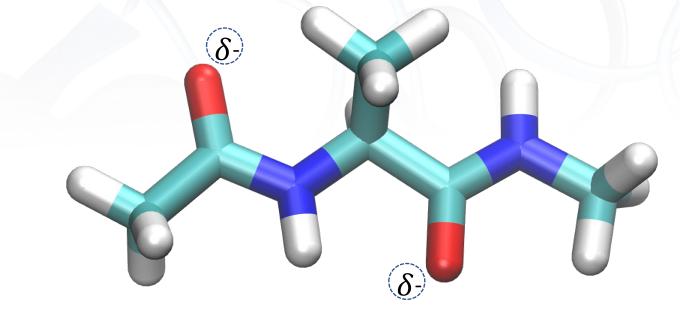
Approximation can have issues:

- Multipole moments are very hard (off-atom charges)
- Not good at very short ranges, no polarization
- Not good for very long ranges, creates more issues
- Condensed phase systems are very sensitive to electrostatics

Electrostatic interactions are most commonly expressed through

some variation of Coulomb's Law.

$$U(r_{ij}) = \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}$$



There are a number of ways to calculate charges but *most commonly* derived from quantum chemical calculation.

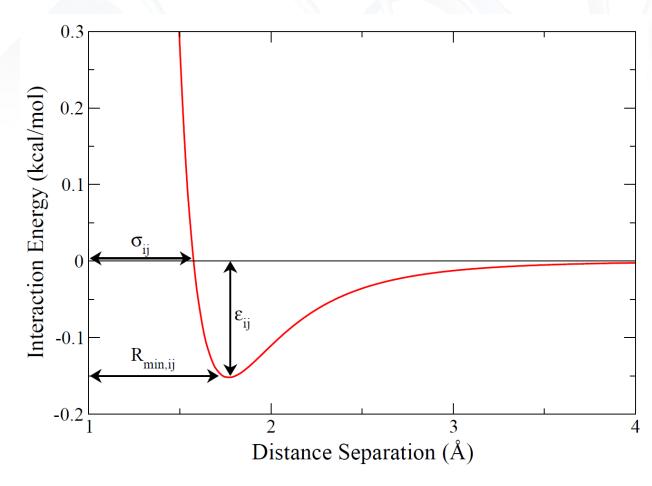
Election dispersion and correlation interactions, commonly called van der Waals interactions, are *most commonly* expressed through the Lennard-Jones (12-6) potential.

$$U(r_{ij}) = 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right]$$

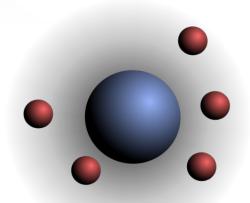
Election dispersion and correlation interactions, commonly called van der Waals interactions, are most commonly expressed through

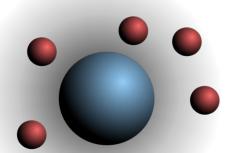
the Lennard-Jones (12-6) potential.

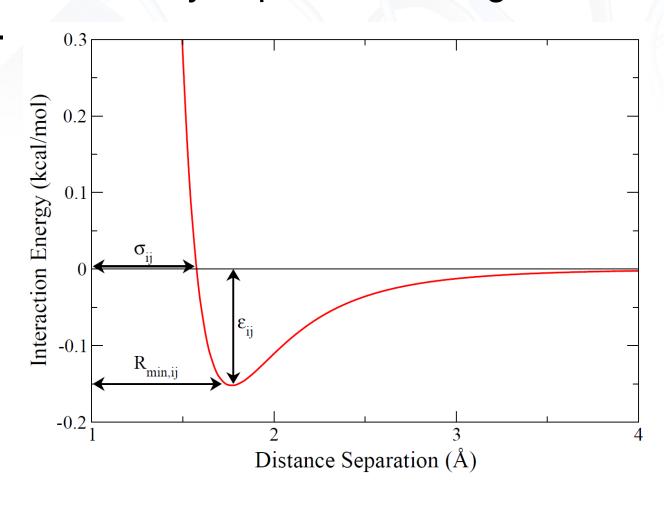
$$U(r_{ij}) = 4 arepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right]^{\frac{0.3}{60.1}}$$



Election dispersion and correlation interactions, commonly called van der Waals interactions, are *most commonly* expressed through

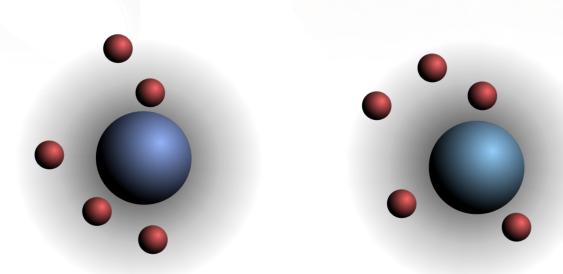


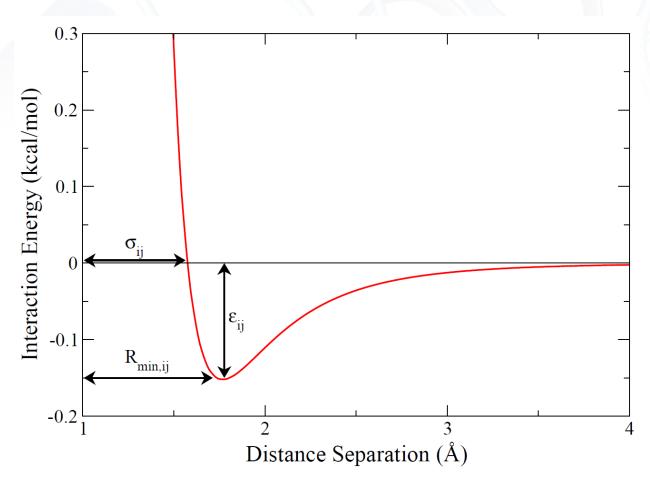




Election dispersion and correlation interactions, commonly called

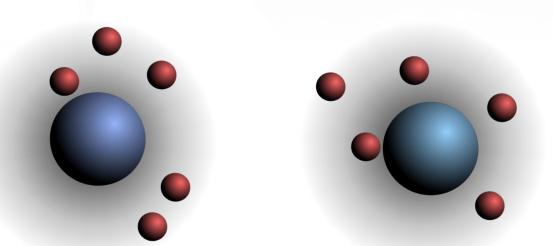
van der Waals interactions, are most commonly expressed through

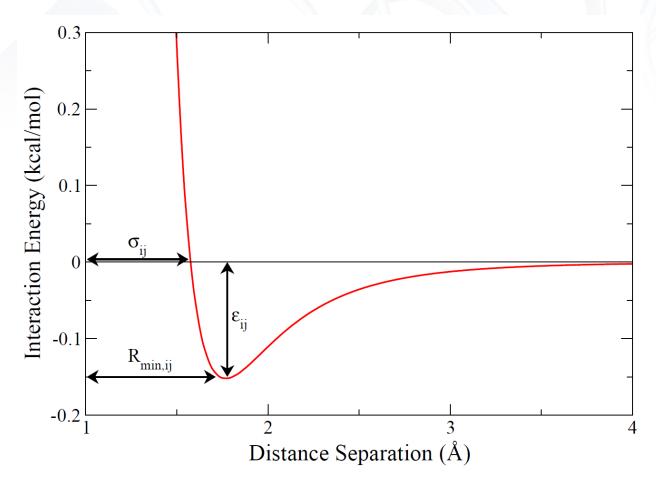




Election dispersion and correlation interactions, commonly called van der Waals interactions, are most commonly expressed through

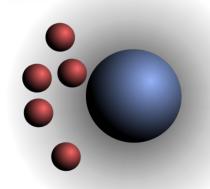
van der Waals interactions, are most commonly expressed through

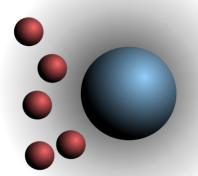


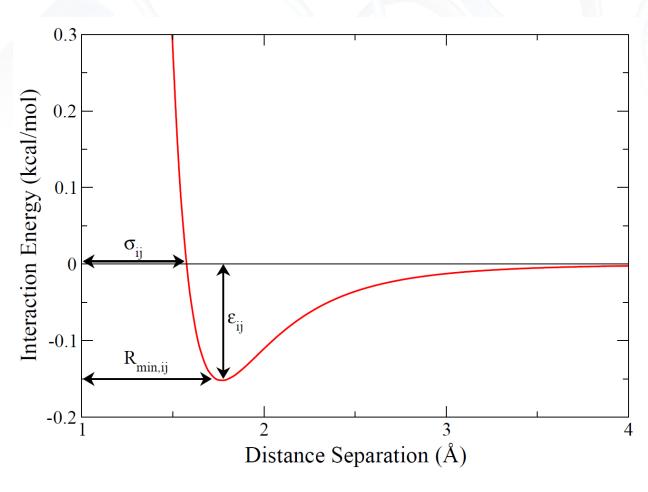


Election dispersion and correlation interactions, commonly called

van der Waals interactions, are most commonly expressed through

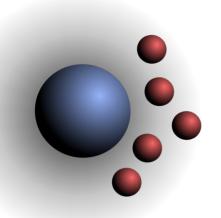


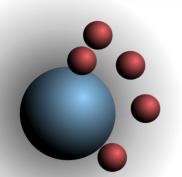


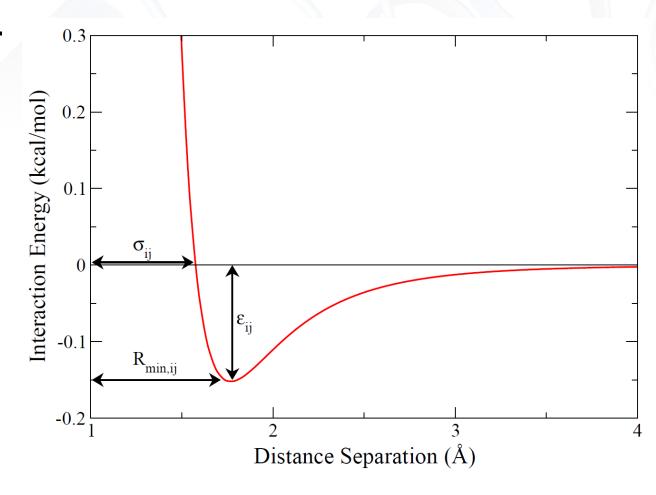


Election dispersion and correlation interactions, commonly called

van der Waals interactions, are most commonly expressed through

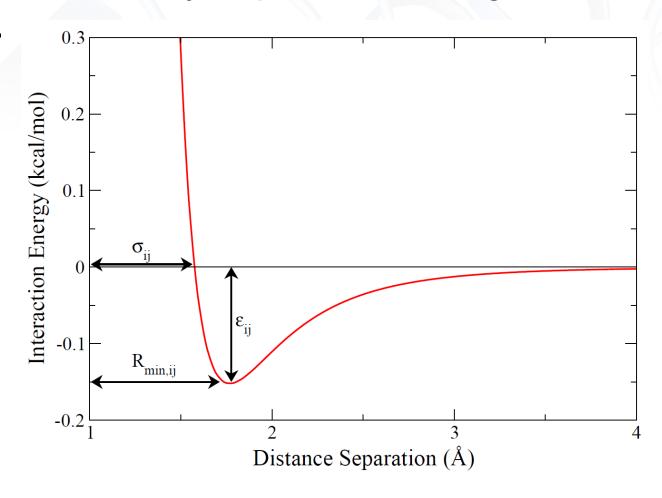






Election dispersion and correlation interactions, commonly called van der Waals interactions, are most commonly expressed through

$$U(r_{ij}) = 4 arepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right]^{\frac{0.3}{60.1}}$$



Adding it all together:

$$U(r^{N})$$

$$= \sum_{bonds} \frac{1}{2} k(r - r_{0})^{2} + \sum_{angles} \frac{1}{2} k(\theta - \theta_{0})^{2}$$

$$+ \sum_{torsions} \frac{1}{2} V_{n} [1 + \cos(n\omega - \gamma)]$$

$$+ \sum_{i=1}^{N} \sum_{j>i}^{N} \left\{ \frac{q_{i}q_{j}}{4\pi\varepsilon_{0}r_{ij}} + 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] \right\}$$

Adding it all together:

$$U(r^{N})$$

$$= \sum_{bonds} \frac{1}{2}k(r - r_{0})^{2} + \sum_{angles} \frac{1}{2}k(\theta - \theta_{0})^{2}$$

$$+ \sum_{torsions} \frac{1}{2}V_{n}[1 + \cos(n\omega - \gamma)]$$

$$+\sum_{i=1}^{N}\sum_{j>i}^{N}\left\{\frac{q_{i}q_{j}}{4\pi\varepsilon_{0}r_{ij}}+4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12}-\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right]\right\}$$

Non-bonded terms

Adding it all together:

$$U(r^{N})$$

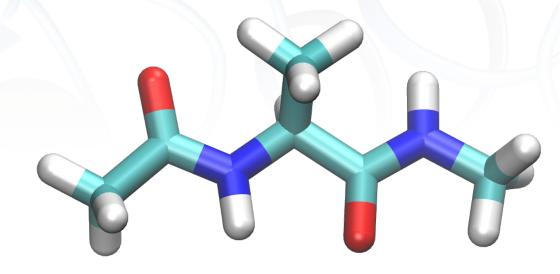
$$= \sum_{bonds} \frac{1}{2}k(r - r_{0})^{2} + \sum_{angles} \frac{1}{2}k(\theta - \theta_{0})^{2}$$

$$+ \sum_{torsions} \frac{1}{2}V_{n}[1 + \cos(n\omega - \gamma)]$$

$$+\sum_{i=1}^{N}\sum_{j>i}^{N}\left\{\frac{q_{i}q_{j}}{4\pi\varepsilon_{0}r_{ij}}+4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12}-\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right]\right\}$$

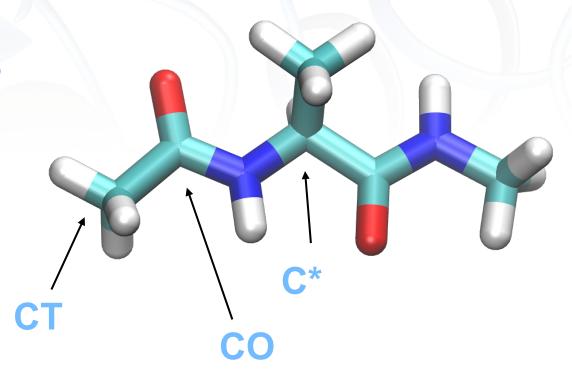
System Parameters: $r_0, \theta_0, k, V, \sigma_{ij}$, etc

System Parameters: $r_0, \theta_0, k, V, q_i, \sigma_{ij}$, etc



System Parameters: $r_0, \theta_0, k, V, q_i, \sigma_{ij}$, etc

To define these parameters, atom-types are defined based on local information.

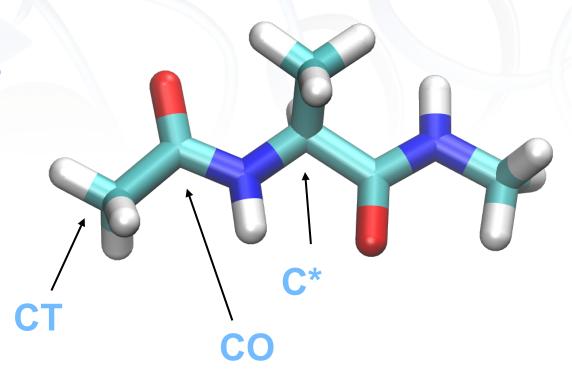


System Parameters: $r_0, \theta_0, k, V, q_i, \sigma_{ij}$, etc

To define these parameters, atom-types are defined based on local information.

Something to think about:

Are more are less atom-types better?



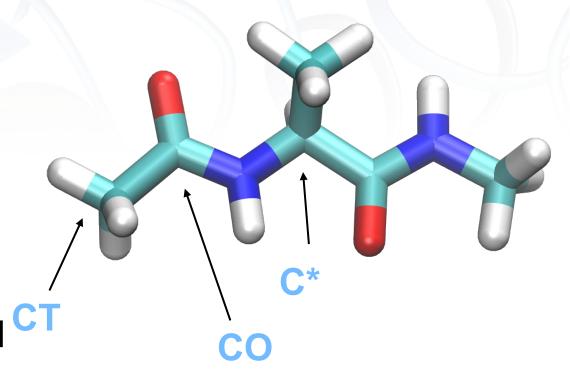
System Parameters: $r_0, \theta_0, k, V, q_i, \sigma_{ij}$, etc

To define these parameters, atom-types are defined based on local information.

Something to think about:

Are more are less atom-types better?

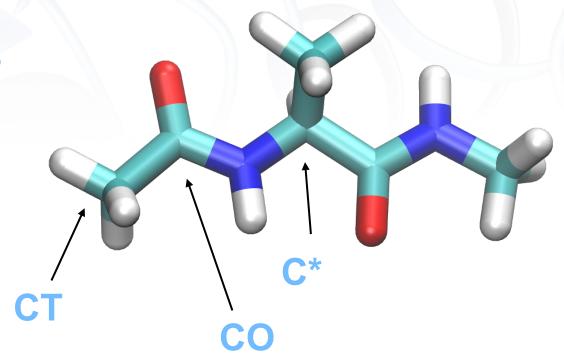
Specific interactions are more accurately modelled but the force field parameterization is harder with each one. The force field is less transferable.

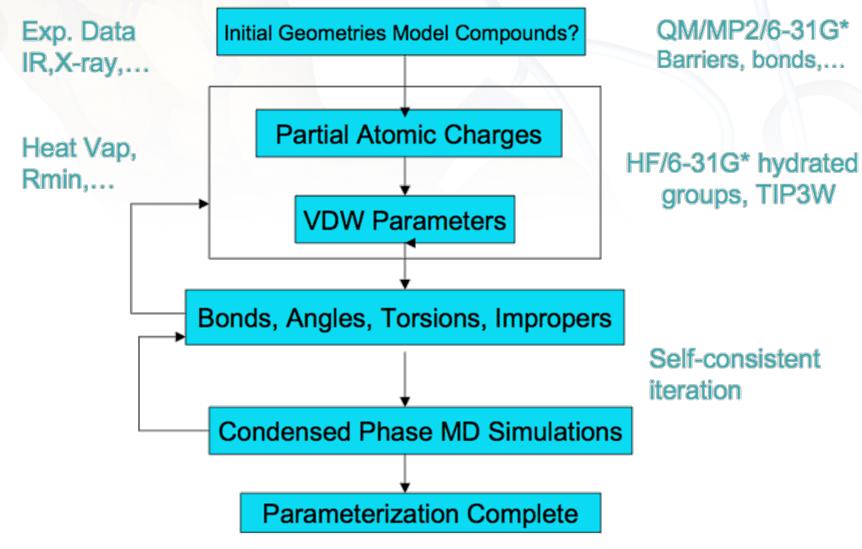


System Parameters: $r_0, \theta_0, k, V, q_i, \sigma_{ij}$, etc

To define these parameters, atom-types are defined based on local information.

Protein and nucleic acid structures are sensitive changes in these parameters.





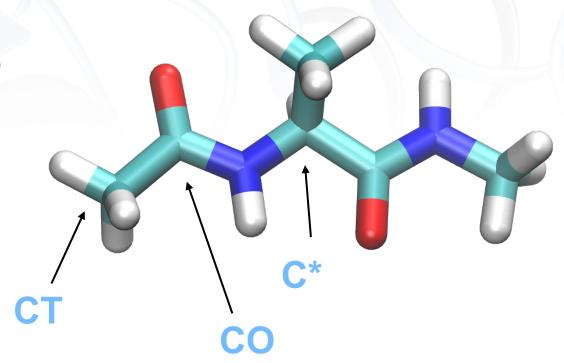
^{*} Summary of MacKerell, JCC v21, 86,105 (2000)

System Parameters: $r_0, \theta_0, k, V, q_i, \sigma_{ij}$, etc

To define these parameters, atom-types are defined based on local information.

Protein and nucleic acid structures are sensitive changes in these parameters.

Furthermore, parameterizations depend on the type of system being studied as well as the solvation model.



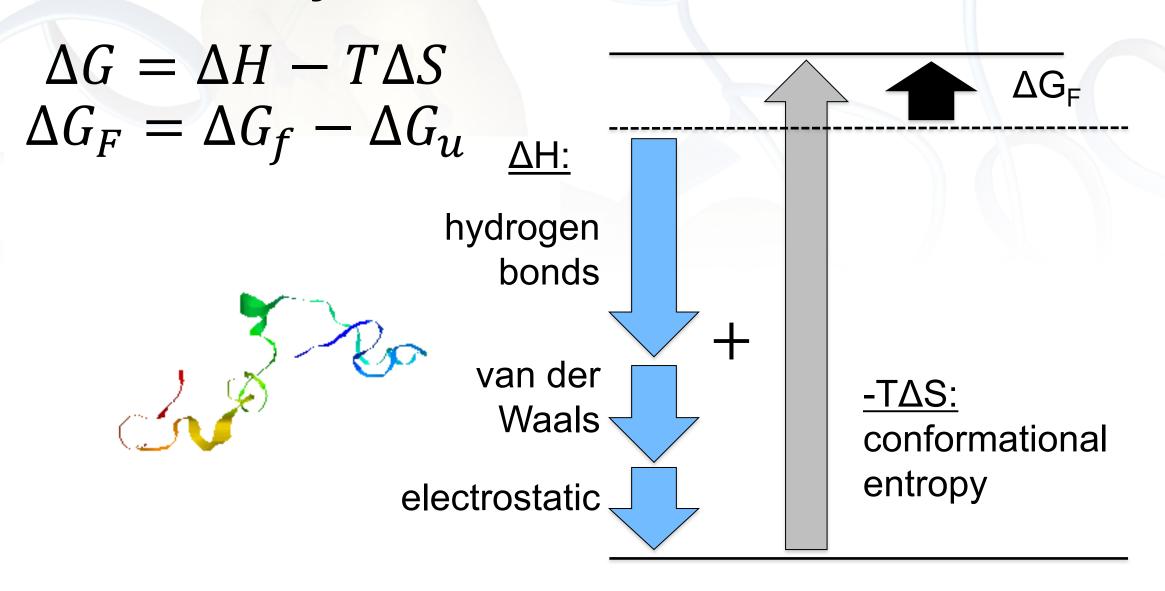
Molecular Dynamics Solvation

$$\Delta G = \Delta H - T \Delta S$$

$$\Delta G_F = \Delta G_f - \Delta G_u$$

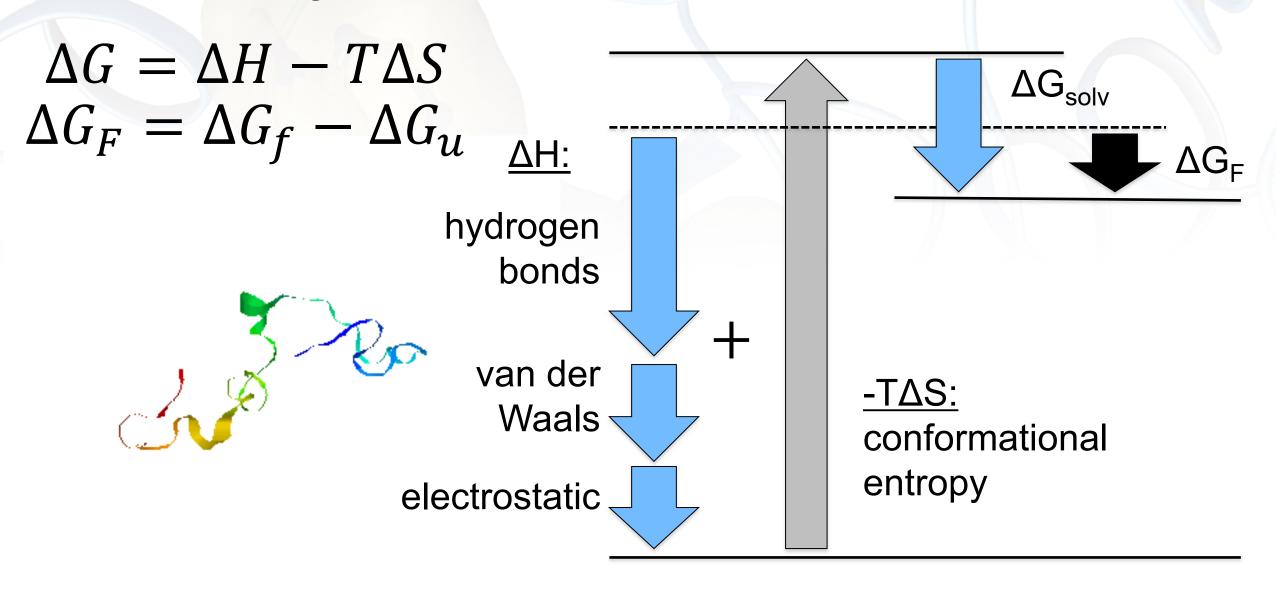


Molecular Dynamics Solvation

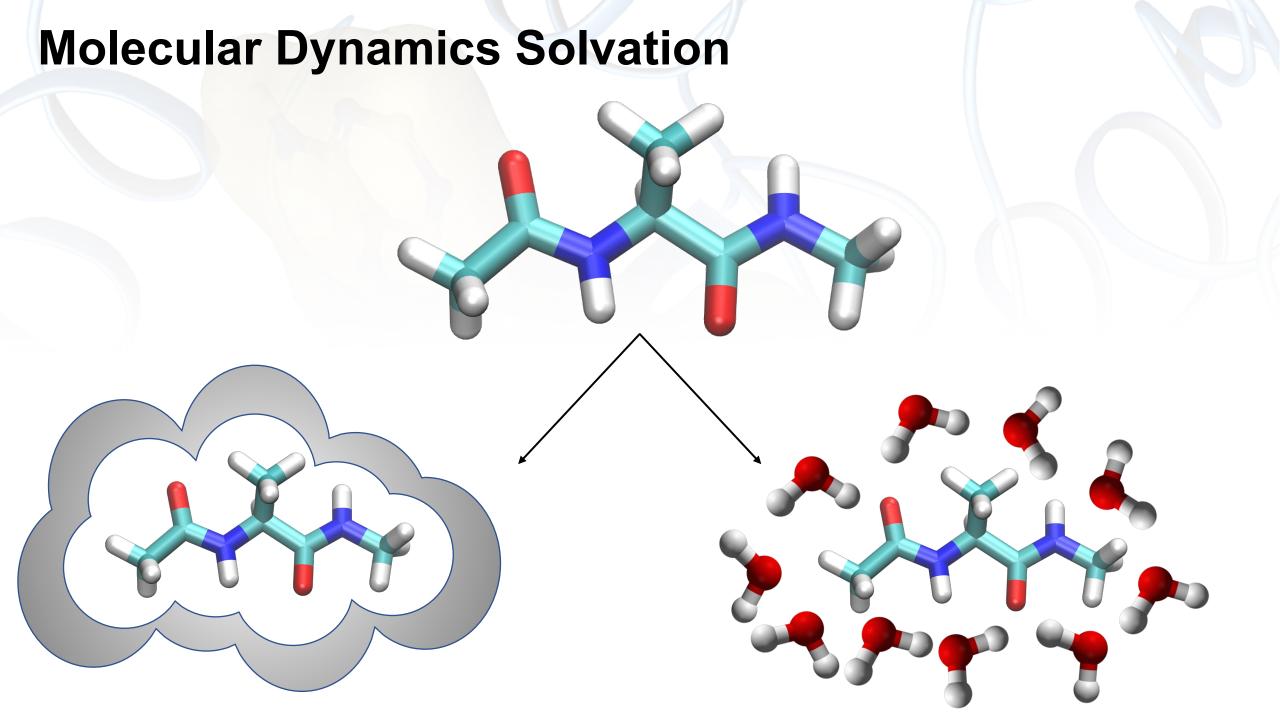


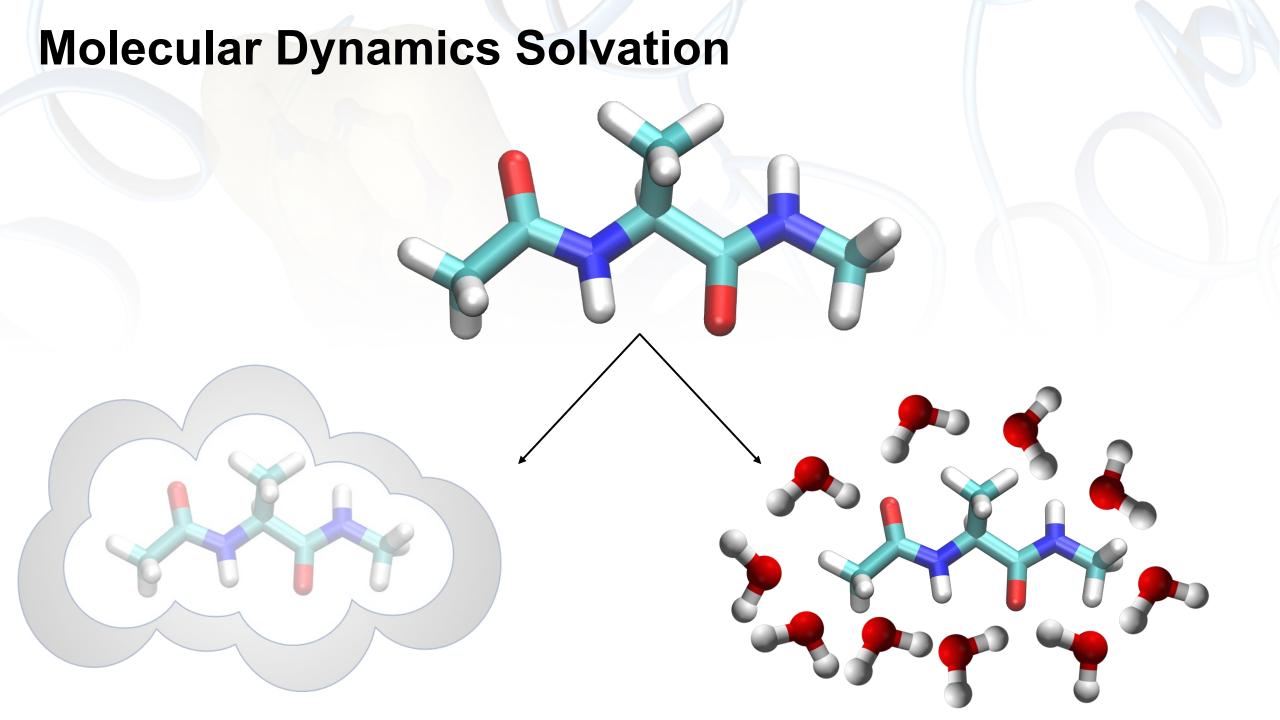
Protein (solute) alone

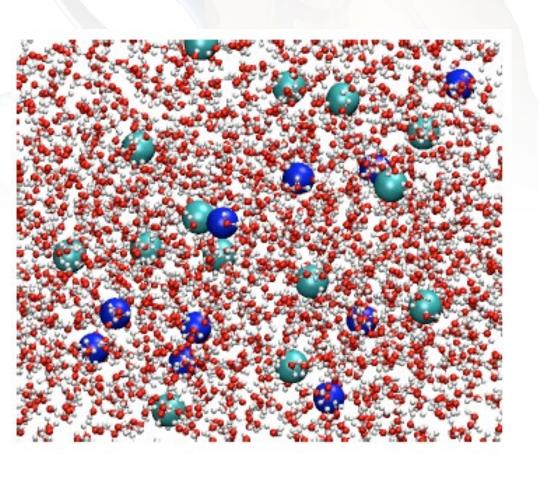
Molecular Dynamics Solvation



Protein (solute) in water







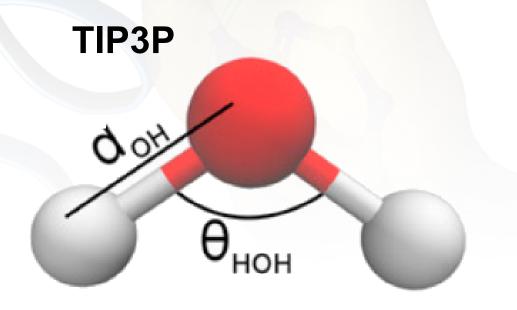
Explicit solvation uses specific water models to simulate bulk properties and electrostatic screening.

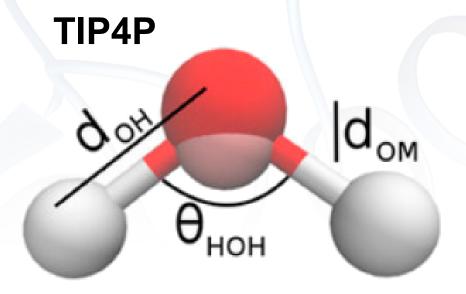
Advantages:

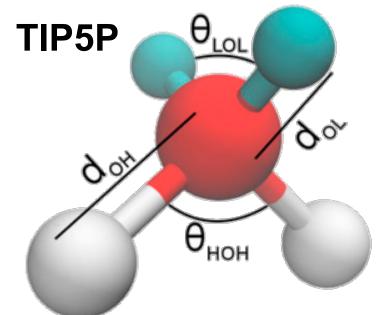
- High level detail (ions, cosolvents, etc)
- Interactions are directly modelled
- No additional terms added to MM FF

Disadvantages:

- High computational cost
- Slow to converge
- Still a model

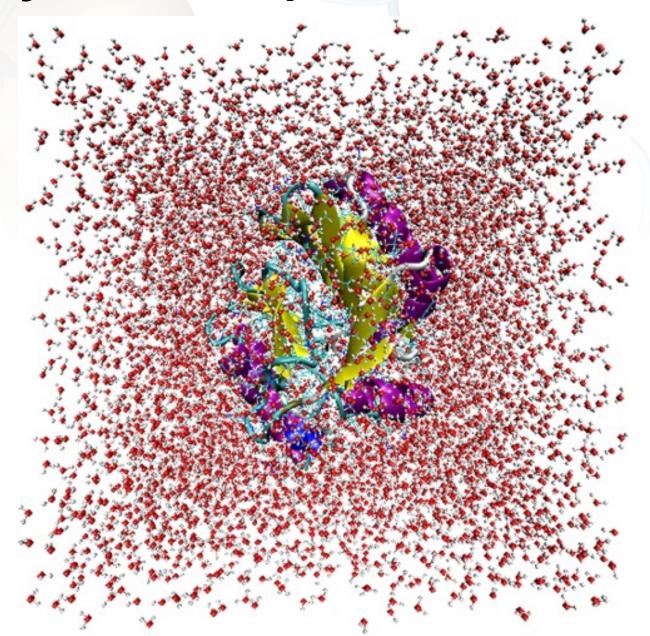


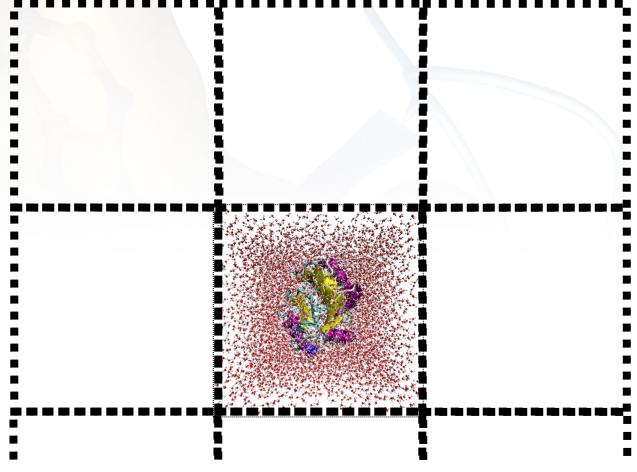




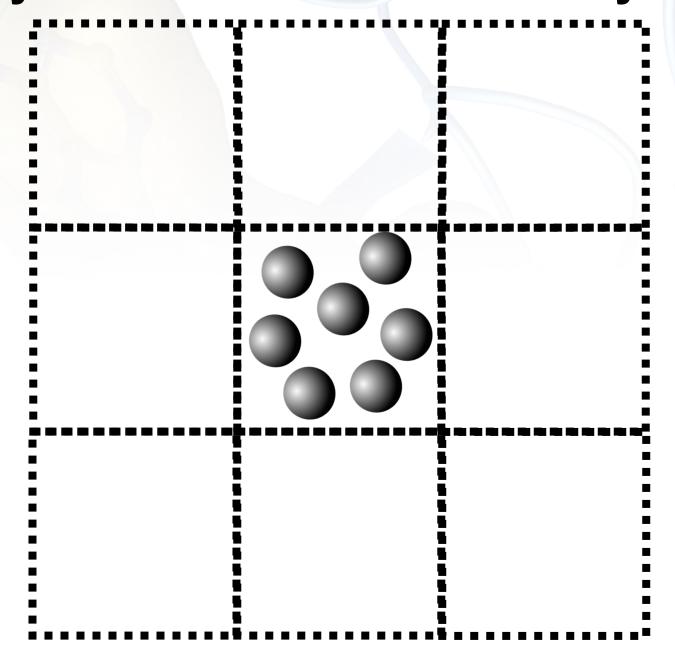
$$d_{oL} = 0.7 \text{\AA}$$
 $d_{oM} = 0.15 \text{\AA}$
 $d_{OH} = 0.9572 \text{Å}$
 $\theta_{LOL} = 109.47^{\circ}$
 $\theta_{HOH} = 104.52^{\circ}$

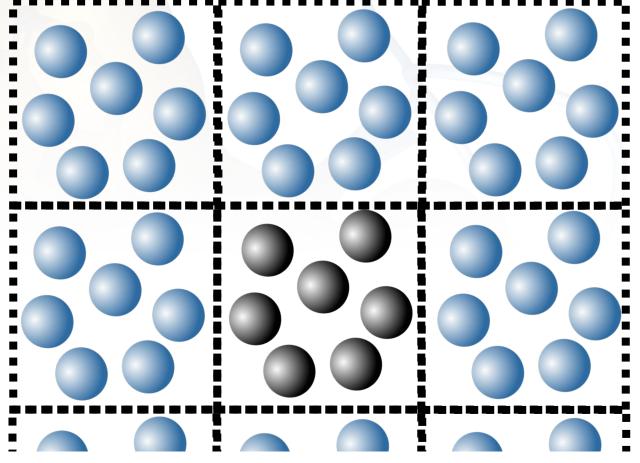
https://doi.org/10.1063/1.5070137



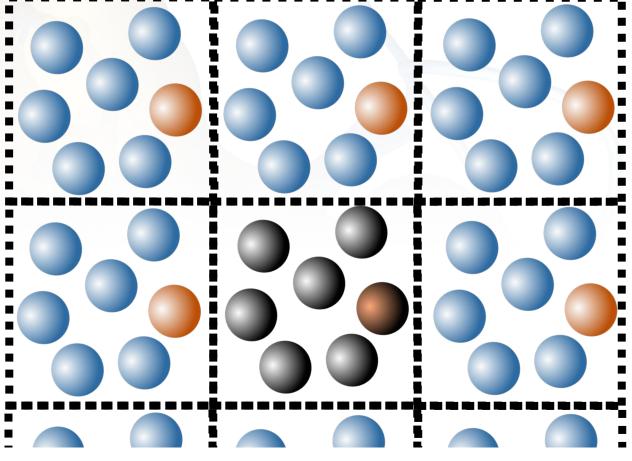


However, when using explicit solvent, periodic boundary conditions are very frequently used

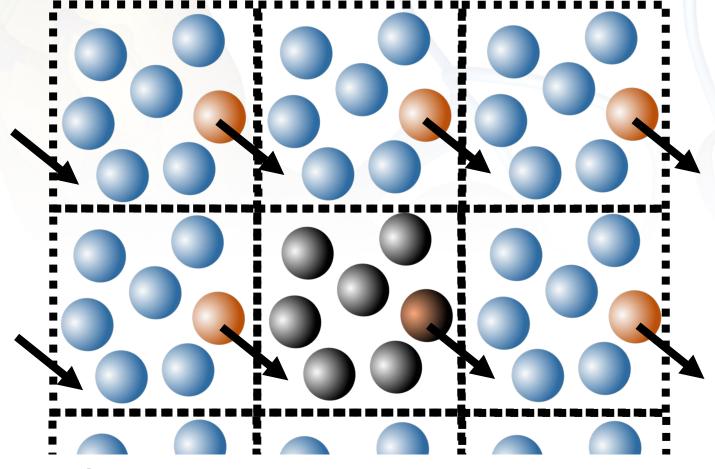




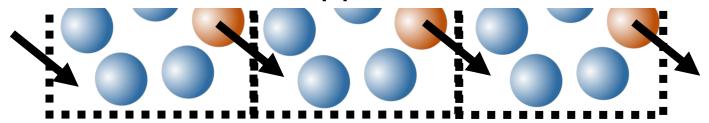
Can be envisioned as virtual images of the primary cell replicated in all directions.

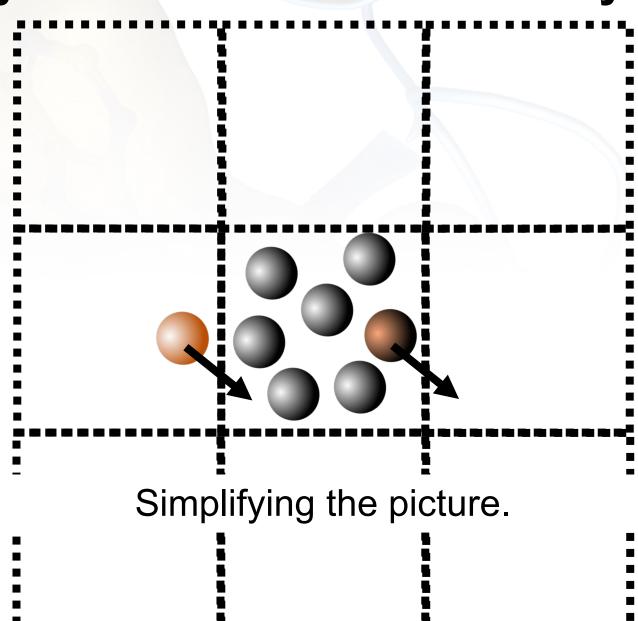


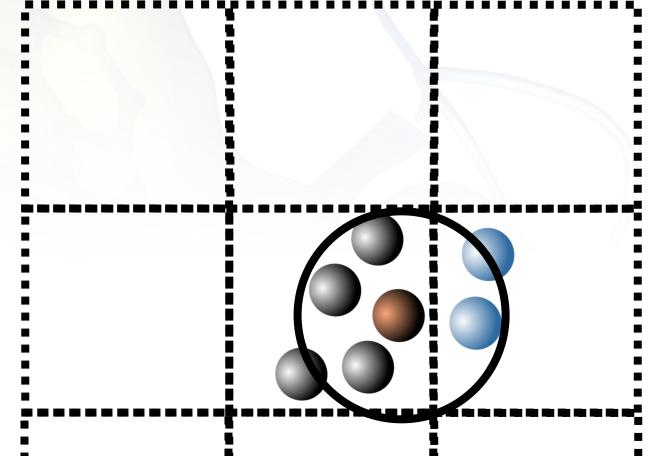
Now we can select one atom in the primary cell, and thus all its images.



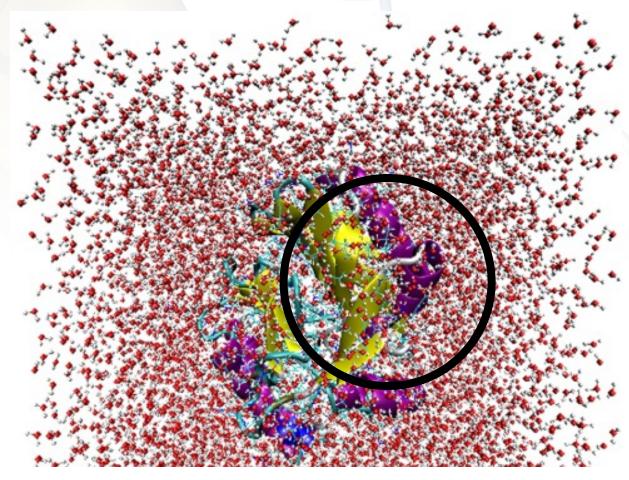
Let's see what happens when it moves.



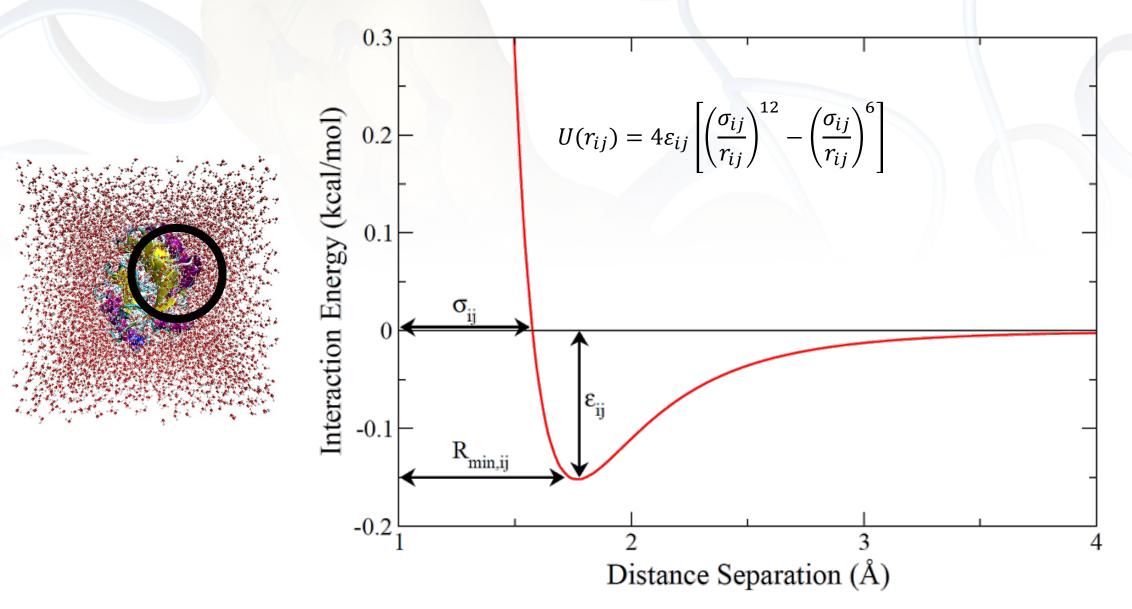


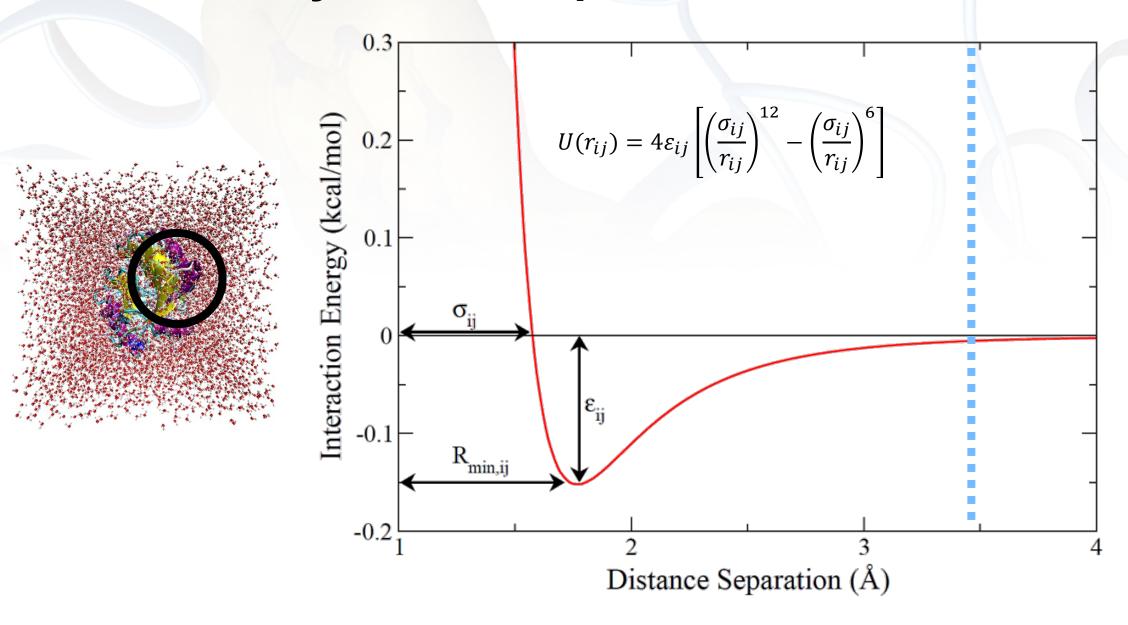


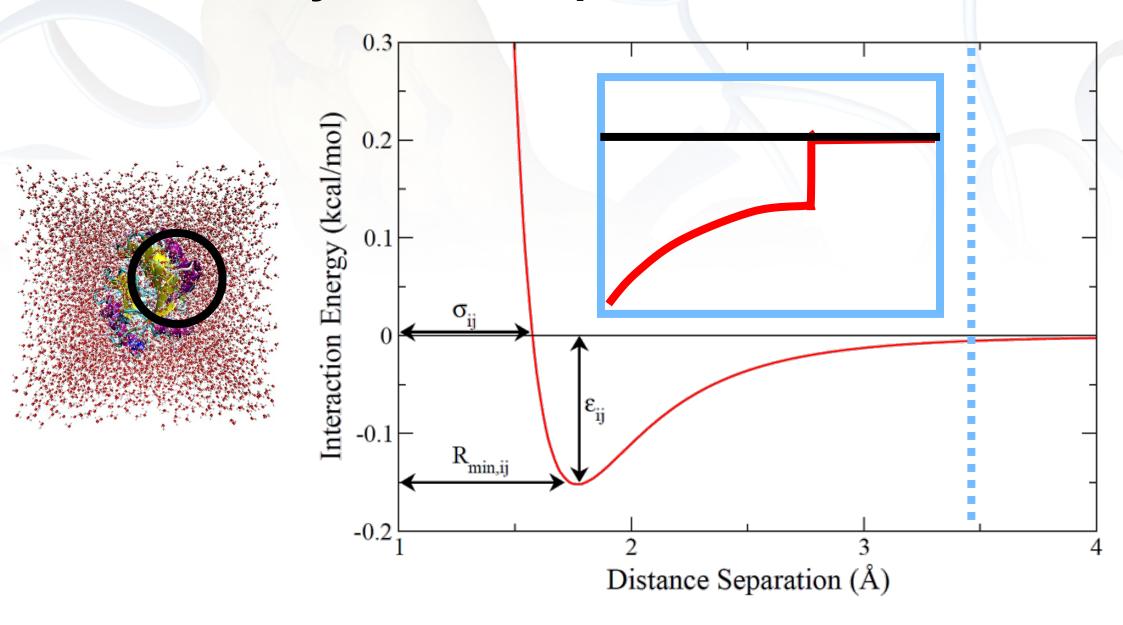
Apply an interaction (*electrostatic* and *non-bonded*) **cutoff**. Here, the **minimum image convention** is shown.

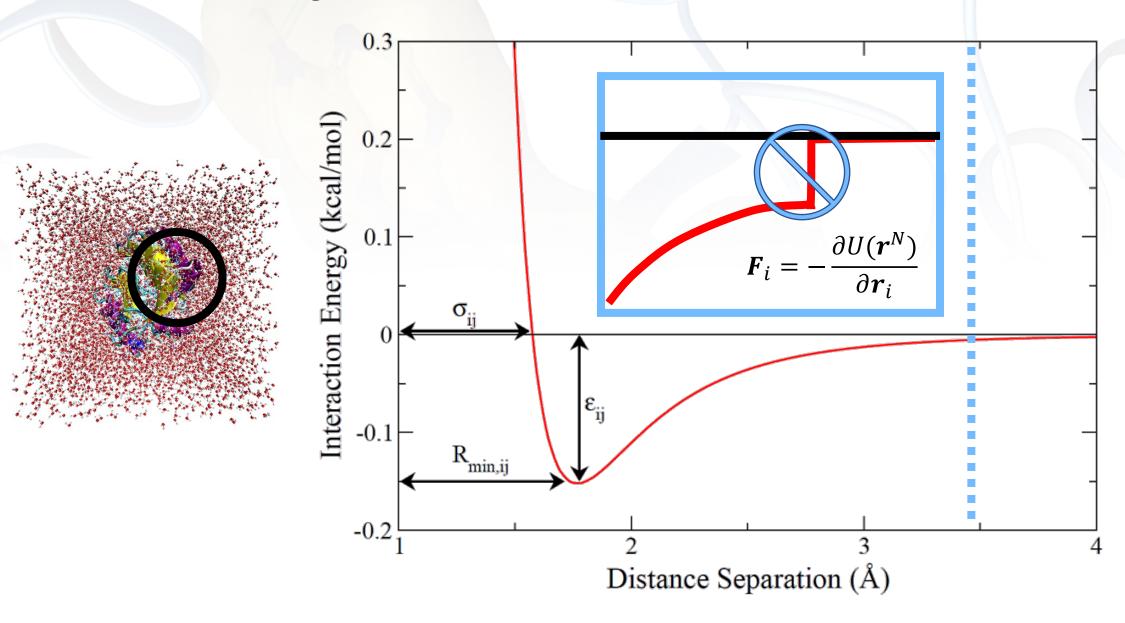


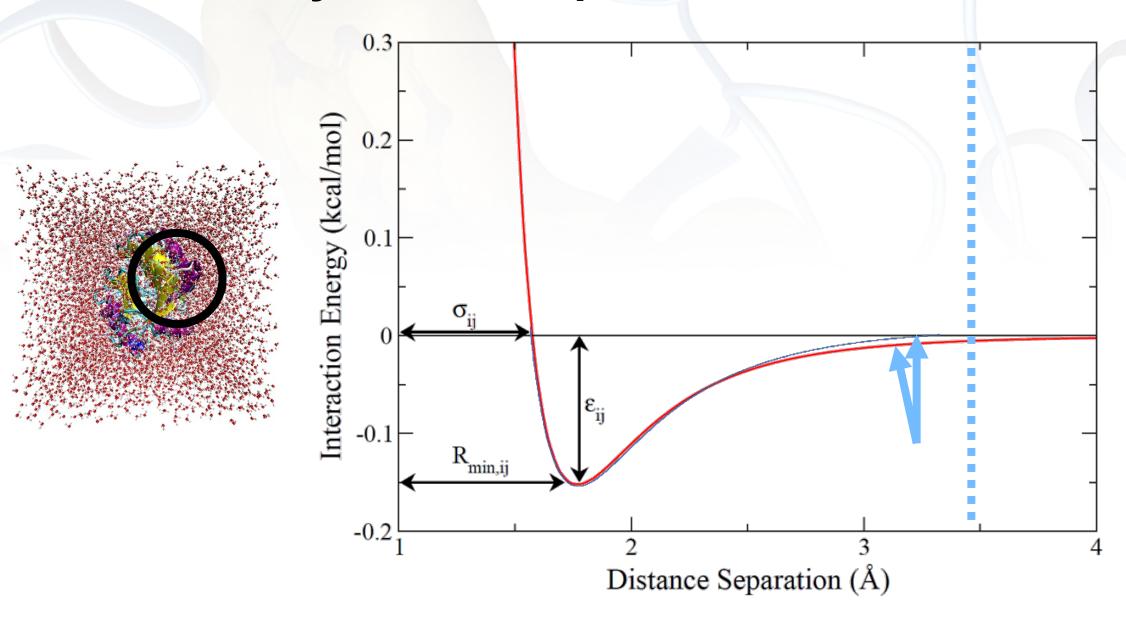
However, even more strict cutoffs *frequently* need to be taken to make a simulation computationally viable.

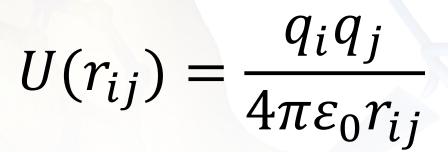


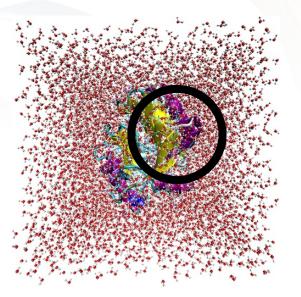




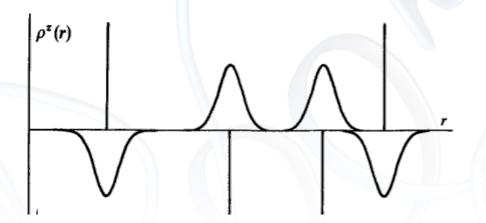




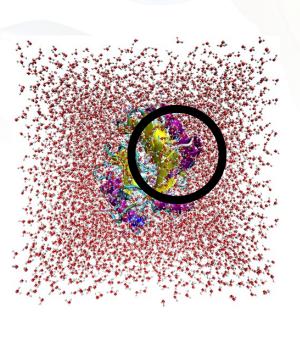


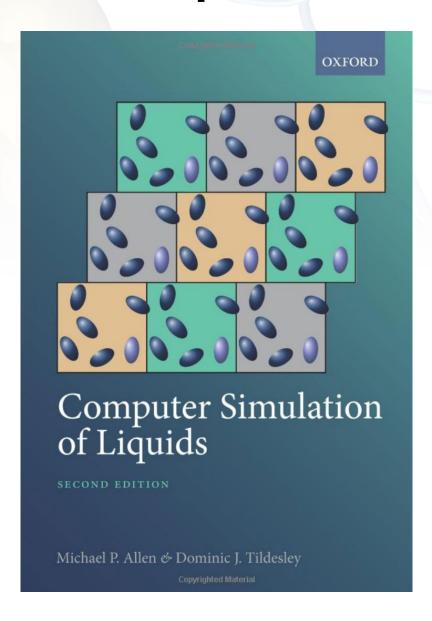


$$U(r_{ij}) = \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}$$



$$\mathcal{V}^{zz}(\varepsilon_{s} = 1) = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \left(\sum_{|\mathbf{n}|=0}^{\infty'} z_{i} z_{j} \frac{\operatorname{erfc}(\kappa |\mathbf{r}_{ij} + \mathbf{n}|)}{|\mathbf{r}_{ij} + \mathbf{n}|} + (1/\pi L^{3}) \sum_{\mathbf{k} \neq 0} z_{i} z_{j} (4\pi^{2}/k^{2}) \exp(-k^{2}/4\kappa^{2}) \cos(\mathbf{k} \cdot \mathbf{r}_{ij}) \right) - (\kappa/\pi^{1/2}) \sum_{i=1}^{N} z_{i}^{2} + (2\pi/3L^{3}) \left| \sum_{i=1}^{N} z_{i} \mathbf{r}_{i} \right|^{2}.$$

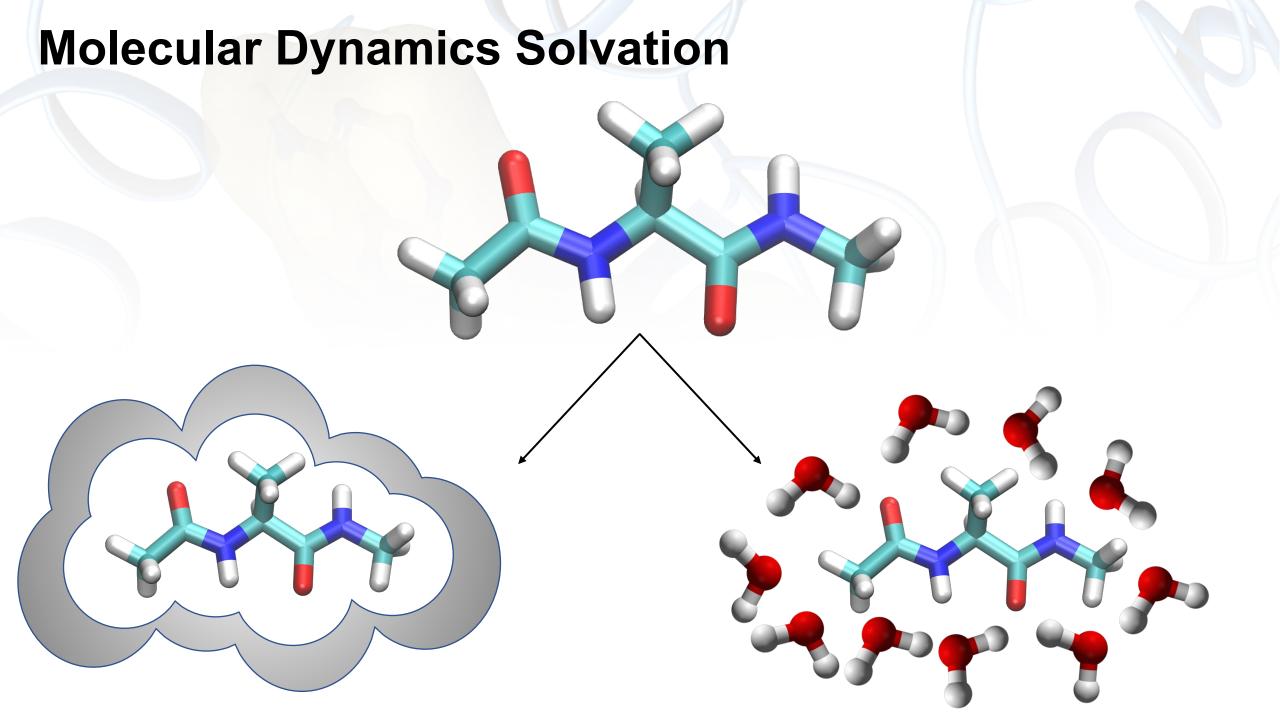


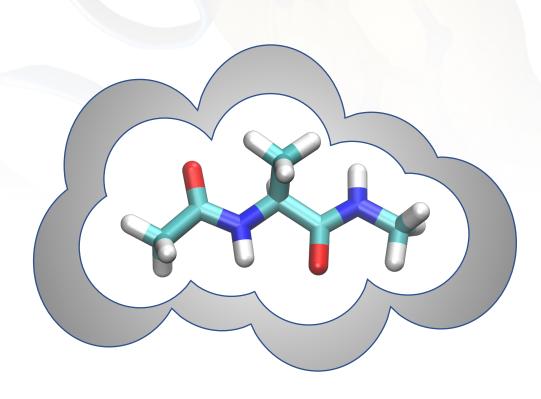


Allen & Tildesley

Computer Simulations

of Liquids





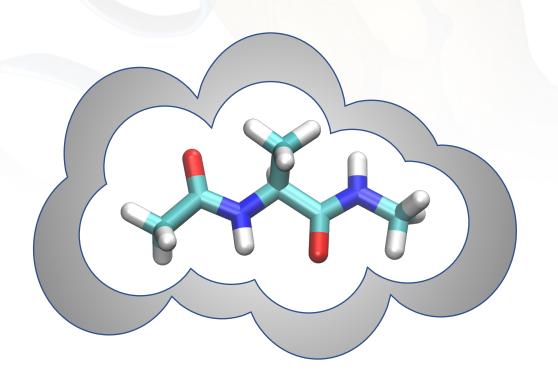
Implicit solvation, or continuum solvent models, uses a bulk polarizable medium for electrostatic screening and non-bonded cavitation.

Advantages:

- Lower computational cost
- Fast convergence
- Allows for high-level solutes

Disadvantages:

- Additional terms added to MM FF
- It's a big hunk of metal



$$U(\mathbf{r}^{N})$$

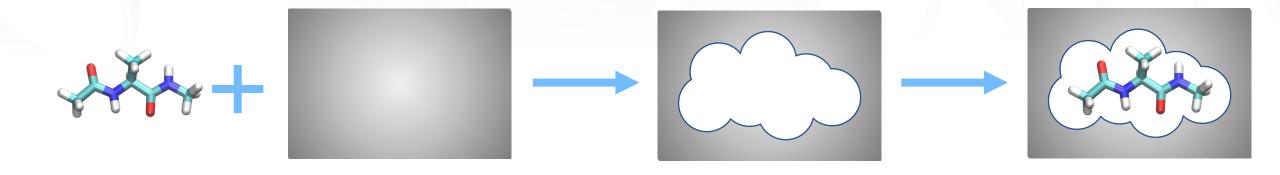
$$= \sum_{bonds} \frac{1}{2} k(r - r_{0})^{2} + \sum_{angles} \frac{1}{2} k(\theta - \theta_{0})^{2}$$

$$+ \sum_{torsions} \frac{1}{2} V_{n} [1 + \cos(n\omega - \gamma)]$$

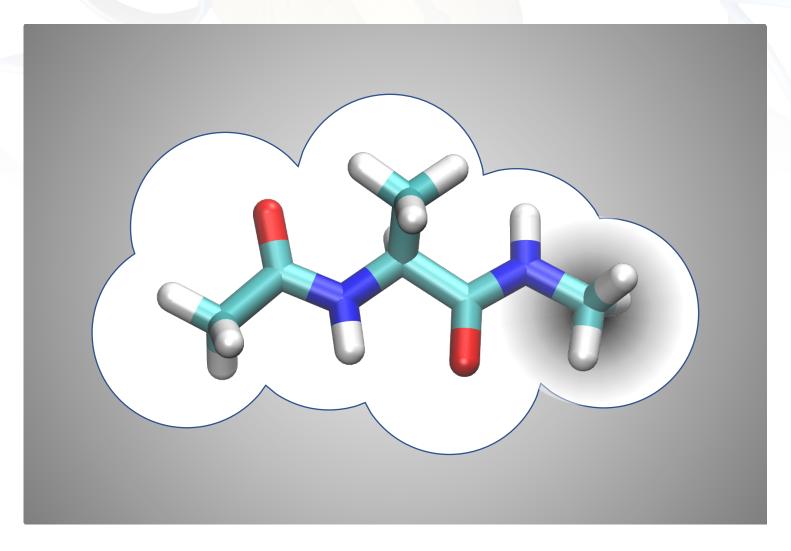
$$+ \sum_{i=1}^{N} \sum_{j>i}^{N} \left\{ \frac{q_{i}q_{j}}{4\pi\varepsilon_{0}r_{ij}} + 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] \right\}$$

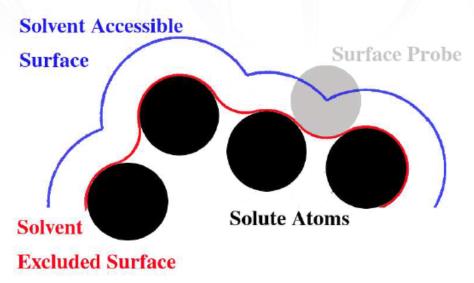
$$+ U_{\text{solvation}}$$

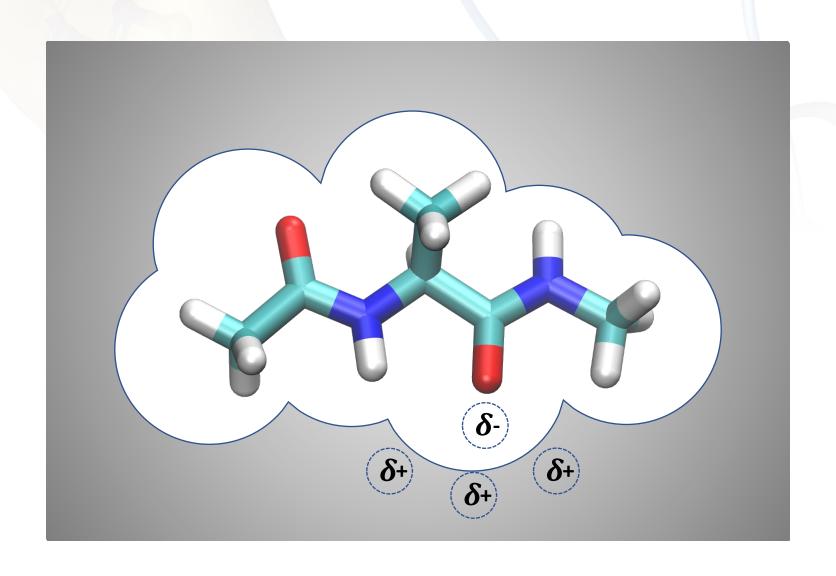
$$\Delta G_{\text{solv}} = \Delta G_{\text{cav}} + \Delta G_{\text{vdW}} + \Delta G_{\text{ele}}$$



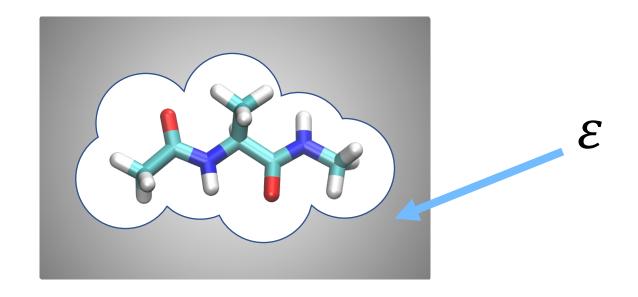
$$\Delta G_{\rm cav} + \Delta G_{\rm vdW}$$



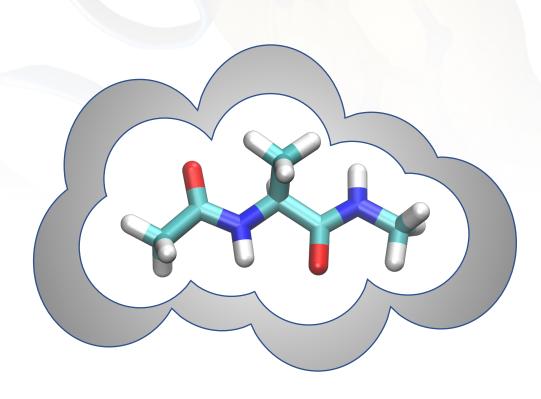




$$\Delta G_{\text{ele}} = \frac{1}{2} \left(1 - \frac{1}{\varepsilon} \right) \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{q_i q_j}{f(r_{ij}, a_{ij})}$$



		Dielectric constant
	Material	ε
	Vacuum	1.0000
	Air (1 atm)	1.0006
	Paraffin	2.2
	Polystyrene	2.6
	Vinyl (plastic)	2–4
	Paper	3.7
	Quartz	4.3
	Oil	4
	Glass, Pyrex	5
	Rubber, neoprene	6.7
	Porcelain	6-8
	Mica	7
	Water (liquid)	80
	Strontium titanate	300



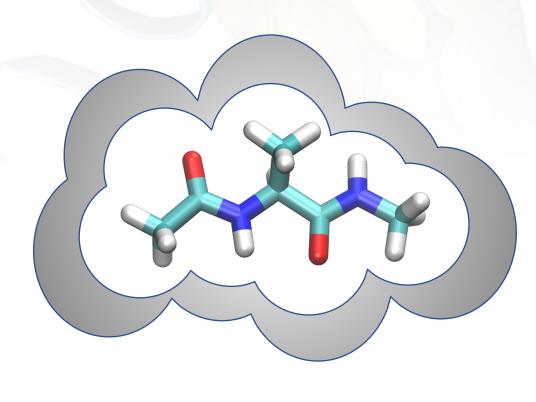
Implicit solvation, or continuum solvent models, uses a bulk polarizable medium for electrostatic screening and non-bonded cavitation.

Advantages:

- Lower computational cost
- Fast convergence
- Allows for high-level solutes

Disadvantages:

- Additional terms added to MM FF
- It's a big hunk of metal



Implicit solvation

Advantages:

- Very quick
- Allows for more detailed/larger systems

Disadvantages:

- Only accounts for solvent in an average way
 - No direct solvent interactions
 - No hydrogen bonding
 - Problems in cavitation

Introduction to biomolecular simulation

Lecture 2: Force Fields & Solvation Models

