Non-bonded interactions

speeding up the number-crunching

Marcus Elstner and Tomáš Kubař

2018, May 11 & 15

Non-bonded interactions – why care?

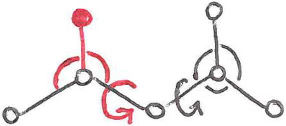
- key to understand biomolecular structure and function
  - binding of a ligand
  - efficiency of a reaction
  - color of a chromophore
- two-body potentials → computational effort of $\mathcal{O}(N^2)$
  - good target of optimization
- solvent (H_2O) – crucial role, huge amount
  - efficient description needed

How many pair-wise interactions are there?

imagine we introduce an additional atom into a system
that already has $N-1$ atoms

bonded interactions

- we add at most (roughly)
  2 bonds, 2 angles, 3 dihedrals
- for N atoms, this is at most $7N$
  – proportional to N: $\mathcal{O}(N)$

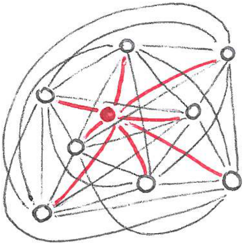


How many pair-wise interactions are there?

imagine we introduce an additional atom into a system
that already has $N - 1$ atoms

non-bonded interactions

- between the new atom
  and each of the previous atoms:
  $N - 1$ interactions!
- for N atoms, this is $N(N-1)/2$
  – proportional to N^2: $\mathcal{O}(N^2)$



How many pair-wise interactions are there?

Let us assume: the calculation of every atom–atom interaction takes the same amount of time

Then, the $\mathcal{O}(N^2)$ evaluation of non-bonded interactions will be the most computationally intensive procedure in the entire simulation (the bottle neck)

Coulomb's law

elstat. interaction energy of point charges q and Q on distance r:

$$E^{\mathrm{el}} = \frac{1}{4\pi\varepsilon_0} \cdot \frac{q \cdot Q}{r}$$

electrostatic potential (ESP) induced at $\vec{r}$ by Q at $\vec{r}_1$:

$$\phi(\vec{r}) = \frac{1}{4\pi\varepsilon_0} \cdot \frac{Q}{|\vec{r} - \vec{r}_1|}$$

ESP induced by a number of point charges Q_i:

$$\Phi(\vec{r}) = \frac{1}{4\pi\varepsilon_0} \sum_i \frac{Q_i}{|\vec{r} - \vec{r}_i|}$$

if we know ESP at $\vec{r}$, and q is placed there, elstat. energy follows as

$$E^{\mathrm{el}}(\vec{r}) = \Phi(\vec{r}) \cdot q$$

→ 'electrostatic potential energy surface'

Coulomb's law

- continuous charge distribution – charge density $\rho = \partial Q / \partial V$
- charge in a volume element V_i: $Q_i = \rho(\vec{r}_i) \cdot V_i = \rho(\vec{r}_i) \cdot \Delta V$
- summing the potential induced by all elements gives

$$\Phi(\vec{r}) = \frac{1}{4\pi\varepsilon_0} \sum_i \frac{\rho(\vec{r}_i) \cdot \Delta V}{|\vec{r} - \vec{r}_i|}$$

- for infinitesimal volume elements (with $\mathrm{d}^3\vec{r} = \mathrm{d}V$):

$$\Phi(\vec{r}) = \frac{1}{4\pi\varepsilon_0} \int \frac{\rho(\vec{r}_1)}{|\vec{r} - \vec{r}_1|} \, \mathrm{d}^3\vec{r}_1$$

- elstat. energy of a charge density $\rho(\vec{r})$ follows as

$$E = \frac{1}{2} \int \Phi(\vec{r}) \cdot \rho(\vec{r}) \, \mathrm{d}V = \frac{1}{8\pi\varepsilon_0} \iint \frac{\rho(\vec{r}_1) \cdot \rho(\vec{r})}{|\vec{r} - \vec{r}_1|} \, \mathrm{d}^3\vec{r} \, \mathrm{d}^3\vec{r}_1$$

Poisson's equation

- needs to be solved to get ESP from charge distribution (differential equation for Φ as a function of $\vec{r}$):

$$\nabla^2 \Phi(\vec{r}) = -\frac{\rho(\vec{r})}{\varepsilon}$$

- if permittivity ε is not constant:

$$\nabla \left(\varepsilon \nabla \Phi(\vec{r}) \right) = -\rho(\vec{r})$$

example: ESP of Gaussian charge density around $\vec{o}$ with width σ:

$$\rho(r) = Q \cdot \frac{1}{\sigma^3 \sqrt{2\pi}^3} \cdot \exp\left[\frac{r^2}{2\sigma^2}\right]$$

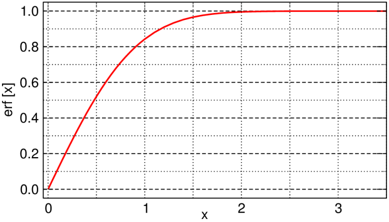
solution of Poisson's equation:

$$\Phi(r) = \frac{1}{4\pi\varepsilon} \cdot \frac{Q}{r} \cdot \operatorname{erf}\left[\frac{r}{\sqrt{2}\sigma}\right]$$

Poisson's equation

solution for Gaussian charge density:

$$\Phi(r) = \frac{1}{4\pi\varepsilon} \cdot \frac{Q}{r} \cdot \mathrm{erf}\left[\frac{r}{\sqrt{2}\sigma}\right]$$



Poisson's equation

solution for Gaussian charge density:

$$\Phi(r) = \frac{1}{4\pi\varepsilon} \cdot \frac{Q}{r} \cdot \mathrm{erf}\left[\frac{r}{\sqrt{2}\sigma}\right]$$

if we move far from the center of charge density (r is large)

- erf converges to 1, ESP equals that
  of a point charge placed in $\vec{o}$
- accordance with experience – a point charge
  and a well-localized charge density
  interact with distant charges in the same way
- actually, we need not go far to see that
  – erf $=$ 0.999 already for $x = 2.4\sigma$

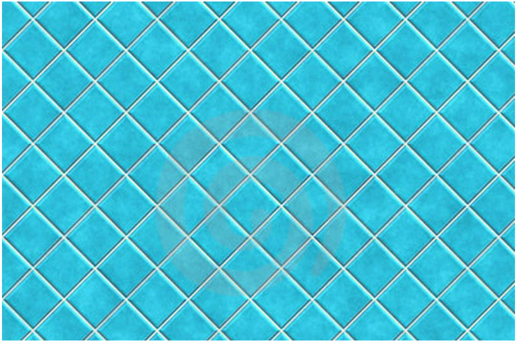
Biomolecule in solution

typical MD simulations – molecular system in aqueous solution
task – make the system as small as possible (reduce cost)
straightforward solution – single molecule of solute (protein, DNA)
with a smallest possible number of H_2O molecules
typical – several thousand H_2O molecules in a cube $n \times n \times n$ nm

issue – everything is close to the surface,
while we are interested in a molecule in bulk solvent
so – we may be simulating something else than what we want

Periodic boundary conditions

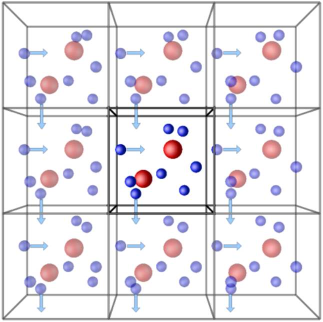
- elegant way to avoid these problems
- molecular system placed in a regular-shaped box
- the box is virtually replicated in all spatial directions



Periodic boundary conditions

- elegant way to avoid these problems
- molecular system placed in a regular-shaped box
- the box is virtually replicated in all spatial directions
- positions (and velocities) of all particles
  are identical in all replicas,
  so we can keep only one copy in the memory
- this way, the system is infinite – no surface!
- the atoms near the wall of the simulation cell interact
  with the atoms in the neighboring replica

PBC – example



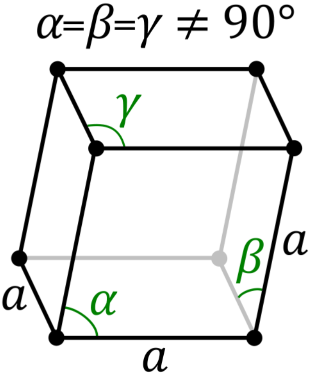
isaacs.sourceforge.net

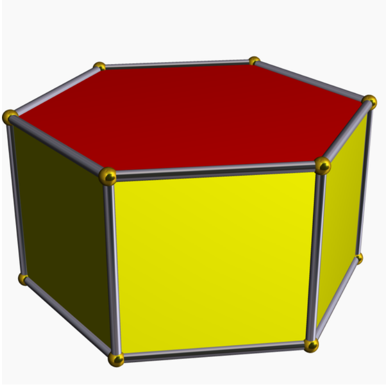
PBC – features

- small problem – artificial periodicity in the system (entropy ☹)
  – still much better than boundary with vacuum
- only coordinates of the unit cell are recorded
- atom that leaves the box enters it on the other side.
- carefull accounting of the interactions of atoms necessary!
  - simplest – minimum image convention:
    - an atom interacts with the nearest copy of every other
      – interaction with two different images of another atom,
      or even with another image of itself is avoided

PBC – box shape

may be simple – cubic or orthorhombic, parallelepiped (specially, rhombohedron), or hexagonal prism

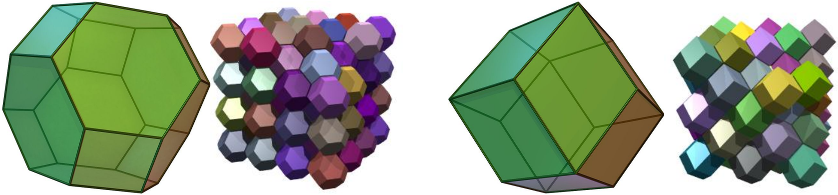




PBC – box shape

. . . but also more complicated

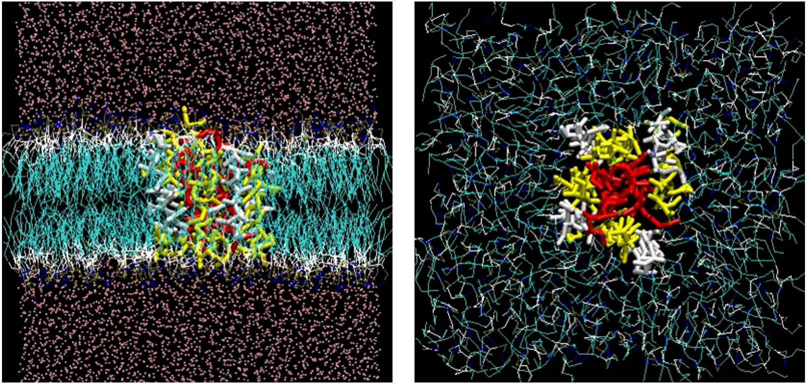
- truncated octahedral or rhombic dodecahedral
- quite complex equations for interactions & eqns of motion



advantage for simulation of spherical objects (globular proteins)
– no corners far from the molecule filled with unnecessary H_2O

PBC – box shape

2D objects – phase interfaces, membrane systems
– usually treated in a slab geometry



Cut-off – simple idea

non-bonded terms – bottleneck of the calculation
with PBC – infinite number of interaction pairs in principle,
  still too many interaction with min. image convention
  but the interaction gets weaker with distance

simplest and crudest approach to limit the number of calculations:
  neglect interaction of atoms further apart than r_c – cut-off

  very good for rapidly decaying LJ interaction $(1/r^6)$ $(r_c = 10\text{ Å})$
  not so good for slowly decaying electrostatics $(1/r)$
    – sudden jump (discontinuity) of potential energy,
      disaster for forces at the cut-off distance

Cut-off – better: shift

shift the whole function by $V(r_c)$ – eliminate the jump at r_c:

$$V^{\text{sh}}(r) = \begin{cases} V(r) - V(r_c), & \text{for} \quad r \leq r_c, \\ 0, & \text{otherwise.} \end{cases}$$

still, the gradients (forces) are at r_c discontinuous!

shift-force potential gets rid of that ($V' \equiv \mathrm{d}V/\mathrm{d}r$):

$$V^{\text{sf}}(r) = \begin{cases} V(r) - V(r_c) - V'(r_c) \cdot (r - r_c), & \text{for} \quad r \leq r_c, \\ 0, & \text{otherwise.} \end{cases}$$

drawback – the Coulomb energy is not quite Coulomb anymore

Cut-off – better: reaction field

reaction field interaction:

assume a constant dielectric environment beyond the cut-off r_c, with a dielectric constant $\varepsilon_{\mathrm{rf}}$ (parameter):

$$\begin{aligned} V^{\mathrm{rf}}(r) &= \frac{1}{r} \cdot \left(1 + \frac{\varepsilon_{\mathrm{rf}} - 1}{2\varepsilon_{\mathrm{rf}} + 1} \cdot \frac{r^3}{r_c^3}\right) - V^{\mathrm{rf}}(r_c) \\ F^{\mathrm{rf}}(r) &= -\frac{1}{r^2} \cdot \left(1 - \frac{2\varepsilon_{\mathrm{rf}} - 2}{2\varepsilon_{\mathrm{rf}} + 1} \cdot \frac{r^3}{r_c^3}\right) \end{aligned}$$

(the force at cut-off is very small, and vanishes with $\varepsilon_{\mathrm{rf}}$)

there is a physical motivation – possible advantage

Cut-off – better: switch

switch off the Coulomb interaction from full strength to zero,
starting from a certain distance r_1,
by multiplication with a function passing from 1 to 0

$$V^{\mathrm{sw}}(r) = \begin{cases} V(r) & \text{for} \quad r < r_1, \\ V(r) \cdot \varphi\left(\frac{r - r_1}{r_c - r_1}\right) & \text{for} \quad r_1 < r < r_c, \\ 0, & \text{otherwise.} \end{cases}$$

– interaction altered in the cut-off region

switch-force: $F^{\mathrm{fsw}}(r) = F(r) \cdot \varphi\left(\frac{r - r_1}{r_c - r_1}\right)$ for $r_1 < r < r_c$,

if needed, obtain energy formally as $V^{\mathrm{fsw}}(r) = \int_{\infty}^{r} F^{\mathrm{fsw}}(r')$

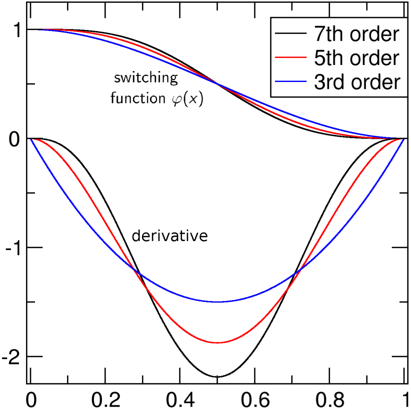
Note: switching function

a general concept to approximation
avoiding abrupt change of a value of the function

$$f_{\mathrm{sw}}(x) = f(x) \cdot \varphi\left(\frac{x - x_1}{x_0 - x_1}\right)$$

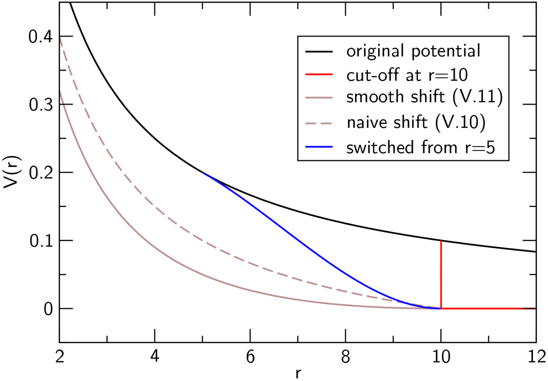
switching function $\varphi(x)$

- defined on interval $(0, 1)$
- goes from 1 to 0
- need continuous derivative?
  - – cubic function
- need cont. 2nd derivative?
  - – 5th-order polynom



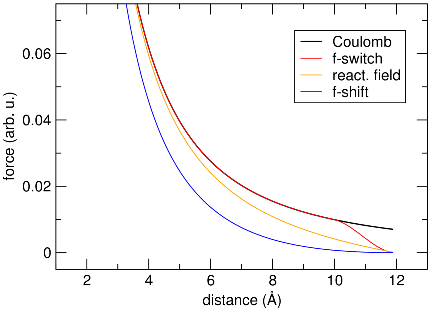
Cut-off – better alternatives

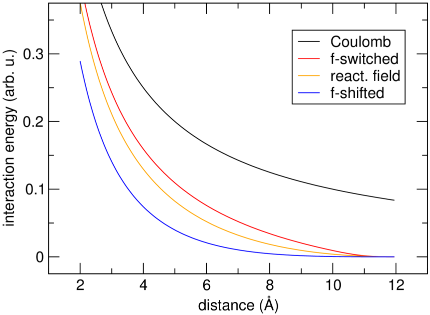
elstat. interaction energy of two unit positive charges



shift / switch – applied here to energy, better apply them to force

Cut-off – better alternatives





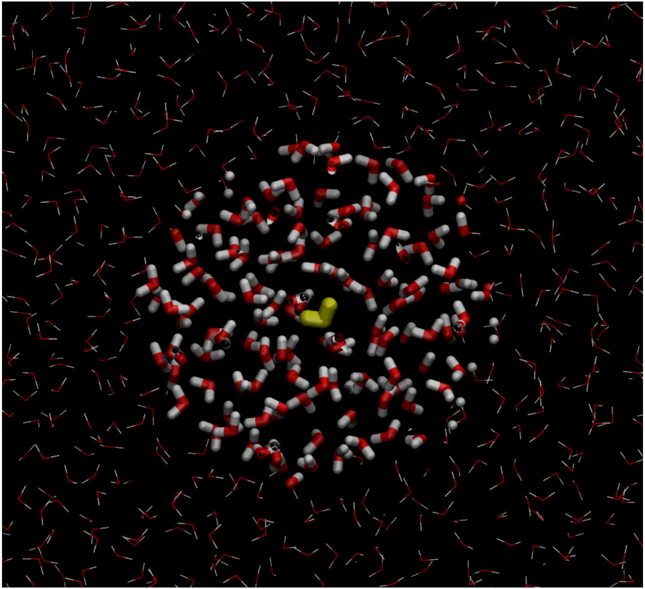
Neighbor lists

cut-off – we still have to calculate the distance for every two atoms
(to compare it with the cut-off distance)
→ we do not win much yet – there are still $\mathcal{O}(N^2)$ distances

observation: pick an atom A.
the atoms that are within cut-off distance r_c around A,
remain within r_c for several consecutive steps of dynamics,
while no other atoms approach A that close

idea: maybe it is only necessary to calculate the interactions
between A and these close atoms – neighbors

Neighbor lists



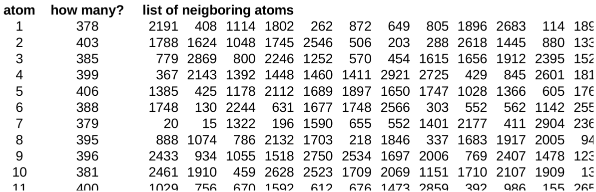
Neighbor lists

what will we do?

calculate the distances for every pair of atoms

less frequently, i.e. every 10 or 20 steps of dynamics, and

record the atoms within cut-off distance in a neighbor list



atom	how many?	list of neigboring atoms											
1	378	2191	408	1114	1802	262	872	649	805	1896	2683	114	189
2	403	1788	1624	1048	1745	2546	506	203	288	2618	1445	880	133
3	385	779	2869	800	2246	1252	570	454	1615	1656	1912	2395	152
4	399	367	2143	1392	1448	1460	1411	2921	2725	429	845	2601	181
5	406	1385	425	1178	2112	1689	1897	1650	1747	1028	1366	605	176
6	388	1748	130	2244	631	1677	1748	2566	303	552	562	1142	255
7	379	20	15	1322	196	1590	655	552	1401	2177	411	2904	236
8	395	888	1074	786	2132	1703	218	1846	337	1683	1917	2005	94
9	396	2433	934	1055	1518	2750	2534	1697	2006	769	2407	1478	123
10	381	2461	1910	459	2628	2523	1709	2069	1151	1710	2107	1909	13
11	400	1029	756	670	1592	612	676	1473	2859	392	986	155	265

then – calculate the interaction for each atom

only with for the atoms in the neighbor list – formally $\mathcal{O}(N)$

note – the build of the neighbor list itself is $\mathcal{O}(N^2)$,

which can be reduced with further tricks ('cell lists')

Accounting of all of the replicas

cut-off – often bad approximation, e.g. with highly charged molecular systems (DNA, some proteins)

artificial forces with switching function

→ e.g. artificial accumulation of ions around cut-off

only way – abandon the minimum image convention and cut-off

– sum up the long-range Coulomb interaction between all the replicas of the simulation cell

introduce $\vec{n}$ running over all the replicas

- for $|\vec{n}| = 0$, we have $\vec{n} = (0, 0, 0)$ – the central unit cell.
- for $|\vec{n}| = L$: $\vec{n} = (0, 0, \pm L)$, $\vec{n} = (0, \pm L, 0)$, $\vec{n} = (\pm L, 0, 0)$ – the six neighboring unit cells.
- continue with $|\vec{n}| = \sqrt{2}L$: 12 cells touching with edge...

Can we sum it up simply?

sum of Coulomb interactions over all replicas:

$$E^{\text{Coul}} = \frac{1}{2} \sum_{i,j} \sum_{\text{replicas } \vec{n}} \frac{q_i \cdot q_j}{|\vec{r_{ij}} + \vec{n}|}$$

i and j run over all atoms in the unit cell (r_{ij} – their distance)

infinite sum with special convergence problems
alternating harmonic series $\sum_n (-1)^n / n$ – conditionally convergent:
  it converges $\sum_{i=1}^{\infty} a_i < \infty$, but
  does not converge absolutely: $\sum_{i=1}^{\infty} |a_i| = \infty$
  – convergence is slow and dependent on the order of summation

BTW: conditionally convergent series

$$\begin{aligned}
\mathsf{I}: \quad S &= 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + \frac{1}{7} - \frac{1}{8} + \ldots \\
\mathsf{II}: \frac{1}{2}S &= \quad + \frac{1}{2} \qquad - \frac{1}{4} \qquad + \frac{1}{6} \qquad - \frac{1}{8} + \ldots \\
\mathsf{I} + \mathsf{II}: \frac{3}{2}S &= 1 \qquad + \frac{1}{3} - \frac{1}{2} + \frac{1}{5} \qquad + \frac{1}{7} - \frac{1}{4} + \frac{1}{9} + \ldots \\
&= 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + \ldots = S \quad \text{(sic!)}
\end{aligned}$$

Sum it up not so simply – Ewald

any smart way to calculate ESP induced by all images of all atoms?

$$\Phi(\vec{r_i}) = \sum_{j} \sum_{\text{replicas } |\vec{n}|} \frac{q_j}{|\vec{r_{ij}} + \vec{n}|}$$

idea: pass to a sum of two series that will converge rapidly:

$$\sum \frac{1}{r} = \sum \frac{f(r)}{r} + \sum \frac{1 - f(r)}{r}$$

may seem awkward, but will work well ☺

how can we implement this idea?

Sum it up not so simply – Ewald

summing over point charges – difficult (convergence problem)
Ewald method uses Gaussian densities of the same magnitude:

$$q_j \;\rightarrow\; q_j \cdot \left(\frac{\alpha}{\sqrt{\pi}}\right)^3 \exp\left[-\alpha^2 \cdot |\vec{r}_j|^2\right]$$

ESP for a Gaussian charge density:

$$\Phi(\vec{r}) = q_j \cdot \frac{\mathrm{erf}\,[\alpha \cdot r]}{r}$$

erf: “error function”

$$\mathrm{erf}[x] = \frac{2}{\sqrt{\pi}} \int_0^x \exp[-t^2]\,\mathrm{d}t \qquad \mathrm{erfc}[x] = 1 - \mathrm{erf}[x]$$

Sum it up not so simply – Ewald

with Ewald: summing ESP induced by all charges, we obtain

$$\Phi(\vec{r_i}) = \sum_j \sum_{\text{replicas } |\vec{n}|} q_j \cdot \frac{\text{erf}\left[\alpha \cdot |\vec{r_{ij}} + \vec{n}|\right]}{|\vec{r_{ij}} + \vec{n}|} \rightarrow \sum \frac{f(r)}{r}$$

do not forget: we have to add a correction – difference of potentials induced by Gaussians and by point charges

$$\rightarrow \sum \frac{1 - f(r)}{r}$$

Sum it up not so simply – Ewald

the full ESP induced by all replicas of all charges:

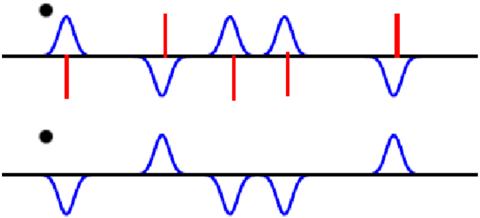
$$\begin{aligned}\Phi(\vec{r}_i) &= \sum_j \sum_{\text{replicas } |\vec{n}|} q_j \cdot \frac{\text{erfc}\left[\alpha \cdot |\vec{r}_{ij} + \vec{n}|\right]}{|\vec{r}_{ij} + \vec{n}|} \\ &+ \sum_j \sum_{\text{replicas } |\vec{n}|} q_j \cdot \frac{\text{erf}\left[\alpha \cdot |\vec{r}_{ij} + \vec{n}|\right]}{|\vec{r}_{ij} + \vec{n}|} \\ &= \Phi^{\text{real}}(\vec{r}_i) + \Phi^{\text{rec}}(\vec{r}_i)\end{aligned}$$

$\Phi^{\text{real}}(\vec{r}_i)$ – real-space contribution

– from a certain, quite small distance (depending on α):
point charges and the charge densities cancel each other
– this contribution vanishes and we can use cut-off here

Ewald – two contributions

real-space contribution to the Ewald sum
– original point charges (red) and Gaussian charge densities (blue)
of the same magnitude but opposite sign



– a distant Gaussian ‘looks much like’ a point charge,
and the difference of ESP goes to zero – cut-off is justified

Ewald – 2nd contribution

reciprocal-space contribution:

the total charge density is periodic → it may be meaningful to Fourier-transform the calculation to the reciprocal space

$$\begin{tikzcd}
\text{(density)} \arrow[r, "\text{Fourier transform}"] \arrow[d, dotted, "\text{solve}"', "\text{(hard)}"] & \text{(transformed density)} \arrow[d, "\text{solve}"', "\text{(easy)}"] \\
\text{(potential)} & \text{(transformed potential)} \arrow[l, "\text{inverse Fourier transform}"']
\end{tikzcd}$$

$$\hat{f}(\xi) = \int_{-\infty}^{\infty} f(x) \cdot \exp[-2\pi i \cdot x \cdot \xi] \, \mathrm{d}x$$

$$\begin{aligned}
\mathrm{Re}\, \hat{f}(\xi) &= \int_{-\infty}^{\infty} f(x) \cdot \cos[x \cdot \xi] \, \mathrm{d}x \\
\mathrm{Im}\, \hat{f}(\xi) &= \int_{-\infty}^{\infty} f(x) \cdot \sin[x \cdot \xi] \, \mathrm{d}x
\end{aligned}$$

Ewald – 2nd contribution

$\Phi^{\mathrm{rec}}(\vec{r}_i)$ – reciprocal-space contribution

- with 'reciprocal' vector $\vec{k} = \left(k_x \cdot \frac{2\pi}{L_x}, k_y \cdot \frac{2\pi}{L_y}, k_z \cdot \frac{2\pi}{L_z}\right),\ k_i \in \mathcal{Z}$
- best evaluated in the form

$$\Phi^{\mathrm{rec}}(\vec{r}_i) = \frac{4\pi}{V} \cdot \sum_{\vec{k} \neq \vec{o}} \frac{1}{\vec{k}^2} \cdot \exp\left[-\frac{|\vec{k}|^2}{4\alpha^2}\right] \cdot \sum_j q_j \cdot \exp[-\mathrm{i} \cdot \vec{k} \cdot \vec{r}_{ij}]$$

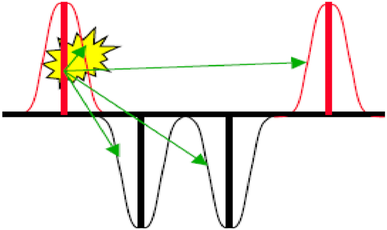
- terms decrease with increasing $|\vec{k}|$ quickly – cut-off possible
- converges fast with large Gaussian width α
- the value of α is a compromise between
  the requirements of real- and reciprocal-space calculations

both contributions – favorable convergence behavior → we can evaluate electrostatic interactions with atoms in all periodic images

Ewald – the last contribution

broadened charge density interacts with itself

and this energy must be subtracted from the final result



Coulomb self-energy of a broadened Gaussian:

$$E^{\mathrm{self}} = \sum_j q_j \cdot \Phi(\vec{o}) = \sum_j q_j \cdot q_j \cdot \frac{\alpha}{\sqrt{\pi}}$$

Ewald – complete expression for energy

3 contributions:

- ‘real-space’

$$E^{\text{real}} = \frac{1}{2} \sum_j q_j \cdot \Phi^{\text{real}}(\vec{r_j})$$

- ‘reciprocal-space’

$$E^{\text{rec}} = \frac{1}{2} \sum_j q_j \cdot \Phi^{\text{rec}}(\vec{r_j})$$

- ‘self-energy’

$$E^{\text{Ewald}} = E^{\text{real}} + E^{\text{rec}} - E^{\text{self}}$$

Ewald – optional additional contribution

Surface / dipole term

- for systems with zero charge and dipole moment $\vec{\mu}_{\text{tot}} \neq \vec{o}$

$$\Phi^{\text{sur}}(\vec{r}_i) = \frac{4\pi}{3} \frac{\vec{r}_i \cdot \vec{\mu}_{\text{tot}}}{V}$$

- this describes the situation with surrounding vacuum
- universal application may lead to problems
  when mobile ions cross the box boundaries
  (abrupt changes of $\vec{\mu}_{\text{tot}}$)
- if this term is not present
  – 'tin-foil' boundary conditions – surrounding $\varepsilon = \infty$

Thinking about Ewald

Ewald summation – correct Coulomb interaction energy at higher computational cost (compared to cut-off):

- scales with the number of atoms as $\mathcal{O}(N^2)$
- with a better algorithm – $\mathcal{O}(N^{\frac{3}{2}})$
- not efficient enough for large-scale simulations
- goal – improved efficiency of the long-range sum (reciprocal-space contribution)

particle–mesh Ewald method (1993)
– combines ideas from crystallography (Ewald method) and plasma physics (particle–mesh method)
– key to success – 3D fast Fourier transform technique

Long–range energy with PME

PME works with a regular grid constructed in the simulation box

step 1

convert the point charges to Gaussian charge densities and spread on the grid in the form of splines

practically, we need to have charges discretized on the grid points

if an atom is close to the edge of the box, a part of its charge must be put to the opposite side of the box (PBC)

Long–range energy with PME

step 2

Fourier transform the charge density on the grid

– discrete 3D fast Fourier transform technique

solve Poisson's eqn in the reciprocal space

→ energy and Fourier transform of potential

$$E^{\mathrm{rec}} = \frac{1}{2} \sum_{k_1=0}^{K_1-1} \sum_{k_2=0}^{K_2-1} \sum_{k_3=0}^{K_3-1} Q(k_1, k_2, k_3) \cdot (\Theta^{\mathrm{rec}} \star Q)(k_1, k_2, k_3)$$

3D-FFT used to calculate the convolution $\Theta^{\mathrm{rec}} \star Q$

– this corresponds to the ESP in reciprocal space

Θ^{rec} depends on box size and character of splines

Long–range energy with PME

step 3

get the potential in real space (inverse Fourier transform),
interpolate its derivative to calculate the forces

– expressed in terms of splines – analytical calculation

step 4

get E^{real} and E^{self} – directly from the presented expressions

step 5

attention – the reciprocal energy/forces include contributions from atom pairs that are connected with bonds
– these have to be subtracted afterwards (excluded)

$$E^{\text{rec}}_{\text{excl}} = -\sum_{i,j}^{\text{list}} \frac{1}{4\pi\varepsilon_0} \frac{q_i \cdot q_j}{r_{ij}} \, \text{erf}\left[\alpha \cdot r_{ij}\right]$$

Long–range energy with PME

PME parameters: spacing of grid ca. 1 Å,
  $\alpha^{-1} \approx$ 2.5 Å $\rightarrow$ short-range cutoff $\leq$ 10 Å possible

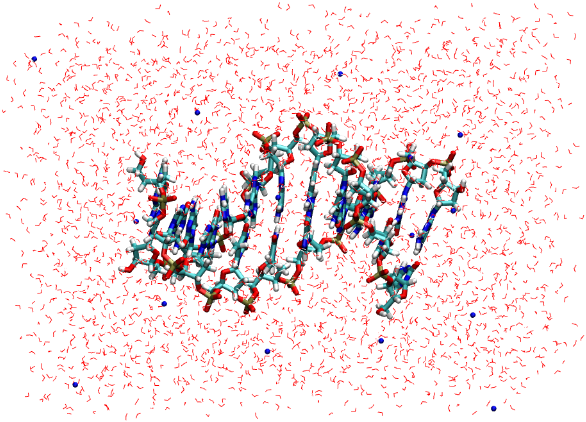
neighbor lists used for the real-space interactions
  $\rightarrow$ linear scaling of real-space calculation ($\mathcal{O}(N)$)

complexity of the long-range PME component:
  $\mathcal{O}(N \cdot \log N)$ due to the efficiency of FFT
  modern implementations – nearly as efficient as cut-off!

Water in biomolecular simulations

most simulations – something in aqueous solutions

H_2O – usually (many) thousands molecules



Water in biomolecular simulations

most simulations – something in aqueous solutions
H_2O – usually (many) thousands molecules

example – simulation of DNA decanucleotide:

- PBC box 3.9 × 4.1 × 5.6 nm (smallest meaningful)
- 630 atoms in DNA, 8346 atoms in water and 18 Na^+
- concentration of DNA: 18 mmol/L – very high!
- of all pair interactions: 86 % are water–water,
  most of the others involve water

Water models

most interactions involve H_2O

→ necessary to pay attention to its description

model of water must be simple enough (computational cost)

and accurate enough, at the same time

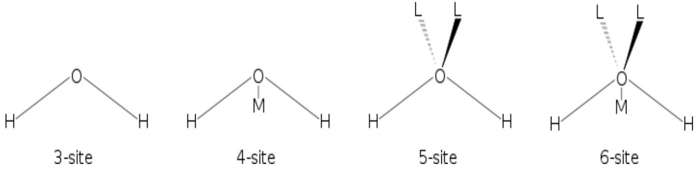
water models – usually rigid

– bond lengths and angles do not vary – constraints

molecule with three sites (atoms in this case), or up to six sites

– three atoms and virtual sites corresponding

to a ‘center’ of electron density or lone electron pairs



Water models

TIP3P (very similar is SPC)

- most frequently used
- 3 atoms with 3 rigid bonds, charge on every atom ($-0.834/+0.417$)
- only the O possesses non-zero LJ parameters (optimization)

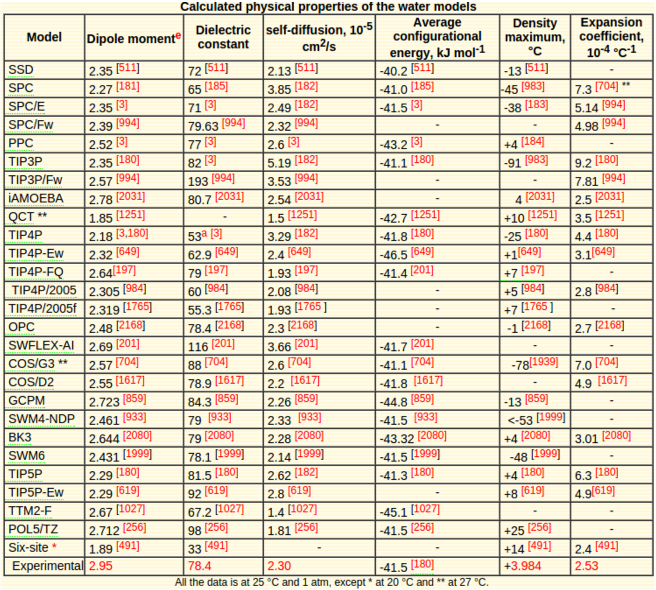
TIP4P

- negative charge placed on virtual site M rather than on the O
- electric field around the molecule described better

TIP5P

- 2 virtual sites L with negative charges near the O – lone pairs
- better description of directionality of H-bonding etc.
  (radial distribution function, temperature of highest density)

Water models



Calculated physical properties of the water models

Model	Dipole moment[e]	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]	-45 [983]	7.3 [704] **
SPC/E	2.35 [3]	71 [3]	2.49 [182]	-41.5 [3]	-38 [183]	5.14 [994]
SPC/Fw	2.39 [994]	79.63 [994]	2.32 [994]	-	-	4.98 [994]
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]	-91 [983]	9.2 [180]
TIP3P/Fw	2.57 [994]	193 [994]	3.53 [994]	-	-	7.81 [994]
iAMOEBA	2.78 [2031]	80.7 [2031]	2.54 [2031]	-	4 [2031]	2.5 [2031]
QCT **	1.85 [1251]	-	1.5 [1251]	-42.7 [1251]	+10 [1251]	3.5 [1251]
TIP4P	2.18 [3,180]	53[a] [3]	3.29 [182]	-41.8 [180]	-25 [180]	4.4 [180]
TIP4P-Ew	2.32 [649]	62.9 [649]	2.4 [649]	-46.5 [649]	+1 [649]	3.1 [649]
TIP4P-FQ	2.64 [197]	79 [197]	1.93 [197]	-41.4 [201]	+7 [197]	-
TIP4P/2005	2.305 [984]	60 [984]	2.08 [984]	-	+5 [984]	2.8 [984]
TIP4P/2005f	2.319 [1765]	55.3 [1765]	1.93 [1765]	-	+7 [1765]	-
OPC	2.48 [2168]	78.4 [2168]	2.3 [2168]	-	-1 [2168]	2.7 [2168]
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]	-78 [1939]	7.0 [704]
COS/D2	2.55 [1617]	78.9 [1617]	2.2 [1617]	-41.8 [1617]	-	4.9 [1617]
GCPM	2.723 [859]	84.3 [859]	2.26 [859]	-44.8 [859]	-13 [859]	-
SWM4-NDP	2.461 [933]	79 [933]	2.33 [933]	-41.5 [933]	<-53 [1999]	-
BK3	2.644 [2080]	79 [2080]	2.28 [2080]	-43.32 [2080]	+4 [2080]	3.01 [2080]
SWM6	2.431 [1999]	78.1 [1999]	2.14 [1999]	-41.5 [1999]	-48 [1999]	-
TIP5P	2.29 [180]	81.5 [180]	2.62 [182]	-41.3 [180]	+4 [180]	6.3 [180]
TIP5P-Ew	2.29 [619]	92 [619]	2.8 [619]	-	+8 [619]	4.9 [619]
TTM2-F	2.67 [1027]	67.2 [1027]	1.4 [1027]	-45.1 [1027]	-	-
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]
Experimental	2.95	78.4	2.30	-41.5 [180]	+3.984	2.53

All the data is at 25 °C and 1 atm, except * at 20 °C and ** at 27 °C.

Continuum electrostatics methods

Situation up to now

- molecules in an explicit solvent
- all interactions between atoms involved
- polarizability / permittivity of the solvent
  - present in the simulation as a consequence of interactions and dynamics
- for instance, solvation free energy is involved "by the way"
  - if desired, may be evaluated with special methods

Continuum electrostatics methods

Example – polypeptide in the α-helix and β-sheet conformations.

The free energy difference of the two structures is given by

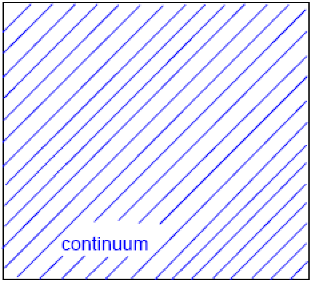
- the difference of internal energies / enthalpies
- the entropic contributions – above all vibrational entropy
- the difference of free energies of solvation

α-helix: much larger dipole moment than β-sheet
$\rightarrow$ α-helix is better solvated in a polar medium (H_2O)
$\rightarrow$ crucial effect of solvation on the equilibrium
between conformations of solvated peptide

Motivation: the amount of solvent becomes excessive easily,
so it may be meaningful to abandon explicit solvent representation,
and apply an implicit model instead

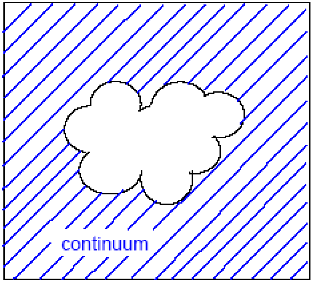
Continuum electrostatics methods

Solvation free energy: $\Delta G_{\text{solv}} = \Delta G_{\text{cav}} + \Delta G_{\text{vdW}} + \Delta G_{\text{ele}}$



Continuum electrostatics methods

Solvation free energy: $\Delta G_{\mathrm{solv}} = \Delta G_{\mathrm{cav}} + \Delta G_{\mathrm{vdW}} + \Delta G_{\mathrm{ele}}$

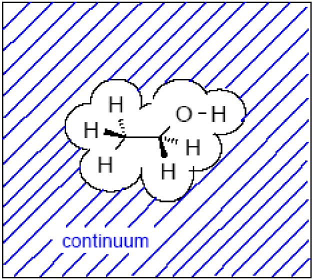


- A cavity in the solvent is formed
  - rearrangement of the solvent molecules

  ΔG_{cav}: decrease of S and loss of solvent–solvent interactions

Continuum electrostatics methods

Solvation free energy: $\Delta G_{\text{solv}} = \Delta G_{\text{cav}} + \Delta G_{\text{vdW}} + \Delta G_{\text{ele}}$



- A cavity in the solvent is formed
  - rearrangement of the solvent molecules

  ΔG_{cav}: decrease of S and loss of solvent–solvent interactions
- solute–solvent interaction – van der Waals and electrostatic

Solvent-accessible surface area

SASA – important concept

- solvent-exposed surface of molecule as a solid body
- reasonable approx.: ΔG_{cav} and ΔG_{vdW} proportional to SASA.
- total surface composed from surfaces of individual atoms S_i
- then: $\Delta G_{\mathrm{cav}} + \Delta G_{\mathrm{vdW}} = \sum_i c_i \cdot S_i$
- alternative: obtain SASA by rolling a ball of a certain diameter (typically 2.8 Å to mimic H_2O) on the molecular surface

Solvent-accessible surface area

When does it work?

- if the electrostatic effect of the surrounding solvent dominates (shielding of solvent-exposed charged side chains of proteins)
- not if there is specific solute–solvent interaction (like hydrogen bonding)

Difficult example: dynamics of small peptides dissolved in water

– competition between various hydrogen-bonding patterns

Continuum electrostatics methods

Big question: how to calculate ΔG_{ele}?

often used is the term “reaction field”

$$\Delta G_{\text{ele}} = q \cdot \Phi_{\text{rf}}(\vec{r})$$

for moving the cavity with the solute from vacuo to the solvent

Born and Onsager models

Born: the work to bring charge q from vacuo
into spherical cavity of radius a
in solvent with dielectric constant ε:

$$\Delta G_{\text{ele}} = -\frac{q^2}{2a}\left(1 - \frac{1}{\varepsilon}\right)$$

ε: 1 for vacuo (thus $\Delta G_{\text{ele}} = 0$), 80 for water, 2 to 20 for protein

Onsager and Kirkwood: model for dipole μ in cavity

$$\begin{aligned}\Phi_{\text{rf}} &= \frac{2(\varepsilon - 1)}{2\varepsilon + 1} \cdot \frac{1}{a^3} \cdot \mu \\ \Delta G_{\text{ele}} &= -\frac{1}{2}\Phi_{\text{rf}} \cdot \mu\end{aligned}$$

Born and Onsager models

- simple models
- implemented in many standard programs
- quite unrealistic approximations even for small molecules

Extensions:

- polarizable continuum model (PCM) –
  arbitrary surfaces constructed
  with the use of vdW radii of individual atoms
- conductor-like screening models (COSMO) –
  polarization of the dielectric (insulating) solvent
  derived from scaled-conductor approximation.

Poisson–Boltzmann equation (PBE)

For big molecules, the simple models may be too simple and inefficient at the same time :-(

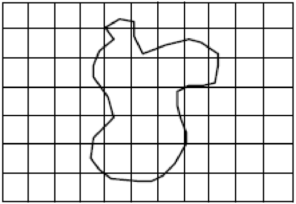
other approximations – starting from Poisson's equation

$$\nabla \varepsilon \nabla \Phi = -4 \pi \rho$$

given – charge distribution ρ and dielectric constant ε
to be found – potential Φ

possibility to solve:

– discretize on a 3D grid,
use finite differences
calc. Φ on every grid point
iteratively



Ions in the solvent

ions are very important – counterions compensate charged solute,
or salt mimicks physiologic conditions

the position of ions depends on the potential:

$$\rho_{\text{ions}} = \sum_i q_i \cdot c_i \cdot \exp\left[-\frac{q_i \cdot \Phi(r)}{k_\text{B} T}\right]$$

or: anions like to be where $\Phi > 0$, and cations like $\Phi < 0$

an additional term appears in Poisson's equation:
linearized Poisson–Boltzmann equation at low ionic strength:

$$\nabla \varepsilon \nabla \Phi = -4\pi\rho + \varepsilon \cdot \kappa^2 \cdot \Phi(r)$$

with the Debye–Hückel parameter $\kappa^2 = \frac{8\pi q^2 I}{\varepsilon \cdot k_\text{B} T}$
(ionic strength $I = \frac{1}{2} \sum_i c_i z_i^2$, c_i concentration, z_i charge of ion i)

Ions in the solvent – PBE

- charge distribution on the protein
  polarizes the dielectric outside ("solvent")
  → screening of any solvent-exposed charges of the protein
  effectively, charges pointing into the solvent will vanish nearly
- solvent ions will distribute to make
  the overall charge distribution more uniform
  if a negative charge points into the solvent,
  a cation will be located close to it

The solvent around a protein should always be taken into account.

PBE – not efficient enough to be calculated in every MD step
→ approximations are necessary

Generalized Born model (GB)

idea – use the simple Born equation for MM atomic charges

$$\Delta G^1_{\text{ele}} = -\left(1 - \frac{1}{\varepsilon}\right) \sum_i \frac{q_i^2}{2a_i}$$

the interaction of individual charges changes in solution

$$\begin{aligned} E_{\text{ele}} &= \frac{1}{2} \sum_{i \neq j} \frac{1}{\varepsilon} \frac{q_i \cdot q_j}{r_{ij}} = \\ &= \frac{1}{2} \sum_{i \neq j} \frac{q_i \cdot q_j}{r_{ij}} - \frac{1}{2} \left(1 - \frac{1}{\varepsilon}\right) \sum_{i \neq j} \frac{q_i \cdot q_j}{r_{ij}} \end{aligned}$$

giving another contribution to solvation free energy

$$\Delta G^2_{\text{ele}} = -\frac{1}{2} \left(1 - \frac{1}{\varepsilon}\right) \sum_{i \neq j} \frac{q_i \cdot q_j}{r_{ij}}$$

solvation free energy $= \Delta G^1_{\text{ele}} + \Delta G^2_{\text{ele}}$

Generalized Born model (GB)

problem 1 – Born's formula holds for interaction of charges located in spherical cavities (with radii a_i)

– only valid for charged bodies of general shapes if $r_{ij} \gg a_i + a_j$

– two extreme cases are covered:

$$E = \begin{cases} \frac{q_i^2}{a_i}, & \text{if } i = j \text{ (`self-interaction, i.e. solvation energy)} \\ \frac{q_i \cdot q_j}{r_{ij}}, & \text{if } i \neq j \text{ and } r_{ij} \to \infty \end{cases}$$

what to do at intermediate distances (2 Å to 10 Å)? interpolate!

$$f(r_{ij}) = \sqrt{r_{ij}^2 + a_i a_j \exp\left[-\frac{r_{ij}^2}{4 a_i a_j}\right]}$$

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

Generalized Born model (GB)

Born's equation holds for a charged particle in contact with solvent

problem 2 – many charges are buried deeply inside the protein,
far from the solvent!
→ solvation free energy may be overestimated heavily

possible solution – scale up a_i in a reasonable way!

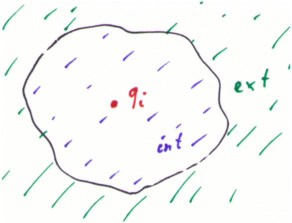
the most important task when using the GB method
– to use/calculate reasonable radii a_i

How to get the radii in GB

approximate interaction energy of a charge q_i in the protein interior with the solvent:

$$\Delta G^i_{\text{ele}} = -\frac{1}{8\pi}\left(1 - \frac{1}{\varepsilon_W}\right)\int_{\text{ext}} \frac{q_i^2}{r^4}\,\mathrm{d}V$$

integration runs over
the 'exterior' of the protein



comparing with the Born formula, we find

$$\frac{1}{a_i} = \frac{1}{4\pi}\int_{\text{ext}} \frac{1}{r^4}\,\mathrm{d}V$$

r – distance from the charge to the 'boundary' of the protein.

How to get the radii in GB

several GB models exist; generally, $\int_{\text{ext}}$ transformed to $\int_{\text{int}}$

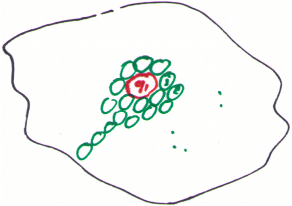
- GB molecular volume – with van der Waals radius α_i:

$$\frac{1}{a_i} = \frac{1}{\alpha_i} - \frac{1}{4\pi} \int_{\text{int},r>\alpha_i} \frac{1}{r^4}\,\mathrm{d}V$$

  – possibly longish calculation time

- pairwise models – the interior ≈ union of atomic spheres

$$\frac{1}{a_i} = \frac{1}{\alpha_i} - \sum_{j\neq i} \frac{1}{4\pi} \int_{\text{sphere}\,j} \frac{1}{r^4}\,\mathrm{d}V$$



  – this is insufficient because of partial overlap / void places

How to get the radii in GB

several GB models exist; generally, $\int_{\mathrm{ext}}$ transformed to $\int_{\mathrm{int}}$

- GB molecular volume – with van der Waals radius α_i:

$$\frac{1}{a_i} = \frac{1}{\alpha_i} - \frac{1}{4\pi} \int_{\mathrm{int}, r>\alpha_i} \frac{1}{r^4} \, \mathrm{d}V$$

  – possibly longish calculation time

- pairwise models – the interior ≈ union of atomic spheres
  empirical formula may be used instead:

$$\begin{aligned}\frac{1}{a_i} &= \frac{1}{\lambda \cdot R_{\mathrm{vdW},i}} - P_1 \frac{1}{R^2_{\mathrm{vdW},i}} - \sum_j^{\mathrm{bond}} \frac{P_2 V_j}{r_{ij}^4} - \sum_j^{\mathrm{angle}} \frac{P_3 V_j}{r_{ij}^4} \\ &- \sum_j^{\mathrm{nonbond}} \frac{P_4 V_j}{r_{ij}^4} \cdot \mathrm{CCF}(P_5, r_{ij})\end{aligned}$$

MM-PBSA

- another application of implicit solvent models
- free energies of binding of ligands to biomolecules
- post-processing approach to evaluate free energies
- a normal MD simulation is run,
  and free energies are computed a posteriori

binding free energy obtained component-wise with various methods
solvation free energy – with Poisson–Boltzmann or so
non-polar contribution – SASA-dependent terms
configurational entropy – normal-mode analysis

very approximative, yet may still give results of good quality