How to model water in simulations

explicitly or implicitly

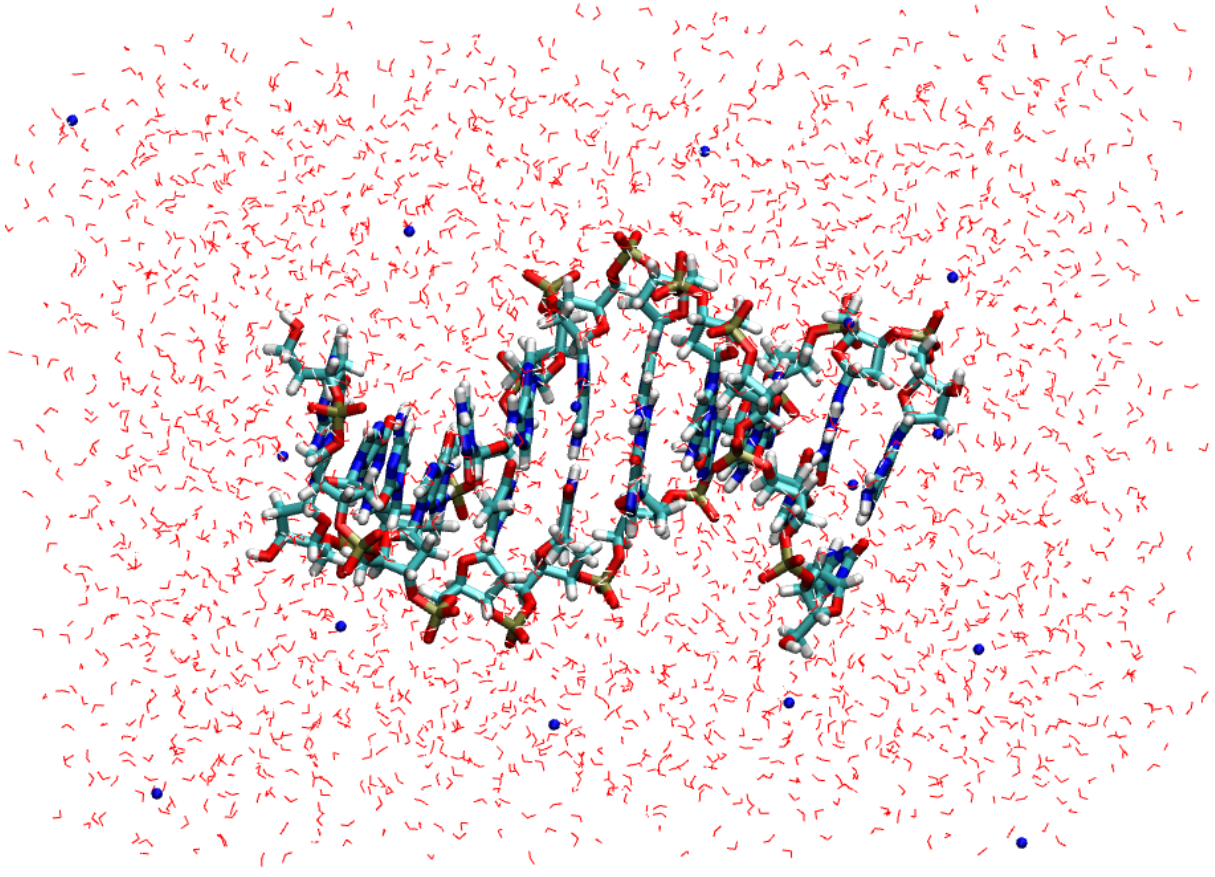
Marcus Elstner and Tomáš Kubař

2019, June 12

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 $3.9 \times 4.1 \times 5.6$ 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

$\rightarrow$ necessary to pay attention to its description

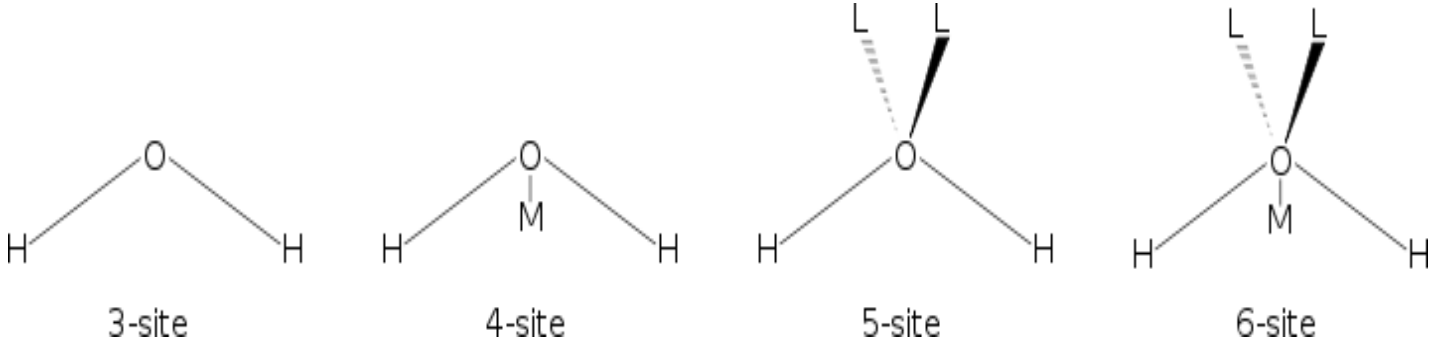
model of water must be simple enough (computational cost)
and accurate enough, at the same time

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

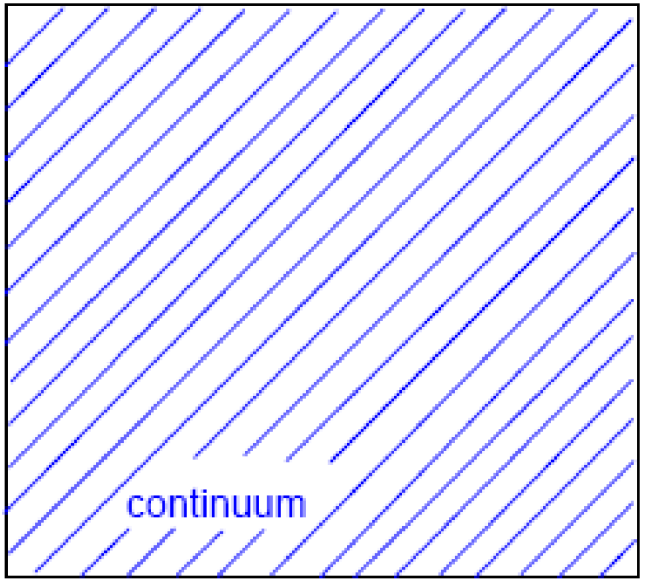
- the difference of internal energies / enthalpies
- the entropic contributions – above all vibrational entropy
- the difference of free energies of solvation

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

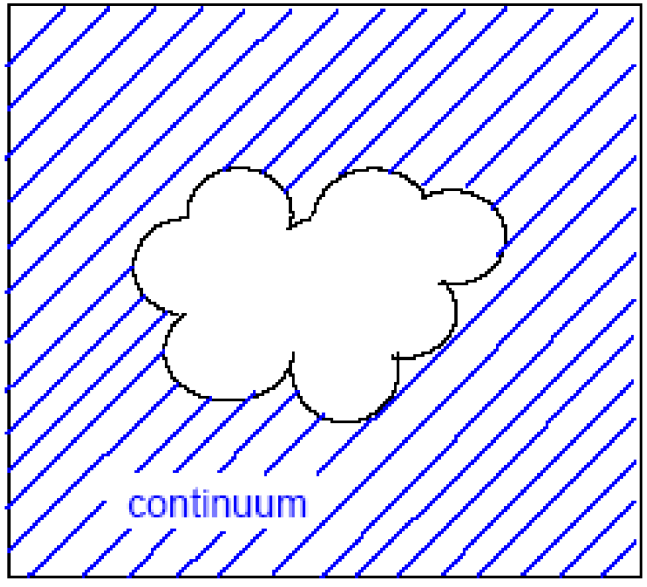
Continuum electrostatics methods

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

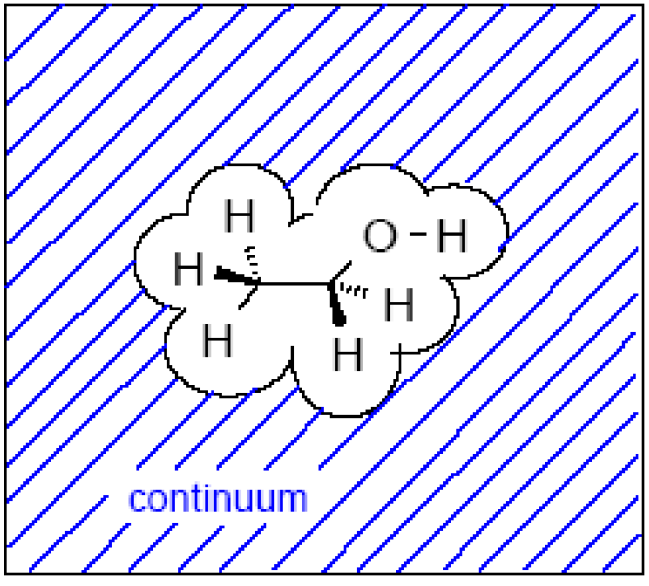


- A cavity in the solvent is formed
  – rearrangement of the solvent molecules

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

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: $\Delta G_{\text{cav}} + \Delta G_{\text{vdW}} = \sum_i c_i \cdot S_i$
- 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 :-(

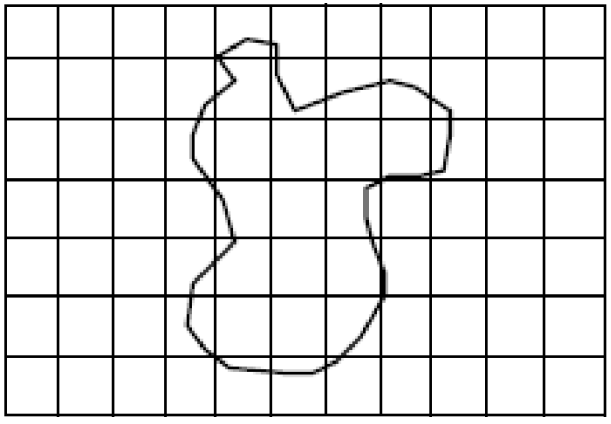
other approximations – starting from Poisson's equation

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

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:

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

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 $\kappa^2 = \frac{8\pi q^2 I}{\varepsilon \cdot k_B T}$
(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

$$E_{\text{ele}} = \frac{1}{2} \sum_{i \neq j} \frac{1}{\varepsilon} \frac{q_i \cdot q_j}{r_{ij}} =$$

$$= \frac{1}{2} \sum_{i \neq j} \frac{q_i \cdot q_j}{r_{ij}} - \frac{1}{2} \left(1 - \frac{1}{\varepsilon}\right) \sum_{i \neq j} \frac{q_i \cdot q_j}{r_{ij}}$$

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:

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

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

$$\Delta G_{\text{ele}} = -\frac{1}{2}\left(1 - \frac{1}{\varepsilon}\right) \cdot \sum_{i,j} \frac{q_i \cdot q_j}{f(r_{ij})} \qquad f(r_{ij}) = \sqrt{r_{ij}^2 + a_i a_j \exp\left[-\frac{r_{ij}^2}{4 a_i a_j}\right]}$$

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

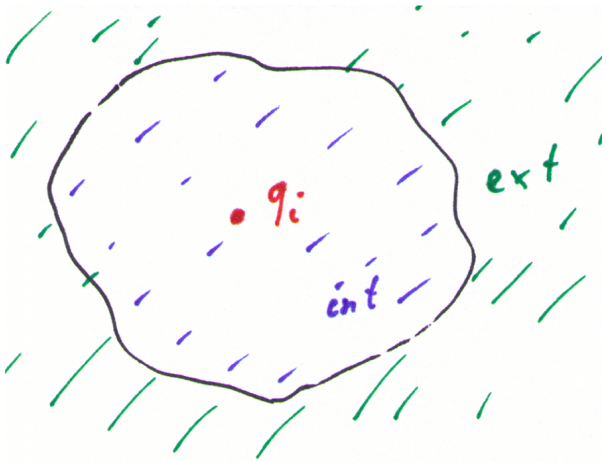
possible solution – scale up a_i in a reasonable way!

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:

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



integration runs over the 'exterior' of the protein

comparing with the Born formula, we find

$$\Delta G^1_{\text{ele}} = -\frac{1}{2}\left(1 - \frac{1}{\varepsilon}\right)\frac{q_i^2}{a_i} \quad \rightarrow \quad \frac{1}{a_i} = \int_{\text{ext}} \frac{1}{r^4}\,\mathrm{d}V$$

r – distance from the charge to the point in the exterior of the protein

How to get the radii in GB

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

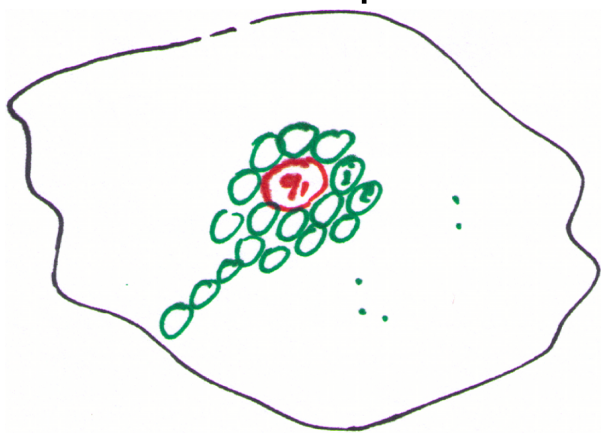
- GB molecular volume – with van der Waals radius α_i:

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

  – possibly longish calculation time

- pairwise models – the interior $\approx$ union of atomic spheres

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



  – this is insufficient because of partial overlap / void places

How to get the radii in GB

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

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

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

  – possibly longish calculation time

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

$$\frac{1}{a_i} = \frac{1}{\lambda \cdot R_{\text{vdW},i}} - P_1 \frac{1}{R^2_{\text{vdW},i}} - \sum_j^{\text{bond}} \frac{P_2 V_j}{r_{ij}^4} - \sum_j^{\text{angle}} \frac{P_3 V_j}{r_{ij}^4}$$

$$- \sum_j^{\text{nonbond}} \frac{P_4 V_j}{r_{ij}^4} \cdot \text{CCF}(P_5, r_{ij})$$

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