Force-field development

Chris Oostenbrink

Institute of molecular modeling and simulation

University of Natural Resources and Life Sciences

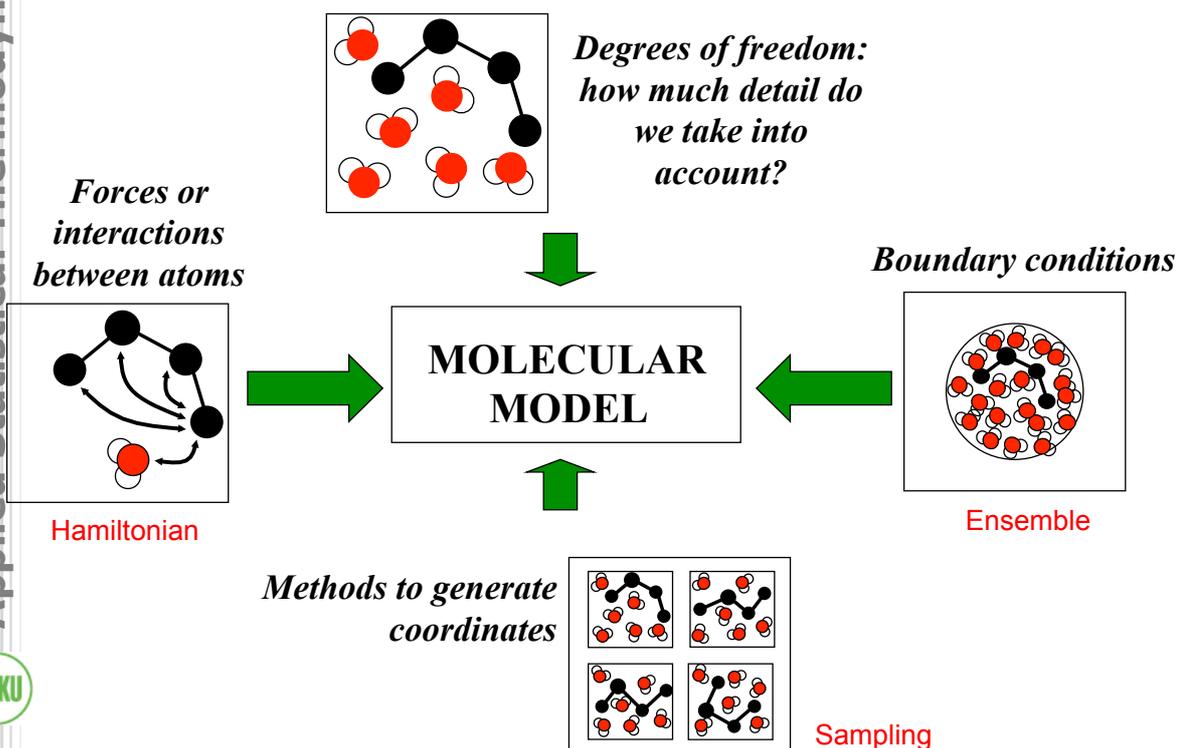**BOKU** Institute of molecular modeling and simulation

---

Schedule

Time	Mon. 20/11/17	Tue. 21/11/17	Wed. 22/11/17	Thur. 23/11/17	Fri. 24/11/17	Mon. 27/11/17	Tue. 28/11/17	Wed. 29/11/17	Thur. 30/11/17	Fri. 01/12/17
9:00	Free Time	Welcome JAG	Free Time	Free Time	Free Time	Free Time	Free Time	Free Time	Free Time	Free Time
9:30		Lecture 1: Overview and Introduction JAG	Lecture 4: Thermodynamics JAG	Lecture 7: Classical Mechanics I JAG	Lecture 10: Ensembles I CO	Lecture 13: Free energies: reaction coordinates JAG	Lecture 16: Calculating properties from simulations JAG	Lecture 19: Electrostatics WvG	Lecture 22: On the ethics of the academic endeavour: where do we go?	Lecture 24: Left-overs/ questions and future perspectives WvG
10:15		Break	Break	Break	Break	Break	Break	Break		Break
10:30		Lecture 2: Molecular Simulations MD/SD/MC CO	Lecture 5: Force-Field Development CO	Lecture 8: Classical Mechanics II JAG	Lecture 11: Ensembles II JAG	Lecture 14: Boundary Conditions I WvG	Lecture 17: Comparison with Experiments WvG	Lecture 20: Polarization WvG	WvG	Lecture 25: Students plans I
11:15		Coffee Break	Coffee Break	Coffee Break	Coffee Break	Coffee Break	Coffee Break	Coffee Break	Coffee Break	Coffee Break
11:45		Lecture 3: How to simulate using GROMOS CO	Lecture 6: Structure Refinement CO	Lecture 9: Analizing with GROMOS CO	Lecture 12: Free energies: alchemistry CO	Lecture 15: Boundary Conditions II WvG	Lecture 18: Searching & Enhanced Sampling WvG	Lecture 21: Multi-resolution simulations WvG	Lecture 23: QM/MM WvG	Lecture 26: Students plans II
12:30		Break for lunch, self-study, discussion.*	Break for lunch, self-study, discussion.*	Break for lunch, self-study, discussion.*	Break for lunch, self-study, discussion.*	Break for lunch, self-study, discussion.*	Break for lunch, self-study, discussion.*	Break for lunch, self-study, discussion.*	Break for lunch, self-study, discussion.*	Tutorial 10: **
14:00	Registration	Tutorial 2: Running MD GROMOS Tutorial	Tutorial 3: Running MD Students Plans	Tutorial 4: Running MD Students Plans	Tutorial 5: Running MD Students Plans	Tutorial 6: Statistical Mechanics exercises	Tutorial 7 Analizing MD: GROMOS Tutorial	Tutorial 8: Analizing MD: Students Plans	Tutorial 9: Analizing MD: Students Plans	How to prepare a barbecue
	Tutorial 1 Linux, NLHPC and OS-dongle installation									
16:00	Coffee Break	Coffee Break	Coffee Break	Coffee Break	Coffee Break	Coffee Break	Coffee Break	Coffee Break	Coffee Break	
17:30	End of session	End of session	End of session	End of session	End of session	End of session	End of session	End of session	End of session	
20:30	Free time	Free time	Free time	Free time	Beer, Science & Friendship Good Stock Bar	Free time	Free time	Free time	Free time	Farewell

Outline

- Biomolecular force fields
  - Interactions: bonded, nonbonded, cutoffs

- What do we parameterize on
  - Structural data, thermodynamic data

- GROMOS parameterization
  - Free energies of solvation
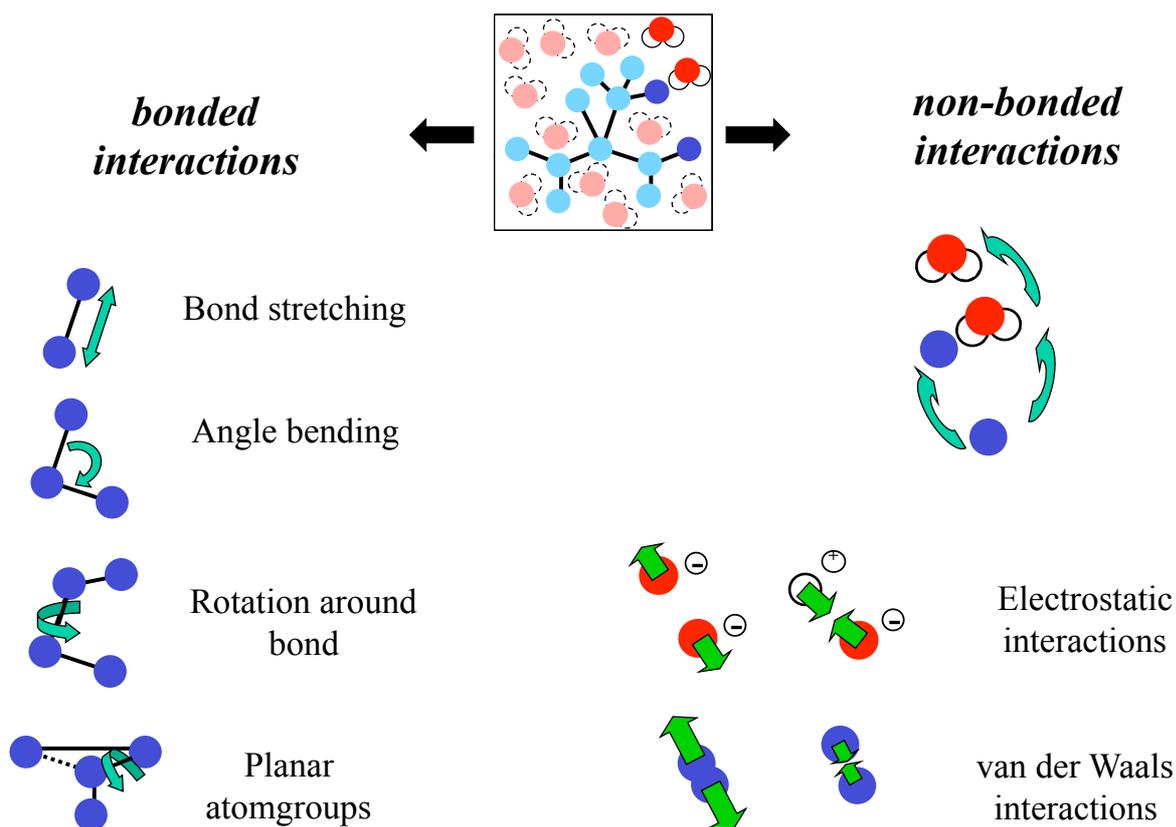

A model for molecular computations

Degrees of freedom: how much detail do we take into account?

Forces or interactions between atoms

MOLECULAR MODEL

Boundary conditions

Hamiltonian

Ensemble

Methods to generate coordinates

Sampling

Born-Oppenheimer approximation

- The motion of atomic nuclei is very much slower than the motion of the electrons

- We can describe the (energy of the) electrons separately, given the positions of the atomic nuclei

- The electrons determine a potential energy surface in which the nuclei move

Molecular structure ⟷ Energy

Force Field: interaction

bonded interactions ⟵ ⟶ *non-bonded interactions*

Bond stretching

Angle bending

Rotation around bond

Planar atomgroups

Electrostatic interactions

van der Waals interactions

Interacting Particles

- **Physical Terms**

$$V^{bond}\left(\mathbf{r}^N\right) = \sum_{bonds\ i} \frac{1}{2} K_i^b \left[b_i\left(\mathbf{r}^N\right) - b_i^0 \right]^2 \qquad V^{v.d.Waals}\left(\vec{r}^N\right) = \sum_{pairs\ i<j} 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6} \right]$$
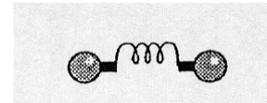
$$V^{angle}\left(\mathbf{r}^N\right) = \sum_{angles\ i} \frac{1}{2} K_i^a \left[\theta_i\left(\mathbf{r}^N\right) - \theta_i^0 \right]^2 \qquad V^{Coulomb}\left(\vec{r}^N\right) = \sum_{pairs\ i<j} \frac{1}{4\pi\varepsilon_0\varepsilon_r} \frac{q_i q_j}{r_{ij}}$$

$$V^{torsion}\left(\mathbf{r}^N\right) = \sum_{torsion\ i} K_i^\varphi \cos\left[m_i \varphi_i\left(\mathbf{r}^N\right) + \delta_i \right] \qquad V^{pol}\left(\vec{r}^N\right) = \text{N-body polarization energy}$$

- **Special interaction Terms**
  - For example restraints on the system
    - From experiments (see lecture 11)
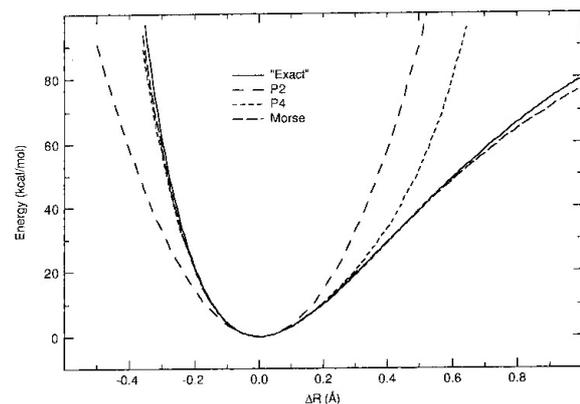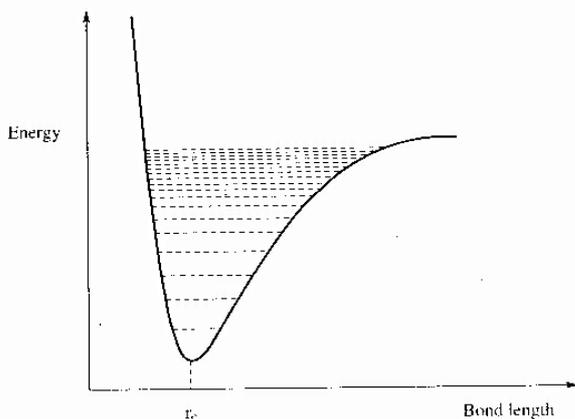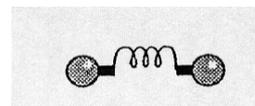   - To bias the simulation

---

Bond stretch term

- The bond stretch term is almost always approximated with a harmonic potential:

$$V^{bond}\left(\mathbf{r}^N\right) = \sum_{bonds\ i} V_i^{bond}\left(\mathbf{r}^N\right)$$

$$V_i^{bond}\left(\mathbf{r}^N\right) = \frac{1}{2} K_i^b \left[b_i\left(\mathbf{r}^N\right) - b_i^0 \right]^2$$

Bond stretch term

- The potential energy function may be made more accurate by including additional terms

$$V_i^{bond}\left(\mathbf{r}^N\right) = \frac{1}{2}K_i^b\left[b_i\left(\mathbf{r}^N\right) - b_i^0\right]^2 + K3_i^b\left[b_i\left(\mathbf{r}^N\right) - b_i^0\right]^3 + K4_i^b\left[b_i\left(\mathbf{r}^N\right) - b_i^0\right]^4$$

- Sometimes people use the accurate Morse potential
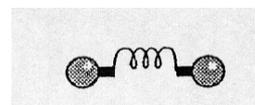
$$V_i^{bond}\left(\mathbf{r}^N\right) = D_i\left[1 - e^{-\alpha_i\left(b_i(\mathbf{r}) - b_i^0\right)}\right]^2$$

- Which is expensive and often not necessary

- In GROMOS we use a quartic bond

$$V_i^{bond}\left(\mathbf{r}^N\right) = \frac{1}{2}K_i^{qb}\left[b_i^2\left(\mathbf{r}^N\right) - \left(b_i^0\right)^2\right]^2$$


Bond stretch term

- Forces between bound atoms are usually large and it takes a lot of energy to deviate the bond length from its optimal value

Bond	b^0 (nm)	K^{qb} (10^6 kJ mol^{-1} nm^{-4})
CHn - CHn	0.153	7.15
C - C	0.139	10.8
C = O	0.123	16.6
C - H	0.109	12.3
CHn - OA	0.143	8.18
NR (His) - Fe	0.221	0.540

- At room temperature, deviations of up to 0.006 nm are possible

- Often, we use SHAKE to constrain the bond lengths

Bond angle term
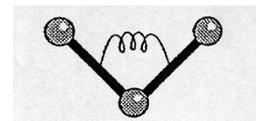
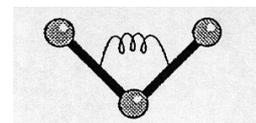- The bond angle term is also often approximated with a *harmonic potential*:

$$V^{angle}\left(\mathbf{r}^N\right) = \sum_{angles\ i} V_i^{angle}\left(\mathbf{r}^N\right)$$

$$V_i^{angle}\left(\mathbf{r}^N\right) = \frac{1}{2} K_i^a \left[\theta_i\left(\mathbf{r}^N\right) - \theta_i^0\right]^2$$

- It may also be made more accurate by including additional terms

- In GROMOS we use

$$V_i^{angle}\left(\mathbf{r}^N\right) = \frac{1}{2} K_i^{ca} \left[\cos\theta_i\left(\mathbf{r}^N\right) - \cos\theta_i^0\right]^2$$

Bond angle term

Angle	θ^0 (deg)	K^{ca} (kJ mol^{-1})
CHn-CHn-CHn	109.5	520
C-C-C	120.0	560
C-C-H	120.0	505
O-C-N	124.0	730
CHn-N-C	117.0	635
CH2-S-S	104.0	490

- At room temperature deviations of about 10 degrees are possible

Torsion term
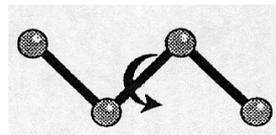
- The dihedral torsional term for φ is approximated by one or several cosine functions:

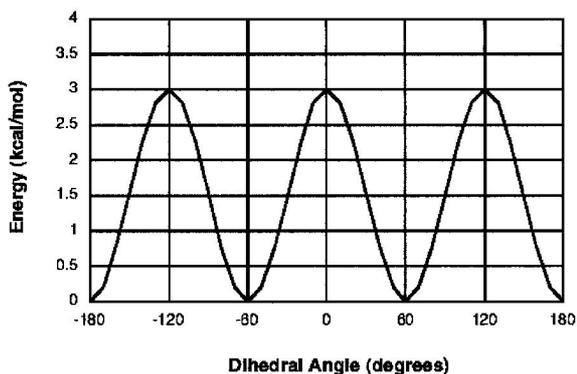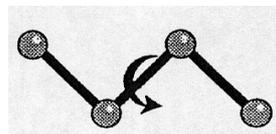$$V^{torsion}\left(\mathbf{r}^{N}\right) = \sum_{torsion\ i} V_{i}^{torsion}\left(\mathbf{r}^{N}\right)$$

$$V_{i}^{torsion}\left(\mathbf{r}^{N}\right) = K_{i}^{\varphi}\cos\left[m_{i}\varphi_{i}\left(\mathbf{r}^{N}\right)+\delta_{i}\right]$$
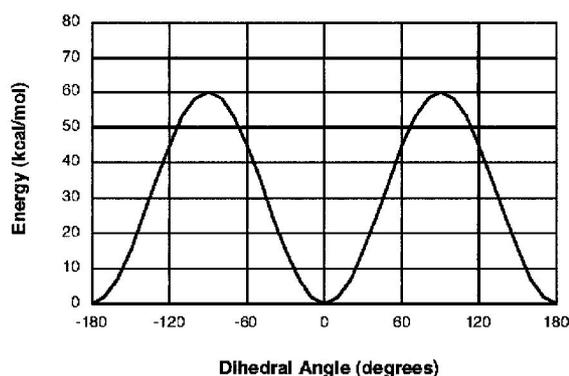
K^{φ} = amplitude or barrier

m_{i} = periodicity (number of minima over 360 °)
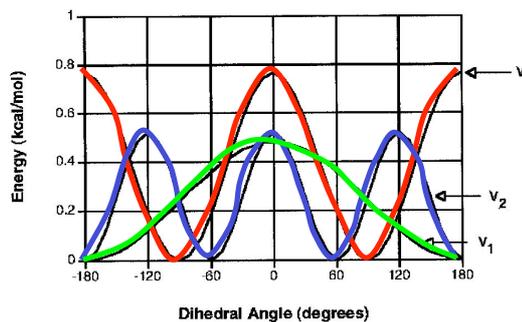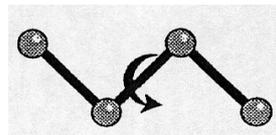
δ_{i} = phase (shifting of the cosine)

ethane: sp3-sp3 bond

CH3-CH3

amplitude = 3 kcal/mol

periodicity = 3

phase = 0 degrees

ethene: sp2-sp2 bond

CH2=CH2

amplitude = 60 kcal/mol

periodicity = 2

phase = 90 degrees

Torsion term

sp3-sp3 bond
Z-CH$_2$-CH$_2$-Z

period. = 3, phase = 0 º
period. = 2, phase = 0 º
period. = 1, phase = 0 º
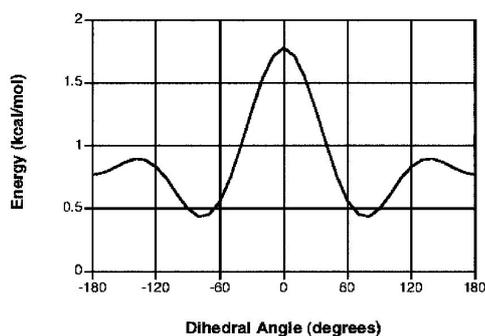
Combined torsional term
for Z-CH$_2$-CH$_2$-Z

Non-bonded interactions
Also interfere

Out-of-plane bending

- Is again approximated by a *harmonic potential:*

$$V^{oop}\left(\mathbf{r}^N\right) = \sum_{oop\ i} V_i^{oop}\left(\mathbf{r}^N\right)$$

$$V_i^{oop}\left(\mathbf{r}^N\right) = \frac{1}{2} K_i^{oop}\left[\chi_i\left(\mathbf{r}^N\right) - \chi_i^0\right]^2$$

- χ can also be defined as a improper dihedral
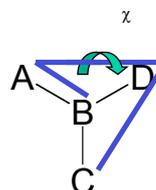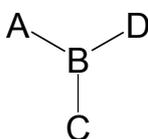
Nonbonded interactions

- All atoms see each other through space
    - ~ ½ N (N – 1)   interactions

- Bound atoms (1,2 neighbours) and their neighbours (1,3 neighbours) are often excluded because the bonded interactions already take these into account
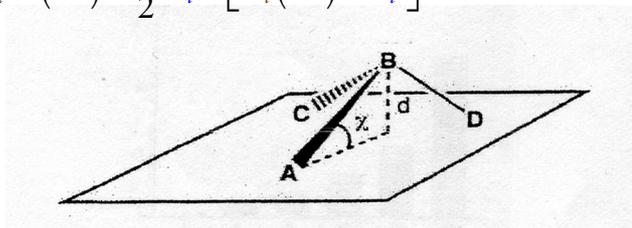- Sometimes 1,4 neighbours are treated differently to keep the proper torsional profile

- Often we work with a cutoff for the nonbonded interaction:

Black particles sees green particles only
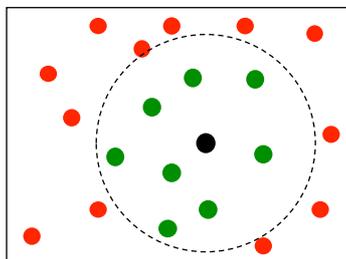
Nonbonded interactions

- For 5000 atoms:

nonbond interactions (10^6)

| cutoff distance (Å) | 6 | 8 | 10 | 12 | 14 | none |
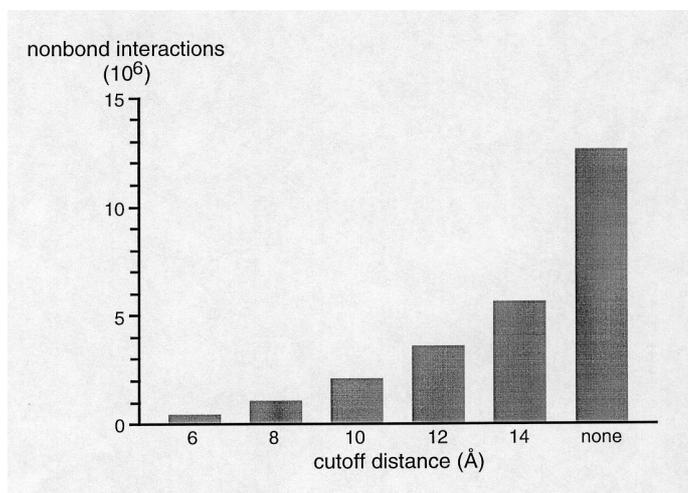
- Nonbonded interactions are usually the rate determining step in simulations

Vanderwaals interactions

- Lennard Jones potential:
  - r^{-6} term from dispersion attraction
  - r^{-12} term empirically (simple)

$$V^{v.d.Waals}\left(\mathbf{r}^{N}\right) = \sum_{pairs\ i<j} 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right]$$
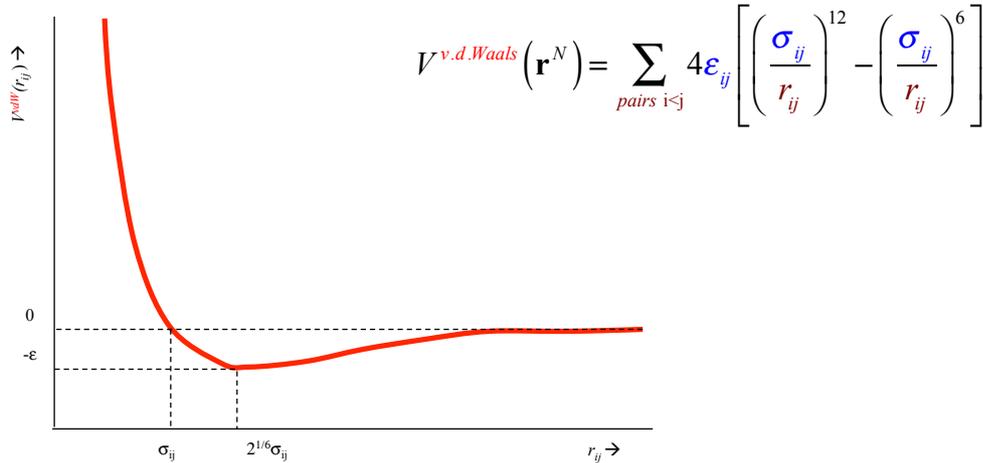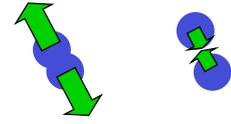
A cutoff of 1.4 nm does not make big errors

Electrostatic interactions

- Coulomb potential

$$V^{Coulomb}\left(\mathbf{r}^{N}\right) = \sum_{pairs\ i<j} \frac{1}{4\pi\varepsilon_0\varepsilon_r}\frac{q_i q_j}{r_{ij}}$$
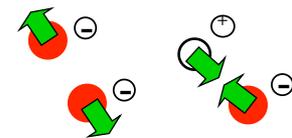
q_i, q_j = partial charges for atoms i and j

ε_0 = dielectrice constant of vacuum

ε_r = relative dielectric constant of the medium

Much larger mistake with a 1.4 nm cutoff

- **Possible improvements**:
  - charge-group based cutoff
  - reaction-field contributions
  - lattice-sum methods to use periodicity

See lectures 14 and 15

Other interactions

- Some force fields show alternative interactions

    - Specific hydrogen bond interactions
      usually a balance between vdw and electrostatics

    - Crossterms, e.g. Energy as function of bond length and angle: $U(r,q)$
      needed to reproduce vibrational spectra

    - Electronic polarization
      various models are possible, usually need multiple iterations

    - …

Outline

- Biomolecular force fields
    - Interactions: bonded, nonbonded, cutoffs

- What do we parameterize on
    - Structural data, thermodynamic data

- GROMOS parameterization
    - Free energies of solvation

(Biomolecular) force field development

- GROMOS force field
  - Parameterized on primary experimental data
  - Simple functional form using as few parameters as possible
    - ~50 atom, bond, bond angle, dihedral types
  - Transferable parameters
    - Experienced users can 'guess' reasonable parameters

- With increased computational power, parameterization became more elaborate
  - small molecule crystallography, infrared spectroscopy
    - bond lengths, angles, force constants
  - simulations of liquids
    - density, heat of vaporisation (D, ε, η, τ_2, α, …)
  - simulations of mixtures
    - Excess properties
    - Free energy of solvation in different media

Latest GROMOS force field

- 52 Bond stretch types: 104 parameters
- 54 Bond angle types: 108 parameters
- 45 Torsional angle types: 135 parameters
- 5 Out-of-plane types: 10 parameters

- 54 Van der Waals types: ~150 parameters
- Charges for ~40 groups: ~60 parameters

                                        ~550 parameters

All need to be parameterised to work consistently together

The force field problem

The Force Field Problem

A very small (free) energy differences ($k_B T$ = 2.5 kJ/mol)
resulting from summation over very many contributions (atoms)

$\downarrow$ $\downarrow$

$10^6 - 10^8$ must be very accurate

B accounting for entropic effects
not only energy minima are of
importance but whole range of
x-values with energies ~$k_B T$

must be included in the
force field parameter calibration

energy
E(x)

may have higher energy
but lower free energy
than

coordinate x

The force field problem

C size problem
the larger the system, the more accurate the individual energy
contributions (from atoms) must be to reach the same overall accuracy

Calibrate force field using *thermodynamic* data for *small* molecules in the *condensed* phase
keep force field physical + simple   transferable
computable

Force field development

Choice of calibration set of data, systems, phase, and properties

Type of data	Type of system	Phase	Type of properties	Force field parameters
structural data (exp.)	small molecules	crystalline solid phase	molecular geometry: bond lengths, bond angles	b_0, θ_0, χ_0
spectroscopic data (exp.)	small molecules	gas phase	intra-molecular vibrations: force constants	K_b, K_θ, K_χ
energy profiles (theor.)	small molecules	gas phase	quantum-chemical calculation of torsional-angle rotational profiles	K_φ, δ, m
electrondensities (theor.)	small molecules	gas phase	quantum-chemical calculation of atom charges	charges q_i (initial)
thermodynamic data (exp.)	small molecules, mixtures, solutions	condensed phase	heat of vaporisation, density, partition coefficients, free energy of solvation	van der Waals $C_{12}(i,j)$, $C_6(i,j)$ q_i (final)
dielectric data (exp.)	small molecules	condensed phase	dielectric permittivity, relaxation	charges q_i
transport data (exp.)	small molecules	condensed phase	transport coefficients: diffusion, viscosity	$C_{12}(i,j)$, $C_6(i,j)$ q_i

---

Outline

Applied Statistical Thermodynamics

- Biomolecular force fields
  - Interactions: bonded, nonbonded, cutoffs

- What do we parameterize on
  - Structural data, thermodynamic data

- GROMOS parameterization
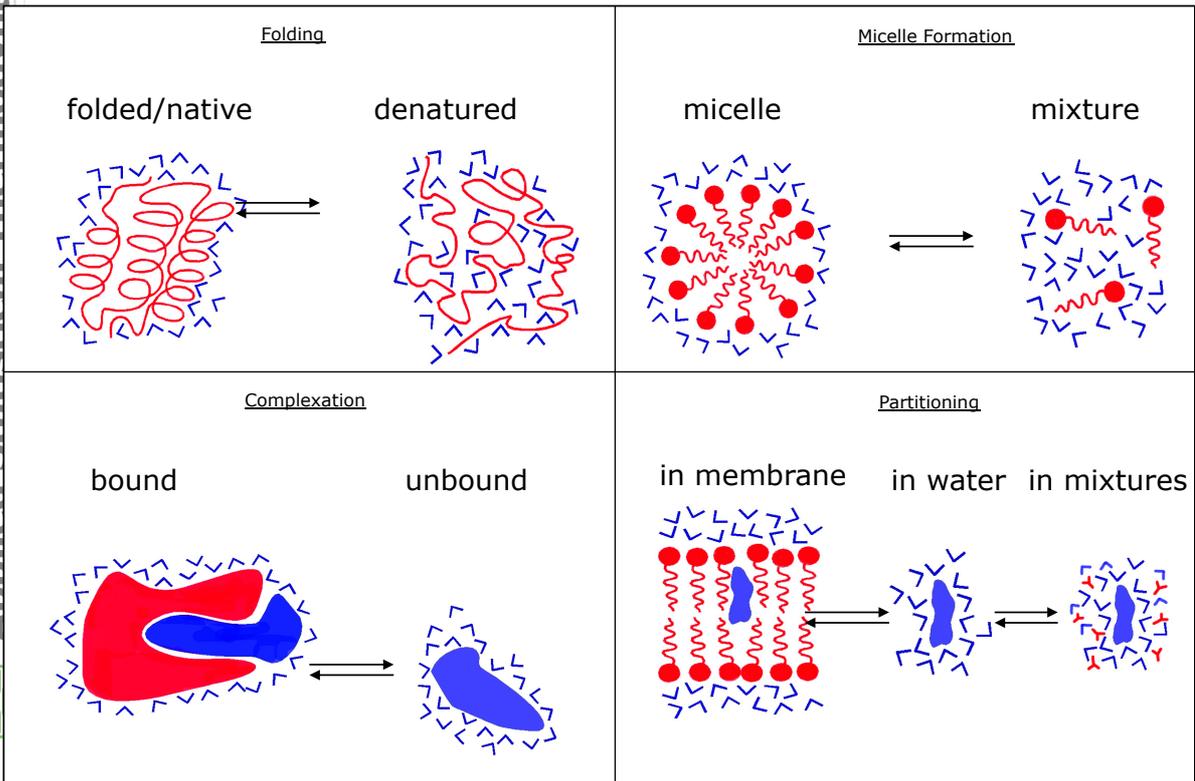 - Free energies of solvation

BOKU

Amino acid solvation

- All polar amino acids were too hydrophobic!
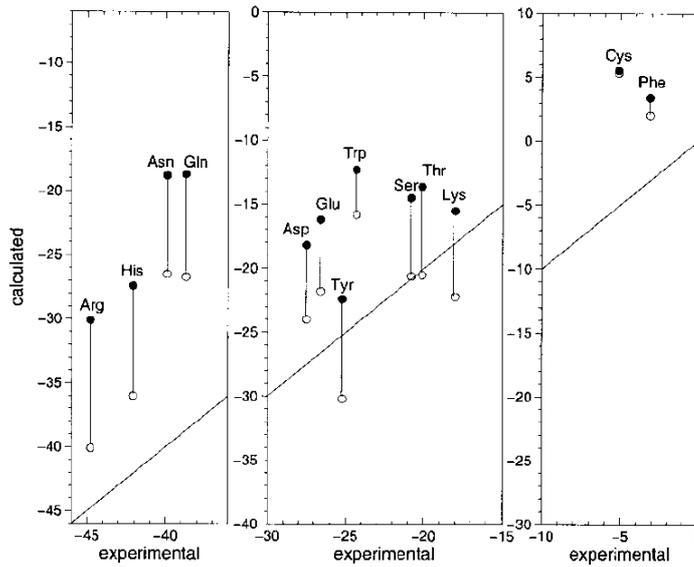
Figure 4. Hydration free energies (kJ mol^{-1}) at T = 293 K: experimental vs. calculated using the GROMOS96 force field (43a2) (●) and the Q1.1 force field (○).

A. Villa, A.E. Mark, J. Comput. Chem. 2002, **23**, 548-553

Processes: Thermodynamic Equilibria

Folding

folded/native   denatured

Micelle Formation

micelle   mixture

Complexation

bound   unbound

Partitioning

in membrane   in water   in mixtures

Force field parameterization

1. structural data

- large molecules: crystal structures
                                 solution structures of proteins

2. thermodynamic data

- small molecules: heat of vaporization, density
                                 in *condensed phase*
                                 free energy of solvation in different solvents
                                 partition coefficients
                                 ε, D, η, etc.

3. theoretical data

- small molecules: electrostatic potential and gradient
                                 in *gas phase*

Determination of FF Parameters

Calibration sets of small molecules

1. Non-polar molecules

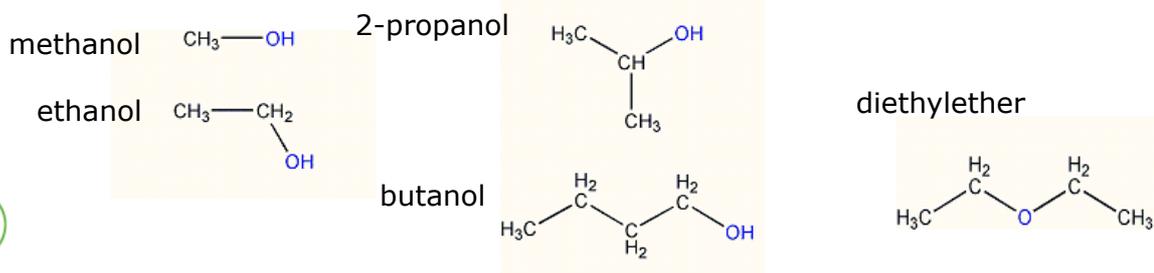   L.D. Schuler, X. Daura, W.F. van Gunsteren, J. Comput. Chem. 2002, **22**, 1205-1218

2. Polar molecules

   C. Oostenbrink, A. Villa, A.E. Mark, W.F. van Gunsteren, J. Comput. Chem. 2004, **25**, 1656-1676

3. Ionic molecules

   M. Reif, P. Hünenberger and C. Oostenbrink, J. Chem. Theory Comp. 2012, **8**, 3705 – 3723

Polar molecules: Calibration set: 21 compounds, e.g.

methanol CH_3—OH   2-propanol

ethanol CH_3—CH_2 diethylether
                     OH

butanol

Determination of Force Field Parameters

force field parameter set: 17 parameters

quantities to reproduce:

for all compounds (liquids)
- heat of vaporisation
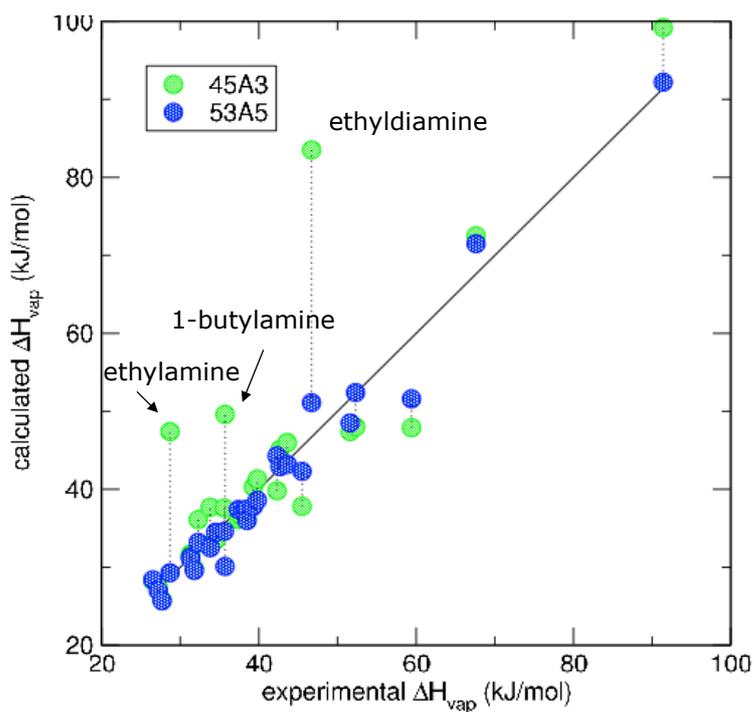- density

for analogues of polar amino acid sidechains:
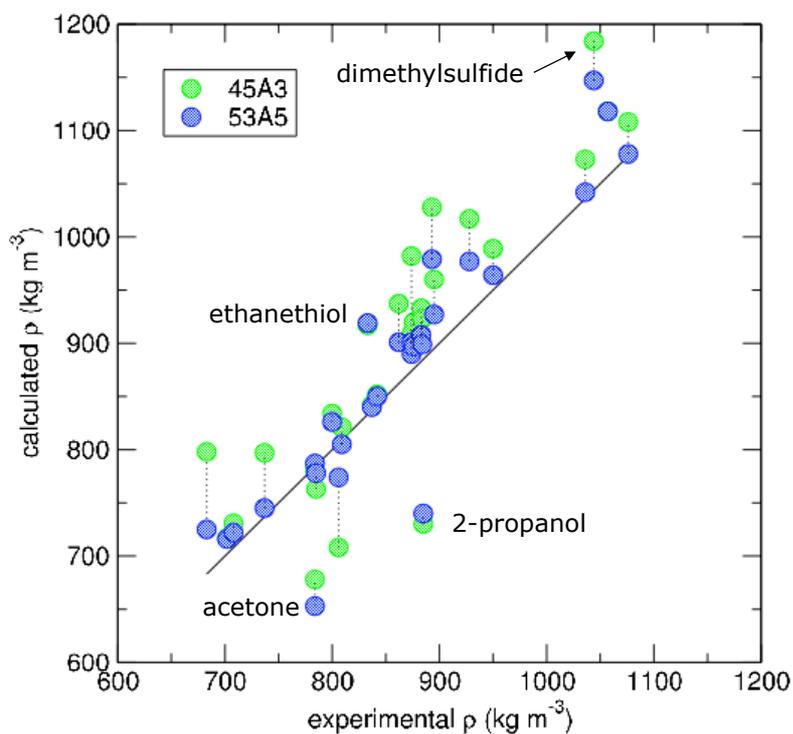- free enthalpy of solvation:

in cyclohexane
in water

C. Oostenbrink, A. Villa, A.E. Mark, W.F. van Gunsteren, J. Comput. Chem. 2004, **25**, 1656-1676
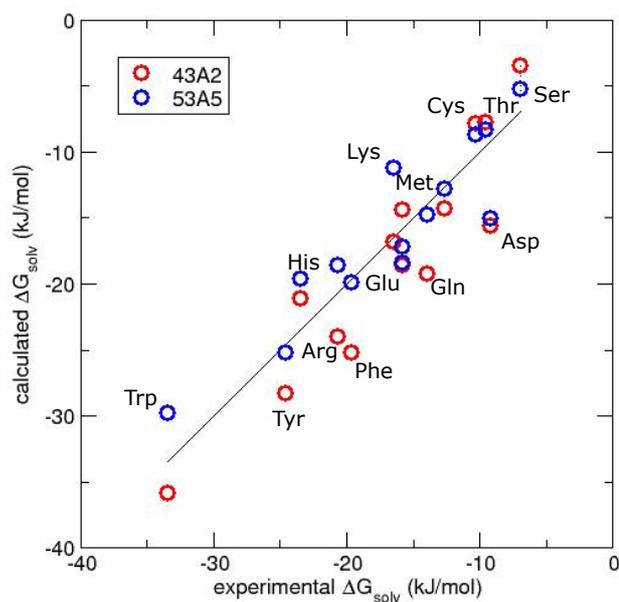

Heat of Vaporization for Pure Liquids

average absolute deviation: 1.9 kJ/mol (53A5)

Density for Pure Liquids

average absolute deviation: 4.0% (53A5)

Free Energy of Solvation in Cyclohexane
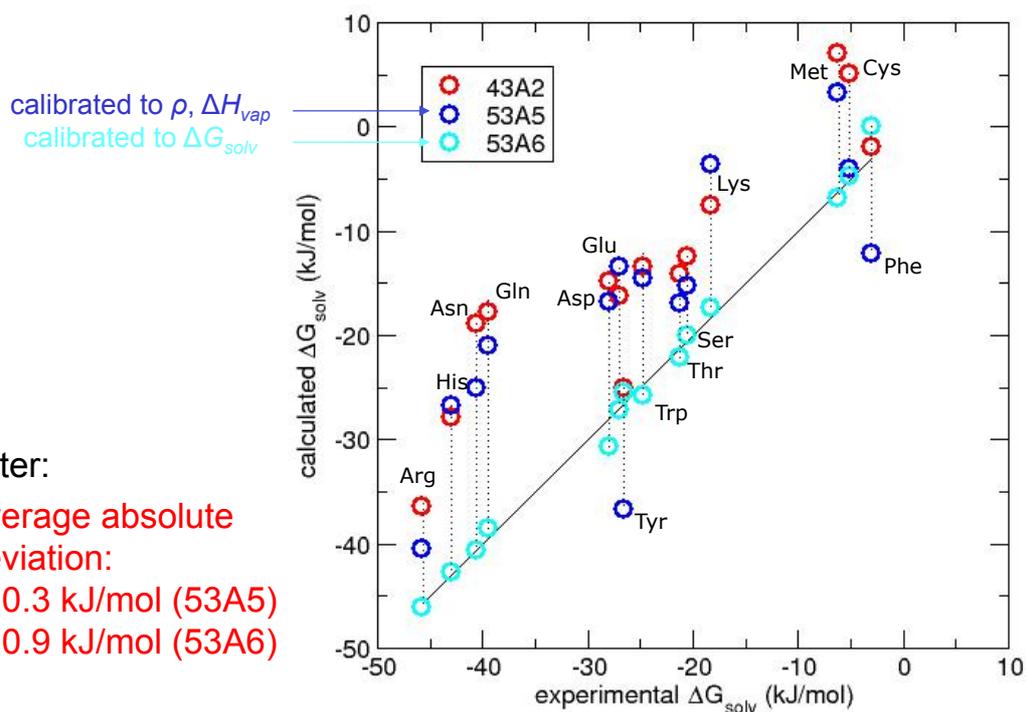

average absolute deviation: 2.2 kJ/mol (53A5)

Applied Statistical Thermodynamics

Free Energy of Solvation in Water

calibrated to ρ, ΔH_{vap}
calibrated to ΔG_{solv}

In water:

Average absolute deviation:
10.3 kJ/mol (53A5)
0.9 kJ/mol (53A6)

C. Oostenbrink, A. Villa, A.E. Mark, W.F. van Gunsteren. *J. Comput. Chem.* **25**, 1656 – 1676 (2004)

GROMOS force fields

- Parameter sets in GROMOS get a code, e.g. 54A8
  - 54 different atom types (multiple C's, O's, N's, etc.)
  - A for condensed phase, B for vacuum simulations
  - 8 sequential number

- Force field history:
  - 26C1   First reported GROMOS parameter set (1982)
  - 37C4   Standard condensed phase in GROMOS87
  - 43A1   Re-optimized alkanes and some solvents
  - 45A4   Re-optimized DNA, Heme, carbohydrates
  - 53A5   All polar groups revised, optimized for pure liquids
  - 53A6   All polar groups revised, optimized for mixtures
  - 54A7   Backbone torsions re-optimized
 - 54A8   Charged groups optimized for free energy of solvation

Different flavours
current developments

- $53A6_{OXY}$ Balanced compromise between 53A5 and 53A6 for oxygen containing compounds
- _carbo   Force field files specifically for carbohydrates
- _beta   Force field filed specifically for beta-peptides

- Inclusion of electronic polarization: charge on a spring (COS)

Induced dipole
-q,+q connected by spring

Atom
q_i (permanent)
Vdw interaction
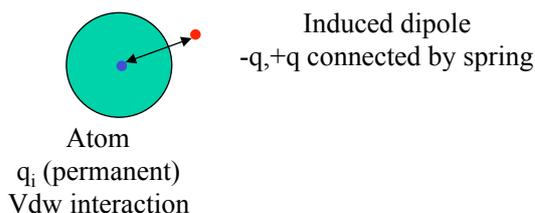
See lecture 17

- Coarse graining: reduce number of degrees of freedom
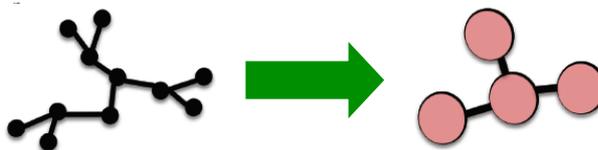
See lecture 19

Building a new topology

- Parameterize your molecule
  - Relevant experimental data available?

- Transferability of parameters:
  - select types according to functional groups

- Manual assignment of parameters

- Automated topology builder
  A.K. Malde, L. Zuo, M. Breeze, M. Stroet, D. Poger, P.C. Nair, C. Oostenbrink and A.E. Mark, J. Chem. Theor. Comput. 2011, **7**, 4026 – 4037

- **Always check the parameters for consistency**

Conclusions

- **Theoretical description of force fields**
  - Bonds, Angles: harmonic oscillator
  - Dihedral angles: cosine series
  - Van der Waals: Lennard-Jones interaction with a cutoff
  - Electrostatic interactions: Coulomb interactions
    - Long range: lattice sum or reaction field

- **Force field parameterisation**
  - Parameterization against primary experimental data
  - Latest versions: free energy of solvation in water and cyclohexane
  - Summary of GROMOS force fields