All-atom Molecular Mechanics

Trent E. Balius

AMS 535 / CHE 535

09/27/2010

Outline

- Molecular models

- Molecular mechanics
- Force Fields
- Potential energy function
  - functional form
  - parameters and parameterization

- Motivations and applications
- Biomolecular Force Fields

Molecular Models

- Quantum Mechanics
  - physical, but expensive
  - Schrödinger equation: $H \Psi = E \Psi$
  - wave functions defines electron density

- Molecular Mechanics
  - less physical --> empirical parameterization
  - cheap and accurate

Empirical Force Fields for Biological Macromolecules: Overview and Issues

ALEXANDER D. MACKERELL, JR.
Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland,
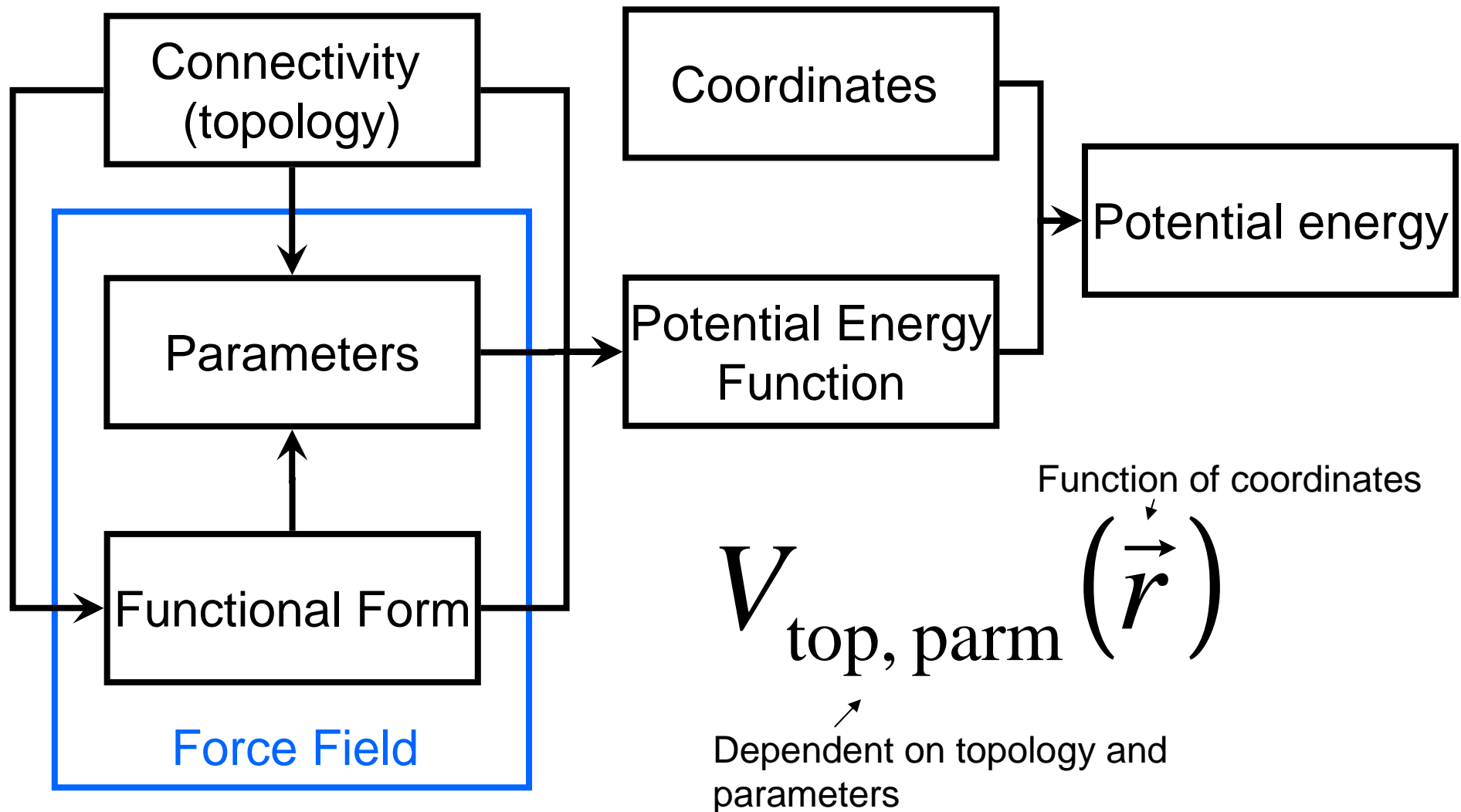*20 Penn Street, Baltimore, Maryland 21201*

Abstract: Empirical force field-based studies of biological macromolecules are becoming a common tool for investigating their structure–activity relationships at an atomic level of detail. Such studies facilitate interpretation of experimental data and allow for information not readily accessible to experimental methods to be obtained. A large part of the success of empirical force field-based methods is the quality of the force fields combined with the algorithmic advances that allow for more accurate reproduction of experimental observables. Presented is an overview of the issues associated with the development and application of empirical force fields to biomolecular systems. This is followed by a summary of the force fields commonly applied to the different classes of biomolecules; proteins, nucleic acids, lipids, and carbohydrates. In addition, issues associated with computational studies on "heterogeneous" biomolecular systems and the transferability of force fields to a wide range of organic molecules of pharmacological interest are discussed.

© 2004 Wiley Periodicals, Inc.   J Comput Chem 25: 1584–1604, 2004

Molecular Mechanics Force Field

Connectivity (topology)

Coordinates

Potential energy

Parameters

Potential Energy Function

Functional Form

Force Field

Function of coordinates

$$V_{\text{top, parm}}(\vec{r})$$

Dependent on topology and parameters

Molecular Mechanics

- Every atom is a sphere with a radius (Lennard Jones)

- Point charge is located at each atomic center (Coulomb's law)

- Bonds and angles are held by springs to ideal lengths
  - eg. $V_{bond} = k_b \left(r - r_0 \right)^2$
  - Hooke's Law, K_b: spring constant, r_0: ideal length

- Dihedrals are represented by sigmoidal function which has energy wells at favorable angles.

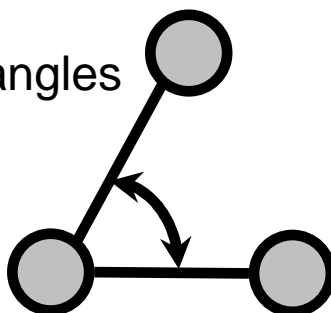- Improper torsions force atoms to be a defined angle to plane.

Class I Potential Energy function

The "Tinker-toy Model"

Bonded terms

bonds

dihedrals
(torsions)

angles

$$V_{bond} = k_b (r - r_0)^2$$

$$V_{angle} = k_\theta (\theta - \theta_0)^2$$

$$V_{dihidral} = K_\chi (1 + \cos(n\chi - \delta))$$
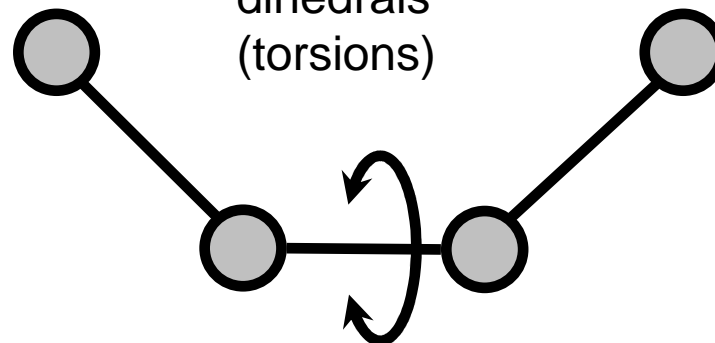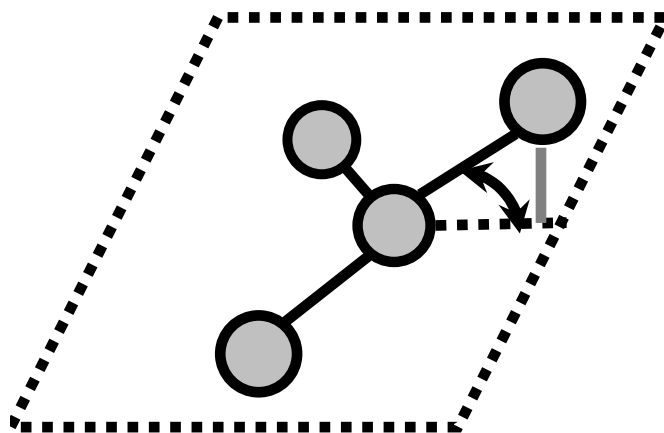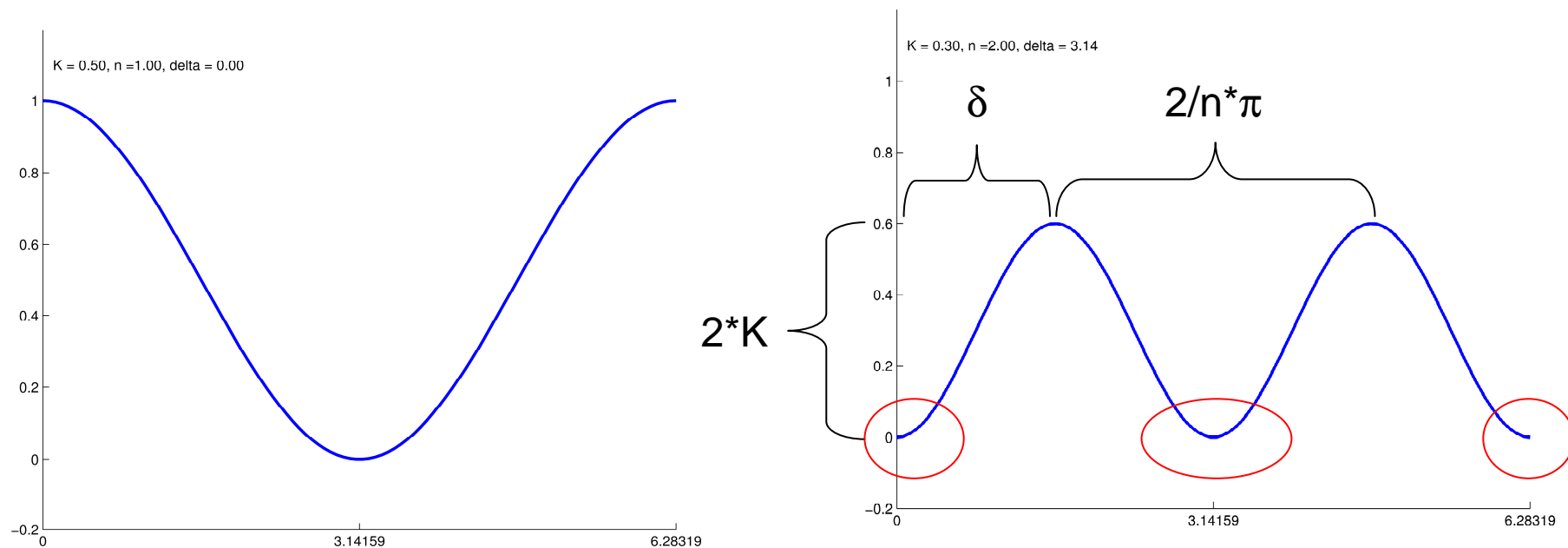
improper
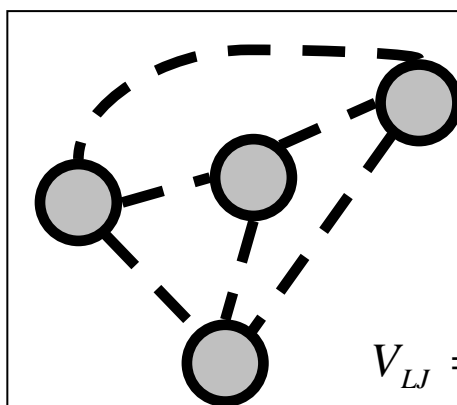
Through space
interactions

$$V_{coul} = \frac{q_i \, q_j}{\varepsilon \, r_{ij}}$$

$$V_{LJ} = \varepsilon_{i,j} \left[\left(\frac{R\min_{i,j}}{r_{i,j}} \right)^{12} - 2 \left(\frac{R\min_{i,j}}{r_{i,j}} \right)^6 \right]$$

$$V_{improper} = k_\varphi (\varphi - \varphi_0)^2$$

$$R\min_{i,j} = \frac{R\min_i + R\min_j}{2} \qquad \varepsilon_{i,j} = \sqrt{\varepsilon_i * \varepsilon_j}$$

Dihedral Term

$$V_{dihidral} = K_\chi \left(1 + \cos(n\chi - \delta)\right)$$

K = 0.50, n =1.00, delta = 0.00
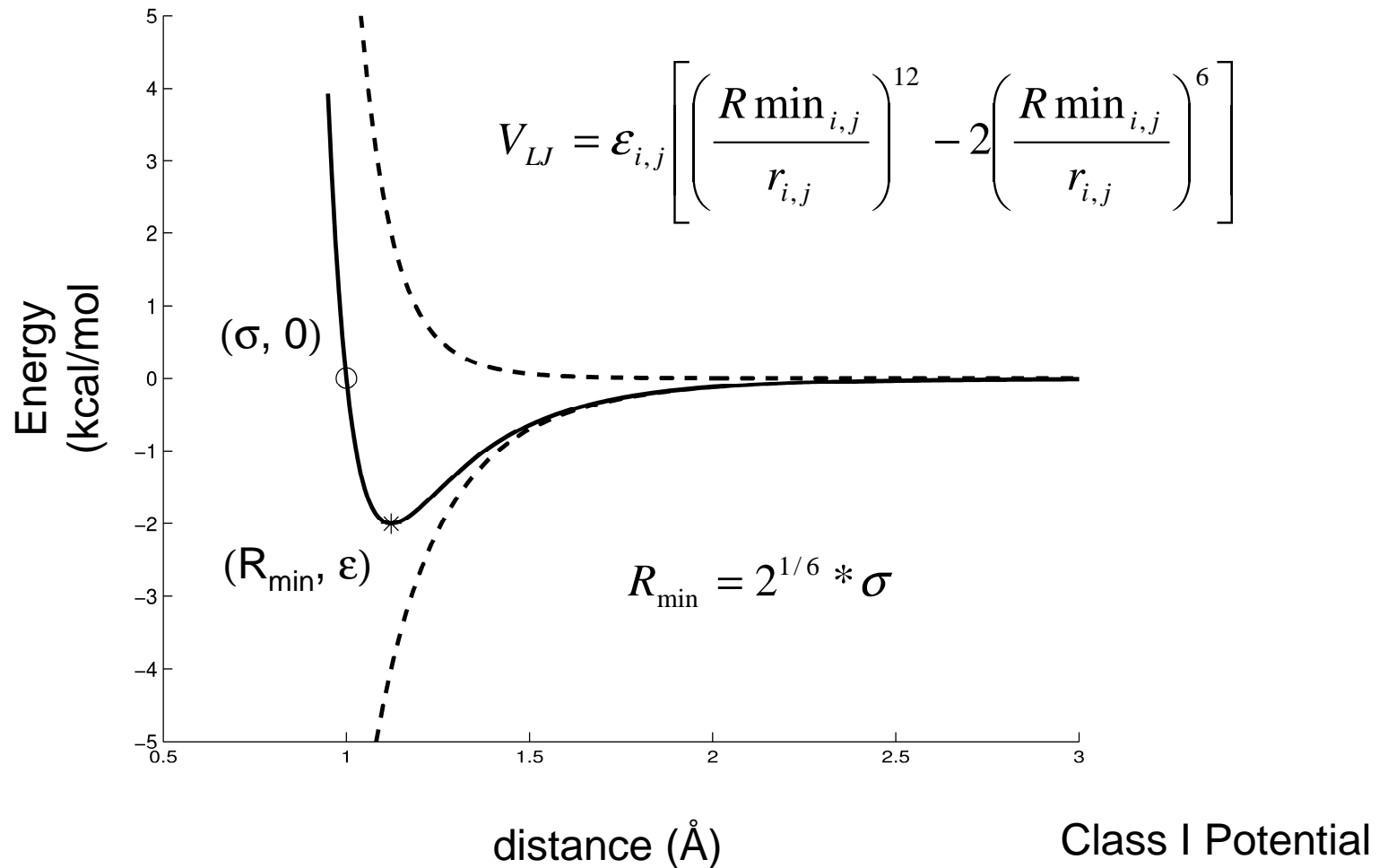
K = 0.30, n =2.00, delta = 3.14

δ

2/n*π

2*K

π (3.14) radians = 180 degrees

Molecular Modelling Principles and applications,
Leach Pearson Prentice hall second edition (chapter 4)

Class I Potential
Energy function

Lennard-Jones Equation

$$V_{LJ} = \varepsilon_{i,j} \left[\left(\frac{R\min_{i,j}}{r_{i,j}} \right)^{12} - 2 \left(\frac{R\min_{i,j}}{r_{i,j}} \right)^{6} \right]$$

$(\sigma, 0)$

(R_{min}, ε)

$$R_{min} = 2^{1/6} * \sigma$$

distance (Å)

Class I Potential
Energy function

Molecular Modelling Principles and applications,
Leach Pearson Prentice hall second edition (chapter 4)

Potential Energy function

$$V = \sum_{bonds} k_b (b - b_0)^2 + \sum_{angles} k_\theta (\theta - \theta_0)^2 + \sum_{impropers} k_\varphi (\varphi - \varphi_0)^2$$

$$+ \sum_{dihedrals} K_\chi (1 + \cos(n\chi - \delta)) + \sum_{i=1}^{N} \sum_{j=i+1}^{N} \varepsilon_{i,j} \left[\left(\frac{R\min_{i,j}}{r_{i,j}} \right)^{12} - 2 \left(\frac{R\min_{i,j}}{r_{i,j}} \right)^{6} \right] + \frac{q_i q_j}{\varepsilon r_{ij}}$$

Different Force-Field

- CHARMM
- AMBER
- GROMOS
- OPLS

Parameterization

- Experimental observables
- Quantum Mechanical calculations

Interdependences among parameters

Molecular Modelling Principles and applications,
Leach Pearson Prentice hall second edition (chapter 4)

Mackerell, Vol. 25, No. 13, Journal of Computational Chemistry

Class I Potential
Energy function

Applications

- Dynamics of Molecules
  - how do proteins work
- Energy: affinity, specificity
  - how do proteins interact with one another and other molecules (drugs, substrates, DNA or RNA)
- Protein design
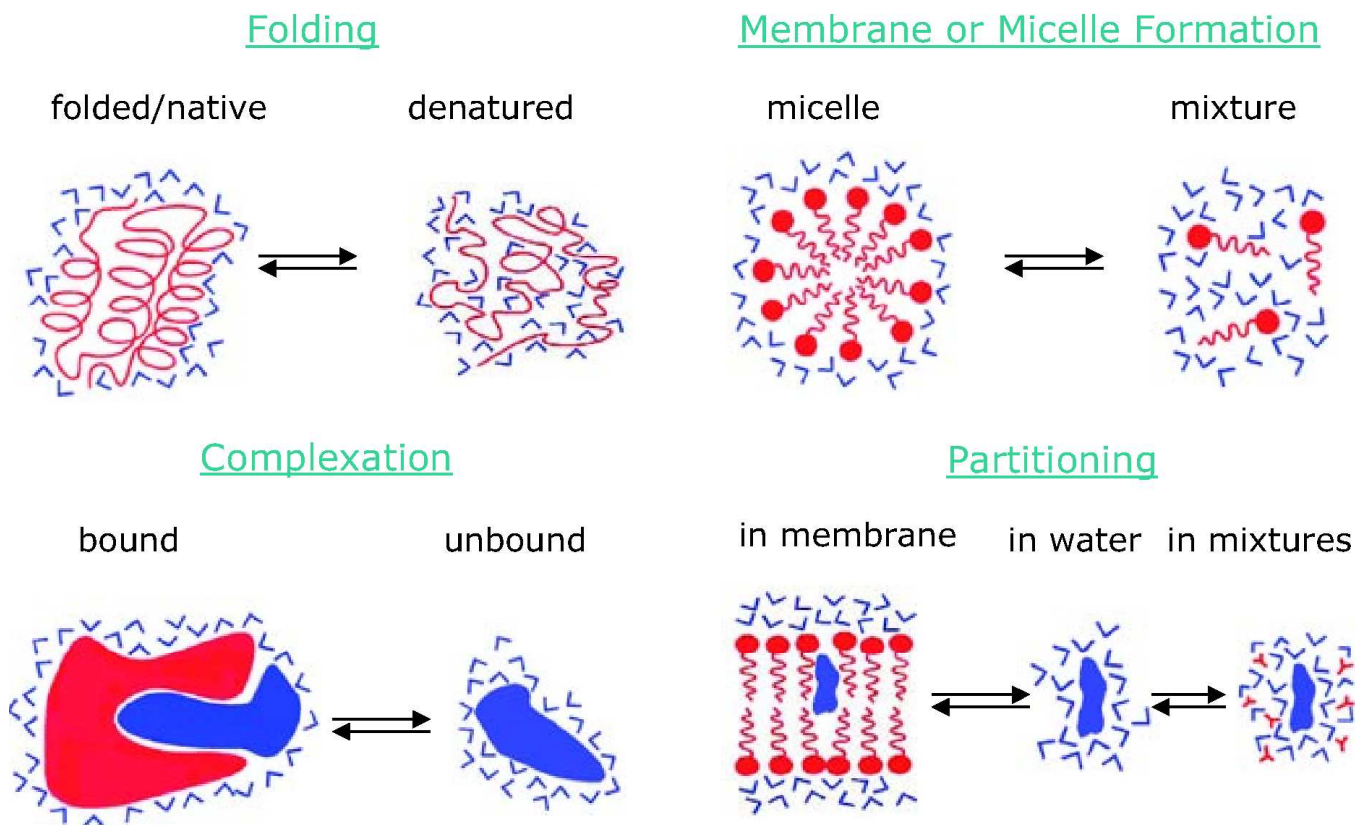 - can proteins be designed to have a specific function

Applications

Folding

folded/native   denatured

Membrane or Micelle Formation

micelle   mixture

Complexation

bound   unbound

Partitioning

in membrane   in water   in mixtures

Figure 1. Four biomolecular processes that are governed by thermodynamic equilibria.

Things to consider

Degrees of freedom: atoms are the elementary particles

size of the system

Boundary conditions

Forces or interactions between atoms

Energy Function

MOLECULAR MODEL

system temperature pressure walls external forces

Force field = physicochemical knowledge

sampling method

Methods to generate configurations of atoms: Newton
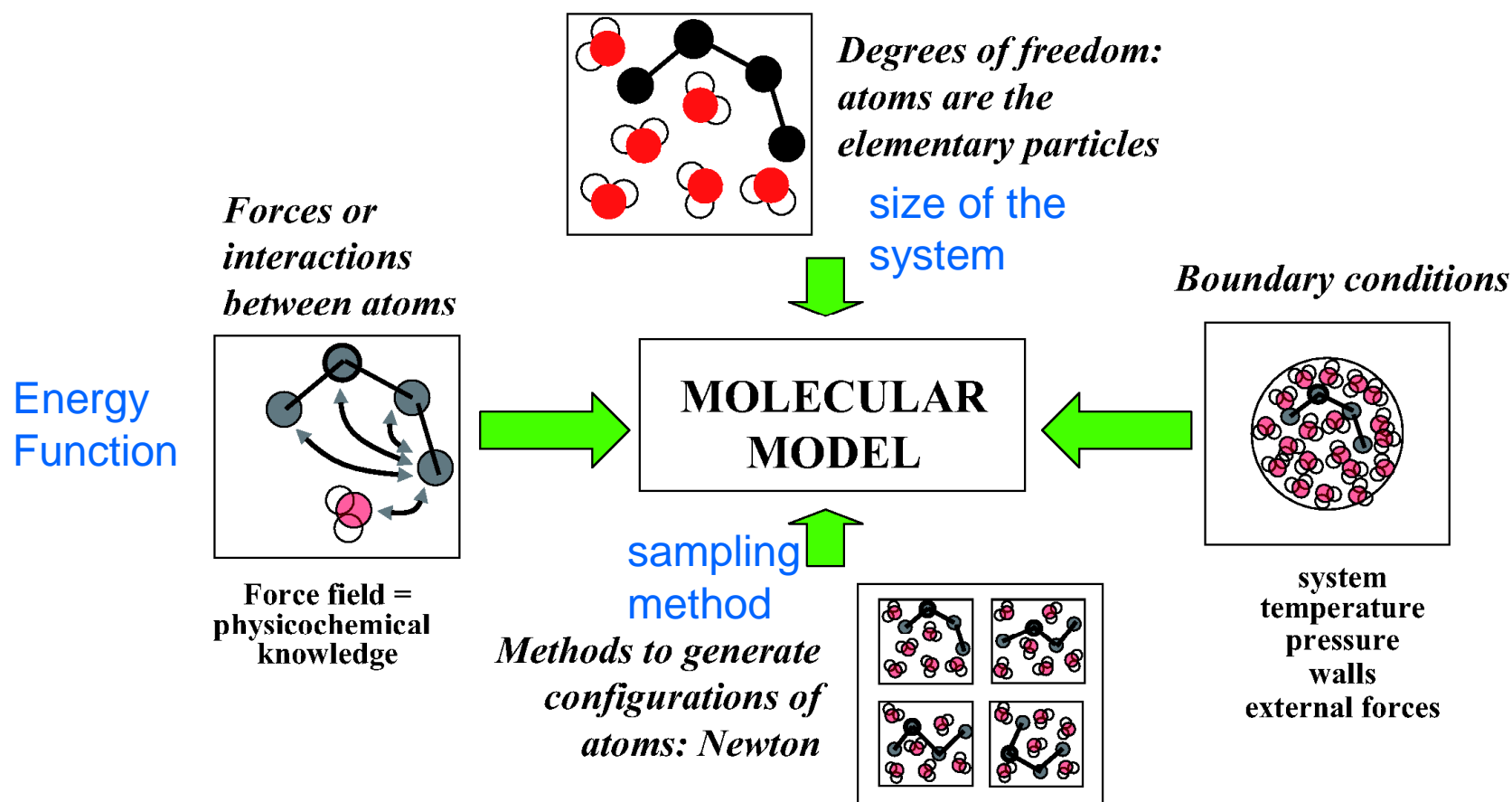
Figure 2. Four basic choices in the definition of a model for molecular simulation.

Angew. Chem. Int. Ed. 2006, 45, 4064 – 4092

Limitations to
Molecular Mechanics ?

Limitations to MM

MM cannot model the following easily:

- Chemical reactions
  - Formation and breaking of bonds

- Polarizability

- Protonation states

Potential Energy Function

- Electronic Polarizability
- Combining Rules
- 1,4 Interactions
- Lone Pairs
- All-Atom vs. United Atom Force Fields
- Treatment of Solvation
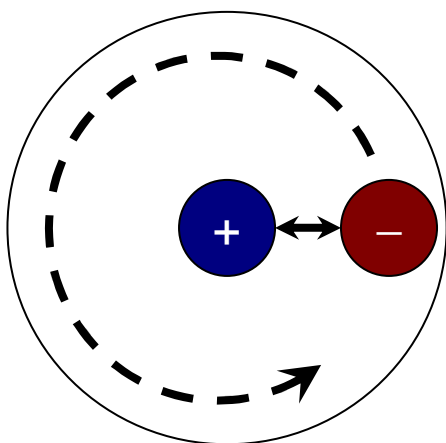- Treatment of Long-Range Interactions

Potential Energy Function

- **Electronic Polarizability**
- Combining Rules
- 1,4 Interactions
- Lone Pairs
- All-Atom vs. United Atom Force Fields
- Treatment of Solvation
- Treatment of Long-Range Interactions

Electronic Polarizability
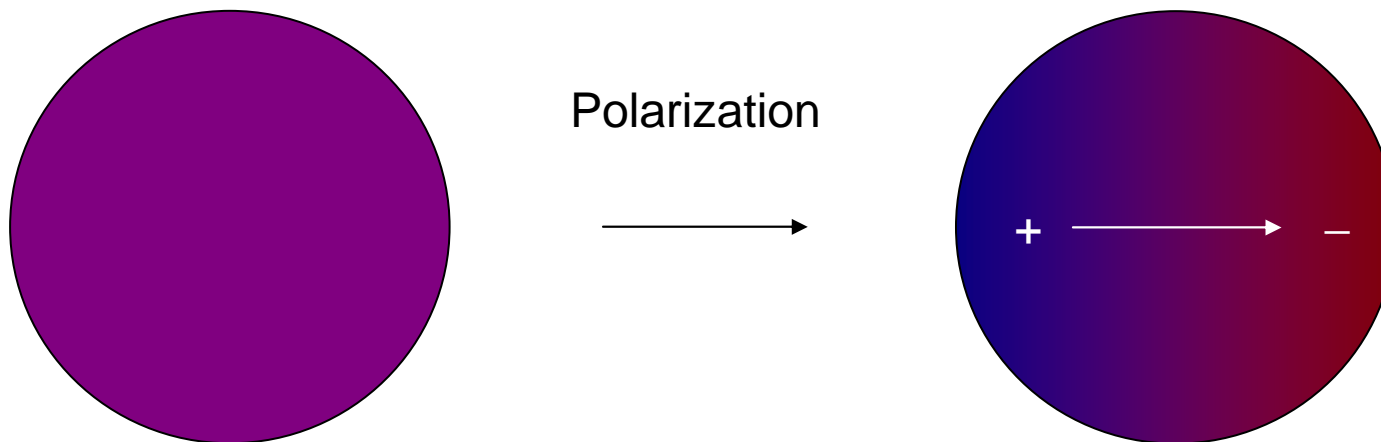
Polarization

Drude particle

$$U_{pol} = \frac{1}{2}\sum_i \mu_i E_i$$

induced dipole models

μ_i is the dipole moment of atom i
E_i is the electrostatic field at atom i

Class II Potential Energy function

Electronic Polarizability

- Common force fields use a fixed charge model for every atom

- Polarizable or non-additive force fields allow for atoms to vary their charges in the context of the electric field

- Drude particle (Drude oscillator method)
  - every atom has two point charges (atom center and Drude particle) connected by a spring (hooke's law)

Potential Energy Function

- Electronic Polarizability
- **Combining Rules**
- 1,4 Interactions
- Lone Pairs
- All-Atom vs. United Atom Force Fields
- Treatment of Solvation
- Treatment of Long-Range Interactions

Combining Rules

- force fields combine parameters differently
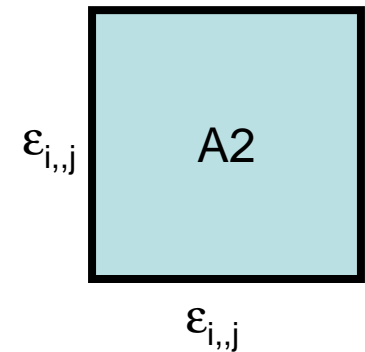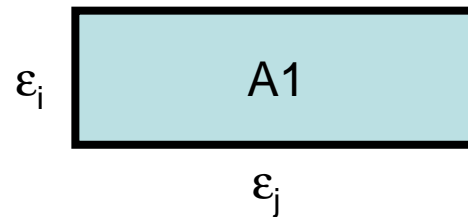 - arithmetic mean

$$R\min_{i,j} = \frac{R\min_i + R\min_j}{2}$$

  - geometric mean

$$\varepsilon_{i,j} = \sqrt{\varepsilon_i * \varepsilon_j}$$

$$V_{LJ} = \varepsilon_{i,j}\left[\left(\frac{R\min_{i,j}}{r_{i,j}}\right)^{12} - \left(\frac{R\min_{i,j}}{r_{i,j}}\right)^{6}\right]$$
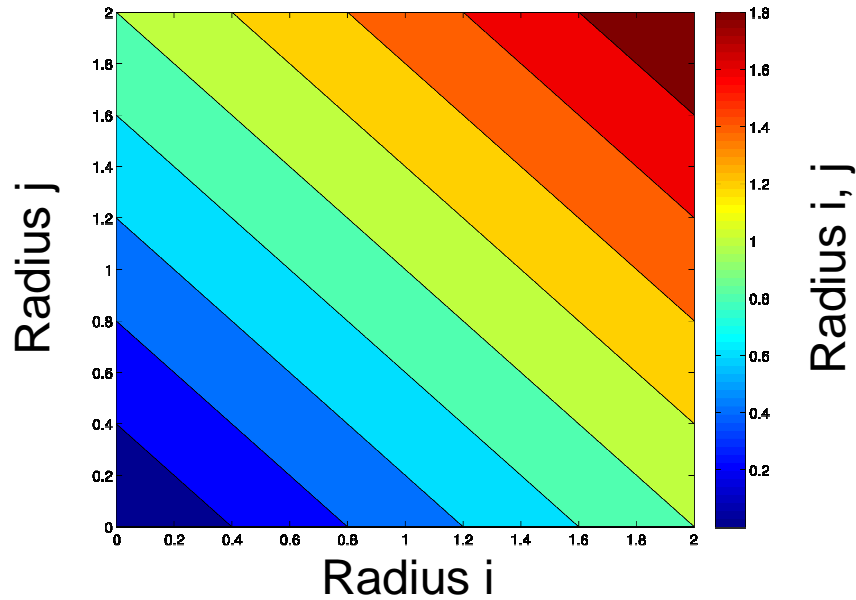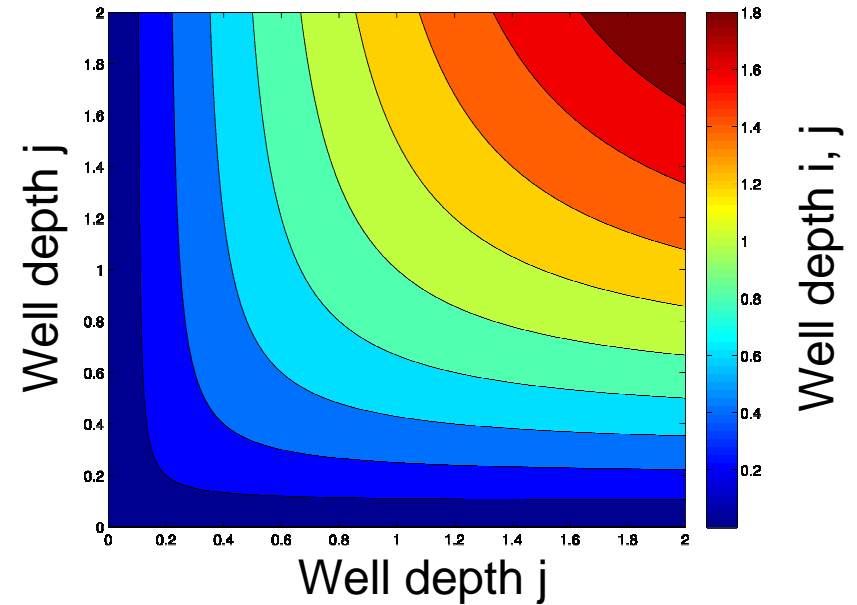
0.1, 0.9

0.5, 0.5

A1 = A2

ε_i | A1 | $\varepsilon_{i,,j}$ | A2

ε_j | $\varepsilon_{i,,j}$
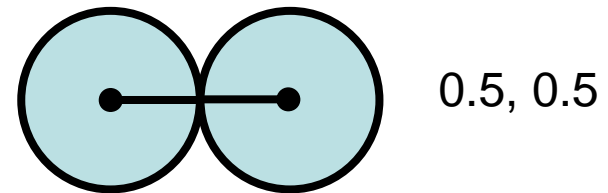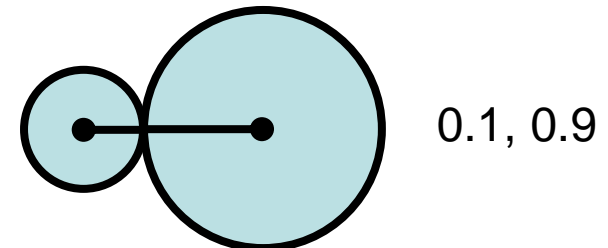
Combining Rules

Arithmetic mean

Geometric mean

(0.1 + 0.9) / 2 = 0.5
(0.5 + 0.5) / 2 = 0.5

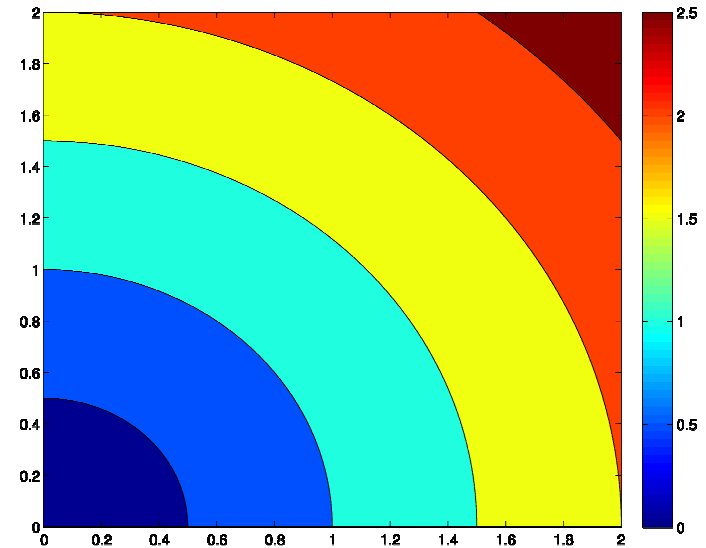(0.1 * 0.9)^(1 / 2) = 0.3
(0.5 * 0.5)^(1 / 2) = 0.5

Combining Rules

Harmonic mean
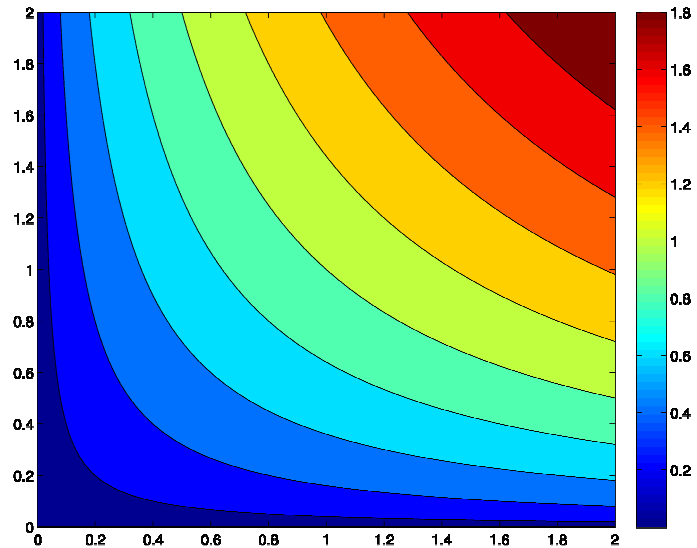
Euclidean relationship

2 / (1 / 0.1 + 1 / 0.9) = 0.18
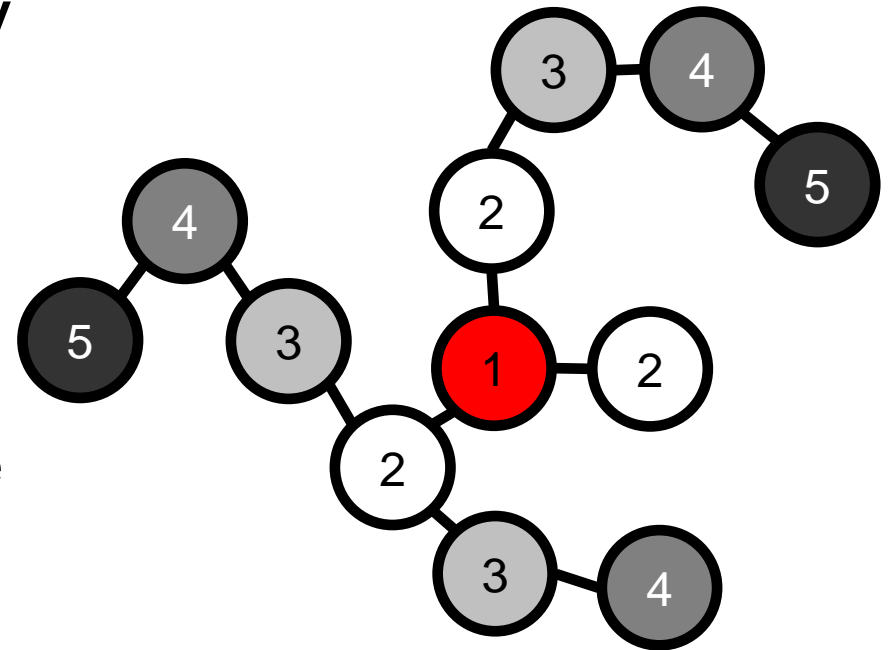2 / (1 / 0.5 + 1 / 0.5) = 0.5

(0.1^2 + 0.9^2)^(1 / 2) = 0.9
(0.5^2 + 0.5^2)^(1 / 2) = 0.7
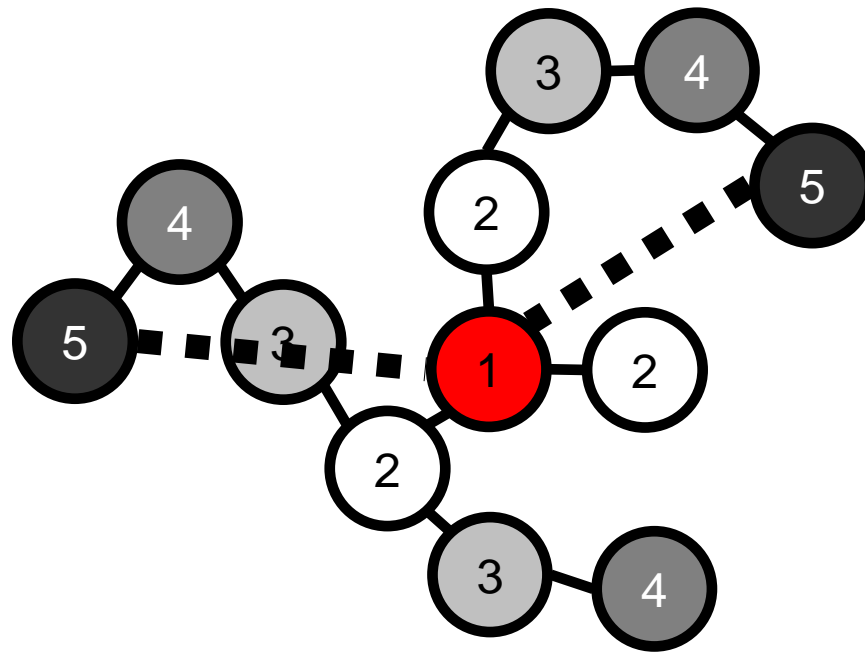
Potential Energy Function

- Electronic Polarizability
- Combining Rules
- **1,4 Interactions**
- Lone Pairs
- All-Atom vs. United Atom Force Fields
- Treatment of Solvation
- Treatment of Long-Range Interactions
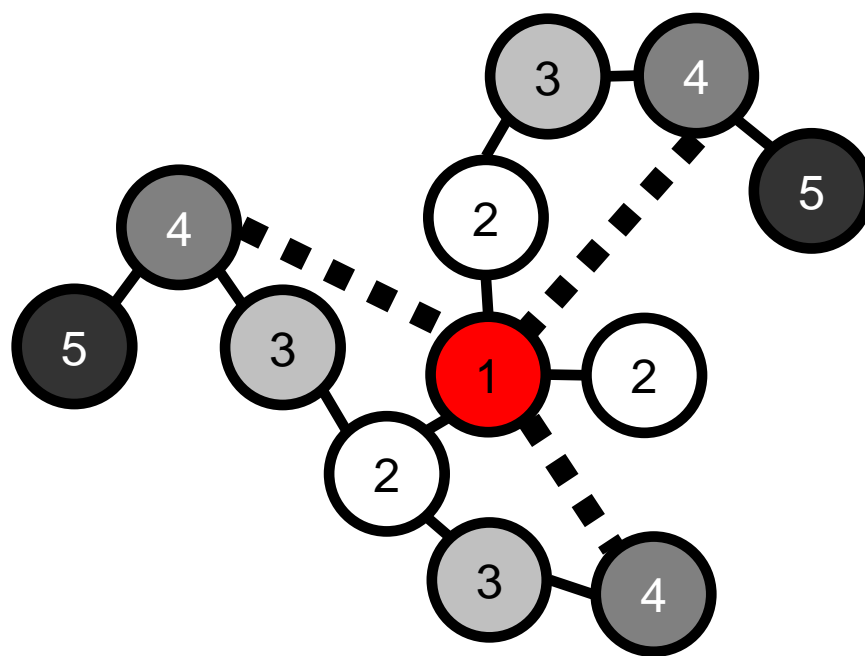
1,4 (non-bonded) Interactions

- 1,2 and 1,3 are normally ignored by force-fields
- 1,4 are scaled
  – CHARMM no scaling
  – AMBER and OPLS scale 0.83
- Bonded terms dominate

1,5 interactions
Not scaled (s = 1)

1,4 interactions
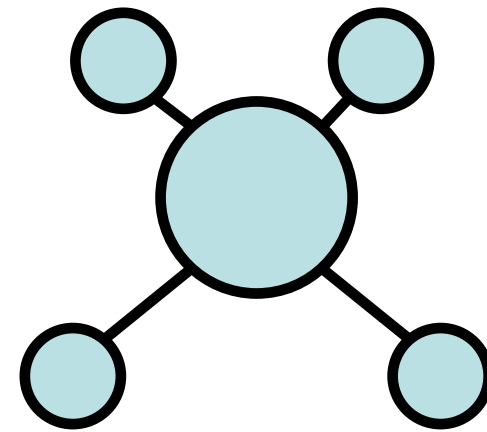# Maybe scaled (amber: s = 0.83)

Note that 1 / 1.2 = 0.83

Potential Energy Function

- Electronic Polarizability
- Combining Rules
- 1,4 Interactions
- **Lone Pairs**
- All-Atom vs. United Atom Force Fields
- Treatment of Solvation
- Treatment of Long-Range Interactions

Lone Pairs

- Force field without lone pairs do well
- Adding more parameters
- Fitting charges from quantum mechanical electrostatic potentials
- TIP5P and ST2 Water models
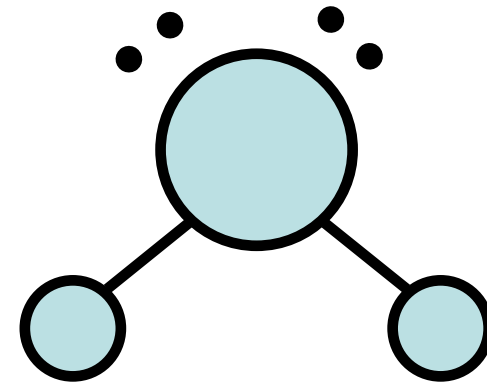
Potential Energy Function

- Electronic Polarizability
- Combining Rules
- 1,4 Interactions
- Lone Pairs
- **All-Atom vs. United Atom Force Fields**
- Treatment of Solvation
- Treatment of Long-Range Interactions

All-Atom vs. United Atom Force Fields

- Hydrogen atoms are not represented explicitly. incorporated in to the bonded heavy atom

- Methyl group will be represented by only one sphere

- Corse grain models

Potential Energy function

- Electronic Polarizability
- Combining Rules
- 1,4 Interactions
- Lone Pairs
- All-Atom vs. United Atom Force Fields
- **Treatment of Solvation**
- Treatment of Long-Range Interactions

Treatment of Solvation

- Explicit water models
  - TIP3P, TIP4P, TIP5P, SPC,ST2
- Implicit water models (Continuum with fixed dielectric)
  - Poisson-Boltzmann (PB) model
  - Generalized Born (GB) model
- This will be discussed in detail in a following class
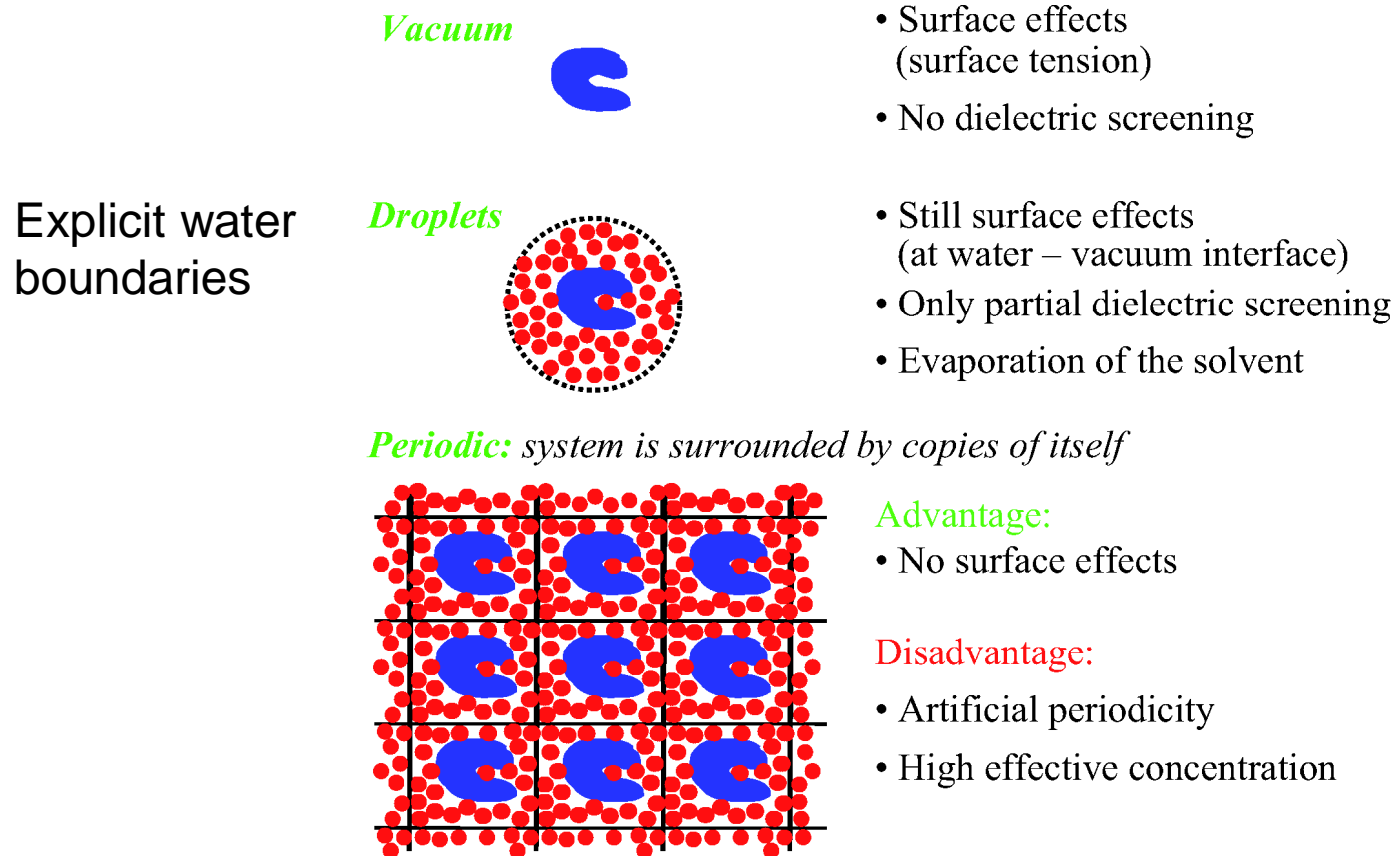
Treatment of Solvation

Vacuum

- Surface effects
  (surface tension)

- No dielectric screening

Explicit water boundaries

Droplets

- Still surface effects
  (at water – vacuum interface)
- Only partial dielectric screening
- Evaporation of the solvent

Periodic: *system is surrounded by copies of itself*

Advantage:
- No surface effects

Disadvantage:
- Artificial periodicity
- High effective concentration

Figure 3. Three types of spatial boundary conditions used in molecular simulation.

Potential Energy Function

- Electronic Polarizability
- Combining Rules
- 1,4 Interactions
- Lone Pairs
- All-Atom vs. United Atom Force Fields
- Treatment of Solvation
- **Treatment of Long-Range Interactions**

Treatment of Long-Range Interactions

- Cut-offs

- Partial Mesh Ewald (PME)

- Smoothing functions

- Re-parameterization with different treatments of long-range interactions

Mackerell, Vol. 25, No. 13, Journal of Computational Chemistry
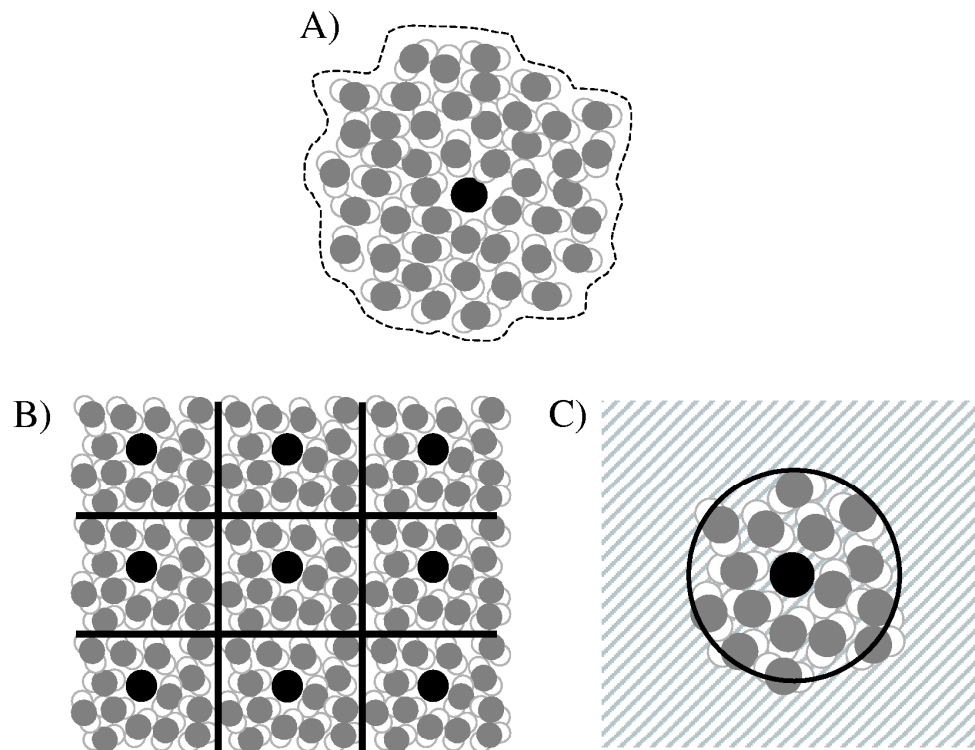
Solvation and long range electrostatics

Figure 6. Two methods for calculating long-range electrostatic energies and forces in a molecular system: A) real system with explicit solvent; B) periodicity used in the Ewald, P^3M, and PME methods, and C) continuum approximations beyond a given cut-off distance.

Biomolecular Force Fields

- Protein
- Nucleic Acid
- Lipid
- Carbohydrates
- Drug-Like Molecules
- Heterogeneous Biomolecular Systems (combining different types of biomolecules)

Conclusions

- Introduced Molecular Mechanics

- Described functional form

- Discussed parameterization

- Considered dependences and other issues

- Discussed applications

Thank you for your attention.

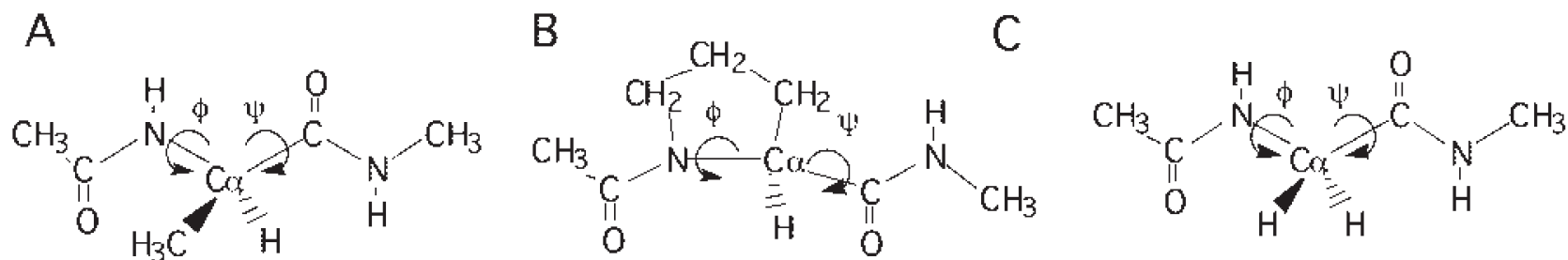## Questions?

Protein Force Field

Figure 1. (A) Alanine, (B) proline, and (C) glycine dipeptides used for parameter development of the protein backbone. Shown in the figures are the ϕ, ψ dihedral angles that define the Ramachandran map.[169]

Force Field Optimization

Table 1. Comparison of Peptide Bond Geometries from QM and Experimental Methods.

	Experimental			MP2/6-31 G(d)[b]		
	Gas[c]	Crystal[d]	Survey[e]	Gas	3H$_2$O	H$_2$O,2FM
Bonds						
C$_m$—C	1.520 (5)	1.515 (3)	1.52 (1)	1.514	1.510	1.512
C—N	1.386 (4)	1.325 (3)	1.33 (1)	1.365	1.339	1.337
N—C$_m$	1.469 (6)	1.454 (3)	1.45 (2)	1.448	1.454	1.454
C=O	1.225 (3)	1.246 (2)	1.23 (1)	1.232	1.255	1.254
Angles						
C$_m$—C—N	114.1 (15)	116.3 (6)	116 (2)	115.3	117.1	116.6
O=C—N	121.8 (4)	121.7 (6)	123 (1)	123.1	122.1	122.6
C$_m$—C=O	124.1	121.9 (6)	121 (4)	121.6	120.9	120.9
C—N—C$_m$	119.7 (8)	121.3 (6)	122 (1)	122.1	121.1	121.3

Bonds and angles in Å and degrees, respectively. C$_m$ indicates a terminal methyl carbon. Values in parenthesis represent the standard deviation in the final digit(s).

[b]From ref. 324, 3H$_2$O indicates two water molecules hydrogen bonding to the carbonyl oxygen and one water molecule hydrogen bonding to the amide proton; H$_2$O,2FM indicates one water molecule and one formamide hydrogen bonding to the carbonyl oxygen and one formamide hydrogen bonding to the amide proton; see original reference for the exact geometries.

[c]Gas phase electron diffraction data from ref. 325.

[d]Crystal values are from ref. 326 for the 0.9 occupancy structure.

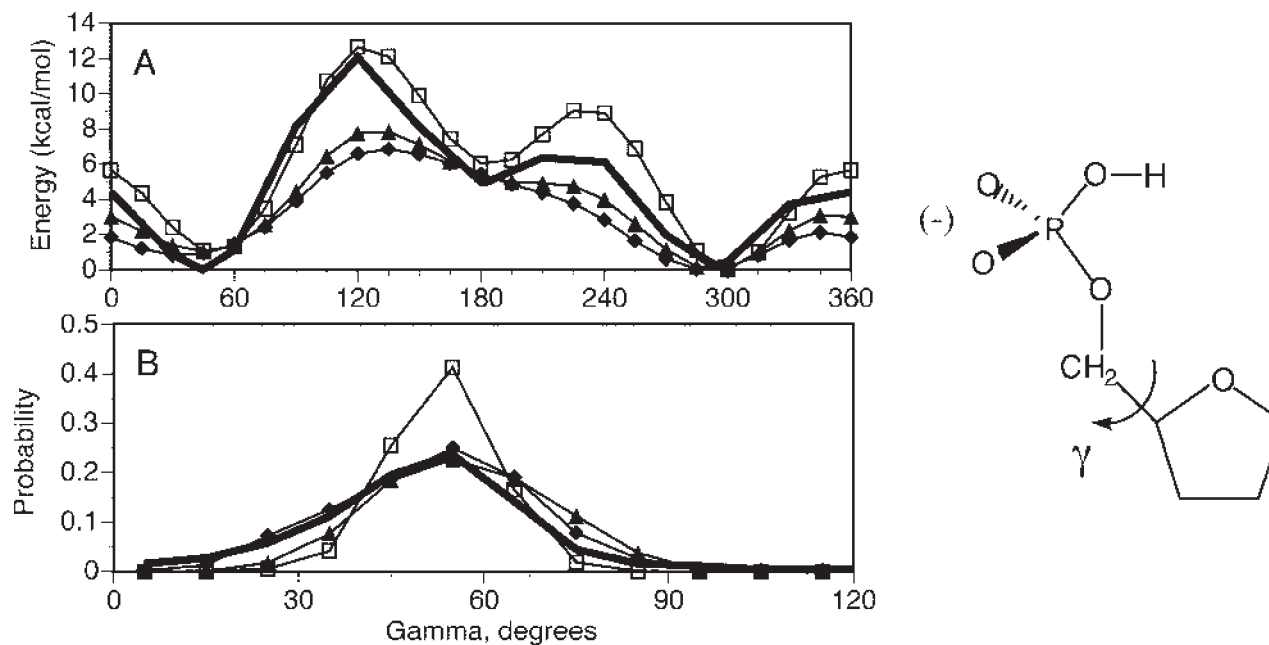[e]Survey of the Cambridge Crystal Data Bank from ref. 161.

Figure 2. Potential energies (A) and probability distributions (B) as a function of the γ dihedral. The potential energy surfaces (A) were obtained using the presented compound at the QM HF/6-31+G* (bold line) level of theory and for three empirical parameter sets designated 1 (open squares), 2 (triangles), and 3 (diamonds). Probability distributions are from the NDB survey (bold line) from crystal simulations of the CGATCGATCG B form decamer using the same three empirical parameter sets 1 (open squares), 2 (triangles), and 3 (diamonds). Note the change in the X-axis upon going from A (0 to 360°) to B (0 to 120°). See ref. 48 for methodological details.
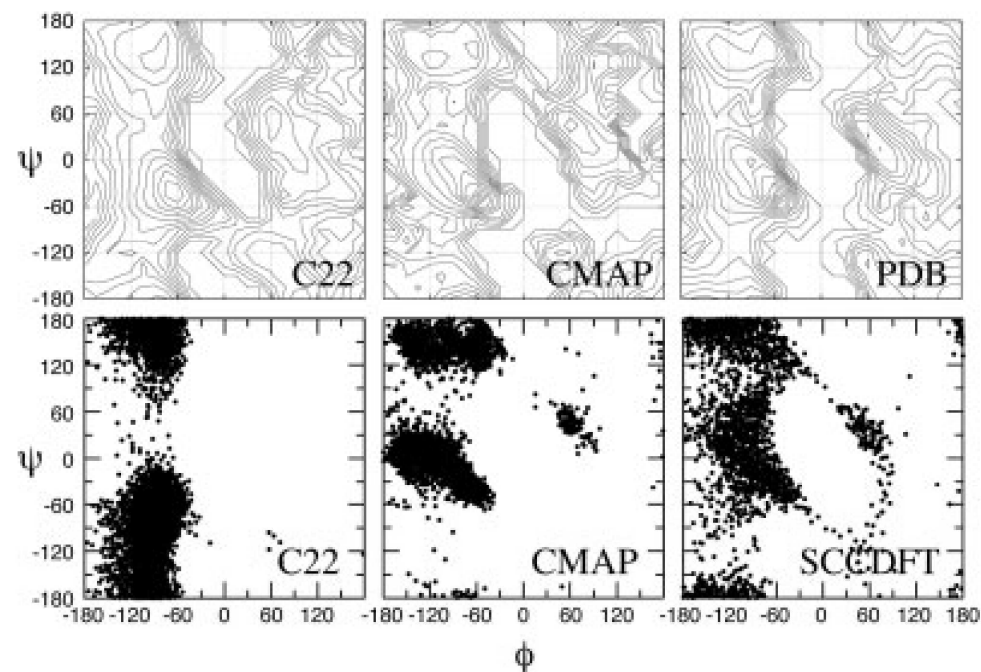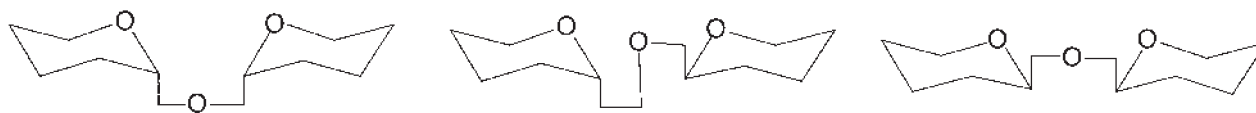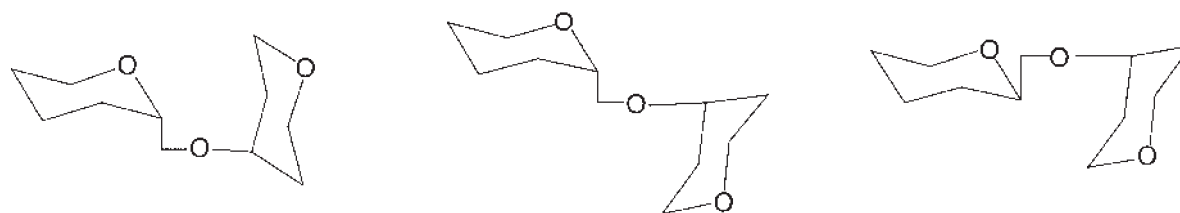
Mackerell, Vol. 25, No. 13, Journal of Computational Chemistry

Figure 3. ϕ,ψ PMFs based on MD simulations using the CHARMM22 and CHARMM22 grid-corrected (CMAP) empirical force fields and from a survey of the PDB (upper frames) and ϕ, ψ distributions from MD simulations of the alanine dipeptide (Ace-Ala-Nme, lower frames) in solution using the CHARMM22[161] and CHARMM22 grid-corrected empirical force fields and previously published data from a QM/MM model (SCCDFTB). PMF contours are in 0.5 kcal/mol increments up to 6 kcal/mol above the global minimum. PMFs were obtained from the respective probability distributions based on a Boltzmann distribution.[199] See ref. 29 for more details. Reproduced with permission from J Am Chem Soc 2004, 126, 698–699. Copyright 2004 Am Chem Soc.
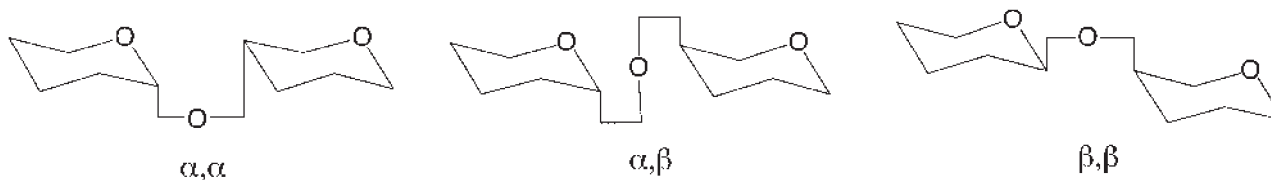
Carbohydrate Force Field

1,1-linkages (e.g. trehalose)

1,3-linkages (e.g. nigerose, laminarabiose)
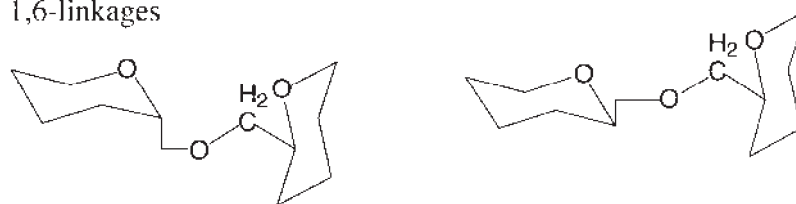
1,4-linkages (e.g. galabiose, maltose, cellobiose)

α,α   α,β   β,β

1,6-linkages

Figure 4. Example glycosyl linkages found in polysaccharides. Hydroxyl groups have been omitted for clarity.