EML

M
C
M

Workshop on Structure-based ligand design: *Lecture 3: Molecular Modeling for COMBINE*

Rebecca C. Wade

European Media Laboratory
Heidelberg

rebecca.wade@eml.villa-bosch.de
http://www.eml.org/english/Research/MCM

E
M
B
L

Workshop schedule

- **Lecture 1:** Introduction to Structure-based Drug Design, (GRID and 'flu)
- **Practical 1:** GRID
- **Lecture 2:** COMBINE Analysis overview
- **Lecture 3:** Molecular modeling for COMBINE
- **Practical 2:** COMBINE Analysis- molecular modeling
- **Lecture 4:** Chemometric analysis for COMBINE
- **Practical 3:** COMBINE Analysis- chemometrics
- **Lecture 5/Demo/Discussion/Practical 4**

Lecture 3: Overview

- **How can the energy components for COMBINE analysis be computed?**
  - **Modeling of complexes**
  - **Energy minimization**
  - **Calculation of energy terms**
  - **Computation of surface-area dependent terms**
  - **Computation of Poisson-Boltzmann terms**

Flowchart For Combine Analysis

Model unbound receptor r

Model unbound ligands l

Model $r{:}l$ complexes

↓

Energy minimize all models

↓

Compute
$\Delta U = E_{lr} - E_r - E_l$
for each l

↓

Partition ΔU into components

↓

Matrix pretreatment
(may include variable selection)

↓

Correlate components of ΔU with measured binding free energies (ΔG) (or activities) by PLS to derive model: $\Delta G \sim \Sigma w_i \Delta u_i^{sel} + C$

↓

Validate model:
-cross-validation
-scrambled ΔG values
-random ΔG values

↓

Make predictions for new l

Influenza neuraminidase inhibitors

- 43 complexes:
  - 29 inhibitors: sialic acid TS and benzoic acid derivatives
  - N9 + N2 subtypes + mutants
- 32 crystal structures
- 11 docked (comparative/AUTODOCK)
- Energy minimize: AMBER

Wang,T., Wade,R.C. *J. Med. Chem.* (2001) **44**, 961-971

Preparation of structures

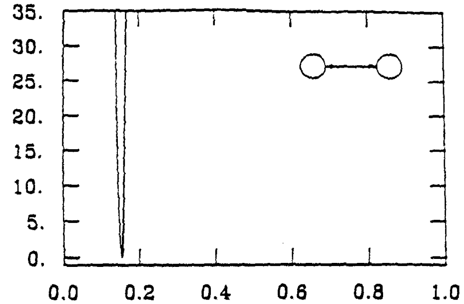
- **Should be done with care!**
- Protein structures may be from comparative models
- Ligands may be docked or modelled into the binding site by analogy
- Hydrogen atom coordinates (at least) will need to be modelled
  - Orient rotatable hydrogens carefully as poor hydrogen positions can adversely affect energies (watch out for His tautomers)
    - E.g. by WHATIF H-bond network optimization

Parameterization of structures

- **Should be done with care!**
- Non-”standard” molecules will need to have new parameters.
  - Bonded terms by analogy
  - Charges by a procedure consistent with protein force field
    - E.g. antechamber in AMBER7 assigns parameters including charges based on MOPAC calculation and adjusted for compatibility with AMBER protein charges.

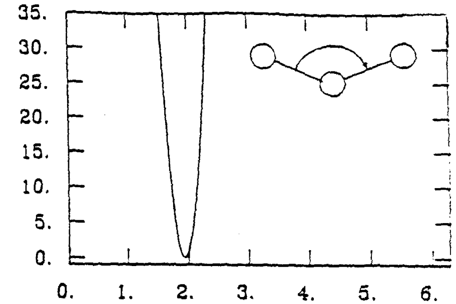
Molecular mechanics force field

Bond



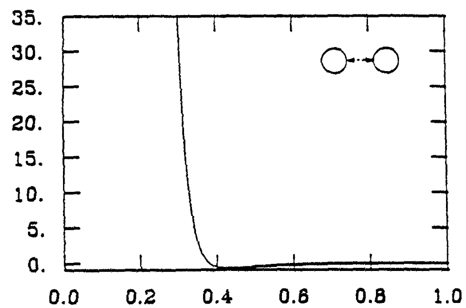
$$\sum_{n=1}^{N_b} \frac{1}{2} C_{b,n}(b_n - b_{0,n})^2$$

Bond angle



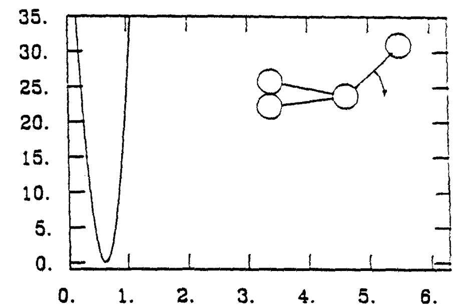
$$\sum_{n=1}^{N_\theta} \frac{1}{2} C_{\theta,n}(\theta_n - \theta_{0,n})^2$$

Lennard-Jones



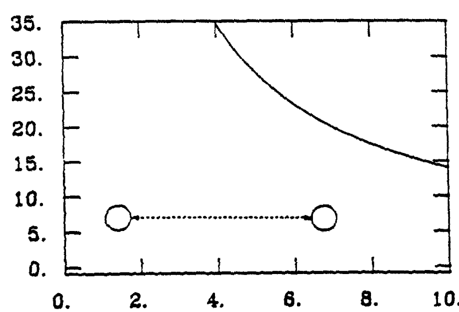
$$\sum_{i<j}^{N} \left[\frac{C_{12,ij}}{r_{ij}^{12}} - \frac{C_{6,ij}}{r_{ij}^{6}} \right]$$

Improper torsion



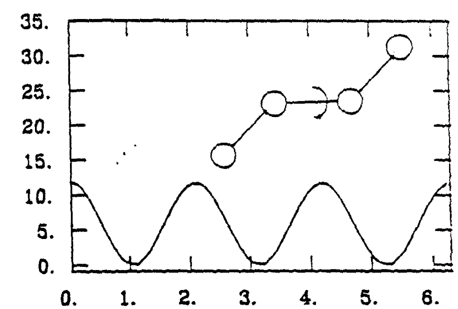
$$\sum_{n=1}^{N_\xi} \frac{1}{2} C_{\xi,n}(\xi_n - \xi_{0,n})^2$$

Electrostatic



$$\sum_{i<j}^{N} \frac{q_i q_j}{4\pi\epsilon_r\epsilon_0 r_{ij}}$$

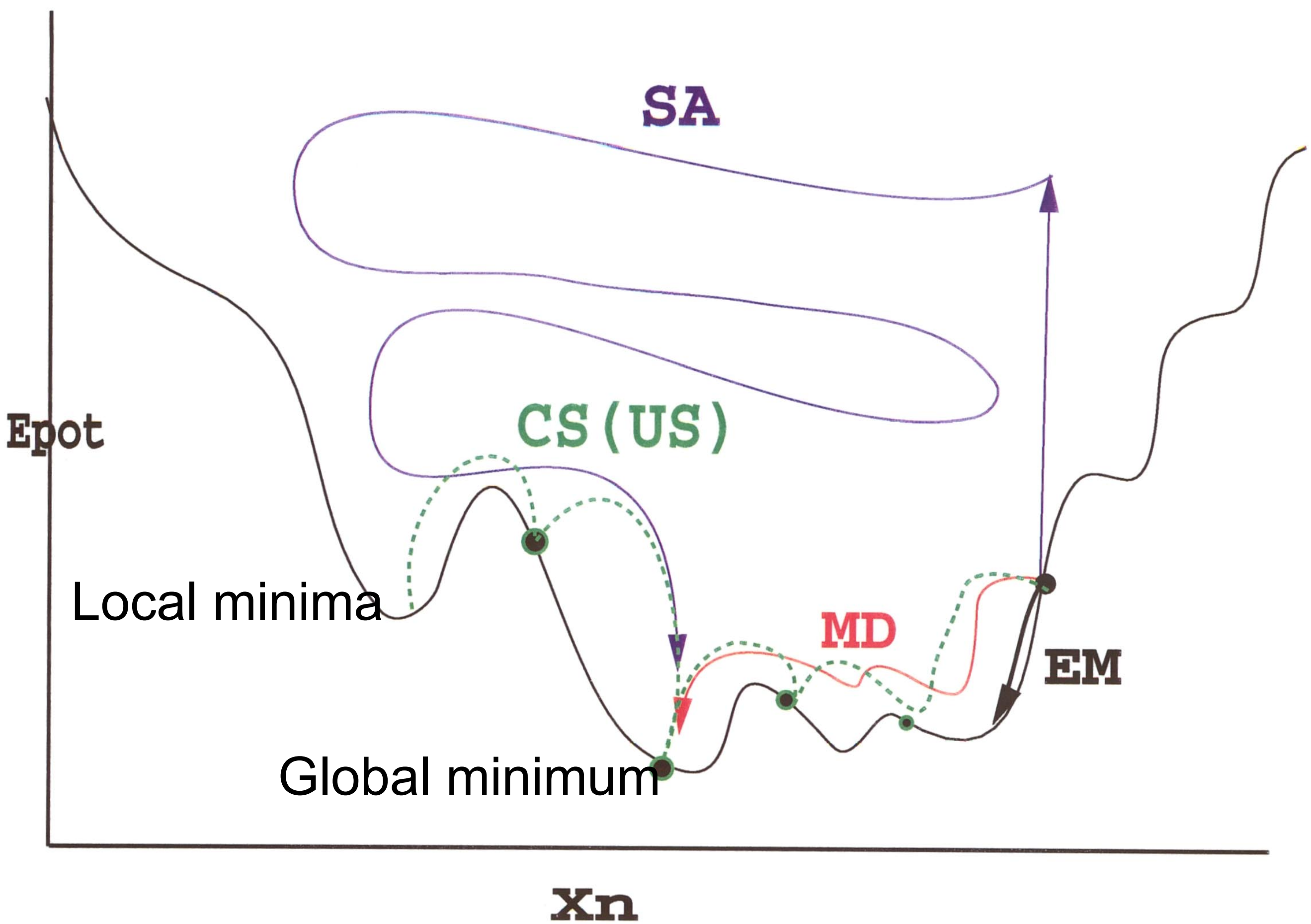
Dihedral angle



$$\sum_{n=1}^{N_\phi} C_{\phi,n}[1 + \cos(m_n\phi_n - \delta_n)]$$

Straatsma, T.

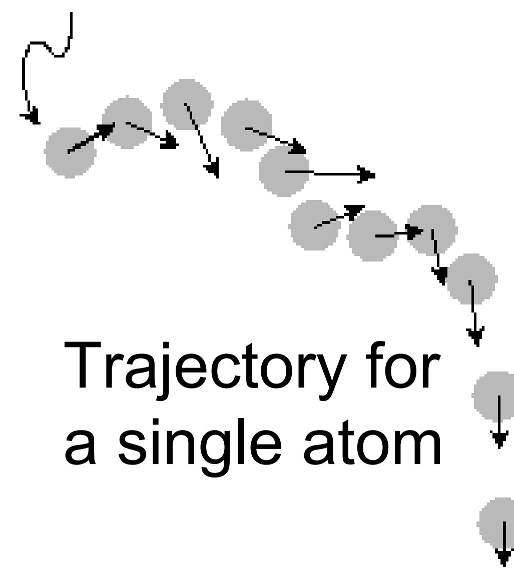
Exploring complex energy landscapes



- Conformational Search
- **Energy Minimization**
- Molecular Dynamics
- Simulated Annealing
- Monte Carlo

To find an Energy Minimum:

- **Minimum:** $\delta U(\mathbf{r})/\delta \mathbf{r}_i = 0; \delta^2 U(\mathbf{r})/\delta \mathbf{r}_i^2 > 0$
- **Taylor series expansion of energy** $U(\mathbf{r}_1, \mathbf{r}_2 \ldots.)$
  - $U(\mathbf{r}) = U(\mathbf{r}_0) + (\mathbf{r} - \mathbf{r}_0)U'(\mathbf{r}_0) + ½(\mathbf{r} - \mathbf{r}_0)^2 U''(\mathbf{r}_0) + \ldots..$
- **Zeroth order**
  - Grid/Systematic/Conformational search
  - Simplex
- **First order**
  - Steepest descent
  - Conjugate gradient
  - Powell
- **Second order**
  - (Adopted basis) Newton-Raphson

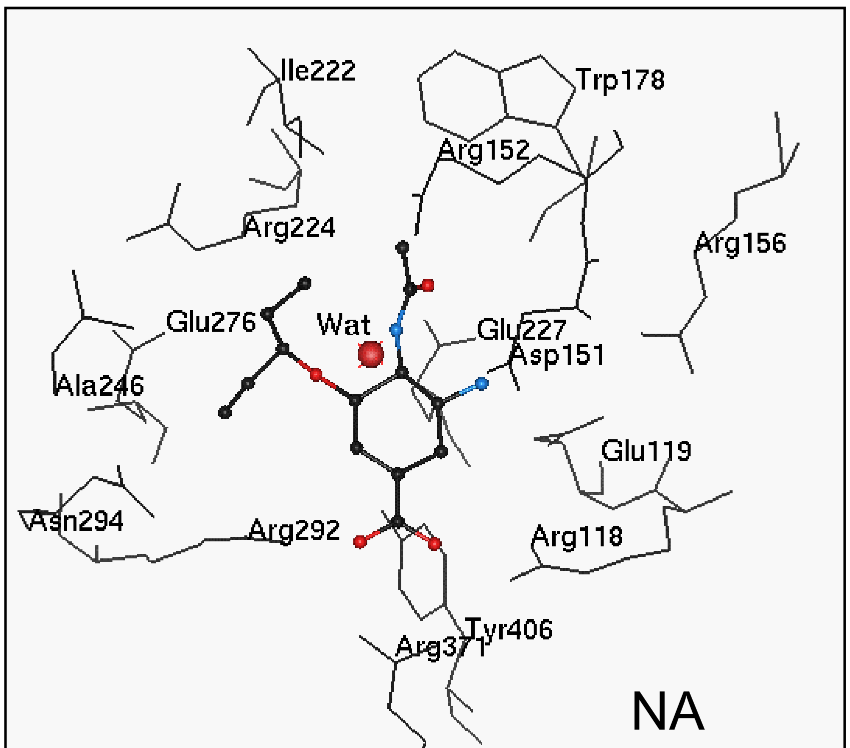


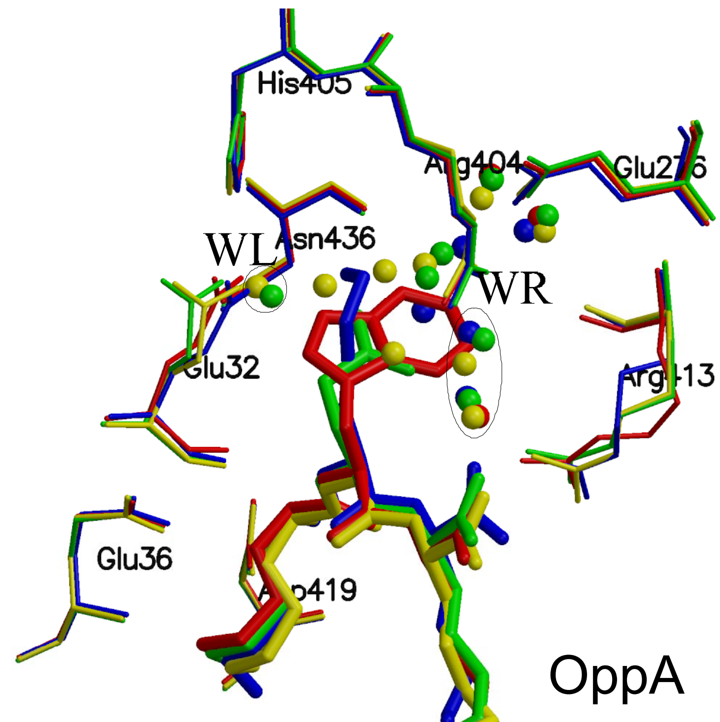
Energy minimization of structures

- Retain crystallographic water molecules (if present) for this (and prior hydrogen coordinate assignment)
- Distance-dependent dielectric constant
  - Better continuum solvent models on the way?
- Minimize modeled atoms before those with experimentally determined positions
- Restraints may be applied to part of system
- Do not minimize so much that the structures obtained reflect artifacts of the force-field
  - E.g. formation of non-natural surface salt-links

Computation of energy components - water

- After minimization, remove water molecules except those that may be important for the COMBINE model. These should be retained and treated, singly or as a group, as “residues”.





Computation of energy components

- Partition system into fragments ("residues")
  - Amino acid residues
  - Nucleic acid bases, phosphate and sugar moieties
  - Ligands may be treated as one "residue" or divided into fragments "residues" according to substituent or atom spatial location
- Compute intermolecular energy terms between residues.
    - E.g. using the Analysis module of AMBER7
- Intramolecular residue-residue interaction terms may be computed similarly

Comparative Binding Energy (COMBINE) Analysis

- COMBINE: data + techniques
  - 3D macromolecular structure + experimental binding data
  - Empirical molecular mechanics energies + chemometric PLS

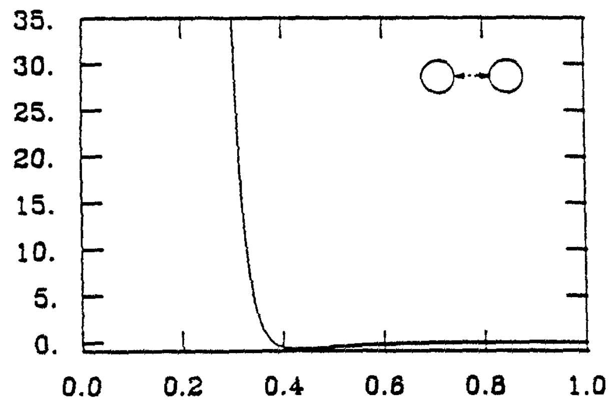
$$\Delta G = \sum_i w_i \, \Delta u_i + C$$

$$\Delta G = \sum_i w_i^{vdw} u_i^{vdw} + \sum_i w_i^{ele} u_i^{ele} + C$$

Ortiz,A.R., Pisabarro,M.T., Gago,F. Wade,R.C. *J. Med. Chem.* (1995) 38, 2681
Wade,R.C., Ortiz,A.R., Gago,F. *Persp. Drug. Disc. & Des.* (1998) 9, 19

Molecular mechanics force field

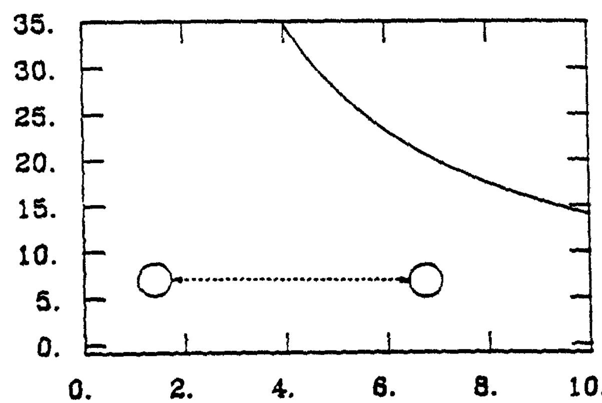
Lennard-Jones



$$\sum_{i<j}^{N}\left[\frac{C_{12,ij}}{r_{ij}^{12}}-\frac{C_{6,ij}}{r_{ij}^{6}}\right]$$

Short-range

Electrostatic



$$\sum_{i<j}^{N}\frac{q_i q_j}{4\pi\epsilon_r\epsilon_0 r_{ij}}$$

Long-range

Further energy terms

- Surface area dependent desolvation
- Electrostatic desolvation – Poisson-Boltzmann model
- PB electrostatic interaction
- Sidechain rotamer conformational entropy
  - Count accessible rotamers
    - E.g. as computed in InsightII
  - No significant contribution in any COMBINE model derived so far
  - Poor model of entropic changes upon binding?
- +++???

Surface area dependent energy terms

- Surface area dependent desolvation
  - Entropic, enthalpic
- May be dependent on whether residue is polar or non-polar

$$\sum_i w_i^p \Delta SA_i^p + \sum_i w_i^{np} \Delta SA_i^{np}$$

- For a amino acid sidechain/nucleic acid base

$$\Delta\Delta G_i^{hyd} = \Delta G_i^{hyd} \left(\Delta SA_i / SA_i^T \right)$$

Cf measured solvation free energies ΔG_i^{hyd}
and total reference solvent-accessible surface areas SA_i^T

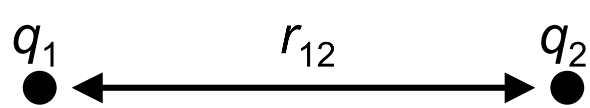
Surface areas computed e.g. with NACCESS (Hubbard)

Tomic et al, J. Med. Chem, 2000

Poisson-Boltzmann terms

- Electrostatic desolvation – Poisson-Boltzmann model
- PB electrostatic interaction
- Coulomb's Law gives the Interaction energy of two point-charges *in vacuo* :

$$U = \frac{332 q_1 q_2}{r_{12}}$$



Energy, U: kcal/mol
Charge, q: electron charges
Distance, r: Angstroms

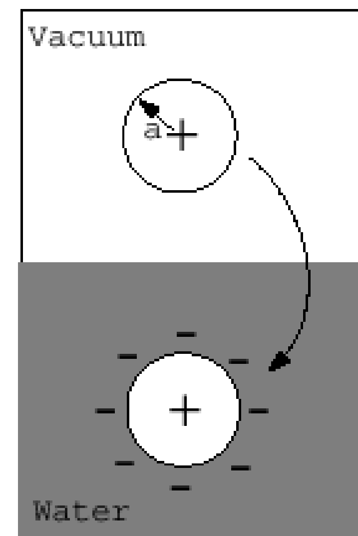
- Screening by ε=r is an empirical approximation of the heterogeneous dielectric environment.

Electrostatic solvation

- In a high dielectric solvent (e.g. water), charges will tend to be repelled from low dielectric solutes
- **Born ion solvation**
  - Work done to transfer Born ion between two dielectrics
  - Born ion is point charge in spherical cavity

$$U_{Born} = \frac{332q^2}{2a}\left(\frac{1}{\varepsilon_2} - \frac{1}{\varepsilon_1}\right)$$

  - E.g. Na+, K+, Cl- in water: a ~ 1.5-2.5 Å
    Free energy of hydration ~ -100- -50 kcal/mol
  - Charge polarizes solvent, which produces reaction field at charge with which charge interacts



M. Gilson

Molecules in solvent

- Total electrostatic free energy of a molecule in water:
  - Sum of Coulombic and solvation terms

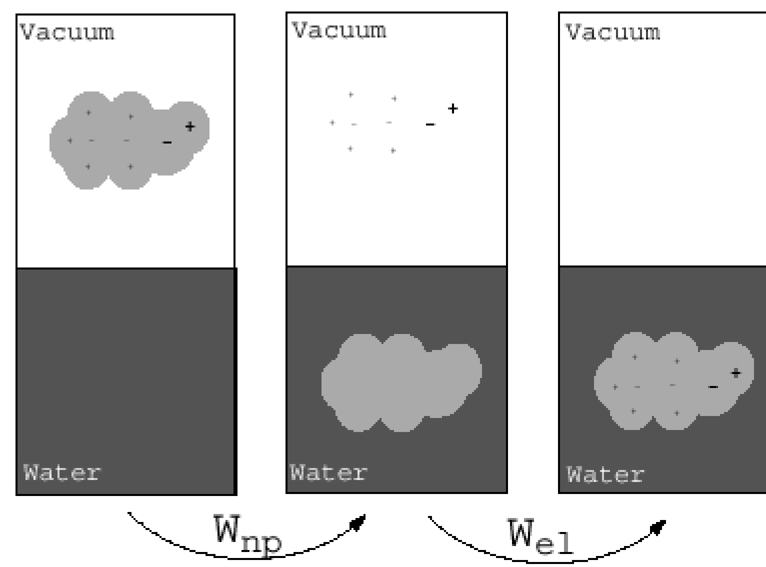
$$U_{elec} = \sum_{i>j} \frac{q_i q_j}{4\pi\varepsilon_o \varepsilon_m r_{ij}} + \frac{1}{2}\sum_i q_i \phi_{RF}$$

- Solvation energy also due to
  - van der Waals interactions
  - Hydrophobic effect

$$U_{solvation} = U_{nonpolar} + U_{elec}$$

  - E.g.

$$U_{nonpolar} = f(surfacearea)$$

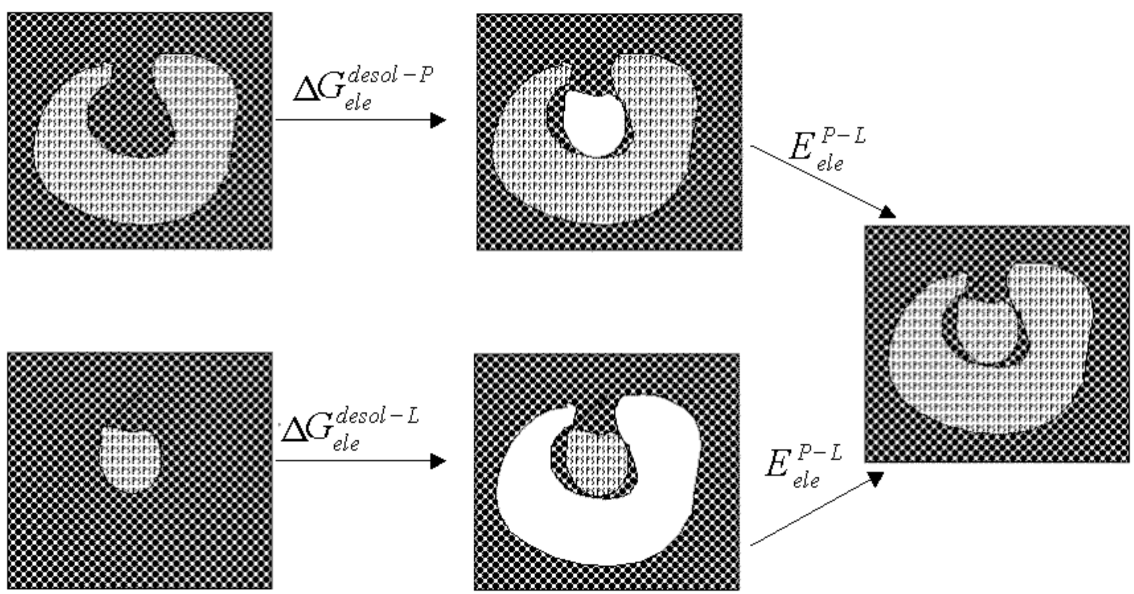


M. Gilson

PB terms for COMBINE analysis

$$\Delta G_{ele}^{bind} = E_{ele}^{P-L} - \Delta G_{ele}^{desol-P} - \Delta G_{ele}^{desol-L}$$

- Need to compute energy such that interaction term E_{ele}^{P-L} can be decomposed
- Use stepwise procedure



Mobile ions in the solvent

- Ionic solution with dissolved ions (electrolyte)
  - Ions redistribute in the presence of a molecule with charges to weaken/screen its electrostatic interactions

- Debye-Hueckel theory
  - Ions assumed to distribute according to the local potential with a Boltzmann factor

$$c_{ion}(\mathbf{r}) = c_{ion,bulk} e^{-\beta\phi(\mathbf{r})q_{ion}}$$

Ion screening

- For a point-charge in ionic solution:

$$\phi(r) = \frac{q}{4\pi\varepsilon_0\varepsilon_r r} e^{-r/r_{Debye}}$$

- Debye length:

$$r_{Debye} = \frac{1}{\kappa}$$

  - Characteristic distance of exponential screening
  - Decreases as ionic strength increases
  - At 150mM, Debye length is ca 8 Å

Poisson-Boltzmann equation

- For electrolyte with N types of ion each with charge q_i and concentration c_i:
  - Non-linear PB equation:

$$-\varepsilon_0 \nabla . \left[\varepsilon_r(\mathbf{r}) \nabla \phi(\mathbf{r}) \right] = \rho^f(r) + \sum_{i=1}^{N} q_i c_{i,bulk}(\mathbf{r}) e^{-\beta\phi(\mathbf{r})q_i}$$

Linearized Poisson-Boltzmann equation

- Taylor series expansion of Boltzmann factor and use terms up to first order in ϕ, assuming exponent is small

$$-\varepsilon_0 \nabla . \left[\varepsilon_r(\mathbf{r}) \nabla \phi(\mathbf{r}) \right] = \rho^f(r) - \varepsilon_0 \varepsilon_r(r) \kappa^2(r) \phi(r)$$

$$\kappa^2(r) = \frac{\beta}{\varepsilon_0 \varepsilon_r} \sum_1^N c_{i,bulk} q_i^2 = \frac{2e^2 N_A}{\varepsilon_0 \varepsilon_r kT} I$$

- Advantages over NLPBE:
  - Fields of charges are superposable
  - Electrostatic energy given by:

$$U = \frac{1}{2} \sum_i q_i \phi_i$$

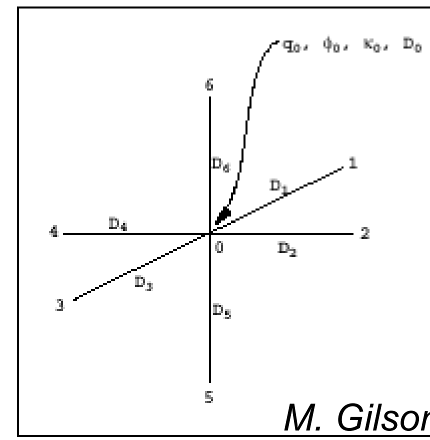
Solving the Poisson-Boltzmann equation

- Analytic solutions
  - Only for simple geometries: spheres, planar boundaries
- Numerical solutions
  - Convert continuous partial differential equation into discretized problem with systems of linear equations that can be solved by matrix methods
  - For example:
    - Boundary element method
      - Tesselate dielectric boundary, solve for surface charge on each boundary element
    - Finite difference method
      - Discretize space into a cubic lattice, solve iteratively to convergence for each grid point by finite differences
      - E.g. **UHBD**, Delphi

Solving the finite difference linearized Poisson-Boltzmann equation

- For a grid with grid spacing h
- Need to satisfy condition at each grid point, i:

$$\phi_i(r) = \frac{\sum_{1}^{6} \varepsilon_o \varepsilon_{r,j} \phi_j + q_i / h}{\sum_{1}^{6} \varepsilon_o \varepsilon_{r,j} + \varepsilon_o \varepsilon_{r,i} \kappa_0^2 h^2}$$



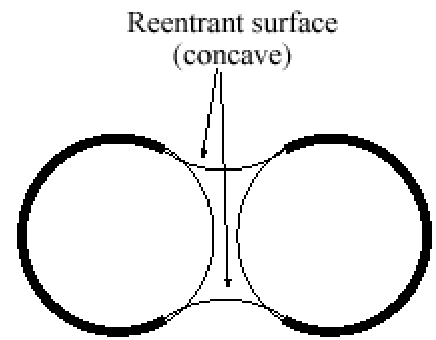
- Solve system of linear equations for all points
- Interpolate potential at any off-grid point from nearest 8 grid points
- Grid points at grid edge are fixed (boundary conditions)
  - E.g. from sum of Debye-Huckel screened potentials due to solute charges

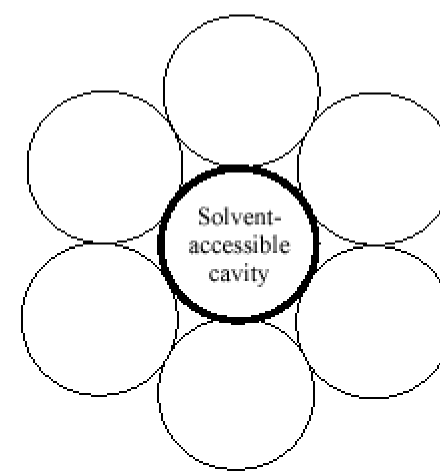
Solving the finite difference linearized Poisson-Boltzmann equation: Practical aspects

- Need to assign:
- Atomic radii and charges
  - From "standard" force field e.g. CHARMM, AMBER
  - Set specifically for continuum electrostatics, parameterized to reproduce small molecule solvation energies e.g. PARSE
  - NB Hydrogens: better convergence when radii ≠ 0
- Grid size and spacing
  - Large enough for boundary potentials to be sufficiently accurate
    - Focusing using nested grids is possible
  - Spacing of 0.5-1 Å for viewing potentials or Brownian dynamics forces
  - Spacing of 0.2-0.3 Å for energies, pK_as

Solving the finite difference linearized Poisson-Boltzmann equation: Practical aspects

- Need to assign:
- Dielectric boundaries
  - Van der Waals surface
  - Molecular surface
  - Smooth dielectric constant over points adjacent to the boundary



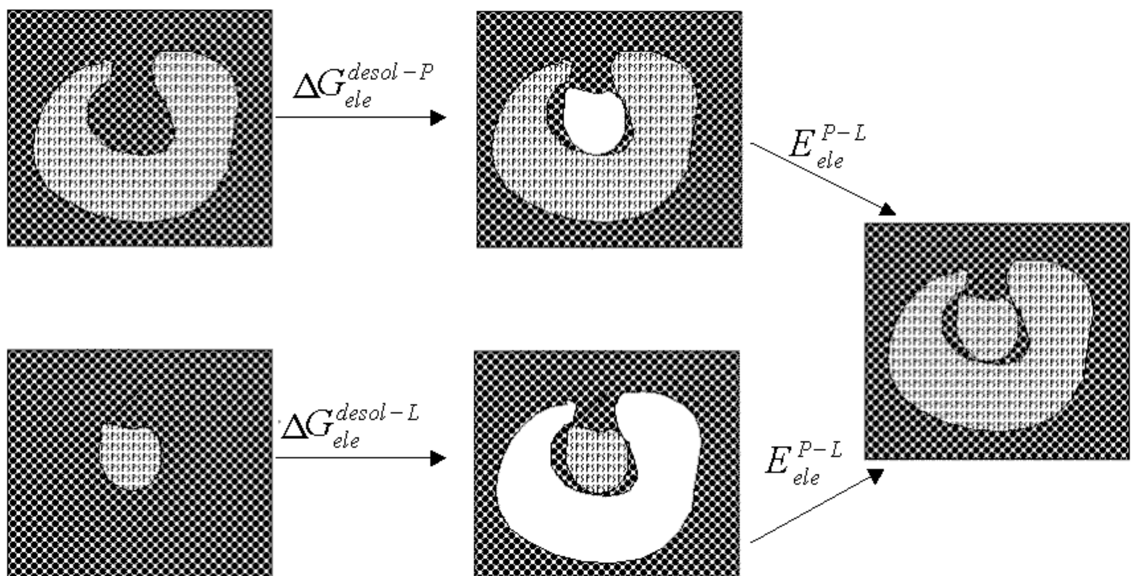


M. Gilson

PB terms for COMBINE analysis

$$\Delta G_{ele}^{bind} = E_{ele}^{P-L} - \Delta G_{ele}^{desol-P} - \Delta G_{ele}^{desol-L}$$

- Need to compute energy such that interaction term E_{ele}^{P-L} can be decomposed
- Use stepwise procedure



Energy terms for COMBINE Analysis

- ***Basic procedure + extras to consider such as:***
- ***Solvation:***
  - Continuum Electrostatics
  - Explicit water molecules
- ***Conformations:***
  - Comparative modelling
  - Docking-*de novo/* comparative
  - Conformational changes on binding
  - Entropy

Lecture 3: Overview

- How can the energy components for COMBINE analysis be computed?
  - Modeling of complexes
  - Energy minimization
  - Calculation of energy terms
  - Computation of surface-area dependent terms
  - Computation of Poisson-Boltzmann terms