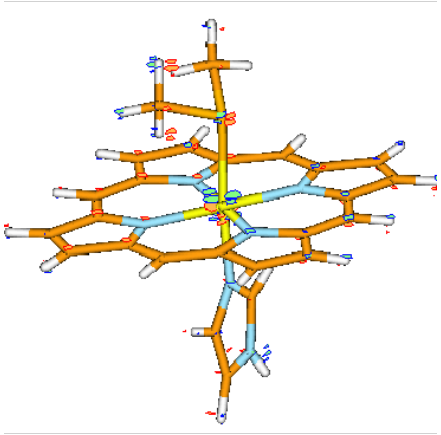
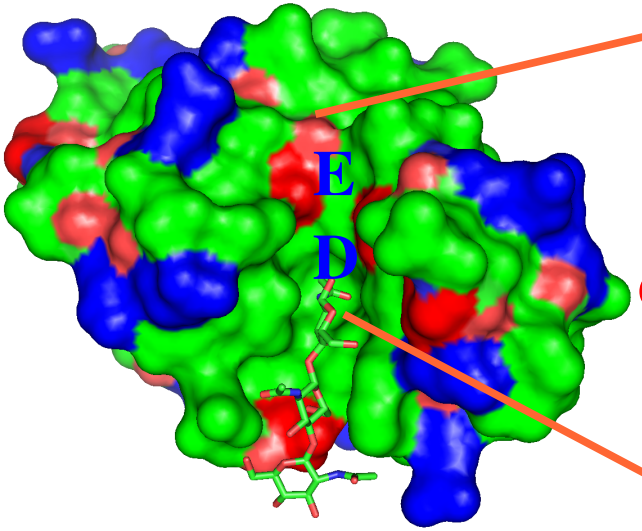
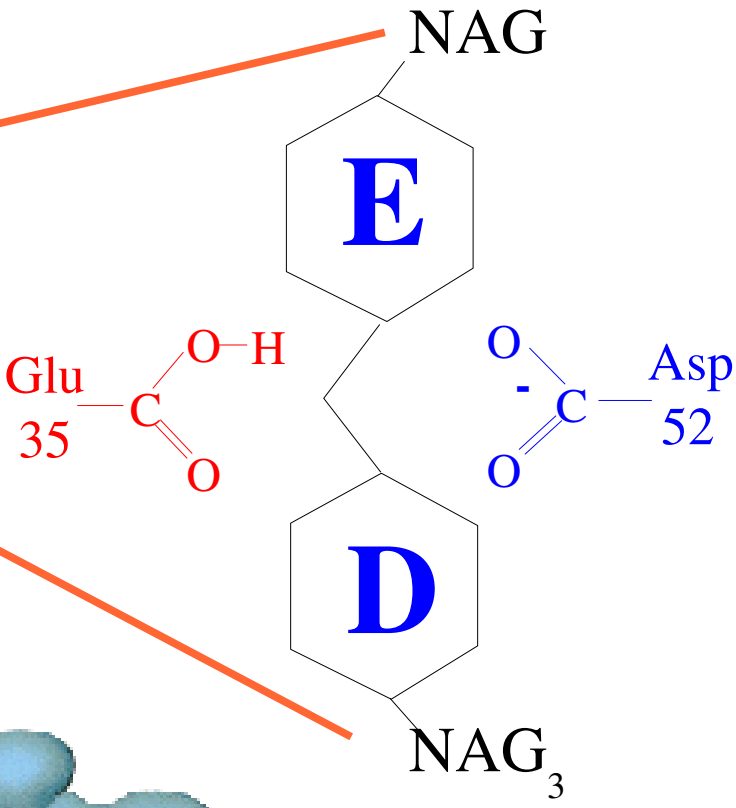
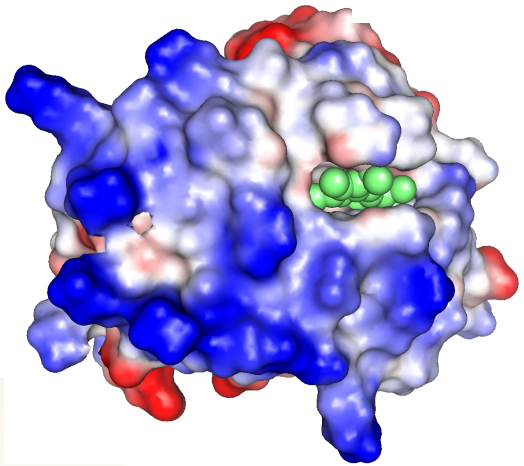
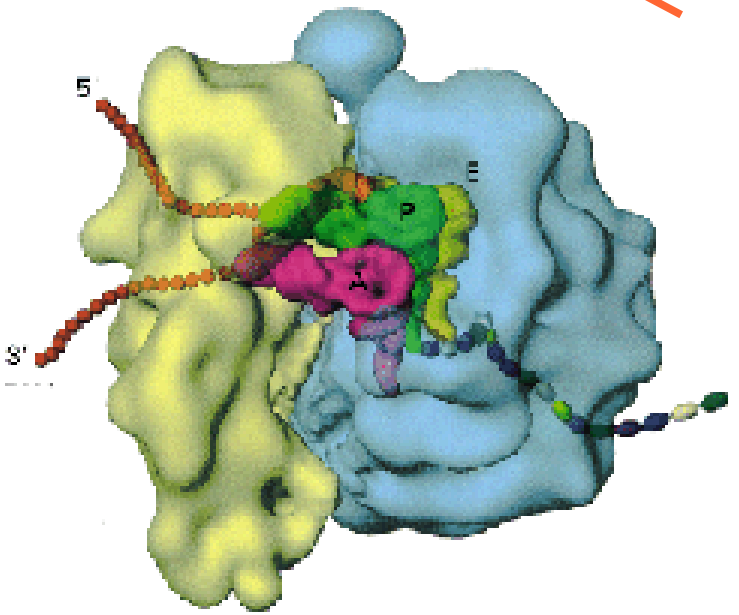
Simple free energy methods for protein electrostatics





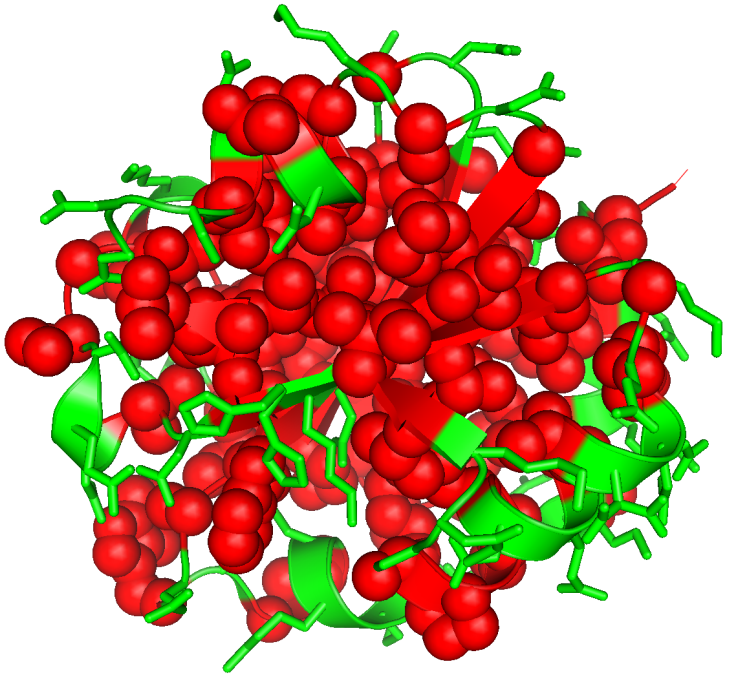






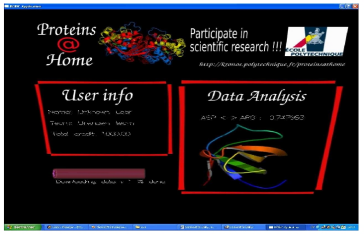
The quest for simple models

As a crude 1st approximation, protein interior = simple, homogeneous dielectric



More generally

Integrate out selected degrees of freedom to obtain “implicit” models
Simple gaussian model for fluctuations: linear response models

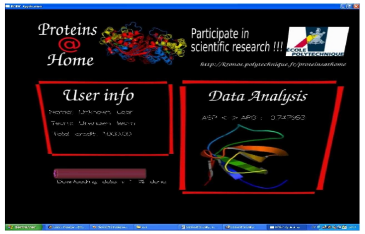


Simple free energy methods for protein electrostatics
Constant pH Monte Carlo for pK_a and redox calculations

Protons and electrons as electrostatic reporters
Dielectric continuum model for selected degrees of freedom
Monte Carlo sampling to obtain free energies
Some general notions concerning electrostatics
Unusual numerics, borrowed from Protein Design

Alexey Aleksandrov, Thomas Simonson
Dept of Biology, Ecole Polytechnique, Paris

Savvas Polydorides, Georgios Archontis
Dept of Physics, University of Cyprus, Nicosia



Constant pH Monte Carlo for pK_a and redox calculations

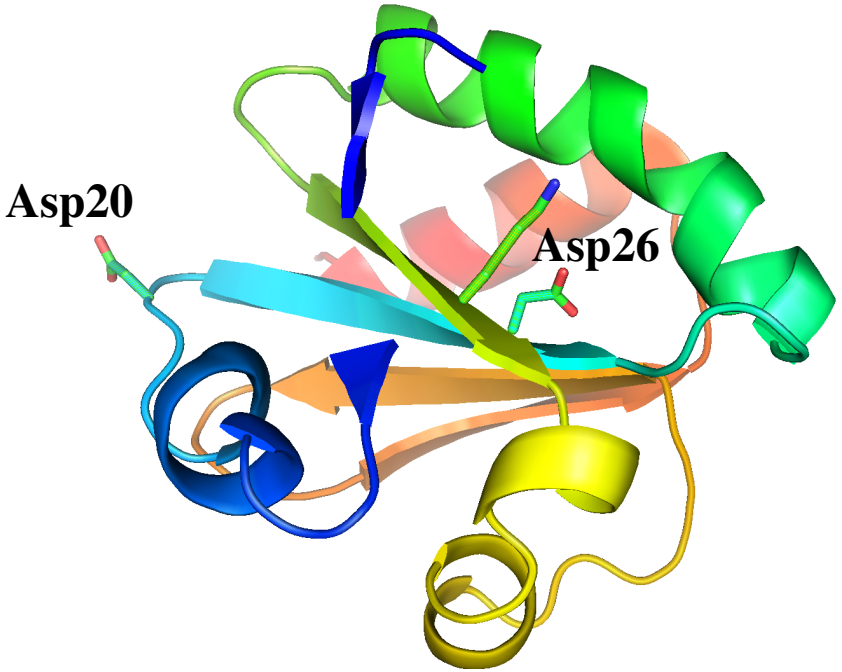
- Simple pKa methods cannot give both energy gaps and reorganization
- Protein design technology; pKa's as a special case
- Constant pH Monte Carlo framework
- Generalized Born (GB) with pairwise complexity
- Test results for pKa's and titration curves

Alexey Aleksandrov, Thomas Simonson
Dept of Biology, Ecole Polytechnique, Paris

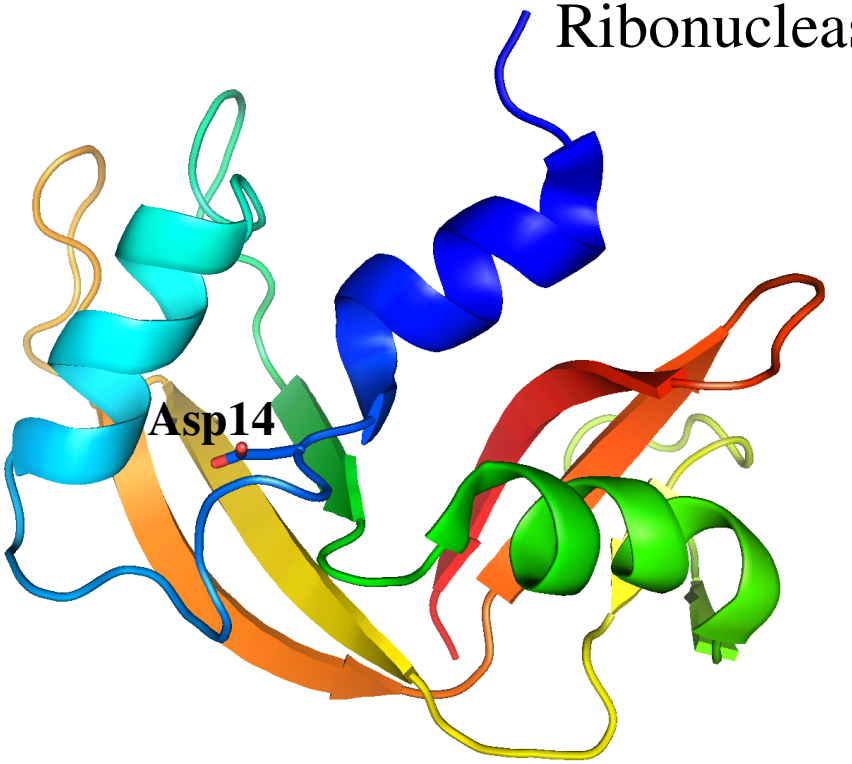
Savvas Polydorides, Georgios Archontis
Dept of Physics, University of Cyprus, Nicosia

pK_a calculations with MD free energy simulations using explicit or implicit (GB) solvent

Thioredoxin



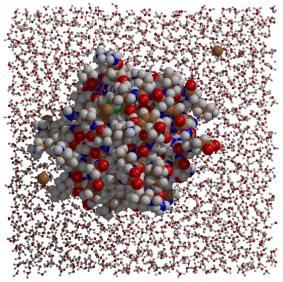
Ribonuclease A

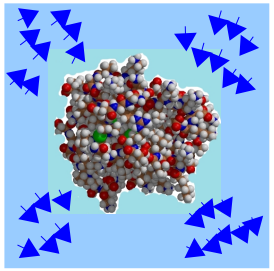


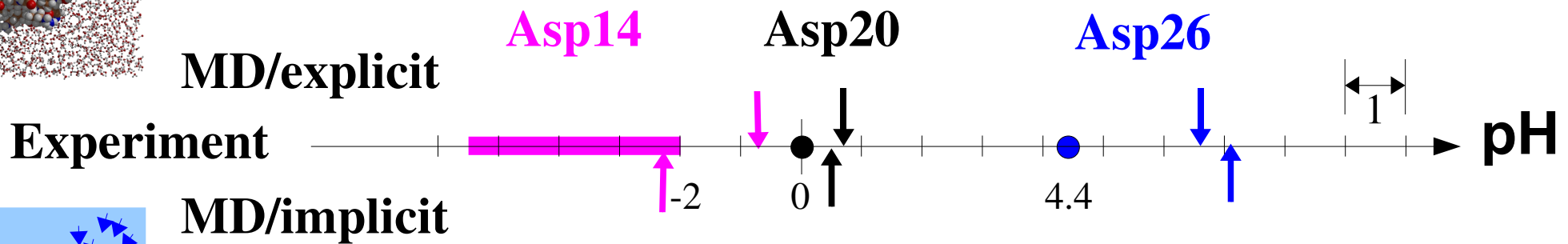
TS, Carlsson, Case (2004) J Amer Chem Soc; Archontis & TS (2005) Biophys J

pK_a calculations with MD free energy simulations

pK_a shifts	Asp14	Asp20	Asp26
experiment	< -2	0	+4.4
MDFE/explicit	-0.9	1.3	+6.7
MDFE/implicit	-2.2	0.5	+7.1







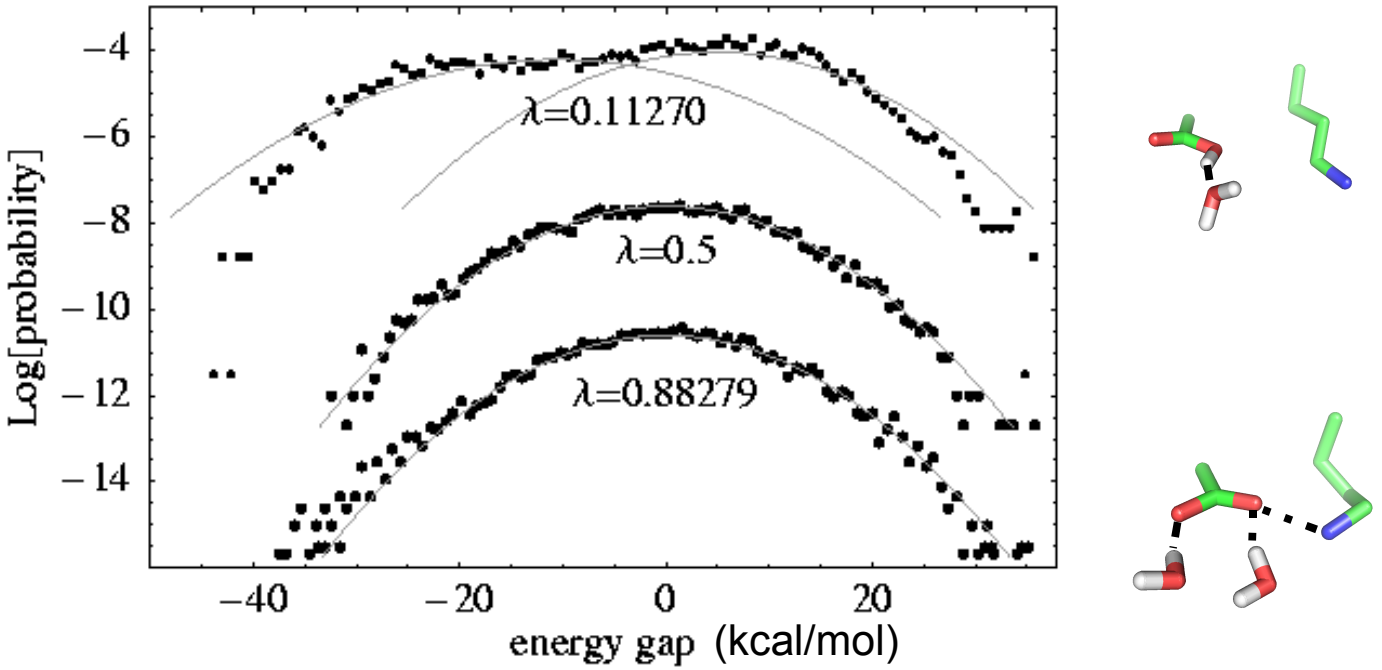
TS, Carlsson, Case (2004) J Amer Chem Soc; Archontis & TS (2005) Biophys J

Compute ΔG for protonation, but also energy gap δH and reorganization energy Λ

$$\Delta G = \langle \delta H \rangle_0 + \Lambda$$

Marcus, Hush, 1960s

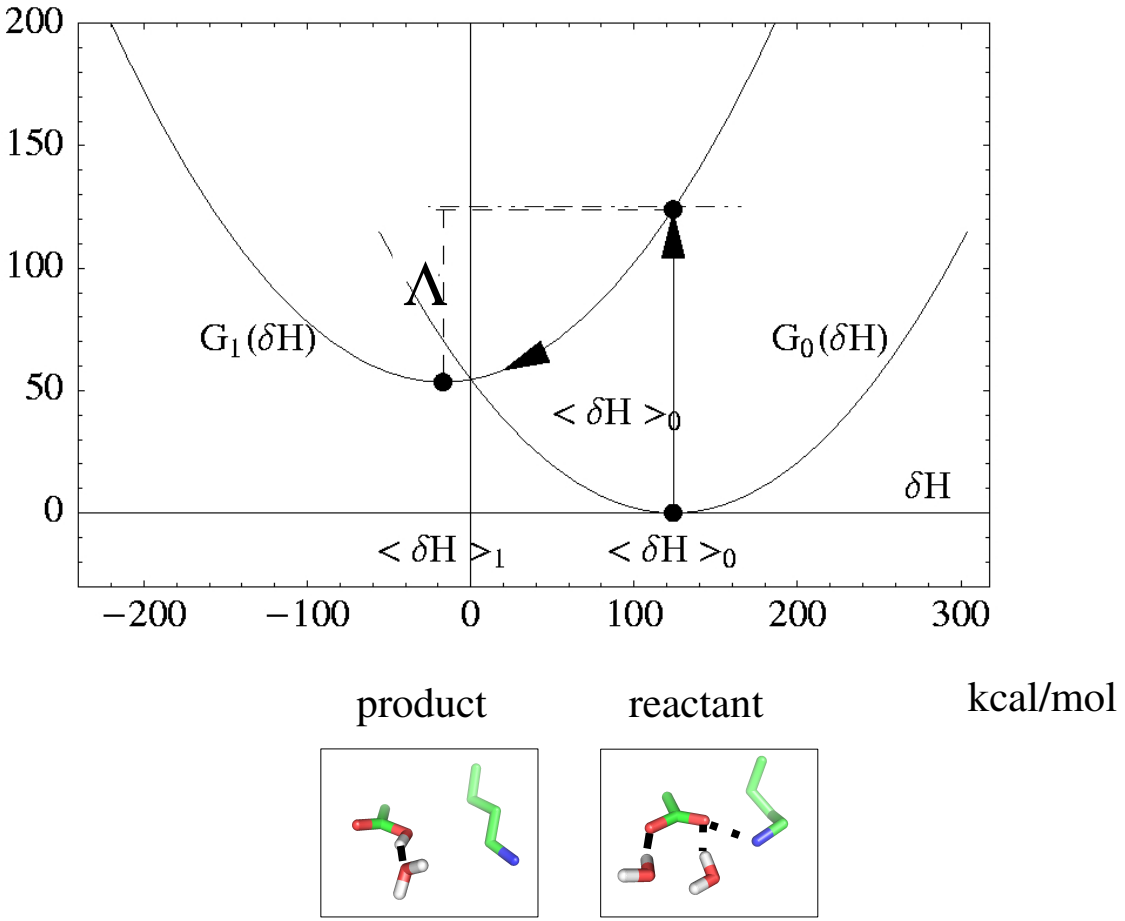
Asp26 energy gap histograms from explicit solvent simulations



The energy gap corresponds to instantaneous protonation of Asp26.
λ = 0 is the protonated state; λ = 1 is the ionized state.

Marcus diabatic free energy curves (Asp26)

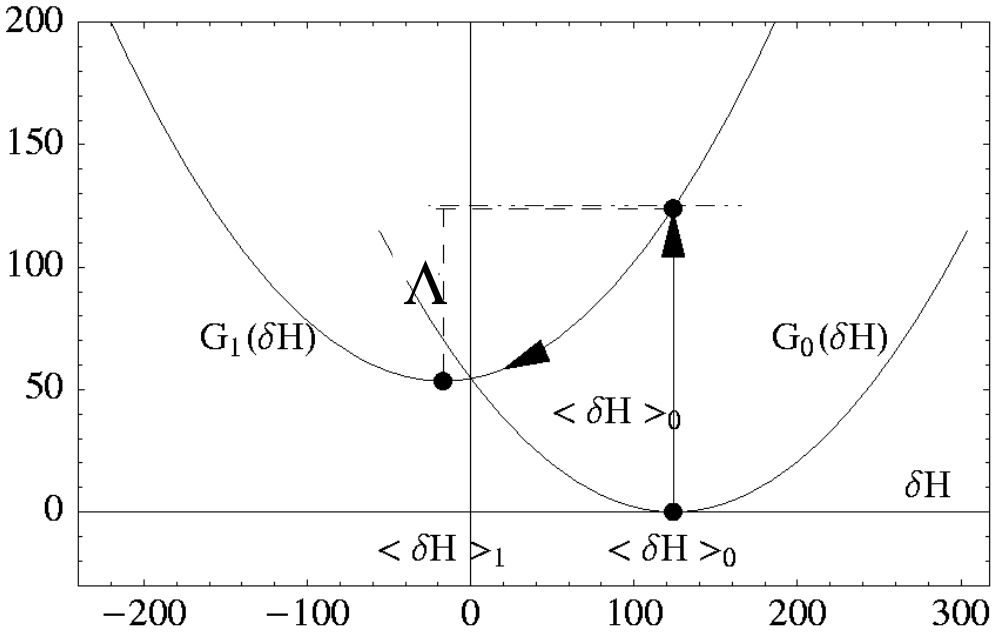
$$\Delta G = \langle \delta H \rangle_0 + \Lambda$$



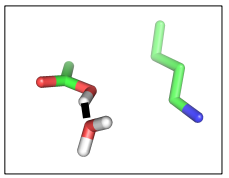
Simonson RepProgPhys 2003, Archontis & Simonson, Biophys J, 2005

Compare to Poisson-Boltzmann approach (Marcus, 1960s)

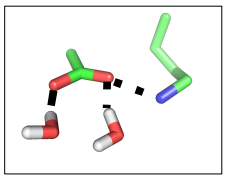
Explicit solvent



product



reactant



kcal/mol

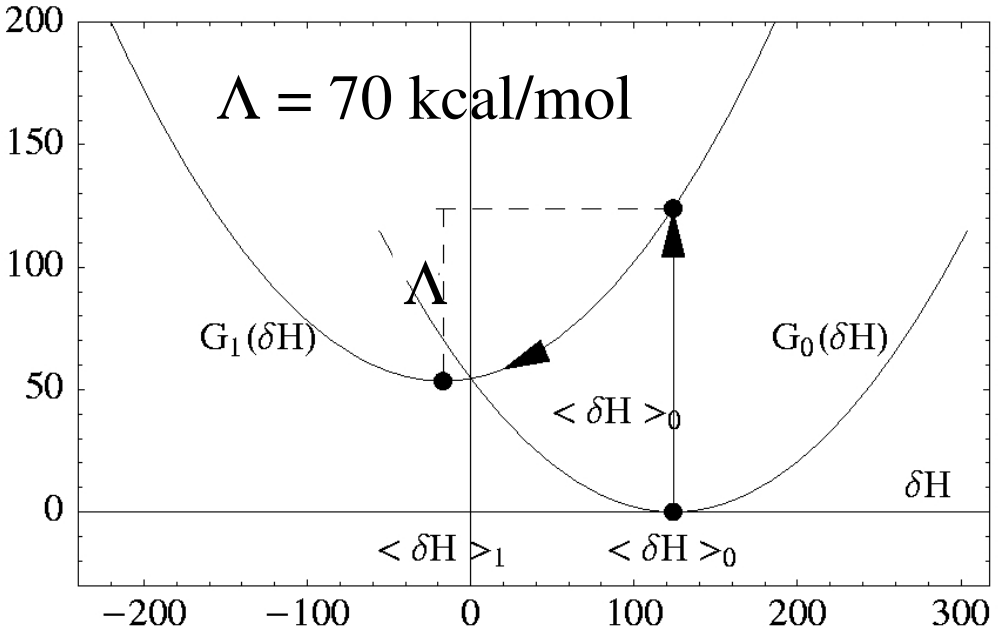
Dielectric continuum

?

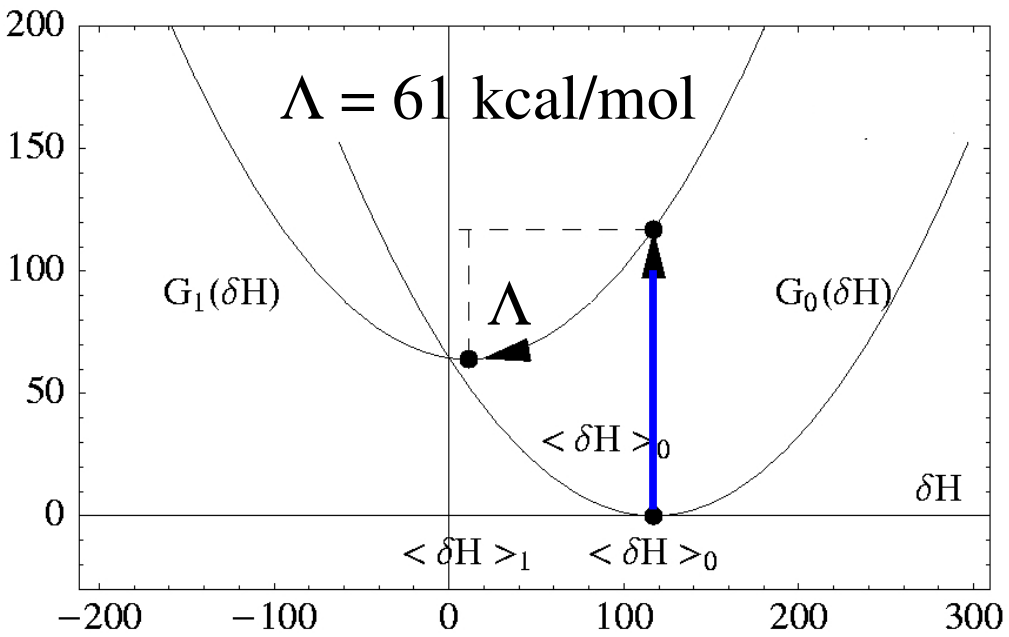
Archontis & Simonson, Biophys J, 2005

Marcus diabatic free energy curves (Asp26)

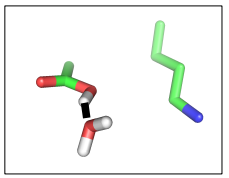
Explicit solvent



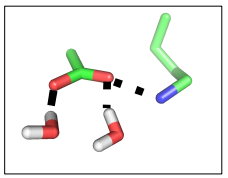
Dielectric continuum



product



reactant



kcal/mol

$\varepsilon_p^{stat} = 2$, $\varepsilon_p^{rlx} = 3$

Good agreement for overall free energy but also for separate components when *two distinct PB models* are used for the "vertical" and "reorganization" steps.

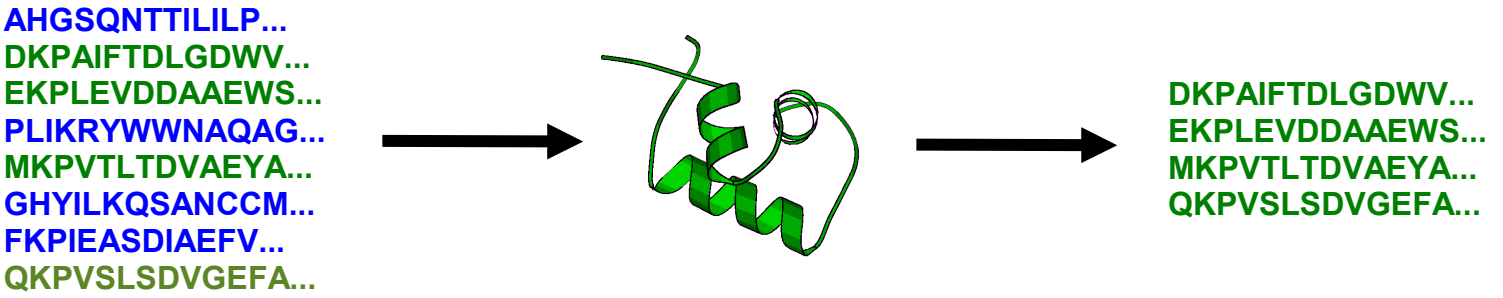
A constant pH Monte Carlo method for pKa's

- pKa methods should not be too simple
- Protein design technology; pKa's as a special case
- GB with pairwise complexity
- Constant pH Monte Carlo framework
- Test results for pKa's and titration curves

Protein design: Search sequence/conformation space for preferred/functional sequences

random mutagenesis + selective pressure

Ponder, Richards (1987); Hellinga, Richards (1994); Mayo (1997) Desjarlais (1998)
Koehl, Levitt (1999); Baker, Serrano, Wodak, Handel (2000-04); others...



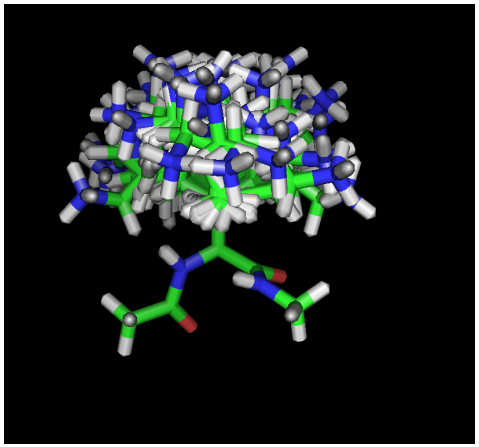
Sequence score: $\Delta G = \Delta G_{folded} - \Delta G_{unfolded}$

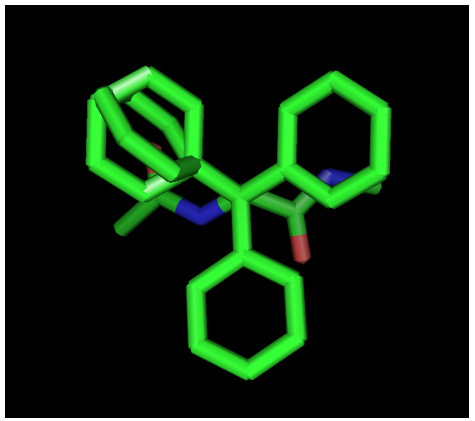
Free energy function

Structural model of both ensembles

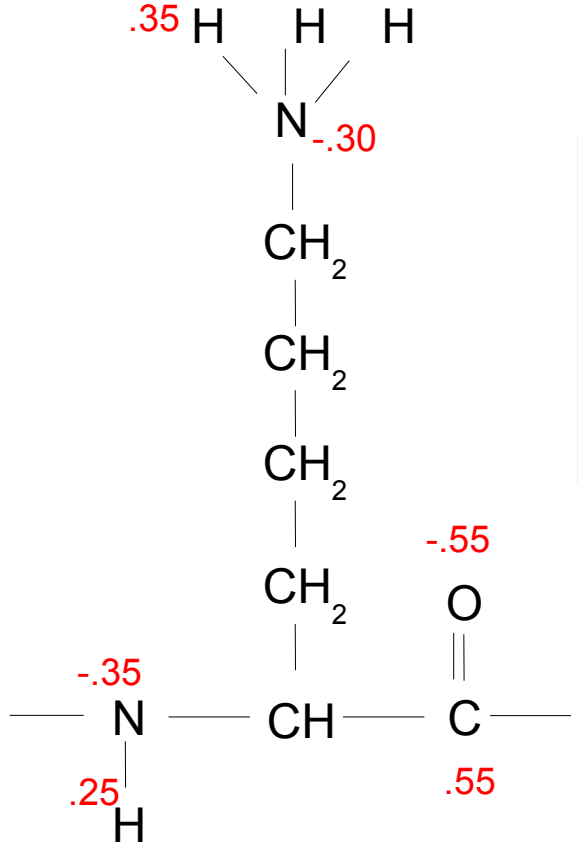
Discrete space + pairwise energy function

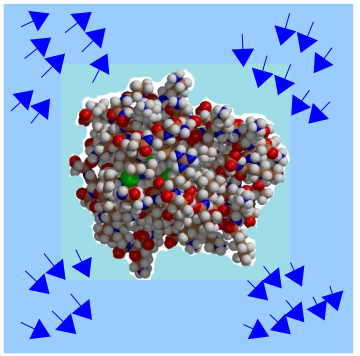
Conformational space
=
fixed backbone
+
sidechain rotamers





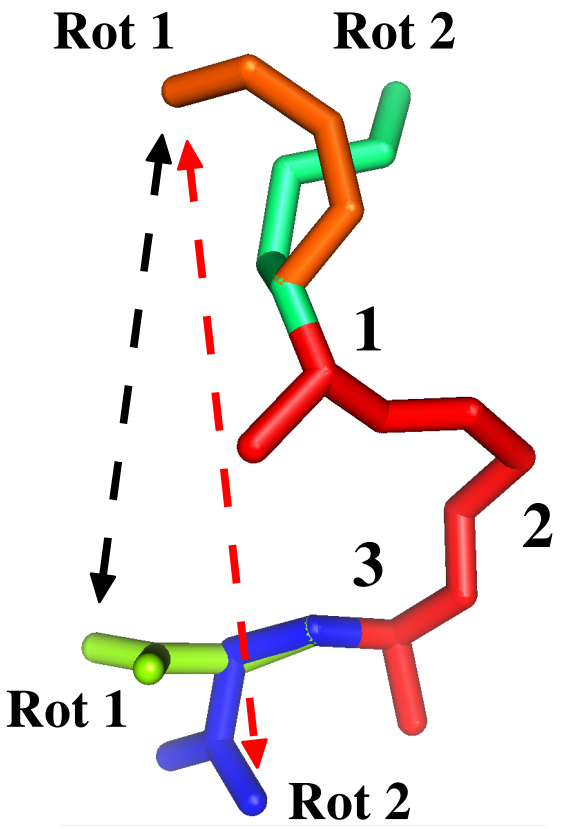
Free energy
=
molecular mechanics
+
simple dielectric shielding





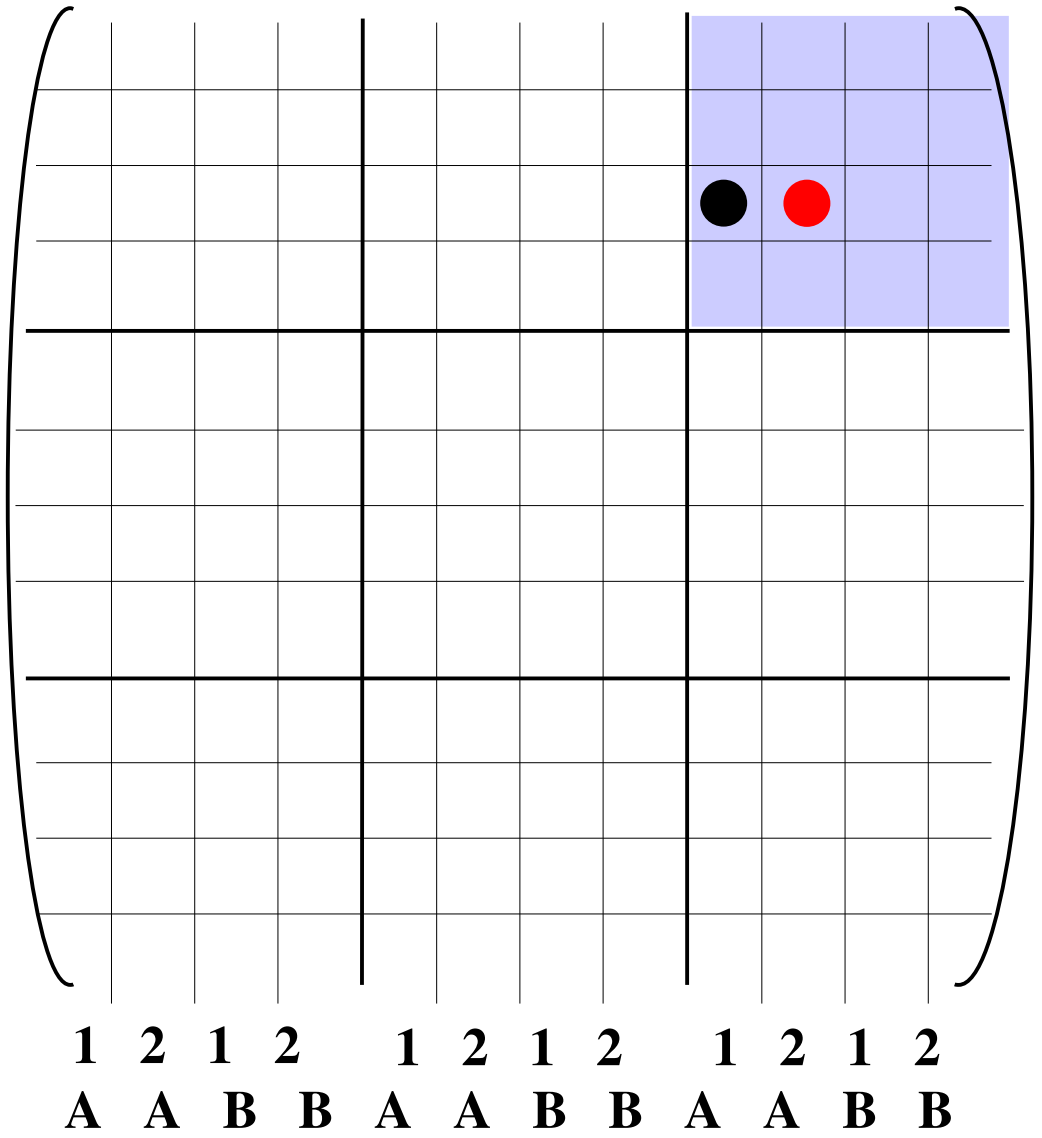
1) Precompute all pair interactions, allowing for all residue types and rotamers (Mayo, 1997)

Structure

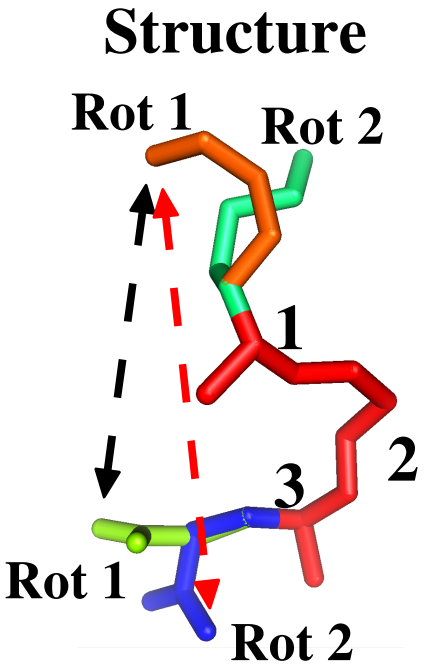


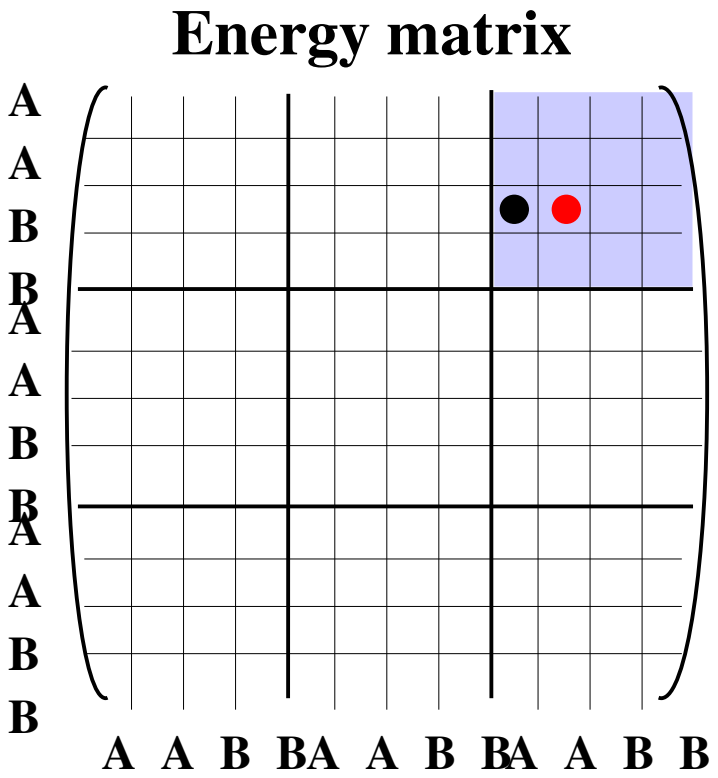
Energy matrix

Position	Type	Rotamer
1	A	1
	A	2
	B	1
	B	2
2	A	1
	A	2
	B	1
	B	2
3	A	1
	A	2
	B	1
	B	2



1) Precompute all pair interactions, allowing for all residue types and rotamers



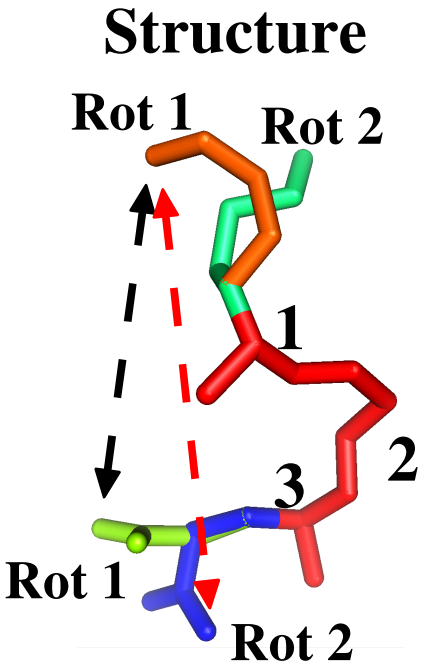


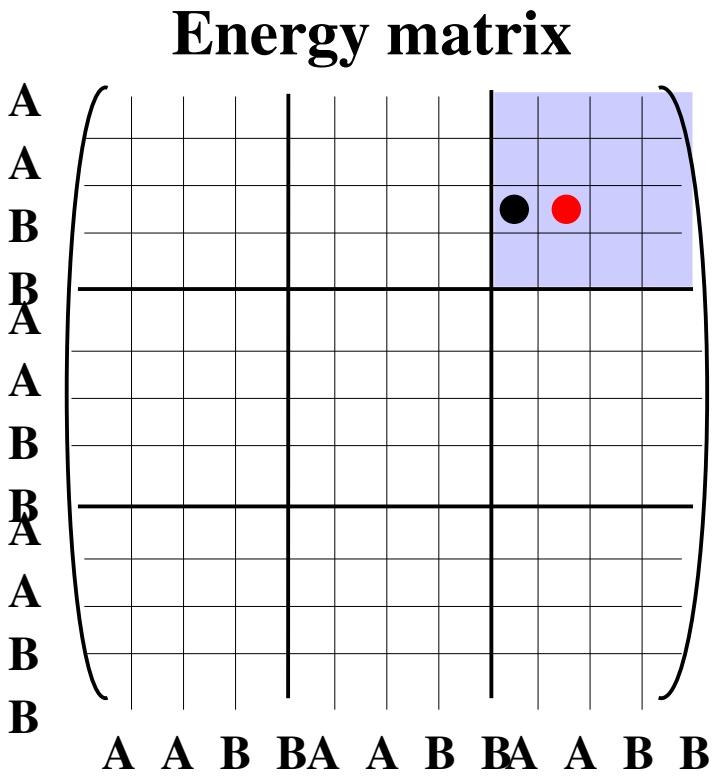
2) Combinatorial exploration of sequence/structure space

Monte Carlo, mean field, branch and bound, heuristic, ...

Mayo, 1997; Schmidt, Lopes, TS et al, 2007-11, J Comp Chem, Proteins, Plos One

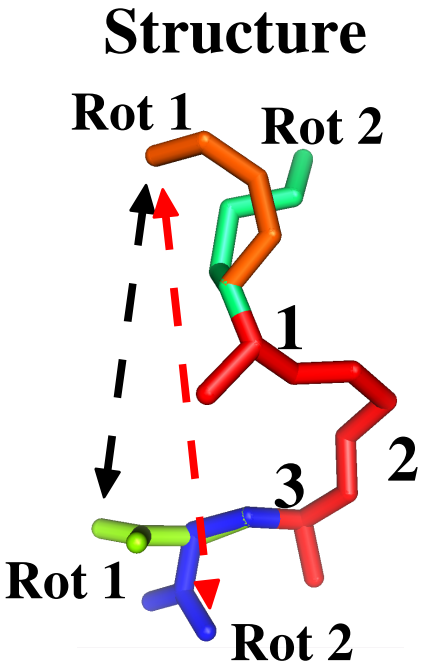
1) Precompute all pair interactions, allowing for all protonation states and rotamers

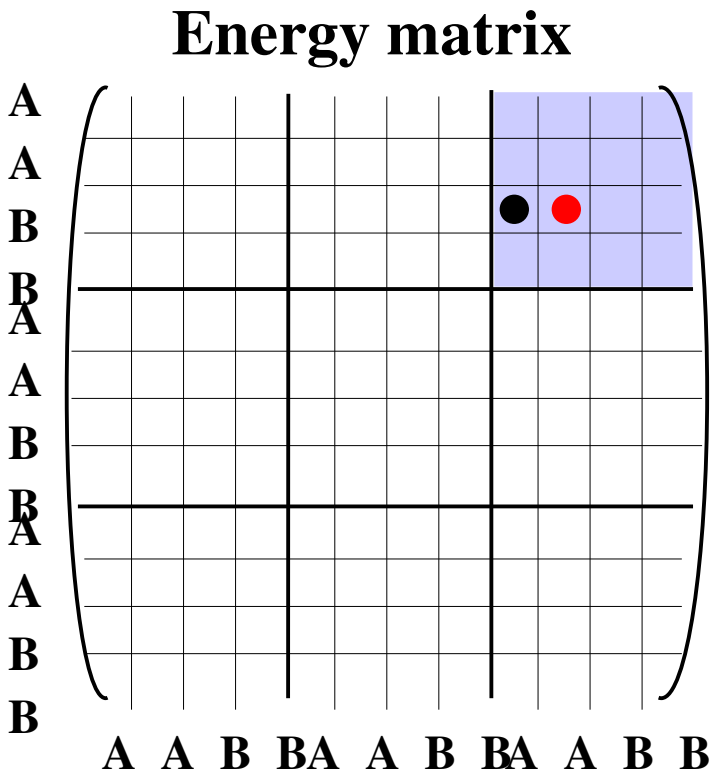




2) Monte Carlo exploration of protonation/structure space

1) Precompute all pair interactions, allowing for all protonation states and rotamers



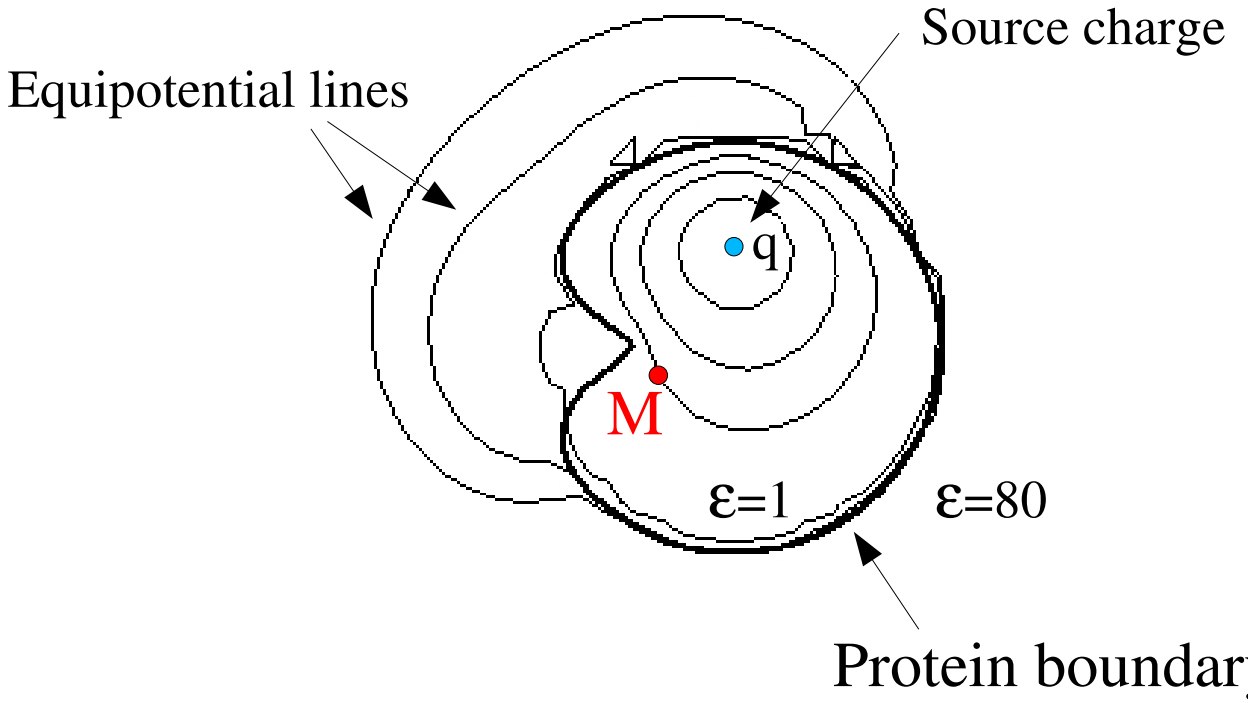


2) Monte Carlo exploration of protonation/structure space

Calculate pK_as with MC in protonation/structure space

Bashford, Case, Gunner, Knapp, ...

The continuum electrostatic solvent model is intrinsically non-pairwise

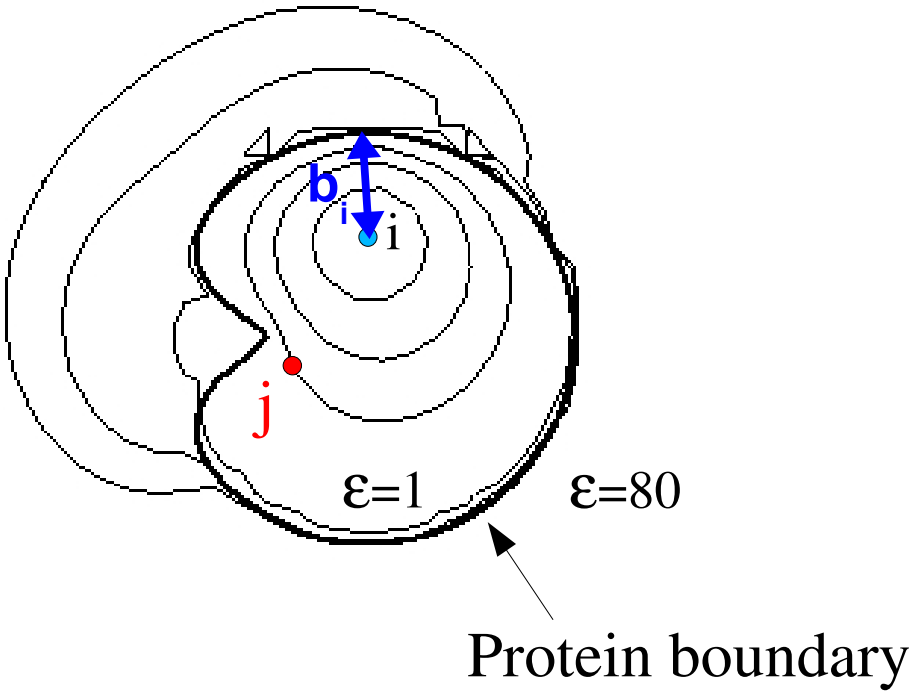


The potential at M depends on the shape of the cavity.

The generalized Born solvent model is intrinsically non-pairwise

$$E_{int}(i,j) = (1/\varepsilon_W - 1/\varepsilon_P)\, q_i\, q_j /(r_{ij}^2 + b_i b_j \exp[-r_{ij}^2/4b_i b_j])^{1/2}$$

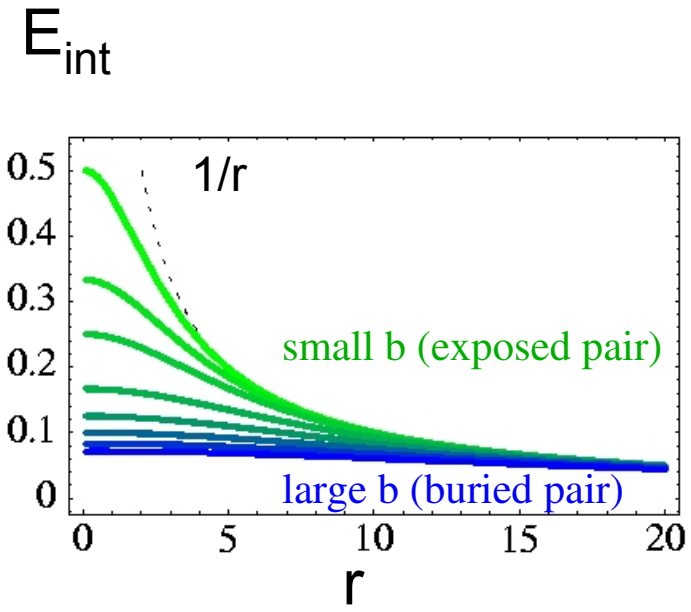
b_i, b_j = "solvation radii", depend on entire structure

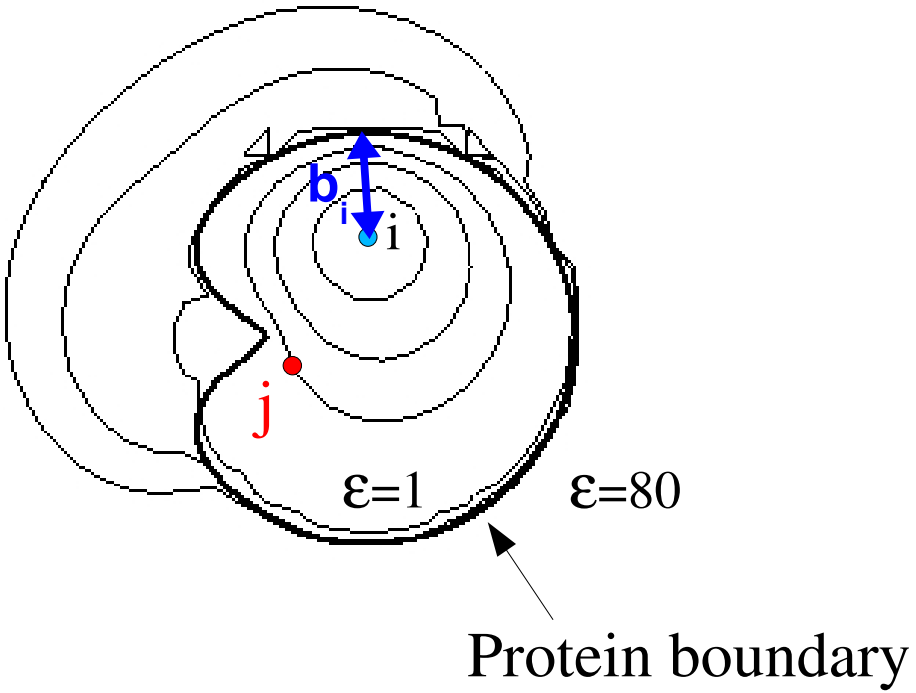


The generalized Born solvent model is intrinsically non-pairwise

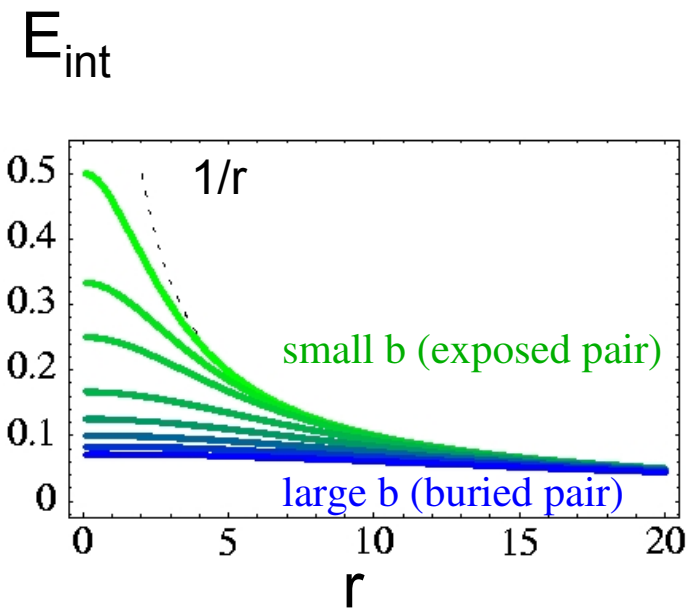
$$E_{int}(i,j) = (1/\varepsilon_W - 1/\varepsilon_P)\, q_i\, q_j /(r_{ij}^2 + b_i b_j \exp[-r_{ij}^2/4b_i b_j])^{1/2}$$

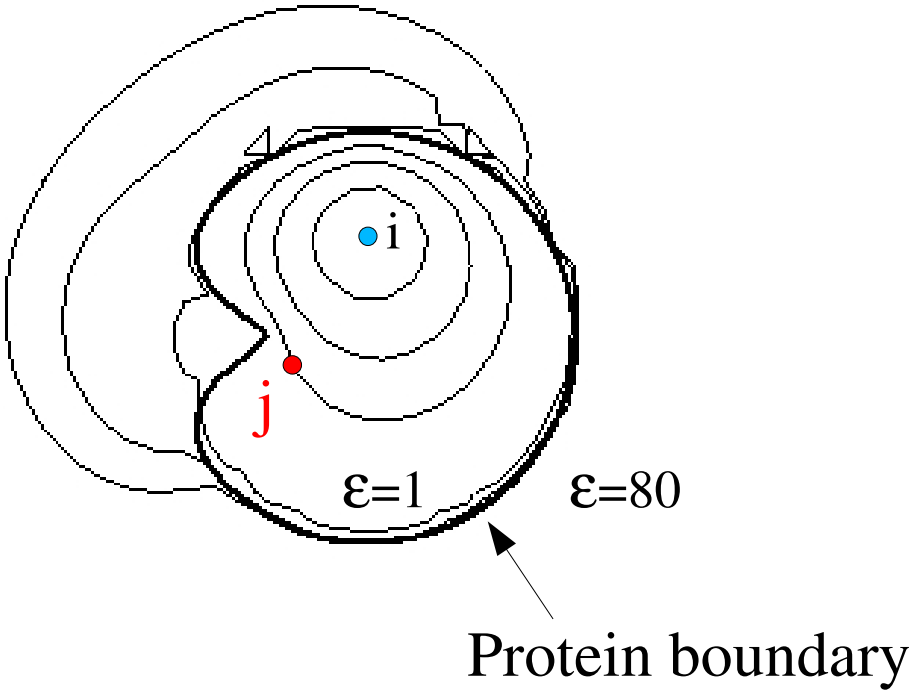
b_i, b_j = "solvation radii", depend on entire structure





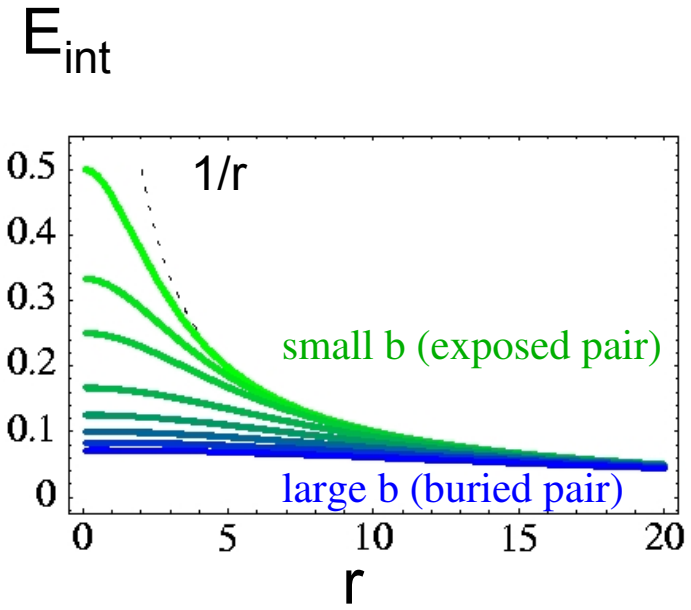
The continuum electrostatic solvent model is intrinsically non-pairwise





Many workers define an average, effective boundary:
Caflisch (docking), Gunner (pKa's), Handel, Mayo, Harbury (design), ...

A generalized Born model with "pairwise complexity", suitable for protein design and pKa calculations



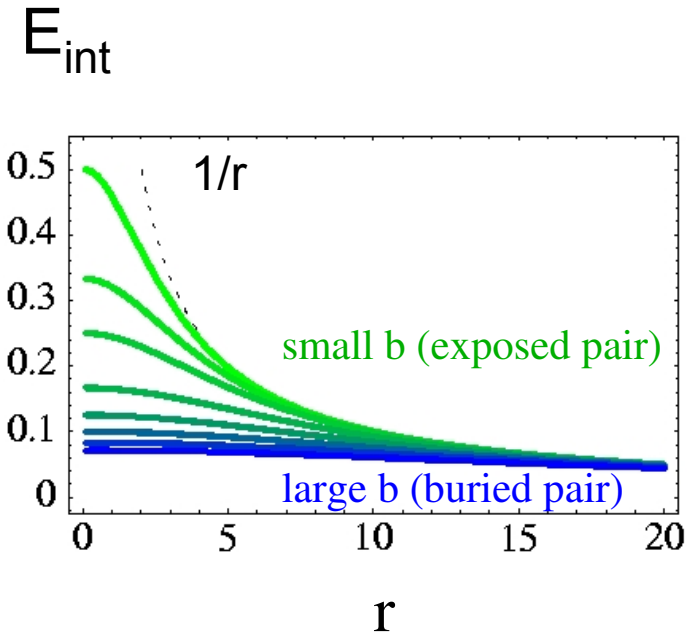
Generalized Born analytical approximation for dielectric screening:

$$\Delta W = \Sigma_{i,j} E_{int}(i,j)$$

$$E_{int}(i,j) = (1/\varepsilon_W - 1/\varepsilon_P)$$

$$q_i q_j /(r_{ij}^2 + b_i b_j \exp[-r_{ij}^2/4b_i b_j])^{1/2}$$

A generalized Born model with "pairwise complexity", suitable for protein design and pKa calculations



Generalized Born analytical approximation for dielectric screening:

$$\Delta W = \Sigma_{i,j} E_{int}(i,j)$$

$$E_{int}(i,j) = (1/\varepsilon_W - 1/\varepsilon_P)$$

$$q_i q_j /(r_{ij}^2 + b_i b_j \exp[-r_{ij}^2/4b_i b_j])^{1/2}$$

Pairwise complexity in two steps:

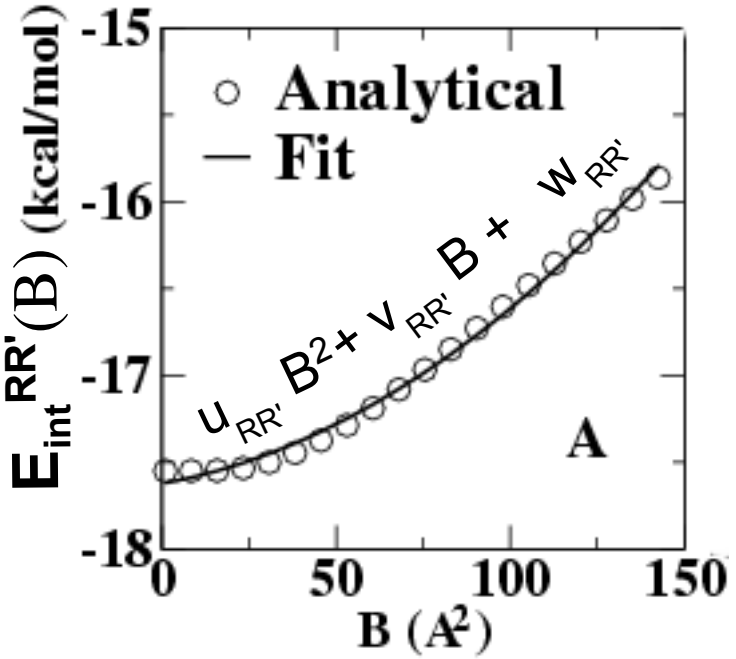
1) switch from atomic b_i's to residue-averaged values, b_R

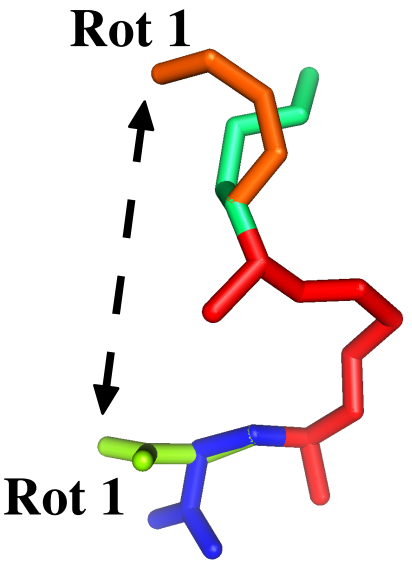
2) for each residue pair, precompute large-b and small-b interaction E_{int}, for future interpolation

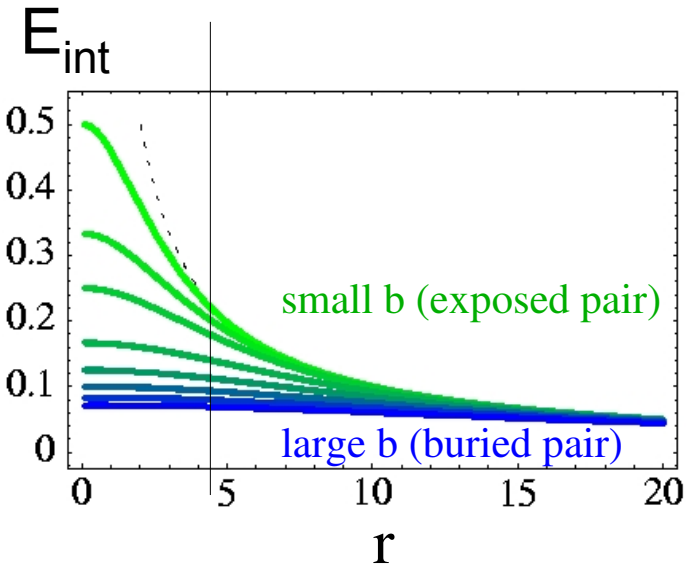
Archontis & Simonson (2005) JPCB, Aleksandrov et al (2010) JPCB

2a) for any pair of residues R, R', precalculate the GB interaction energy $E_{int}^{RR'}$ for small and large values of the product $B = b_R b_{R'}$

2b) calculate (and store) the fitting coefficients needed to interpolate for any intermediate B value:





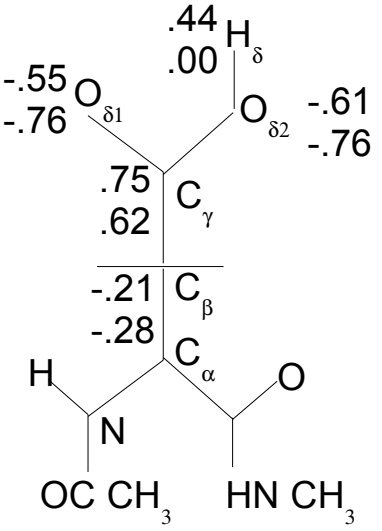


To describe all possible GB interactions for the entire system, we need 3 coefficient matrices: $(u_{RR'})$, $(v_{RR'})$, $(w_{RR'})$. For any protein conformation, we can reconstruct any R, R' contribution to the GB energy from these.

A multiconformation method for pKa calculations

- Need for sophisticated pKa methods
- Protein design technology; pKa's as a special case
- GB with pairwise complexity
- Constant pH Monte Carlo framework
- Test results for pKa's and titration curves

Constant pH Monte Carlo framework



Partition the degrees of freedom into explicit/implicit

Explicit: rotamer changes, protonation/deprotonation

Implicit: solvent + protein backbone motions and electronic polarizability

$$P_J(N+1;V,T,\mu) / P_J(N;V,T,\mu) = \exp\{-\beta[W_J(N+1)-W_J(N)]\}\exp(\beta\mu)$$

W = potential of mean force

$$\Delta W_J(\text{prot}) = W_J(N+1) - W_J(N)$$

Protein-Asp^- ⟶ Protein-AspH

Asp^- ⟶ AspH

$\Delta W(\text{model})$

$$\mu = \mu^0 - 2.303\ kT\ pH$$

$$\Delta W(\text{model}) = \mu^0 - 2.303\ kT\ pK_{model}$$

Aleksandrov et al (2010) JPCB

Choice of protein dielectric constant ε_P?

Most pKa methods use a high value, 20-80

Partition of the degrees of freedom into explicit/implicit

Explicit: rotamer changes, protonation/deprotonation

Implicit: solvent + protein backbone motions and electronic polarizability

a) Fit ε_P to experimental data

b) Do MC with different ε_P values; compare fluctuations to MD

c) Connect protein fluctuations to a dielectric constant; compare MD/MC

$$\langle \Delta M_{LF}^2 \rangle / kTR^3 = G = \{ f(\varepsilon_P,\varepsilon_W)(\varepsilon_P-1) - f(\varepsilon_P^{HF},\varepsilon_W^{HF})(\varepsilon_P^{HF}-1) \} / f(\varepsilon_P^{HF},\varepsilon_W)$$

$$f(\varepsilon_P,\varepsilon_W) = 3\varepsilon_W / (\varepsilon_P + 2\,\varepsilon_W)$$

Compute fluctuations with Monte Carlo + GB(ε_P)

Compare to MD in explicit solvent

$$\varepsilon_p \sim 3 - 4$$

Simonson RepProgPhys 2003

6 test proteins: barnase, BPTI, ovomucoid, protein G, lysozyme, thioredoxin
78 pKa's

AMBER force field and GB implementation

fixed backbone; Tuffery rotamer library for sidechains

protein dielectric $\varepsilon_p = 4$

pH increased gradually

10^7 Monte Carlo moves in protonation/rotamer space for each pH value

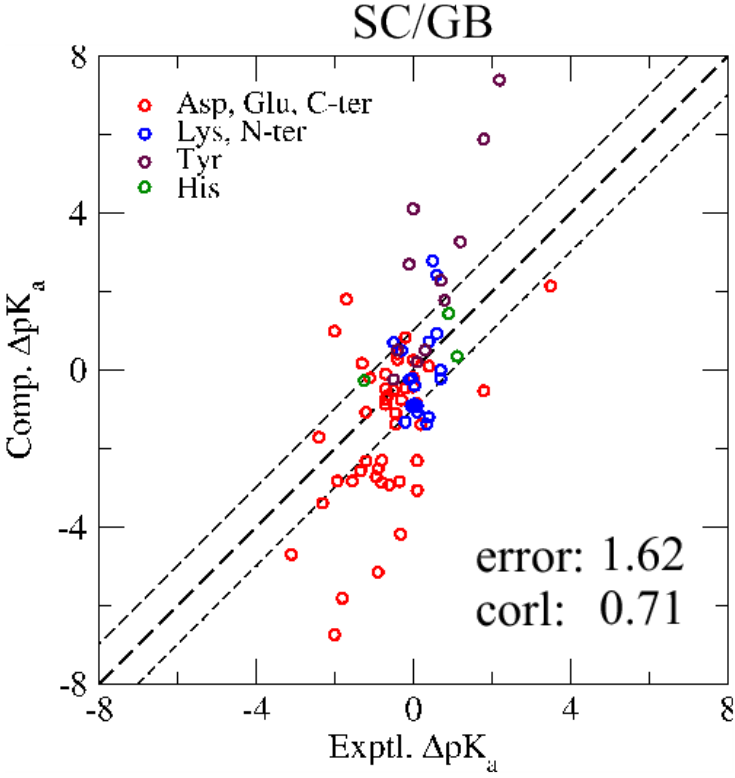
MC moves involve either one or two sidechains

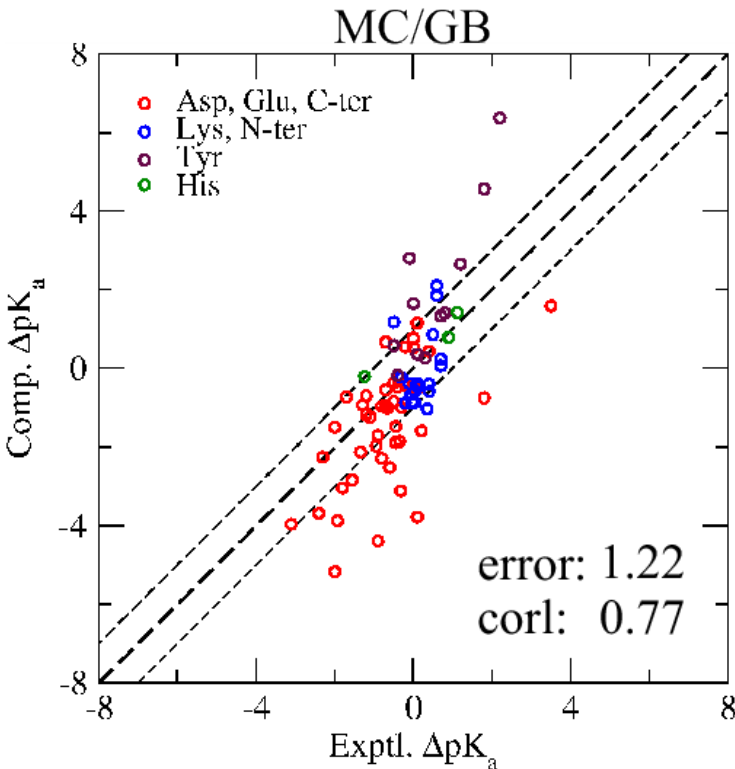
alternate conformational and protonation moves

~ ½ day per protein to scan whole pH range

Comparing to other approaches

Method	Error rms (max)	corr
Null	1.07 (3.5)	–
PROPKA	0.88 (4.4)	0.74
sc-PB	2.34 (10.8)	0.67
sc-GB	1.62 (4.3)	0.71
mc-GB	**1.22 (3.9)**	**0.77**

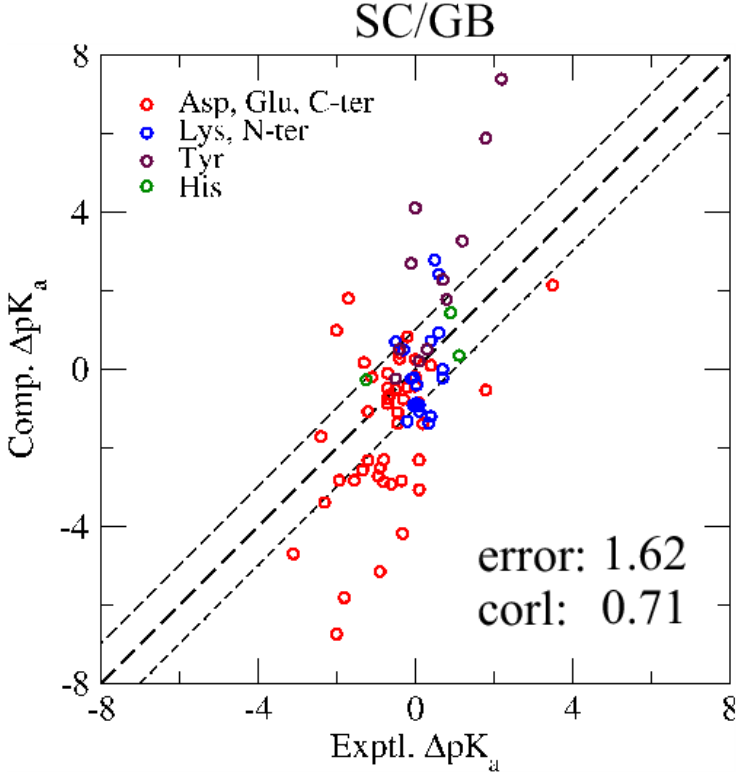




Comparing to other approaches

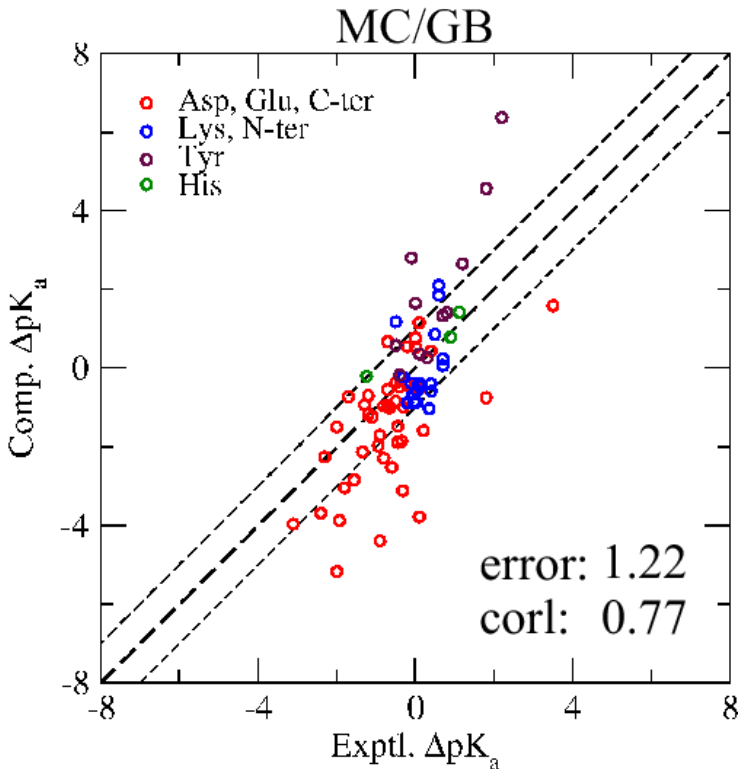
Method	Error rms (max)	corr
Null	1.07 (3.5)	–
PROPKA	0.88 (4.4)	0.74
sc-PB	2.34 (10.8)	0.67
sc-GB	1.62 (4.3)	0.71
mc-GB	**1.22 (3.9)**	**0.77**

27	ASP	1.32
15	GLU	
11	TYR	1.91
3	HIS	0.62
17	LYS	0.91
4	Cter	
1	Nter	



Compare groups with large or small pKa shifts

Exp shift	Number of groups	Method MC-GB	SC-GB	SC-PB	PROPKA
|pKa| < 1	57	1.22 (3.9)	1.52 (4.3)	2.06 (9.7)	0.86 (4.4)
|pKa| > 1	21	1.22 (2.8)	1.85 (4.1)	2.96 (10.8)	0.95 (2.3)
|pKa| > 2	8	0.93 (1.9)	1.47 (3.5)	1.91 (4.4)	0.85 (2.3)
All pKa's	78	1.22 (3.9)	1.62 (4.3)	2.34 (10.8)	0.88 (4.4)



Conclusions

Dielectric continuum picture is a surprisingly good first approximation

Can be obtained by field theory + a mean field approximation
(Orland and collaborators)

Care must be taken to choose a physically sensible implementation

Compare to constant pH MD (Brooks, Case, Baptista, Hunenberger, ...)

“Residue pairwise” GB is highly efficient yet it remains essentially exact

Applicable to redox changes and sidechain mutations: underway

Include protein-solvent dispersion interactions: analytical, GB-like model

Ligand binding; significant new difficulties