Laboratory of Computational Biophysical Chemistry

Electrostatic interactions *in* and *between* biomolecules:

From fundamentals concepts to applications

Part 2

BIOINFORMÁTICA USP

STAMiNA Global Network

Fernando Luís BARROSO da Silva

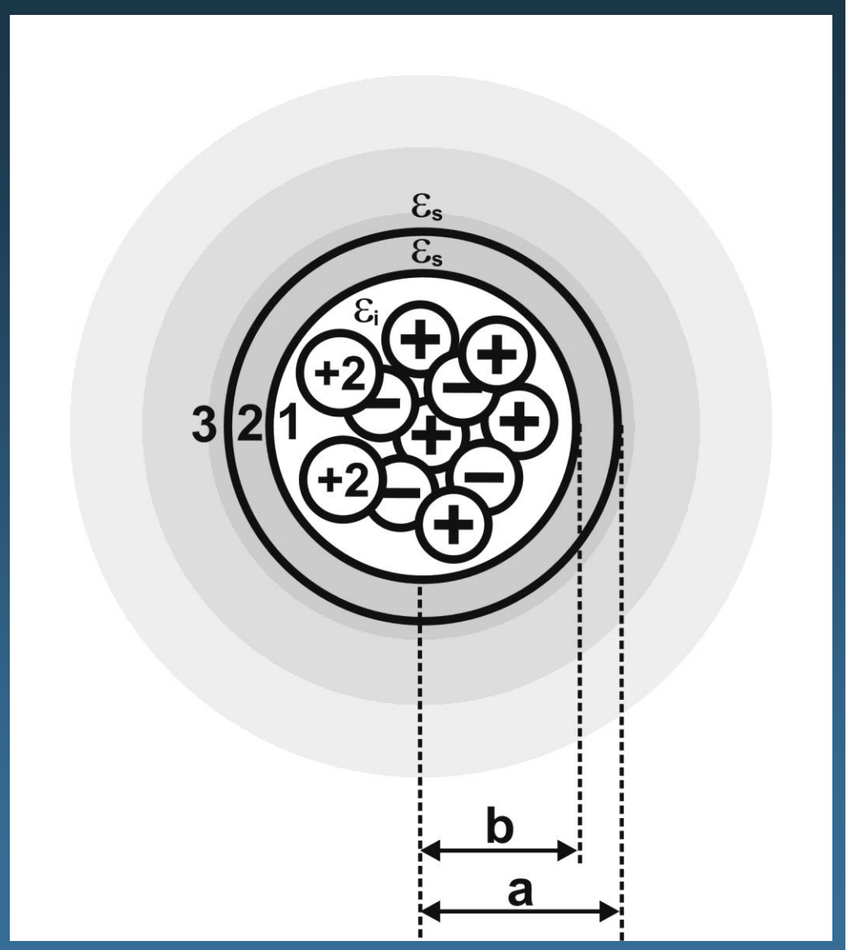
Department of Biomolecular Sciences, FCFRP/USP, Brazil

flbarroso@usp.br

March 14, 2020

63

Tanford-Kirkwood model



64

Fern

OM PROTEINSTOFFERNES IONISATION

AF

K. LINDERSTRØM-LANG.

Det er en kendt Sag, at Adsorptionsteorien, som har bidraget saa meget til at uddybe og udvide vort Kendskab til Stoffernes indbyrdes Vekselvirkning og til de Forhold, som betinger Fasernes Koeksistens, efterhaanden er taget i Brug ved Behandlingen af saa forskelligartede Problemer, at det ikke har kunnet undgaas, at den har draget Spørgsmaal ind under sig, som burde være taget op paa bredere Basis, — jeg sigter her til Proteinstoffernes Forhold.

Overalt, hvor Adsorptionsteorien bringes i Anvendelse, benyttes Formlen: »Koncentrationsændringer i Grænseoverfladen« til Beskrivelse af de Associationsfænomener, der optræder, naar man suspenderer Partikler af et Medium i et andet, men man kan ikke se bort fra, at en saadan Betragtningsmaade rummer en Ensidighed, der, saa vidt jeg kan se, stammer fra det empirisk-termodynamiske Udspring, Adsorptionsteorien har. Nogen fysisk Mening har denne Synsmaade først, naar Partiklerne er saa store, at de hver for sig staar i Faseligevægt med Dispersionsmidlet og har en Overflade i egentlig Forstand. Dette er imidlertid næppe Tilfældet ved Proteinstofferne; sættes nemlig efter S. P. L. Sørensen[1]) Ægalbuminpartiklernes Molekulstørrelse lig 35000, kommer hver Partikel til at indeholde ca. 3000 Atomer, og den formentlige Overflade kun ca. 1000, hvilket for det første udgør en Tredjedel af alle Atomerne i Partiklen og for det andet naturligvis er et for ringe Antal til at danne Overflade i almindelig Forstand.

Imidlertid er det ofte ikke helt let at forstaa den skarpe

[1]) Proteinstudier, p. 348. Carlsberg Lab. Med. **12** (1917).

JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

(Registered in U. S. Patent Office) (© Copyright, 1957, by the American Chemical Society)

VOLUME 79 OCTOBER 22, 1957 NUMBER 20

PHYSICAL AND INORGANIC CHEMISTRY

[CONTRIBUTION NO. 1415 FROM STERLING CHEMISTRY LABORATORY, YALE UNIVERSITY]

Theory of Protein Titration Curves. I. General Equations for Impenetrable Spheres

BY CHARLES TANFORD[1] AND JOHN G. KIRKWOOD

RECEIVED MAY 8, 1957

For many years the theory of titration curves of impenetrable proteins has been based on a model which represents the protein molecule as a sphere with a continuous and uniform distribution of charge on its surface. In this paper this model is replaced by a more realistic one in which the charges are taken to be discrete unit charges located at fixed positions. General equations are obtained which express the titration curve as a function of the locations of ionizable sites and of their intrinsic properties. It is concluded that the intrinsic properties may themselves be quite sensitive to the location of the dissociable site with respect to the surface of the protein molecule.

Introduction

Hydrogen ion titration curves of proteins can be obtained experimentally with considerable accuracy. Their most prominent feature is a strong electrostatic interaction which results in the fact that the titration curve of a protein containing acidic and basic side chains in any given proportions is considerably flatter than the titration curve of a mixture of corresponding simple acids and bases in the same proportions

ment been replaced by a single kind of ion with continuously variable charge, $\bar{Z}$.

With these assumptions one obtains the result that the titration curve of a protein is a superposition of the curves for the individual types of groups: the fraction α of dissociated groups of any type being given by the relation

$$\text{pH} - \log \frac{\alpha}{1-\alpha} = \text{p}K_{\text{int}} - \frac{1}{2.303kT}\frac{\partial \mathbf{W}}{\partial Z} \qquad (1)$$

The Tanford-Kirkwood model

$$G^{el} = \frac{1}{2}\sum_{i=1}^{N} q_i \phi(r_i)$$

$$G^{el} = \frac{1}{2}\sum_{i=1}^{N} (ez_i) \phi(r_i)$$

$$G^{el} = \frac{e^2}{8\pi\varepsilon_0}\sum_{i=1}^{m}\sum_{j=1}^{m} z_i z_j \left(A_{ij} - B_{ij} - C_{ij}\right)$$

salt

solvent

Coulombic interactions

$$G^{el} = \frac{e^2}{8\pi\varepsilon_0} \sum_{i=1}^{m} \sum_{j=1}^{m} z_i z_j \left(A_{ij} - B_{ij} - C_{ij} \right)$$

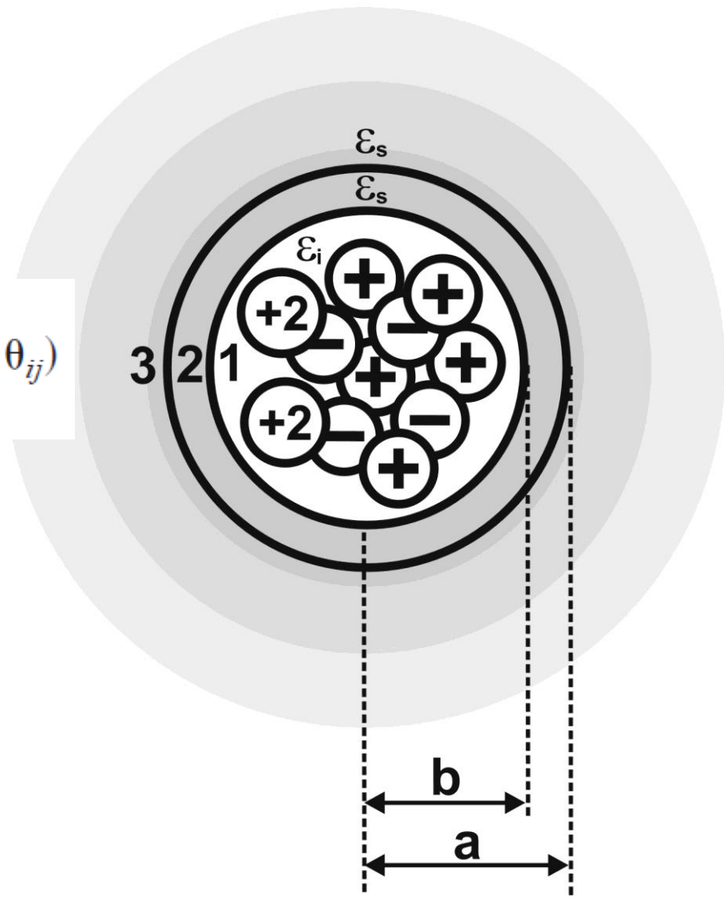
salt

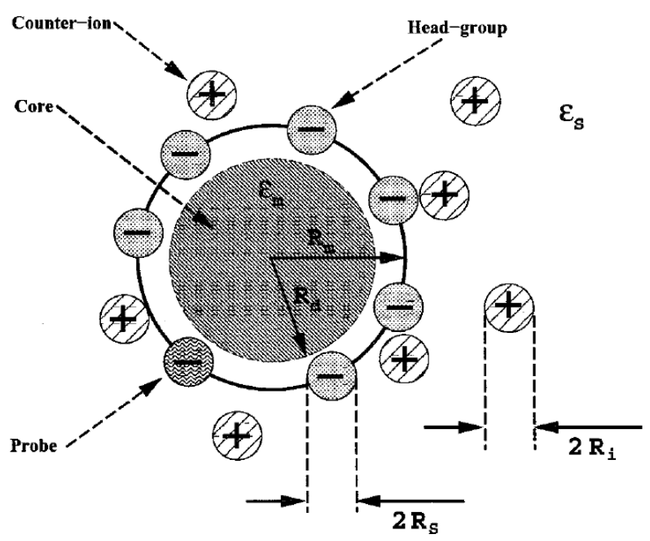
solvent with ≠ ε´s

Coulombic interactions

TK

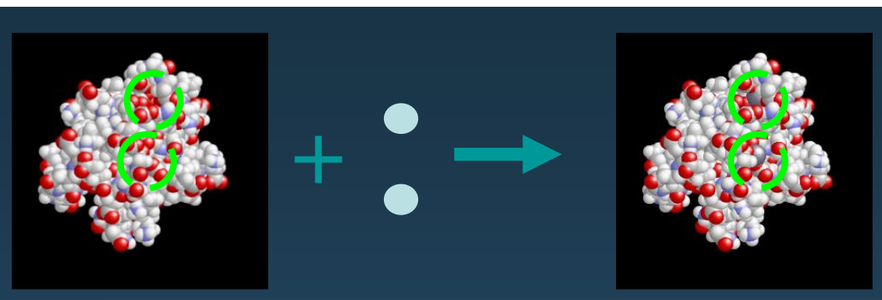
$$B_{ij} = \frac{1}{\epsilon_p R_d} \sum_{n=0}^{\infty} \frac{(\epsilon_s - \epsilon_p)}{(\epsilon_s + \epsilon_p n \,/\, (n+1))} \left(\frac{r_i r_j}{R_d^{\,2}} \right)^n P_n(\cos\,\theta_{ij})$$





Fernando Barroso, FCFRP/USP
SAIFR - March 9-15, 2020

67



$$M + L \xrightarrow{K} ML$$

$$\Delta G^{ele} = \Delta G(ML) - \Delta G(M) - \Delta G(L)$$

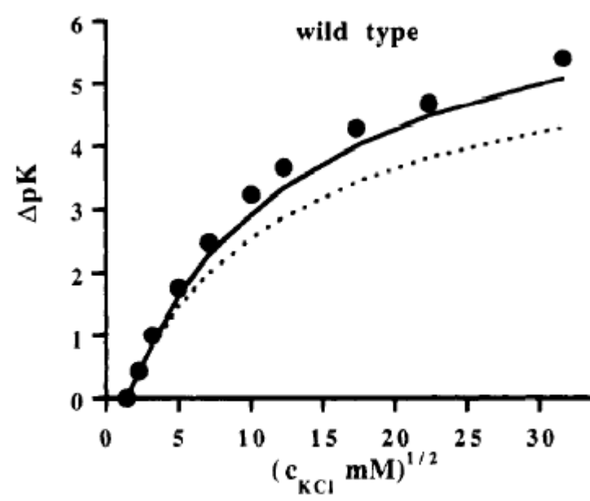


FIGURE 2: Experimental versus simulated ΔpK for wild-type calbindin D_{9k} in KCl solution. Symbols represent experimental data, and solid lines refer to MC simulations. The wild-type protein at 2 mM KCl has been taken as the reference. The dashed curve shows the results from the TK model.

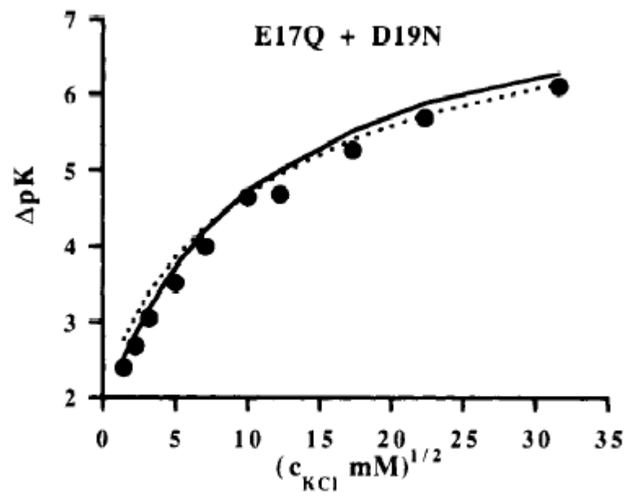


FIGURE 4: Experimental versus simulated ΔpK for the E17Q + D19N mutant of calbindin D_{9k} KCl solution. Symbols are as in Figure 2.

Biochemistry, Vol. 33, No. 47, 1994 14173

Protein Electrostatics
Prof. *Fernando Luís Barroso da Silva* (fernando@fcfrp.usp.br)

B.P.C. – DCBM/FCFRP - USP
SAIFR - March 9-15, 2020

68

TK

A critical investigation of the Tanford-Kirkwood scheme by means of Monte Carlo simulations

FERNANDO LUÍS B. DA SILVA,[1,3] BO JÖNSSON,[1] AND ROBERT PENFOLD[2]

[1]Theoretical Chemistry, Chemical Centre, Lund University, S-221 00 Lund, Sweden
[2]Institute of Food Research, Norwich Research Park, Colney, Norwich NR4 7UA, United Kindgom
[3]Group of Biomolecular Physics, Department of Physics, Faculty of Science, UNESP/Bauru, 17033-360 Bauru, São Paulo, Brazil

Protein Science (2001), 10:1415–1425

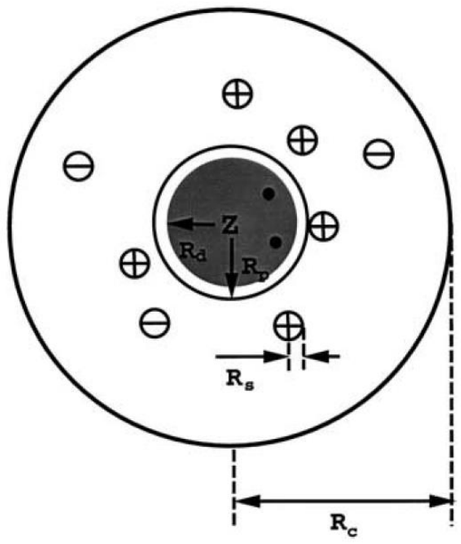
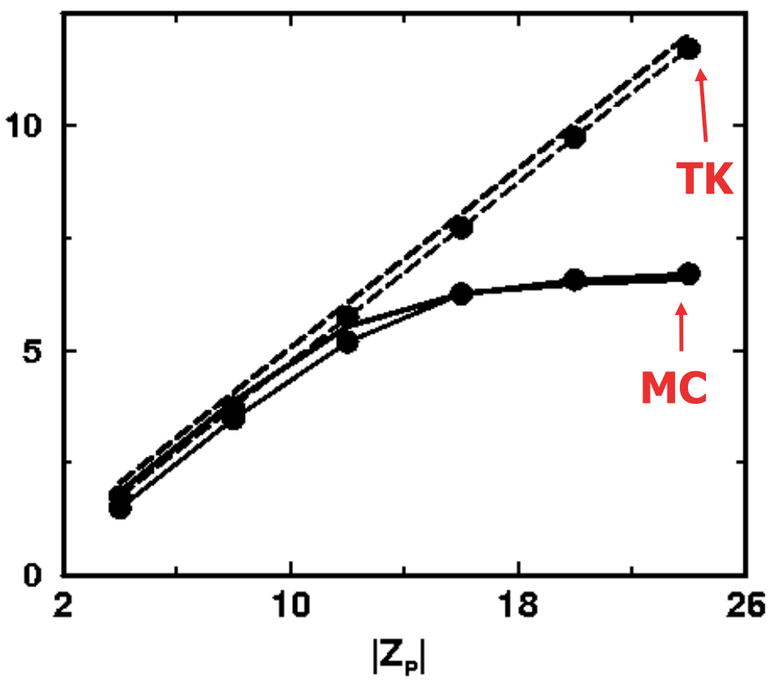
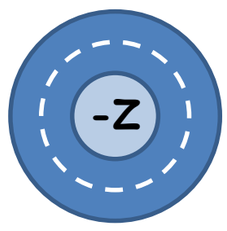


Fig. 1. Schematic representation of the model systems (see also Table 1). A spherical protein in a spherical cell with radii R_p and R_C, respectively. The protein interior with a low dielectric permittivity is shown as a shaded region of radius $R_d < R_p$. A central charge of valency Z and two binding sites marked as black dots.

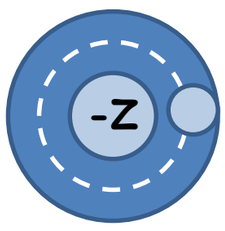


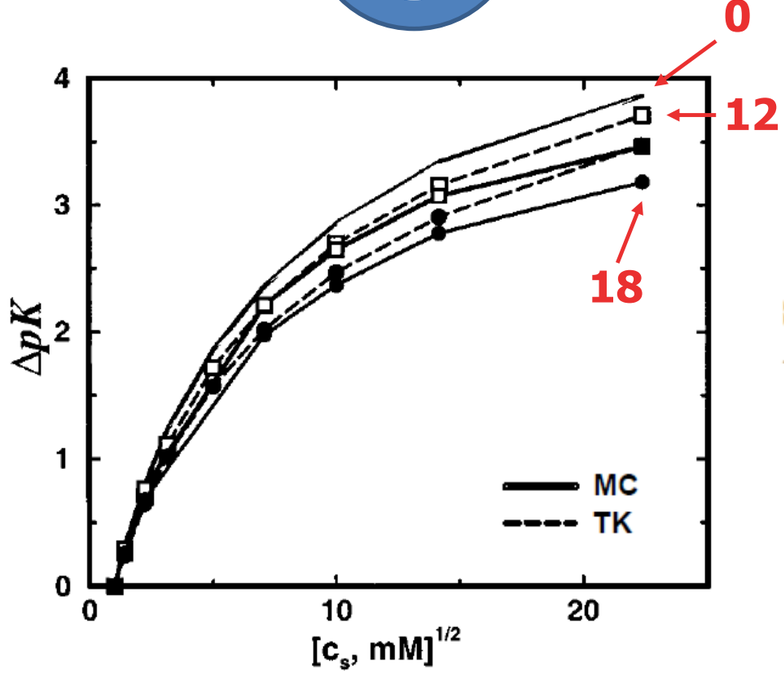
69

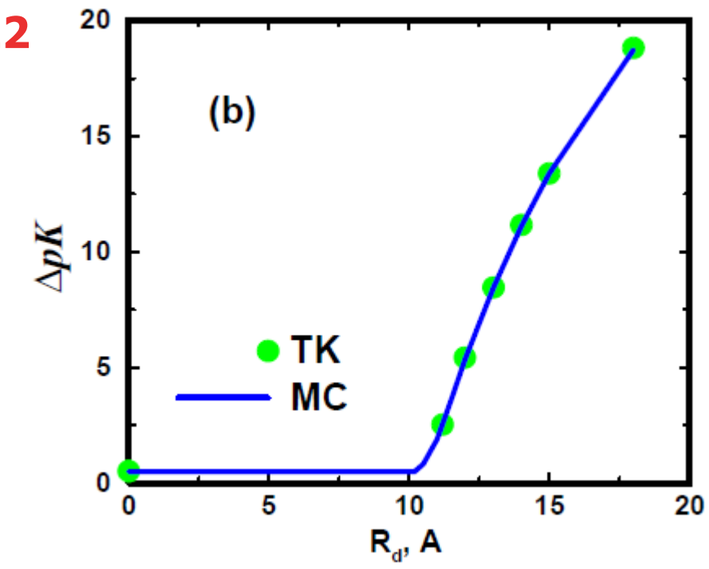
TK



0







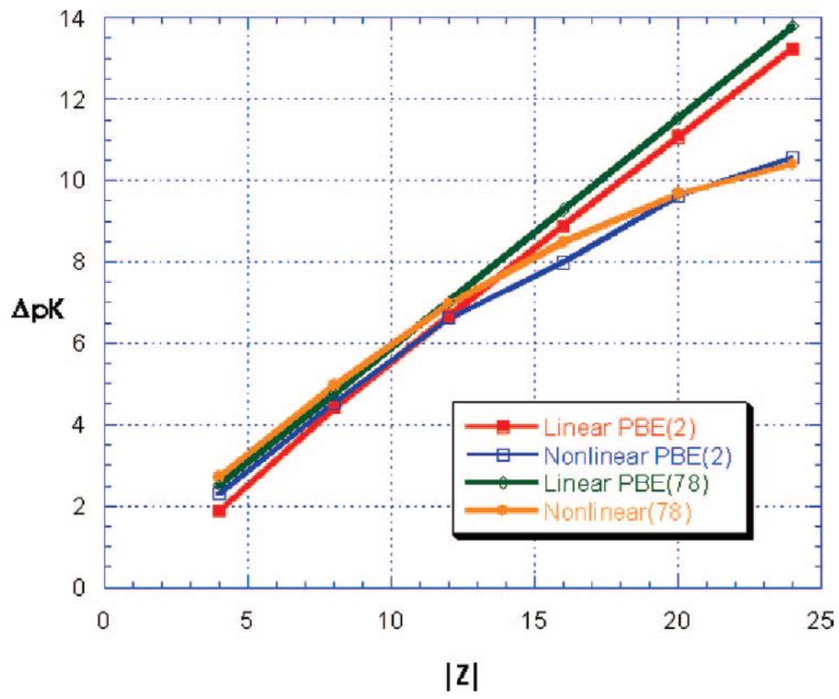
Protein Science (2001), 10:1415–1425

70

Protein−Ion Binding Process on Finite Macromolecular Concentration. A Poisson−Boltzmann and Monte Carlo Study

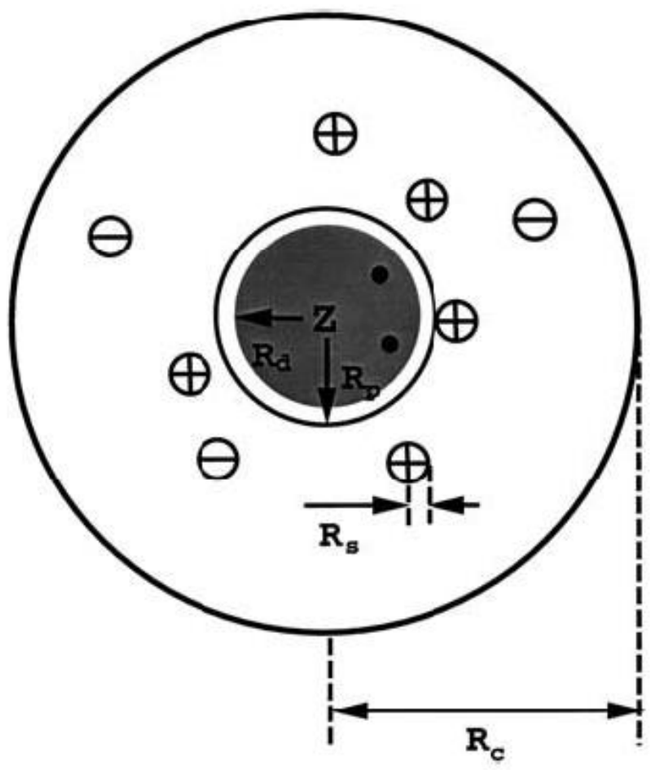
Sidney Jurado de Carvalho,[†] Márcia O. Fenley,[‡] and Fernando Luís Barroso da Silva*[,§]

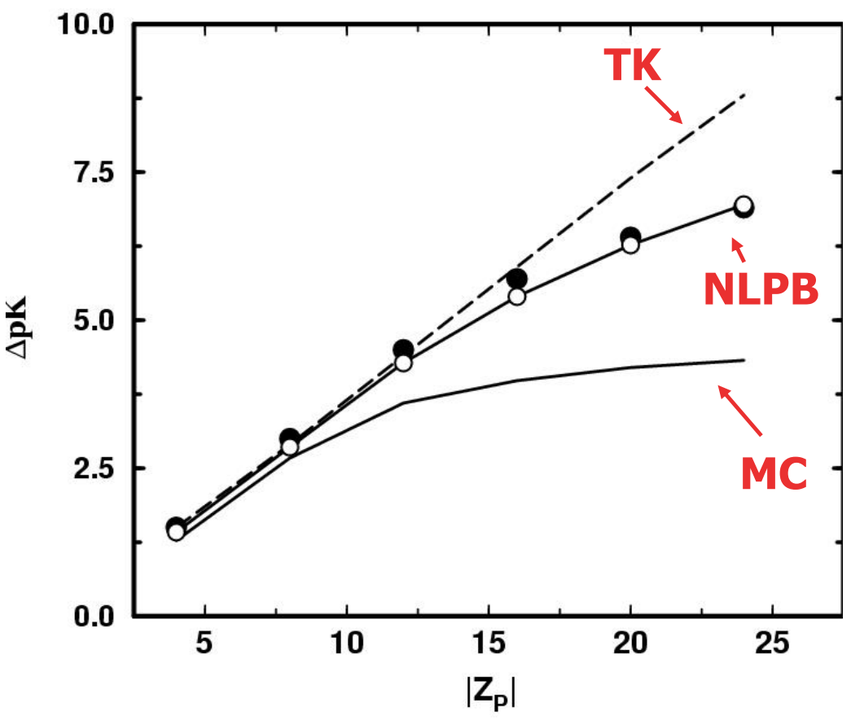
Department of Physics, IBILCE/Unesp, 15054-000 - Rua Cristovão Colombo, 2265, Jd. Nazareth, São José do Rio Preto − SP, Brazil, Department of Physics and Institute of Molecular Biophysics, Florida State University, Tallahassee, Florida 32306, and Department of Physics and Chemistry, Faculdade de Ciências Farmacêuticas de Ribeirão Preto (FCFRP) − USP, 14040−903 Avenida do café, s/no., Ribeirão Preto, SP, Brazil



71

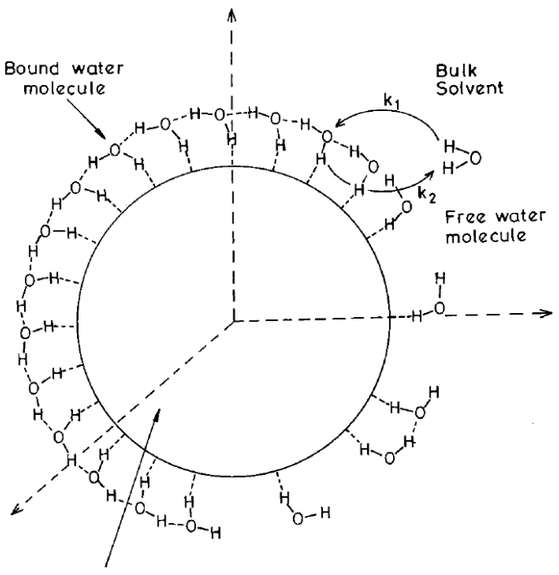
MC vs PB vs TK

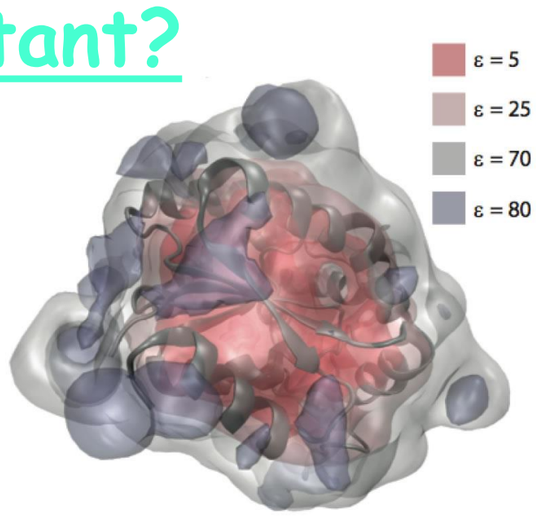


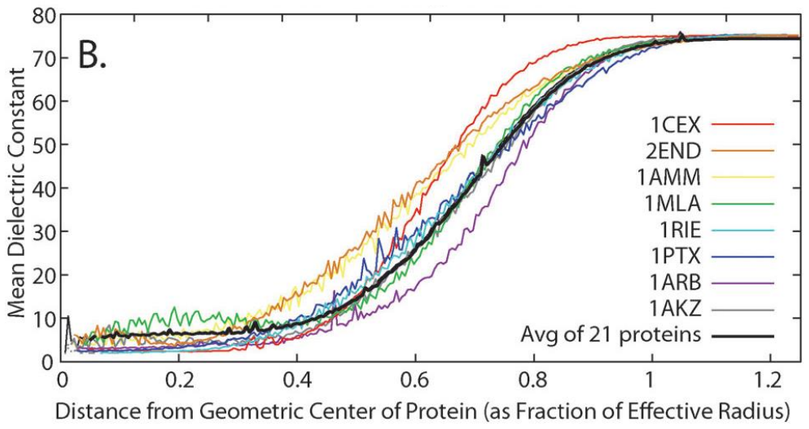


J. Phys. Chem. B **2008,** *112,* 16766–16776

Uniform dielectric constant?







• http://www.physics.ubc.ca/~steve/research/8-DielectricProperties.html

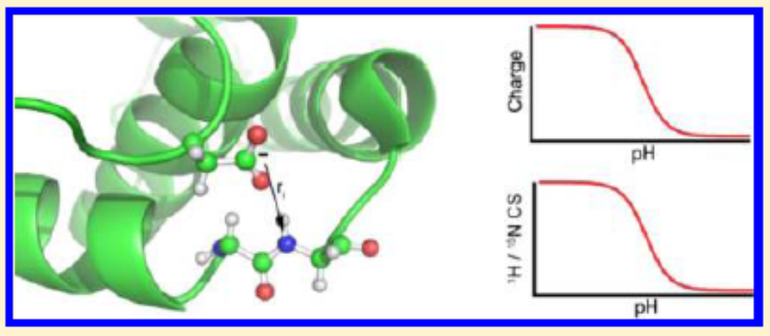
Uniform dielectric constant?

Protein Dielectric Constants Determined from NMR Chemical Shift Perturbations

Predrag Kukic,[†,∥] Damien Farrell,[†] Lawrence P. McIntosh,[‡] Bertrand García-Moreno E.,[§] Kristine Steen Jensen,[⊥] Zigmantas Toleikis,[⊥] Kaare Teilum,[⊥] and Jens Erik Nielsen[†,*,▽]

J. Am. Chem. Soc. 2013, 135, 16968−16976

ABSTRACT: Understanding the connection between protein structure and function requires a quantitative understanding of electrostatic effects. Structure-based electrostatic calculations are essential for this purpose, but their use has been limited by a long-standing discussion on which value to use for the dielectric constants (ε_{eff} and ε_p) required in Coulombic and Poisson−Boltzmann models. The currently used values for ε_{eff} and ε_p are essentially empirical parameters calibrated against thermodynamic properties that are indirect measurements of protein electric fields. We determine optimal values for ε_{eff} and ε_p by measuring protein electric fields in solution using direct detection of NMR chemical shift perturbations (CSPs). We measured CSPs in 14 proteins to get a broad and general characterization of electric fields. Coulomb's law reproduces the measured CSPs optimally with a protein dielectric constant (ε_{eff}) from 3 to 13, with an optimal value across all proteins of 6.5. However, when the water−protein interface is treated with finite difference Poisson−Boltzmann calculations, the optimal protein dielectric constant (ε_p) ranged from 2 to 5 with an optimum of 3. It is striking how similar this value is to the dielectric constant of 2−4 measured for protein powders and how different it is from the ε_p of 6−20 used in models based on the Poisson−Boltzmann equation when calculating thermodynamic parameters. Because the value of $\varepsilon_p = 3$ is obtained by analysis of NMR chemical shift perturbations instead of thermodynamic parameters such as pK_a values, it is likely to describe only the electric field and thus represent a more general, intrinsic, and transferable ε_p common to most folded proteins.

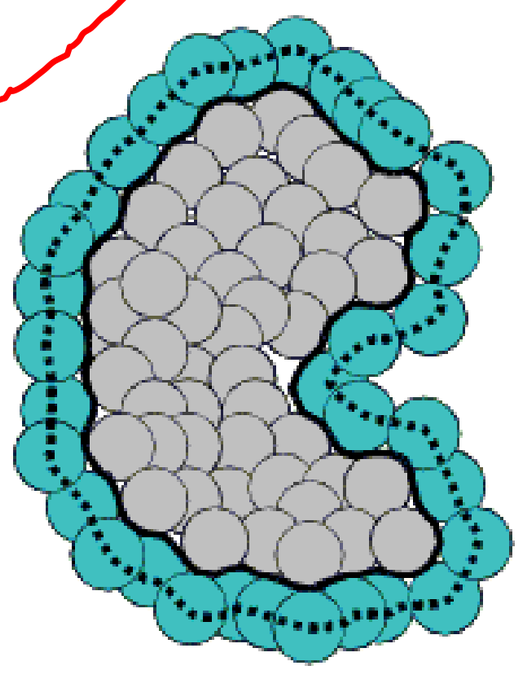


Molecular dielectric coefficients

- A heterogeneous molecule like a biomolecule shouldn't really be represented by a continuum dielectric…
- …however, that doesn't keep people from trying

Including me!

- Multiple dielectric values:
  - 1 = vacuum
  - 2-4 = atomic polarizability (solid)
  - 4-10 = some libration, minor sidechain rearrangement
  - 10-20 = significant internal rearrangement
- Multiple surface definitions:
  - van der Waals
  - Splines
  - Molecular surface



[Baker, 2012]

75

But,….

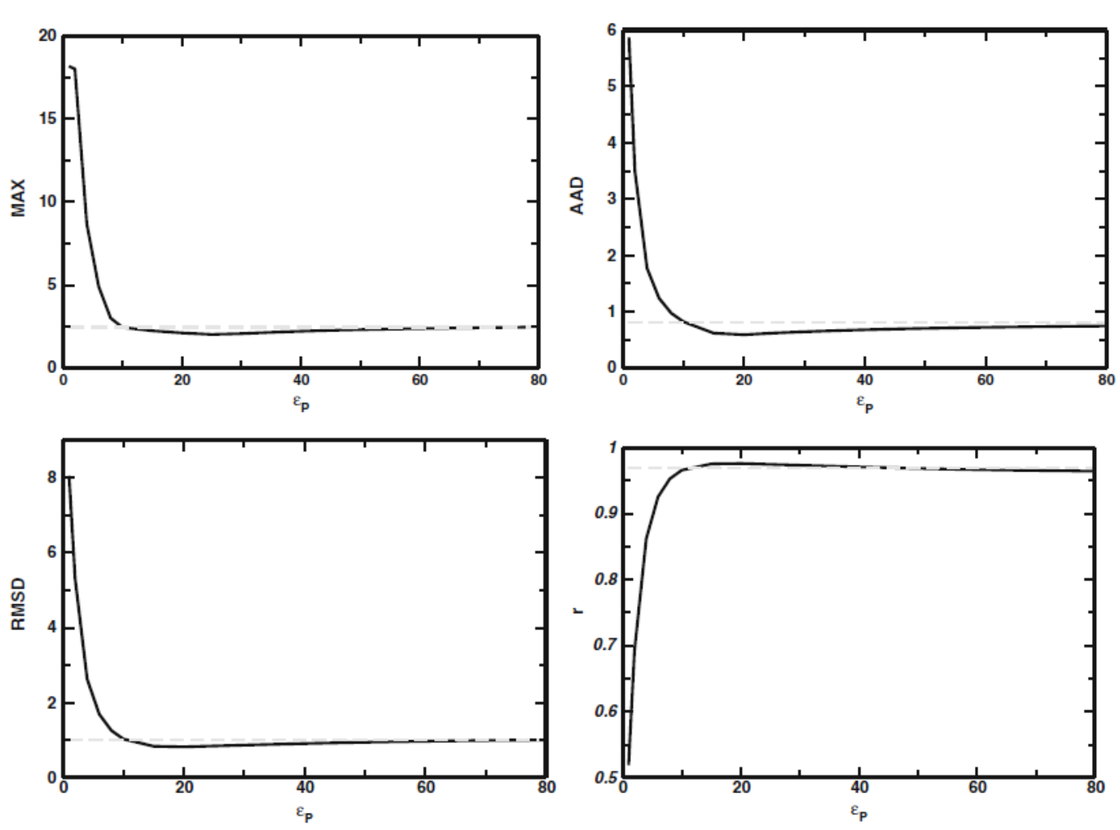


Fig. 5 Standard descriptors to measure the quality of calculated pK_a values by the PB equation as a function of the protein dielectric constant ϵ_p. The "NULL model" predictions are given by the dashed gray line. For these calculations, the pK_0s were taken from ref. Bashford and Karplus (1990) where simulation details are also described for $\epsilon_p = 4$. The PB calculations were carried out with the package MEAD 2.2.9. (Bashford 1997)

Biophys Rev (2017) 9:699–728

76

Basic ideas of the PB

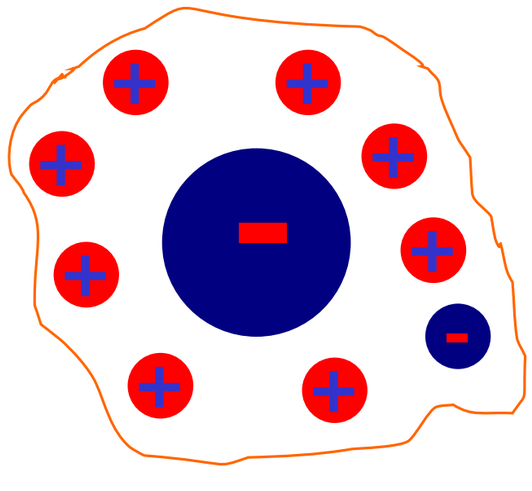
Fernando Barroso, FCFRP/USP
SAIFR – March 9-15, 2020

Particle properties Field properties

	Particle properties	Inter-relations	Field properties
Vector	Force $\mathbf{F} = \frac{1}{4\pi\epsilon_0}\frac{q_1 q_2}{\lvert r\rvert^2}\hat{\mathbf{r}}$	$\mathbf{F} = q\mathbf{E}$	Electric field $\mathbf{E} = \frac{1}{4\pi\epsilon_0}\frac{q}{\lvert r\rvert^2}\hat{\mathbf{r}}$
Inter-relations	$\mathbf{F} = -\nabla U$		$\mathbf{E} = -\nabla V$
Scalar	Potential energy $U = \frac{1}{4\pi\epsilon_0}\frac{q_1 q_2}{\lvert r\rvert}$	$U = qV$	Potential $V = \frac{1}{4\pi\epsilon_0}\frac{q}{\lvert r\rvert}$

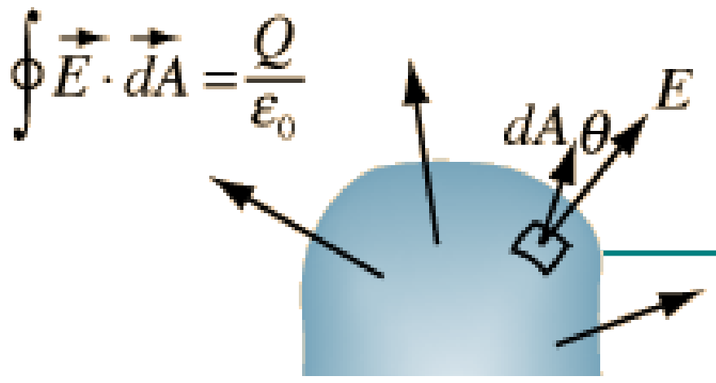
From the physics,

Gauss Law

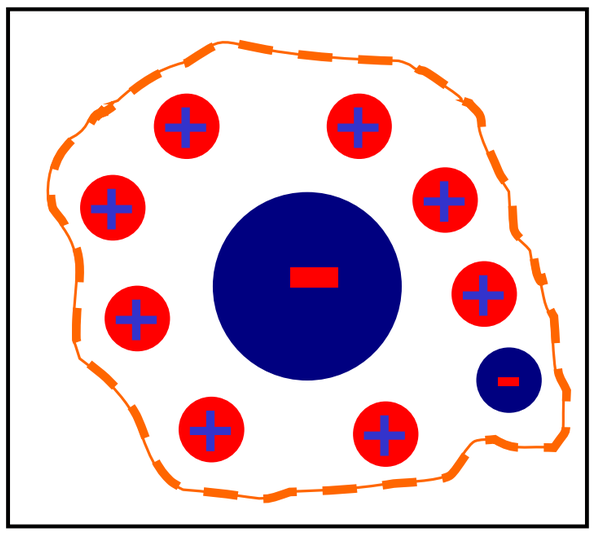


$$\Phi = \oint \vec{E}(\vec{x}).d\vec{s} = \frac{Q_{\text{inside}}}{\varepsilon_o}$$

$$\nabla.\vec{E} = \frac{\rho_e}{\varepsilon\varepsilon_0}$$



For a high number of particles



Gauss Law

$$\nabla.\vec{E} = \frac{\rho_e}{\varepsilon\varepsilon_0}$$

ρ_e = *charge/volume*

But,

$$\vec{E} = -\vec{\nabla}\phi$$

$$\vec{\nabla}.(-\vec{\nabla}\phi) = \frac{\rho_e}{\varepsilon\varepsilon_0}$$

Poisson equation

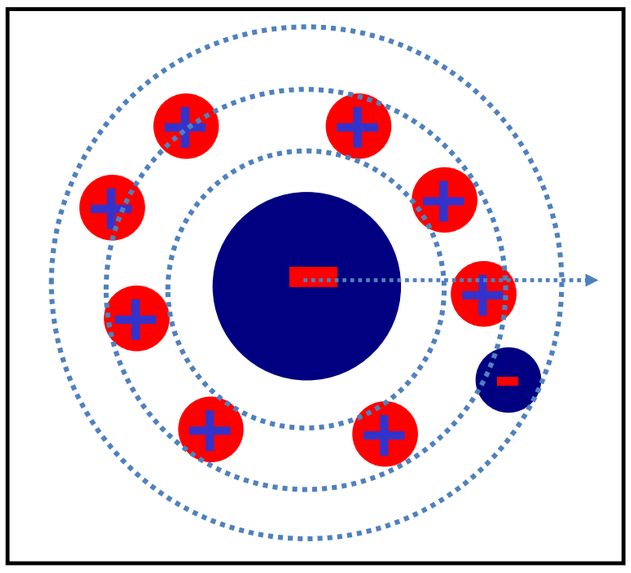
$$\nabla^2\phi = -\frac{\rho_e}{\varepsilon\varepsilon_0}$$

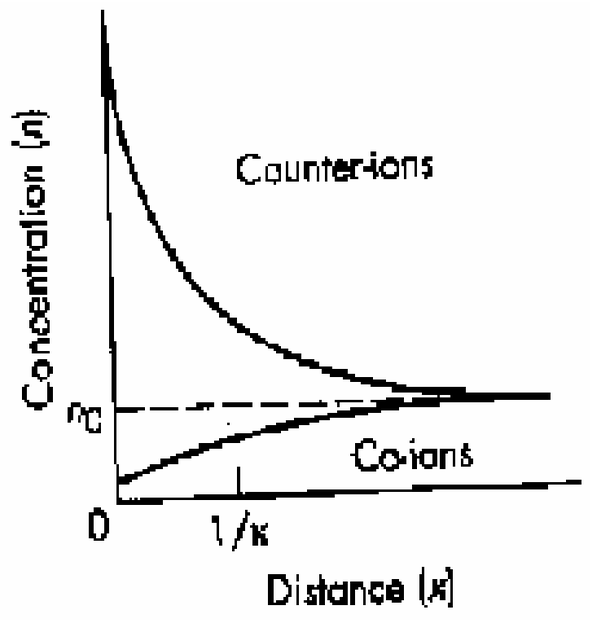
Poisson

$$\nabla^2\phi = -\frac{\rho_e}{\varepsilon\varepsilon_0}$$

Boltzmann

$$n_k = n_{0,k}\exp\left[\frac{-z_k e\psi_k}{k_B T}\right]$$





[Ilustração do livro do Shaw]

Poisson-Boltzmann equation

[Ilustração WikiPedia]

Poisson equation

$$\nabla^2\phi = -\frac{\rho_e}{\epsilon_0}$$

solvent →

Poisson equation

$$\nabla^2\phi = -\frac{\rho_e}{\epsilon_0\,\epsilon_s}$$

→

$$n_k = n_{0k}\exp\left[\frac{-z_k\,e\,\psi_k}{kT}\right]$$

$$\rho_e = z\,e\,(n_+ - n_-) \quad = \quad z\,e\,n_0\left(\exp\left[\frac{-z_+\,e\,\psi_+}{kT}\right] - \exp\left[\frac{-z_-\,e\,\psi_-}{kT}\right]\right)$$

Poisson-Boltzmann equation

[Ilustração WikiPedia]

$$\nabla^2\phi = -\frac{\rho_f}{\epsilon_0\,\epsilon_s} \overset{\psi=\phi}{\approx} \frac{2\,z\,e\,n_0}{\epsilon_0\,\epsilon_s}\,\sinh\left(\frac{z\,e\,\phi}{kT}\right)$$

non-linear equation

Analitical solution only for very simple system
(e.g. infinite charged planar surface – Gouy-Chapman)

$$\Psi(x) = \Psi_o e^{-\kappa x}$$

Debye-Hückel

[Ilustração WikiPedia]

Debye and Hückel introduced a *linearization* in the PB equation making it more tractable. In short, they expanded the exponential of the PB equation as

$$\exp(x) = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots$$

which for small values of x becomes

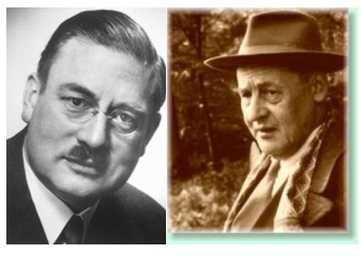
$$\exp(x) \approx 1 + x$$

For $x = (-ez_k\phi/kT)$, we obtain

$$\exp\left(\frac{-ez_k\phi}{kT}\right) = 1 - \frac{ez_k\phi}{kT}$$

$$n_k = n_{0k}\,\exp\left[\frac{-z_k\,e\,\psi_k}{kT}\right]$$

Debye-Hückel



[Ilustração WikiPedia]

Then, the PB equation becomes

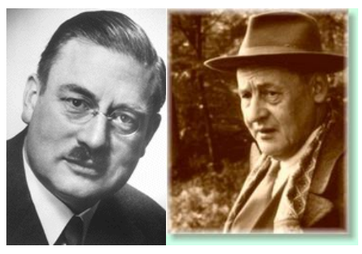
$$\nabla^2\phi = -\frac{1}{\epsilon_0\,\epsilon_s}\left(\sum_{k=1}^{N} e\,z_k\,n_{0k} \;-\; \sum_{k=1}^{N}\frac{(e\,z_k)^2\,n_{0k}\,\phi}{kT}\right)$$

The term $\sum_{k=1}^{N} e z_k n_{0k}$ must be zero to keep the electrolyte solution neutral. Therefore,

$$\nabla^2\phi = -\sum_{k=1}^{N}\frac{(e\,z_k)^2\,n_{0k}}{\epsilon_0\,\epsilon_s}\,\frac{\phi}{kT}$$

which is the *linearized Poisson–Boltzmann equation* (LPB) also known as the *Debye–Hückel* equation.

Debye-Hückel: some useful definitions



[Ilustração WikiPedia]

(a) The ionic strength, I (mol/l):

$$I = \frac{1}{2}\sum_{k=1}^{N}\frac{n_{0k}\,(z_k)^2}{1000\,N_a}$$

(Note that the factor $1000\,N_a$ [15] is due to the units of I).

(b) The Debye–Hückel screening length, κ^{-1} (m):

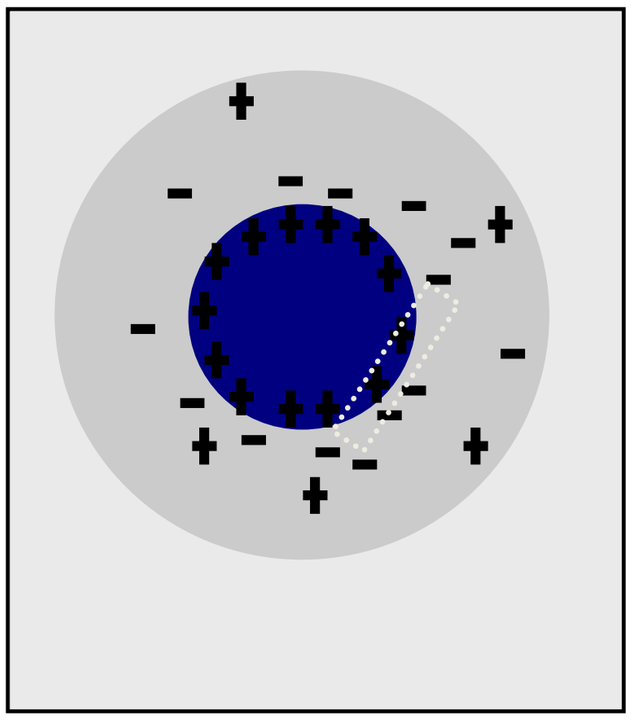
[Ilustração de fonte desconhecida]

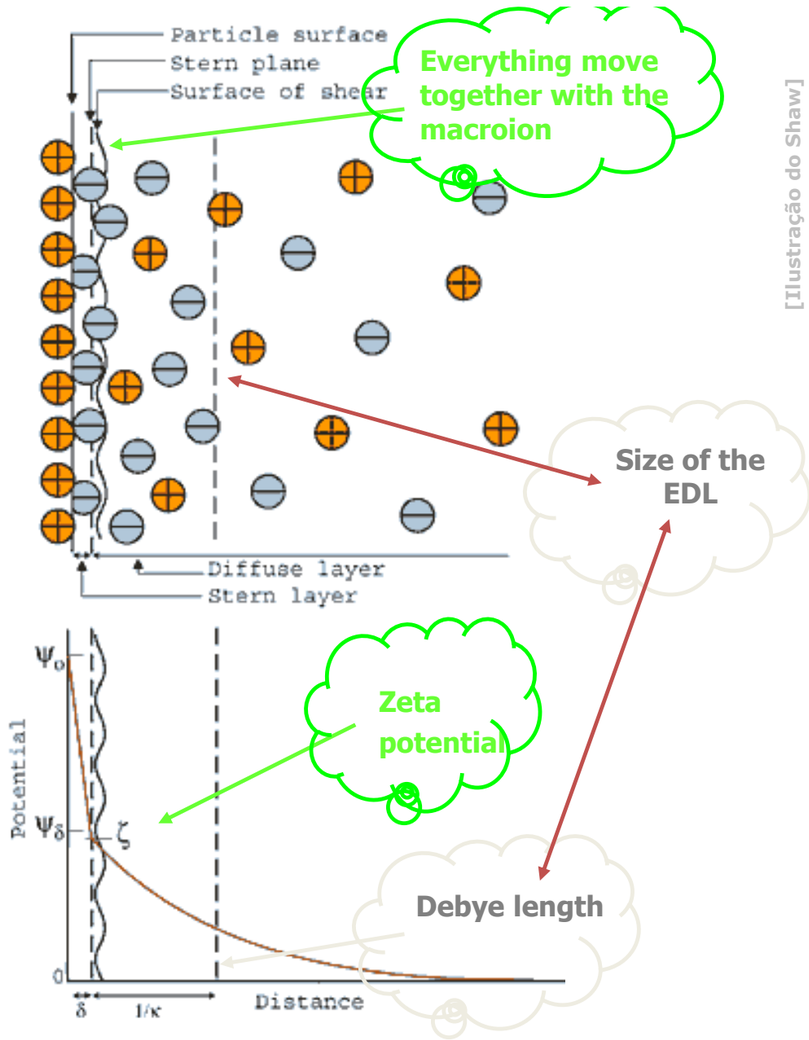
$$\kappa^{-1} = \left[\frac{\epsilon_0\,\epsilon_s\,kT}{1000\,N_a\,e^2\,2\,I}\right]^{\frac{1}{2}} \quad = \quad \left[\frac{\epsilon_0\,\epsilon_s\,kT}{e^2\,\sum_{k=1}^{N} n_{0k}\,(z_k)^2}\right]^{\frac{1}{2}}$$

(c) The Bjerrum length λ_b (m):

$$\lambda_b = \frac{z_i\,z_j\,e^2}{kT\,4\pi\epsilon_0\,\epsilon_s}$$

Electrical double layer





[Ilustração do Shaw]

Current methods for solution of the Poisson-Boltzmann equation

- Finite difference (box volume)
  - Multigrid solvers
  - Other
- Boundary element
  - Smaller problem domain
  - Easier interaction evaluation
- Finite element
  - *A priori* error estimation and discretization
  - Two-level solver
- Others

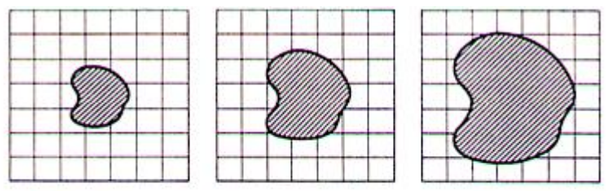
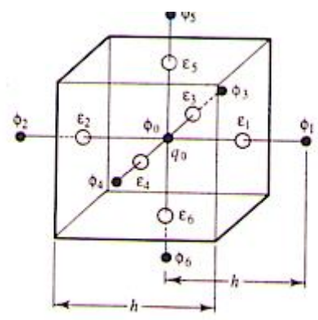


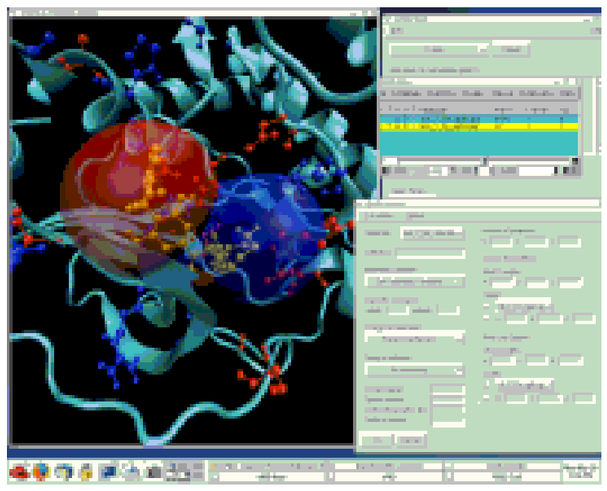
Fig. 9.21 Focussing can improve the accuracy of finite difference Poisson–Boltzmann calculations.

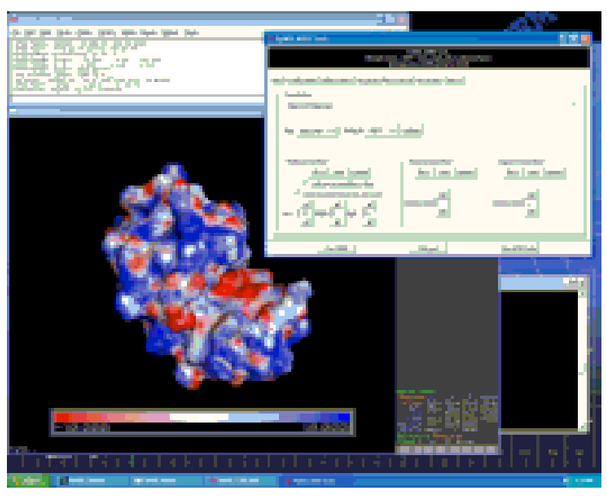


[Baker, 2012]

Implicit solvent tools

- APBS (http://apbs.sf.net/)
  - PB electrostatics calculations
  - Freely available
  - Fast finite element (FEtk) and multigrid (PMG) solvers from Holst group (http://fetk.org)
  - Works with most popular visualization software (VMD, PMV, PyMOL)
    - Links with CHARMM, AMBER, TINKER*
- PDB2PQR (http://pdb2pqr.sf.net/)





[Baker, 2012]

Baker NA, et al *Proc Natl Acad Sci USA*, **98**, 10037, 2001;
*Schnieders MJ, et al. *J Chem Phys*, **126**, 124114, 2007.

89

biophysics.cs.vt.edu

H++

Virginia Tech

HOME
PROCESS A STRUCTURE
CONTACT US
FAQ
EXAMPLES
CREDITS

LOGIN:
User Name:
Password:

WELCOME

H++ is an automated system that computes pK values of ionizable groups in macromolecules and adds missing hydrogen atoms according to the specified pH of the environment. Given a (PDB) structure file on input, H++ outputs the completed structure in several common formats (PDB, PQR, AMBER inpcrd/prmtop) and provides a set of tools for analysis of electrostatic-related molecular properties. Why H++

If this is your first time using H++, please read the FAQ carefully.

H++ in more detail. Citations

Please report problems to the H++ team: ramu@vt.edu

NEWS:

7/2015 - Most common ions can now be retained for pK calculations and will be included in the output topology/parameter and coordinate files. See FAQ for details.

7/2015 - H++ now uses the latest version of AmberTools and Amber force field parameters: AmberTools 15 using ff14SB force field parameters for proteins. with OL3 modifications for RNA, and OL1+OL4 modifications for DNA. See FAQ for details.

90

Molecular simulation	Poisson–Boltzmann
• *exact* solution	• mean-field *approximation*
• numerically stable	• the non-linear form is not stable
• simpler protein models	• explicit ions are absent
• protein concentration is an input parameter	• protein at infinite dilution
• hard to simulate high salt and low protein concentrations	• breaks down when ionic correlations increase
• time consuming	• memory consuming

[FLBDS, 2000]

$$\Xi = 2\pi\, z_k^3 l_b^2 \sigma_S$$

Laboratory of Computational Biophysical Chemistry

The "breakdown" of the Poisson-Boltzmann

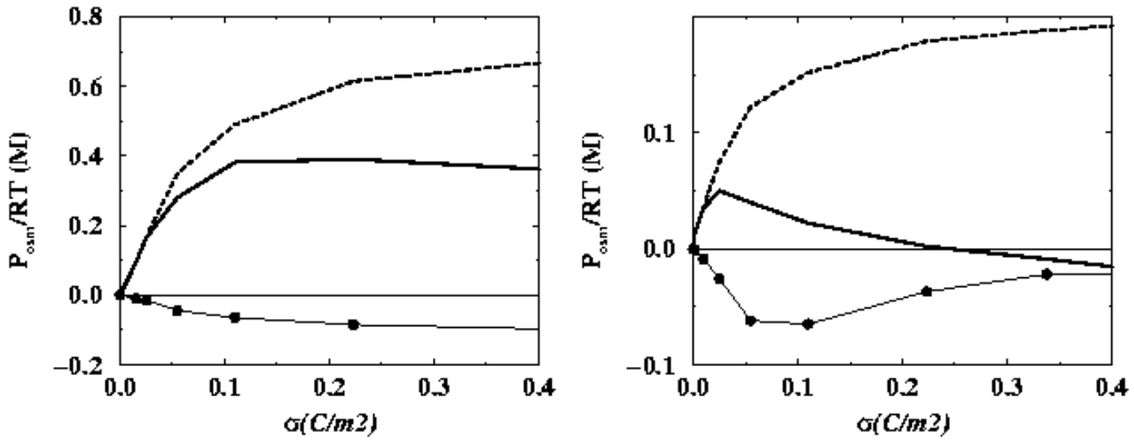


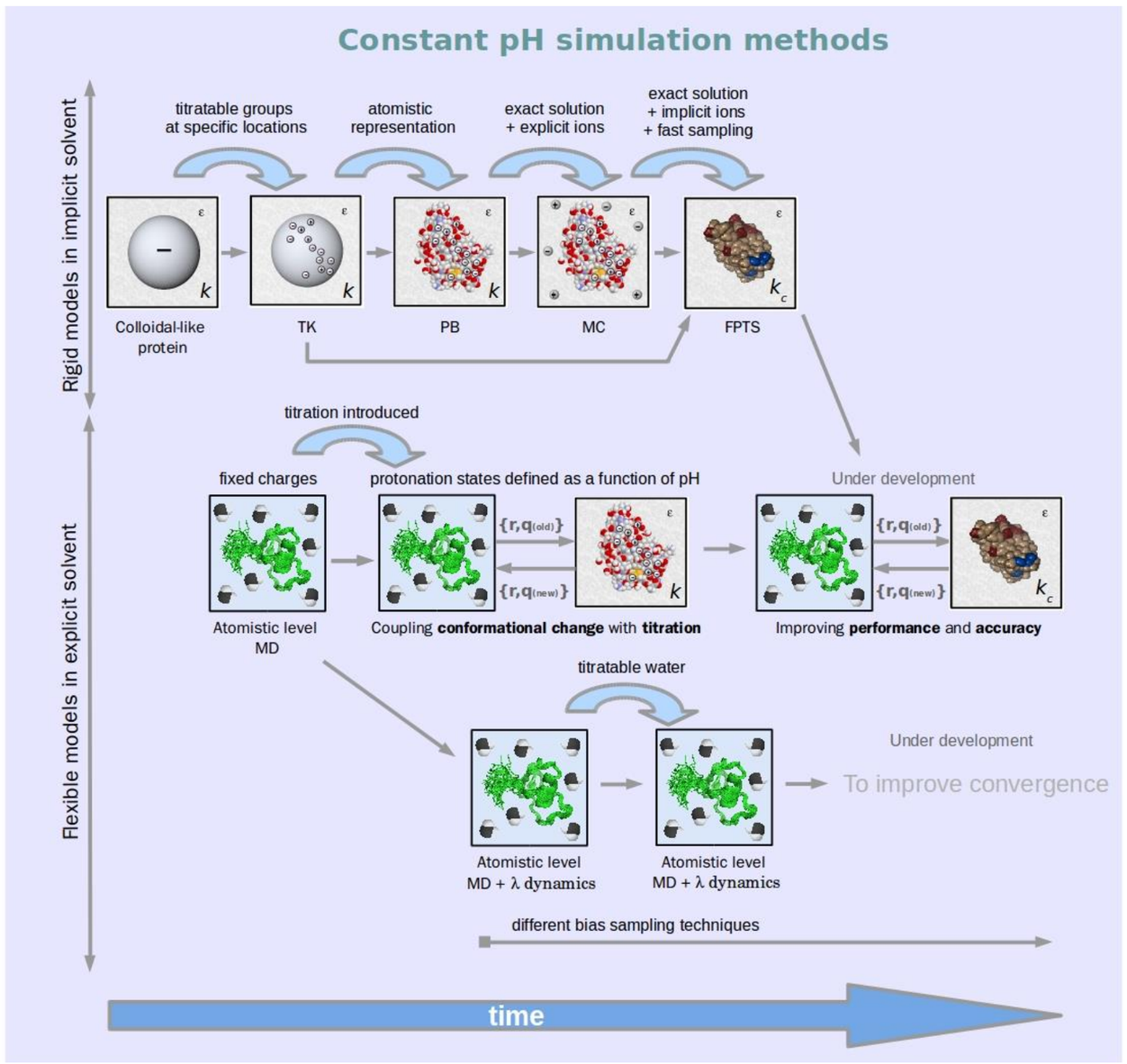
Figure 5: a) The osmotic pressure as a function of surface charge density for two planar double layers with neutralizing monovalent counterions. The solid line is from MC simulations and the dashed line from the PB equation. The thin line with symbols show the attractive contribution to the total pressure. b) The same as a) but with divalent counterions.

(L. Guldbrand *et al.*, *J. Chem. Phys.*, **80** (1984) 2221)

Laboratory of Computational Biophysical Chemistry

Other methods for pKa calculations

1. Molecular dynamics at constant pH
   (time-dependent protonation states)
2. Monte Carlo methods at constant pH
3. Empirical methods

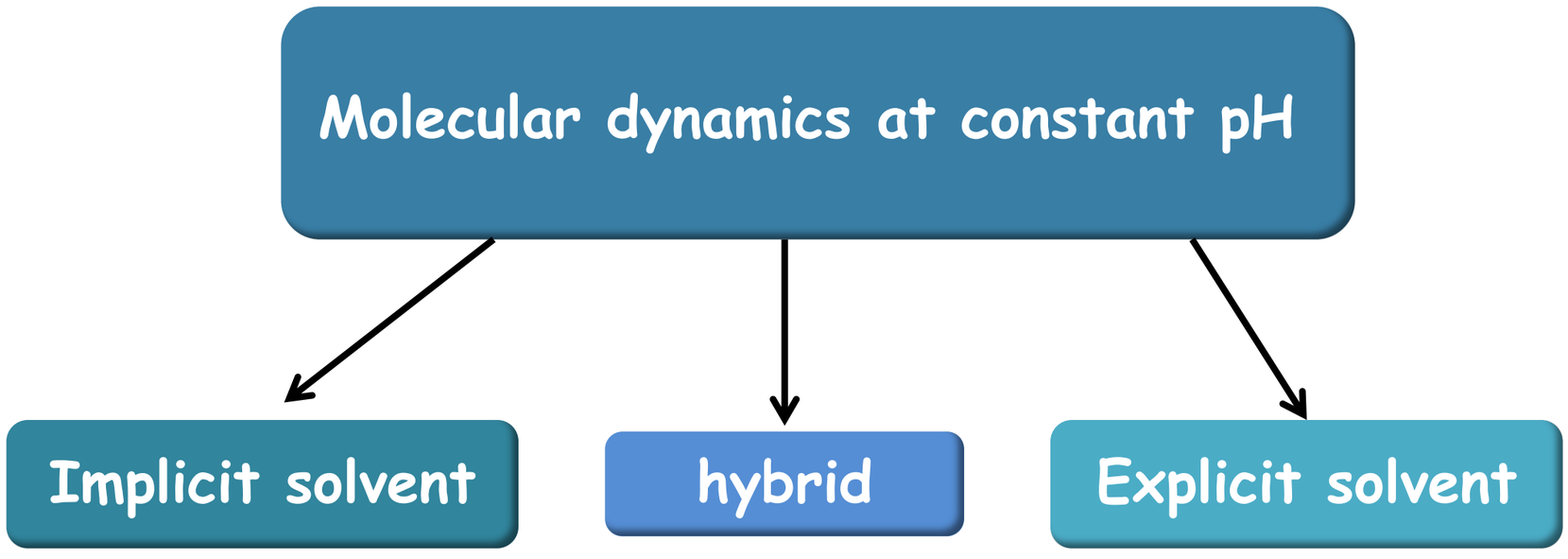


Laboratory of Computational Biophysical Chemistry

Constant-pH computational methods

Fernando Barroso, FCFRP/USP

Laboratory of Computational Biophysical Chemistry



Fernando Barroso, FCFRP/USP
SAIFR - March 9-15, 2020

Laboratory of Computational Biophysical Chemistry

PROTEINS: Structure, Function, and Genetics 27:523–544 (1997)

Simulation of Protein Conformational Freedom as a Function of pH: Constant-pH Molecular Dynamics Using Implicit Titration

António M. Baptista, Paulo J. Martel, and Steffen B. Petersen*
MR-Center, SINTEF UNIMED, N-7034 Trondheim, Norway

Laboratory of Computational Biophysical Chemistry

λ-dynamics: A new approach to free energy calculations

Xianjun Kong and Charles L. Brooks III

Citation: The Journal of Chemical Physics **105**, 2414 (1996); doi: 10.1063/1.472109

Protonation becomes the dynamics of a λ-particle!

$$m_\lambda \frac{d^2\lambda}{dt^2} = -\frac{\partial V(x,\lambda)}{\partial \lambda}$$

$$V(x,\lambda) = (1-\lambda)V^A(x) + \lambda V^B(x) + U(\lambda) + \Delta G_{chem}$$

$$\Delta G_{chem} = \lambda RT \ln(10)[\,pK_0 - pH\,] + \Delta G_{MM}^{corr}(\lambda)$$

Laboratory of Computational Biophysical Chemistry

Constant pH Molecular Dynamics in Explicit Solvent with Enveloping Distribution Sampling and Hamiltonian Exchange

Juyong Lee,*,† Benjamin T. Miller,† Ana Damjanović,†,‡ and Bernard R. Brooks†

J. Chem. Theory Comput. 2014, 10, 2738–2750

Charge-Neutral Constant pH Molecular Dynamics Simulations Using a Parsimonious Proton Buffer

Serena Donnini,† R. Thomas Ullmann,‡ Gerrit Groenhof,*,§ and Helmut Grubmüller*,‡

J. Chem. Theory Comput. 2016, 12, 1040–1051

Constant-pH Hybrid Nonequilibrium Molecular Dynamics–Monte Carlo Simulation Method

Yunjie Chen and Benoît Roux*

J. Chem. Theory Comput. 2015, 11, 3919–3931

Results are promising!

Biophysical Journal Volume 105 August 2013 L15 L17

Introducing Titratable Water to All-Atom Molecular Dynamics at Constant pH

Wei Chen, Jason A. Wallace, Zhi Yue, and Jana K. Shen*
Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland, Baltimore, Maryland

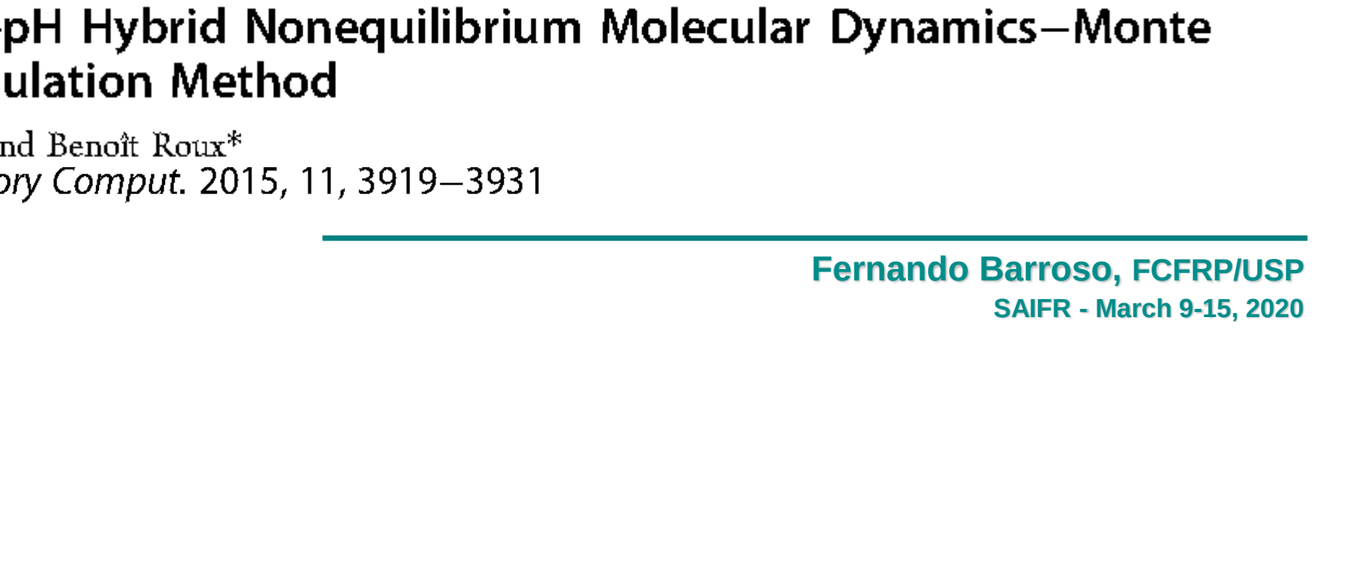
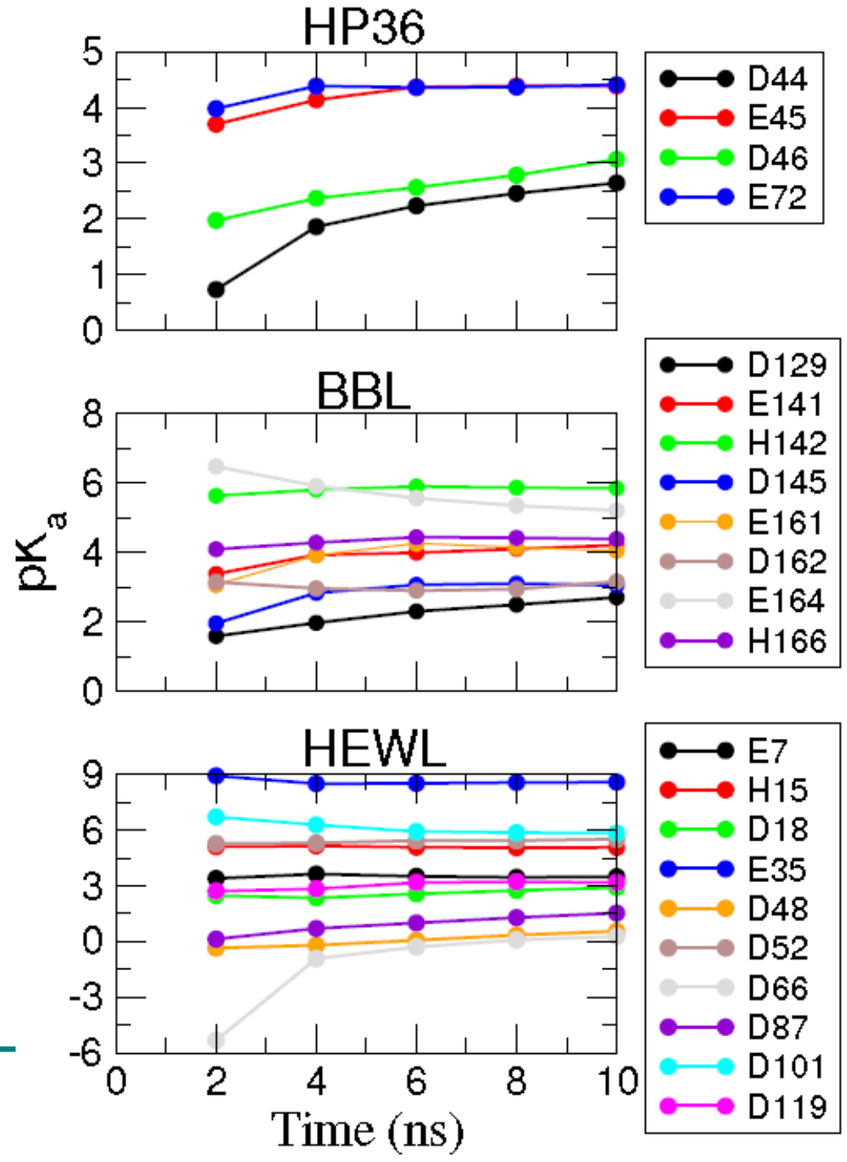


Figure S4: Comparison between calculated and experimental pK_a's for different proteins (*left*) and residues (*right*). Calculated pK_a's are from the entire 10-ns simulations. The solid line represents the linear regression to the data (slope=1.20 and R^2=0.76).

But,
convergence is still a problem!



101

Laboratory of Computational Biophysical Chemistry

But, convergence is still a problem!

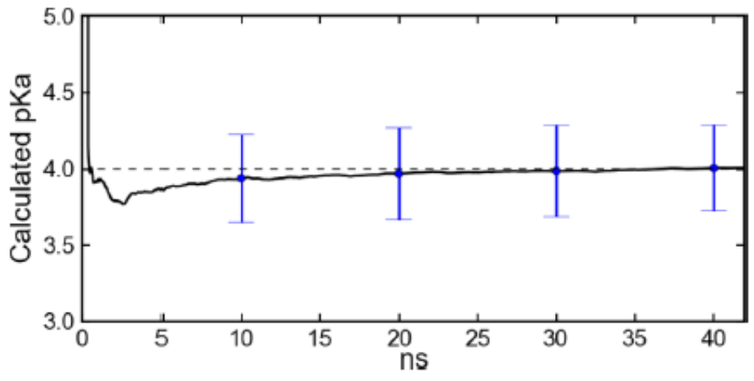


Figure 3. Evolution of the calculated pK_a for the aspartic dipeptide in water. The flat dashed line is the experimental pK_a of the system. The black line is calculated from the time average of all 15 trajectories, and the error bars show the standard deviation. Each round of hybrid neMD−MC simulation uses 0.2 ps for the equilibrium MD and 10 ps for the neMD switch. The time axis reflects the total computational cost of the hybrid simulation by including the number of time steps used for the equilibrium MD and for the neMD simulation.

J. Chem. Theory Comput. 2015, 11, 3919−3931

Laboratory of Computational Biophysical Chemistry

A MC fast proton titration scheme

Laboratory of Computational Biophysical Chemistry

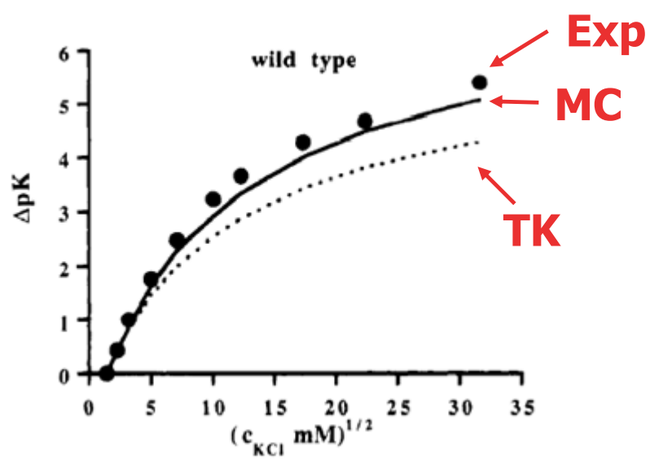
Molecular modeling

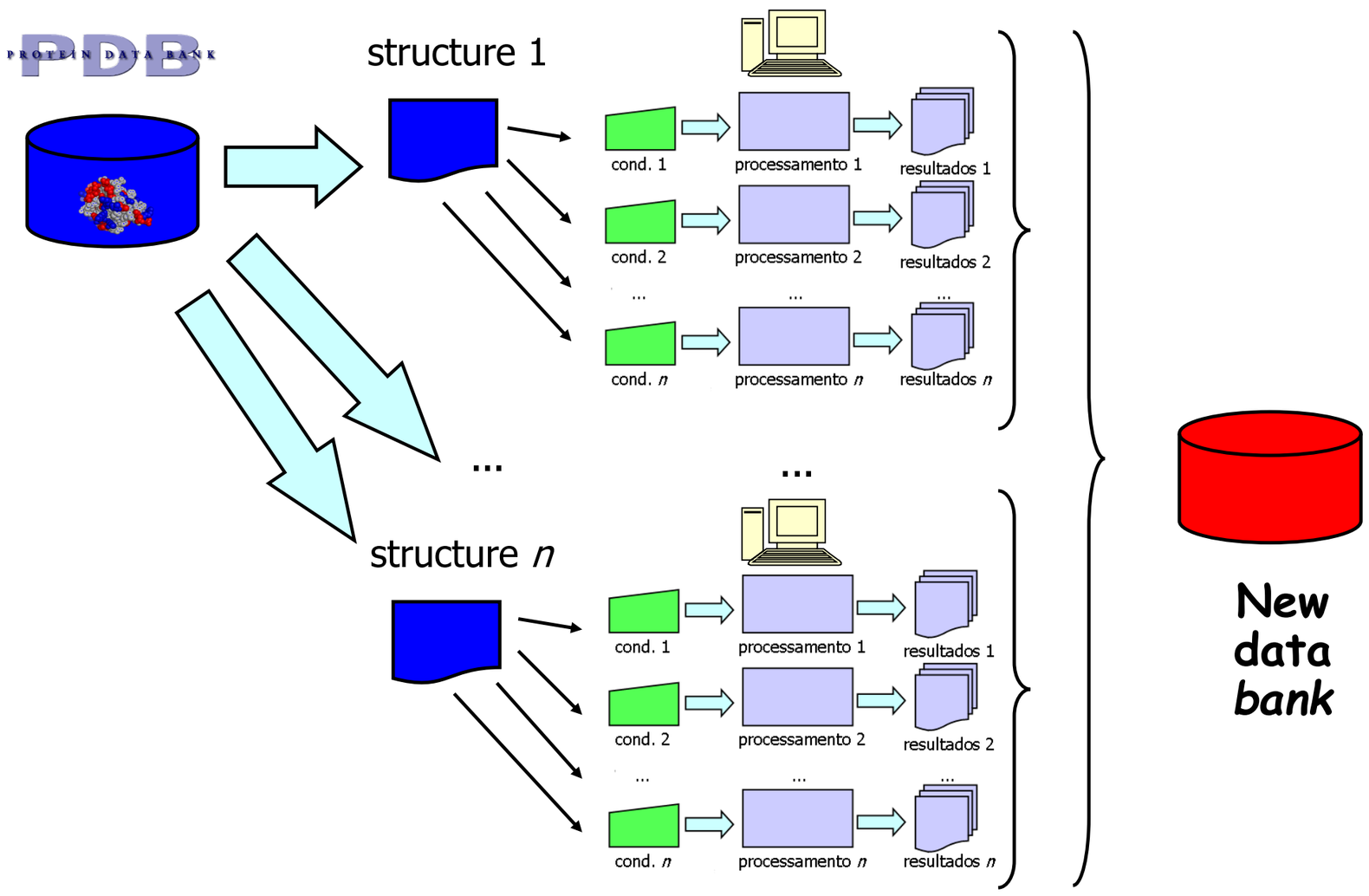
- ✓ **Electrostatic interactions are the main driving force**
- ✓ **Be able to describe pH, ionic strength and mutational effects**
- ✓ **Free energy as a function of the separation [A(r)]**
- ✓ **Repeat the calculations in several conditions**
- ✓ **(small $$$budgets$$$)**

$$\beta A(R) \approx \frac{l_B \langle Q_A \rangle \langle Q_B \rangle}{R} - \frac{l_B^2}{2R^2}\left[C_A C_B + C_A \langle Q_B \rangle^2 + C_B \langle Q_A \rangle^2\right]$$

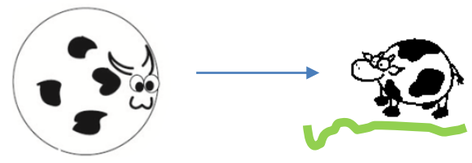
direct Coulomb term

induced charge-induced charge
charge-induced charge terms

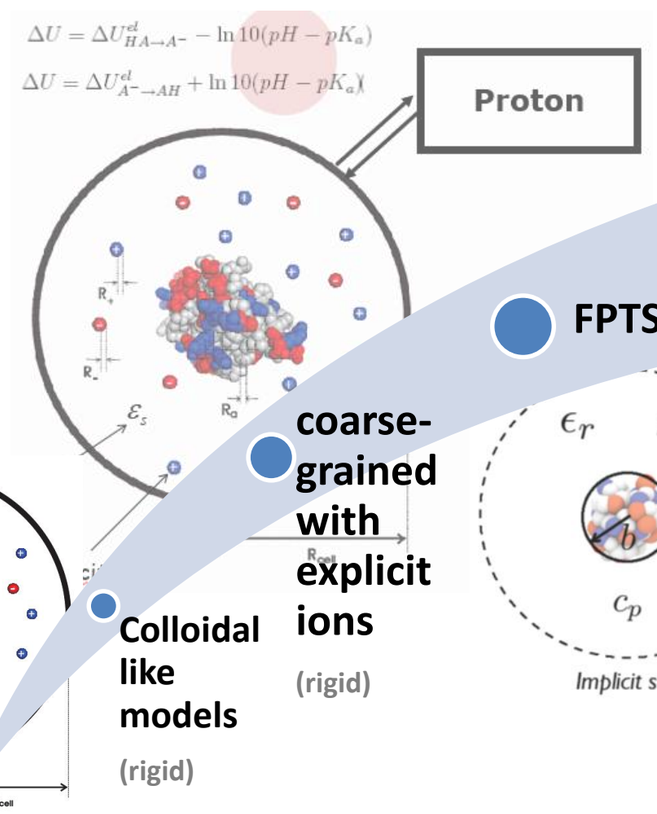




Development of CpH methods (CG)



Protein interactions



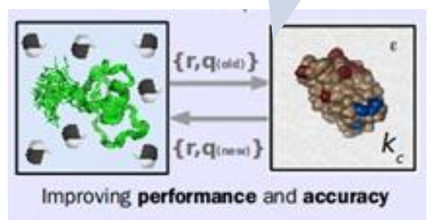
Colloidal like models (rigid)

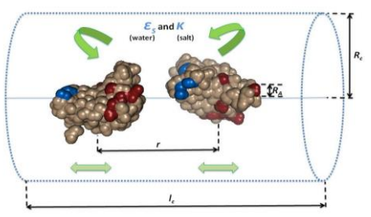
coarse-grained with explicit ions (rigid)

FPTS (rigid)

FPTS+OPEP (flexible)

More to come

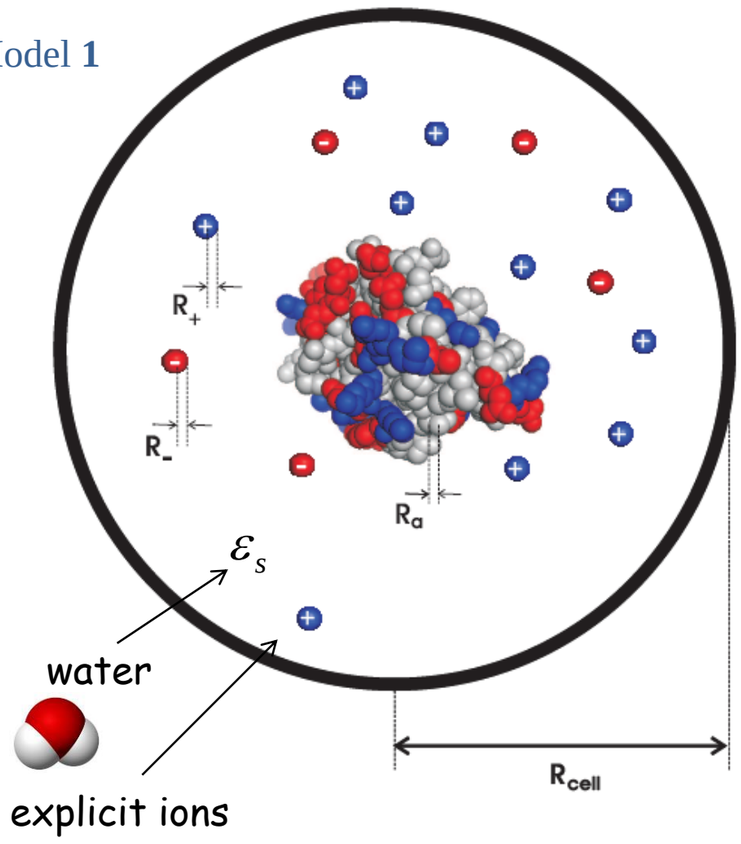




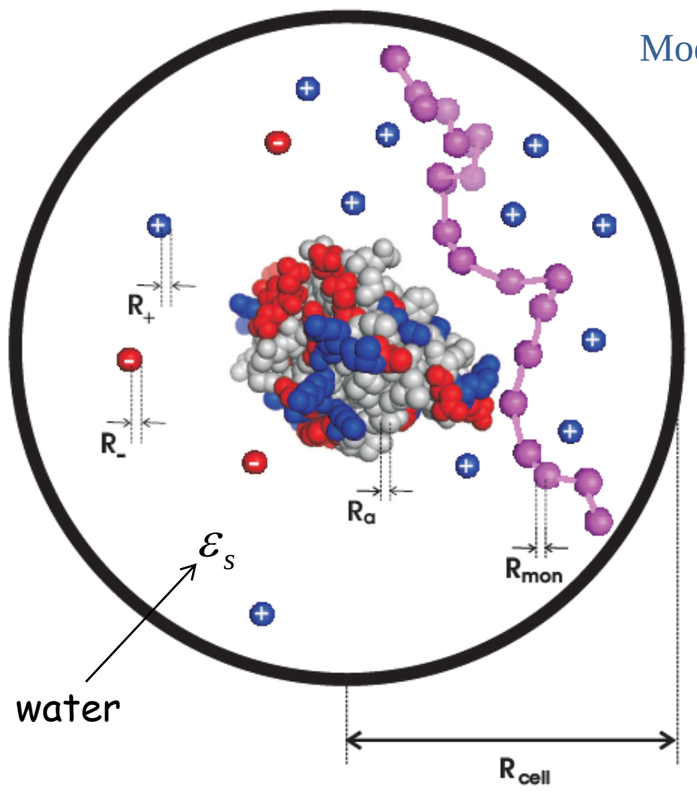
Laboratory of Computational Biophysical Chemistry

We have developed several different "spherical cows"...

Model **1**

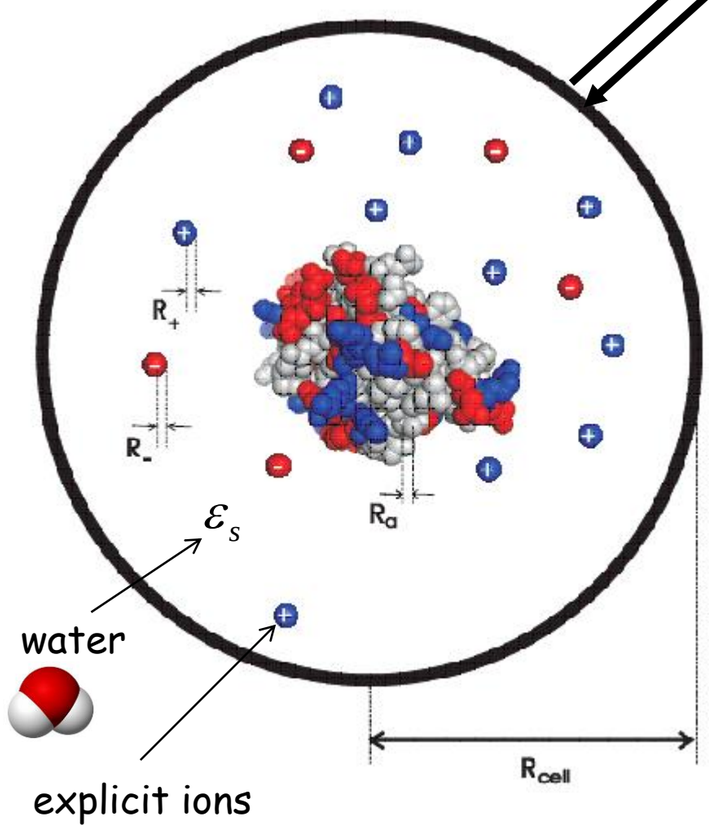


Model **2**

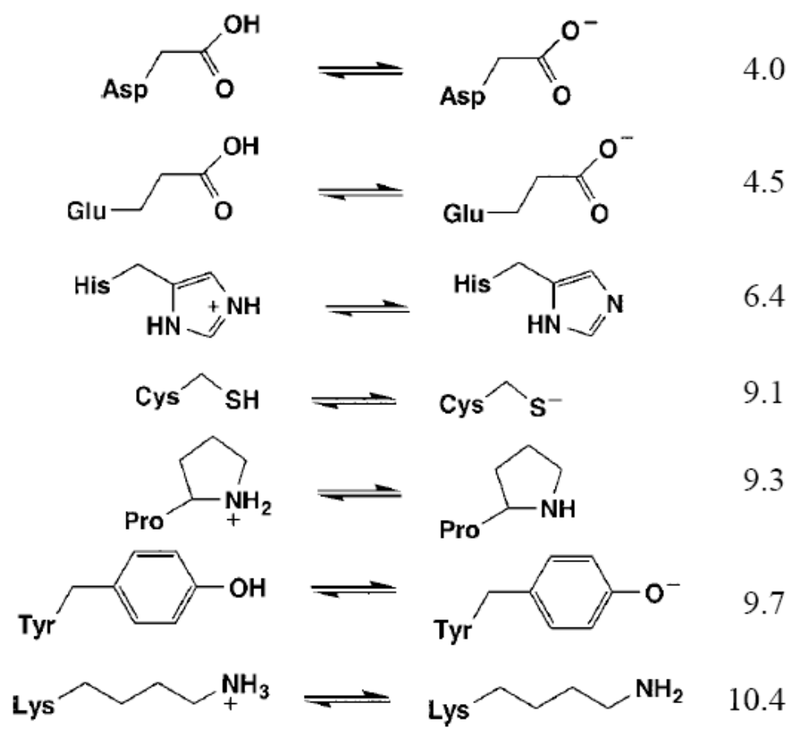


$$\Delta U = \Delta U^{el}_{HA \rightarrow A^-} - \ln 10(pH - pK_a)$$

$$\Delta U = \Delta U^{el}_{A^- \rightarrow AH} + \ln 10(pH - pK_a)$$



$$u(r_{ij}) = \begin{cases} \infty \\ \dfrac{q_i \, q_j}{4 \, \pi \, \varepsilon_0 \, \varepsilon_s \, r_{ij}} \end{cases}$$



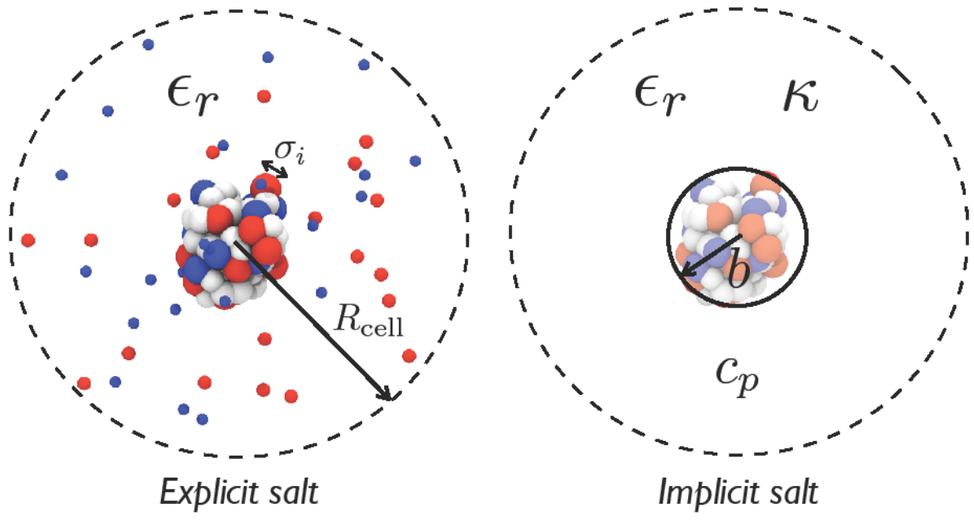
Fast Proton Titration Scheme for Multiscale Modeling of Protein Solutions

Andre Azevedo Reis Teixeira,[†] Mikael Lund,[‡] and Fernando Luís Barroso da Silva*[,†,‡]

JCTC Journal of Chemical Theory and Computation

J. Chem. Theory Comput. **2010,** *6,* 3259–3266

How to reduce cpu time in order to simulate large protein aggregates?



FPTS

$$w_{TK} \approx \frac{e^2}{4\pi\varepsilon_0\varepsilon_r}\left(\sum_{i>j}^{n}\frac{z_i z_j}{r_{ij}} - \frac{Z_P^2\kappa}{2(1+\kappa b)}\right)$$

Laboratory of Computational Biophysical Chemistry

Acid-base equilibrium

Dissociation of a weak acid (HA)

$$HA_{(aq)} \Leftrightarrow H^+_{(aq)} + A^-_{(aq)}$$

$$K_a = \frac{a_{H^+}a_{A^-}}{a_{HA}}$$

activities

thermodynamic equilibrium constant

$$a = \gamma c$$

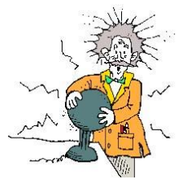
For na *ideal* system,

$$K_s = \frac{c_{H^+}c_{A^-}}{c_{HA}}$$

concentrations

stoichiometric equilibrium constant

Acid-base equilibrium (2)



It follows that

$$a = \gamma c$$

$$K_a = \frac{a_{H^+} a_{A^-}}{a_{HA}} = K_\gamma K_s = \frac{\gamma_{H^+} \gamma_{A^-}}{\gamma_{HA}} \times \frac{c_{H^+} c_{A^-}}{c_{HA}}$$

which can be re-written as

$$-\log K_a = -\log\left(\gamma_{H^+} c_{H^+}\right) - \log\left(\frac{\gamma_{A^-}}{\gamma_{HA}}\right) - \log\left(\frac{c_{A^-}}{c_{HA}}\right)$$

$$pK_a = pH - \log\left(\frac{\gamma_{A^-}}{\gamma_{HA}}\right) - \log\left(\frac{c_{A^-}}{c_{HA}}\right)$$

Acid-base equilibrium (3)

$$pK_a = pH - \log\left(\frac{\gamma_{A^-}}{\gamma_{HA}}\right) - \log\left(\frac{c_{A^-}}{c_{HA}}\right)$$

or,

$$-\ln\left(\frac{c_{A^-}}{c_{HA}}\right) = -\ln\left(\frac{\gamma_{HA}}{\gamma_{A^-}}\right) - \left(pH - pK_a\right)\ln 10$$

$$\beta \Delta A_{HA \to A^-}$$

$$\therefore w = \Delta E - \left(pH - pK_a\right)\ln 10$$

Acid-base equilibrium (4)

$$w = \Delta E - \left(pH - pK_a \right) \ln 10$$

$$G^{el} = \frac{e^2}{8\pi\varepsilon_0} \sum_{i=1}^{Np} \sum_{j=1}^{Np} z_i z_j \left(A_{ij} - C_{ij} \right)$$

$$\mu_F^{ex} = -\frac{\kappa z^2 e^2}{8\pi\varepsilon_0 \varepsilon_s K_B T \left(1 + 2\kappa R_s \right)}$$

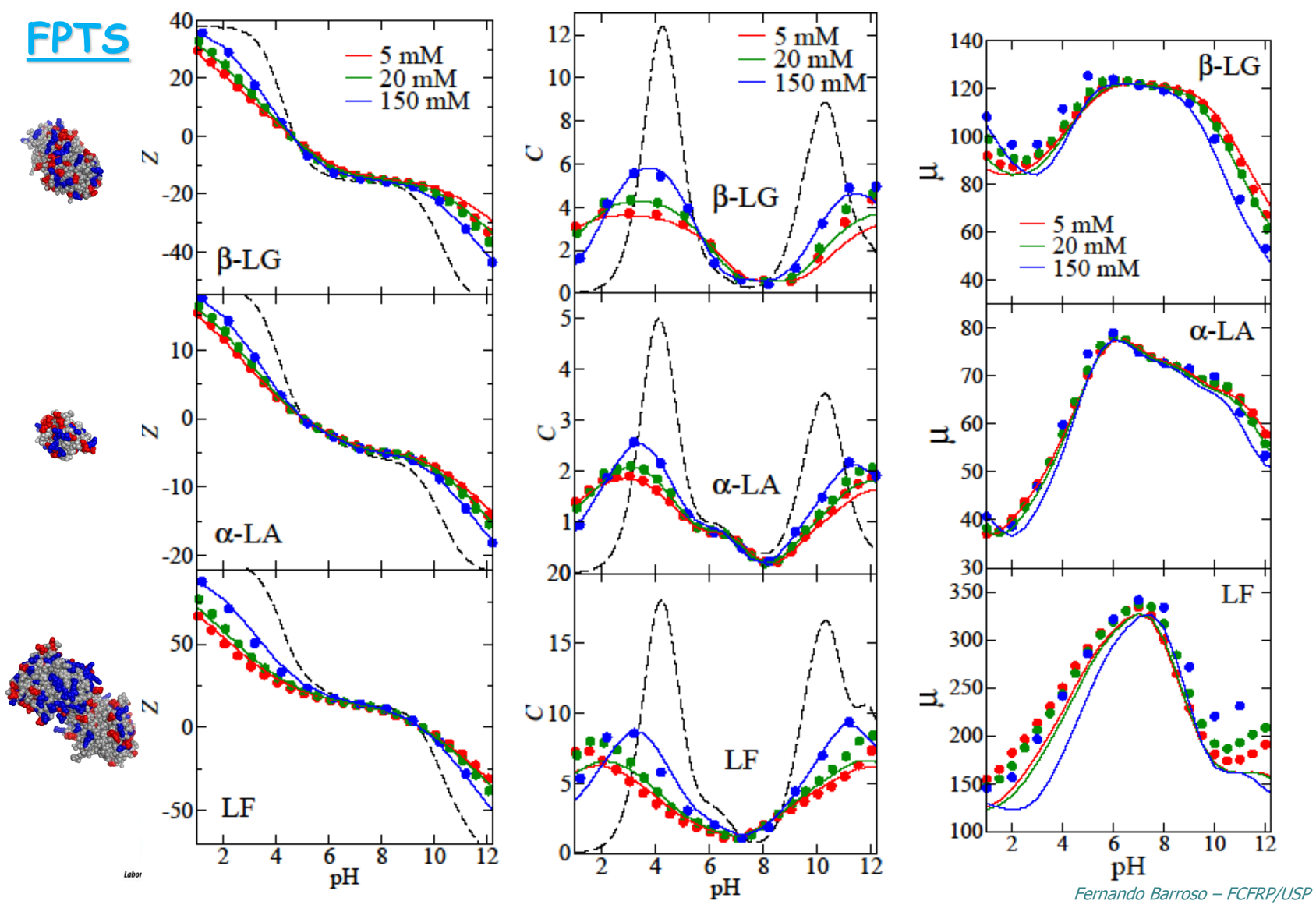
salt

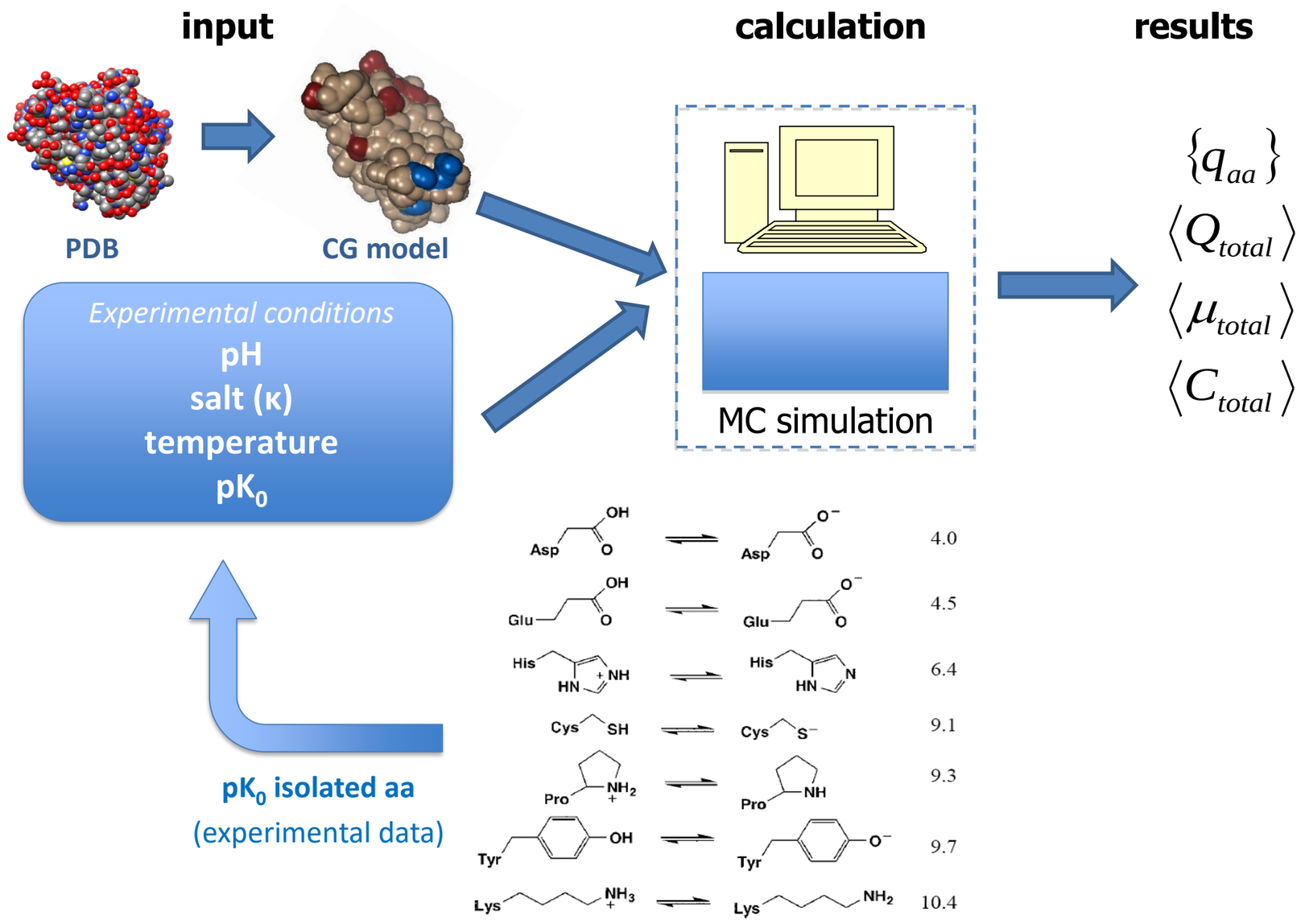
$$A_{ij} = \frac{1}{\varepsilon_s r_{ij}}$$

Coulombic interactions

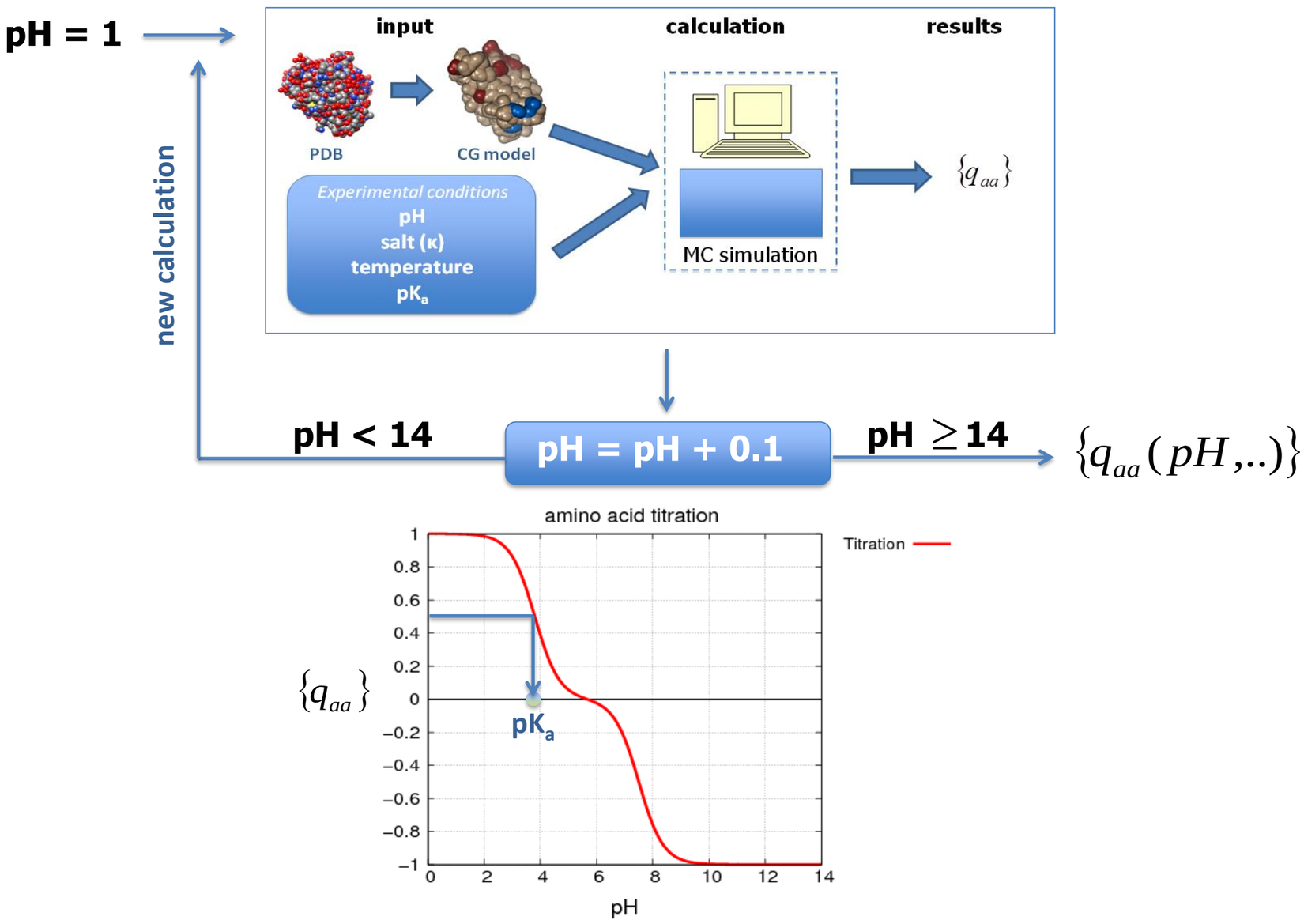
$$w_{TK} = \frac{e^2}{4\pi\varepsilon_0\varepsilon_s} \left[\sum_{i>j}^{Np} \frac{z_i z_j}{r_{ij}} - \frac{Z_P^2 \kappa_c}{2\left(1+\kappa_c b\right)} \right] + \lambda \left(pH - pKa \right) \ln(10)$$

FPTS




input
calculation
results
PDB
CG model
Experimental conditions
pH
salt (κ)
temperature
pK0
MC simulation
{q_aa}
⟨Q_total⟩
⟨μ_total⟩
⟨C_total⟩
pK0 isolated aa
(experimental data)
Asp 4.0
Glu 4.5
His 6.4
Cys 9.1
Pro 9.3
Tyr 9.7
Lys 10.4

115


pH = 1
input
calculation
results
PDB
CG model
Experimental conditions
pH
salt (κ)
temperature
pKa
MC simulation
{q_aa}
new calculation
pH < 14
pH = pH + 0.1
pH ≥ 14
{q_aa(pH,..)}
amino acid titration
Titration
{q_aa}
pKa
pH

116

Laboratory of Computational Biophysical Chemistry

Benchmarking

FPTS

Net charge profiles

[LD & FLBDS, *Food Hydrocolloids*, 2016]

	α-LA	β-LG	LF	αs1-CS	β-CS	k-CS
pI	5.0 (4.2-5.0)	4.8 (4.6-5.5)	9.6 (8.7-9.0)	5.0 (4.4–4.8)	5.0 (4.8–5.1)	5.9 (5.3–5.8)

(*) Experimental data between parentesis: J. Dairy Sci. 87:1641–1674; Thompson, A., Boland, M., & Singh, H. (2009). Milk proteins : From expression to food. USA: Academic Press/Elsevier.

[FLBDS & DMK, JCTC, 2017]

Table 1 Calculated and experimental pKa values of three proteins.

Residue	Experiment	GB 0-1	All-atom CpHMD 0-5	5-10	0-10	Implicit method
HP36						
Asp44	3.10(1)	3.2(1)	2.0	3.0	2.6(5)	3.7
Glu45	3.95(1)	3.5(1)	4.3	4.5	4.4(1)	4.5
Asp46	3.45(12)	3.5(1)	2.4	3.7	3.1(6)	3.8
Glu72	4.37(3)	3.5(1)	4.4	4.4	4.4(0)	3.5
BBL						
Asp129	3.88(2)	3.2(0)	2.2	3.2	2.7(5)	3.7
Glu141	4.46(4)	4.3(0)	4.0	4.4	4.2(2)	3.8
His142	6.47(4)	7.1(0)	5.9	5.8	5.8(0)	6.5
Asp145	3.65(4)	2.8(2)	3.0	3.1	3.1(0)	3.7
Glu161	3.72(5)	3.6(3)	4.2	3.9	4.0(2)	4.1
Asp162	3.18(4)	3.4(3)	2.9	3.5	3.2(3)	3.2
Glu164	4.50(3)	4.5(1)	5.7	4.6	5.2(6)	3.8
His166	5.39(2)	5.4(1)	4.4	4.4	4.4(0)	6.0
HEWL						
Glu7	2.6(2)	2.6(1)	3.6	3.4	3.5(1)	3.3
His15	5.5(2)	5.3(5)	5.1	5.1	5.1(0)	5.6
Asp18	2.8(3)	2.9(0)	2.5	3.3	2.9(4)	2.8
Glu35	6.1(4)	4.4(2)	8.5	8.7	8.6(1)	3.5
Asp48	1.4(2)	2.8(2)	0.1	1.1	0.6(6)	3.4
Asp52	3.6(3)	4.6(0)	5.4	5.6	5.5(1)	3.3
Asp66	1.2(2)	1.2(4)	0.6	0.8	0.3(7)	3.0
Asp87	2.2(1)	2.0(1)	0.8	2.1	1.5(7)	3.2
Asp101	4.5(1)	3.3(3)	6.1	5.7	5.9(2)	2.9
Asp119	3.5(3)	2.5(1)	3.0	3.3	3.2(1)	3.2
Maximum absolute deviation		1.8	2.4	2.6	2.5	2.6
Average absolute deviation (RMS deviation)		0.5	1.0	0.6	0.7	0.5

FPTS

Proteins

Table 2 Calculated and experimental pKa values of Roux proteins.

FPTS [FLBDS & DMK, JCTC, 2017]

Proteins

Residue	Experiment	FTSP	Roux	FTSP-exp	Roux-exp
Turkey ovomuc.					
ASP-7	2.7	3.34	3.4	0.64	0.73
ASP-27	2.3	3.15	4.3	0.85	1.97
GLU-10	4.1	3.99	4.0	-0.12	-0.06
GLU-19	3.2	3.38	3.5	0.18	0.33
GLU-43	4.8	4.12	4.4	-0.68	-0.41
Maximum absolute deviation		0.85	1.97		
Average absolute deviation (RMS deviation)		0.57	0.97		

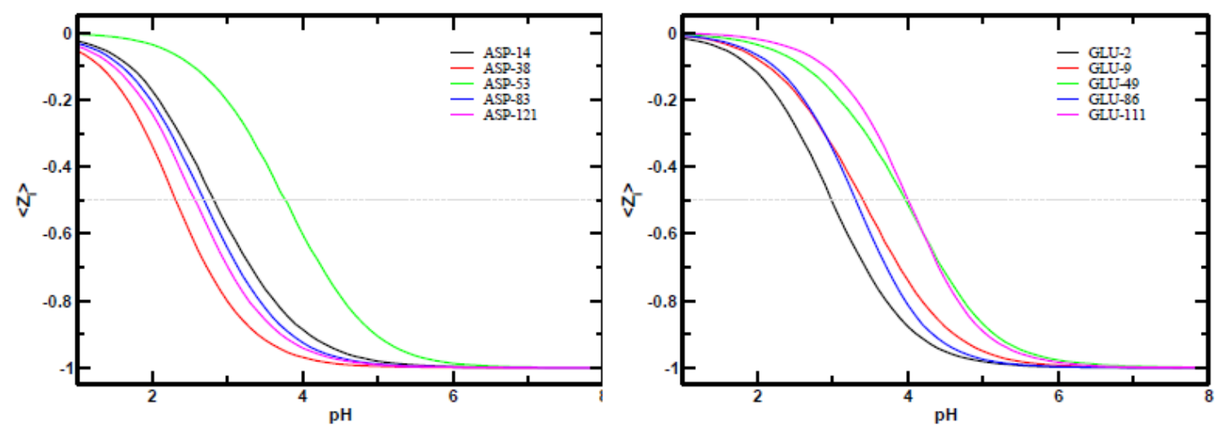


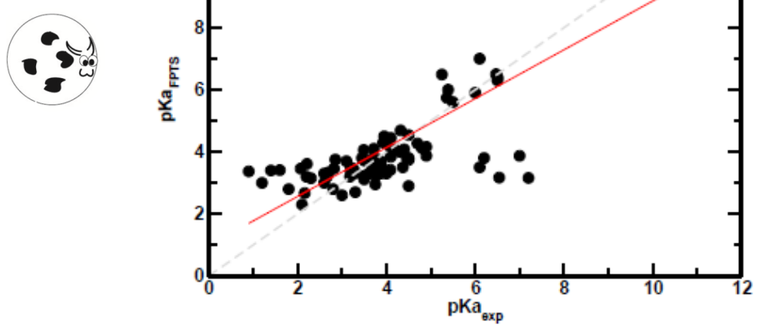
Figure 2: Computed titration plots of the acid aminoacid residues ASP (left panel) and GLU (right panel) of ribonuclease A at 60mM salt concentration. The dashed gray lines indicate the half of the protonated states which is used to predicted the theoretical pK_a. Data are taken from the titration simulations with the FPTS. The intrinsic pK_0 values of the aminoacid model compounds are 4.0 and 4.4, respectively, for ASP and GLU.[27]

/USP
2020

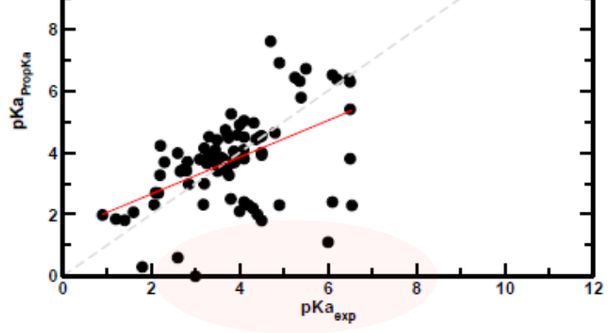
119

[FLBDS & DMK, JCTC, 2017]

FPTS



PropKa



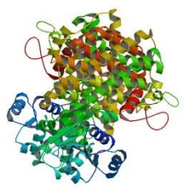


Figure 4: Correlation between experimental and calculated pK_as using the FPTS (top panel) and PropKa (botton panel) methods. All points were taken from the previous tables (1–6) for the cases where there was both an experimental and predicted data ($n = 87$ for FPTS and $n = 81$ for PropKa). Fitted regression lines are given in red. The linear correlation coefficients and slopes are equal to 0.68 and 0.79, for FPTS; 0.51 and 0.60, for PropKa.

Performance	PB	FPTS
6-phosphogluconate dehydrogenase (PDB id 2zyg)	9185.2s AMD Opteron 2356 processador (8 cores and 2.3 GHz)	96s Intel i7-3630QM and 2.40 GHz

FPTS

[FLBDS, SP & PD, *JCP*, 2017]

Table 2 Calculated pKa values for the lead-dependent ribozyme (LDZ) in 100 mM NaCl. All data for the experimental, NLPB, GNLPB and CPHMD$^{MS\lambda D}$ calculations were obtained, respectively, from refs.[4,5,6] and[7]. The mean and standard deviations of the calculated pKa values for the present work were obtained from the results of all 25 NMR structures available in the PDB coordinates (PDB id 1ldz). Precision is defined here as the mean standard deviations. AUE is the unsigned average error as used in refs.[7,8].

FPTS

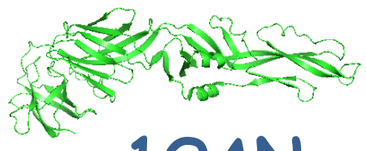
RNA

		NLPB		GNLPB		CPHMD$^{MS\lambda D}$		Present work	
Nucleotide	exp. pKa	pKa	error	pKa	error	pKa	error	pKa	error
C2		2.1±1.5				4.4±0.2		5.62±0.12	
A4	¡3.1	¡3		3.9±0.8		0.6±0.1		4.64±0.07	
C5		3.0± 2.0				3.5±0.4		5.67±0.15	
C6		2.8±2.4				3.0±0.3		5.77±0.12	
A8	4.3 ± 0.3	4.9±0.8	-0.6	4.7±0.5	-0.4	3.7±0.3	+0.6	4.99±0.13	-0.69
C10		1.4±1.5				1.1±0.3		5.82±0.10	
C11		3.7±1.5				1.3±0.9		5.66±0.08	
A12	¡3.1	¡3		4.0±0.3		0.7±0.3		4.69±0.05	
C14		4.6± 1.0				3.2±0.3		5.81±0.05	
A16	3.8±0.4	3.4±1.1	+0.4	4.3±0.7	-0.5	2.6±0.1	+1.2	4.17±0.05	-0.37
A17	3.8±0.4	2.4±1.3	+1.4	3.8±0.7	0.0	0.9±0.5	+2.9	4.73±0.10	-0.93
A18	3.5±0.6	3.6±0.9	-0.1	4.1±0.3	-0.6	3.8±0.1	-0.3	4.92±0.07	-1.42
A25	6.5±0.1	7.3±1.8	-0.8	5.7±0.5	0.8	4.8±0.5	+1.7	5.16±0.12	+1.34
C28		3.1±0.7				3.7±0.1		5.59±0.05	
C30		5.0±2.0				4.8±0.3		5.15±0.11	
Max abs dev			1.4		0.8		2.9		1.4
AUE			0.7		0.5		1.3		1.0
Precision		1.4		0.5		0.3		0.09	

USP

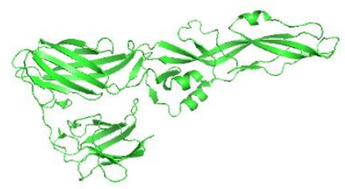
Protein interactions

Dengue 2 virus envelope protein



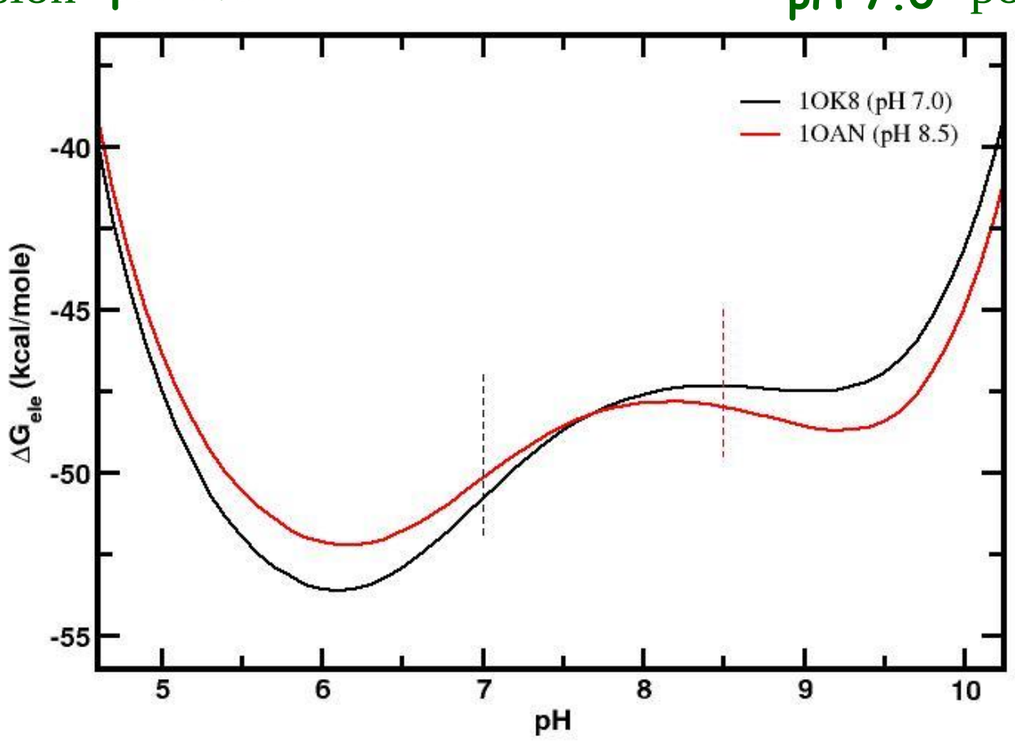
1OAN

prefusion pH 8.5



1OK8

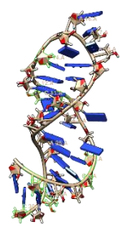
pH 7.0 post fusion



[**FLBDS** & CE, in preparation]

Laboratory of Computational Biophysical Chemistry

Fast convergence for any system!



[FLBDS, SP & PD, JCP, 2017]

FPTS

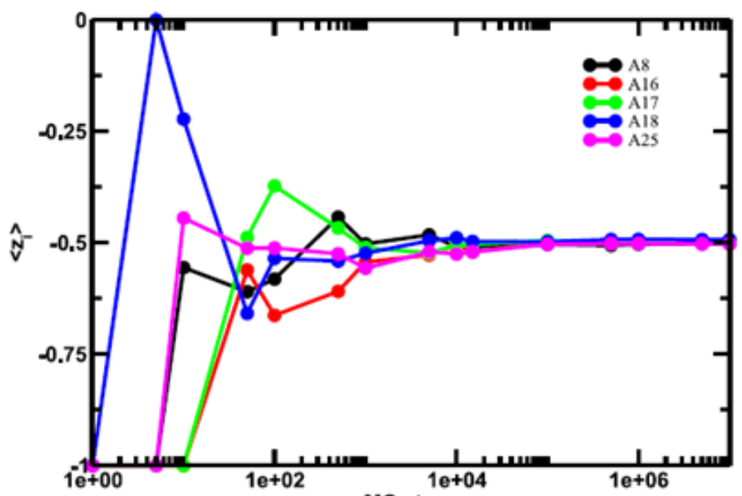


FIG. 2. Numerical convergence of the average valency of adenosines (A8, A16, A17, A18, and A25) as a function of the number of MC steps. Data for LDZ at 100 mM NaCl using the first NMR structure available in the PDB file (PDB id 1LDZ). Each curve was obtained at the pH correspondingly to the pK_a of the nucleobase.

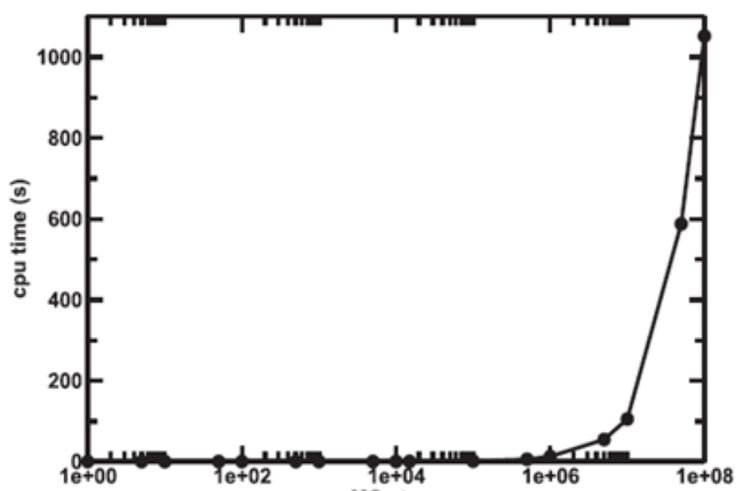


FIG. 3. Estimated cpu time for the numerical convergence of the FPTS as a function of the number of MC steps. Data for LDZ at pH 4.5 and 100 mM NaCl using the first NMR structure available in the PDB file (PDB id 1LDZ).