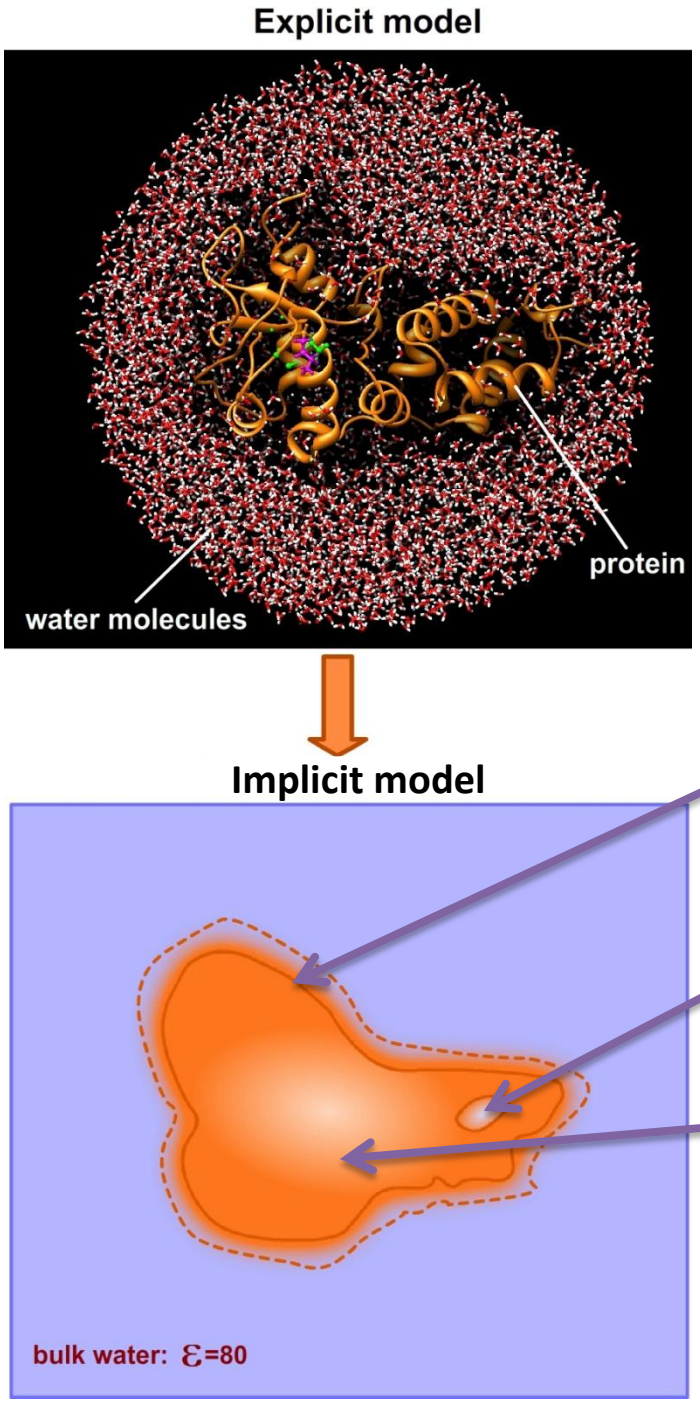
Implementing smooth dielectric function to pKa predictions and other applications

Lin Li, Chuan Li, Emil Alexov

Computational Biophysics and Bioinformatics Lab
Department of Physics and Astronomy
Clemson University



Framework of continuum electrostatics

Three important problems:

(A) How to treat the solvent-solution boundary

(B) How to treat the cavities inside biomolecules

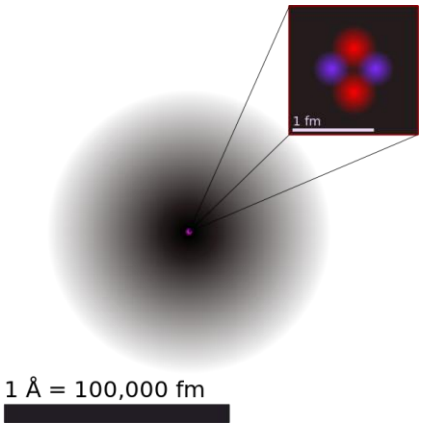
(C) How to treat the dielectric property of a biomolecule

How to assign dielectric constant values to different regions **automatically**?

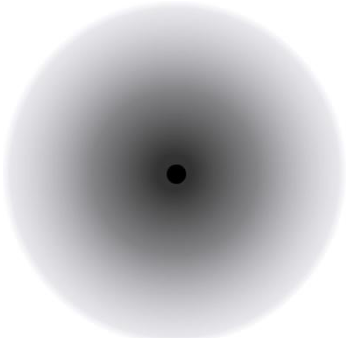
How to model smooth dielectric function

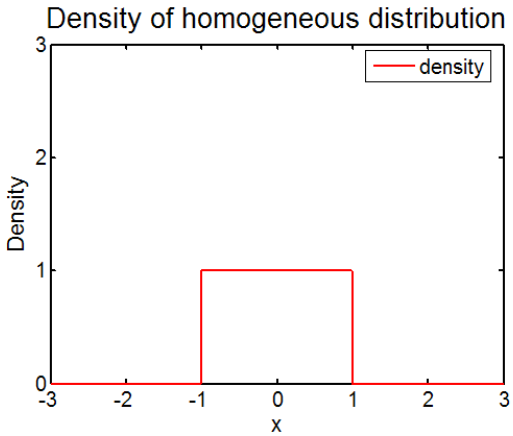
Grant, J. A.; Pickup, B. T.; Nicholls, A. Journal of computational chemistry 2001, 22(6), 608-640.

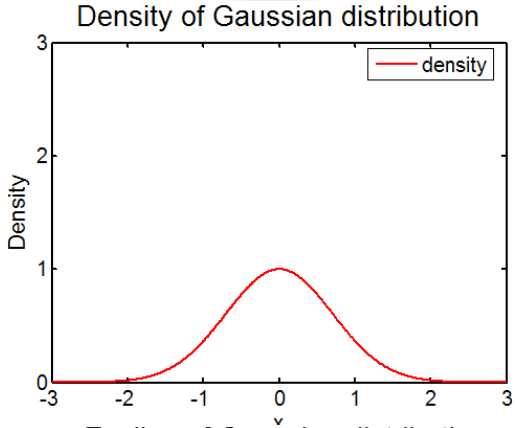
What is an atom?

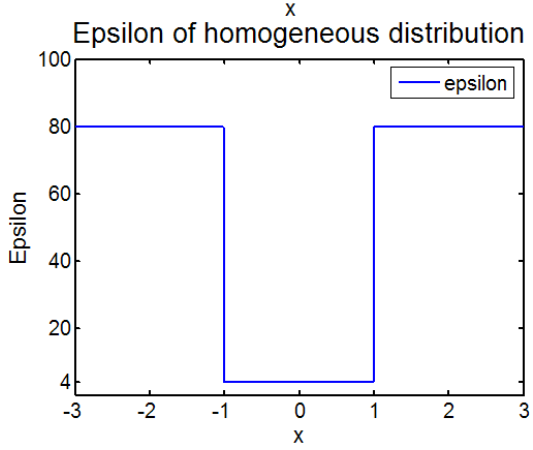


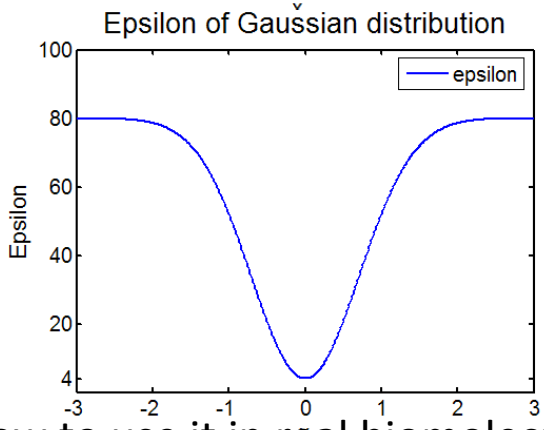
Density and epsilon of an atom











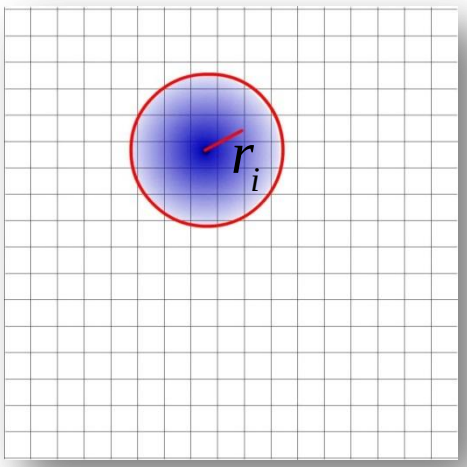
How to use it in real biomolecules?

How to model the density and epsilon of molecules

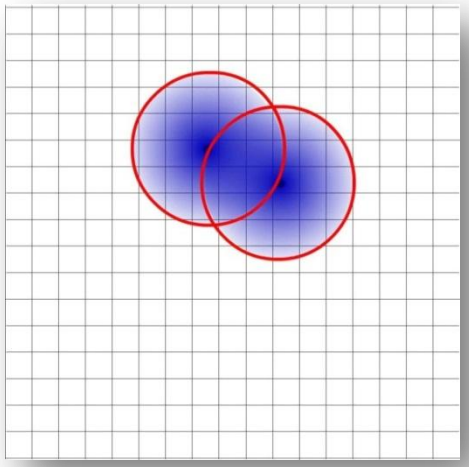
$$\rho_i(r) = e^{-\frac{r_i^2}{\sigma^2 \cdot R_i^2}}$$

$$\rho_{mol}(r) = 1 - \prod_i [1 - \rho_i(r)]$$

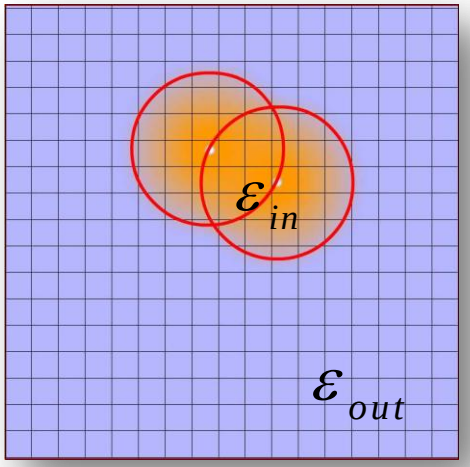
$$\varepsilon = \rho \cdot \varepsilon_{in} + (1 - \rho) \cdot \varepsilon_{out}$$



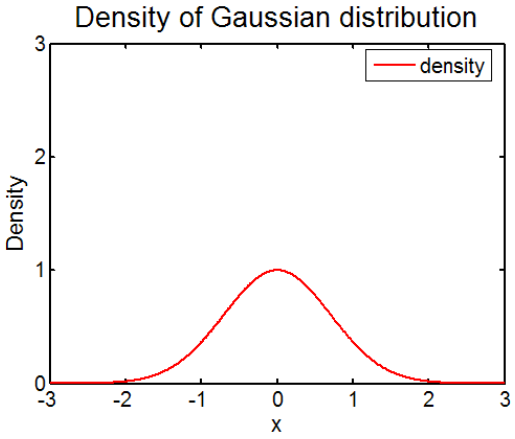
Density of single atom

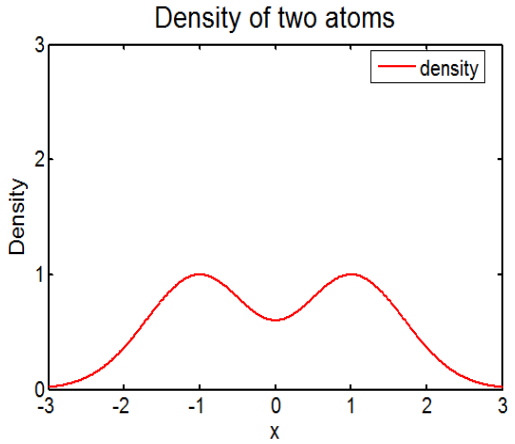


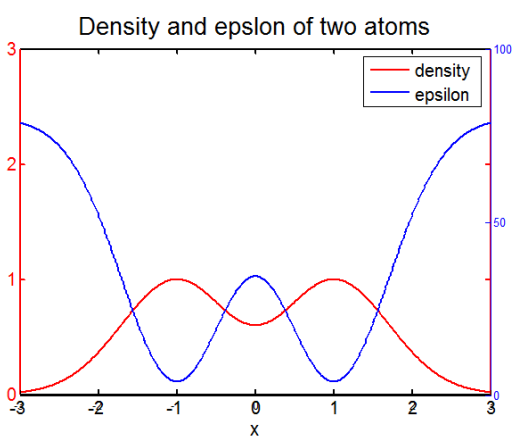
Density of multiple atoms



Convert density to epsilon



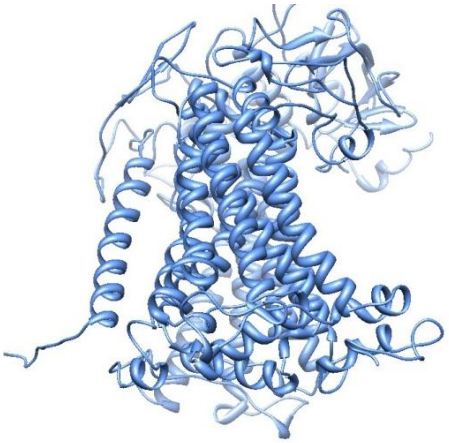


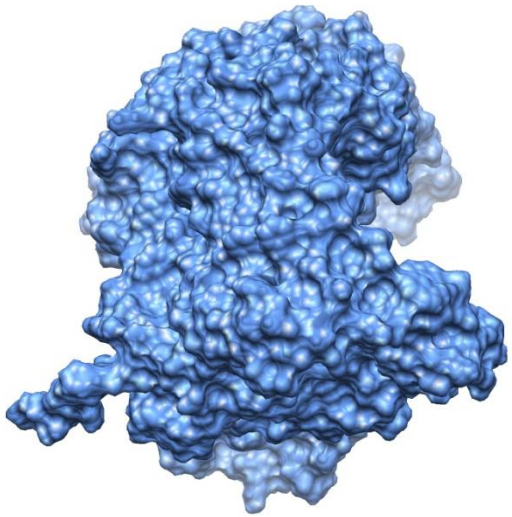


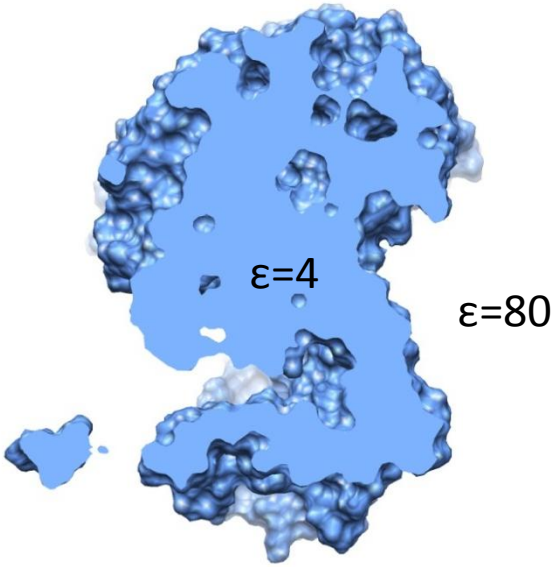
Grant, J. A.; Pickup, B. T.; Nicholls, A. Journal of computational chemistry 2001, 22(6), 608-640.

Dielectric distributions from different methods

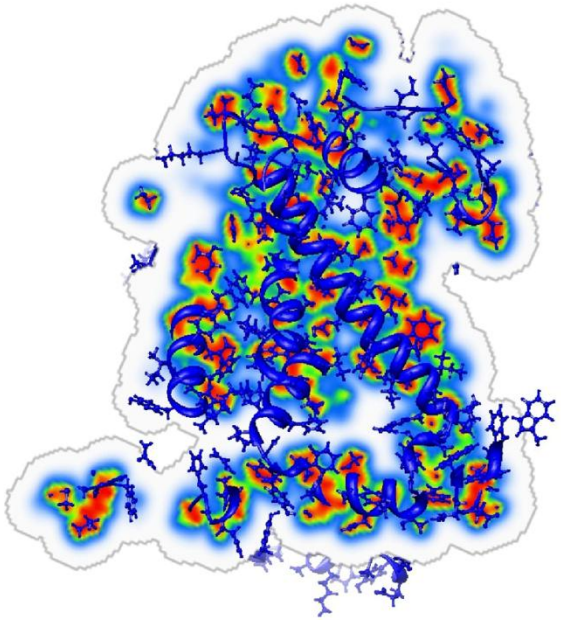
Homogeneous DelPhi:

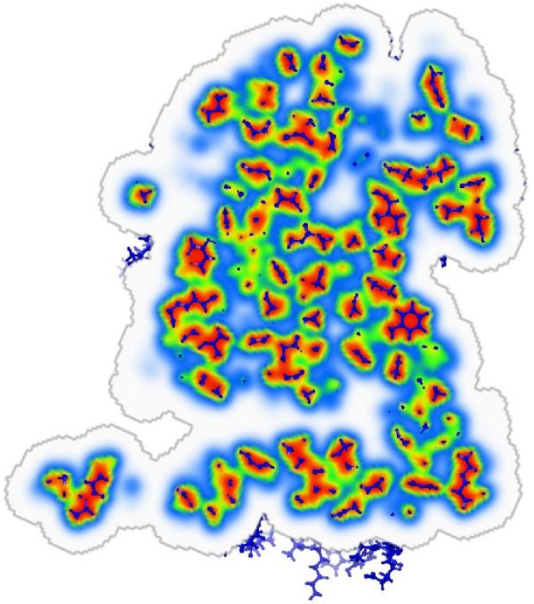


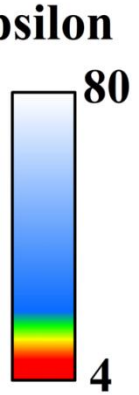




Gaussian DelPhi:

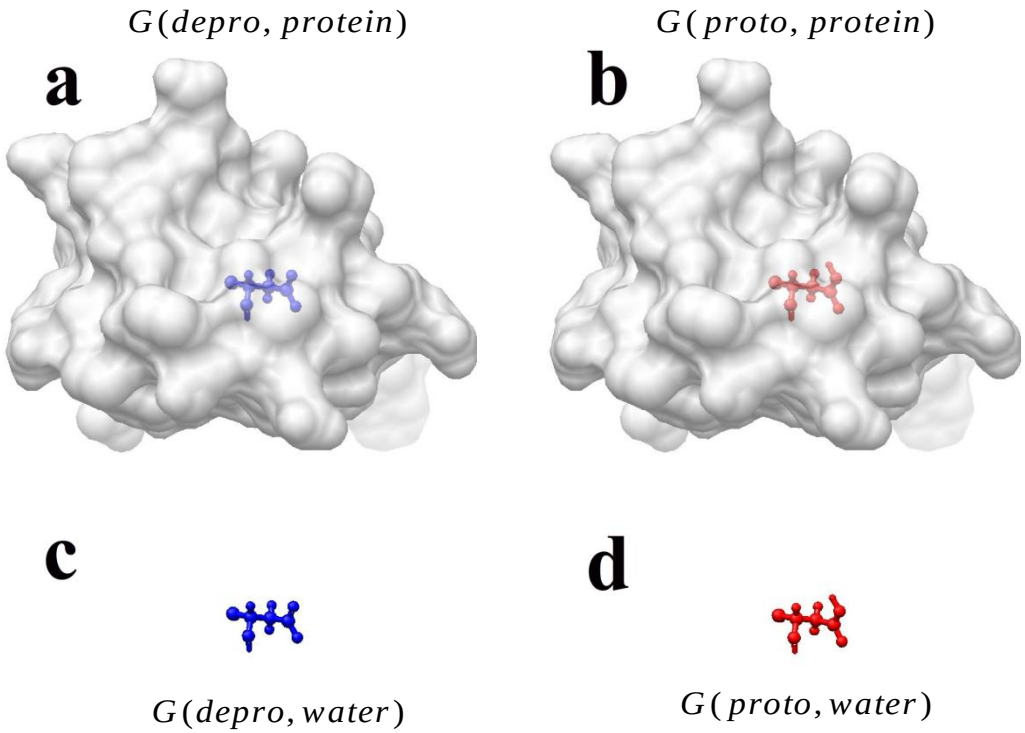






Movie

pKa calculation



Homogeneous DelPhi:

The $\varepsilon_{protein}$ value was varied from 1.0 to 20.0 with increment 1.0.

Gaussian DelPhi:

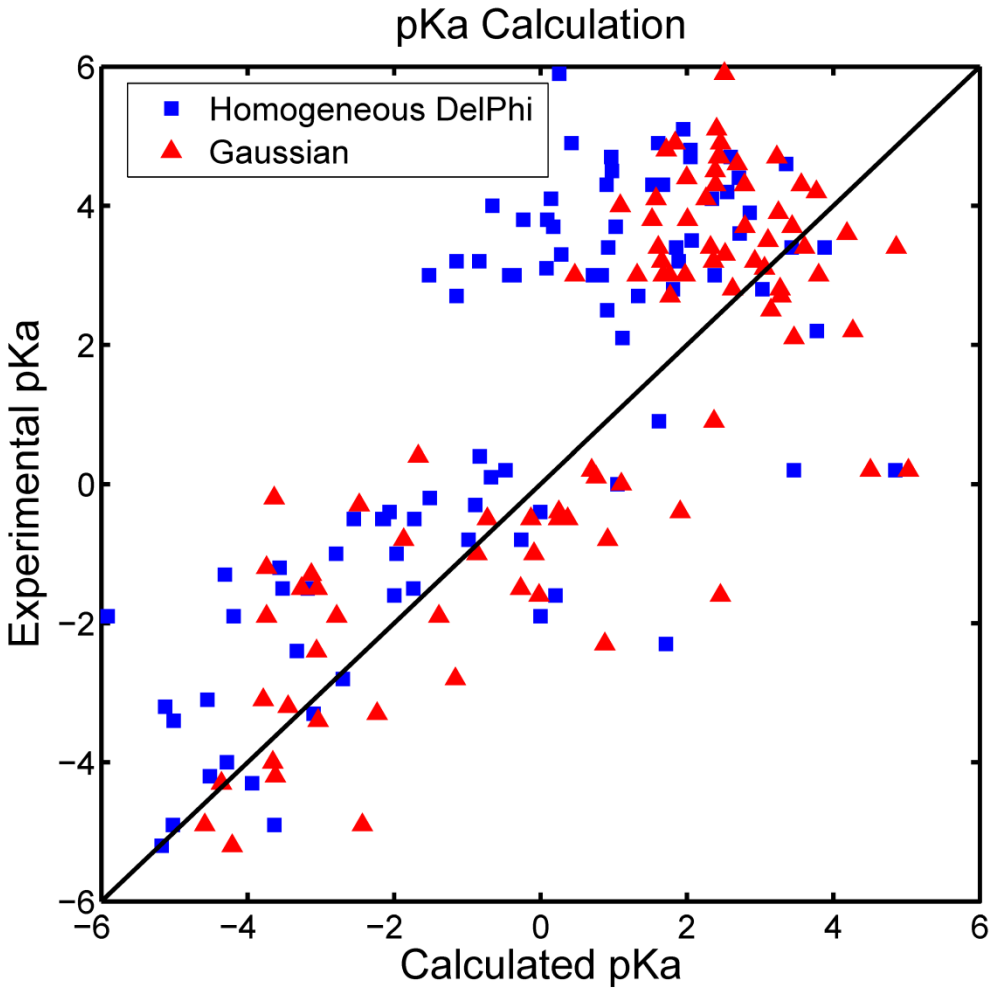
The reference ε_{in} varied from 1.0 to 10.0 with increment 1.0
σ varied from 0.80 to 1.20 with increment 0.01.

$$\Delta pK_a = [G(depro, protein) - G(proto, protein) - G(depro, water) + G(proto, water)] / 2.3$$

No “surface” is needed, it’s the difference of grid energies.

pKa calculation

The dataset used is from Garcia-Moreno lab (http://pkacoop.org/wordpress/?p=28),
This dataset is comprised of **89** pKa's staphylococcal nuclease (SNase).
which is used in pKa-cooperative (http://pkacoop.org/wordpress/).



Homogeneous DelPhi: RMSD=2.62

$\varepsilon_{protein}$ = 10.0

Gaussian DelPhi: RMSD=1.77

ε_{in} = 4.0 and σ =0.93

Is it reasonable physics?

Dielectric distribution analysis

Dataset from PDB bank

Three filtering steps were performed to all of the protein structures of entire PDB bank:

1. Only structures determined by X-Ray with resolution < 1.5 Å were selected.
2. The sequence similarity > 30% were removed.
3. structures with cofactors which are not made of regular residues were removed

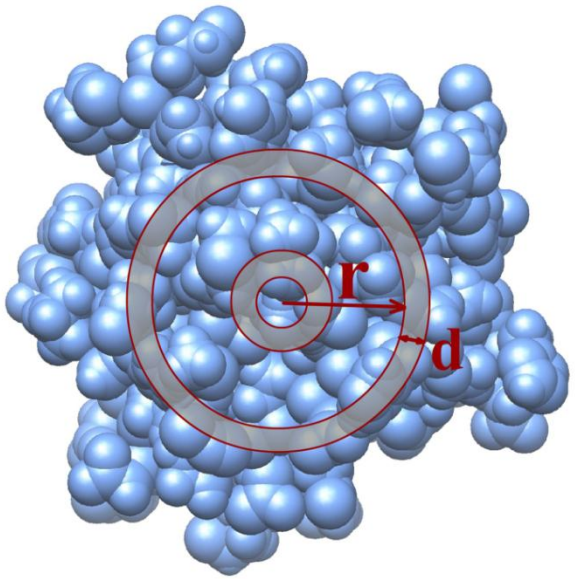
The final dataset is made of **91 proteins**:

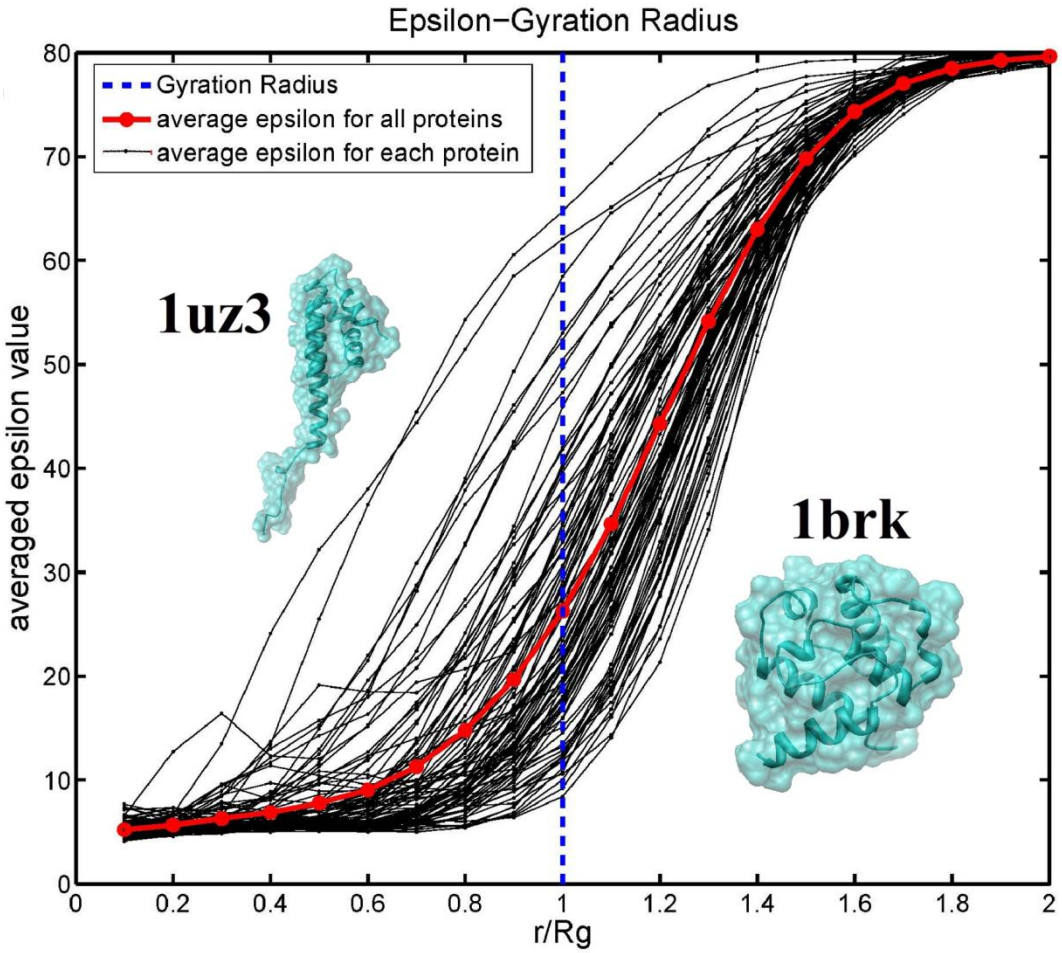
Based on this dataset, we analyzed:

(1) the average ε distribution against radii of proteins;

(2) the average ε distribution against residue types.

The average ε distribution with respect to radii of proteins

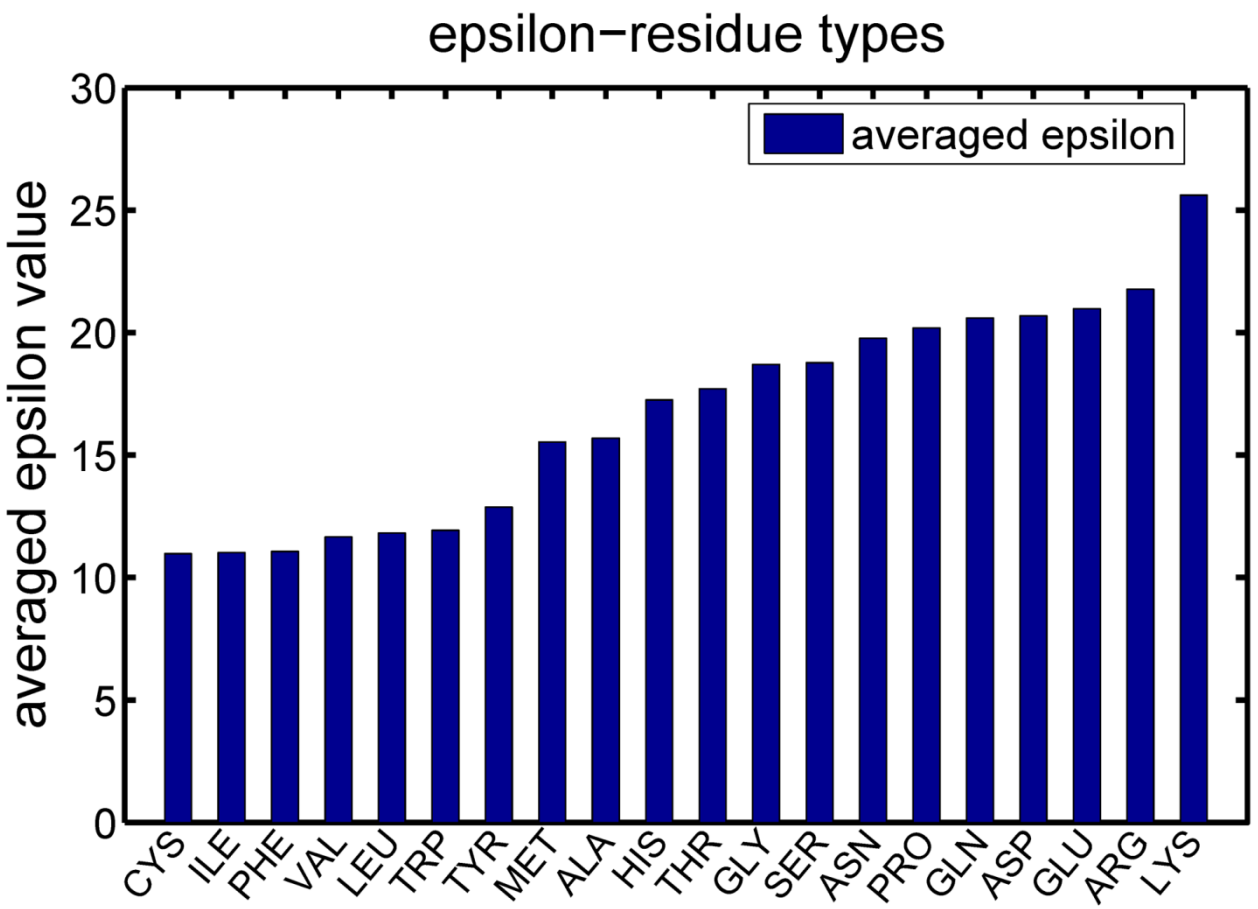




Similar tendency has also been observed by analysis of MD simulation with explicit model:

Simonson, T. & Perahia, D. Internal and interfacial dielectric properties of cytochrome c from molecular dynamics in aqueous solution. ***Proceedings of the National Academy of Sciences*** *92, 1082-1086 (1995).*

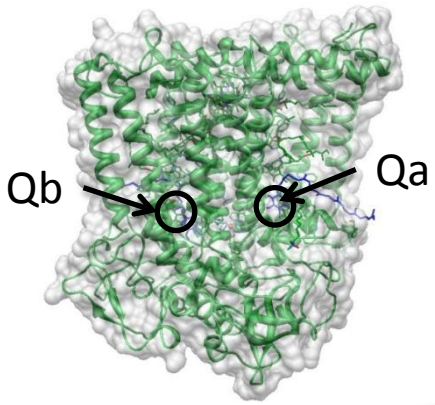
The average ε with respect to residue type distribution

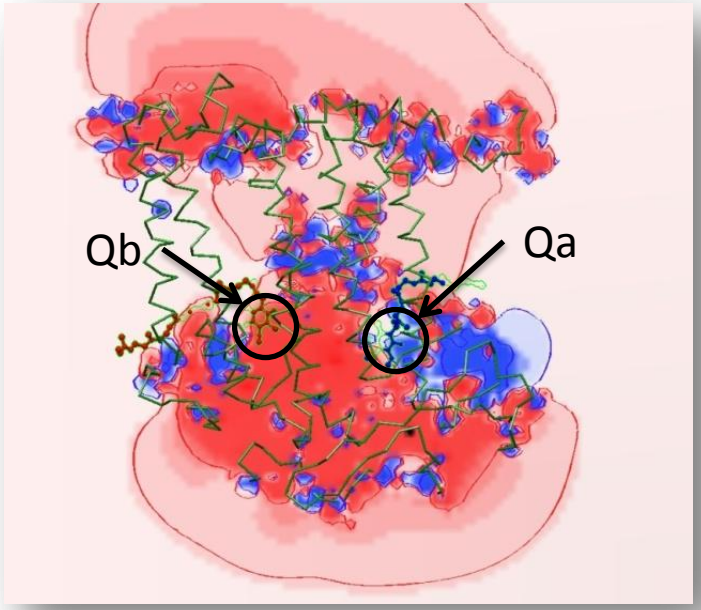


The average dielectric constant per residue was calculated using only the side chain atoms. Then a sphere of radius 5 Å was drown around each side-chain atom and the dielectric constant of all mid-grid points within the sphere were summed and averaged. Further these average dielectric constants were summed over all atoms of the side-chain and averaged again to obtain the average dielectric constant per side-chain. Finally the average dielectric constant for each type of residues was obtained from all residues with the 91 protein set.

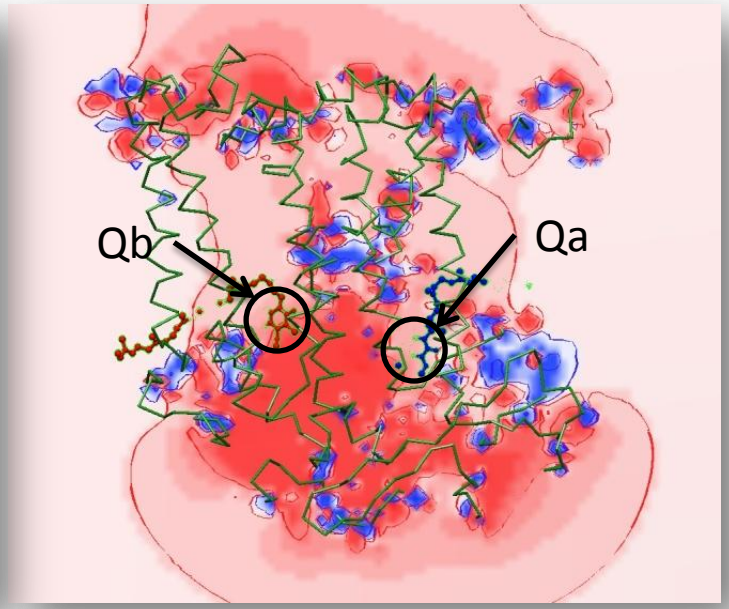
Electrostatic potential calculation for reaction center protein

PDB ID: 1AIJ





Homogeneous ε_{in} Delphi:
Epsilon: 4.0 or 80.0



Gaussian DelPhi:
Epsilon: from 4.0 to 80.0

Similar result has been obtained

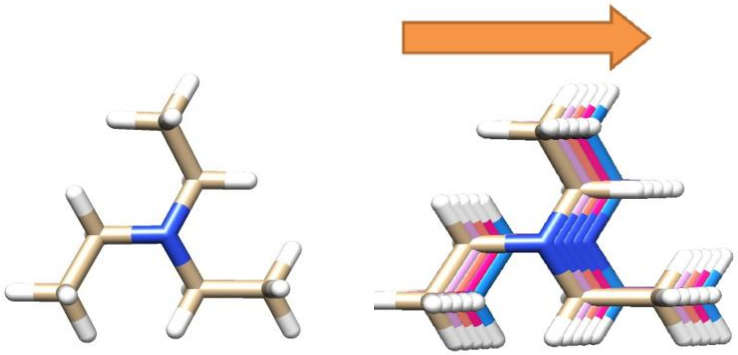
Alexov, E. and Gunner, M., Biochemistry, (1999), 38 (26) 8253-8270.

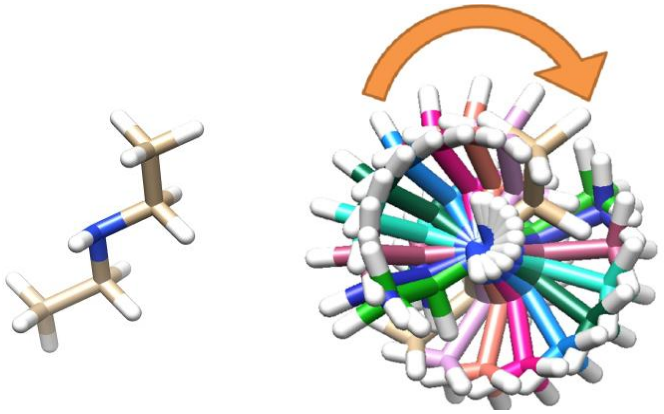
Rocchia, W., Alexov, E. and Honig, B. , J. Phys. Chem. (2001), 105, 6507-6514.

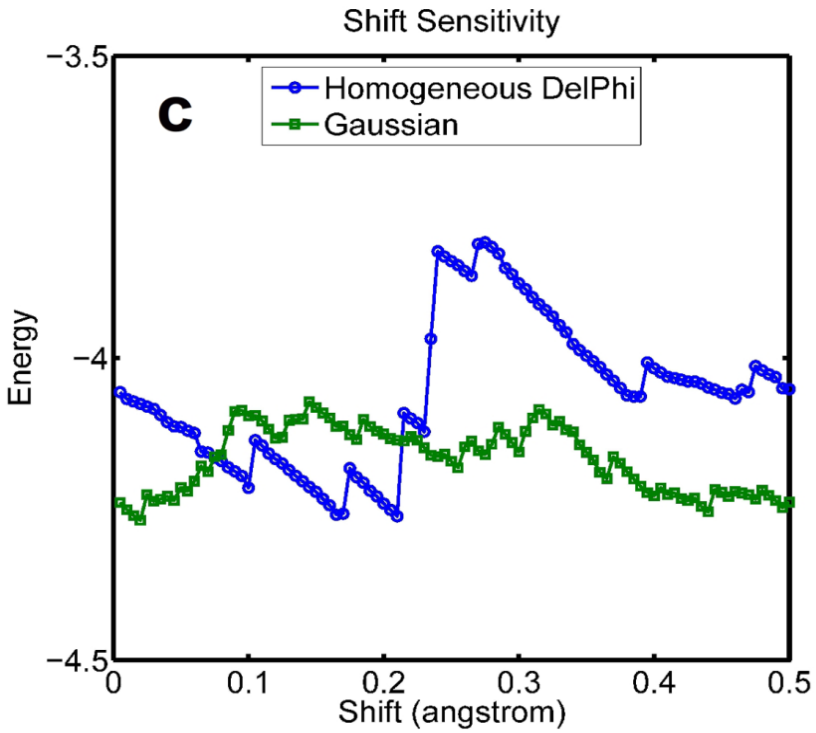
Assign different dielectric values **manually**
Based on analysis of **Monte-Carlo** simulation.

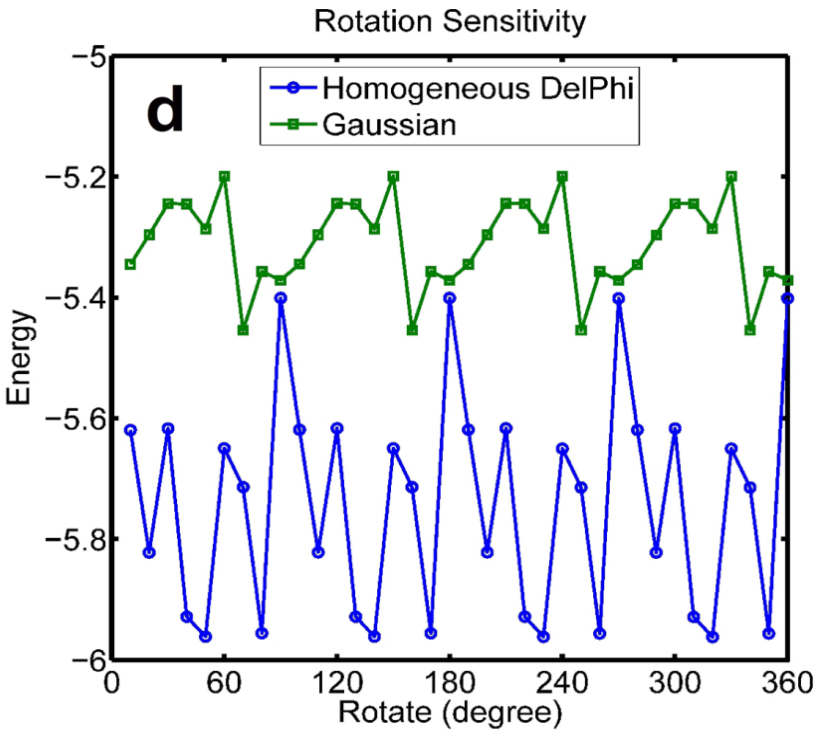
Our method:
Assign dielectric values **automatically**
No MD or MC simulations

Reducing grid dependence



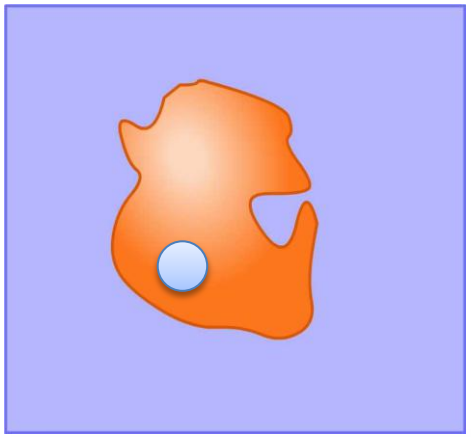






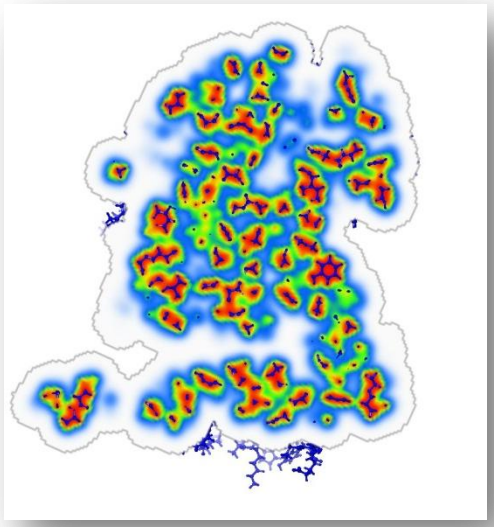
Conclusion

Go back to the considerations at the beginning



Three important problems:

(A) how to treat the solvent-solution boundary
(B) How to treat the cavities inside proteins
(C) how to treat the macromolecule dielectric properties



The first layer of water shell surround protein:
$\varepsilon < 80$

Cavities:
$4 < \varepsilon < 80$

Inside protein:
ε is inhomogeneous

Li L, Li C, Zhang Z, Alexov E., J Chem Theory Comput. 2013 Apr 9;9(4):2126-2136

Current concern:

Side chains at surfaces and cavities:

1. More flexible
2. Some of them are highly charged

How to model the dielectric property of these regions?
Should the dielectric constant be always less than in water?

$$\varepsilon = \begin{cases} \varepsilon_{out}, & \text{when } \rho = 0 \\ \rho \cdot \varepsilon_{in} + (1-\rho) \cdot \varepsilon'_{out}, & \text{when } \rho > 0 \end{cases}$$

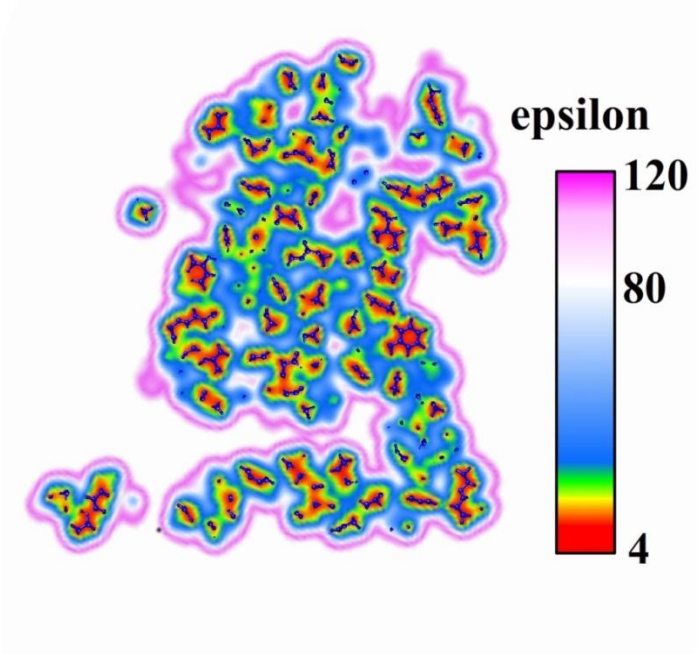
ε_{out} = 80

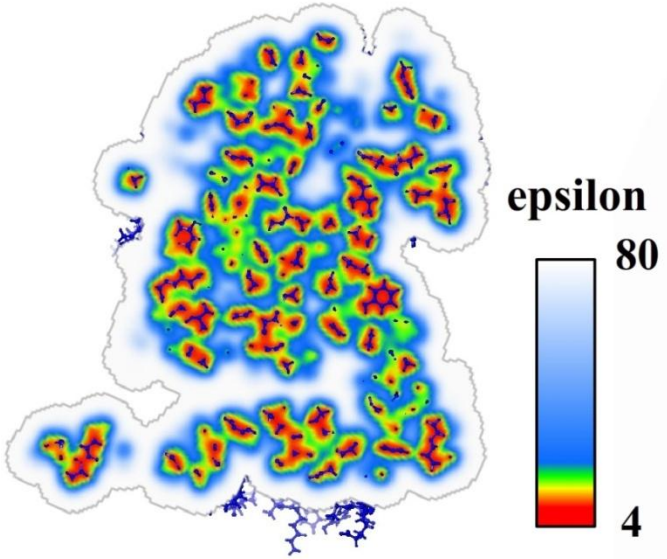
ε'_{out} = 60, 100, 120

The reference ε_{in} varied from 1.0 to 10.0 with increment 1.0
σ varied from 0.80 to 1.20 with increment 0.01.

Results show that ε'_{out} prefers to be larger than 80.0

Eps(in)	sigma	Eps'(out)	RMSD
8.0	1.00	120	1.69099
8.0	0.99	120	1.69786
8.0	1.00	100	1.71363
6.0	0.96	120	1.71674
8.0	1.01	120	1.71711
6.0	0.97	120	1.72060
6.0	0.97	100	1.72069
8.0	1.01	100	1.72650
4.0	0.93	120	1.72790
6.0	0.96	100	1.72814
4.0	0.93	100	1.72884
8.0	0.99	100	1.72993
8.0	0.98	120	1.73627
4.0	0.94	100	1.74421
4.0	0.94	120	1.74899
6.0	0.95	120	1.76269
8.0	1.02	100	1.76664
6.0	0.98	100	1.76722





previous calculation (without ε'_{out})

Eps(in)	sigma	Eps'(out)	RMSD
4.0	0.93	---	1.77

Acknowledgement

- Thank Barry Honig for the continuous support
- Thank Chuan Li and Prof. Emil Alexov and our lab (Computational Biophysics and Bioinformatics Lab, Clemson University)
- Thank the work from Anthony Nicholls‘s group which is published on J.C.C. 2001
- Grant: NIH R01 GM093937

Thank you!