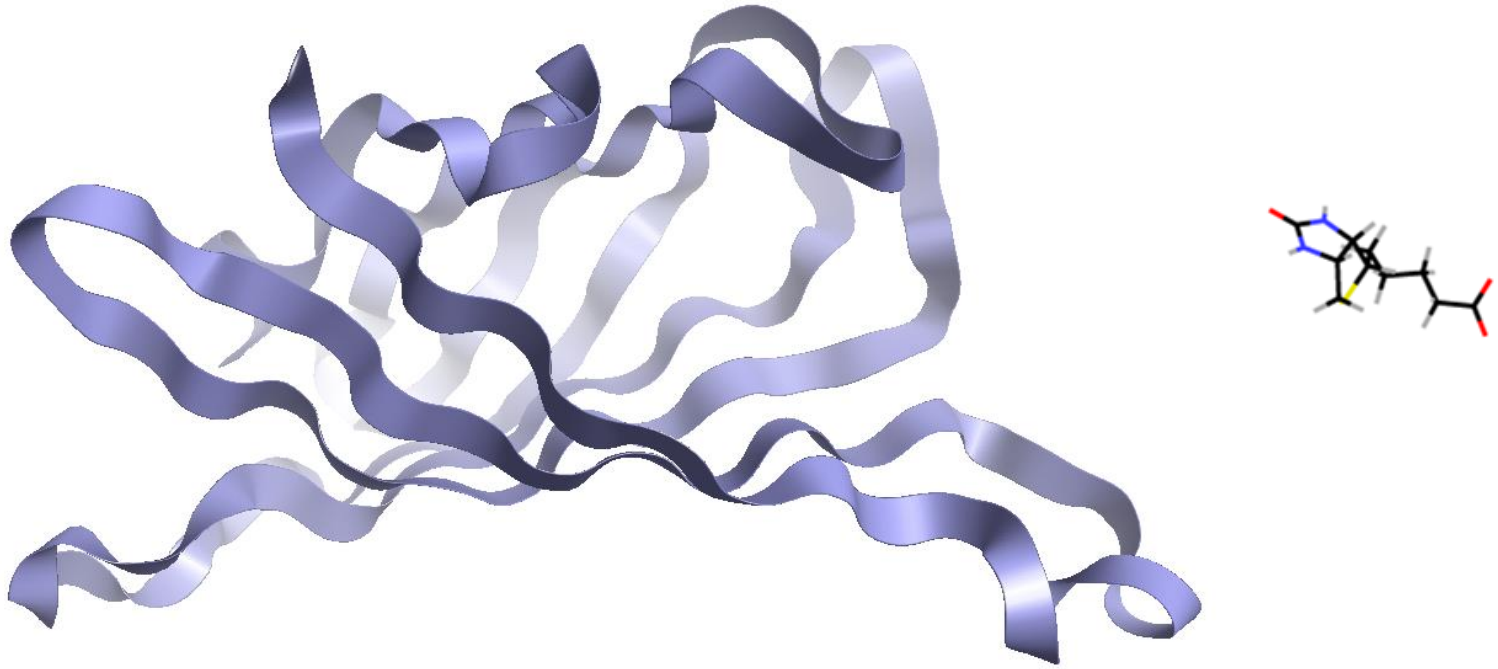
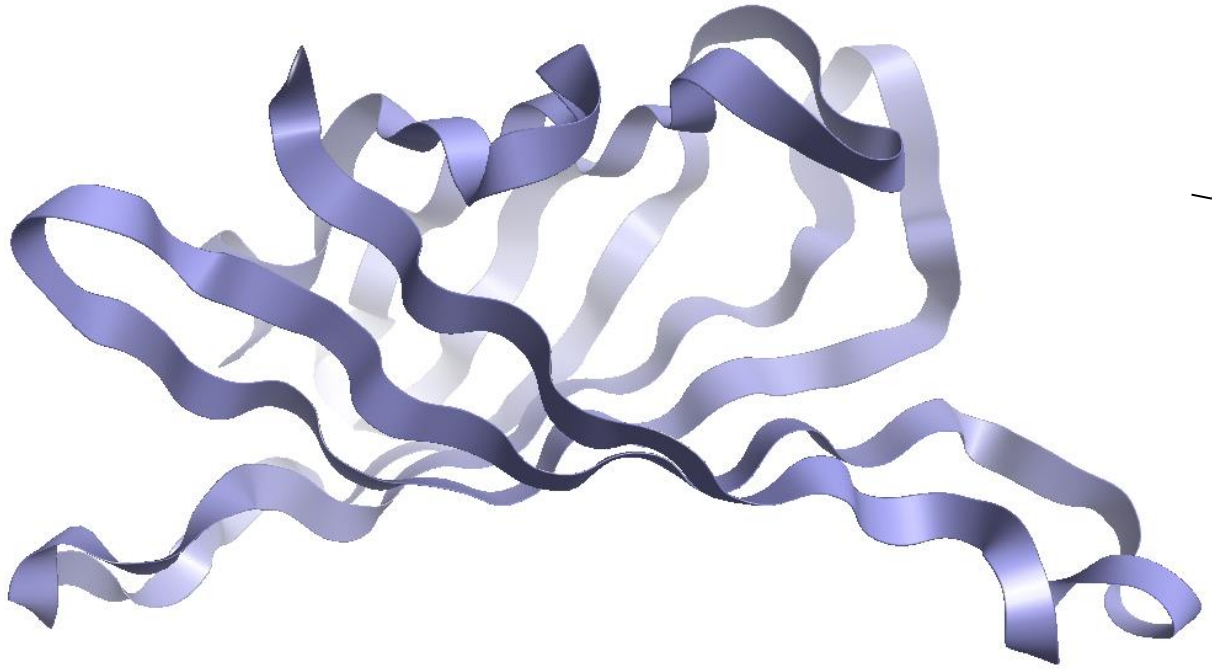
3D-RISM – effects of improved electrostatic models

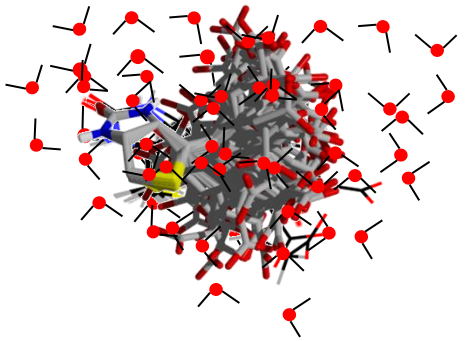
Paolo Tosco

Ligand binding to a protein

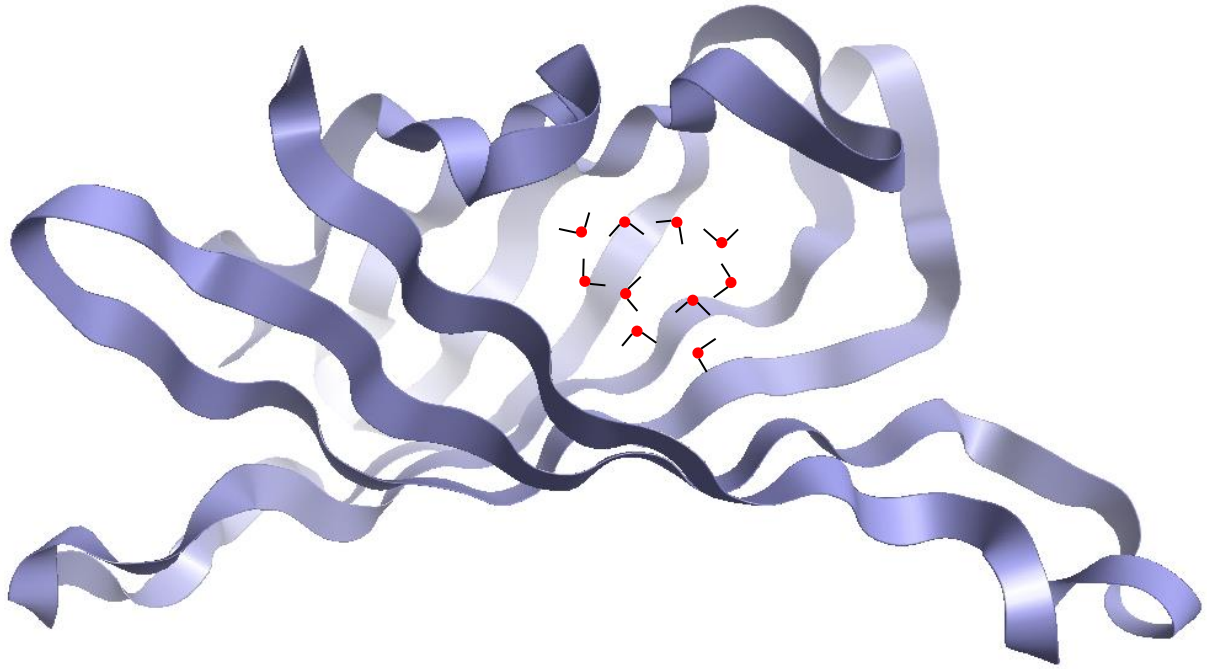


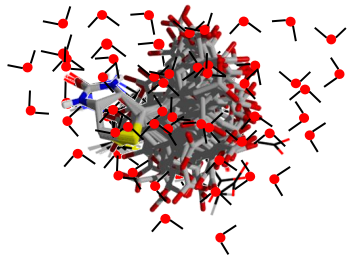
Oops, forgot about conformational entropy and desolvation!

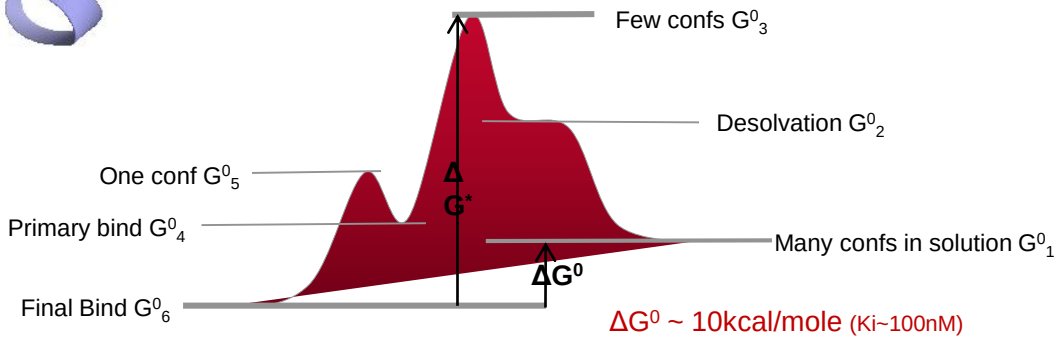




Ligand binding is dominated by solvation effects







Water in proteins

- Every ligand binding event displaces water from the protein
  - How many waters?
  - Which ones?
  - How much did that cost (or gain) in ΔG?

3D-RISM

> Analytical method for working out where water goes (Ornstein-Zernike equation)

> Conceptually equivalent to running an infinite-time MD simulation on the solvent and extracting the solvent particle densities

$$h(r_{12}) = c(r_{12}) + \int dr_3 c(r_{13})\rho(r_3)h(r_{23})$$

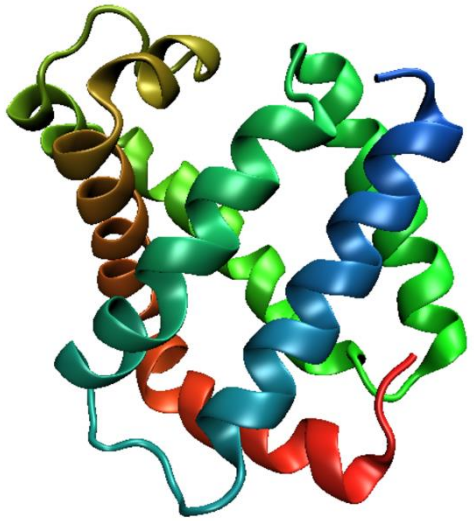
Total correlation function
'What is the distribution of solvent around the solute?'

Direct correlation function
'How does a solvent molecule interact with the solute?'

Indirect influence through all possible chains of mediating third particles
'What is the effect of a solvent molecule interacting with another solvent molecule which is interacting with the solute?'

3D-RISM

> Analytical method for working out where water goes (Ornstein-Zernike equation)
> Conceptually equivalent to running an infinite-time MD simulation on the solvent and extracting the solvent particle densities
> Output is grid containing particle densities (for water, O and H densities)
> Thermodynamic analysis to assign 'happiness' to each position on the grid



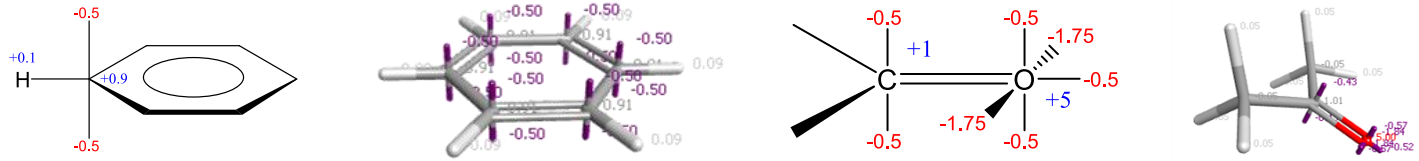
$$
\begin{aligned}
\Delta c^{\mathrm{PSE}-n} &= \Delta\mu^{\mathrm{PSE}-n} - \delta_T\Delta\mu^{\mathrm{PSE}-n} \\
&= -kT\sum_{\gamma}\rho_{\gamma}\int h_{\gamma}(\mathbf{r})\delta_T h_{\gamma}(\mathbf{r}) - \delta_T c_{\gamma}(\mathbf{r}) \\
&\quad -\frac{1}{2}[\{\delta_T h_{\gamma}(\mathbf{r})\}c_{\gamma}(\mathbf{r}) + h_{\gamma}(\mathbf{r})\delta_T c_{\gamma}(\mathbf{r})] \\
&\quad -\frac{t_{\gamma}^{*n}(\mathbf{r})}{n!}[\beta u(\mathbf{r}) + \delta_T h_{\gamma}(\mathbf{r}) - \delta_T c_{\gamma}(\mathbf{r})]\Theta(h_{\gamma}(\mathbf{r}))\mathrm{d}\mathbf{r}.
\end{aligned}
\tag{10}
$$

Problems

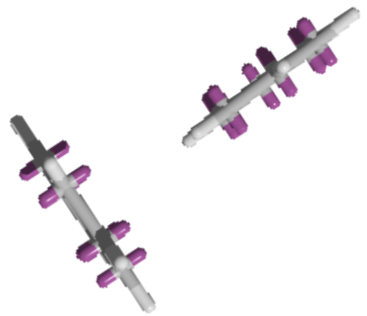
- Fixed solute
  - No accounting for protein movement
- Can’t solve equations exactly
  - Need to use a 'bridge function' – unclear what the correct functional form is
- Total solvation ΔG values only have moderate accuracy
  - 3D-RISM gives a poor estimate of the cavity creation term, so you have to apply parameterised correction factors
  - However, we are interested in the relative partitioning of the solvation ΔG, so this error can be neglected
- Results depend on the interaction potential $U(r)$ used by the closure function
  - In practise, this means vdW + electrostatics
  - Results only as good as your potential functions
- Can an improved description of electrostatics give better results?

Electrostatics from Molecular Mechanics

- XED force field – eXtended Electron Distribution
  - Multipoles via additional monopoles



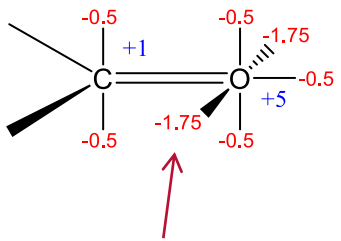
  - Hückel
    - separation of π and σ charges – substituent effects
    - find bond orders and assign hybridization – analogue N atoms
  - Full MM Force Field with excellent coverage of organic chemistry and proteins
    - Minimization, Conformations etc.
    - Additional atoms cost more than ACC
    - Cheaper than other multipole methods
    - Local polarization
    - In development for >20 years



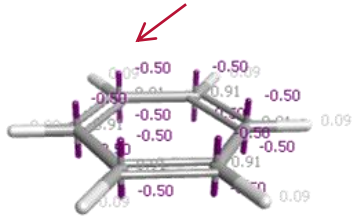
'Extended Electron Distributions Applied to the Molecular Mechanics of some Intermolecular Interactions', J.G. Vinter, J. Comput.-Aided Mol. Des., 8, 653-668, 1994

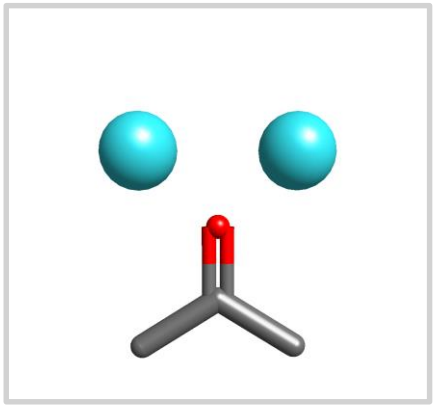
Detailed Electrostatics from XED

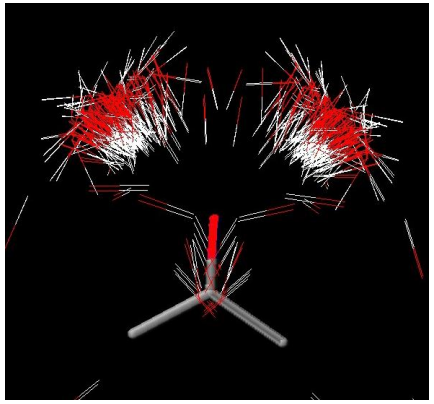
> eXtended Electron Distribution designed to give detailed electrostatic interaction patterns



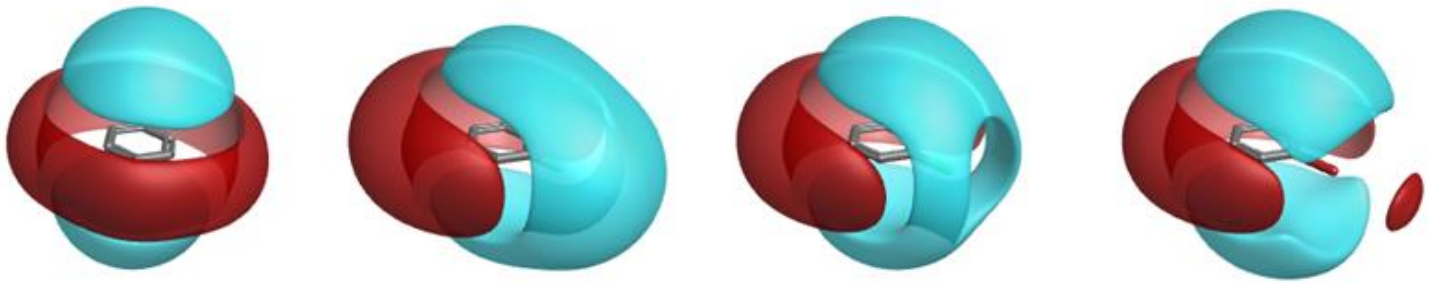
XED adds extra charges to get detailed representation of atoms







Interaction of Acetone and Any-OH from small molecule crystal structures

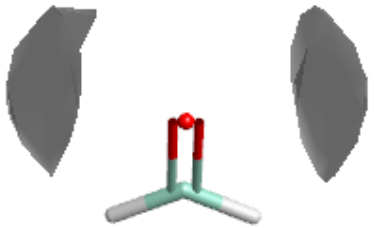


benzene fluorobenzene chlorobenzene bromobenzene

Force field comparison

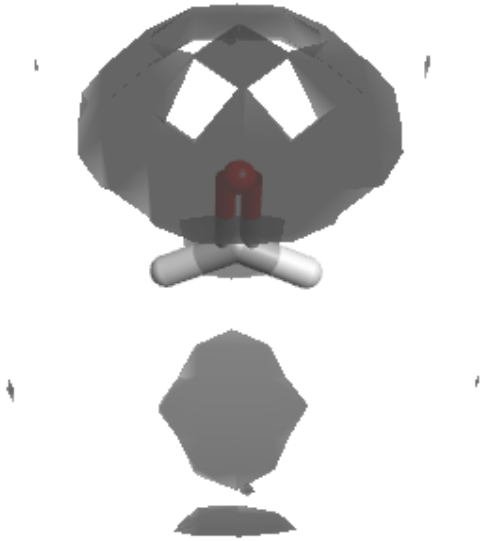
- The most commonly-used force field for 3D-RISM calculations on proteins is AMBER
- Compare XED charge model to the AMBER/GAFF AM1/BCC charge model

Comparing XED with GAFF – Hydrogen Density



formaldehyde_x:1

XED

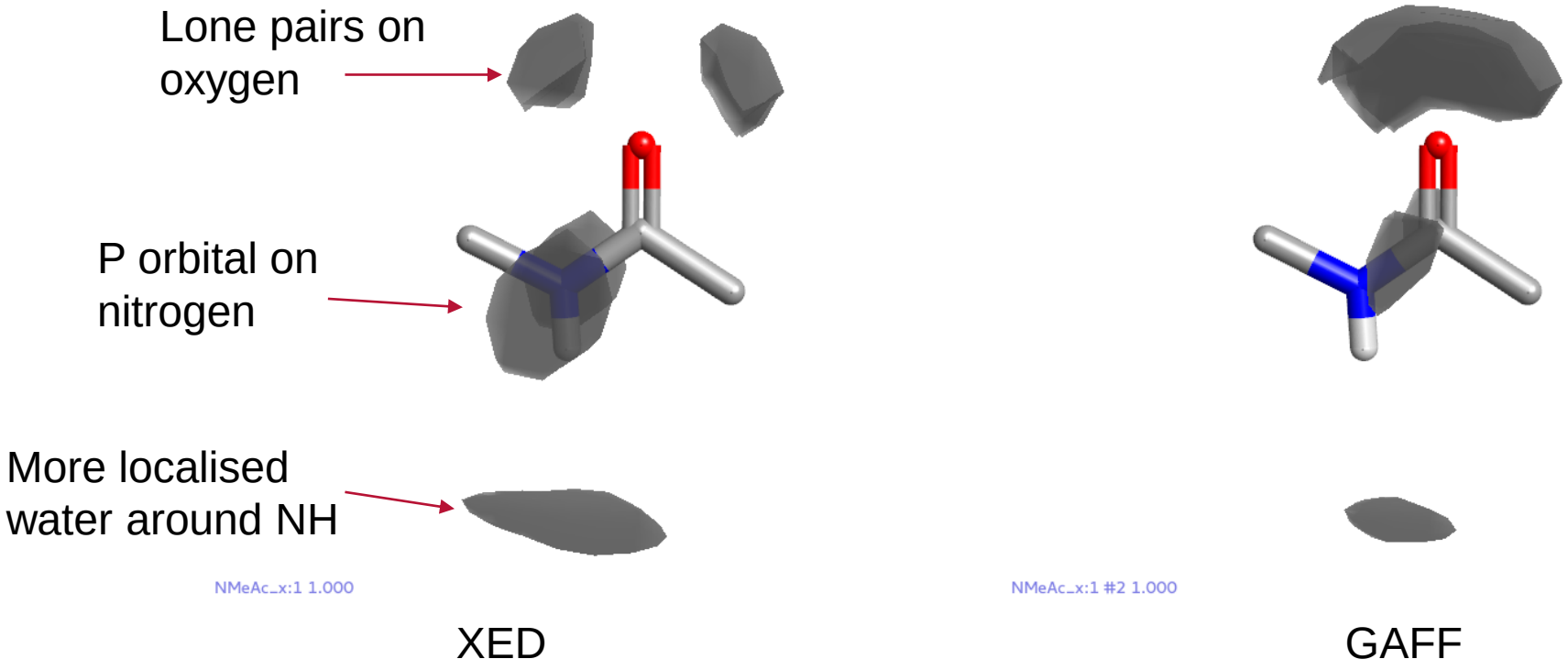


MOL 1.000

GAFF

Symmetric distribution around oxygen – no lone pairs!

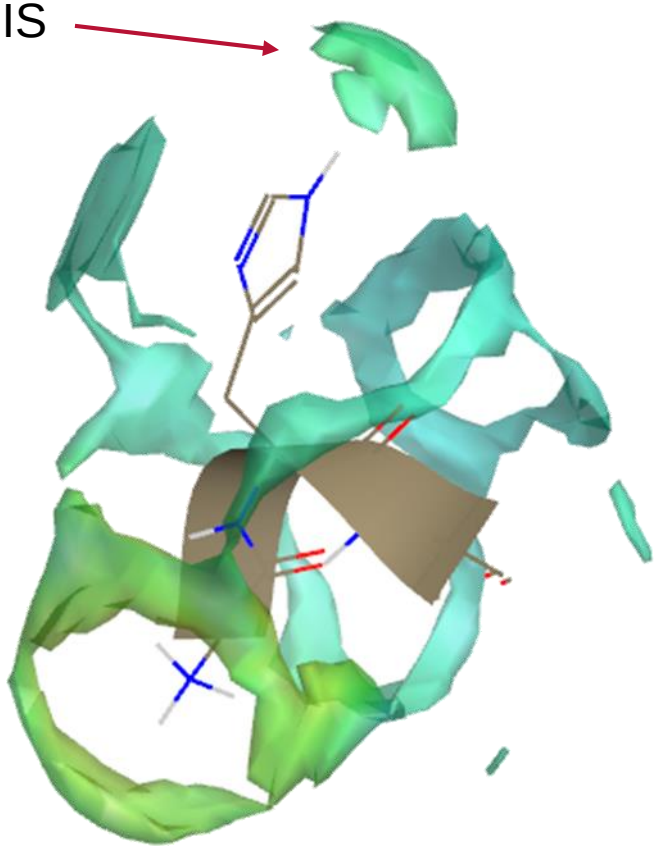
Comparing XED with GAFF – Hydrogen Density



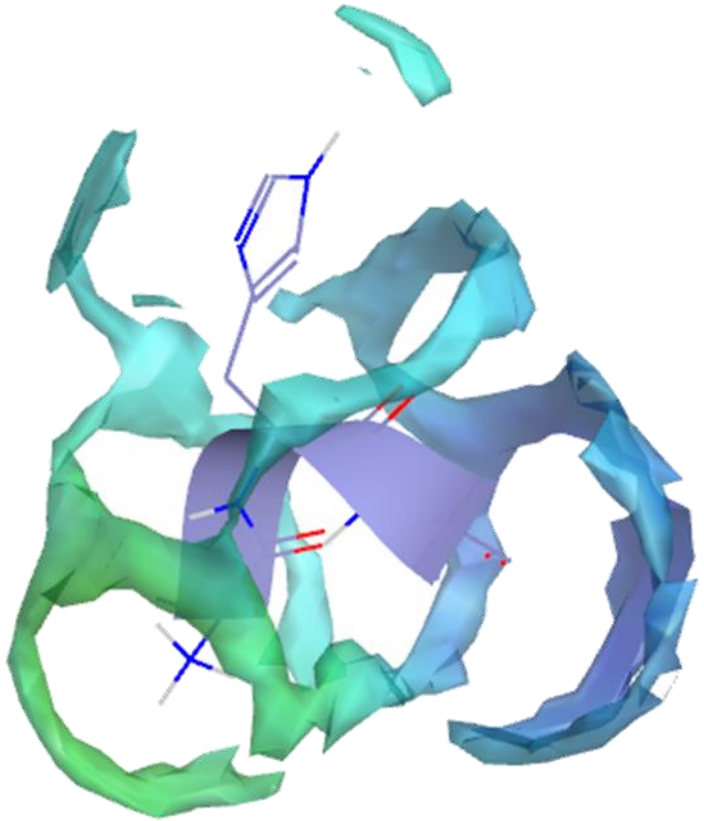
XED GAFF

More complex systems: O density

More polarization
of imidazole in HIS



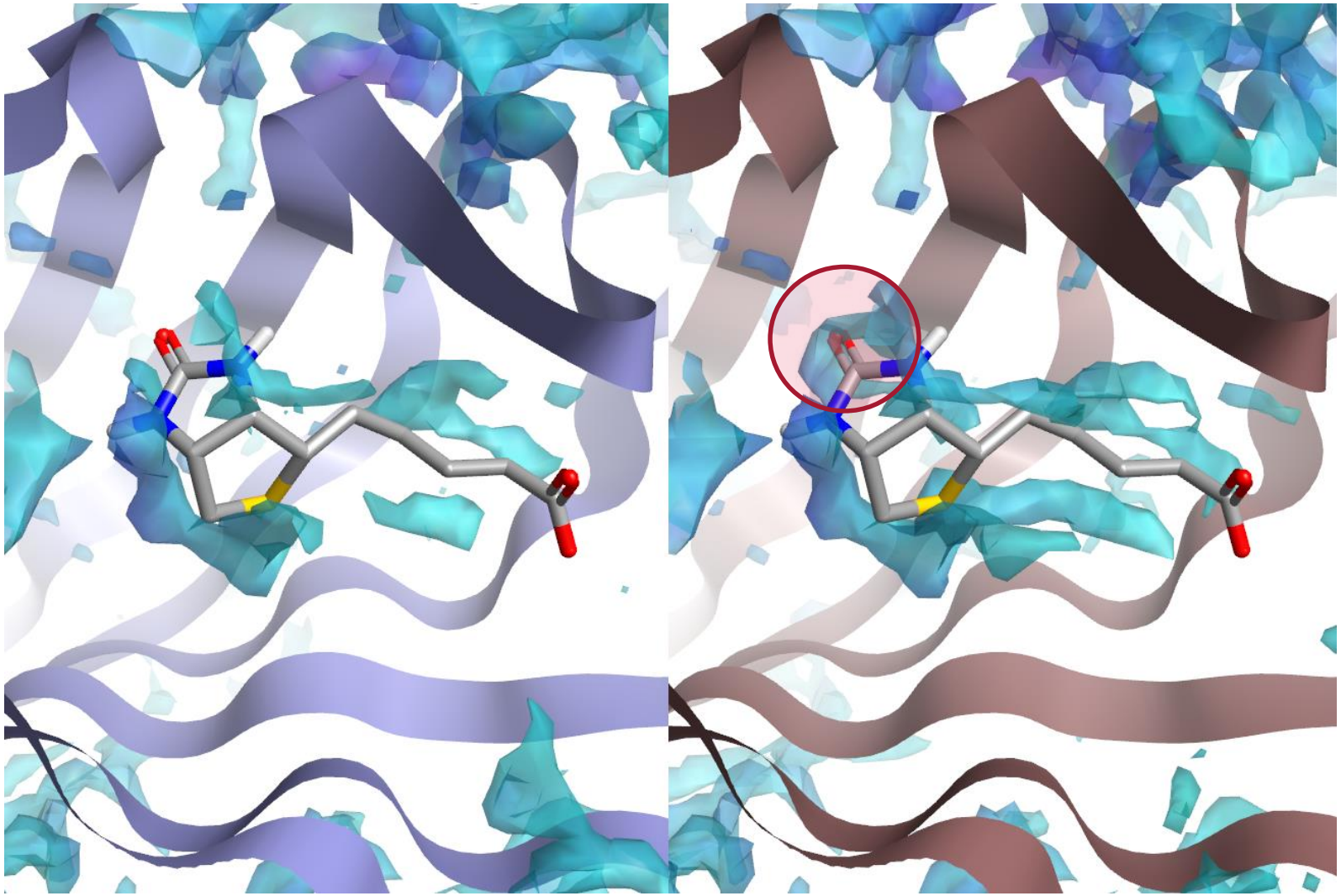
XED



AMBER

Extend to proteins – biotin/streptavidin

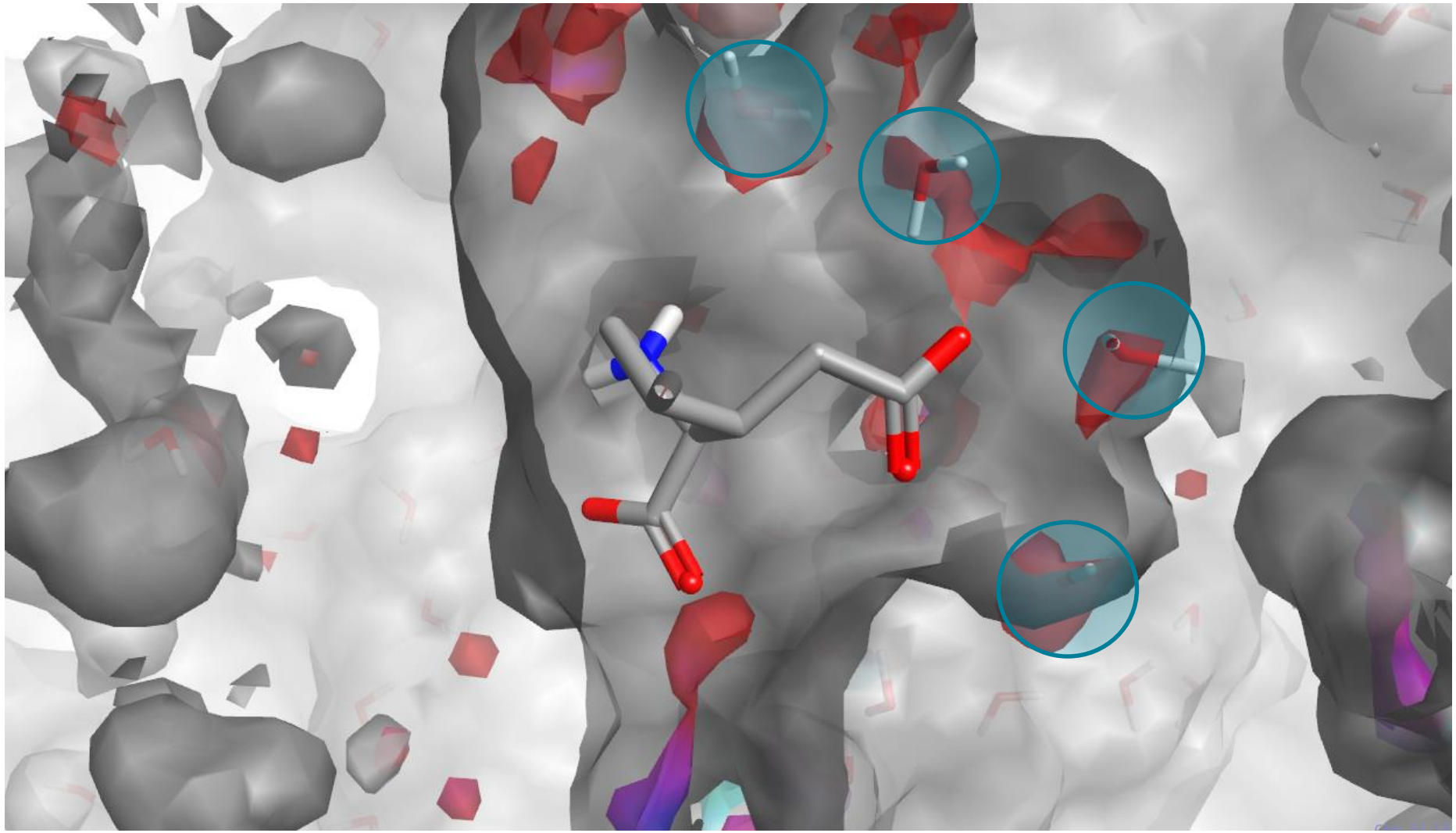
XED
unfavorable
water



AMBER
unfavorable
water

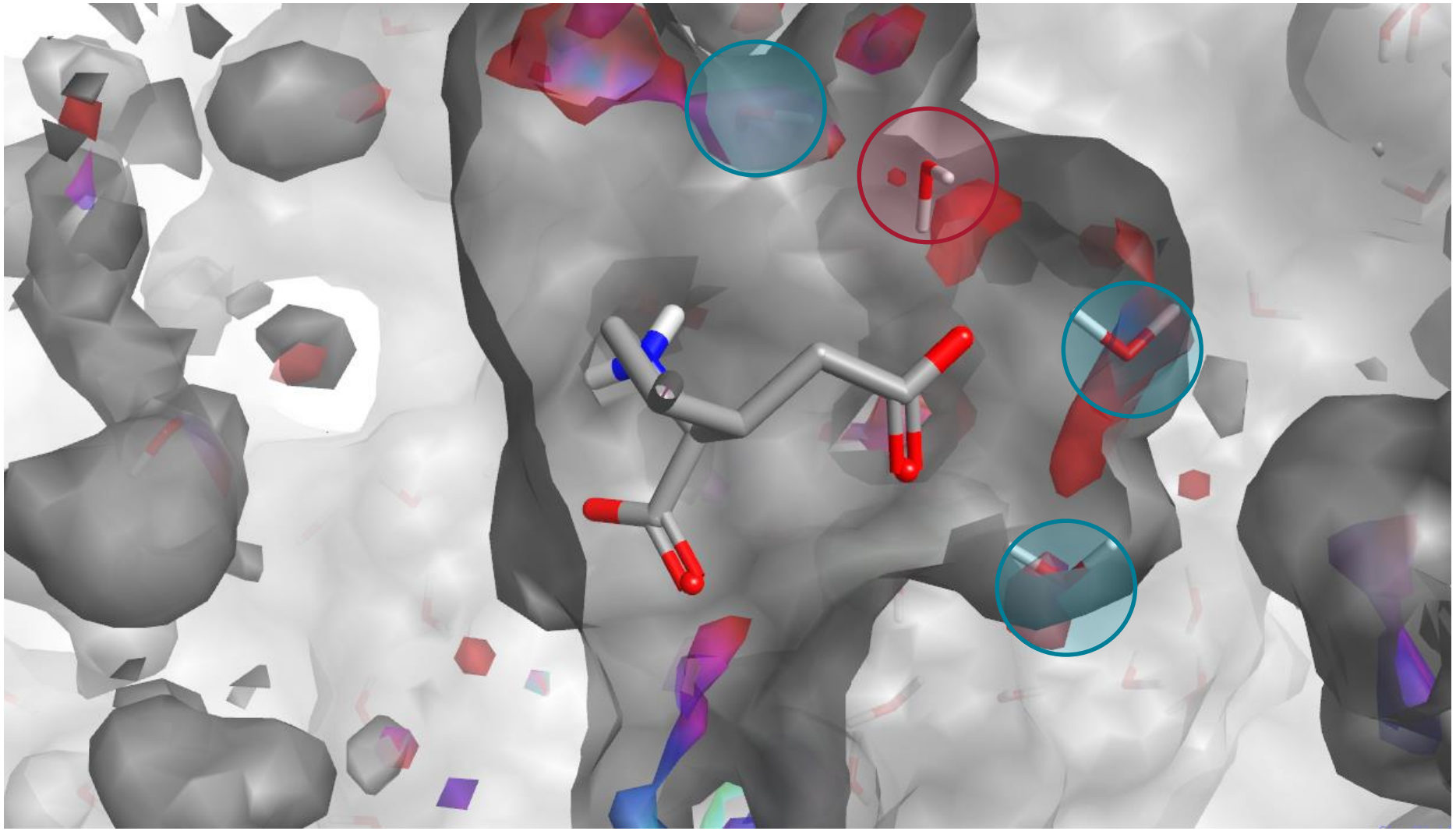
RISM density match to experimental waters - XED

1TT1



RISM density match to experimental waters - AMBER

1TT1



Validation

> Positional validation looks good (RISM density matches exptl water positions)

Validation

- Positional validation looks good (RISM density matches exptl water positions)
- Energetic validation is difficult
  - solvation energy partitioning is not an experimental observable

Validation

- Positional validation looks good (RISM density matches exptl water positions)
- Energetic validation is difficult
  - solvation energy partitioning is not an experimental observable
- Partial validation against QM results?

Compare 3D-RISM water energetics to QM values

- Run ONETEP calculations (linear-scaling DFT) on several proteins
  - Single explicit water molecules in implicit solvent
- Results proved very difficult to interpret due to complex protein environment
- Look at a few model systems instead

International Journal of Quantum Chemistry WWW.Q-CHEM.ORG FULL PAPER

Large-Scale DFT Calculations in Implicit Solvent—A Case Study on the T4 Lysozyme L99A/M102Q Protein

Jacek Dziedzic,[a]† Stephen J. Fox,[a] Thomas Fox,[b] Christofer S. Tautermann,[b] and Chris-Kriton Skylaris*[a]

Recently, variants of implicit solvation models for first principles electronic structure calculations based on a direct solution of the nonhomogeneous Poisson equation in real space have been developed. These implicit solvation models are very elegant from a physical point of view as the solute cavity is defined directly via isosurfaces of the electronic density, and the molecular charge is polarized self-consistently by the reaction field of the dielectric continuum which surrounds the solute. Nevertheless, the implementation of these models is technically complex and requires great care. A certain level of care is required from users of such models as a number of numerical parameters need to be given appropriate values to obtain the most accurate and physically relevant results. Here, we describe in what parts of the solvent model each of these numerical parameters is involved and present a detailed study of how they can affect the calculation, using the solvation model which has been implemented in the ONETEP program for linear-scaling density functional theory (DFT) calculations. As ONETEP is capable of DFT calculations with thousands of atoms, we focus our investigation of the numerical parameters with a case study on protein–ligand complexes of the entire 2602-atom T4 Lysozyme L99/M102Q protein. We examine effects on solvation energies and binding energies, which are critical quantities for computational drug optimization and other types of biomolecular simulations. We propose optimal choices of these parameters suitable for routine "production" calculations. © 2012 Wiley Periodicals, Inc.

DOI: 10.1002/qua.24075

Introduction

Chemistry, biochemistry, and materials and interfacial processes typically take place in and require the presence of solvent. Therefore, simulations at the atomic level must include a description of the solvent. Implicit solvent models, which describe the solvent as a dielectric continuum, have proved very effective in this task and have been an active area of research with many improvements over the years, both within atomistic classical force field simulation methods and in first principles quantum chemistry methods. These models are particularly effective in the context of quantum chemistry calculations, as the reaction field of the dielectric is included directly in the Hamiltonian operator and polarizes the density during the self-consistent solution of the quantum mechanical model. Notable variants of such self-consistent implicit solvation models are the polarizable continuum model (PCM) of Tomasi and coworkers,[1] the COSMO model[2] as well as the very accurate but heavily parameterized SMD model of Truhlar and coworkers,[3] which is founded in the integral equation formalism[4] of the PCM model. Although the physical principles on which these models are based are very elegant, the actual implementation can depend on a large number of parameters which need careful determination by fitting to experimental or theoretical data.

Recently, Fattebert and Gygi[5] proposed a new model of continuum solvation, where the dielectric is defined as a functional of the electronic density of the solute. This model was further extended by Scherlis *et al.*[6] to include the calculation of the cavitation energy, by defining it in terms of the quantum surface of the solute. This model is particularly attractive, as it retains the elegance of the implicit solvent philosophy, as the reaction field is obtained by direct solution of the nonhomogeneous Poisson equation (NPE) in real space:

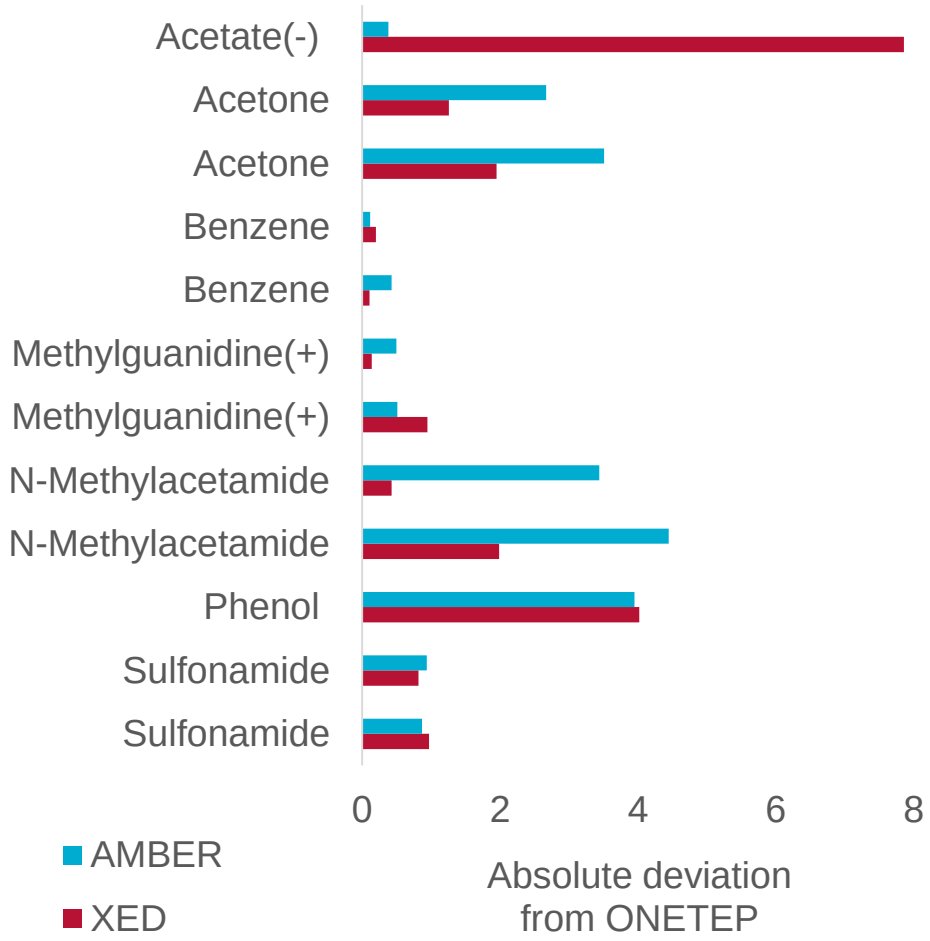
$$\nabla \cdot (\varepsilon[\rho]\nabla\phi_{\mathrm{NPE}}(\mathbf{r})) = -4\pi\rho_{\mathrm{tot}}(\mathbf{r}), \qquad (1)$$

where $\rho(\mathbf{r})$ is the electronic density and $\rho_{\mathrm{tot}}(\mathbf{r})$ is the total density due to electrons and nuclei (or ionic cores in the case of pseudopotentials). Despite this, results obtained with this model in its original formulation were reasonable but significantly less accurate than the conventional approaches such as PCM, especially for charged molecules. We have recently shown[7] how this limitation can be overcome using appropriate boundary conditions, including dispersion interactions with the solvent and redetermining appropriately the two parameters in the functional $\varepsilon[\rho]$. The solvent model by Dziedzic *et al.* has been validated on an extensive set of more than 130 molecules (a representative selection of 20 neutral, 20 cationic, and 20 anionic molecules from Ref. [8], and 71 larger neutral molecules from Refs. [9, 10]) and produces solvation energies that agree with experimental

[a] J. Dziedzic, S. J. Fox, C.-K. Skylaris
School of Chemistry, University of Southampton, Highfield, Southampton SO17 1BJ, United Kingdom
E-mail: c.skylaris@soton.ac.uk
[b] T. Fox, C. S. Tautermann
Lead Identification and Optimization Support, Boehringer Ingelheim Pharma GmbH & Co. KG, 88397 Biberach, Germany
† Also at Faculty of Technical Physics and Applied Mathematics, Gdansk University of Technology, Poland

Comparison of water interaction energies

System	ID	ΔE (kcal/mol)		
		ONETEP	XED	AMBER
Acetate(-)	1.X	-7.39	-15.25	-7.77
Acetone	2.A	-3.65	-1.70	-0.14
	2.X	-3.81	-2.55	-1.14
Benzene	2.A	-1.01	-0.90	-0.58
	2.X	-0.72	-0.52	-0.60
Methylguanidine(+)	3.A	-0.87	-1.82	-1.38
	4.A	-0.45	-0.31	-0.95
***N*-Methylacetamide**	2.A	-4.94	-2.95	-0.49
	2.X	-4.30	-3.87	-0.86
Phenol	2.X/2.A	-5.22	-1.20	-1.27
Sulfonamide	2.X	-1.88	-0.91	-1.01
	4.X	-0.12	0.70	0.82



RISM with XED conclusions

- Water patterns around small molecules look better with XED
- In proteins, XED provides better water patterns for most cases
  - A few limitations: it seems to over-polarise charged residues
- Validation of energetics is difficult – no direct experimental observables

Acknowledgements

- Cresset
  - Tim Cheeseright
  - Mark Mackey
  - Susana Tomásio
- Southampton
  - Chris Skylaris
  - Max Phipps