Design of reduced point charge models for proteins

Leherte, Laurence; Vercauteren, Daniel P.

Publication date:
2014

Document Version
Peer reviewed version

Design of reduced point charge models for proteins

Laurence Leherte, Daniel P. Vercauteren
University of Namur – Belgium

ACS San Francisco – August 2014

Introduction

We report reduced point charge models (RPCMs) for proteins obtained from topological analyses of smoothed charge density (CD) and electron density (ED) distribution functions. The RPCMs differ from the well-known approximation of a single point charge per amino acid ($q = 0, \pm 1$). The applicability of various RPCMs in approximating all-atom electrostatic properties and in molecular dynamics (MD) calculations is assessed. Applications are reported for the protein Ubiquitin.

Method

1. Smoothing of the electron density

$$\rho_A(r) = \sum_{a\in A}\sum_{i=1}^{3\text{or}5} \alpha_{a,i} e^{-\beta_{a,i} r^2}$$

Unsmoothed molecular ED generated using the Promolecular Atomic Shell Approximation (PASA) [1]

$$\rho_{A,s}(r) = \sum_{a\in A}\sum_{i=1}^{3\text{or}5} \alpha_{a,i,s} e^{-\beta_{a,i,s} r^2}$$

Smoothed molecular ED (s = smoothing factor) [2]

with $\alpha_{a,i,s} = Z_a w_{a,i}\left(\frac{2\varsigma_{a,i}}{\pi}\right)^{3/2} \frac{1}{(1+8\varsigma_{a,i}s)^{3/2}}$ and $\beta_{a,i,s} = \frac{2\varsigma_{a,i}}{(1+8\varsigma_{a,i}s)}$

The smoothing factor s can be seen as the product of a diffusion coefficient D with time t $Dt \approx s$ or as an overall isotropic thermal displacement factor $B \approx s$ [3].

2. Smoothing of the Coulomb potential

$$V_A(\mathbf{r}) = \sum_{a\in A} \frac{q_a}{|\mathbf{r}-\mathbf{R}_a|}$$

Unsmoothed molecular electrostatic potential (MEP). Amber99 atomic charges [4] are assigned to atoms using PDB2PQR [5].

$$V_{A,s}(\mathbf{r}) = \sum_{a\in A} \frac{q_a}{|\mathbf{r}-\mathbf{R}_a|} erf\left(\frac{|\mathbf{r}-\mathbf{R}_a|}{2\sqrt{s}}\right)$$

Smoothed MEP [6]

The Poisson equation is applied to generate the corresponding smoothed atomic charge density (CD) distribution function, $\rho_{a,s}$:

$$-\nabla^2 V_{a,s}(\mathbf{r}) = \frac{\rho_{a,s}}{\varepsilon_0} \longrightarrow \frac{\rho_{a,s}}{\varepsilon_0} = \frac{q_a}{(4\pi s)^{3/2}} e^{-r^2/4s}$$

3. Examples

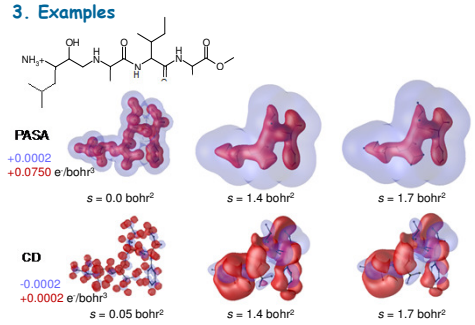


Fig. 1 Smoothed ED (Top) and CD (Bottom) of a peptide-like molecule.

4. Location of Critical Points (CP) in $\rho_{A,s}$

A hierarchical merging algorithm, based on the idea of Leung *et al.* [7], is used to locate local extrema in $\rho_{A,s}$.
- At scale s = 0, each atom of a molecular structure is considered as a starting point of the merging procedure.
- As s increases, each point moves along a gradient path to reach a location in the 3D space where:

$$|\nabla\rho_{CP}(s)| = 0$$

These trajectories are defined by:

$$\mathbf{r}_{CP}(s) = \mathbf{r}_{CP}(s-\Delta s) + \frac{\Delta}{\rho_{CP}(s-\Delta s)} \nabla\rho_{CP}(s)$$

Δ = displacement step

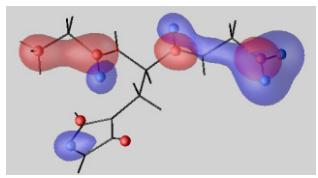


Fig. 2 Isocontours of the CD of Gly-Hisδ -Gly smoothed at s = 1.7 bohr². -0.005 +0.005 e⁻/bohr³

- Charges are fitted at the critical points of
  - the CD smoothed at 1.7 bohr²
  - the PASA ED smoothed at 1.4 bohr²

vs. unsmoothed Amber99 MEPs, considering various amino acid rotamers [8], with constraints: total electric charge & total dipole moment.
- Side chain charges are first assigned [9,10], then backbone charges are fitted using the side chain charge values as constraints.

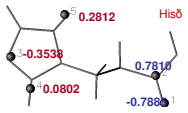


Fig. 3 Reduced point charge model 'mCD' of Hisδ.

Templates are thus obtained for each amino acid residue. A third model, 'mCDa', is similar to 'mCD' but most of the point charges are forced to be located on atoms.

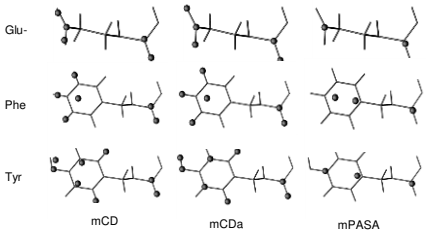


Fig. 4 Reduced point charge models of three amino acid residues.

5. Application to Ubiquitin

1UBQ.pdb - 76 amino acids - 1231 atoms

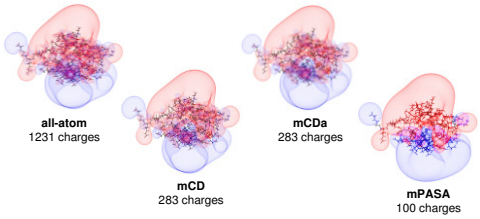
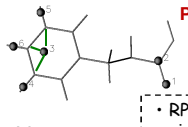


Fig. 5 MEP isocontours of Ubiquitin: -0.05, +0.05 e⁻/bohr.

Molecular dynamics of Ubiquitin

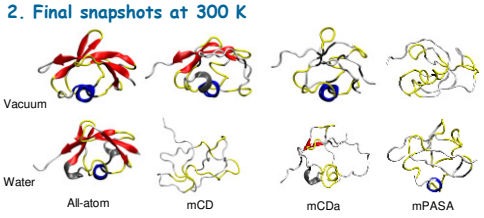
1. Simulation conditions

Gromacs 4.5.5 [11]
Amber99SB and TIP4P-Ew force fields, PME
Non-atomic point charges = virtual sites
Other force field terms are preserved → no RPCM-to-all-atom conversion is needed.

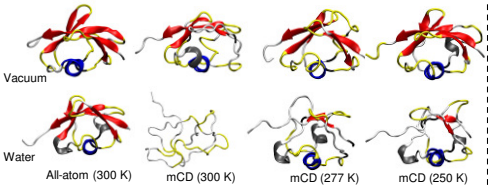


Equilibration : 30 ns
Production : 10 ns
NVT (vacuum), NPT (water)
T = 300 (277 and 250 K)

2. Final snapshots at 300 K



3. Effect of temperature



4. Correlation E_{RPCM} – $E_{all\text{-}atom}$

- Acceptable correlation values obtained for mCD and mCDa.
- Better energy agreement for the mCDa model which allows a better representation of the $Coulomb_{1\text{-}4}$ energy term.
- mPASA has no such terms.

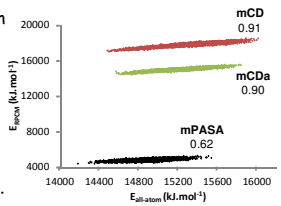


Fig. 6 Potential energy of the RPCM *vs.* all-atom models for identical protein conformations.

5. Intra-molecular H bonds

Distance and angle distributions present the same trends both in vacuum and in water. Angle distribution is strongly affected when a RPCM is used.

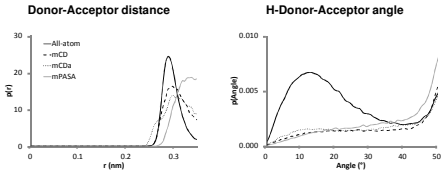


Fig. 7 Distance and angle distributions of intra-molecular H-bonds calculated for solvated Ubiquitin.

6. Protein-water interface

Distance and angle distributions present trends similar to the all-atom case, except for the coarser RPCM, *i.e.*, mPASA.

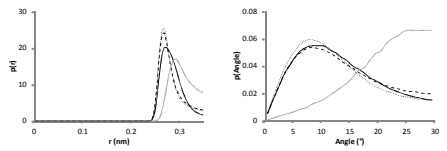


Fig. 8 Distance and angle distributions of Ubiquitin-water H-bonds.

Table 1. Selected properties of water molecules located in a layer of thickness 0.35 nm surrounding Ubiquitin.

	All-atom (300 K)	mCD (300 K)	mCDa (300 K)	mPASA (300 K)	mCD (277 K)	mCD (250 K)
# of H_2O	358.5	546.1	505.7	329.3	482.7	450.7
% of H-bonded H_2O	83	58	62	28	56	57
D (10^{-5} cm²/s)	1.28	0.64	0.55	1.70	0.34	0.10

Conclusions

- RPCMs allow the approximation of the MEP of rigid proteins. They also allow simulations of flexible structures by MD provided (i) they involve a good description of the $Cb_{1\text{-}4}$ energy terms, (ii) the coarse-grain level is not too reduced, *i.e.*, a dipole moment is preserved both on the backbone and the side chain of the residues.
- Secondary structure elements are destroyed under the influence of water or at higher temperatures. This, in combination with previous studies [12] may indicate a lowering of energy barriers (useful for conformational sampling).
- RPCMs involve modifications for the interface water structure and dynamics.

[1] Amat *et al.* J. Chem. Inf. Chem. Sci. 40 (2000) 1188; [2] Kostrowicki *et al.* J. Phys. Chem. 95 (1991) 4113; [3] Leherte Acta Cryst D60 (2004) 1254; [4] Duan *et al.* J. Comput. Chem. 24 (2003) 1999; [5] pdb2pqr.sourceforge.net; [6] Hart *et al.* J. Comput. Chem. 21 (2000) 531; [7] Leung *et al.* IEEE T. Pattern Anal. 22 (2000) 1396; [8] Simms *et al.* Prot. Eng. Des. Select. 21 (2008) 369, www.dynameomics.org; [9] Leherte *et al.* J. Chem. Theory Comput. 5 (2009) 3279; [10] Leherte *et al.* J. Comput.-Aided Mol. Des. 25 (2011) 913; [11] Pronk *et al.* Bioinformatics 29 (2013) 845; [12] Leherte *et al.* J. Mol. Graphics Model. 47 (2014) 44