Determination of protein coarse-grain charges from smoothed molecular electrostatic potentials

Leherte, Laurence; Vercauteren, Daniel

Publication date:
2008

Document Version
Peer reviewed version

Determination of protein coarse grain charges from smoothed molecular electrostatic potentials

Laurence Leherte, Daniel P. Vercauteren
Laboratoire de Physico-Chimie Informatique, Groupe de Chimie Physique Théorique et Structurale,
University of Namur (FUNDP)

Introduction

The design of coarse grain (*CG*) models [1] and their corresponding potential functions [2] for protein computational studies is currently an active field of research, especially in solving long-scale dynamics problems such as protein folding, protein-protein docking, … For example, to eliminate fast degrees of freedom, it has been shown that one can rely on *CG* representations only, or on mixtures of *CG* and more detailed descriptions [3,4] in order to significantly increase the time step in molecular dynamics (MD) simulations. Among the parameters involved in *CG* potentials, the electrostatic interactions are of major importance [5] since they govern local and global properties such as their stability [6], their flexibility [7], …

In this poster, we present an approach to design and evaluate *CG* electrostatic point charges from smoothed molecular electrostatic potentials (MEP). In a previous approach [8], electron density (ED)-based *CG* were determined through a merging/clustering procedure of atom trajectories generated in progressively smoothed ED distribution functions. In the present work, atoms are clustered according to their trajectories defined in a smoothed MEP function, more particularly the Amber potential reported in [9]. A fitting algorithm is applied to evaluate *CG* charges.

1. Location of *CG* points

A hierarchical merging algorithm, based on the idea of Leung *et al.* [10], is used to locate local maxima and minima in a MEP function *V*, as a function of the degree of smoothing *t*:

1. At scale $t = 0$, each atom of a molecular structure is considered as a local maximum or minimum of *V*. All atoms are thus considered as the starting points of the merging procedure.
2. As t increases from ~ 0.0 to a given maximal value, each point moves continuously along a gradient path to reach a location in the 3D space where:

$$\vec{\nabla} V(t) = 0$$

On a practical point of view, this consists in following the trajectory of the points within the MEP function calculated at *t* according to Equation:

$$\vec{r}_{V(t)} = \vec{r}_{V(t-\Delta t)} + \frac{\Delta}{V(t)} \vec{\nabla} V(t)$$

2. Molecular electrostatic potential

$$V_A(\vec{r}) = \sum_{a \in A} \frac{Z_a}{\left|\vec{r} - \vec{R}_a\right|}$$

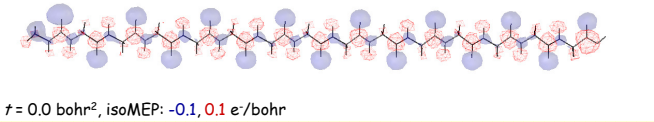
$$V_{A,t}(\vec{r}) = \sum_{a \in A} \frac{Z_a}{\left|\vec{r} - \vec{R}_a\right|} erf\left(\frac{\left|\vec{r} - \vec{R}_a\right|}{2\sqrt{t}}\right)$$

3. Determination of *CG* charges

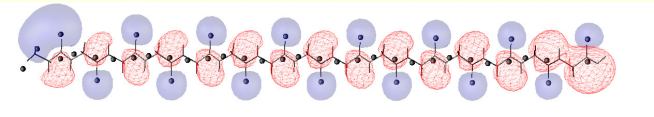
This is achieved through the program QFIT [11] to get *CG* point charges fitted from an unsmoothed MEP grid, considering the following constraints: the total molecular charge and dipole.

4. Determination of backbone *CG* charges

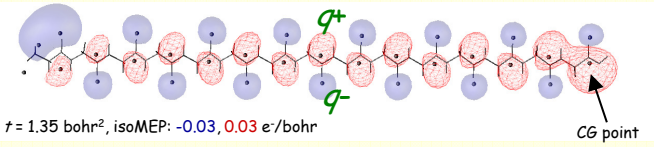
a) Construction of Gly_{15} in an extended conformation (Ω = 180°, Φ = -139°, Ψ = 135°) using SMMP05 [12], a Monte carlo/Simulated Annealing program.
b) Application of the hierarchical merging/clustering algorithm



t = 0.0 bohr², isoMEP: -0.1, 0.1 e⁻/bohr



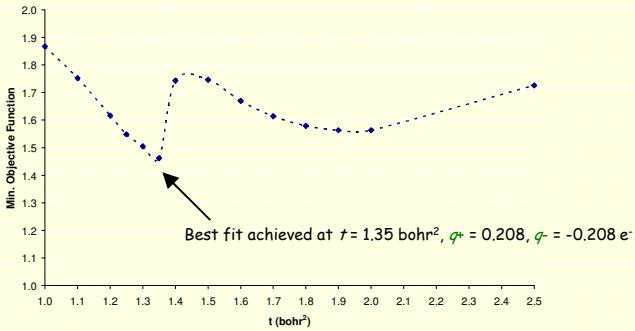
t = 0.95 bohr², isoMEP: -0.03, 0.03 e⁻/bohr



t = 1.35 bohr², isoMEP: -0.03, 0.03 e⁻/bohr

CG point

b) Charge fitting *vs.* unsmoothed MEP as a function of *t*:



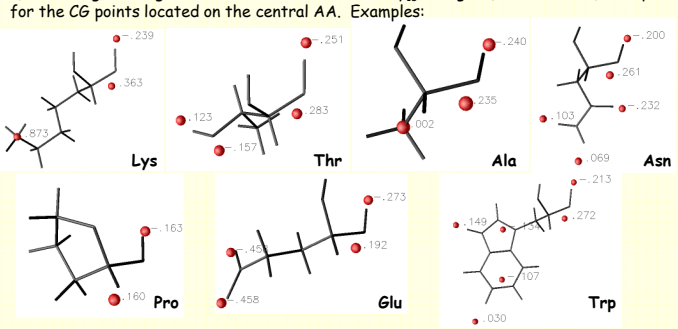
5. Determination of *CG* charges of amino acid (*AA*) side chains

a) Construction of Gly7-*AA*-Gly7 in an extended conformation (Ω = 180°, Φ = -139°, Ψ = 135°) with various *AA* rotamers [13] using SMMP05 [12]. Examples:

	Conformation	χ1 (°)	χ2 (°)	χ3 (°)	χ4 (°)	Occurrence (%)
Arg	*g-, t, g-, g-*	300	180	300	300	9.5
	g-, t, g-, t	300	180	300	180	11.9
	g-, t, g+, t	300	180	60	180	12.2
	g-, t, t, t	300	180	180	180	12.2
Asn	*t, Nt*	180	0			11.1
	t, Og-	180	300			21.3
	t, Og+	180	60			23.6

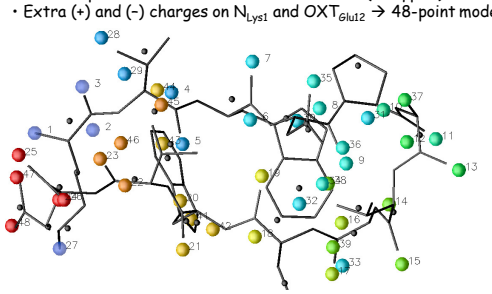
b) Use of *CG* points obtained at $t = 1.35$ bohr² (see sections 1 and 2)

c) *CG* Charge fitting with additional constraints: Gly_{15} charges (see section 3) except for the *CG* points located on the central *AA*. Examples:



6. Application to 12-residue β-hairpin HP7 [14]

- (2+)Lys-Thr-Trp-Asn-Pro-Ala-Thr-Gly-Lys(+)-Trp-Thr-Glu(2-)
- Positioning of *CG* points through QUATFIT, a superposition algorithm [15], using the above templates and the PDB structure of HP7 (2evq.pdb)
- Extra (+) and (-) charges on N_{Lys1} and OXT_{Glu12} → 48-point model



- **48-point model** (color-coded by *AA*)
  rmsd V = 4.63 kcal/mol
  rmsd μ = 5.51 D
- **Fitted 48-point model**
  rmsd V = 1.56 kcal/mol
  rmsd μ = 0.21 D
- **Fitted Basdevant's model [14]** (black spheres)
  rmsd V = 5.45 kcal/mol
  rmsd μ = 1.57 D

7. Conclusions, perspectives

- A *CG* model built from a smoothed MEP
  - seems to be a more significant electrostatic model than a description based on *AA* centers-of-mass as reported in [14]
  - can be derived for any set of point charges (Amber, Gromos, ECEPP, …)
- Transferability has to be confirmed

References

[1] Tozzini, Curr. Opin. Struct. Biol. 2005, 15, 144. [2] Nielsen *et al.* J. Phys.: Condens. Matter 2004, 16, R481. [3] Neri *et al.* Phys. Rev. Lett. 2005, 95, 218102/1. [4] Colombo *et al.* Theor. Chem. Acc. 2006, 116, 75. [5] Vizcarra *et al.* Curr. Opin. Chem. Biol. 2005, 9, 622. [6] Stigter *et al.* Proc. Natl. Acad. Sci. USA 1991, 88, 4176. [7] Kumar *et al.* IBM. J. Res. & Dev. 2001, 45, 499. [8] Leherte *et al.* in The Quantum Theory of Atoms in Molecules - From Solid State to DNA and Drug Design; Matta, C. F.; Boyd, R. J., Eds.; Wiley-VCH, 2007, 285. [9] Duan *et al.* Comput. Chem. 2003, 24, 1999. [10] Leung *et al.* IEEE T. Pattern Anal. 2000, 22, 1396. [11] Borodin *et al.* Force Field Fitting Toolkit; University of Utah; www.eng.utah.edu/~qdsmith/fff.html. [12] Eisenmenger *et al.* Comp. Phys. Comm. 2006, 174, 422; www.smmp05.net/. [13] Simms *et al.* Prot. Eng. Design & Selection 2008, 21, 369; www.dynameomics.org/. [14] Basdevant *et al.* Phys. Chem. B 2007, 111, 9390. [15] Heisterberg, D.J., Technical report, Ohio Supercomputer Center. Translation from FORTRAN to C and Input/Output by Jan Labanowski, Ohio Supercomputer Center, 1990.