

## RESEARCH OUTPUTS / RÉSULTATS DE RECHERCHE

### Reduced point charge models of proteins: Assessment based on molecular dynamics simulations

Leherte, Laurence

*Published in:*  
Molecular Simulation

*DOI:*  
[10.1080/08927022.2015.1044452](https://doi.org/10.1080/08927022.2015.1044452)

*Publication date:*  
2016

*Document Version*  
Peer reviewed version

[Link to publication](#)

*Citation for published version (HARVARD):*  
Leherte, L 2016, 'Reduced point charge models of proteins: Assessment based on molecular dynamics simulations', *Molecular Simulation*, vol. 42, no. 4, pp. 289-304. <https://doi.org/10.1080/08927022.2015.1044452>

#### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

#### Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

**Reduced point charge models of proteins: Assessment  
based on molecular dynamics simulations**

Journal:	<i>Molecular Simulation/Journal of Experimental Nanoscience</i>
Manuscript ID:	GMOS-2014-0417.R1
Journal:	Molecular Simulation
Date Submitted by the Author:	n/a
Complete List of Authors:	Leherte, Laurence; University of Namur, Chemistry
Keywords:	molecular electrostatic potential, smoothing of molecular fields, reduced point charge model, Ubiquitin, Barnase–Barstar

SCHOLARONE™  
Manuscripts

View Only

1  
2  
3 **Reduced point charge models of proteins: Assessment based on**  
4 **molecular dynamics simulations**  
5  
6

7  
8 Laurence LEHERTE  
9

10  
11 Laboratoire de Physico-Chimie Informatique  
12

13  
14 Unité de Chimie Physique Théorique et Structurale  
15

16  
17 Department of Chemistry  
18

19  
20 Namur MEDicine & Drug Innovation Center (NAME DIC)  
21

22  
23 University of Namur, Rue de Bruxelles 61, B-5000 Namur (Belgium)  
24

25  
26 [laurence.leherte@unamur.be](mailto:laurence.leherte@unamur.be)  
27

28  
29 Tel. +32-81-72.45.60  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## Reduced point charge models of proteins: Assessment based on molecular dynamics simulations

A reduced point charge distribution is used to model Ubiquitin and two complexes, Vps27 UIM-1–Ubiquitin and Barnase–Barstar. It is designed from local extrema in charge density distributions obtained from the Poisson equation applied to smoothed molecular electrostatic potentials. A variant distribution is built by locating point charges on atoms. Various charge fitting conditions are selected, i.e., from either electrostatic Amber99 Coulomb potential or forces, considering reference grid points located within various distances from the protein atoms, with or without separate treatment of main and side chain charges. The program GROMACS is used to generate Amber99SB molecular dynamics (MD) trajectories of the solvated proteins modelled using the various reduced point charge models (RPCM) so obtained. Point charges that are not located on atoms are considered as virtual sites. Some RPCMs lead to stable MD trajectories. They however involve a partial loss in the protein secondary structure and lead to a less structured solute solvation shell. The model built by fitting charges on Coulomb forces calculated at grid points ranging between 1.2 and 2.0 times the van der Waals radius of the atoms, with a separate treatment of main chain and side chain charges, appears to best approximate all-atom MD trajectories.

**Keywords:** molecular electrostatic potential; smoothing of molecular fields; reduced point charge model; Ubiquitin; Barnase–Barstar

### 1. Introduction

Nowadays, Molecular Dynamics (MD) simulations are a common tool to interpret and/or predict the energetic, dynamical, and structural properties of protein structures.

Well-known force fields, such as Amber, CHARMM, or OPLS, that are currently used, are still the subject of validation tests and modifications.[1,2] In addition to the so-called bonding terms, the force fields commonly include non-bonding terms such as van der Waals and Coulomb contributions, the latter involving partial atomic charges.

1  
2  
3 To reduce the number of degrees of freedom of a molecular system, coarse-  
4 grained representations and their associated force fields are an active field of  
5 research.[3,4] Besides the use of unit charges as in references [5-9], an approach to  
6 assign a partial charge to a molecular fragment or to the corresponding pseudo-atom,  
7 named here coarse grain, is to sum over the atomic charges involved in the fragment  
8 [10,11]. In the work of DeVane *et al.*, [9] unit values are scaled down to compensate  
9 for the solvent that is represented by uncharged spheres. Advanced approaches involve  
10 the assignment of multipolar contributions to ellipsoids.[12] In that last work, dedicated  
11 to the modelling of the amino acids, the charge distribution is represented by point  
12 multipolar expansions fitted to reproduce all-atom energy profiles.

13  
14  
15 Charge assignment methods usually consist in a least-square fitting of the coarse  
16 grain potential parameters so as to reproduce at best the all-atom potential values even if  
17 the size of the system and its conformational dependency may raise problems. [13,14]  
18 Terakawa and Takada [15] proposed a method to fit non-integer charges on the C $\alpha$   
19 beads of surface amino acids of a protein through an approximation of the all-atom  
20 Poisson-Boltzmann electrostatic potential, a procedure adopted earlier by Basdevant *et*  
21 *al.* [16] when constructing a reduced version of the Amber force field. The mimic of  
22 all-atom electrostatic interactions using a limited set of point charges can also be  
23 achieved through a genetic algorithm procedure.[12,17]

24  
25  
26 In our approach used so far to generate point charges of the amino acids,[18] the  
27 program QFIT [19] was used to assign, through a least square fitting algorithm, charge  
28 values to a reduced amount of points taking into account various molecular  
29 conformations. Following perspectives mentioned in a previous work regarding the  
30 revision of the calculation of the point charge values, [20] we apply here the idea of  
31 force fitting, as forces are the driving property in MD simulations. Charge values are  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 fitted on electrostatic interaction forces rather than on electrostatic interaction  
4 potentials, e.g., as achieved by Wang *et al.* for the design of a coarse-grained force field  
5 based on MD trajectories [21].  
6  
7  
8

9  
10 In the present work, we calculate new point charge values for the two kinds of  
11 charge distributions obtained previously, named mCD and mCDa in reference [22], and  
12 we determine how well these models approximate the all-atom one through the analysis  
13 of MD trajectories. The first model, mCD, based on charges located at critical points of  
14 smoothed charge density (CD) distribution functions of amino acids, calculated from  
15 Amber99 [23] atomic values, involves two point charges on the main chain of each  
16 amino acid, precisely located on atoms C and O, and up to six charges for the side  
17 chain. In the second model, most of the point charges observed in the first model were  
18 set at selected atom positions rather than being located away from atom positions. In  
19 model IIIa, only residues his<sup>+</sup>, phe, and trp present a non-atomic charge. Both models  
20 involve the same amount of point charges and are displayed in Figure 1 where amino  
21 acid (AA) residues are represented with the particular main chain atoms (C=O)<sub>AA</sub>(N-  
22 H)<sub>AA+1</sub> as the two main chain charges originate from that particular moiety [18].  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40

41 Various other charge fitting conditions are selected in the present work, *i.e.*,  
42 based on electrostatic potential or forces, considering reference grid points located  
43 within various distance ranges from the protein atoms, with or without separate  
44 treatment of main chain and side chain charges. They lead to diverse sets of charge  
45 values which are implemented and evaluated versus results obtained with the original  
46 all-atom Amber99 point charge distribution.  
47  
48  
49  
50  
51  
52  
53  
54

55 Applications are given for three biological systems, *i.e.*, Ubiquitin [24], and the  
56 two protein complexes Vps27 UIM-1–Ubiquitin [25] and Barnase–Barstar [26].  
57  
58  
59  
60

## 2. Materials and methods

### 2.1 Charge fitting conditions

The spatial distribution of the original reduced set of point charges (Figure 1 Top) was obtained through a topological analysis of the CD distribution functions of each amino acids, where the CDs are obtained from the Poisson equation applied to smoothed Coulomb potentials. From the mathematical formalism given in references [18,20,22], the smoothed analytical CD distribution function of an atom  $\rho_{a,s}(r)$  can be expressed as:

$$\rho_{a,s}(r) = \frac{q_a}{(4\pi s)^{3/2}} e^{-r^2/4s} \quad (1)$$

where  $s$  is the smoothing factor and  $q_a$  stand for the atomic charge.

To follow the pattern of local maxima and minima in a CD field, as a function of the degree of smoothing, the following strategy is adopted. First, each atom of a molecule is considered as a starting point. As the smoothing degree increases, each point moves along a path to reach a location where the CD gradient value vanishes. Convergence of trajectories leads to a reduction of the number of points.

Charge values were determined using the charge fitting program QFIT [19] applied to best approximate molecular electrostatic potentials (MEP) or molecular electrostatic forces taking into account various molecular conformations. All reference MEP grids were built using the Amber99 [23] point charges, assigned to the amino acid atoms using the software PDB2PQR [27,28], with a grid step of 0.5 Å. Fittings were first achieved by considering MEP grid points located at distances between 1.4 and 2.0 times the van der Waals radius of the atoms.[18] These two limiting distance values were selected after the so-called Merz-Singh-Kollman scheme.[29] Another range of limiting values, between 2.0 and 5.0 times the van der Waals radius of the atoms, was also applied to include points located at distances involving atoms separated by three

1  
2  
3 successive chemical bonds, i.e., 1-4 distances, and beyond. This last choice results from  
4  
5 an earlier observation showing that Coulomb 1-4 interactions should best approximate  
6  
7 the corresponding all-atom force field term.[22] Point charge values were first  
8  
9 generated for the side chain only, and a second fitting procedure was applied for the  
10  
11 whole amino acid considering the side chain charge values previously obtained. A  
12  
13 single fit carried out over all charges, main chain and side chain ones, was tested earlier  
14  
15 when working with MEP maps with less efficiency than when working separately on  
16  
17 side chain and main chain points.[18] Such fitting conditions were nevertheless tested  
18  
19 again in the present work. In all fittings, the total electric charge and the magnitude of  
20  
21 the molecular dipole moment were constrained to be equal to the corresponding all-  
22  
23 atom Amber99 values. All dipole moment components were calculated with the origin  
24  
25 of the atom coordinates set to (0. 0. 0.).  
26  
27  
28  
29  
30

31  
32 In order to fit charges on molecular electrostatic forces rather than on MEP  
33  
34 maps, three reference grids of forces acting along the x, y, and z axes were generated  
35  
36 during the charge fitting procedure by numerical differentiation of the MEP values  $V$ .  
37

38  
39  $\alpha$  (alpha) For each direction  $\alpha$ , a five point first derivative formula was applied to calculate the  
40  
41 force  $V_i^{(1)}$  at grid point  $i$ :  
42  
43  
44

$$V_i^{(1)} = (-V_{i+2} + 8V_{i+1} - 8V_{i-1} + V_{i-2}) / 12h \quad (2)$$

45  
46 where  $h$  stands for the grid step. This prevents the need to initially calculate and store  
47  
48 three reference all-atom force maps. The charge fitting was carried out so as to  
49  
50 minimize the error function  $y$ :  
51  
52  
53  
54  
55

$$y = \sum_{m=1}^{N_m} w_m \frac{1}{N_g} \sum_{\alpha=1}^3 \sum_{i=1}^{N_g} (V_{i,\text{ref}}^{(1)} - V_{i,\text{model}}^{(1)})^2 \quad (3)$$



1  
2  
3 where  $N_m$ ,  $w_m$ , and  $N_g$  stand for the number of molecular conformations and their  
4  
5 weight, and the number of valid grid points, respectively.  $V_{i,\text{ref}}^{(1)}$  is the reference all-atom  
6  
7 force acting at point  $i$  while  $V_{i,\text{model}}^{(1)}$  is the force calculated using the reduced set of point  
8  
9 charges.  
10  
11  
12

13 All reduced point charge models (RPCM) built considering the various  
14  
15 combinations of charge fitting conditions were first tested on the protein Ubiquitin as  
16  
17 reported below.  
18  
19  
20

## 21 22 **2.2 Reduced point charge models and Molecular Dynamics simulations**

23

24 To allow MD simulations, the non-atomic point charges of the RPCMs were  
25  
26 implemented in the topological files of the GROMACS package [30,31] as virtual sites  
27  
28 characterised by a nul mass and radius. The corresponding parameters of models mCD  
29  
30 (named here model II) and mCDa (named here model IIIa) described in Table 1 were  
31  
32 given in reference [22]. All other parameters and charge values associated with the six  
33  
34 models generated using the new charge fitting conditions are reported in SI 1 to SI 6 for  
35  
36 models IV, VI, XII, Va, VIIa, and IXa, respectively. Models II, IV, VI, and XII involve  
37  
38 point charges located at critical points of CD distributions, while models IIIa, Va, VIIa,  
39  
40 and IXa are characterised by charges mostly located on atoms. The number occurring  
41  
42 in the code name of the models is arbitrary. In order to assign charges to the C-terminal  
43  
44 residue of the proteins, a same value is considered for both oxygen atoms of the  
45  
46 backbone while an equivalent but positive charge value is assigned to the N atom of the  
47  
48 N-terminal amino acid. All other terms of the original Amber99SB force field, e.g.,  
49  
50 bonded and van der Waals terms, are left unchanged.  
51  
52  
53  
54  
55  
56  
57

58 A large set of fitting conditions were tested but only those that allowed relatively  
59  
60 stable MD trajectories for the initially tested system, Ubiquitin, are reported (Table 1).

1  
2  
3 From Table 1, it can be seen that the fit to forces for the model where the point charges  
4 are not located on atoms, i.e., model IV, allows to decrease the extreme values among  
5 all amino acid charges, ranging them between  $-0.80$  and  $1.03$   $|e^-|$ . Larger ranges are  
6 indeed observed for models where charges were fitted from MEP values, i.e., II and XII,  
7 or when the valid grid value range is extended, like in model VI. Models VI and XII  
8 are characterised by the largest range of possible values, i.e., between  $-0.84$  and  $1.53$ ,  
9 and between  $-0.87$  and  $1.92$   $|e^-|$ , respectively. They are also associated with the largest  
10 mean absolute charge for the main chain. Locating charges on atoms, such as in models  
11 Va and VIIa, allows to further reduce the amount of charges of high magnitudes. For  
12 example, the range of charge values now extends from  $-0.76$  to  $1.03$   $|e^-|$  for Va. Models  
13 Va and IXa are characterised by the lowest main chain charges, with a mean absolute  
14 value of  $0.64$  and  $0.62$   $|e^-|$ , respectively.

15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32 The MD protocol used to simulate the protein systems under the various RPCMs  
33 is briefly given hereafter. The equilibration stage was doubled versus previous  
34 works.[20,22] MD trajectories of the systems were run using the GROMACS 4.5.5  
35 program package [30,31] with the Amber99SB force field [32] under particle mesh  
36 Ewald periodic boundary conditions. Long-range dispersion corrections to energy and  
37 pressure were applied. The initial configurations were retrieved from the Protein Data  
38 Bank [33] (PDB IDs: 1UBQ [24], 1Q0W [25], 1BRS [26]) and solvated using TIP4P-  
39 Ew (an all-atom four-site model) [34] water molecules so as protein atoms lie at least at  
40 1.2 nm from the cubic box walls. For a same protein system, the number of water  
41 molecules may slightly vary with the model (Table 2). The systems were first  
42 approximately optimized to eliminate large forces and then heated to 50 K through a 10  
43 ps canonical (NVT) MD, with a time step of 2 fs and LINCS constraints acting on  
44 bonds involving H atoms. The trajectory was followed by two successive 20 ps heating  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 stages, at 150 K and at the final temperature, i.e., 300 K, under the same conditions.  
4  
5  
6 Next, each system was equilibrated during 50 ps in the NPT ensemble to relax the  
7  
8 solvent molecules. Finally, two successive 20 ns MD simulations were performed in  
9  
10 the NPT ensemble. The 'V-Rescale' and 'Parrinello-Rahman' algorithms were selected  
11  
12 to constrain T and P, respectively. A final production run of 20 ns was performed for  
13  
14 the evaluation of energetic, structural, and dynamical properties of the systems.  
15  
16  
17 Trajectory data were saved every 2 ps.

18  
19  
20 The total number of point charges to be considered in the protein representations  
21  
22 is reduced by a factor that is slightly larger than 4 for the three systems under study  
23  
24 (Table 2). For instance, structure 1UBQ that consists of 1231 atoms is characterised by  
25  
26 283 point charges only when using a RPCM. One also notices that the RPCMs provide  
27  
28 dipole moment values, for the initially optimized protein structure, that are of the same  
29  
30 order of magnitude as for the all-atom models (Table 2). Most RPCM dipole moment  
31  
32 values are slightly larger than their corresponding all-atom value, except for three  
33  
34 critical point-based models of structure 1Q0W, i.e., II, VI, and XII, with values of  
35  
36 205.2, 210.1, and 208.6 D, respectively.  
37  
38  
39  
40  
41  
42

### 43 **2.3 Protein systems**

44  
45 Ubiquitin (PDB ID: 1UBQ [24]) is a reference protein system that has already been  
46  
47 studied by MD simulations as, for examples, in references [32,35-39]. It involves 76  
48  
49 amino acid residues (1231 atoms) and its secondary structure is characterised by a  $\beta$ -  
50  
51 sheet made of five strands as well as two  $\alpha$ -helices formed by residues 23 to 34 and 56  
52  
53 to 59. The his residue of Ubiquitin is in its his $\epsilon$  state, thus leading to a net protein  
54  
55 charge of 0  $|e^-|$  [37]. It was the first system considered to test the various point charge  
56  
57 models we developed (Table 1). Sets of charges II and IIIa were already applied to  
58  
59 Ubiquitin in reference [20], with a shorter equilibration stage. In the present paper, the  
60

1  
2  
3 equilibration stage was increased by 20 ns. In that previous paper reporting MD  
4  
5 simulation results of solvated and isolated Ubiquitin, the effect of locating point charges  
6  
7 away from or on the atoms, as well as the effect of the solvent force field selected to  
8  
9 model water, were discussed. For both models, one observed a progressive loss in the  
10  
11 secondary structure of the proteins at room temperature. At 300 K, model IIIa better  
12  
13 preserved some secondary elements, due to a better description of the 1-4 Coulomb and  
14  
15 short-range Lennard-Jones energy terms. Nevertheless, at lower temperatures, MD  
16  
17 simulations carried out with model II provided results that were essentially similar to  
18  
19 the all-atom model. TIP4P-Ew was best to maintain the protein structure in a  
20  
21 conformation close to the all-atom one and is more structuring at low temperature,  
22  
23 possibly due to low self-diffusion coefficients versus the water force field SPC.  
24  
25  
26  
27  
28

29  
30 In the Vps27 UIM-1–Ubiquitin complex (PDB ID: 1Q0W [25]), the two  
31  
32 partners are composed of 24 (394 atoms) and 76 (1227 atoms) amino acid residues.  
33  
34 They are numbered 255 to 278 and 1 to 76 in the PDB file, respectively. The total  
35  
36 numbers of atoms in 1UBQ and in bound Ubiquitin differ due to slight changes in the  
37  
38 amino acid content of the two structures. Pro19, glu24, ala28, and his68 of structure  
39  
40 1UBQ are replaced by ser19, asp24, ser28, and his+68 in structure 1Q0W. Vps27 UIM-  
41  
42 1, a short  $\alpha$ -helical structure, is known to interact with the five-stranded  $\beta$ -sheet of  
43  
44 Ubiquitin.[25,40] As specified in [25], the his residue of Ubiquitin is fully protonated  
45  
46 (his+ state). Two Na<sup>+</sup> ions were added to cancel the net charge of the system. The  
47  
48 complex system already studied previously [22] was again studied here using the  
49  
50 various RPCMs described in Table 1.  
51  
52  
53  
54

55  
56 The Barnase–Barstar protein complex is a benchmark system whose close-fitting  
57  
58 interface is largely studied through molecular modelling techniques.[41-46] It is, in the  
59  
60 present paper, studied for the first time with our RPCMs. Barnase is a 110-residue

1  
2  
3 protein (numbered 1 to 110, with 1727 atoms) whose functions are inhibited by Barstar,  
4  
5 a 90-residue polypeptide (numbered 111 to 199, with 1434 atoms) bound to it through  
6  
7 an  $\alpha$ -helix that sterically blocks the active site of Barnase (see Figures 1 in references  
8  
9 [42,44]). Many H-bonds are involved between the two partners which strongly interact  
10  
11 through electrostatic interactions [41-45] and undergoes an important role of water  
12  
13 molecules [45-47]. The atom coordinates for the Barnase–Barstar complex were  
14  
15 retrieved from the Protein Data Bank (PDB ID: 1BRS [26]). Histidine residues were  
16  
17 protonated his $\delta$  except for his102 in Barnase, protonated his $\epsilon$  as in reference [43]. Four  
18  
19 Na<sup>+</sup> ions were added to cancel the net charge of the system.  
20  
21  
22  
23  
24  
25

### 26 **3. Results and discussion**

#### 27 28 29 *3.1 Molecular dynamics trajectories*

30  
31  
32 As already mentioned, first tests made on protein Ubiquitin (PDB ID: 1UBQ [24]) led  
33  
34 to a selection of six RPCMs which allowed relatively stable MD trajectories over the  
35  
36 chosen simulation time, and in some cases, a close agreement with the all-atom  
37  
38 Amber99SB MD trajectories. More precisely, models IV, Va, VI, VIIa, IXa, and XII  
39  
40 were retained and were latter examined together with the original II- and IIIa-based MD  
41  
42 trajectories (Table 1).  
43  
44

45  
46 As discussed in references [20,22], the decrease in the MD calculation time is  
47  
48 limited by two factors, i.e., the conservation of all original terms in the Amber99SB  
49  
50 force field except for the Coulomb interactions that act on a reduced number of point  
51  
52 charges, and the all-atom description of the solvent molecules. A reduction factor of  
53  
54 about 15 % is observed for the solute alone for calculations performed on two 2.66 GHz  
55  
56 processors, while the gain in time is insignificant when the solvent is considered. Let us  
57  
58 mention that if working with an all-atom description of the protein structure limits the  
59  
60

1  
2  
3 gain of calculation time, it nevertheless allows to very easily switch from a RPCM to an  
4  
5 all-atom protein representation, as illustrated in reference [22].  
6  
7

8 A plot of the time evolution of the root mean square deviation (RMSD)  
9  
10 calculated over all atoms of the systems versus the initially optimized protein structures  
11  
12 is displayed in SI 7. Mean values and their standard deviation are reported in Table 3.  
13  
14 Regardless of the protein system, all mean RMSD values are larger than corresponding  
15  
16 all-atom values when a RPCM is applied. Among the RPCMs, IV and Va appear to be  
17  
18 characterised by the lowest RMSD values, with however a slight discrepancy to this  
19  
20 rule for model XII applied to Ubiquitin with  $\text{RMSD} = 0.47$  nm, and for model IXa  
21  
22 applied to Barnase–Barstar with  $\text{RMSD} = 0.48$  nm. The highest values, 1.17, 1.45, and  
23  
24 1.63 nm, are observed for structure 1Q0W modelled with XII, VIIa and IXa,  
25  
26 respectively. They are due to a progressive decomplexation of the two protein partners  
27  
28 as illustrated using snapshots of the last MD frame (Figure 2) and lead to higher  
29  
30 standard deviations of 0.29, 0.10, and 0.55 nm, respectively (Table 3). In the case of  
31  
32 1BRS, the structure modelled using II and XII differs the most from the starting protein  
33  
34 conformation, with a mean  $\text{RMSD} = 1.00$  nm. The simulated structure is however  
35  
36 relatively stable, with a standard deviation of 0.04 nm in each case. There actually is a  
37  
38 slight interpenetration of Barnase into the structure of Barstar, due to the strong  
39  
40 deconstruction of the complex structure with model II, while, with model XII, one  
41  
42 observes a strong unfolding of the Barstar amino acid sequence 190 to 199 interacting  
43  
44 along the Barnase segment 37 to 30 (SI 8).  
45  
46  
47  
48  
49  
50  
51  
52  
53

### 54 **3.2 Structure analysis**

55  
56  
57 The analysis of maps reporting the mean shortest residue-residue distances (SI9), shows  
58  
59 that the least deconstructing model is Va, i.e., the model constructed with charges fitted  
60  
on all-atom Coulomb forces calculated at grid points ranging between 1.2 and 2.0 times

1  
2  
3 the van der Waals radius of the atoms, with a separate treatment of main chain and side  
4  
5 chain charges. In structure 1UBQ modelled using II and IIIa, some of the close  
6  
7 contacts, especially those occurring between amino acid residues 30 to 40 and 69 to 75  
8  
9 or between residues 15 to 25 and 40 to 50, i.e.,  $\beta$ -strands, are missing. Models IV and  
10  
11 Va allow to retain the main features of the 3D folds, as also observed for the two other  
12  
13 protein systems, while II, IXa, and XII show a deconstruction of Ubiquitin in 1QOW as  
14  
15 well as a displacement of Vps27 UIM-1 away from Ubiquitin. In the case of 1BRS, IIIa  
16  
17 and IXa also provide distance maps that are similar to the all-atom results, with slight  
18  
19 discrepancies for Barnase and Barstar, respectively. With IIIa, amino acid sequences 20  
20  
21 to 38 and 38 to 50 have a reduced number of close contacts due to the unfolding of the  
22  
23 sequence 20 to 50 into a loose loop, while with IXa, sequences 111 to 131 and 170 to  
24  
25 199 have a reduced number of close contacts due to a deconstruction of the involved  
26  
27 helices and strands (SI 8). Nevertheless, secondary structures (SI 10) as well as final  
28  
29 snapshots of the MD trajectories (Figure 2) show at least a partial conservation of the  
30  
31 molecular structure. In the case of structure 1UBQ, IV, Va, VI, and XII seem to favour  
32  
33 the helix moiety versus the other representations while all selected models but II, IIIa,  
34  
35 Va, and VIIa, let appear rather well preserved  $\beta$ -strands (SI 10). Additionally, the mean  
36  
37 gyration radius of structure 1UBQ, calculated from the IV- and Va-based MD  
38  
39 trajectories, 1.27 nm in both cases, is closer to the all-atom value, 1.18 nm (Table 4).  
40  
41 They are also associated with relatively low standard deviation values. Snapshots taken  
42  
43 at the final MD step (Figure 2) show that IV, Va, IXa, and XII preserve some of the  
44  
45 regular secondary structure elements of Ubiquitin, i.e.,  $\alpha$  and  $\beta$  structural elements. The  
46  
47 corresponding RMSD value calculated versus the initially optimized structure using  
48  
49 VMD [48] adopt the lowest RMSD values, i.e., below or close to 0.5 nm (Table 5). For  
50  
51 example, IV and Va present values of 0.484 and 0.421 nm, respectively. Model XII  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 seems to even better preserve the global shape of the protein with a RMSD value of  
4  
5 0.355 nm.  
6  
7

8 In the case of structure 1Q0W, II and VI appear to completely miss the helix  
9  
10 structures (SI 10), while XII misses the  $\beta$ -strands. Models IIIa and IV have a stronger  
11  
12 trend to preserve these two kinds of secondary structure elements, while Va and VIIa  
13  
14 still show a progressive loss of the secondary structure. Model IXa appears to preserve  
15  
16 part of the secondary structure of Ubiquitin, as also seen from Figure 2, while the helix  
17  
18 structure of the ligand is, in all RPCMs, strongly deconstructed. As for uncomplexed  
19  
20 Ubiquitin, the RMSD value of the final protein conformation calculated versus the  
21  
22 initially optimized structure stay close to 0.5 nm when using IV and Va (Table 5).  
23  
24  
25

26  
27 Regarding the Barnase–Barstar complex, models Va and IXa allow to maintain a  
28  
29 number of  $\alpha$ -helical structures, especially the very first helix of Barnase, as well as a  
30  
31 higher number of  $\beta$ -strands than the other RPCMs (Figure 2 and SI 10). Structures  
32  
33 simulated by these two models are very stable, especially when using Va. Additionally,  
34  
35 for these two RPCMs, regions of the distance maps involving the first 40 Barnase  
36  
37 residues let appear close contacts, similarly to the all-atom case (SI 9). Again, such  
38  
39 more satisfying models come with the lowest mean RMSD values, below 0.5 nm (Table  
40  
41 3) and with gyration radii  $r_G$  that are the closest to the corresponding all-atom values  
42  
43 (Table 4). The closest agreement between  $r_G$  values is provided by Va, with a value of  
44  
45 1.86 versus 1.76 nm for the all-atom model. It also appears to be the less varying value  
46  
47 during the 20 ns MD trajectory, with the lowest standard deviation value, 0.02 nm  
48  
49 (Table 4). Finally, the RMSD value that is associated with the final frame is close to  
50  
51 0.5 nm, as already observed for the best models of the two other protein systems (Table  
52  
53 5).  
54  
55  
56  
57  
58  
59  
60



### 3.3 Backbone dynamics

An analysis of the C $\alpha$  root mean square fluctuations (RMSF) shows that the motions of the amino acid residues can be strongly enhanced when one selects a RPCM (Figure 3). Large deviations, calculated as the RMSD between the RPCM and all-atom RMSFs, are even observed for models like IXa and XII when applied to structure 1Q0W (Table 6).

In that case, RMSD values of 0.873 and 0.651 nm are obtained, respectively.

Nevertheless, the RMSD values reported in Table 6 are among the lowest ones for IV and Va, with values of 0.055 and 0.111 nm for structures 1UBQ and 1Q0W, and 0.116, 0.158, and 0.095 nm, for 1UBQ, 1Q0W, and 1BRS, respectively. Correlation

coefficients  $\kappa$  between the all-atom and RPCM-based RMSF values are calculated using:

$$\kappa = \left( \frac{1}{N} \sum_{i=1}^{\text{No. of residues}} (u_{\text{all-atom}} - u_{\text{RPCM}})_i - (\bar{u}_{\text{all-atom}} - \bar{u}_{\text{RPCM}}) \right) / (\sigma_{\text{all-atom}} \sigma_{\text{RPCM}}) \quad (4)$$

where  $u$  stands for the RMSF values.  $\bar{u}$  and  $\sigma$  are the average and the standard deviation of the  $u$  values for a given protein structure, respectively. As reported in Table 6, correlation coefficient values can be well below 1, especially for the two complex systems 1Q0W and 1BRS. This illustrates that the fluctuation pattern of the values  $u$  calculated for the all-atom trajectory is not systematically well reproduced by the RPCMs. However,  $\kappa$  has the highest values when obtained from II-, IV-, and IXa-based MD simulations, for 1UBQ, 1Q0W, and 1BRS, respectively. On the whole, IV and Va that are built using the same fitting conditions (Table 1) provide correlation coefficients that rank among the highest values for each protein system, with values of 0.910, 0.840, 0.590, and 0.835, 0.708, and 0.577, respectively. Contrarily, II and IIIa that are, generally, characterised by high RMSD and low  $\kappa$  values, are more likely to favour conformational changes, as illustrated in reference [22] for the Vps27 UIM-1-Ubiquitin

1  
2  
3 complex system.  
4  
5  
6

### 7 **3.4 Hydrogen bond networks**

8

9  
10 The default parameters of hydrogen bonds in the GROMACS analysis tools are, for the  
11 H-acceptor distance and the donor-H-acceptor angle, set to 0.35 nm and 30°,  
12 respectively. The analysis of intra- and intermolecular H-bonds occurring in all protein  
13 structures provided results that are given in Table 7. The Table shows that, consistently  
14 with values obtained previously,[20,22] the number of intramolecular H-bonds is  
15 drastically reduced when using RPCMs. For structure 1UBQ, one observes a reduction  
16 factor of 7 between the number of H-bonds in the all-atom model, i.e., 55.9, and in  
17 model II, i.e., 7.9. For all three structures, IV, Va, and IXa, are the least disagreeing  
18 models versus the all-atom ones. The decrease in the number of H-bonds mainly  
19 originates from the absence of any charge on the N and H atoms of the main chain, and  
20 on selected atoms of side chains, e.g., arg and lys. In addition, the absence of any clear  
21 maximum in the intramolecular H-bond angle distribution functions originates from a  
22 loss in the orientational character of the intra- H-bonds (Figure 4 and SI 11). The  
23 features presented in Figure 4 for 1UBQ only are also valid for the other protein  
24 structures and RPCMs, as illustrated in SI 11. Contrarily, RPCMs lead to an apparent  
25 increase in the number of protein-water H-bonds. This is related to the less structured  
26 water network as shown by radial distribution functions which illustrate a less well  
27 defined first solvation shell (Figure 5 and SI 12). The features presented in Figure 5 for  
28 1UBQ are generalised to the other two protein structures and RPCMs, as illustrated in  
29 SI 12. However, the number of such H-bonds, 191.9 in the all-atom case of structure  
30 1UBQ, is almost preserved when one considers the standard deviations of the numbers  
31 obtained with IV, Va, and IXa. Reduced point charge distributions Va and IXa are also  
32 appropriate to model 1Q0W and 1BRS.  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

The distribution of water molecules in the vicinity of the protein surface is illustrated using radial distribution functions (Figure 5 and SI 12). As expected for the all-atom models,  $g(\text{surf-O}_w)$  indeed lets appear two peaks, the first one being located at about 0.2 nm which originates from the closest water molecules interacting through H-bonds with the protein surface atoms, and a second peak, at about 0.26 nm [36,38]. Those two peaks define the first solvation shell of the proteins. In the RPCM results, the first peak of  $g(\text{surf-O}_w)$  clearly vanishes but is still present in the  $g(\text{surf-H}_w)$  distributions (Figure 5). The layer of the closest  $\text{O}_w$  atoms appears to be displaced towards larger distances and is overlapped by the second peak of  $\text{O}_w$  atoms. A high amount of water molecules are thus oriented differently when a RPCM is used.

The dynamics of protein-water H-bonds can be characterised through the so-called H-bond autocorrelation functions:

$$C(t) = \langle h(0)h(t) \rangle / \langle h \rangle \quad (5)$$

where  $h(t)$  is assigned a value of 1 or 0 if a particular pair of atoms is H-bonded or not.

The approach that was applied to evaluate overall correlation times  $\tau$  associated with  $C(t)$ , is:

$$\tau = \int_0^{\infty} C(t) dt \quad (6)$$

Values of  $\tau$  are reported in Table 7. They show that protein-water H-bonds are best approximated by Va and IXa for the three protein systems. For examples, mean values of 459.3 and 477.1 are provided by those two models, respectively, and compare rather well to the all-atom value of 452.9. As reported before [20],  $\tau$  is largely increased when using a RPCM, regardless of the protein structure. It illustrates a slower H-bond dynamics, most probably due to the higher packing of water at the protein

1  
2  
3 surface or to the greater short-range electrostatic interactions occurring due to large  
4  
5 partial charges [20]. Besides the fact that the mean numbers of protein-ligand H-bonds  
6  
7 are reduced versus the all-atom case (Table 7), their associated values of  $\tau$  have no  
8  
9 definite trends in common to the two complexes. They however tend to show an  
10  
11 increased lifetime for such H-bonds in the case of structure 1BRS, all  $\tau$  being larger  
12  
13 than the all-atom value of 146.3 ps. A deeper analysis of the effect of the RPCM on the  
14  
15 interface solvent molecules can be seen as a perspective to the present work by avoiding  
16  
17 any changes in the protein conformations from one simulation to another. This can be  
18  
19 achieved by simulating rigid protein structures.  
20  
21  
22  
23  
24  
25

### 26 **3.5 Energetics**

27  
28 For each MD frame generated using a RPCM, the corresponding all-atom values of  
29  
30 various energy terms were obtained through post-processing calculations. Linear  
31  
32 regression calculations were then achieved for the RPCM versus all-atom energy terms:  
33  
34

$$35 E_{RPCM} = S E_{all-atom} + I \quad (7)$$

36  
37 where  $S$  and  $I$  stand for the slope and the intercept of the linear equations, respectively.  
38  
39

40  
41 The determination coefficient  $R$ ,  $S$  and  $I$  are reported in SI 13 to SI 15, respectively.  
42  
43

44  
45 Examination of the data shows that the  $Cb_{14}$  terms, i.e., the Coulomb interaction  
46  
47 potentials between atoms separated by three chemical bonds, are the most affected  
48  
49 contributions. Indeed, the  $R$  and  $S$  values that are associated with those contributions  
50  
51 are largely below 1. This implies that if one study, for instance, rigid systems by  
52  
53 freezing dihedrals, the RPCMs should be well suited for electrostatic calculations, as  
54  
55 already shown in our work about potassium ion channels [18,49]. Coulomb short-range  
56  
57 ( $Cb_{SR}$ ) regression data behave a lot better, with  $R$  and  $S$  close to 1. One even notices  
58  
59 that while the intramolecular protein-protein  $Cb_{SR}$  ( $p-p$ ) slope is almost always lower  
60

1  
2  
3 than 1, the intermolecular protein-non protein  $Cb\_SR$  ( $p-np$ ) slope can be larger than 1,  
4  
5 which means that these energy terms can be slightly over-estimated, especially when  
6  
7 using models with charges located away from the atom locations, i.e., II, IV, VI, and  
8  
9 XII. On the whole,  $R$  and  $S$  associated with the total energy  $E_{tot}$  are almost always of  
10  
11 the order of 0.99. Exceptions occur for 1Q0W modelled with VI and XII for which  $R$   
12  
13 and  $S$  can be slightly lower, about 0.98. It is uneasy to classify the models as more or  
14  
15 less satisfying based on the  $R$  and  $S$  values. One can however notices that the best  
16  
17 models so far, Va and IXa, all have a  $Cb\_SR$  ( $p-np$ ) slope that is lower than 1, contrarily  
18  
19 to all other models. To inspect the deviation of the energy values from their all-atom  
20  
21 counterpart, intercept values of the linear regressions were also analysed (SI 15).  
22  
23 Models Va and IXa almost systematically present the lowest absolute intercept values.  
24  
25 It is actually always the case for  $Cb\_14$ ,  $Cb\_SR$ ,  $E_{pot}$ , and  $E_{tot}$ . This may be related to  
26  
27 the fact that both Va and IXa have similar distributions of main chain charge values  
28  
29 (Table 1). In conclusion, a better approximation of the  $Cb\_14$  term occurs when  
30  
31 charges are set on the atoms of the proteins and are fitted from Coulomb forces rather  
32  
33 than from potentials.  
34  
35  
36  
37  
38  
39  
40

41 More generally, sets of force-fitted charges like IV and Va allow to  
42  
43 systematically better approximate all-atom forces than II and IIIa at very short distances  
44  
45 from the protein atoms, i.e., between 1.0 and 1.4 times the van der Waals radius of the  
46  
47 atoms. Indeed, the error function  $y$  defined in equation (3) presents an averaged  
48  
49 decrease of 14 and 18 % for model IV versus II and Va versus IIIa, respectively. In the  
50  
51 range of distances between 1.4 and 10.0 times the van der Waals radius of the atoms,  
52  
53 potential- and force-based charges behave similarly when evaluating forces, with a  
54  
55 slight averaged increase of 6 and 4 % for IV versus II and Va versus IIIa, respectively.  
56  
57  
58  
59

60 Finally, increasing the distance range of force values away from the protein

atoms should apparently be combined with a single consideration of all charges in the fitting procedure, as in the case of IXa.

Model Va always performs the best for intra- and intermolecular Coulomb short-range terms, i.e.,  $Cb\_SR(p-p)$  and  $Cb\_SR(p-np)$ , respectively. For example, the  $Cb\_14$  intercept of Va for structure 1UBQ is  $5587.5 \text{ kJ.mol}^{-1}$  versus 9169.0 for model IIIa.

Model IV is also among the best model to consider when using charges that are located away from the molecular skeleton. Again, for structure 1UBQ, the intercept value is 9467.7 versus 12097.1 for model II. On the whole, intercept values are lower for the CDa-based models than they are for the CD-based ones. Among the CD-based models, IV performs the best for all energy terms and all protein systems except for the reciprocal term  $Cb\_recip$ . Models II and XII, as well as IIIa and VIIa are, on the whole, the less favourable models to consider in the CD-based and CDa-based family, respectively. This confirms the high potency of II and IIIa to rapidly provide various protein conformations, as studied in reference [22] for the Vps27 UIM-1–Ubiquitin complex system.

### 3.6 Protein-ligand contacts

A detailed study of the contacts between protein partners in Vps27 UIM-1–Ubiquitin and Barnase–Barstar complexes is illustrated by Figure 6 and SI 16 that present the mean shortest distance between the amino acid residues of both partners averaged over the 20 ns MD trajectories. In the first case, one clearly distinguishes three regions extended along the Vps27 UIM-1 chain. The first region corresponds to the contacts occurring between the segment of amino acids 4 to 17 (259 to 272) of Vps27 UIM-1 and the  $\beta$ -strand 4 to 10 of Ubiquitin, while the second and third regions are due to contacts with  $\beta$ -strands 40 to 45 and 48 to 49, and  $\beta$ -strand 66 to 72, respectively. A pattern similar to the 1Q0W all-atom one was obtained when using IV. Model Va also

1  
2  
3 presents the three regions but the first region appears to be more extended in the sense  
4 that almost all residues are in close contact with Ubiquitin, except for the central amino  
5 acids. This is due to the bending of Vps27 UIM-1 through its central residues (Figure  
6  
7  
8  
9  
10  
11 2). Highest occurrence frequency values of the protein-protein H-bonds, calculated  
12 using VMD [48] with cut-off distance and angle of 0.35 nm and 30°, are given in Table  
13  
14  
15 8. Models for which no values were obtained are not reported. The Table shows, as  
16 already reported in Table 7, that H-bonds are less frequent and less numerous for the  
17 RCPMs than they are in the all-atom case. Model IV is however characterised by three  
18 Vps27 UIM-1– Ubiquitin H-bonds, i.e., glu273-lys6, leu271-ser65, glu273-hip68,  
19 occurring in regions 1 and 3. Models VI and VIIa present the three regions too, with  
20 reduced area (SI 16) due, respectively, to a drastic bending or extension of the ligand  
21 (Figure 2), while IXa and XII are strongly limited in their number of contacts due to the  
22 decomplexation of the ligand. In model XII, Vps27 UIM-1 still interacts with Ubiquitin  
23 through its C-terminal residue.  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35

36 Model Va also allows to reproduce the main features of the Barnase–Barstar  
37 contact map pattern (Figure 6, SI 16). These features form a set of eight regions and are  
38 determined from the observed shortest distances (Figure 6). Among the eight areas  
39 reported in Table 9, region #3 is not listed in the Contact Map Database ABC<sup>2</sup> [50]. It  
40 actually involves looser contacts observed along the MD trajectory. Contrarily, contacts  
41 detected in ABC<sup>2</sup> and also appearing in the all-atom MD simulation have disappeared  
42 from the RPCM simulations. Those are lys27-thr152 (region #1), arg59-glu186 (region  
43 #5), arg83-tyr139 (region #7), and hie102-tyr140 (region #8). When one focusses on  
44 the protein-ligand H-bonds occurring with a frequency larger than 10 %, one notices  
45 that one or more H-bonds identified by ABC<sup>2</sup> are detected using the all-atom MD  
46 simulation (Table 9), e.g., for region #1, the lys27-thr152 H-bond occurs with a  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 frequency of 57.0 %. Models IXa, and XII let appear H-bonds in five different regions,  
4  
5 but Va presents relatively high frequency values in the four areas it covers. For  
6  
7 examples, regions #2, #4, #7, and #8 are characterised by H-bond occurrence frequency  
8  
9 values of 39.4, 89.9, 46.9, and 46.9 %, respectively. H-bonds between gly52 and  
10  
11 asp193 as well as between gly53 and glu190 are also found with XII. They appear  
12  
13 along the extended amino acid sequence 190-199 of Barstar as illustrated earlier (SI 8).  
14  
15  
16  
17

#### 18 19 **4. Conclusions and perspectives**

20  
21 Two reduced point charge distributions were considered for Molecular Dynamics (MD)  
22  
23 simulations of three protein systems, i.e., Ubiquitin, Vps27 UILM-1–Ubiquitin, and  
24  
25 Barnase–Barstar. The first distribution, based on charges located at critical points of  
26  
27 smoothed amino acid charge density distribution functions calculated from Amber99  
28  
29 atomic charge values, involves two point charges on the main chain of each amino acid,  
30  
31 precisely located on atoms C and O, and up to six charges for the side chain, mostly  
32  
33 located away from atomic positions. In the second distribution, most of the charges are  
34  
35 set at selected atom positions. Several sets of charge values were obtained by using  
36  
37 different charge fitting conditions, i.e., based on electrostatic potential or forces,  
38  
39 considering reference grid points located within various distance ranges from the  
40  
41 protein atoms, with or without separate treatment of main chain and side chain charges.  
42  
43  
44  
45  
46

47  
48 The MD simulations were carried out using the program GROMACS with the  
49  
50 Amber99SB force field, in TIP4P-Ew water, at 300 K. Energetic, structural, and  
51  
52 dynamical information were retrieved from the analysis of the MD trajectories of the  
53  
54 reduced point charge models (RPCMs) and discussed versus the all-atom model and  
55  
56 available literature data. An emphasis was put on the global fold, the secondary  
57  
58 structure elements of the proteins, their energetics and fluctuations, and the  
59  
60 characterisation of H-bonds within the protein and with the solvent.



1  
2  
3 On a structural point of view, one observed a progressive loss in the secondary  
4 structure of the proteins when RPCMs are used. They also lead to an increase of the  
5 gyration radius. Among the eight charge sets used in the paper, a model based on the  
6 use of Coulomb forces as reference values for charge fitting, i.e., model Va, better  
7 preserves some secondary elements, due to a better description of the short range 1-4  
8 Coulomb energy terms, and limit the increase of the gyration radius. Precisely, charges  
9 of Va were fitted on all-atom Coulomb forces calculated at grid points ranging between  
10 1.2 and 2.0 times the van der Waals radius of the atoms, with a separate treatment of  
11 main chain and side chain charges. Model Va is also seen as one of the best to  
12 approximate energy values and is among the models that limit the increase of the  
13 backbone dynamics observed with RPCMs. Model IXa, built by fitting all point charge  
14 values on Coulomb forces calculated at grid points ranging between 2.0 and 5.0 times  
15 the van der Waals radius of the atoms, also appears to be a reliable model. However, it  
16 leads to strong structural changes of the Vps UIM-1 helix. Fitting charges on a limited  
17 number of points is more efficient when electrostatic forces are taken as reference  
18 values most likely because it systematically improves the approximation of all-atom  
19 forces at short separations, thus leading to MD trajectories that better approximate the  
20 all-atom ones. Additionnally, it appears that Coulomb energy values are also closer to  
21 the all-atom ones.

22  
23  
24 The RPCMs do not favour the formation of a first hydration shell as clearly as  
25 the all-atom model does. They however allow the formation of solute-solvent H-bonds  
26 with geometrical properties similar to the all-atom case. Intra-protein H-bonds are  
27 differently described with no well-defined angle distributions. The mean number of  
28 intra-protein H-bonds is largely reduced versus the corresponding all-atom values, due  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 to the decrease in the number of point charges, while the opposite trend is observed for  
4  
5 the solute-solvent H-bonds, due to less-structured first solvation shells.  
6  
7

8 Following the work presented above, we will further focus on the RPCMs that  
9  
10 allow major conformational changes in the protein structure, i.e., II and IIIa. Indeed, a  
11  
12 work achieved on structure 1Q0W [22] showed that these charge models allow to  
13  
14 generate particular conformations that appear to be stable ones through all-atom MD  
15  
16 simulations.  
17  
18

19 It is also planned, as a longer term perspective, to combine a RPCM with a  
20  
21 coarse-grained description of the protein structures.  
22  
23  
24  
25

## 26 Acknowledgments

27  
28 The author thanks the referees for their useful comments regarding the manuscript, and Daniel  
29  
30 Vercauteren, Director of the ‘Laboratoire de Physico-Chimie Informatique’ at the University of  
31  
32 Namur, for fruitful discussions. Frédéric Wautelet and Laurent Demelenne are gratefully  
33  
34 acknowledged for program installation and maintenance. This research used resources of the  
35  
36 ‘Plateforme Technologique de Calcul Intensif (PTCI)’ (<http://www.ptci.unamur.be>) located at  
37  
38 the University of Namur, Belgium, which is supported by the F.R.S.-FNRS. The PTCI is  
39  
40 member of the ‘Consortium des Équipements de Calcul Intensif (CÉCI)’ ([http://www.ceci-  
hpc.be](http://www.ceci-hpc.be)).  
41  
42

## 43 References

- 44  
45 [1] Lindorff-Larsen K, Maragakis P, Piana S, Eastwood MP, Dror RO, Shaw DE.  
46 Systematic validation of protein force fields against experimental data. PLoS  
47 ONE. 2012;7:e321131/1-e321131/6.  
48  
49 [2] Beauchamp KA, Lin Y-S, Das R, Pande VS. Are protein force fields getting  
50 better? A systematic benchmark on 524 diverse NMR measurements. J. Chem.  
51 Theory Comput. 2012;8:1409-1414.  
52  
53 [3] Voth GA, editor. Coarse-graining of condensed phase and biomolecular  
54 systems. Boca Raton (FL): CRC Press; 2009.  
55  
56 [4] Hills RD Jr, Lu L, Voth GA. Multiscale coarse-graining of the protein energy  
57 landscape. PLoS Comput. Biol. 2010;6:e1000827/1-e1000827/12.  
58  
59  
60

- 1  
2  
3 [5] Imai K, Mitaku S. Common pattern of coarse-grained charge distribution of  
4 structurally analogous proteins. *Chem-Bio. Inform. J.* 2003;3:194-200.  
5  
6 [6] Skepö M, Linse P, Arnebrant T. Coarse-grained modeling of proline rich protein  
7 1 (PRP-1) in bulk solution and adsorbed to a negatively charged surface. *J. Phys.*  
8 *Chem. B.* 2006;110:12141-12148.  
9  
10 [7] Pizzitutti F, Marchi M, Borgis D. Coarse-graining the accessible surface and the  
11 electrostatics of proteins for protein-protein interactions. *J. Chem. Theory*  
12 *Comput.* 2007;3:1867-1876.  
13  
14 [8] Monticelli L, Kandasamy SK, Periole X, Larson RG, Tieleman DP, Marrink SJ.  
15 The MARTINI coarse-grained forcefield: Extension to proteins. *J. Chem.*  
16 *Theory Comput.* 2008;4:819-834.  
17  
18 [9] DeVane R, Shinoda W, Moore PB, Klein ML. Transferable coarse grain  
19 nonbonded interaction model for amino acids. *J. Chem. Theory Comput.*  
20 2009;5:2115-2124.  
21  
22 [10] Nielsen SO, Lopez CF, Srinivas G, Klein ML. Coarse grain models and the  
23 computer simulation of soft materials. *J. Phys.: Condens. Matter.* 2004;16:R481-  
24 R512.  
25  
26 [11] Leherte, Guillot B, Vercauteren DP, Pichon-Pesme V, Jelsch C, Lagoutte A,  
27 Lecomte C. Topological analysis of proteins as derived from medium and high  
28 resolution electron density: Applications to electrostatic properties. In: Matta  
29 CF, Boyd RJ, editors. *The Quantum Theory of Atoms in Molecules - From Solid*  
30 *State to DNA and Drug Design.* Weinheim (D): Wiley-VCH; 2007, p. 285-316.  
31  
32 [12] Shen H, Li Y, Ren P, Zhang D, Li G. Anisotropic coarse-grained models for  
33 proteins based on Gay-Berne and electric multipole potentials. *J. Chem. Theory*  
34 *Comput.* 2009;5:2115-2124.  
35  
36 [13] Francl MM, Chirlian LE. The plus and minuses of mapping atomic charges to  
37 electrostatic potentials. *Rev. Comput. Chem.* 2000;14:1-31.  
38  
39 [14] Curcó D, Nussinov R, Alemán C. Coarse-grained representation of  $\beta$ -helical  
40 protein building blocks. *J. Phys. Chem. B.* 2007;111:10538-10549.  
41  
42 [15] Terakawa T, Takada S. RESPAC: Method to determine partial charges in  
43 coarse-grained protein model and its application to DNA-binding proteins. *J.*  
44 *Chem. Theory Comput.* 2014;10:711-721.  
45  
46 [16] Basdevant N, Borgis D, Ha-Duong T. A coarse-grained protein-protein potential  
47 derived from an all-atom force field. *J. Phys. Chem. B.* 2007;111:9390-9399.  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

- 1  
2  
3 [17] Berardi R, Muccioli L, Orlandi S, Ricci M, Zannoni C. Mimicking electrostatic  
4 interactions with a set of effective charges: A genetic algorithm. *Chem. Phys.*  
5 *Lett.* 2004;389373-389378.  
6  
7  
8 [18] Leherte L, Vercauteren DP. Coarse point charge models for proteins from  
9 smoothed molecular electrostatic potentials. *J. Chem. Theory Comput.*  
10 2009;5:3279-3298.  
11  
12 [19] Borodin O, Smith GD. Force Field Fitting Toolkit. The University of Utah.  
13 <http://www.eng.utah.edu/~gdsmith/fff.html> (Accessed 26 Aug. 2009).  
14  
15 [20] Leherte L, Vercauteren DP. Comparaison of reduced point charge models of  
16 proteins: Molecular dynamics simulations of Ubiquitin. *Sci. China Chem.*  
17 2014;57:1340-1354.  
18  
19 [21] Wang Y, Noid WG, Liu P, Voth GA. Effective force coarse-graining. *Phys.*  
20 *Chem. Chem. Phys.* 2009;11:2002-2015.  
21  
22 [22] Leherte L, Vercauteren DP. Evaluation of reduced point charge models of  
23 proteins through molecular dynamics simulations: Application to the Vps27  
24 UIM-1–Ubiquitin complex. *J. Mol. Graphics Model.* 2014;47:44-61.  
25  
26 [23] Wang J, Cieplak P, Kollman PA. How well does a restrained electrostatic  
27 potential (RESP) model perform in calculating conformational energies of  
28 organic and biological molecules, *J. Comput. Chem.* 2000;21:1999-2012.  
29  
30 [24] Vijay-Kumar S, Bugg CE, Cook WJ. Structure of Ubiquitin refined at 1.8 Å  
31 resolution. *J. Mol. Biol.* 1987;194:531-544.  
32  
33 [25] Swanson KA, Kang RS, Stamenova SD, Hicke L, Radhakrishnan I. Solution  
34 structure of Vps27 UIM–Ubiquitin complex important for endosomal sorting  
35 and receptor downregulation. *EMBO J.* 2003;22:4597-4606.  
36  
37 [26] Buckle AM, Schreiber G, Fersht AR. Protein-protein recognition: Crystal  
38 structural analysis of a Barnase–Barstar complex at 2.0 Å resolution.  
39 *Biochemistry* 1994;33:8878-8889.  
40  
41 [27] Dolinsky TJ, Nielsen JE, McCammon JA, Baker NA. PDB2PQR: An automated  
42 pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucl.*  
43 *Acids Res.* 2004;32:W665-W667.  
44  
45 [28] PDB2PQR, An automated pipeline for the setup, execution, and analysis of  
46 Poisson-Boltzmann electrostatics calculations. 2007. SourceForge Project Page.  
47 <http://pdb2pqr.sourceforge.net/> (Accessed 31 Aug. 2009)  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

- 1  
2  
3 [29] Singh UC, Kollman PA. An approach to computing electrostatic charges for  
4 molecules. *J. Comput. Chem.* 1984;5:129-145.  
5  
6 [30] Hess B, Kutzner C, van der Spoel D, Lindahl E. GROMACS 4: Algorithms for  
7 highly efficient, load-balanced, and scalable molecular simulation. *J. Chem.*  
8 *Theory Comput.* 2008;4:435-447.  
9  
10 [31] Pronk S, Páll S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, Shirts MR,  
11 Smith JC, Kasson PM, van der Spoel D, Hess B, Lindahl E. GROMACS 4.5: A  
12 high-throughput and highly parallel open source molecular simulation toolkit.  
13 *Bioinformatics* 2013;29:845-854.  
14  
15 [32] Showalter SA, Brüschweiler R. Validation of molecular dynamics simulations of  
16 biomolecules using NMR spin relaxation as benchmarks: Application to the  
17 AMBER99SB force field. *J. Chem. Theory Comput.* 2007;3:961-975.  
18  
19 [33] Bernstein FC, Koetzle TF, Williams GJ, Meyer Jr EE, Brice MD, Rodgers JR,  
20 Kennard O, Shimanouchi T, Tasumi M. The Protein Data Bank: A computer-  
21 based archival file for macromolecular structures. *J. Mol. Biol.* 1977;112:535-  
22 542.  
23  
24 [34] Horn HW, Swope WC, Pitera JW, Madura JD, Dick TJ, Hura GL, Head-Gordon  
25 T. Development of an improved four-site water model for biomolecular  
26 simulations: TIP4P-Ew. *J. Chem. Phys.* 2004;120:9665-9678.  
27  
28 [35] Abscher R, Schreiber H, Steinhauser O. The influence of a protein on water  
29 dynamics in its vicinity investigated by molecular dynamics simulation. *Proteins*  
30 1996;25:366-378.  
31  
32 [36] Dastidar SG, Mukhopadhyay C. Structure, dynamics, and energetics of water at  
33 the surface of a small globular protein: A molecular dynamics simulation. *Phys.*  
34 *Rev.* 2003;68:021921.  
35  
36 [37] Schröder C, Rudas T, Boresch S, Steinhauser O. Simulation studies of the  
37 protein-water interface. I. Properties at the molecular resolution. *J. Chem. Phys.*  
38 2006;124:234907/1-234907/18.  
39  
40 [38] Virtanen JJ, Makowski L, Sosnick TR, Freed KF. Modeling the hydration  
41 layer around proteins: HyPred. *Biophys. J.* 2010;99:1611-1619.  
42  
43 [39] Ganoth A, Tsfadia Y, Wiener R. Ubiquitin: Molecular modeling and  
44 simulations. *J. Mol. Graphics Mod.* 2013;46:29-40.  
45  
46 [40] Hurley JH, Lee S, Prag G. Ubiquitin-binding domains. *Biochem. J.*  
47 2006;399:361-372.  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

- 1  
2  
3 [41] Lee L-P, Tidor B. Barstar is electrostatically optimized for tight binding to  
4 Barnase. *Nature Struct. Biol.* 2001;8:73-76.  
5  
6 [42] Dong F, Vijayakumar M, Zhou H-X. Comparison of calculation and experiment  
7 implicates significant electrostatic contributions to the binding stability of  
8 Barnase and Barstar. *Biophys. J.* 2003;85:49-60.  
9  
10 [43] Wang T, Tomic S, Gabdoulina RR, Wade RC. How optimal are the binding  
11 energetics of Barnase and Barstar? *Biophys. J.* 2004;87:1618-1630.  
12  
13 [44] Ababou A, van der Vaart A, Gogonra V, Merz Jr KM. Interaction energy  
14 decomposition in protein-protein association: A quantum mechanical study of  
15 Barnase–Barstar study. *Biophys. Chem.* 2007;125:221-236.  
16  
17 [45] Hoefling M, Gottschalk KE. Barnase–Barstar: From first encounter to final  
18 complex. *J. Struct. Biol.* 2010;171:52-63.  
19  
20 [46] Ulucan O, Jaitly T, Helms V. Energetics of hydrophilic protein-protein  
21 association and the role of water. *J. Chem. Theory Comput.* 2014;10:3512-3524.  
22  
23 [47] Urakubo Y, Ikura T, Ito N. Crystal structure analysis of protein-protein  
24 interactions drastically destabilized by a single mutation. *Prot. Sci.*  
25 2008;17:1055-1065.  
26  
27 [48] Humphrey W, Dalk A, Schulten K. VMD - Visual Molecular Dynamics. *J. Mol.*  
28 *Graphics* 1996;14:33-38.  
29  
30 [49] Leherter L, Vercauteren DP. Charge density distributions derived from smoothed  
31 electrostatic potential functions: Design of protein reduced point charge models.  
32 *J. Comput. Aided Mol. Des.* 2011;25:913-930.  
33  
34 [50] Peter W, Sam A, Volkhard H. ABC (Analyzing Biomolecular Contacts)-  
35 database. *J. Integr. Bioinf.* 2007;4:50-58.  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## Figure captions

Figure 1. Location of point charges (black spheres) of the amino acid residues on (Top) critical points of smoothed charge density distribution functions, and (Bottom) selected atoms.

Figure 2. Final snapshots of the protein structures obtained from the last frames of 20 ns AMBER99SB-TIP4P-Ew MD trajectories at 300 K. Secondary structure elements are color-coded as follows: Coil (white),  $\alpha$ -helix (blue),  $\pi$  helix (purple),  $3_{10}$  helix (grey),  $\beta$ -sheet (red),  $\beta$ -bridge (black), bend (green), turn (yellow). For an interpretation of the references to colour, the reader is referred to the web version of the article.

Figure 3. RMSF of the  $C\alpha$  atoms of structures 1UBQ, 1Q0W, and 1BRS, obtained from 20 ns AMBER99SB-TIP4P-Ew MD trajectories at 300 K. (Plain line) All-atom, (dashed line) model IV, (dotted line) model Va. Residues of the protein complexes are numbered 1 to 24 (Vps27 UIM-1) and 25 to 100 (Ubiquitin) for 1Q0W, and 1 to 110 (Barnase) and 111 to 199 (Barstar) for 1BRS.

Figure 4. Distance and angle distributions of the Ubiquitin (1UBQ)-water H-bonds obtained from 20 ns AMBER99SB-TIP4P-Ew MD trajectories at 300 K. (Plain line) All-atom, (dotted line) model Va.

Figure 5. Radial distribution functions of the Ubiquitin (1UBQ) surface atoms versus the water atoms,  $g(P-O_w)$  and  $g(P-H_w)$ , obtained from 20 ns AMBER99SB-TIP4P-Ew MD trajectories at 300 K. (Plain line) All-atom, (dotted line) model Va.

Figure 6. Mean shortest protein-ligand distance maps as calculated from 20 ns AMBER99SB-TIP4P-Ew MD trajectories at 300 K for model Va. Encircled areas correspond to regions described in Tables 8 and 9. Distances are given in nm in the colour scale. For an interpretation of the references to colour, the reader is referred to the web version of the article.

Table 1. Charge fitting conditions applied to generate various sets of reduced point charge models (RPCM) based on the previously developed models mCD (model II) and mCDa (model IIIa) [22].

RPCM <sup>a</sup>	Charge fitting conditions			Range of charge values (e <sup>-</sup> )	Average and standard deviation of the absolute charge values of the main chain (e <sup>-</sup> )	Charges and virtual site parameters
	Reference grid <sup>b</sup>	Separate treatment of main and side chain charges	Range of valid grid values (Å)			
<b>CD-based models</b>						
II	MEP	yes	1.2 – 2.0	-0.85 – 1.35	0.77 ± 0.09	[22]
IV	MEF	yes	1.2 – 2.0	-0.80 – 1.03	0.69 ± 0.08	SI 1
VI	MEF	yes	2.0 – 5.0	-0.84 – 1.53	0.77 ± 0.09	SI 2
XII	MEP	yes	2.0 – 5.0	-0.87 – 1.92	0.79 ± 0.10	SI 3
<b>CDA-based models</b>						
IIIa	MEP	yes	1.2 – 2.0	-0.81 – 1.03	0.73 ± 0.09	[22]
Va	MEF	yes	1.2 – 2.0	-0.76 – 1.03	0.64 ± 0.07	SI 4
VIIa	MEF	yes	2.0 – 5.0	-0.79 – 1.09	0.73 ± 0.09	SI 5
IXa	MEF	no	2.0 – 5.0	-0.84 – 1.03	0.62 ± 0.10	SI 6

<sup>a</sup>CD and CDA stand for models where point charges are located at the critical points of the charge density (CD) and at atoms, respectively.

<sup>b</sup>MEP and MEF stand for molecular electrostatic potential and molecular electrostatic force, respectively.



Table 2. Description of the protein systems simulated by Amber99SB-TIP4P-Ew MD using various point charge models.

	All-atom	Point charge model							
		II	IV	VI	XII	IIIa	Va	VIIa	IXa
<b>1UBQ</b>									
# H <sub>2</sub> O	10369	10366	10366	10366	10366	10368	10366	10368	10368
# Point charges	1231	283	283	283	283	283	283	283	283
# Non-atomic point charges	0	84	84	84	84	2	2	2	2
Simulation box (nm)	6.89	6.89	6.89	6.89	6.89	6.89	6.89	6.89	6.89
Dipole moment (D) of the optimized protein structure	217.8	221.9	230.1	226.1	222.6	231.2	236.1	231.6	237.6
<b>1Q0W</b>									
# H <sub>2</sub> O	10553	10542	10542	10542	10542	10551	10551	10551	10551
# Point charges	1623	382	382	382	382	382	382	382	382
# Non-atomic point charges	0	112	112	112	112	3	3	3	3
Simulation box (nm)	6.95	6.95	6.95	6.95	6.95	6.95	6.95	6.95	6.95
Dipole moment (D) of the optimized protein structure	210.2	205.2	216.3	210.1	208.6	212.9	218.3	212.5	218.4
<b>1BR5</b>									
# H <sub>2</sub> O	18738	18916	18916	18723	18912	18740	18740	18739	18740
# Point charges	3161	786	786	786	786	786	786	786	786
# Non-atomic point charges	0	272	272	272	272	12	12	12	12
Simulation box (nm)	8.43	8.45	8.45	8.43	8.45	8.43	8.43	8.43	8.43
Dipole moment (D) of the optimized protein structure	215.5	219.7	228.6	222.5	220.9	224.0	231.2	224.1	222.3

Table 3. Mean RMSD values (nm) calculated versus the initially optimized structures, and their standard deviation, obtained from the analysis of the last 20 ns of the solvated Amber99SB-based MD trajectories at 300 K. All atoms are considered in the calculations.

	1UBQ	1Q0W	1BRS
All-atom	$0.23 \pm 0.02$	$0.34 \pm 0.04$	$0.19 \pm 0.01$
<b>CD-based models</b>			
II	$0.86 \pm 0.09$	$0.70 \pm 0.07$	$1.00 \pm 0.04$
IV	$0.57 \pm 0.03$	$0.59 \pm 0.02$	$0.60 \pm 0.05$
VI	$0.53 \pm 0.06$	$1.00 \pm 0.07$	$0.77 \pm 0.09$
XII	$0.47 \pm 0.03$	$1.17 \pm 0.29$	$1.00 \pm 0.04$
<b>CDA-based models</b>			
IIIa	$0.74 \pm 0.18$	$0.95 \pm 0.05$	$0.78 \pm 0.04$
Va	$0.51 \pm 0.06$	$0.71 \pm 0.05$	$0.49 \pm 0.01$
VIIa	$0.63 \pm 0.09$	$1.45 \pm 0.10$	$0.73 \pm 0.07$
IXa	$0.61 \pm 0.03$	$1.63 \pm 0.55$	$0.48 \pm 0.04$

Table 4. Mean gyration radii  $r_G$  (nm), and their standard deviation obtained from the analysis of the last 20 ns of the Amber99SB-TIP4P-Ew MD trajectories at 300 K.

	1UBQ	1Q0W	1BRS
All-atom	$1.18 \pm 0.01$	$1.37 \pm 0.01$	$1.76 \pm 0.01$
<b>CD-based models</b>			
II	$1.40 \pm 0.03$	$1.61 \pm 0.04$	$1.93 \pm 0.03$
IV	$1.27 \pm 0.01$	$1.47 \pm 0.02$	$1.98 \pm 0.04$
VI	$1.30 \pm 0.01$	$1.58 \pm 0.03$	$1.99 \pm 0.05$
XII	$1.32 \pm 0.02$	$1.94 \pm 0.25$	$1.99 \pm 0.05$
<b>CDa-based models</b>			
IIIa	$1.45 \pm 0.07$	$1.44 \pm 0.02$	$2.00 \pm 0.03$
Va	$1.27 \pm 0.02$	$1.39 \pm 0.02$	$1.86 \pm 0.02$
VIIa	$1.34 \pm 0.02$	$1.74 \pm 0.06$	$1.97 \pm 0.06$
IXa	$1.33 \pm 0.01$	$2.11 \pm 0.31$	$1.93 \pm 0.02$

Table 5. RMSD (nm) of the final protein structure calculated versus the initially optimized structures using VMD [48] from the last 20 ns of the Amber99SB-TIP4P-Ew MD trajectories at 300 K. Only backbone atoms are considered in the calculations.

	1UBQ	1Q0W	1BRS
All-atom	0.123	0.236	0.136
<b>CD-based models</b>			
II	0.891	0.844	0.895
IV	0.484	0.537	0.705
VI	0.530	0.949	0.940
XII	0.355	1.727	0.905
<b>CDa-based models</b>			
IIIa	0.864	0.904	0.636
Va	0.421	0.543	0.409
VIIa	0.596	1.238	0.573
IXa	0.525	2.132	0.526

Table 6. RMSD (nm) and correlation coefficient  $\kappa$  calculated between the simulated Ca RMSF values (RPCM versus all-atom) obtained from the analysis of the last 20 ns of the Amber99SB-TIP4P-Ew MD trajectories at 300 K.

	RMSD			$\kappa$		
	1UBQ	1Q0W	1BRS	1UBQ	1Q0W	1BRS
<b>CD-based models</b>						
II	0.130	0.204	0.214	0.919	0.599	0.192
IV	0.055	0.111	0.251	0.910	0.840	0.590
VI	0.086	0.407	0.360	0.846	0.528	0.310
XII	0.104	0.651	0.457	0.868	0.589	0.505
<b>CDA-based models</b>						
IIIa	0.308	0.100	0.209	0.380	0.680	0.191
Va	0.116	0.158	0.095	0.835	0.708	0.577
VIIa	0.145	0.415	0.267	0.880	0.720	0.334
IXa	0.071	0.873	0.102	0.849	0.681	0.655

Table 7. Mean number of H-bonds and their standard deviation obtained from the analysis of the last 20 ns of the Amber99SB-TIP4P-Ew MD trajectories at 300 K. Integration times  $\tau$  are given for the protein-water and protein-ligand H-bonds.

	Mean number of H-bonds			$\tau$ (ps)	
	intramolecular	protein-water	protein-ligand	protein-water	protein-ligand
<b>1UBQ</b>					
All-atom	55.9 ± 3.9	191.9 ± 7.2	-	104.6	-
<b>CD-based models</b>					
II	7.9 ± 2.6	252.9 ± 8.4	-	287.4	-
IV	18.9 ± 3.1	203.2 ± 7.5	-	218.6	-
IV	11.8 ± 2.7	237.0 ± 8.3	-	353.5	-
XII	12.6 ± 2.9	236.8 ± 7.8	-	372.2	-
<b>CDa-based models</b>					
IIIa	9.0 ± 2.7	245.3 ± 10.2	-	158.8	-
Va	17.8 ± 3.6	202.2 ± 7.8	-	123.9	-
VIIa	13.2 ± 2.8	236.0 ± 7.5	-	240.1	-
IXa	15.2 ± 3.0	205.8 ± 7.0	-	158.0	-
<b>1Q0W</b>					
All-atom	73.2 ± 4.3	282.6 ± 8.6	4.7 ± 1.2	94.6	592.8
<b>CD-based models</b>					
II	13.6 ± 3.4	344.2 ± 11.7	0.8 ± 0.9	215.7	572.5
IV	23.5 ± 3.6	287.9 ± 9.9	2.8 ± 0.9	212.6	795.7
IV	12.6 ± 3.6	340.7 ± 10.2	1.5 ± 0.6	308.8	1151.3
XII	17.6 ± 3.3	324.9 ± 9.4	0.7 ± 0.8	257.8	301.7
<b>CDa-based models</b>					
IIIa	15.0 ± 3.2	336.6 ± 9.3	0.9 ± 0.9	291.8	519.8
Va	21.2 ± 4.0	290.8 ± 9.1	0.9 ± 0.9	192.8	415.1
VIIa	15.9 ± 3.3	336.5 ± 9.3	0.4 ± 0.6	200.0	351.4
IXa	22.0 ± 3.6	298.0 ± 9.0	0.0(2) ± 0.1	129.1	43.8
<b>1BRS</b>					
All-atom	163.6 ± 5.5	452.9 ± 10.1	12.1 ± 1.6	147.9	146.3
<b>CD-based models</b>					
II	39.4 ± 5.1	589.6 ± 15.8	3.8 ± 1.4	319.3	758.6
IV	48.1 ± 5.0	534.5 ± 12.8	1.8 ± 1.1	290.1	359.7
IV	33.4 ± 4.5	589.6 ± 12.5	3.4 ± 1.5	315.6	561.4
XII	34.8 ± 4.4	591.0 ± 12.9	4.5 ± 1.7	333.3	326.7
<b>CDa-based models</b>					
IIIa	53.1 ± 4.8	572.1 ± 12.2	5.1 ± 1.6	255.1	762.8
Va	70.9 ± 5.7	459.3 ± 12.0	4.8 ± 1.4	263.2	725.1
VIIa	54.3 ± 6.1	564.2 ± 14.2	4.1 ± 2.1	247.3	750.7
IXa	69.3 ± 5.7	477.1 ± 11.2	2.2 ± 1.2	212.1	952.3

Table 8. Vps27 UIM-1–Ubiquitin intermolecular H-bonds occurring with an occurrence frequency larger than 10 % during the last 20 ns of the Amber99SB-TIP4P-Ew MD trajectories at 300 K. Contacts also observed with ABC<sup>2</sup> [50] are underlined.

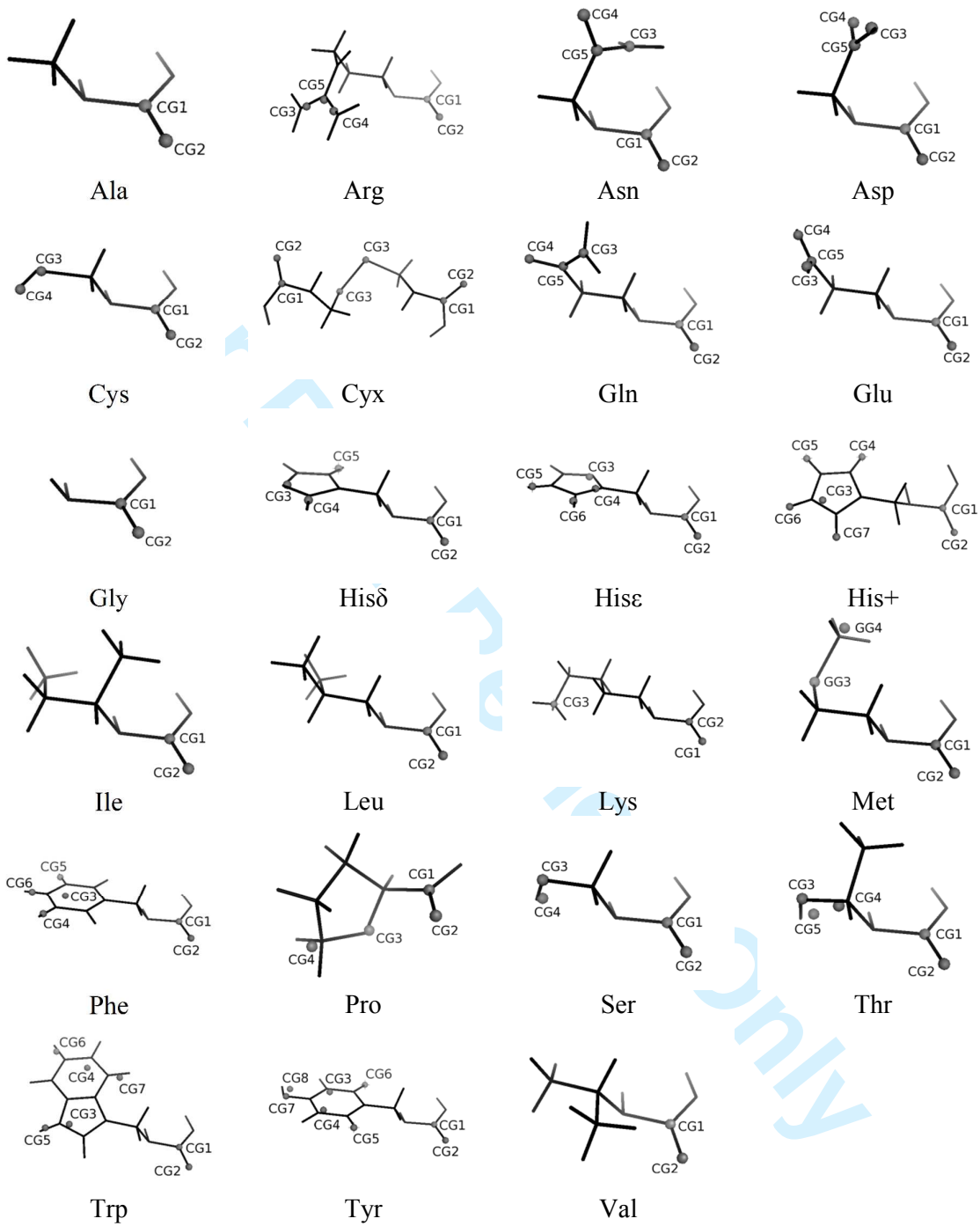
Vps27 - UIM-1	Ubiquitin	All-atom	II	IV	VI	IIIa
<b>Region 1</b>						
glu273	lys6	46.4		20.7		
<b>Region 2</b>						
glu260	arg42	84.8				
<u>ser270</u>	<u>gly47</u>	79.6				
<b>Region 3</b>						
leu271	ser65			60.2		
ser274	lys63		25.9			
glu268	hip68				64.4	
<u>glu273</u>	<u>hip68</u>	85.7		70.5		16.2

Table 9. Barnase–Barstar (residues 1-110 and 111-199) intermolecular H-bonds occurring with an occurrence frequency larger than 10 % during the last 20 ns of the Amber99SB-TIP4P-Ew MD trajectories at 300 K. Contacts also observed with ABC<sup>2</sup> [50] are underlined.

Barnase	Barstar	All-atom	II	IV	VI	XII	IIIa	Va	VIIa	IXa
<b>Region 1</b>										
lys27	asp149				15.6					
<u>lys27</u>	<u>thr152</u>	57.0								
<b>Region 2</b>										
ser38	gly153								50.0	
ser38	<u>glu156</u>						39.4			20.2
<b>Region 3</b>										
lys27	asp193					15.4				
ser28	glu190		13.0		28.8					
gly34	glu192					12.0				
<b>Region 4</b>										
ile55	trp148				21.3					
<u>phe56</u>	<u>asp145</u>		10.1							
ser56	asp149		14.0							
ser57	asp145		85.9	59.4	26.3	67.5	55.2	89.9		
<u>arg59</u>	<u>asp145</u>	77.2			13.8		58.1	53.3	12.9	
<u>arg59</u>	<u>trp148</u>	41.4	40.9	19.5			13.9	44.3		13.9
<u>glu60</u>	<u>leu144</u>	22.5	10.7		10.5					
<u>glu60</u>	<u>asp145</u>		46.6	16.3					27.0	
lys62	asp145									
lys62	leu147		10.8							
<b>Region 5</b>										
<u>arg59</u>	<u>glu186</u>	96.1								
<b>Region 6</b>										
phe82	tyr139			12.1						
<u>arg83</u>	<u>tyr139</u>	36.3								37.5
ser85	<u>tyr139</u>		33.9			10.2	55.1			
<b>Region 7</b>										
<u>phe82</u>	<u>trp154</u>					13.2				
<u>arg83</u>	<u>asp149</u>	66.5				21.6		46.9		
<u>arg83</u>	<u>gly153</u>	15.6								
ser85	asp149						58.7			
arg87	tyr139								44.5	
<u>arg87</u>	<u>asp149</u>	98.8								26.1
<b>Region 8</b>										
<u>hie102</u>	<u>tyr139</u>		27.8			21.6		46.9		
<u>hie102</u>	<u>tyr140</u>	15.5								
<u>hie102</u>	<u>gly141</u>	86.7					32.8	12.3		15.8
<u>hie102</u>	<u>asn143</u>	62.2						15.3	15.3	
<u>hie102</u>	<u>asp149</u>	92.2					73.9		12.5	17.6
<u>tyr103</u>	<u>asn143</u>		10.1	13.0	13.0		17.4			
<u>tyr103</u>	<u>asp149</u>					53.8	30.1			11.1
<u>gln104</u>	<u>asn143</u>								10.9	
<b>Additional H-bonds</b>										
gly52	asp193					15.9				
gly53	glu190					21.3				



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



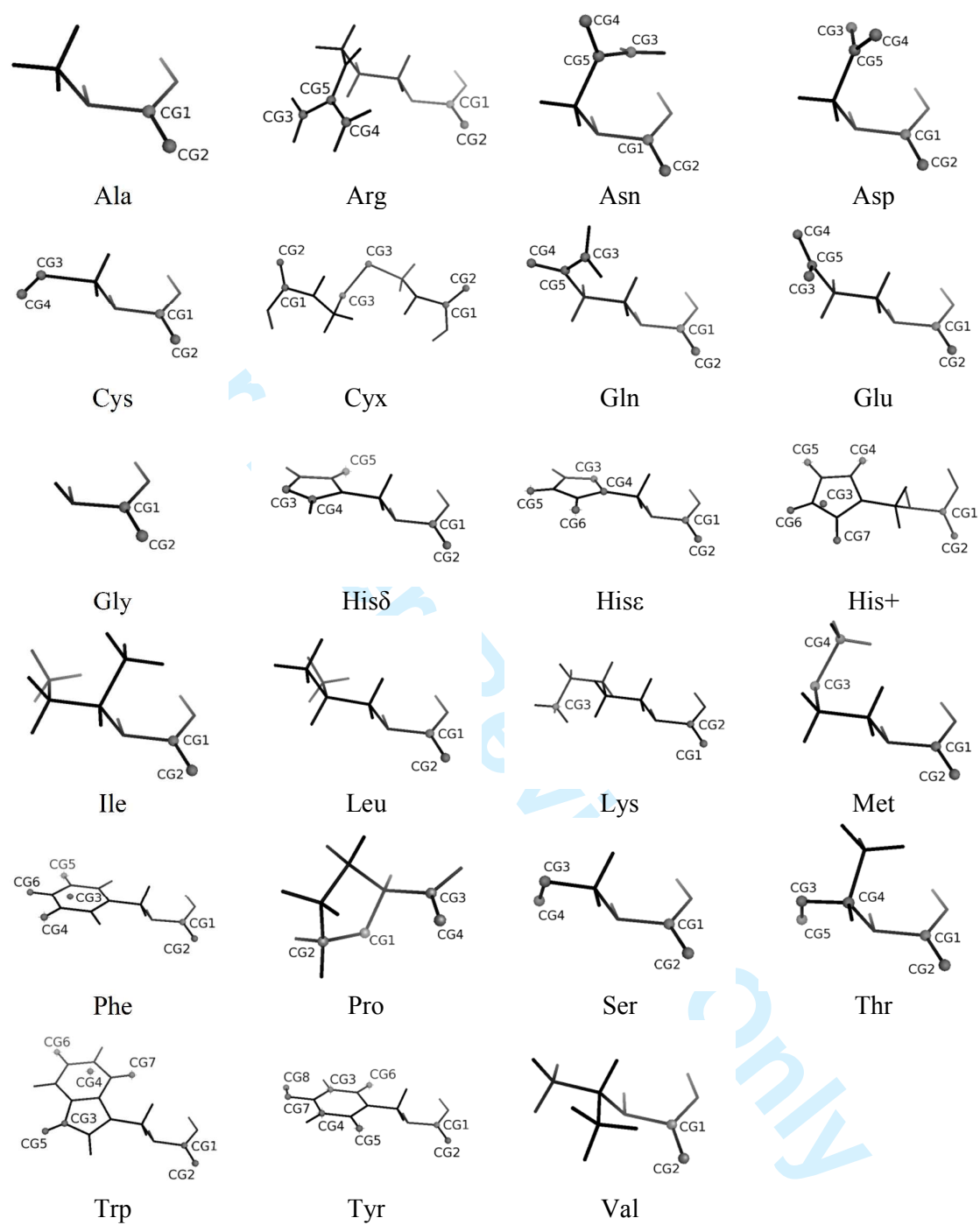


Figure 1.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

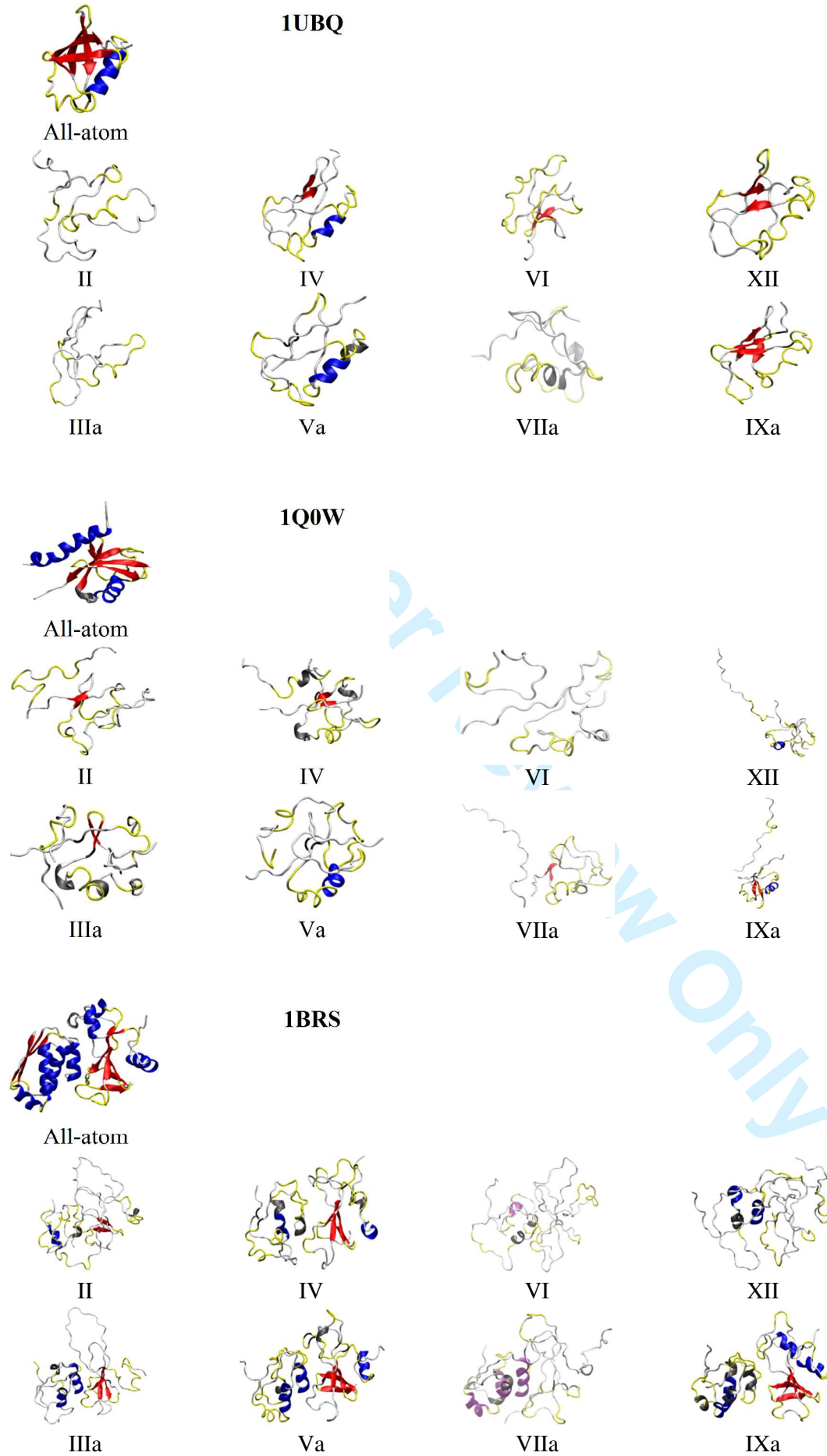


Figure 2.

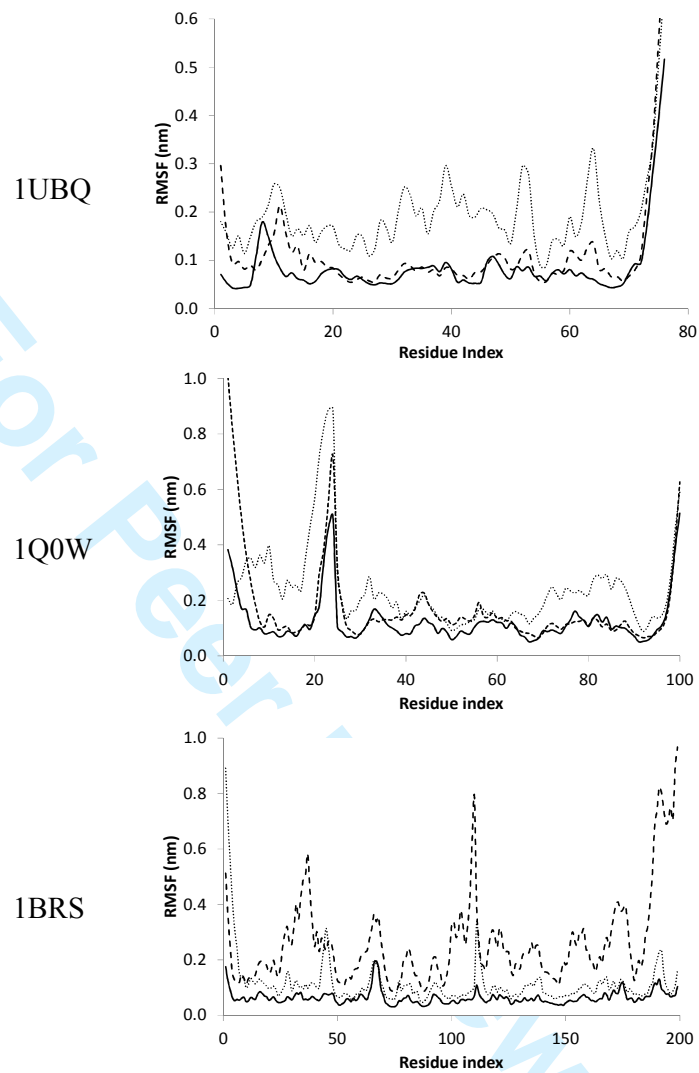


Figure 3.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

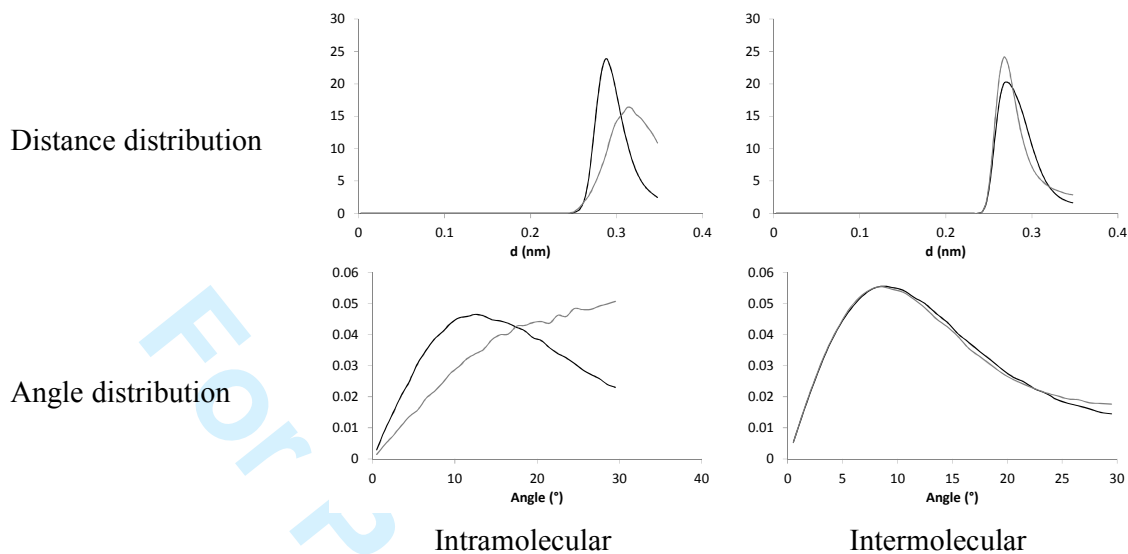


Figure 4.

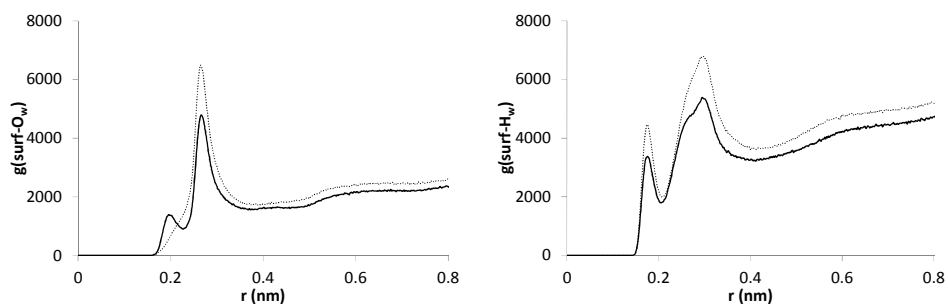


Figure 5.

For Peer Review Only

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

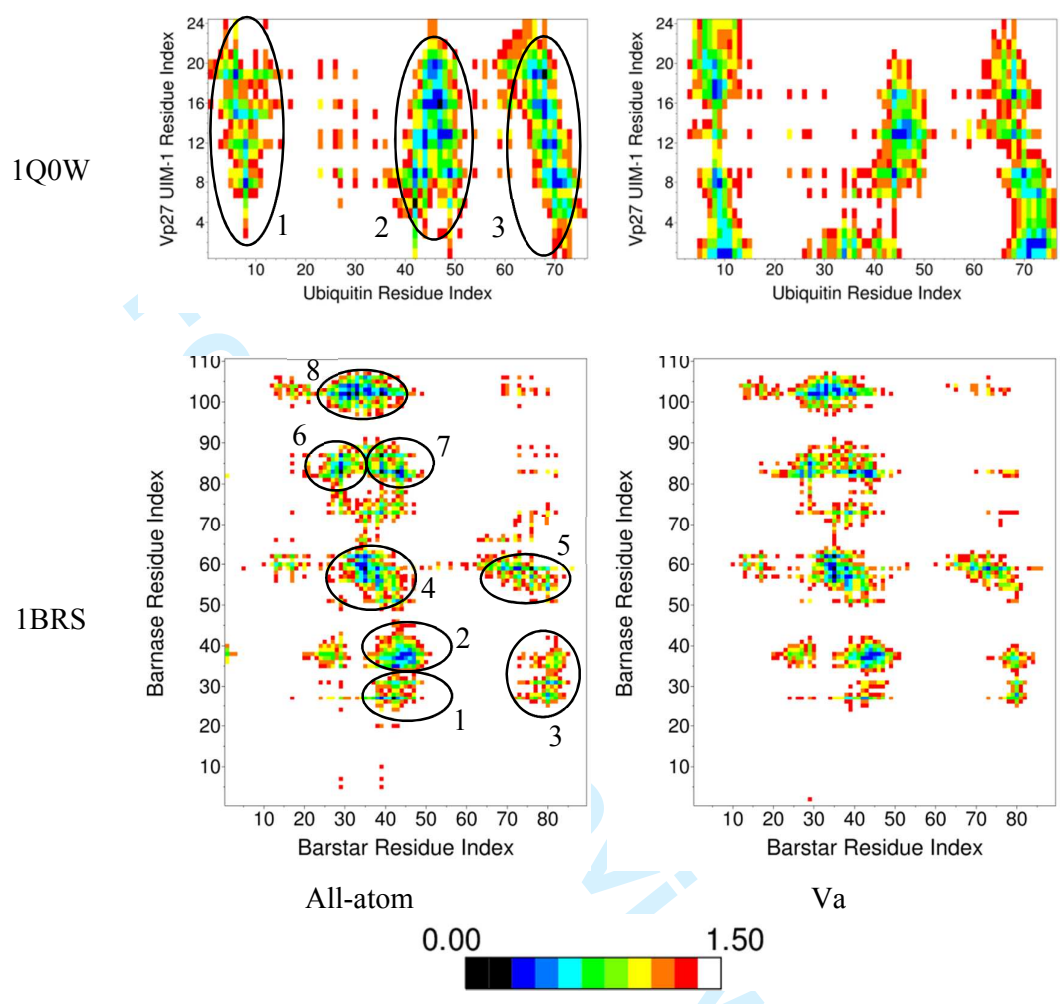


Figure 6.

## Reduced point charge models of proteins: Assessment based on molecular dynamics simulations

Laurence LEHERTE  
 Department of Chemistry  
 Unité de Chimie Physique Théorique et Structurale  
 Laboratoire de Physico-Chimie Informatique  
 Namur MEDicine & Drug Innovation Center (NAMEIDIC)  
 University of Namur, Rue de Bruxelles 61, B-5000 Namur (Belgium)

**SI 1.** Point charge representation of the Amber99-based model IV. Charges not located on atoms are defined as virtual sites (vs) versus reference atoms (Atom\_n). The parameters *a*, *b*, and *c*, are determined such as :

$$\mathbf{r}_{\text{vs}} = \mathbf{r}_1 + a\mathbf{r}_{12} + b\mathbf{r}_{13} + c(\mathbf{r}_{12} \times \mathbf{r}_{13})$$

When no parameters are given, the charge CGx is located on its corresponding Atom\_1.

Residue code	Charge location	Reference atoms			GROMACS virtual site parameters			Charge value (e)
		Atom_1	Atom_2	Atom_3	a	b	c	
ALA	C							0.7211
	O							-0.7211
ARG	C							0.7199
	O							-0.7561
	CG3	CZ	NH1	NH2	0.782449441	0.048673435	0.0000298755	0.3292
	CG4	CZ	NH1	NH2	0.048755496	0.782565421	0.0000292582	0.3299
	CG5	CZ	NH1	NH2	0.107102377	0.14967938	0.003881976	0.3777
ASN	C							0.7266
	O							-0.7219
	CG5	CG						0.2987
	CG4	OD1						-0.5444
	CG3	ND2						0.2410
ASP	C							0.5942



1									
2									
3									
4									
5		O							-0.7362
6		CG3	CG	OD1	OD2	0.118215768	0.867367856	0.001275297	-0.7939
7		CG4	CG	OD1	OD2	0.86752938	0.1184395	0.002555006	-0.7939
8		CG5	CG						0.7299
9									
10	CYS	C							0.7512
11		O							-0.7319
12		CG3	CB	SG	HG	0.917100073	-0.00000025231	-0.000019307	-0.1452
13		CG4	CB	SG	HG	-0.02909584	0.985199139	0.005564586	0.1259
14	CYX	C							0.7139
15		O							-0.7088
16		CG3	S						-0.0051
17	GLN	C							0.6751
18		O							-0.7076
19		CG5	CD						0.4463
20		CG4	OE1						-0.6013
21		CG3	NE2						0.1874
22									
23	GLU	C							0.6188
24		O							-0.7367
25		CG3	CD	OE1	OE2	0.11818935	0.8711497	-0.00048537	-0.7987
26		CG4	CD	OE1	OE2	0.870186314	0.116540793	-0.00024918	-0.7987
27		CG5	CD						0.7153
28	GLY	C							0.7301
29		O							-0.7301
30									
31	HID	C							0.6939
32		O							-0.7014
33		CG3	ND1	NE2	CD2	0.916320707	0.005196893	-0.00000063655	-0.4024
34		CG4	ND1	NE2	CD2	-0.08027802	1.223375206	0.0000958419	0.1336
35		CG5	ND1	NE2	CD2	-0.11571336	-0.27227644	-0.00244624	0.2763
36	HIE	C							0.6641
37		O							-0.7006
38		CG3	ND1	NE2	CD2	0.061007215	0.04812896	-0.0001865	-0.3728
39		CG4	ND1	NE2	CD2	-0.46203366	0.963349084	0.000106117	0.0918
40									
41									
42									
43									
44									
45									
46									
47									
48									
49									

5		CG5	ND1	NE2	CD2	1.559582406	-0.25855641	0.000186767	0.2596
6		CG6	ND1	NE2	CD2	-0.09843092	1.350020156	-0.000060777	0.0578
7	HIP	C							0.8247
8		O							-0.7673
9		CG3	ND1	NE2	CD2	0.615886733	0.271621075	-0.000024757	-0.3486
10		CG4	HD1						0.3615
11		CG5	HD2						0.2917
12		CG6	HE2						0.3743
13		CG7	HE1						0.2637
14									
15	ILE	C							0.7271
16		O							-0.7271
17									
18	LEU	C							0.7208
19		O							-0.7208
20									
21	LYS	C							0.6904
22		O							-0.7163
23		CG3	NZ						1.0259
24	MET	C							0.6439
25		O							-0.6940
26		CG3	CG	SD	CE	0.959905842	0.039588611	0.000398866	-0.1213
27		CG4	CG	SD	CE	-0.20419381	1.18524695	0.004993253	0.1714
28									
29	PHE	C							0.6940
30		O							-0.7130
31		CG3	CE1	CZ	CE2	-0.30176242	0.650745242	0.0000307846	-0.2683
32		CG4	CE1	CZ	CE2	0.624252298	0.972278293	0.000431521	0.1208
33		CG5	CE1	CZ	CE2	0.640402301	-0.6043884	0.000446	0.1208
34		CG6	CE1	CZ	CE2	2.077589666	-0.53877972	0.000204656	0.0458
35									
36	PRO	C							0.2745
37		O							-0.5145
38		CG3	N	CG	CD	0.099466032	-0.1612276	0.037641373	0.0646
39		CG4	N	CG	CD	-0.08850188	1.318442433	0.004996816	0.1754
40									
41	SER	C							0.6418
42		O							-0.6967

1									
2									
3									
4									
5		CG3	OG					-0.3564	
6		CG4	CB	OG	HG	-0.03421213	0.922118716	-0.00905747	0.4114
7	THR	C							0.6520
8		O							-0.7046
9		CG3	CB	OG1	HG1	0.982727085	-0.01308135	-0.02208364	-0.6101
10		CG4	CB	OG1	HG1	-0.18228543	0.361268843	-0.05627379	0.2405
11		CG5	CB	OG1	HG1	-0.08275132	0.791948979	-0.02786446	0.4222
12	TRP	C							0.6766
13		O							-0.7062
14		CG3	CD1	NE1	CE2	0.491758761	0.151567209	0.000125363	-0.2282
15		CG4	CH2	CZ3	CE3	-0.00267307	0.528431507	-0.000094374	-0.4027
16		CG5	CD1	NE1	CE2	1.870902148	-0.40053691	-0.00062521	0.3142
17		CG6	CH2	CZ3	CE3	0.424622379	-0.46909705	-0.000041485	0.1481
18		CG7	CH2	CZ3	CE3	0.188659273	1.148274727	0.016218941	0.1982
19	TYR	C							0.6709
20		O							-0.6992
21		CG3	CZ	CD1	CD2	-0.16654319	0.578653277	0.0000858705	-0.1152
22		CG4	CZ	CD1	CD2	0.578716566	-0.16633363	-0.0000063462	-0.1152
23		CG5	CZ	CD1	CD2	1.492593637	-0.278735759	0.000115166	0.0899
24		CG6	CZ	CD1	CD2	-0.26746548	1.502654876	-0.00491085	0.0899
25		CG7	CZ	OH	HH	0.913216448	0.0000457516	-0.000013018	-0.4272
26		CG8	CZ	OH	HH	-0.15052509	0.849156187	-0.0025143	0.5061
27	VAL	C							0.7206
28		O							-0.7206
29									
30									
31									
32									
33									
34									
35									
36									
37									
38									
39									
40									
41									
42									
43									
44									
45									
46									
47									
48									
49									

## Reduced point charge models of proteins: Assessment based on molecular dynamics simulations

Laurence LEHERTE  
 Department of Chemistry  
 Unité de Chimie Physique Théorique et Structurale  
 Laboratoire de Physico-Chimie Informatique  
 Namur MEDicine & Drug Innovation Center (NAMEDIC)  
 University of Namur, Rue de Bruxelles 61, B-5000 Namur (Belgium)

**SI 2.** Point charge representation of the Amber99-based model VI. Charges not located on atoms are defined as virtual sites (vs) versus reference atoms (Atom\_n). The parameters *a*, *b*, and *c*, are determined such as :

$$\mathbf{r}_{vs} = \mathbf{r}_1 + a\mathbf{r}_{12} + b\mathbf{r}_{13} + c(\mathbf{r}_{12} \times \mathbf{r}_{13})$$

When no parameters are given, the charge CGx is located on its corresponding Atom\_1.

Residue code	Charge location	Reference atoms			GROMACS virtual site parameters			Charge value (e <sup>-</sup> )
		Atom_1	Atom_2	Atom_3	a	b	c	
ALA	C							0.8215
	O							-0.8215
ARG	C							0.7905
	O							-0.8268
	CG3	CZ	NH1	NH2	0.782449441	0.048673435	0.0000298755	-0.2466
	CG4	CZ	NH1	NH2	0.048755496	0.782565421	0.0000292582	-0.2466
	CG5	CZ	NH1	NH2	0.107102377	0.14967938	0.003881976	1.5295
ASN	C							0.8293
	O							-0.8246
	CG5	CG						0.3047
	CG4	OD1						-0.5826
	CG3	ND2						0.2732
ASP	C							0.6941

1									
2									
3									
4									
5		O							-0.8361
6		CG3	CG	OD1	OD2	0.118215768	0.867367856	0.001275297	-0.7504
7		CG4	CG	OD1	OD2	0.86752938	0.1184395	0.002555006	-0.7504
8		CG5	CG						0.6426
9	CYS	C							0.8127
10		O							-0.7934
11		CG3	CB	SG	HG	0.917100073	-0.00000025231	-0.000019307	-0.1233
12		CG4	CB	SG	HG	-0.02909584	0.985199139	0.005564586	0.1040
13	CYX	C							0.7368
14		O							-0.7317
15		CG3	S						-0.0051
16	GLN	C							0.7645
17		O							-0.7970
18		CG5	CD						0.5078
19		CG4	OE1						-0.6647
20		CG3	NE2						0.1894
21	GLU	C							0.7213
22		O							-0.8391
23		CG3	CD	OE1	OE2	0.11818935	0.8711497	-0.00048537	-0.7659
24		CG4	CD	OE1	OE2	0.870186314	0.116540793	-0.00024918	-0.7659
25		CG5	CD						0.6496
26	GLY	C							0.8092
27		O							-0.8092
28	HID	C							0.8114
29		O							-0.8189
30		CG3	ND1	NE2	CD2	0.916320707	0.005196893	-0.00000063655	-0.3509
31		CG4	ND1	NE2	CD2	-0.08027802	1.223375206	0.0000958419	0.0753
32		CG5	ND1	NE2	CD2	-0.11571336	-0.27227644	-0.00244624	0.2831
33	HIE	C							0.7802
34		O							-0.8167
35		CG3	ND1	NE2	CD2	0.061007215	0.04812896	-0.0001865	-0.3603
36		CG4	ND1	NE2	CD2	-0.46203366	0.963349084	0.000106117	0.0786
37									
38									
39									
40									
41									
42									
43									
44									
45									
46									
47									
48									
49									

5		CG5	ND1	NE2	CD2	1.559582406	-0.25855641	0.000186767	0.2716
6		CG6	ND1	NE2	CD2	-0.09843092	1.350020156	-0.000060777	0.0466
7	HIP	C							0.8320
8		O							-0.7747
9		CG3	ND1	NE2	CD2	0.615886733	0.271621075	-0.000024757	-0.2755
10		CG4	HD1						0.3763
11		CG5	HD2						0.3017
12		CG6	HE2						0.3160
13		CG7	HE1						0.2241
14	ILE	C							0.8199
15		O							-0.8199
16	LEU	C							0.8330
17		O							-0.8330
18	LYS	C							0.6671
19		O							-0.6930
20		CG3	NZ						1.0259
21	MET	C							0.6751
22		O							-0.7252
23		CG3	CG	SD	CE	0.959905842	0.039588611	0.000398866	-0.1007
24		CG4	CG	SD	CE	-0.20419381	1.18524695	0.004993253	0.1508
25	PHE	C							0.7791
26		O							-0.7981
27		CG3	CE1	CZ	CE2	-0.30176242	0.650745242	0.0000307846	-0.1799
28		CG4	CE1	CZ	CE2	0.624252298	0.972278293	0.000431521	0.1105
29		CG5	CE1	CZ	CE2	0.640402301	-0.6043884	0.000446	0.1105
30		CG6	CE1	CZ	CE2	2.077589666	-0.53877972	0.000204656	-0.0221
31	PRO	C							0.2896
32		O							-0.5296
33		CG3	N	CG	CD	0.099466032	-0.1612276	0.037641373	0.0901
34		CG4	N	CG	CD	-0.08850188	1.318442433	0.004996816	0.1499
35	SER	C							0.7066
36		O							-0.7616

1									
2									
3									
4									
5		CG3	OG					-0.3445	
6		CG4	CB	OG	HG	-0.03421213	0.922118716	-0.00905747	0.3995
7	THR	C							0.7441
8		O							-0.7967
9		CG3	CB	OG1	HG1	0.982727085	-0.01308135	-0.02208364	-0.6236
10		CG4	CB	OG1	HG1	-0.18228543	0.361268843	-0.05627379	0.2482
11		CG5	CB	OG1	HG1	-0.08275132	0.791948979	-0.02786446	0.4280
12	TRP	C							0.7733
13		O							-0.8029
14		CG3	CD1	NE1	CE2	0.491758761	0.151567209	0.000125363	-0.2072
15		CG4	CH2	CZ3	CE3	-0.00267307	0.528431507	-0.000094374	-0.4989
16		CG5	CD1	NE1	CE2	1.870902148	-0.40053691	-0.00062521	0.3135
17		CG6	CH2	CZ3	CE3	0.424622379	-0.46909705	-0.000041485	0.1697
18		CG7	CH2	CZ3	CE3	0.188659273	1.148274727	0.016218941	0.2525
19	TYR	C							0.7770
20		O							-0.8052
21		CG3	CZ	CD1	CD2	-0.16654319	0.578653277	0.0000858705	-0.1357
22		CG4	CZ	CD1	CD2	0.578716566	-0.16633363	-0.0000063462	-0.1357
23		CG5	CZ	CD1	CD2	1.492593637	-0.278735759	0.000115166	0.0966
24		CG6	CZ	CD1	CD2	-0.26746548	1.502654876	-0.00491085	0.0966
25		CG7	CZ	OH	HH	0.913216448	0.0000457516	-0.000013018	-0.3892
26		CG8	CZ	OH	HH	-0.15052509	0.849156187	-0.0025143	0.4956
27	VAL	C							0.8302
28		O							-0.8302
29									
30									
31									
32									
33									
34									
35									
36									
37									
38									
39									
40									
41									
42									
43									
44									
45									
46									
47									
48									
49									

## Reduced point charge models of proteins: Assessment based on molecular dynamics simulations

Laurence LEHERTE  
 Department of Chemistry  
 Unité de Chimie Physique Théorique et Structurale  
 Laboratoire de Physico-Chimie Informatique  
 Namur MEDicine & Drug Innovation Center (NAMEIDIC)  
 University of Namur, Rue de Bruxelles 61, B-5000 Namur (Belgium)

**SI 3.** Point charge representation of the Amber99-based model XII. Charges not located on atoms are defined as virtual sites (vs) versus reference atoms (Atom\_n). The parameters *a*, *b*, and *c*, are determined such as :

$$\mathbf{r}_{\text{vs}} = \mathbf{r}_1 + a\mathbf{r}_{12} + b\mathbf{r}_{13} + c(\mathbf{r}_{12} \times \mathbf{r}_{13})$$

When no parameters are given, the charge CGx is located on its corresponding Atom\_1.

Residue code	Charge location	Reference atoms			GROMACS virtual site parameters			Charge value (e)
		Atom_1	Atom_2	Atom_3	a	b	c	
ALA	C							0.8516
	O							-0.8516
ARG	C							0.8109
	O							-0.8470
	CG3	CZ	NH1	NH2	0.782449441	0.048673435	0.0000298755	-0.4412
	CG4	CZ	NH1	NH2	0.048755496	0.782565421	0.0000292582	-0.4412
	CG5	CZ	NH1	NH2	0.107102377	0.14967938	0.003881976	1.9185
ASN	C							0.8594
	O							-0.8547
	CG5	CG						0.3265
	CG4	OD1						-0.5946
	CG3	ND2						0.2634
ASP	C							0.7231



1									
2									
3									
4									
5		O							-0.8651
6		CG3	CG	OD1	OD2	0.118215768	0.867367856	0.001275297	-0.7292
7		CG4	CG	OD1	OD2	0.86752938	0.1184395	0.002555006	-0.7292
8		CG5	CG						0.6004
9	CYS	C							0.8235
10		O							-0.8042
11		CG3	CB	SG	HG	0.917100073	-0.00000025231	-0.000019307	-0.1163
12		CG4	CB	SG	HG	-0.02909584	0.985199139	0.005564586	0.0970
13									
14	CYX	C							0.7703
15		O							-0.7652
16		CG3	S						-0.0051
17	GLN	C							0.7903
18		O							-0.8230
19		CG5	CD						0.5543
20		CG4	OE1						-0.6859
21		CG3	NE2						0.1641
22									
23	GLU	C							0.7480
24		O							-0.8658
25		CG3	CD	OE1	OE2	0.11818935	0.8711497	-0.00048537	-0.7494
26		CG4	CD	OE1	OE2	0.870186314	0.116540793	-0.00024918	-0.7493
27		CG5	CD						0.6165
28									
29	GLY	C							0.8381
30		O							-0.8381
31	HID	C							0.8212
32		O							-0.8287
33		CG3	ND1	NE2	CD2	0.916320707	0.005196893	-0.00000063655	-0.3290
34		CG4	ND1	NE2	CD2	-0.08027802	1.223375206	0.0000958419	0.0566
35		CG5	ND1	NE2	CD2	-0.11571336	-0.27227644	-0.00244624	0.2799
36									
37	HIE	C							0.7990
38		O							-0.8355
39		CG3	ND1	NE2	CD2	0.061007215	0.04812896	-0.0001865	-0.3689
40		CG4	ND1	NE2	CD2	-0.46203366	0.963349084	0.000106117	0.0948
41									
42									
43									
44									
45									
46									
47									
48									
49									

5		CG5	ND1	NE2	CD2	1.559582406	-0.25855641	0.000186767	0.2784
6		CG6	ND1	NE2	CD2	-0.09843092	1.350020156	-0.000060777	0.0322
7	HIP	C							0.8674
8		O							-0.8101
9		CG3	ND1	NE2	CD2	0.615886733	0.271621075	-0.000024757	-0.2522
10		CG4	HD1						0.3769
11		CG5	HD2						0.3113
12		CG6	HE2						0.2878
13		CG7	HE1						0.2189
14	ILE	C							0.8454
15		O							-0.8454
16	LEU	C							0.8562
17		O							-0.8562
18	LYS	C							0.6731
19		O							-0.6990
20		CG3	NZ						1.0259
21	MET	C							0.6838
22		O							-0.7339
23		CG3	CG	SD	CE	0.959905842	0.039588611	0.000398866	-0.0989
24		CG4	CG	SD	CE	-0.20419381	1.18524695	0.004993253	0.1490
25	PHE	C							0.8020
26		O							-0.8210
27		CG3	CE1	CZ	CE2	-0.30176242	0.650745242	0.0000307846	-0.1734
28		CG4	CE1	CZ	CE2	0.624252298	0.972278293	0.000431521	0.1232
29		CG5	CE1	CZ	CE2	0.640402301	-0.6043884	0.000446	0.1232
30		CG6	CE1	CZ	CE2	2.077589666	-0.53877972	0.000204656	-0.0540
31	PRO	C							0.2966
32		O							-0.5366
33		CG3	N	CG	CD	0.099466032	-0.1612276	0.037641373	0.0935
34		CG4	N	CG	CD	-0.08850188	1.318442433	0.004996816	0.1465
35	SER	C							0.7291
36		O							-0.7841

1									
2									
3									
4									
5		CG3	OG						-0.3261
6		CG4	CB	OG	HG	-0.03421213	0.922118716	-0.00905747	0.3811
7	THR	C							0.7732
8		O							-0.8258
9		CG3	CB	OG1	HG1	0.982727085	-0.01308135	-0.02208364	-0.6106
10		CG4	CB	OG1	HG1	-0.18228543	0.361268843	-0.05627379	0.2652
11		CG5	CB	OG1	HG1	-0.08275132	0.791948979	-0.02786446	0.3980
12	TRP	C							0.7997
13		O							-0.8293
14		CG3	CD1	NE1	CE2	0.491758761	0.151567209	0.000125363	-0.1891
15		CG4	CH2	CZ3	CE3	-0.00267307	0.528431507	-0.000094374	-0.6069
16		CG5	CD1	NE1	CE2	1.870902148	-0.40053691	-0.00062521	0.3052
17		CG6	CH2	CZ3	CE3	0.424622379	-0.46909705	-0.000041485	0.2164
18		CG7	CH2	CZ3	CE3	0.188659273	1.148274727	0.016218941	0.3040
19	TYR	C							0.8059
20		O							-0.8341
21		CG3	CZ	CD1	CD2	-0.16654319	0.578653277	0.0000858705	-0.1703
22		CG4	CZ	CD1	CD2	0.578716566	-0.16633363	-0.0000063462	-0.1703
23		CG5	CZ	CD1	CD2	1.492593637	-0.278735759	0.000115166	0.1147
24		CG6	CZ	CD1	CD2	-0.26746548	1.502654876	-0.00491085	0.1146
25		CG7	CZ	OH	HH	0.913216448	0.0000457516	-0.000013018	-0.3306
26		CG8	CZ	OH	HH	-0.15052509	0.849156187	-0.0025143	0.4701
27	VAL	C							0.8598
28		O							-0.8598
29									
30									
31									
32									
33									
34									
35									
36									
37									
38									
39									
40									
41									
42									
43									
44									
45									
46									
47									
48									
49									

## Reduced point charge models of proteins: Assessment based on molecular dynamics simulations

Laurence LEHERTE  
 Department of Chemistry  
 Unité de Chimie Physique Théorique et Structurale  
 Laboratoire de Physico-Chimie Informatique  
 Namur MEDicine & Drug Innovation Center (NAMEDIC)  
 University of Namur, Rue de Bruxelles 61, B-5000 Namur (Belgium)

**SI 4.** Point charge representation of the Amber99-based model Va. Charges not located on atoms are defined as virtual sites (vs) versus reference atoms (Atom\_n). The parameters *a*, *b*, and *c*, are determined such as :

$$\mathbf{r}_{vs} = \mathbf{r}_1 + a\mathbf{r}_{12} + b\mathbf{r}_{13} + c(\mathbf{r}_{12} \times \mathbf{r}_{13})$$

When no parameters are given, the charge CGx is located on its corresponding Atom\_1.

Residue code	Charge location	Reference atoms			GROMACS virtual site parameters			Charge value (e <sup>-</sup> )
		Atom_1	Atom_2	Atom_3	a	b	c	
ALA	C							0.6656
	O							-0.6656
ARG	C							0.6629
	O							-0.6991
	CG3	NH1						0.2845
	CG4	NH2						0.2845
	CG5	CZ						0.4672
ASN	C							0.6675
	O							-0.6628
	CG3	ND2						0.2445
	CG4	OD1						-0.5204
	CG5	CG						0.2712
ASP	C							0.5195



5		CG5	HE2						0.2266
6		CG6	HD2						0.0647
7	HIP	C							0.7382
8		O							-0.6809
9		CG3	ND1	NE2	CD2	0.615886733	0.271621075	-0.000024757	-0.0640
10		CG4	HD1						0.3382
11		CG5	HE1						0.2042
12		CG6	HE2						0.2779
13		CG7	HD2						0.1864
14	ILE	C							0.6576
15		O							-0.6576
16	LEU	C							0.6636
17		O							-0.6636
18	LYS	C							0.6395
19		O							-0.6654
20		CG3	NZ						1.0259
21	MET	C							0.5884
22		O							-0.6385
23		CG3	SD						-0.1476
24		CG4	CE						0.1977
25	PHE	C							0.6366
26		O							-0.6556
27		CG3	CE1	CZ	CE2	-0.30176242	0.650745242	0.0000307846	-0.2196
28		CG4	HE1						0.0940
29		CG5	HE2						0.0940
30		CG6	HZ						0.0506
31	PRO	C							0.2358
32		O							-0.4758
33		CG3	N						0.0173
34		CG4	CD						0.2227
35	SER	C							0.5909
36		O							-0.6409



## Reduced point charge models of proteins: Assessment based on molecular dynamics simulations

Laurence LEHERTE  
 Department of Chemistry  
 Unité de Chimie Physique Théorique et Structurale  
 Laboratoire de Physico-Chimie Informatique  
 Namur MEDicine & Drug Innovation Center (NAME DIC)  
 University of Namur, Rue de Bruxelles 61, B-5000 Namur (Belgium)

**SI 5.** Point charge representation of the Amber99-based model VIIa. Charges not located on atoms are defined as virtual sites (vs) versus reference atoms (Atom\_n). The parameters *a*, *b*, and *c*, are determined such as :

$$\mathbf{r}_{\text{vs}} = \mathbf{r}_1 + a\mathbf{r}_{12} + b\mathbf{r}_{13} + c(\mathbf{r}_{12} \times \mathbf{r}_{13})$$

When no parameters are given, the charge CGx is located on its corresponding Atom\_1.

Residue code	Charge location	Reference atoms			GROMACS virtual site parameters			Charge value (e <sup>-</sup> )
		Atom_1	Atom_2	Atom_3	a	b	c	
ALA	C							0.7768
	O							-0.7768
ARG	C							0.7539
	O							-0.7901
	CG3	NH1						0.0244
	CG4	NH2						0.0244
	CG5	CZ						1.0850
ASN	C							0.7825
	O							-0.7778
	CG3	ND2						0.2736
	CG4	OD1						-0.5579
	CG5	CG						0.2796
ASP	C							0.6316





	CG5	HE2						0.2363
	CG6	HD2						0.0526
HIP	C							0.8236
	O							-0.7663
	CG3	ND1	NE2	CD2	0.615886733	0.271621075	-0.000024757	-0.0205
	CG4	HD1						0.3404
	CG5	HE1						0.1783
	CG6	HE2						0.2448
	CG7	HD2						0.1997
ILE	C							0.7750
	O							-0.7750
LEU	C							0.7864
	O							-0.7864
LYS	C							0.6369
	O							-0.6628
	CG3	NZ						1.0259
MET	C							0.6336
	O							-0.6837
	CG3	SD						-0.1221
	CG4	CE						0.1722
PHE	C							0.7330
	O							-0.7520
	CG3	CE1	CZ	CE2	-0.30176242	0.650745242	0.0000307846	-0.1380
	CG4	HE1						0.0806
	CG5	HE2						0.0806
	CG6	HZ						-0.0042
PRO	C							0.2491
	O							-0.4891
	CG3	N						0.0430
	CG4	CD						0.1970
SER	C							0.6704
	O							-0.7254



## Reduced point charge models of proteins: Assessment based on molecular dynamics simulations

Laurence LEHERTE  
 Department of Chemistry  
 Unité de Chimie Physique Théorique et Structurale  
 Laboratoire de Physico-Chimie Informatique  
 Namur MEDicine & Drug Innovation Center (NAMEIDIC)  
 University of Namur, Rue de Bruxelles 61, B-5000 Namur (Belgium)

**SI 6.** Point charge representation of the Amber99-based model IXa. Charges not located on atoms are defined as virtual sites (vs) versus reference atoms (Atom\_n). The parameters *a*, *b*, and *c*, are determined such as :

$$\mathbf{r}_{vs} = \mathbf{r}_1 + a\mathbf{r}_{12} + b\mathbf{r}_{13} + c(\mathbf{r}_{12} \times \mathbf{r}_{13})$$

When no parameters are given, the charge CGx is located on its corresponding Atom\_1.

Residue code	Charge location	Reference atoms			GROMACS virtual site parameters			Charge value (e <sup>-</sup> )
		Atom_1	Atom_2	Atom_3	a	b	c	
ALA	C							0.6656
	O							-0.6656
ARG	C							0.8763
	O							-0.7415
	CG3	NH1						0.3549
	CG4	NH2						0.3549
	CG5	CZ						0.1554
ASN	C							0.5703
	O							-0.6359
	CG3	ND2						0.1558
	CG4	OD1						-0.6213
	CG5	CG						0.5281
ASP	C							0.3820



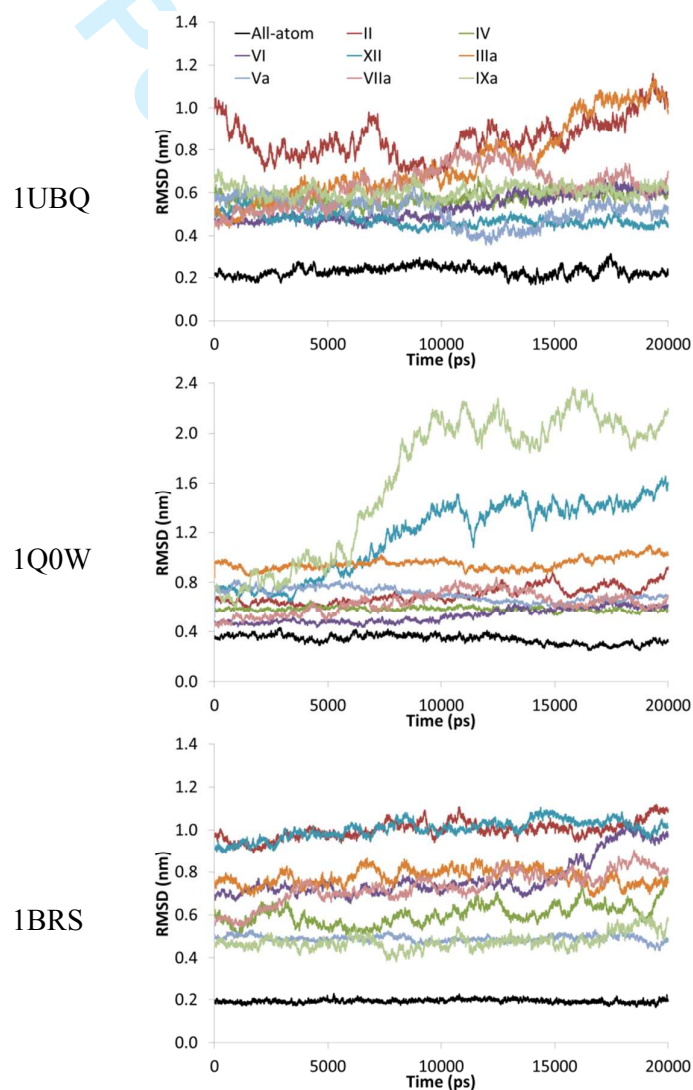
5		CG5	HE2						0.2131
6		CG6	HD2						0.0542
7	HIP	C							0.7493
8		O							-0.6781
9		CG3	ND1	NE2	CD2	0.615886733	0.271621075	-0.000024757	-0.0628
10		CG4	HD1						0.3595
11		CG5	HE1						0.1949
12		CG6	HE2						0.2780
13		CG7	HD2						0.1592
14	ILE	C							0.6576
15		O							-0.6576
16	LEU	C							0.6636
17		O							-0.6636
18	LYS	C							0.6395
19		O							-0.6654
20		CG3	NZ						1.0259
21	MET	C							0.6316
22		O							-0.6508
23		CG3	SD						-0.1645
24		CG4	CE						0.1837
25	PHE	C							0.6492
26		O							-0.6592
27		CG3	CE1	CZ	CE2	-0.30176242	0.650745242	0.0000307846	-0.2229
28		CG4	HE1						0.0925
29		CG5	HE2						0.0924
30		CG6	HZ						0.0480
31	PRO	C							0.2179
32		O							-0.4698
33		CG3	N						0.0547
34		CG4	CD						0.1972
35	SER	C							0.6856
36		O							-0.6689



## Reduced point charge models of proteins: Assessment based on molecular dynamics simulations

Laurence LEHERTE  
 Department of Chemistry  
 Unité de Chimie Physique Théorique et Structurale  
 Laboratoire de Physico-Chimie Informatique  
 Namur MEDICINE & Drug Innovation Center (NAMEDIC)  
 University of Namur, Rue de Bruxelles 61, B-5000 Namur (Belgium)

**SI 7.** RMSD (nm) of the protein atoms calculated versus the initially optimized protein structure. Time evolution is obtained from the analysis of 20 ns AMBER99SB-TIP4P-Ew MD trajectories at 300 K.

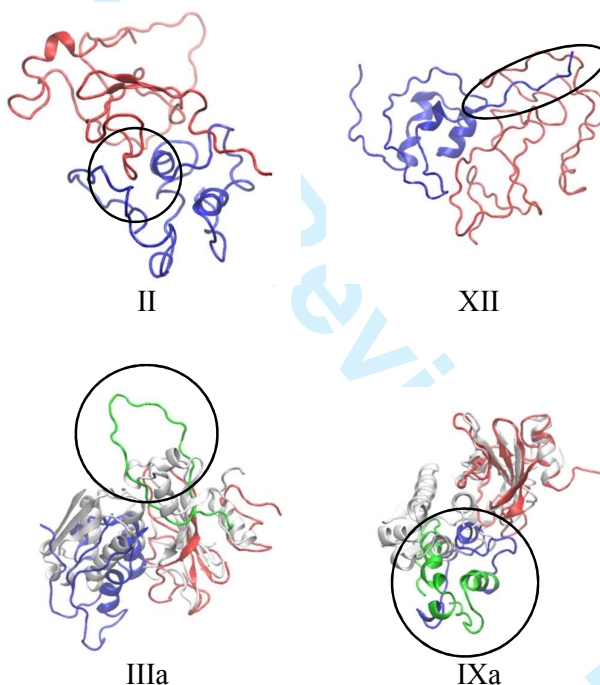




## Reduced point charge models of proteins: Assessment based on molecular dynamics simulations

Laurence LEHERTE  
 Department of Chemistry  
 Unité de Chimie Physique Théorique et Structurale  
 Laboratoire de Physico-Chimie Informatique  
 Namur MEDICINE & Drug Innovation Center (NAMEDIC)  
 University of Namur, Rue de Bruxelles 61, B-5000 Namur (Belgium)

**SI 8.** End frames of the MD trajectories built with models II, XII, IIIa, and IXa for the Barnase (red) - Barstar (blue) complex simulated at 300 K using the Amber99-TIP4P-Ew FFs. Areas mentioned in the manuscript are encircled.



IIIa  
 AA sequence 20 to 50 is  
 displayed in green.

IXa  
 AA sequences 111 to  
 131 and 170 to 199 are  
 displayed in green.

White ribbons correspond to the initially optimized  
 structure.

# Reduced point charge models of proteins: Assessment based on molecular dynamics simulations

Laurence LEHERTE

Department of Chemistry

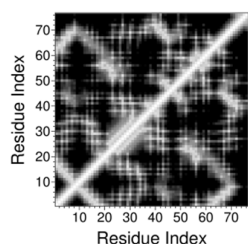
Unité de Chimie Physique Théorique et Structurale

Laboratoire de Physico-Chimie Informatique

Namur Medicine & Drug Innovation Center (NAMEDIC)

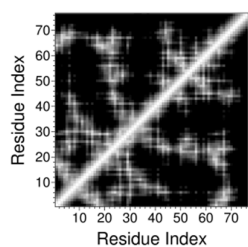
University of Namur, Rue de Bruxelles 61, B-5000 Namur (Belgium)

**SI 9.** Residue-residue mean shortest distance maps calculated from 20 ns AMBER99SB-TIP4P-Ew MD trajectories at 300 K. Residues of the protein complexes are numbered 1 to 24 (Vps27 UIM-1) and 25 to 100 (Ubiquitin) for 1Q0W, and 1 to 110 (Barnase) and 111 to 199 (Barstar) for 1BRS. White to black color-code stands for distances ranging from 0 to 1.5 nm (step = 0.15 nm).

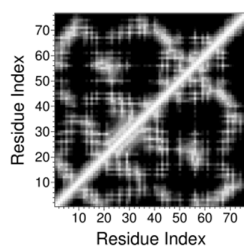


1UBQ

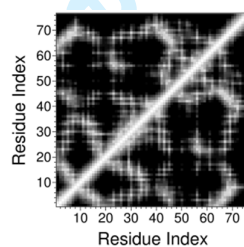
All-atom



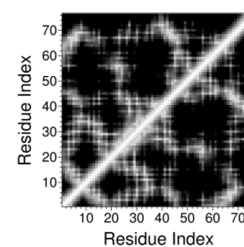
II



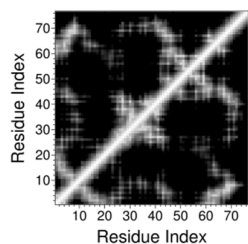
IV



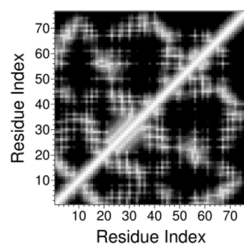
VI



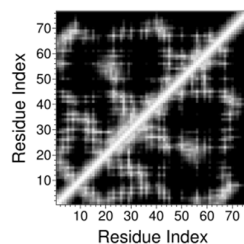
XII



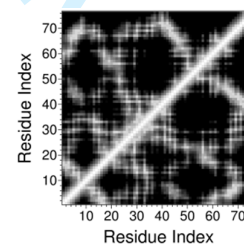
IIIa



Va

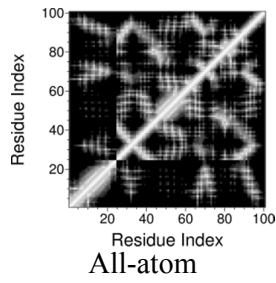


VIIa

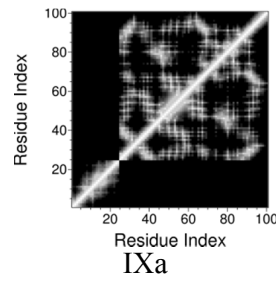
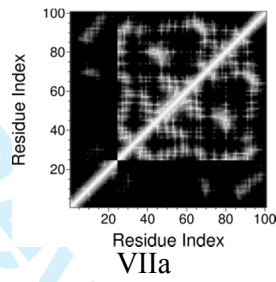
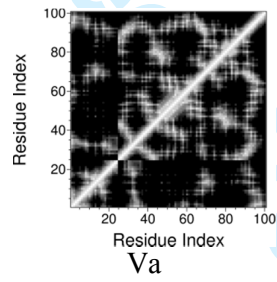
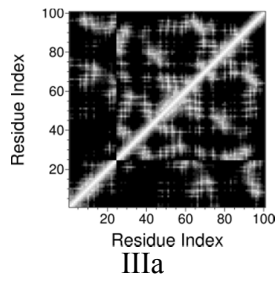
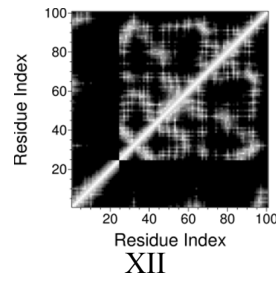
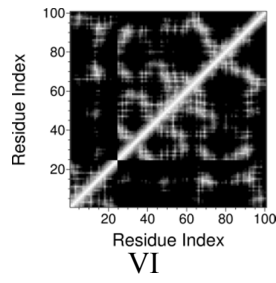
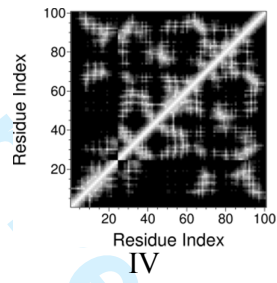
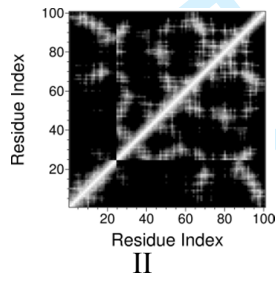


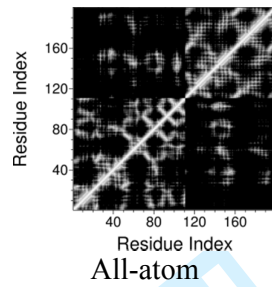
IXa

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

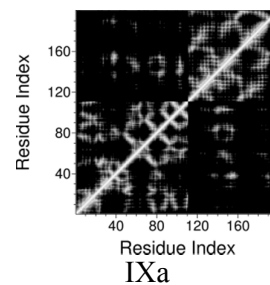
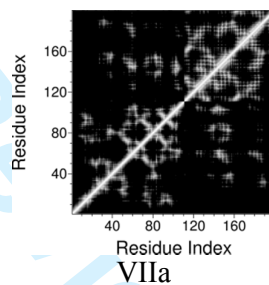
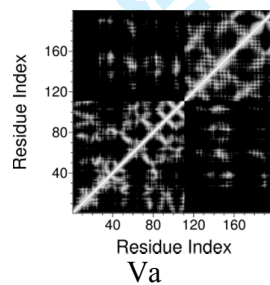
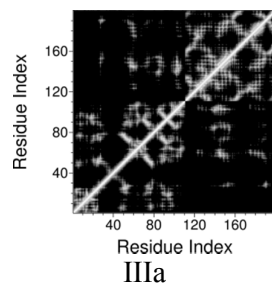
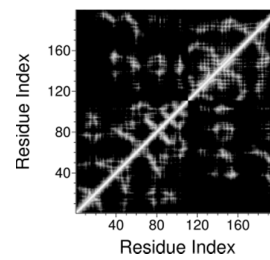
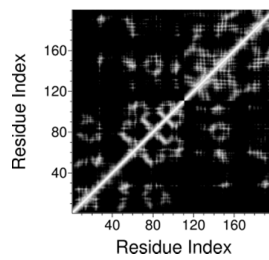
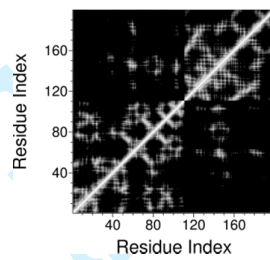
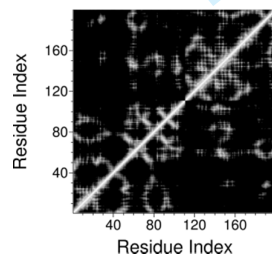


1Q0W





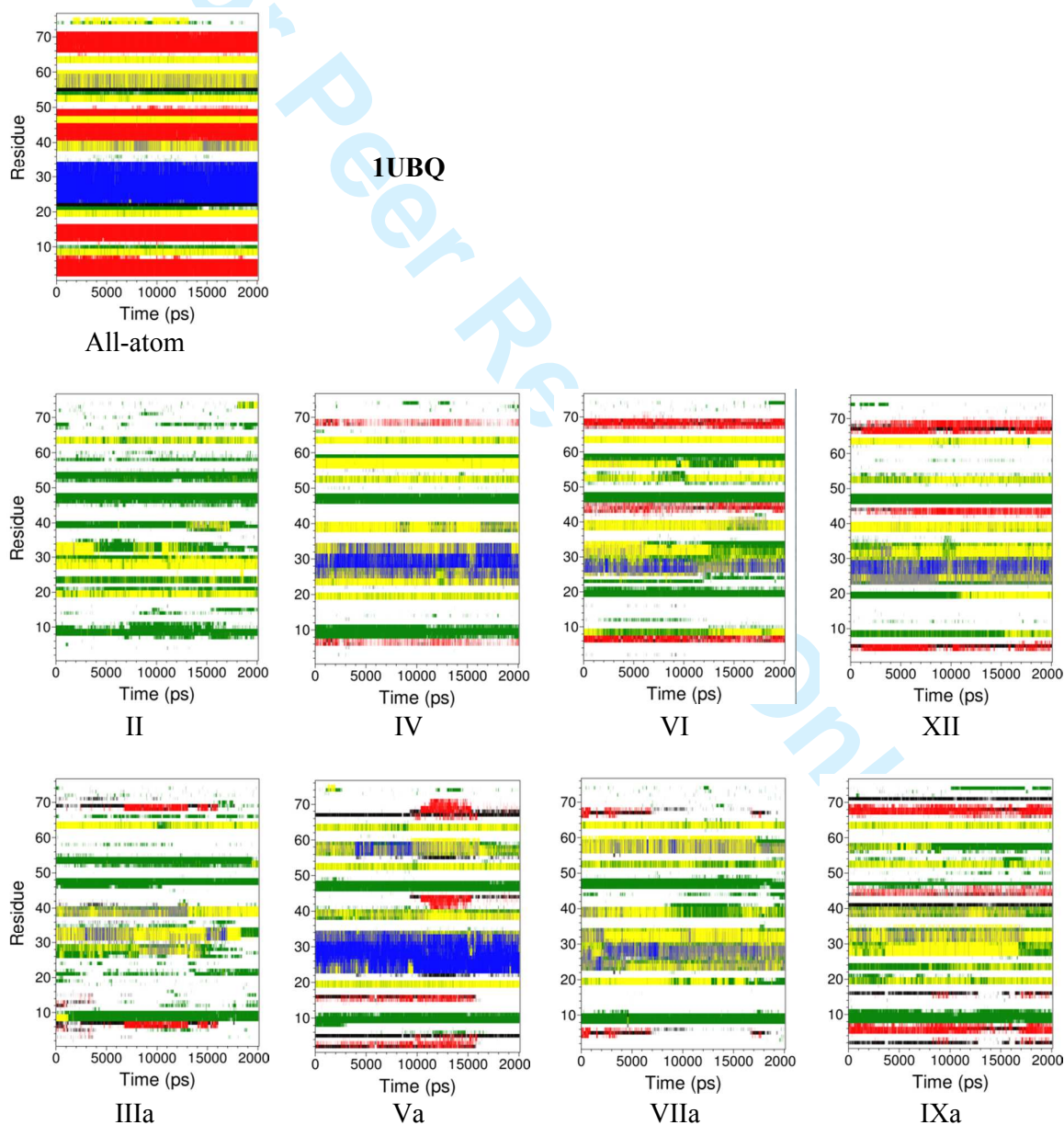
1BRS

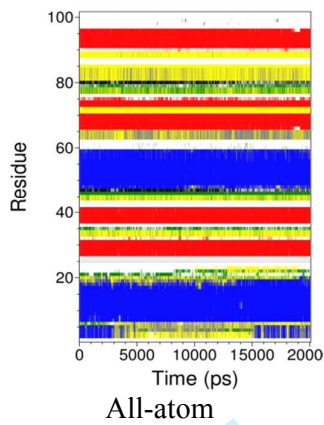


# Reduced point charge models of proteins: Assessment based on molecular dynamics simulations

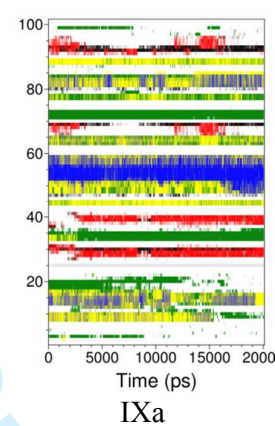
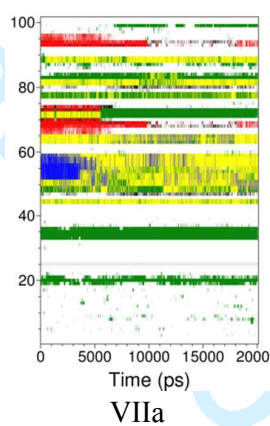
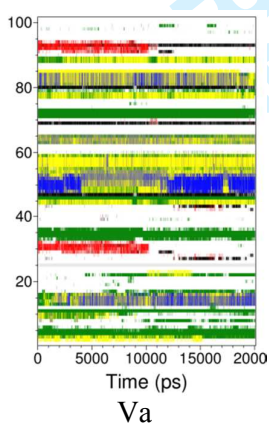
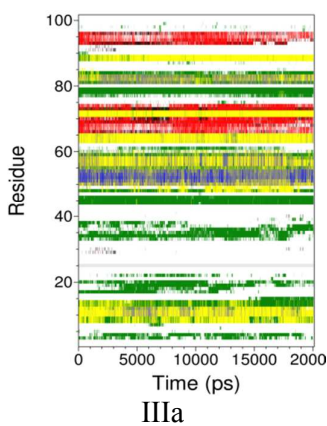
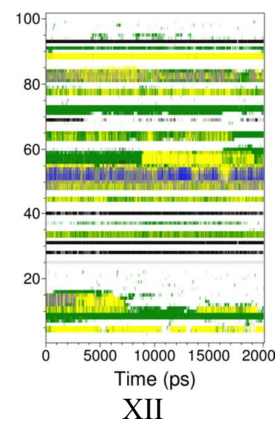
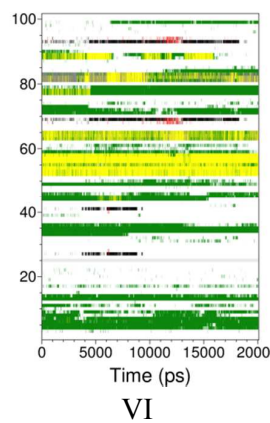
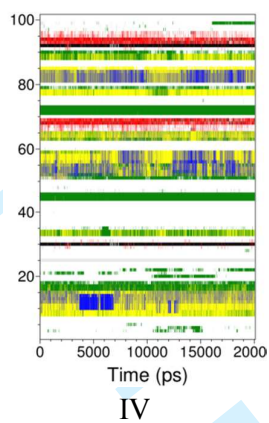
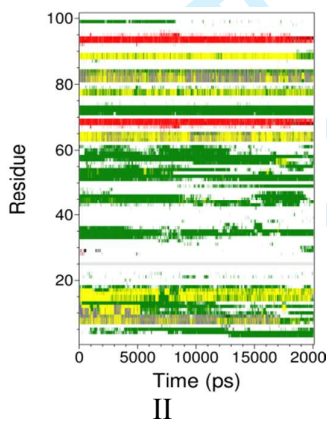
Laurence LEHERTE  
 Department of Chemistry  
 Unité de Chimie Physique Théorique et Structurale  
 Laboratoire de Physico-Chimie Informatique  
 Namur Medicine & Drug Innovation Center (NAMEDIC)  
 University of Namur, Rue de Bruxelles 61, B-5000 Namur (Belgium)

**SI 10.** Secondary structures elements determined from 20 ns AMBER99SB-TIP4P-Ew MD trajectories at 300 K. Secondary structure elements are color-coded as follows: Coil (white),  $\alpha$ -helix (blue),  $\pi$  helix (purple),  $3_{10}$  helix (grey),  $\beta$ -sheet (red),  $\beta$ -bridge (black), bend (green), turn (yellow).

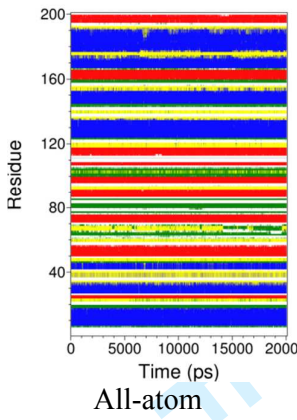




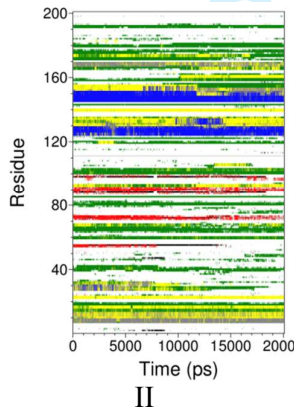
1Q0W



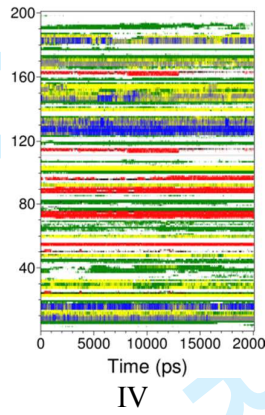
1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



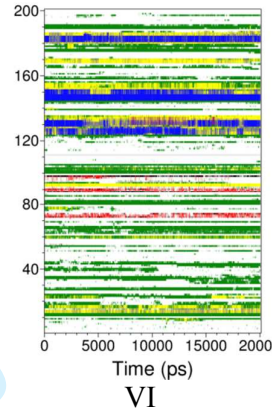
1BRS



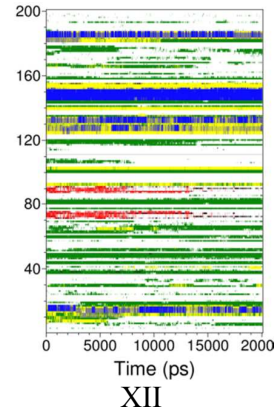
II



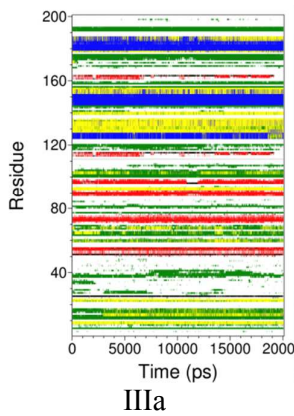
IV



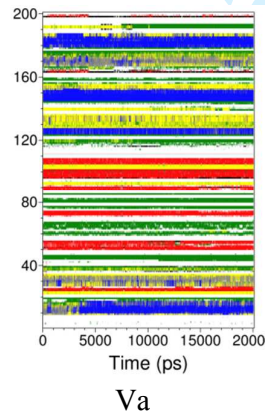
VI



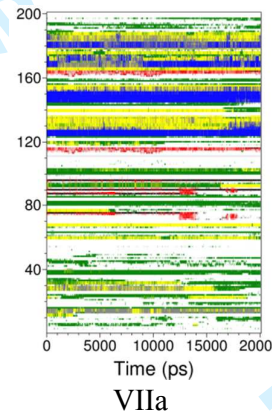
XII



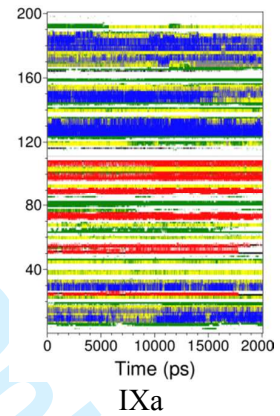
IIIa



Va



VIIa



IXa

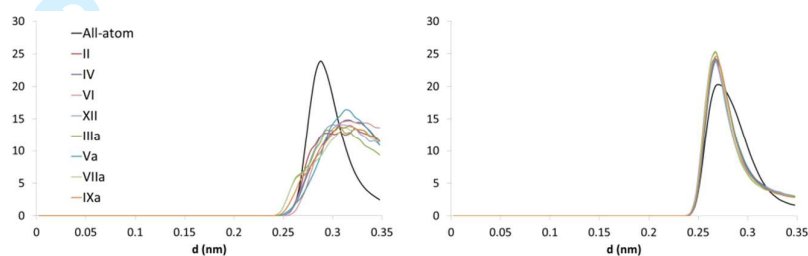
# Reduced point charge models of proteins: Assessment based on molecular dynamics simulations

Laurence LEHERTE  
 Department of Chemistry  
 Unité de Chimie Physique Théorique et Structurale  
 Laboratoire de Physico-Chimie Informatique  
 Namur MEDICINE & Drug Innovation Center (NAMEDIC)  
 University of Namur, Rue de Bruxelles 61, B-5000 Namur (Belgium)

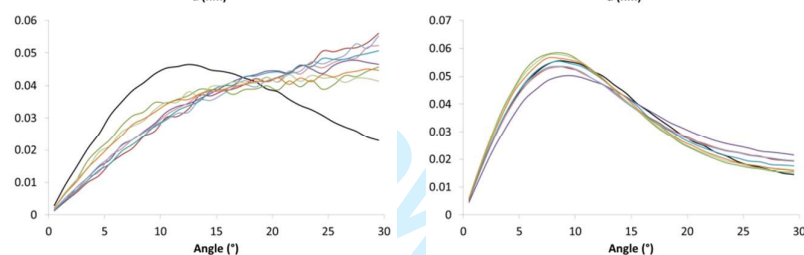
**SI 11.** Distance and angle distributions of the protein-water H-bonds obtained from 20 ns AMBER99SB-TIP4P-Ew MD trajectories at 300 K.

**1UBQ**

Distance distribution



Angle distribution



Intramolecular

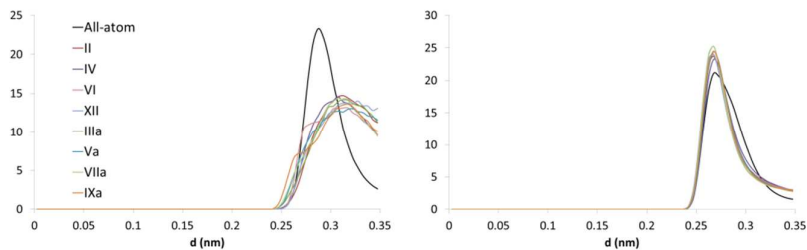
Intermolecular



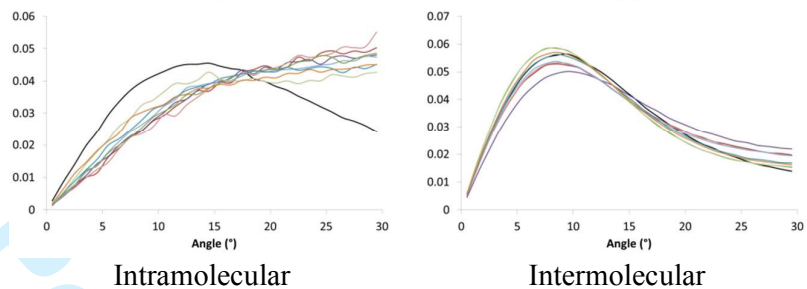
1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**1Q0W**

Distance distribution

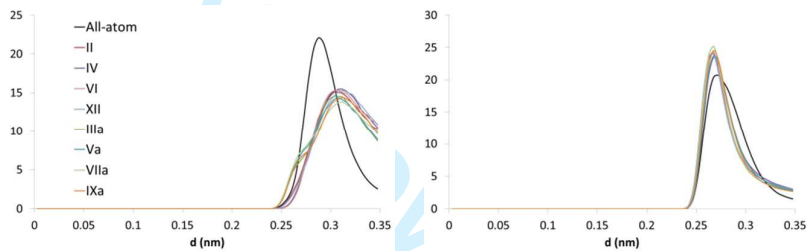


Angle distribution

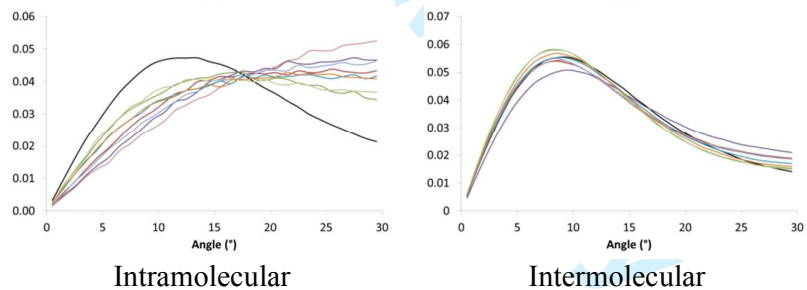


**1BRS**

Distance distribution



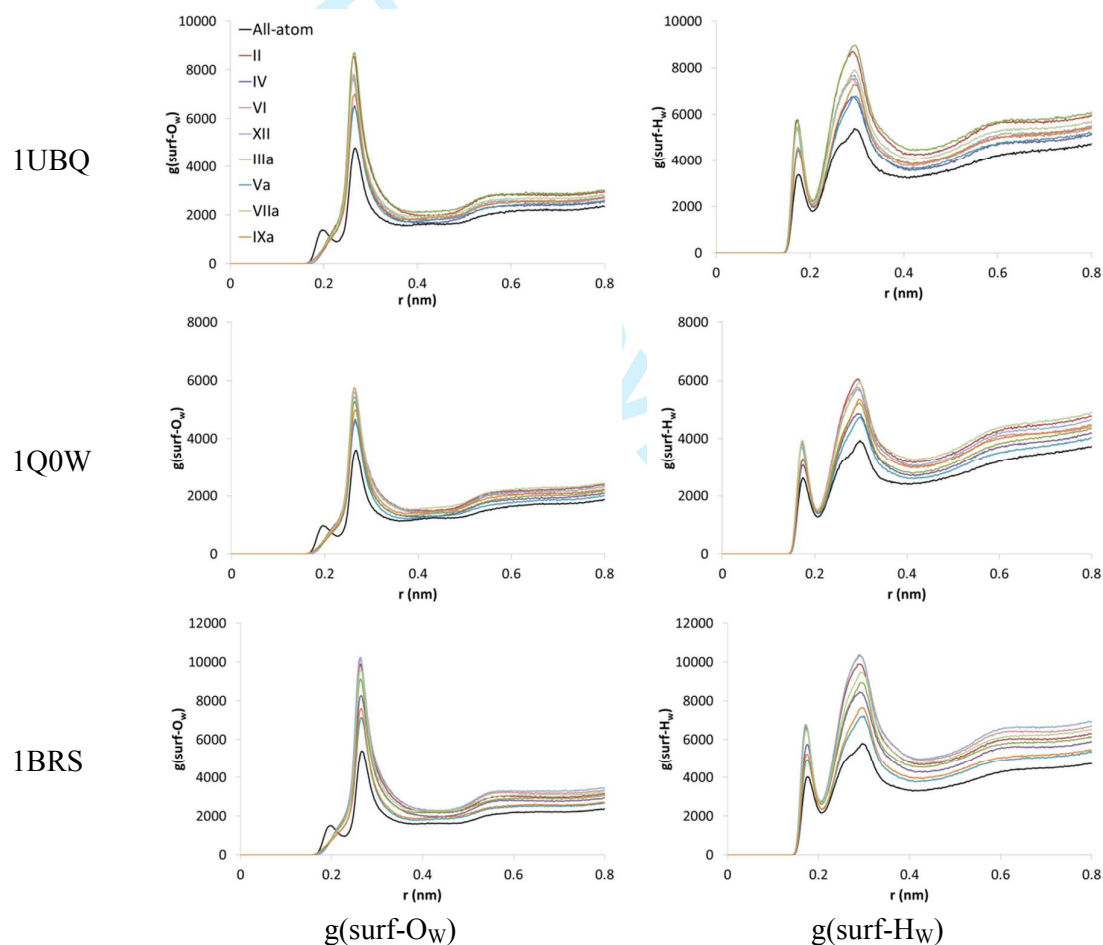
Angle distribution



# Reduced point charge models of proteins: Assessment based on molecular dynamics simulations

Laurence LEHERTE  
 Department of Chemistry  
 Unité de Chimie Physique Théorique et Structurale  
 Laboratoire de Physico-Chimie Informatique  
 Namur MEDICINE & Drug Innovation Center (NAMEDIC)  
 University of Namur, Rue de Bruxelles 61, B-5000 Namur (Belgium)

**SI 12.** Radial distribution functions of the protein surface atoms versus the water atoms,  $g(\text{P-O}_w)$  and  $g(\text{P-H}_w)$ , as obtained from 20 ns AMBER99SB-TIP4P-Ew MD trajectories at 300 K.



# Design of reduced point charge models of proteins: Assessment based on molecular dynamics simulations

Laurence LEHERTE

Department of Chemistry

Unité de Chimie Physique Théorique et Structurale

Laboratoire de Physico-Chimie Informatique

Namur MEDicine & Drug Innovation Center (NAMEDIC)

University of Namur, Rue de Bruxelles 61, B-5000 Namur (Belgium)

**SI 13.** Determination coefficients  $R$  associated with the linear regressions carried out on RPCM energy terms as functions of all-atom contributions. Only the energy terms that are affected by the point charge model are considered.

	Cb_14	Cb_SR	Cb_recip	Epot	Etot	Cb_SR (p-p)	Cb_SR (p-np)
<b>1UBQ</b>							
<b>CD_based models</b>							
II	0.106	0.993	0.472	0.983	0.989	0.605	0.926
IV	0.145	0.995	0.583	0.987	0.992	0.582	0.913
VI	0.201	0.995	0.611	0.984	0.990	0.551	0.929
XII	0.182	0.994	0.621	0.983	0.990	0.474	0.887
<b>CDa_based models</b>							
IIIa	0.078	0.990	0.424	0.983	0.990	0.485	0.965
Va	0.276	0.997	0.508	0.991	0.994	0.651	0.953
VIIa	0.305	0.994	0.568	0.987	0.992	0.411	0.923
IXa	0.282	0.996	0.630	0.990	0.994	0.446	0.939
<b>1Q0W</b>							
<b>CD_based models</b>							
II	0.366	0.992	0.531	0.976	0.986	0.570	0.946
IV	0.330	0.994	0.597	0.982	0.989	0.677	0.945
VI	0.018	0.983	0.204	0.970	0.982	0.138	0.901
XII	0.271	0.989	0.520	0.972	0.983	0.476	0.921
<b>CDa_based models</b>							
IIIa	0.280	0.994	0.537	0.983	0.990	0.567	0.938
Va	0.298	0.995	0.499	0.985	0.991	0.636	0.949
VIIa	0.232	0.988	0.530	0.978	0.987	0.341	0.942
IXa	0.388	0.995	0.655	0.987	0.992	0.536	0.954
<b>1BRS</b>							
<b>CD_based models</b>							
II	0.220	0.988	0.461	0.972	0.987	0.484	0.961
IV	0.270	0.994	0.617	0.979	0.990	0.531	0.922
VI	0.164	0.992	0.509	0.979	0.990	0.453	0.880
XII	0.334	0.991	0.557	0.976	0.989	0.456	0.910
<b>CDa_based models</b>							
IIIa	0.232	0.994	0.550	0.983	0.992	0.568	0.938
Va	0.360	0.996	0.611	0.987	0.994	0.699	0.961
VIIa	0.330	0.991	0.512	0.982	0.992	0.505	0.967
IXa	0.386	0.995	0.660	0.987	0.994	0.666	0.954

Cb\_14 = Coulomb interactions between atoms separated by three successive bonds; Cb\_SR = short-range Coulomb interactions, Cb\_recip = Cb interactions in the reciprocal space; Epot = potential energy; Etot = total energy; p-p = protein-protein interactions; p-np = protein-non protein interactions

# Reduced point charge models of proteins: Assessment based on molecular dynamics simulations

Laurence LEHERTE

Department of Chemistry

Unité de Chimie Physique Théorique et Structurale

Laboratoire de Physico-Chimie Informatique

Namur Medicine & Drug Innovation Center (NAMEDIC)

University of Namur, Rue de Bruxelles 61, B-5000 Namur (Belgium)

**SI 14.** Slope  $S$  associated with the linear regressions carried out on RPCM energy terms as functions of all-atom contributions. Only the energy terms that are affected by the point charge model are considered.

	Cb_14	Cb_SR	Cb_recip	Epot	Etot	Cb_SR (p-p)	Cb_SR (p-np)
<b>1UBQ</b>							
<b>CD_based models</b>							
II	0.278	0.997	0.727	0.989	0.995	0.884	1.060
IV	0.237	0.998	0.706	0.989	0.993	0.850	1.064
VI	0.351	0.998	0.822	0.991	0.995	0.910	1.083
XII	0.411	0.998	0.874	0.989	0.994	0.673	1.049
<b>CD_based models</b>							
IIIa	0.254	0.998	0.639	0.989	0.993	0.572	1.053
Va	0.283	0.999	0.570	0.993	0.995	0.774	0.998
VIIa	0.319	0.997	0.613	0.992	0.996	0.638	1.028
IXa	0.355	0.997	0.706	0.991	0.994	0.533	0.966
<b>1Q0W</b>							
<b>CD_based models</b>							
II	0.513	0.998	0.734	0.983	0.991	0.858	1.053
IV	0.376	0.997	0.698	0.986	0.992	0.955	1.033
VI	0.108	0.991	0.426	0.982	0.990	0.501	1.054
XII	0.437	0.995	0.726	0.974	0.983	0.816	1.105
<b>CDA_based models</b>							
IIIa	0.348	0.997	0.670	0.990	0.996	0.739	0.991
Va	0.276	0.995	0.564	0.986	0.992	0.767	0.985
VIIa	0.351	0.998	0.651	0.988	0.994	0.551	1.067
IXa	0.399	0.996	0.677	0.986	0.992	0.702	0.987
<b>1BRS</b>							
<b>CD_based models</b>							
II	0.392	0.999	0.649	0.986	0.995	0.709	1.117
IV	0.332	0.997	0.726	0.988	0.996	0.778	1.042
VI	0.307	0.996	0.722	0.985	0.994	0.697	1.054
XII	0.577	0.997	0.872	0.981	0.990	1.047	1.139
<b>CDA_based models</b>							
IIIa	0.348	0.997	0.715	0.985	0.995	0.786	1.049
Va	0.340	0.996	0.685	0.989	0.994	0.740	0.983
VIIa	0.434	0.998	0.644	0.990	0.998	0.820	1.095
IXa	0.383	0.998	0.694	0.990	0.994	0.710	0.979

Cb\_14 = Coulomb interactions between atoms separated by three successive bonds; Cb\_SR = short-range Coulomb interactions, Cb\_recip = Cb interactions in the reciprocal space; Epot = potential energy; Etot = total energy; p-p = protein-protein interactions; p-np = protein-non protein interactions

# Reduced point charge models of proteins: Assessment based on molecular dynamics simulations

Laurence LEHERTE  
 Department of Chemistry  
 Unité de Chimie Physique Théorique et Structurale  
 Laboratoire de Physico-Chimie Informatique  
 Namur Medicine & Drug Innovation Center (NAMEDIC)  
 University of Namur, Rue de Bruxelles 61, B-5000 Namur (Belgium)

**SI 15.** Intercept  $I$  associated with the linear regressions carried out on RPCM energy terms as functions of all-atom contributions. Only the energy terms that are affected by the point charge model are considered.

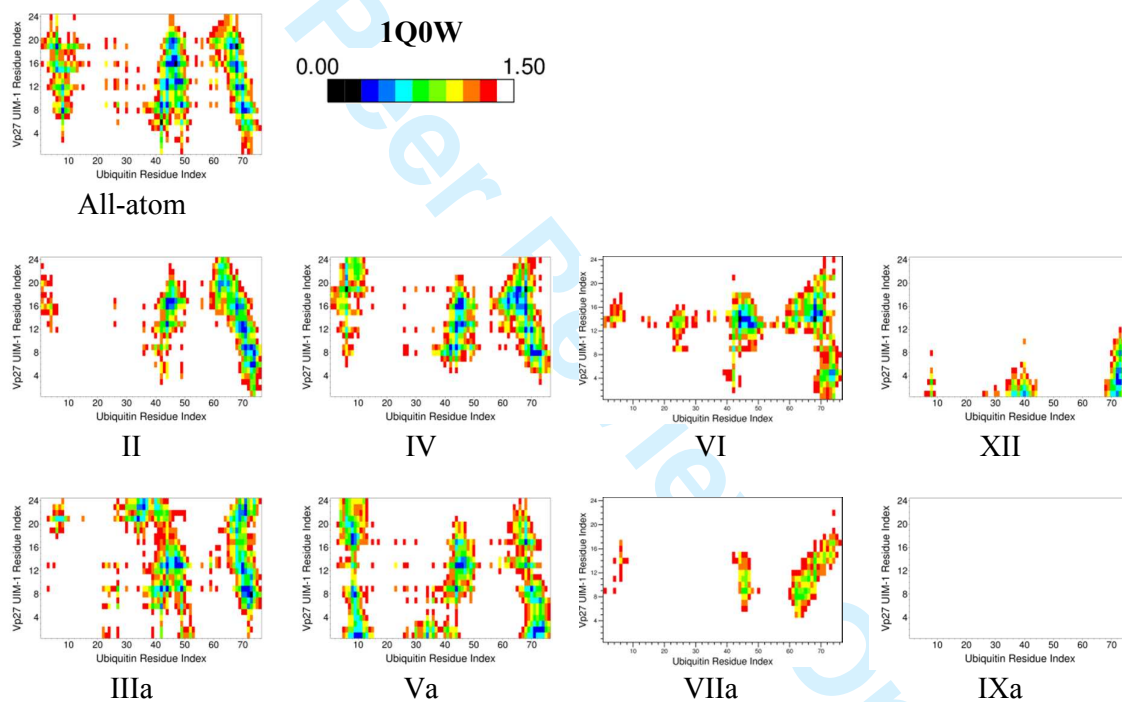
	Cb_14	Cb_SR	Cb_recip	Epot	Etot	Cb_SR (p-p)	Cb_SR (p-np)
<b>1UBQ</b>							
<b>CD_based models</b>							
II	12097.1	-4428.5	-17783.1	-6395.2	-3885.4	-909.4	-1447.6
IV	9467.7	-1847.9	-15509.5	-4521.8	-2541.7	-148.1	-467.0
VI	11023.8	-3213.2	-11948.3	-5283.4	-3369.4	-715.2	-1070.8
XII	11227.4	-3688.1	-9847.5	-6730.4	-4162.1	-1827.4	-1673.4
<b>CD_based models</b>							
IIIa	9169.0	-2361.5	-20319.4	-5652.1	-3532.5	-1287.3	-683.0
Va	5587.5	151.2	-20839.9	-1697.7	-367.2	214.7	-335.7
VIIa	8307.5	-2641.2	-21798.8	-4095.6	-2220.2	-1148.0	-876.7
IXa	4028.4	-1020.6	-11683.8	-2189.2	-308.8	-505.0	-629.2
<b>1Q0W</b>							
<b>CD_based models</b>							
II	11256.0	-4634.7	-19908.8	-10916.4	-6530.3	-1234.9	-1934.5
IV	9672.7	-2455.7	-17710.0	-6236.9	-3204.4	314.5	-963.1
VI	18068.4	-8083.5	-40560.7	-10692.5	-6840.6	-2928.2	-1777.7
XII	13806.6	-6928.2	-21575.3	-15680.3	-11104.0	-1948.6	-1322.6
<b>CDA_based models</b>							
IIIa	10544.4	-2902.9	-21004.0	-5515.4	-2146.2	-1083.2	-1536.1
Va	7380.4	-2052.0	-23229.5	-4422.3	-1793.3	171.7	-550.7
VIIa	10185.2	-2506.1	-21964.8	-6522.5	-3233.5	-1814.8	-452.5
IXa	4738.7	-886.5	-14216.3	-3734.3	-797.9	28.4	-526.9
<b>1BRS</b>							
<b>CD_based models</b>							
II	27166.1	-6666.8	-46702.2	-16207.6	-8376.5	-3527.2	-2221.5
IV	20464.4	-3789.9	-27341.3	-9052.7	-1496.2	-481.9	-1878.4
VI	30133.9	-9789.0	-36835.0	-16484.4	-8958.9	-3725.4	-3734.1
XII	22999.4	-10245.6	-20997.1	-21981.4	-13911.5	-1259.9	-2137.9
<b>CDA_based models</b>							
IIIa	21320.8	-5631.3	-31686.9	-14310.2	-5319.2	-1427.5	-1783.5
Va	12447.1	-1852.1	-25253.8	-4965.3	-408.2	424.5	-1142.6
VIIa	17982.9	-4192.5	-40233.3	-9173.2	-2830.1	-1220.2	-674.6
IXa	9096.0	402.4	-21367.5	-3066.8	1073.1	480.6	-1399.9

Cb\_14 = Coulomb interactions between atoms separated by three successive bonds; Cb\_SR = short-range Coulomb interactions, Cb\_recip = Cb interactions in the reciprocal space; Epot = potential energy; Etot = total energy; p-p = protein-protein interactions; p-np = protein-non protein interactions

# Reduced point charge models of proteins: Assessment based on molecular dynamics simulations

Laurence LEHERTE  
 Department of Chemistry  
 Unité de Chimie Physique Théorique et Structurale  
 Laboratoire de Physico-Chimie Informatique  
 Namur Medicine & Drug Innovation Center (NAMEDIC)  
 University of Namur, Rue de Bruxelles 61, B-5000 Namur (Belgium)

**SI 16.** Mean shortest protein-ligand distance maps as calculated from 20 ns AMBER99SB-TIP4P-Ew MD trajectories at 300 K. Distances are given in nm in the colour scale.



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

