Evaluation of Reduced Point Charge Models of Proteins through Molecular Dynamics Simulations: Application to the Vps27 UIM-1 – Ubiquitin Complex

Leherte, Laurence; Vercauteren, Daniel

Published in:
Journal of Molecular Graphics and Modelling

Publication date:
2014

Document Version
Peer reviewed version

AUTHOR QUERY FORM

ELSEVIER	**Journal:** JMG **Article Number:** 6354	**Please e-mail or fax your responses and any corrections to:** **E-mail:** corrections.esch@elsevier.thomsondigital.com **Fax:** +353 6170 9272

Dear Author,

Please check your proof carefully and mark all corrections at the appropriate place in the proof (e.g., by using on-screen annotation in the PDF file) or compile them in a separate list. Note: if you opt to annotate the file with software other than Adobe Reader then please also highlight the appropriate place in the PDF file. To ensure fast publication of your paper please return your corrections within 48 hours.

For correction or revision of any artwork, please consult http://www.elsevier.com/artworkinstructions.

Any queries or remarks that have arisen during the processing of your manuscript are listed below and highlighted by flags in the proof. Click on the 'Q' link to go to the location in the proof.

Location in article	Query / Remark: click on the Q link to go Please insert your reply or correction at the corresponding line in the proof
Q1 Q2	Please confirm that given names and surnames have been identified correctly. Figs. 2, 5 to 11 will appear in black and white in print and in color on the web. Based on this, the respective figure captions have been updated. Please check, and correct if necessary. Please check this box or indicate your approval if you have no corrections to make to the PDF file ☐

Thank you for your assistance.

Journal of Molecular Graphics and Modelling xxx (2013) xxx–xxx

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

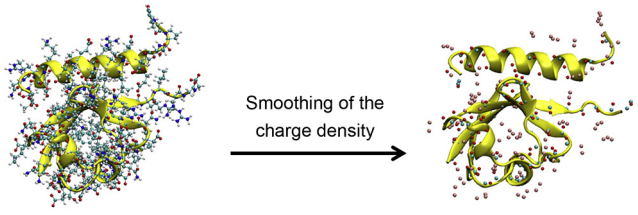
journal homepage: www.elsevier.com/locate/JMGM

Graphical Abstract

Journal of Molecular Graphics and Modelling xxx (2013) xxx–xxx

Evaluation of reduced point charge models of proteins through Molecular Dynamics simulations: Application to the Vps27 UIM-1–Ubiquitin complex

Laurence Leherte*, Daniel P. Vercauteren



All-atom point charge model

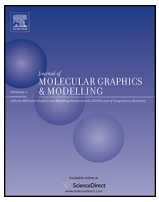
Reduced point charge model

Journal of Molecular Graphics and Modelling xxx (2013) xxx–xxx

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM



Highlights

Journal of Molecular Graphics and Modelling xxx (2013) xxx–xxx

Evaluation of reduced point charge models of proteins through Molecular Dynamics simulations: Application to the Vps27 UIM-1–Ubiquitin complex

Laurence Leherte*, Daniel P. Vercauteren

- Reduced point charge models provide stable Molecular Dynamics trajectories of the protein complex.
- H-bond networks are progressively modified as the reduction degree increases.
- The models allow to probe local potential hyper-surface minima that are similar to all-atom ones.
- The models allow to sample protein conformations more rapidly than the all-atom case due to a lowering of energy barriers.
- Implementation of point charges as virtual sites requires attention to the reference atoms and to the Cb-14 energy terms.

Journal of Molecular Graphics and Modelling xxx (2013) xxx–xxx

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM

Evaluation of reduced point charge models of proteins through Molecular Dynamics simulations: Application to the Vps27 UIM-1–Ubiquitin complex

Laurence Leherte*, Daniel P. Vercauteren

Unité de Chimie Physique Théorique et Structurale, University of Namur, Rue de Bruxelles 61, B-5000 Namur, Belgium

ARTICLE INFO

Article history:
Accepted 31 October 2013
Available online xxx

Keywords:
Molecular electrostatic potential
Electron density
Smoothing of molecular fields
Critical points
Point charge model
Protein
Ubiquitin complex

ABSTRACT

Reduced point charge models of amino acids are designed, (i) from local extrema positions in charge density distribution functions built from the Poisson equation applied to smoothed molecular electrostatic potential (MEP) functions, and (ii) from local maxima positions in promolecular electron density distribution functions. Corresponding charge values are fitted *versus* all-atom Amber99 MEPs. To easily generate reduced point charge models for protein structures, libraries of amino acid templates are built. The program GROMACS is used to generate stable Molecular Dynamics trajectories of an Ubiquitin-ligand complex (PDB: 1Q0W), under various implementation schemes, solvation, and temperature conditions. Point charges that are not located on atoms are considered as virtual sites with a nul mass and radius. The results illustrate how the intra- and inter-molecular H-bond interactions are affected by the degree of reduction of the point charge models and give directions for their implementation; a special attention to the atoms selected to locate the virtual sites and to the Coulomb-14 interactions is needed. Results obtained at various temperatures suggest that the use of reduced point charge models allows to probe local potential hyper-surface minima that are similar to the all-atom ones, but are characterized by lower energy barriers. It enables allows to generate various conformations of the protein complex more rapidly than the all-atom point charge representation.

1. Introduction

Numerous models have already been proposed in literature regarding the coarse-graining of biomolecules and their corresponding interaction potentials. Several references were already given in [1]. Reviews [2–6] as well as software [7,8] are regularly published on the subject.

In recent publications, we described reduced point charge models [9,10] built from critical point (CP) analyses of smoothed molecular properties, and their applications to Molecular Dynamics (MD) simulations of proteins [1]. These simulations were achieved in vacuum using the program package TINKER [11] wherein point charges were considered as masses attached to the protein structure through harmonic bonds. The mass that was associated with the charges was set to a value of $m = 2$ in order to limit the mass increase of the protein structure and to allow a time step value of 1 fs, a lower value of m implying too strong a decrease of the time step to get stable MD trajectories. All other terms of the selected force field (FF), Amber99 [12], were calculated at the all-atom level, *i.e.*, as in the original FF version.

In the present work, the design of reduced point charge models is not intended to lead to a coarse-grained model *per se*. It is part of a more global project regarding the analysis of low resolution molecular properties such as electron density (ED) and molecular electrostatic potential (MEP). From the very first MD applications of a reduced point charge model to protein structures [1], it was found that the secondary structure of the proteins was only partly lost during the simulations while the overall three-dimensional (3D) fold remained stable. Additionally, a decrease of about a factor 2 of the calculation time was observed for the simulations in vacuum. As a perspective to the work reported in [1], it was expected to adopt other implementation approaches to get rid of the non-zero mass assigned to the charges. This perspective led to the present paper.

In the present work, two molecular properties leading to different reduced point charge models are considered. First, as in [1,10], a limited number of point charges is obtained through the search for the maxima and minima of a smoothed version of the charge density (CD) generated by the atomic charges defined in Amber99 (or Amber99SB) FF [12]. Second, the point charges are obtained through a search of the maxima of the full promolecular

* Corresponding author. Tel.: +32 81 724560; fax: +32 81 725466.
E-mail address: laurence.leherte@unamur.be (L. Leherte).

1093-3263/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jmgm.2013.10.011

ED of the molecular structure and, as in the first approach, charge values are assigned to those maxima using a least-square charge fitting procedure. This last molecular property is easily calculated using the so-called Promolecular Atom Shell Approximation (PASA) formalism that was developed by Amat and Carbó-Dorca [13,14]. Those two approaches led to various implementations which are discussed in the present paper. The program GROMACS [15,16] was selected due to its implementation towards shorter calculation times and the possibility to define virtual sites, *i.e.*, particles with nul mass and radius that are coupled to the molecular structure through geometrical rules.

The aim of the paper is to go deeper in the modelling of protein structures and dynamics using reduced point charge models, with an application to a particular protein complex. Such a system was selected to provide information on the usefulness of the models to simulate short peptides and proteins as well as their mutual interactions. The effect of the point charge distributions on both the intra- and inter-molecular interactions in various simulation conditions, such as temperature and solvation, is also investigated. At this stage of our research work, only the electrostatic part of the FF is modified. Energetic, structural, and dynamical properties are calculated and compared to the all-atom ones, which is easily achieved as no conversion stage is required between the reduced and all-atom models. However, keeping a significant number of all-atom contributions to the FF limits, for the moment, the possible gain in calculation time.

In the next section, we detail how the models and their implementation were designed. Then, MD simulation results for the protein complex involving the Ubiquitin Interacting Motif UIM-1 of protein Vps27 and Ubiquitin (PDB access code 1Q0W) with various point charge models, and under different simulation conditions (temperature, solvation state), are analysed and discussed.

2. Background theory

In this section, we present the mathematical formalism that was used to design a molecular reduced point charge representation and its corresponding charge values. As all these aspects were already detailed before [9,10], we only provide a short overview. First, the smoothing algorithm is briefly described. Then, the approach applied to locate the point charges is presented, as well as the procedure to assign charge values. Finally, the automation procedure that is implemented to rapidly determine the point charge locations for any protein structure is explained.

2.1. Smoothing of a molecular property

In the present approach to generate smoothed 3D functions, a smoothed CD or ED map is a lower resolution version that is directly expressed as the solution of the diffusion equation according to the formalism presented by Kostrowicki *et al.* [17]. From the formalism given in [1], the smoothed analytical CD distribution function $\rho_{a,s}(r)$ that is obtained from an atomic charge q_a and the Poisson equation is expressed as:

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

where a, s, and r, stand for the atom index, the smoothing factor (in Bohr^2, 1 Bohr = 0.52918×10^{-10} m), and the distance *versus* the atom position, respectively. The full promolecular ED is calculated as:

$$\rho_{a,s}(r) = \sum_{i=1}^{3} \sigma_{a,i} \quad \text{where} \quad \sigma_{a,i} = \alpha_{a,i} e^{-\beta_{a,i} r^2} \quad (2)$$

with

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

where Z_a, $w_{a,i}$ and $\zeta_{a,i}$, are the atomic number of atom a, and the two fitted parameters, respectively. Unsmoothed functions are obtained by imposing $s = 0$ Bohr^2.

2.2. Search for critical points

An algorithm initially described by Leung *et al.* [18] was implemented to follow the trajectories of CPs, more specifically, the maxima and/or minima in a CD or ED function, as a function of the degree of smoothing. As already reported before [9], we adapted their idea to 3D molecular property functions, f, such as:

$$\mathbf{r}_{f(s)} = \mathbf{r}f(s - \Delta s) + \frac{\nabla f(s) \cdot \Delta}{f(s)} \quad (4)$$

where $\mathbf{r}$ stands for the location vector of a point in a 3D function, such as a molecular scalar field, and $\Delta/f(s)$ is the step length.

The various steps of the resulting merging/clustering algorithm are as follows. First, at scale $s = 0$, each atom of a molecular structure is considered either as a local maximum (peak) or minimum (pit) of the scalar field f. All atoms are consequently taken as the starting points of the merging procedure. Second, as s increases from 0 to a given maximal value s_{max}, each point moves continuously along a gradient path to reach a location in the 3D space where $\nabla f(s) = 0$. On a practical point of view, this consists in following the trajectory of the peaks and pits on the molecular property surface calculated at s according to Eq. (4). The trajectory search is stopped when $|\nabla f(s)|$ is lower or equal to a limit value, $grad_{lim}$. Once all peak/pit locations are found, close points are merged if their inter-distance is lower than the initial value of $\Delta^{1/2}$. The procedure is repeated for each selected value of s. If the initial Δ value is too small to allow convergence towards a local maximum or minimum within the given number of iterations, its value is doubled (a scaling factor that is arbitrarily selected) and the procedure is repeated until final convergence.

2.3. Charge calculation

To stay consistent with the analytical expression of the Amber99 FF, only point charge values are assigned to each of the CPs of a 3D molecular property field. In recent literature, one also finds coarse-grained electrostatic energy terms which also involve dipolar terms, such as in the work of Spiga *et al.* [19]. The charge fitting program QFIT [20] was used as detailed in [9]. All MEP grids were built using the Amber99 [12] atomic charges which were assigned using the software PDB2PQR [21,22], with a grid step of 0.5 Å. Fittings were achieved by considering MEP grid points located between 1.4 and 2.0 times the van der Waals (vdW) radius of the atoms. These two limiting distance values were selected after the so-called Merz-Singh-Kollman scheme [23]. Side chains and main chains of the amino acids (AA) were treated separately, as discussed in [9].

In all fittings, the total electric charge and the magnitude of the molecular dipole moment were constrained to be equal to the corresponding all-atom Amber99 values. All dipole moment components were calculated with the origin of the atom coordinates set to (0 0 0).

Please cite this article in press as: L. Leherte, D.P. Vercauteren, Evaluation of reduced point charge models of proteins through Molecular Dynamics simulations: Application to the Vps27 UIM-1–Ubiquitin complex, J. Mol. Graph. Model. (2013), http://dx.doi.org/10.1016/j.jmgm.2013.10.011

3. Design of amino acid reduced point charge models

Reduced point charge representations of each of the twenty AAs were obtained by considering the AAs in specific conformational states. Except for Gly and Ala, most recurrent rotamers were generated by considering the angular constraints given in Table 2 of [9]. All AAs were considered as electrically neutral except for Arg+, His+, Lys+, Asp−, and Glu−. Histidine was also modelled in its neutral protonated states Hisδ and Hisε.

3.1. CD-based templates

From extended pentadecapeptide chains Gly_7-AA-Gly_7 generated using SMMP05 [24,25] only the central AA was kept with main chain atoms $(C\alpha{-}C{=}O)_{AA}(N{-}H)_{AA+1}$. Then, the design of the AA point charge templates was achieved in four stages, as follows. First, isolated AA structures were assigned Amber99 atom charges using PDB2PQR [21,22]. Side chain extrema were located using our merging/clustering algorithm applied to the CD distribution functions smoothed at $s = 1.7$ Bohr^2, with $\Delta_{init} = 10^{-4}$ Bohr^2 and $grad_{lim} = 10^{-6}$ e$^-$ Bohr^{-2}. This was carried out separately for the positively and negatively charged atoms. Second, the charge values of the resulting peaks and pits together were fitted *versus* the all-atom MEP generated from the side chain atoms only. In this procedure, several rotamer descriptions were considered according to their occurrence probability (see Table 2 of [9]). Third, the main chain point charges were located in accordance with the motif found for Gly8 in an extended Gly_{15} strand [9] and, fourth, a second charge fitting procedure, now carried out *versus* the MEP calculated using all the AA atoms, was achieved to determine the charge values of the two main chain point charges while preserving the side chain point charge values first obtained.

All main chain point charges, observed to be located very close to the C and O atoms, were set exactly on those atoms [1] (Supplementary Information SI 1). All AA bear side chain charges except Ala, Gly, Ile, Leu, and Val. AA models are given in SI 2 and were discussed with details in [1]. In the present work, an extra protonation state for His was generated, *i.e.*, His+. It is characterized by the highest number of point charges, *i.e.*, 6, *versus* the other histidine residues, *i.e.*, 4 and 5. As most of these point charges of His+ are close to H atoms, they were set to be located exactly on these atoms to facilitate the implementation of the point charge model. For the end residues, a charge of +0.9288 or −0.9288 e$^-$ is set on the N and OXT atoms, respectively [9,10]. In the further parts of this paper, the model will be referred to as model *mCD*.

A second point charge description was derived from the model described above. In this second model, to fully facilitate the implementation of the AA models in GROMACS, most of the point charges were set exactly on atoms of the residues, and a charge fitting algorithm was again applied. Results are presented in SI 3. This implies that only three AA residues, His+, Phe, and Trp, have a point charge that is not located on an atom of their structure. In that model, one obtains end charge values of +0.7705 and −0.7705 for the end N and OXT atoms. In the further parts of this paper, the second model will be referred to as model *mCDa*.

A last model based on the point charge distribution *mCD* was considered similarly to the approach adopted in [1] with the program package TINKER [11]. A mass $m = 2$ was assigned to each point charge and harmonic constraints were applied to bonds and angles presented in SI 2. The model will be referred to as model *mCDh*.

3.2. PASA-based templates

CP searches of the PASA ED distribution functions were carried out to generate still coarser charge descriptions for the AAs. Indeed, with the CD distribution functions depicted above, it is not possible to obtain less than two main chain point charges per residue, *i.e.*, one negative and one positive charge associated with the O and C atoms, respectively. Within the framework of the PASA, the ED depends only on the atomic number Z_a of the atoms, not on their charge (Eqs. (2) and (3)). The use of the merging/clustering algorithm was carried out with $\Delta_{init} = 10^{-4}$ Bohr^2 and $grad_{lim} = 10^{-5}$ e$^-$ Bohr^{-2}. This limit is an order of magnitude greater than in the CD case as too fine a $grad_{lim}$ threshold increases the possibility to miss the recognition of duplicate CPs during the search. A smoothing degree of $s = 1.4$ Bohr^2 was considered. An exception occurs for Trp for which one observes two side chain CPs at s below or equal to 1.05 Bohr^2. We selected that value to better differentiate that residue from the other aromatic residues characterized by only one CP. As each AA main chain involves only one CP, its charge value was directly set as the sum over the corresponding atomic charges. Then, for AAs involving more than one side chain CP, a charge fitting procedure was applied as for the CD-based models. For end residues, one observed no main chain CP on the terminal NH_3^+ group. The charge of the main chain CP of the N-terminal residue is thus incremented by +1, while the main chain charge of the C-terminal residue is incremented by −1. Individual AA point charge representations are given in Fig. 1. Regarding the side chains, most of the residues are characterised by only one CP. Their location is mostly determined by the atoms with the highest atomic number, *i.e.*, S, O, N, and C. The H atoms do not significantly affect the ED distribution functions and this makes all His residues looking alike. In the further parts of this paper, the model will be referred to as model *mPASA*. Its implementation within the program GROMACS is detailed in SI 4.

3.3. Automated point charge generation procedure

The four point charge templates described above are established for isolated AA structures. Their properties are thus independent on the neighbourhood occurring in a particular protein of a complex structure. This presents a great advantage when those properties are transferable. In a previous work [9], one indeed suggested, through a good approximation of MEP profiles of ion channels and numerous free energy calculations, that transferability occurs for rigid protein structures. Additionally, in their paper regarding the optimal number of coarse-grained sites in biomolecular complexes, Sinitskiy *et al.* concluded that the transferability of individual protein properties between unbound and bound states is supported by the possibility to coarse-grain complex partners independently one from each other [27].

To study large protein structures, an automation stage was developed to rapidly locate point charges on the structure. It is fully based on the application of a superimposition algorithm of CP templates of each AA onto their corresponding all-atom structure of the protein under study. We used the program QUATFIT [28,29] to, first, superimpose a limited set of atoms from the template on the studied structure, and then use the resulting transformation matrix to generate the corresponding point charge coordinates. The templates obtained from CD distribution functions were already made available in Table 3 of [9]. The His+ template newly studied in the present work is provided in SI 5. Additionally, the templates obtained from the analysis of the PASA ED distribution functions are reported in SI 6. The GROMACS topology file, wherein point charges are defined as virtual sites, is further generated through an in-house program that outputs geometrical parameters as reported in SI 1, 3, and 4, for the *mCD*, *mCDa*, and *mPASA* models, respectively.

When a GROMACS virtual site is generated, a number of atoms, two or three, are selected to determine its location in space. The choice of the reference atoms for each point charge is not unique. We have most of the time considered the closest atoms by excluding H ones, except for the alcohol functions of Ser, Thr, and Tyr, to allow

Please cite this article in press as: L. Leherte, D.P. Vercauteren, Evaluation of reduced point charge models of proteins through Molecular Dynamics simulations: Application to the Vps27 UIM-1–Ubiquitin complex, J. Mol. Graph. Model. (2013), http://dx.doi.org/10.1016/j.jmgm.2013.10.011

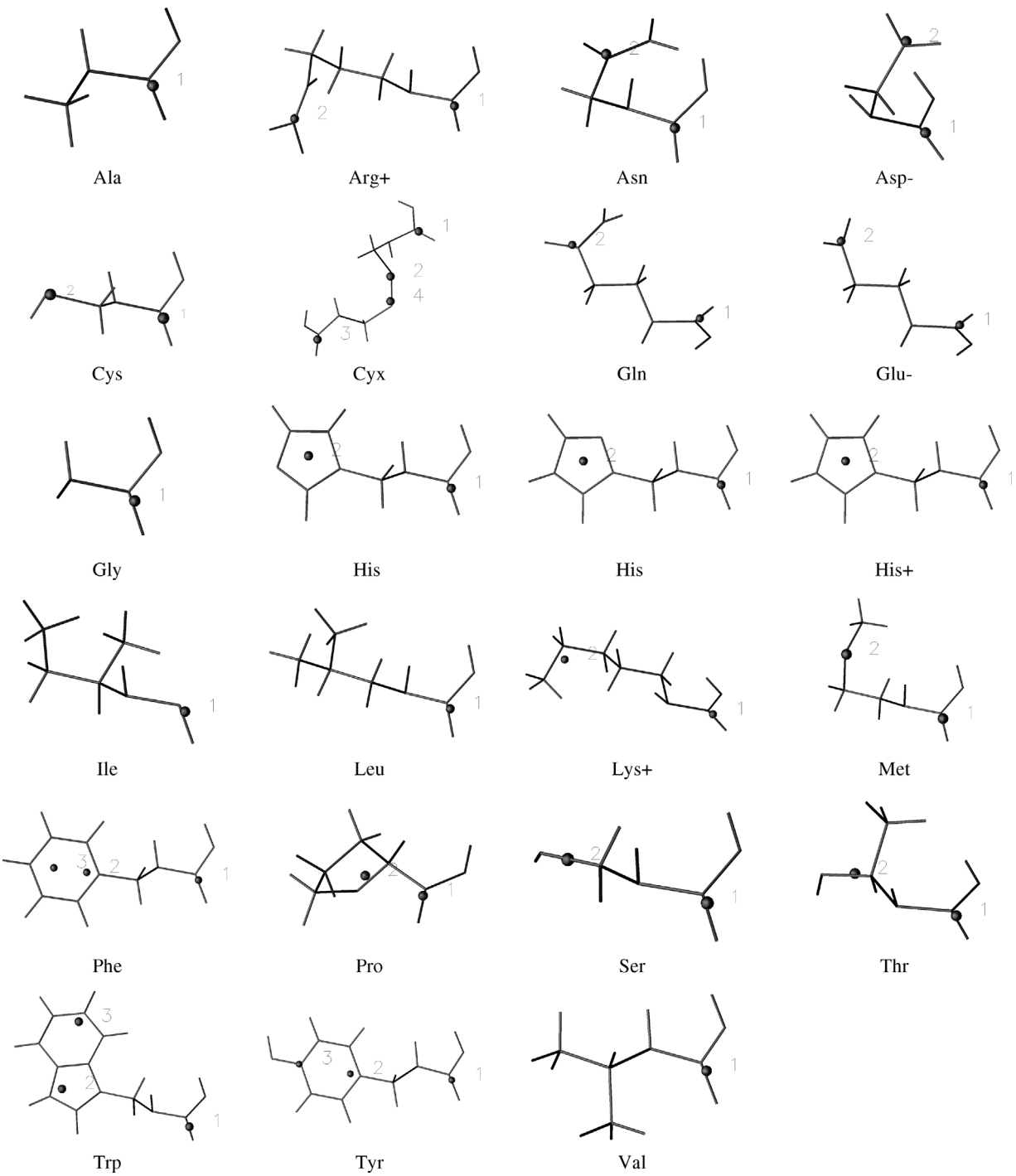


Fig. 1. Point charge model for the 20 AA residues as established at $s = 1.4$ Bohr^2 from the hierarchical merging/clustering algorithm applied to the all-atom PASA ED function. Point charges are numbered as in Supporting Information SI 4. Figures were generated using OpenDX [26].

rotations around the C—O bond. Other models might obviously be tested, for example for neutral His residues, that are not used in the present studies, especially to determine their influence on H-bond interactions. During a MD simulation, the forces acting on the virtual sites are redistributed among their reference atoms. Force redistribution was partly limited in model *mCDa* by locating most of the point charges on atoms.

As already mentioned, point charges are considered as virtual sites that act only through Coulomb interactions. An additional implementation strategy was considered where point charges are seen as masses interacting with the protein structure through restrained harmonic bonds. The advantage of such an approach lies in the fact that the electrostatic forces acting on the charges are not redistributed among the atoms, but the model is biased by artificial masses added to the system.

4. Application to the MD of the Vps27 UIM-1–Ubiquitin complex

Vps27 UIM-1, a short α-helical structure made of 24 AA residues (numbered 255–278 in the PDB), is known to interact with the five-stranded β-sheet of Ubiquitin [30,31]. It consists mainly of

Please cite this article in press as: L. Leherte, D.P. Vercauteren, Evaluation of reduced point charge models of proteins through Molecular Dynamics simulations: Application to the Vps27 UIM-1–Ubiquitin complex, J. Mol. Graph. Model. (2013), http://dx.doi.org/10.1016/j.jmgm.2013.10.011

Table 1
Description of the point charge models used for the Amber99SB-based MD simulations of the Vps27 UMI-1–Ubiquitin complex.

	No. of water molecules	Total no. of point charges associated with both complex partners (Vps27 UIM-1/Ubiquitin)	No. of non-atomic point charges	Box size (nm) (from final snapshot)
All-atom	10,553	394/1227	0	6.915
All-atom-2	13,269	394/1227	0	7.442
mCD	10,542	96/286	112	6.909
mCDa	10,551	96/286	3	6.899
mCDh	10,542	96/286	112	6.916
mPASA	10,551	36/102	136	6.919

hydrophilic residues, encompassing a hydrophobic motif that contains residues Leu262, Ile263, Ala266, Ile267, Leu269, and Leu271, that interact with residues Leu8, Ile44, Val70, and Ala56 of Ubiquitin [30,32]. According to Swanson *et al.* [30], electrostatic interactions are expected to occur between negatively charged Glu$^-$ residues of Vps27 UMI-1 (Glu$^-$257, Glu−259, Glu−260, and Glu−261) and Arg+ residues of Ubiquitin (Arg+42, Arg+72, and Arg+74), as well as between Glu$^-$273 and His+68, while H-bonds are observed between Ala266 and His+68, Ser270 and Ala46 and Gly47, as well as between Gly47 and His+68 [33].

The study of such a protein-protein system using MD approaches is not new [33–35] and the applications presented in this paper are intended to test and assess, *versus* their all-atom counterpart, the point charge reduced models developed above. Our reference works are thus the data available in literature as well as our own all-atom MD simulation results.

Molecular simulation conditions were kept as close as possible of those proposed by Showalter and Brüschweiler in their work about the Amber99SB FF [36]. MD trajectories of the system were run using the GROMACS 4.5.5 program package [15,16] with the Amber99SB FF [37] under particle mesh Ewald periodic boundary conditions. Long-range dispersion corrections to energy and pressure were applied. The initial configurations were retrieved from the Protein Data Base (PDB: 1Q0W) and solvated, if required, using TIP4P-Ew water molecules [38] so as protein atoms lie at least at 1.2 nm from the cubic box walls. The Vps27 UMI-1 and Ubiquitin partners involve each 394 and 1227 atoms, respectively. To cancel the net charge of structure 1Q0W, two Na^+ ions were added to the system using the ion generator tool of Gromacs. As specified in [30], the His residue of Ubiquitin is fully protonated (His+ state). The systems were first optimized and then heated to 50 K through a 10 ps canonical (NVT) MD, with a time step of 2 fs and LINCS constraints acting on bonds involving H atoms. The trajectory was followed by two successive 20 ps heating stages, at 150 and 300 K, under the same conditions. Next, each system was equilibrated during 50 ps in the NPT ensemble to relax the solvent molecules. Finally, a 20 ns MD simulation was performed in the NPT ensemble, for solvated systems. In vacuum, only the NVT ensemble was used. The 'V-Rescale' and 'Parrinello-Rahman' algorithms were selected to perform NVT and NPT simulations, respectively. In case of obvious lack of equilibration, an extra production run of 20 ns was performed. When considering model *mCDh*, the constraints acting on the bonds involving H atoms had to be removed and the time step was set equal to 1 fs, thus leading to twice the number of MD iterations as in the all-atom, *mCD*, and *mCDa* simulations. Snapshots were saved every 2 ps, *i.e.*, twice the value considered by Showalter and Brüschweiler [36]; that choice did not significantly alter the results. A description of the systems under study is presented in Table 1. The total number of point charges to be considered for the protein complex is reduced by a factor of 4.2 and 11.7 for the CD- and PASA-based models, respectively. Depending upon the implementation, the number of non-atomic charges is largely variable. For instance, there are only three of such point charges in model *mCDa*, which originate from the His+ and Phe residues of Ubiquitin. There are, in each simulation, approximately 10,500 water molecules that are not coarse-grained. A larger solvation box was used for the system named *All-atom-2* which corresponds to a highly different structure of the complex. This particular case will be described later in the paper, in Section 4.4.

It is clear that a real gain in simulation time will be possible if coarse-graining occurs at the solvent level. A review of coarse-grained water models can, for example, be found in [39,40]. Working with an implicit solvent representation is another efficient way to largely reduce the calculation time as discussed in [41,42]. As a recent example, let us mention the approach adopted in the coarse-grained FF PRIMO by Kar *et al.* [43]. At the present stage of our work, no information is available regarding the radius value to be assigned to the non-atomic point charges. The approach we employed earlier to calculate free energy of solvation using the program APBS [44], *i.e.*, to assign a nul radius to the point charges [10], appeared not to be effective with GROMACS. Thus, interfacing our reduced point charge models with coarse-grained or implicit solvent representations is a perspective to bring to the present work.

4.1. Molecular electrostatic maps

MEP maps are displayed in Fig. 2 for the four optimized complexes, *i.e.*, the starting points of the MD simulations, described using the all-atom, *mCD*, *mCDa*, and *mPASA* models. Solvent molecules and ions were not included in the calculations. It is noticed that the contours displayed at Fig. 2, *i.e.*, a negative MEP contour for Vps27 UIM-1 ($-0.2\ e^-\ Bohr^{-1}$) and a positive MEP contour for Ubiquitin ($+0.1\ e^-\ Bohr^{-1}$), show similar 3D shapes for each of the four models. Similar behaviours were already put forward in the study of rigid systems such as ion channels [9]. At short range, *e.g.*, in the case of intramolecular interactions, the point charge models are expected to affect the dynamical behaviour of the molecules as detailed further in the paper. To preliminary illustrate that assumption, MEP maps are displayed for two individual AAs, Glu$^-$ and His+ (Fig. 3). In the case of glutamate, one observes higher MEP values at the two negative side chain charges for model *mCD* and, obviously, the absence of such a separation for model *mPASA* which involves only one side chain negative charge. The all-atom representation of the main chain is the only one to let appear two negatively charged sites. One may thus expect structural changes, especially in secondary structure elements. For protonated histidine, the positive MEP areas look similar, except for the main chain and the *mPASA* model. One thus assumes that a change in the point charge model will affect, at least, the formation of H-bonds during a MD simulation. As seen later during the analysis of the MD trajectories, changes in the secondary structure elements will be observed, as well as in the fold of the complex in some cases. As a tentative to eliminate the effect of Ubiquitin structural changes on Vps27 UIM-1, separate MD simulations were carried out by restraining Ubiquitin to its crystal structure. Such trials did not prevent Vps27 UIM-1 to alter its shape and these MD simulations will not be presented in the paper. All results presented below concerns fully

Please cite this article in press as: L. Leherte, D.P. Vercauteren, Evaluation of reduced point charge models of proteins through Molecular Dynamics simulations: Application to the Vps27 UIM-1–Ubiquitin complex, J. Mol. Graph. Model. (2013), http://dx.doi.org/10.1016/j.jmgm.2013.10.011

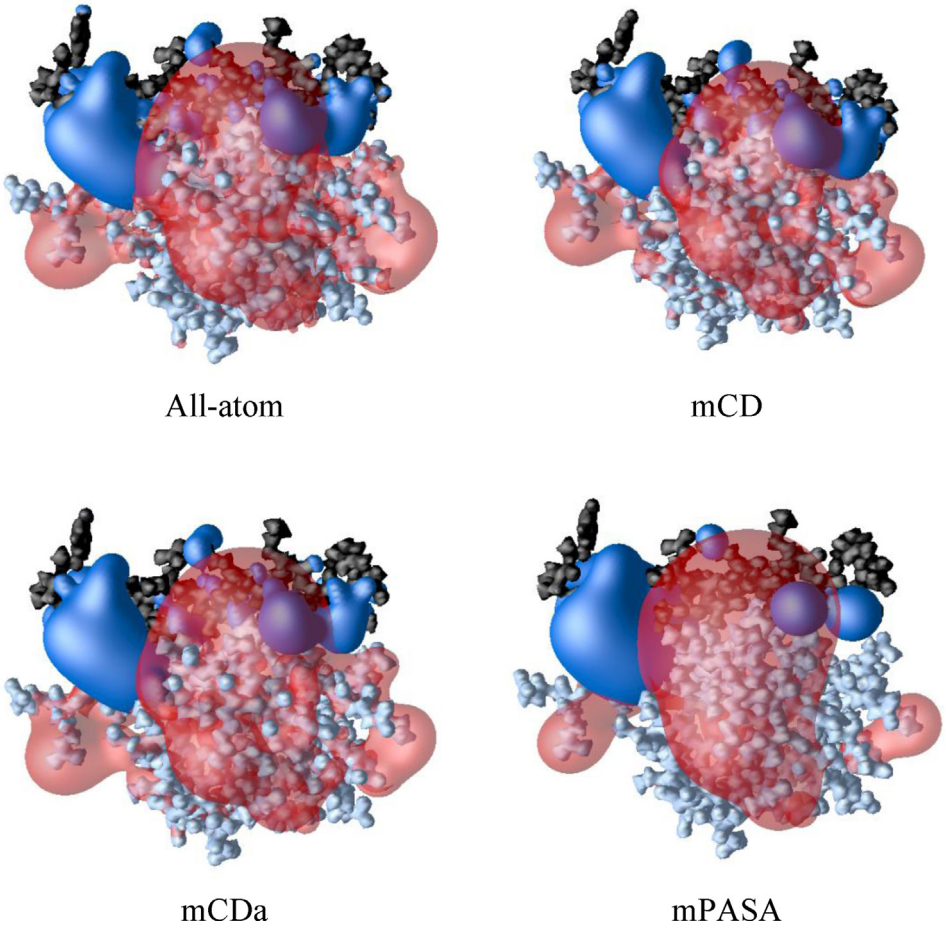


Fig. 2. MEP contours of Ubiquitin (red: +0.1 e^- $Bohr^{-1}$) and Vps27 UIM-1 (blue: −0.2 e^- $Bohr^{-1}$) in their initial optimized configuration, as obtained using the *All-atom*, *mCD*, *mCDa*, and *mPASA* models. Ubiquitin and its ligand are displayed using light blue and black spheres, respectively. Figures were generated using OpenDX [26]. (For an Q2 interpretation of the references to colour in the artwork, the reader is referred to the web version of the article).

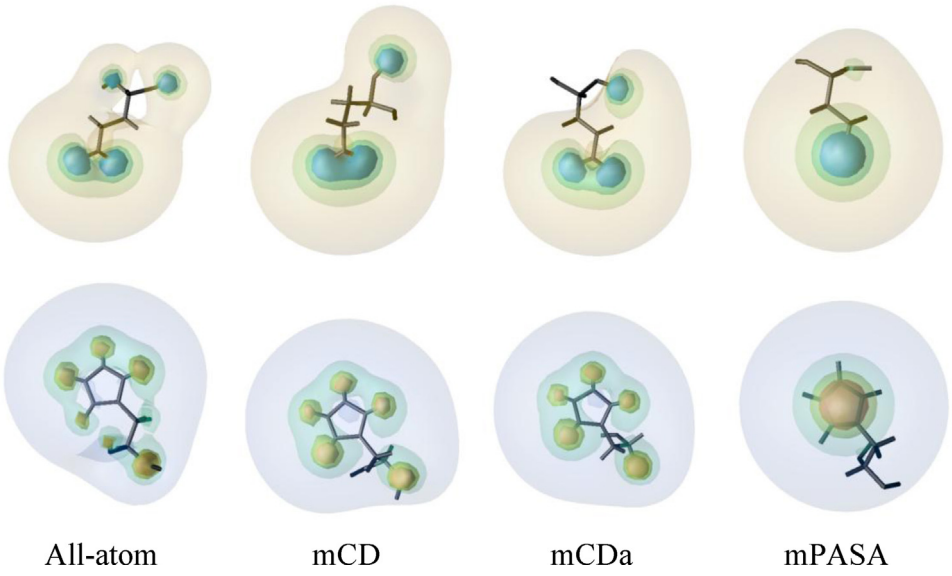


Fig. 3. MEP contours of (top) Glu−273 (−0.4 to −0.1 e^- $Bohr^{-1}$) and (bottom) His+68 (0.1 to 0.5 e^- $Bohr^{-1}$) as obtained using the *All-atom*, *mCD*, *mCDa*, and *mPASA* models. Increment = 0.1 e^- $Bohr^{-1}$. Figures were generated using OpenDX [26].

Please cite this article in press as: L. Leherte, D.P. Vercauteren, Evaluation of reduced point charge models of proteins through Molecular Dynamics simulations: Application to the Vps27 UIM-1–Ubiquitin complex, J. Mol. Graph. Model. (2013), http://dx.doi.org/10.1016/j.jmgm.2013.10.011

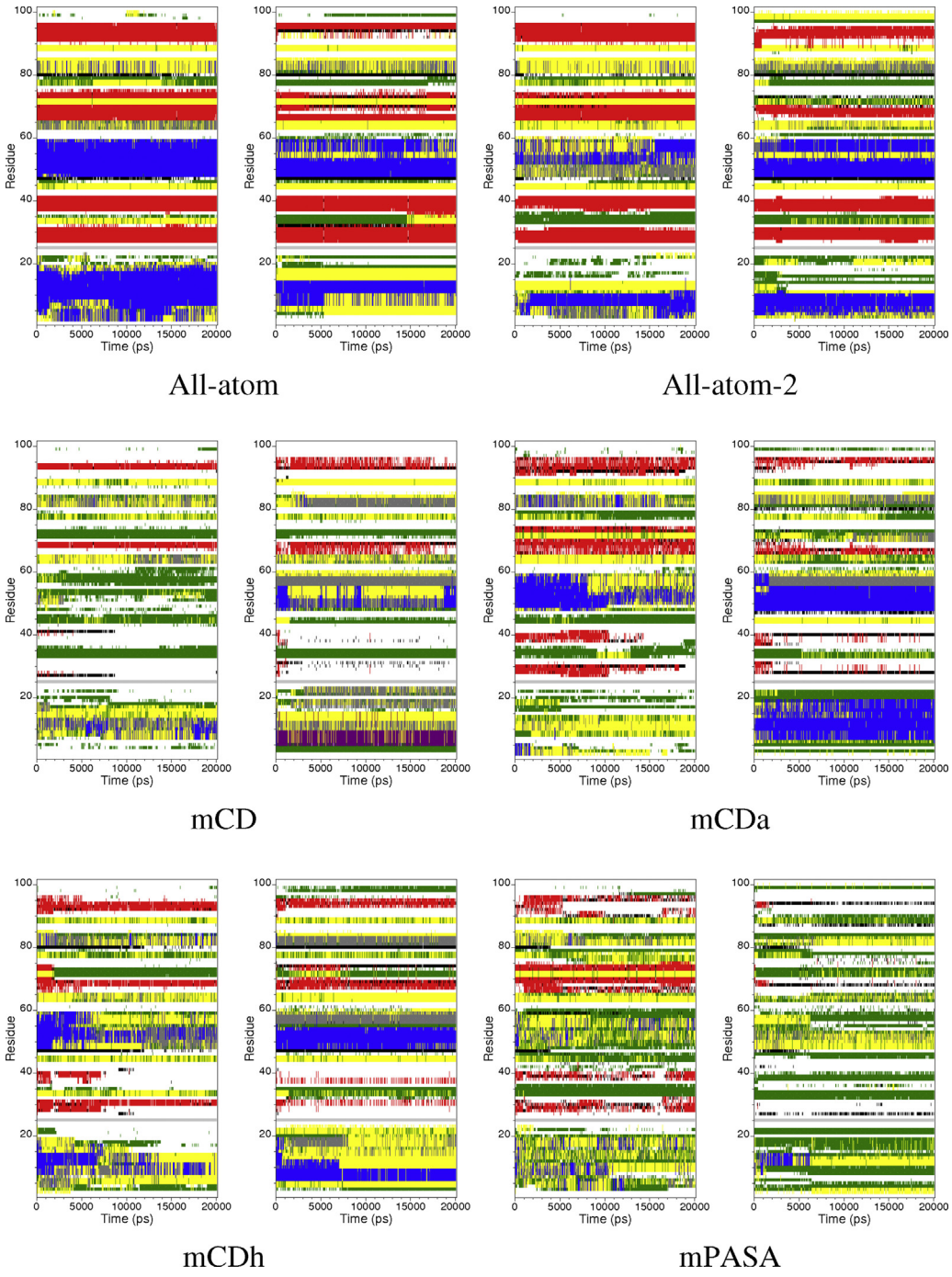


Fig. 4. Secondary structure of the complex Vps27 UIM-1–Ubiquitin observed during the last 20 ns Amber99SB-based MD trajectories in water (left) and in vacuum (right) at 300 K, as obtained using the *All-atom*, *All-atom-2*, *mCD*, *mCDa*, *mCDh*, and *mPASA* models. Vps27 UIM-1 and Ubiquitin involve the first 24 and last 76 amino acid residues, respectively. Secondary structure elements are colour-coded as follows: Coil (white), α-helix (blue), π-helix (purple), 3_{10} helix (grey), β-sheet (red), β-bridge (black), bend (green), turn (yellow), chain separation (light grey).

flexible systems, except for constraints applied to bonds involving H atoms as mentioned above.

4.2. All-atom MD trajectories

A first analysis of the MD trajectories obtained for the model named *All-atom* dealt with the secondary structure of the whole complex and showed that it is slightly more conserved in water (Fig. 4 left) than in vacuum (Fig. 4 right), due to the stabilizing contribution of water on the protein structure. In vacuum, all secondary structure elements like α-helices and β-strands, except β-strands 2–7 and 12–16 of Ubiquitin, are shorter. In both cases, the helical structure of the UIM-1 unit is loosen, to a larger extend in vacuum (Figs. 4 and 5).

A study of the 3D fold of the protein complex was achieved to especially determine the binding of Vps27 UIM-1 to Ubiquitin. Distance maps between the atoms of the two partners were built by considering the minimal distances between the AA atoms of the two partners and are reported in Fig. 6. The distances were calculated over the last 10 ns of the MD trajectories to minimize the effect of a slow secondary structure relaxation process that occurs for the solvated systems modelled, as shown later, using

Please cite this article in press as: L. Leherte, D.P. Vercauteren, Evaluation of reduced point charge models of proteins through Molecular Dynamics simulations: Application to the Vps27 UIM-1–Ubiquitin complex, J. Mol. Graph. Model. (2013), http://dx.doi.org/10.1016/j.jmgm.2013.10.011

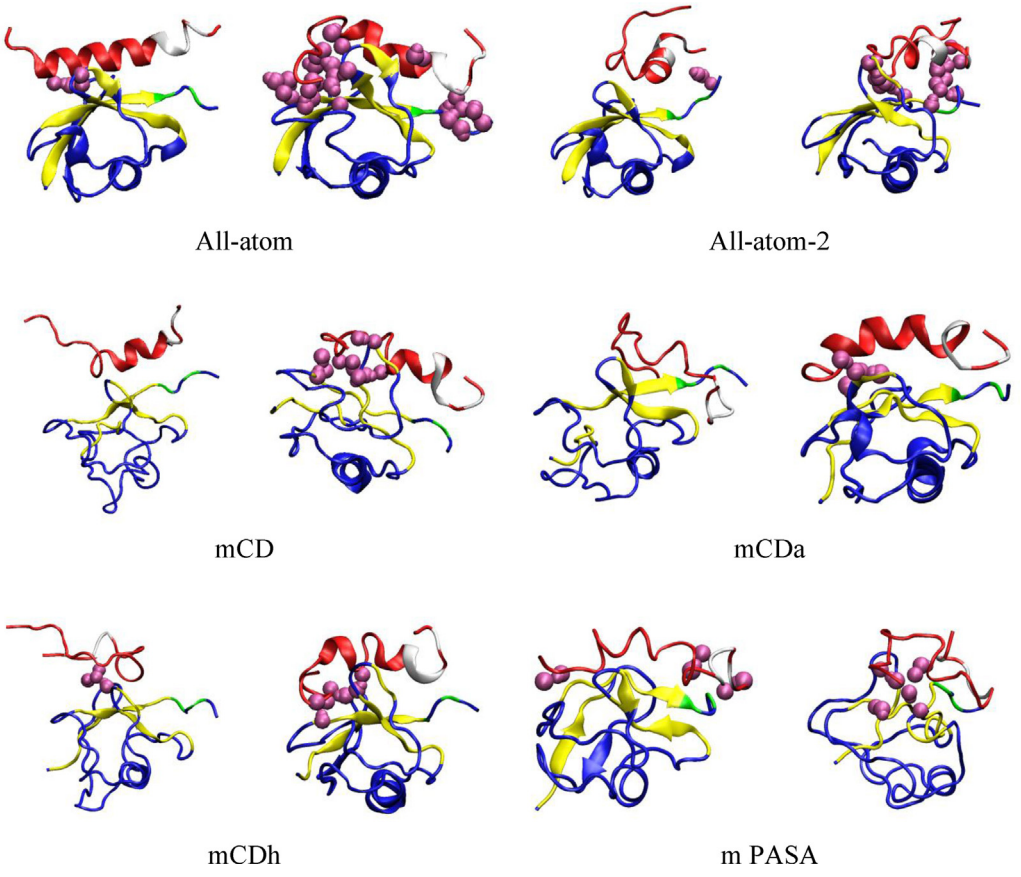


Fig. 5. Final snapshots of the protein complex Vp27 UIM-1 (red)–Ubiquitin (blue) obtained from the last 20 ns Amber99SB-based MD trajectories in water (left) and in vacuum (right) at 300 K, as generated using the *All-atom*, *All-atom-2*, *mCD*, *mCDa*, *mCDh*, and *mPASA* models. Structural elements are colour-coded as follows: β-strands of Ubiquitin interacting closely with Vps27 UIM-1 (yellow), Arg+72 and Arg+74 (green), Glu−257 and Glu−259 to Glu−261 (white), H-bonds between the two partners (purple spheres). Figures were generated using VMD [45]. (For an interpretation of the references to colour in the artwork, the reader is referred to the web version of the article).

mCD, *mCDa*, and *mCDh*. In the map generated by the analysis of the solvated *All-atom* MD trajectory, one clearly distinguishes three regions extended along the Vps27 UIM-1 chain. This extension is due to the spatial alignment of Vps27 UIM-1 with a number of β-strands of Ubiquitin. The first region corresponds to the contacts occurring between segment 259 to 272 of Vps27 UIM-1 and β-strand 4–10 of Ubiquitin, while the second and third regions are due to contacts with β-strands 40–45 and 48–49 and β-strand 66–72, respectively. The shortest distances, below 0.3 nm, appear in these two last areas. There is a last region of interest, generated by the contacts between Glu−257 and Glu−259 of Vps27 UIM-1 and Arg+ residues located at the C-terminal segment of Ubiquitin. In vacuum, these four regions are enhanced due to the closer location of Vps27 UIM-1 *versus* Ubiquitin. They are larger as they now involve the C-terminal residues of Vps27 UIM-1 with emphasized electrostatic "contacts" between the N-terminal Glu− residues (257, 259–261) of UIM-1 and N-terminal Arg+ residues (72 and 74) of Ubiquitin.

The analysis of selected energy terms averaged over the last 10 ns of the MD trajectories provided the results reported in Table 2. The absolute value of the interaction energy between the two partners of the complex, $|E_{12}|$, occurs with a ratio of 4.6 and 13.6% *versus* the total solute potential energy $(E_1 + E_2)$, in water and in vacuum, respectively. When Coulomb interactions are concerned, the corresponding ratio in water, 2.2%, also increases to reach a value of 7.8% in vacuum. The higher relative contribution of $|E_{12}|$ and $|Cb_{12}|$ supports the higher compactness of the complex as just discussed from distance maps. Intra-molecular total and Coulomb potential energies of Vps27 UIM-1, *i.e.*, $|E_1|$ and $|Cb_1|$ terms, are rather constant. They contribute to 25.9 and 24.7%, and 26.2 and 23.6%, respectively in water and in vacuum (Table 2). The vacuum-induced compactness involves a decrease in the mobility of the UIM atoms *versus* Ubiquitin as illustrated by the Root Mean Square Fluctuations (RMSF) of the Vps27 UIM-1 atoms with

Table 2
Energy ratios calculated from values obtained from Amber99SB-based MD trajectories (reported in SI 7) for the Vps27 UIM-1 Ubiquitin system. Indices '1' and '2', stand for Vps27 UIM-1 and Ubiquitin, respectively. *E* and *Cb* stand for total (Coulomb + Lennard–Jones) and Coulomb potential energy, respectively.

	$\frac{E_{12}}{E_1+E_2}$	$\frac{Cb_{12}}{Cb_1+Cb_2}$	$\frac{E_1}{E_1+E_2}$	$\frac{Cb_1}{Cb_1+Cb_2}$
Solvated				
All-atom	−4.6	−2.2	25.9	24.7
All-atom-2	−2.1	−0.6	26.8	25.1
mCD	−2.2	−1.0	21.1	20.5
mCDa	−3.7	−2.0	24.2	22.8
mCDh	0.001	−363.6	18.5	53.4
mPASA	34.6	34.8	12.9	29.9
mCD (277 K)	−2.3	−1.4	22.4	21.2
mCD (250 K)	−2.1	−1.0	22.6	21.3
mCD (150 K)	−2.4	−0.9	21.5	20.9
Vacuum				
All-atom	−13.6	−7.8	26.2	23.6
All-atom-2	−11.4	−6.6	24.7	22.7
mCD	−9.6	−5.9	19.3	18.0
mCDa	−11.4	−7.0	24.2	22.1
mCDh	−5.8	−620.4	18.1	−3.0
mPASA	24.0	23.3	21.3	27.9
mCD (277 K)	−6.4	−3.4	18.5	17.2
mCD (250 K)	−8.4	−4.8	19.5	18.3
mCD (150 K)	−7.2	−4.0	18.5	17.2

Please cite this article in press as: L. Leherte, D.P. Vercauteren, Evaluation of reduced point charge models of proteins through Molecular Dynamics simulations: Application to the Vps27 UIM-1–Ubiquitin complex, J. Mol. Graph. Model. (2013), http://dx.doi.org/10.1016/j.jmgm.2013.10.011

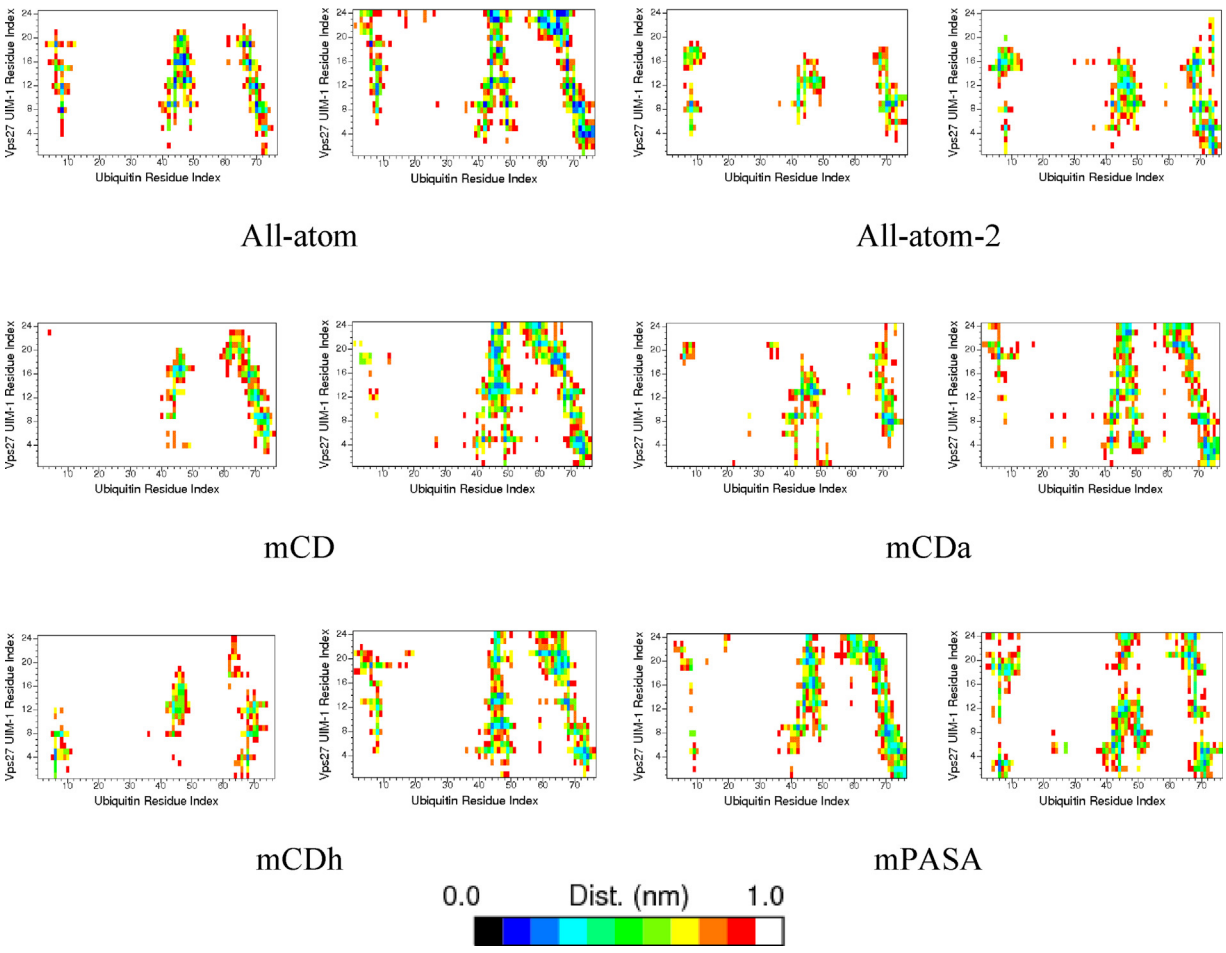


Fig. 6. Distance maps of the protein complex Vps27p-Ubiquitin established during the last 10 ns of the Amber99SB-based MD trajectories in water (left) and in vacuum (right) at 300 K, as obtained using the *All-atom*, *All-atom-2*, *mCD*, *mCDa*, *mCDh*, and *mPASA* models. Scale is coloured using a distance increment of 0.1 nm. (For an interpretation of the references to colour in the artwork, the reader is referred to the web version of the article).

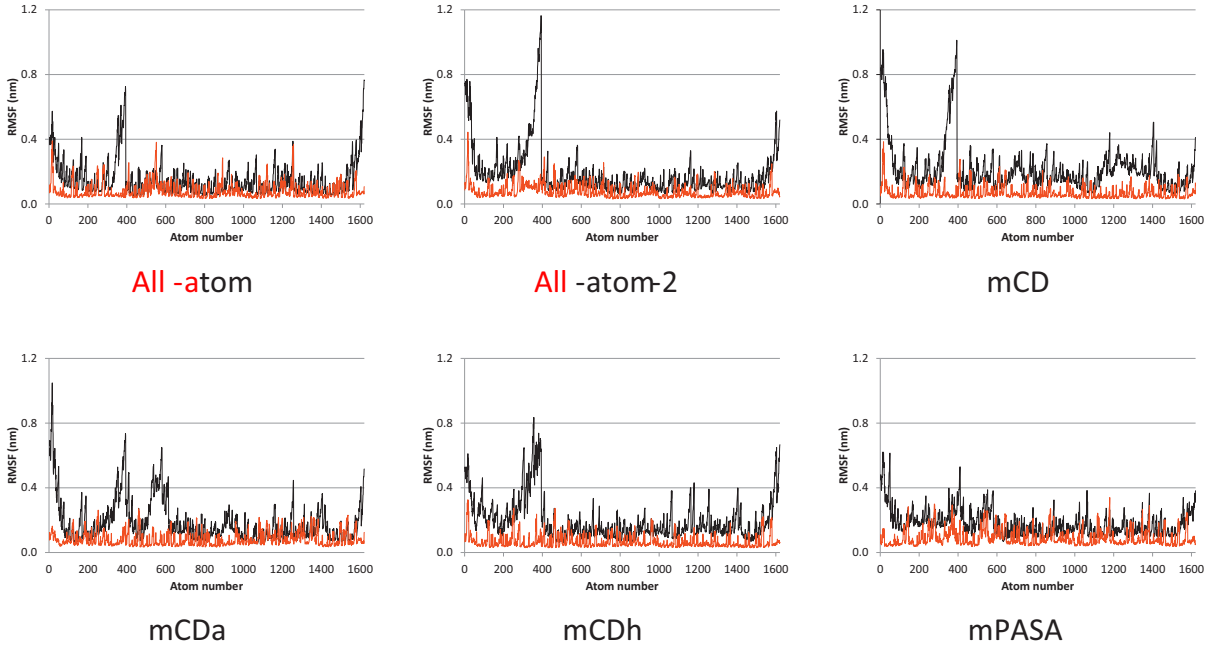


Fig. 7. RMSF of the Vps27 UIM-1 and Ubiquitin atoms (atoms 1–394 and 395–1621, respectively) in water (black) and in vacuum (red) calculated from the last 10 ns of the Amber99SB-based MD trajectories at 300 K, as obtained using the *All-atom*, *All-atom-2*, *mCD*, *mCDa*, *mCDh*, and *mPASA* models. (For an interpretation of the references to colour in the artwork, the reader is referred to the web version of the article).

Please cite this article in press as: L. Leherte, D.P. Vercauteren, Evaluation of reduced point charge models of proteins through Molecular Dynamics simulations: Application to the Vps27 UIM-1–Ubiquitin complex, J. Mol. Graph. Model. (2013), http://dx.doi.org/10.1016/j.jmgm.2013.10.011

Table 3
Average numbers of H-bonds occurring between various components of the Vps27 UIM-1–Ubiquitin system, as obtained from the analysis of the last 10 ns of the Amber99SB-based MD trajectories at 300 K.

	UIM–Ubiquitin		Complex–water					UIM–water
	Solvated	Vacuum	Total	Main chain	Side chains	N—H	C=O	
All-atom	3.6 ± 1.7	14.0 ± 1.4	287.8 ± 8.3	105.2 ± 5.0	182.5 ± 7.3	20.9 ± 2.9	69.1 ± 3.7	84.5 ± 4.6
All-atom-2	0.9 ± 1.0	10.4 ± 1.1	300.1 ± 8.9	119.6 ± 5.1	180.5 ± 7.4	9.3 ± 1.6	95.2 ± 1.3	97.4 ± 4.9
mCD	1.6 ± 1.0	7.7 ± 1.5	330.2 ± 9.5	197.0 ± 6.0	133.2 ± 7.1	23.8 ± 3.9	158.2 ± 4.2	90.0 ± 5.0
mCDa	0.6 ± 0.8	5.7 ± 1.7	333.2 ± 11.4	175.1 ± 7.8	157.9 ± 6.8	23.2 ± 3.8	138.7 ± 5.9	101.6 ± 4.8
mCDh	2.0 ± 1.2	4.7 ± 1.4	300.4 ± 9.3	176.2 ± 6.7	124.2 ± 7.2	20.8 ± 3.7	140.1 ± 6.0	89.6 ± 5.0
mPASA	2.3 ± 1.4	3.2 ± 1.4	116.3 ± 8.2	21.7 ± 4.1	94.6 ± 6.7	6.5 ± 2.6	6.2 ± 2.3	42.2 ± 4.8
mCD (277 K)	1.2 ± 0.9	1.0 ± 0.9	328.4 ± 9.2	188.3 ± 5.3	140.1 ± 7.4	24.1 ± 3.6	148.6 ± 3.6	104.8 ± 6.3
mCD (250 K)	0.6 ± 0.7	5.5 ± 1.2	308.5 ± 8.5	161.7 ± 8.1	146.8 ± 7.2	18.0 ± 3.3	128.4 ± 6.0	101.6 ± 4.9
mCD (150 K)	3.1 ± 0.4	3.2 ± 0.7	264.2 ± 5.4	118.0 ± 2.3	146.2 ± 4.8	9.3 ± 1.6	95.2 ± 1.3	81.1 ± 2.8

time (Fig. 7). One clearly distinguishes a drastic decrease in the RMSF values associated with the atoms of the end segments of the peptide UIM-1, *i.e.*, around atoms 1–100 and 300–394. This decrease in mobility is correlated to a higher average number of H-bonds occurring between the two partners of the system, 14.0 ± 1.4, about four times the number of H-bonds in water, *i.e.*, 3.6 ± 1.7, as reported in Table 3. H-bonds are determined based on cut-off values of 30° and 0.35 nm for the angle Hydrogen–Donor–Acceptor and the distance Donor–Acceptor, respectively. In water, very low numbers of H-bonds can be observed between the two partners. For instance, the solvated complex lets appear only two H-bonds at t = 20 ns, which are formed by N—H(Gly47)···OG(Ser270) and NE2-HE2(His+68)···OE2(Glu−273) atoms. These two H-bonds are characterized by a percentage of occurrence above 80% during the simulation time and are thus the most persistent ones as emphasized in SI 8 where the occupancy of the H-bonds formed between the two protein partners is reported. H-bond occupancy along MD trajectories was calculated using VMD1.9.1 [45] with threshold distance and angle values of 3.5 Å and 30°, respectively. Contrarily, in vacuum, up to 15 H-bonds appear at t = 20 ns (Fig. 5), which mostly involve the end segments of UIM with Glu– residues that fold towards the Arg+ residues of Ubiquitin. In vacuum, a larger number of H-bonds appear as a substitute to the protein-solvent H-bond network observed in the solvated state.

Protein hydration can be studied through the analysis of Radial Distribution Functions (RDF) as plotted in Fig. 8. As explained in the paper by Virtanen *et al.* [46], a first hydration shell occurs between 0.1 and 0.2 nm from the protein atoms, and is followed by a second sharply marked shell just below 0.3 nm (Fig. 3 of [46]). In the present work, RDF between oxygen atoms of water and the protein atoms show a slight first hydration shell at a distance of 0.192 nm with a contact distance of 0.154 nm. This shell involves a very limited number of water molecules; integration under the first peak of the RDF function leads to a value of 84 H_2O molecules. A second shell appears at about 0.280 nm, a distance that was actually identified as the first hydration shell of crystalline proteins by Chen *et al.* [47]. The second shell identified by Chen *et al.*, corresponding to water interactions with protein non-polar atoms is, in our simulated protein system, located at 0.372 nm.

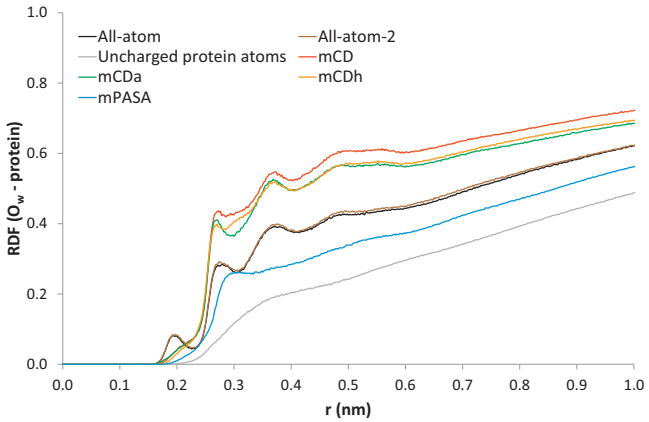


Fig. 8. RDF of water oxygen–protein atom pairs of the solvated system Vps27 UIM-1–Ubiquitin calculated from the last 10 ns of the Amber99SB-based MD trajectories at 300 K, as obtained using the *All-atom*, *All-atom-2*, *mCD*, *mCDa*, *mCDh*, *mPASA*, and neutral protein atom models. (For an interpretation of the references to colour in the artwork, the reader is referred to the web version of the article).

Besides a difference in the UIM shape between the solvated and vacuum states (Fig. 5), one additionally notices that the gyration radius r_G of Ubiquitin is affected by the environment. Indeed, mean values of 1.20 ± 0.01 and 1.15 ± 0.01 nm are obtained in water and in vacuum, as reported in Table 4 wherein averages were calculated over the last 10 ns of the final MD trajectories as drifts in r_G were still appearing during the first 10 ns. The slight structure contraction observed in vacuum is due to the lack of interactions with surrounding water molecules and comes with reduced atom fluctuations (Fig. 7).

Regarding the solvent itself, a calculation of the mean square displacement as a function of time, carried out for molecules located within 0.35 nm, and between 0.35 and 1.2 nm from the protein structure, shows that all sets of water molecules behave as a Brownian fluid with a similar self-diffusion coefficient D of (2.38 ± 0.01) × 10^{-5} and (2.40 ± 0.02) × 10^{-5} $cm^2\ s^{-1}$, respectively (Table 4). The very slight decrease in D for water molecules interacting closely with the solute does not appear to be significant; all D values stay close to the self-diffusion coefficient of water calculated with the TIP4P-Ew potential, *i.e.*, (2.4 ± 0.06) × 10^{-5} $cm^2\ s^{-1}$ [38].

4.3. *mCD MD trajectories*

As explained later when discussing the shape of the Ubiquitin partner, a second 20 ns MD production stage was carried out for the solvated complex when using the *mCD* model. All results discussed below were obtained from the analysis of that additional run.

The study of the secondary structure of the complex in water and in vacuum directly shows an enhanced loss of the secondary structure elements of the system *versus* the *All-atom* simulation results (Fig. 4). Part of the β-strands disappears and helices are the motifs that are the most perturbed during the simulations. However, in vacuum, regular motifs, especially the helix of the Ubiquitin structure, appear to be slightly more preserved. Indeed, in water, the electrostatic interaction of the protein with the solvent molecules is particularly modified as illustrated by the number of H-bonds occurring between the main chain and the side chains of the solute and water (Table 3). Surprisingly, a larger average number of main chain H-bonds, 197.0 ± 6.0 *versus* 105.2 ± 5.0, is obtained for *mCD* despite the absence of charges on N and H atoms. It appears to be due to the C=O groups that, with their different charge distribution

Please cite this article in press as: L. Leherte, D.P. Vercauteren, Evaluation of reduced point charge models of proteins through Molecular Dynamics simulations: Application to the Vps27 UIM-1–Ubiquitin complex, J. Mol. Graph. Model. (2013), http://dx.doi.org/10.1016/j.jmgm.2013.10.011

Table 4
Mean gyration radius r_G of Ubiquitin and self-diffusion coefficient D of water molecules solvating the Vps27 UIM-1–Ubiquitin system as obtained from the analysis of the last 10 ns of the Amber99SB-based MD trajectories at 300 K.

	r_G (nm)		D ($\times 10^{-5}$ cm^2 s^{-1})	
	Solvated	Vacuum	Within 0.35 nm of the solute	Between 0.35 and 1.20 nm from the solute
All-atom	1.20 ± 0.01	1.15 ± 0.01	2.38 ± 0.01	2.40 ± 0.02
All-atom-2	1.19 ± 0.01	1.12 ± 0.00	2.46 ± 0.08	2.39 ± 0.04
mCD	1.36 ± 0.02	1.15 ± 0.00	2.35 ± 0.05	2.31 ± 0.02
mCDa	1.30 ± 0.01	1.12 ± 0.00	2.20 ± 0.02	2.33 ± 0.02
mCDh	1.28 ± 0.01	1.13 ± 0.00	2.37 ± 0.03	2.36 ± 0.05
mPASA	1.17 ± 0.01	1.14 ± 0.01	2.46 ± 0.08	2.47 ± 0.05
mCD (277 K)	1.25 ± 0.01	1.13 ± 0.00	1.21 ± 0.01	1.26 ± 0.04
mCD (250 K)	1.20 ± 0.01	1.12 ± 0.00	0.38 ± 0.00	0.43 ± 0.00
mCD (150 K)	1.20 ± 0.00	1.12 ± 0.00	~10^{-5}	~10^{-5}

versus the *All-atom* case, affects the formation of such a type of interaction. Even in the absence of charges on the N and H atoms of the AA main chains, the average number of H-bonds is not drastically modified, with 23.8 ± 3.9 *versus* 20.9 ± 2.9 bonds observed for the *mCD* and *All-atom* models, respectively. Contrarily, the number of H-bonds involving the side chain atoms is reduced to an average value of 133.2 ± 7.1 *versus* 182.5 ± 7.3 for *mCD* and *All-atom* models. The consequence of those changes in the number of H-bonds formed with the solvent is illustrated in Fig. 8, where it is clearly seen that the very first hydration shell is reduced to a weak shoulder in the RDF curve with model *mCD*. Nevertheless, a study of the distance and angle values adopted by the H-bonds shows that the distributions for model *mCD* are similar to those that are valid for the *All-atom* model, *i.e.*, centred around 0.27 nm and 10.5° (Fig. 9 top). It is however noticed that, regarding intra-protein H-bonds (Fig. 9 bottom), if the distance distribution is rather similar to the all-atom one, there is a significant displacement of the maximum of the angle distribution towards higher angle values, *i.e.*, about 25° *versus* 15° for the *All-atom* model. These last trends are observed for both the solvated and isolated systems.

As visualized in Fig. 5, both ends of the solvated UIM-1 get separated from Ubiquitin while the central hydrophobic segment (Ile263 to Leu271) stays at the proximity of the two stable β-strands of Ubiquitin, *i.e.*, segments 39–44 and 67–70. The consequence of such a configuration change is illustrated by the distance maps displayed in Fig. 6 where a contact area involving the first residues of Ubiquitin fades away due to the loss of the two first β-strand elements occurring along the Ubiquitin chain, *i.e.*, segments 2–7 and 12–16. On the contrary, β-strands 40–45 and 48–49, as well as strand 66–72 are strongly preserved but the H-bonds they form are characterized by occurrence percentage values lower than 20%. Nevertheless, the lower average number of H-bonds formed between the two partners adopt a similar behaviour as with the *All-atom* model, in the sense that it is very limited in water while

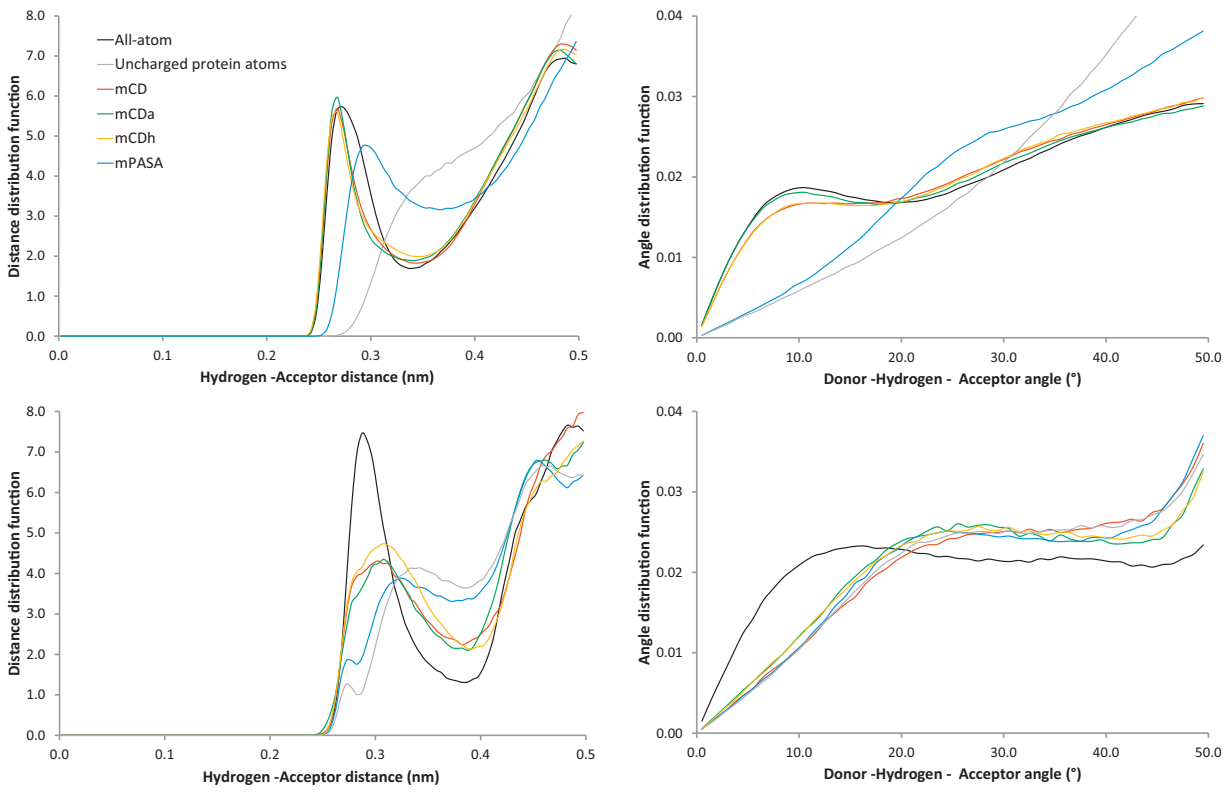


Fig. 9. (Left) Distance and (right) angle distribution functions of the solvated system Vps27 UIM-1–Ubiquitin calculated from the last 10 ns of the Amber99SB-based MD trajectories at 300 K, as obtained using the *All-atom*, *mCD*, *mCDa*, *mCDh*, *mPASA*, and neutral protein atom models. (Top) Protein-water H-bonds, (bottom) intra-molecular H-bonds. (For an interpretation of the references to colour in the artwork, the reader is referred to the web version of the article).

Please cite this article in press as: L. Leherte, D.P. Vercauteren, Evaluation of reduced point charge models of proteins through Molecular Dynamics simulations: Application to the Vps27 UIM-1–Ubiquitin complex, J. Mol. Graph. Model. (2013), http://dx.doi.org/10.1016/j.jmgm.2013.10.011

Table 5
Total and inter-protein energy values (kJ mol^{-1}) of the optimized configurations of system Vps27 UIM-1–Ubiquitin in vacuum used as starting points for Amber99SB-based MD simulations. Subscripts '1' and '2' stand for Vps27 UIM-1 and Ubiquitin, respectively. '14' denotes interactions between atoms separated by 3 chemical bonds.

	All-atom	mCD	mCDa	mCDh	mPASA
Stretching	334.9	287.4	153.2	193.6	150.9
Bending	619.8	593.2	741.0	557.1	663.5
Torsion	3780.0	3774.0	3845.5	3773.7	3777.6
Improper	43.2	45.1	55.5	47.5	23.9
LJ-14	1823.4	1824.8	1574.0	1785.6	1471.7
Cb-14	17,025.5	20,693.4	16,909.2	20,670.8	–
LJ	−1651.8	−1631.1	−2473.6	−1774.7	−2773.3
Cb	−26,887.7	−28,909.8	−26,792.9	−32,096.4	−2687.9
Cb_{12}	−996.9	−1038.2	−1239.5	−1285.7	−932.7
LJ_{12}	13.1	15.8	−130.0	−3.6	−155.4

it increases in vacuum (Table 3). As an example, there are no H-bonds occurring at $t = 20$ ns in the solvated system (Fig. 5). The apparent separation of Vps27 UIM-1 from Ubiquitin does however not come with a drastic change of the number of H-bonds formed between the UIM and the solvent (Table 3). On the average, there is an increase of only 5.5 H-bonds, from 84.5 ± 4.6 to 90.0 ± 5.0. Indeed, the expected larger increase of the number of H-bonds due to the configuration change is compensated by the decrease in the possibility to form H-bonds due to the reduced point charge model.

Energy values obtained with the different point charge models are hardly comparable one to each other. We therefore chose to consider energy ratios such as $|E_{12}|/(E_1 + E_2)$ and $|Cb_{12}|/(Cb_1 + Cb_2)$ to evaluate the proportion of the protein–protein interaction energy *versus* intra-molecular energy values (Table 2). In water, the corresponding energy ratios, 2.2 and 1.0% respectively, indicate a lower relative importance of the *mCD* inter-molecular potential energy *versus* the *All-atom* model, i.e., 4.6 and 2.2%, respectively, with, however, a similar ratio $[E_{12}/(E_1 + E_2)]/[Cb_{12}/(Cb_1 + Cb_2)]$. Similar trends are observed in vacuum. To allow more detailed comparisons between energy contributions from the various point charge models, a detailed decomposition of the total potential energy was achieved for the optimized initial structures of the complex in vacuum to avoid any solvent contribution (Table 5). All energy terms of these conformationally close 3D structures are of the same orders of magnitude, but the bond energy is lower *versus* the all-atom contribution, 287.4 *versus* 334.9 kJ mol^{-1}, and the Coulomb term involving atoms separated by three bonds, Cb-14, is higher and destabilizing, 20,693.4 *versus* 17,025.5 kJ mol^{-1}.

Similarly to the *All-atom* model, the RMSF function clearly emphasizes the greater mobility of the Vps27 UMI-1 ends *versus* Ubiquitin in water (Fig. 7). Ubiquitin itself is also affected by the change in the point charge representation. This can be shown, for example, by an analysis of its gyration radius r_G which progressively increases during the first 20 ns production stage. As already mentioned, an additional 20 ns simulation was performed and confirmed a higher value of the gyration radius r_G for Ubiquitin in water than in vacuum, with 1.36 ± 0.02 and 1.15 nm, respectively (Table 4). As actually seen further in the paper, all r_G values obtained in water are higher than those in vacuum, regardless of the point charge model used. Model *mCD* leads to an increase of r_G by about 12% (from 1.20 to 1.36 nm), while the change in vacuum is imperceptible (1.15 nm in both cases) even if the secondary structure elements are affected. It indicates that the solvent may serve as an intermediate in the modification of the protein structure.

Having observed that the structural stability of the system is modified with model *mCD versus* the *All-atom* representation, additional MD simulations were achieved at three lower temperatures, *i.e.*, 277, 250, and 150 K. The two last temperature values were not selected to reflect a physical state for water (they are both below the freezing point of the solvent) but were chosen to locally probe the potential energy hyper-surface of the system. The analyses of the 150 K trajectory clearly show stable protein structures, as illustrated by the time evolution of the secondary structure (Fig. 10). At higher temperature values, a deconstruction of the secondary structure elements, particularly the helices, is observed, with a slow down as the temperature decreases. Simultaneously, the RMSF of the atoms of both partners also decreases and, at 150 K, do not show any maxima at protein ends.

At the lowest temperatures, *i.e.*, 250 and 150 K, the time-dependency of r_G for Ubiquitin shows little fluctuations with both an average value of 1.20 nm (Table 4). Such a value is strictly comparable to the mean *All-atom* value, *i.e.*, 1.20 nm, obtained at 300 K. Contrarily, a contraction of Ubiquitin in vacuum is observed at temperatures lower than 300 K. Indeed, mean values of 1.15 and 1.12 nm are obtained at 300 and 250 K, respectively. As the secondary structure is preserved at low temperatures, one assumes that the potential minimum occurring at the all-atom level also exists for the *mCD* point charge model. This can also be deduced from vacuum simulations carried out at temperatures below 300 K. The analysis of the secondary structure elements in such conditions shows a very stable structure (Fig. 10) with preserved helices and β-strands, that are however slightly shorter than in the corresponding *All-atom* simulation.

The evolution of the number of H-bonds formed between Vps27 UIM-1 and Ubiquitin is described in Table 3. In water as well as in vacuum, the trend is not monotonic, *i.e.*, the lowest number of H-bonds, *i.e.*, 0.6 ± 0.7 and 1.0 ± 0.9, is not observed at the lowest temperature but at 250 and 277 K, respectively. Below and beyond those values, the numbers increase extremely fast in vacuum but smoothly in water. It might be due, at lower temperatures, to a freezing of the structure favouring a persistence of the H-bonds and, at higher temperatures, to an increased probability to form polar contacts with diverse residues due to the increased mobility of the atoms. One clearly distinguishes a loss in persistent H-bonds when using model *mCD* in water *versus* the corresponding *All-atom* case. All H-bonds are now characterized by an occurrence degree below 20%. Such values are increased only when working in vacuum and/or by reducing the temperature. Indeed, in water at $T = 150$ K, one find three types of H-bonds, *i.e.*, Leu73···Glu−259, Gly47···Ser270, His+68···Glu−273. The two last are identical to the *All-atom* H-bonds (SI 8).

Studying the behaviour of model *mCD* at low temperatures allows to re-evaluate the H-bond ratios calculated from values reported in Table 3. Let us first consider that the *mCD* system at the low temperature of 150 K probes similar conformations than the *All-atom* system. The proportion of main chain H-bonds represents 44.7% of the total number of H-bonds. That value stays higher than the corresponding *All-atom* one, *i.e.*, 36.6%. Thus, one concludes that the change in the point charge model indeed leads to an increase of the main chain H-bonds, regardless of a change in the conformation.

A specific study of the intra-molecular H-bonds occurring in solvated Ubiquitin shows that a temperature decrease tends to an increase in the number of such H-bonds, going from averages of 11.5 ± 2.7, 16.0 ± 2.8, 23.3 ± 3.1, and 32.9 ± 2.7 at $T = 300$, 277, 250,

Please cite this article in press as: L. Leherte, D.P. Vercauteren, Evaluation of reduced point charge models of proteins through Molecular Dynamics simulations: Application to the Vps27 UIM-1–Ubiquitin complex, J. Mol. Graph. Model. (2013), http://dx.doi.org/10.1016/j.jmgm.2013.10.011

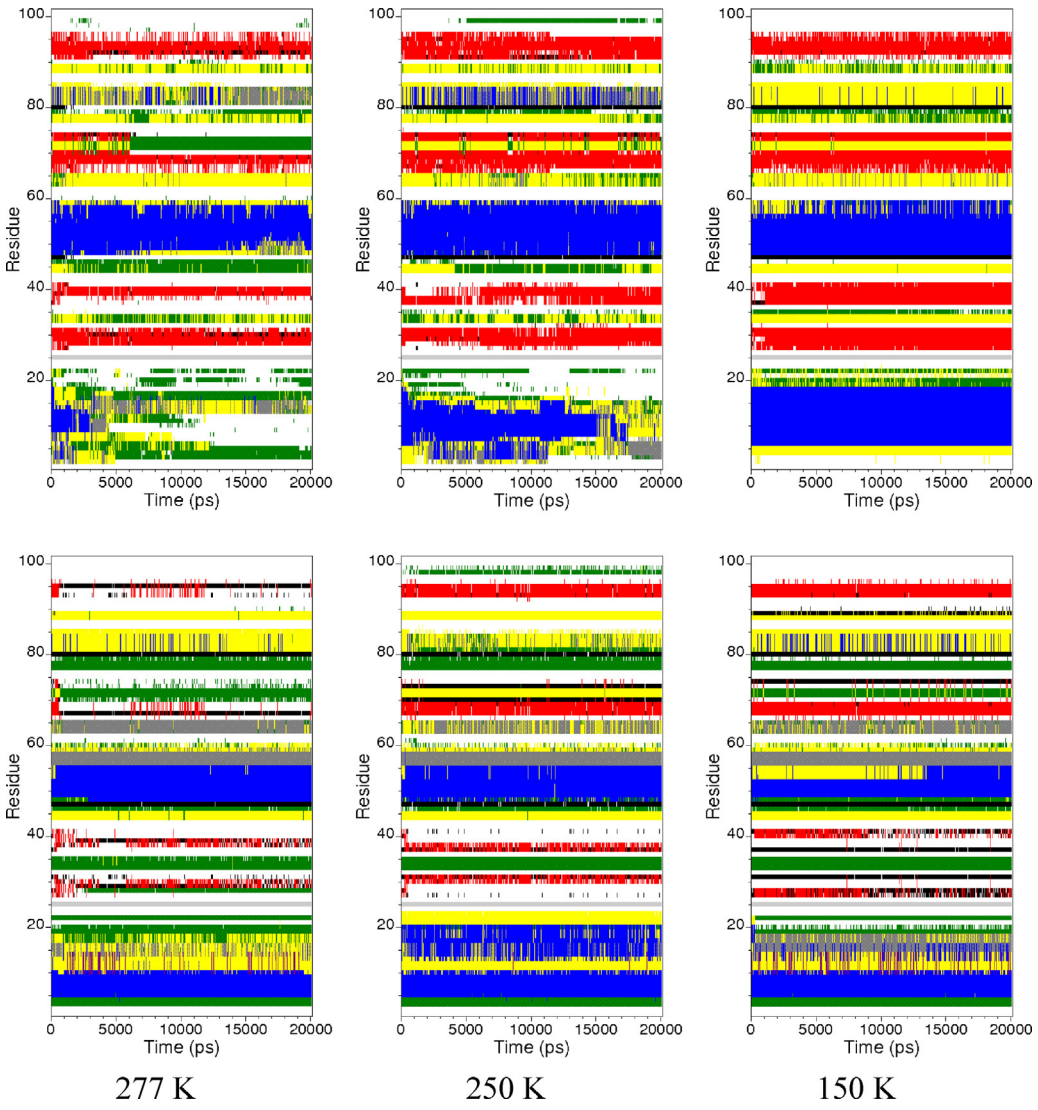


Fig. 10. Secondary structure of the complex Vps27 UIM-1–Ubiquitin observed during the last 20 ns *mCD* Amber99SB-based MD trajectories in water (top) and in vacuum (bottom) at various temperatures. Vps27 UIM-1 and Ubiquitin involve the first 24 and last 76 amino acid residues, respectively. Secondary structure elements are colour-coded as follows: coil (white), α-helix (blue), π helix (purple), 3_{10} helix (grey), β-sheet (red), β-bridge (black), bend (green), turn (yellow), chain separation (light grey). (For an interpretation of the references to colour in the artwork, the reader is referred to the web version of the article).

and 150 K, respectively (Table 6), *i.e.*, one comes closer and closer to the value of 48.0 ± 3.2 obtained for the *All-atom* model. The trend is slightly different for Vps27 UIM-1 with a minimum number of H-bonds, 1.9 ± 1.3, observed at 277 K. As interpreted earlier in the paper, higher numbers of H-bonds observed at lower temperatures might be due to an increased persistence of the H-bonds while, at higher temperatures, to an increased mobility of the atoms.

As the protein structures modelled with *mCD* reorganize at room temperature, both in water and in vacuum, one concludes that the potential hyper-surface corresponding to the reduced point charge model is, at least locally, characterized by lower energy barriers than with the all-atom model [6].

In conclusion, the *mCD* model allows to obtain stable MD trajectories without any separation of the complex partners. It has a contracting effect in vacuum that does not occur in water, leading, in that last case, to a more mobile peptide than in the all-atom case. The overall 3D folding of the complex is preserved especially for the largest and globular partner (this may be partly due to the presence of all-atom vdW contributions to the FF) while the secondary structure is significantly dismantled for the helix-shaped Vps27 UIM-1 peptide. This effect is cancelled at lower temperatures, *i.e.*, at about 250 K, which implies that the change in the point charge representation does not affect the location of the energy minimum on the potential hyper-surface but only affects its shape.

Table 6
Average numbers of intra-molecular H-bonds occurring in Vps27 UIM-1 and in Ubiquitin as obtained from the analysis of the last 10 ns of the Amber99SB-based MD trajectories at 300 K.

	Vps27 UIM-1		Ubiquitin	
	Solvated	Vacuum	Solvated	Vacuum
All-atom	16.2 ± 2.8	28.4 ± 2.1	48.0 ± 3.2	81.8 ± 3.7
All-atom-2	12.0 ± 2.4	26.7 ± 2.2	50.1 ± 3.4	81.7 ± 3.8
mCD	4.6 ± 1.5	13.9 ± 2.4	11.5 ± 2.7	31.3 ± 3.5
mCDa	2.5 ± 1.3	13.7 ± 2.2	14.3 ± 3.3	42.6 ± 3.7
mCDh	6.0 ± 2.1	12.0 ± 2.0	16.5 ± 3.1	30.5 ± 3.7
mPASA	3.8 ± 2.0	5.3 ± 1.7	14.5 ± 3.6	19.7 ± 3.3
mCD (277 K)	1.9 ± 1.3	13.0 ± 2.3	16.0 ± 2.8	34.9 ± 3.7
mCD (250 K)	5.6 ± 1.6	11.8 ± 2.2	23.3 ± 3.1	41.5 ± 3.6
mCD (150 K)	11.5 ± 1.4	14.2 ± 1.9	32.9 ± 2.7	39.4 ± 3.2

Please cite this article in press as: L. Leherte, D.P. Vercauteren, Evaluation of reduced point charge models of proteins through Molecular Dynamics simulations: Application to the Vps27 UIM-1–Ubiquitin complex, J. Mol. Graph. Model. (2013), http://dx.doi.org/10.1016/j.jmgm.2013.10.011

4.4. *mCDa* MD trajectories

Model *mCDa* was designed to facilitate the implementation of the reduced point charge model within GROMACS and to get rid, as much as possible, of the effect of a force redistribution onto the atoms used to define the virtual sites.

A first observation of the protein secondary structure in water showed that there is a slow structure loss of the UIM peptide. The initial production stage was thus continued for another 20 ns, which also depicts a progressive loss of the regular motifs of Ubiquitin (Fig. 4). The final configuration obtained from that second run is illustrated in Fig. 5, and additionally shows a preservation of the two same β-strands as with model *mCD*. In water, the UIM-1 peptide adopts a rather extended conformation, and the corresponding distance map displayed in Fig. 6 is consequently strongly modified with a decrease of all interaction areas. Contrarily, in vacuum, the two helices are well preserved (Fig. 4) and the UIM-1 orientation *versus* Ubiquitin is similar to the *All-atom* case (Fig. 6).

The mean number of H-bonds formed between the two partners in water during the simulation is still lower than for model *mCD*, *i.e.*, 0.6 ± 0.8 rather than 1.6 ± 1.0 (Table 3), due to a displacement of both end segments of Vps27 away from Ubiquitin. Particularly, there is no H-bond observed in the final complex configuration (Fig. 5). Additionally, the 3D configuration of the complex is drastically modified, leading to a rotation of the UIM-1 chain of about 90° *versus* the two preserved β-strands of Ubiquitin. This leads to a higher number of H-bonds occurring between the UIM-1 and the solvent, *e.g.*, 109 at $t = 20$ ns with an average of 101.6 ± 4.8, higher than the *All-atom* value of 84.5 ± 4.6 (Table 3). As for model *mCD*, the very first hydration shell is reduced to a weak shoulder in the RDF of oxygen-protein pairs (Fig. 8). Nevertheless, a study of the distance and angle values adopted by the protein-water H-bonds shows that the distributions are very similar to those that are valid for the *All-atom* model, *i.e.*, centered around 0.27 nm and 10.5° (Fig. 9).

In vacuum, a higher number of inter-protein H-bonds are detected, that maintain the 3D configuration of the complex close to the original PDB one, with a mean value of 5.7 ± 1.7, a value that nevertheless stays lower than for the *All-atom* (14.0 ± 1.4) and *mCD* (7.7 ± 1.5) models (Table 3). For example, three H-bonds are observed between both partners in the final configuration, *i.e.*, OG-HG(Ser277)···O(Glu−47), NZ-HZ1(Lys+6)···OE2(Glu−273), ND1-HD1(His+68)···OE2(Glu−273) (Fig. 5). Despite the large conformational change, energy ratios $|E_{12}|/(E_1 + E_2)$ and $|Cb_{12}|/(Cb_1 + Cb_2)$, *i.e.*, 3.7 and 2.0%, respectively, are closer to the values obtained for the *All-atom* model than they are for model *mCD* (Table 2). Corresponding $|E_1|/(E_1 + E_2)$ and $|Cb_1|/(Cb_1 + Cb_2)$ values are also comparable with the *All-atom* model. The energy terms of the initial optimized structure reported in Table 5 illustrate that the bond term is still lower in energy *versus* the *All-atom* and *mCD* contributions, 153.2 kJ mol^{-1}, while the Lennard–Jones (LJ) term involving atoms separated by 3 bonds, *LJ-14*, is stabilizing *versus* the corresponding *All-atom* value, 1574.0 *versus* 1823.4 kJ mol^{-1}. Additionally, the *Cb-14* term, 16,909.2 kJ mol^{-1}, is lower than in the *mCD* case, 20,693.4 kJ mol^{-1}. Those two terms indicate why the secondary structure is better preserved with *mCDa* than with *mCD*. The less good agreement of interaction energy values Cb_{12} and LJ_{12}, *i.e.*, −1239.5 and −130.0 kJ mol^{-1}, respectively, with the corresponding *All-atom* ones explains the less good reproduction of the orientation of Vps27 UIM-1 *versus* Ubiquitin.

As for the secondary structure displayed in Fig. 4, the time evolution of r_G calculated for the solvated Ubiquitin showed that the system is not fully equilibrated yet. The structure appears to expand slowly and reaches a value of 1.30 ± 0.01 nm (larger than the *All-atom* value, 1.20 ± 0.01 nm, similarly to *mCD*) while, in vacuum, it seems to be contracting to a value of 1.12 nm (Table 4).

The drastic change in the UIM structure and orientation *versus* Ubiquitin was verified using different simulation conditions. NVT conditions were used to generate a 20 ns trajectory at 300 K in water. A detailed analysis of the trajectory is not given here but a complete change in the peptide structure was also observed, particularly a hairpin-like shape, with a marked bend at the level of residues Ile267 to Leu269 as seen in Fig. 5. This particular conformation was considered as the starting point of a new all-atom MD simulation carried out in the same conditions as described above, *i.e.*, same equilibration stages, followed by a 20 ns NPT calculations at 300 K in water and in vacuum. The model will be referred to as *All-atom-2* further in the text. The analysis of the 20 ns trajectory showed that secondary structure elements like helices and β-strands re-appear (Fig. 4), but the bend persists both in water and in vacuum (Fig. 5). Additionally, r_G of solvated Ubiquitin recovers the mean value of the original *All-atom* simulation, *i.e.*, 1.19 ± 0.01 nm *versus* 1.20 ± 0.01 nm (Table 4), while it stays only slightly lower in vacuum, 1.12 nm *versus* 1.15 nm.

In water, the distance map (Fig. 6) that is associated with that particular configuration of the UIM-1 *versus* the larger partner is characterized by smaller contact areas and looser distances than in the original *All-atom* case; no minimal distance below 0.2 nm is observed. Also, very few UIM–Ubiquitin H-bonds, *i.e.*, an average of 0.9 ± 1.0, occur between the two partners (Table 3). No H-bonds with an occurrence degree higher than 30% is reported (SI8). This is compensated by a larger average number of UIM-water H-bonds, *i.e.*, 97.4 ± 4.9 *versus* 84.5 ± 4.6 for the original *All-atom* model (Table 3). It explains the relative stability of that configuration for the complex. That mean value actually hides a decrease, from about 105 to 95 H-bonds, around 7.5 ns. It corresponds to the closure of the Vps27 UIM-1 hairpin structure, as illustrated by the time evolution of the N(Tyr255)···C(Ala278) distance (Fig. 11), without any significant change in the secondary structure (Fig. 4), and is illustrated by snapshots taken at 4 and 7 ns (Fig. 11). In vacuum, the opening/closure of the peptide structure is limited in the absence of any possibility to form H-bonds with a solvent (Fig. 5). Besides that, the numbers of intra-molecular H-bonds in the original *All-atom* simulation, *i.e.*, 48.0 ± 3.2 for Ubiquitin and 16.2 ± 2.8 for UIM, are almost recovered in the hairpin-shaped UIM-1 complex with values of 50.1 ± 3.4 and 12.0 ± 2.4, respectively (Table 6).

In conclusion, model *mCDa* does not appear to allow a good preservation neither of the expected (PDB) secondary structure nor of the tertiary structure of the complex in water. Nevertheless, it is more efficient in the modelling of some secondary structure elements *versus* model *mCD*, especially in vacuum. It leads to a very slow relaxation of the system and does not favour H-bonding between the AA residues, leading consequently to an increased number of H-bonds with the solvent. The MD results suggest that a careful choice of the point charge location for a protein model may be required, and that choice can be appropriate when based on CD topology criteria. However, configurations generated by the model cannot be rejected as unphysical ones; they may be considered as local minimum configurations of the all-atom potential hyper-surface and the use of the model is a way to generate such configurations by MD simulations more rapidly than with the all-atom model.

4.5. *mPASA* MD trajectories

The use of model *mPASA*, described in Section 3.2, which is characterized by the lowest number of point charges, all located away from atom centres (except for the S atom), leads to a complete loss of the helix components of the complex, regardless of the solvent presence (Fig. 4). In water, the β-strand motifs are somewhat preserved, while they completely disappear in vacuum. However, in water, the loss in secondary structure elements does not imply

Please cite this article in press as: L. Leherte, D.P. Vercauteren, Evaluation of reduced point charge models of proteins through Molecular Dynamics simulations: Application to the Vps27 UIM-1–Ubiquitin complex, J. Mol. Graph. Model. (2013), http://dx.doi.org/10.1016/j.jmgm.2013.10.011

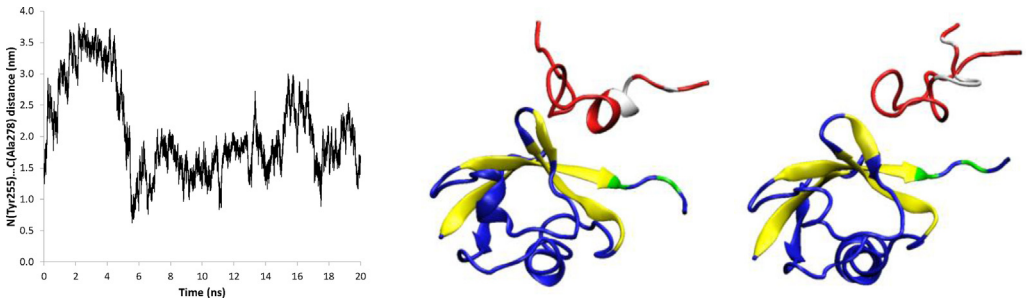


Fig. 11. (Left) Time-dependence of the distance N(Tyr255)···C(Ala278) of the protein system Vps27–UIM-1–Ubiquitin from 20 ns *All-atom-2* Amber99SB-based MD trajectories in water at 300 K. Snapshots taken at (middle) 4.0 and (right) 7.0 ns. Structural elements are colour-coded as in Fig. 5. Figures were generated using VMD [45]. (For an interpretation of the references to colour in the artwork, the reader is referred to the web version of the article).

the unfolding of the Ubiquitin structure (Fig. 5) but is accompanied by its contraction with a r_G value of 1.17 ± 0.01 nm only (Table 4). Among the solvated reduced point charge models studied so far at 300 K, *mPASA* is characterized by the highest number of H-bonds occurring between both protein partners, with a mean value of 2.3 ± 1.4 (Table 3). For example, four H-bonds are identified for the solvated complex at 20 ns, *i.e.*, OG-HG(Ser277)···O(Ala46), NH2-HH22(Arg+275)···O(Thr66), NZ-HZ(Lys+6)···OE2(Glu−273), and NE2-HE2(His+68)···O(Glu−273) (Fig. 5), but these H-bonds are characterized by occupancy degrees lower than 30%. The behaviour of the complex modelled using model *mPASA* strongly differs from the other models. As said hereabove, more H-bonds appear for the solvated system, all with a low degree of occurrence. Contrarily, the mean numbers of H-bonds formed between the two partners and water are the lowest, with average values of 42.2 ± 4.8 and 116.3 ± 8.2 H-bonds, for UIM and the complex, respectively (Table 3). Such decreases are expected due to the absence of any dipole on the AA main chains. The lowest number of main chain-water H-bonds observed at 300 K, *i.e.*, 21.7 ± 4.1 comes along with the lowest number of side chain-water H-bonds, *i.e.*, 94.6 ± 6.7. Those H-bonds are characterized by geometries that differ from the conventional ones (Fig. 9). Indeed, the first occurrence peak in the distribution function is seen at 0.29 nm, a value that is slightly larger than for the other models, and no clear second maxima is observed below 0.5 nm. Additionally, no clear trends appear for the angle distribution function that shows a shoulder around 27° rather than a maximum at 10.5° as it was for the all-atom and other reduced point charge models. Moreover, the RDF of the water oxygen–protein atom pairs totally lacks the very first hydration peak, which was still partly visible for the other reduced point charge models (Fig. 8). A progressive reduction of the number of point charges thus leads to RDFs that come closer to the results a hypothetical uncharged model provides (MD simulation results not discussed in this paper) where the successive hydration shells fade away (Fig. 8). Intra-protein H-bonds are also the most dissimilar *versus* the all-atom and the other reduced point charge models (Fig. 8). The effect on the solvent is seen at the level of the self-diffusion coefficient D of the closest water molecules, with the highest observed value of $(2.46 \pm 0.08) \times 10^{-5}$ cm^2 s^{-1}, which can be correlated to the clear limitations of the protein model to form H-bonds. Energy contributions are also deviating strongly (Table 2), notably due to negative intra-molecular potential energies due to the cancelling of *Cb-14* energy terms (Table 5) in that particular implementation of the model (SI 4). One also notices a very low |*Cb*| term, *i.e.*, 2687.9 kJ mol^{-1}, *versus* all the other models.

In conclusion, model *mPASA*, regardless of its limits, still allows some preservation of the 3D folding when solvation is used. It is only at a very low temperature, *i.e.*, 150 K, that helices and extended strands appear to be stable (MD results not shown here). The interaction with the solvent is deeply perturbed, leading to a slight modification in the self-diffusion coefficient of the water molecules and a drastic change in their ability to form "conventional" H-bonds with the solute. The use of the H-bond concept might be revisited. In a study about combining all-atom and coarse-grained water and MARTINI models [48], the authors mention that charged and polar solutes in water still represent a major challenge [41]. Coarse-grained water models cannot represent H-bonding properly, and the authors suggest to possibly consider supplementary H-bonding energy terms as, for example, achieved in the coarse-grained potential PRIMO developed by Kar *et al.* [43]. This is not without any relation to the conclusion presented by Sinitskiy *et al.* regarding the waste a high resolution model can be when used to represent some parts of a biomolecular system if other parts are modelled too coarsely [27].

4.6. mCDh MD trajectories

The implementation of model *mCDh* in a GROMACS topology file was achieved to provide a protein representation as similar as possible to the model used previously with the program package TINKER [11]. Contrarily to the reduced models analysed so far, it is not based on the definition of virtual sites (see Section 3.1). As for the results obtained with TINKER for isolated proteins, that led to a rather good preservation of secondary structure elements and of the 3D fold of the simulated proteins, the analysis of GROMACS MD trajectories also showed that the secondary structure elements of the protein complex are preserved, even if shorter. However, this appears to be less true in water where some elements like β-strands and helices deteriorate with time (Figs. 4 and 5). As illustrated in Fig. 5, the final configuration of the solvated system is very different from the original PDB structure, and the peptide adopts a hairpin-like structure as already observed with model *mCDa* with a low average number of H-bonds formed between the two partners, *i.e.*, 2.0 ± 1.2 (Table 3). Such a particular configuration leads to distance map with a topology rather similar to the *All-atom-2* case, *i.e.*, with limited contact areas involving very few contacts shorter than 0.4 nm (Fig. 6). Particularly, Arg+ residues of Ubiquitin and Glu− residues of the UIM-1 stay located far apart (Fig. 5).

As models *mCD*, *mCDa*, and *mCDh* have the same charge description for the main chain, there is no large change in the number of H-bonds created between the solute main chains and water (Table 3). Their average numbers remain larger than the number of H-bonds observed with the all-atom models. Protein–water H-bonds are characterized by the same geometrical parameters as model *mCD*, *i.e.*, distance and angle distributions that stay close to the all-atom ones, while protein–protein H-bonds are geometrically similar as to the other reduced point charge models (Fig. 9).

On an energy point of view, model *mCDh*, like model *mPASA*, differs strongly from the all-atom and other reduced point charge models (Table 2 and SI 7). Intra-molecular Cb_1 *and* Cb_2 terms are more stabilizing than for the other CD-based models, with respective values of -56.6 ± 286.9 and -49.3 ± 0.7 kJ mol^{-1}

Please cite this article in press as: L. Leherte, D.P. Vercauteren, Evaluation of reduced point charge models of proteins through Molecular Dynamics simulations: Application to the Vps27 UIM-1–Ubiquitin complex, J. Mol. Graph. Model. (2013), http://dx.doi.org/10.1016/j.jmgm.2013.10.011

in solution, and 1.7 ± 23.9 and -58.5 ± 0.4 kJ mol^{-1} in vacuum. Contrarily, the inter-molecular E_{12} and Cb_{12} terms of the solvated system are destabilizing, for example with 0.13 ± 99.7 and 384.9 ± 18.4 kJ mol^{-1}, respectively.

As for *mCD* and *mCDa*, r_G of Ubiquitin is larger than the *All-atom* value, *i.e.*, one gets a mean value of 1.28 ± 0.01 rather than 1.20 ± 0.01 nm, but the gyration radius value obtained in vacuum is similar to the corresponding *All-atom* value, *i.e.*, 1.13 *versus* 1.15 ± 0.01 nm (Table 4). In water, one can additionally notice steps in the increase of r_G, due to the progressive loss of, first, the helix of Ubiquitin, followed by the disappearance of β-strands, occurring below 2.5 and at 10 ns, respectively (Fig. 4). MD trajectories generated with model *mCDh* suggest that, in water, the system is still evolving towards a different hyper-surface energy minimum. Thus, in addition to a lowering of the integration step that is required to carry out the MD simulations, the equilibration of the system seems very slow.

5. Conclusions and perspectives

Two reduced point charge models have been considered for Molecular Dynamics (MD) simulations of a protein complex, Vps27 UIM-1–Ubiquitin, using the program package GROMACS [15,16]. The first model, based on charges located at critical points (CP) of smoothed amino acid (AA) charge density (CD) distribution functions calculated from Amber99 atomic values, involves two point charges on the main chain of each AA, precisely located on atoms C and O, and up to six charges for the side chain. The second model, built by assigning charges to the maxima of AA smoothed promolecular electron density (ED) distribution functions, considers one point charge on the main chain and no more than two charges on the side chain.

For the first model, three different implementations were considered. In a first stage, the model is applied as is by considering charges as virtual sites in the system (model *mCD*). Second, rather than being located away from atom positions, most of the charges are set at selected atom positions. Their values are recalculated accordingly (model *mCDa*). Third, the charges are considered as additional masses attached to the system through harmonic bonds (model *mCDh*), as done in a previous work using the program package TINKER [11]. For the second model, only the first kind of implementation was considered (model *mPASA*).

MD simulations were carried out using the program GROMACS with the Amber99SB force field (FF), in water and in vacuum. The selected temperature was 300 K, except for model *mCD* where three lower temperature values, 277, 250, and 150 K, were also considered. The equilibration stages of *mCD*- and *mCDa*-based MD simulations were lengthened due to the increased ability of such models to sample various regions of the energy hyper-surface. Regarding the all-atom simulations, two starting configurations were selected: the PDB crystal structure (model *All-atom*) and a protein complex configuration obtained from a simulation using model *mCDa* (model *All-atom-2*). Energetic, structural, and dynamical information were retrieved from the analysis of the MD trajectories and discussed *versus* the *All-atom* model and available literature data. An emphasis was put on the secondary structure elements of the proteins, the conformation/configuration of UIM-1 *versus* Ubiquitin, and the characterization of H-bonds within the complex and with the solvent.

Regarding Ubiquitin, all three-dimensional (3D) folds remained rather similar, whatever the model used, during the simulations. However, the gyration radius, number and geometry of H-bonds, as well as the nature of the secondary structure elements varied. Vps27 UIM-1 was the most sensitive partner to the choice of the point charge model. Its conformation and orientation *versus* Ubiquitin were highly variable.

In vacuum, all models but the original *All-atom* one presented a better tendency than in water to preserve the secondary structure elements of the complex. In water, only the β-strands of Ubiquitin that are in closer contact with UIM-1 were always preserved. In both environments, model *mCD* led to the best fold description but strong deconstruction of secondary structure elements, while the inverse was observed for model *mCDa*. Model *mCD* is thus expected to provide long-range electrostatic interaction energy closer to the all-atom model, while model *mCDa*, which limits local conformational changes, is helpful to better preserve secondary structure elements of the proteins. Nevertheless, it can lead to a strong Vps27 UIM-1 deformation, such as a α-helix to hairpin transition, which however corresponds to an energetically probable conformation (model *All-atom-2*).

If the reduced point charge models do not favour the formation of a first hydration shell as clearly as with the all-atom model, they however allow the formation of solute-solvent H-bonds with geometrical properties similar to the all-atom case. Additionally, the large increase in the number of solute–solvent H-bonds is due to the C=O groups of atoms, except for model *mPASA* with only one charge of the AA main chain, while less side chain-water H-bonds are detected for all reduced point charge models. Intra-protein H-bonds are differently described with an angle distribution shifted towards higher angle values. In such aspects, the use of an all-atom description for the solvent molecules may still be meaningful.

MD simulations carried out with model *mCD* at various temperatures below 300 K led to the conclusion that this particular point charge model is able, at low temperature, to provide results that are essentially similar to the all-atom model. At 300 K, as results vary significantly from all-atom ones, one might find there a clue to conclude that, with reduced point charge models, energy barriers of the potential well are lowered, conformations can be perturbed more easily, but the location of that potential well on the energy hyper-surface is similar. In agreement with that conclusion, the deconstructed geometry obtained with model *mCDa*, and used as a starting point for an all-atom simulation, appeared to also probe a local energy well.

Model *mCDh* leads to the most time-consuming simulations as it involves the lowest time step value and seems to require longer equilibration stage.

Model *mPASA* led to the largest differences *versus* the all-atom model in terms of energetic, structural, and dynamical properties of the system. First, the implementation of the *mPASA* model is such as no *Cb-14* contributions to the potential energy are involved. Second, the number of point charges is too low to allow a first hydration shell as in the all-atom and in the other reduced point charge models. Geometrical parameters, that are associated with the existing H-bonds, adopt larger distance and angle values. The structure of the complex is particularly deconstructed in vacuum, contrarily to the trends followed by the other models. Even in water, Ubiquitin undergoes the more important contraction effect with the smallest gyration radius, while the other point charge reduced model systematically lead to an increase of the radius. Considering such facts, the use of an all-atom description for the solvent molecules together with a *mPASA* description for the solute appears to be meaningless.

Properties that can still be described using reduced point charge models are the overall 3D fold of Ubiquitin and, depending on the degree of point charge reduction, H-bond interactions with water molecules. On the contrary, local intra-molecular geometries are reproduced with less success, leading for some models, to a loss of most regular secondary structure elements. This could be further investigated in order to refine the model. Indeed, a good modelling of the *Cb-14* interactions seems to favour the preservation of

Please cite this article in press as: L. Leherte, D.P. Vercauteren, Evaluation of reduced point charge models of proteins through Molecular Dynamics simulations: Application to the Vps27 UIM-1–Ubiquitin complex, J. Mol. Graph. Model. (2013), http://dx.doi.org/10.1016/j.jmgm.2013.10.011

the secondary structure. Also, an adequate choice of the reference atoms that are associated with the virtual sites seems important. A new implementation of the models should follow those criteria. On the whole, locating point charges on molecular field extrema appears to be a sensible choice.

As reduced point charge models describe long-range electrostatic interactions rather efficiently, they are well suited to model rigid systems, as well as protein–protein interactions. To model protein systems, one may also imagine a combination of several levels of description, like in hybrid systems. Discussions have recently appeared on the subject [41,42]. The present models would be very easily implemented as they differ only by the number and location of point charges, without any change in atom types and sizes and FF formulae. Let us however mention that no trial was brought to these models to adapt the other FF parameters with the degree of point charge reduction.

Acknowledgments

The authors acknowledge the “Fonds National de la Recherche Scientifique” (FNRS-FRFC), the “Loterie Nationale” (convention no. 2.4578.02), and the University of Namur (UNamur) for the use of the Plateforme Technologique de Calcul Intensif (PTCI). They gratefully acknowledge F. Wautelet and L. Demelenne for program installation and maintenance.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jmgm.2013.10.011.

References

[1] L. Leherte, D.P. Vercauteren, Implementation of a protein reduced point charge model towards Molecular Dynamics applications, J. Phys. Chem. A 115 (2011) 12531–12543.
[2] S. Cranford, M.J. Buehler, Coarse-graining parametrization and multiscale simulation of hierarchical systems. Part I: theory and model formulation, in: P. Derosa, T. Cagin (Eds.), Multiscale Modeling: From Atoms to Devices, CRC Press, Boca Raton, FL, USA, 2010, pp. 13–34.
[3] V. Tozzini, Multiscale modeling of proteins, Acc. Chem. Res. 43 (2010) 220–230.
[4] M.G. Saunders, G.A. Voth, Coarse-graining of multiprotein assemblies, Curr. Opin. Struct. Biol. 22 (2012) 144–150.
[5] H. Shen, Z. Xia, G. Li, P. Ren, A review of physics-based coarse-grained potentials for the simulations of protein structure and dynamics, Annu. Rep. Comput. Chem. 8 (2012) 129–148.
[6] M.G. Saunders, G.A. Voth, Coarse-graining methods for computational biology, Annu. Rev. Biophys. 42 (2013) 73–93.
[7] A. Arnold, O. Lenz, S. Kesselheim, R. Weeber, F. Fahrenberger, D. Roehm, P. Košovan, C. Holm, ESPResSo 3.1 – Molecular Dynamics software for coarse-grained models, in: M. Griebel, M.A. Schweitzer (Eds.), Meshfree Methods for Partial Differential Equations VI, Lecture Notes in Computational Science and Engineering, vol. 89, Springer, Berlin, 2013, pp. 1–23.
[8] A. Mirzoev, A.P. Lyubartsev, MagiC: software package for multiscale modeling, J. Chem. Theory Comput. 9 (2013) 1512–1520.
[9] L. Leherte, D.P. Vercauteren, Coarse point charge models for proteins from smoothed molecular electrostatic potentials, J. Chem. Theory Comput. 5 (2009) 3279–3298.
[10] L. Leherte, D.P. Vercauteren, Charge density distributions derived from smoothed electrostatic potential functions: Design of protein reduced point charge models, J. Comput.-Aided Mol. Des. 25 (2011) 913–930.
[11] TINKER – Software Tools for Molecular Design, v. 5.0, http://dasher.wustl.edu/tinker/ (accessed 10.06.2013).
[12] J. Wang, P. Cieplak, P.A. Kollman, How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules, J. Comput. Chem. 21 (2000) 1999–2012.
[13] L. Amat, R. Carbó-Dorca, Molecular electronic density fitting using elementary Jacobi rotations under atomic shell approximation, J. Chem. Inf. Comput. Sci. 40 (2000) 1188–1198.
[14] L. Amat, R. Carbó-Dorca, Quantum similarity measures under atomic shell approximation: first order density fitting using elementary Jacobi rotations, J. Comput. Chem. 18 (1997) 2023–2039. Parameters are available at: http://iqc.udg.es/cat/similarity/ASA/funcset.html (accessed 24.06.2013).
[15] B. Hess, C. Kutzner, D. van der Spoel, E. Lindahl, GROMACS 4 algorithms for highly efficient, load-balanced, and scalable molecular simulation, J. Chem. Theory Comput. 4 (2008) 435–447.
[16] S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M.R. Shirts, J.C. Smith, P.M. Kasson, D. van der Spoel, B. Hess, E. Lindahl, GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit, Bioinformatics 29 (2013) 845–854.
[17] J. Kostrowicki, L. Piela, B.J. Cherayil, H.A. Scheraga, Performance of the diffusion equation method in searches for optimum structures of clusters of Lennard–Jones atoms, J. Phys. Chem. 95 (1991) 4113–4119.
[18] Y. Leung, J.-S. Zhang, Z.-B. Xu, Clustering by scale-space filtering, IEEE Trans. Pattern Anal. Mach. Intell. 22 (2000) 1396–1410.
[19] E. Spiga, D. Alemani, M.T. Degiacomi, M. Cascella, M. Dal Peraro, Electrostatic-consistent coarse-grained potentials for molecular simulations of proteins, J. Chem. Theory Comput. 9 (2013) 3515–3526.
[20] O. Borodin, G.D. Smith, Force Field Fitting Toolkit, The University of Utah, http://www.eng.utah.edu/~gdsmith/fff.html (accessed 26.08.2009).
[21] T.J. Dolinsky, J.E. Nielsen, J.A. McCammon, N.A. Baker, PDB2PQ.R. An automated pipeline for the setup of Poisson–Boltzmann electrostatics calculations, Nucleic Acids Res. 32 (2004) W665–W667.
[22] PDB2PQR, an automated pipeline for the setup, execution, and analysis of Poisson-Boltzmann electrostatics calculations, 2007, SourceForge Project Page, http://pdb2pqr.sourceforge.net/ (accessed 24.06.2013).
[23] U.C. Singh, P.A. Kollman, An approach to computing electrostatic charges for molecules, J. Comput. Chem. 5 (1984) 129–145.
[24] F. Eisenmenger, U.H.E. Hansmann, S. Hayryan, C.-K. Hu, An enhanced version of SMMP-open-source software package for simulation of proteins, Comput. Phys. Commun. 174 (2006) 422–429.
[25] Simple Molecular Mechanics for Proteins, http://developer.berlios.de/projects/smmp/ (accessed 24.06.2013).
[26] OpenDX, The Open Source Software Project Based on IBM's Visualization Data Explorer, Visualization and Imagery Solutions, Inc., http://www.opendx.org (accessed 20.08.2013).
[27] A.V. Sinitskiy, M.G. Saunders, G.A. Voth, Optimal number of coarse-grained sites in different components of large biomolecular complexes, J. Phys. Chem. B 116 (2012) 8363–8374.
[28] D.J. Heisterberg, Ohio Supercomputer Center, Translation from FORTRAN to C and input/output by J. Labanowski, Ohio Supercomputer Center, Labanowski, 1990 (technical report).
[29] CCL quaternion-mol-fit, 1999, Computational Chemistry List, Ltd, http://www.ccl.net/cca/software/SOURCES/C/quaternion-mol-fit/ (accessed 24.06.2013).
[30] K.A. Swanson, R.S. Kang, S.D. Stamenova, L. Hicke, I. Radhakrishnan, Solution structure of Vps27 UIM-ubiquitin complex important for endosomal sorting and receptor downregulation, EMBO J. 22 (2003) 4597–4606.
[31] J.H. Hurley, S. Lee, G. Prag, Ubiquitin-binding domains, Biochem. J. 399 (2006) 361–372.
[32] N.G. Sgourakis, M.M. Patel, A.E. Garcia, G.I. Makhatadze, S.A. McCallum, Conformational dynamics and structural plasticity play critical roles in the ubiquitin recognition of a UIM domain, J. Mol. Biol. 396 (2010) 1128–1144.
[33] JunGoo Jee, Unambiguous determination of intermolecular hydrogen bond of NMR structure by Molecular Dynamics refinement using all-atom force field and implicit solvent model, Bull. Korean Chem. Soc. 31 (2010) 2717–2720.
[34] Y.C. Kim, G. Hummer, Coarse-grained models for simulations of multiprotein complexes: application to ubiquitin binding, J. Mol. Biol. 375 (2008) 1416–1433.
[35] Y.C. Kim, R.B. Best, J. Mittal, Macromolecular crowding effects on protein–protein binding affinity and specificity, J. Chem. Phys. 133 (2010) 7, 205101.
[36] S.A. Showalter, R. Brüschweiler, Validation of Molecular Dynamics simulations of biomolecules using NMR spin relaxation as benchmarks: application to the AMBER99SB force field, J. Chem. Theory Comput. 3 (2007) 961–975.
[37] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, C. Simmerling, Comparison of multiple Amber force fields and development of improved protein backbone parameters, Proteins 65 (2006) 712–725.
[38] H.W. Horn, W.C. Swope, J.W. Pitera, J.D. Madura, T.J. Dick, G.L. Hura, T. Head-Gordon, Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew, J. Chem. Phys. 120 (2004) 9665–9678.
[39] K.R. Hadley, C. McCabe, Coarse-grained molecular models of water: a review, Mol. Simul. 38 (2012) 671–681.
[40] L. Darré, M.R. Machado, S. Pantano, Coarse-grained models of water, WIRES: Comput. Mol. Sci. 2 (2012) 921–930.
[41] T.A. Wassenaar, H.I. Ingólfsson, M. Priess, S.J. Marrink, L.V. Schäfer, Mixing MARTINI: electrostatic coupling in hybrid atomistic-coarse-grained biomolecular simulations, J. Phys. Chem. B 117 (2013) 3516–3530.
[42] K. Meier, A. Choutko, J. Dolenc, A.P. Eichenberger, S. Riniker, W.F. van Gunsteren, Multi-resolution simulation of biomolecular systems: a review of methodological issues, Angew. Chem. Int. Ed. 52 (2013) 2820–2834.
[43] P. Kar, S.M. Gopal, Y.-M. Cheng, A. Predeus, M. Feig, PRIM.O., A transferable coarse-grained force field for proteins, J. Chem. Theory Comput. 9 (2013) 3769–3788.
[44] N.A. Baker, D. Sept, S. Joseph, M.J. Holst, J.A. McCammon, Electrostatics of nanosystems: application to microtubules and the ribosome, Proc. Natl. Acad. Sci. U.S.A. 98 (2001) 10037–10041.

Please cite this article in press as: L. Leherte, D.P. Vercauteren, Evaluation of reduced point charge models of proteins through Molecular Dynamics simulations: Application to the Vps27 UIM-1–Ubiquitin complex, J. Mol. Graph. Model. (2013), http://dx.doi.org/10.1016/j.jmgm.2013.10.011

[45] W. Humphrey, A. Dalk, K. Schulten, VMD – Visual Molecular Dynamics, J. Mol. Graph. 14 (1996) 33–38.
[46] J.J. Virtanen, L. Makowski, T.R. Sosnick, K.F. Freed, Modeling the hydration layer around proteins: HyPred, Biophys. J. 99 (2010) 1–9.
[47] X. Chen, I. Weber, R.W. Harrison, Hydration water and bulk water in proteins have distinct properties in radial distributions calculated from 105 atomic resolution crystal structures, J. Phys. Chem. B 112 (2008) 12073–12080.
[48] S.J. Marrink, D.P. Tieleman, Perspective on the Martini model, Chem. Soc. Rev. (2013), http://dx.doi.org/10.1039/C3CS60093A.

Please cite this article in press as: L. Leherte, D.P. Vercauteren, Evaluation of reduced point charge models of proteins through Molecular Dynamics simulations: Application to the Vps27 UIM-1–Ubiquitin complex, J. Mol. Graph. Model. (2013), http://dx.doi.org/10.1016/j.jmgm.2013.10.011