



# LUND UNIVERSITY

## Conformational dependence of charges in protein simulations

Söderhjelm, Pär; Ryde, Ulf

*Published in:*  
Journal of Computational Chemistry

*DOI:*  
[10.1002/jcc.21097](https://doi.org/10.1002/jcc.21097)

2009

*Document Version:*  
Peer reviewed version (aka post-print)

[Link to publication](#)

*Citation for published version (APA):*  
Söderhjelm, P., & Ryde, U. (2009). Conformational dependence of charges in protein simulations. *Journal of Computational Chemistry*, 30(5), 750-760. <https://doi.org/10.1002/jcc.21097>

*Total number of authors:*  
2

*Creative Commons License:*  
Unspecified

### General rights

Unless other specific re-use rights are stated the following general rights apply:  
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

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

Read more about Creative commons licenses: <https://creativecommons.org/licenses/>

### Take down policy

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

LUND UNIVERSITY

PO Box 117  
221 00 Lund  
+46 46-222 00 00

**Conformational dependence of charges  
in protein simulations**

**Pär Söderhjelm & Ulf Ryde \***

Department of Theoretical Chemistry, Lund University, Chemical Centre, P. O. Box 124,  
SE-221 00 Lund, Sweden

Correspondence to Ulf Ryde, E-mail: [Ulf.Ryde@teokem.lu.se](mailto:Ulf.Ryde@teokem.lu.se),

Tel: +46 – 46 2224502, Fax: +46 – 46 2224543

2008-06-12

## **Abstract**

We have studied the conformational dependence of molecular mechanic atomic charges for proteins by calculating charges fitted to the quantum mechanical (QM) electrostatic potential (ESP) for all atoms in complexes between avidin and seven biotin analogues for 20 snapshots from molecular dynamics simulations. We have studied how various other charge sets reproduce those charges. The QM charges, even if averaged over all snapshots or all residues, in general have a larger magnitude than standard Amber charges, indicating that the restraint towards zero in the restrained ESP method is too strong. This has a significant influence on the electrostatic conformational energies and the interaction energy between the biotin ligand and the protein, giving a difference between the QM and Amber charges of 43 and 8 kJ/mol for the negatively charged and neutral biotin analogues, respectively (3–4 %). However, this energy difference is strongly reduced if the solvation energy (calculated by the Poisson–Boltzmann or Generalised Born methods) is added, viz. to 7 kJ/mol for charged and 3 kJ/mol for uncharged ligand. In fact, charges need to be recalculated with a QM method only for residues within 7 or 4 Å of the ligand, if the error should be less than 4 kJ/mol. Unfortunately, the QM charges do not give significantly better MM/PBSA estimates of ligand-binding affinities than standard Amber charges.

**Key Words:** electrostatic potential charges, molecular dynamics simulations, MM/PBSA, generalized Born, conformation dependence

## Introduction

During the latest decades, computational methods have established themselves as an important complement to experiments for the study of the structure and function of proteins and other macromolecules. They are typically performed by molecular mechanics (MM) methods, based on an empirical force field. One of the most important and critical parts of this force field is the treatment of electrostatic interaction. In most commonly used macromolecule force fields, these interactions are treated by partial point charges, one for each atom. Of course, the performance of the force field strongly depends on the quality of these point charges, which normally are determined by quantum mechanical (QM) methods [1,2]. A typical procedure is to fit the charges to the QM electronic potential (ESP), calculated at a large number of points around the molecule of interest [3,4,5].

A problem with the ESP charges is that they quite strongly depend on the conformation (geometry) of the molecule of interest [6,7,8,9,10]. Several methods have been suggested to reduce this conformational dependence, e.g. by fixing some charges [11], fitting to several conformations simultaneously [12,13], fitting to the ESP of one conformation with restraints to the dipole moment of another molecule [12], restraining the charges towards zero [13,14] (or to Mulliken or other less conformational-dependent charges [15,16]), or averaging the charges obtained for different conformations [17]. Alternative methods to determine geometry-dependent charges, not based on ESP fits, also exist [18,19]. However, the only method that has been extensively employed for general-purpose macromolecule force fields is the restrained ESP (RESP) method [13], which restrains the charges towards zero using a hyperbolic restraint. It is the standard method to obtain charges in the Amber force field, typically using also several conformations for the fit [20].

The conformational dependence of ESP charges has mostly been discussed in terms of the values of the charges. This information is of restricted interest, because it is well-known that large variations in the charges, especially for buried atoms, may give rise essentially equally good fits to the ESPs [1,2,6,14]. In some cases, the effect of the charges on various calculated energies has also been discussed. For example, it has been shown that interaction energies between glycerylphosphorylcholine (with a net charge of  $+1 e$ ) and individual water molecules in the first solvation shell may vary by up to 20 kJ/mol depending on what conformation was used in the charge fitting [7]. For a neutral sucrose molecule, the maximum error was slightly smaller, 13 kJ/mol, and the mean absolute deviation was 3 kJ/mol for the first-shell water molecules [14]. For the difference in free energy of solvation between ethanol and propanol, different charge sets may differ by up to 9 kJ/mol, depending on what conformation was used in the charge fit [8]. It is often observed that the effect of the charges is greatly reduced in water solution, compared to gas phase [21,22].

However, all these studies are restricted to simulations of a small molecule in water solution. Little is known about the effect of the conformational dependence of charges in simulations of macromolecules. Therefore, we have determined charges for all atoms in all amino acids in 20 snapshots from a molecular dynamics simulation of the protein avidin, binding seven different biotin analogues [23]. We study the variation of these  $\sim 1.1$  million unique charges, and their influence on the total electrostatic interaction energy, as well as the binding energy of the biotin analogues, both in vacuum and in water solution. Moreover, we study the distance dependence of the interaction energy and compare the charges with standard Amber charges [20], as well as charges obtained by averaging the ESP charges in various ways. This gives a detailed picture of the conformational dependence of ESP charges for proteins.

## Methods

### *The protein*

All calculations in this investigation are based on the X-ray structure of biotin complexed with avidin (PDB code 1avd [24]). Avidin is a tetramer, composed of four identical subunits. The complete tetramer was included explicitly in all calculations. When interaction energies between the ligand and the protein were studied, the calculations were either made on all four sites separately or only on one of the sites, in which case the other three sites were considered as a part of the protein (the four biotin sites are independent). The treatment of the protein was exactly the same as in our previous investigation [23]. In particular, all Asp and Glu residues were assumed to be negatively charged and all Lys and Arg residues were positively charged, whereas the single His residue in each subunit was protonated on the ND1 atom.

We considered biotin (Btn1) and six analogues (Btn2–Btn7); they are shown in Figure 1. The Btn2 and Btn5 biotin analogues include a guanidinium group that is positively charged in neutral aqueous solution. However, it has been shown that it is the neutral form of the ligand that binds to the protein [25,26]. Therefore, we simulated only the neutral form of these molecules (as in previous investigations [23,27]) and the corresponding experimental binding affinities were corrected for the fact that only the neutral form of the ligand binds to the protein [25,26]. The analogues were built into the active site as described before [23].

### *MD simulations and MM/PB(GB)SA calculations*

The molecular dynamics (MD) simulations were run with the Amber 8 software [28], as described before [23]. The full biotin tetramer with four identical ligands was solvated in an octahedral box, extending at least 10 Å outside the protein on all sides (~10845 water molecules, giving a total of ~40375 atoms). The protein was minimised, equilibrated with a constant pressure for 70 ps, and then equilibrated for an additional 200 ps with a constant volume, before coordinates were sampled every 10 ps for 200 ps. In the new simulations presented in this paper, only the final constant-volume equilibration was performed and the sampling was extended to 1700 ps.

The binding free energy of the biotin analogues to avidin was estimated by the MM/PBSA (molecular mechanics combined with Poisson–Boltzmann and surface area) method [29]. In this, the free energy of the ligand, the protein, and the complex are estimated as the sum of four terms:

$$G = \langle E_{MM} \rangle + \langle G_{Solv} \rangle + \langle G_{np} \rangle - T \langle S_{MM} \rangle \quad (1)$$

where  $G_{Solv}$  is the polar solvation energy of the molecule, estimated by the solution of the Poisson–Boltzmann (PB) equation [30] or by the default Generalised Born (GB) method in Amber 8, viz. GB<sup>OBC</sup> with  $\alpha$ ,  $\beta$ , and  $\gamma$  set to 1.0, 0.8, and 4.85, respectively [31] (i.e. strictly giving the MM/GBSA method),  $G_{np}$  is the non-polar solvation energy, estimated from the solvent-accessible surface area (SASA) of the molecule, according to  $G_{np} = 0.0227 \text{ SASA (in } \text{Å}^2) + 3.85 \text{ kJ/mol}$  [32],  $T$  is the temperature,  $S_{MM}$  is the entropy of the molecule, estimated from a normal-mode analysis of harmonic frequencies calculated at the molecular mechanics (MM) level, and  $E_{MM}$  is the MM energy of the molecule, i.e. the sum of the internal energy of the molecule and the electrostatics and van der Waals interactions. All the terms in Eqn. (1) are averages of energies obtained from a number of snapshots taken from MD simulations. In order to reduce the time-consumption and to obtain stable energies, the same geometry was used for all three reactants (complex, ligand, and protein) [27,33]. Thereby, the internal MM energy cancels out in the calculation of  $\Delta G_{bind}$ . The MM/PBSA calculations were run with the DelPhi II [34] (PB term) and Amber 8 [28] (all the other terms) software in the same way as in our previous investigation [23].

### *Charge sets*

In this paper, we will discuss the results of six different sets of charges, which are defined here: The first is the QM ESP charges, which were calculated for all atoms in the protein and the biotin analogue for 20 different snapshots, taken from seven 200-ps MD simulations (one for each analogue) of the biotin tetramer, based on the Amber 1994 force field [20] (the 94oh simulation in 23). Charges for all residues in the complex were calculated with the Hartree–Fock (HF) method and the 6-31G\* basis set with the Gaussian 03 software [35] (to be compatible with the Amber 1994 force field). For each snapshot, a separate set of charges was calculated. The protein was divided into dipeptides (i.e. each residue was capped by CH<sub>3</sub>CO– and –NHCH<sub>3</sub> groups) and the charges were calculated for these using the Merz–Kollman (MK) scheme [4]. The ESP charges on the capping groups were then discarded, whereas the charge on the CA atom was adapted so that the whole residue had the proper integer charge (more sophisticated methods exist for this, e.g with Lagrange multipliers [13], but they would have made the fit more complicated). The calculations took ~150 CPU-days in total for the 70 000 calculations of all residues in all complexes and snapshots (~3 CPU-minutes per residue on the average). This charge set will be called QM and consists of  $\sim 7830 * 7 * 20 = 1\,095\,760$  unique charges (there are 7708 atoms in the avidin tetramer and 12–41 atoms in the biotin analogues). These charges are considered to be the correct ones, because they are calculated for the right conformation of all atoms. The aim of this paper is to see how well other charge sets reproduce these charges and energies calculated from them.

At the other end of the spectrum we have the original Amber 1994 (Cornell et al.) charges [20]. In this set, which is called Amber below, there is one charge for each symmetrically distinct atom (hydrogen atoms bound to the same carbon, as well as other symmetry-related atoms have the same charges) in each amino acid. This sums up to 301 distinct charges for avidin. They were obtained by the RESP method [13], based on HF/6-31G\* ESPs, selected according to the MK scheme [4], and averaged over a number of conformations. The charges for the seven biotin analogues (208 charges) were obtained in a similar way [23], but only for a single conformation, taken from a HF/6-31G\*\* geometry optimisation started from the conformation in the protein.

Between these two endpoints, we have constructed four additional sets of charges. First, we simply took the QM charges from the first snapshot and used it for all the other snapshots. This set of charges is called QM1 and thus consists of  $\sim 7830 * 7 = 54\,788$  charges. Compared to the QM charges, the computer load to determine the QM1 charges is reduced by a factor of 20.

Next, we constructed a set of charges by simply averaging the charges of each protein atom over the 140 snapshots and each atom in the biotin analogues over the 20 snapshots. This gives  $7708 + 208 = 7916$  charges, but in this case, there is no saving of computer load; instead, we reduce the number of data and make the charges more stable. This set of charges is called Aver.

The number of charges can be further reduced by averaging all the charges of the same atoms in the same amino-acid residues in the Aver charge set. This gave  $388 + 208 = 596$  (protein+ligands) distinct charges (no attempt was made to make symmetry related atoms identical). This consensus set of charges is called Cons below.

Finally, we can use the same procedure on the QM1 charges to obtain a set of  $\sim 30+388$  for each of the ligands (giving a total of  $7*388 + 208 = 2924$  distinct charges). This set of charges is called Cons1. The six sets of charges are summarised in Table 1.

## Result and Discussion

### *Variation of the charges*

We have determined QM Merz–Kollman ESP charges [4] for all atoms in avidin and the seven biotin analogues for the actual conformation obtained in 20 snapshots from MD simulations in an explicit water solvent. These charges were originally used in MM/PBSA estimations of the binding affinity of the seven biotin analogues to avidin, but no improvement in the estimated binding affinities were observed compared to standard Amber charges [23]. In this paper, we analyse the charges and their conformational dependence and compare them to other sets of charges.

First, we simply compared the QM charges for all the 7708 atoms in the protein obtained from the  $20 * 7 = 140$  snapshots. The atom that had the largest charge was the CZ atom of various Arg residues (A88, A115, C115, D27, D112, or D115; the letter indicates the subunit in the avidin tetramer), with a charge of 1.31–1.43  $e$ . The same atom has also the largest charge in the Amber force field, 0.83  $e$ . On the other hand, the atom with the most negative charge varied more. In the simulations of two of the biotin analogues, it was the ND2 atom of Asn-A58 or Asn B89 (–1.27 and –1.23  $e$ ) and in one simulation it was the NH2 atom of Arg-A88. These atoms have large negative charges also in the Amber force field, –0.92 and –0.86  $e$ , but the one with the most negative charge is actually NE2 of Gln (–0.94  $e$ ). However, in the simulations with the other four biotin analogues, the atom with the most negative charge was actually CA of Ile-C42 or C104 (–1.58 to –1.38  $e$ ), which has a charge of only +0.09  $e$  in the Amber force field. This indicates that there are some problems with poorly determined charges [1,2,6,13,14] when we use the MK scheme [4], instead of the Amber RESP procedure [13]. However, this should not be a problem as long as the charges are only employed for the same conformation for which they were calculated: The MK charges reproduce the QM ESP as well as possible (in the least-squares sense, even if the charges may seem to have a too large magnitude), whereas the restraints towards zero in the RESP charges will make the fit to the QM ESP slightly worse [14].

The same is seen if we look at the variation of the charges. The largest ranges (1.38–1.83  $e$ ) are found for CA atoms of various residues (Arg-27, Lys-59, Lys-72, Glu-75, Met-97, and Ile-107). This shows that the CA atoms are poorest determined by our method (probably because we have capped the residues with  $\text{CH}_3\text{-CO-NH-CHR-CO-NH-CH}_3$ , around the CA atom). The atoms that show the smallest variation are the HG atom of Ser-A71 or B103, or the HE1, HH2, or HZ2 atoms of Trp-71, 103, and 111. All these atoms show an extremely small range of 0.03  $e$  in the simulations of one biotin analogue, and slightly more (0.05  $e$ ) if all the seven simulations are taken together.

If we compare the charges for the same atoms in the same type of amino acid anywhere in the sequence (i.e. for the 388 unique protein atoms in the Cons charge set), we obtain similar results. The largest range is found for CA in Ile (2.15  $e$ ) and the smallest range is found for HH in Tyr (0.10  $e$ ). In general, the range is smallest for the hydrogen atoms (~0.1  $e$  for aromatic hydrogen atoms 0.10–0.25  $e$  for polar hydrogen atoms, 0.18–0.37  $e$  for H, and 0.19–0.59  $e$  for non-polar hydrogen atoms), followed by the oxygen atoms (0.11–0.24  $e$  for unprotonated and 0.17–0.29 for protonated oxygen). The two sulphur atoms in Cys (always cross-linked cystine) and Met have ranges of 0.25 and 0.29  $e$ . Nitrogen atoms (0.19  $e$  for NE1 in His and 0.37–1.1  $e$  for the other protonated nitrogen atoms) and carbon atoms have the largest ranges (0.36–0.82  $e$  for aromatic carbons, 0.38–0.64  $e$  for carbonyl carbons, 0.41–0.83  $e$  for C, 0.58–1.5  $e$  for aliphatic carbons, and 0.71–2.2  $e$  for CA). Thus, the ranges follow quite closely how hidden each atom is (the number of valences of the atoms).

Consequently, there is a clear indication that buried charges give strongly varying values, which sometimes are quite large. This may explain why we got problems when we used the QM charges in MD simulations [23]. A simple way to avoid this problem is to average each

charge over the 140 snapshots in the seven trajectories to obtain the Aver charge set. This gave much more stable results: The largest positive Aver charge was 1.10  $e$  for CZ in Arg-A115 and the most negative one was  $-1.09 e$  for ND2 of Asn-A89, exactly as in the Amber force field. The largest absolute charge on a CA atom was  $-0.65 e$  for Ile-D107. Thus, averaging over trajectories is an effective way to suppress poorly determined charge, as also has been suggested before [17].

Figure 2 shows the variation of the Aver protein charges, collected after the atom name and residue type, and colour coded after the element. It can be seen that there still is some variation in the individual charges, especially for the hidden carbon and nitrogen atoms (with four neighbouring atoms). To decide how much of this variation is actually caused by the conformational dependence of the charges, we performed a Student  $t$  test for the mean values (over the 140 snapshots and 4 subunits) of each pair of identical atoms in different residues (assuming normal distribution and an equal standard deviation). In fact, it turned out that 71% of the 7478 pairs of atoms had significantly different averages at the 99.9% level, showing that the conformational dependence is significant.

In Figure 2, we have also included the standard Amber charge for each atom. It can be seen that they in general are smaller in magnitude than the QM charges, as can be expected from the restraint towards zero in the RESP procedure. However, it can be seen that they almost always are outside the range of the corresponding average QM charges, also for the quite well determined oxygen and hydrogen atoms. It can be questioned if this is really desirable: The RESP procedure is a way to take into account the conformational dependence in an average way. However, the present results show that this is not really the case.

By averaging the charges over all residues and snapshots, we constructed the Cons charge set. This represents an alternative method to obtain standard charges for each residue, viz. to include an extensive configurational sampling, rather than introducing an artificial restraint for the charges towards zero. However, it should be noted that the sampling is not even, because the number of the residues varies from two for the amino terminals (subunits A and C start with Lys-3, whereas subunits B and D start with Arg-2) to 80 for Thr. In the Cons charge set, the largest positive charge is 1.03  $e$  for CZ of Arg, as in the Amber force field, although the charge is 0.22  $e$  more positive. The most negative charge is  $-1.06 e$  for NE2 of Gln, again exactly as in the Amber force field, although the charge is 0.12  $e$  more negative. Thus, the averaging leading to the Cons charges has automatically suppressed the large charges. For example, the Cons charges on the CA atoms range from  $-0.34$  to  $+0.41 e$  (Ile to Cys), whereas those in Amber range from  $-0.24 e$  (Lys) to  $+0.04 e$  (Cys).

Figure 3 shows the correlation between the Amber and the Cons charges for each atom. It can be seen that there is a good correlation between the two sets of charges ( $r^2 = 0.91$ ), although the Cons charges are typically slightly larger in magnitude than the Amber charges. However, there are two pronounced outliers, viz. the N atoms of the amino terminal Lys and Arg residues, which have a charge of  $-0.73$  and  $-0.68 e$  in our calculations, but  $+0.10$  and  $+0.13 e$  in the Amber force field (the Amber charge of the similar side-chain NZ atom in Lys is  $-0.39 e$ ). This indicates that these Amber charges are questionable.

If these two outliers are ignored, the average difference between the Cons and Amber charges is 0.09  $e$  with a standard deviation is 0.08  $e$ . The maximum difference is 0.37  $e$  for CA in Cys, and an additional seven atoms have differences above 0.3  $e$  (N in Cys, N and CA in Lys, N and CA in Arg, as well as CG in Glu and Gln). More experience is needed before it can be conclusively decided whether the Amber charges should be improved.

The Cons charges do not vary much if they are instead determined for each of the seven simulations separately: The largest difference is 0.16  $e$ , but it is found for the amino- and carboxy-terminals and the single His residue, for which the statistics is poorer than for the other residues. If these residues are excluded, the largest difference is only 0.05  $e$  (for CB in Met) calculations. This shows that the variation among the seven simulations is almost entirely statistical.



## *Energies*

In principle, the individual charges are not very interesting, because it is well-known that rather different charges may give similar fits to the QM ESPs [1,2,14]. Therefore, we next looked at various energies obtained from the charges. The total electrostatic energy within the avidin tetramer (excluding the ligand) is widely different for the Amber charges (−66 MJ/mol) and the other charge sets (−102 to −104 MJ/mol; the interaction energy was calculated without any cut-off, but excluding 1,2 and 1,3 interactions, and scaling the 1,4 interactions with a factor of 1.2, as in the standard Amber force field [20]). However, this difference is caused mainly by a constant offset that is irrelevant for actual energies (only relative MM energies make a physical sense), even if it may be relevant for large conformational changes, e.g. between unfolded and folded forms of a protein.

If we instead compare the total interaction energies for the whole protein in each snapshot relative to the average energy over all snapshots, the difference between the various charge sets is much reduced. From Figure 4, it can be seen that the five charge sets give roughly parallel relative energy curves (note that we can no longer compare with the QM charges, which are correct only for one snapshot). However, there are still differences of up to 150 kJ/mol between the various charge sets, which is actually 17 % of the total variation in the electrostatic interaction energies over the 140 snapshots (906 kJ/mol for the Amber charges). This is a very large energy difference for conformational energies and therefore can be expected to have a significant influence on the ensemble of sampled structures, assuming that the remaining energy terms of the force field are not changed.

On the other hand, the Cons and Cons1 charge sets are very similar to each other (the MAD is only 4 kJ/mol) and also to the Aver charges (MADs of 10 kJ/mol), showing that the averaging makes the charges very similar. However, the QM1 charges give an appreciably larger MAD, e.g. 18 kJ/mol compared to the Aver charges, with the largest deviation (79 kJ/mol) for the snapshot for which the charges were determined (the first one).

## *Ligand-interaction energies*

Still more interesting are the interaction energies between the ligand and the surrounding protein. Table 2 shows how the electrostatic interaction energy between the ligand and the surrounding protein differs between the various charge sets. For example, for the 20 snapshots of the simulation of Btn1, the electrostatic interaction energies from the QM charges (different for each snapshot) and Amber charges differ by 19–74 kJ/mol (46 kJ/mol on average). It is notable that the Amber charges always give a more negative interaction energy than the QM charges (although the latter normally have larger absolute values).

If we concentrate on the snapshot that gave the largest deviation, the five largest contributions to this difference come from residues Thr-40 (21 kJ/mol with a closest distance to biotin of 2.2 Å), Ala-39 (12 kJ/mol, 1.8 Å), Tyr-33 (−10 kJ/mol, 1.7 Å), Lys-71 (7 kJ/mol, 8.0 Å), and Ser-102 (6 kJ/mol, 8.0 Å). Thus, the largest differences (and also the largest interaction energies) come from residues close to biotin. This is clearly illustrated in Figure 5, which shows that over 80% of the total energy difference is obtained already when all residues with a shortest distance of 8 Å from the ligand are considered. However, the convergence is rather slow, and the total energy difference does not converge to within 4 kJ/mol until all residues within 29 Å of the ligand are included and even for the most distant residue in the protein (Thr-C125 at a distance of 43 Å), the difference in the interaction energies between the QM and Amber charges is still 0.2 kJ/mol.

For the Cons charges, the difference to the QM charges is similar (18–69 kJ/mol) and they still always give more negative electrostatic interaction energies. However, for the Aver charges, the difference to the QM charges is appreciably smaller (0–52 kJ/mol, average 25 kJ/mol), and one snapshot gives a more positive interaction energy. The convergence of these charges is also better: The cumulative difference between the Aver and QM charges is less than

4 kJ/mol for all residues outside 14 Å from Btn1 on the average, and 25 Å in the worst case. The Cons charges have a slightly worse convergence (average and maximum distance for convergence is 18 and 26 Å, respectively).

The results are similar for the other two charged biotin analogues (Btn2 and Btn3): The average difference to the Amber and Cons charges is 39–43 kJ/mol, but only 23–25 kJ/mol for the Aver charges. The Aver and Cons charges are converged at 12–16 Å (21–23 Å in the worst cases).

For the four neutral biotin analogues, the results are quite different: The total electrostatic interaction energies are almost an order of magnitude smaller (e.g. 105–172 kJ/mol compared to 1209–1330 kJ/mol for the QM charges). Consequently, the average difference from the QM charges is also smaller, 4–14 kJ/mol for all the charge sets (Table 2). Owing to the small magnitude of the difference, the charges also converge much more rapidly and they are converged (within 4 kJ/mol) when all residues within 8 Å of the ligand are considered for the Amber charges and within 6 Å for the Cons and Aver charges. Thus, we can conclude that the conformational dependence of the charges is crucial only for interaction energies involving charged ligands (although the maximum difference is still 25–27 kJ/mol for the neutral ligands and the Amber, Cons, and Aver charges).

Finally, we have also tested how well one set of QM charges (QM1) performs compared to the right QM charges. Such a procedure actually gives as good results as the Aver charges: the average MAD for the charged and neutral biotin analogues are 16 and 8 kJ/mol, respectively, and with convergence after 15 and 3 Å on the average. Likewise, the Cons1 charges give results that are similar to those of the Cons charges.

### *Solvent-screened interaction energies*

It has been observed that solvation effects decrease the conformational dependence of charges [21,22] and that the solvent tends to decrease the effective electrostatic interaction between the protein and the ligand. In fact, it is normally argued that the protein on the average behaves like a continuum solvent with an effective dielectric constant of 2–20 [36,37], and for the interaction with solvent-exposed charged residues, even larger effective dielectric constants are frequently used [38]. Therefore, we have studied also solvent-screened ligand–protein electrostatic interaction energies.

Unfortunately, it is not fully straight forward to estimate the effect of solvent screening of various charge sets in an effective and informative way. For example, explicit solvent energies converge very slowly. Therefore, we have instead employed a simplified approach, commonly used in theoretical estimates of ligand-binding affinities, viz. the MM/GBSA approach [29], in which the electrostatic solvation effects are estimated by the generalised Born (GB) implicit solvent model [31] (the Poisson–Boltzmann approach gave similar, but less smooth results). Thus, we have performed standard MM/GBSA calculations for the seven biotin analogues with either the correct QM charges for each snapshot or with the other charge sets. To investigate the distance dependence of the effect, we have also used mixed charge sets, where the charges of all residues outside a certain distance from the ligand (we used 16 different distances, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 30, 35, and 40 Å for the charged biotin analogues and 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 15, 20,25, 30, 35, and 40 Å for the others) are changed to the approximate charges.

In Table 3, we show how much the solvent-screened protein–ligand interaction energies differ when calculated by the five charge sets, compared to the QM charges. It can be seen that the difference is strongly reduced: For the charged ligands, inclusion of the solvation energy strongly reduces the MAD in the protein–ligand interaction energy to only 6–8 kJ/mol. For the neutral ligands, the MAD is reduced to 2–4 kJ/mol, with little difference between the five charge sets.

Table 3 also shows the average and maximum distance at which the interaction energy for a

set of charges agrees with that obtained with the QM charges to within 4 kJ/mol. It can be seen that even for the charged ligands, the Amber charges converge on the average after 3 Å with GB solvation (29 Å without the solvation). This clearly illustrates the extensive screening effect of solvation. In fact, even the worst snapshot (Md in Table 3), has converged within 7 Å for the Amber charges. For the neutral ligands, the corresponding average and maximum values are 2 and 6 Å without, and 1 and 4 Å with GB solvation. The other four charge sets show very similar convergence for the solvent-screened interactions. This suggests that once the electrostatics interactions are screened, the sensitivity to the charges is strongly reduced. Of course, this is a great advantage in real applications.

#### *MD simulations and MM/PBSA energies*

Finally, we tested whether the various charge sets could be used in MD simulations for new MM/PBSA calculations. The motivation for this is that we observed in our original investigation [23] that MD simulations based on the QM1 charges crashed with SHAKE problems, most likely originating from too strong electrostatic interactions when the ESP charges were allowed to move (showing that the charges are not fully compatible with the corresponding van der Waals parameters). Even without SHAKE, very large interaction energies and unreliable structures were obtained.

Therefore, we also ran MD simulations based on the other three charge sets (Aver, Cons, and Cons1). Quite satisfactorily, no problems were encountered with any of those calculations, showing that it was the charges that caused the problems and that they are solved by any sort of averaging, which reduce the magnitude of the buried charges.

Next, we ran a 200 ps equilibration and 1700 ps data sampling for each charge set and biotin analogue, and performed a MM/PBSA calculation based on those 171 snapshots. The results of these calculations are collected in Table 4. It can be seen that the results for these three charge sets are similar to those of the original Amber simulation: The MADs are 12, 21, and 21 kJ/mol for the Aver, Cons, and Cons1 charges, respectively, whereas the original Amber result was 14 kJ/mol. It is unlikely that these differences are significant, considering that the standard deviation of the estimated binding affinities is 5–15 kJ/mol [23]. The results are similar to those obtained with other reasonable simulation methods and force fields (MADs of 7–19 kJ/mol) [23]. If a different charge set is used in the MM/PBSA energy calculations compared to the MD simulation, the results are slightly worse, with two exceptions, as was also observed before [23]. In particular, energies calculated with the QM charges on the Amber MD snapshots gave a MAD of 20 kJ/mol [23]. Thus, there is no clear-cut improvement in the MM/PBSA method of using conformation-dependent charges, probably because the accuracy is limited by other factors and the standard deviation is too high.

#### **Conclusions**

We have studied the conformational dependence of charges in standard MM calculations by calculating QM MK charges for all atoms in 20 snapshots of seven biotin–avidin complexes from MD calculations. We have compared these charges with those in the standard Amber 1994 force field (which uses the same QM method to obtain charges [20]) and various averages of the QM charges. In particular, we have concentrated on how large influence the conformational dependence has on ligand-binding energies, both with and without water solvation. This has provided much useful information.

- The conformational dependence of the charges is significant (i.e. the average charges on the same type of atom at different places in the sequence are significantly different).
- Amber charges of the amino-terminal N atoms (at least of Lys and Arg) seem to be erroneous (they are around +0.1  $e$  in Amber, but around -0.7  $e$  for the averaged QM charges).

- The Amber RESP procedure restrains charges stronger towards zero than simple averaging of the charges, i.e. the Amber charges are most often outside the range of the QM charges.
- The relative energies in a MD simulation, i.e. the conformational energies of the protein, depend significantly on the charge set employed (up to 150 kJ/mol or 17 % of the total variation in electrostatic energy. This should have a noticeable effect on the generated ensemble of structures.
- Standard Amber charges give electrostatic ligand-binding energies that differ from those obtained by QM charges by 43 and 8 kJ/mol on average for negatively charged and neutral ligands, respectively (3–4 % of the total energy).
- Fortunately, these differences are to a great extent compensated by solvation effects. For the sum of the electrostatic and solvation interaction energy, the difference between the Amber and QM charges is reduced to 7 and 3 kJ/mol, respectively.
- Averaging of QM charges over snapshots provides an effective way to solve the buried-atom problem and obtain stable energies that are still close to the QM results. However, as expected from the significant conformational dependency of the charges, the averaging should preferably be done over the set of conformations that is dynamically available for each particular residue (Aver) and not include other occurrences of the same amino acid in the protein (Cons; cf. Table 2).
- On the other hand, there are many instances in which a consensus set of charges (i.e. with the same charge for all occurrences of the same amino acids) is more useful. We have shown that such set can be generated from the QM charges by averaging out the conformational dependence (as has been pointed out before [17]). The resulting energies do not diverge too much from the results with QM charges (especially not in solvent), and in variance to QM charges, the consensus charges are useful in MD simulations. We provide in the supplementary material (also available at <http://www.teokem.lu.se/~ulf/cons.in> in Amber input format) the first such set of charges for all amino acids (but only cystine and Hid), obtained by averaging over 140 snapshots and 2–80 different residues in the protein (i.e. the Cons charges). We invite the reader to test this seemingly less biased and more accurate set of atomic charges. In future investigations, we will determine similar charges also for the missing Cys, Hie, and Hip residues and examine if the charges are biased by the fact that they are determined for only a single protein.

Thus, the conformational dependence has a significant effect on conformational energies and ligand-binding energies, even in real, solvent-screened, situations. If the desired accuracy is better than ~7 kJ/mol, as is normally the case in studies of ligand-binding affinities (6 kJ/mol corresponds to one order of magnitude in dissociation constants), these effects need to be considered. However, it is not necessary to calculate the charges for all atoms in the protein in all snapshots: Our investigations of the distance dependence of the interaction energies in Tables 2–3 show that if the desired accuracy is 4 kJ/mol, we only need to recalculate the charges in residues within 7 Å of a charged ligand (~40 residues), and within 4 Å of a neutral ligand (~20 residues) in the worst case. The rest of the protein can be treated by the Amber force field. At the HF/6-31G\* level, this is not a major effort – one snapshot of the whole protein typically takes 30 CPU hours and each residue takes 3 minutes on average (and the calculations can be run in parallel). There does not seem to be any significant advantage of using QM energies obtained from a single snapshot (QM1 or Cons1) instead of the Amber charges.

Unfortunately, the estimated ligand-binding affinities are not significantly improved compared to experiments by the recalculation of the charges, at least not at the MM/PBSA level. This most likely indicates that the limiting factor of that method is not the quality of the charges, but rather the other terms, e.g. the entropy, solvation, non-polar solvation, van der Waals interactions, or even the approximations of the approach, e.g. the use of the same

geometry for the complex, receptor, and ligand, the ignorance of polarisation, or a too restricted sampling of snapshots. Still, it is clear that the conformational dependence of the MM charges are significant and important problem, and needs to be treated properly if really accurate results are to be obtained.

Of course, it can be questioned if HF/6-31G\* is a proper level to obtain accurate results. We have employed this level to obtain charges that are compatible with the Amber charges, but it may well be that this approach (especially the neglect of explicit polarisation) is a more severe approximation than the neglect of conformational dependence of the charges. However, we expect that our results are applicable also for better QM methods. For example, a full multicentre–multipole expansion plus anisotropic polarisabilities can be calculated for a whole protein at the B3LYP/aug-cc-pVTZ level in about 20 CPU weeks (one snapshot) [39], which is not prohibitive with today's computer resources.

Finally, we want to point out that the present results are not restricted to ligand-binding affinities. Similar results are most likely also obtained if the catalytic mechanism of enzymes is studied, for example (in that case, the active site takes the place of the ligand), so the results have strong bearings on all types of accurate calculations on proteins.

### Acknowledgements

This investigation has been supported by AstraZeneca, the Crafoord foundation, and grants from the Swedish research council. It has also been supported by computer resources of Lunarc at Lund University.

### References

- 1 Bachrach, S. M. *Rev. Comp. Chem.*, 1994, 5, 171.
- 2 Sigfridsson, E.; Ryde, U. *J. Comp. Chem.* 1998, 19, 377-395.
- 3 Chirlain, L. E.; Francl, M. M. *J. Comp. Chem.*, 6, 894 (1987).
- 4 Besler, B. H.; Merz, K. M.; Kollman, P. A. *J. Comput. Chem.* 11, 431 (1990).
- 5 Breneman, C. M.; Wiberg, K. B. *J. Comp. Chem.*, 11, 361 (1990).
- 6 Williams, D. E. *Biopol.* 1990, 29, 1367-1386.
- 7 Stouch, T. R.; Williams, D. E. *J. Comput. Chem.* 1992, 13, 622-632.
- 8 Reynolds, C. A.; Essex, J. W.; Richards, W. G. *Chem. Phys. Lett.* 1992, 199, 257-260
- 9 Shirts, R. B.; Stolworthy, L. D. *J. Inclusion Phenom. Mol. Recogn.* 1995, 20, 297-321.
- 10 Pacios, L. F.; Gómez, P. C. *J. Mol. Struct. (Theochem)* 2001, 544, 237-251.
- 11 Stouch, T. R.; Williams, D. E. *J. Comput. Chem.* 1993, 14, 858-866.
- 12 Reynolds, C. A.; Essex, J. W.; Richards, W. G. *J. Am. Chem. Soc.* 1992, 114, 9075-9079.
- 13 Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* 1993, 97, 10269-10280.
- 14 Sigfridsson, E.; Ryde, U. *J. Comp. Chem.*, 2002, 23, 351-364.
- 15 Lévy, B.; Enescu, M. *J. Mol. Struct. (Theochem)* 1998, 432, 235.
- 16 Berente, I.; Czinki, E.; Náray-Szabó, G. *J. Comput. Chem.* 2007, 28, 1936-1942.
- 17 Basma, M.; Sundara, S.; Calgan, D.; Vernali, T.; Woods, R. J. *J. Comput. Chem.* 2001, 22, 1125-1137.
- 18 Rappé, A. K.; Goddard, W. A. *J. Phys. Chem.* 1991, 95, 3358.
- 19 Dinur, U.; Hagler, A. T. *J. Comput. Chem.* 1995, 16, 154-170.
- 20 Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollman, P. A. *J. Am. Chem. Soc.* 1993, 115, 9620-9631.
- 21 Urban, J. J.; Famini, G. R. *J. Comput. Chem.* 1993, 14, 353-362.
- 22 Resat, H.; Maye, P. V.; Mezei, M. *Biopol.* 1997, 41, 73-81.
- 23 Weis, A.; Katebzadeh, K.; Söderhjelm, P.; Nilsson, I.; Ryde, U. *J. Med. Chem.* 2006, 49,

- 
- 6596-6606.
- 24 L. Pugliese, A. Coda, M. Malcovati, M. Bolognesi, 3-Dimensional Structure of the Tetragonal Crystal Form of Egg-White Avidin in Its Functional Complex With Biotin At 2.7 Å Resolution, *J. Mol. Biol.* 1993, 231, 698-710.
- 25 Green, N. M. Thermodynamics of the binding of biotin and some analogues by avidin. *Biochem. J.* **1966**, 101, 774-780.
- 26 N. M. Green, Avidin, *Adv. Protein Chem.* 1975, 29, 85-133.
- 27 B. Kuhn, P. A. Kollman, Binding of a Diverse Set of Ligands to Avidin and streptavidin: An Accurate Quantitative Prediction of Their Relative Affinities by a Combination of Molecular Mechanics and Continuum Solvent Models, *J. Med. Chem.* 2000, 43, 3786-3791.
- 28 D. A. Case, T. A. Darden, T. E. Cheatham, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, K. M. Merz, B. Wang, D. A. Pearlman, M. Crowley, S. Brozell, V. Tsui, H. Gohlke, J. Mongan, V. Hornak, G. Cui, P. Beroza, C. Schafmeister, J. W. Caldwell, W. S. Ross, and P. A. Kollman (2004) AMBER 8, University of California, San Francisco.
- 29 P. A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D. A. Case, T. E. Cheatham, Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. *Acc. Chem. Res.* 2000, 33(12), 889-897.
- 30 Gilson, M. K; Honig, B. Calculations of the total electrostatic energy of a macromolecular system: Solvation energies, binding energies, and conformational analysis. *Proteins, Struct. Funct. Gen.* **1998**, 4, 7-18.
- 31 A. Onufriev, D. Bashford, D. A. Case, Exploring protein native states and large-scale conformational changes with a modified generalized Born model, *Proteins* 2004, 55, 383-394.
- 32 Hermann, R. B. Theory of hydrophobic bonding. II. Correlation of hydrocarbon solubility in water with solvent cavity surface area. *J. Phys. Chem.* 1972, 76, 2754-2759.
- 33 Swanson, J. M. J.; Henchman, R. H.; McCammon, J. A. Revisiting free energy calculations: A theoretical connection to MM/PBSA and direct calculation of the association free energy. *Biophys. J.* **2004**, 86, 67-74.
- 34 W. Rocchia, E. Alexov, B. Honig.; Extending the Applicability of the Nonlinear Poisson-Boltzmann Equation: Multiple Dielectric Constants and Multivalent Ions, *J. Phys. Chem. B* 2001, 105:6507-6514.
- 35 Gaussian 03, Revision, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; and Pople, J. A.; Gaussian, Inc., Wallingford CT, 2004.
- 36 Sharp K. A. *Annu. Rev. Biophys. Biophys. Chem.* 1990, 19, 301-332.
- 37 Honig B. *Science* 1995, 268, 1144-1149.
- 38 Schutz, C. N.; Warshel, A. *Proteins, Struct. Funct. Gen.* 2001, 44, 400-417.
- 39 Söderhjelm, P.; Husberg, C.; A. Strambi; M. Olivucci; Ryde, U., manuscript.

**Table 1.** Description of the six charge sets discussed in this paper.

Charge set	# Distinct charges	Description	Are the charges different for		
			snapshots	Btn simulations	same residue
QM	1095760	QM ESP (MK) charges	yes	yes	yes
QM1	54788	QM charges from first snapshot	no	yes	yes
Aver	7916	Average over snapshots	no	yes	yes
Cons1	2924	QM1, averaged over residues	no	yes	no
Cons	596	Aver, averaged over residues	no	no	no
Amber	509	Amber FF94 charges	no	no	no

**Table 2.** Mean absolute, minimum, and maximum difference (MAD, Min, and Max) between the various charge sets and the QM charges for ligand–protein electrostatic interaction energy in the 20 snapshots of the simulations of the seven biotin analogues (a negative sign indicates that that charge set gives more negative interaction energies than the QM charges). In addition, the average (Av) and maximum (Md) distance at which the residue-wise cumulative energy difference is converged to within 4 kJ/mol are given. The three last lines give the averages, maximum, or minimum over the three negatively charged biotin analogues (1–3), the four neutral biotin analogues (4–7), and over all seven biotin analogues (1–7).

	Amber					Cons					Cons1					Aver					QM1				
	MAD	Min	Max	Av	Md	MAD	Min	Max	Av	Md	MAD	Min	Max	Av	Md	MAD	Min	Max	Av	Md	MAD	Min	Max	Av	Md
Btn1	46.4	-74.3	-18.8	27.3	29	44.3	-68.6	-18.5	18.2	26	48.5	-73.3	-22.4	16.3	25	25.4	-51.8	0.6	13.5	25	15.5	-40.8	5.8	13.2	24
Btn2	43.1	-88.6	-5.9	26.4	28	40.9	-88.1	-3.8	12.9	23	41.5	-89.0	-3.3	13.6	23	25.2	-68.6	14.6	11.6	23	18.4	-59.2	18.9	12.9	25
Btn3	39.5	-72.1	-6.5	26.7	29	39.0	-70.6	-2.1	16.5	22	35.1	-66.8	3.7	18.9	22	22.8	-52.8	13.5	14.3	21	14.9	-30.6	29.1	17.7	24
Btn4	7.6	-10.6	16.9	3.7	7	7.0	-16.0	4.5	1.8	5	9.8	-20.0	1.1	2.4	6	5.6	-16.8	5.4	1.6	5	8.9	-21.8	0.3	3.3	7
Btn5	11.0	-16.2	25.0	2.0	8	9.4	-27.1	22.2	1.0	6	9.5	-24.2	24.8	1.5	7	9.0	-24.6	22.1	0.7	5	13.9	-15.7	36.5	3.8	11
Btn6	6.9	-8.5	15.3	0.5	4	5.4	-14.4	4.2	0.8	4	5.0	-14.6	8.5	1.1	7	4.2	-11.9	7.5	1.0	4	5.0	-11.8	17.9	2.9	7
Btn7	5.2	-6.4	15.0	0.1	2	4.5	-10.1	9.3	0.4	3	4.5	-10.6	8.6	0.7	3	3.8	-9.7	10.6	0.2	3	4.2	-5.3	12.0	0.3	3
1–3	43.0	-88.6	-5.9	26.8	29	41.4	-88.1	-2.1	15.8	26	41.7	-89.0	3.7	16.3	25	24.5	-68.6	14.6	13.1	25	16.3	-59.2	29.1	14.6	25
4–7	7.7	-16.2	25.0	1.6	8	6.6	-27.1	22.2	1.0	6	7.2	-24.2	24.8	1.4	7	5.7	-24.6	22.1	0.8	5	8.0	-21.8	36.5	2.5	11
1–7	22.8	-88.6	25.0	12.4	29	21.5	-88.1	22.2	7.3	26	22.0	-89.0	24.8	7.8	25	13.7	-68.6	22.1	6.1	25	11.6	-59.2	36.5	7.7	25



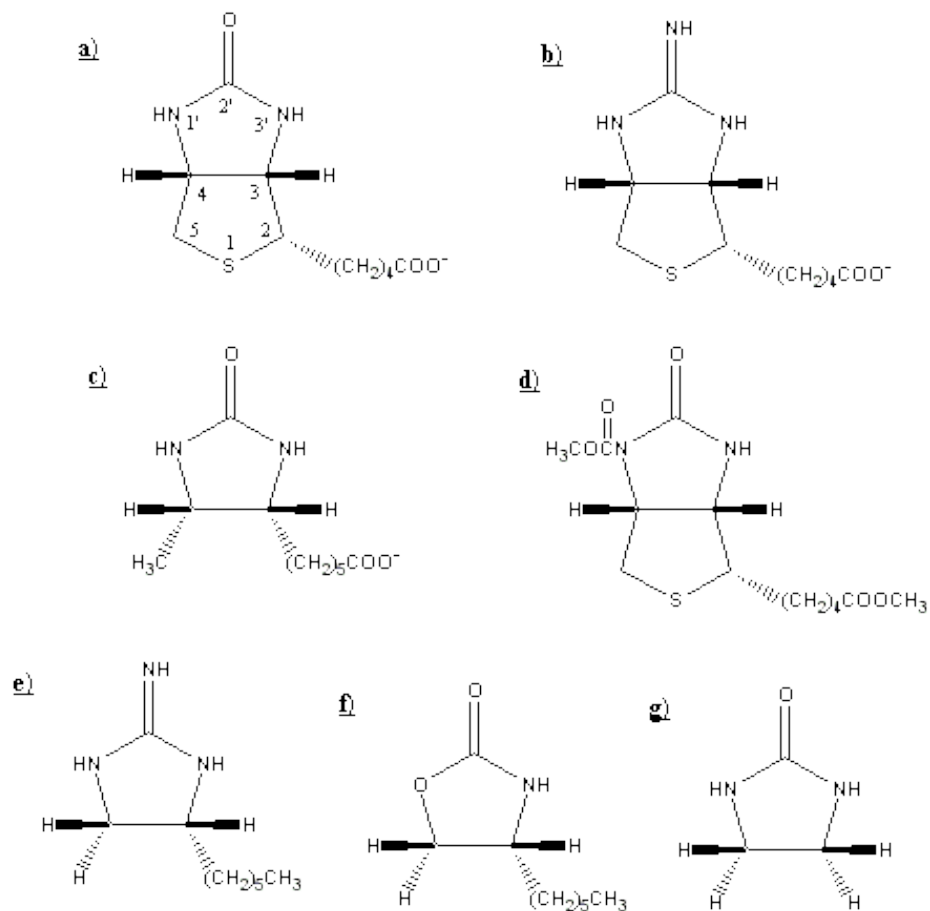
**Table 3.** Mean absolute, minimum, and maximum difference (MAD, Min, and Max) between the various charge sets and the QM charges for the sum of the ligand–protein electrostatic interaction energy and the generalised Born solvation energy in the 20 snapshots of the simulations of the seven biotin analogues (a negative sign indicates that that charge set gives more negative interaction energies than the QM charges). In addition, the average (Av) and maximum (Md) distance at which the residue-wise cumulative energy difference is converged to within 4 kJ/mol are given. The three last lines give the averages, maximum, or minimum over the three negatively charged biotin analogues (1–3), the four neutral biotin analogues (4–7), and over all seven biotin analogues (1–7).

	Amber					Cons					Cons1					Aver					QM1				
	MAD	Min	Max	Av	Md	MAD	Min	Max	Av	Md	MAD	Min	Max	Av	Md	MAD	Min	Max	Av	Md	MAD	Min	Max	Av	Md
Btn1	6.2	-15.2	7.1	2.6	5	9.1	-23.3	0.8	2.7	7	9.4	-23.4	0.9	2.7	7	7.2	-17.2	4.0	2.6	5	3.6	-11.4	7.6	1.8	7
Btn2	6.9	-17.7	12.7	3.3	7	7.5	-22.7	12.1	3.6	7	8.1	-23.2	11.6	3.2	7	8.5	-26.3	10.9	3.3	7	8.4	-23.3	10.0	2.9	7
Btn3	6.6	-24.6	7.5	2.8	5	8.5	-30.3	6.3	2.6	7	7.7	-28.4	8.8	2.6	7	8.3	-29.8	6.2	2.8	5	5.5	-21.7	11.5	1.5	3
Btn4	2.5	-6.7	2.9	1.6	3	2.6	-5.6	4.2	1.3	3	2.5	-6.7	2.9	0.9	3	3.1	-8.4	2.7	1.6	3	2.5	-7.7	3.8	0.7	3
Btn5	3.4	-3.9	12.8	1.3	4	3.0	-6.1	9.1	1.4	3	3.4	-3.9	12.8	1.6	3	3.2	-8.0	6.9	1.3	4	5.0	-1.4	15.6	1.5	4
Btn6	4.4	-10.4	0.5	1.8	3	2.0	-4.1	4.1	0.4	3	2.4	-4.1	4.8	0.5	3	2.4	-5.9	2.8	1.8	3	4.8	-0.8	8.9	1.8	3
Btn7	2.1	-1.1	6.5	0.7	3	2.2	-1.1	6.4	0.3	2	2.3	-1.0	6.4	0.3	2	1.7	-2.9	4.4	0.7	3	2.3	-2.1	5.5	0.6	3
1-3	6.5	-24.6	12.7	2.9	7	8.4	-30.3	12.1	3.0	7	8.4	-28.4	11.6	2.8	7	8.0	-29.8	10.9	2.9	7	5.8	-23.3	11.5	2.0	7
4-7	3.1	-10.4	12.8	1.3	4	2.4	-6.1	9.1	0.8	3	2.6	-6.7	12.8	0.8	3	2.6	-8.4	6.9	1.3	4	3.7	-7.7	15.6	1.1	4
1-7	4.6	-24.6	12.8	2.0	7	5.0	-30.3	12.1	1.7	7	5.1	-28.4	12.8	1.7	7	4.9	-29.8	10.9	2.0	7	4.6	-23.3	15.6	1.5	7

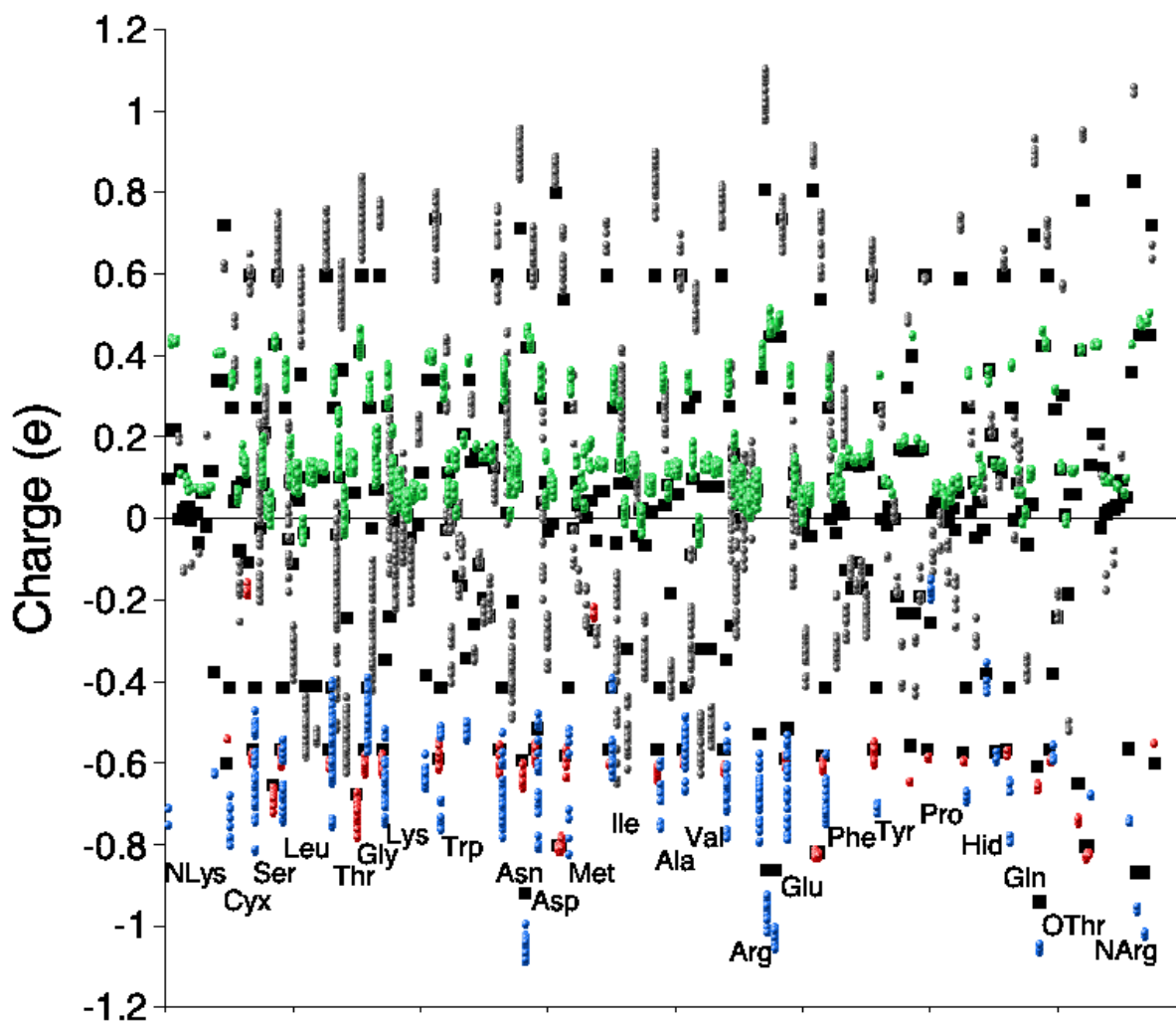
**Table 4.** MM/PBSA results with various charges. Four MD simulations were performed, with the Amber [23], Aver, Cons, and Cons1 charge sets, and then the MM/PBSA energies were calculated with all six charge sets for the Amber simulation, and with the native and Amber charge set for the other three simulations. In the table, the error of each calculated biotin analogue is given (compared to the experimental values, which are  $-85.4$ ,  $-59.8$ ,  $58.6$ ,  $-36.8$ ,  $-34.3$ ,  $-20.9$ , and  $-18.8$  kJ/mol, respectively [26]). In addition, the mean absolute deviation (MAD) and the correlation coefficient ( $r^2$ ) are given. The Amber simulation involved 20 snapshots, whereas the other three simulations involved 171 snapshots. The differences between MM/PBSA calculations based on the same MD snapshots but with different charge sets in the energy calculations are mainly caused by the entropy term.

MD	Amber					Cons		Cons1		Aver	
	Amber	QM1	Aver	Cons	Cons1	Cons	Amber	Cons1	Amber	Aver	Amber
Btn1	-28.9	-40.7	-14.4	-26.1	-38.0	-25.9	-19.3	-10.4	-3.8	-5.1	-26.2
Btn2	-21.6	-5.7	-23.3	-11.7	-10.7	-46.1	-39.9	-49.5	-43.9	-17.2	-8.6
Btn3	-11.6	-18.8	-34.7	-5.9	-12.7	-0.4	7.0	-12.3	-5.8	-22.8	-24.4
Btn4	13.4	41.7	39.4	21.3	23.7	22.0	8.1	28.6	20.7	-10.3	-3.4
Btn5	8.5	35.3	14.9	5.0	27.0	10.0	11.4	12.5	11.0	9.0	21.7
Btn6	-0.6	19.1	22.1	33.3	-1.5	17.0	7.2	15.9	-14.7	6.6	10.1
Btn7	10.1	8.6	16.7	6.5	15.6	22.4	18.3	18.7	16.4	12.7	16.4
MAD	13.5	24.3	23.6	15.7	18.5	20.6	15.9	21.1	16.6	11.9	15.8
$r^2$	0.96	0.87	0.85	0.94	0.91	0.89	0.87	0.81	0.72	0.91	0.95

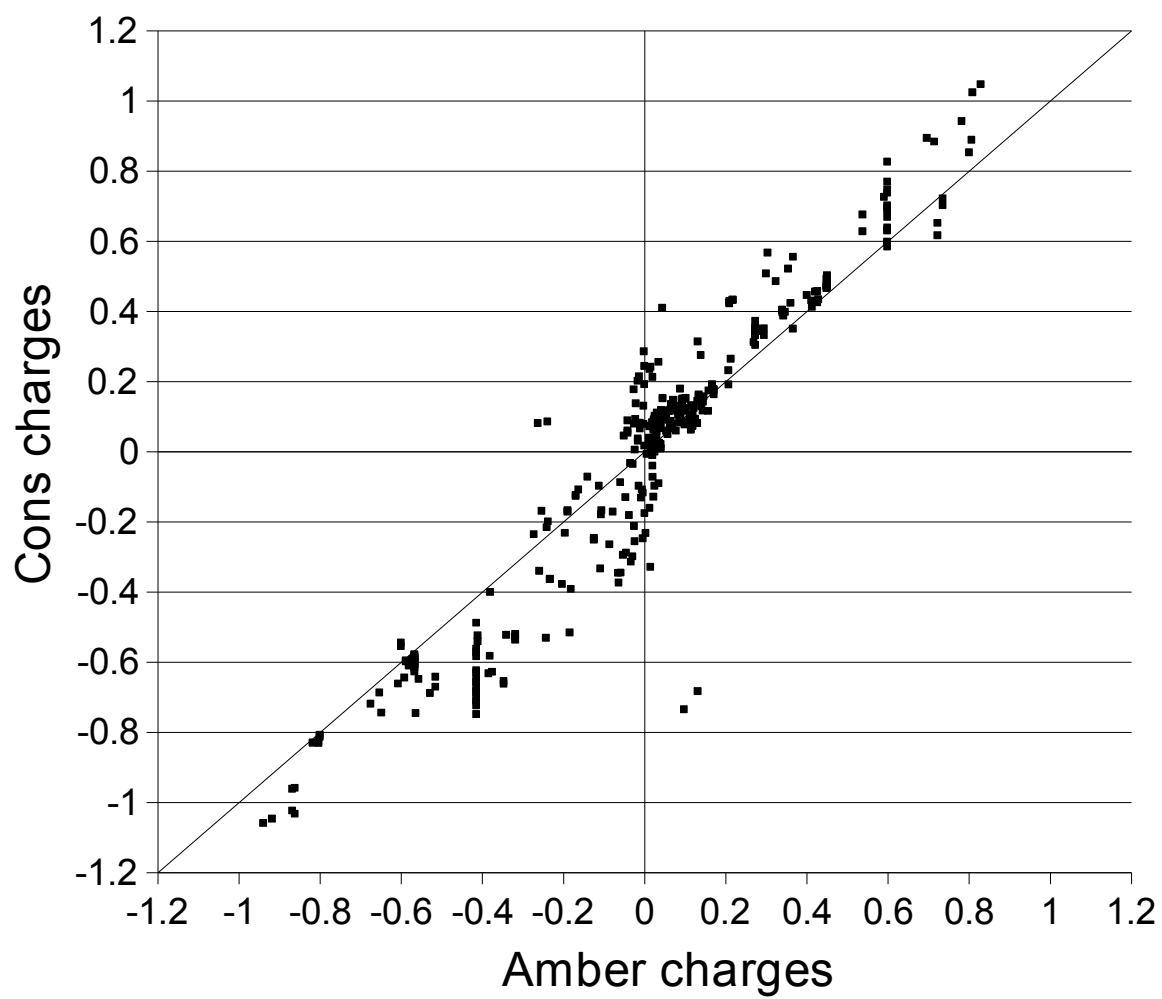
**Figure 1.** Structures of the seven biotin analogues studied: a) biotin (Btn1), b) 2'-iminobiotin (Btn2), c) desthiobiotin (Btn3), d) 1'-N-methoxycarbonylbiotin methyl ester (Btn4), e) *D*-4-*n*-hexyl-2-iminoimidazolidine (Btn5), f) *D*-4-*n*-hexyloxazolidone (Btn6), and g) imidazolidone (Btn7).



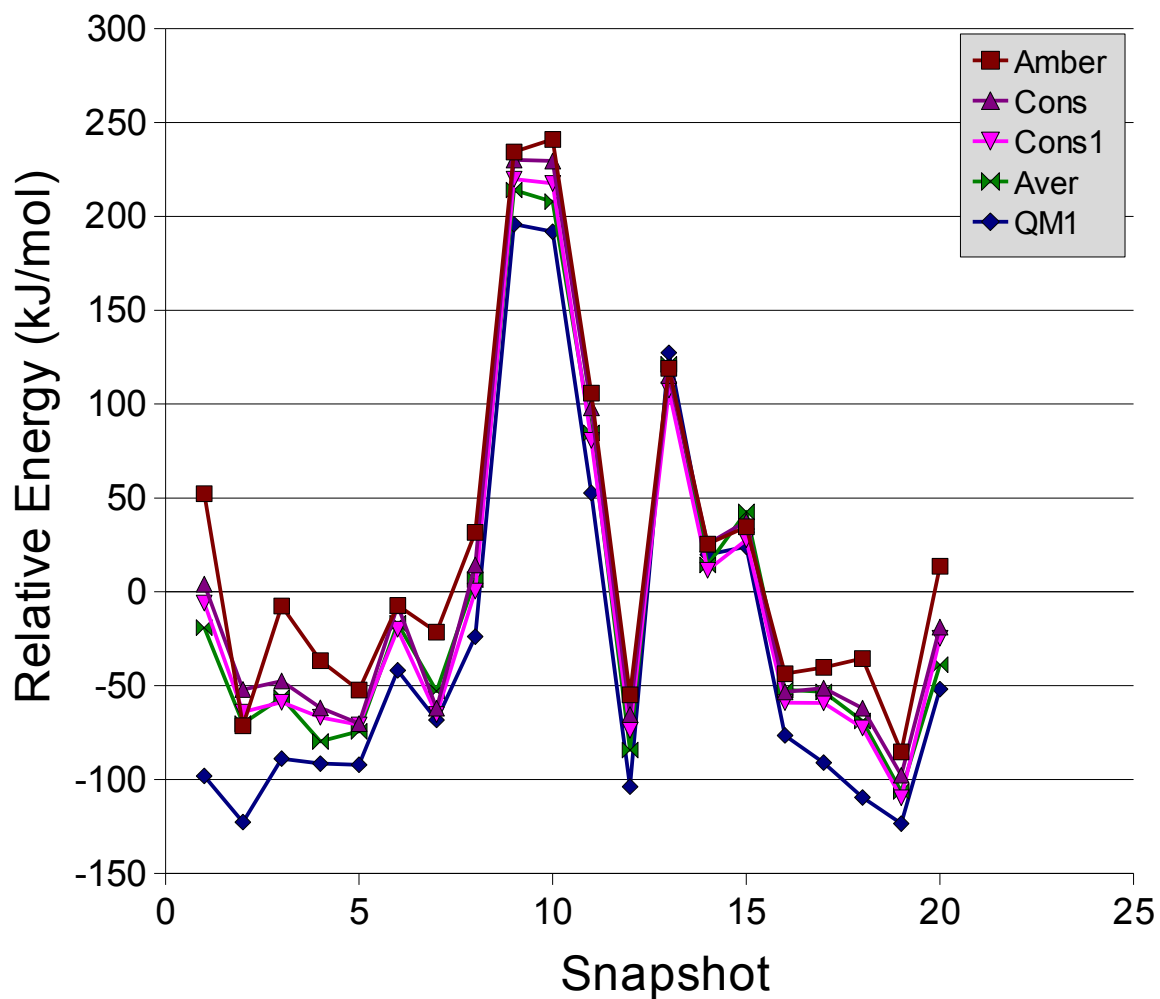
**Figure 2.** Variation of the Aver charges (only protein atoms) collected according to the residue type and atom name (residue-wise starting with N and ending with O; NLys and NArg are N-terminal residues; OThr is the carboxy-terminal residue). The charges are colour coded after the element: carbon – grey, oxygen and sulphur – red, nitrogen – blue, and hydrogen – green. The corresponding Amber 1994 charges are marked by black squares.



**Figure 3.** The correlation between the Amber and Cons charges. The solid line represents the  $x = y$  line.



**Figure 4.** Total electrostatic interaction energy within the avidin tetramer (excluding the ligand and water molecules) for the 20 snapshots of the Btn1 simulation, relative to the average energy of all snapshots.



**Figure 5.** Distance dependence of the individual and cumulative difference in the electrostatic interaction energies between the Amber and QM charges for biotin and avidin in the snapshot with the largest total difference. Energies are plotted for each residue in the avidin tetramer as a function of the shortest distance between any atom in that residue and any atom in the ligand.

