Conformational Dependence of Isotropic Polarizabilities

Söderhjelm, Pär; Kongsted, Jacob; Ryde, Ulf

Published in:
Journal of Chemical Theory and Computation

DOI:
10.1021/ct100714e

2011

Conformational dependence of isotropic polarisabilities

Pär Söderhjelm [1], Jacob Kongsted [2] & Ulf Ryde [3] *

[1] Department of Chemistry and Applied biosciences - Computational Science, ETH Zürich, Via Giuseppe Buffi 13, CH-6900 Lugano, Switzerland

[2] Department of Physics and Chemistry, University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark

[3] 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,

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

2012-02-03

Abstract

We perform a statistical and energetic analysis of atomic polarisabilities obtained with the LoProp approach for all atoms in the avidin tetramer for seventy snapshots from molecular dynamics simulations with seven different biotin analogues, and from the crystal structure of the photosynthetic reaction centre (in total 560698 individual polarisabilities). Dynamic effects give a variation of the polarisabilities of 0.09 $Å^3$ on average. Atoms at different positions in the sequence show a variation of 0.14 $Å^3$ on average, caused by the conformational dependence of the polarisabilities. This variation gives errors of 2 and 1 kJ/mol for relative conformational and ligand-binding induction energies. Averaged element-wise or atom-type polarisabilities give larger errors, e.g. 9 and 7 kJ/mol, respectively, for the relative conformational energies. Therefore, we recommend that polarisabilities should be assigned atom wise (i.e. individual polarisabilities for each atom in all residues), in the same way as for charges. We provide such a set of extensively averaged polarisabilities (xAvPol) for all atoms in avidin and the photosynthetic reaction centre, applicable at the B3LYP/aug-cc-pVTZ level, which is converged with respect to the basis-set limit.

Key Words: atomic polarisabilities, molecular mechanics, force fields, conformational dependence

Introduction

During the latest decades, molecular simulations have become a powerful alternative and complement to experiments to obtain information about the structure and function of macromolecules. Such simulations are mainly based on the molecular mechanics (MM) approach, employing empirical force fields.[1] One of the most crucial issues in these force fields is the treatment of electrostatics. The great majority of such MM force fields for macromolecules employ a simple Coulomb interaction between atom-centred fixed partial charges. The atomic charges are typically obtained from quantum mechanics (QM) calculations, by fitting them to reproduce either the QM electrostatic potential or intermolecular interaction energies.[2,3,4,5]

It has long been recognised that this provides a quite crude description of the electrostatics. In particular, induction effects are completely ignored or treated in an implicit average sense, although it is well know that polarisation typically constitutes 6–30 % of the electrostatic interaction energy.[6,7,8,9,10,11] Consequently, there has been a great interest to incorporate induction effects in the MM force field,[11,12,13,14,15] e.g. by using fluctuation charges,[16,17] induced dipoles,[18,19,20] or Drude oscillators.[21,22,23] The first polarisable force field appeared as early as in the mid 1970ies,[19] and specialised and accurate force fields like SIBFA, EFP, and NEMO also early employed polarisabilities (and higher-order multipoles).[13,24,25] During the last decade, polarised variants of the more widely used macromolecules force fields have started to appear, e.g. Amber-02, PFF, and Amoeba,[11,26,27,28] all three based on atomic isotropic dipole polarisabilities.

Naturally, the accuracy of polarisable force fields depends on the accuracy of the atomic polarisabilities employed. As for atomic partial charges,[2] atomic polarisabilities are not observables, meaning that there are no reference values that could be obtained from experiments or QM calculations.[11] Instead, atomic polarisabilities have to be determined by some (arbitrary) method that is optimised in a specific way. Several methods to obtain distributed polarisabilities from QM calculations have been suggested.[11] For example, the atomic polarisabilities can be obtained by partitioning molecular polarisabilities, either in real space (e.g. the atoms-in-molecules approach[29]) or in terms of the basis set.[30,31] Moreover, there are also several ways to apply the perturbing field.[32,33,34] Alternatively, the polarisabilities can be determined by fitting to a property calculated by QM methods, e.g. the molecular polarisabilities or induction energy.[18,35,36,37,38,39,40,41,42]

There are several sets of atomic polarisabilities available. Some of them are listed in Table 1.[18,19,20,26,27,28,35,43,44,45] Apparently, there is little agreement in the values used or how the polarisabilities should be assigned. Thole and van Duijnen have argued that good reproduction of molecular polarisabilities can be obtained by a single isotropic polarisability for each element,[20,46] and Warshel simply uses 0.5 $Å^3$ for hydrogen atoms and 1 $Å^3$ for all other atoms.[19] Other force fields use 8–15 atom types, with 1–4 different polarisabilities for each element for the normal amino acids. This is in sharp contrast to atomic charges, for which most general-purpose macromolecular force fields today employ individual charges on each distinct (by symmetry) atom in each amino acid. In fact, Woods and coworkers have shown that improved accuracy is obtained using specific atomic polarisabilities, rather than polarisabilities determined by the atom type.[42] They also tested the conformational dependence of the fitted polarisabilities and showed that it was quite small, ~1%.

In this paper, we address these issues in a more systematic way. In previous investigations of the influence of the protein electrostatics on excitation and ligand-binding energies, we have calculated polarisabilities for all atoms in several proteins with QM calculations,[47,48] using the LoProp approach.[34] Here, we analyse those data, collecting statistics over the polarisabilities of each atom in the sequence. Thereby, we can address questions such as: How large is the conformational dependence of atomic polarisabilities? How are polarisabilities best assigned: by element, by atom type, or by atom? Can transferable polarisabilities be obtained by simply averaging over all calculated values?

Methods

In this paper, we analyse polarisabilities calculated in two studies, viz. a study of the binding affinity of seven biotin analogues to the protein avidin[48] and new calculations for the photosynthetic reaction centre (PRC) from *Rhodobacter sphaeroides*. Both these studies employed a multicentre–multipole expansion up to quadrupoles and anisotropic polarisabilities, obtained with the LoProp approach[34] using the Molcas software.[49] The LoProp method has been shown to be better than other related methods to calculate polarisabilities.[50] The calculations were performed at the density functional B3LYP[51] level, using either the 6-31G*,[52] aug-cc-pVDZ, aug-cc-pVTZ, or aug-cc-pVQZ basis sets.[53] These basis sets are of sizes smaller, similar, larger, and much larger, respectively, than the popular Sadlej basis set designed for the calculation of polarisabilities.[54] Each basis set was turned into the atomic natural orbital form (as required by the LoProp procedure) by a linear transformation that does not affect the orbital optimisation.

The properties were calculated for the whole protein by dividing it into the individual amino-acid residues, which were capped with $CH_3CO–$ and $–NHCH_3$ groups (dipeptides). The effect of the capping groups were removed by calculating the properties also of the overlapping $CH_3CONHCH_3$ fragments and subtracting them from the properties of the corresponding dipeptides, the molecular fractionation with conjugate caps approach,[55] which has been shown to give errors of 1 kJ/mol or less.[10] A separate calculation was performed on every residue in the structure, with the actual geometry obtained either from the crystal structure (PRC) or from ten snapshots from a MD simulation with the Amber-02 force field (avidin[56]).

In the standard LoProp approach, anisotropic polarisabilities are obtained both for atoms and for bond isocentres. To facilitate the present comparison, we restricted the study to isotropic polarisabilities, because this is the form used in the Amber-02, PFF, and Amoeba force field. The isotropic polarisabilities were obtained as the average of the three diagonal elements of the anisotropic tensor. Moreover, only atomic polarisabilities were considered by partitioning the bond polarisabilities equally on the two bonded atoms.

Interaction energies were calculated with the Amber-10 software,[57] using Amber exclusion rules, i.e. that polarisation between atoms separated by one or two bonds is ignored, whereas for atoms separated by three bonds, the electric field was scaled by a factor of 1.2.[26] The induction energy was calculated iteratively until successive estimates of the induced dipoles agreed within 0.0001 Debye, using a second-order extrapolation scheme (indmeth=1).

The exclusion rules are important, because they influence the molecular polarisability resulting from a given set of atomic polarisabilities. Therefore, polarisabilities derived with a specific set of rules are in principle not comparable to those derived with other rules, and they cannot be directly transferred. Nevertheless, such transferability have sometimes been assumed, as in the development of the Amber 2002 force field,[26] in which Applequist polarisabilities, derived using coupling between all atoms, were adopted into the much more restricted coupling scheme of Amber. One can therefore expect that these polarisabilities are too small.

The same problem also occurs in this investigation, because the LoProp polarisabilities add up to the molecular polarisability and thus should not be coupled within the molecule used to calculate them, in our case a protein residue. Thus, when they are used with the Amber exclusion rules or numerically compared to Amber polarisabilities, they should in principle be scaled down to reproduce the (isotropic) molecular polarisability. To investigate the magnitude of this effect, we assumed a uniform scaling over all atoms in a molecule and calculated the required scale factor for each of the 991 molecules used to compute the LoProp polarisabilities for an avidin snapshot. On average, this factor was 0.987, with a standard deviation of 0.007. Because the influence of such scaling on the results would be negligible, we did not modify the polarisabilities. It should also be noted that the choice of exclusion rules also has an effect on the polarisation caused by the static charges. However, in the

Amber polarisable force field, the charges are derived by taking the statically induced dipoles into account so that the major part of this effect is cancelled. Because of this connection, we did not specifically study this issue.

We studied the binding of the seven biotin analogues (BTN1–BTN7) in Figure 1 to avidin. The set-up of the molecular dynamics simulations has been described before.[10,56] We used ten snapshots (sampled every 20 ps) for each analogue taken from this investigation, performed by the polarisable Amber 2002 force field[11,26] (the 02ohp simulation in ref. 56).

Result and Discussion

Polarisabilities

First, we studied the conformational dependence of the polarisabilities calculated with the LoProp approach[34] for all atoms in 10 snapshots from MD simulations of avidin bound to the seven different biotin analogues in Figure 1 using the B3LYP/6-31G* method. The LoProp polarisabilities range from 0.05 to 2.45 Å^3 (H in Phe-70 to SG in Cyx-452; Cyx denotes Cys in cystine linkages). For individual atoms, the range of the polarisability (i.e. the maximum minus the minimum value of the polarisability of the same atom) among the 70 snapshots varies from 0.008 to 0.35 Å^3 (for HH2 in Trp-219 and CD2 in Trp-68; average 0.09 Å^3). This illustrates the expected variation of the polarisabilities caused by dynamic effects. There is little similarity between the calculated polarisabilities and those in the Amber-02 force field: In fact, for 6796 of the 7708 protein atoms (88%), the Amber value is outside the range of the calculated polarisabilities in the various snapshots.

An interesting question is how polarisabilities are best assigned to atoms in a protein. Are they the same for each element, for each atom type, or should they be assigned atom-wise, like point charges? Statistics for elemental polarisabilities are given in Table 2. It can be seen that the LoProp polarisabilities of all elements show a quite large variation, ranging from 0.27 Å^3 for H and O to ~0.75 Å^3 for N and C. Thus, it does not seem to be a good idea to assign polarisabilities only on the basis of the element. For all elements, except sulphur, the averaged LoProp polarisabilities are higher than the corresponding Amber values. For H, C, and O, the Amber values are within the calculated range, but for C and S, at least some of the Amber values are outside the range of the LoProp values. The same applies to all the other sets of polarisabilities in Table 1, although with different elements.

The corresponding statistics for the Amber-02 atom types are shown in Tables 3 and 4. Amber-02 employs 27 atom types for a normal protein, which are all included and described in Table 4. However, most of the Amber-02 atom types of the same element use the same polarisabilities. In fact, there are only ten distinct polarisabilities in Amber (taken from Applequist;[18] three for C and H, two for O, and one for N and S). These are shown in Table 3. It can be seen that the LoProp polarisabilities still show large ranges, e.g. up to 0.77 Å^3 for carbon, and 0.57 and 0.75 Å^3 for S and N. Hydrogen has the lowest ranges (0.10–0.23 Å^3), followed by oxygen (0.23–0.27 Å^3). There is a fair correlation between the average calculated values and the Amber values (r^2 = 0.78), but only for three atom types (CT, H other, and O+O2) are the Amber-02 polarisabilities within the range of the calculated ones.

The corresponding statistics for all the 27 Amber-02 atom types are given in Table 4. It can be seen that the range is still large for most atom types, up to 0.77 Å^3 for CT (sp^3 carbon). In fact, the range is below 0.1 Å^3 only for three of the Amber atom types, H4, H5, and HP (explained in Table 3). For 20 of the 27 atom types, the Amber polarisabilities are outside the range of the calculated ones. In many cases, it is obvious that the Amber atom types still are too crude to give accurate and transferable polarisabilities. This is clearly illustrated for CA atoms of Gly and Asp (which share the same Amber atom type), shown in Figure 2, where the frequencies of the LoProp polarisabilities are shown for the 70 snapshots and the 44 and 22 atoms of each type, respectively. It is obvious that the two distributions are distinct and

essentially non-overlapping, so that different polarisabilities are appropriate for the CA atom in these two amino acids.

Finally, we calculated the average of the polarisabilities for the same atom in the same residue anywhere in the sequence and over the 70 snapshots. This suppressed some of the variation. Now, the average range was 0.14 Å^3. 229 of the 388 distinct atoms (59%) showed a range of less than 0.15 Å^3 and only 28 atoms showed a range over 0.3 Å^3, with CD2 of Trp showing the largest ranges (0.46 Å^3). Other atoms with large ranges are always carbon and nitrogen atoms, as well as the two sulphur atoms. These are also the atoms with the highest polarisabilities. In fact, there is a good correlation between the size of the polarisabilities and the range (r^2 = 0.76), as is shown in Figure 3. This shows that there is a significant conformational dependence of the polarisabilities (23% on average), much larger than for small model compounds (1%).[42] In fact, 70% of the polarisabilities of all possible pairs of atoms from the same residue at different places in the sequence were statistically different at the 95% level according to a simple *t* test.

The largest polarisabilities are those of the two S atoms in Cys and Met (2.27 and 2.04 Å^3). Then, come carbon atoms, typically CA atoms in various residues, but also some CB and CG atoms (up to 1.39 Å^3 for CA in Asp). The smallest C polarisability is the CG atoms of Val (0.90 Å^3). The largest nitrogen polarisability is that of the back-bone amide in Pro (1.14 Å^3) and the smallest one is the side-chain NZ of Lys (0.62 Å^3). The largest oxygen polarisability is that of the OH group in Tyr (0.64 Å^3). The smallest one is the amide back-bone O of Cyx (0.43 Å^3). The hydrogen polarisabilities are well separated from those of the other elements. The largest one is that of HH2 in Trp (0.32 Å^3) and the smallest is the amide back-bone H of Phe (0.16 Å^3).

There are several obvious groups of the calculated polarisabilities. For O, they are distinct and not overlapping: hydroxyl and back-bone carbonyl groups (0.50–0.55 Å^3), side-chain carbonyl groups and all carboxyl groups (0.56–0.60 Å^3), and the hydroxyl group of Tyr (0.64 Å^3). The same applies to N atoms, although the ranges are larger: N in Lys side chains, and in NH of Arg (0.62–0.69 Å^3), N in side-chain amides (0.71 Å^3), N in His and NE in Arg (0.86–0.91 Å^3), N in the back-bone amides and NE in Trp (0.84–1.06 Å^3), and N in Pro (1.14 Å^3). However, for the hydrogen atoms, the ranges are large and overlapping: H in amide and NH_3^+ groups (0.14–0.18 Å^3), H in hydroxyl groups (0.16–0.17 Å^3), H in side-chain amide groups (0.18–0.22 Å^3), HC with electron-withdrawing neighbours (0.19–0.26 Å^3), H in aromatic groups (0.26–0.32 Å^3), and other HC (0.22–0.28 Å^3).

Finally, for carbon atoms, it becomes even harder to find natural groups: Methyl groups, as well as CB and CD in Pro and CE1 in His have 0.90–0.96 Å^3, CD2 in His, CD in Arg, CG and CD in Lys have 0.97–1.06 Å^3, C in side-chain carbonyl groups and all carboxyl groups have 1.04–1.13 Å^3, C in back-bone carbonyl groups, as well as CD2 in Hie and Hip, and CD1 in Trp give 1.10–1.20 Å^3. However, the remaining aliphatic and aromatic C atoms still give large and overlapping ranges (1.03–1.39 and 1.07–1.34 Å^3, respectively), without any obvious grouping.

Figure 4 shows the correlation between the atomic polarisabilities and the Amber polarisabilities. It can be seen that there is some correlation (r^2 = 0.72), but there is room for a significant improvement, in particular for the carbon, nitrogen, and sulphur atoms. Apparently, the polarisabilities of the atoms are very sensitive to their neighbouring atoms in a way that hard to describe without introducing very many atom types. Therefore, we suggest that for accurate results, it is better to assign separate polarisabilities to each atom in every amino acid, rather than using atom types, as is done for the charges in most force fields, including Amber. In analogy with the extensively averaged electrostatic potential (xAvESP) charges obtained in a similar way,[58,59] we call these averaged LoProp atomic polarisabilities from avidin xAvPol1 in the following and they are provided in the supplementary material, Table S1.

Basis-set dependence

It is well-known that calculated polarisabilities are sensitive to the specific electronic-structure method and the one-electron basis sets.[60] Owing to the presence of the electric-dipole operator in the second-order perturbation theory expression for the dipole–dipole polarisability, use of diffuse basis functions in accurate calculations of polarisabilities is usually of great importance. In the avidin calculations, we have used the B3LYP density functional combined with the middle-sized 6-31G* basis set. In order to check the reproducibility of these results, we need to assure that polarisabilities calculated with other methods are not widely different. Fortunately, we have also polarisabilities calculated at the B3LYP/aug-cc-pVTZ level for one snapshot of two of the biotin analogues (Btn1 and Btn7; the results for the two ligands are very similar). Therefore, we can make a direct comparison of the polarisabilities obtained with this more accurate but much more expensive method. The polarisabilities calculated with the two methods differ by 0.12 Å^3 on average, the larger basis set giving larger polarisabilities (only for ~5% of the atoms does the calculation with the smaller basis set gives larger polarisabilities, and only by up to 0.04 Å^3). As expected, the largest differences are obtained for the negatively charged carboxylate groups and for the sulphur atoms: The difference is 0.61 Å^3 for SD in Met, 0.44 Å^3 for SG in Cyx, 0.48–0.55 Å^3 for the carboxylate O atoms, and 0.42–0.51 Å^3 for the carboxylate C atoms (with slight differences between Asp, Glu, and the carboxy terminals). Other atoms with large differences are OE1 of Gln (0.31 Å^3), CE1 and NE2 of Hid (0.29 Å^3), OD1 of Asn, CH2 and CZ3 of Trp (0.28 Å^3).

Again, there is a significant variation between the various atoms, which is impossible to describe element-wise and also hard to describe by atom types. Instead, it is best described by atomic polarisabilities. Then, the differences are highly reproducible: Only three atomic polarisabilities give difference over 0.01 Å^3 between the Btn1 and Btn7 simulations (SD in Met, OD1 in Asp, and C in the carboxy-terminal, with differences of 0.04, 0.02, and 0.02 Å^3, respectively). Thus, the effect of the basis set is quite small and highly consistent and therefore the polarisabilities can quite easily be extrapolated to the larger basis set. This will increase the polarisabilities for all except five atoms (CA in Lys and Arg, CB in Ile and Val, as well as CG in Leu). Therefore, the difference towards the Amber polarisabilities will increase, except for the two S atoms, which become similar with the larger basis set, 2.65 and 2.67 Å^3, for SD in Met and SG in Cyx, respectively (2.9 Å^3 in Amber).

To further study the basis-set dependence of the polarisabilities, we performed some additional calculations with the aug-cc-pVDZ, aug-cc-pVTZ, and aug-cc-pVQZ basis sets (still with the B3LYP method) for the groups that showed the largest dependence with respect to the basis set, Cys, Cyx, Met, Asp, and a carboxy terminal. The results show that the polarisabilities are reasonable converged at the aug-cc-pVTZ level: The polarisabilities calculated at the aug-cc-pVTZ and aug-cc-pVQZ differ by only 0.02 Å^3 on average, with a maximum difference of 0.09 Å^3 for SD in Met (the polarisability decreases when the basis set is increased). The SG atoms in Cys and Cyx also show rather large differences, 0.04–0.08 Å^3, whereas the polarisabilities of the carboxylate O atom change by only 0.03 Å^3 (but that of the carboxylate C atom change by 0.05 Å^3). Besides these atoms, the largest change is 0.04 Å^3 for some carbonyl O atoms. In fact, the polarisabilities are fairly converged already at the aug-cc-pVDZ level, with average and maximum differences of 0.03 and 0.15 Å^3 (again SD of Met gives the largest change) towards the aug-cc-pVQZ data. This shows that it is probably better to calculate the polarisabilities with the aug-cc-pVDZ or Sadlej basis set than with 6-31G*.

On the other hand, it is normally assumed that polarisabilities in the condensed phase are lower than those calculated in vacuum,[27,61,62] e.g. by 7–9 % for water. Therefore, the Friesner group uses a basis set without diffuse functions (cc-pVTZ-f[63]) for the calculation of polarisabilities, whereas MacKerell and coworkers scale down polarisabilities by a factor of 0.724.[62] However, the primary aim of this paper is not to establish a proper level to calculate polarisabilities, but rather to quantify the extent and effect of conformational dependence of

polarisabilities in proteins.

Different proteins

Next, we performed the same analysis for another protein, viz. the photosynthetic reaction centre from *Rhodobacter sphaeroides*. We calculated the LoProp isotropic atom-centred polarisabilities for each atom (in total 12818), but only for a single structure (the protonated crystal structure). From these, we calculated atomic polarisabilities by averaging over all residues of each type in the protein (xAvPol2; also included in Table S1). For the 325 atoms that are common to avidin, the average difference between the two sets is only 0.02 $Å^3$, indicating that the LoProp atomic polarisabilities are remarkably transferable between different proteins. In particular, the largest differences (up to 0.13 $Å^3$), were observed for C and N atoms in Hid and Tyr residues, for which there is only one occurrences in the avidin monomer, showing that the deviation is mainly statistical in the nature (but it also indicates that there is a significant conformational dependence of the polarisabilities).

Finally, we constructed a set of atomic polarisabilities by averaging over the two proteins, weighting the average after the number of residues of each type in the monomer of each protein. For example, there are 79 Ala residues in PRC and 4 in the avidin monomer, so we summed the polarisability from PRC multiplied by 79 and that of avidin multiplied by 4 and divided the sum by 83. Note, however, that this weighting of the average has a maximum effect of 0.06 $Å^3$, so it is of little importance. This averaged set of atomic polarisabilities will be called xAvPol3 in the following. We also constructed a forth set of polarisabilities by extrapolating the xAvPol3 polarisabilities to the aug-cc-pVTZ basis set with the atomic correction factors obtained in the previous section. The resulting set, xAvPol4, is also included in Table S1.

Induction energies

Up to now, we have only discussed the actual value of the polarisabilities. To put these into a more interesting perspective, we studied how these differences in the polarisabilities affect electrostatic interaction energies. Therefore, we have calculated three types of energies for avidin and its complexes with the seven biotin analogues in Figure 1. We tested 13 different sets of polarisabilities viz. the original LoProp polarisabilities for avidin (LoProp), polarisabilities averaged over the ten snapshots (Aver), xAvPol1, xAvPol2, xAvPol3, xAvPol4, the average elemental polarisabilities in Table 2 (Element), the averaged polarisabilities for the 27 Amber atom types in Table 4 (Type), as well as the Amber02, Amber09, Charmm, Amoeba, and Enzymix polarisabilities listed in Table 1. The polarisabilities are shortly described in Table 5. All the other MM parameters, including the atomic charges, were identical in the calculations. The calculations were performed with the Amber software[57] and the Amber-02 charges.[26]

First, we studied the total induction energy within the whole avidin tetramer without any ligand and water molecules in the 70 snapshots. The absolute energies are not comparable, because different polarisabilities are used, but the fluctuations around the average value should be similar if the different force fields are to sample the same configurational space. Interestingly, all polarisabilities give fluctuations with a range (maximum minus minimum value among the 70 snapshots) of 1515–1651 kJ/mol. The force fields based on the LoProp B3LYP/6-31G* polarisabilities give a smaller range (1515–1533 kJ/mol) than the other polarisabilities (1553–1595 kJ/mol), and Enzymix gives the largest range (1651 kJ/mol).

Next, we compared these relative interaction energies for each snapshot, using the Aver polarisabilities as a reference (we cannot use the LoProp polarisabilities as a reference, because they change for each snapshot). Several conclusions can be drawn from the results presented in Table 6. First, the various force fields give mean absolute differences (MADs) of

2–65 kJ/mol in the order xAvPol1, xAvPol3, Type, Element, xAvPol2, Amber-02, xAvPol4, Amber-09, Charmm, Amoeba, and Enzymix. Thus, the polarisabilities are much less sensitive to the conformation than charges: The MAD between the Aver and xAvPol1, xAvPol2, or xAvPol3 sets is only 2 kJ/mol, and both the Type and Element polarisabilities give MADs less than 10 kJ/mol, which may be acceptable in many applications.

Second, the B3LYP/6-31G* polarisabilities are clearly not converged, because the B3LYP/aug-cc-pVTZ polarisabilities (xAvPol4) give induction energies that differ by 27 kJ/mol on the average. This shows that larger basis sets should be used for the calculation of the polarisabilities or they should be corrected in the same way as for xAvPol4.

Third, different standard force fields give widely differing results, differing from Aver by 26–65 kJ/mol, or up to 4% of the total variation. In most cases, the crude Enzymix polarisabilities give the largest difference. Of course, some of this difference may be caused by the fact that the Aver polarisabilities are based on calculations with a too small basis set. Therefore, we have added an extra line in Table 6 (MAD') where we instead use the xAvPol4 results (which are close to the basis-set limit) as the reference. It can be seen that the MAD for Amber09, Amoeba, and Enzymix are reduced to 15, 35, and 42 kJ/mol, whereas the MAD for Charmm is not changed and that of Amber02 actually increases. This shows that there still are extensive differences between the polarisabilities of the various force fields, far beyond what is caused by the conformational dependence.

Finally, we note that the variation in the relative induction energies is appreciably smaller than the corresponding variation in relative electrostatic energies when the atomic charges were varied in a similar manner (up to 150 kJ/mol).[58] This is in accordance with the observation that induction energies typically are 6–30% of the electrostatic energies.[6,7,8,9,10,11] Still, differences of over 10 kJ/mol in relative energies may have a strong influence on the phase space visited during a MD simulation.

Ligand binding energies

Next, we studied the induction contribution to the binding energy of the seven biotin analogues in Figure 1 with ten snapshots for each ligand and the same 13 sets of polarisabilities (and still with the same Amber-02 charges). The energy was calculated as the difference between the interaction energies in the complex, the protein, and the ligand:

$$E(\mathrm{PL}) - E(\mathrm{P}) - E(\mathrm{L})$$

without any solvation. Only one of the biotin ligands in the tetramer (the fourth) was considered, whereas the other three were considered as a part of the protein. The results in Table 7 shows that the Aver polarisabilities give induction contributions to the binding energies that are most similar to those obtained with the LoProp polarisabilities, with a MAD of 1 kJ/mol and a maximum error of 3 kJ/mol for the three charged ligands (Btn1–Btn3), and 0.3 and 0.9 kJ/mol for the neutral ligands, respectively. The xAvPol1 polarisabilities also give excellent results with only slightly higher deviations. If the xAvPol2, xAvPol3, or even the atom-type polarisabilities are instead used, the MADs increase to 2 kJ/mol and 1 kJ/mol, respectively, and the maximum errors increase to 5 and 2–3 kJ/mol. On the other hand, the elemental polarisabilities give much worse results, with a MAD of up to 8 kJ/mol for the charged ligands (but only 1 kJ/mol for the neutral ligands). Recalculating the polarisabilities with a larger basis set (xAvPol4) has a major effect on the interaction energies, with MADs of 19 and 6 kJ/mol, respectively, again indicating that 6-31G* is a too small basis set for polarisabilities.

Among the various standard force fields, Amber-02 polarisabilities give results that are closest to the LoProp results, with MADs of 4–5 kJ/mol and maximum errors of 11–12 kJ/mol. The other force fields give larger differences, e.g. MADs of 25–41 kJ/mol for the charged ligands and 3–16 kJ/mol for the neutral ligands. If we instead compare to the xAvPol4 results (available only for Btn1 and Btn7), the results for all force fields are

improved (to 2 –8 kJ/mol average deviation for Btn7 and 8 –17 kJ/mol for Btn1), except for Amber02. This indicates that the Amber02 polarisabilities are not compatible with high-level QM calculations, presumably because the force field employs artificially restrictive exclusion rules, as discussed in the method section.

Previously, we have observed that effects of variations of the charges are strongly screened by solvation.[58] Therefore, we studied the effect of solvation also for the polarisabilities. Unfortunately, neither of the continuum-solvation models available in Amber is compatible with a polarisable force field. Therefore, we instead simply included all explicit solvent molecules in the calculation of the energy terms for the complex and the free protein. Of course, this is not a fully consistent method, but it at least gives an indication of how much solvation may screen the effect of differences in the polarisabilities. The results in Table 8 show that solvation has a small effect on the induction-energy part of the ligand-binding energies. In particular, no clear screening by solvation is seen. In fact, if different solvation models are used in the calculations (i.e. polarisabilities for the explicit water molecules that are consistent with the respectively force field), the differences are typically increased, whereas if the same (LoProp) water polarisabilities are used in all calculations, the results are similar to that obtained without solvation.

Conclusions

In this paper, we have made a statistical and energetic analysis of isotropic atom-centred polarisabilities calculated individually for all atoms in two different proteins and for seventy snapshots from molecular dynamics simulations (in total 560698 individual polarisabilities). As mentioned in the introduction, atomic polarisabilities are not observables, so there are no true reference values of these. It is also well-known that polarisabilities strongly depend on the method and basis sets used for their calculation and that polarisabilities in the condensed phase are different from those in gas phase.[27,60,61,62] Moreover, the polarisabilities are closely connected to the model used for the permanent electrostatics and exclusion rules used in the force field.[11] Therefore, it is not meaningful to discuss whether one set of polarisabilities is better than another without defining all the other components of the force field. Instead, this article is concerned with more general aspects of the polarisabilities, viz. their variation with conformation and chemical environment, and how polarisabilities are best assigned (by element, by atom type, or by individual atoms).

First, we show that dynamic effects induce a variation in the polarisabilities of individual atoms of 0.01–0.35 Å^3, with an average of 0.09 Å^3 for the 7827 atoms in the avidin tetramer. The standard deviation ranges from 0.002 to 0.07 Å^3 (average 0.02 Å^3), indicating that up to 50 snapshots are needed to obtain a standard error of less than 0.01 Å^3 for all polarisabilities. This clearly shows that it is not enough to calculate polarisabilities for a single structure.

Second, we show that it is very hard to assign transferable polarisabilities by element or atom types. Element-wise polarisabilities would have an uncertainty of up to 0.77 Å^3, i.e. 50% of the magnitude of the polarisabilities themselves. This would induce errors of up to 36 kJ/mol in relative conformational induction energies and of up to 11 kJ/mol in ligand-binding energies. Likewise, polarisabilities assigned by the 27 Amber protein atom types would still have an uncertainty of up to 0.77 Å^3, and it would induce errors of up to 17 kJ/mol in relative energies and of up to 5 kJ/mol for ligand-binding energies (7 and 2 kJ/mol on average). We have also tried to design better groups of atom types, but this is very hard, in particular for aliphatic and aromatic carbon atoms, for which the range is up to 0.36 Å^3.

Therefore, we suggest that polarisabilities should be assigned the same was as for charges, i.e. atom wise. This suppressed the variation of the polarisabilities to 0.14 Å^3 on average, with a maximum of 0.46 Å^3. The average and maximum standard deviations are 0.01 and 0.07 Å^3. This remaining variation reflects the conformational dependence of the polarisabilities and it cannot be further suppressed unless the conformational dependence is explicitly modelled.

The variation is related to the size of the polarisabilities, with an average of 23%. The conformational dependence induces average and maximum errors of 2 and 5 kJ/mol for relative conformational energies, and of 1 and 4 kJ/mol for ligand-binding energies. Polarisabilities calculated in the same way for a different protein (the photosynthetic reaction centre) give similar results: 2 and 9 kJ/mol average and maximum error for relative conformational energies and 1 and 5 kJ/mol for ligand-binding energies.

On the other hand, the polarisabilities strongly depend on the basis sets used in the QM calculations. Clearly, the 6-31G* basis set is too small to give converged polarisabilities. Instead, at least the aug-cc-pVDZ (and preferably, the aug-cc-pVTZ) basis set should be used in the calculations. Fortunately, the atomic correction factors between the 6-31G* and aug-cc-pVTZ basis sets are transferable, so the results can be easily extrapolated from bulk calculations with the 6-31G* basis set. In the supplementary material, we present a set of such polarisabilities (xAvPol4), averaged over 70 molecular-dynamics snapshots for avidin and over two different proteins, and finally extrapolated to the aug-cc-pVTZ basis set. These are the best atomic polarisabilities obtained in this paper.

Acknowledgements

This investigation has been supported by grants from the Swedish research council (project 2010-5025) and by computer resources of Lunarc at Lund University. J.K. thanks the Danish Natural Science Research Council / The Danish Councils for Independent Research and the Lundbeck Foundation for financial support.

References

1 Ponder, J. W.; Case, D. A. *Adv. Protein Chem*. 2003, *66*, 27-85.
2 Bachrach, S.M. *Rev. Comp. Chem*. **1994**, *5*, 171-227
3 Sigfridsson, E.; Ryde, U. *J. Comp. Chem*. **1998**, *19*, 377-395.
4 Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem*. **1993**, *97*, 10269-10280.
5 Brooks, B. R.; Brooks, C. L.; MacKerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archonitis, G.; Bartels, C.; Boersch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazardis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Veneable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. *J. Comput. Chem*. **2009**, *30*, 1545-1614.
6 Gao, J.; Xia, X. *Science* **1992**, *258*, 631-635.
7 Gao, J.; Habibollazadeh, D.; Shao, L. *J. Phys. Chem*. **1995**, *99*, 16460-16467.
8 Gao, J.; Pavelites, J. J.; Habibollazadeh, D. *J. Phys. Chem*. **1996**, *100*, 2689-2697.
9 Xie, W.; Pu, J.; MacKerell, A. D.; Gao, J. *J. Chem. Theory Comput*. **2007**, *3*, 1878-1889.
10 Söderhjelm., P.; Ryde, U. *J. Phys. Chem. A* **2009**, *113*, 617-627.
11 Cieplak, P.; Dupradeau, F.-Y.; Duan, Y.; Wang, J. *J. Phys. Condens. Matter* **2009**, *21*, 333102.
12 Halgren, R.A.; Damm, W. *Curr. Opin. Struct. Biol*. **2001**, *11*, 236-242.
13 Gresh, N.; Cisneros, G.A.; Darden, T.D.; Piquemal, J.-P. *J. Chem. Theory Comput*. **2007**, *3*, 1960-1986.
14 Warshel A; Kato, M.; Pisliakov, A.V. *J. Chem. Theory Comput*. **2007**, *3*, 2034-2045.
15 Lopes, P. E. M.; Roux, B.; MacKerell, A. D. *Theor. Chem. Acc*. **2009**, *124*, 11-28.
16 Rappé, A. K.; Goddard, W. A., III. *J. Phys. Chem*. **1991**, *95*, 3358-3363.
17 Rick, S. W.; Stuart, S. J.; Berne, B. J. *J. Chem. Phys*. **1994**, *101*, 6141-6156.
18 Applequist, J.; Carl, J. R.; Fung, K.-K. *J. Am. Chem. Soc*. **1972**, *94*, 2952-2960.
19 Warshel, A.; Levitt, M. *J. Mol. Biol*. **1976**, *103*, 227-249.
20 Thole, B. T. *Chem. Phys*. **1981**, *59*, 341-350.
21 Dick, B. G.; Overhauser, A. W. *Physical Rev*. **1958**, *112*, 90-103.
22 Cao, J.; Berne, B. J. *J. Chem. Phys*. **1993**, *99*, 6998-7011.
23 Lamoureux, G.; Roux, B. *J. Chem. Phys*. **2003**, *119*, 3025-3039.
24 Day, P. N.; Jensen, J. H.; Gordon, M. S.; Webb, S. P.; Stevens, W. J.; Kraus, M.; Garmer, D.; Basch, H.; Cohen, D. *J. Chem. Phys*. **1996**, *105*, 1968-1986.
25 Engkvist, O.; Åstrand, P.-O.; Karlström, G. *Chem. Rev*. **2000**, *100*, 4087-4108.
26 Cieplak, P., Caldwell, J., Kollman, P. *J. Comput. Chem*. **2001**, *22*, 1048-1057.
27 Maple, J. R.; Cao, Y.; Damm, W.; Halgren, T. A.; Kaminski, G. A.; Zhang, L.Y.; Friesner. R. A. *J. Chem. Theory Comput.* **2005**, *1*, 694-715.
28 Ren, P.; Ponder, J. W. *J. Comput. Chem*. **2002**, *23*, 1497-1506.
29 Ángyán, J. G.; Jansen, G.; Loos, M.; Hättig, C.; Hess. B. A. *Chem. Phys. Lett*. **1994**, *219*, 267-273.
30 Stone, A. J. *Mol. Phys*. **1985**, *56*, 1065-1082.
31 Le Sueur, C. R.; Stone, A. J. *Mol. Phys*. **1994**, *83*, 293-307.
32 Dehez F.; Chipot, C.; Millot, C., Ángyán, J. G. *Chem. Phys. Lett*. **2001**, *338*, 180-188
33 Williams, G. J., Stone, A. J. *J. Chem. Phys*. **2003**, *119*, 4620-4628.
34 Gagliardi, L.; Lindh, R.; Karlström, G. *J. Chem. Phys*. **2004**, *121*, 4494-4500.
35 Stout, J. M.; Dykstra, C. E. *J. Am. Chem. Soc*. **1995**, *117*, 5127-5132.
36 Zhou, T.; Dykstra, C. E. *J. Phys. Chem. A* **2000**, *104*, 2204-2210.
37 Nakagawa, S.; Kosugi, N. P *Chem. Phys. Lett*. **1993**, *210*, 180-186.
38 Alkorta, I.; Bachs, M.; Perez, J. J. *Chem. Phys. Lett*. **1994**, *224*, 160-165.

39 Celebi, N.; Ángyán, J. G.; Dehez, F.; Millot, C.; Chipot, C. *J. Chem. Phys*. **2000**, *112*, 2709-2717.
40 Dehez, F.; Soetens, J. C.; Chipot, C.; Ángyán, J. G.; Millot, C. *J. Phys. Chem. A* **2000**, *104*, 1293-1303.
41 Soteras, I.; Curutchet, C.; Bidon-Chanal, A.; Dehez, F.; Ángyán, J. G.; Orozco, M.; Chipot, C.; Luque, F. J. *J. Chem. Theory. Comput*. **2007**, *3*, 1901-1913.
42 Elking, D.; Darden, T.; Woods, R. J. *J. Comput. Chem*. **2007**, *28*, 1261-1274.
43 Vogel, A. I. *J. Chem. Soc*. **1948**, 1833-1847.
44 Miller, K. J. *J. Am. Chem. Soc* **1990**, *112*, 8533-8542.
45 Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, J. S.; Karplus, M. *J. Comput. Chem*. **1983**, *4*, 187-217.
46 van Duijnen, P. T.; Swart, M. *J. Phys. Chem. A* **1998**, *102*, 2399-2407.
47 Söderhjelm, P; Husberg, C.; Strambi, A.; Olivucci, M.; Ryde, U. *J. Chem. Theory Comput*. **2009**, *5*, 649-658
48 Söderhjelm, P.; Kongsted, J.; Ryde, U. *J. Chem. Theory Comput*. **2010**, *6*, 1726-1737
49 Karlström, G.; Lindh, R. Malmqvist, P.-Å.; Roos, B. O.; Ryde, U.; Veryazov, V.; Widmark, P.-O.; Cossi, M.; Schimmelpfennig, B.; Neogrady, P.; Seijo, L. *Comput. Materials Sci*. **2003**, *28*, 222-239.
50 Söderhjelm, P.; Krogh, J. W.; Karlström, G.; Ryde, U.; Lindh, R. *J. Comput. Chem.* **2007**, *28*, 1083-1090
51 Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372-1377.
52 Hariharan, P. C.; Pople, J. A. *Theor. Chim. Acta* **1973**, *28*, 213-222.
53 Dunning, T. H., *J. Chem. Phys*. **1989**, *90*, 1007-1023.
54 Sadlej, A. J. *Collect. Czech. Chem. Commun*. **1988**, 53, 1995-2016.
55 Zhang, D. W.; Zhang, J. Z. H. *J. Chem. Phys*. **2003**, *119*, 3599-3605.
56 Weis, A.; Katebzadeh, K.; Söderhjelm, P.; Nilsson, I.; Ryde, U.; *J. Med. Chem.* **2006**, *49*, 6596-6606.
57 Case, D. A.; Darden, T. A.; Cheatham III, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Crowley, M.; Walker, R. C.; Zhang, W.; Merz, K. M.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossvary, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Kollman, P. A. Amber 10, University of California, San Francisco, **2008**.
58 Söderhjelm, P.; Ryde, U. *J. Comput. Chem.* **2009**, *30*, 750-760.
59 Genheden, S.; Söderhjelm, P.; Ryde, U. *Int. J. Quant. Chem.* **2011**, in press.
60 Giese, T. J.; York, D. M. *J. Chem. Phys.* **2004**, *120*, 9903-9906.
61 Morita, A. *J. Comput. Chem.* **2002**, *23*, 1466-1471.
62 Anisimov, V. M.; Lamoureux, G.; Vorobyov, I. V.; Huang, N.; Roux, B.; MacKerell, A. D. *J. Chem. Theory Comput.* **2005**, *1*, 153-168.
63 Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A. *J. Phys. Chem. A* **2004**, *108*, 621-627.

Table 1. Comparison of a 10 different sets of atomic polarisabilities.

Atom	Vogel	Applequist	Thole	Dykstra	Enzymix	Charmm	Amber02	Amoeba	PFF	Amber09
	43	18	20	35	19	45[a]	26[b]	28[c]	27[d]	11
HC alkyl			0.514	0.00	0.5	0.044	0.135	0.496	0.25	0.443
HC aromatic	0.407	0.135	0.514	0.00	0.5	0.10	0.167	0.800	0.39	0.443
HO alcohol	0.405	0.135	0.514	0.00	0.5	0.044	0.135	0.496	0.22	0.443
HN amides		0.161	0.514	0.00	0.5	0.044	0.161	0.496	0.24	0.443
HN amines			0.514	0.00	0.5	0.044	0.135	0.496	0.24	0.443
HN in RNH_3^+			0.514	0.00	0.5	0.044	0.135	0.496	0.24	0.443
C alkyl	1.027	0.878	1.405	1.87	1.0	0.98	0.878	1.334	1.22	0.920
C aromatic			1.405	1.61	1.0	2.07	0.360	1.334	1.49	1.298
C amide	1.027	0.616	1.405	1.88	1.0	1.65	0.616	1.334	0.83	1.298
C in COO^-			1.405	1.88	1.0	1.65	0.616	1.334	0.82	1.298
N amine			1.105	1.64	1.0	1.10	0.530	1.073	1.33	0.934
N aromatic			1.105	1.29	1.0	1.10	0.530	1.073	1.42	0.934
N amide		0.530	1.105	1.29	1.0	1.10	0.530	1.073	1.15	0.934
OH aliphatic alcohol	0.604	0.465	0.862	0.75	1.0	0.84	0.465	0.834	0.77	0.606
OH aromatic alcohol			0.862	0.75	1.0	0.84	0.465	0.873	0.77	0.593
O back-bone amide	0.841	0.434	0.862	0.25	1.0	0.84	0.434	0.837	0.91	0.593
O side-chain amide			0.862	0.25	1.0	0.84	0.434	0.834	0.91	0.593
O in COO^-			0.862	0.25	1.0	2.14	0.434	0.837	0.97	0.593
S					1.0	0.34	2.900	3.300	2.872	3.183

[a] Listed data for CHARMM are from an old but complete listing.[13] Newer developments for alcohols, alkanes, and amides[7,8,9] have either used slightly modified Applequist parameters[18] or the Thole parameters.[20]

[b] Data from the parm99.dat file in the Amber10 distribution.

[c] Data from the amoebapro.prm files in the Amber10 distribution.

[d] Data from Table 8 in ref. no. 27.

Table 2. Polarisabilities calculated for each element in the 70 snapshots of avidin (only protein atoms). # is the number of individual polarisabilities obtained for each element. Aver, Stdev, Min, Max, and Range are the average, standard deviation, minimum and maximum values for each element. Range is Max – Min. For comparison, the Min and Max values of the Amber-02 polarisabilities are also included.

		LoProp					Amber-02	
Element	#	Aver	Stdev	Min	Max	Range	Min	Max
H	267820	0.22	0.04	0.05	0.33	0.27	0.14	0.17
C	169400	1.13	0.13	0.82	1.59	0.77	0.36	0.88
N	48160	0.91	0.13	0.49	1.24	0.75	0.53	0.53
O	53060	0.54	0.03	0.41	0.68	0.27	0.43	0.47
S	1120	2.16	0.13	1.88	2.45	0.57	2.90	2.90

Table 3. Statistics for LoProp polarisabilities over the Amber-02 atom types that have distinct polarisabilities. The columns have the same meaning as in Table 2. The atom types are explained in Table 4.

Atom type	#	Aver	Stdev	Min	Max	Range	Amber
C	41020	1.15	0.06	0.94	1.39	0.45	0.62
CT	103040	1.12	0.15	0.82	1.59	0.77	0.88
C other	25340	1.17	0.09	0.88	1.52	0.64	0.36
H	59920	0.17	0.02	0.05	0.23	0.18	0.16
HA, H4, H5	17080	0.28	0.02	0.23	0.33	0.10	0.17
H other	190820	0.24	0.03	0.09	0.32	0.23	0.14
N	48160	0.91	0.13	0.49	1.24	0.75	0.53
O, O2	44380	0.54	0.03	0.41	0.68	0.27	0.43
OH	8680	0.53	0.03	0.45	0.68	0.23	0.47
S	1120	2.16	0.13	1.88	2.45	0.57	2.90

Table 4. Statistics for the LoProp polarisabilities over all the Amber-02 atom types for proteins. The columns have the same meaning as in Table 2.

Atom type	#	Aver	Stdev	Min	Max	Range	Amber	Description
C	41020	1.15	0.06	0.94	1.39	0.45	0.62	sp^2 C in carbonyl groups
CA	20020	1.15	0.08	0.98	1.51	0.53	0.36	aromatic C
CB	1120	1.25	0.06	1.01	1.47	0.46	0.36	CD2 in Trp
CC	280	1.19	0.03	1.11	1.26	0.15	0.36	CG in His
CN	1120	1.27	0.05	1.10	1.41	0.31	0.36	CE2 in Trp
CR	280	0.96	0.03	0.89	1.03	0.14	0.36	CE1 in His
CT	103040	1.12	0.15	0.82	1.59	0.77	0.88	sp^3 aliphatic C
CV	280	0.97	0.04	0.88	1.07	0.19	0.36	C2D in Hid
CW	1120	1.15	0.04	1.02	1.25	0.23	0.36	CD2 in Hie and Hip, CD1 in Trp
C*	1120	1.34	0.05	1.19	1.52	0.34	0.36	CG in Trp
H	59920	0.17	0.02	0.05	0.23	0.18	0.16	H bound to N
H1	57540	0.23	0.03	0.15	0.30	0.15	0.14	aliphatic H bound to C with one electron-withdrawing group
H4	1400	0.28	0.02	0.23	0.30	0.07	0.17	HD1 in Trp, HD2 in Hid
H5	280	0.29	0.01	0.28	0.30	0.03	0.17	HE1 in Hid
HA	15400	0.28	0.02	0.23	0.33	0.10	0.17	aromatic H
HC	119840	0.25	0.02	0.18	0.32	0.13	0.14	aliphatic H bound to C without electron-withdrawing groups
HO	8680	0.16	0.02	0.09	0.21	0.12	0.14	H in hydroxyl groups
HP	4760	0.22	0.01	0.17	0.27	0.09	0.14	HE in Lys
N	37660	0.96	0.09	0.64	1.24	0.61	0.53	sp^2 N in amide groups
N2	6300	0.74	0.12	0.57	1.02	0.44	0.53	NE and NH in Arg
N3	2520	0.64	0.02	0.49	0.70	0.21	0.53	NZ in Lys
NA	1400	0.94	0.07	0.76	1.17	0.41	0.53	protonated N in aromatic rings
NB	280	0.87	0.04	0.79	0.97	0.18	0.53	non-protonated N in aromatic rings
O	37660	0.54	0.03	0.41	0.64	0.23	0.43	O in carbonyl groups
O2	6720	0.58	0.04	0.42	0.68	0.26	0.43	O in carboxyl groups
OH	8680	0.53	0.03	0.45	0.68	0.23	0.47	O in hydroxyl group
S	1120	2.16	0.13	1.88	2.45	0.57	2.90	S

Table 5. Description of the various sets of polarisabilities considered in the work.

Charge set	# Distinct polarisabilities	Description	Polarisabilities different for Snapshots	Same residue	Based on protein
LoProp	547 880	LoProp atomic polarisabilities	Yes	Yes	Avidin
Aver	7916	Average over snapshots	No	Yes	Avidin
xAvPol1	459	Aver, averaged over residues	No	No	Avidin
xAvPol2	309	Like xAvPol1 but from PRC	No	No	PRC
xAvPol3	521	Weighted average over xAvPol1 and xAvPol2	No	No	Avidin, PRC
xAvPol4	395	xAvPol3 corrected to aug-cc-pVTZ basis	No	No	Avidin, PRC
Element	5	LoProp averaged over elements (Table 2)	No	No	
Type	27	LoProp averaged over atom types (Table 4)	No	No	
Amber02	10	Amber FF02 polarisabilities[26]	No	No	
Amber09	7	New Amber polarisabilities[11]	No	No	
Charmm	9	CHARMM polarisabilities[45]	No	No	
Amoeba	8	Amoeba polarisabilities[28]	No	No	
Enzymix	2	Enzymix polarisabilities[19]	No	No	

.

Table 6. Differences in relative polarisation energies relative to Aver.

	xAvPol1	xAvPol2	xAvPol3	xAvPol4	Element Type	Amber02	Amber09	Charmm	Amoeba	Enzymix
MAD	2	2	2	27	9	7	26	35	36	58	65
Min	-5	-10	-9	-72	-36	-12	-69	-97	-133	-142	-151
Max	4	6	5	61	22	17	53	91	99	166	181
Range	9	15	14	134	57	29	122	188	232	308	333
MAD' [a]							42	15	34	35	42

[a] Mean absolute deviation from the xAvPol4 results.

Table 7. Differences in ligand-interaction polarisation energies, compared to LoProp. Mean absolute (MAD) and maximum differences (Max) compared to those obtained with the LoProp polarisabilities are listed, calculated either over all seven ligands, or over the charged (1–3) or the neutral ligands (4–7).

	Aver	xAvPol1	xAvPol2	xAvPol3	xAvPol4	Element Type		Amber02	Amber09	Charmm	Amoeba	Enzymix
MAD	0.6	0.8	1.1	1.1	12.6	4.1	1.5	4.8	17.2	12.4	24.9	25.7
Max	2.6	3.6	4.8	4.6	30.5	10.8	5.4	11.9	39.3	36.3	54.2	61.1
MAD1-3	1.0	1.4	1.8	1.7	19.4	7.8	2.1	4.2	28.7	24.6	36.9	40.6
Max1-3	2.6	3.6	4.8	4.6	30.5	10.8	5.4	10.6	39.3	36.3	54.2	61.1
MAD4-7	0.3	0.4	0.6	0.6	5.8	1.3	1.0	5.2	8.6	3.4	15.9	14.5
Max4-7	0.9	1.4	2.2	2.0	12.9	4.6	3.2	11.9	23.1	8.1	37.1	33.2
MAD'[a]								12.7	6.9	5.2	14.5	15.7
Max'[a]								32.7	14.9	12.7	24.7	30.6

[a] Deviations from the xAvPol4 results (only for Btn1 and Btn7).

Table 8. Differences in ligand-interaction polarisation energies, compared to LoProp, with explicit solvent. Mean absolute (MAD) and maximum differences (Max) compared to those obtained with the LoProp polarisabilities are listed, calculated either over all seven ligands, or over the charged (1–3) or the neutral ligands (4–7).

	Aver	xAvPol1	xAvPol2	xAvPol3	xAvPol4	Element Type	Amber02	Amber09	Charmm	Amoeba	Enzymix
Force-field specific water polarisabilities											
MAD	0.6	0.9	1.0	0.9	23.9	4.2	1.5	5.3	38.5	17.8	23.1	47.9
Max	3.2	4.2	4.9	4.8	54.4	19.6	4.8	13.8	171.4	47.9	94.3	245.5
MAD1-3	1.0	1.5	1.5	1.5	41.2	7.0	1.9	7.4	59.2	36.6	35.7	77.6
Max1-3	3.2	4.2	4.9	4.8	54.4	19.6	4.8	13.8	171.4	47.9	94.3	245.5
MAD4-7	0.3	0.4	0.6	0.5	6.7	2.0	1.1	3.8	23.0	3.3	13.7	25.6
Max4-7	0.9	1.3	1.8	1.7	13.5	5.3	3.3	11.1	47.3	9.0	27.7	63.0
LoProp water polarisabilities											
MAD	0.6	0.9	1.0	0.9	15.7	3.6	1.5	4.8	14.5	27.6	22.0	25.7
Max	3.2	4.2	4.9	4.8	32.8	8.3	4.8	13.5	36.6	82.5	55.3	60.3
MAD1-3	1.0	1.5	1.5	1.5	25.4	6.2	1.9	5.6	22.2	58.7	31.5	40.6
Max1-3	3.2	4.2	4.9	4.8	32.8	8.3	4.8	13.5	36.6	82.5	55.3	60.3
MAD4-7	0.3	0.4	0.6	0.5	6.0	1.6	1.1	4.2	8.7	3.6	14.9	14.6
Max4-7	0.9	1.3	1.8	1.7	12.8	4.8	3.3	11.4	23.2	9.0	36.2	33.1

Figure 1. The seven biotin analogues used in this study. a) Btn1 (biotin), b) – g) Btn2–Btn7.

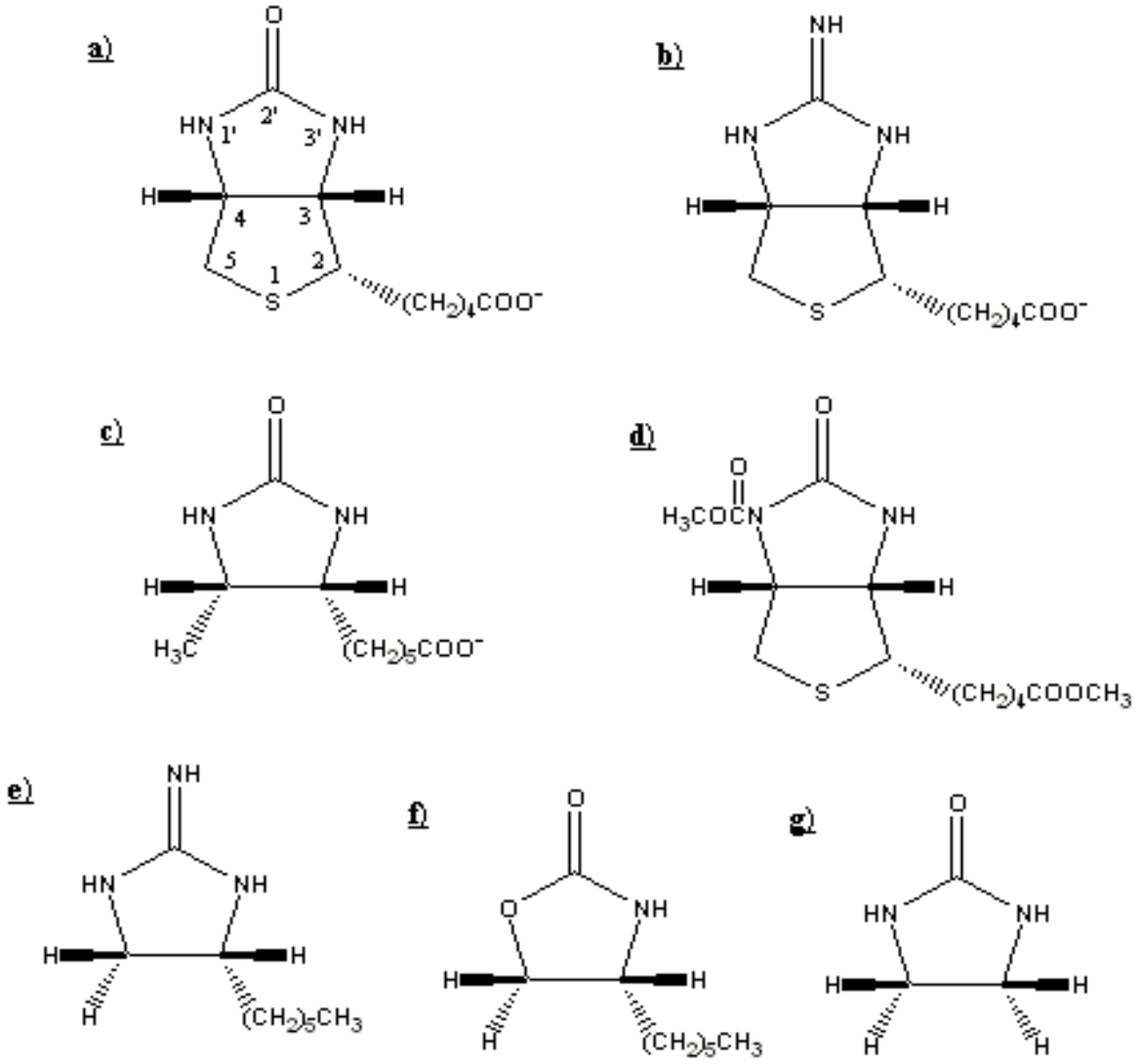

Figure 2. Frequency plot for the LoProp polarisabilities (Å^3) of the CA atom in Gly and Asp in avidin (3080 and 1400 individual polarisabilities, respectively).

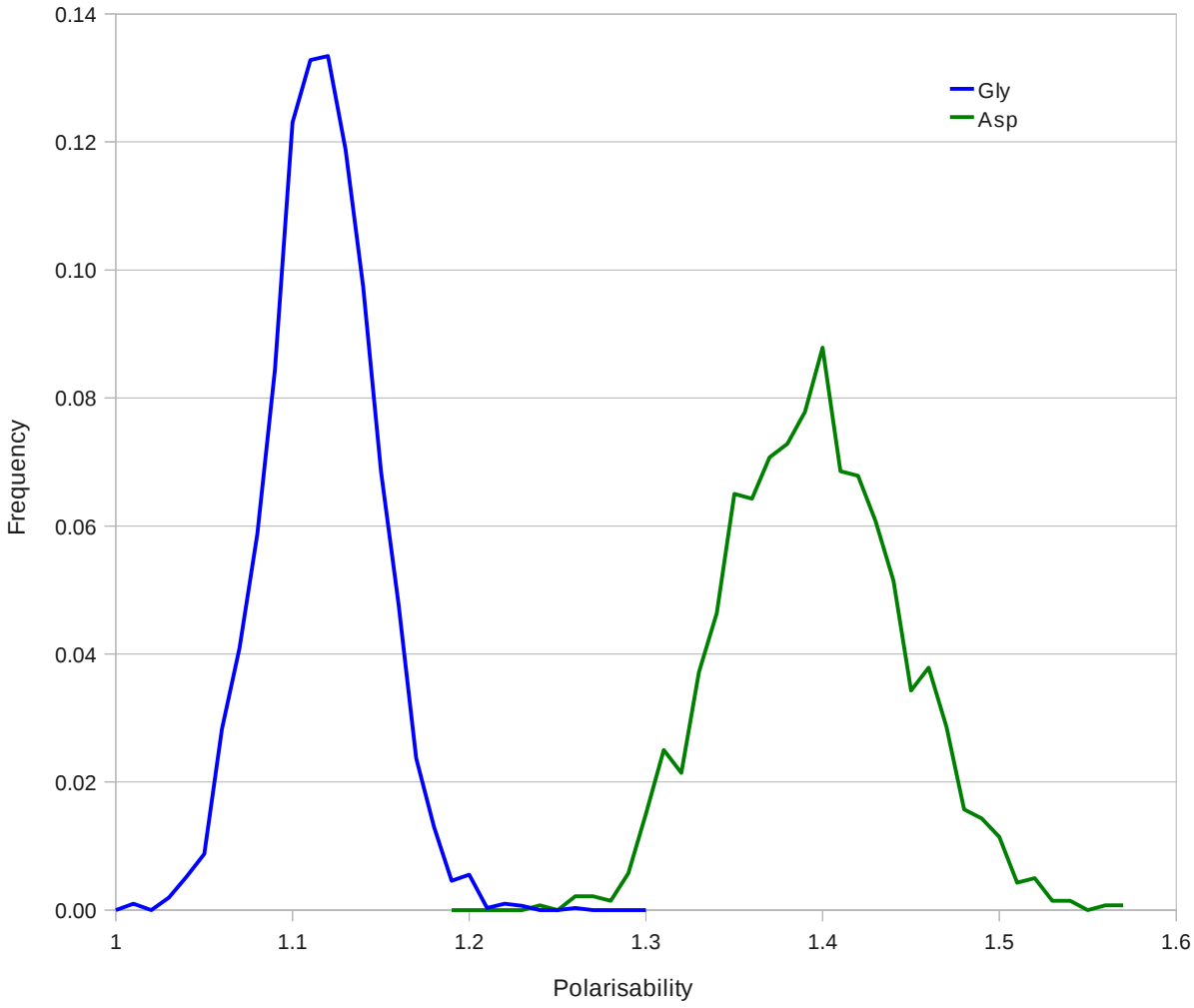

Figure 3. Correlation between the average size of the LoProp atomic polarisabilities and their range (both in units of Å^3). The points are coded after the element: (H – green squares, C – black diamonds, N – blue lying triangles, O – red standing triangles, and S – yellow double triangles).

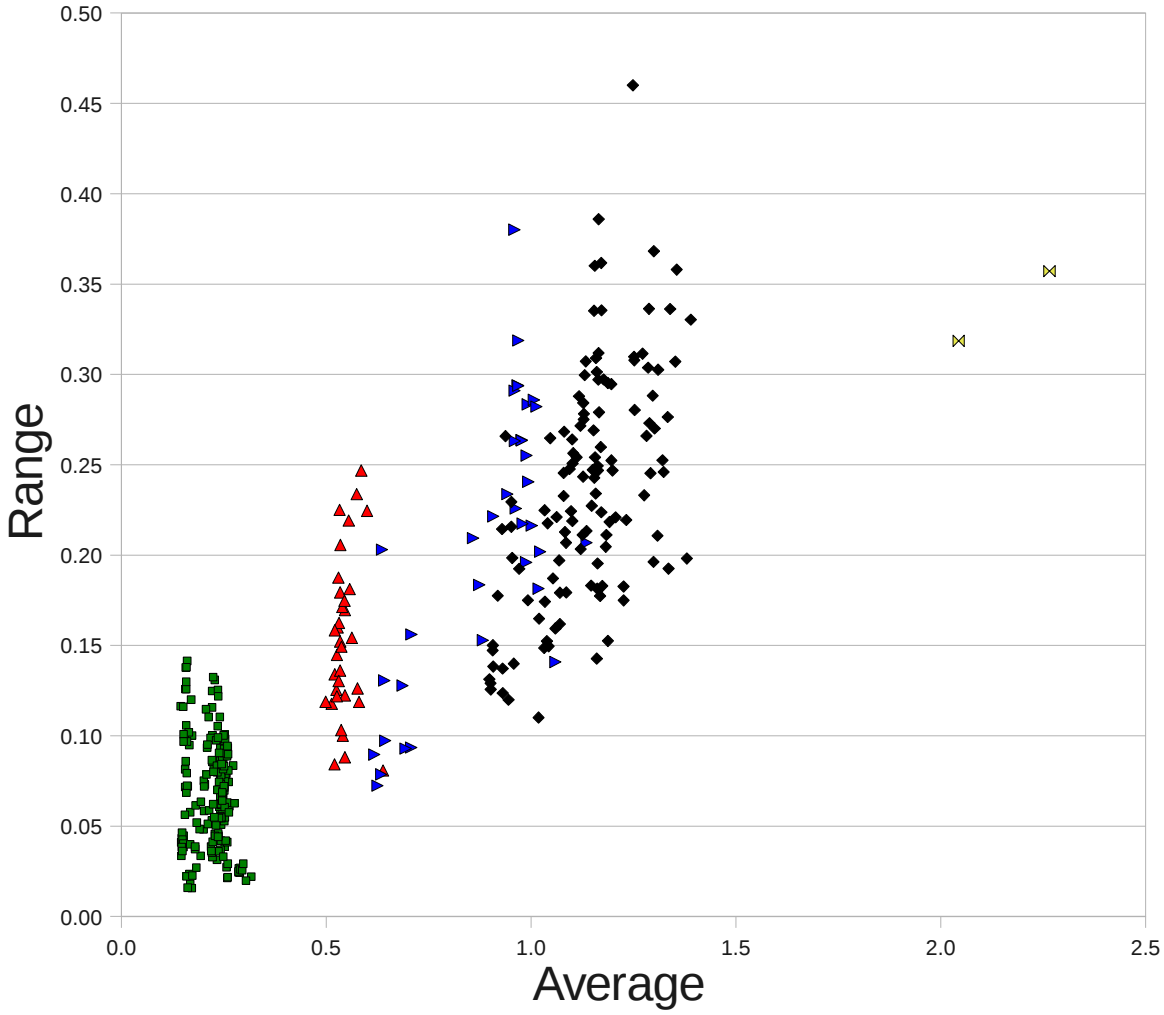

Figure 4. Comparison between the atomic LoProp and the Amber polarisabilities (both in units of Å^3). The points are coded after the element: (H – green squares, C – black diamonds, N – blue lying triangles, O – red standing triangles, and S – yellow double triangles). The line is $x = y$.