



LUND UNIVERSITY

Restrained point-charge models for disaccharides

Sigfridsson, E; Ryde, Ulf; Bush, BL

Published in:
Journal of Computational Chemistry

DOI:
[10.1002/jcc.10024](https://doi.org/10.1002/jcc.10024)

2002

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

[Link to publication](#)

Citation for published version (APA):
Sigfridsson, E., Ryde, U., & Bush, BL. (2002). Restrained point-charge models for disaccharides. *Journal of Computational Chemistry*, 23(3), 351-364. <https://doi.org/10.1002/jcc.10024>

Total number of authors:
3

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

Restrained point-charge models for disaccharides

Emma Sigfridsson¹, Ulf Ryde^{1*}, and Bruce L. Bush²

¹Department of Theoretical Chemistry,
Lund University
Chemical Centre, P.O.B 124
S-221 00 Lund, Sweden

²Merck Research Laboratories
Molecular Systems Department
Rahway NJ 07065
New Jersey, USA

* Correspondence to
Ulf.Ryde@teokem.lu.se
Tel: 46-46-2224502
Fax: 46-46-2224543

2017-04-01

Key words: atomic charges, chemical transferability, conformational transferability, restrained electrostatic potential charges, rank-deficiency problems.

Abstract

Various methods for deriving atomic partial charges from the quantum chemical electrostatic potential and moments have been tested for the sucrose molecule. We show that if no further information is used, the charges on some carbon atoms become large and charge patterns involving these atoms are badly determined and poorly transferable. Adding lone-pairs on the ether oxygen atoms or dividing the molecule into smaller fragments did not cure the instabilities. We develop a method, CHELP-BOW0, that restrains charges toward zero with different weights for different atoms. These harmonic restraints preserve the linear form of the least-squares equations, which are solved in a single step using singular-value decomposition. CHELP-BOW0 improves the chemical transferability of the charges compared to unrestrained methods, and slightly improves their conformational transferability. It introduces a modest degradation of the fit compared to unrestrained CHELP-BOW (mean average deviation of the potential 0.00016 v.s. 0.00010 a.u.). A second new method, CHELP-BOWC, avoids the need for restraints by including several conformations in the fit, weighting each according to its estimated energy in solution. CHELP-BOWC charges are more transferable than CHELP-BOW or CHELP-BOW0 charges to conformations not included in the training set. Restraints to zero charge do not further improve transferability of the CHELP-BOWC charges. We therefore recommend CHELP-BOW charges for rigid molecules and CHELP-BOWC charges for flexible molecules.

Introduction

Theoretical simulations using molecular mechanics, molecular dynamics, and Monte Carlo methods have become an important tool in many areas of chemistry. In all such methods, a proper description of the electrostatic interactions in the system is of key importance. They can be treated at various levels of sophistication in the simulations. However, in most widely used force fields, the simplest possible representation is employed, an atom-centred point-charge model. Thus, most classical simulation methods require a set of point charges for all molecules of interest.

Atomic charges are parameters, not observables that can be measured in experiments or determined unambiguously by quantum chemical calculations. Therefore, a wide variety of methods have been suggested for the estimation of point-charge parameters [1,2]. At present, the most widely used methods derive point charges from a least-squares fit to the electrostatic potential (ESP) calculated by quantum chemical methods in a large number of points around the molecule of interest. There are several examples of this kind of methods, such as CHELP [3], CHELPG [4], RESP [5], and the Merz-Kollman scheme [6,7]. They differ mainly in how they select the points where the electrostatic potential is calculated [2] (e.g. the density of the points, the geometry, and the outer and inner exclusion radii, outside which no points are sampled). For some systems, e.g. transition-metal complexes, different methods give rather different charges [2].

In an attempt to avoid this arbitrariness, we recently developed the CHELP-BOW method [2], where the potential points in principle can be selected by any mechanism and at any distance from the molecule (provided that the point density is high enough) because they are weighted by a Boltzmann factor taken from the force-field in which the charges will be used. The philosophy behind this procedure is that the charges should be as accurate as possible in terms of the total interaction energy in actual simulations. Therefore, points within the Van der Waals envelope of the molecule, which are rarely encountered in actual simulations should be weighted down, whereas other points should have full weight. Moreover, we emphasise the importance of reproducing the potential far from the molecule, because far interactions vastly outnumber the close interaction and therefore are more important in terms of total energy [2]. To avoid including ESP points to infinite distance, the

charges are constrained to reproduce the total charge, dipole moment, and possibly also the quadrupole moment, which determine the electrostatics of the molecule at large distances. Charges calculated with this method have been shown to reproduce quantum chemical moments and potentials well [2].

In all ESP methods, charges are determined by a least-squares fit to the potential. Such a fit does not always behave properly [8,9]. Even if the number of potential points is much greater than the number of charges to be fitted, it is not certain that the system is overdetermined. On the contrary, potentials at neighbouring points are strongly correlated. Furthermore, the sampled potentials may contain little information about certain patterns of buried charges. If this is the case, the least-squares system may even be underdetermined (rank-deficient), yielding nonsense charges. For example, Francl et al. showed that the least-squares matrices for 16 typical small and medium-sized molecules were rank-deficient [8]. The CHELP-BOW method avoids this by using an improved least-squares procedure [10], which employs pseudo-inverses calculated by singular-value decomposition (which gives a full control of the rank of the least-squares matrix) [2]. Random placement of the sampling points further reduces the near-exact symmetries which give rise to near-zero singular values. Using this method, no rank-deficiency problem was encountered even for molecules with up to 84 atoms.

In this article we study various point-charge models for sucrose, as a typical example of the class of disaccharides. Francl et al. have recently studied also this system and suggested that only 24 of the 45 atom charges can be unambiguously determined [9]. We would prefer to say that only 24 of the 45 *singular components* are well-determined, to make clear that *all* the fitted atomic charge parameters depend on each other. More than 21 - indeed, all 45 atoms - may be involved in the 21 poorly-determined singular components. Conversely, a single charge cannot be poorly determined, because the fit is constrained to reproduce the total molecular charge. If the fit is further constrained to reproduce the molecular dipole moment vector, then any instability must involve at least five atoms

In this study, we show that CHELP-BOW charges for sucrose and several related molecules can be calculated without any rank-deficiency problems. However, the resulting charges are neither chemically intuitive nor transferable to the isolated glucose and fructose molecules (the monomer constituents of sucrose). Through studies of fragment molecules we

show that the problem is not inherent in the chemical substructures. Rather, in the large disaccharide, certain patterns of charge are buried or, more precisely, generate small potential changes in the sampling region and are therefore poorly determined. We discuss how the CHELP-BOW procedure can be cured of this problem and study the chemical and conformational transferability of various types of charges.

Methods

Quantum chemical calculations

All quantum chemical calculations were performed at the ab initio Hartree-Fock (HF) level with the 6-31G* basis set [11] using the Gaussian-94 software [12]. The geometry of the sucrose molecule was fully optimised at this level. It was then divided into its fructose and glucose constituents (or other fragments) without further optimisation of the geometry. This way, any influence on the charges from the geometry of the molecules is avoided. The fragments were saturated with hydrogen atoms at standard distances. The sucrose molecule and its atom labels are shown in Figure 1. Primed numbers refer to the fructose moiety (and are used also in the isolated fructose molecule). All calculations were performed on SGI Octane or Origin 2000 workstations.

CHELP-BOW fits

The CHELP-BOW procedure (charges from Boltzmann-weighted electrostatic potentials) has been described and tested in detail before [2]. Here, we only give a short overview for the reader's convenience. The electrostatic potential was calculated for 2000 points/atom (90 000 points for sucrose), sampled at random out to a maximum distance of 8 Å from any atom. Charges on each atom were estimated by a weighted least-squares fit to these potentials. The least-squares equations were solved using pseudo-inverses calculated with the stable singular-value composition method [10]. The charges were constrained to reproduce the total charge and the calculated three components of the dipole moment for the molecule. In addition, the charges on hydrogens bonded to the same heavy atom (H6, H1', and H6') were constrained to be equal. In some calculations, further constraints or restraints were introduced, as is described in detail below.

In the fit, the potentials were weighted by a Boltzmann factor,

$$w = \exp(-E / R T) \quad (1)$$

where R is the gas constant, T is the temperature (298 K), and E is the Lennard-Jones interaction energy between the molecule and a probe molecule (a TIP3P water) at the potential point. Energy is independent of probe orientation because only the oxygen atom of this molecule participates in the steric Lennard-Jones interaction. We consider only the repulsive part of the Lennard-Jones potential and use potential parameters of the Amber force field [13].

Criteria for the quality of charge sets

The criteria used to compare charges obtained with different methods are of great importance [2]. As was discussed in the introduction, we aim at reproducing the total interaction energy in actual simulations as accurately as possible. Therefore, our main criterion of accuracy (labelled "m.a." in the tables) is the Boltzmann weighted mean absolute deviation from the potential calculated by quantum chemistry. The potential multiplied by a probe charge gives directly the electrostatic interaction energy; the Boltzmann weight accounts for the likelihood of the encounter. For convenience, we also list another commonly used measure of how well the quantum chemical potential is reproduced, the Boltzmann weighted root-mean-squared (r.m.s) deviation.

We will also list the average absolute difference of all spherical harmonic terms for each of the four lowest multipole moments (dipole, quadrupole, octupole, and hexadecapole, the latter three abbreviated 4-, 8-, and 16-pole in the tables). This gives a good estimate of how well the electrostatics is reproduced, especially far from the molecule. Charges and potentials are given in atomic units (a.u.), whereas multipole moments are in D (dipole), DÅ (quadrupole), etc., as customary.

Interaction energy with neutral molecules will reflect some cancellation of deviations in potential if the deviation varies slowly in space. We therefore report energies of electrostatic interaction 3681 water molecules in a 30-Å sphere around sucrose. Positions were obtained after a 160-ps simulation molecular dynamics simulation at 300 K, using the CHELP-BOWC charges (see below) for sucrose and the Amber force field for all other parameters [13]. Interaction energies are based on TIP3P charges for water (+0.417 e on hydrogen). Reference

values of interaction energies were calculated directly from the quantum chemical potential calculated at the atomic positions of each water molecule.

Conformational and chemical transferability

Transferability is another important quality. Charges should be transferable both chemically (i.e. charges obtained for one compound can be used to generate charges for corresponding atoms of a similar compound) and geometrically (i.e. charges obtained for one conformation of the molecule are useful also for other conformations of the same molecule). The latter criterion is extremely important in conformational search, molecular dynamics, or Monte Carlo simulations, for example.

We made three tests of chemical transferability of charges between sucrose and its constituents glucose and fructose, and vice versa. First, we tested stability by comparing the charges obtained for sucrose with those of its glucose and fructose constituents, at the same geometry. Atoms O1/O2', HO1 and HO2' were not included in these comparisons, since their chemical functions differ.

Second, the charges from glucose and fructose were used directly in sucrose to see how well they reproduce the calculated electric moments and potentials. The charge on O1 was determined to make the sucrose charge neutral.

Third, the charges from sucrose were used directly in glucose or fructose to test if they reproduce the calculated electric moments and potentials. The charges on O1 and HO1 or O2' and HO2' were determined by a normal CHELP-BOW fit to the respective target molecule with only the total charge as constraint. This procedure is lenient, as it uses a small amount of information from quantum calculations on the target molecules.

Conformational transferability was tested on 23 conformations of sucrose. First, we added a random displacement with a norm of up to 0.05 or 0.10 Å to each atom. Second, we ran a 1-ns molecular dynamics simulation of sucrose in vacuum at 400 K. Conformations were sampled each 0.1 ns and five of these were optimised by quantum chemistry. Five further conformations were obtained by manually changing the internal hydrogen-bond

interactions in the original sucrose molecule

Fitting to multiple conformations: the CHELP-BOWC procedure

Finally, we modified the CHELP-BOW method to incorporate several different conformations of a molecule simultaneously in the fit. One set of charges was fitted to Boltzmann-weighted potentials of several conformations using the total charge and equal charges on equivalent atoms as the only constraints. The dipole moment was not constrained, since it differs for the various conformations. Each conformation was globally weighted using Eqn. (1), but with E corresponding to its total energy in water solution, as estimated by the conductor polarised continuum model (CPCM) in Gaussian-98 [14]. This method will be denoted CHELP-BOWC below.

RESP fits

The calculated charges were compared with charges obtained by the restrained electrostatic potential model (RESP). The RESP charges were calculated with the standard two-stage fit using the Amber 5.0 software [5,15]. To minimise differences caused by the sampling scheme of the ESP points, we used the same points for the CHELP-BOW and RESP fits. However, since RESP cannot weight the points with a Boltzmann factor, we employed for RESP only points with a weight higher than 0.1 (~99 % of the points). Test calculations showed that this gave slightly better charges than if all points were included and considerably better charges than if only ESP points with a full Boltzmann weight were used.

The default restraint weights used with RESP were calibrated to the Merz-Kollman sampling of ESP points. They may not be optimal for CHELP-BOW sampling. However, the differences are rather small. Restraints affect predominantly the calculated potentials close to the molecular surface, especially when the fit is constrained to reproduce the molecular dipole moment. CHELP-BOW sampling within 2.4 Å of sucrose includes 3800 points with total weight 1738.3, or weight 4.5 per Å². This is rather similar to 4 shells of one point per Å² as was used in the original RESP paper [5].

It is also possible to fit RESP charges to several conformations. Such fits are denoted

RESP-C below and employ the same conformations and Boltzmann weights as those used for the CHELP-BOWC charges. The ESP points were sampled as above and a two-stage RESP fit was performed as usual.

To test the influence of the hyperbolic restraints, we obtained one set of charges by turning off these restraints by setting the hyperbolic weight factor (qwt) to zero. This method is denoted ESP-C below, since it is a normal ESP fit (without any restraints) to several conformations. It differs from the CHELP-BOWC method only in that it does not Boltzmann weight the potentials.

Result and discussion

Variability in sucrose charges

We started by calculating charges for sucrose, glucose, and fructose using the standard CHELP-BOW method with the results shown in Table I. The large charges on C1 ($-1.11 e$) and C2 ($1.08 e$) in sucrose indicate instability in the CHELP-BOW method (rather than a physical displacement of the electron density toward C1). Furthermore, atom charges for the glucose moiety of sucrose differ strongly from those of the isolated glucose molecule, and likewise for fructose; the average and maximum differences are 0.26 and $1.29e$ (C1) for glucose and 0.06 and $0.20 e$ (C5') for fructose, respectively. The deviation is most pronounced for the carbon atoms (average difference $0.67 e$ for glucose, compared to 0.12 and $0.10 e$ for oxygen and hydrogen atoms, respectively). Closer examination of the sucrose charges reveals a strong pattern of alternating charge around the glucose ring. This confirms our observation that unstable patterns must involve several atoms, forming groups with cancelling monopole (charge) and dipole moments. Since glucose and fructose have the same geometry as in sucrose, these differences do not result from conformational variation, nor do they appear to have a chemical origin such as electron resonance. We conclude that the discrepancies result mainly from the instability in fitting the sucrose charges to electrostatic potentials, rather than to large shifts of electron density within the molecule.

To investigate the problem, we compared charges obtained by five fits to different sets of potentials (90 000 points chosen at random in each set) for sucrose. The result in Table II shows an appreciable variation in the charges. The average and maximum difference between the various sets are 0.01 and 0.10 e (for the C1 atom). At such a high point density, the charges would be expected to be converged within 0.01 e [16].

Even if the five charge sets are numerically different, they fit the various potential sets equally well. For example, all five charge sets give the same m.a. and r.m.s. deviation to the first ESP point set (0.00010, and 0.00027 a.u., respectively) and from Table II it can be seen that they also give similar deviations from the quantum chemical moments. Thus, these five sets of charges are equivalent (i.e. of the same quality). This shows that the charges are not over-fitted in the meaning that they reproduce the selected potential points well, but not neighbouring points omitted from the fit. The result does not change if we constrain the charges to reproduce the quadrupole, octupole, and even the hexadecapole moments.

Similar, but larger, differences in the charges were obtained if a small random displacement was added to each atom in sucrose before calculation of new ESP points and CHELP-BOW charges. With a random displacement of up to 0.05 Å for each atom, the average and maximum deviations were 0.03 and 0.16 e (Table II). Twice as large displacement gave twice as large difference. Again, the maximum change was seen for the C1 atom.

Together these results show that the least-squares fit for sucrose is inherently unstable, i.e. that small perturbations in the structure or in the sampling of potential points give rise to a large variations in the fitted charges. This also means that the problem cannot be solved by brute force methods, e.g. by increasing the number of potential points.

Interestingly, in none of these calculations did the singular-value decomposition indicate any rank deficiency of the least-squares matrix; the smallest condition number was 0.0032, which is 300 times larger than the normal criterion for rank deficiency (10^{-5}) [8]. Since the condition number depends entirely on the geometry of atom and potential points [17] and not on any electronic effects, it appears that the dense random sampling of potential points prevents rank deficiency without insuring stability of fitted charges [5].

It is notable that the problem of instability is not conspicuous for glucose or fructose.

Either the problem is inherent in the glycosidic linkage, or it results indirectly from inadequate potential sampling, i.e. from the burial of some atoms of sucrose.

Fragment molecules

The only way to sample the potential closer to the buried atoms, while retaining the CHELP-BOW philosophy, is to divide the molecule of interest into smaller fragments, where the problematic atoms are less buried. The resulting charges of the fragment molecules could then be used as constraints or restraints for the corresponding atoms in the charge fit of the full molecule. We tested four fragment molecules, shown in Figure 2, as models of the buried glycosidic linkage. Fragments 4 and 1 were chosen as a minimal and a realistic model of the glycosidic linkage, respectively, whereas fragments 2 and 3 represent the glucose and fructose parts of the linkage.

The CHELP-BOW charges of these four molecules are shown in Table III. It can be seen that for the fructose atoms, fragments 1 and 3 give rather similar results (up to 0.08 e difference for C1'). However for C1 and O1, the discrepancy is larger (0.21 e for C1), indicating that these atoms in fragment 1 are still involved in buried charge patterns. A large discrepancy for the charges of C2' and O5' between fructose and fragment 3 (0.21-0.22 e) indicates that the original fructose charges might not be so well-determined as they originally appeared to be.

To transfer information from the better-determined linker fragments to the larger molecules, we fitted charges to sucrose, glucose, and fructose with five different sets of constraints to the fragment molecule charges, viz. the charges of each of the fragments 1-4 alone, or a composite of fragments 2-4 (C1, H1, and O5 from fragment 2, C1', O1', H1', HO1', C2', and O5' from fragment 3, and O1 from fragment 4). In no case did we use charges from the methyl groups of the fragments. As seen in Table IV, constraints to the fragment molecule charges do not improve the sucrose charges significantly; the maximum deviation between the sucrose and glucose charges for the same atoms are still very large, 0.51-1.35 e .

Lone-pair centres

An alternative explanation for unstable charges around the glycosidic bond of sucrose is

that each ether oxygen atom (O5, O2', and O5') has two lone-pair orbitals, which are not represented in the fit. The neighbouring atoms might try to compensate for these, yielding patterns of charges with large magnitude and opposing signs. If this is correct, introduction of lone-pair centres on the ether oxygen atoms would improve the fit and might allow more reasonable charge separation along the bonds involving adjacent carbons.

Lone-pair orbitals were modelled as two extra charge centres placed 0.65 Å from each of the ether oxygen atoms at an angle of 109.5° between them. The two lone-pairs on each atom were constrained to have the same charge.

The charges of sucrose, fructose, and glucose obtained with such a fit involving lone-pairs are shown in Table I. They are not more chemically reasonable, nor do they show any improved transferability. The maximum and average deviation between glucose and sucrose do not change (0.24 and 1.29 e for C1) and between fructose and sucrose the deviations increase slightly compared to the values obtained without lone-pairs (0.12 and 0.92 e for O5'). The ether oxygen atoms in sucrose, O1 and O5', exhibit large charges (-1.45 and -1.27 e); worse, the charges on the lone-pair sites of O1 and O5' are positive (0.36 and 0.31 e), whereas those of O5 are negative (-0.30 e). Clearly, a lone-pair orbital is expected to be negative.

Consequently, we can conclude that adding additional charge centres (such as the lone pairs) can never correct the instability problem; on the contrary it will worsen it (i.e. the condition number will always decrease if the number of parameters is increased), as was seen for O1 and O5' above. However, since the lone-pair charges provides some additional degrees of freedom, they give an improved fit (which can be seen from the slightly improved potentials and moments shown in Table V).

Restraining charges to zero: The RESP and CHELP-BOW0 methods

The simplest way to suppress instabilities, using essentially no chemical information, is to restrain charges to zero [5]. The justification for this procedure is that the poorly determined singular components of the fit may assume almost any value without changing the quality of the fit. The restraints add more information to the fit, giving unambiguous values also to the poorly determined components. Details of restraint methods vary; for example, it is possible to employ several restraint weights or to restrain the charges toward previously

determined physically reasonable values. All such methods reduce over-fitting of a single set of data; it is often observed that the restrained charges become more transferable [5].

Among the commonly used ESP methods, the RESP method was constructed with this philosophy [5]. Here, the charges are restrained to zero using a hyperbolic restraint. The fit is usually performed in two stages. In the first stage, all atoms are restrained with a weak restraint. Then charges on all atoms are fixed except carbons with two or more hydrogens, which are fitted with a stronger restraint, constraining equivalent hydrogen atoms to have the same charge.

Charges calculated with RESP for sucrose, glucose, and fructose are shown in Table I. The agreement between sucrose and glucose charges is much improved, but not yet satisfactory. The average deviation for fructose is $0.05 e$, but the maximum deviation is still large, $0.42 e$ (C2'). The agreement with glucose is even worse; the C1 atom shows a $0.82 e$ difference. Moreover, compared with CHELP-BOW charges the RESP charges reproduce sucrose molecular moments with 2-13 fold greater error and the CHELP-BOW potentials with 3-fold greater mean absolute deviation, as can be seen in Table V.

Consequently, it would be desirable to find a new method that combines the transferability of the RESP charges with the good reproduction of the electrostatic properties achieved by the CHELP-BOW charges. In designing the CHELP-BOW0 method, we decided to employ a harmonic term (a penalty on the square of each atomic charge), since that is compatible with the stable singular-value decomposition method for least-squares fits.

Mathematically, such a harmonic penalty adds a positive offset to every singular value before the calculation of the pseudo-inverse. If all atoms are given the same weight in the restraint penalty, each singular value is increased by the same offset. If one chooses an offset larger than the smallest (ill-determined) singular values, but much smaller than the large (well-determined) singular values, the well-determined charge patterns are little affected, but the ill-determined ones are pulled toward zero. The constraints on total molecular moments (charge, dipole, etc.) employed in CHELP-BOW further reduce the distortion of the large-scale distribution of charge.

As a first attempt, we used a target value of zero for all constraints, and chose a single weight (restraint force constant) w for all atoms. Several different values were tested for sucrose, fructose, and glucose, as is shown in Table V. It can be seen that weight factors of 0.3

or larger give rise to unacceptably large errors in the electrostatic moments and potentials. On the other hand, a weight of 0.03 does not sufficiently improve the transferability (similarity of corresponding charges). Restraining all charges to zero with the weight 0.1 yields charges that are both more transferable and reproduce moments and potentials better than the RESP method.

CHELP-BOW0 restraints to zero charge

A restraint to zero charge with a single weight on all atoms will tend to reduce the charges of both polar and non-polar atoms by the same proportion. Since the polar regions of the molecule is normally not buried, the polar atoms are often well determined and therefore do not need to be restrained. Furthermore, in practice the charges of polar regions are more important to electrostatic energies because they are typically large. Thus, instead of accepting the same *relative* distortion in charges, it may be advisable to introduce a smaller *absolute* distortion into the charges of polar groups than of non-polar groups. This is the reason why RESP uses hyperbolic restraints, which has a restraint weight that decreases with charge. The disadvantage with hyperbolic restraints is that the least-squares problem is no longer linear, so that the equations have to be solved iteratively instead of by a single matrix pseudo-inversion.

CHELP-BOW may easily be modified to achieve similar results by placing a smaller restrain on polar atoms (w_O and w_H) than on non-polar (w_C) atoms. We have tried a number of different combinations, as shown in Table V. Starting with the uniform value $w = 0.1$, we found that removing the restraint on the hydrogen atoms ($w_H = 0$) does not change the charges significantly. Similarly, the weight on the oxygen atoms (w_O) could be reduced to 0.05 without any large change in the quality of the charges. It is notable that in all these fits, the largest deviation between sucrose and fructose or glucose is obtained for carbon atoms. Therefore, we increased the weight on the carbon atoms (w_C) to 0.2, enough to reduce the average and maximum difference to 0.02 and 0.06 e for both glucose and fructose. Further changes in the various weights did not improve the transferability significantly, but the quality of the charges deteriorated slightly. Thus, the best charges seems to be obtained with weights of 0.2, 0.05, and 0 on the carbon, oxygen, and hydrogen atoms, respectively. A fit with these weights will be called CHELP-BOW0 below.

The charges for sucrose, glucose, and fructose obtained with this method are shown in

Table I. It can be seen that the constraints have strongly decreased the charges on the carbon atom; they are all less than 0.05 e . This is appreciably less than for the RESP charges, where the maximum charge is 0.71 e . More important, the chemical transferability between equivalent rings has greatly improved, generally to within 0.03 e , and the charges are reasonable around the ether oxygen O1 formed by condensation of hydroxyls in the monosaccharides.

The restrained CHELP-BOW0 method increases the condition number for sucrose by a factor of 6 (to 0.013) and those for fructose and glucose by a factor of 3 (to 0.039 and 0.033). This indicates that the condition number actually reflects the stability of the fit, but that a more appropriate threshold for well-determined charges should be about 0.01.

Restraints to consensus charges

One may expect even better results by imposing physically reasonable target values (instead of zero), letting the information on potentials (and molecular multipoles) adjust them as needed to account for induction effects. Thus, we need to assign a consensus charge to each atom in the molecule, and we therefore divided the atoms in the sucrose molecule into five types: carbon, polar and non-polar hydrogen (HO and HC), and hydroxide and ether oxygen (OH and OC). We tested two different sets of consensus charges, both taken from the CHELP-BOW0 result in the previous section (i.e. with weights of 0.2, 0.05, and 0), viz. the average or the minimum charges of each type of atom for sucrose, fructose, and glucose. The target values are given in the legend to Table V. We used the same weight (w) for all atom types, varying it between 1.0 and 0.01.

Results are included in Table V. A compromise between transferability and the reproduction of the moments and potentials is obtained for a weight of around 0.2-0.3 (there is no significant difference between the two sets of consensus charges). Yet, the results are not clearly better than those obtained for restraints to a zero charge. Since the consensus charge method involves more parameters, we see no advantage of this method.

Chemical transferability

CHELP-BOW0 restraints produce charges for sucrose that are reasonably similar to its constituents, glucose and fructose (Table V). It remains to assess the quality of the potential

produced by building up charges for sucrose from the charges fitted to the constituent monomers. Such a build-up procedure will be necessary as subject molecules become too large for quantum calculations. It may also furnish an alternative way to overcome stability problems in the subject molecule if stable charges can be fitted to its the constituents. Most fundamentally, it assures us that the atomic-charge parameters have intuitive meaning.

Thus, we merge the charges of fructose and glucose, ignoring those of HO1 and HO2' and adapting the charge of O1 so that the molecule becomes neutral. The results of such a test are shown in Table VI for five types of charges (CHELP-BOW, RESP, consensus charges, and CHELP-BOW0 with and without lone-pair centres on the ether oxygen atoms; after some tests we decided to assign to the lone pairs the same weight as for hydrogen atoms, whereas the ether oxygens got the same weight as the carbon atoms).

Interestingly, all charge sets reproduce the moments and potential 2-14 fold worse than the original CHELP-BOW charges fitted to the sucrose data. Moreover, restraining charges to zero or to consensus charges does not improve the transferability very much (compared to the CHELP-BOW charges). The RESP charges seem to give the best result.

This unexpected behaviour of the build-up procedure led us to test the reverse process, i.e. to see how well the sucrose charges reproduce the potential and moments of the isolated fructose and glucose molecules. All charges except O1 and HO1 on glucose and O2' and HO2' on fructose were constrained to the values found in sucrose, and the remaining two atoms were determined by a normal CHELP-BOW fit, with only the total charge as an additional constraint.

The results in Table VII show that the sucrose charges determined with restraints are clearly more transferable than unrestrained (CHELP-BOW) sucrose charges to the monomers. In fact, the restraints to consensus or zero charges decreased all measures of deviation from monomer potentials by a factor of 2-3 (there is no clear difference between these two restraints). Charges obtained with the RESP method are also more transferable than CHELP-BOW charges, but clearly inferior to the CHELP-BOW0 charges. CHELP-BOW0 charges with ether lone-pairs are clearly less transferable than those without lone-pair centres. Thus, there is a small, but significant, improvement of the chemical transferability of the charges produced by restraints to zero or to consensus charges.

Conformational transferability

Conformational transferability is essential to allow simulations of flexible molecules, even in cases where charges can be determined directly for each subject molecule.

Conformational transferability has in general attracted less interest than chemical transferability, but is discussed in some studies [18-20]. We have addressed this problem in several ways.

First, we tested two random displacements (up to 0.05 or 0.10 Å) of the sucrose coordinates, calculating charges for each conformation. As we saw in Table II, this gave rise to quite large differences between the three sets of charges (up to 0.32 e). However, considering the results of the preceding section, it is more interesting to see how the new charges reproduce the original moments and potentials. Table VIII shows that the deviations of both moments and potentials increases 2-4 fold as an effect of the displacement. It is possible, however, that such effects may partially average out as the distortions oscillate around equilibrium.

To study larger but energetically accessible conformational changes, including bond rotations, we developed a test set of 21 conformations of sucrose obtained by molecular dynamics simulations, reoptimisations, and variations in the hydrogen-bond pattern as is described in the Methods section. For all these new geometries, we calculated how well the original CHELP-BOW charges reproduce the quantum chemical moments and potentials of the various conformations. The results are collected in Table IX as the average deviation over the entire test set. The results for individual test conformations are similar, except that the potentials and moments for the manually built alternative hydrogen-bond patterns are typically better reproduced than those of the other variations by a factor of up to two.

From these results, it can be seen that the original charges reproduce the potential and moments of the various conformations quite poorly. All deviations increase by a factor of 4-26 compared to charges fitted to the correct conformation. The the original CHELP-BOW charges are clearly worst, followed by the consensus and RESP charges, but even the restrained CHELP-BOW0 method gives only a marginal improvement of the charges (less than a factor of two for the deviations) compared to what is obtained by a fit to the actual

conformation. The conclusion must therefore be that the fixed point charges from all methods reproduce the potential and moments of a flexible molecule very poorly.

Fitting to multiple conformations: the CHELP-BOWC method

No method of fitting a single conformation seems to produce charges usable over a wide variety of conformations. This indicates that we should include several different conformations when fitting the charges. This can be done by a minor modification of the CHELP-BOW program as is described in the Methods section. This method will be called CHELP-BOWC below. Since we aim at reproducing the total energy in actual simulations as accurately as possible, it is natural to weight the various conformations by a Boltzmann factor obtained from the energy of the conformation. It remains to settle how this energy should be calculated. At this preliminary stage, a reasonable energy was obtained from a quantum chemical calculation of the sucrose molecule in a continuum aqueous solvent.

Charges for sucrose obtained with the CHELP-BOWC method using the six structures optimised by quantum mechanics are listed in Table I. The six conformations have total solvation energies within 4 kJ/mole of the original structure and therefore weights in the range 0.2-1.0. All the other structures had a higher energy (10-352 kJ/mole) and would therefore contribute by a small weight in the fit (<0.02). From Table I, it can be seen that the charges are quite sensible, but some of the carbon atoms have still rather large charges, e.g. 0.61 e for C5. Note that no restraints towards a consensus or zero charge were used in the fit. We wish to investigate whether the additional information in a few conformations can make such restraints unnecessary.

Table X shows how well these conformation-weighted charges reproduce the moments and potentials of the six conformations included in the fit. It can be seen that the moments and potentials are fairly well reproduced and that the deviations of the moments and potentials of individual conformations follow approximately the weight of the conformation in the fit. The average deviations are 4-fold smaller than those produced by the CHELP-BOW charges obtained from one conformation and 2-fold smaller than for restrained CHELP-BOW0

charges obtained from one conformation. However, the errors are still 2-4 times larger than those obtained by a direct fit to the correct conformation.

If one of the conformations is removed from the fit (we tested the one with the second largest weight, 0.62), the resulting charges changed by 0.02 e on average (maximum change 0.10 e for C1). However, the average quality of the fit to all conformations did not change significantly (cf. Table X), and even the fit to the excluded conformation is only slightly degraded. This indicates that the fit is self-consistent, which is very important considering that we can never include all relevant conformations in the fit. However, we still have to accept an uncertainty in the charges of at least 0.1 e .

If we test against a set of ten (more dissimilar) conformations not included in the fit, viz. those obtained by a molecular dynamics simulation of sucrose, then the average deviation in the moments and potentials increases by a factor of two. Table XI shows that the conformation-weighted CHELP-BOWC charges are still better than what was obtained by a CHELP-BOW fit on a single conformations, with or without restraints.

We also tested whether the transferability of the charges improves if they are restrained towards zero with the CHELP-BOWC0 method. Interestingly, the results in Table X and XI show that such restraints do not significantly change the deviation of the moments or potentials. Thus, there is no reason to restrain the charges if multiple conformations are used.

It is possible to include multiple conformations also with the RESP method. The results of such a fit is also included in Tables X and XI (row RESP-C). It can be seen that the result is slightly worse than for CHELP-BOWC. As for CHELP-BOWC0, the results improve slightly if the hyperbolic restraints to zero charge are removed (row ESP-C). The latter results are simply a normal ESP fit with several conformations. It is similar to the CHELP-BOWC results, except that slightly fewer ESP points were used in the fit and that the potentials were not Boltzmann weighted.

Simultaneous fitting with CHELP-BOWC over only 6 conformations significantly improved the transferability of the charges. To treat more flexible molecules, it may be advisable to use many more conformations, or to employ a divide-and-conquer strategy by

building up the molecule from several fragments, thoroughly exploring the conformations of each fragment. Although the current CHELP-BOW software is not designed for very large training sets, such a simultaneous fit can be handled by the covariance-matrix procedure [17].

Interaction energies with water probes

The preceding results indicate that a fixed point-charge model for sucrose can give rise to quite appreciable errors in the potential for a flexible molecule. For example, a 0.0008-a.u. mean absolute deviation of the potential corresponds to an average error in the interaction energy with a charged group of ~ 2 kJ/mole, or ~ 1 RT . Neutral probes such as water or uncharged polar groups may experience smaller errors, especially beyond the first hydration shell, as the errors in potential appear to vary slowly in space and to decrease roughly as the square of distance.

To test this and to get an estimate of the actual error from the point-charge model in a real simulation, we performed a molecular dynamics simulation of a sucrose molecule in the centre of a 30-Å sphere of (3681) water molecules. At regular times, we sampled the coordinates and calculated the quantum chemical potential at the position of each water atom caused by the sucrose molecule. Then, we compared how well various point-charge models reproduce the interaction energies between the TIP3P water molecules and the sucrose molecule calculated from the quantum chemical potentials. This way, we estimate the errors introduced by the fixed point-charge model, but avoid for example errors introduced by the pair-potential approximation.

The resulting deviations are presented in Table XII. As usual we distinguish between charge sets fitted to the actual conformation and those fitted to other conformations of sucrose. First, we see that CHELP-BOW gives the best charges if the molecule is not flexible (i.e. if the charges are fitted to the correct conformation); restraining charges with the CHELP-BOW0 or RESP methods increase the average and maximum error by up to 40 %. Adding lone-pairs on the ether oxygens, slightly improve the fit, as expected from the increased number of charge centres.

Charges fitted to other conformations give 4-7 times larger mean average errors than

CHELP-BOW charges fitted to the correct conformation. CHELP-BOW with or without lone pairs give the largest mean average deviations (0.07 kJ/mole), which are lowered by restraints to zero charge (0.06 kJ/mole). Including several conformations in the fit gives even lower mean average deviations (0.05 kJ/mole). It is notable, however that the maximum error seems to go in the other direction, increasing from 8 kJ/mole for CHELP-BOW, via 12 kJ/mole for CHELP-BOW0 to 13 kJ/mole for CHELP-BOWC. All methods fitting charges to several conformations (CHELP-BOWC, CHELP-BOWC0, RESP-C, and ESP-C) give similar results, and there is not clear degradation of the performance when constraints to zero charge are included.

The magnitude of the average absolute error for the best method is 0.05 kJ/mole. However, errors of different signs cancel and the mean *signed* error is only 0.004 kJ/mole (15 kJ/mole for all water molecules). It is similar for all methods (0.002-0.011 kJ/mole). The maximum error follows fairly well the average error, but it has a larger statistical uncertainty. It amounts to 4-5 kJ/mole for charges fitted to the correct conformation and to 8-13 kJ/mole for the other charges. The water molecule with the maximum deviation is always found in the first hydration shell of sucrose. The mean average deviation in interaction energies for all water molecules in the first hydration shell is 0.9-1.4 kJ/mole for charges fitted to the correct conformation and 2.2-3.3 kJ/mole for charges fitted to other conformations. Together all these results gives us a quite clear picture of errors expected in actual simulations of polar molecules.

Concluding remarks

In this paper we have tested various methods to improve the quality and transferability of charges for the sucrose molecule. This molecule was selected for its biochemical interest and because it allows chemical transferability to be studied in a well-defined way, without complications from electron delocalisation, since it is formed by condensation of glucose and fructose. Nonetheless, large fluctuations and clearly unstable charge parameters result from standard methods for fitting electrostatic potential.

The instability depends on purely geometric factors, reflected in the singular values of the distance matrix between the atoms of the subject molecule and the potential points. Such

problems cannot be solved by refining the model with further parameters, such as lone-pair charges; they must be addressed by reducing the number of parameters through constraints, or by increasing the amount of information in the fit by using restraints to improve the conditioning of the matrix. We find that for these sugar molecules, the large charges can be suppressed with a modest loss of precision simply by restraining each atomic charge to zero. A further refinement uses smaller restraint weights on polar atoms (0.2 for non-polar atoms, 0.05 for polar heavy atoms, and 0 for all hydrogen atoms). This method, which we call CHELP-BOW0, produces relatively small changes in the charges of polar groups; indeed, restraining the charges with a single restraint weight to a set of consensus charges gives a similar result.

We have tested the usefulness of CHELP-BOW0 charges in terms of chemical and conformational transferability. Building up the sucrose disaccharide by transferring charges from the monomer sugar produces a 3-fold degradation in accuracy of the potential, possibly because this glycoside linkage allows only one degree of freedom to adjust the sucrose charges, leading to errors even in the total molecular dipole.

Poor conformational transferability is observed for atomic charges fitted to a single conformation. The CHELP-BOWC method, fitting simultaneously to several conformations, improves the transferability significantly, to the point that further application of restraints (CHELP-BOWC0) is not necessary. However the accuracy of a single set of atomic charges is limited to an average error 0.0008 a.u. in the potential, or a maximum error of 13 kJ/mole in interaction energy with a solvent water. Visual examination shows a complex pattern of errors centred around hydroxyl oxygens and dying off rapidly at larger distances. These errors might be mitigated by a more realistic charge model, including lone pairs on the hydroxyls, but our results with lone-pairs on the ether oxygens do not point in that direction. For high accuracy, however, the point charge model must most likely be abandoned in favour of models including higher multipoles and a physical model of how they vary with the conformation [21].

Of course, all constraints will slightly degrade the quality of the charges. Therefore, no restraints should be used if we are interested in charges for a given conformation (a fixed molecule), e.g. for a single-point energy calculation using a fixed crystal structure. Even if the unrestrained charges may seem against chemical intuition, they give the best possible

reproduction of the moments and potentials around this fixed molecule. Thus, restraints should not be used for aesthetic reasons, only to improve transferability.

For simulations of a flexible molecule, we recommend an unrestrained CHELP-BOWC fit including many different conformations of the molecule. It is much more important to include many conformations in the fit than to restrain the charges to zero, and if several conformations are included, such restraints will actually degrade the accuracy of fit. Thus, the only situation we recommend a restrained fit (CHELP-BOW0 or RESP) is when only a single conformation of a flexible molecule can be studied, and the application can tolerate the resulting degradation of accuracy.

Acknowledgements

This investigation has been supported by computer resources of the Swedish Council for Planning and Coordination of Research (FRN), Paralleldatorcentrum (PDC) at the Royal Institute of Technology, Stockholm, the High Performance Computing Centre North (HPC2N) at the University of Umeå, and Lunarc at Lund university.

References

1. Bachrach S. M. *Rev Comput Chem* 1994, 5, 171-227.
2. Sigfridsson, E.; Ryde U., *J Comput Chem* 1998, 19, 377-395.
3. Chirlian, L. E.; Francl, M. M. *J Comput Chem* 1987, 6, 894-905.
4. Breneman, C. M. ;Wiberg K. B. *J Comput Chem* 1990, 11, 361-373.
5. Bayly, C. I.; Cieplak, P.; Cornell, W. D. ; Kollman, P. A. *J Phys Chem* 1993, 97, 10269-10280.
6. Singh, U. C.; Kollman, P. A. *J Comput Chem* 1984, 5, 129-145.
7. Besler, B. H. ;Merz, K. M.; Kollman P. A. *J Comput Chem* 1990, 11, 431-439.
8. Francl, M. M.; Carey, C.; Chirlian, L. E.; Gange, D. M. *J Comput Chem* 1996, 17, 367-383.
9. Carey, C.; Chirlian, L.E.; Francl, M.M.; Gange, D.M. *Glycoconjugate J* 1997, 14, 501-505.
10. Hinsen, K.; Roux, B. *J Comput Chem* 1997, 18, 368-380.
11. Hehre, W. J.; Radom, L. ;Scheyer, P. V. R.; Pople, J. A. *Ab Initio Molecular orbital theory*, Wiley-Interscience, New York 1986.
12. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T.; Petersson, G. A.; Montgomery, J. A.; Raghavachari, K.; Al-Laham, M. A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Cioslowski, J.; Stefanov, B. B.; Nanayakkara, A.; Challacombe, M.; Peng, C. Y.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Binkley, J. S.; Defrees, D. J.; Baker, J.; Stewart, J. P.; Head-Gordon, M.; Gonzalez, C.; Pople, J. A. *Gaussian 94, Revision D.1*, Gaussian, Inc., Pittsburgh PA, 1995.
13. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J Am Chem Soc* 1995, 117, 5179-5197.
14. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Knudin, K. N.; Strain, M. C; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo,

- C.; Clifford, S.; Ochterski, J.; Petersson G. A.; Ayala P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari K.; Foresman J. B.; Cioslowski J.; Ortiz J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A; Piskorz, P; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; Chen, W.; Wong, M. W.; Andres, J. L.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. 1998. Gaussian 98, Revision A.5, Gaussian, Inc, Pittsburgh PA.
15. Pearlman, D. A. ; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham III, T. E.; Ferguson, D. M.; Seibel, G. L.; Singh, U. C.; Weiner P. K.; Kollman, P. A. AMBER 4.1, University of California, San Francisco, 1995.
 16. Woods, R. J.; Khalil, M.; Pell, W.; Moffat S. H.; Smith, V. H. J Comput Chem 1990 11, 297-310.
 17. Bush, B.L; Bayly, C. I.; Halgren, T. A. J Comp Chem 1999, 20 ,14, 1495-1516.
 18. Williams, D. E. Biopolymers 1990, 29, 1367-1386.
 19. Stouch, T. R.; Williams, D. E. J Comput Chem 1992, 13, 622-632.
 20. Reynolds, C. A.; Essex, J. W.; Richards, W. G. Chem Phys Lett 1992, 199, 257-260.
 21. Engkvist, O.; Åstrand, P.-O.; Karlström, G. Chem Rev 2000, 100, 4087-4108.

Legends to the Figures

Figure 1. The structure and atom names of sucrose. Glucose is obtained by replacing C2' by HO1, whereas fructose is formed by replacing C1 by HO2' and O1 by O2'.

Figure 2. The structure and atom names (and therefore the correspondence to sucrose) in the four fragment molecules 1-4.

Table I. Charges for sucrose, fructose and glucose calculated with the CHELP-BOW method with and without lone-pairs on the ether oxygens and with the RESP, CHELP-BOW0, and CHELP-BOWC methods.

Ato	CHELP-BOW			CHELP-BOW			RESP			CHELP-BOW0			CHELP-
	with lone-pairs												BOWC
m	Sucr	Fruct	Gluc	Sucr	Fruct	Gluc	Sucr	Fruct	Gluc	Sucr	Fruct	Gluc	Sucr
C1	-1.11	0.46	0.18	-1.22	0.47	0.07	-0.56	0.49	0.25	-0.03	0.45	0.01	-0.10
H1	0.46		0.10	0.51		0.08	0.34		0.09	0.21		0.17	0.28
O1	-0.45	-0.72	-0.62	-1.45	-0.72	-0.66	-0.26	-0.75	-0.62	-0.21	-0.57	-0.56	-0.39
C2	1.08		0.21	1.16		0.25	0.71		0.20	0.04		0.01	0.05
H2	0.00		0.09	-0.03		0.06	0.03		0.11	0.20		0.19	0.12
O2	-0.78		-0.65	-0.77		-0.65	-0.72		-0.66	-0.59		-0.58	-0.59
HO2	0.48		0.42	0.47		0.41	0.44		0.44	0.43		0.43	0.45
C3	-0.37		0.25	-0.42		0.21	0.00		0.14	0.03		0.02	0.34
H3	0.18		0.01	0.19		0.00	0.06		0.01	0.12		0.13	-0.03
O3	-0.66		-0.68	-0.66		-0.67	-0.69		-0.64	-0.60		-0.59	-0.70
HO3	0.48		0.44	0.48		0.44	0.46		0.43	0.43		0.43	0.45
C4	0.51		0.02	0.52		0.11	0.22		0.26	0.03		0.00	0.10
H4	0.08		0.16	0.07		0.11	0.10		0.07	0.19		0.20	0.04
O4	-0.74		-0.64	-0.74		-0.66	-0.69		-0.69	-0.63		-0.62	-0.64
HO4	0.46		0.42	0.46		0.43	0.43		0.43	0.43		0.42	0.38
C5	-0.32		0.04	-0.41		-0.07	0.08		0.10	-0.01		0.00	0.61
H5	0.15		0.09	0.15		0.04	0.11		0.05	0.09		0.11	-0.11
O5	0.00		-0.27	0.90		0.81	-0.26		-0.38	-0.26		-0.20	-0.54
C6	0.39		-0.04	0.41		-0.03	0.23		0.11	0.04		-0.01	0.55
H6	-0.01		0.12	-0.01		0.11	0.02		0.07	0.09		0.11	-0.11
O6	-0.69		-0.62	-0.70		-0.64	-0.70		-0.68	-0.61		-0.61	-0.70
HO6	0.43		0.43	0.44		0.44	0.46		0.46	0.43		0.43	0.39
C1'	0.18	0.07	0.41	0.09	0.18	0.43	0.11	0.09	0.40	0.01	-0.01	0.41	0.12
H1'	0.07	0.10		0.12	0.07		0.10	0.08		0.13	0.15		0.08
O1'	-0.72	-0.68		-0.70	-0.72		-0.70	-0.67		-0.63	-0.63		-0.61
HO1'	0.46	0.45		0.47	0.47		0.47	0.45		0.46	0.45		0.39
C2'	0.48	0.52		0.58	0.46		0.10	0.52		0.00	0.06		0.37
C3'	0.06	-0.11		0.13	-0.05		0.10	0.03		0.00	0.01		-0.07
H3'	0.10	0.18		0.09	0.16		0.12	0.12		0.22	0.22		0.13
O3'	-0.72	-0.68		-0.71	-0.69		-0.66	-0.65		-0.66	-0.66		-0.62
HO3'	0.50	0.44		0.49	0.45		0.43	0.42		0.46	0.45		0.40
C4'	0.45	0.33		0.40	0.34		0.28	0.26		0.04	0.02		0.48
H4'	0.07	0.11		0.09	0.11		0.12	0.10		0.22	0.21		0.03
O4'	-0.78	-0.75		-0.77	-0.76		-0.73	-0.72		-0.65	-0.65		-0.76
HO4'	0.47	0.46		0.48	0.47		0.47	0.45		0.45	0.44		0.45
C5'	-0.13	0.08		-0.02	0.08		-0.03	0.01		0.00	0.01		0.04
H5'	0.14	0.10		0.15	0.11		0.13	0.12		0.18	0.18		0.09
O5'	-0.45	-0.50		-1.27	-0.35		-0.36	-0.49		-0.32	-0.29		-0.40
C6'	0.41	0.31		0.44	0.24		0.43	0.36		0.00	0.02		0.12
H6'	-0.01	0.00		-0.01	0.02		-0.04	-0.02		0.11	0.11		0.06
O6'	-0.69	-0.68		-0.69	-0.66		-0.70	-0.70		-0.59	-0.63		-0.62
HO6'	0.47	0.42		0.47	0.41		0.47	0.44		0.43	0.43		0.44
Lp1				0.36									
Lp5				-0.30		-0.37							
Lp5'				0.31	-0.09								

Table II. Quality of charges for sucrose obtained by CHELP-BOW using five different sets of ESP points (ESP 1-5) or adding a random displacement with a maximum norm of 0.05 or 0.10 Å to the sucrose coordinates.

Case	Average deviation of moments ($D \text{ Å}^{n-1}$)			Deviation of potential		Deviation of charges	
	4-pole	8-pole	16-pole	(a.u.)		relative to ESP 1	
				m.a.	r.m.s.	average	maximum
ESP 1	0.073	0.95	10.4	0.00010	0.00027		
ESP 2	0.079	0.95	10.6	0.00010	0.00027	0.01	0.04 C5
ESP 3	0.068	1.01	11.0	0.00010	0.00027	0.02	0.10 C1
ESP 4	0.073	0.84	10.6	0.00009	0.00026	0.01	0.03 C3
ESP 5	0.067	0.88	10.1	0.00009	0.00027	0.02	0.10 C1
Random 0.05 Å	0.064	0.94	10.8	0.00009	0.00026	0.03	0.16 C1
Random 0.10 Å	0.054	0.94	11.1	0.00010	0.00026	0.06	0.32 C1

Table III. Charges of the heavy atoms in the fragment molecules and in sucrose, glucose, and fructose all obtained by the CHELP-BOW method.

Atom	Fragment 1	Fragment 2	Fragment 3	Fragment 4	Sucrose	Glucose	Fructose
C1	-0.05	0.16	-0.10 ^a	-0.10 ^a	-1.11	0.18	
O1	-0.42	-0.34	-0.44	-0.38	-0.45		
C5	-0.28 ^a	-0.45 ^a			-0.32	0.04	
O5	-0.27	-0.26			0.00	-0.27	
C1'	-0.03		0.05		0.18		0.07
O1'	-0.64		-0.64		-0.72		-0.68
C2'	0.37	-0.18 ^a	0.31	0.18 ^a	0.48		0.52
C5'	-0.43 ^a		-0.37 ^a		-0.13		0.08
O5'	-0.26		-0.28		-0.45		-0.50

^a A methyl group in the fragment

Table IV. Quality of charges for sucrose calculated without or with constraints to the charges of fragment molecules 1-4, and comparison to constrained fructose and glucose charges.

	Average deviation of			Deviation of		Deviation of		Deviation of	
	moments ($\text{D } \text{\AA}^{n-1}$)			potential (a.u.)		fructose charges		glucose charges	
Constraints	4-pole	8-pole	16-pole	m.a.	r.m.s	av.	max.	av.	max.
No extra	0.073	0.95	10.4	0.00010	0.00027	0.06	0.20 C5'	0.26	1.29 C1
Fragment 1	0.053	1.30	12.2	0.00013	0.00034	0.06	0.26 C3'	0.14	0.56 C6
Fragment 2	0.081	1.13	11.8	0.00012	0.00032	0.08	0.40 C2'	0.13	0.52 C5
Fragment 3	0.051	1.34	12.5	0.00014	0.00035	0.06	0.24 C3'	0.17	0.59 C2
Fragment 4	0.074	0.94	10.4	0.00010	0.00027	0.06	0.21 C5'	0.26	1.35 C1
Fragments 2+3+4	0.049	1.19	12.6	0.00014	0.00036	0.05	0.42 C3'	0.11	0.51 C5

Table V. Quality of charges for sucrose obtained with various restraints. The original CHELP-BOW method contains no constraints to zero charge. The RESP method [r5] uses hyperbolic constraints to zero charge. Harmonic constraint weights on all atoms (w); different weights on carbon, oxygen, and hydrogen atoms (w_C , w_O , and w_H ; if no w_H is given, it was 0). Consensus charge set A: C 0.00, HC 0.09, HO 0.42, OH -0.56, and OC -0.20; and set B: C 0.00, HC 0.15, HO 0.44, OC -0.31, and OH -0.62. Lp denotes calculations where lone pair centres on ether oxygens were included.

Method	Average deviation of moments (D Å ⁿ⁻¹)			Deviation of potential (a.u.)		Fructose deviation		Glucose deviation	
	4-pole	8-pole	16-pole	m.a.	r.m.s.	av.	max.	av.	max.
CHELP-BOW	0.073	0.95	10.4	0.00010	0.00027	0.06	0.20 C5'	0.26	1.29 C1
CHELP-BOW + Lp	0.051	0.91	9.4	0.00009	0.00026	0.12	0.92 O5'	0.24	1.29 C1
RESP ^a	0.935	3.83	20.8	0.00033	0.00056	0.05	0.42 C2'	0.11	0.82 C1
$w=1.0$	0.537	21.04	66.9	0.00119	0.00259	0.03	0.14 HO6'	0.03	0.12 O5
$w=0.3$	0.217	5.71	30.0	0.00045	0.00110	0.04	0.19 HO6'	0.03	0.12 O5
$w=0.1$	0.092	1.6	14.0	0.00017	0.00043	0.03	0.11 C2'	0.03	0.12 C6
$w=0.03$	0.087	1.11	10.6	0.00011	0.00029	0.05	0.24 C2'	0.14	0.68 C1
$w_C=0.1$ $w_O=0.1$	0.098	1.71	13.0	0.00016	0.00040	0.03	0.13 C2'	0.04	0.14 C6
$w_C=0.1$ $w_O=0.05$	0.105	1.67	12.2	0.00014	0.00035	0.02	0.15 C2'	0.04	0.15 C6
$w_C=0.2$ $w_O=0.05$	0.111	1.85	13.1	0.00016	0.00037	0.01	0.06 C2'	0.02	0.06 O5
(CHELP-BOW0)									
$w_C=0.3$ $w_O=0.05$	0.112	1.89	13.3	0.00016	0.00038	0.01	0.04 O5'	0.01	0.06 O5
$w_C=0.3$	0.113	1.94	13.1	0.00016	0.00038	0.01	0.04 O5'	0.01	0.07 O5
$w_C=0.5$ $w_O=0.05$	0.113	1.92	13.4	0.00016	0.00038	0.01	0.05 O5'	0.01	0.06 O5
$w_C=0.3$ $w_O=w_H=0.05$	0.111	1.86	13.6	0.00016	0.00038	0.01	0.05 O6'	0.01	0.05 O5
$w_C=0.1$	0.106	1.66	12.2	0.00014	0.00034	0.02	0.14 C2'	0.04	0.16 C6
$w_C=0.05$	0.097	1.3	11.4	0.00012	0.00030	0.04	0.19 C2'	0.1	0.39 C1
Consensus set A $w=1.0$	0.277	4.66	15.2	0.00028	0.00058	0.01	0.02 H1'	0.01	0.02 H5
Consensus set A $w=0.3$	0.103	2.02	15.0	0.00017	0.00041	0.01	0.04 O6'	0.01	0.03 H5
Consensus set A $w=0.2$	0.106	1.89	13.9	0.00016	0.00038	0.01	0.05 C2'	0.02	0.04 C1
Consensus set A $w=0.1$	0.105	1.66	12.2	0.00014	0.00035	0.02	0.15 C2'	0.04	0.15 C1
Consensus set A $w=0.01$	0.077	0.97	10.2	0.00010	0.00027	0.06	0.20 C5'	0.24	1.15 C1
Consensus set B $w=1.0$	0.306	2.77	19.1	0.00020	0.00044	0.01	0.02 HO6'	0.01	0.03 O5
Consensus set B $w=0.3$	0.115	2.08	14.7	0.00016	0.00038	0.01	0.03 O6'	0.01	0.05 O5
Consensus set B $w=0.1$	0.107	1.71	12.4	0.00014	0.00034	0.02	0.13 C2'	0.04	0.14 C1
CHELP-BOW0 + Lp	0.118	1.94	13.4	0.00016	0.00038	0.01	0.05 C2'	0.02	0.05 C6

^a In the RESP method, the dipole moment is not a constraint. It deviates by 0.095 D from the quantum chemical value.

Table VI. Chemical transferability of glucose and fructose charges to sucrose, compared to CHELP-BOW charges, obtained from a direct fit.

Method	Average deviation of moments ($\text{D } \text{\AA}^{n-1}$)				Deviation of potential (a.u.)	
	dipole	4-pole	8-pole	16-pole	m.a.	r.m.s.
CHELP-BOW fit to sucrose	0.000	0.073	0.95	10.4	0.00010	0.00027
<i>Transfer from monomers:</i>						
CHELP-BOW	0.232	0.627	3.52	23.6	0.00045	0.00067
RESP	0.095	0.935	3.83	20.8	0.00033	0.00085
CHELP-BOW0	0.204	0.630	3.14	22.4	0.00042	0.00056
Consensus A $w=0.2$	0.195	0.643	3.07	21.2	0.00041	0.00063
CHELP-BOW0 + Lp	0.210	0.618	3.27	22.5	0.00040	0.00061

Table VII. Chemical transferability of the sucrose charges to its monomers, compared to CHELP-BOW charges obtained from direct fit.

Glucose						
Method	Average deviation of moments				Deviation of potential	
	dipole	(D Å ⁿ⁻¹)			(a.u.)	
		4-pole	8-pole	16-pole	m.a.	r.m.s.
CHELP-BOW fit to glucose		0.19	2.56	8.8	0.00018	0.00042
<i>Transfer from sucrose:</i>						
CHELP-BOW	0.58	1.06	5.04	22.4	0.00134	0.00191
RESP	0.37	0.92	4.84	13.9	0.00086	0.00078
CHELP-BOW0	0.19	0.46	3.03	11.9	0.00043	0.00070
Consensus A $w=0.2$	0.15	0.35	2.95	11.0	0.00036	0.00064
CHELP-BOW0 + Lp	0.30	0.84	4.72	18.1	0.00063	0.00121

Fructose						
Method	Average deviation of moments				Deviation of potential	
	dipole	(D Å ⁿ⁻¹)			(a.u.)	
		4-pole	8-pole	16-pole	m.a.	r.m.s.
CHELP-BOW fit to fructose		0.11	1.69	11.4	0.00014	0.00033
<i>Transfer from sucrose:</i>						
CHELP-BOW	0.60	1.09	4.29	14.4	0.00125	0.00197
RESP	0.39	0.68	4.83	16.2	0.00078	0.00124
CHELP-BOW0	0.10	0.23	3.20	15.9	0.00038	0.00072
Consensus A $w=0.2$	0.15	0.21	2.83	15.1	0.00041	0.00074
CHELP-BOW0 + Lp	0.14	0.32	3.24	16.3	0.00040	0.00077

Table VIII. Transferability of charges for sucrose obtained with various random displacements. It is investigated how well the three sets of charges obtained with different displacements reproduce the moments and potentials of molecule without any displacement.

Random displacement (\AA)	Average deviation of moments ($\text{D } \text{\AA}^{n-1}$)				Deviation of potential (a.u.)	
	dipole	4-pole	8-pole	16-pole	m.a.	r.m.s.
0.00	0.000	0.073	0.95	10.4	0.00010	0.00027
0.05	0.197	0.420	2.41	13.8	0.00032	0.00052
0.10	0.385	0.847	5.00	19.1	0.00057	0.00093

Table IX. Transferability of fixed sets of sucrose charges to 20 other test conformations, expressed as averages over the test set.

Method	Average deviation of moments ($\text{D } \text{\AA}^{n-1}$)				Deviation of potential (a.u.)	
	dipole	4-pole	8-pole	16-pole	m.a.	r.m.s.
Correct conformation	0.000	0.07	0.95	10.4	0.00010	0.00027
CHELP-BOW	0.949	1.55	6.74	38.4	0.00147	0.00220
RESP	0.713	1.75	8.45	40.6	0.00124	0.00174
CHELP-BOW0	0.521	1.67	5.41	38.3	0.00093	0.00145
Consensus A $w=0.2$	0.578	1.81	5.40	42.4	0.00111	0.00162

Table X. Quality of charges produced by conformation-weighted fitting, tested on the six conformations included in the fit alone and as an average.

Method	Conformation		Wei-ght	Geometric r.m.s.d. (Å)	Deviation of moments (D Å ⁿ⁻¹)				Deviation of potential (a.u.)	
	Fitted	Tested			dipole	4-pole	8-pole	16-pole	m.a.	r.m.s.
CHELP-BOW	1	1			0.000	0.07	0.95	10.4	0.00010	0.00027
CHELP-BOW	1	1-6			1.110	1.85	6.69	36.4	0.00148	0.00249
CHELP-BOW0	1	1-6			0.542	1.66	5.76	38.4	0.00086	0.00151
CHELP-BOW0Lp	1	1-6			0.469	1.38	5.27	36.4	0.00088	0.00135
CHELP-BOWC	1-6	1	0.37	2.31	0.083	1.11	5.82	27.7	0.00046	0.00077
CHELP-BOWC	1-6	2	0.60	0.72	0.151	0.49	1.59	31.3	0.00033	0.00059
CHELP-BOWC	1-6	3	1.0	0.00	0.146	0.46	3.19	28.6	0.00027	0.00048
CHELP-BOWC	1-6	4	0.36	0.57	0.126	0.28	3.12	28.5	0.00028	0.00050
CHELP-BOWC	1-6	5	0.24	0.87	0.428	0.55	8.16	42.9	0.00064	0.00092
CHELP-BOWC	1-6	6	0.62	0.71	0.152	0.50	1.60	31.3	0.00033	0.00059
CHELP-BOWC	1-5	6	0.00	0.71	0.195	0.57	2.26	33.1	0.00040	0.00066
CHELP-BOWC	1-6	1-6			0.181	0.56	3.91	31.7	0.00038	0.00064
CHELP-BOWC	1-5	1-6			0.181	0.56	3.91	31.7	0.00038	0.00064
CHELP-BOWC0	1-6	1-6			0.192	0.49	4.52	33.8	0.00041	0.00070
RESP-C	1-6	1-6			0.172	0.51	5.04	31.1	0.00042	0.00073
ESP-C	1-6	1-6			0.186	0.56	4.16	32.5	0.00039	0.00065

Table XI. Quality of charges produced by conformation-weighted fitting, tested on ten conformations *not* included in the fit.

Method	Conformations		Average deviation of moments ($\text{D } \text{\AA}^{n-1}$)				Average deviation of potential (a.u.)	
	Fitted	Tested	dipole	4-pole	8-pole	16-pole	m.a.	r.m.s.
CHELP-BOW ^a	1	1	0.000	0.07	0.95	10.4	0.00010	0.00027
CHELP-BOW	1	7-16	1.023	2.18	7.18	45.8	0.00179	0.00245
CHELP-BOW0	1	7-16	0.537	2.11	5.78	46.4	0.00106	0.00159
CHELP-BOW0+Lp	1	7-16	0.521	2.09	5.62	47.5	0.00105	0.00157
CHELP-BOWC	1-6	7-16	0.434	1.09	5.97	44.8	0.00080	0.00115
CHELP-BOWC0	1-6	7-16	0.429	1.11	5.95	47.5	0.00081	0.00118
RESP-C	1-6	7-16	0.452	1.11	6.18	45.7	0.00085	0.00124
ESP-C	1-6	7-16	0.401	1.11	5.58	43.9	0.00074	0.00108

Table XII. Interaction energies (kJ/mole) with 3681 water probe molecules for sucrose charge sets, compared with exact energy calculated from the quantum chemical potentials.

Charges fitted to the correct conformation.

Method	Deviations in interaction energies (kJ/mole)				
	m.a.	r.m.s.	maximum	m.a. ^a	mean signed
CHELP-BOW	0.01	0.105	3.2	1.0	+0.003
CHELP-BOW + Lp	0.009	0.100	3.2	0.9	+0.003
CHELP-BOW0	0.012	0.135	4.3	1.4	+0.005
CHELP-BOW0 + Lp	0.012	0.136	4.3	1.4	+0.005
RESP ^a , w<0.1	0.014	0.109	3.2	1.0	+0.001

Charges fitted to other conformations

Method	Deviations in interaction energies (kJ/mole)				
	m.a.	r.m.s.	maximum	m.a. ^a	mean signed
CHELP-BOW	0.068	0.370	-8.4	3.0	-0.010
CHELP-BOW + Lp	0.068	0.374	-9.0	3.1	-0.010
CHELP-BOW0	0.060	0.308	11.9	2.2	-0.004
CHELP-BOW0 + Lp	0.062	0.313	12.0	2.3	-0.004
RESP	0.057	0.295	10.6	2.3	-0.005
CHELP-BOWC	0.050	0.336	13.4	2.7	+0.004
CHELP-BOWC0	0.046	0.343	12.8	2.9	+0.006
RESP-C	0.046	0.380	13.6	3.3	+0.011
ESP-C	0.049	0.327	13.2	2.5	+0.004

^a Mean average deviation for water molecules in the first hydration shell (i.e. within 3.0 Å from any sucrose atom).