Glycoconjugate Journal (1997) **14:** 501–505

EGC1

Beyond CHELP: improved potential derived charges for sugars

Christina Carey[1], Lisa Emily Chirlian[1], Michelle Miller Francl[1]* and David M. Gange[2]

[1] *Department of Chemistry; Bryn Mawr College; 101 N. Merion Avenue; Bryn Mawr, PA 19010-2899*
[2] *American Cyanamid Company; Agricultural Research Division; P.O. Box 400; Princeton, NJ 08543-0400*

The partitioning of the overall molecular charge distribution into atom centered monopole charges, while quantum mechanically ill-defined, is nevertheless a technique which finds applications in several broad classes of chemical problems. Charges derived from fits to electrostatic potentials have an intuitive appeal since, in principle, these could be derived from either theoretical or experimental data. It has been noted, however, that such potential derived charges can be conformationally dependent in ways that do not appear to reflect the changes in the molecular wavefunction. Both the algorithm used for selecting points at which the molecular electrostatic potential will be fit and the density of points used in the fit have been suggested to influence the resultant charges. Recently [Stouch TR, Williams DE (1992) *J Comp Chem* 13: 622–32; Stouch TR, Williams DE (1993) *J Comp Chem* 14: 858–66] it has been noted that numerical difficulties may make it impossible to fit all the atomic charges in a molecule. Singular value decomposition (SVD) of the linear least squares matrices used in fitting atom based monopoles to molecular electrostatic potentials provides a tool for evaluating the integrity of the calculated charges. Based on the SVD analysis for a selected group of molecules we have noted particularly that increasing the molecular size reduces the fraction of charges which can be validly assigned. Users of PD derived charges, especially those who are using those charges for tasks other than reproduction of the MEP, should be aware that there is a high probability that a significant portion of those charges are statistically unreliable. Therefore, charges in many biological molecules, such as sugars, prove to be difficult to obtain by potential derived (PD) methods such as CHELP or CHELPG. Results from the SVD can be used to both assess PD charges and to generate an improved, albeit incomplete, set. Improved PD fits are presented for a series of simple saccharides.

***Keywords*: atomic charges, potential derived charges, ESP charges, molecular electrostatic potential, ab initio MO theory, CHELP, CHELPG**

***Abbreviations*: HF, Hartree-Fock; LLS, linear least squares; MEP, molecular electrostatic potential; PD, potential derived; SVD, singular value decomposition**

Introduction

Atomic charges are one of the simplest and most widely used reactive indices. From the beginning of their careers, students are taught elementary schemes for dividing up the charge in a molecule. These schemes are compelling in that they allow us to use a very basic principle – like repels like – to sort through complex chemistry. The principal drawback of any charge assignment scheme, from trivial ones based on relative electronegativity to complex schemes such as Bader's charge partitioning [1, 2], is that atomic charge is not a quantum mechanical observation. It is not an experimentally measurable parameter. Concomitantly, it is difficult to assess the relative merits of any charge assignment scheme [3, 4].

One popular method for assigning charges fits an atom centered monopole potential to the molecular electrostatic potential (MEP) [5–10]. These schemes are intuitively appealing, since the MEP is experimentally measurable and hence charges from quantum chemical calculations can be compared to charges derived from experimental measurements [11]. Despite their widespread use, there are some difficulties in assigning charges using these potential derived (PD) charge schemes. PD charges are notoriously sensitive to changes in molecular conformation. A small change in conformation can lead to wild swings in charges. Another frequently encountered problem with PD charges results in atoms which are equivalent by symmetry being assigned substantially different charges. Both problems have been attributed at various times to low point density and/or defects in the point selection scheme [10]. Large molecules pose problems of their own. Stouch and Williams [12, 13]

---

* To whom correspondence should be addressed.

0282–0080

have noted that for at least some large molecules, the least squares data used in the fits is highly correlated making it impossible to fit all the atoms in the molecule. While none of these difficulties affect the reproduction of the MEP, they do muddy the chemical interpretation of the charges and prevent the transfer of charge data from one molecule to another.

We [14] have recently shown that the numerical problems noted by Stouch and Williams [12, 13] may also occur in smaller molecules. We have also demonstrated that the symmetry problem in PD charges assignment schemes, particularly in the CHELP [5, 6] and CHELPG [10] programs, is not entirely a function of the point density or point selection scheme. Further, we show that increasing the density of points at which the fit is done does not necessarily alleviate the difficulties.

One might naively assume that when fitting a set of n atomic charges to the electrostatic potential at N points in space, as long as $N \gg n$ all n charges can be determined. Unfortunately, the linear least squares (LLS) matrix is ill-conditioned and in fact generally fewer than n charges can be assigned. In order to assess this ill-conditioning we have used the singular value decomposition (SVD). The SVD analysis can be used simply as a diagnostic tool, *eg* to give an evaluation of the linear dependencies in any given least squares matrix, but has the additional advantage that the decomposition can also be employed to find a new, better conditioned least squares matrix, and therefore a more stable solution to the problem. We use the SVD of the LLS matrix to make an estimate of the number of charges that can be recovered from the data. A reduced set of charges can then be fit to the MEP. Various schemes can be used to select the subset, in this work we have chosen chemically significant non-hydrogen atoms, up to the maximum number that the SVD has shown to be assignable. The SVD is a standard numerical technique and as such we will refrain from reviewing it here; instead, we refer the interested reader to our work [14] and the literature cited therein [15–17] for further information.

Our previous work with the SVD [14] shows that: (1) in general a numerically significant value of the charge cannot be obtained for every atom in a molecule; (2) the larger the molecule, the smaller the fraction of atoms that can be assigned charges; and (3) that eliminating some atoms from the fit using information from the SVD can improve the remaining charge assignments and does not affect the overall reproduction of the MEP. Again, we refer the interested reader to our earlier work.

PD charges are often used in parameter sets for molecular modeling and molecular dynamics calculations on large molecules, including sugars. If not all the charges are statistically significant, then PD charges used in parameter sets may be suspect. In other words, for some atoms, in some molecules, PD charges are not much more than (expensively generated) random numbers. Since the problems described above are most accentuated in large systems, we have compared standard PD charges to charges derived using information from the SVD of the appropriate matrices for several simple sugars in order to assess the transferability of charge data in these molecules.

Methods

A modified version of CHELP [5, 6], which we will refer to as CHELP-SVD [14], was used to compute the potential derived charges. The modified code has a new algorithm for selecting the points at which the electrostatic potential will be fit and can make use of information obtained from the singular value decomposition (SVD) of the linear least squares matrix to improve the stability of the fit.

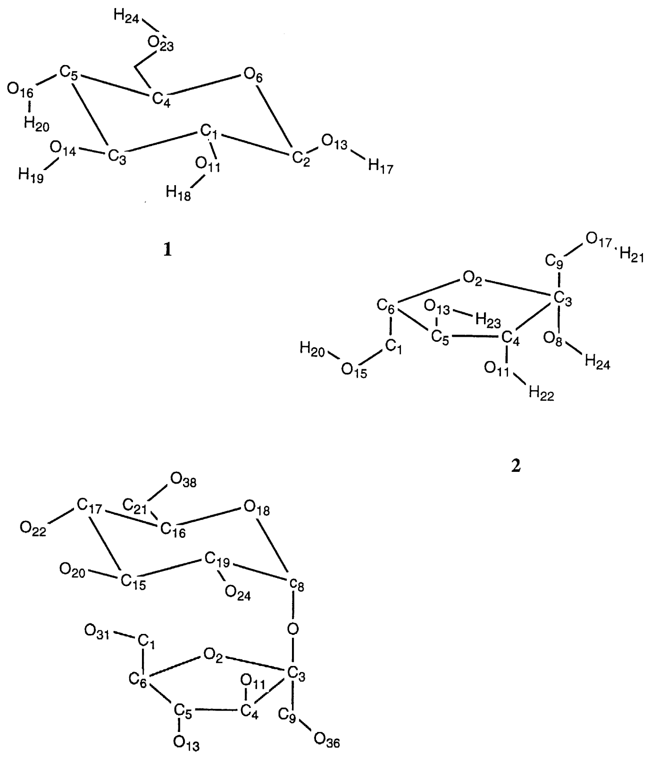
Rather than selecting points based on a regular grid, *eg* on shells around the van der Waals radius (CHELP) or on a rectangular grid (CHELPG), the point selection algorithm has been modified to produce as unbiased a point distribution as possible. We have followed the selection procedure of Woods *et al.* [18] and selected points at random within a rectangular box enclosing the molecule. This type of algorithm should ameliorate the unexpected asymmetries in assigned charges which can arise when a molecule's principal symmetry element is not aligned with the grid, resulting in a different distribution of points around equivalent atoms. Using a random selection of points, as opposed to a grid, can also reduce the collinearity of the matrix. Those points which fall within the van der Waals envelope of the molecule are discarded, unless otherwise noted. We take the van der Waals envelope to be the surface formed by the superposition of spheres centered at the atoms. Atomic radii were taken from sphere fits to electron density surfaces [19].

The only constraint imposed on the charges was that the sum of the atomic charges be equal to the total molecular charge. Molecular wavefunctions were computed using Spartan [20]. Structures were fully optimized using AM1 [21]. HF/3-21G(*) [22, 23] single point calculations using these structures were used to generate the electrostatic potentials.

CHELP and CHELP-SVD charges were obtained for two monosaccharides, D-glucose and D-fructose; and a disaccharide, sucrose (see Figure 1).

Results and discussion

In general we find that charges assigned using CHELP-SVD are more 'sensible'. That is, symmetrical atoms are assigned symmetrical charges; atoms that are expected to bear some charge, do; and bonds are not polarized beyond that expected from chemical intuition. For molecules of this size, we typically find that only between 60 and 70% of the atoms can be assigned numerically valid charges. For those atoms which cannot be assigned a charge, we set the charge to be zero. In essence, we subsume those charges into those of their neighbors. This is not physically implausible,



1

2

3

Figure 1. Numbering scheme for **1** glucose, **2** fructose, **3** sucrose. Not all atoms shown. Most hydrogens eliminated for clarity.

as classically all the charge on a charged object is found on the surface, none on the interior. The fraction of charges that that can be fit falls off roughly with the surface area to volume ratio of the molecules.

Glucose

Values for charges assigned by CHELP and CHELP-SVD for glucose are found in Table 1. Note that only 16 of the total 24 atoms can be fit. Improvements in the prediction of the charges in buried atoms can be seen in these results. CHELP predicts the charges on C2 and C4 to be 0.36 and 0.09 respectively. While the trend is certainly what one might expect from simple electronegativity arguments (C2 more positive than C4), the nearly neutral character of C4 is unanticipated. CHELP-SVD charges display the trend (0.64 compared to 0.28) but the delta positive nature of C2 is now revealed.

Fructose

In Table 2 we compare charges from a 3-21G(*) MEP using a structure obtained from AM1 to charges from 3-21G(*) MEP using an HF/3-21G(*) MEP using an HF/3-21G structure. In neither case can all the atoms be assigned numerically valid charges, only 17 of 24 are assignable. The results

Table 1. Comparison of CHELP and CHELP-SVD charges for glucose. Charges are from electrostatic potential calculations at the HF/3-21G(*) level. Atom numbering scheme given in Figure. 1. See text for additional details.

Atom	*CHELP*	*CHELP-SVD*
C2	0.36	0.64
C3	0.36	0.35
C4	0.09	0.28
C5	0.29	0.36
C7	0.26	0.28
O6	– 0.58	– 0.66
O11	– 0.73	– 0.74
H18	0.44	0.43
O13	– 0.78	– 0.82
H17	0.51	0.50
O14	– 0.81	– 0.80
H19	0.48	0.48
O16	– 0.81	– 0.85
H20	0.50	0.51
O23	– 0.79	– 0.71
H24	0.48	0.44
rms	1.10	1.46

Table 2. Comparison of CHELP and CHELP-SVD charges for fructose. Charges are from electrostatic potential calculations at the HF/3-21G(*) level. Atom numbering scheme given in Figure 1. See text for additional details.

Atom	*HF/3-21G(*)//AM1*		*HF/3-21G(*)// HF/3-21G(*)*	
	CHELP	*CHELP-SVD*	*CHELP*	*CHELP-SVD*
C1	0.16	0.35	0.20	0.32
O2	– 0.71	– 0.58	– 0.78	– 0.76
C3	0.69	0.20	0.90	0.76
C4	0.37	0.40	0.23	0.33
C5	0.13	0.39	0.11	0.26
C6	0.34	0.30	0.47	0.40
H7	0.05		0.01	
O8	– 0.72	– 0.62	– 0.81	– 0.77
C9	– 0.09	0.40	0.12	0.30
H10	0.03		0.06	
O11	– 0.80	– 0.77	– 0.81	– 0.80
H12	0.13		0.09	
O13	– 0.72	– 0.74	– 0.74	– 0.77
H14	0.01		0.04	
O15	– 0.72	– 0.71	– 0.81	– 0.82
H16	0.10		0.04	
O17	– 0.68	– 0.74	– 0.73	– 0.73
H18	0.14		0.06	
H19	0.10		0.05	
H20	0.43	0.40	0.47	0.46
H21	0.45	0.46	0.43	0.41
H22	0.45	0.45	0.48	0.47
H23	0.41	0.41	0.45	0.46
H24	0.44	0.41	0.48	0.46

Table 3. Comparison of CHELP and CHELP-SVD charges for sucrose. Charges are from electrostatic potential calculations at the HF/3-21G(*) level. Atom numbering scheme given in Figure. 1. See text for additional details.

Atom	*CHELP*	*CHELP-SVD*
C1	0.25	0.46
O2	−0.56	−0.50
C3	0.26	0.54
C4	0.19	0.23
C5	0.37	0.18
C6	0.22	0.24
H7	0.09	
C8	0.16	0.40
C9	0.28	0.18
H10	0.08	
O11	−0.73	−0.54
H12	−0.01	
O13	−0.84	−0.73
H14	0.06	
C15	0.33	0.28
C16	0.33	0.16
C17	0.28	0.74
O18	−0.65	−0.95
C19	0.13	−0.18
O20	−0.79	−0.26
C21	0.14	0.15
O22	−0.80	−0.54
H23	0.06	
O24	−0.71	−0.31
H25	0.02	
H26	0.07	
H27	0.01	
H28	0.47	
H29	0.13	
H30	0.03	
O31	−0.71	−0.80
H32	−0.01	
H33	0.43	0.49
H34	0.49	0.46
H35	0.44	0.20
O36	−0.79	−0.28
H37	0.07	
O38	−0.72	−0.59
H39	0.11	
H40	0.04	
H41	0.44	0.20
H42	0.47	0.20
H43	0.44	
H44	0.42	

do suggest that the charges are sensitive to the changes in structure, but not overly so. The only major difference appears to be in the charge assigned to C3, which is 0.20 using the AM1 structure and 0.76 when the 3-21G(*) structure is used.

Again, as seen in glucose, CHELP charges on carbons bonded to oxygen tend to be smaller than expected. For example, C9 is predicted to be slightly negative by CHELP. CHELP-SVD results seem to be more in line with expectations. CHELP-SVD assignments show C9 to be somewhat positively charged.

Sucrose

CHELP and CHELP-SVD charge assignments for sucrose can be found in Table 3. Charges assigned by CHELP to hydroxyl hydrogens appear to be anomalously small. For example, H14 is assigned a charge of only 0.06, while charges of roughly 0.4 to 0.5 (for example see H44) are to be expected for hydroxyl hydrogens. The SVD of the LLS matrix suggests that many fewer charges can fit than there are atoms, a total of only 27 of 45. Note that for the larger molecule, almost 10% fewer charges can be validly assigned.

We have found it best to fit only to the hydroxyl oxygens in these cases; essentially subsuming the charge on the hydroxyl hydrogen into the charge on oxygen. This would be expected to lead to somewhat reduced charges on the oxygens. For example, O13 drops from a charge of −0.84 in the CHELP model to −0.73 in the CHELP-SVD model.

Conclusions

(1) Only a fraction of the total atoms can be assigned charges using PD methods. This fraction gets smaller as the molecule gets larger.

(2) Charges fit to appropriately sized subsets appear to be more ‘sensible’ than those fit to the entire set of atoms.

(3) Users of PD derived charges, especially those who are using those charges for tasks other than reproduction of the MEP, should be aware that there is a high probability that a significant portion of those charges are statistically unreliable.

Acknowledgements

M.M.F. would like to thank Ann Dixon for her technical assistance.

References

1 Bader RFW (1991) *Chem Rev* **91**: 893–928.
2 Bader RFW, Popelier PLA, Keith TA (1994) *Angew Chem Intl Ed Engl* **33**: 620–31.
3 Bachrach S (1994) In *Reviews in Computational Chemistry*, Lipkowitz KB, Boyd DB (eds) **5**: 171–228.
4 Wiberg K, Rablen P (1993) *J Comp Chem* **14**: 1504–18.
5 Chirlian LE, Francl MM (1987) *J Comp Chem* **8**: 894–904.
6 Chirlian LE, Francl MM (1987) *QCPE* **7**: 39.
7 Cox SR, Williams DE (1981) *J Comp Chem* **2**: 304–23.

8 Singh UC, Kollman PA (1984) *J Comp Chem* **5**: 129–45.

9 Besler BH, Merz KM, Kollman PA (1990) *J Comp Chem* **11431**: 431–39.

10 Breneman CM, Wiberg KB (1990) *J Comp Chem* **11**: 361–73.

11 Su Z (1993) *J Comp Chem* **14**: 1036–41.

12 Stouch TR, Williams DE (1992) *J Comp Chem* **13**: 622–32.

13 Stouch TR, Williams DE (1993) *J Comp Chem* **14**: 858–66.

14 Carey C, Chirlian LE, Gange DM, Francl MM (1996) *J Comp Chem* **17**: 367–83.

15 Golub GH, van Loan CF (1989) *Matrix Computations, 2nd edition* pp 571–6 Baltimore: Johns Hopkins University Press.

16 Lawson CL, Hanson RJ (1974) *Solving Least Squares Problems*; Englewood Cliffs, New Jersey: Prentice-Hall.

17 Press WH, Flannery BP, Teukolsky SA, Vetterling WT (1986) *Numerical Recipes*, pp 52–65. Cambridge: Cambridge University Press.

18 Woods RJ, Khalil M, Pell W, Moffat SH, Smith VH Jr (1990) *J Comp Chem* **11**: 297–310.

19 Francl MM, Hout RF Jr, Hehre WJ (1984) *J Am Chem Soc* **106**: 563–70.

20 Spartan; Wavefunction, Inc.: Irvine, CA.

21 Dewar MJS, Zoebisch EG, Healy EF, Stewart JJP (1985) *J Am Chem Soc* **107**: 3902–9.

22 Binkley JS, Pople JA, Hehre WJ (1980) *J Am Chem Soc* **102**: 939–47.

23 Pietro WJ, Francl MM, Hehre WJ, Defrees DJ, Pople JA, Binkley JS (1982) *J Am Chem Soc* **104**: 5039–48.