

FULL PAPER

## Improved Parameters for Generating Partial Charges: Correlation with Observed Dipole Moments

Péter Bagossi<sup>1</sup>, Gábor Zahuczky<sup>1</sup>, József Tözsér<sup>1</sup>, Irene T. Weber<sup>2</sup>, and Robert W. Harrison<sup>2</sup>

<sup>1</sup>Department of Biochemistry and Molecular Biology, University Medical School of Debrecen, H-4012 Debrecen, Hungary. E-mail: peter@indi.biochem.dote.hu

<sup>2</sup>Department of Microbiology and Immunology, Kimmel Cancer Institute, Thomas Jefferson University, Philadelphia, PA 19107, USA.

Received: 29 April 1999/ 3 August 1999/ Published: 9 September 1999

**Abstract** The Universal Force Field was initially combined with the Charge Equilibrium Scheme in the molecular mechanics program AMMP in order to generate partial charges for protein atoms. A new parameter set with improved generation of partial charges has been derived in order to analyse a wider range of atoms and compounds. The electrostatic parameters were modified to achieve better correlation with experimental dipole moments for a training set of 160 compounds, which included alkali metal halogenides, saturated and unsaturated hydrocarbons, alcohols, ethers, amines, thiols, sulphides, oxo compounds, aromatic and heteroaromatic molecules. The correlation coefficient for the calculated and experimental dipole moments was improved from 0.57 to 0.98. The new parameters were tested for another 149 compounds and the correlation coefficient increased from 0.48 to 0.85. The newly generated Modified Parameter Set for AMMMP (MOPSA) improves the predictive power of the program, especially in the area of the macromolecular modelling and drug design, where the nonbonded energies play a crucial role.

**Keywords** UFF, Partial charges, Dipole moments, Molecular mechanics, Force field

### Introduction

Most molecular mechanics force fields are suitable only for limited sets of molecules, e.g. inorganic or organic compounds, small molecules or polymers. The Universal Force Field (UFF) developed by Rappé et al. [1] is one of the exceptions, because its parameter set contains values for every atom in the periodic table. This makes UFF and UFF-de-

rivative force fields especially useful for problems like drug design and unnatural biopolymers because the force fields have terms for unusual molecules. Calculated bond lengths, angles and torsions are in good agreement with experimental values for inorganic [1, 2] and organic compounds [1, 3]. However, non-bonded (van der Waals and electrostatic) interactions are also crucial for determining the energetics of the molecules and estimating chemical properties such as heats of solvation and vaporisation, crystal lattice energies, and interaction energies. The non-bonded interactions are especially important for the study of proteins and nucleic acids. Biologically important interactions between molecules

Correspondence to: P. Bagossi

or segments of molecules are primarily mediated through non-covalent interactions. A poor quality molecular potential will interfere with the study of relatively simple phenomena like base-pairing and ligand binding, let alone difficult problems like predicting protein structure.

Most molecular mechanics methods assume that the electrostatic energy can be expressed as a sum of the coulombic forces between the localised point charges (partial charges) of the atoms [4]. The concept of partial charges works well within the level of error of the molecular mechanics programs [5]. Because UFF does not contain default values for partial charges, they must be derived from other methods in order to apply the UFF force field to macromolecules or to calculate non-bonded energy values related to macroscopic properties of compounds. Unfortunately the partial charges are neither directly measurable by experimental means, nor is there an operator determine the values from quantum mechanics to directly. There are several methods for deriving partial charges that range from extraction from the experimentally determined dipole moments of the diatomic molecules [6] to calculation from the quantum mechanical wavefunction [7, 8] and even the combination of these methods [9]. Each method has some advantages and disadvantages. The wavefunction-derived charges are computationally rather expensive, while other methods are suitable only for a limited set of molecules. The AMBER all atom set [10], which was specifically developed for proteins and nucleic acids, contains values for partial charges derived from quantum mechanics and was used to supplement the UFF for AMMP [11] molecular dynamics simulations on human immunodeficiency virus type 1 protease [12]. Reparameterisations of the potential set to fit covalent bond geometry and IR spectral data better are described by Weber and Harrison [13, 14]. However, the validity of UFF supplemented with partial charges from AMBER is limited to those molecules which have a value in the AMBER set. The authors of UFF suggested [1] the use of partial charges generated by the Charge Equilibrium Scheme (QEq) [15]. This combination was confirmed by tests showing that QEq and AMBER have comparable values for partial charges of amino acids [15]. However, Gundertofte et al. [16] recently tested several molecular mechanics force fields for accuracy in conformational calculations and found that the UFF without any charge calculation performed better than the UFF combined with QEq. Here we test UFF combined with QEq as implemented in AMMP [17], and present a modification of the parameter set that achieves better correlation between the calculated and the experimental dipole moments of the selected set of molecules. The new parameters are referred to as Modified Parameter Set for AMMMP (MOPSA).

QEq is an electronegativity equalisation algorithm [15]. The energy of placing a partial charge on a given atom is parameterised in terms of the electronegativity, the electron affinity (the energy of the negative ion), and the ionisation potential (the energy of the positive ion). The first derivative of the energy of charging a given atom is approximately the electronegativity, and the second derivative is approximately the difference between the ionisation potential and the electron affinity. Therefore, a Newton-Raphson method could be

used to determine the charges by finding values which sat-

$$\text{isfy } \frac{dE}{dq} = q \frac{d^2E}{dq^2}.$$

In practice, a constraint potential is applied to ensure that the total charge is the desired value for the molecule, and the Hessian matrix of second derivatives is used rather than a single derivative. The off-diagonal elements of the Hessian, which determine the interactions between charges on two atoms, are estimated by the average of the individual atomic terms corrected for screening by a function of the distance between the atoms. Both AMMP [17] and QEq [15] use an s-orbital approximation to the electron density and use the Coulomb integral for the correction. However the exact functional forms are different, so caution should be used when transferring parameters between the data sets.

The primary advantage of an empirical or semi-empirical method like QEq for determining charges is its speed. While it is possible to perform *ab initio* quantum mechanical calculations for small molecules, and determine approximate charges by fitting the molecular electrostatic potential, quantum mechanics is expensive. So while it is useful to perform accurate quantum mechanics when a limited number of kinds of molecules are being studied, it could be difficult to apply to a large number of kinds of molecules. The charges will also depend on the molecular conformation, since the molecular electrostatic field depends on the conformation. So even though the charges derived from quantum methods may be highly accurate for a given conformation they will undoubtedly be less accurate for other conformations. On the other hand, a relatively fast empirical method could be used to estimate charges as a function of molecular conformation. The primary disadvantage of an empirical or semi-empirical method is, that if it is badly parameterised, it will give bad values for the charges with no warning that the values are incorrect. Therefore it is critically important to demonstrate that the parameterisation reproduces experimental data on a large number of molecular systems in order to establish the accuracy of the approach.

---

## Computational methods

### Programs

Calculations were done by AMMP [11] using a combination of parameters from UFF and from a modified version of QEq [17]. Calculations were done in vacuum. The gas phase dipole moments were calculated at the optimum geometry determined by UFF minimisation, and the minimisation was continued until the maximum force was less than  $10^{-6}$  kcal mol<sup>-1</sup> or the number of the conjugate gradients iteration was greater than  $10^4$ . Templates for molecules were created in AMMP-readable ASCII file format. The structures were visually checked on a Silicon Graphics Indigo2 computer, using the program Sybyl [18]. During the adjustment only the elec-

**Table 1** Atom type definitions

Atom type	Description
H_:	hydrogen
H_d[a]:	hydrogen attached to F/Cl/Br/I/N/O/S atom
C_3:	tetrahedral (sp <sup>3</sup> ) carbon
C_3x1[a]:	tetrahedral (sp <sup>3</sup> ) carbon attached to one F/Cl/Br/I/N/O/S atom
C_3x2[a]:	tetrahedral (sp <sup>3</sup> ) carbon attached to two F/Cl/Br/I/N/O/S atoms
C_3x3[a]:	tetrahedral (sp <sup>3</sup> ) carbon attached to three or more F/Cl/Br/I/N/O/S atoms
C_2:	trigonal (sp <sup>2</sup> ) carbon
C_2x1[a]:	trigonal (sp <sup>2</sup> ) carbon attached to one F/Cl/Br/I/N/O/S atom
C_2x2[a]:	trigonal (sp <sup>2</sup> ) carbon attached to two or more F/Cl/Br/I/N/O/S atoms
C_1:	linear (sp) carbon
C_1x[a]:	linear (sp) carbon attached to F/Cl/Br/I/N/O/S atom
C_r:	resonant (aromatic) carbon
C_rx[a]:	resonant (aromatic) carbon attached to F/Cl/Br/I/N/O/S atom(s)
O_3:	tetrahedral (sp <sup>3</sup> ) oxygen
O_3o[a]:	tetrahedral (sp <sup>3</sup> ) oxygen in carboxyl group
O_2:	trigonal (sp <sup>2</sup> ) oxygen
O_2am[a]:	oxygen in amide group
O_2n[a]:	oxygen in nitro group
N_3:	tetrahedral (sp <sup>3</sup> ) nitrogen
N_3am[a]:	nitrogen in amide group
N_2:	trigonal (sp <sup>2</sup> ) nitrogen
N_1:	linear (sp) nitrogen
N_n[a]:	nitrogen in nitro group
N_r5[a]:	resonant (aromatic) nitrogen in five membered ring
N_r6[a]:	resonant (aromatic) nitrogen in six membered ring
S_32:	tetrahedral (sp <sup>3</sup> ) sulfur with two valence bonds
P_33:	tetrahedral (sp <sup>3</sup> ) phosphorous with three valence bonds
F_, Cl, Br, I_:	halogens
F_c[a], Clc[a], Brc[a], I_c[a]:	halogens attached to carbon

[a] new atom type

tronegativity (EN) values, which were the major determinants of the partial charges, were modified. Modifications were done manually and/or automatically with the help of in-house programs.

#### Atom Types

It was found that the original choice of atom types was insufficient to reproduce the experimental data accurately. Therefore, the parameterisation was extended to include new atom types defined by the local chemical environment of the atoms. While this increased the number of degrees of freedom used to fit the data, it also resulted in a much improved fit. Even with this increase in parameters, there are far less parameters in the molecular potential than in a typical molecular mechanics potential. When an atom was surrounded by other, more electronegative atoms, the dipole moments were fit better with an altered apparent electronegativity. For ex-

ample, atom type C\_3 is an sp<sup>3</sup> carbon bonded to hydrogen or carbon atoms. C\_3x1 has one higher electronegativity atom (e.g. in CH<sub>3</sub>OH), C\_3x2 has two higher electronegativity atoms (e.g. FCH<sub>2</sub>OH), and C\_3x3 has three or more higher electronegativity atoms bound to it (e.g. CHCl<sub>3</sub>, CF<sub>2</sub>Cl<sub>2</sub>). Similar distinctions were found to be necessary with other hybridisation states of carbon and with nitrogen atoms, and for the halogens where bonding to carbon changed the apparent electronegativity. Some of the changes were introduced to fit bond-angle geometry better than the original set of atoms. These include special cases like the amide nitrogen (N\_3am), nitrate groups, and aromatic carbons. Description of the modified atom types can be found in Table 1.

- Group 1 contains hydrogen, alkali metal and halogen relatives for adjustment of F\_, Cl, Br, I\_, Li, Na, K\_ and H\_d atom types.

- Group 2 contains halogen substituted hydrocarbons for adjustment of H\_, C\_3, C\_3x1, C\_3x2, C\_3x3, F\_c, Clc, Brc, I\_c atom types.

**Table 2** Atom types and electronegativity values for AMMP

	UFF/Qeq		MOPSA	%change from UFF/QEq
H_:	4.528	H_:	6.20	+36.9
		H_d:	5.70	+25.3
C_:	5.343	C_3:	6.90	+29.1
		C_3x1:	5.88	+10.0
		C_3x2:	5.08	-4.9
		C_3x3:	4.00	-25.1
		C_2:	7.10	+32.9
		C_2x1:	6.20	+16.0
		C_2x2:	4.78	-10.5
		C_1:	8.52	+59.5
		C_1x:	4.38	-18.0
		C_r:	6.56	+22.8
		C_rx:	5.05	-5.5
O_:	8.741	O_3:	8.741	0.0
		O_3o:	7.42	-15.1
		O_2:	8.60	-1.6
		O_2am:	9.30	+6.4
		O_2n:	7.36	-15.8
N_:	6.889	N_3:	8.58	+24.5
		N_3am:	6.889	0.0
		N_2:	8.22	+19.3
		N_1:	9.26	+34.4
		N_n:	7.30	+5.9
		N_r5:	8.08	+17.3
		N_r6:	9.24	+34.1
S_:	6.928	S_32:	7.47	+7.8
		S_2:	7.32	+5.6
P_:	5.463	P_33:	6.84	+25.2
F_:	10.874	F_:	8.88	-18.3
		F_c:	7.88	-27.5
Cl:	8.564	Cl:	7.44	-13.1
		Clc:	7.23	-15.6
Br:	7.790	Br:	6.96	-10.7
		Brc:	7.17	-7.9
I_:	6.882	I_:	6.64	-2.7
		I_c:	6.96	+1.1
Li:	3.006	Li:	3.18	+5.8
Na:	2.843	Na:	2.88	+1.3
K_:	2.421	K_:	2.44	+0.8

- Group 3 contains monofunctional organic compounds e.g. alcohols, ethers, amines, thiols and sulphides for adjustment of O\_3, N\_3, S\_32 and P\_33 atom types.

- Group 4 contains double bonded and aromatic carbon compounds for adjustment of C\_2, C\_2x1, C\_2x2, C\_r and C\_rx atom types.

- Group 5 contains oxo compounds e.g. aldehydes, ketones, carbonic acids, esters and amides for adjustment of O\_2, O\_3o, O\_2am and N\_3am atom types. The out-of-plane parameter of the N\_3am atom type was also set to achieve the planar geometry of the amide group.

- Group 6 contains double bonded nitrogen, sulphur, triple bonded carbon and some heterocyclic molecules for adjustment of C\_1, C\_1x, N\_1, O\_2n, N\_n, N\_2, S\_2, N\_r5, N\_r6 atom types. The out-of-plane parameter of the N\_n atom type was also set to achieve the planar geometry of the nitro group.

The groups were selected to include at least three representative molecules for each atom type. The training set (Group 1-6) counts 160 molecules. The experimental gas phase dipole moments were taken from handbooks [19, 20] or from the literature [9, 21].

**Table 3** Slope of the best fitted line using different parameter sets.

	UFF/QEq	MOPSA	Gasteiger-Marsili	Pullman
Group1	1.16	1.01	0.48	-
Group2	3.34	0.95	0.36	0.65
Group3	1.56	0.88	0.34	0.37
Group4	5.33	0.92	1.31	0.61
Group5	0.92	1.02	1.30	0.82
Group6	2.00	0.93	0.92	0.30
Group1-6	1.05	1.00	1.07	0.57
Group7	1.92	0.92	1.22	0.68

**Table 4** Intersection (D) of the best fitted line using different parameter sets

	UFF/QEq	MOPSA	Gasteiger-Marsili	Pullman
Group1	0.54	-0.11	0.41	-
Group2	1.12	-0.01	0.93	0.28
Group3	0.15	0.05	0.76	0.55
Group4	-0.78	0.15	0.11	0.38
Group5	2.22	-0.12	-0.58	-0.06
Group6	0.36	-0.06	1.03	1.27
Group1-6	2.74	-0.09	0.15	0.47
Group7	1.61	0.11	0.11	0.54

**Table 5** Regression coefficient of the best fitted line for different parameter sets

	UFF/QEq	MOPSA	Gasteiger-Marsili	Pullman
Group1	0.97	0.99	0.54	-
Group2	0.82	0.95	0.48	0.48
Group3	0.49	0.86	0.31	0.27
Group4	0.87	0.93	0.61	0.55
Group5	0.55	0.98	0.96	0.94
Group6	0.51	0.92	0.43	0.34
Group1-6	0.57	0.98	0.62	0.61
Group7	0.48	0.85	0.70	0.63

**Table 6** Root mean square deviations (D) of calculated dipole moments from the experimental dipole moments using different parameter sets

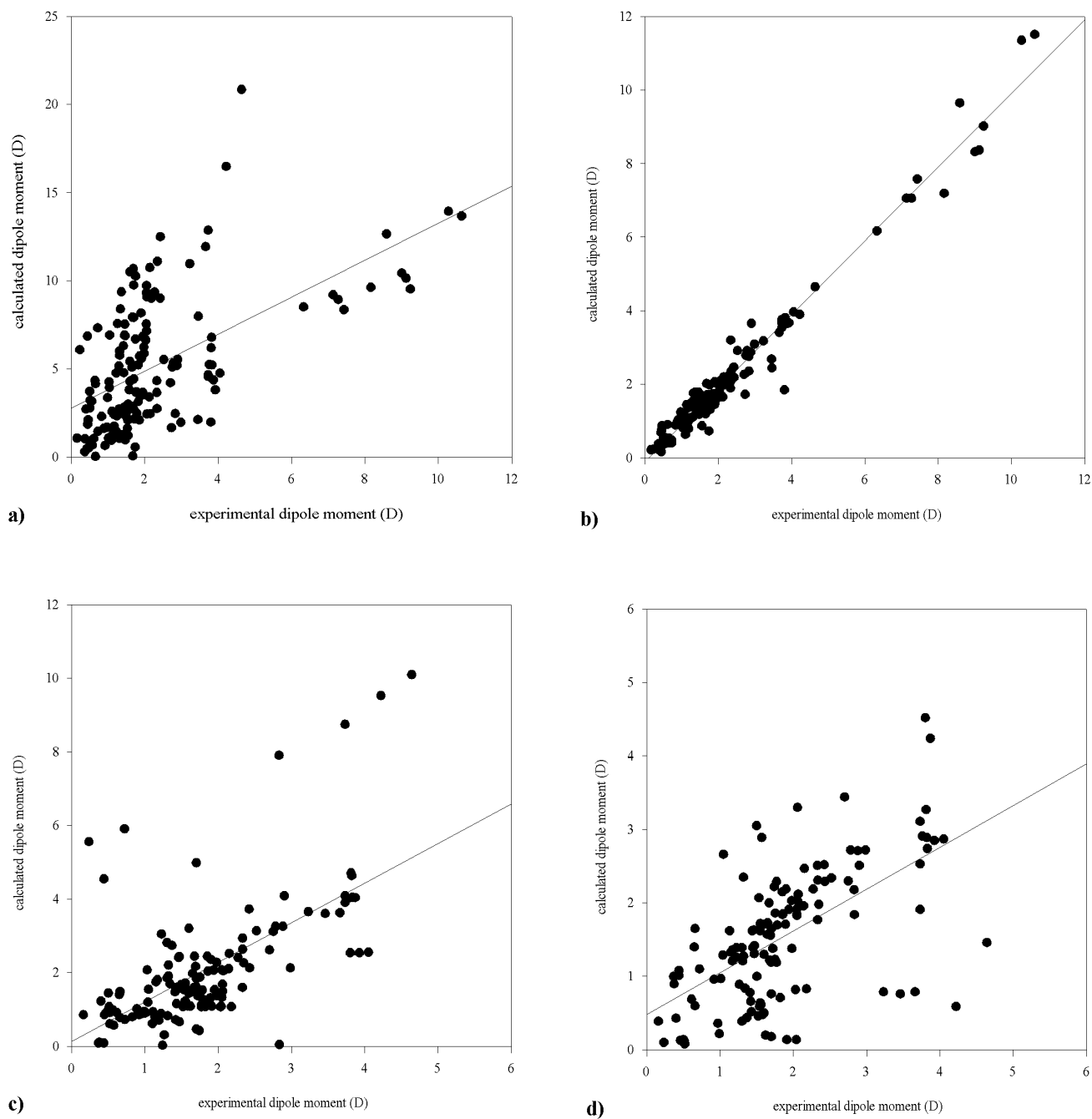
	UFF/QEq	MOPSA	Gasteiger-Marsili	Pullman
Group1	1.77	0.55	0.47	-
Group2	5.16	0.21	0.56	0.80
Group3	1.31	0.22	0.46	0.55
Group4	5.94	0.21	1.15	0.58
Group5	2.36	0.19	0.47	0.63
Group6	4.85	0.52	2.44	1.34
Group1-6	4.13	0.37	1.33	0.90
Group7	5.07	0.59	1.39	0.94

### Benchmarks

A test group (Group 7) including 149 molecules was also created to check the validity of the modified parameters. Two other methods for calculation of dipole moments were used for comparison. The Gasteiger-Marsili and Pullman methods

were used as built in features of Sybyl [18]. The dipole moments were calculated at the above mentioned optimised geometry. The number of  $\pi$ -calculations was set to 20.

The conformational energies of halocyclohexanes (using the modified parameters) and tetrapeptides (using the MOPSA parameters merged with the latest AMMP parameter set de-

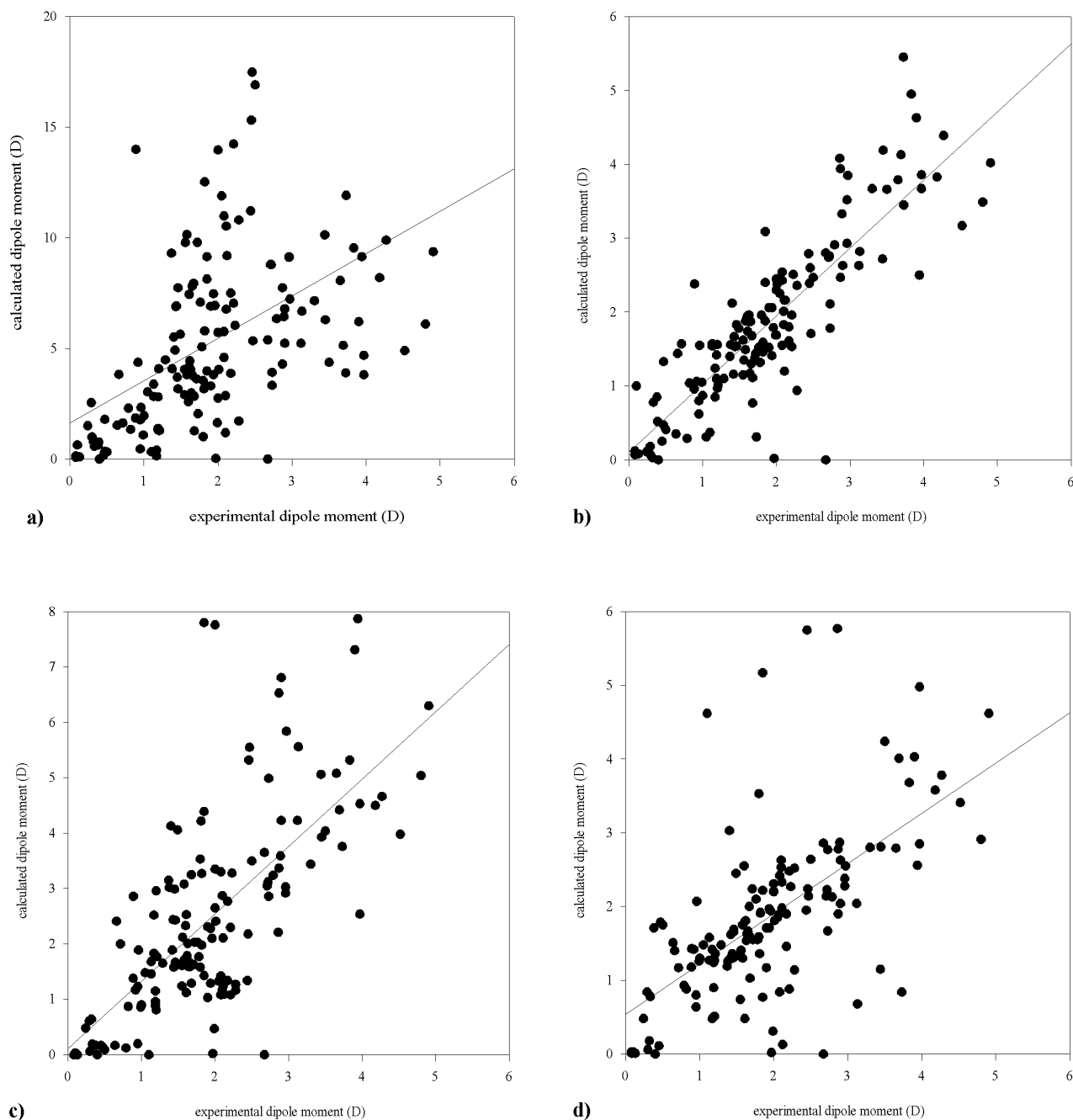


**Figure 1** Calculated dipole moments versus experimental values for the training set (Groups 1-6). a) UFF/Qeq. b) MOPSA. c) Gasteiger-Marsili. d) Pullman.

veloped for proteins) were tested. The tetrapeptide test suite [22] was retrieved from the author's web site [23], and the structures of halocyclohexanes were retrieved from the Computational Chemistry List web site [24] and used without any modification.

## Results and discussion

Calculation of the dipole moments with the original QEq values showed that the computed values correlated with the empirical ones, but it seemed possible to improve the correlation with adjustments of the ENs. We focused on improve-



**Figure 2** Calculated dipole moments versus experimental values for the test compounds (Group 7). a) UFF/Qeq. b) MOPSA. c) Gasteiger-Marsili. d) Pullman

ment of parameters for C, H, N, O atoms, due to our interest in proteins, and due to the limited availability of experimental dipole moments. Molecules included in Groups 1-6 formed homologous series with a wide variety of functional groups. The database contained inorganic and organic compounds, hydrocarbons, molecules with heteroatoms, molecules with single and double bonds, as well as linear and cyclic com-

pounds. We chose the simplest members from each family of molecules to eliminate the interaction of several functional groups. No similar strategy was applied for the test group (Group 7). The size of the database was chosen to be large enough to represent a wide variety of molecules, but to be small enough to do several calculations needed for the iterative procedure: totals of 160 and 149 molecules are included

in the training database (Groups 1-6) and the test database (Group 7), respectively. During the protocol, EN was handled as a free variable without any physically meaningful property. However, there were some series where changes in the EN and/or in the self coulomb energy were not sufficient to get a good correlation. In these cases, new atom types were defined as shown in Table 1. This concept is somewhat different to those of QEq and UFF, but these atom types have a chemical sense, and make the dipole moments more predictable. A similar approach was used for parameters of bonded molecular properties [13]. The ENs of the different series were set in stepwise manner from Group 1 to Group 6, which means that the new ENs from one group of atom types were used in the next stage to set EN for another group of atom types. This method was applied because of the limited resources of the computation.

Dipole moments of inorganic halogenides were slightly overestimated using the original parameter set. The correlation coefficient ( $r$ ) between calculated and measured dipole moments as well as slope ( $m$ ) and intersection ( $b$ ) of the fitted line were selected to be optimised: parameters were adjusted to increase the correlation coefficient, and to make an ideal slope of 1.0 with zero intercept. The EN of hydrogen halogens and alkali metal atom types was adjusted using these criteria. The changes of EN needed to achieve better correlation were in the range from -18.3% ( $F_-$ ) to +25.3% ( $H_d$ ) and they varied gradually according to the position of the elements in the periodic table (Table 2). Atoms from the first row of the periodic table often show some special properties, therefore it is not surprising that the values for these two atoms differ most from their original values based on the periodic table. This tendency also appeared in Group 2, where the organic halogenides were used to adjust the EN of tetrahedral carbon and the halogens which were bonded to it: the changes ranged from -27.5% to +36.9% in case of  $F_c$  and  $H_+$  atom types, respectively. Not surprisingly, the parameters of halogens in inorganic salts differ from the values of the halogens in organic compounds. The carbons bonded to halogens also had to be defined as separate atom types (Table 1), and the ENs of these carbons were strongly dependent on the number of halogen atoms attached to them (Table 2). The dipole moments using the original parameters are greatly overestimated in this group: the slope of the fitted line is 3.34, and the intersection is 1.12 D. With the addition of the new atom types, the values for the slope and intercept changed to 0.95 (Table 3) and -0.01 D (Table 4), respectively, and the correlation coefficient increased from 0.82 to 0.95 (Table 5). By further analysing dipole moments in Group 1 and 2, the original parameters were shown to work well for the inorganic compounds, but larger errors were observed when they were applied for organic molecules. The overestimation was compensated by increasing the EN values of the hydrogen and the carbon atom by approximately 30% (Table 2).

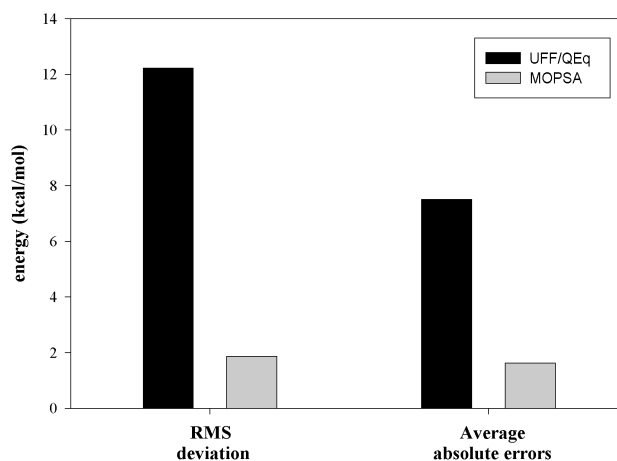
Molecules of Group 3 were used to adjust EN for single bonded oxygen, nitrogen, sulphur and phosphorous atom types. The original parameters worked well for the amines, but the dipole moments of alcohols, thiols, ethers and sulphides were slightly overestimated. This tendency reversed,

when new parameters for carbon and hydrogen atoms were taken into account. Therefore, the value for the  $N_3$  atom type was the most altered to achieve a better correlation (Table 2). The dipole moments of phosphorous halogenides were greatly overestimated, and in addition to the adjustments of halogenides described in the previous step, an increase of the EN of phosphorous atom was also needed. These changes raised the correlation coefficient from 0.49 to 0.86 (Table 5).

In case of trigonal ( $sp^2$ ) and resonant (aromatic) carbon atom containing molecules (Group 4), it was also necessary to divide the original atom types into two different types: carbons attached to atoms with low or high EN (Table 1). The fit was relatively good using the original parameters, but the slope (5.33) was substantially different from unity. Therefore, the EN was increased for  $C_r$ ,  $C_2$ ,  $C_{2x1}$  and decreased for  $C_{rx}$ ,  $C_{2x2}$ . The new slope after adjustment was 0.92 and the new correlation coefficient was 0.93.

Using the original parameters for calculation of dipole moments of oxo compounds, the molecules were separated into three groups, based on the chemical nature of them: 1. amides; 2. acids, esters; and 3. aldehydes, ketones. The calculated dipoles correlated relatively well with experimental ones in their own groups, but no single line could be fitted. The difficulties were not resolved by adjusting the existing atom types, and new oxygen and nitrogen types needed to be defined, e.g. nitrogen in the amide group, oxygen in the carboxyl group (Table 1). The resulting dipole moments correlated much better than the original ones (Table 5). Some separations were also seen but they agreed with the experimental characteristics of dipole moments of these compounds.

Group 6 contains heterocyclic compounds, molecules with double bonded heteroatoms and triple bonded carbon atoms. The requirement of the definition of a separate atom type for linear carbon atoms attached to atoms with large EN, was also obvious in this group. The EN of the tetrahedral, trigonal, and linear carbon atoms decreases systematically with increased number of connected atoms with large EN. Not



**Figure 3** Summary of conformational energy calculation on halocyclohexanes



**Table 7** Results of the tetrapeptide benchmark.

	SP4	SP5
Average RMSD on geometry (nm)	0.055	0.036
Number with RMSD > 0.06 nm	4	0
Energy RMSD (kcal/mol)	1.58	1.21
Number of Pairwise errors > 3 kcal/mol	8	2
Maximum Pairwise error (kcal/mol)	4.62	3.13

surprisingly, nitrogen in the nitro group needed to be defined as a separate type. Interestingly, the largest differences were seen in the EN of nitrogen in five membered (8.08) and six membered (9.24) heteroaromatic rings. However, these values have the greatest uncertainty because of the limited numbers of accessible experimental dipole moments. A correlation coefficient of  $r=0.92$  was achieved with the inclusion of new atom types (Table 5).

#### Performance on a test set

Figure 1 shows the calculated dipole moments versus the experimental ones for the 160 molecules in the training set (Groups 1-6) using the original and the modified parameter set of AMMP. In order to test the new parameters, another 149 molecules were built and their dipole moments were calculated (Figure 2). The modified parameters resulted in better values in every criteria (regression, slope, intersection, rms) as compared to the original ones (Table 3-6). However, in some cases the method failed to predict the correct values. In order to evaluate whether this feature was characteristic of the method, or it was because some special molecular properties cannot be explained in molecular mechanics theory, we compared the method to two other widely used methods for calculation of the partial charges. The Gasteiger-Marsili (GM) and Pullman (P) methods implemented in the Sybyl program package were chosen. Not surprisingly, MOPSA was the best among the four methods for the molecules in the training set (Table 3-6, Figure 1). More importantly, MOPSA was also the best parameter set for the test compounds (Table 3-6, Figure 2). The largest deviations from the experimental dipole moments were obtained in the cases of halogenous compounds calculated by UFF/QEq, Gasteiger-Marsili and Pullman methods, but these errors were generally corrected by MOPSA.

The modified parameter set with adjustments of EN and the addition of 19 new atoms types to the 19 original types has produced an increase from 0.57 to 0.98 in correlation coefficient for the calculated and observed dipole moments of the training set of 160 compounds and an increased correlation from 0.48 to 0.85 for the 149 test molecules. The modified parameter set gave better predictions of the dipole moments of the studied molecules than the original QEq parameter set, the Gasteiger-Marsili parameter set, and the Pullman parameter set. Of course, this result is valid only for the majority of the molecules which are characterised in this study, and not for every molecule.

#### Performance improvement on a realistic benchmarks

The MOPSA parameters dramatically improved the fit between calculated and observed dipole moments. They also improved the quality of the molecular mechanics potential set. UFF with charges showed poor result in an extensive comparison of several force fields tested by Gundertofte et. al. [16], especially in case of halogenated cyclohexanes. The structures were downloaded from the CCL's test suite [24], and were used without further minimisation. Our calculations showed that our modified parameters performed better than the original one (Figure 3), however the accuracy was not yet reached the level of some mature force fields (e.g. MM2 [16]).

The MOPSA parameter set was merged with the latest AMMP parameter set (set SP4) to generate a new standard set SP5. The SP4 had been independently optimised from the initial UFF values to improve the agreement with experimental geometry, IR spectral data, and isotope effects for molecular systems related to biological molecules [13, 14, 17]. The Beachy et. al. [22] tetrapeptide benchmark was retrieved from the author's web site [23] and both the SP4 and SP5 potentials were tested in this benchmark. The charges were generated for the extended conformation (conformation 1 in the benchmark) and the structures were extensively optimised with conjugate gradients. The results are shown in Table 7. While the SP4 set performed reasonably well, the improved potential set which uses the MOPSA parameters performed dramatically better and results were comparable with the best potential sets described by Beachy et. al. [22].

These results suggest that the improvement of the charge generation parameters not only improved the ability to reproduce dipole moments, but also improved the accuracy of the potential set. The new parameters will be valuable for future studies on the electrostatic contribution of nonbonded interactions which plays an important role in the macromolecular modelling, e.g. in the enzyme-substrate, receptor-ligand, protein-protein, protein-nucleic acid, etc. interactions.

**Acknowledgements** This work was supported in part by the Hungarian Science and Research Fund OTKA F25807 (to P.B.), OTKA T22670 (to J.T.), the Hungarian Ministry of Culture and Education FKFP 1318/97 (to J.T.), the United States Public Health Service Grant AI41380 (to I.T.W. and R.W.H.) and by the Fogarty International Center AIDS-FIRCA grant TW 01001 (to I.T.W. and J.T.).

**Supplementary Material Available** Dipole moments of molecules of Group 1-7 calculated by UFF, MOPSA, Gasteiger-Marsili and Pullman method along with the experimental values, and conformational energies of halocyclohexanes calculated by AMMP using UFF/QEq and MOPSA force fields along with the experimental values can be found in formatted ASCII tables.

---

## References

1. Rappé, A.K.; Casewit, C.J.; Colwell, K.S.; Goddard III, W.A.; Skiff, W.M. *J. Am. Chem. Soc.* **1992**, *114*, 10024.
2. Casewit, C.J.; Colwell, K.S.; Rappé, A.K. *J. Am. Chem. Soc.* **1992**, *114*, 10046.
3. Casewit, C.J.; Colwell, K.S.; Rappé, A.K. *J. Am. Chem. Soc.* **1992**, *114*, 10035.
4. Straatsma, T.P.; McCammon, J.A. *Annu. Rev. Phys. Chem.* **1992**, *43*, 407.
5. Wiberg, K.B.; Rablen, P.R. *J. Comp. Chem.* **1993**, *14*, 1504.
6. Coppens, P. *Annu. Rev. Phys. Chem.* **1992**, *43*, 663.
7. Mulliken, R.S. *J. Chem. Phys.* **1962**, *36*, 3428.
8. Bader, R.W.F. *Acc. Chem. Res.* **1985**, *18*, 9.
9. Storer, J.W.; Giesen, D.J.; Cramer, C.J.; Truhlar, D.G. *J. Comp. Aided Mol. Des.* **1995**, *9*, 87.
10. Weiner, S.J.; Kollman, P.A.; Nguyen, D.T.; Case, D.A. *J. Comp. Chem.* **1986**, *7*, 230.
11. Harrison, R.W. *J. Comp. Chem.*, **1993**, *14*, 1112. AMMP Internet site: <http://asterix.jci.tju.edu/ammp.html>
12. Harrison, R.W.; Weber, I.T. *Prot. Eng.* **1994**, *7*, 1353.
13. Weber, I.T.; Harrison, R.W. *Prot. Eng.* **1996**, *8*, 679.
14. Weber, I.T.; Harrison, R.W. *Prot. Sci.* **1997**, *6*, 2365.
15. Rappé, A.K.; Goddard III, W.A. *J. Phys. Chem.* **1991**, *95*, 3358.
16. Gundertoft, K.; Liljefors, T.; Norrby, P.; Petterson, I. *J. Comp. Chem.* **1996**, *17*, 429.
17. Harrison, R.W.; Chatterjee, D.; Weber, I.T. *Proteins Struct. Funct. Genet.* **1995**, *23*, 463.
18. Sybyl molecular modeling package, version 6.3, Tripos Inc., St. Louis, MO.
19. Weast, R.C. *Handbook of chemistry and physics*, 69th edition, CRC Press: Boca Raton, Florida, 1988.
20. Weast, R.C. *Handbook of chemistry and physics*, 72th edition, CRC Press: Boca Raton, Florida, 1991.
21. Gerber, P. R. *J. Comp. Aided Mol. Des.*, **1998**, *12*, 37.
22. Beachy, M.D.; Chasman, D.; Murphy, R.B.; Halgren, T.A.; Friesner, R.A. *J. Am. Chem. Soc.* **1997**, *119*, 5908.
23. <http://www.chem.columbia.edu/~beachy/structures.html>
24. Computational Chemistry List, Force Field Evaluation Suite: [http://www.ccl.net/cca/data/ff\\_evaluation\\_suite/](http://www.ccl.net/cca/data/ff_evaluation_suite/)