J Mol Model (2011) 17:1435–1444
DOI 10.1007/s00894-010-0845-y

ORIGINAL PAPER

Hybridization-displaced charges for amino-acids: a new model using two point charges per atom along with bond-center charges

Neha Agnihotri · P. C. Mishra

Received: 2 July 2010 / Accepted: 1 September 2010 / Published online: 19 September 2010

Abstract A new charge distribution is proposed for the amino acids where each atom is associated with two point charges while each bond center is associated with one point charge. Centroids of charges arising due to atomic orbital hybridization called hybridization-displaced charges (HDC) and those located at the atomic sites and bond centers obtained by a modified form of the Mulliken scheme were combined. The density matrix calculations required for this analysis were performed at the B3LYP/6-31G** level of density functional theory. The combination of HDC centroids with the modified Mulliken charges was found to yield dipole moments and surface molecular electrostatic potentials (MEP) of the amino acids in good agreement with those obtained by rigorous DFT calculations or those obtained using the MEP-fitted CHelpG charges. This study shows that the combination of HDC centroids with the modified Mulliken charges is significantly superior to the conventional Mulliken charges.

Keywords Atomic orbital hybridization · Charge distribution · Dipole moment · Hybridization-displaced charges · Molecular electrostatic potential

Introduction

Molecular electrostatic properties and intermolecular interactions can be conveniently studied using point charges which are widely used in different types of studies [1–4]. Such charges located at the atomic sites are obtained by theoretical or experimental methods or by fitting to surface molecular electrostatic potentials (MEP) [5–10]. The various theoretical methods partition the total electron density distribution in different ways. A commonly used method of this category is the Mulliken population analysis scheme [5]. Various experimental models are used to interpret experimental observations on X-ray diffraction intensity distribution, IR intensity distribution and NMR spectral features and yield point charges in molecules [9]. Theoretical methods partition the continuous electron density and place point charges exactly at the nuclear sites each. It cannot be justified since on molecular formation, the electronic charge centers may be significantly shifted from the nuclear sites. The existence of lone pairs clearly shows that atomic site-based point charge schemes are deficient. Consideration of charges at off-nuclear sites has been found to produce improved results [2, 3]. A detailed theoretical study has revealed that many covalently bonded atoms of V–VII groups have their surfaces associated with regions of both positive and negative MEP [11]. It shows that the usual practice to associate single point charges with atoms in molecules is not always valid. In the new charge distribution scheme discussed here, every atom in molecules is associated with two point charges in stead of one.

The theoretical electron density partitioning schemes do not reproduce molecular electrostatic properties satisfactorily and consistently. A point charge distribution cannot be considered to be satisfactory until it reproduces both molecular dipole moments and surface MEP features satisfactorily. Surface MEP serves as a sensitive electrostatic property to test accuracy of charge distribution [12, 13]. MEP-fitted charges generally reproduce rigorously calculated (e.g.

Electronic supplementary material The online version of this article (doi:10.1007/s00894-010-0845-y) contains supplementary material, which is available to authorized users.

N. Agnihotri · P. C. Mishra (✉)
Department of Physics, Banaras Hindu University,
Varanasi 221 005, India
e-mail: pcmishra_in@yahoo.com

SCF) molecular dipole moments and surface MEP satisfactorily. However, these charges are not satisfactory for homonuclear diatomic and π-electron systems. For such systems, inclusion of off-atomic site charges is necessary [14]. We have introduced the concept of hybridization-displaced charges (HDC) and used the same to calculate electrostatic properties of several molecules including radicals and ions [2, 14–21]. HDC arise due to atomic orbital hybridization in the molecular environments [2]. It has been shown that an improved point charge distribution can be obtained by combining HDC with Mulliken charges [15, 20, 21].

It would be desirable to modify the Mulliken scheme so that better charges than the conventional Mulliken charges may be obtained and HDC may be combined with the same. There are two, six and 12 HDC components associated with each hydrogen atom, second and third row heavy atoms respectively when the 6-31G** basis set is used and the procedure adopted here is employed [15, 19–21]. With larger basis sets, the numbers of HDC components would increase. While increased numbers of HDC components would be expected to correspond to increased accuracy of charge distribution, it would be inconvenient in certain applications, e.g., simulation studies, as larger numbers of HDC components would require significantly increased computational effort. For C, N and H containing molecules, e.g., the amino acids, the HDC components are located at small distances from the atoms [15, 20, 21]. Therefore, in such cases, the HDC components associated with the different atoms may be replaced by a single point charge located at their centroids each. However, this replacement can be accepted only if it is found not to affect the accuracy of description of molecular electrostatic properties with respect to that obtained using individual HDC components significantly. The present study shows that this approximation is valid for the amino acids. In turn, it implies that single point charges located at the HDC centroids near the atoms can be used in place of the various individual HDC components. However, use of charge centroids in place of the individual HDC components would not be justified for metal-containing molecules where some of the HDC distances from the corresponding atomic sites are quite large [15, 19]. In a previous study performed on amino acids using the AM1 semiempirical method [16], HDC were combined with Löwdin charges. In this work [16], for certain sites, quite appreciable differences were found between the MEP values obtained using the (HDC + Löwdin) charges and the CHelpG charges computed at the HF/3-21G level. Therefore, an improved charge distribution including HDC was highly desirable. This objective is met in the present study.

Computational methodology

Definitions

The total dipole moment (μ^t) of a molecule is a vector sum of two components, i.e., μ^{net} and μ^h [22]. The net charges located at the atomic sites give rise to μ^{net} while the shifted charges from the atomic sites due to atomic orbital hybridization give rise to μ^h [22]. While μ^{net} can be obtained using the Mulliken approximation or a variant of it, μ^h can be obtained as discussed below. The α^{th} Cartesian component of the total contribution of a given atom to the hybridization dipole moment μ^h of the given molecule can be written as follows [2, 14–21]

$$\mu_\alpha^h = (KD_\alpha)(Q_\alpha/K) = D_\alpha' Q_\alpha' \quad (\alpha = x, y, z) \tag{1a}$$

where,

$$D_\alpha = (ns|\alpha|mp_\alpha) \tag{1b}$$

(n = 1, 2, 3 etc. and m = 2, 3, 4 etc.)

$$Q_\alpha = -P_{ns,mp_\alpha} \tag{1c}$$

In the above equations, n and m are the principal quantum numbers associated with the s and p_α atomic orbitals of the given atom respectively, while P is the density matrix obtained from the linear combination coefficients of atomic orbitals in molecular orbitals. K in Eq. (1a) is a parameter that may be adjusted to reproduce the value of a selected physical property. D_α' is the distance of charge Q_α' along the direction α (α=x, y or z). The magnitude of the total hybridization dipole moment μ^h can be obtained as

$$\mu^h = \sqrt{\left(\left(\mu_x^h\right)^2 + \left(\mu_y^h\right)^2 + \left(\mu_z^h\right)^2 \right)} \tag{2}$$

where, μ_x^h, μ_y^h and μ_z^h are the three Cartesian components of μ^h.

The resultant displacement D of the Cartesian components D_x, D_y and D_z would be obtained as

$$D = \sqrt{\left(D_x^2 + D_y^2 + D_z^2\right)}. \tag{3}$$

Using contractions of Gaussians, one can calculate the above-mentioned quantities accurately. However, we evaluated these quantities using Slater type atomic orbitals (STOs) for the sake of computational convenience. Using STOs, we get

$$D_\alpha = \frac{2^a \zeta_s^b \zeta_{p_\alpha}^c (a!)}{\sqrt{3}\sqrt{(2n(s)!)}\sqrt{(2m(p_\alpha)!)}d^{a+1}} \quad (\alpha = x,y,z) \qquad (4)$$

where $a = n(s) + m(p_\alpha) + 1$, $b = n(s) + 1/2$, $c = m(p_\alpha) + 1/2$, $d = \zeta_s + \zeta_{p\alpha}$. The principal quantum numbers of the s and p_α (α=x,y,z) atomic orbitals are given by n(s) and $m(p_\alpha)$ while the corresponding Slater exponents are represented by ζ_s and $\zeta_{p\alpha}$ respectively. As Eq. (4) reveals, D_α (α=x,y,z) depends only on the exponents of the s and p_α STOs for given n(s) and $m(p_\alpha)$ values. Thus, for given s and p_α orbitals having specific Slater exponents, D_α (α=x,y,z) and hence the resultant D are fixed. Since $\zeta_{px}=\zeta_{py}=\zeta_{pz}$, we get $D_x=D_y=D_z$. Then we get the total HDC denoted by Q as

$$Q = \frac{\mu^h}{D} = \sqrt{\frac{\left(Q_x^2 + Q_y^2 + Q_z^2\right)}{3}}. \qquad (5)$$

Thus we get Q as the root-mean-square of Q_x, Q_y and Q_z. Eq. (5) would be valid even if the p_x, p_y and p_z orbitals belong to a different quantum number than that to which the s orbital belongs.

Equation (1a) shows that for a given atom type, μ_α^h is proportional to Q_α as D_α is constant. The direction of displacement of the electronic charge Q (total HDC) with respect to the atom under consideration can be obtained using the components of the corresponding vector μ_α^h of the atom which is the basic quantity. Thus the direction of displacement of Q with respect to the given atom can be obtained in terms of the following angles

$$\phi = \tan^{-1}\left(\frac{\mu_y}{\mu_x}\right), \qquad (6)$$

And

$$\theta = \cos^{-1}\left(\frac{\mu_z}{\mu^h}\right). \qquad (7)$$

Computational procedure

Geometries of the canonical forms of the amino acids [23] were optimized at the B3LYP/6-31G** level of density functional theory [24–26]. Density matrices and overlap integrals for the amino acids thus obtained were used to calculate HDC. A modified form of the Mulliken scheme [21] (m-Mulliken) where 40% of the total overlap charge associated with a bond is assigned to the bond center and 30% of the same is placed at each of the bonded atoms, was employed, and the charges so obtained were combined with the HDC centroids. We also attempted using the point charges obtained by natural population analysis (NPA) [7] in place of the Mulliken charges, but this approach did not yield better results than those obtained using the Mulliken charges.

With the 6-31G** basis set, HDC consist of two point charges associated with each hydrogen atom and six point charges associated with each heavy atom (C or N). In this study, in the HDC calculations, mixing of only s and p types of atomic orbitals was considered, and d or higher atomic orbitals were not included. As an approximation, to calculate HDC, the density matrix elements were considered as having been obtained using STOs as basis functions. This approximation is justified as STOs are similar to the corresponding contractions of Gaussians. For the STOs of the innermost (n=1) shells of the different atoms, the Slater exponents were taken to be the same as those given in the literature [27]. However, the ζ values of the higher shells were adjusted in order to obtain the best possible agreement between the dipole moments of several molecules computed using the combination of HDC and modified Mulliken charges and those obtained by rigorous DFT calculations. It must be emphasized that we made no changes in the exponents of Gaussians used in the DFT calculations. We only adjusted ζ values of STOs for HDC calculations. The parameter K was fixed for different atoms employing the criterion of best possible agreement between the MEP values at the van der Waals surfaces of the molecules obtained using the combination of m-Mulliken charges with HDC and those obtained using the MEP-derived CHelpG charges. The adjusted values of K and ζ for different atoms were the same as those reported earlier [19]. It is to be noted that due to the displacement of the negative charge (Q) from an atomic site, an equal amount of positive charge would be acquired by it so that the overall charge state (e.g., neutral, anionic or cationic) of the molecule under consideration is not changed.

A limitation of the present HDC scheme has to be noted. It preserves the contribution of each atom to the total molecular dipole moment approximately (due to the use of STOs along with the parameters ζ), but it does not preserve the atomic contributions to the higher molecular electric moments, e.g., quadrupole moments. Therefore, wherever quadrupole and/or higher electric moments would play an important role, the present HDC scheme would show weakness. Nonetheless, it would emulate the chemical

environments of atoms in molecules in a much better way than what is possible by the atomic site-based point charges. The Windows version of the Gaussian 03 program (G03W) [28] was used to obtain density matrices and overlap integrals while the GaussView program [29] was used to visualize molecular structures and surface MEP maps obtained using the continuous electron density distribution. HDC and surface MEP maps based on (HDC + m-Mulliken) charges were obtained using programs developed in our laboratory.

Results and discussion

The structures of the canonical forms of all 20 amino acids [23] along with the adopted atomic numbering scheme are shown in Fig. 1. The canonical forms of the amino acids were considered here as they combine with one another to form polypeptides. The zwitterionic forms of the amino acids which are formed from the canonical forms in aqueous media [23] were not considered since they are not the building blocks of polypeptides. Use of the HDC

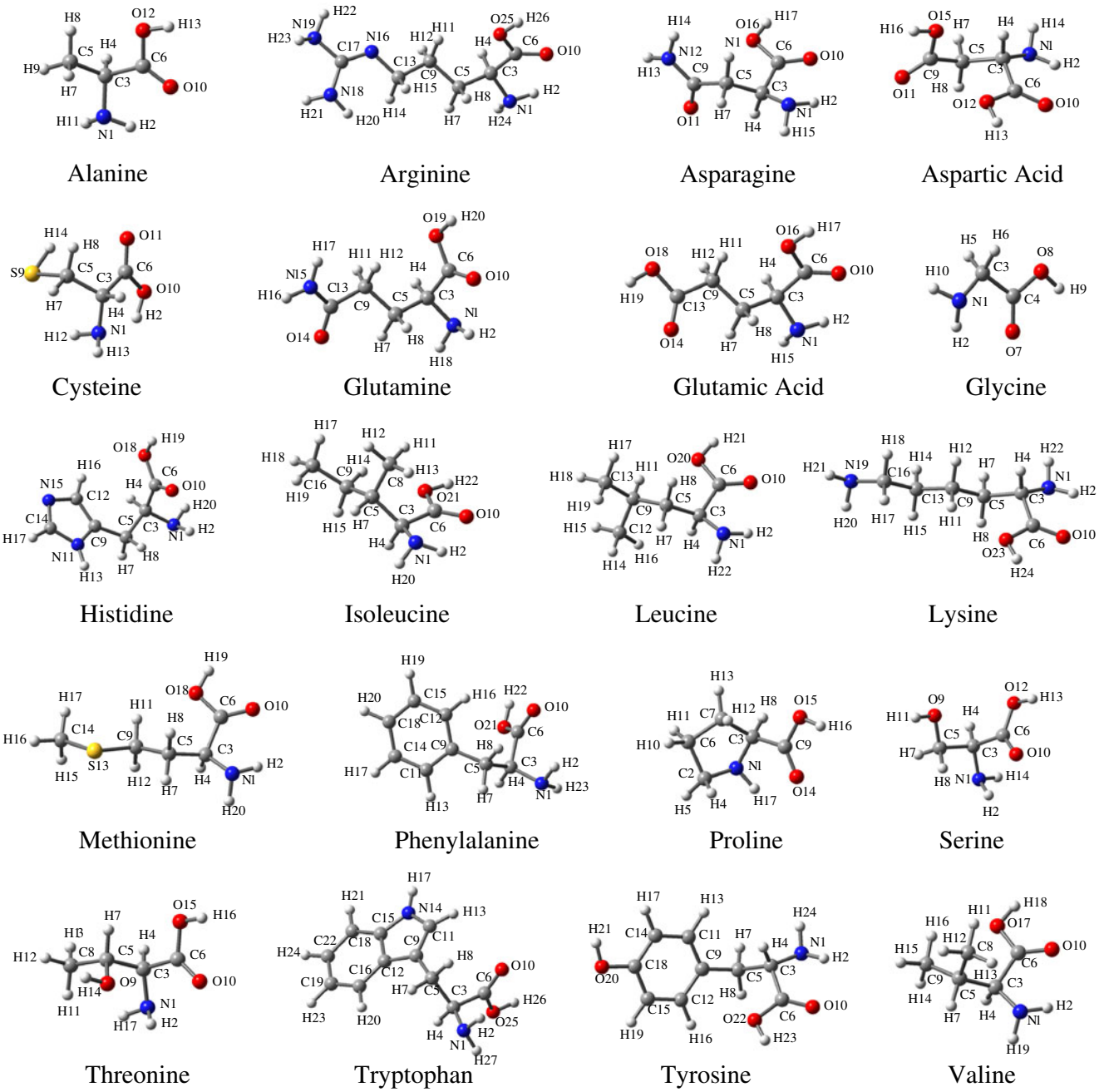

Fig. 1 Structures of the amino acids optimized at the B3LYP/6-31G** level of theory and the adopted atomic numbering scheme

centroids in place of all the HDC components was found not to produce any significant difference in the calculated dipole moments or MEP values. Therefore, the results obtained using the HDC centroids only are presented here. The calculated dipole moments and minimum surface MEP values of the amino acids presented in Tables 1 and 2 respectively reveal the following features.

Dipole moments

The calculated dipole moments (μ) of all 20 amino acids are presented in Table 1. In this table, the DFT dipole moments (μ^{DFT}) were obtained as expectation values of the electric dipole operator at the B3LYP/6-31G** level of theory while the other dipole moments were obtained using the different types of point charges (CHelpG, Mulliken, m- Mulliken) and the classical definition of dipole moment. The last column of Table 1 presents the percentage changes in dipole moments obtained in going from the Mulliken to the (HDC + m-Mulliken) charges. The sign of the %change was taken to be positive if the change ($\mu^{m\text{-}Mulliken+HDC}$- $\mu^{Mulliken}$) was in the direction of μ^{DFT} or μ^{CHelpG} with respect to $\mu^{Mulliken}$ (increasing or decreasing), otherwise it was taken to be negative.

From the dipole moments presented in Table 1, we note the following: (i) In 14 out of the 20 cases, the % changes are positive implying that in these cases, in going from Mulliken to (HDC + m-Mulliken) charges, the dipole moments are corrected. However, in four of these cases (glutamic acid, histidine, lysine and phenylalanine), the corrections are somewhat excessive. (ii) In five cases (alanine, glycine, leucine, isoleucine, valine), the % changes are negative and ~-10% each while in one case (arginine), the %change is negative but quite small. (iii) In some cases, the %changes are positive and quite large. For example, in going from Mulliken to (HDC + m-Mulliken) charges, in the cases of methionine, serine and tryptophan, the %changes are ~48%, ~43% and ~251% respectively though the desirable changes in the latter two cases were even larger. These results show that the (HDC + m-Mulliken) charges are usually much better than the Mulliken charges.

Table 1 Dipole moments (μ^x, x=DFT, CHelpG, Mulliken, m-Mulliken + HDC) (Debye) of Amino acids calculated using various methods or charges at the B3LYP/6-31G** level of theory in gas phase[a]

S. No.	Amino acid[b]	μ^{DFT}	μ^{CHelpG}	$\mu^{m\text{-}Mull+HDC}$	μ^{Mull}	% Change[c]
1.	Alanine	1.96	1.98	2.43	2.23	-9.0
2.	Arginine	3.10	3.14	3.19	3.14	-1.6
3.	Asparagine	2.80	2.80	2.60	2.50	4.0
4.	Aspartic acid	2.55	2.59	2.54	2.19	16.0
5.	Cysteine	4.45	4.46	4.33	4.25	1.9
6.	Glutamine	4.03	4.02	3.87	3.19	21.3
7.	Glutamic acid	2.83	2.85	3.13	2.51	24.7
8.	Glycine	2.00	1.99	2.45	2.22	-10.4
9.	Histidine	2.49	2.44	2.61	2.25	16.0
10.	Isoleucine	1.87	1.87	2.18	1.99	-9.6
11.	Leucine	1.93	1.96	2.40	2.20	-9.1
12.	Lysine	2.42	2.45	2.61	2.12	23.1
13.	Methionine	3.04	3.03	2.97	2.01	47.8
14.	Phenylalanine	1.45	1.34	1.60	1.16	37.9
15.	Proline	1.58	1.53	1.94	2.05	5.4
16.	Serine	0.38	0.38	0.87	1.53	43.1
17.	Threonine	3.79	3.79	3.97	4.05	2.0
18.	Tryptophan	2.43	2.35	1.37	0.39	251.3
19.	Tyrosine	2.55	2.58	2.63	2.07	27.1
20.	Valine	1.84	1.87	2.19	2.00	-9.5

[a] Mull and m-Mull stand for Mulliken and modified Mulliken respectively. In calculating $\mu^{m\text{-}Mull+HDC}$, centroids of HDC were used

[b] For structures of amino acids and atomic numbering scheme, see Fig. 1

[c] Percentage changes in dipole moment in going from Mulliken to (HDC + m-Mulliken) charges were defined as $100 \times \left(\mu^{m-Mull+HDC} - \mu^{Mull}\right)/\mu^{Mull}$. The sign of this % change was taken to be positive if the change ($\mu^{m\text{-}Mull+HDC}$ - μ^{Mull}) was in the direction of μ^{CHelpG} with respect to μ^{Mull}.

Table 2 Minimum surface MEP values (V^x, x=CHelpG, Mulliken and m-Mulliken + HDC) (kcal mol^{-1}) on the van der Waals surfaces of amino acids calculated using different charge distributions obtained at the B3LYP/ 6-31G** level of theory in gas phase[a]

S. No.	Amino acid	Min. MEP near atom[b]	MEP			%Change[c]	Change of MEP min. site[d]
			V^{CHelpG}	$V^{m\text{-}Mull+HDC}$	V^{Mull}		
1.	Alanine	N1	-49.1	-52.0	-43.3(O10)[e]	20.1	O10→N1
2.	Arginine	N16	-53.1	-53.3	-43.7(O10)[e]	22.0	O10→N16
3.	Asparagine	O11	-52.4	-52.7	-49.1(O11)	7.3	No change
4.	Aspartic acid	O11	-46.4	-44.8	-42.3(O11)	5.9	No change
5.	Cysteine	O11	-50.0	-50.5	-49.8(O11)	1.4	No change
6.	Glutamine	O14	-52.7	-53.1	-50.0(O14)	6.2	No change
7.	Glutamic acid	N1	-47.9	-47.8	-41.8(O14)[e]	14.4	O14→N1
8.	Glycine	N1	-49.0	-48.6	-42.5(O7)[e]	14.4	O7→N1
9.	Histidine	N15	-50.9	-53.9	-44.7(N15)	20.6	No change
10.	Isoleucine	N1	-46.9	-53.2	-42.8(O10)[e]	24.3	O10→N1
11.	Leucine	N1	-50.3	-50.9	-42.9(O10)[e]	18.7	O10→N1
12.	Lysine	N19	-52.3	-52.7	-42.2(O10)[e]	24.9	O10→N19
13.	Methionine	N1	-49.2	-49.8	-41.5(O10)[e]	20.0	O10→N1
14.	Phenylalanine	N1	-44.3	-44.3	-42.4(O10)[e]	4.5	O10→N1
15.	Proline	O14	-42.7	-47.6	-43.5(O14)	9.4	No change
16.	Serine	O9	-44.9	-54.4	-45.4(O9)	19.8	No change
17.	Threonine	N1	-55.3	-55.9	-48.4(O10)[e]	15.5	O10→N1
18.	Tryptophan	N1	-45.5	-41.4	-39.7(O10)[e]	4.3	O10→N1
19.	Tyrosine	N1	-48.7	-47.8	-41.4(O10)[e]	15.5	O10→N1
20.	Valine	N1	-46.6	-53.7	-42.9(O10)[e]	25.2	O10→N1

[a] Mull and m-Mull stand for Mulliken and modified Mulliken respectively. In calculating $V^{m\text{-}Mull+HDC}$, centroids of HDC were used

[b] For structures of amino acids and atomic numbering scheme, see Fig. 1

[c] Percentage changes in MEP in going from Mulliken charges to (HDC + m-Mulliken) were defined as $100 \times \left(V^{m-Mull+HDC} - V^{Mull}\right)/V^{Mull}$. The sign of this % change was taken to be positive if the change ($V^{m\text{-}Mull+HDC}$ - V^{Mull}) was in the direction of V^{CHelpG} with respect to V^{Mull}

[d] Change in site of surface MEP minimum in going from Mulliken charges to (HDC + m-Mulliken) charges

[e] In these cases, wrong locations of MEP minima were obtained using Mulliken charges

Surface MEP values

MEP is a more sensitive property to charge distribution than dipole moment. As in a molecule there can be several local surface MEP minima, the site of the lowest surface MEP minimum is important and serves as a good point of comparison between the different charge distributions. Therefore, MEP offers a superior test for accuracy of a charge distribution than dipole moment.

The lowest surface MEP minima for the amino acids are presented in Table 2. The MEP values obtained using three charge distributions, i.e., CHelpG, Mulliken and (HDC + m-Mulliken), are presented in this table. The sites of the lowest surface MEP minima and %changes of MEP, the latter defined in the same way as %changes of dipole moment, are also presented in Table 2. The last column of this table shows the changes of sites of lowest surface MEP minima which occur in some cases in going from the Mulliken to the (HDC + m-Mulliken) charges. The MEP values presented in Table 2 reveal the following features: (i) Out of the 20 amino acids, only in seven cases (asparagine, aspartic acid, cysteine, glutamine, histidine, proline, serine), Mulliken charges predict correct locations of lowest surface MEP minima. In the remaining 13 cases (alanine, arginine, glutamic acid, glycine, isoleucine, leucine, lysine, methionine, phenylalanine, threonine, tryptophan, tyrosine and valine), locations of the lowest surface MEP minima were corrected from O7, O10 or O14 to N1, N16 or N19 in going from the Mulliken to the (HDC + m-Mulliken) charges (Table 2). (ii) The %changes in the lowest surface MEP values in going from the Mulliken to the (HDC + m-Mulliken) charges are positive in all 20 cases, these changes in some cases being as large as ~25%. (iii) The lowest surface MEP values obtained using the CHelpG and (HDC + m-Mulliken) charges are within 1% and 10% of each other in seven and 16 cases out of the 20 cases respectively. In four cases, the differences between the MEP values obtained using the two charge distributions lie in the range 10–20%. Thus the (HDC + m-

Mulliken) charges produce the surface MEP values fairly accurately.

The calculated MEP maps of glycine, taken as an example, are presented in Fig. 2. In this figure, the maps presented in the parts (a), (b) and (c) were obtained by B3LYP/6-31G** calculation (i.e., using continuous electron density), and using CHelpG and (HDC + m-Mulliken) point charges respectively. In part (a), the red-colored patches located near the nitrogen atom of the amino group and the oxygen atom of the CO group correspond to the surface MEP value of -43.9 kcal mol^{-1}. Further, in (a), the most negative surface MEP value located near the nitrogen atom of the amino group is ~50 kcal mol^{-1}. The most negative surface MEP values obtained using the CHelpG and (HDC + m-Mulliken) charges near this site are also similar (~-49 kcal mol^{-1}) (Table 2). Further, the surface MEP maps presented in the parts (b) and (c) of Fig. 2 also have clear negative MEP regions near the two sites mentioned above.

Locations of HDC centroids and total HDC values

In previous studies related to HDC, locations of the different HDC components associated with different atoms in molecules were given [15, 20, 21]. Displacements of HDC from the corresponding atoms depend only on the ζs of atomic orbitals as discussed above. Therefore, for the same type of atom (e.g., C, N or H) in the same or different molecules, the HDC distances are the same as reported in the previous studies [15, 20, 21]. However, HDC centroids do not depend only on the ζs, but also on the values of the HDC components which may be quite different for the same type of atom (e.g., C, N or H) in the same or different molecules. Cartesian coordinates of atoms, HDC centroids and bond centers and the net charges located at these sites in all 20 amino acids are presented in Table S1 (Supporting information). With the help of these coordinates, one can easily locate the HDC centroids with respect to the atoms in the amino acids. The values of total HDC and the locations

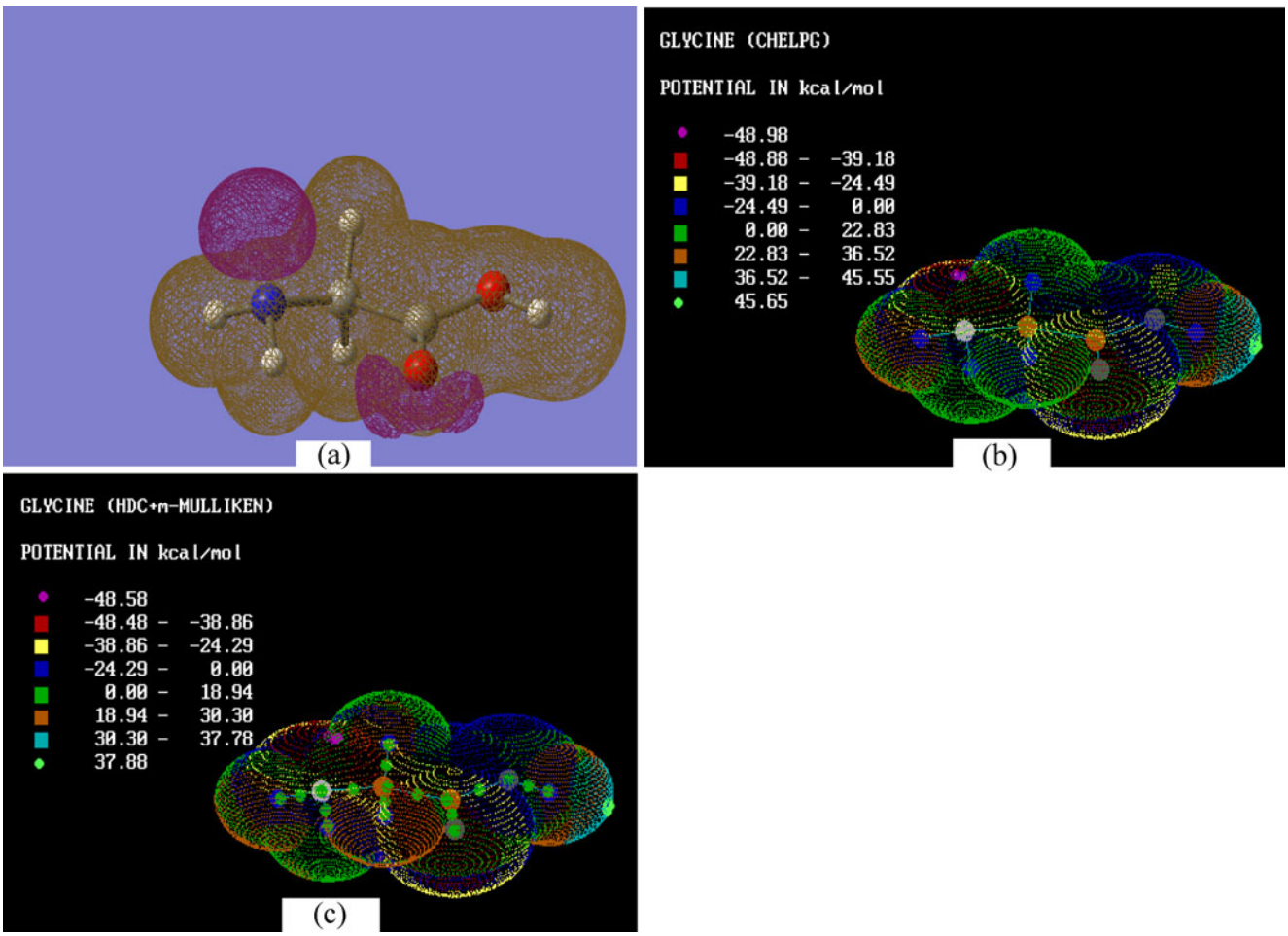

Fig. 2 Surface MEP maps of glycine. The maps of (a), (b) and (c) were obtained by B3LYP/6-31G** calculations, and using CHelpG and (HDC + m-Muliken) charges respectively

Table 3 Locations of HDC centroids in glycine, alanine and lysine. T is the symbol for HDC, Q stands for total HDC value located at the HDC centroid and D is the distance of the HDC centroid from the corresponding atom. Distances are given in Å while angles are given in degrees

Amino acid[a]	Atom No.	HDC value (Q)	Distance (D)	Angle[b]	Dihedral angle[b]
Glycine	N1	-2.691	0.034	C3N1T=108.4	C3N1H2T=117.7
	H2	-0.029	0.039	N1H2T=5.4	N1H2H10T=5.0
	C3	-0.537	0.061	N1C3T=136.6	N1C3C4T=-166.8
	C4	-0.913	0.077	C3C4T=16.2	C3C4O8T=1.6
	H5	-0.030	0.028	C3H5T=4.1	C3H5H6T=5.0
	H6	-0.030	0.029	C3H6T=1.8	C3H6H5T=-2.5
	O7	-2.381	0.020	C4O7T=178.0	C4O7O8T=-179.7
	O8	-2.112	0.020	C4O8T=114.4	C4O8O7T=0.1
	H9	-0.040	0.041	O8H9T=1.3	O8H9O7T=0.3
	H10	-0.032	0.035	N1H10T=3.5	N1H10H2T=-5.6
Alanine	N1	-2.786	0.033	C3N1T=105.6	C3N1H2T=115.4
	H2	-0.029	0.039	N1H2T=5.9	N1H2H11T=4.5
	C3	-0.539	0.061	N1C3T=142.9	N1C3H4T=-157.3
	H4	-0.030	0.028	C3H4T=2.8	C3H4C5T=3.1
	C5	-0.326	0.064	C3C5T=39.5	C3C5H7T=33.0
	C6	-1.027	0.077	C3C6T=13.5	C3C6O10T=-4.0
	H7	-0.028	0.029	C5H7T=2.5	C5H7C3T=3.7
	H8	-0.029	0.029	C5H8T=2.6	C5H8C3T=3.8
	H9	-0.030	0.028	C5H9T=3.8	C5H9C3T=-2.4
	O10	-2.320	0.020	C6O10T=177.5	C6O10O12T=179.6
	H11	-0.031	0.036	N1H11T=3.3	N1H11H2T=-5.3
	O12	-2.116	0.020	C6O12T=114.7	C6O12O10T=0.1
	H13	-0.040	0.041	O12H13T=1.3	O12H13C6T=0.3
Lysine	N1	-2.792	0.033	C3N1T=105.6	C3N1H2T=115.3
	H2	-0.029	0.039	N1H2T=5.7	N1H2H22T=4.0
	C3	-0.551	0.070	N1C3T=148.8	N1C3H4T=177.6
	H4	-0.030	0.028	C3H4T=2.8	C3H4N1T=-4.1
	C5	-0.341	0.067	C3C5T=67.6	C3C5C9T=-17.2
	C6	-1.025	0.079	C3C6T=14.1	C3C6O10T=-3.6
	H7	-0.030	0.028	C5H7T=4.5	C5H7H8T=-0.2
	H8	-0.029	0.028	C5H8T=2.9	C5H8H7T=-1.1
	C9	-0.164	0.081	C5C9T=62.3	C5C9C13T=38.8
	O10	-2.309	0.020	C6O10T=177.5	C6O10O23T=179.9
	H11	-0.028	0.029	C9H11T=3.8	C9H11H12T=6.3
	H12	-0.030	0.027	C9H12T=1.8	C9H12H11T=1.7
	C13	-0.231	0.082	C9C13T=58.3	C9C13C16T=36.2
	H14	-0.029	0.028	C13H14T=4.5	C13H14H15T=-6.7
	H15	-0.030	0.027	C13H15T=2.4	C13H15H14T=-1.3
	C16	-0.455	0.079	C13C16T=42.3	C13C16N19T=18.7
	H17	-0.031	0.027	C16H17T=1.5	C16H17H18T=-0.9
	H18	-0.030	0.027	C16H18T=1.9	C16H18H17T=-2.8
	N19	-2.762	0.034	C16N19T=104.6	C16N19H20T=-116.5
	H20	-0.030	0.036	N19H20T=3.7	N19H20H21T=-4.8
	H21	-0.031	0.035	N19H21T=3.8	N19H21H20T=4.8
	H22	-0.031	0.036	N1H22T=3.2	N1H22H2T=-5.1
	O23	-2.077	0.020	C6O23T=114.6	C6O23O10T=0.3
	H24	-0.040	0.041	O23H24T=1.2	O23H24C6T=0.1

[a] Structures of the amino acids and the adopted atomic numbering scheme are given in Fig. 1

[b] The angles (XYT) and dihedral angles (XYZT) are given with respect to the atoms X,Y and X,Y,Z respectively

of HDC centroids associated with the atoms in three amino acids, i.e., glycine, alanine and lysine, taken as examples, are presented in Table 3. All the HDC values (in the unit of magnitude of electronic charge) are negative as they represent electronic charges (Table 3). In Table 3, the location of each HDC centroid (denoted by T) is given using the distance (YT) from the corresponding atom (Y), the bond angle (XYT) and the dihedral angle (XYZT) where X and Z are atoms located near Y. We make the following observations from the data presented in Table 3:

(i) The HDC values (Q) associated with the hydrogen atoms are in the range -0.04 to -0.03, those associated with the oxygen atoms are in the range -2.3 to -2.1 while those associated with the nitrogen atoms are in the range ~-2.8 to -2.7. The HDC values associated with the carbon atoms vary over a wide range lying in the range ~-1 to -0.2 which conforms to the fact that the electronic environments of the different carbon atoms are quite different.

(ii) The distances of the HDC centroids (D) from the corresponding oxygen and nitrogen atoms are ~0.02 and ~0.03 Å respectively while the corresponding distances for hydrogen atoms lie in the range 0.028 - 0.041 Å (Table 3). For carbon atoms, the distances of the HDC centroids (D) from the corresponding atomic sites lie in the range 0.061 - 0.082 Å (Table 3). Thus the distances of HDC centroids from the corresponding atoms for the different atom types follow the order O<N≤H<C.

(iii) The C3N1T angles in the different cases are more than 105 deg. while the C3N1H2T dihedral angles are more than 115 deg. (Table 3). Similarly, in lysine, the angle C16N19T is ~105 deg. while the dihedral angle C16N19H20T is ~-117 deg. These angles belong to pyramidal amino groups of the molecules and the HDC centroid (T) is located near the N1 atom outside the pyramid in each case. The dihedral angles show that most of the HDC centroids are located in cis or trans position (dihedral angles close to 0 or 180 deg.) with respect to the atoms given in Table 3. However, in some cases (e.g., those belonging to C5 in both of alanine and lysine, and those belonging to C9, C13 and C16 in lysine), the locations of HDC centroids are significantly away from the cis or trans position (Table 3). On the whole, as may be expected on the basis of differences in the structures of the molecules, there does not seem to be a simple prescription possible for the locations of the HDC centroids in the different amino acids.

It may be mentioned that the values of HDC and their distances from the corresponding atomic sites obtained earlier for the amino acids using a semiempirical approach [16] were significantly different from the present ones. As an example, we may compare the HDC associated with the carbonyl and hydroxyl group oxygen atoms of glycine. In the present work, the charges located at these atomic sites (the quantities corresponding to the hydroxyl group oxygen atom being given in parentheses) are 2.035(1.739) while those at the HDC centroids are -2.381(-2.112) and the distances of the HDC centroids from the corresponding atomic sites are both 0.02 Å. The corresponding charges located at the atomic and HDC sites obtained using the semiempirical approach [16] were -0.003(1.119) and -0.379 (-1.372) respectively while the corresponding distances were both 0.392 Å. The lowest surface MEP value for glycine was found to be located near the carbonyl oxygen atom [16] while it should be located near the amino group nitrogen atom as discussed earlier in this work (Table 2). Appreciable differences are also noted between the present and previous semiempirical results [16] with regard to the values and locations of HDC and surface MEP minima for the other amino acids. Such differences between the semiempirical and present results are not surprising in view of the facts that the semiempirical methods only consider valence electrons explicitly and the results obtained using them are strongly dependent on the parameters used. Thus the present results are much more reliable than those obtained by the semiempirical approach [16].

Conclusions

The present study shows that the combination of HDC centroids with the m-Mulliken charges provides a much better charge distribution than that consisting of the conventional Mulliken charges alone for calculating electrostatic properties of amino acids. The same is expected to be true for other molecules containing H, N, C atoms also. In going from the individual HDC components to their centroids, the number of charges is drastically reduced. Thus, in this approach, each atom is associated with only two point charges, one located at the atomic site and the other located near the same, while each bond center is associated with a single point charge. Therefore, the combination of HDC centroids with the m-Mulliken charges may be used in preference to Mulliken charges. Locations of the HDC centroids obtained in the present work appear to be useful in the sense that in MEP-based charge fitting schemes, charges may be fitted at the atomic as well as these sites.

Acknowledgments The authors are thankful to the University Grants Commission (New Delhi) and the Council of Scientific and Industrial Research (New Delhi) for financial support.

References

1. Frauenheim T, Seifert G, Elstner M, Hajnal Z, Jungnickel G, Porezag D, Suhai S, Scholz R (2000) In: Deák P, Frauenheim T, Pederson MR (eds) Computer Simulation of Materials at Atomic Level. Wiley, Berlin
2. Kumar A, Mishra PC (1995) Top Curr Chem 174:27–43
3. Dixon RW, Kollman PA (1997) J Comput Chem 18:1632–1646
4. Kowal M, Roszak S, Leszczynski J (2001) J Chem Phys 114:8251–8256
5. Mulliken RS (1955) J Chem Phys 23:1833–1840
6. Löwdin PO (1950) J Chem Phys 18:365–375
7. Reed AE, Weinstock R, Weinhold F (1985) J Chem Phys 83:735–746
8. Singh UC, Kollman PA (1984) J Comput Chem 5:129–145
9. Fliszar S, Cardinal G, Baykara NA (1986) Can J Chem 64:404–412
10. Breneman CM, Wiberg KB (1990) J Comput Chem 11:361–373
11. Murray JS, Lave P, Clark T, Politzer P (2007) J Mol Model 13:1033–1038
12. Politzer P, Murray JS (2002) Theor Chem Accts 108:134–142
13. Politzer P, Murray JS (1996) Molecular Electrostatic Potentials: Concepts and Applications. In: Murray JS, Sen K (eds) Theoretical and computational chemistry. Book series. Elsevier, Amsterdam, pp 649–657
14. Upadhyay DM, Shukla MK, Mishra PC (2000) J Mol Struct THEOCHEM 531:249–266
15. Jena NR, Mishra PC (2005) J Mol Struct THEOCHEM 719:75–84
16. Kushwaha PS, Kumar A, Mishra PC (1999) Int J Quantum Chem 74:271–289
17. Kumar A, Bhattacharjee AK, Mishra PC (1992) Int J Quantum Chem 43:579–589
18. Mohan CG, Kumar A, Mishra PC (1996) Int J Quantum Chem 60:699–708
19. Singh AK, Kumar A, Mishra PC (2004) J Mol Struct THEOCHEM 682:201–213
20. Singh AK, Mishra PC (2007) Int J Quantum Chem 107:597–608
21. Tiwari S, Shukla PK, Mishra PC (2008) J Mol Model 14:631–640
22. Pople JA, Beveridge DL (1970) Approximate Molecular Orbital Theory. Mc-Graw Hill, New York
23. Lehninger AL (1982) Principles of Biochemistry. Worth, New York
24. Becke AD (1993) J Chem Phys 98:5648–5652
25. Lee C, Yang W, Parr RG (1998) Phys Rev B 37:785–798
26. Becke AD (1993) J Chem Phys 98:1372–1377
27. Warren JH, Radom L, Schleyer PvR, Pople JA (1986) Ab Initio Molecular Orbital Theory. Wiley, New York
28. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Montgomery JA, Vreven T, Kudin KN, Burant JC, Millam JM, Iyengar SS, Tomasi J, Barone V, Mennucci B, Cossi M, Scalmani G, Rega N, Petersson GA, Nakatsuji H, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Klene M, Li X, Knox JE, Hratchian HP, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Ayala PY, Morokuma K, Voth GA, Salvador P, Dannenberg JJ, Zakrzewski VG, Dapprich S, Daniels AD, Strain MC, Farkas O, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Ortiz JV, Cui Q, Baboul AG, Clifford S, Cioslowski J, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Gonzalez C, Pople JA (2004) Gaussian 03, Revision D.01. Gaussian Inc, Wallingford
29. Frisch AE, Dennington RD, Keith TA, Neilsen AB, Holder AJ (2003) GaussView, Rev. 3.09. Gaussian Inc, Pittsburg