

Charge calculations in molecular mechanics. Part 8* Partial atomic charges from classical calculations

Raymond J. Abraham**, Guy H. Grant, Ian S. Haworth and Paul E. Smith

The Robert Robinson Laboratories, The Chemistry Department, University of Liverpool, Liverpool L69 3BX, U.K.

Received 5 June 1990

Accepted 30 June 1990

Key words: Partial charge; Charge calculation; Molecular mechanics; Force field comparison

SUMMARY

The CHARGE2 programme, which involves the classical calculation of both the inductive and resonance contributions to the partial atomic charges in molecules is described, and the charges and electrostatic potentials obtained presented for some illustrative examples.

In substituted methanes (CH_3X , CF_3X , CCl_3X) the effects of varying the electronegativity of the substituents and the α - and β -substituent contributions are clearly illustrated for a variety of substituent groups X.

The problems involved in the inclusion of silicon into this scheme are detailed, together with the methods of overcoming them. The partial atomic charges (σ and π contributions) and electrostatic potentials for some silicon oxygen compounds are presented and discussed.

The partial atomic charges from CHARGE2 for all the natural amino acids as their *N*-acetyl, *N*-methyl-amides are given and compared with those obtained from the AMBER and ECEPP/2 force fields. Considerable differences in these figures are observed, with the AMBER charges consistently much larger than those from the other two methods.

The CHARGE2 partial atomic charges and electrostatic potentials for the four common nucleic acids, adenine, cytosine, guanine and thymine, are given and compared with those derived from other calculations. Again there is general similarity but also there are considerable differences, with those from the AMBER force field somewhat larger than the other methods.

INTRODUCTION

The concept of partial atomic charges in molecules is central to much chemical thinking, yet the actual estimation of the partial atomic charges in molecules is subject to very considerable uncer-

*For previous parts in this series, see Refs. 1–7.

**To whom correspondence should be addressed.

tainty and ambiguity. In principle, quantum mechanics should provide the answer, as an accurate molecular wavefunction does give the electron distribution of the molecule at all points in space. However, the problems involved in condensing this into a generally acceptable partial atomic charge distribution have not yet been satisfactorily solved. The Mulliken population analysis was previously the method of choice, but this has been increasingly questioned recently and is also very basis-set dependent. Recent work has concentrated on obtaining atomic charges (or atomic multipoles) from the quantum-mechanical derived molecular electrostatic potential. A fundamental problem in this is that the charges obtained do depend on the surface adopted for the calculations, as well as on any other assumptions made in the derivation (e.g., the existence or absence of lone pairs, the inclusion of atomic multipoles, etc.).

With these forbidding problems one may well ask why try to impose this essentially subjective view of the molecular electron distribution? The reason lies in the increasing utility and importance of molecular mechanics calculations and molecular modelling in general. Some way of estimating the electrostatic interactions both within and between large molecules is of the utmost importance in any quantitative modelling package, and as virtually all molecular mechanics calculations involve atom-centred steric interactions this almost of necessity imposes the same constraints on the calculation of the electrostatic interactions, i.e. that they result from atom-centred partial charges.

It was on this basis that the CHARGE2 programme has evolved. This is an attempt to obtain a computationally fast and simple method of estimating partial atomic charges in molecules which is both based on chemical intuition and also parameterised on an experimentally observable quantity.

The requirement for computational speed and simplicity is obvious when the possible applications for such a scheme to biomolecules are considered. At present the CHARGE2 programme is parameterised to handle molecules of up to 200 atoms and this could easily be increased. The computational time required is so small (for the molecules so far studied) that the programme runs virtually instantaneously on an IBM PC.

The programme is based on two very fundamental chemical concepts, the inductive effect for saturated molecules combined with Hückel MO calculations for π systems. The inductive effect operates via the atomic electronegativity and polarisability and the Hückel scheme through the appropriate Coulomb and resonance integrals. In contrast to the similar empirical scheme of Gas-teiger and co-workers [8, 9], CHARGE2 does not include the variation of orbital electronegativity with atomic charge except for ions. It also does not include any direct interaction between the σ - and π -calculations, or hyperconjugation. All these could quite easily be included within the programme but this would require re-parameterisation.

The parameterisation of the programme is based on the observed molecular dipole moments. The molecular dipole moment was selected since it is a quantity which is both sensitive to the partial atomic charges in the molecule and is available for a wide variety of molecules.

In previous articles in this series [1–7], the development of the CHARGE2 programme has been given, together with the observed and calculated molecular dipole moments for the molecules studied. In this article we concentrate instead on the partial atomic charges obtained by this method, and also show examples of the electrostatic potentials obtained from these charges, after a brief summary of the relevant theory.

THEORY

If we consider an atom I in a 4-atom fragment I-J-K-L, the partial atomic charge on I is due to 3 effects; an α -effect from atom J, a β -effect from atom K and a γ -effect from L.

The charge (q_i) on atom I resulting from atom J (the α -effect) is given by Eq. 1:

$$q_i = (E_j - E_i)/a_{ij} \quad (1)$$

where E_i and E_j are the electronegativities of atoms I and J, respectively; a_{ij} is a constant depending on the exchange and overlap integrals for the bond I-J. Initially only 3 values were used for a_{ij} , one if the bond was to hydrogen, one for bonds between first-row atoms, and one for bonds involving a second-row (or higher) atom. This method has been extended so that there now exists a set of parameters for all the bonding pairs being considered. These are described as AXY parameters where X and Y are the atoms involved in the bond.

The β -effect, from atoms two bonds removed (I-J-K), is considered to be proportional to both the electronegativity of atom K and the polarizability of atom I. Taking hydrogen as the base substituent the β -effect was defined as:

$$q_i = (E_K - E_H)P_i/b \quad (2)$$

In order to account for the variation of polarizability with charge, the β -effect calculation is carried out iteratively according to Eq. 3

$$P_i = P_i^0(1.0 + 3.0 \times (q_i^0 - q_i)) \quad (3)$$

where P_i is the polarizability of atom I with charge q_i , and P_i^0 and q_i^0 are the corresponding initial values.

The γ -effect is considered to be directly proportional to the β -effect and is given by Eq. 4,

$$q_i(\gamma) = \beta_i/5.0 \quad (4)$$

the constant (5.0) is obtained from NMR data.

The total charge is given by Eq. 5:

$$q_i = q_i(\alpha) + q_i(\beta) + q_i(\gamma) \quad (5)$$

For π -systems a Hückel subroutine has been incorporated to calculate the π -charges which are then added to the σ -charges (above) to give the total atomic charge. The standard Coulomb and resonance integrals for the Hückel routine are given by Eqs. 6,

$$\begin{aligned} \alpha_r &= \alpha_o + h_r \beta_o \\ \beta_{rs} &= k_{rs} \beta_o \end{aligned} \quad (6)$$

where α_o and β_o are the Coulomb and resonance integrals for a carbon $2p_z$ atomic orbital and h_r and k_{rs} the factors modifying these integrals for orbitals other than sp^2 carbon. These factors were again obtained from the observed dipole moments.

Note that in CHARGE2 the σ - and π -charges are not interactive, as of course they are in any SCF treatment.

The standard Hückel routine was later modified by the inclusion of the ω -technique (see later), to model the very polar π -systems of the nucleic acid bases. This required a refinement of the previous parameterisation of some of the earlier data. Also in earlier versions of CHARGE the β -effect of the halogens Cl, Br and I was obtained via Eq. 2 but with an artificially enhanced electro-

negativity. This has now been removed. Thus the atomic charges given here differ in these instances from those given previously. The effect on the molecular dipole moments is, however, minimal.

SUBSTITUTED METHANES

The partial atomic charges in a series of substituted methanes CH_3X , CF_3X and CCl_3X calculated by CHARGE2 are given in Table 1 for a range of substituents X, and this series both illustrates the generality of the CHARGE2 programme and also the relative importance of the α - and β -substituent effects.

The charge on the hydrogen atom in methane (34 m.e.) is equivalent to a C–H bond dipole of 0.18 D, which is within the accepted range of this bond dipole; values of 0.1–0.3 D have been quoted previously.

In CH_3X compounds the α -effect determines the charge on the carbon atom, whilst the β -effect determines the charge on the hydrogen atom. There is a good linear correlation between the calculated charge on the hydrogen atom and the proton chemical shift, with a slope of ca. 150 p.p.m./electron. Again, this is within the accepted range for this type of correlation. Gasteiger obtained 106.4 p.p.m./electron unit from an extensive correlation of proton shifts and the PEOE-calculated atomic charges [10], and a similar value was obtained previously using CHARGE charges in

TABLE 1
CHARGE2 CHARGES FOR Y_3CX MOLECULES*

X $\alpha\beta$	H_3CX				F_3CX				Cl_3CX			
	H	C	α	β	F	C	α	β	Cl	C	α	β
II	34	–136	34		–165	389	107		–125	302	73	
F	62	35	–222		–144	574	–144		–82	426	–179	
Cl	48	39	–184		–154	513	–50		–103	411	–103	
Br	48	30	–172		–155	485	–21		–104	389	–77	
I	46	8	–146		–156	442	27		–106	352	–33	
OH	55	–36	–458	329	–149	423	–306	331	–92	321	–373	331
OCH_3^a	55	–33	–266	–33	–149	471	–183	–8	–92	347	–227	–10
SH	43	–21	–237	128	–158	425	–90	140	–110	342	–150	139
NH_2	47	–62	–505	212	–155	346	–322	220	–104	278	–405	220
NO_2	61	7	477	–333	–145	542	503	–305	–85	361	508	–308
CN	49	–45	362	–464	–154	521	367	–426	–102	365	370	–429
CH_3	37	–110	–110	37	–163	354	–26	54	–121	285	–81	53
CF_3	54	–26	354	–163	–150	450	450	–150	–94	283	453	–151
CCl_3	53	–81	285	–121	–151	453	283	–94	–96	289	289	–96
SiH_3	31	–128	350	–105	–167	411	333	–80	–130	321	316	–82
SiCH_3^b	33	–118	77	–118	–166	281	181	–86	–127	235	117	–88
COCH_3^c	46	–61	190	–342	–157	487	193	–310	–107	340	195	–313
COOH^d	49	–44	285	–373	–154	481	310	–342	–101	324	314	–344
a		55(H)				55(H)				55(H)		
b		33(H)				33(H)				33(H)		
c		–61(C), 46(H)				–37(C), 46(H)				–39(C), 46(H)		
d		–346(O), 330(H)				–318(O), 330(H)				–320(O), 330(H)		

*In m.e.

amino acids. Interestingly there is also a good correlation between the calculated charges on the carbon atoms in CH_3X with the ^{13}C shifts, though here it is necessary to exclude CN and the heavy halogens (Cl, Br, I) which have pronounced anisotropic effects (e.g. δ for CH_3I is -20.7). Here the slope is 468 p.p.m./e.u. and this is quite different from the value obtained previously using CHARGE and amino acid chemical shifts (220 p.p.m./e.u.). The values obtained from the various correlations of ^{13}C chemical shifts in aromatic compounds with π (and total) electron densities are ca. 180 p.p.m./e.u., thus the value obtained from the substituted methanes is clearly not comparable with these other investigations. The CH_3X correlation is somewhat different in that the atom under consideration is directly bonded to the variable substituent, and this may accentuate the response of the ^{13}C chemical shift to the electron density. What is of interest is the remarkably good correlation for $\text{X}=\text{H}, \text{CH}_3, \text{SiMe}_3, \text{COCH}_3, \text{NH}_2, \text{OH}, \text{F}$ and this obviously requires further investigation, as a general prediction of ^1H and ^{13}C chemical shifts using the CHARGE electron densities would be of considerable utility.

Although the charge on the carbon atom spans a considerable range, from -136 m.e. in methane to $+574$ m.e. in CF_4 , these charges are much less than a simple dipole model would predict, due to the compensating β -effect of the electronegative groups. For example this β -effect reduces the charge on fluorine in CF_4 from -222 m.e. (the value in CH_3F) to -144 m.e. and this decreases the charge on the carbon atom from $+888$ m.e. (on a simple additive scheme) to $+574$ m.e.

The β -effect of the substituents is also clearly illustrated in the contrast between the CF_3X (and CCl_3X) and CH_3X charges, atom X in CF_3X being as expected, much more positive (or less negative) than in CH_3X . The β -effect of the fluorine atoms combined with the large polarisability of iodine results in the iodine atom in CF_3I (but not CCl_3I) having a nett positive charge.

The γ -effect of the electronegative CF_3 group is however, much smaller and in many cases hardly significant, though in CHARGE2, in complete contrast to CNDO, there is no alternation of electron charge along a chain, the α -, β - and γ -effects falling off smoothly as in the chemical concept of an inductive effect.

APPLICATION TO SILICON COMPOUNDS

In order that an element may be included in the scheme, it is necessary to obtain values for the electronegativity and polarizability of that element, in whichever hybridisation state is appropriate. Using this data it is then possible to calculate suitable a_{ij} values for each of the Si-X bond pairs under consideration.

For silicon the above information initially proved problematical, there being some dispute as to the actual value of the Si electronegativity with respect to hydrogen (see below). It was also impossible to find an appropriate value for the Si polarizability as this was not available on the scale used in our previous calculations. This latter problem was overcome by extrapolation from the values obtained for sp^3 carbon as there appeared to be an additive relationship between first- and second-row atoms (cf. O, 0.912; S, 1.748; F, 0.697; Cl, 1.518). The final value chosen for Si was 1.877.

Although Si is a Group IV element it is not possible to consider it as being analogous to C; there appear to be many deviations both in terms of observed geometries of Si-containing compounds and in the size of the observable dipole moments with respect to those of the analogous carbon compounds.

It is possible to understand this anomalous behavior if one considers that silicon has d-orbitals which are of low enough energy to make their use in bonding a serious possibility [11]. We will consider the consequences of this $d_{\pi}-p_{\pi}$ bonding in greater detail later as it is obvious that a bonding pair which utilizes this type of link cannot be treated in the same way as a typical σ -system. In order to account for this feature, any bond between silicon and an atom which has lone pairs of electrons which are available for donation must necessarily be treated as a π -system.

In previous work [1–3], the electronegativity values used were taken from the work of Hinze and Jaffe [12] whose scale was based upon the following equation developed by Mulliken [13]:

$$\chi^A = (I^A + E^A)/2 \quad (7)$$

where I^A is the ionisation potential of atom A and E^A its electron affinity.

Hinze and Jaffe attempted to correlate the electronegativities calculated according to the method devised by Mulliken [13] with those from the scale developed by Pauling [14]. Although the latter did not consider different valence states, the orbital electronegativity used in the correlation was that calculated for sp^3 hybridisation. Although a reasonable agreement was found for the correlation according to the following equation:

$$0.168(\chi_M - 1.23) = \chi_P \quad (8)$$

it was found that silicon produced an anomalous result with χ_M appearing to be too large. This result is important as it assigns silicon to be more electronegative than hydrogen, in opposition to other electronegativity scales [14–16] which consider silicon to be the more electropositive.

As part of a series of recent calculations on silicon-containing molecules, Luke et al. [17] have plotted substituent electronegativity vs. the bond strength of that substituent to silicon. The electronegativity scale used for this correlation was that of Allred and Rochow [18]. In this scale the value of the silicon electronegativity is close to that given by Pauling, and later, Huggins, i.e. silicon is more electropositive than hydrogen.

This theoretical evidence for the polarity of the Si–H bond has also been supported by experimental evidence in the form of Si–H bond dipoles [19] and the direction of the dipole moment in

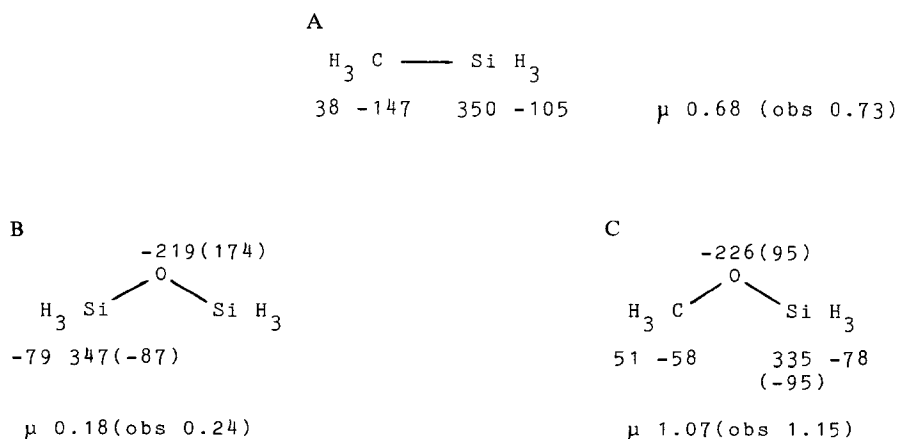


Fig. 1. Calculated partial atomic charges (π -contribution), m.e. and calculated and observed dipole moments (D) in methyl silane (A) and silyl ethers (B and C).

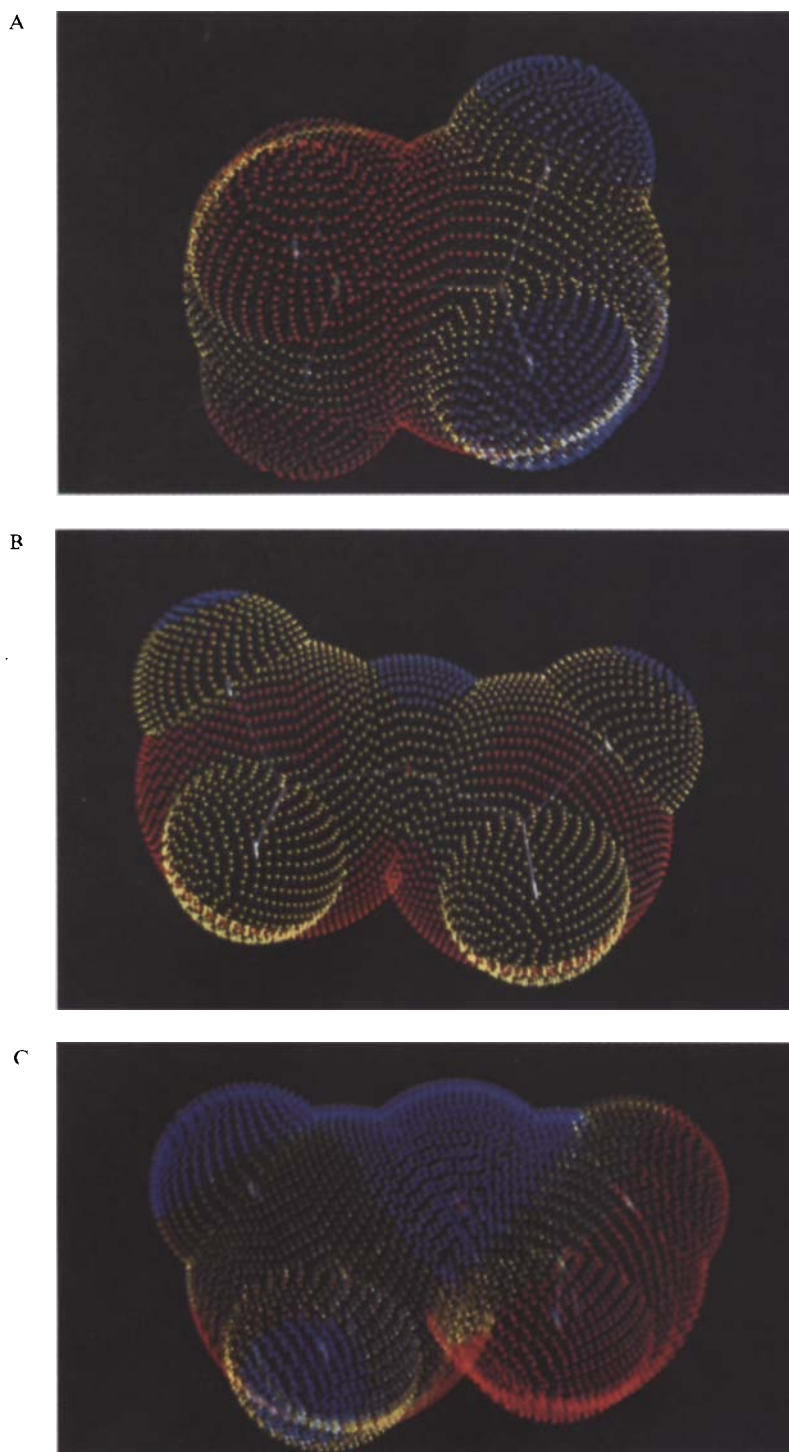


Fig. 2. Electrostatic potentials of the molecules of Fig. 1. Red $> +10$ kcal·mol⁻¹; yellow between $+10$ and -10 kcal·mol⁻¹; blue < -10 kcal·mol⁻¹.

various alkyl silanes [20, 21]. The direction of the Si–H bond dipole is considered to be from the silicon to the hydrogen.

The important conclusion that can be drawn from the above is that the value of χ_{Si} cannot be taken from Hinze and Jaffe [12]. We, therefore, decided that the most appropriate value would be that given by Huggins [15], in a refinement of the Pauling scale, and then, using the relationship given in Eq. 8 convert that value to one which complements our other electronegativity values. The value quoted by Huggins for Si is 1.90 which was then converted to 6.27.

With the values of the electronegativity and polarisability for silicon obtained, the appropriate values of the bonding integrals A_{ij} involving silicon can be obtained from the observed molecular dipole moments. It was found, as mentioned previously that for Si–X bonds in which X has lone pairs of electrons (notably oxygen), the observed dipole moments can only be realised by a sizeable donation of π -charge from oxygen to silicon. This can easily be accommodated in the CHAR-GE2 Hückel scheme and this parameterisation has been given [5].

As illustrative examples, Fig. 1 gives the calculated total (and π) charges for methyl silane and two silyl ethers, and Fig. 2 the electrostatic potential calculated from these charges for these compounds.

In methyl silane there is no need to invoke any π -contribution, however, the molecule is quite polar (cf. chloroform $\mu = 1.09$ D) with the CH_3 and SiH_3 hydrogens being of opposite polarity. We note that most of the observed dipole moment comes from the Si–H ($\mu_{\text{Si-H}} 0.75$ D) and C–H bonds ($\mu_{\text{C-H}} 0.20$ D) and very little from the C–Si bond.

The results for disilyl ether and methyl silyl ether, in contrast, show considerable donation of π -charge from the oxygen atom to the silicon, and these compounds are, in consequence, much less polar than may have been anticipated. This is illustrated by both the electrostatic potential and by the observed dipole moment, particularly in the case of disilyl ether. This has important consequences in zeolite chemistry in that the charges on the silicon and oxygen atoms in the Si–O framework are predicted to be much less than the integral Si^{4+} O^{2-} charges used in recent calculations. Indeed, our results suggest that the Si–O framework in zeolites should be more appropriately considered as a non-polar extended π -system rather than an ionic one.

PARTIAL ATOMIC CHARGES OF AMINO ACIDS

Molecular modelling and molecular mechanics calculations are now used extensively in modelling peptide and protein conformations, their interactions with substrates, and solvation energies. Thus it is of some interest to derive the partial atomic charges of the amino acids from CHAR-GE2, and also to compare them with other commonly used routines. Here we give as examples the AMBER and ECEPP/2 charges. These are given in Table 2 which uses the nomenclature of Fig. 3.

Table 2 lists the partial atomic charges for all the amino acids as calculated by the various methods for the *N*-acetyl-*N'*-methanamide derivatives. The AMBER charges have been determined by scaling point charges to fit the ab-initio-derived electrostatic potential [22, 23]; and may be used with a dielectric constant of 1.0, for calculations in the gas phase and when including solvent molecules explicitly, or with a distance-dependent dielectric constant ($D = R_{ij}$) to approximate solvation effects. In both cases the 1,4 electrostatic interactions (i.e. between atoms separated by 3 covalent bonds) are halved to improve agreement with experiment [22]. The ECEPP/2 charges have

been derived from CNDO/2(ON) calculations and are used with a dielectric constant of 2.0 giving an effective dielectric constant of 4.0 [24].

The CHARGE2-derived partial atomic charges are very similar to the charges obtained by the

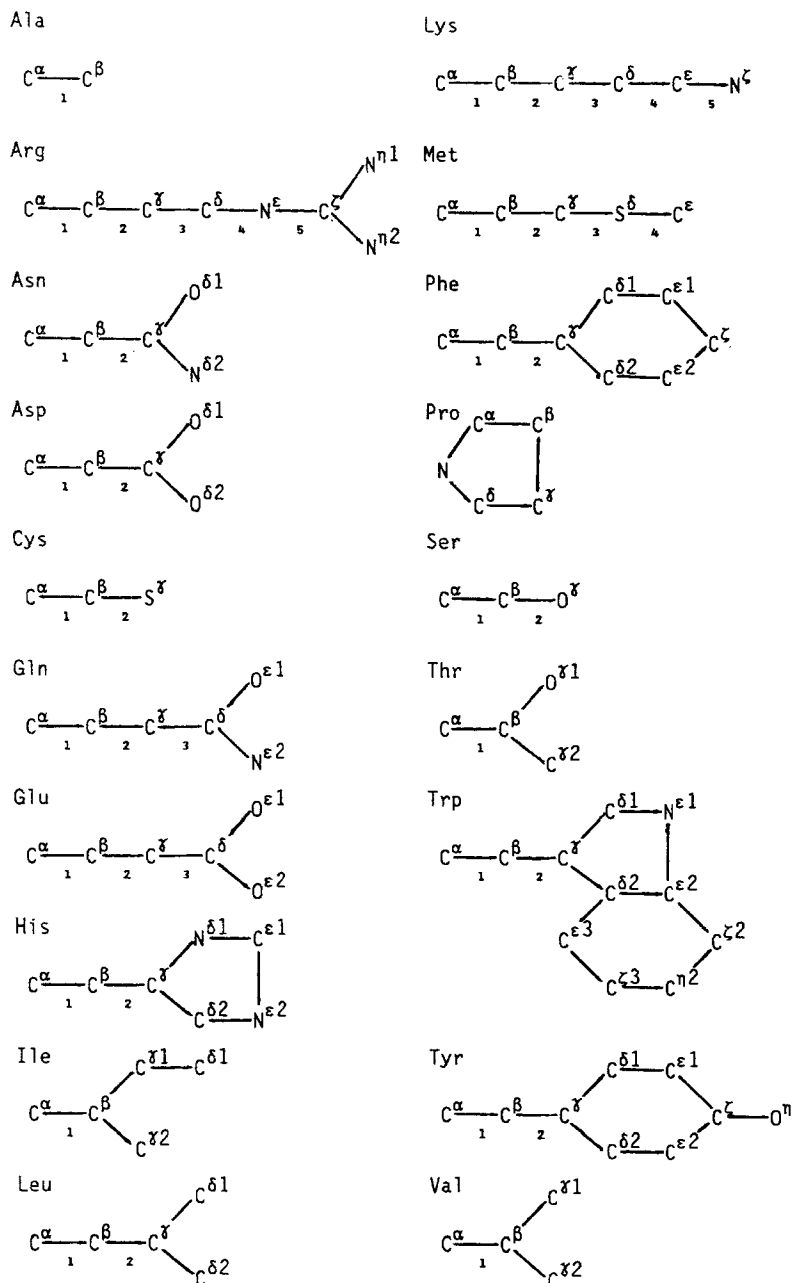


Fig. 3. Amino acid side-chain nomenclature.

TABLE 2
PARTIAL ATOMIC CHARGES OF AMINO ACIDS

Atom ^a	N	H	C ^{α}	H ^{α}	C ^{γ}	O	C ^{β}	H ^{β}
<i>Ala^a</i>								
AMBER	-0.463	0.252	0.035	0.048	0.616	-0.504	-0.098	0.038
CHARGE2	-0.259	0.224	0.035	0.072	0.275	-0.401	-0.087	0.047
ECEPP/2	-0.356	0.176	0.064	0.020	0.450	-0.384	-0.090	0.040
<i>Arg(+)</i>								
AMBER	-0.463	0.252	0.035	0.048	0.616	-0.504	-0.080	0.056
CHARGE2	-0.258	0.224	0.040	0.072	0.276	-0.401	-0.057	0.050
ECEPP/2	-0.356	0.176	0.064	0.020	0.450	-0.384	-0.030	0.015
<i>Asn</i>								
AMBER	-0.463	0.252	0.035	0.048	0.616	-0.504	-0.086	0.038
CHARGE2	-0.257	0.224	0.054	0.073	0.280	-0.401	-0.010	0.061
ECEPP/2	-0.356	0.176	0.064	0.020	0.450	-0.384	-0.120	0.055
<i>Asp(-)</i>								
AMBER	-0.463	0.252	0.035	0.048	0.616	-0.504	-0.398	0.071
CHARGE2	-0.257	0.224	0.052	0.073	0.279	-0.401	-0.018	0.060
ECEPP/2	-0.356	0.176	-0.060	0.024	0.450	-0.384	-0.170	-0.020
<i>Cys^b</i>								
AMBER	-0.463	0.252	0.035	0.048	0.616	-0.504	-0.060	0.038
CHARGE2	-0.254	0.224	0.052	0.075	0.279	-0.401	-0.013	0.064
ECEPP/2	-0.356	0.176	-0.060	0.024	0.450	-0.384	-0.105	0.055
<i>Gln</i>								
AMBER	-0.463	0.252	0.035	0.048	0.616	-0.504	-0.098	0.038
CHARGE2	-0.258	0.224	0.042	0.072	0.277	-0.401	-0.045	0.051
ECEPP/2	-0.356	0.176	0.064	0.020	0.450	-0.384	-0.030	0.020
<i>Glu(-)</i>								
AMBER	-0.463	0.252	0.035	0.048	0.616	-0.504	-0.184	0.092
CHARGE2	-0.258	0.224	0.042	0.072	0.277	-0.401	-0.047	0.051
ECEPP/2	-0.356	0.176	0.064	0.020	0.450	-0.384	-0.120	0.020
<i>Gly</i>								
AMBER	-0.463	0.252	0.035	0.032	0.616	-0.504		
CHARGE2	-0.262	0.224	0.027	0.070	0.275	-0.402		
ECEPP/2	-0.344	0.176	-0.008	0.055	0.450	-0.384		
<i>His(o)^c</i>								
AMBER	-0.463	0.252	0.035	0.048	0.616	-0.504	-0.098	0.038
CHARGE2	-0.257	0.224	0.048	0.073	0.278	-0.401	-0.027	0.056
ECEPP/2	-0.356	0.176	0.064	0.020	0.450	-0.384	-0.040	0.015
<i>His(+)</i>								
AMBER	-0.463	0.252	0.035	0.048	0.616	-0.504	-0.098	0.086
CHARGE2	-0.257	0.224	0.048	0.073	0.278	-0.400	-0.027	0.056
ECEPP/2	-0.356	0.176	0.064	0.020	0.450	-0.384	-0.050	0.065
<i>Ile</i>								
AMBER	-0.463	0.252	0.035	0.048	0.616	-0.504	-0.012	0.022
CHARGE2	-0.258	0.224	0.043	0.073	0.277	-0.401	-0.042	0.053
ECEPP/2	-0.356	0.176	0.064	0.020	0.450	-0.384	-0.005	0.025

TABLE 2
PARTIAL ATOMIC CHARGES OF AMINO ACIDS

<i>Arg (+)</i>	C ^γ	H ^γ	C ^δ	H ^δ	N ^ε	H ^ε	C ^ζ	N ^η	H ^η
AMBER	-0.103	0.074	-0.228	0.133	-0.324	0.269	0.760	-0.624	0.361
CHARGE2	-0.074	0.049	-0.042	0.058	-0.143	0.217	0.447	-0.153	0.173
ECEPP/2	-0.030	0.030	0.120	0.010	-0.300	0.230	0.580	-0.390	0.280
<i>Asn</i>	C ^γ	O ^{δ1}	N ^{δ2}	H ^{δ2}					
AMBER	0.675	-0.470	-0.867	0.344					
CHARGE2	0.258	-0.408	-0.378	0.222					
ECEPP/2	0.460	-0.375	-0.445	0.200					
<i>Asp (-)</i>	C ^γ	O ^δ							
AMBER	0.714	-0.721							
CHARGE2	0.018	-0.545							
ECEPP/2	0.500	-0.570							
<i>Cys^b</i>	S ^γ	H ^γ							
AMBER	0.827	0.135							
CHARGE2	-0.218	0.128							
ECEPP/2	0.015	0.010							
<i>Gln</i>	C ^γ	H ^γ	C ^γ	O ^{ε1}	N ^{ε2}	H ^{ε2}			
AMBER	-0.102	0.057	0.675	-0.470	-0.867	0.344			
CHARGE2	-0.037	0.057	0.252	-0.408	-0.378	0.222			
ECEPP/2	-0.110	0.050	0.470	-0.380	-0.430	0.185			
<i>Glu(-)</i>	C ^γ	H ^γ	C ^δ	O ^ε					
AMBER	-0.398	0.071	0.714	-0.721					
CHARGE2	0.043	0.056	0.009	-0.545					
ECEPP/2	-0.170	-0.040	0.500	-0.570					
<i>His(o)^c</i>	C ^γ	N ^{δ1}	H ^{δ1}	C ^{δ2}	H ^{δ2}	C ^{ε1}	H ^{ε1}	N ^{ε2}	
AMBER	-0.032	-0.146	0.228	0.195	0.018	0.241	0.036	-0.502	
CHARGE2	0.003	-0.137	0.224	-0.079	0.102	0.126	0.109	-0.398	
ECEPP/2	0.055	-0.250	0.150	0.080	0.035	0.190	0.020	-0.240	
<i>His(+)</i>	C ^γ	N ^{δ1}	H ^{δ1}	C ^{δ2}	H ^{δ2}	C ^{ε1}	H ^{ε1}	N ^{ε2}	H ^{ε2}
AMBER	0.058	-0.058	0.306	-0.037	0.153	0.114	0.158	-0.058	0.306
CHARGE2	0.032	-0.053	0.211	-0.005	0.101	0.402	0.107	-0.055	0.210
ECEPP/2	0.150	-0.020	0.275	0.100	0.135	0.275	0.150	-0.200	0.265
<i>Ile</i>	C ^{γ1}	H ^{γ1}	C ^{γ2}	H ^{γ2}	C ^δ	H ^δ			
AMBER	-0.049	0.027	-0.085	0.029	-0.085	0.028			
CHARGE2	-0.088	0.047	-0.114	0.044	-0.124	0.044			
ECEPP/2	-0.020	0.015	-0.075	0.025	-0.075	0.025			

TABLE 2
PARTIAL ATOMIC CHARGES OF AMINO ACIDS (*continued*)

Atom ^a	N	H	C ^α	H ^α	C'	O	C ^β	H ^β
<i>Leu</i>								
AMBER	−0.463	0.252	0.035	0.048	0.616	−0.504	−0.061	0.033
CHARGE2	−0.258	0.224	0.040	0.072	0.276	−0.401	−0.058	0.051
ECEPP/2	−0.356	0.176	0.064	0.020	0.450	−0.384	−0.030	0.022
<i>Lys(+)</i>								
AMBER	−0.463	0.252	0.035	0.048	0.616	−0.504	−0.098	0.038
CHARGE2	−0.258	0.224	0.040	0.072	0.266	−0.401	−0.060	0.050
ECEPP/2	−0.356	0.176	0.064	0.020	0.450	−0.384	−0.030	0.015
<i>Met^d</i>								
AMBER	−0.463	0.252	0.035	0.048	0.616	−0.504	−0.151	0.027
CHARGE2	−0.258	0.224	0.044	0.072	0.277	−0.401	−0.046	0.053
ECEPP/2	−0.356	0.176	0.064	0.020	0.450	−0.384	−0.010	0.030
<i>Phe</i>								
AMBER	−0.463	0.252	0.035	0.048	0.616	−0.504	−0.100	0.108
CHARGE2	−0.258	0.224	0.045	0.073	0.277	−0.401	−0.039	0.054
ECEPP/2	−0.356	0.176	0.064	0.020	0.450	−0.384	−0.040	0.025
<i>Pro</i>								
AMBER	−0.229	0.035	0.048	0.526	−0.500	−0.115	0.061	−0.021
CHARGE2	−0.175	0.047	0.073	0.278	−0.401	−0.055	0.050	−0.071
ECEPP/2	−0.285	0.050	0.040	0.455	−0.385	−0.025	0.015	−0.050
<i>Ser</i>								
AMBER	−0.463	0.252	0.035	0.048	0.616	−0.504	0.018	0.199
CHARGE2	−0.253	0.224	0.055	0.075	0.280	−0.401	−0.002	0.067
ECEPP/2	−0.356	0.176	−0.060	0.024	0.450	−0.384	0.130	0.020
<i>Thr</i>								
AMBER	−0.463	0.252	0.035	0.048	0.616	−0.504	0.170	0.082
CHARGE2	−0.253	0.224	0.057	0.076	0.280	−0.401	0.004	0.069
ECEPP/2	−0.356	0.176	0.064	0.020	0.450	−0.384	0.160	0.015
<i>Trp</i>								
AMBER	−0.463	0.252	0.035	0.048	0.616	−0.504	−0.098	0.038
CHARGE2	−0.257	0.224	0.045	0.073	0.278	−0.401	−0.037	0.054
ECEPP/2	−0.356	0.176	0.064	0.020	0.450	−0.384	−0.030	0.015
<i>Tyr</i>								
AMBER	−0.463	0.252	0.035	0.048	0.616	−0.504	−0.098	0.038
CHARGE2	−0.258	0.224	0.045	0.073	0.277	−0.401	−0.039	0.054
ECEPP/2	−0.356	0.176	0.064	0.020	0.450	−0.384	−0.040	0.025
<i>Val</i>								
AMBER	−0.463	0.252	0.035	0.048	0.616	−0.504	−0.012	0.024
CHARGE2	−0.258	0.224	0.042	0.073	0.277	−0.401	−0.044	0.052
ECEPP/2	−0.356	0.176	0.064	0.020	0.450	−0.384	0.008	0.016

^a See Fig. 3 for nomenclature. AMBER charges from Ref. 23 and ECEPP/2 charges from Ref. 24.

^b Cysteine. AMBER force field has lone pairs on sulphur carrying a charge of −0.481 each.

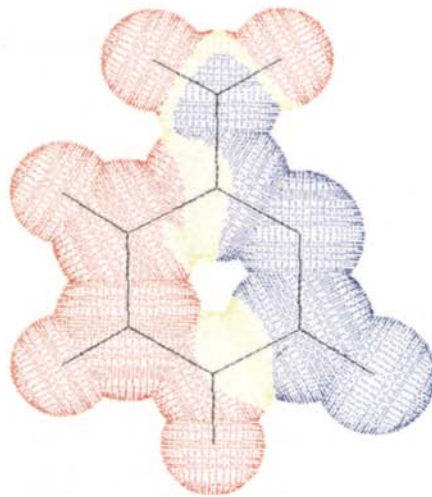
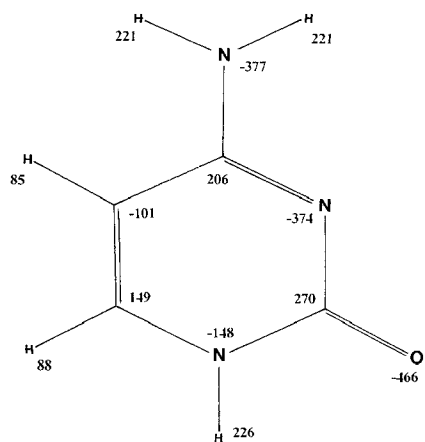
TABLE 2
PARTIAL ATOMIC CHARGES OF AMINO ACIDS (*continued*)

<i>Leu</i>	C ^γ	H ^γ	C ^δ	H ^δ				
AMBER	-0.010	0.031	-0.107	0.034				
CHARGE2	-0.067	0.049	-0.122	0.044				
ECEPP/2	-0.011	0.025	-0.074	0.025				
<i>Lys</i> (+)	C ^γ	H ^γ	C ^δ	H ^δ	C ^ε	H ^ε	N ^ζ	H ^ζ
AMBER	-0.160	0.166	-0.180	0.122	-0.038	0.098	-0.138	0.294
CHARGE2	-0.082	0.047	-0.080	0.049	-0.039	0.059	0.140	0.253
ECEPP/2	-0.025	0.020	-0.115	0.100	0.050	0.120	-0.320	0.320
<i>Met</i> ^d	C ^γ	H ^γ	S ^δ	C ^ε	H ^ε			
AMBER	-0.054	0.065	0.737	-0.134	0.065			
CHARGE2	-0.035	0.061	-0.218	-0.060	0.058			
ECEPP/2	-0.120	0.045	0.035	-0.190	0.055			
<i>Phe</i>	C ^γ	C ^δ	H ^δ	C ^ε	H ^ε	C ^ζ	H ^ζ	
AMBER	-0.100	-0.150	0.150	-0.159	0.150	-0.150	0.150	
CHARGE2	-0.042	-0.087	0.092	-0.091	0.091	-0.091	0.091	
ECEPP/2	0.020	-0.015	0.010	-0.015	0.010	0.015	0.005	
<i>Pro</i>	H ^γ	C ^δ	H ^δ					
AMBER	0.063	-0.012	0.060					
CHARGE2	0.049	-0.028	0.060					
ECEPP/2	0.025	0.100	0.010					
<i>Ser</i>	O ^γ	H ^γ						
AMBER	-0.550	0.310						
CHARGE2	-0.443	0.329						
ECEPP/2	-0.310	0.170						
<i>Thr</i>	O ^{γ1}	H ^{γ1}	C ^{γ2}	H ^{γ2}				
AMBER	-0.550	0.310	-0.191	0.065				
CHARGE2	-0.437	0.329	-0.093	0.048				
ECEPP/2	-0.310	0.170	-0.095	0.030				
<i>Trp</i>	C ^γ	C ^{δ1}	H ^{δ1}	C ^{δ2}	N ^{ε1}	H ^{ε1}	C ^{ε2}	C ^{ε3}
AMBER	-0.135	0.044	0.093	0.146	-0.352	0.271	0.154	-0.173
CHARGE2	-0.108	-0.008	0.101	-0.043	-0.192	0.223	0.031	-0.093
ECEPP/2	-0.025	0.085	0.000	0.005	-0.280	0.130	0.150	-0.080
	H ^{ε3}	C ^{γ2}	H ^{γ2}	C ^{γ3}	H ^{γ3}	C ^{η2}	H ^{η2}	
AMBER	0.086	-0.168	0.084	-0.066	0.057	-0.077	0.074	
CHARGE2	0.092	-0.095	0.094	-0.113	0.091	-0.103	0.091	
ECEPP/2	0.080	-0.060	0.030	-0.020	0.010	-0.005	0.010	
<i>Tyr</i>	C ^γ	C ^δ	H ^δ	C ^ε	H ^ε	C ^ζ	O ^η	H ^η
AMBER	-0.030	-0.002	0.064	-0.264	0.102	0.462	-0.528	0.334
CHARGE2	-0.058	-0.080	0.092	-0.089	0.095	0.041	-0.377	0.330
ECEPP/2	0.020	-0.010	0.010	-0.060	0.030	0.225	-0.330	0.165
<i>Val</i>	C ^γ	H ^γ						
AMBER	-0.091	0.031						
CHARGE2	-0.115	0.044						
ECEPP/2	-0.072	0.025						

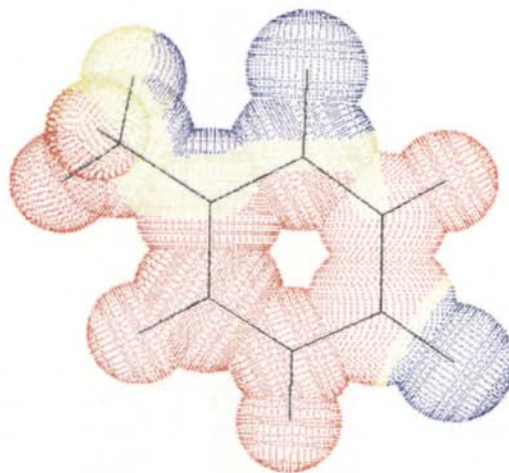
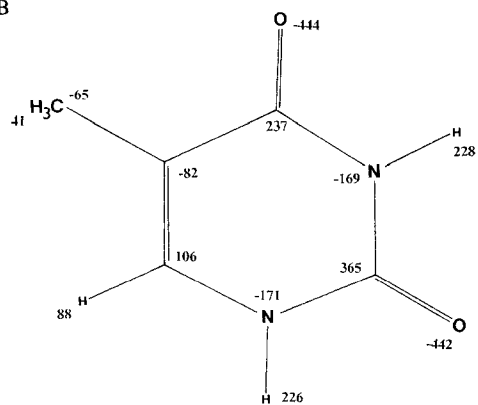
^cH(δ1) form.

^dAMBER force field has lone pairs on sulphur bearing a charge of -0.381 each.

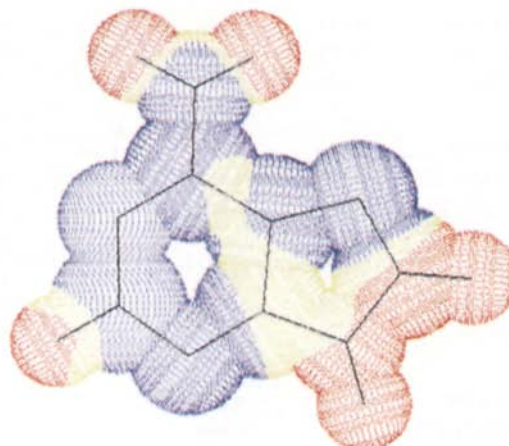
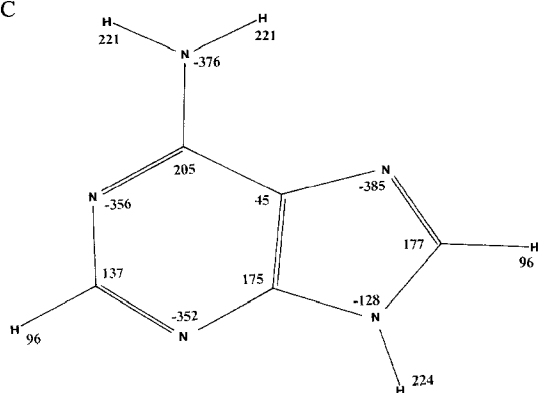
A



B



C



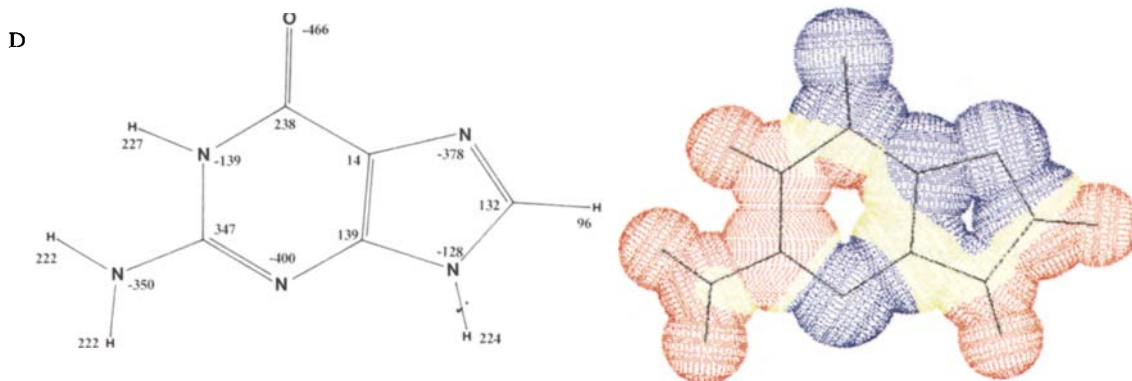


Fig. 4. CHARGE2 partial atomic charges for (A) cytosine, (B) thymine, (C) adenine and (D) guanine with the corresponding electrostatic potential (colour coding as in Fig. 2).

previous version of CHARGE which used CNDO/2 calculated π -densities for the aromatic residues [3]. These partial atomic charges were shown to give a better correlation of the ^1H and ^{13}C NMR chemical shifts with charge compared with the ECEPP/2 charges. The only significant difference between the former scheme and CHARGE2 arises for RCO_2 groups and aromatics where the π -densities are now calculated using the Hückel procedure. This does not change the good correlation obtained previously between the calculated charges and the ^1H and ^{13}C chemical shifts.

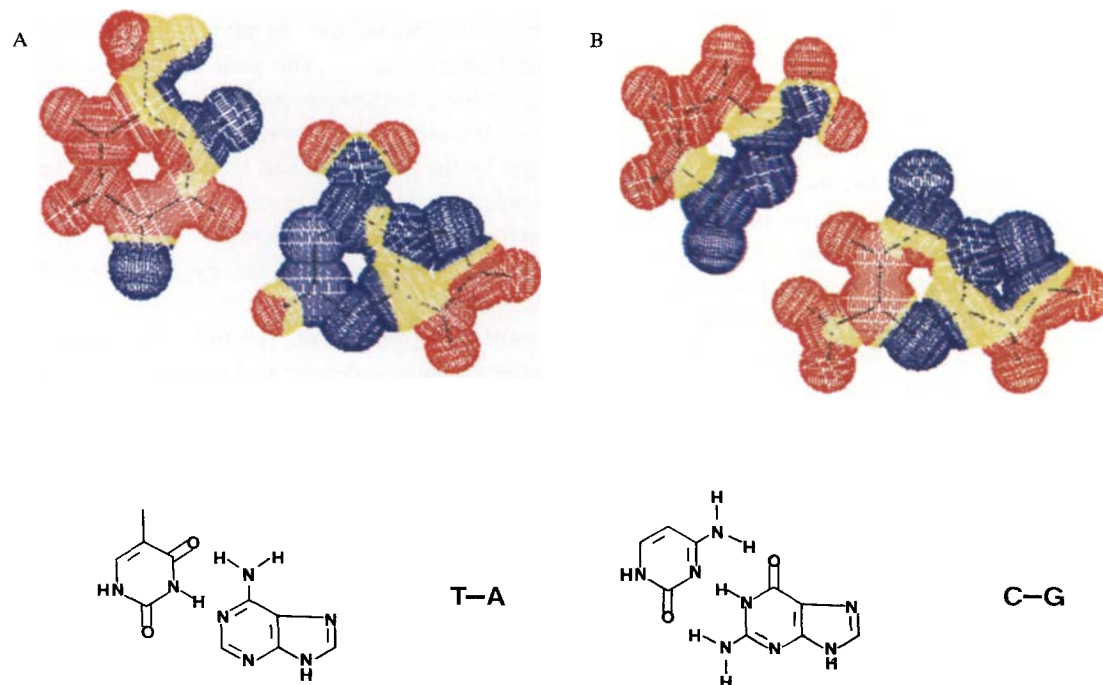


Fig. 5. Base-pairing complementarity in A-T and C-G base pairs.

Examination of Table 2 shows considerable disagreement between the methods concerning the backbone charges (N, H, C(α), C' and O). As was previously found for the nucleic acid bases, the AMBER charges are consistently larger than the other methods with the CHARGE2 calculations giving generally lower partial atomic charges. As expected all three methods give very similar charges for the hydrocarbon side chains e.g. Ala, Ile, Leu and Val, with the large differences occurring for ionised side chains i.e. C(γ) in Asp(−), C(δ) in Glu(−) and N(ζ) in Lys(+). The latter case is very interesting as AMBER and ECEPP/2 predict the N(ζ) atom to be negatively charged, even though it carries a formal positive charge, whereas CHARGE2 predicts a positive nitrogen atom.

The AMBER force field includes two lone pairs on sulphur atoms which carry a charge of −0.381 each for Met and −0.481 each for Cys. Therefore sulphur atoms in the AMBER force field have a large positive charge compared with an essentially neutral sulphur in ECEPP/2, and a negative sulphur from CHARGE2 calculations. Parameterisation of the charge scheme for sulphur compounds did not necessitate the use of lone-pair dipoles in the present scheme. Lastly, according to the ECEPP/2 charges the polarity of the O–H bond is substantially lower than for either the AMBER or CHARGE2 charges (see Ser, Thr and Tyr).

NUCLEIC ACID BASES

The importance of the electrostatic contribution to the conformational energy of DNA can hardly be overemphasised. Indeed as hydrogen bonding is largely electrostatic, the dominant complexing interaction in the DNA molecule must be due to electrostatic forces.

This is emphasised by the very polar nature of the nucleic acid bases. They are very polar systems with molecular dipole moments of ca. 4–6 D and possess extensive π -systems. This initially proved a problem for the CHARGE parameterisation, as the Hückel routine gave, as might have been anticipated, much too large π -charges and hence dipole moments. This well-known characteristic of the Hückel method was overcome by the use of the ω -technique, in which the Coulomb integral for each atom varies according to the charge on the atom. This provides a simple and effective method of reducing the excessive π -charges given by the unconstrained Hückel method. In CHARGE2 this ω -technique operates only if the atom has a π -charge in excess of 200 m.e. Thus any π -system with lower calculated π -charges than this for any atom is unaffected.

With this modification the observed dipole moments of a variety of pyridines, pyrimidines and purines, including the nucleic acid bases were well reproduced by CHARGE2.

Here we show as examples (Fig. 4) the calculated partial atomic charges (σ - and π -charges together) obtained for the four nucleic acid bases cytosine, thymine, adenine and guanine together with the electrostatic potential map obtained from these charges, using the CHEM-X routine [25].

The relationship between the atomic charges and the corresponding electrostatic potentials is of interest, as it is not always as obvious as may have been expected. For example, in cytosine the electrostatic potential around the NH₂ group clearly illustrates the contrast between the positive hydrogen atoms and the negative nitrogen as expected. However, the electrostatic potential around the equally polar carbonyl group does not reflect the positive charge on the carbon, as the effect of this positive charge is more than compensated by the large negative charges on the adjoining nitrogen and oxygen atoms. On the electrostatic potential map, the most positive carbon

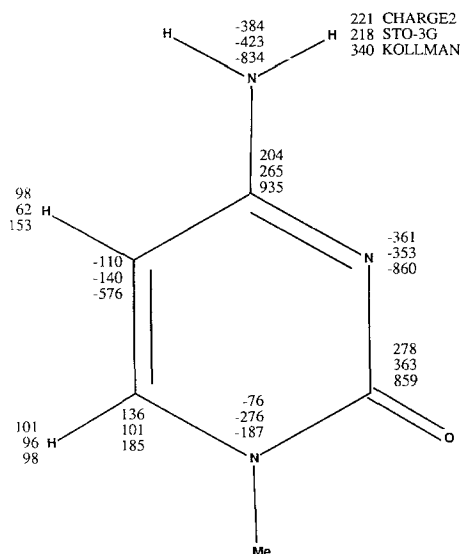


Fig. 6. Comparison of atomic charges from CHARGE2, STO-3G and Kollman for 1-methyl cytosine.

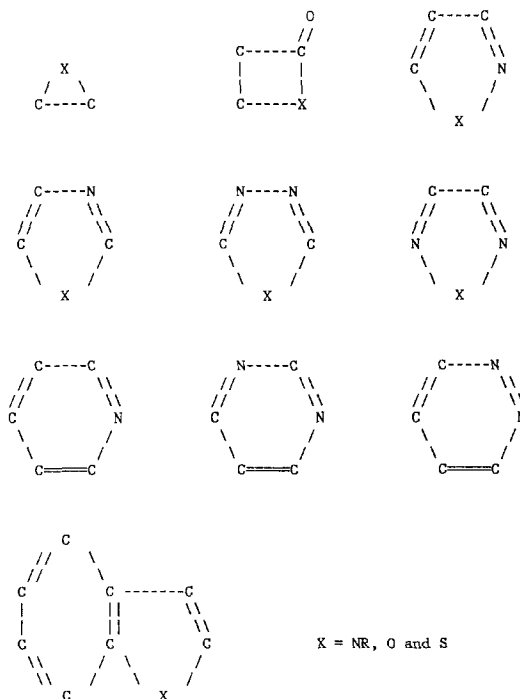


Fig. 7. Cycle systems capable of inclusion into CHARGE2: Naphthalenes, nucleotide bases and sugars.

atoms are the olefine carbons C_5 and C_6 , not the carbonyl carbon. This 'smearing out' of the charge distribution is seen throughout these illustrations.

The electrostatic potential does, however, beautifully illustrate the complementarity of the A-T and C-G base pairs (shown in Fig. 5). The electrostatic interaction energy between these base pairs in the Watson-Crick model has been calculated from the partial atomic charges given, and compared with both the observed and other calculated values. The electrostatic energies given from the CHARGE2 charges are in reasonable agreement with the observed interaction energies (-7.7 and -20.6 kcal·mol $^{-1}$ for A-T and G-C, respectively compared to -13.0 and -21.0 observed), and compare very favourably with other calculated values. In particular the calculated values using partial atomic charges from MO methods are invariably much less than the observed values. Only the energies given by the Kollman charges (-10.1 and -17.0 kcal·mol $^{-1}$) compare reasonably with the observed values, although the charges given by Kollman are invariably much greater than in all other methods. This is illustrated in Fig. 6 which gives the partial atomic charges for 1-methyl cytosine calculated by CHARGE2, STO-3G (Mulliken Population Analysis) and Kollman. The first two methods are in general agreement, but the Kollman charges are often very much larger, some unreasonably large (C_4 935 m.e., N_3 -860 m.e., etc.).

TABLE 3

TYPES OF COMPOUNDS FOR WHICH CHARGES CAN BE CALCULATED USING CHARGE2^a

	Alkanes	Alkenes	Alkynes	Benzenes	Pyrroles	Furans	Thiophenes
-R	*	*	*	*	!	!	*
-F	*	*		*	!	!	!
-Cl	*	*		*	!	!	*
-Br	*	*		*	!	!	*
-I	*	!		!	!	!	!
-OR	*	*		*	!	!	!
-SR	*	*		*	!	!	!
-NR ₂	*	!		*	!	!	!
-NH·CO·R	*	!		!	!	!	!
-CO·OR	*	!		*	!	!	!
-COR	*	*		*	!	!	!
-SiR ₃	*	!		!	!	!	!
-OSiR ₃	*	!		!	!	!	!
-CN	*	*		*	!	!	*
-CCH	*	!		*	!	!	!
-CO·X	*	*		*	!	!	!
-NO ₂	*	*		*	!	!	*
-CO·NH ₂	*	!		!	!	!	!
-O·CO·R	*	!		!	!	!	!
-Ph	*	*		!	!	!	!
-C=S·R	*	!		!	!	!	!

^aThe program should stop on finding unparameterised atoms or systems.

R = H or alkyl; X = F, Cl or Br.

*Compounds have been tested.

! Compounds have not been tested but can be run.

CONCLUSIONS

The above examples serve to illustrate the general utility of the CHARGE2 procedure for calculating partial atomic charges in molecules. At present this routine can handle a wide variety of molecules and atoms quickly and easily. An illustrative summary of the types of functional groups and ring systems which can be input into CHARGE2 is given in Table 3 and Fig. 7. All these molecules can be run in essentially seconds on an IBM PC making this routine very useful for simple calculations or as an input to more extensive molecular modelling routines.

An incidental but considerable advantage of CHARGE2 over the more complex M.O. calculations is that it is a connectivity-based programme, i.e. the calculated charges are functions only of the molecular connectivity, not of the molecular geometry (with the single exception of conjugation between overlapping π -systems which is dependent on the dihedral angle between them). Obviously the calculated dipole moments will depend on the molecular geometry. This facility means that an accurate molecular geometry is not essential for CHARGE2 to function.

ACKNOWLEDGEMENTS

G.H.G., I.S.H. and P.E.S. acknowledge SERC/CASE awards with I.C.I. (New Science Group), I.C.I. (Advanced Materials) and Roche Products Ltd respectively.

REFERENCES

- 1 Abraham, R.J., Griffiths, L. and Loftus, P., *J. Comp. Chem.*, 3 (1982) 407.
- 2 Abraham, R.J. and Hudson, B.D., *J. Comp. Chem.*, 5 (1984) 562.
- 3 Abraham, R.J. and Hudson, B.D., *J. Comp. Chem.*, 6 (1985) 173.
- 4 Abraham, R.J. and Smith, P.E., *J. Comp. Chem.*, 9 (1988) 288.
- 5 Abraham, R.J. and Grant, G.H., *J. Comp. Chem.*, 9 (1988) 244.
- 6 Abraham, R.J. and Smith, P.E., *Nucleic Acids Res.*, 16 (1988) 2639.
- 7 Abraham, R.J. and Smith, P.E., *J. Comput.-Aided Mol. Design*, 3 (1989) 175.
- 8 Gasteiger, J. and Marsili, M., *Tetrahedron*, 36 (1980) 3219.
- 9 Gasteiger, J. and Saller, H., *Angew. Chem. Int. Ed. Engl.*, 24 (1985) 687.
- 10 Gasteiger, J. and Marsili, M., *Org. Mag. Reson.*, 25 (1981) 353.
- 11 Varma, R., McDiarmid, A.G. and Miller, J.G., *Inorg. Chem.*, 3 (1964) 1754.
- 12 Hinze, J. and Jaffe, H.H., *J. Am. Chem. Soc.*, 84 (1962) 540.
- 13 Mulliken, R.S., *J. Chem. Phys.*, 2 (1934) 782.
- 14 Pauling, L., *J. Am. Chem. Soc.*, 54 (1932) 3570.
- 15 Huggins, M.L., *J. Am. Chem. Soc.*, 75 (1953) 4123.
- 16 Laurie, V.W., *J. Chem. Phys.*, 26 (1957) 1359.
- 17 Luke, B.T., Pople, J.A., Krogh-Jespersen, M., Apeloig, Y., Chandrasakekhar, J. and Schleyer, R.P. von, *J. Am. Chem. Soc.*, 108 (1986) 260.
- 18 Allred, A.L. and Rochow, E.G., *J. Inorg. Nucl. Chem.*, 1 (1955) 112.
- 19 Pola, J. and Chvalovsky, V., *Coll. Czech. Chem. Comm.*, 42 (1977) 3581.
- 20 Muentner, J.S. and Laurie, V.W., *J. Chem. Phys.*, 45 (1966) 855.
- 21 Shiki, Y., Kuginuki, Y., Hasegawa, A. and Hayashi, H., *J. Mol. Spectrosc.*, 73 (1978) 9.
- 22 Weiner, S.J., Kollman, P.A., Case, D.A., Singh, U.C., Ghio, C., Alagona, G., Profeta Jr., S. and Weiner, P., *J. Am. Chem. Soc.*, 106 (1984) 765.
- 23 Weiner, S.J., Kollman, P.A., Nguyen, D.T. and Case, D.A., *J. Comp. Chem.*, 7 (1986) 230.
- 24 (a) Momany, F.A., McGuire, R.F., Burgess, A.W. and Scheraga, H.A., *J. Phys. Chem.*, 79 (1975) 2361; (b) Nemethy, G., Pottle, M.S. and Scheraga, H.A., *J. Phys. Chem.*, 87 (1983) 1883.
- 25 Chem-X, developed and distributed by Chemical Design Ltd., Oxford, U.K.