Automated site-directed drug design: An assessment of the transferability of atomic residual charges (CNDO) for molecular fragments

P.-L. Chau and P.M. Dean*

Department of Pharmacology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QJ, U.K.

SUMMARY

In this paper a database of atomic residual charges has been constructed for all the molecular fragments defined previously in a combinatorial search of the Cambridge Structural Database. The charges generated for the atoms in each fragment are compared with charges calculated for whole molecules containing those fragments. The fragment atomic charges lie within 1 S.D. of the mean for 68%, and within 2 S.D. for 91%, of the atoms whose charges were computed for whole molecules. The actual charges on any atom are strongly influenced by the adjacent connected atoms. There is a large spread of atomic residual charge within the fragments database.

INTRODUCTION

In previous papers [1,2] we discussed the creation of a complete database of frequently occurring molecular fragments for use in site-directed drug design. Each fragment has its geometry determined from a statistical analysis using the Cambridge Structural Database (CSD). This paper describes the assignment of residual atomic charges to the atoms of each fragment and analyses statistically whether the atomic charges calculated for the isolated fragment, whose geometry had been previously determined [2], are consistent with those obtained from a population of fragments contained in known crystal structures.

Many quantum chemistry packages are available for the calculation of electronic properties of molecules. More recently, a rapid procedure for calculating partial atomic charges has been proposed in a series of pioneering papers by Abraham et al. [3]. Their procedure is an alternative to

---

* To whom correspondence should be addressed.

methods based on Mulliken population analysis and charges derived from the electrostatic potential computed from quantum chemistry. In our work, the method chosen must be applicable to the generation of large molecules, since the electron distribution on large ligand molecules has to be calculated for future purposes using a consistent method. Only this strategy would enable us to obtain a direct comparison of the electronic charge distributions in the fragment in isolation and the fragment when it exists in a larger molecule. The method also has to be efficient because a large number of calculations has to be performed. The choice, therefore, was limited to the semi-empirical methods; ab initio methods are so costly in computer time that they become impracticable for the calculation of electronic distribution of hundreds of molecules with more than 10 atoms or with second-row elements. CNDO was chosen for 3 reasons. Firstly, it is one of the most widely used packages in drug design. Secondly, average protein residual atomic charge data are available, calculated using CNDO from the ECEPP programme [4]. In subsequent work we would like to examine the electrostatic complementarity between receptor sites and ligands built up from these molecular fragments. Thirdly, the CNDO residual charge output can be used with VSS [5] to calculate the molecular van der Waals surface electrostatic potential. Therefore our main justification for using CNDO charges is primarily to maintain consistency within our evolving programs for automated site-directed drug design.

COMPUTING METHODS

Our approach has been to determine the electron distributions, using CNDO, by two methods on the commonly occurring fragments [1]. Firstly we used a statistically derived geometry for each fragment [2]; hydrogen atoms were added to the empty valencies where appropriate using standard bond lengths, bond angles and torsion angles. Atomic residual charges were then computed using CNDO [6] on these fragments. Secondly the fragments were used to search through the CSD for molecules comprised of less than 50 atoms containing the fragment. Uncharged molecules only were selected for the neutral fragments. Retrieved molecules not containing hydrogen atoms were discarded. After these screens the atomic Cartesian coordinates of the retained molecules were used for input to the CNDO program. Results were accepted if the energy optimization converged successfully within 25 iterations.

The CSD provides a facility for storing the coordinates of the fragment and the retrieved molecule separately. A matching program was written to identify and extract the intramolecular fragment and its associated atomic residual charges from the fragment. In cases where a molecule contained more than one analogous fragment, only the first was chosen to ensure a wide statistical sampling. Furthermore, at least 10 screened fragments were used for statistical analysis.

The output from CNDO gives the valence shell electron density on each atom in the molecule used for input. The residual atomic charge was calculated by subtracting the valence shell electron density from the number of valence shell electrons of the single neutral atom in the unexcited state. Comparisons were then made between the residual atomic charges on the fragment derived from averaged geometries (called the predicted fragment atomic charge) and those obtained from a large number of molecules containing each fragment (called the observed fragment atomic charge). The fractional deviation of each atom is then defined as

$$D_f = |q^p - q^o|/\sigma$$

where q^p is the predicted charge, q^o is the mean observed charge and σ is the standard deviation in the charge distribution for that atom in the fragment.

RESULTS

The fractional deviation D_f shows by how much the predicted value differs from the standard deviation. The measure can therefore be used to classify the predictability of charges calculated for an isolated fragment compared with a sample of residual charges for that fragment within a population of molecules. A gross statistical analysis of the predictability of fragment atom charges is shown in Figs. 1 and 2. The fractional deviation is compared both with the mean atomic residual charge for C, N, O, H, F, Cl and S (Fig. 1A), and with the standard deviation of atomic residual charge (Fig. 1B). For C, N and O the atom types were further divided into connectivity classes for comparison. An example for oxygen is shown in Figs. 2A and B. These figures provide a general overview for the predictability of CNDO-derived charges.

Atoms with very well predicted charges lie to the left, and below a fractional deviation of 1.0, of the scattergrams in Figs. 1B and 2B. Charges on the oxygen atoms are clearly dependent on the connectivity; if the oxygen has one double bond the charge is more negative and better predicted than an oxygen atom linked to neighbours by two single bonds. Carbon atoms have charges in the range of -0.2 to $+0.6$ although some clustering is apparent around 0.0 (Fig. 1A); there are some very well predicted charges. The charges on nitrogen atoms are less well predicted; the atoms are predominantly negatively charged within the range -0.3 to $+0.1$. Oxygen atoms are scattered in

A Predictability of atomic residual electronic charge   B   Constancy of atomic residual electronic charge
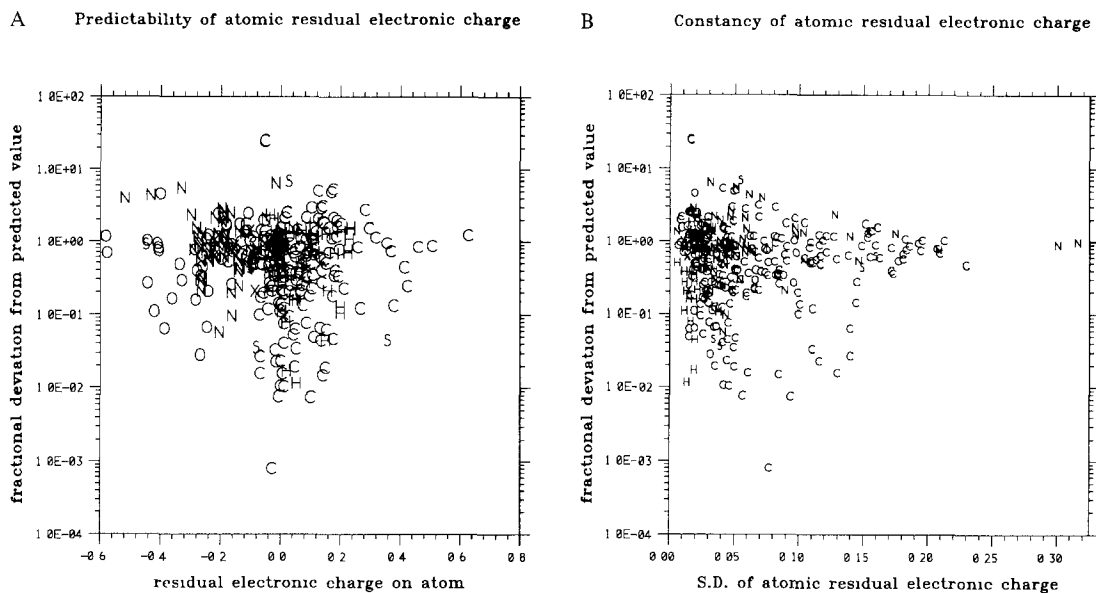
Fig. 1. (A) Predictability of atomic residual charges from molecular fragments. The predicted value is derived from an idealized geometry of the fragment, and is compared with the statistical mean of values calculated from molecules containing these fragments in the Cambridge Structural Database (CSD). (B) Fractional deviation of the predicted charge, as defined above, plotted against the standard deviation of the atomic residual electronic charge for atoms considered in this study.

410

Fig. 2 (A) Predictability of atomic residual charges for oxygen atoms from molecular fragments. The label 1 denotes that the oxygen atom is attached to one other atom only; the label 2 indicates that the oxygen has two s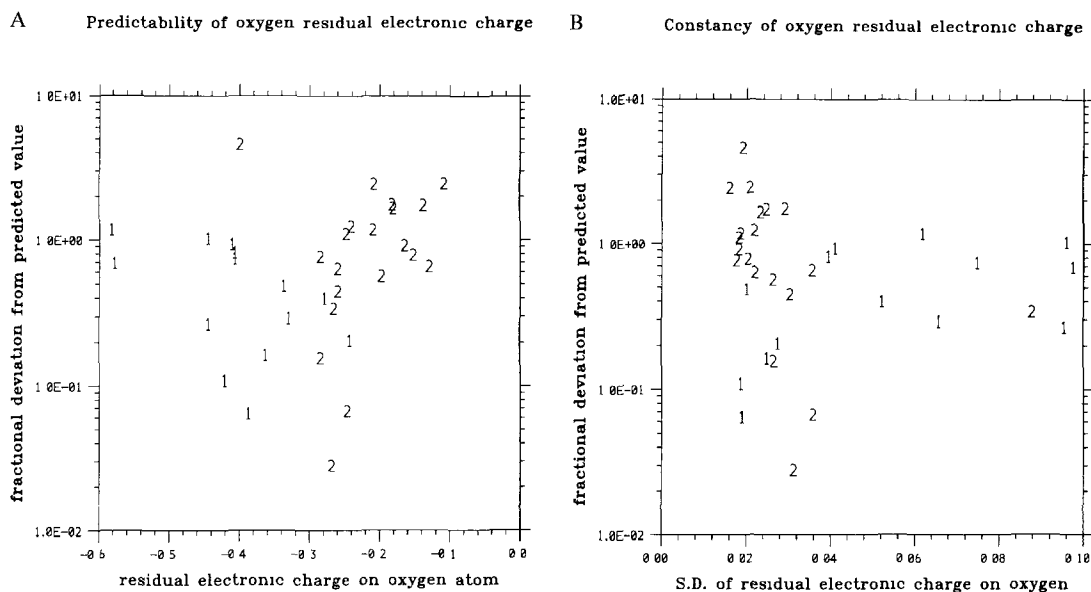ingle bonds attached to it. The predicted value is derived from an idealized geometry of the fragment, and is compared with the statistical mean of values calculated from molecules containing these fragments in the Cambridge Structural Database (CSD). (B) Fractional deviation of the predicted charge, as defined above, plotted against the standard deviation of the atomic residual electronic charge for atoms considered in this study.

the range -0.6 to -0.1. Hydrogen atoms usually have a very small positive charge but they are found in the range -0.05 to $+0.25$. Halogen atoms are negatively charged with fluorine -0.2 and chlorine -0.1. Sulphur atoms show outliers at -0.4 and $+0.4$ but most are slightly negative.

A full statistical analysis of all frequently occurring fragments is given in Tables 1–8. Table 1 provides residual atomic charges on all non-charged 3-atom fragments. Of the 16 fragments considered statistically, 5 fragments (3.k2, 3.m, 3.nl, 3.o and 3.q) have badly predicted charges compared with those obtained from the statistical analysis; these fragments contain O or S. Table 2 shows the data for non-charged 4-atom fragments containing 1 double bond. Of the 18 fragments considered statistically only 3 fragments (4.2.b, 4.2.j, 4.2.q) contain poorly predicted charges; these contain nitrogen atoms. In contrast 4-atom fragments containing only single bonds show poor charge predictability (Table 3). The output for 5-atom fragments is shown in Table 4; only one fragment (5.g), out of 17, shows poor predictability. The data for charged fragments are given in Table 5; of the 7 fragments considered statistically only two (Q.b, Q.h) show badly predicted atoms. Residual atomic charges for aromatic rings are shown in Tables 6–8. Five-membered (Table 6) and 6-membered (Table 7) aromatic rings all show well predicted charges. With fused bicyclic aromatic rings, isoindole and purine rings are the only ones badly predicted (Table 8).

DISCUSSION

The 113 molecular fragments considered here were chosen using objective geometric criteria

TABLE 1
ATOMIC RESIDUAL CHARGES ON 3-ATOM FRAGMENTS[a]

Fragment number	Atom	Calc.	Obs. ± S.D.	n	Comment
3.a	C	0.007	− 0.027 ± 0.028	45	< 2 S.D.
	C	− 0.115	− 0.052 ± 0.044		< 2 S.D.
	C	− 0.004	0.098 ± 0.115		< 1 S.D.
3.b	C	0.095	0.092 ± 0.016	21	< 1 S.D.
	N	− 0.165	− 0.159 ± 0.031		< 1 S.D.
	C	− 0.024	0.004 ± 0.081		< 1 S.D.
3.c	C	− 0.066	− 0.106 ± 0.028	22	< 2 S.D.
	C	− 0.058	− 0.014 ± 0.022		< 2 S.D.
	H	0.064	0.067 ± 0.006		< 1 S.D
3.d	C	0.251		< 10	
	S	− 0.328			
	N	− 0.119			
3.e	N	− 0.121		< 10	
	C	0.066			
	C	0.077			
3.f	N	0.001	− 0.086 ± 0.059	17	< 2 S.D.
	C	0.022	0.171 ± 0.126		< 2 S.D.
	C	− 0.210	0.015 ± 0.175		< 2 S.D.
3.g	N	− 0.014	− 0.061 ± 0.064	61	< 1 S.D.
	C	− 0.004	0.053 ± 0.089		< 1 S.D.
	C	− 0.195	− 0.120 ± 0.039		< 2 S.D.
3.h	N	0.002	− 0.059 ± 0.056	62	< 2 S.D.
	C	0.028	0.104 ± 0.081		< 1 S.D.
	O	− 0.198	− 0.213 ± 0.026		< 1 S.D.
3.ı	N	− 0.180		< 10	
	C	0.090			
	H	0.097			
3.j	N	0.166		< 10	
	N	− 0.332			
	N	− 0.180			
3.k1	O	− 0.204		< 10	
	Csp2	0.166			
	Csp2	0.161			
3.k2	O	− 0.208	− 0.247 ± 0.016	227	> 2 S.D.
	Csp2	0.469	0.409 ± 0.049		< 2 S.D.
	Csp3	0.130	0.127 ± 0.034		< 1 S.D.

TABLE 1 (continued)

Fragment number	Atom	Calc.	Obs. ± S.D.	n	Comment
3.k3	O	− 0.210	− 0.232 ± 0.019	31	< 2 S.D.
	Csp3	0.136	0.156 ± 0.050		< 1 S.D.
	Csp3	0.139	0.139 ± 0.084		< 1 S.D.
3.1	O	− 0.138	− 0.181 ± 0.025	35	< 2 S.D.
	C	0.129	0.222 ± 0.155		< 1 S.D.
	N	− 0.193	0.077 ± 0.301		< 1 S.D.
3.m	O	− 0.108	− 0.159 ± 0.021	14	> 2 S.D.
	C	0.135	0.374 ± 0.081		> 2 S.D.
	O	− 0.165	− 0.148 ± 0.018		< 1 S.D.
3.nl	O	− 0.241	− 0.268 ± 0.022	196	< 2 S.D.
	Csp2	0.175	0.394 ± 0.041		> 2 S.D.
	H	0.151	0.180 ± 0.022		< 2 S.D.
3.n2	O	− 0.260	− 0.274 ± 0.022	198	< 1 S.D.
	Csp3	0.133	0.135 ± 0.044		< 1 S.D.
.	H	0.157	0.161 ± 0.021		< 1 S.D.
3.o	O	− 0.180	− 0.219 ± 0.023	98	< 2 S.D.
	N	− 0.190	− 0.064 ± 0.047		> 2 S.D.
	H	0.149	0.149 ± 0.019		< 1 S.D.
3.p	S	− 0.082	− 0.080 ± 0.039	68	< 1 S.D.
	C	− 0.063	0.022 ± 0.121		< 1 S.D.
	C	− 0 059	− 0.040 ± 0.099		< 1 S.D.
3.q	O	− 0.182	− 0.233 ± 0.029	23	< 2 S.D.
	C	0.148	0.149 ± 0.049		< 1 S.D.
	S	0.029	0.406 ± 0.055		> 2 S.D.
3.r	S	− 0.045	− 0.065 ± 0.038	24	< 1 S.D.
	C	− 0.045	− 0.032 ± 0.110		< 1 S.D.
	S	− 0.093	− 0.057 ± 0.031		< 2 S.D.

a The fragments are drawn in Fig. 1 of Chau and Dean [1]. Charges are calculated for the isolated fragment atoms (Calc.) and charges are obtained from a statistical sample of molecules containing that fragment (Obs. ± S.D., n = number of observations). The atoms follow the order in Fig. 1 [1]: central atom, left atom, right atom.

from a larger database generated combinatorially [1]. The average geometry of each fragment was determined from an extensive set of larger molecules containing the fragment in question [2]. In this work, we calculated the atomic residual charge for each fragment atom using CNDO. We also calculated the atomic residual charge for fragment atoms when the fragment is part of a larger molecule. Using a large sample of constituent molecules, we were able to determine statistically the variation in atomic residual charge of the fragments when they exist in different molecules.

TABLE 2

ATOMIC RESIDUAL CHARGES ON 4-ATOM FRAGMENTS CONTAINING 1 DOUBLE BOND[a]

Fragment number	Atom	Calc.	Obs. $\pm$ S.D.	n	Comment
4.2.a	C	0.066	-0.026 ± 0.082	20	< 2 S.D.
	C	-0.087	0.018 ± 0.109		< 1 S.D.
	C	-0.012	0.206 ± 0.212		< 2 S.D.
	C	-0.014	0.018 ± 0.180		< 1 S.D.
4.2.b	C	0.200	0.041 ± 0.090	46	< 2 S.D.
	N	-0.330	-0.047 ± 0.051		> 2 S.D.
	C	-0.053	0.168 ± 0.156		< 2 S.D.
	C	-0.054	0.089 ± 0.194		< 1 S.D.
4.2.c	C	0.265	0.261 ± 0.035	57	< 1 S.D
	O	-0.279	-0.300 ± 0.052		< 1 S.D
	C	-0.071	-0.067 ± 0.139		< 1 S.D.
	C	-0.071	$-0.069 \pm 0\ 129$		< 1 S.D.
4.2.d	C	0.178	0.150 ± 0.083	23	< 1 S.D.
	C	-0.177	$-0\ 120 \pm 0.109$		< 1 S.D.
	C	-0.035	0.108 ± 0.209		< 1 S.D.
	N	-0.238	-0.160 ± 0.146		< 1 S.D.
4.2.e	C	$0\ 210$	0.225 ± 0.065	14	< 1 S.D.
	C	-0.165	-0.143 ± 0.084		< 1 S.D.
	C	-0.041	0.067 ± 0.229		< 1 S.D.
	O	-0.246	-0.244 ± 0.036		< 1 S.D.
4.2.f	C	0.024	-0.052 ± 0.064	58	< 2 S.D.
	C	-0.057	0.036 ± 0.080		< 2 S.D.
	C	-0.007	-0.191 ± 0.153		< 2 S.D.
	H	0.006	0.020 ± 0.015		< 1 S.D.
4 2.g	C	0.316	0.262 ± 0.047	10	< 2 S D.
	N	-0.432	-0.164 ± 0.061		< 2 S D.
	C	-0.070	-0.017 ± 0.108		< 1 S.D.
	N	-0.250	-0.210 ± 0.058		< 1 S.D.
4.2.h	C	0.367	0.355 ± 0.017	99	< 1 S.D.
	O	-0.362	-0.358 ± 0.025		< 1 S.D.
	C	-0.088	$-0\ 028 \pm 0.104$		< 1 S.D.
	N	-0.248	-0.210 ± 0.047		< 1 S.D.
4.2.ı	C	0.347		< 10	
	N	-0.405			
	C	-0.073			
	O	-0.266			

TABLE 2 (continued)

Fragment number	Atom	Calc.	Obs. ± S.D.	n	Comment
4.2.j	C	0.176	0.030 ± 0.069	25	> 2 S.D.
	N	− 0.298	− 0.001 ± 0.128		> 2 S.D.
	C	− 0.046	0.151 ± 0.182		< 2 S.D.
	H	− 0.034	0.012 ± 0.020		> 2 S.D.
4.2.k	C	0.413	0.404 ± 0.020	256	< 1 S.D.
	O	− 0.329	− 0.349 ± 0.066		< 1 S.D.
	C	− 0.094	− 0.038 ± 0.116		< 1 S.D.
	O	− 0.266	− 0.296 ± 0.088		< 1 S.D.
4.2.l	C	0.257	0.245 ± 0.014	20	< 1 S.D.
	O	− 0.242	− 0.248 ± 0.028		< 1 S.D.
	C	− 0.069	− 0.079 ± 0.059		< 1 S.D.
	H	− 0.040	− 0.045 ± 0.015		< 1 S.D.
4.2.m	C	0.299		< 10	
	C	− 0.301			
	N	− 0.238			
	N	− 0.239			
4.2.n	C	0.145	0.043 ± 0.180	19	< 1 S.D.
	C	− 0.169	− 0.006 ± 0.207		< 1 S.D.
	N	− 0.219	0.091 ± 0.316		< 1 S.D.
	H	− 0.001	0.021 ± 0.026		< 1 S.D.
4.2.o	C	0.163		< 10	
	C	− 0.127			
	O	− 0.239			
	H	0.008			
4.2.p	C	− 0.036	− 0.065 ± 0.057	66	< 1 S.D.
	C	− 0.023	0.027 ± 0.063		< 1 S.D.
	H	0.027	0.016 ± 0.016		< 1 S.D.
	H	0.019	0.018 ± 0.019		< 1 S.D.
4.2.q	C	0.423	0.411 ± 0.048	22	< 1 S.D.
	N	− 0.518	− 0.235 ± 0.071		> 2 S.D.
	N	− 0.267	− 0.211 ± 0.104		< 1 S.D.
	N	− 0.265	− 0.246 ± 0.089		< 1 S.D.
4.2.r	C	0.459	0.449 ± 0.011	34	< 1 S.D.
	O	− 0.421	− 0.419 ± 0.019		< 1 S.D.
	N	− 0.271	− 0.229 ± 0.049		< 1 S.D.
	N	− 0.269	− 0.253 ± 0.054		< 1 S.D.
4.2.s	C	0.296		< 10	
	N	− 0.405			
	N	− 0.241			
	H	− 0.042			

TABLE 2 (continued)

Fragment number	Atom	Calc.	Obs. $\pm$ S.D.	n	Comment
4.2.t	C	0.506	0.500 $\pm$ 0.006	45	< 1 S.D.
	O	$-$ 0.387	$-$ 0.386 $\pm$ 0.019		< 1 S.D.
	O	$-$ 0.285	$-$ 0.272 $\pm$ 0.018		< 1 S.D.
	N	$-$ 0.274	$-$ 0.236 $\pm$ 0.039		< 1 S.D.
4.2.u	C	0.354	0.347 $\pm$ 0.008	10	< 1 S.D.
	O	$-$ 0.337	$-$ 0.327 $\pm$ 0.020		< 1 S.D.
	N	$-$ 0.236	$-$ 0.187 $\pm$ 0.058		< 1 S.D.
	H	$-$ 0.040	$-$ 0.037 $\pm$ 0.013		< 1 S.D.
4.2.v	C	0.109		< 10	
	C	$-$ 0.048			
	C	$-$ 0.014			
	Cl	$-$ 0.162			
4.2.w	C	0.296	0.282 $\pm$ 0.009	32	< 2 S.D.
	S	$-$ 0.448	$-$ 0.462 $\pm$ 0.015		< 1 S.D.
	N	$-$ 0.193	$-$ 0.148 $\pm$ 0.046		< 1 S.D.
	N	$-$ 0.188	$-$ 0.161 $\pm$ 0.046		< 1 S.D.

[a] The fragments are drawn in Fig. 2 of Chau and Dean [1]. Charges are calculated for the isolated fragment atoms (Calc.) and charges are obtained from a statistical sample of molecules containing that fragment (Obs. $\pm$ S.D., n = number of observations). The atoms follow the order in Fig. 2 [1]: central atom, other atom on bond 1, other atom on bond 2, other atom on bond 3.

The variation is found to be generally small, and this means that the charges have a high transferability from our developing database of fragments for further use in drug design. Approximately 91% of the atoms studied in all the fragments lie within 2 S.D. of the mean value found in a sample of molecules containing the fragment; this is in line with statistical expectations from a normal distribution.

An examination of Fig. 1A reveals, as expected, that there is a considerable spread in atomic residual charge in the range -0.6 to $+0.6$. However, if the atoms are considered by their type and connectivity there is still a spread. For example, with oxygen double bonded to another atom the charge lies in the range -0.58 to -0.24; oxygen attached by single bonds to its neighbours can be found to have a charge in the range -0.4 to -0.1 (Fig. 2A). There is a similar spread of charge and overlap with other atoms with different connectivities. What emerges clearly, by careful scrutiny of the data in Tables 1–4, is the overriding effect of adjacent bonded atoms on the charge of the central atom of the fragment. Furthermore, if an atom in a fragment is terminal, for example a carbonyl oxygen atom, then the other atoms of the fragment have an influence on that terminal atom charge. In other words, if we are to create a database of charges for predictive use in drug design, by assembling atoms on molecular graphs generated in the site, then we must consider fragments as basic building units.

It could be argued that the fragments used here are themselves too small and that one should consider a bigger fragment; the combinatorics of this suggestion would be too large to allow an

TABLE 3
ATOMIC RESIDUAL CHARGES ON 4-ATOM FRAGMENTS CONTAINING SINGLE BONDS ONLY[a]

Fragment number	Atom	Calc.	Obs. ± S.D.	n	Comment
4.1.a	N	− 0.138	− 0.146 ± 0.005	15	< 2 S.D.
	C	0.084	0.087 ± 0.030		< 1 S.D.
	C	0.084	0.061 ± 0.022		< 2 S.D.
	C	0.085	0.067 ± 0.009		> 2 S.D.
4.1.b	N	− 0.085		< 10	
	N	− 0.214			
	C	0.088			
	C	0.088			
4.1.c	N	− 0.060		< 10	
	O	− 0.202			
	C	0.102			
	C	0.088			
4.1.d	N	− 0.205	− 0.239 ± 0.038	14	< 1 S.D.
	H	0.112	0.141 ± 0.016		< 2 S.D.
	Csp2	0.125	0.386 ± 0.051		> 2 S.D.
	Csp2	0.188	0.334 ± 0.155		< 2 S.D.
4.1.e	N	− 0.205	− 0.203 ± 0.041	106	< 1 S.D.
	H	0.112	0.134 ± 0.017		< 2 S.D.
	Csp3	0.099	0.099 ± 0.093		< 1 S.D.
	Csp3	0.140	0.364 ± 0.068		> 2 S.D.
4.1.f	N	− 0.187	− 0.213 ± 0.023	14	< 2 S.D.
	H	0.094	0.122 ± 0.022		< 1 S.D.
	Csp3	0.085	0.111 ± 0.075		< 1 S.D.
	Csp3	0.084	0.107 ± 0.044		< 1 S.D.
4.1.g	N	− 0.140	− 0.130 ± 0.041	64	< 1 S.D.
	H	0.107	0.130 ± 0.019		< 2 S.D.
	N	− 0.209	− 0.090 ± 0.097		< 2 S.D.
	C	0.125	0.323 ± 0.096		> 2 S.D.
4.1.h	N	− 0.260	− 0.250 ± 0.034	105	< 1 S.D.
	H	0.115	0.137 ± 0.032		< 1 S.D.
	H	0.114	0.135 ± 0.020		< 2 S.D.
	C	0.143	0.331 ± 0.104		< 2 S.D.
4.1.i	N	− 0.136	− 0.147 ± 0.026	33	< 1 S.D.
	H	0.084	0.089 ± 0.019		< 1 S.D.
	H	0.070	0.096 ± 0.018		< 2 S.D.
	N	− 0.219	− 0.119 ± 0.047		> 2 S.D.
4.1.j	N	− 0.170		< 10	
	C	0.107			
	S	− 0.649			
	H	0.130			

[a] The fragments are drawn in Fig. 3 of Chau and Dean [1]. Charges are calculated for the isolated fragments (Calc.) and charges are obtained from a statistical sample of molecules containing the fragments (Obs. ± S.D., n = number of observations). The atoms follow the order of Fig. 3 [1]: central atom, other atom of bond 1, other atom of bond 2 and other atom of bond 3.

TABLE 4
ATOMIC RESIDUAL CHARGES ON 5-ATOM FRAGMENTS[a]

Fragment number	Atom	Calc.	Obs. $\pm$ S.D.	n	Comment
5.a	C	0.063	0.024 ± 0.046	26	< 1 S.D.
	C	0.013	0.154 ± 0.159		< 1 S.D.
	C	0.047	0.038 ± 0.139		< 1 S.D
	C	0.074	0.011 ± 0.113		< 1 S.D.
	C	0.083	0.011 ± 0.116		< 1 S.D.
5.b	C	0.160	0.096 ± 0.041	22	< 2 S.D.
	C	-0.021	0.118 ± 0.183		< 1 S.D.
	C	-0.009	0.012 ± 0.144		< 1 S.D.
	C	-0.020	0.074 ± 0.177		< 1 S.D.
	N	-0.289	-0.131 ± 0.208		< 1 S.D.
5.c	C	0.193	0.179 ± 0.040	44	< 1 S.D.
	C	-0.024	0.043 ± 0.171		< 1 S.D.
	C	-0.020	-0.017 ± 0.110		< 1 S.D.
	C	-0.018	-0.015 ± 0.115		< 1 S.D.
	O	-0.284	-0.280 ± 0.026		< 1 S.D.
5.d	C	0.041	0.010 ± 0.044	42	< 1 S.D.
	C	-0.006	0.082 ± 0.108		< 1 S.D.
	C	-0.005	0.058 ± 0.173		< 1 S.D.
	C	-0.005	0.035 ± 0.144		< 1 S.D.
	H	-0.001	0.012 ± 0.024		< 1 S.D.
5.e	C	0.242		< 10	
	C	-0.038			
	C	-0.038			
	N	-0.302			
	N	-0.303			
5.f	C	0.287		< 10	
	C	-0.048			
	C	-0.058			
	N	-0.297			
	O	-0.283			
5.g	C	0.140	0.048 ± 0.049	70	< 2 S.D.
	C	-0.021	0.180 ± 0.195		< 2 S.D.
	C	-0.026	0.133 ± 0.185		< 1 S.D.
	N	-0.281	-0.132 ± 0.097		< 2 S.D.
	H	-0.023	0.020 ± 0.021		> 2 S.D.
5.h	C	0.325		< 10	
	C	-0.066			
	C	-0.066			
	O	-0.282			
	O	-0.281			

TABLE 4 (continued)

Fragment number	Atom	Calc.	Obs. ± S.D.	n	Comment
5.i	C	0.170	0.133 ± 0.033	54	< 2 S.D.
	C	− 0.035	0.106 ± 0.115		< 2 S.D.
	C	− 0.031	0.098 ± 0.148		< 1 S.D.
	O	− 0.269	− 0.268 ± 0.031		< 1 S.D.
	H	− 0.014	0.006 ± 0.025		< 1 S.D.
5.j	C	0.015	− 0.034 ± 0.044	104	< 2 S.D.
	C	− 0.004	0.051 ± 0.110		< 1 S.D.
	C	− 0.003	0.178 ± 0.192		< 1 S.D.
	H	0.001	0.024 ± 0.026		< 1 S.D.
	H	0.001	0.024 ± 0.026		< 1 S.D.
5.k	C	0.245		< 10	
	C	− 0.045			
	N	− 0.294			
	N	− 0.291			
	H	− 0.041			
5.l	C	0.114	0.071 ± 0.049	62	< 1 S.D.
	C	− 0.025	0.115 ± 0.171		< 1 S.D.
	N	− 0.276	− 0.112 ± 0.139		< 2 S.D.
	H	− 0.019	0.023 ± 0.030		< 2 S.D.
	H	− 0.005	0.019 ± 0.027		< 1 S.D
5.m	C	0.307		< 10	
	C	− 0.062			
	O	− 0.270			
	O	− 0.275			
	H	− 0.034			
5.n	C	0.150	0.135 ± 0.029	146	< 1 S.D.
	C	− 0.028	0.060 ± 0.138		< 1 S.D.
	O	− 0.260	− 0.246 ± 0.030		< 1 S.D.
	H	− 0.009	0.003 ± 0.024		< 1 S.D.
	H	− 0.016	0.001 ± 0.018		< 1 S D.
5.o	C	− 0.033	− 0.074 ± 0.049	482	< 1 S.D.
	C	0.002	0.175 ± 0.153		< 2 S.D.
	H	0.012	0.028 ± 0.019		< 1 S.D.
	H	0.012	0.029 ± 0.020		< 1 S.D.
	H	0.012	0.028 ± 0.020		< 1 S.D.
5.p	C	0.626	0.589 ± 0.030	21	< 2 S.D.
	C	− 0.116	0.142 ± 0.151		< 2 S.D.
	F	− 0.223	− 0.212 ± 0.010		< 2 S.D.
	F	− 0.223	− 0.215 ± 0.013		< 1 S.D.
	F	− 0.222	− 0.213 ± 0.011		< 1 S.D.

TABLE 4 (continued)

Fragment number	Atom	Calc.	Obs. ± S.D.	n	Comment
5.q	C	0.216		< 10	
	N	− 0.286			
	N	− 0.286			
	H	− 0.018			
	H	− 0.029			
5.r	C	0.072	0.063 ± 0.019	103	< 1 S.D.
	N	− 0.264	− 0.126 ± 0.103		< 2 S.D.
	H	− 0 008	0.012 ± 0.016		< 2 S.D.
	H	0.003	0.012 ± 0.020		< 1 S.D.
	H	− 0.006	0.012 ± 0.018		< 2 S.D.
5.s	C	0.292		< 10	
	O	− 0.265			
	O	− 0.260			
	H	− 0.027			
	H	− 0.029			
5.t	C	0.109	0.110 ± 0.014	334	< 1 S.D.
	O	− 0.248	− 0.228 ± 0.018		< 2 S.D.
	H	0.003	0.006 ± 0.011		< 1 S.D.
	H	− 0.004	0.005 ± 0.012		< 1 S.D.
	H	0.003	0.004 ± 0.011		< 1 S.D.
5.u	C	0.137		< 10	
	C	− 0.011			
	C	− 0.008			
	C	− 0.011			
	Cl	− 0.206			
5.v	C	0.096		< 10	
	C	− 0.007			
	C	− 0.008			
	H	0.024			
	Cl	− 0.177			
5.w	C	0.186	0.160 ± 0.022	21	< 2 S.D.
	C	0.001	0.145 ± 0.121		< 2 S.D.
	Cl	− 0.086	− 0.083 ± 0.014		< 1 S.D.
	Cl	− 0.102	− 0.081 ± 0.015		< 2 S.D.
	Cl	− 0.090	− 0 082 ± 0.013		< 1 S.D.
5.x	C	0.055	− 0.006 ± 0.042	30	< 2 S.D.
	C	− 0.002	0.248 ± 0.161		< 2 S.D.
	H	0.031	0.054 ± 0.017		< 2 S.D.
	H	0.031	0.053 ± 0.017		< 2 S.D.
	Cl	− 0.154	− 0.129 ± 0.029		< 1 S.D.

TABLE 4 (continued)

Fragment number	Atom	Calc.	Obs. ± S.D.	n	Comment
5.y	C	− 0.021	− 0.058 ± 0.045	59	< 1 S.D.
	S	− 0.117	− 0.018 ± 0.159		< 1 S.D.
	C	0.014	0.109 ± 0.162		< 1 S.D.
	H	0.042	0.045 ± 0.018		< 1 S.D.
	H	0.038	0.044 ± 0.019		< 1 S.D.
5.z	C	− 0.090	− 0.102 ± 0.019	67	< 1 S.D.
	S	− 0.100	− 0.035 ± 0.149		< 1 S.D.
	H	0.051	0.051 ± 0.013		< 1 S.D.
	H	0.058	0.054 ± 0.011		< 1 S.D.
	H	0.052	0.054 ± 0.015		< 1 S.D.

[a] The fragments are drawn in Fig. 4 of Chau and Dean [1]. Charges are calculated for the isolated fragment atoms (Calc.) and charges are obtained from a statistical sample of molecules containing that fragment (Obs. ± S.D., n = number of observations). The atoms follow the order in Fig. 4 [1]: central atom, other atom of bond 1, other atom of bond 2, other atom of bond 3, other atom of bond 4.

adequate sampling of molecular data. In our judgement there is no need to use bigger fragments since 91% of the predicted charges, derived from the idealized geometries, lie within 2 S.D. of the mean charges found in a sample population of molecules; other atom connections beyond the fragment could be considered to be randomly distributed. If the fragment size had been too small, then the predicted atomic residual charges would not have correlated so well with statistical expectations of a normal distribution. This paper has provided an analysis of the population fragment atomic charges and the computed charges of the idealized fragment atoms. The difference between the idealized fragment charges and the mean is an indication of the predictive error of using the idealized fragments. If the charges lie close to the population mean the predictive error will be low; the further the charges are from the mean then the error will increase. Therefore, those charges lying within 1 S.D. would be expected to be highly predictive. In subsequent papers in preparation we will demonstrate that the charges documented here, for fragment atoms, are sufficient to predict the electrostatic potential with a high correlation coefficient compared with the potential calculated de novo by CNDO on the whole molecule.

Automated site-directed drug design is an extremely complex problem that still awaits even an approximate solution. Conceptually it is plagued with combinatoric sub-problems at two levels: the generation of molecular graphs in the site, and the assembly of atoms onto those graphs once they are positioned in the site. If atoms are to be assembled on an existing graph, with evolving properties complementary to the site, then one is faced with either recalculating the charges for each atom substitution made, or with using fragment-derived charges. Clearly the latter choice is a much faster procedure if the charges to be selected have satisfactorily predictable properties.

In summary, these three papers have presented a chemical fragments database composed of all possible 3-, 4- and 5-atom molecular fragments, together with some aromatic fragments, from a specified set of atoms. All the fragments are frequently occurring amongst organic molecules. We have calculated the average geometric and electronic properties of each fragment; they are found

TABLE 5
ATOMIC RESIDUAL CHARGES ON THE CHARGED FRAGMENTS[a]

Fragment number	Atom	Calc.	Obs. $\pm$ S.D.	n	Comment
Q.a	P	0.377		< 10	
	O	− 0.427			
	O	− 0.434			
	O	− 0.397			
	O	− 0.214			
	C	0.381			
	H	0.120			
Q. b	P	0.262	0.251 $\pm$ 0.010	19	< 2 S.D.
	O	− 0.444	− 0.418 $\pm$ 0.096		< 1 S.D.
	O	− 0.445	− 0.346 $\pm$ 0.096		< 2 S.D.
	C	− 0.134	− 0.061 $\pm$ 0.072		< 2 S.D.
	O	− 0.400	− 0.312 $\pm$ 0.019		> 2 S.D.
	H	0.202	0.165 $\pm$ 0.027		< 2 S.D.
Q.c	S	0.357	0.356 $\pm$ 0.034	45	< 1 S.D.
	O	− 0.410	− 0.372 $\pm$ 0.041		< 1 S.D.
	O	− 0.407	− 0.375 $\pm$ 0.039		< 1 S.D.
	O	− 0.407	− 0.351 $\pm$ 0.075		< 1 S.D.
	C	− 0.094	− 0.021 $\pm$ 0.129		< 1 S.D.
Q.d	C	0.377	0.380 $\pm$ 0.023	140	< 1 S.D
	O	− 0.582	− 0.509 $\pm$ 0.062		< 2 S.D.
	O	− 0.578	− 0.510 $\pm$ 0.098		< 1 S.D.
	C	− 0.104	0.017 $\pm$ 0.118		< 2 S.D.
Q.e	N	0.000	− 0.019 $\pm$ 0.024	117	< 1 S.D.
	H	0.225	0.205 $\pm$ 0.026		< 1 S.D.
	H	0.229	0.204 $\pm$ 0.023		< 2 S.D.
	C	0.047	0.038 $\pm$ 0.049		< 1 S.D.
	H	0.225	0.205 $\pm$ 0.029		< 1 S.D.
Q.f	N	0.041	− 0.006 $\pm$ 0.031	28	< 2 S.D.
	C	0.048	0.075 $\pm$ 0.034		< 1 S.D.
	H	0.199	0 203 $\pm$ 0.038		< 1 S.D.
	C	0.048	0.079 $\pm$ 0.051		< 1 S.D.
	H	0.199	0.203 $\pm$ 0.034		< 1 S.D
Q.g	N	0.072	0.028 $\pm$ 0.032	22	< 2 S.D.
	C	0.050	0.066 $\pm$ 0.054		< 1 S.D.
	C	0.050	0.048 $\pm$ 0.049		< 1 S.D.
	C	0.050	0.063 $\pm$ 0.033		< 1 S.D.
	H	0.179	0.207 $\pm$ 0.040		< 1 S.D.
Q.h	N	− 0.016	− 0.219 $\pm$ 0.031	23	> 2 S.D
	C	0.282	0.442 $\pm$ 0.058		> 2 S.D.
	H	0.234	0.193 $\pm$ 0.026		< 2 S.D.
	H	0.235	0.186 $\pm$ 0.033		< 2 S.D.

[a] The fragments are drawn in Fig. 5 of Chau and Dean [1]. Charges are calculated for the isolated fragment atoms (Calc.) and charges are obtained from a statistical sample of molecules containing that fragment (Obs. $\pm$ S.D., n = number of observations). The atoms follow the order in Fig. 5 [1]: central atom, followed by the undescribed atoms in the increasing bond number.

TABLE 6
ATOMIC RESIDUAL CHARGES FOR 5-MEMBERED RING FRAGMENTS[a]

Fragment number	Atom	Calc.	Obs. $\pm$ S.D.	n	Comment
R5.a	N	-0.101	-0.079 ± 0.028	15	< 1 S.D.
	C	0.027	0.047 ± 0.050		< 1 S.D.
	C	-0.032	-0.022 ± 0.048		< 1 S.D.
	C	-0.032	-0.041 ± 0.039		< 1 S.D.
	C	0.032	0.062 ± 0.084		< 1 S.D.
R5.b	N	-0.059	-0.016 ± 0.043	16	< 2 S.D.
	N	-0.142	-0.155 ± 0.034		< 1 S.D.
	C	0.088	0.073 ± 0.047		< 1 S.D.
	C	-0.073	-0.063 ± 0.099		< 1 S.D.
	C	0.073	0.087 ± 0.100		< 1 S.D.
R5.c	N	-0.120	-0.092 ± 0.021	19	< 2 S.D.
	C	0.152	0.151 ± 0.015		< 1 S.D.
	N	-0.185	-0.209 ± 0.019		< 2 S.D.
	C	0.044	0.039 ± 0.025		< 1 S.D.
	C	0.073	0.021 ± 0.030	•	< 1 S.D.
R5.d	N	-0.077	-0.042 ± 0.046	10	< 1 S.D.
	N	-0.164	-0.168 ± 0.043		< 1 S.D.
	C	0.174	0.176 ± 0.050		< 1 S.D.
	N	-0.218	-0.249 ± 0.051		< 1 S.D.
	C	0.182	0.215 ± 0.084		< 1 S.D.
R5.e	N	-0.140		< 10	
	C	0.136			
	N	-0.133			
	N	-0.131			
	C	0.138			
R5.f	N	-0.082		< 10	
	N	-0.052			
	N	-0.013			
	N	-0.150			
	C	0.158			
R5.g	O	-0.152	-0.168 ± 0.020	40	< 1 S.D.
	C	0.101	0.132 ± 0.028		< 2 S.D.
	C	-0.038	-0.057 ± 0.027		< 1 S.D.
	C	-0.039	-0.013 ± 0.027		< 1 S.D.
	C	0.106	0.119 ± 0.038		< 1 S.D.
R5.h	O	-0.130	-0.153 ± 0.036	12	< 1 S.D.
	N	-0.080	-0.141 ± 0.077		< 1 S.D.
	C	0.106	0.200 ± 0.086		< 2 S.D.
	C	-0.091	-0.132 ± 0.081		< 1 S.D.
	C	0.160	0.206 ± 0.065		< 1 S.D.

[a] The fragments are drawn in Fig. 6 of Chau and Dean [1]. Charges are calculated for the isolated fragment (Calc.) and charges are obtained from a statistical sample of molecules containing that fragment (Obs. $\pm$ S.D., n = number of observations). The atoms are numbered in sequence in the order of the bonds.

TABLE 7
ATOMIC RESIDUAL CHARGES FOR 6-MEMBERED AROMATIC RING FRAGMENTS[a]

Fragment number	Atom	Calc.	Obs. ± S.D.	n	Comment
R6.a	C	0.006	0.032 ± 0.066	47	< 1 S.D.
	C	0.009	0.007 ± 0.041		< 1 S.D.
	C	0.009	0.008 ± 0.045		< 1 S.D.
	C	0.011	0.038 ± 0.070		< 1 S.D.
	C	0.009	0.003 ± 0.030		< 1 S.D.
	C	0.008	0.004 ± 0.040		< 1 S.D.
R6.b	N	− 0.143	− 0.163 ± 0.036	49	< 1 S.D.
	C	0.101	0.139 ± 0.076		< 1 S.D.
	C	− 0.030	− 0.030 ± 0.077		< 1 S.D.
	C	0.045	0.044 ± 0.035		< 1 S.D.
	C	− 0.028	− 0.037 ± 0.038		< 1 S.D.
	C	0.103	0.112 ± 0.042		< 1 S.D.
R6.c	N	− 0.072		< 10	
	N	− 0.076			
	C	0.070			
	C	0.007			
	C	0.013			
	C	0.072			
R6.d	N	− 0.172	− 0.217 ± 0.055	22	< 1 S.D.
	C	0.194	0.251 ± 0.073		< 1 S.D.
	N	− 0.173	− 0.220 ± 0.058		< 1 S.D.
	C	0.134	0.173 ± 0.068		< 1 S.D.
	C	− 0.068	− 0.096 ± 0.043		< 1 S.D.
	C	0.136	0.147 ± 0.041		< 1 S.D.
R6.e	N	− 0.109	− 0.129 ± 0.042	19	< 1 S.D.
	C	0.064	− 0.091 ± 0.073		< 1 S.D.
	C	0.064	0.079 ± 0.068		< 1 S.D.
	N	− 0.109	− 0.120 ± 0.026		< 1 S.D.
	C	0.065	0.076 ± 0.059		< 1 S.D.
	C	0.065	0.087 ± 0.060		< 1 S.D.
R6.f	N	− 0.198		< 10	
	C	0.223			
	N	− 0.199			
	C	0.223			
	N	− 0.199			
	C	0.221			

[a] The fragments are drawn in Fig. 7 of Chau and Dean [1]. Charges are calculated for the isolated fragment (Calc.) and charges are obtained from a statistical sample of molecules containing that fragment (Obs. ± S.D., n = number of observations). The atoms are numbered in sequence in the order of the bonds.

TABLE 8
ATOMIC RESIDUAL CHARGES FOR FUSED AROMATIC RING FRAGMENTS[a]

Fragment number	Atom	Calc.	Obs. $\pm$ S.D.	n	Comment
Rf.a.	C	-0.028	-0.042 ± 0.013	23	< 2 S.D.
	C	0.010	0.005 ± 0.017		< 1 S.D.
	C	-0.008	-0.009 ± 0.056		< 1 S.D.
	C	-0.001	0.000 ± 0.044		< 1 S.D.
	C	0.008	-0.010 ± 0.014		< 2 S.D.
	C	0.088	0.095 ± 0.010		< 1 S.D.
	N	-0.142	-0.122 ± 0.036		< 1 S.D.
	C	0.069	0.062 ± 0.025		< 1 S.D.
	C	-0.062	-0.043 ± 0.040		< 1 S.D.
Rf.b	C	-0.007	0.020 ± 0.023	27	< 2 S.D.
	C	0.003	0.014 ± 0.028		< 1 S.D.
	C	0.002	0.006 ± 0.029		< 1 S.D.
	C	-0.006	0.021 ± 0.024		< 2 S.D.
	C	0.021	-0.018 ± 0.015		> 2 S.D.
	C	0.020	-0.016 ± 0.014		> 2 S.D.
	C	-0.055	0.343 ± 0.016		> 2 S.D.
	N	-0.049	-0.169 ± 0.058		> 2 S.D.
	C	-0.054	0.344 ± 0.016		> 2 S.D.
Rf.c	C	-0.022		< 10	
	C	0.002			
	C	-0.003			
	C	-0.015			
	C	0.077			
	C	0.068			
	N	-0.153			
	C	0.175			
	N	-0.209			
Rf.d	N	-0.196	-0.247 ± 0.018	36	> 2 S.D.
	C	0.189	0.210 ± 0.013		< 2 S.D.
	N	-0.189	-0.264 ± 0.034		> 2 S.D.
	C	0.102	0.243 ± 0.048		> 2 S.D.
	C	0.013	-0.049 ± 0.029		> 2 S.D.
	C	0.187	0.202 ± 0.008		< 2 S.D.
	N	-0.168	-0.126 ± 0.017		> 2 S.D.
	C	0.172	0.167 ± 0.028		< 1 S.D.
	N	-0.191	-0.212 ± 0.017		< 2 S.D.
Rf.e	C	-0.002		< 10	
	C	0.002			
	C	0.005			
	C	-0.005			
	C	0.022			
	C	0.017			
	C	0.014			

TABLE 8 (continued)

Fragment number	Atom	Calc.	Obs ± S.D.	n	Comment
	O	− 0.084			
	C	0.003			
Rf.f	C	− 0.006	0.041 ± 0.090	37	< 1 S.D.
	C	0.007	− 0.007 ± 0.065		< 1 S.D.
	C	0.007	− 0.001 ± 0.031		< 1 S.D.
	C	− 0.002	− 0.008 ± 0.046		< 1 S.D.
	C	0.027	0.024 ± 0.023		< 1 S.D.
	C	− 0.027	− 0.021 ± 0.029		< 1 S D.
	C	− 0.005	− 0.001 ± 0.032		< 1 S.D.
	C	0.008	− 0.002 ± 0.024		< 1 S.D.
	C	0.008	0.005 ± 0.045		< 1 S.D.
	C	− 0.003	− 0.003 ± 0.041		< 1 S.D.
Rf. g	N	− 0.159	− 0.188 ± 0.034	26	< 1 S.D.
	C	0.104	0.152 ± 0.063		< 1 S.D.
	C	− 0.030	− 0.056 ± 0.024		< 2 S.D.
	C	0.031	0.040 ± 0.035		< 1 S.D.
	C	0.001	− 0.006 ± 0.017		< 1 S.D.
	C	0.109	0.193 ± 0.025		< 1 S.D.
	C	− 0.023	0.053 ± 0.090		< 1 S.D.
	C	0 015	− 0.019 ± 0.044		< 1 S.D.
	C	0 001	0.011 ± 0.051		< 1 S.D.
	C	0.004	0.003 ± 0.060		< 1 S D.
Rf.h	N	− 0.128		< 10	
	C	0.071			
	C	0.065			
	N	− 0.127			
	C	0.082			
	C	0.083			
	C	− 0.015			
	C	0.008			
	C	0.009			
	C	− 0.014			

ᵃ The fragments are drawn in Fig. 8 of Chau and Dean [1]. Charges are calculated for the isolated fragment (Calc.) and charges are obtained from a statistical sample of molecules containing that fragment (Obs. ± S.D., n = number of observations). The atoms are numbered in sequence in the order of the bonds.

to be transferable from molecule to molecule. This database could form a reliable tool for attacking the problem of site-directed structure assembly.

ACKNOWLEDGEMENTS

REFERENCES

1 Chau, P.-L. and Dean, P.M., J. Comput.-Aided Mol. Design, 6 (1992) 385.
2 Chau, P.-L. and Dean, P.M., J. Comput.-Aided Mol. Design, 6 (1992) 397.
3 Abraham, R.J., Grant, G.H., Haworth, I.S. and Smith, P.E., J. Comput.-Aided Mol. Design, 5 (1991) 21.
4 Momany, F.A., McGuire, R.F., Burgess, A.W. and Scheraga, H.A., J. Phys. Chem., 79 (1975) 2361.
5 Giessner-Prettre, C. and Pullman, A., Theor. Chim. Acta, 25 (1972) 83.
6 Dobosh, P.A., Q.C.P.E., 9 (1968) program number 141.