Journal of Computer-Aided Molecular Design, 6 (1992) 461–474
ESCOM

J-CAMD 168

Ligand atom partial charges assignment for complementary electrostatic potentials

S.L. Chan[a], P.-L. Chau[b,*] and J.M. Goodman[c]

[a]*Department of Molecular and Medical Genetics, University of Toronto, Canada M5S 1A8*
[b]*School of Clinical Medicine, University of Cambridge, Cambridge CB2 2SP, U.K.*
[c]*Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, U.K.*

Received 3 February 1992
Accepted 5 June 1992

Key words: Drug design; Molecular electrostatic potential

SUMMARY

The design of molecules to fit into the active site of receptors is a rapidly developing area of pharmacology and medicinal chemistry. A good ligand needs a suitable geometry and also appropriate electrostatic properties. The electrostatic properties of the ligand should complement those of the receptor. We present a method for the assignment of atom-centred point charges for a ligand, based on the electrostatic potential of the receptor. These point charges are chosen to give the best possible complementarity to the receptor electrostatic potential over the van der Waals surface of the ligand. We demonstrate that point charges can be chosen to give good electrostatic complementarity, and suggest that a molecule with similar electrostatic properties should bind well to the receptor.

INTRODUCTION

The design of ligands to fit in a receptor site is an important problem that is being approached from several different angles. The geometry of a good ligand may be determined from a database of shapes [1–3] or found by a 'building up' process [4–6]. In this study, we assume that a reasonable geometry has already been found using these methods. The electrostatic properties of a good ligand must be carefully chosen to give a ligand that binds well in a receptor site. This paper presents a general method of assigning partial charges to a known chemical skeleton, so that the resulting electrostatic potential generated is optimal for ligand binding.

METHOD

Our program calculates atom-centred point charges for a ligand, starting with the position of

* To whom correspondence should be addressed at Department of Pharmacology, University of Cambridge, Cambridge CB2 1QJ, U.K.

the ligand atoms relative to the binding site and the electrostatic potential due to the receptor at the ligand surface. There are methods in the literature for finding suitable positions for the ligand atoms [4–6]. The potential due to the receptor at the ligand surface can be calculated in various ways. In this study we assigned partial charges to all the receptor atoms, using ECEPP charges [7,8] which are based on CNDO/2 calculations [9,10]. These partial charges were used to calculate the electrostatic potential at the ligand surface. Other schemes for calculating the electrostatic potential are available; these may also be used to provide data for our program.

It has been postulated that the electrostatic potential due to the receptor will be complementary to the electrostatic potential due to the ligand, and so there will be a region of low electrostatic potential between the ligand and the receptor [11–13]. Other factors may also be important. To a first approximation, the potential due to the ligand may be calculated from the partial charges assigned at the centres of the ligand atoms. What are the best possible partial charges to place on the ligand to do this? We can calculate these partial charges using a least-square fit to the polarity-reverse of the receptor electrostatic potential on the van der Waals surface of the ligand. This method, which we call the image charge method, was inspired by CHELP [14], a program for finding partial charges from ab initio calculated electrostatic potentials, which was itself based on earlier programs [15–17]. The program we have written is called YING.

The choice of the surface over which the program tries to achieve electrostatic complementarity is of great importance. If the ligand fits loosely into the binding site of the receptor, then there is a wide range of surfaces which could be considered, ranging from the van der Waals surface of the receptor to that of the ligand. Usually, however, ligands will fit tightly into the binding site, and this was true of all the cases we considered. The van der Waals surfaces of the ligands and the receptors were very close to each other, and so the choice of surface was very restricted. For convenience, we chose to use the van der Waals surface of the ligand.

The choice of the points on this surface at which to compare the electrostatic potentials is also important. We chose a series of points evenly spread over the ligand surface using an icosahedral tessellation method [18]. This method produces points more evenly distributed than those from a longitude-latitude method. Two hundred and fifty-two points were chosen on the van der Waals sphere of each atom, and then all points within the van der Waals surface of the molecule were removed. In all of our examples, some of the points lay within the receptor van der Waals surface. These points were also deleted.

Not all the points on the van der Waals surface were used because some of the points were not facing any of the protein atoms. If we imagine placing a light at each receptor atom, areas of the ligand's van der Waals surface may not be illuminated because there are no receptor atoms in their lines of sight. All points that lay within these dark areas were deleted. This was only a small percentage in the examples which we considered.

Occasionally, there is a gap between the ligand and the receptor. If a water molecule can fit into this gap, then the electrostatic effect of the receptor on the ligand will be greatly attenuated at this position. A water molecule can be treated as a sphere of radius 1.4 Å [19]. All points on the ligand surface greater than 2.8 Å from any point on the receptor surface were discarded. The number of points removed for these three reasons amounts to about 10% of the total number on the van der Waals surface of the ligands, except in the case of 2′-GMP and ATP, where respectively 15% and 25% were removed.

The electrostatic potential due to the receptor at these points was calculated, assuming that the

charge distribution on the receptor molecule can be represented by partial charges at the receptor atomic centres. We assume that the ligand will fit best when the electrostatic potential due to the ligand complements that due to the protein, over the van der Waals surface of the ligand. We express this complementarity mathematically by adding the potential P due to the protein and the potential S due to the ligand at each point i, squaring the total at each point, and then summing the squared quantities from all these points. In equation form, the complementarity measure F is given by:

$$F = \sum_{i=1}^{n_p} (P_i + S_i)^2$$

where n_p is the number of points. This number is a measure of the goodness of fit of the ligand in the protein, and we want to choose partial charges on the ligand to make it as small as possible. It is possible to solve this problem analytically, and the details of the equations are given in the Appendix.

RESULTS

The crystallographic datasets used were dihydrofolate reductase with ligand NADPH (Brookhaven Protein Databank code 3DFR), *p*-hydroxybenzoate hydroxylase with FAD and 3,4-dihydroxybenzoate (1PHH), catabolite gene activator protein with two cAMP molecules (3GAP), phosphoglycerate kinase with ATP (3PGK) and ribonuclease T_1 and 2′-GMP (1RNT). All these ligands have a nonzero formal charge. It would have been interesting to test a nonpolar ligand as well, but no suitable receptors were available in the Brookhaven Database. The following calculations were performed on each dataset:

- We took the atomic positions of the ligands and calculated the valence shell electronic density of each atom using CNDO/2 [9,10]. From this, we can evaluate the partial charge for each atom using a Mulliken analysis. We call these the *calculated partial atomic charges.*
- The van der Waals surface of each ligand was constructed according to the method described above.
- The electrostatic potential due to the calculated partial atomic charge on the ligand van der Waals surface was calculated using the formula:

$$E = \sum_{i=1}^{n} \frac{q_1 q_i}{4\pi\varepsilon_o r_i}$$

  where E is the potential energy, q_1 is the charge of the probe (set to be a proton charge in this case), q_i the partial charge on the i-th atom, ε_o the permittivity of vacuum and r_i the distance from the i-th atom to the point where the potential is summed over the n atoms of the ligand. We call this electrostatic potential the *calculated ligand potential.*
- The partial atomic charges for the protein atoms were assigned from ECEPP [7,8]. The electrostatic potential due to the assigned protein atomic charges on the ligand's van der Waals surface was calculated from these partial charges. This potential is called the *calculated protein potential.*

Program testing

After setting up the datasets and various potentials, we tested the program by reversing the polarity of the calculated ligand potential. Using this potential as the input, the YING program was asked to find a set of charges located at the positions of the ligand atoms which would give the best fit to complement this potential. No constraints on the charges were imposed. The program should calculate the partial atomic charges of the ligand atoms which were used to generate the potential, and it did this successfully.

Unconstrained calculations

The protein potential is the electrostatic potential on the *ligand* surface generated by the protein atoms. The program was given the protein potential and the atomic positions of the ligand. The partial atomic charges at each ligand atom position which would give the electrostatic potential most complementary to the protein potential were calculated. These partial atomic charges were not constrained.

To test if the charges assigned to the ligand atom positions would generate a complementary electrostatic potential on the ligand's van der Waals surface, we evaluated the electrostatic potential generated by the assigned charges (called the *ying* charges) using the method described above. This electrostatic potential is called the *ying* potential. The correlation between this potential and the protein potential was calculated. If the assigned charges generated a perfectly complementary potential, we would expect both the correlation and rank correlation coefficient to be -1. A linear regression would show a slope of -1, with an intercept of 0. The quantitative statistical expressions of complementarity were explained in Ref. 13. The results from these tests are shown in Table 1.

We can see that the *ying* potentials are highly complementary to the protein potentials. The scattergram for the 2'-GMP site and the *ying* ligand is shown in Fig. 1. The R_{rank} ranges from -0.947 to -0.991, while the range of r is from -0.918 to -0.990. The slope of regression varies from -0.870 to -0.983. All of these values are very near to -1, and so the complementarity is nearly perfect. The regression constant c_r ranges from -0.124 kJ/mol to -39.5 kJ/mol, but this

TABLE 1
COMPLEMENTARITY BETWEEN THE CALCULATED PROTEIN POTENTIAL AND THE *YING* POTENTIAL, WHERE THE LATTER IS EVALUATED WITH NO CONSTRAINTS

L/R	R_{rank}	r	$m_r \pm$S.E.	$c_r \pm$S.E.	e_c (%)
cAMP(a) 3GAP	−0.957	−0.959	−0.921 ± 0.006	−13.2 ± 1.1	2.86
cAMP(b) 3GAP	−0.959	−0.959	−0.920 ± 0.006	−15.2 ± 1.3	3.38
NADPH 3DFR	−0.991	−0.990	−0.983 ± 0.002	− 8.45 ± 1.0	0.90
3,4-DHB 1PHH	−0.947	−0.918	−0.870 ± 0.001	−39.5 ± 3.5	9.58
FAD 1PHH	−0.968	−0.966	−0.936 ± 0.003	−21.7 ± 1.2	4.19
ATP 3PGK	−0.976	−0.973	−0.956 ± 0.004	−22.2 ± 2.3	3.16
2'-GMP 1RNT	−0.977	−0.978	−0.956 ± 0.004	− 0.124 ± 0.64	0.02

cAMP(a) and cAMP(b) are the two distinct cAMP molecules that bind to the two receptor sites on the catabolite gene activator protein, 3GAP. L/R = ligand/receptor complex using the Brookhaven Protein Databank code; R_{rank} = Spearman's rank correlation coefficient; r = correlation coefficient; m_r = slope of regression; c_r = regression constant (in kJ/mol); S.E. = standard error; e_c = c_r as fraction of total range of *ying* potential.

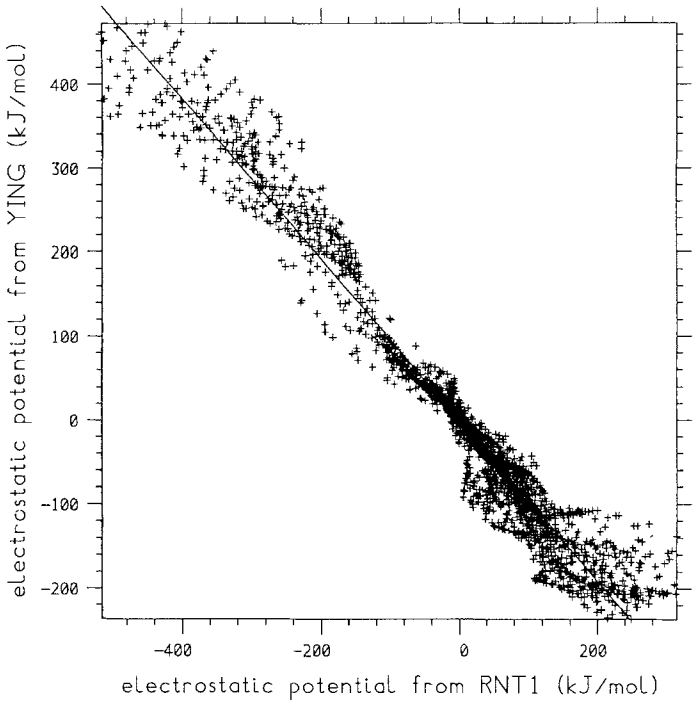

Fig. 1. Comparison of the electrostatic potential from *ying* and from the protein, at 2449 points on the van der Waals surface of 2'-GMP.

must be considered with the range of electrostatic potential. We define the fractional regression constant, e_c, as the regression constant divided by the total range of values on the y-axis (i.e., the *ying* ligand potential). This is never bigger than 10%. All this shows that the *ying* potential we generated from the charges is highly complementary to the protein potential.

However, since we constrained neither the total formal charge on the molecule, nor the partial atomic charge on each individual atom, the partial charges sometimes attained values which do not appear to correspond to physical reality, as shown in Table 2. For example, the *ying* charges for the NADPH skeleton have a total formal charge of −2.27, with the highest partial charge as-

TABLE 2
TOTAL FORMAL CHARGE OF THE *YING* COMPOUND, TOGETHER WITH THE LOWEST AND HIGHEST PARTIAL CHARGES OF THE ATOMS, ALL IN ATOMIC UNITS

Ligand	Total formal charge	Lowest partial charge	Highest partial charge
cAMP(a)	−0.536	−0.446	0.471
cAMP(b)	−0.587	−0.463	0.331
NADPH	−2.271	−0.564	1.073
3,4-DHB	−0.755	−0.382	0.516
FAD	−1.628	−0.478	0.742
2'-GMP	−0.004	−0.479	0.475
ATP	−1.710	−0.586	0.760

signed to an atom as +1.073. This is much larger than is possible using the CNDO/2 method of calculating charge. A survey of commonly-occurring organic molecules shows that the atomic CNDO/2 charges almost never exceed the range of ±0.6 electronic charge [20]. In the next section, we describe calculations where the magnitudes of the charges were constrained.

Constrained calculations

In the previous section, we see that unconstrained calculations produce results which do not appear to correspond to physical reality. In order to remedy this, the calculations were repeated with each partial atomic charge constrained within the range of ±0.6 electronic charge. The value 0.6 was chosen because it was the largest value commonly found in a survey of organic molecules, using the CNDO/2 method of calculating partial charges [20]. The total charge on the ligand was not constrained. The introduction of these constraints meant that it was no longer possible to find analytical solutions to the equations for finding partial charges. Solutions could still be found using numerical methods, which are described in the Appendix. The results of these calculations are given in Table 3. This constraint affected only 3 of the 7 cases.

These results differed from those previously obtained with no charge constraints in only 3 compounds: ATP, FAD and NADPH. The correlations for these pairs of ligand and receptor were slightly reduced, as expected, but the correlations were still good. There was no change in the results for the other compounds since the partial atomic charges were already within the ±0.6 range.

However, the total formal charge on the molecule attains noninteger values as follows: cAMP(a) −0.536, cAMP(b) −0.587, NADPH −2.271, 3,4-DHB −0.755, FAD −1.628, ATP −1.706, 2′-GMP −0.004. This means that the partial atomic charges assigned to each atom produce a nonphysical solution. To overcome this problem, further calculations were performed to investigate the best-fit charges that have integral sums. The atomic partial charges were limited to ±0.6, while the total formal charges on each molecule were set to integers. In all the compounds chosen, the ligand was totally enclosed by the protein, and so the total formal charge on the ligand was likely to be integral. The total charges chosen were the integers immediately larger or smaller than the total formal charge generated in the previous calculation. The results are shown in Table 4.

As expected, an additional constraint worsens the complementarity statistics. However, all but one of the charge assignments have r and R_{rank} better than −0.7. The choice of the total formal charge is also of importance. For example, in the case of 2′-GMP, for a total formal charge of 0,

TABLE 3
COMPLEMENTARITY BETWEEN THE CALCULATED PROTEIN POTENTIAL AND THE *YING* POTENTIAL WITH CONSTRAINTS ON THE ATOMIC PARTIAL CHARGES LIMITED TO ±0.6 ATOMIC UNITS[a]

L/R	R_{rank}	r	$m_r \pm$S.E.	$c_r \pm$S.E.	e_c(%)
NADPH 3DFR	−0.991	−0.990	−0.982±0.002	− 8.81±1.1	0.94
FAD 1PHH	−0.968	−0.966	−0.936±0.003	−21.8 ±1.1	4.21
ATP 3PGK	−0.977	−0.972	−0.955±0.004	−22.8 ±2.3	3.25

[a] Only those entries different from the previous table are shown. For abbreviations, see Table 1.

TABLE 4
COMPLEMENTARITY BETWEEN THE CALCULATED PROTEIN POTENTIAL AND THE *YING* POTENTIAL WITH CONSTRAINTS ON BOTH THE TOTAL FORMAL CHARGE AND THE ATOMIC PARTIAL CHARGES

L/R	q	R_{rank}	r	$m_r \pm$S.E.	$c_r \pm$S.E.	e_c (%)
cAMP(a) 3GAP	0	−0.805	−0.817	−0.928 ± 0.014	−126 ± 2.7	19.1
	−1	−0.827	−0.843	−0.913 ± 0.012	−135 ± 2.4	24.8
cAMP(b) 3GAP	0	−0.546	−0.563	−0.703 ± 0.022	77 ± 4.3	11.2
	−1	−0.865	−0.860	−0.873 ± 0.011	−128 ± 2.4	26.7
NADPH 3DFR	−2	−0.985	−0.984	−1.01 ± 0.003	45 ± 1.3	4.46
	−3	−0.945	−0.943	−0.906 ± 0.004	−157 ± 2.4	17.5
3,4-DHB 1PHH	0	−0.791	−0.666	−0.730 ± 0.023	208 ± 7.6	48.2
	−1	−0.865	−0.882	−0.912 ± 0.014	−121 ± 4.5	28.3
FAD 1PHH	−1	−0.885	−0.903	−1.11 ± 0.007	129 ± 2.5	18.9
	−2	−0.921	−0.923	−0.828 ± 0.005	−111 ± 1.6	22.0
ATP 3PGK	−1	−0.927	−0.894	−1.18 ± 0.011	181 ± 6.0	13.9
	−2	−0.928	−0.939	−0.87 ± 0.006	−98 ± 3.3	14.8
2′-GMP 1RNT	0	−0.977	−0.978	−0.956 ± 0.004	−0.859 ± 0.64	0.12
	−1	−0.726	−0.727	−0.951 ± 0.018	−224 ± 2.8	24.0

q = total formal charge in atomic units; for other abbreviations, see Table 1.

r = −0.977, while R_{rank} = −0.978 and e_c = 0.12%. If the total formal charge is changed to −1, r = −0.726, R_{rank} = −0.727 and e_c = 24.0%. This is a very significant change in complementarity, and it appears that the choice of total formal charge is of importance to the results of the optimization. Nevertheless, if we focus our attention on the total formal charge assignment where the complementarity statistics are better, it can be seen that neither r nor R_{rank} is ever bigger than −0.8, while e_c is generally below 25%. This shows that the YING program can give us a charge assignment with a highly complementary electrostatic potential, without exceeding the usual limits for CNDO/2-derived charges.

It is interesting to note that some of these 'molecules' are actually more complementary to the receptor site than the natural compounds. The comparison is shown in Table 5. The *ying* ligand chosen has the higher complementarity.

The fact that some *ying* ligands are more complementary than naturally occurring ones does not necessarily mean that the natural ligands bind to the receptor site less well than the *ying* ligands. The main reason for this is probably due to solvent effects, which have not been fully considered. Many of the protein receptors bind more than one ligand simultaneously, so a ligand may bind the receptor site best with a co-ligand. The enzyme phosphoglycerate kinase 3PGK binds 3 ligands simultaneously: Mg^{2+}, 3-phosphoglycerate and ATP [21]. This may explain why the slope of regression between the natural ligand ATP and the enzyme is about −2, but that the total charge assigned by our method is less than the −4 formal charge on ATP. The 'missing' charges could be accounted for by the Mg^{2+} ion.

TABLE 5
COMPARISON OF THE COMPLEMENTARITY BETWEEN THE CALCULATED PROTEIN POTENTIAL AND THE *YING* POTENTIAL, AND THAT BETWEEN THE CALCULATED PROTEIN POTENTIAL AND THE CALCULATED LIGAND POTENTIAL

L/R	t	R_{rank}	r	$m_r \pm$S.E.	$c_r \pm$S.E.	e_c(%)
cAMP(a) 3GAP	y	−0.827	−0.843	−0.913±0.012	−135 ± 2.4	24.8
	n	−0.702	−0.700	−0.968±0.021	−143 ± 4.1	23.4
cAMP(b) 3GAP	y	−0.865	−0.860	−0.873±0.011	−128 ± 2.4	26.7
	n	−0.677	−0.678	−0.881±0.020	−148 ± 4.4	25.8
NADPH 3DFR	y	−0.985	−0.984	−1.01 ±0.003	45 ± 1.3	4.46
	n	−0.863	−0.825	−1.18 ±0.011	269 ± 6.0	21.0
3,4-DHB 1PHH	y	−0.865	−0.882	−0.912±0.014	−121 ± 4.5	28.3
	n	−0.768	−0.825	−1.27 ±0.025	−8.05 ± 8.0	1.40
FAD 1PHH	y	−0.921	−0.923	−0.828±0.005	−111 ± 1.6	22.0
	n	−0.491	−0.607	−1.31 ±0.023	14.5 ± 8.0	1.62
ATP 3PGK	y	−0.928	−0.939	−0.87 ±0.006	−98 ± 3.3	14.8
	n	−0.641	−0.586	−1.91 ±0.005	−110 ±27	6.99
2′-GMP 1RNT	y	−0.977	−0.978	−0.956±0.004	−0.859± 0.64	0.12
	n	−0.729	−0.622	−1.10 ±0.028	−622 ± 4.4	59.3

t = ligand type; y = *ying* ligand; n = natural ligand; for other abbreviations, see Table 1.

Figure 2 shows the chemical structure of one of the ligands studied, 2′-GMP. Table 6 shows a comparison of the 'natural' partial charge and the *ying*-assigned partial charge of each atomic position. A similar structure and comparison is shown for NADPH in Fig. 3 and Table 7. It can be seen that the assigned partial charges are rather different from the 'natural' partial charges.

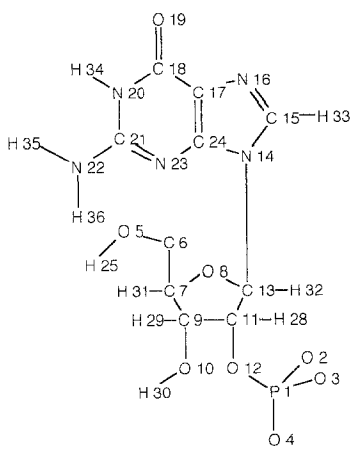

Fig. 2. Atom numbering for 2′-GMP.

TABLE 6
COMPARISON OF THE PARTIAL CHARGES ON 2'-GMP

Atom	q_n	q_y	Atom	q_n	q_y
P1	0.161	0.197	O19	−0.444	−0.094
O2	−0.597	−0.212	N20	−0.209	−0.088
O3	−0.578	−0.053	C21	0.360	−0.346
O4	−0.541	−0.081	N22	−0.240	0.226
O5	−0.265	−0.013	N23	−0.278	−0.040
C6	0.112	−0.060	C24	0.233	0.480
C7	0.119	−0.252	H25	0.145	0.005
O8	−0.295	0.095	H26	−0.016	0.044
C9	0.122	0.328	H27	−0.009	0.017
O10	−0.273	−0.070	H28	−0.049	0.065
C11	0.153	−0.099	H29	−0.019	−0.074
O12	−0.289	−0.165	H30	0.116	−0.016
C13	0.232	0.221	H31	−0.018	0.023
N14	−0.076	−0.482	H32	−0.048	−0.074
C15	0.170	0.207	H33	−0.023	−0.109
N16	−0.217	−0.110	H34	0.102	0.435
C17	−0.097	−0.192	H35	0.109	0.055
C18	0.352	0.055	H36	0.095	0.178

q_n = partial charge on natural ligand in atomic units; q_y = partial charge on *ying* ligand in atomic units.

DISCUSSION

The problem of site-directed ligand design can be broken down into several parts, one of which is to generate a ligand with an electrostatic potential complementary to that of the receptor. Electrostatic potential complementarity has been demonstrated in many ligand–receptor pairs [11–13]. Nevertheless, no objective method has yet been devised to assign partial charges to ligand atoms, so that the ligand will have a complementary electrostatic potential. We present a method for the calculation of a set of ligand partial charges, given the ligand atom positions and the receptor electrostatic potential.

The partial charges found by this method were rather large, and it might not be possible to synthesize molecules which correspond precisely to these 'ideal' ligands. Receptors might be able to bind to ligands even better if the ligands were not restricted to a quite narrow range of charges. We are interested in finding charge distributions that might correspond to real molecules, so we need a way of limiting the charges on the ligand.

The *ying* ligand charges are also different from the naturally occurring ligand charges. This is partly because potential-derived charges tend to have a different distribution from population analysis-derived charges. This also suggests that other factors may be important in determining the ligand–receptor interaction, e.g., hydrophobicity. The best possible ligand probably has a charge distribution which is a compromise between these factors. It also suggests that different atom types could be used to produce novel ligands which will bind to the site.

The image charge method has a number of potential uses. The charge distributions it produces

TABLE 7
COMPARISON OF THE PARTIAL CHARGES ON NADPH

Atom	q_n	q_y	Atom	q_n	q_y
N1	−0.302	−0.187	O38	−0.272	0.012
C2	0.173	0.108	O39	−0.277	0.020
N3	−0.186	−0.244	N40	−0.156	0.012
C4	0.202	0.232	C41	0.123	0.367
C5	−0.115	−0.398	C42	−0.097	−0.600
C6	0.243	−0.004	C43	0.092	0.274
N7	−0.275	0.107	C44	−0.081	0.012
C8	0.217	0.031	C45	0.086	−0.240
N9	−0.078	0.031	C46	0.315	0.204
N10	−0.231	0.044	O47	−0.376	0.083
C11	0.193	−0.315	N48	−0.290	−0.270
C12	0.129	0.081	H49	−0.066	0.012
C13	0.071	0.600	H50	−0.016	−0.086
C14	0.151	0.137	H51	0.082	−0.038
O15	−0.265	−0.065	H52	0.113	0.019
O16	−0.315	−0.080	H53	−0.021	0.178
O17	−0.256	−0.091	H54	−0.048	−0.157
P18	0.188	0.206	H55	−0.042	−0.265
O19	−0.564	−0.243	H56	0.019	−0.181
O20	−0.575	−0.159	H57	0.125	−0.339
O21	−0.578	−0.340	H58	−0.030	−0.217
C22	0.104	0.305	H59	−0.040	−0.089
O23	−0.262	−0.267	H60	−0.015	−0.105
P24	0.444	0.129	H61	−0.037	−0.039
O25	−0.724	−0.127	H62	−0.040	0.065
O26	−0.414	−0.090	H63	0.000	0.075
O27	0.002	−0.119	H64	−0.014	−0.039
P28	0.397	0.124	H65	0.014	0.086
O29	−0.679	−0.079	H66	0.114	−0.016
O30	−0.395	−0.229	H67	0.125	−0.113
O31	−0.313	−0.171	H68	−0.002	−0.031
C32	0.146	0.277	H69	−0.021	0.038
C33	0.226	−0.238	H70	−0.026	0.036
C34	0.120	0.064	H71	0.009	0.062
C35	0.129	−0.066	H72	0.039	0.071
C36	0.109	−0.122	H73	0.130	0.274
O37	−0.252	0.007	H74	0.116	0.076

For abbreviations, see Table 6.

could be used to guide organic chemists in the design and synthesis of receptor molecules. The correspondence between the charge distribution of these molecules and the ideal YING charges will, necessarily, be approximate, but these molecules may then be evaluated as ligands using the same YING method. We could also use the method to evaluate partial charges for ligands for which the binding affinity is known, and qualitatively compare the experimental results with the calculated binding affinity, as judged by the root-mean-square deviation for the electrostatic po-

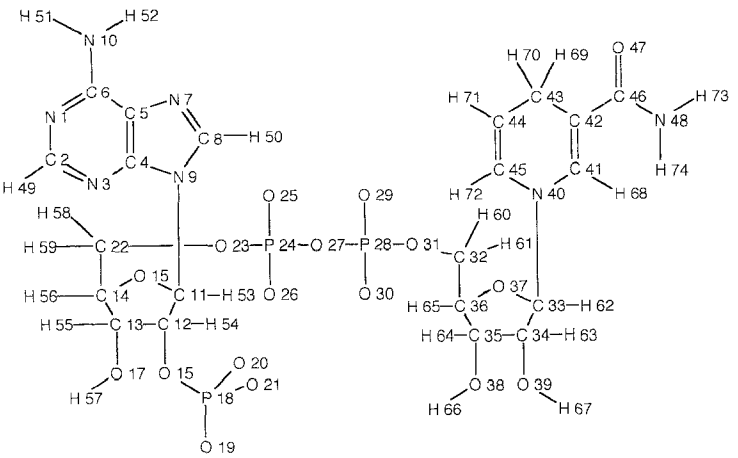

Fig. 3. Atom numbering for NADPH.

tential over the surface. Eventually, the image charge method might be used to evaluate partial charges for ligands and predict their binding affinity, thus using the method to assist in the rational design of new drugs and biologically active compounds.

Although we have managed to create complementary electrostatic potentials on the ligand's van der Waals surface, it should be borne in mind that this surface depends on the atoms of the ligand. Some atoms have larger van der Waals radii, e.g., phosphorus, while some atoms have smaller radii, e.g., carbon. When we carried out the YING calculations, we used the van der Waals radii of the atoms that occurred in the naturally occurring ligand. When we assign atoms to fit the *ying* charges, the atoms assigned may not correspond to the natural ligand atoms, so the van der Waals surface changes slightly. This would introduce minor inaccuracies into our work. However, the effect should only be small because the variation in atomic van der Waals radii, for nonhydrogen atoms, is only from 1.5 Å (carbon), to 1.8 Å (phosphorus), a variation of 20%. The surface could also be refined when the atom types are fixed.

The electrostatic potential from the protein will not have been calculated very accurately. This is because we use Mulliken point charges to calculate the electrostatic potential around the receptor, but Mulliken charges are not ideal for this purpose [16], because they were designed as a simple way of summing electron density over space. The difficulty of calculating the electrostatic potential of a protein may be overcome by using more accurate methods. The essential parts of the image charge method could be used without changes.

CONCLUSION

In our view, the problem of finding the best ligand to fit a receptor site can be broken down into four parts:

(1) Calculating a geometry for the ligand skeleton.

(2) Assigning an optimal charge to each of the atoms on the skeleton.

(3) Refining the ideal solution to allow for hydrogen bonding.

(4) Assigning chemical fragments to the chemical skeleton and charge distribution scheme to make the best synthesizable molecule(s).

The first step has already been investigated [1–6]. In this paper, we have addressed the second step of this approach. Based on a given geometry of the ligand, we have described an algorithm to determine optimal atom-centred partial charges for a ligand, using a least-squares fit to the electrostatic potential of the receptor. The optimal charges were chosen such that the electrostatic potential of the ligand and that of the receptor on the interfacial ligand surface were most complementary [11].

The program produces sets of charges for ligands which have electrostatic potentials highly complementary to the electrostatic potentials of the receptors. Molecules rarely have partial charges of more than ± 0.6 on any of their atoms, when the charges are assessed using CNDO/2. The program limits partial charges to less than ± 0.6. The total charge on the ligands was also limited to integral multiples of the electronic charge. These constraints slightly reduced the correlation between the ligand and the receptor electrostatic potentials, but the correlation remained encouragingly high. In the cases investigated where a natural ligand exists for the receptor, the electrostatic complementarity of our 'ideal' ligand compares favourably with that of the natural ligand. The YING method we have developed is a powerful procedure in drug design for assigning optimal charges to vertices of a known molecular skeleton.

ACKNOWLEDGEMENTS

The authors wish to thank Kai-Meng Hock for the original idea of the image charge, and Philip Dean for helpful discussions. S.L.C. and P.L.C. thank the Croucher Foundation for generous support for part of this work. J.M.G. thanks Clare College, Cambridge for a Research Fellowship.

REFERENCES

1 Kuntz, I.D., Blaney, J.M., Oatley, S.J., Langridge, R. and Ferrin, T.E., J. Mol. Biol., 161 (1982) 269.
2 DesJarlais, R.L., Sheridan, R.P., Dixon, J.S., Kuntz, I.D. and Venkataraghavan, R., J. Med. Chem., 29 (1986) 2149.
3 DesJarlais, R.L., Sheridan, R.P., Seibel, G.L., Dixon, J.S., Kuntz, I.D. and Venkataraghavan, R., J. Med. Chem., 31 (1988) 722.
4 Lewis, R.A. and Dean, P.M., Proc. R. Soc. Lond., B236 (1989) 125.
5 Lewis, R.A. and Dean, P.M., Proc. R. Soc. Lond., B236 (1989) 141.
6 Lewis, R.A., J. Comput.-Aided Mol. Design, 4 (1990) 205.
7 Browman, M.J., Carruthers, L.M., Kashuba, K.L., Momany, F.A., Pottle, M.S., Rosen, S.P. and Rumsey, S.M., QCPE, 11 (1975) programme 286.
8 Némethy, G., Pottie, M.S. and Scheraga, H.A., J. Phys. Chem., 87 (1983) 1883.
9 Pople, J.A., Santry, D.P. and Segal, G.A., J. Chem. Phys., 43 (1965) S129.
10 Segal, G.A. and Pople, J.A., J. Chem. Phys., 43 (1965) S136.
11 Nakamura, H., Komatsu, K., Nakagawa, S. and Umeyama, H., J. Mol. Graphics, 3 (1985) 2.
12 Náray-Szabó, G., J. Mol. Graphics, 7 (1989), 76.
13 Dean, P.M., Chau, P.-L. and Barakat, M.T., J. Mol. Struct., 256 (1992) 75.
14 Chirlian, L.E. and Francl, M.M., J. Comput. Chem., 8 (1987) 894.
15 Momany, F.A., J. Phys. Chem., 82 (1978) 592.
16 Cox, S.R. and Williams, D.E., J. Comput. Chem., 2 (1981) 304.
17 Singh, U.C. and Kollman, P.A., J. Comput. Chem., 5 (1984) 129.
18 Chau, P.-L. and Dean, P.M., J. Mol. Graphics, 5 (1987) 97.
19 Lee, B. and Richards, F.M., J. Mol. Biol., 55 (1971) 379.

20 Chau, P.L. and Dean, P.M., J. Comput.-Aided Mol. Design, 6 (1992) 407.
21 Watson, H.C., Walker, N.P.C., Shaw, P.J., Bryant, T.N., Wendell, P.L., Fothergill, L.A., Perkins, R.E., Conroy, S.C., Dobson, M.J., Tuite, M.F., Kingsman, A.J. and Kingsman, S.M., EMBO J., 1 (1982) 1635.

APPENDIX

The electrostatic potential due to the protein at a point on the surface p, S_i, can be expressed as follows:

$$S_i = \sum_{j=1}^{n_q} \frac{1}{4\pi\varepsilon_o} \frac{q_j}{r_{ij}}$$

where n_q is the number of charges to be fitted, ε_o the permittivity of vacuum, q_j the charge on atom j, and r_{ij} the distance from atom j to a point on surface p.

Let P_i be the electrostatic potential due to the receptor at the same point on surface p, and n_p be the number of points on this surface. The objective function to be minimized is F,

$$F = \sum_{i=1}^{n_p} (P_i + S_i)^2$$

Using the method of Lagrangian multipliers, we obtain a function L to be minimized:

$$L = \sum_{i=1}^{n_p} (P_i + S_i)^2 + \lambda \left[\left(\sum_{j=1}^{n_q} q_j \right) - q_m \right]$$

The first term in L is to minimize the sum of potentials. In the last term, q_m is the total formal charge on the molecule, and this term constrains the total formal charge on the whole molecule.

We differentiate L with respect to the charges and λ:

$$\frac{\partial L}{\partial q_k} = \sum_{i=1}^{n_q} \frac{P_i + S_i}{2\pi\varepsilon_o r_{ik}} + \lambda$$

$$\frac{\partial L}{\partial \lambda} = \left(\sum_{j=1}^{n_q} q_j \right) - q_m$$

By setting all these quantities to 0, we can determine the charges that would give us an electrostatic potential most complementary to the potential from the receptor subject to the condition that the sum of the charges equals the preset value q_m.

This method, however, does not allow us to set constraints on individual charges. To do that, we used another method but using the same objective function.

$$F = \sum_{i=1}^{n_p} (P_i + S_i)^2 = \left(\frac{1}{4\pi\varepsilon_o} \right)^2 \|\mathbf{Aq} - \mathbf{h}\|^2$$

where

$$\mathbf{A}_{jp}=\frac{1}{|\mathbf{r}_{jp}|}$$

$$\mathbf{h}_j=-4\pi\varepsilon_o\mathbf{P}_j$$

The NAG FORTRAN routine E04NCF was employed to minimize F. A brief description of the minimization algorithm is given below. For a detailed description, refer to the NAG manual.

Minimization was carried out using an iterative procedure in the n-dimensional space spanned by the q_i's. The conjugate gradient method was employed for the iteration. The first derivatives (vector **g**) and the matrix of the second derivatives (the Hessian matrix **H**) of F were calculated at the present point, and the new iterated point is the point where the first derivative is expected to be zero. Mathematically, the iterative vector **s** is given by

$$\mathbf{Hs}=-\mathbf{g}$$

The constraints on the absolute limit of the charges were imposed explicitly: the point of iteration was confined to the space allowed by the constraints. If the point predicted by the conjugate gradient method lies outside the allowed space, the iterative vector would be scaled down.

When a constraint is imposed on the total charge, the allowed space will be a plane (of dimension $n-1$) in the n-dimensional space defined by

$$\sum q_i=q$$

The iterative step would be the projection of the iterative vector (predicted by the method of conjugate gradient) onto this plane.