Journal of Computer-Aided Molecular Design, 8 (1994) 527–544
ESCOM

J-CAMD 261

Electrostatic complementarity between proteins and ligands. 2. Ligand moieties

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

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

Received 3 March 1994
Accepted 17 May 1994

Key words: Drug design; Molecular electrostatic potential; MEP

SUMMARY

Drug design strategies consider factors governing intermolecular interactions to build up putative ligands. In many strategies, the ligand is constructed using fragments which are placed in the site sequentially. The optimization is then performed with each fragment. We would like to examine if this optimization strategy could generate ligands with optimal electrostatic interactions. The electrostatic complementarities between constituent moieties and the receptor site have been calculated. The whole-ligand complementarity does not appear to be the mathematical mean of the individual complementarities, nor have we found a simple relationship between the moiety and whole-ligand complementarities. The results demonstrate clearly that, using a simple model, it is very difficult to predict the electrostatic potential complementarity of the whole ligand from the complementarities of its constituent chemical moieties. This means that ligand design strategies must optimize the electrostatic complementarity globally, and not moiety by moiety.

INTRODUCTION

In the previous paper [1] we developed a statistical method to analyze electrostatic potential complementarity, and found that it is not substantially changed by including the interfacial region only, nor is the pattern modified by changing the homogeneous dielectric to a distance-dependent one. In this paper, we examine the complementarity of the whole ligand and its constituent moieties.

Drug design strategies take into account the geometry of the molecules, the hydrogen bonds, the electrostatic interactions and the hydrophobic effect. In many drug design strategies, the ligand is built up of fragments which are placed in the site sequentially, and the optimization is performed with each fragment. This method works well if the major contribution to the binding energy is a localized interaction, e.g., hydrogen bonds. However, if a substantial amount of binding energy is provided for by less localized effects, e.g., electrostatic interaction, then this placement method would not generate the optimal ligand for the site. To find such a ligand, it is necessary not only to

*To whom correspondence should be addressed.

perform local optimizations on each fragment, but also a global optimization for the whole ligand.

Electrostatic interactions are long-range interactions. In this work, we examine this 'local design' strategy by calculating the electrostatic complementarity between the moieties and the site. If electrostatic potential complementarity were achieved by piecing together complementary fragments, then one would observe complementarity between the ligand and the site, as well as between the constituent moieties and the site. If complementarity were a global effect, then one would only observe complementarity between the ligand and the site, not between the moieties and the site. We would like to investigate to what extent whole-ligand electrostatic potential complementarity is made up of local constituent moiety electrostatic complementarities.

In this work, we aim to compare the whole-ligand electrostatic complementarity and that between the different parts of the ligand and the receptor site. Specifically, we would like to address this question: Can we predict the whole molecular electrostatic complementarity from knowledge of the complementarity of the constituent moieties of the ligand? If so, then what is the relative contribution of each constituent part of the ligand? If not, then what are the important factors that determine whole-ligand electrostatic potential complementarity?

METHODS

The electrostatic potentials of whole ligands were calculated as described in the previous paper [1]. The ligands were then split into several parts and hydrogen atoms were added so that each part could be considered a molecular moiety, whose partial charges were recalculated. Figure 1 shows how the ligands are divided into numerous moieties. Then, the whole-ligand van der Waals surface was divided into regions corresponding to each part of the molecule. The electrostatic potentials were calculated for each part on each region. This electrostatic potential pattern is compared statistically with that on the same region generated by the whole receptor.

The size of each moiety is such that they are molecular units of the size that a drug designer might use as basic building blocks. Moreover, they must be small enough so that each part is substantially different from the whole ligand and the different chemical groups are separated from each other. Yet, they must be large enough so that delocalized systems are not disrupted either in this splitting process. The typical size is from 10 to 20 atoms in each moiety.

An analysis of the contribution of large structural groups to complementarity has been made. In all, 24 different ligands have been partitioned into chemical moieties. Some of the ligands occur twice in different co-crystals, so the Brookhaven reference codes of the co-crystal are given in brackets after the ligand. The 24 ligands are grouped according to the site they bind to:

(1) HIV-1 protease ligands: SKB-Va (1AAQ) [2], SB204144 (1HOS) [3], MVT-101 (4HVP) [4], L-700417 (4PHV) [5], acetyl-pepstatin (5HVP) [6], JG-365 (7HVP) [7] and U-85548e (8HVP) [8].

(2) AMP (1AK3) [9].

(3) Biliverdin IX-γ (1BBP) [10].

(4) Citrate synthase ligands: carboxymethylcoenzyme A (1CSC and 2CSC) and acetylcoenzyme A (3CSC and 4CSC) [11].

(5) Dihydrofolate reductase sites for folate (1DHF and 7DFR) [12,13] and $NADP^+$ (6DFR and 7DFR) [13]. 5-Deazafolate (2DHF) [12] binds at the folate site, but the ligand is different. The keto form of folate is only very weakly complementary ($r_A = -0.164$, $r_B = -0.160$). All the data refer to the enol form of the ligand.

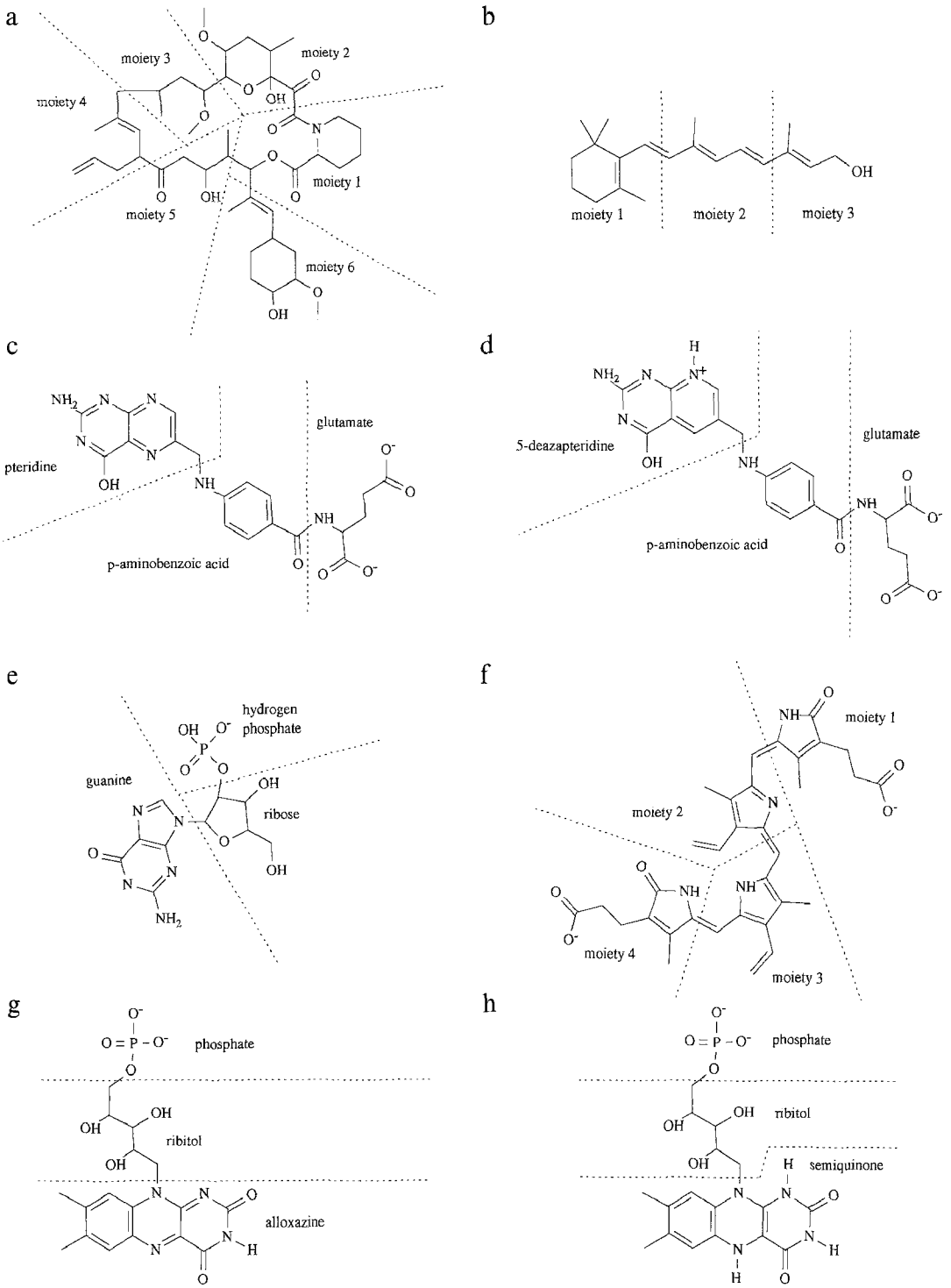

Fig. 1. Chemical structures of some ligands and their division into constituent moieties. (a) FK-506; (b) retinol; (c) folate; (d) 5-deazafolate; (e) GMP; (f) biliverdin; (g) oxidized FMN (alloxazine form); (h) reduced FMN (semiquinone form).

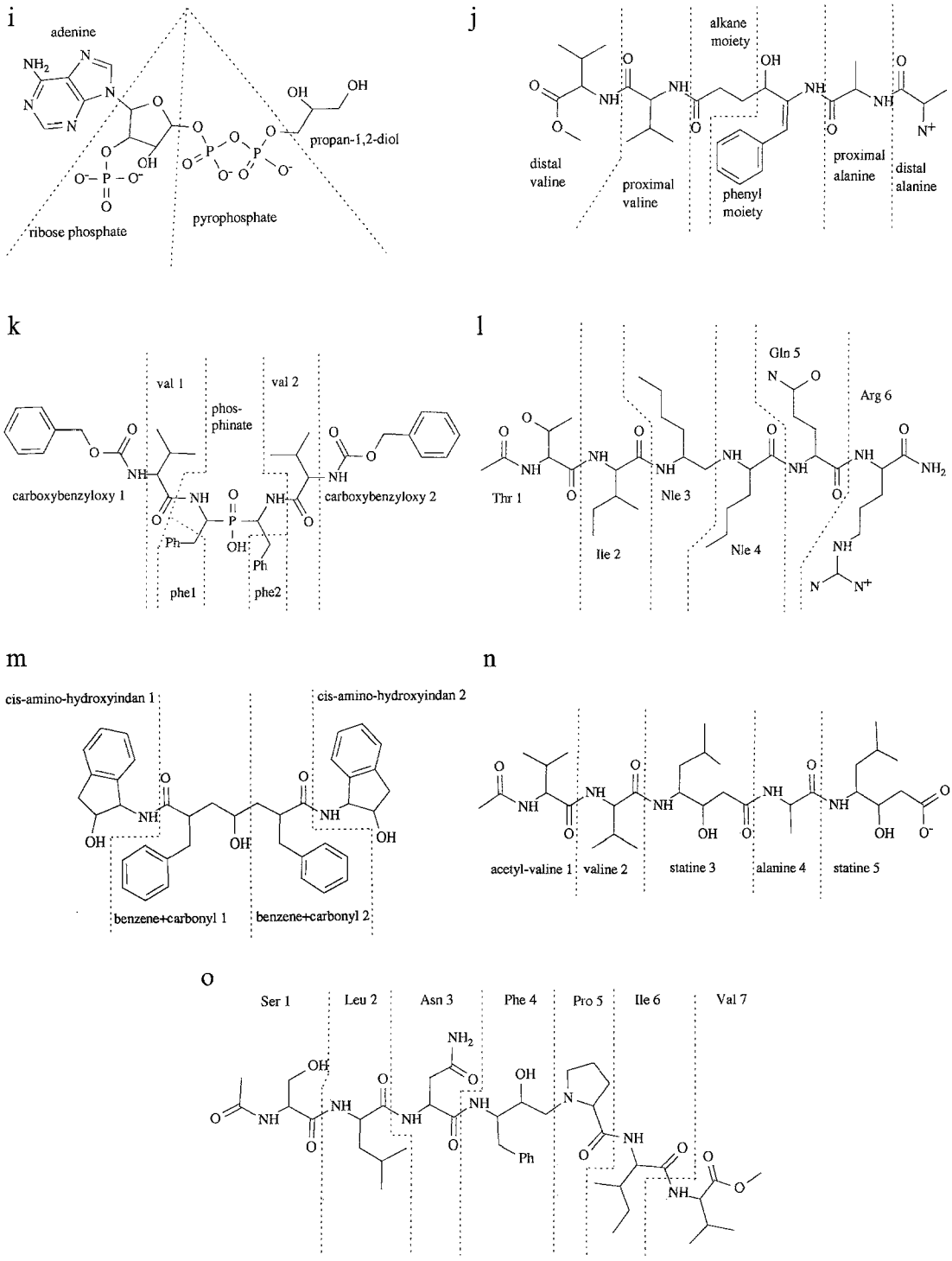

Fig. 1. (continued) (i) incomplete NADP$^+$; (j) SKB-Va; (k) SB204144; (l) MVT-101; (m) L-700417; (n) acetyl-pepstatin; (o) JG-365.

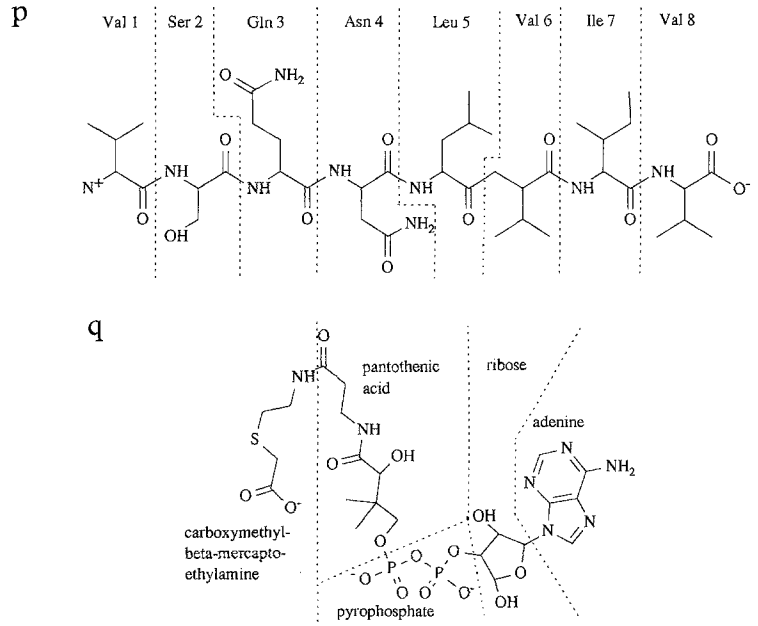

Fig. 1. (continued) (p) U-85548e; and (q) carboxymethylcoenzyme A.

(6) FK-506 (1FKF) [14].
(7) FAD (1FNR) [15].
(8) FAD (1PHH) [16].
(9) NADH (1GD1) [17].
(10) FMN (1GOX), with FMN in the alloxazine form [18].
(11) FMN (3FXN), with FMN in the semiquinone form [19].
(12) NAD^+ (1LDM) [20].
(13) Retinol (1RBP) [21].
(14) 2'-GMP (1RNT) [22].
(15) Cyclic AMP (3GAP) [23].
(16) *Streptomyces griseus* inhibitor site: ace-Pro-Ala-Pro-Phe-OH (4SGA) and ace-Pro-Ala-Pro-Tyr-OH (5SGA) [24].

RESULTS

We divide these ligand–receptor complex pairs into two groups: those where the site does not carry any formal charges (1FKF and 1RBP), and those where the site carries at least one formal charge (the rest). For the second group, we can divide the ligand–receptor complexes into two subgroups: those where the correlation coefficient is reasonable (r more negative than −0.7), and those where the correlation coefficient is poor ($r > -0.7$).

For the case where $r < -0.7$, 49% of the variance of one potential is correlated with that of the other. This is about half, and the criterion $r < -0.7$ is chosen to divide the complexes into two: those where more than half of the variances of the potentials are correlated, and those where they are not.

Complexes with no formal charges at the site

Table 1 shows the two complexes with no formal charge in the receptor site, 1FKF and 1RBP. FK-506 does not exhibit any significant degree of electrostatic potential complementarity to the receptor site on 1FKF; the statistics show a weak complementarity, with a very shallow slope of regression. When the ligand is divided into six moieties (Fig. 1a), only moiety 4 shows a very weak degree of electrostatic potential complementarity to the site, with correlation coefficients better than –0.5, but the slope of regression is still shallow. The statistics for moieties 3 and 5 show that the correlation is not significantly different from zero. Moieties 1 and 6 show no significant degree of complementarity, i.e., low correlation coefficients, while moiety 2 actually exhibits a trend towards electrostatic similarity.

In the case of retinol and 1RBP, the whole ligand shows weak electrostatic potential similarity to the receptor site. Of the three constituent moieties (Fig. 1b), moieties 1 and 2 exhibit an even higher degree of similarity, while moiety 3 does not show any significant degree of complementarity to the site. Significance levels are set at $p<0.0005$ ($1200<n<9000$).

Complexes with at least one formal charge at the site: $r<-0.7$

Tables 2 to 4 show examples from 10 ligand–receptor complexes where the correlation coefficient is better than –0.7.

Folate shows a good degree of electrostatic potential complementarity towards its receptor site on 1DHF (Table 2). It is divided into three moieties: glutamate, *p*-aminobenzoic acid and pteridine (Fig. 1c). The potential of glutamate is not significantly correlated to 1DHF. Pteridine shows a weak degree of complementarity, while *p*-aminobenzoic acid exhibits a good degree of complementarity, which is comparable to that of the whole ligand.

In another dihydrofolate reductase data set, 5-deazafolate is highly complementary to its receptor site (Table 2). It is divided into three moieties: glutamate, *p*-aminobenzoic acid and 5-deazapteridine (Fig. 1d). Glutamate shows a weak degree of electrostatic potential complemen-

TABLE 1
REGRESSION ANALYSIS OF LIGAND MOIETY ELECTROSTATIC POTENTIALS[a]

Ligand	r	r_s	m	b
FK-506 and 1FKF				
FK-506	–0.125	–0.215	–0.122	19
moiety 1	–0.172	–0.204	–1.732	–96
moiety 2	0.262	0.160	0.228	50
moiety 3	0.003[b]	–0.114	0.002	38
moiety 4	–0.541	–0.547	–0.538	–22
moiety 5	–0.069[c]	–0.180	–0.172	16
moiety 6	–0.282	–0.316	–0.707	–24
Retinol and 1RBP				
retinol	0.069	0.081	0.053	20
moiety 1	0.103	0.224	6.017	–682
moiety 2	0.237	0.480	5.589	–619
moiety 3	–0.070	–0.072[d]	–0.105	33

[a] None of the sites carry any formal charges.
[b] r not significantly different from 0, $p=0.902$.
[c] r not significantly different from 0, $p=0.018$.
[d] r and r_s not significantly different from 0, $p=0.024$ for r and $p=0.021$ for r_s.

tarity to its site, while *p*-aminobenzoic acid shows good complementarity. The potentials of 5-deazapteridine do not exhibit any significant statistical correlation with the site potentials.

The third dihydrofolate reductase data set is $NADP^+$ and 7DFR (Table 2). This is the cofactor site, not the substrate site as in the previous two cases. $NADP^+$ shows a high degree of electrostatic potential complementarity to its site on 7DFR. The ligand is divided into five moieties: adenine, ribose phosphate, pyrophosphate, ribose and nicotinamide. The potentials of adenine do not exhibit any significant statistical correlation with the site potentials. Ribose phosphate, ribose and nicotinamide all show some electrostatic potential similarity. Only the pyrophosphate group is very weakly complementary to the site.

FAD is highly complementary to its site on 1FNR (Table 3). This ligand is divided into five moieties: adenine, pyrophosphate, ribose, isoalloxazine and ribitol. The pyrophosphate moiety is weakly complementary to the site. Adenine shows negative correlation coefficients, but the slope is very steep. Ribose shows weak electrostatic similarity to the site. The potentials of both isoalloxazine and ribitol do not exhibit any significant degree of statistical correlation with the site potentials.

FAD is also complementary to its site on 1PHH (Table 3). In this case, only ribose shows some electrostatic similarity to the site, but the slope is very shallow. The potentials of all the other moieties do not exhibit any significant degree of statistical correlation with the site potentials.

In the case of NADH, only the interaction of subunit Q is studied, because it has the best complementarity of the four subunits. NADH shows a good degree of electrostatic complementarity with its site on 1GD1, subunit Q (Table 3). The molecule is divided into five moieties: adenine, the adenine-bound ribose, pyrophosphate, the nicotinamide-bound ribose and nicotinamide. Both adenine and adenine-bound ribose show electrostatic similarity to the site. The pyrophosphate moiety ex-

TABLE 2
REGRESSION ANALYSIS OF LIGAND MOIETY ELECTROSTATIC POTENTIALS[a]

Ligand	r	r_s	m	b
Folate and 1DHF (subunit A)				
folate	−0.771	−0.844	−1.988	−328
glutamate	−0.101	0.044[b]	−3.589	−78
p-aminobenzoic acid	−0.772	−0.717	−0.608	82
pteridine	−0.323	−0.267	−0.166	3
5-Deazafolate and 2DHF (subunit A)				
5-deazafolate	−0.788	−0.879	−2.808	−35
glutamate	−0.404	−0.414	−1.224	−594
p-aminobenzoic acid	−0.666	−0.665	−0.842	88
5-deazapteridine	0.173	0.036[c]	0.161	312
$NADP^+$ and 7DFR				
$NADP^+$	−0.817	−0.838	−1.338	−118
adenine	0.126	0.010[d]	0.118	−32
ribose phosphate	0.616	0.638	3.142	−2394
pyrophosphate	−0.144	−0.418	−2.874	723
ribose	0.217	0.270	0.112	−17
nicotinamide	0.442	0.537	1.459	148

[a] All sites carry at least one formal charge and the correlation coefficient $r < -0.7$ for the whole ligand.
[b] r_s not significantly different from 0, $p = 0.103$.
[c] r_s not significantly different from 0, $p = 0.164$.
[d] r_s not significantly different from 0, $p = 0.730$.

hibits a weak degree of electrostatic complementarity. The potentials of neither nicotinamide-bound ribose nor nicotinamide show any significant degree of statistical correlation with the site potentials.

2'-GMP is highly complementary to its site 1RNT (Table 4). The ionization state of 2'-GMP is assigned according to the reference paper [22]. It is divided into three moieties: guanine, ribose and hydrogen phosphate (HPO_4^-, see Fig. 1e). The guanine group exhibits a weak degree of complementarity, while the ribose group is hardly complementary. The hydrogen phosphate group shows good electrostatic potential complementarity to the site, but the correlation and regression are not as good as for the whole ligand.

Cyclic AMP shows a good degree of electrostatic potential complementarity with its site on dimeric 3GAP (Table 4). It is divided into two moieties: adenine and cyclic ribose phosphate. Adenine is very weakly complementary to the site. Cyclic ribose phosphate shows a reasonable degree of complementarity, but not as high as that of the whole ligand.

In these eight cases, i.e., folate (1DHF), 5-deazafolate (2DHF), FAD (1FNR and 1PHH), NADH (1GD1), 2'-GMP (1RNT), cAMP (3GAP) and $NADP^+$ (7DFR), all constituent chemical moieties are less complementary than the whole ligand. This is the most common finding.

TABLE 3
REGRESSION ANALYSIS OF LIGAND MOIETY ELECTROSTATIC POTENTIALS[a]

Ligand	r	r_s	m	b
FAD and 1FNR				
FAD	−0.796	−0.855	−1.385	−433
adenine	−0.134	−0.227	−14.46	70
pyrophosphate	−0.517	−0.482	−0.750	−769
ribose	0.184	0.140	2.753	−117
isoalloxazine	−0.199	0.028[b]	−0.231	−20
ribitol	0.077	−0.003[c]	0.058	20
NADH and 1GD1 (subunit Q)				
NADH	−0.725	−0.749	−2.075	464
adenine	0.354	0.321	0.344	−124
ribose (adenine-bound)	0.154	0.191	0.081	−6
pyrophosphate	−0.429	−0.507	−1.876	161
ribose (nicotinamide-bound)	0.133	0.058[d]	1.028	−420
nicotinamide	0.090[e]	0.248	0.165	−66
FAD and 1PHH				
FAD	−0.713	−0.702	−1.483	551
adenine	0.004	−0.057[f]	0.003	5
pyrophosphate	0.007	−0.039[g]	0.012	−953
ribose	0.202	0.290	0.095	−44
isoalloxazine	−0.040	0.052[h]	−6.561	3649
ribitol	−0.125	−0.086[i]	−0.048	60

[a] All sites carry at least one formal charge and the correlation coefficient $r < -0.7$ for the whole ligand.
[b] r_s not significantly different from 0, $p = 0.223$.
[c] r and r_s not significantly different from 0, $p = 0.005$ for r and $p = 0.923$ for r_s.
[d] r_s not significantly different from 0, $p = 0.045$.
[e] r not significantly different from 0, $p = 0.002$.
[f] r and r_s not significantly different from 0, $p = 0.893$ for r and $p = 0.054$ for r_s.
[g] r and r_s not significantly different from 0, $p = 0.829$ for r and $p = 0.254$ for r_s.
[h] r and r_s not significantly different from 0, $p = 0.073$ for r and $p = 0.022$ for r_s.
[i] r_s not significantly different from 0, $p = 0.002$.

TABLE 4
REGRESSION ANALYSIS OF LIGAND MOIETY ELECTROSTATIC POTENTIALS[a]

Ligand	r	r_s	m	b
GMP and 1RNT				
2'-GMP	−0.860	−0.833	−0.962	412
guanine	−0.257	−0.361	−1.262	880
ribose	−0.164	−0.265	−0.103	85
hydrogen phosphate	−0.645	−0.683	−1.093	496
Cyclic AMP and 3GAP (both subunits)				
cAMP	−0.708	−0.743	−1.751	276
adenine	−0.175	−0.143	−0.241	66
ribose phosphate	−0.609	−0.656	−1.945	329
Ace-Pro-Ala-Pro-Phe-OH and 4SGA				
inhibitor	−0.772	−0.731	−1.165	210
acetyl-Pro	−0.075[b]	0.067	−21.39	5523
Ala	−0.613	−0.227	−0.818	282
Pro	0.003[c]	0.125	0.003	17
Phe-OH	−0.801	−0.873	−1.352	290
Ace-Pro-Ala-Pro-Tyr-OH and 5SGA				
peptide	−0.763	−0.724	−1.155	207
acetyl-Pro	−0.070[d]	0.101	−25.67	6635
Ala	−0.692	−0.257	−0.900	302
Pro	−0.005[e]	0.132	−0.005	21
Tyr-OH	−0.789	−0.839	−1.287	269

[a] All sites carry at least one formal charge and the correlation coefficient $r < -0.7$ for the whole ligand.
[b] r not significantly different from 0, $p = 0.002$.
[c] r not significantly different from 0, $p = 0.929$.
[d] r not significantly different from 0, $p = 0.004$.
[e] r not significantly different from 0, $p = 0.880$.

The other two cases are ace-Pro-Ala-Pro-Phe-OH (4SGA) and ace-Pro-Ala-Pro-Tyr-OH (5SGA) (Table 4). Both ligands are divided into four moieties, corresponding to the constituent amino acids.

In the case of 4SGA, ace-Pro-Ala-Pro-Phe-OH exhibits good electrostatic potential complementarity. The potentials of neither acetyl-proline nor proline show any significant statistical correlation with the site, while alanine is reasonably complementary to the site. Phe-OH is highly complementary to the site, with better correlation coefficients and regression results than for the whole ligand.

5SGA and ace-Pro-Ala-Pro-Tyr-OH exhibit good electrostatic potential complementarity. The statistics are remarkably similar to those for 4SGA; the potentials of neither acetyl-proline nor proline show any significant statistical correlation with the site, while alanine is reasonably complementary to the site. Tyr-OH is highly complementary to the site, with better correlation coefficients and regression results than for the whole ligand.

In these two cases, namely, ace-Pro-Ala-Pro-Phe-OH/4SGA and ace-Pro-Ala-Pro-Tyr-OH/5SGA, one of the constituent moieties is slightly more complementary than the whole ligand, while the other moieties are all less complementary.

Complexes with at least one formal charge at the site: $-0.7 \leq r < -0.4$

Tables 5 and 6 show the seven ligand–receptor complexes where $-0.7 \leq r < -0.4$ for the whole

ligand: AMP (1AK3), biliverdin IX-γ (1BBP), FMN (1GOX), reduced FMN (3FXN), incomplete $NADP^+$ (6DFR) and folate (7DFR apo-enzyme).

In the case of biliverdin IX-γ, one of the moieties exhibits an electrostatic complementarity better than that of the whole molecule (Fig. 1f). This complex is a tetramer, and the subunit C interactions are chosen for our study because it exhibits the highest complementarity of the four subunits. Biliverdin IX-γ shows a weak degree of electrostatic potential complementarity to the site on 1BBP. The ligand is broken down into four moieties. Moieties 1 to 3 all exhibit a lower degree of complementarity, but moiety 4 shows a good degree of complementarity with the site.

In all other five cases, the whole-ligand complementarity is always better or equal to the complementarities of the constituent moieties.

AMP exhibits a reasonable degree of electrostatic potential complementarity to its site 1AK3 (Table 5). It is broken down into three moieties: adenine, ribose and phosphate. Ribose shows electrostatic potential similarity with the site, while the potentials of neither adenine nor phosphate exhibit statistically significant correlations with the site.

FMN (oxidized form) shows a reasonable degree of electrostatic potential complementarity with its site in 1GOX (Table 5). The ligand is divided into three moieties: phosphate, isoalloxazine and ribitol (Fig. 1g). The phosphate group exhibits a weak degree of complementarity to the site, while both isoalloxazine and ribitol show some electrostatic similarity.

FMN (reduced form) exhibits a good degree of electrostatic complementarity with its site in 3FXN (Table 6). The ligand is divided into three moieties: phosphate, semiquinone and ribitol (Fig. 1h). The semiquinone group shows electrostatic similarity with the site, while the other two moieties are weakly complementary.

$NADP^+$ (incomplete) is reasonably complementary to its site in 6DFR (Table 6). It is divided into

TABLE 5
REGRESSION ANALYSIS OF LIGAND MOIETY ELECTROSTATIC POTENTIALS[a]

Ligand	r	r_s	m	b
AMP and 1AK3 (subunit A)				
AMP	−0.684	−0.733	−3.223	1548
adenine	−0.044	−0.085[b]	−0.018	19
ribose	0.667	0.721	0.399	−252
phosphate	−0.073	−0.097[c]	−12.71	8736
Biliverdin IX-γ and 1BBP (subunit C)				
biliverdin IX-γ	−0.455	−0.536	−0.633	−282
moiety 1	−0.365	−0.532	−9.290	1033
moiety 2	−0.057	0.029[d]	−0.041	423
moiety 3	−0.403	−0.375	−0.287	474
moiety 4	−0.716	−0.758	−1.514	501
FMN and 1GOX				
FMN	−0.490	−0.361	−2.194	1219
phosphate	−0.416	−0.350	−0.651	−478
isoalloxazine	0.257	0.432	0.226	−129
ribitol	0.138	0.181	0.042	−14

[a] All sites carry at least one formal charge and the correlation coefficient $-0.7 \le r < -0.4$ for the whole ligand.
[b] r and r_s not significantly different from 0, $p = 0.169$ for r and $p = 0.008$ for r_s.
[c] r and r_s not significantly different from 0, $p = 0.084$ for r and $p = 0.022$ for r_s.
[d] r and r_s not significantly different from 0, $p = 0.041$ for r and $p = 0.300$ for r_s.

TABLE 6
REGRESSION ANALYSIS OF LIGAND MOIETY ELECTROSTATIC POTENTIALS[a]

Ligand	r	r_s	m	b
Reduced FMN and 3FXN				
FMN	−0.674	−0.739	−2.796	−696
phosphate	−0.305	−0.305	−4.150	−662
semiquinone	0.404	0.192	0.226	59
ribitol	−0.226	−0.254	−0.292	8
Incomplete $NADP^+$ and 6DFR				
incomplete $NADP^+$	−0.663	−0.734	−2.471	18
adenine	0.047	0.003[b]	0.030	−3
pyrophosphate	−0.059[c]	−0.241	−1.780	6
propan-1,2-diol	0.529	0.553	0.447	−126
ribose phosphate	0.507	0.506	3.970	−2655
Folate and 7DFR apo-enzyme				
folate	−0.626	−0.695	−2.842	−30
glutamate	−0.219	−0.204	−0.940	−773
p-aminobenzoic acid	−0.704	−0.632	−0.653	−72
pteridine	−0.007	0.048[d]	−0.006	13

[a] All sites carry at least one formal charge and the correlation coefficient $-0.7 \leq r < -0.4$ for the whole ligand.
[b] r and r_s not significantly different from 0, $p = 0.128$ for r and $p = 0.924$ for r_s.
[c] r not significantly different from 0, $p = 0.096$.
[d] r and r_s not significantly different from 0, $p = 0.780$ for r and $p = 0.063$ for r_s.

four moieties: adenine, pyrophosphate, propan-1,2-diol and ribose phosphate (Fig. 1i). The potentials of adenine and pyrophosphate do not exhibit any statistically significant correlation with the site potentials. The potentials of propan-1,2-diol and ribose phosphate are similar to the site potentials.

Folate shows a reasonable degree of electrostatic potential complementarity to its site in the 7DFR apo-enzyme (Table 6). This ligand is divided, as described, into glutamate, *p*-aminobenzoic acid and pteridine. Glutamate is weakly complementary to the site. The *p*-aminobenzoic acid moiety has a more negative correlation coefficient than the whole ligand, but the Spearman's rank correlation coefficient is less negative, and the slope is less steep. The potential of pteridine does not exhibit any significant statistical correlation with the site potentials.

Complexes with at least one formal charge at the site: $0 < r \leq -0.4$

There is only one site, but seven ligands, where $0 < r \leq -0.4$, namely the HIV-1 protease site. The ligands are SKB-Va (1AAQ), SB204144 (1HOS), MVT-101 (4HVP), L-700417 (4PHV), acetyl-pepstatin (5HVP), JG-365 (7HVP) and U-85548e (8HVP). Three of these ligands, SB204144, L-700417 and acetyl-pepstatin, can be crystallized in two alternative forms. In this work only one form is studied. The statistical analysis is displayed in Tables 7 and 8.

For SKB-Va and the 1AAQ site (Table 7), the correlation coefficients are slightly negative. There is probably a very weak degree of electrostatic potential complementarity. The complex is divided into six moieties: 'distal' alanine, 'proximal' alanine, an alkane moiety, a phenyl moiety, 'proximal' valine and 'distal' valine (Fig. 1j). 'Distal' alanine has a slightly better correlation coefficient; the slope of regression is also much closer to −1. The potentials of 'proximal' alanine do not bear any statistically significant correlation with the potentials of the receptor; the slope of regression is extremely shallow. The alkane segment also has a weakly complementary correlation, but the slope

TABLE 7
REGRESSION ANALYSIS OF LIGAND MOIETY ELECTROSTATIC POTENTIALS[a]

Ligand	r	r_s	m	b
SKB-Va and 1AAQ				
SKB-Va	−0.285	−0.252	−3.275	−1018
distal alanine	−0.448	−0.453	−1.164	−32
proximal alanine	0.002	−0.013[b]	0.001	30
alkane moiety	−0.449	−0.558	−0.497	−205
phenyl moiety	−0.301	−0.412	−0.280	284
proximal valine	−0.131	−0.033	−0.125	−45
distal valine	0.361	0.391	0.117	67
SB204144 (form A) and 1HOS				
SB204144	−0.208	−0.289	−0.130	−17
carbobenzyloxy 1	−0.057[c]	−0.211	−5.321	−1491
valine 1	−0.112	−0.091	−0.486	−118
phenyl 1	−0.482	−0.570	−0.031	7
phosphinate	0.177	0.288	0.127	63
phenyl 2	−0.428	−0.536	−0.044	5
valine 2	−0.418	−0.442	−0.433	−120
carbobenzyloxy 2	0.103	0.149	6.148	1708
MVT-101 and 4HVP				
MVT-101	−0.097	−0.125	−1.498	−1230
threonine 1	−0.385	−0.341	−0.193	−161
isoleucine 2	−0.430	−0.598	−0.215	−179
norleucine 3	0.111	0.113	0.007	30
norleucine 4	−0.288	−0.456	−0.095	−64
glutamine 5	0.667	0.760	0.637	667
arginine 6	−0.315	−0.350	−0.926	−477
L-700417 (form 1) and 4PHV				
L-700417	−0.232	−0.421	−0.067	−17
cis-aminohydroxyindan 1	0.107	0.157	0.035	38
benzene + carbonyl 1	−0.333	−0.592	−0.062	−8
benzene + carbonyl 2	−0.199	−0.373	−0.033	3
cis-aminohydroxyindan 2	−0.013[d]	0.102	−0.005	17

[a] All sites carry at least one formal charge and the correlation coefficient $0 < r \leq -0.4$ for the whole ligand.
[b] r and r_s not significantly different from 0, $p = 0.963$ for r and $p = 0.747$ for r_s.
[c] r not significantly different from 0, $p = 0.024$.
[d] r not significantly different from 0, $p = 0.619$.

of regression is far from −1. Both the phenyl moiety and the 'proximal' valine show little complementarity, with shallow slopes of regression. The 'distal' valine exhibits electrostatic similarity.

Form A of SB204144 is very weakly complementary to its site in 1HOS (Table 7). SB204144 is a symmetric molecule, so the identical moieties are designated 1 and 2 to differentiate them. The seven moieties are: carbobenzyloxy 1, valine 1, phenyl 1, phosphinate, phenyl 2, valine 2 and carbobenzyloxy 2 (Fig. 1k). The potentials of carbobenzyloxy 1 show no statistically significant correlation with its site. Valine 1 has slightly negative correlation coefficients. The correlation coefficients of phenyl groups 1 and 2 are better, but the slope of regression is very shallow in both cases. Both phosphinate and carbobenzyloxy 1 exhibit electrostatic potential similarity. Valine 2 has reasonably negative correlation coefficients, and its slope of regression is also acceptable; it is more complementary than the whole ligand.

MVT-101 is very weakly complementary to its site in 4HVP (Table 7). It is divided into its constituent amino acids: threonine 1, isoleucine 2, norleucine 3, norleucine 4, glutamine 5 and arginine 6 (Fig. 1l). Threonine 1 is weakly complementary to the site, with a reasonable slope of regression. Isoleucine 2 has slightly more negative correlation coefficients, but its slope of regression is very shallow. Norleucines 3 and 4 have positive and negative correlation coefficients, respectively, but their slopes of regression are extremely shallow. Glutamine 5 shows a considerable degree of electrostatic potential similarity. Arginine 6 is weakly complementary to the site, with a slope very close to –1.

Form 1 of L-700417 shows very little complementarity to its site 4PHV (Table 7); the slope of regression is so shallow that the negative correlation coefficients are useless for assessing the degree of complementarity. It is divided into four moieties: *cis*-aminohydroxyindan 1, benzene + carbonyl 1, benzene + carbonyl 2 and *cis*-aminohydroxyindan 2 (Fig. 1m). The *cis*-aminohydroxyindan 1 moiety shows positive correlation coefficients, while the two benzene + carbonyl moieties exhibit negative correlation coefficients. The potentials of *cis*-aminohydroxyindan 2 do

TABLE 8
REGRESSION ANALYSIS OF LIGAND MOIETY ELECTROSTATIC POTENTIALS[a]

Ligand	r	r_s	m	b
Acetyl-pepstatin (form 1) and 5HVP				
acetyl-pepstatin	–0.128	0.016[b]	–0.140	–244
acetyl-valine 1	–0.100	–0.115	–0.032	13
valine 2	–0.214	–0.250	–0.066	–21
statine 3	–0.351	–0.456	–0.066	–39
alanine 4	0.060[c]	0.256	0.034	48
statine 5	–0.572	–0.451	–2.440	–1922
JG-365 and 7HVP				
JG-365	–0.006	–0.005[d]	–0.004	29
serine 1	0.287	0.226	0.638	–112
leucine 2	–0.307	–0.455	–0.087	49
asparagine 3	0.226	0.262	0.904	–148
phenylalanine 4	0.428	0.171	0.109	13
proline 5	0.199	0.211	0.442	–43
isoleucine 6	–0.535	–0.145	–0.195	49
valine 7	0.368	0.572	0.221	–31
U-85548e and 8HVP				
U-85548e	–0.198	–0.219	–7.364	1090
valine 1	–0.606	–0.583	–3.903	922
serine 2	–0.094[e]	–0.157	–0.270	63
glutamine 3	–0.492	–0.642	–0.377	122
asparagine 4	0.082	0.124	1.580	–204
leucine 5	0.572	0.261	0.126	21
valine 6	0.008[f]	–0.192	0.007	30
isoleucine 7	–0.500	–0.352	–0.211	52
valine 8	0.104	0.102	6.920	–2060

[a] All sites carry at least one formal charge and the correlation coefficient $0 < r \leq -0.4$ for the whole ligand.
[b] r_s not significantly different from 0, $p = 0.152$.
[c] r not significantly different from 0, $p = 0.089$.
[d] r and r_s not significantly different from 0, $p = 0.560$ for r and $p = 0.624$ for r_s.
[e] r not significantly different from 0, $p = 0.006$.
[f] r not significantly different from 0, $p = 0.785$.

not show any statistically significant correlation with the site potentials. In all five cases, the slopes of regression are very shallow; there is probably no correlation at all between the ligand or moiety potentials and the site potentials.

The potentials of form 1 of acetyl-pepstatin do not exhibit any statistically significant correlation to the site potentials from 5HVP (Table 8); r_s is not significantly different from zero, and the slope of regression is shallow. The compound is divided into four moieties: acetyl-valine 1, valine 2, statine 3, alanine 4 and statine 5 (Fig. 1n). The correlation coefficients for acetyl-valine 1, valine 2 and statine 3 are negative, but the slope of regression in all three cases is very shallow. The potential of alanine 4 does not show any statistically significant correlation with the site potentials, and the slope is very shallow. Statine 5 is the only moiety to show some degree of complementarity with an acceptable slope of regression.

TABLE 9
REGRESSION ANALYSIS OF LIGAND MOIETY ELECTROSTATIC POTENTIALS[a]

Ligand	r	r_s	m	b
Carboxymethylcoenzyme A and 1CSC				
carboxymethyl-CoA	0.207	0.152	3.379	−3164
adenine	0.125	0.156	0.117	−52
ribose phosphate	0.604	0.592	5.794	−3471
pyrophosphate	−0.116	−0.106	−0.316	−710
pantothenic acid	−0.102	−0.295	−0.068	68
carboxymethyl-β-mercaptoethylamine	−0.100	−0.152	−0.071	−361
Carboxymethylcoenzyme A and 2CSC				
carboxymethyl-CoA	0.200	0.146	3.418	−3186
adenine	0.126	0.155	0.097	−43
ribose phosphate	0.613	0.604	5.900	−3530
pyrophosphate	−0.189	−0.173	−0.454	−606
pantothenic acid	−0.102	−0.321	−0.072	72
carboxymethyl-β-mercaptoethylamine	−0.081[b]	−0.132	−0.055	−369
Acetylcoenzyme A and 3CSC				
acetyl-CoA	0.333	0.305	4.056	−3410
adenine	0.108	0.135	0.077	−33
ribose phosphate	0.573	0.560	6.206	−3699
pyrophosphate	−0.089	−0.043[c]	−0.200	−793
pantothenic acid	−0.034[d]	−0.132	−0.019	18
acetyl-β-mercaptoethylamine	−0.089[e]	−0.136	0.018	12
Acetylcoenzyme A and 4CSC				
acetyl-CoA	0.326	0.304	4.167	−3462
adenine	0.121	0.142	0.097	−44
ribose phosphate	0.633	0.630	5.570	−3378
pyrophosphate	0.597	0.602	0.496	−744
pantothenic acid	−0.012[f]	−0.128	0.006	27
acetyl-β-mercaptoethylamine	−0.186	−0.095[g]	−0.029	40

[a] All sites carry at least one formal charge and the correlation coefficient $r \geq 0$ for the whole ligand.
[b] r not significantly different from 0, $p = 0.004$.
[c] r and r_s not significantly different from 0, $p = 0.011$ for r and $p = 0.217$ for r_s.
[d] r not significantly different from 0, $p = 0.128$.
[e] r not significantly different from 0, $p = 0.002$.
[f] r not significantly different from 0, $p = 0.569$.
[g] r_s not significantly different from 0, $p = 0.001$.

The potential of JG-365 does not exhibit any statistically significant correlation with the site potentials from 7HVP (Table 8); the slope of regression is extremely shallow. It is divided into seven moieties: serine 1, leucine 2, asparagine 3, phenylalanine 4, proline 5, isoleucine 6 and valine 7 (Fig. 1o). Serine 1 and asparagine 3 both show comparable degrees of electrostatic similarity with the site. Leucine 2 has negative correlation coefficients, but the slope of regression is very shallow. Phenylalanine 4, proline 5 and valine 7 exhibit varying degrees of electrostatic similarity, but the slope of regression is shallow. Isoleucine 6 shows a low degree of electrostatic complementarity with a shallow slope.

U-85548e is weakly complementary to its site in 8HVP (Table 8), and the slope of regression is rather steep. It is divided into eight moieties: valine 1, serine 2, glutamine 3, asparagine 4, leucine 5, valine 6, isoleucine 7 and valine 8 (Fig. 1p). Valine 1 is reasonably complementary to the site, with still a steep slope of regression. The potentials of serine 2 and valine 6 do not exhibit any statistically significant degree of correlation with the site potentials. Glutamine 3 and isoleucine 7 are weakly complementary to the site, with shallow slopes of regression. Asparagine 4, leucine 5 and valine 8 show varying degrees of electrostatic potential similarity, but the correlation coefficients are small in the case of asparagine 4, or the slope of regression is far from −1 in the other two cases.

In summary, there is probably not any significant degree of complementarity between any of these seven ligands and their receptor sites. In the cases of L-700417 (4PHV), JG-365 (7HVP) and U-85548e (8HVP), neither the whole ligands nor the constituent moieties exhibit any complementarity. In the other four cases, one of the moieties possibly shows more complementarity than the whole ligand. The original ligands and these moieties are: SKB-Va ('distal' alanine), SB204144 (valine 2), MVT-101 (arginine 6) and acetyl-pepstatin (statine 5). However, the increase in complementarity is extremely small, and compared against a background of noncomplementarity, the difference may not be significant.

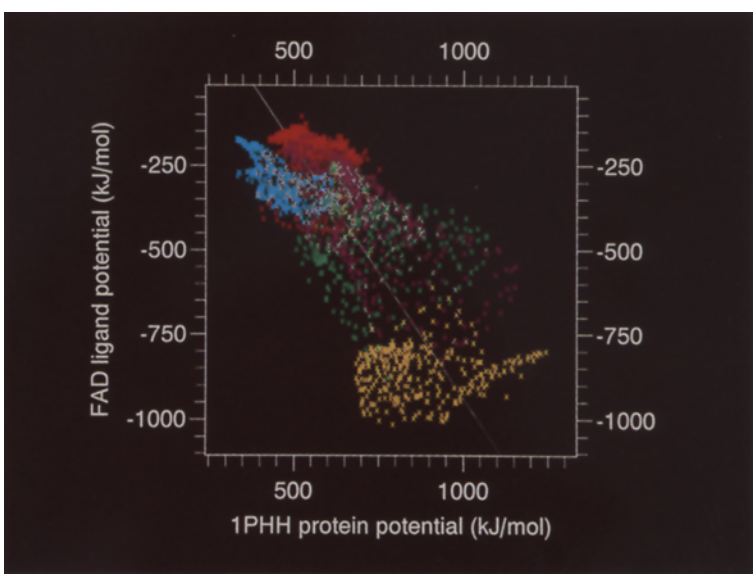

Fig. 2. The scattergram of FAD against 1PHH. The differently coloured points correspond to different moieties of the ligand van der Waals surface. Yellow: pyrophosphate; green: ribose; light blue: adenine; red: alloxazine; magenta: ribitol. For the whole ligand, $r = -0.713$, $r_s = -0.702$, $m = -1.483$ and $b = 551$.

Complexes with at least one formal charge at the site: $r \geq 0$

There are five ligand–receptor complexes in this group: carboxymethylcoenzyme A and 1CSC, carboxymethylcoenzyme A and 2CSC, acetylcoenzyme A and 3CSC, acetylcoenzyme A and 4CSC, NAD and 1LDM. Tables 9 and 10 display the statistical analysis.

Carboxymethylcoenzyme A in both 1CSC and 2CSC exhibits a weak degree of electrostatic potential similarity (Table 9). The ligand is broken down into five moieties: adenine, ribose phosphate, pyrophosphate, pantothenic acid and carboxymethyl-β-mercaptoethylamine (Fig. 1q). In both 1CSC and 2CSC, adenine shows a very weak degree of similarity with a shallow slope of regression, while ribose phosphate exhibits a higher degree of similarity with a steep slope of regression. The pyrophosphate moiety is weakly complementary in both 1CSC and 2CSC, with an acceptable slope of regression. In 1CSC as well as 2CSC, both pantothenic acid and carboxymethyl-β-mercaptoethylamine have slightly negative correlation coefficients with very shallow slopes of regression. In fact, for carboxymethyl-β-mercaptoethylamine and 2CSC, the correlation is not statistically significant.

Acetylcoenzyme A in both 3CSC and 4CSC shows a low degree of electrostatic potential similarity, with rather steep slopes (Table 9). The ligand is divided into five moieties: adenine, ribose phosphate, pyrophosphate, pantothenic acid and acetyl-β-mercaptoethylamine. In both 3CSC and 4CSC, adenine shows a very weak degree of similarity with a shallow slope of regression, while ribose phosphate exhibits a higher degree of similarity with a steep slope of regression. The potential of the pyrophosphate moiety in 3CSC does not show any statistically significant correlation with the site potential. However, in 4CSC the pyrophosphate moiety exhibits a reasonable degree of electrostatic similarity. The potentials of pantothenic acid and acetyl-β-mercaptoethylamine in neither 3CSC nor 4CSC show any statistically significant correlation with the site potentials.

NAD^+ is weakly similar to its site 1LDM (Table 10), with an acceptable slope of regression. It is divided into five moieties: adenine, adenine-bound ribose, pyrophosphate, nicotinamide-bound ribose and nicotinamide. Adenine is weakly similar to the site, while adenine-bound ribose is even more weakly similar with an extremely shallow slope of regression. The pyrophosphate moiety is weakly complementary to the site. The potentials of nicotinamide-bound ribose do not exhibit any statistically significant correlation with the site potentials. The nicotinamide moiety is reasonably complementary, but the slope of regression is just acceptable.

In summary, neither carboxymethylcoenzyme A, acetylcoenzyme A, nor any of their constituent

TABLE 10
REGRESSION ANALYSIS OF LIGAND MOIETY ELECTROSTATIC POTENTIALS FOR THE NAD^+–1LDM COMPLEX[a]

Ligand	r	r_s	m	b
NAD^+	0.261	0.337	0.719	−599
Adenine	0.340	0.300	0.344	−72
Ribose (adenine-bound)	0.139	0.179	0.045	11
Pyrophosphate	−0.255	−0.255	−0.789	−436
Ribose (nicotinamide-bound)	−0.042	−0.002[b]	−0.020	30
Nicotinamide	−0.627	−0.627	−0.587	950

[a] All sites carry at least one formal charge and the correlation coefficient $r \geq 0$ for the whole ligand.
[b] r and r_s not significantly different from 0, $p = 0.159$ for r and $p = 0.938$ for r_s.

chemical moieties exhibit any significant electrostatic complementarity against the receptor site. In the case of NAD^+ (1LDM), both the pyrophosphate and nicotinamide moieties are slightly more complementary than the whole ligand, which shows no significant complementarity.

DISCUSSION

We have taken ligands from the co-crystals of the Brookhaven PDB, and examined the electrostatic potential complementarity between the ligand and the site, and also between the constituent moieties of the ligand and the site. Many ligands show good complementarity to the site, but their constituent moieties exhibit a complementarity that is less good than the whole ligand.

Of the 16 ligand–receptor complexes with demonstrable complementarity, i.e., $r<-0.4$, 13 of the ligands show higher electrostatic complementarity than any of the constituent moieties. In the other three cases, one of the moieties is more complementary than the whole ligand. However, the whole-ligand complementarity does not appear to be the mathematical mean of the individual complementarities, nor does it seem to be a simple relationship between the moiety and whole-ligand complementarities. For the other five sites, the ligand does not exhibit any significant electrostatic complementarity, so it would be difficult to compare the case of the whole ligand with those of ligand moieties.

The results shown above demonstrate clearly that it is very difficult to predict, using an additive model, the electrostatic potential complementarity of the whole ligand from the complementarity of its constituent chemical moieties. Take, for example, the case of FAD and 1PHH, and that of SKB-Va and 1AAQ. If one adds up the Pearson's correlation coefficients for all the five moieties of FAD, one would get $r=0.048$, but the r value for the whole ligand is -0.713. In this case, the moieties have poor electrostatic potential complementarity, but the whole ligand has a much higher complementarity. In the case of SKB-Va, if one repeats this process, one would obtain a combined-moiety r value of -0.966, but the actual whole-ligand correlation coefficient is only -0.285. In this case, the individual moieties have weak complementarity, but the whole-ligand complementarity is very weak.

Figure 2 shows the scattergram for the case of FAD and 1PHH. The points on the graph are coloured according to the moiety they come from. The reason why electrostatic potential complementarity is not simply an additive property can be readily seen. The distribution of points on a scattergram can be noncomplementary for one moiety, but they could combine to produce a complementary distribution. It is extremely difficult to predict the whole-ligand distribution from the moiety complementarity statistics; one would need detailed information of the potentials of the moiety to calculate the whole-ligand complementarity statistics. Complementarity is not simply an additive or multiplicative parameter. It should be noted in passing that, in most cases studied, all constituent moieties exhibit a lower complementarity than the whole ligand.

This has very important implications for drug design. Many ad hoc drug design strategies are still based on the idea that complementary moieties that fit different parts of a binding site can be pieced together to form a drug molecule that fits the site as a whole. The medicinal chemist would have a qualitative idea of the electrostatic potential of common chemical fragments, and would place the appropriate fragments one by one to fit the properties of the binding site. Whilst this local optimization strategy is clearly indispensable in optimizing hydrogen bonds, which are rather local interactions, it would not be appropriate for assigning fragments to give a resultant

ligand that exhibits electrostatic potential complementarity. The optimization must be performed globally, because electrostatic potential complementarities of small molecules bear no simple relationship to the complementarity of a large molecule made up of them.

It would be interesting to know why some ligands or moieties are complementary to the site, while some are not. What are the necessary and sufficient conditions for electrostatic potential complementarity? Is it possible to predict the complementarity from structural properties of the ligand–receptor complexes? The following paper [23] discusses both questions.

ACKNOWLEDGEMENTS

P.L.C. thanks the Croucher Foundation for support of part of this work, Rhône-Poulenc Rorer for a postdoctoral research fellowship and New Hall, Cambridge for a Research Fellowship. P.M.D. thanks the Wellcome Trust for personal financial support through the Principal Research Fellowship Scheme. Part of this work was carried out in the Cambridge Centre for Molecular Recognition, supported by the SERC.

REFERENCES

1 Chau, P.-L. and Dean, P.M., J. Comput.-Aided Mol. Design, 8 (1994) 513.
2 Dreyer, G.B., Lambert, D.M., Meek, T.D., Carr, T.J., Tomaszek, T.A., Fernandez, A.V., Bartus, H., Cacciavillani, E., Hassell, A.M., Minnich, M., Petteway, S.R. and Metcalf, B.W., Biochemistry, 31 (1992) 6646.
3 Abdel-Meguid, S.S., Zhao, B., Murthy, D.H.M., Winborne, E., Choi, J.-K., DesJarlais, R.L., Minnich, M.D., Culp, J.S., Debouck, C., Tomaszek, T.A., Meek, T.D. and Dreyer, G.B., Biochemistry, 32 (1993) 7972.
4 Miller, M., Schneider, J., Sathyanarayana, B.K., Toth, M.V., Marshall, G.R., Clawson, L., Selk, L., Kent, S.B.H. and Wlodawer, A., Science, 246 (1989) 1149.
5 Bone, R., Vacca, J.P., Anderson, P.S. and Holloway, M.K., J. Am. Chem. Soc., 113 (1991) 9382.
6 Fitzgerald, P.M.D., McKeever, B.M., Van Middlesworth, J.F., Springer, J.P., Heimbach, J.C., Leu, C.-T., Herbert, W.K., Dixon, R.A.F. and Drake, P.L., J. Biol. Chem., 265 (1990) 14209.
7 Swain, A.L., Miller, M.M., Green, J., Rich, D.H., Schneider, J., Kent, S.B.H. and Wlodawer, A., Proc. Natl. Acad. Sci. USA, 87 (1990) 8805.
8 Jaskólski, M., Tomasselli, A.G., Sawyer, T.K., Staples, D.G., Heinrikson, R.L., Schneider, J., Kent, S.B.H. and Wlodawer, A., Biochemistry, 30 (1991) 1600.
9 Diederichs, K. and Schulz, G.E., J. Mol. Biol., 217 (1991) 541.
10 Huber, R., Schneider, M., Mayr, I., Müller, R., Deutzmann, R., Suter, F., Zuber, H., Falk, H. and Kayser, H., J. Mol. Biol., 198 (1987) 499.
11 Karpusas, M., Holland, D. and Remington, S.J., Biochemistry, 30 (1991) 6024.
12 Davies, J.F., Delcamp, T.J., Prendergast, N.J., Ashford, V.A., Freisheim, J.H. and Kraut, J., Biochemistry, 29 (1990) 9467.
13 Bystroff, C., Oatley, S.J. and Kraut, J., Biochemistry, 29 (1990) 3263.
14 Van Duyne, G.D., Standaert, R.F., Karplus, P.A., Schreiber, S.L. and Clardy, J., Science, 252 (1991) 839.
15 Karplus, P.A., Daniels, M.J. and Herriott, J.R., Science, 251 (1991) 60.
16 Schreuder, H.A., Van der Laan, J.M., Hol, W.G.J. and Drenth, J., J. Mol. Biol., 199 (1988) 637.
17 Skarżyński, T., Moody, P.C.E. and Wonacott, A.J., J. Mol. Biol., 193 (1987) 171.
18 Lindqvist, Y., J. Mol. Biol., 209 (1989) 151.
19 Smith, W.W., Burnett, R.M., Darling, G.D. and Ludwig, M.L., J. Mol. Biol., 117 (1977) 195.
20 Abad-Zapatero, C., Griffith, J.P., Sussman, J.L. and Rossmann, M.G., J. Mol. Biol., 198 (1987) 445.
21 Cowan, S.W., Newcomer, M.E. and Jones, T.A., Protein Struct. Funct. Genet., 8 (1990) 44.
22 Arni, R., Heinemann, U., Tokuoka, R. and Saenger, W., J. Biol. Chem., 263 (1988) 15358.
23 Weber, I.T. and Steitz, T.A., J. Mol. Biol., 198 (1987) 311.
24 James, M.N.G., Sielecki, A.R., Brayer, G.D., Delbaere, L.T.J. and Bauer, C.-A., J. Mol. Biol., 144 (1980) 43.
25 Chau, P.-L. and Dean, P.M., J. Comput.-Aided Mol. Design, 8 (1994) 545.