Journal of Computer-Aided Molecular Design, 8 (1994) 513–525
ESCOM

J-CAMD 260

Electrostatic complementarity between proteins and ligands. 1. Charge disposition, dielectric and interface effects

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; Dielectric

SUMMARY

Electrostatic interactions have always been considered an important factor governing ligand–receptor interactions. Previous work in this field has established the existence of electrostatic complementarity between the ligand and its receptor site. However, this property has not been treated rigorously, and the description remains largely qualitative. In this work, 34 data sets of high quality were chosen from the Brookhaven Protein Databank. The electrostatic complementarity has been calculated between the surface potentials; complementarity is absent between adjacent or neighbouring atoms of the ligand and the receptor. There is little difference between complementarities on the total ligand surface and the interfacial region. Altering the homogeneous dielectric to distance-dependent dielectrics reduces the complementarity slightly, but does not affect the pattern of complementarity.

INTRODUCTION

Electrostatic theory has played an important rôle in the development of models for molecular recognition and their consequent application in drug design. This importance stems from two factors: firstly, the display of the molecular electrostatic potential, and its visual inspection, is simple; secondly, the Coulombic component in the interaction energy between molecules can readily be understood. Comparison of the electrostatic potential maps between a ligand and its site often reveals what appear to be tantalizingly complementary dispositions. However, a critical analysis of electrostatic complementarity has not been made. Probably the principal reason for this omission has been the difficulty of quantifying complementarity in any satisfactory way; consequently, hand-waving arguments have prevailed. This difficulty stems from the fact that the notion of complementarity, used by different scientists, can be developed along two lines of thought: complementarity in pattern, where peaks and troughs in potential are identified, and

*To whom correspondence should be addressed.

complementarity where values of potential are compared at specified positions. In this paper we analyze in detail the electrostatic complementarity between 41 ligands co-crystallized with their proteins from the Brookhaven Protein Data Bank. The second paper in this series [1] is concerned with a more detailed analysis of complementarity in terms of component parts of the molecular structures; the research attempts to investigate partitioning of the electrostatic potential into molecular moieties and reveals in many cases a dominant charge effect. The third paper [2] examines the relative disposition of dominant ionized atoms in the ligand and site in electrostatic complementarity. Analytical expressions are derived which describe, in simple terms, the contribution from dominant charges in the interaction between ligands and proteins.

A model for studying electrostatic complementarity was proposed by Nakamura et al. [3]. They outlined a projection of molecular electrostatic potential from a site to the ligand surface and compared this projection with the ligand's potential on its own surface, the so-called host-on-guest and guest-on-guest method. The technique has the advantage that comparisons of potential are made at identical points on the ligand surface. To a large extent this surface will be at the interface between the ligand and the site. At any point p on the ligand surface, there will be a potential e_p^G generated by the guest and a potential e_p^H generated by the host. Their analysis of complementarity was dependent on defining a complementarity parameter C_p^{GH}:

$$C_p^{GH} = \text{sign}\left(e_p^G e_p^H\right)\sqrt{\left|e_p^G e_p^H\right|}$$

Negative values for C_p^{GH} denote complementarity. The difficulty with the method is that a pair of values such as $e_p^H = 4$, $e_p^G = -4$ would give the same value for complementarity as $e_p^H = 1$, $e_p^G = -16$. This is unsatisfactory if many values of C_p^{GH} are to be summed over all points p of the surface; furthermore, the method is not scale-free and so makes it impossible to compare different protein–ligand complexes.

In this work and the following two papers we preserve the host-on-guest and guest-on-guest model, but use a different form of analysis employing a rigorous statistical measure for quantifying electrostatic complementarity. Correlation coefficients, used before for a description of similarity between two ligands [4], are retained in exactly the same way to quantify the complementarity in potential between the site and the ligand at the ligand surface. However, where the correlation coefficients approach +1 for a measure of perfect similarity, the coefficients approach −1 for complementarity.

Common assumptions about electrostatic complementarity between the ligand and its site can be readily tested. One of the frequent assertions in drug design is that adjacent atoms between the ligand and the site have complementary partial charges. How true is that assertion? To what extent is electrostatic complementarity modified by dielectric effects? Is the complementarity at the interface between ligand and site different from that computed for the whole of the ligand surface? Is the relative disposition of ionized atoms in the ligand and site dominant in electrostatic complementarity? Can electrostatic complementarity be partitioned into contributions from component molecular moieties? All these questions have an important bearing on the development of novel computing methods using electrostatic interactions within automated techniques for drug design.

In the current paper, attention is focussed on four of these questions:

(1) The site-point ligand-point paradigm, employed in automated methods for drug design,

uses ligand seed points in the site region complementary to a site point. Whilst this is sensible for hydrogen-bonding seeds in retaining complementary features, the idea of complementarity between ligand-point site-point charges needs to be assessed from ligand–protein co-crystal data. An extensive statistical analysis of adjacent charges in the site and on the ligands is performed in this paper.

(2) Electrostatic complementarity is examined by correlating the two potentials computed at common points on the ligand surface. A theoretical justification for this correlation procedure is provided in the accompanying paper [2].

(3) The modification of the electrostatic potential by dielectric changes in the site has been studied using a linear distance-dependent model with two gradients.

(4) One problem that has been of concern to drug designers trying to mimic electrostatic similarity in a putative molecule compared with an active structure is: to what extent should the whole surface of the model molecule be compared to that of the active molecule's surface? In ligand–protein co-crystal complexes, it is possible to assess how much of the ligand surface is at the interface and furthermore, the contribution of the interfacial electrostatic complementarity can be compared with the complementarity of the whole ligand in the site.

METHODS

Data selection

The data sets chosen from the Brookhaven Protein Data Bank (PDB) conform to the following requirements:

(1) the atomic coordinates have been determined experimentally by X-ray crystallography;

(2) a resolution better than 2.5 Å;

(3) the ligand is bound to the site by noncovalent interactions only;

(4) no metallic ion is involved in the ligand–receptor interactions.

The major difficulty with the data analysis is that the ionization status of ligands, and of amino acids at the site, is often unknown. In some cases, other studies have been performed to determine the ionization state, and we follow the findings of that research to assign the correct ionization state. Often, however, the ionization state is not determined, and in such cases we decided to choose the ionization state that gives the best complementarity.

The ligand–receptor complexes chosen can be divided into two main groups: those where the receptor site carries no formal charges (see Table 1) and those where the receptor site carries at least one formal charge (see Table 2). The tables also include references to the original crystallographic papers.

TABLE 1
COMPLEXES IN WHICH THE RECEPTOR SITE CARRIES NO FORMAL CHARGES

PDB code	Ligand	Receptor	Ref.
1FKF	FK-506	FK-506-binding protein	[5]
1RBP	Retinol	Retinol-binding protein	[6]
5ABP	α-D-Galactose	*E. coli* arabinose-binding protein	[7]
6ABP	α-L-Arabinose	*E. coli* Met^{108} arabinose-binding protein	[7]
7ABP	α-D-Fucose	*E. coli* Met^{108} arabinose-binding protein	[7]
8ABP	α-D-Galactose	*E. coli* Met^{108} arabinose-binding protein	[7]

Calculation of electrostatic potential

Our model for studying electrostatic complementarity uses a host-on-guest and guest-on-guest paradigm where the host is the receptor site and the guest is the ligand. Hydrogen atoms were added to the heavy atoms using standard bond lengths and angles. The electrostatic potential complementarity did not vary significantly with the positions of the hydrogen atoms. This was tested on all rotatable hydrogen-bonding groups in seven ligand–protein co-crystals. Mulliken point charges for the ligand atoms were calculated using CNDO/2 [30], while the charges for the protein atoms were assigned using ECEPP; the ECEPP charges were originally calculated using CNDO/2 [31]. The electrostatic potentials on the van der Waals surface of the ligand were calculated from the Mulliken charges using the VSS program [32].

Surface points must be as equally spaced as possible to provide a satisfactory statistical sample. A gnomonic projection of a tessellated icosahedron is used to generate a set of points on each ligand atom [33]. Interior points, where van der Waals spheres overlap, are removed.

TABLE 2
COMPLEXES IN WHICH THE RECEPTOR SITE CARRIES AT LEAST ONE FORMAL CHARGE

PDB code	Ligand	Receptor	Ref.
1AAQ	SKB-Va	HIV-1 protease	[8]
1HOS	SB204144	HIV-1 protease	[9]
4HVP	MVT-101	HIV-1 protease	[10]
4PHV	L-700417	HIV-1 protease	[11]
5HVP	Acetyl-pepstatin	HIV-1 protease	[12]
7HVP	JG-365	HIV-1 protease	[13]
8HVP	U-85548e	HIV-1 protease	[14]
1AK3	AMP	Bovine adenylate kinase	[15]
1BBP	Biliverdin IX-γ	*Pieris brassicae* bilin-binding protein	[16]
1CSC	L-Malate and carboxymethyl-coenzyme A	Chicken citrate synthase	[17]
2CSC	D-Malate and carboxymethyl-coenzyme A	Chicken citrate synthase	[17]
3CSC	L-Malate and acetylcoenzyme A	Chicken citrate synthase	[17]
4CSC	D-Malate and acetylcoenzyme A	Chicken citrate synthase	[17]
1DHF	Folate	Human dihydrofolate reductase	[18]
2DHF	5-Deazafolate	Human dihydrofolate reductase	[18]
1FNR	FAD	Ferredoxin-$NADP^+$ reductase	[19]
1GD1	NADH	*Bacillus stearothermophilus* hologlyceraldehyde-3-phosphate dehydrogenase	[20]
1GOX	Oxidized FMN	Spinach glycolate oxidase	[21]
1LDM	NAD^+	Dogfish M_4 apo-lactate dehydrogenase	[22]
1PHH	FAD and 3,4-dihydroxybenzoate	*Pseudomonas fluorescens* p-hydroxybenzoate hydroxylase	[23]
1RNT	2'-GMP	*Aspergillus oryzae* ribonuclease T_1	[24]
2YPI	2-Phosphoglycolate	Triosephosphate isomerase	[25]
3FXN	Reduced FMN	*Clostridium MP* flavodoxin	[26]
3GAP	Cyclic AMP	*E. coli* catabolite gene activator protein	[27]
4SGA	Ace-P-A-P-F	*Streptomyces griseus* protease A	[28]
5SGA	Ace-P-A-P-Y	*Streptomyces griseus* protease A	[28]
6DFR	Incomplete $NADP^+$	*E. coli* dihydrofolate reductase	[29]
7DFR	$NADP^+$ and folate	*E. coli* dihydrofolate reductase	[29]

Two sets of ligand points were used: points on the whole of the ligand surface and points from the interfacial region only. An interface point is defined as lying within 2.8 Å of a site van der Waals surface atom. This is the diameter of a water molecule as defined by Lee and Richards [34]. For each surface point, there are two electrostatic potential values: one derived from the partial charges of the ligand, and the other derived from the partial charges of the protein.

Dielectric values were taken to be constant with $\varepsilon = \varepsilon_0$, or two linearly varying dielectrics:

(1) ε increasing between ε_0 and $4\varepsilon_0$ over a range of 10 Å and equal to $4\varepsilon_0$ for distances greater than 10 Å. Thus, for $d < 10$ Å, $\varepsilon = \varepsilon_0 (1 + (3d/10))$, where d is the distance in Å, and for $d \geq 10$ Å, $\varepsilon = 4\varepsilon_0$;

(2) ε increasing between ε_0 and $20\varepsilon_0$ over the same range, and equal to $20\varepsilon_0$ for distances greater than 10 Å. Thus, for $d < 10$ Å, $\varepsilon = \varepsilon_0 (1 + (19d/10))$, where d is the distance in Å, and for $d \geq 10$ Å, $\varepsilon = 20\varepsilon_0$.

Statistical analysis

Statistical analysis has been performed using parametric and nonparametric methods. Pearson's product-moment correlation coefficient, r, Spearman's rank correlation coefficient, r_s, and linear regression parameters have been calculated for all data sets.

Pearson's product-moment correlation coefficient

If two sets of potential values, each containing n samples, are compared, the similarity between these values can be obtained by Pearson's product-moment correlation coefficient [35], usually abbreviated to r. An r value of +1 means that the variance of one set of potentials is perfectly correlated with the variance of the other. An r value of −1 means that the variance of one set of potentials is perfectly anticorrelated with the variance of the other.

Spearman's rank correlation coefficient

Spearman's rank correlation, r_s, is a measure of the similarity in pattern between two sets of potentials [35]. This measure is independent of the actual values, so it is valuable in assessing the trend of electrostatic complementarity, especially in cases where the relationship is monotonic but may not be perfectly linear. A value of $r_s = +1$ suggests that the potential values have a similar pattern, and a value of $r_s = -1$ suggests that the pattern is complementary.

Perpendicular least-squares regression

Regression analysis between the two sets of potentials has been performed with a perpendicular least-squares procedure [36]. In this method, neither set of potentials is treated as the independent variable, and error is considered to come from both sets of potentials. The gradient of the linear regression is m and the intercept on the ordinate is b.

Perfect complementarity occurs when $r = -1$, $r_s = -1$, $m = -1$ and $b = 0$.

RESULTS

Nearest-neighbour analysis of partial charges

Atoms within 3.6 Å of each other between the ligand and its receptor site were obtained and

their corresponding partial charges analyzed statistically. At this separation distance the pair of atoms will be in contact with each other. A typical example of this analysis is shown in Fig. 1 for the protein part of 1FNR and its ligand FAD. The abscissa displays the partial charges on the binding-site atoms, while the ordinate shows the partial charges on the adjacent ligand atoms. It can be seen that there is no correlation between the partial charges on neighbouring atoms. Statistical analysis reveals that in this case, $r=0.027$ (the probability that $r=0$ being 0.662) and $r_s = -0.011$ (the probability that $r_s=0$ being 0.858). In all 41 cases studied, the probability that the correlation coefficient is zero is at least 0.130, while the probability that Spearman's rank correlation coefficient is zero is at least 0.022.

This study was repeated for atoms within 5.89 Å of each other. This is the shortest separation distance between a pair of atoms with a layer of atoms between them. Figure 2 shows the data for 1FNR and its ligand FAD. Statistical analysis shows that $r=-0.015$ (the probability that $r=0$ being 0.444) and $r_s=-0.023$ (the probability that $r_s=0$ being 0.239). In all cases studied, the probability that the correlation coefficient is zero is at least 0.168, while the probability that Spearman's rank correlation coefficient is zero is at least 0.202.

From the results obtained with these 34 data sets, one can conclude that no complementarity between partial charges on adjacent atoms or neighbouring atoms on the ligand–receptor complex appears to be present. The implications of this conclusion for drug design are discussed in a later section.

Effect of interface on complementarity

Complementarity has been calculated at two regions of the ligand surface, i.e., the whole of the ligand surface and the interfacial region of the surface. The interface is defined as the surface of the ligand which is adjacent to the surface of the protein and lying within 2.8 Å of a receptor

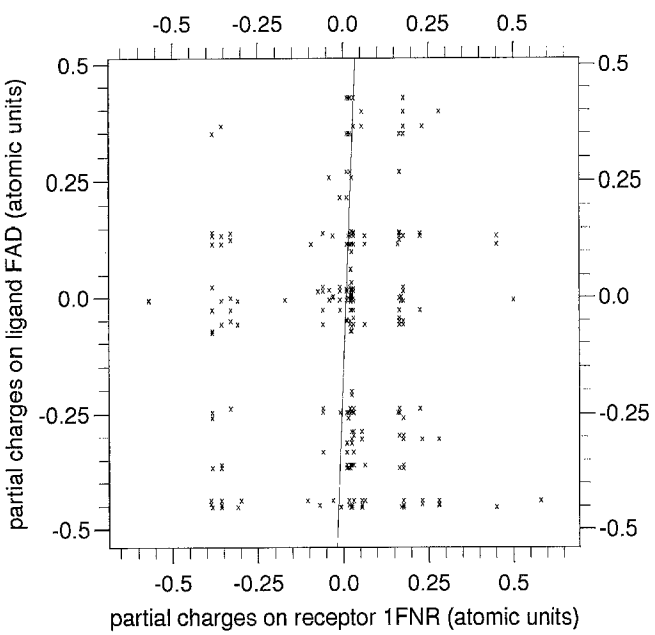

Fig. 1. Partial charges on adjacent atoms of 1FNR and FAD. The atoms are within 3.6 Å of each other.

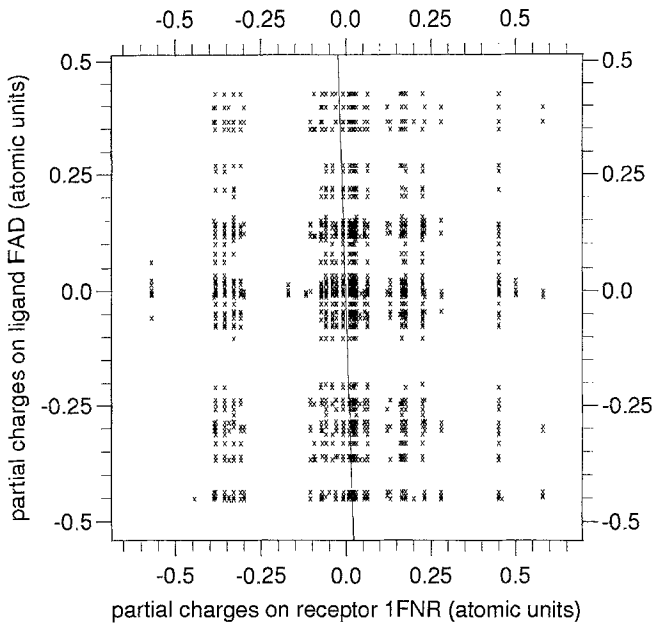

Fig. 2. Partial charges on neighbouring atoms of 1FNR and FAD. The atoms are within 5.89 Å of each other.

protein van der Waals surface atom. The results are shown in Tables 3 and 4. For symmetrical multi-subunit proteins, data for only one site and one ligand are shown.

We have considered 41 ligand–protein pairs. The interface comprises 59.5% (FK506) to 99.7% (α-L-arabinose) of the total ligand surface area. For 35 complexes, the ratio r_w / r_{if} is between 0.899 and 1.045, where r_w is the correlation coefficient considering the whole ligand van der Waals surface, and r_{if} is the correlation coefficient considering only the interfacial region. There appears to be strong evidence that including only the interfacial region has little effect on the complementarity. Twelve of these 35 ligand–protein complexes have $r_w < -0.7$ (L-malate and 1CSC, folate and 1DHF, 5-deazafolate and 2DHF, FAD and 1FNR, oxamate(a) and 1LDM, FAD and 1PHH, 3,4-dihydroxybenzoate and 1PHH, 2'-GMP and 1RNT, cyclic AMP and 3GAP, ace-Pro-Ala-Pro-Phe-OH and 4SGA, ace-Pro-Ala-Pro-Tyr-OH and 5SGA, $NADP^+$ and 7DFR

TABLE 3
ELECTROSTATIC COMPLEMENTARITY OF WHOLE LIGAND AND INTERFACIAL SURFACE[a]

Ligand	Receptor	r_w	r_{if}	f	r_w / r_{if}
FK506	1FKF	−0.125	−0.022*	0.595	5.682
Retinol	1RBP	0.069	0.074	0.973	0.932
α-D-Galactose	5ABP	0.190	0.183	0.938	1.038
α-L-Arabinose	6ABP	−0.259	−0.261	0.997	0.992
α-D-Fucose	7ABP	0.346	0.350	0.937	0.989
α-D-Galactose	8ABP	0.318	0.309	0.925	1.029

[a] None of the sites carry any formal charges.
r_w = whole surface r; r_{if} = interfacial region r; f = interface as a fraction of the whole surface. An * indicates that the correlation is not significantly different from 0, $p < 0.0005$.

TABLE 4
ELECTROSTATIC COMPLEMENTARITY OF WHOLE LIGAND AND INTERFACIAL SURFACE[a]

Ligand	Receptor	r_w	r_{if}	f	r_w / r_{if}
SKB-Va	1AAQ	–0.306	–0.294	0.911	1.041
AMP(a)	1AK3(a)	–0.684	–0.701	0.914	0.967
Biliverdin(a)	1BBP(a)	–0.334	–0.329	0.970	1.015
L-Malate	1CSC	–0.714	–0.715	0.913	0.999
Cm-CoA	1CSC	0.207	0.077	0.757	2.688
Folate(a)	1DHF(a)	–0.771	–0.770	0.950	1.001
FAD	1FNR	–0.796	–0.832	0.764	0.957
NADH(o)	1GD1(o)	–0.660	–0.660	0.877	1.000
FMN	1GOX	–0.490	–0.468	0.967	1.025
SB204144(a)	1HOS	–0.208	–0.199	0.898	1.045
NAD^+	1LDM	0.261	0.256	0.942	1.020
Oxamate(a)	1LDM	–0.904	–0.904	0.982	1.000
FAD	1PHH	–0.713	–0.701	0.936	1.017
3,4-DHB	1PHH	–0.855	–0.858	0.973	0.997
2'-GMP	1RNT	–0.860	–0.857	0.854	1.004
D-Malate	2CSC	–0.432	–0.422	0.973	1.024
Cm-CoA	2CSC	0.200	0.065	0.758	3.077
5-Deazafolate(a)	2DHF(a)	–0.788	–0.788	0.984	1.000
2-PGA(a)	2YPI(a)	–0.410	–0.456	0.843	0.899
L-Malate	3CSC	–0.271	–0.261	0.959	1.038
Ace-CoA	3CSC	0.333	0.253	0.770	1.316
FMN	3FXN	–0.674	–0.688	0.866	0.980
cAMP(a)	3GAP(a)	–0.708	–0.715	0.939	0.990
D-Malate	4CSC	–0.499	–0.489	0.972	1.020
Ace-CoA	4CSC	0.326	0.229	0.760	1.423
MVT-101	4HVP	–0.097	–0.094	0.870	1.032
L-700417(1)	4PHV	–0.232	–0.217	0.922	1.069
Peptide 1	4SGA	–0.772	–0.773	0.765	0.999
Ace-pep (2)	5HVP	–0.032	–0.034	0.901	0.941
Peptide 2	5SGA	–0.763	–0.766	0.753	0.996
$NADP^+$(part)	6DFR	–0.663	–0.668	0.918	0.993
Folate	7DFR	–0.626	–0.638	0.923	0.981
$NADP^+$	7DFR	–0.817	–0.824	0.878	0.992
JG-365	7HVP	–0.006*	0.010*	0.862	–0.600
U-85548e	8HVP	–0.198	–0.193	0.813	1.026

[a] All sites carry at least one formal charge. The legend is identical to that in Table 3.

apo-enzyme). This high degree of complementarity in electrostatic potential is retained in the r_{if}. In fact, for these 12 ligands only, the ratio r_w / r_{if} lies between 0.957 and 1.024, meaning that the correlation coefficient changes by less than ±5% when only the interface is considered.

Eighteen of these 35 ligand–protein complexes have both r_w and r_{if} significantly different from zero, but larger (i.e., less negative) than –0.7 (α-L-arabinose and 6ABP, SKB-Va and 1AAQ, SB204144 and 1HOS, MVT-101 and 4HVP, L-700417 and 4PHV, acetyl-pepstatin and 5HVP, U-85548e and 8HVP, AMP and 1AK3, biliverdin IX-γ and 1BBP, NADH and 1GD1, FMN and 1GOX, D-malate and 2CSC, L-malate and 3CSC, D-malate and 4CSC, 2-phosphoglycolate and 2YPI, FMN and 3FXN, incomplete $NADP^+$ and 6DFR, folate and 7DFR apo-enzyme). The r_w / r_{if} ratio in these cases ranges from 0.899 to 1.045. This shows that for less complementary ligands, the decrease in complementarity seems larger. However, the change is small and is still within the ±10% range.

Five of the 35 ligand–protein pairs exhibit some degree of electrostatic similarity (retinol and 1RBP, α-D-galactose and 5ABP, α-D-fucose and 7ABP, α-D-galactose and 8ABP, NAD^+ and 1LDM). The r_w / r_{if} ratio for these complexes is between 0.932 and 1.038. They are included to show that in some cases, the similarity is not affected by including only the interface.

Six ligand–protein pairs have very different values for r_w and r_{if}, i.e., carboxymethylcoenzyme A and 1CSC, carboxymethylcoenzyme A and 2CSC, FK506 and 1FKF, acetylcoenzyme A and 3CSC, acetylcoenzyme A and 4CSC, JG-365 and 7HVP. The cases of JG-365 and 7HVP and of FK-506 and 1FKF have no statistical significance. The remaining four cases are co-crystals containing citrate synthase. Two of them have very low r_{if} values, so although they are significantly different from zero, the correlation is too weak for us to draw any meaningful conclusion. The other two cases do not exhibit any electrostatic complementarity. In summary, using only the interfacial region does not appear to alter the complementarity significantly.

Effect of distance-dependent dielectric

In this study, the distance-independent dielectric was changed into a distance-dependent one. The interior of a protein is not a homogeneous dielectric, and it can be argued that the electrostatic complementarity would be reduced if the protein were a strong dielectric medium. Two kinds of distance-dependent dielectrics were used to investigate their effects on electrostatic potential complementarity.

Tables 5 and 6 show the effect of varying the dielectric linearly over a range of 10 Å, with the dielectric being ε_0 at 0 Å and increasing to a plateau of $4\varepsilon_0$ at ≥ 10 Å or to a plateau of $20\varepsilon_0$ at ≥ 10 Å. The results are for complementarity computed over the whole ligand surface. The correlation coefficient calculated for the first type of distance-dependent dielectric is defined to be r_4, while the correlation coefficient evaluated for the second type of dielectric is defined to be r_{20}.

Thirty of the ligand–protein pairs show an r_w / r_4 ratio of between 0.832 and 1.244 and an r_w / r_{20} ratio of between 0.756 and 1.468 (α-D-galactose and 5ABP, α-L-arabinose and 6ABP, α-D-fucose and 7ABP, α-D-galactose and 8ABP, SKB-Va and 1AAQ, SB204144 and 1HOS, L-700417 and 4PHV, U-85548e and 8HVP, AMP and 1AK3, biliverdin IX-γ and 1BBP, L-malate and 1CSC, D-malate and 2CSC, D-malate and 4CSC, folate and 1DHF, 5-deazafolate and 2DHF, FAD and 1FNR, NADH and 1GD1, FMN and 1GOX, oxamate(a) and 1LDM, FAD and 1PHH, 3,4-

TABLE 5
COMPARISON OF COMPLEMENTARITY WITH CONSTANT AND DISTANCE-DEPENDENT DIELECTRIC ON THE WHOLE LIGAND SURFACE[a]

Ligand	Receptor	r_w	r_4	r_w / r_4	r_{20}	r_w / r_{20}
FK506	1FKF	−0.125	0.023*	−5.435	0.059	−2.119
Retinol	1RBP	0.069	0.022*	3.136	−0.005*	−13.800
α-D-Galactose	5ABP	0.190	0.165	1.152	0.137	1.387
α-L-Arabinose	6ABP	−0.259	−0.251	1.032	−0.227	1.141
α-D-Fucose	7ABP	0.346	0.348	0.994	0.329	1.052
α-D-Galactose	8ABP	0.318	0.299	1.064	0.293	1.085

[a] None of the sites carry any formal charges.

$r_w = r$ calculated with constant ε; $r_4 = r$ calculated with distance-dependent dielectric which peaks at 4ε; $r_{20} = r$ calculated with distance-dependent dielectric which peaks at 20ε. An * indicates that the correlation is not significantly different from 0, $p < 0.0005$.

TABLE 6
COMPARISON OF COMPLEMENTARITY WITH CONSTANT AND DISTANCE-DEPENDENT DIELECTRIC ON THE WHOLE LIGAND SURFACE[a]

Ligand	Receptor	r_w	r_4	r_w / r_4	r_{20}	r_w / r_{20}
SKB-Va	1AAQ	−0.285	−0.286	0.997	−0.302	0.944
AMP(a)	1AK3(a)	−0.684	−0.650	1.052	−0.572	1.196
Biliverdin(a)	1BBP(a)	−0.334	−0.363	0.920	−0.345	0.968
L-Malate	1CSC	−0.714	−0.658	1.085	−0.582	1.227
Cm-CoA	1CSC	0.207	−0.018*	−3.726	−0.105	−1.971
Folate(a)	1DHF(a)	−0.771	−0.686	1.124	−0.615	1.254
FAD	1FNR	−0.796	−0.781	1.019	−0.727	1.095
NADH(o)	1GD1(o)	−0.660	−0.570	1.158	−0.510	1.294
FMN	1GOX	−0.490	−0.492	0.996	−0.486	1.008
SB204144(a)	1HOS	−0.208	−0.250	0.832	−0.275	0.756
NAD^+	1LDM	0.261	0.062	4.210	−0.067	−3.900
Oxamate(a)	1LDM	−0.904	−0.895	1.010	−0.876	1.032
FAD	1PHH	−0.713	−0.670	1.064	−0.592	1.204
3,4-DHB	1PHH	−0.855	−0.826	1.035	−0.732	1.168
2'-GMP	1RNT	−0.860	−0.819	1.050	−0.758	1.135
D-Malate	2CSC	−0.432	−0.393	1.099	−0.349	1.238
Cm-CoA	2CSC	0.200	−0.021*	−9.524	−0.109	−1.835
5-Deazafolate(a)	2DHF(a)	−0.788	−0.716	1.101	−0.661	1.192
2-PGA(a)	2YPI(a)	−0.410	−0.373	1.099	−0.323	1.270
L-Malate	3CSC	−0.271	−0.212	1.278	−0.141	1.922
Ace-CoA	3CSC	0.333	0.054	6.167	−0.082	−4.061
FMN	3FXN	−0.674	−0.542	1.244	−0.459	1.468
cAMP(a)	3GAP(a)	−0.708	−0.610	1.161	−0.554	1.278
D-Malate	4CSC	−0.499	−0.461	1.082	−0.415	1.202
Ace-CoA	4CSC	0.326	0.025*	13.040	−0.109	−2.991
MVT-101	4HVP	−0.097	−0.173	0.561	−0.235	0.413
L-700417(1)	4PHV	−0.232	−0.261	0.889	−0.258	0.899
Peptide 1	4SGA	−0.772	−0.681	1.134	−0.602	1.282
Peptide 2	5SGA	−0.763	−0.670	1.139	−0.593	1.287
Ace-pep(2)	5HVP	−0.032	−0.035*	0.914	−0.020*	1.600
$NADP^+$(part)	6DFR	−0.663	−0.553	1.199	−0.463	1.432
Folate	7DFR	−0.626	−0.602	1.040	−0.528	1.186
$NADP^+$	7DFR	−0.817	−0.744	1.098	−0.655	1.247
JG-365	7HVP	−0.006*	−0.025*	0.240	−0.048	0.125
U-85548e	8HVP	−0.198	−0.231	0.857	−0.230	0.861

[a] All sites carry at least one formal charge. The legend is identical to that in Table 5.

dihydroxybenzoate and 1PHH, 2'-GMP and 1RNT, 2-phosphoglycolate and 2YPI, FMN and 3FXN, cyclic AMP and 3GAP, ace-Pro-Ala-Pro-Phe-OH and 4SGA, ace-Pro-Ala-Pro-Tyr-OH and 5SGA, incomplete $NADP^+$ and 6DFR, folate and 7DFR, $NADP^+$ and 7DFR). Of these 30 complexes, 12 exhibit an r_w of less than −0.7 (L-malate and 1CSC, folate and 1DHF, 5-deazafolate and 2DHF, FAD and 1FNR, oxamate(a) and 1LDM, FAD and 1PHH, 3,4-dihydroxybenzoate and 1PHH, 2'-GMP and 1RNT, cyclic AMP and 3GAP, ace-Pro-Ala-Pro-Phe-OH and 4SGA, ace-Pro-Ala-Pro-Tyr-OH and 5SGA, $NADP^+$ and 7DFR). The r_w / r_4 ratios for these complexes are between 1.010 and 1.139, and the r_w / r_{20} ratios between 1.095 and 1.287. There is a small but clear decrease in complementarity when the homogeneous dielectric is changed to a distance-dependent one.

Fifteen of these 30 ligand–protein pairs exhibit r_w values in the range from −0.7 to 0 (α-L-arabinose and 6ABP, SKB-Va and 1AAQ, SB204144 and 1HOS, L-700417 and 4PHV, U-85548e and 8HVP, AMP and 1AK3, biliverdin IX-γ and 1BBP, D-malate and 2CSC, D-malate and 4CSC, NADH and 1GD1, FMN and 1GOX, 2-phosphoglycolate and 2YPI, FMN and 3FXN, incomplete $NADP^+$ and 6DFR, folate and 7DFR). The range of the r_w/r_4 ratio for these complexes is between 0.832 and 1.244, while the range for the r_w/r_{20} ratio is 0.756–1.432. It can be seen that the range is larger than for the more complementary ligands. Moreover, there is a decrease as well as an increase in the complementarity. This increase is usually observed in complexes where $-0.3 < r_w < 0$, where no substantial electrostatic potential complementarity exists.

Amongst these 30 complexes, there are three where $r_w > 0$, namely, α-D-galactose and 5ABP, α-D-fucose and 7ABP, α-D-galactose and 8ABP. The range of r_w/r_4 is 0.994 to 1.152, while that of r_w/r_{20} is 1.052 to 1.387. It appears that electrostatic similarity between the ligand and the site is not changed to a significant degree by the change in permittivity.

Eleven of the original 41 ligand–protein pairs have very large positive or negative r_w/r_4 or r_w/r_{20} ratios. In seven of these cases (FK506 and 1FKF, retinol and 1RBP, carboxymethylcoenzyme A and 1CSC, carboxymethylcoezyme A and 2CSC, acetylcoenzyme A and 4CSC, acetyl-pepstatin and 5HVP, JG365 and 7HVP), at least one of the correlation coefficients is not significantly different from zero. Of the remaining four cases, two have very small correlation coefficients (NAD^+ and 1LDM, MVT-101 and 4HVP). The other two are from citrate synthase data sets (L-malate and 3CSC, acetylcoenzyme A and 3CSC); the reasons for this observation are not known.

It can be concluded that the effect of using a distance-dependent dielectric is to decrease moderately the electrostatic complementarity between ligands and receptor sites. However, the general pattern of complementarity is not abolished by the change in permittivity.

DISCUSSION

Electrostatic complementarity between ligands and their sites has been postulated to exist for some time. Weiner and co-workers [37] calculated the electrostatic potential on the surface of a number of molecules, and by dividing the potential into several bands they were able to visualize complementarity. Nevertheless, they did not attempt to quantify this electrostatic complementarity.

Nakamura and co-workers [2,38] were the first to attempt quantification. Nevertheless, their method suffers from one drawback. They defined a complementarity parameter which gives a falsely high complementarity value when one of the pairs of potentials is strongly negative or positive, and the other one close to zero (see Ref. 4 for a detailed discussion).

The work described here has important implications for ligand design. We have demonstrated that it is electrostatic potential complementarity, not partial charges complementarity, that exists between many ligands and their receptor sites. This is consistent with the fact that electrostatic interactions are long-range ones; they vary as $1/r$, and so the complementarity is not limited to local partial charges, but it also includes an effect from partial charges further away. In ligand design, it is important to consider the ligand–receptor interaction as a whole, and not divide it into local interactions.

Moreover, we have also demonstrated that electrostatic potential complementarity does not change significantly if only the interface is taken into account. This means that if a lead com-

pound is found for a receptor, it is possible to modify it without necessarily sacrificing any electrostatic complementarity. The extra surface created need not matter if the additional atoms do not grossly change the potential at the interface.

The electrostatic potential complementarity is decreased by changing the homogeneous dielectric at $\varepsilon=\varepsilon_0$ to a distance-dependent one, but the pattern of complementarity is still observed. This shows that the electrostatic potential complementarity is an effect observed independent of the dielectric of the medium.

We used CNDO/2 partial charges [30] and the VSS program [32] to calculate electrostatic potentials in this work. It would be interesting to study these co-crystals using an alternative charge model and including the polarization effects.

In this work, we have established the existence, and examined the properties, of electrostatic potential complementarity. The effect of solvation on ligand–receptor interactions is still poorly understood, and therefore we resist the temptation to relate electrostatic complementarity to the total energy of interaction. We would now like to turn our attention from studying electrostatic complementarity to examining how complementarity is generated. In many ligand design strategies, the ligand is built up from constituent fragments. Thus, the next questions we would like to ask are these: what are the relative contributions of different parts of the ligand to this electrostatic potential complementarity? How can we generate optimal electrostatic potential complementarity from chemical structures? In the next two papers in this series, we attempt to answer these questions.

ACKNOWLEDGEMENTS

We thank Dr. M.T. Barakat for making her programs available. 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) 527.
2 Chau, P.-L. and Dean, P.M., J. Comput.-Aided Mol. Design, 8 (1994) 545.
3 Nakamura, H., Komatsu, K., Nakagawa, S. and Umeyama, H., J. Mol. Graphics, 3 (1985) 2.
4 Namasivayam, S. and Dean, P.M., J. Mol. Graphics, 4 (1986) 46.
5 Van Duyne, G.D., Standaert, R.F., Karplus, P.A., Schreiber, S.L. and Clardy, J., Science, 252 (1991) 839.
6 Cowan, S.W., Newcomer, M.E. and Jones, T.A., Protein Struct. Funct. Genet., 8 (1990) 44.
7 Vermersch, P.S., Lemon, D.D., Tesmer, J.J.G. and Quiocho, F.A., Biochemistry, 30 (1991) 6861.
8 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.
9 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.
10 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.
11 Bone, R., Vacca, J.P., Anderson, P.S. and Holloway, M.K., J. Am. Chem. Soc., 113 (1991) 9382.

12 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.
13 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.
14 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.
15 Diederichs, K. and Schulz, G.E., J. Mol. Biol., 217 (1991) 541.
16 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.
17 Karpusas, M., Holland, D. and Remington, S.J., Biochemistry, 30 (1991) 6024.
18 Davies, J.F., Delcamp, T.J., Prendergast, N.J., Ashford, V.A., Freisheim, J.H. and Kraut, J., Biochemistry, 29 (1990) 9467.
19 Karplus, P.A., Daniels, M.J. and Herriott, J.R., Science, 251 (1991) 60.
20 Skarżyński, T., Moody, P.C.E. and Wonacott, A.J., J. Mol. Biol., 193 (1987) 171.
21 Lindqvist, Y., J. Mol. Biol., 209 (1989) 151.
22 Abad-Zapatero, C., Griffith, J.P., Sussman, J.L. and Rossmann, M.G., J. Mol. Biol., 198 (1987) 445.
23 Schreuder, H.A., Van der Laan, J.M., Hol, W.G.J. and Drenth, J., J. Mol. Biol., 199 (1988) 637.
24 Arni, R., Heinemann, U., Tokuoka, R. and Saenger, W., J. Biol. Chem., 263 (1988) 15358.
25 Lolis, E. and Petsko, G.A., Biochemistry, 29 (1990) 6619.
26 Smith, W.W., Burnett, R.M., Darling, G.D. and Ludwig, M.L., J. Mol. Biol., 117 (1977) 195.
27 Weber, I.T. and Steitz, T.A., J. Mol. Biol., 198 (1987) 311.
28 James, M.N.G., Sielecki, A.R., Brayer, G.D., Delbaere, L.T.J. and Bauer, C.-A., J. Mol. Biol., 144 (1980) 43.
29 Bystroff, C., Oatley, S.J. and Kraut, J., Biochemistry, 29 (1990) 3263.
30 Dobosh, P.A., QCPE Program No. 141, Bloomington, IN, 1968.
31 Scheraga, H.A., QCPE Program No. 286, Bloomington, IN, 1975.
32 Giessner-Prettre, C., QCPE Program No. 249, Bloomington, IN, 1974.
33 Chau, P.-L. and Dean, P.M., J. Mol. Graphics, 5 (1987) 97.
34 Lee, B. and Richards, F.M., J. Mol. Biol., 55 (1971) 379.
35 Press, W.H., Teukolsky, S.A., Vetterling, W.T. and Flannery, B.P., Numerical Recipes in FORTRAN, 2nd ed., Cambridge University Press, Cambridge, 1992, pp. 630 and 633.
36 Riggs, D.S., Guarnieri, J.A. and Addelman, S., Life Sci., 22 (1978) 1305.
37 Weiner, P.K., Langridge, R., Blaney, J.M., Schaefer, R. and Kollman, P.A., Proc. Natl. Acad. Sci. USA, 79 (1982) 3754.
38 Ito, W., Nakamura, H. and Arata, Y., J. Mol. Graphics, 7 (1989) 60.