Journal of Computer-Aided Molecular Design, 9 (1995) 448–456
ESCOM

J-CAMD 300

The atom assignment problem in automated de novo drug design. 4. Tests for site-directed fragment placement based on molecular complementarity

M.T. Barakat and P.M. Dean*

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

Received 1 February 1995
Accepted 23 May 1995

Keywords: De novo drug design; Molecular electrostatic potential; Molecular complementarity; Molecular similarity

Summary

Three previous papers in this series have outlined an optimization method for atom assignment in drug design using fragment placement. In this paper the procedure is rigorously tested on a selection of five ligand–protein co-crystals. The algorithm is presented with the molecular graph of the ligand, and the electrostatic/hydrophobic potential of the site, with the aim of creating a placement on the molecular graph which is as electrostatically complementary or hydrophobically similar to the site as possible. Various designer options were tested, including, where appropriate, hydrogen bonding and a restricted number of halogens. In most cases, the placement obtained was at least as good as the native ligand, if not significantly better.

Introduction

An atom assignment method has been outlined in three previous papers as part of an automated de novo drug design strategy [1–3]. The method is based on a graph perception technique for placing atoms onto 3D molecular skeletons. The objective function utilises small fragments to convert the skeleton into a molecular structure, using either complementarity or similarity for combinatorial placement. The constraints for placement are electrostatic potential, hydrogen-bond type and hydrophobicity. The objective function is optimized by simulated annealing. Extensive self-placement tests have been performed to test the methodology [3]. If a skeleton with an associated van der Waals surface is given, together with a property computed from the target molecule projected onto that surface, how good is the algorithm at consistently recreating the original structure? This is an unambiguous test of the method. Inconsistencies were predominantly found for multiple halogen placements with some test structures; these failings could be rectified by imposing a limit on the number of halogens allowed in the placement. The procedure must now be tested further: (i) by tackling the more difficult situation of placing fragments onto a molecular graph such that the resulting surface potential of the molecule created will be complementary (or similar in the case of hydrophobic potential) to the potential of the site; and (ii) by taking the super-surface envelope of similar molecules and placing atoms onto a molecular graph contained within the envelope by the creation of a potential similar to the envelope. Our objective is to make a careful analysis of the performance of the strategy for atom assignment used in this work. The complementarity of the natural ligand provides a reference point with which to compare novel atom assignments. This paper considers the first of these tests; the second test approach is presented in the following paper [4].

A suitable test model is a ligand from a Brookhaven Protein Data Bank (PDB) co-crystal, where the ligand can be treated as the molecular graph on which fragment placement will be performed. The van der Waals surface of the ligand is generated, and the electrostatic (or hydrophobic where available) potential from the protein is projected onto it; a new molecule is created whose surface potential will complement that of the protein. It is important to note that there is no 'correct answer' for this test case, for the following reasons:

(1) The native ligand may not be the most complemen-

*To whom correspondence should be addressed.

tary molecule, and so the target aim of reproducing the original ligand does not necessarily apply for many sites;

(2) the measure of complementarity/similarity between the new molecule created and the site depends on the method of assessment used (e.g., Pearson's or Spearman's correlation coefficient): a high value in one coefficient may correspond to a lower value in another;

(3) without calculating the actual charges of the new arrangement of atoms on the graph, the structure may fall into the small group of molecules whose residual charge predictability with fragment charges is poor, thereby giving a false correlation with the site potential.

Bearing these caveats in mind, five PDB co-crystals have been used to study the performance of the fragment placement algorithm. Comparison is needed between the results generated in the current paper and those considered in the previous paper [3].

Methods and Results

If a hydrogen-bonding heavy atom is present within 3.6 Å of a heavy atom in the native ligand, it is assumed that there is a functional hydrogen bond with the site, and an electronegative atom is specified at the vertex of the molecular graph which can contribute to a hydrogen bond.

In the annealing process, it is computationally favourable to have as small a number of surface points as possible. The tessellation frequency for all graph surfaces was set to 1, since it has already been noted that a tessellation frequency of 1 gives statistically equivalent results to a frequency of 5 [3]. Different combinations of options were investigated; in particular, the performance of three assessment methods of the objective function (either electrostatic potential complementarity or hydrophobic potential similarity) were assessed. The three correlating methods were Pearson's correlation coefficient (r), Spearman's rank correlation coefficient (r_s) and the summation of absolute pair sums for complementarity (or absolute pair differences for similarity) of potential (*eo*). With each set of conditions, 10 replicate runs were performed on a Sun SPARCstation IPX. In some cases, the algorithm failed to converge due to locking of the algorithm in a local minimum in which the interfragment connectivity is imperfect, while the temperature is too low to replace the offending fragments.

It is important to emphasize that the results listed in tables for each of the following sections comprise the direct output of the program for placement. The correlation coefficients represent the correlation between the potential set that is given as input, and the potential set calculated in the program from fragment residual charges or atomic hydrophobic parameters. In the complementarity runs, new atom assignments may result in more complementary structures than the native ligands.

Further processing of the best structures for each co-crystal has been performed, and involves recalculation of the residual charges (CNDO) on new atom arrangements. The true correlation coefficients with the site are then calculated using the surface of the new structure (the tessellation frequency is now set to 5). However, no conformational minimization, apart from the assignment of standard bond lengths and repositioning back onto the original graph, has been performed.

Testing with co-crystals from the Brookhaven PDB

The set of ligand co-crystals has been described previously [3].

1AK3 and AMP

The crystal structure of the complex between bovine mitochondrial matrix adenylate kinase (1AK3) and its substrate AMP has been refined at 1.85 Å resolution [5]. Some protein amino acids were defined as charged, namely Arg^{41}, Lys^{62} and Arg^{92}. Hydrogen atoms were added, and amino acid residual charges were assigned to the atoms of the protein. For the hydrogen-bonding information when the site was considered, heavy atoms from the site were located within 3.6 Å of AMP heavy atoms O4 (Arg^{41} $N_{\eta 1}$, 2.65 Å), O12 (Lys^{62} O, 2.65 Å), N16 (Ser^{36} O_γ, 2.86 Å; Gly^{89} O, 3.38 Å; Arg^{92} $N_{\eta 2}$, 3.48 Å), N19 (Gln^{96}

TABLE 1
ANNEALING RESULTS USING AMP AS A MOLECULAR GRAPH IN THE 1AK3 SITE UNDER VARYING ANNEALING CONDITIONS

Method of calculation	Method of assessment	Mean value of the final objective function	Time (s)	Ratio perfect runs/ total runs	New atoms
Electrostatic potential	r	−0.765±0.005	34.2±1.6	9/10	2
	r_s	−0.790±0.007	47.7±2.5	3/9	2.3±0.5
	eo	42 540±34 930	31.2±6.6	0/10	3.7±2.0
Electrostatic potential with hydrogen bonds specified	r	−0.767±0.000	35.5±6.9	10/10	0
	r_s	−0.792±0.007	44.0±6.2	6/10	1.0±0.0
	eo	37 420±261.8	24.9±3.0	8/10	1.0±0.0

The values in each row represent averages over 10 runs for a given set of conditions; where no convergence was reached, the values were omitted from the analysis. The fifth column represents the number of runs in which AMP itself was reproduced, while the sixth column represents the average number of atoms which differ from AMP in the remaining runs. *eo* refers to the summation of absolute sums in complementarity runs.

TABLE 2
ANNEALING RESULTS USING cAMP_a AS A MOLECULAR GRAPH IN THE 3GAP SITE UNDER VARYING ANNEALING CONDITIONS

Method of calculation	Method of assessment	Mean value of the final objective function	Time (s)	Ratio perfect runs/ total runs	New atoms
Electrostatic potential	r	−0.581 ± 0.005	40.0 ± 10.1	0/10	1.0 ± 0.0
	r_s	−0.606 ± 0.007	47.9 ± 10.7	0/10	1.0 ± 0.0
	eo	36 440 ± 328.5	44.2 ± 10.0	0/10	2.6 ± 0.5
Electrostatic potential with hydrogen bonds specified	r	−0.579 ± 0.005	47.2 ± 5.0	0/10	1.0 ± 0.0
	r_s	−0.596 ± 0.023	61.8 ± 5.5	2/10	1.0 ± 0.0
	eo	33 950 ± 464.6	51.1 ± 7.4	9/10	1

See the footnote to Table 1 for an explanation of the parameters.

$O_{\varepsilon 1}$, 2.75 Å; Gly^{89} O, 2.93 Å; Gln^{96} $N_{\varepsilon 2}$, 3.53 Å), N20 (Gln^{96} $N_{\varepsilon 2}$, 2.85 Å) and N22 (Ile^{64} N, 3.3 Å); these AMP vertices were specified as requiring electronegative atoms involved in hydrogen bonding. The number of surface points used was 136. The electrostatic potential complementarity values for AMP and its site were computed using a tessellation frequency of 1.

The results for the annealing of fragments onto the molecular graph of AMP are shown in Table 1. In some of the runs, hydrogen-bonding vertices were specified using information from the site (electronegative atoms specified in the graph). For each set of conditions (e.g., using Pearson's correlation coefficient, and taking hydrogen bonding into account), 10 runs were performed.

In the case of self-placement, the program successfully replicated AMP using either Pearson's or Spearman's correlation coefficients as the method of assessment of similarity [3]. Using the site electrostatic potential, the results suggested that AMP itself was the molecule most complementary to the site. This is well illustrated by the 10 replicates when hydrogen-bonding information was given. In the run employing Spearman's rank correlation coefficient, one structure (having a chloro-adenine) seemed at first to be more complementary, with a Spearman's rank of −0.796 using the fragment charges. However, once charges were recalculated, the correlations on the new surface with a tessellation frequency of 5 dropped to values below those of the native AMP: new values for the chloro-adenine derivative were $r = -0.709$, $r_s = -0.740$ and $\tau = -0.535$; AMP correlations were $r = -0.717$, $r_s = -0.751$ and $\tau = -0.546$. The error function complement (erfc) for the two Pearson's correlation coefficients was 0.554. This non-zero value suggested that the complementarities of AMP and its chloro-adenine derivative to 1AK3 were not statistically different. If hydrogen bonds are specified on input, the native ligand is reproduced more frequently, the near misses being chloroderivatives. Algorithm CPU times were comparable with those of the self-placement tests.

3GAP and cAMP

The structure of the *Escherichia coli* catabolite gene activator protein (3GAP) was used for complementarity tests with cAMP [6]. Three amino acids were charged in both the A and B subunits of the protein: Glu^{72}, Arg^{82} and Arg^{123}. Electronegative heavy-atom contacts within 3.6 Å of the protein for cAMP_a included O2 ($SerA^{83}$ O_γ, 2.96 Å), O3 ($ArgA^{82}$ $N_{\eta 1}$, 3.04 Å), O11 ($GlyA^{71}$ N, 3.06 Å; $GluA^{72}$ $O_{\varepsilon 2}$, 2.42 Å) and N18 ($SerB^{128}$ O_γ, 2.44 Å). The corresponding vertices were defined as electronegative contributors to hydrogen bonding in the molecular graphs of cAMP_a in the complementarity tests. Using a tessellation frequency of 1, 114 surface points were found for cAMP_a. The annealing results for the molecular graph of cAMP_a are shown in Table 2.

The self-placement results showed that cAMP_a could be easily replicated [3]; thus, the fragment placement method was successful for these graphs. Complementarity

TABLE 3
ANNEALING RESULTS USING FOLATE AS A MOLECULAR GRAPH AND 1DHF AS THE SITE UNDER VARYING ANNEALING CONDITIONS

Method of calculation	Method of assessment	Mean value of the final objective function	Time (s)	Ratio perfect runs/ total runs	New atoms
Electrostatic potential	r	−0.789 ± 0.045	81.4 ± 8.3	0/10	10.7 ± 3.0
	r_s	−0.832 ± 0.041	102.4 ± 13.0	0/10	7.5 ± 2.5
	eo	19 280 ± 3110	54.9 ± 7.0	0/10	5.8 ± 1.3
Electrostatic potential with hydrogen bonds specified	r	−0.767 ± 0.047	79.2 ± 5.1	0/10	9.9 ± 2.8
	r_s	−0.858 ± 0.043	106.8 ± 13.5	0/10	8.7 ± 3.4
	eo	21 280 ± 3606	57.5 ± 5.8	0/10	6.0 ± 1.8

See the footnote to Table 1 for an explanation of the parameters.

derived placements were slightly faster than self-placement tests. The runs using the electrostatic potential of the site yielded mostly a chloro-adenine derivative of cAMP. In the case of one of the best placements for the cAMP_a molecular graph, the correlations and regression values with the site using re-calculated charges and a tessellation frequency of 5 were: $r=-0.717$, $r_s=-0.782$, $\tau=-0.575$, gradient $=-1.675$ and intercept $=698.5$ kJ mol^{-1} (compare the native cAMP_a values of $r=-0.656$, $r_s=-0.691$, $\tau=-0.499$, gradient $=-1.962$ and intercept $=852.8$ kJ mol^{-1}). The error function complement is close to zero (5.91×10^{-5}), indicating that the complementarity between the two structures and the site is significantly different. The hydrogen-bond option had a marked effect only on the absolute difference summation procedure.

1DHF and folate

The crystal structure of recombinant human dihydrofolate reductase (1DHF) with one of its substrates, folate (FOL), has been solved at 2.3 Å resolution [7]. The A structure was used in this work, and the following amino acids were assigned charges in the protein: Arg^{28}, Glu^{30}, Arg^{70}, Lys^{108} and Asp^{110}. When adjacent heavy atoms in the protein were considered, only three vertices in folate were found to be within 3.6 Å: N3 (Glu^{30} $O_{\varepsilon1}$, 2.68 Å), N4 (Glu^{30} $O_{\varepsilon2}$, 2.77 Å), and O31 (Arg^{70} $N_{\eta2}$, 2.93 Å), which were subsequently specified as electronegative vertices in the complementarity placement test. The number of surface points used was 181. The results of the annealing runs are shown in Table 3.

The runs with the site produced some interesting results. The best structure generated (using Pearson's correlation coefficient for assessment) had its charges recalculated, and the true correlation coefficients with the site, using a tessellation frequency of 5, were: $r=-0.770$, $r_s=-0.806$ and $\tau=-0.614$. This was a marginal improvement (in terms of r and τ) compared to the native ligand, folate, which had the following correlations: $r=-0.726$, $r_s=-0.815$ and $\tau=-0.593$. The erfc for the two Pearson's correlation coefficients was 2.52×10^{-5}, showing that the difference between them is statistically significant. However, the new structure differs from folate in having 10 chlorines, an aldehyde group instead of one of the carboxylates, a carbonyl instead of group N14-H47, and a

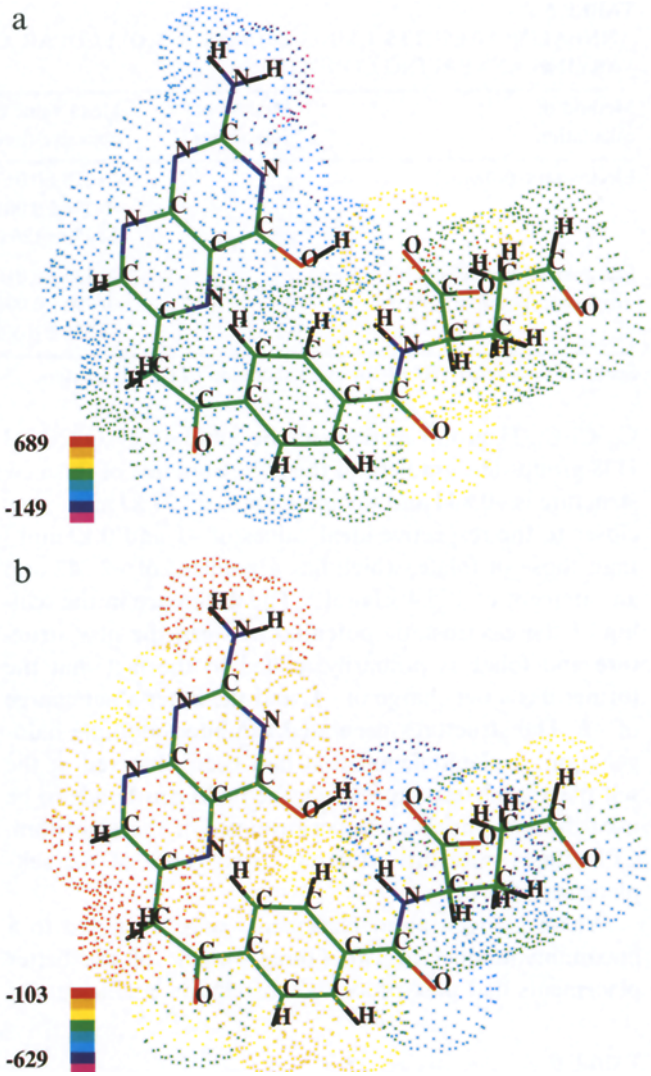

Fig. 1. An electrostatically complementary structure in the site of the protein 1DHF. Values of potential are in kJ mol^{-1}. A tessellation frequency of 5 was used for the van der Waals surface, giving 3541 points. (a) The electrostatic potential from 1DHF projected onto the surface of the new structure; (b) the electrostatic potential from the new structure projected onto the same surface. $r=-0.768$, $r_s=-0.807$ and $\tau=-0.617$.

TABLE 4
ANNEALING RESULTS USING FOLATE AS A MOLECULAR GRAPH WITH THE 1DHF SITE UNDER VARYING ANNEALING CONDITIONS

Method of calculation	Method of assessment	Mean value of the final objective function	Time (s)	Ratio perfect runs/ total runs	New atoms
Electrostatic potential	r	−0.770±0.031	65.8±7.5	0/10	3.8±1.0
	r_s	−0.820±0.019	91.9±7.8	0/10	2.2±1.6
	eo	18 080±1972	52.5±8.6	0/10	3.8±1.3
Electrostatic potential with hydrogen bonds specified	r	−0.785±0.012	72.8±6.6	0/10	3.0±0.9
	r_s	−0.836±0.019	91.5±12.0	0/10	1.9±0.9
	eo	18 410±1391	59.5±7.9	0/10	2.6±1.5

See the footnote to Table 1 for an explanation of the parameters. The number of halogen atoms to be added was limited to a maximum of two.

TABLE 5
ANNEALING RESULTS USING DHB AS A MOLECULAR GRAPH AND ITS SITE COMPOSED OF 1PHH AND FAD UNDER VARYING ANNEALING CONDITIONS

Method of calculation	Method of assessment	Mean value of the final objective function	Time (s)	Ratio perfect runs/ total runs	New atoms
Electrostatic potential	r	−0.828±0.017	4.3±1.0	4/10	1.0±0.0
	r_s	−0.796±0.001	4.7±0.7	0/10	1.0±0.0
	eo	14470±126.0	2.9±0.3	9/10	3
Electrostatic potential with hydrogen bonds specified	r	−0.835±0.016	4.5±0.6	6/10	1.0±0.0
	r_s	−0.796±0.000	3.7±0.5	0/10	1.0±0.0
	eo	14520±165.6	3.0±0.2	8/10	3.0±0.0

See the footnote to Table 1 for an explanation of the parameters.

$C_{sp^2}Cl$-$C_{sp^2}H$ group instead of the amide C21 O22-N23 H48 group. The gradient of the regression line of this new structure is −0.844 and the intercept is 27.77 kJ mol^{-1}, i.e., closer to the respective ideal values of −1 and 0 kJ mol^{-1} than those of folate, which has a gradient of −2.242 and an intercept of 175.4 kJ mol^{-1}. The difference in the scaling of the electrostatic potential between the new structure and folate is primarily related to the fact that the former has a net charge of −1, and the latter a net charge of −2. The structures generated exhibited multiple halogenation, a phenomenon that had been observed in the self-placement tests; specifying the hydrogen bonds to be used did not improve the performance of the algorithm. CPU times were little different from those of the self-placement tests.

Where halogen atom placements were restricted to a maximum number of two (Table 4), one of the better placements had post-processed correlation coefficients of $r=-0.768$, $r_s=-0.807$, and $\tau=-0.617$, a gradient of −1.066 and an intercept of 81.09 kJ mol^{-1}, i.e., values comparable both to those of the chlorinated placement ($r=-0.770$, $r_s=-0.806$, $\tau=-0.614$, gradient = −0.844 and intercept = 27.77 kJ mol^{-1}) and to those of the native ligand folate ($r=-0.726$, $r_s=-0.815$, $\tau=-0.593$, gradient = −2.242 and intercept = 175.4 kJ mol^{-1}). When the error function complement was calculated for the Pearson's correlation coefficients, the new structure was found to be significantly different to folate ($erfc = 5.18\times10^{-5}$), but statistically equivalent to the halogenated placement ($erfc = 0.805$). The nonhalogenated structure has an aldehyde group instead of one of the carboxylates of folate, and a carbonyl instead of the N14-H47 of folate (Fig. 1). The net charge of −1 has resulted in a gradient and intercept which are approximately half those for the native folate, which has a net charge of −2. The same is true of the multiply halogenated structure.

TABLE 6
ANNEALING RESULTS USING RETINOL AS A MOLECULAR GRAPH AND 1RBP AS THE SITE

Method of calculation	Method of assessment	Mean value of the final objective function	Time (s)	Ratio perfect runs/ total runs	New atoms
Electrostatic potential	r	−0.664±0.054	300.6±15.3	0/10	18.0±2.5
	r_s	−0.811±0.079	396.0±40.8	0/10	16.0±1.4
	eo	56930±5398	245.4±41.5	0/10	20.1±2.4
Hydrophobic potential	r	0.157±0.045	206.3±13.5	0/10	12.2±1.4
	r_s	0.113±0.046	325.6±18.9	0/10	12.2±2.3
	eo	164.8±22.24	148.2±11.3	0/10	6.1±2.2
Electrostatic potential and hydrophobic potential	electrostatic 0.75 with hydrophobic 0.25; r	−0.491±0.201	238.7±11.8	0/10	14.7±2.4
	electrostatic 0.5 with hydrophobic 0.5; r	0.186±0.078	203.6±12.4	0/10	12.5±2.0
	electrostatic 0.25 with hydrophobic 0.75; r	0.148±0.077	192.4±11.1	0/10	11.8±2.1

See the footnote to Table 1 for an explanation of the parameters. Alternative objective functions using hydrophobic potentials and mixed electrostatic and hydrophobic potentials are shown.

1PHH+FAD and DHB

The crystal structure of the enzyme *p*-hydroxybenzoate hydroxylase (1PHH) from *Pseudomonas fluorescens* complexed with its product, 3,4-dihydroxybenzoate (DHB), and FAD has been determined at a resolution of 2.3 Å [8]. The amino acids in the protein which were assigned charges were: Arg^{42}, Arg^{44} and Arg^{214}. With the site, the 3.6 Å heavy-atom contacts were: O4 (Pro^{293} O, 2.37 Å), O6 (Tyr^{201} O_η, 2.48 Å), O10 (Tyr^{222} O_η, 2.56 Å) and O11 (Ser^{212} O_γ, 2.69 Å; Arg^{214} $N_{\eta 2}$, 2.69 Å). These vertices were defined as electronegative contributors to hydrogen bonding in the complementarity. The complementarity values were found for DHB and the combined electrostatic potential from 1PHH and FAD using a tessellation frequency of 1. The results of the annealing runs using the DHB molecular graph are shown in Table 5.

In the complementarity runs, one of the better structures (with a fluorine atom in the position of H12 of DHB) gave the following correlation coefficients after obtaining recalculated charges and employing a tessellation frequency of 5: $r=-0.877$, $r_s=-0.792$ and $\tau=-0.606$. These values were marginally better than those of the native ligand DHB ($r=-0.861$, $r_s=-0.765$ and $\tau=-0.581$), although the difference between the two Pearson's correlation coefficients was not significant (erfc = 0.105). The regression results for the new structure and DHB were similar: the new structure had a gradient of −1.644 and an intercept of 647.1 kJ mol^{-1}, while the corresponding values for DHB were −1.481 and 544.8 kJ mol^{-1}. The specification of hydrogen bonds did not materially affect the results. CPU times were comparable to those found in the self-placement tests.

1RBP and retinol

Human serum retinol-binding protein (1RBP) has been crystallized with retinol, and the structures have been refined at 2 Å resolution [9]. The ligand retinol is neutral and mostly a hydrocarbon, except for a terminal hydroxyl group. No amino acids in the protein were charged. No heavy atom in the protein was found within 3.6 Å of the ligand, and so no hydrogen-bond information on the retinol graph was specified in the complementarity testing of the placement algorithm.

The annealing runs against both the electrostatic potential and the hydrophobic potential of the site are given in Table 6. The electrostatic complementarity correlations for the native retinol with the site, using a tessellation frequency of 1, were uncorrelated, while the unexpected hydrophobic correlations were: $r=-0.235$, $r_s=-0.204$ and *eo* (summation of absolute pair differences) = 115.4. Curiously, in the case of electrostatic potential complementarity, one of the best structures produced (Fig. 2) was a zwitterion, with both a sulphate and a quaternary nitrogen group; it also had two trifluoro groups and two chlorines. With recalculated charges from this structure and a tessellation frequency of 5, the electrostatic complementarity with the site was markedly elevated to: $r=-0.756$, $r_s=-0.738$ and $\tau=-0.545$. The gradient and intercept of the regression line of this new structure, however, were high: gradient = −7.563 and intercept = 2193 kJ mol^{-1}.

TABLE 7
ANNEALING RESULTS USING RETINOL AS A MOLECULAR GRAPH AND 1RBP AS THE SITE WITH RESTRICTIONS

Method of calculation	Method of assessment	Mean value of the final objective function	Time (s)	Ratio perfect runs/ total runs	New atoms
Electrostatic potential	*r*	−0.243 ± 0.140	229.2 ± 78.3	0/10	6.4 ± 1.3
	r_s	0.037 ± 0.239	309.4 ± 42.4	0/10	5.8 ± 1.8
	eo	56 833.5 ± 2521.7	188.3 ± 56.3	0/10	7.6 ± 1.6
Hydrophobic potential	*r*	−0.101 ± 0.039	246.4 ± 69.7	0/10	5.4 ± 1.4
	r_s	−0.040 ± 0.037	440.9 ± 173.1	0/10	6.6 ± 1.9
	eo	155.2 ± 18.53	184.7 ± 55.3	0/10	6.2 ± 1.1
Electrostatic potential and hydrophobic potential	electrostatic 0.75 with hydrophobic 0.25; *r*	−0.088 ± 0.165	316.5 ± 123.2	0/10	6.5 ± 2.6
	electrostatic 0.5 with hydrophobic 0.5; *r*	−0.120 ± 0.039	286.1 ± 54.5	0/8	6.3 ± 2.3
	electrostatic 0.25 with hydrophobic 0.75; *r*	−0.157 ± 0.067	212.9 ± 50.8	0/7	4.6 ± 1.9

The algorithm was given two restrictions: (i) the number of halogen atoms was limited to two; (ii) only neutral fragments could be placed. See the footnote to Table 1 for an explanation of the parameters. Alternative objective functions using hydrophobic potentials and mixed electrostatic and hydrophobic potentials are shown.

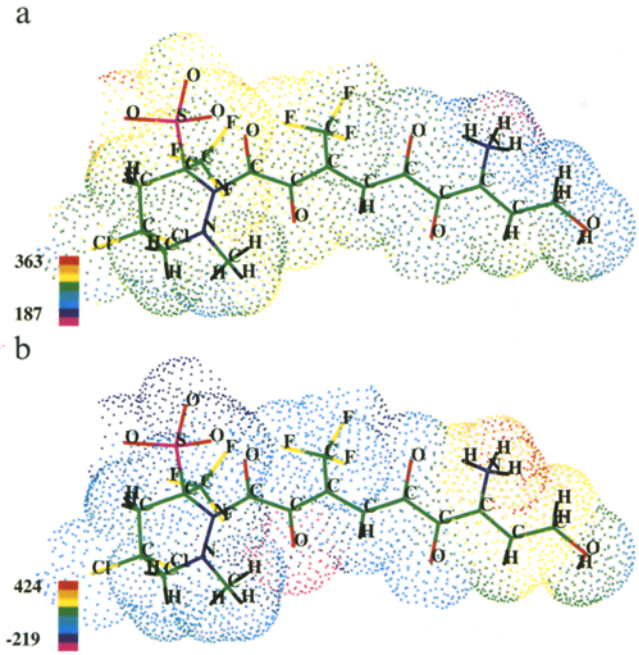

Fig. 2. A structure that is more electrostatically complementary to retinol in the site of the protein 1RBP. Values of potential are in kJ mol^{-1}. A tessellation frequency of 5 was used for the van der Waals surface, giving 3525 points. (a) The electrostatic potential from 1RBP projected onto the surface of the new structure; (b) the electrostatic potential from the new structure projected onto the same surface. $r=-0.756$, $r_s=-0.738$ and $\tau=-0.545$.

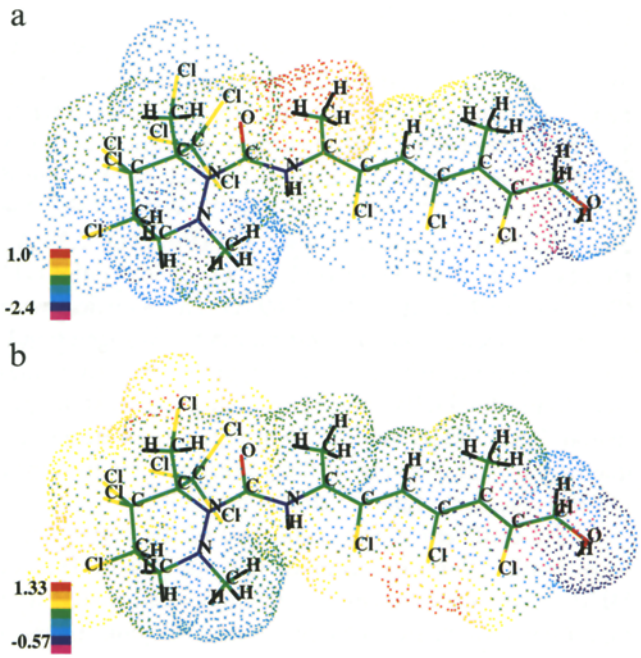

Fig. 3. A structure that is more hydrophobically similar to retinol in the site of the protein 1RBP. A tessellation frequency of 5 was used for the van der Waals surface, giving 3273 points. (a) The hydrophobic potential (Φ) from 1RBP projected onto the surface of the new structure; (b) the hydrophobic potential (Φ) from the new structure projected onto the same surface. $r=0.341$, $r_s=0.377$ and $\tau=0.228$. The potential values are in –kJ mol^{-1}.

These values reflected the relatively higher positive electrostatic potential of the new structure compared to that of the site.

The hydrophobicity annealing results with the site were not as good as those of the electrostatic runs. One of the better structures generated in terms of hydrophobic similarity, using a surface with a tessellation frequency of 5, had correlation and regression values of: $r=0.341$, $r_s=0.377$, $\tau=0.228$, gradient $=0.249$ and intercept $=-0.628$ kJ mol^{-1} (see Fig. 3). Although the correlations were poor, they were better than the corresponding values of the native retinol ($r=-0.255$, $r_s=-0.219$, $\tau=-0.142$, gradient $=-3.297$ and intercept $=-1.917$ kJ mol^{-1}). The new structure is multiply halogenated (10 chlorines) and has an amide group and an aliphatic heterocycle.

Table 7 shows the results for fragment placements where two placement options were used; the number of halogens was restricted to two and only neutral fragments were allowed to be placed. In general, the results showed a much smaller number of atoms differing from the native ligand compared with the unrestricted placement.

Comparisons between Tables 6 and 7 are very informative. When restrictions are imposed, the correlation coefficients r and r_s are greatly reduced for electrostatic complementarity. An analogous effect is observed with composite electrostatic potentials and hydrophobic potentials. These coefficient values are nearer to those observed for the native ligand in the 1RBP site. The restrictions drastically reduce the number of new atoms, to between four and eight.

Discussion and Conclusions

This paper describes the rigorous testing of the fragment placement procedure in site-directed atom assignment. The test models used were five protein–ligand co-crystals from the Brookhaven PDB. These were the same test models used previously in a precise self-placement test [3]. In site-directed design, there is no unambiguous test for optimization of the algorithm for atom assignment; one can never be sure that the global optimum has been found, despite the fact that the search procedure is in theory ergodic. Multiple replicates, using different sequences of random numbers, can reduce the uncertainty and provide classes of solutions with closely similar values for the global optimum. The procedure developed here was designed to test how complementary, in terms of the electrostatic potential, the atom assignment was to the site. At the same time, it would be possible to assess the similarity between the new site-directed assignment and the atom disposition on the natural ligand. In all five test cases, the algorithm generated atom assignments, based on electrostatic potential complementarity with the site, which were equivalent to or better than the natural li-

gand. In the most difficult case, with 10^{21} possible assignments, the objective function converged within 7 min of CPU time.

The molecular graph for AMP has 4×10^8 possible fragment placements. Self-placement tests showed consistency for a correct placement in 100% of the replications using r or r_s as the objective function [3]. Site-directed placement, from 1AK3, reproduced AMP in 90% of the replications using r, and 33% of the replications using r_s; if the hydrogen-bond option was specified, replication was improved to 100% and 60%, respectively. In the nonreplicated structures with the hydrogen-bond option specified, only one atom was different from AMP, giving rise to a chloro-adenine. However, the difference in correlation coefficients for the electrostatic potentials between the two alternative structures was insignificant.

Self-placement tests for cAMP gave complete reproducibility [3]. About 10^8 fragment placements are possible. Site-directed placement from 3GAP consistently modified the structure to a chloro-adenine derivative in all but two replicates for objective functions composed of r or r_s. In this case the electrostatic potential of the structure was significantly better, in terms of complementarity with the site, than was the natural ligand. However, the structure, although showing a better complementarity, was still very similar to the native ligand; only one atom was different.

Fragments could be placed on the folate molecular graph in 10^{11} ways. Self-placement tests were rarely successful and usually had three to five atoms misplaced [3]. Site-directed placement from 1DHF created structures with electrostatic potentials more complementary to the site than the native folate. The placements have five to ten atoms different from folate. Most structures were multiply halogenated. However, when halogen atom placement was restricted to two atoms, the resulting structures differed in an average of two to four atoms from folate. In this case the correlation coefficients for electrostatic complementarity were better than those of folate, but not significantly different from those of the multiply halogenated structures. This suggests that alternative structures to folate can be created with equivalent electrostatic potentials. Multiple halogenation appears to create difficulties in site-directed placement; however, these can be overcome effectively, without a significant loss in efficiency, by a restricted placement.

The molecular graph for dihydroxybenzoate presents only a trivial problem for fragment placement, since the total number of placements is only 2000. Not surprisingly, self-placement tests were entirely consistent [3]. However, site-directed placement from 1PHH and the co-enzyme FAD created either the native ligand or a monofluoro derivative. The complementarity in the electrostatic potential of the two structures with the site was not significantly different.

The retinol/retinol-binding protein co-crystal presents a fascinating challenge to the algorithm, principally because the native ligand/protein complex shows no complementarity in electrostatic potential. Furthermore, although there are many hydrophobic groups in the ligand and the site, the hydrophobic potentials of ligand and site are also dissimilar. The retinol self-placement problem is extremely hard: there are 10^{21} possible arrangements of fragments [3]. In the self-placement tests, the best method gave 50% reproducibility. In this case, the hydrogen bonds were specified and the objective function was composed of equal contributions from the electrostatic and hydrophobic potentials. In the nonreplicating structures, two to three atoms were misplaced. It is therefore not surprising that the site-directed placement algorithm generates atom assignments which are considerably different from the natural ligand. It is possible to take the molecular graph of retinol and use the site 1RBP to create a structure with much greater electrostatic complementarity, with 16–20 atom assignments different from retinol. If hydrophobic similarity with the site is used as an objective function, then 6–13 atoms differ; one structure was more hydrophobically similar than the native retinol. Mixed objective functions give intermediate differences. The structure with the best electrostatic complementarity was zwitterionic. Curiously, one paper in the literature has suggested that the retinol binding site is charged [10], resulting in the pH-induced release of retinol from retinol-binding protein. Once more, multiple halogenation plagued the unrestricted assignment. However, when restrictions were imposed to limit the number of halogen atoms and to prevent placement with charged fragments, the resultant structures had fewer new atoms differing from retinol. These restrictions had a marked degrading effect on the values of the electrostatic objective function and they appeared to be more equivalent to those with the native ligand and the site.

The lessons to be drawn from the work with retinol-binding protein are: (i) the algorithm will create complementary electrostatic structures, even though complementarity may not exist between the site and the natural ligand; (ii) a similar case occurs with an objective function driven by hydrophobic similarity; and (iii) options to limit the number of halogen atoms, and to prevent the algorithm from using charged fragments, may lead to structures which show more resemblance to the ligand. This last option should be used with care and perhaps only where information about the site merits it.

On further examination of the site, three water molecules appear to be present within 4 Å of retinol and the site. These are in prime positions to form hydrogen bonds (classical type) with the following residues in the protein: Asp^{103} O, His^{104} $N_{\delta 1}$, Tyr^{133} O_{η}, Leu^{35} O, Gln^{98} $N_{\varepsilon 2}$, Arg^{121} $N_{\eta 1}$ and Phe^{96} O. The water molecule next to Phe^{96} is sufficiently close (2.73 Å O ··· O) to hydrogen bond with the hydroxyl oxygen of retinol, while the other two waters

may be involved in weak hydrogen bonds with the hydrocarbon chain of retinol. It is possible that these water molecules play an essential role in the binding of retinol to its binding site. Since the electrostatic complementarity and hydrophobic similarity correlations and annealing results were performed without considering the effects of the waters, the values may be deficient. This work needs to be expanded to regard the site as the composite of all the protein atoms and water molecules around the retinol.

The simulated annealing algorithm converged satisfactorily in self-placement tests [3] and in the site-directed placement it provided structures with significant levels of similarity to the natural ligand. In examples where the natural ligand showed high degrees of complementarity, the algorithm usually converged well; in this case the target potential is defined within a well-spread range of maxima and minima. Where extremes of potential are closer together, complementarity with a natural ligand is weaker, but the algorithm can create structures with greater complementarity. However, complementarity is achieved by unusual atom combinations, such as multiple halogenation or zwitterion formation. If these arrangements are limited by input options, complementarity can be drastically reduced. It would seem to be a good course of action to apply these restrictions routinely, provided that the site does not prohibit it. In the following paper, the algorithm is tested further by creating structures with similarity in their electrostatic potential to a target structure. Envelope-directed design is treated as an extension of site-directed design, but using the same formalism for atom assignment.

Acknowledgements

The authors would like to thank the Wellcome Trust for financial support through the Wellcome Prize Studentship (M.T.B.) and Wellcome Principal Research Fellowship (P.M.D.) schemes. Part of this work was carried out in the Cambridge Centre for Molecular Recognition.

References

1 Barakat, M.T. and Dean, P.M., J. Comput.-Aided Mol. Design, 9 (1995) 341.
2 Barakat, M.T. and Dean, P.M., J. Comput.-Aided Mol. Design, 9 (1995) 351.
3 Barakat, M.T. and Dean, P.M., J. Comput.-Aided Mol. Design, 9 (1995) 359.
4 Barakat, M.T. and Dean, P.M., J. Comput.-Aided Mol. Design, 9 (1995) 457.
5 Diederichs, K. and Schultz, G.E., J. Mol. Biol., 217 (1991) 541.
6 Weber, I.T. and Steitz, T.A., J. Mol. Biol., 198 (1987) 311.
7 Davies, J.F., Delcamp, T.L., Prendergast, N.J., Ashford, V.A., Freisheim, J.H. and Kraut, J., Biochemistry, 29 (1990) 9467.
8 Schreuder, H.A., Van der Laan, J.M., Hol, W.G.J. and Drenth, J., J. Mol. Biol., 199 (1988) 637.
9 Cowan, S.W., Newcomer, M.E. and Jones, T.A., Protein Struct. Funct. Genet., 8 (1990) 44.
10 Ptitsyn, O.B., Zanotti, G., Denesyuk, A.L. and Bychkova, V.E., FEBS Lett., 317 (1993) 181.