Molecular recognition: optimized searching through rotational 3-space for pattern matches on molecular surfaces

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

Department of Pharmacology, University of Cambridge, Hills Road, Cambridge CB2 2QD, UK
*Department of Physics, Cavendish Laboratory, Madingley Road, Cambridge CB3 OHE, UK

This paper provides a method for searching for pattern matches between parameters mapped onto molecular surfaces. A window, of any specified shape, is projected on to one molecule and a second molecule is rotated to find discrete matches within the projected windows. This search is optimized at two levels. Level 1 finds feasible regions for matching. Level 2 ascertains the best match in a feasible region through an intermediate application of cluster analysis to level 1 data; this is followed by further extensive minimization to locate distinct matching orientations. The method has been tested for the accessible surface and the electrostatic potential surrounding the neurotoxins saxitoxin and tetrodotoxin.

Keywords: surface pattern-matching, accessible surface, electrostatic potential, optimized searching, saxitoxin, tetrodotoxin

Received 1 July 1987
Accepted 14 July 1987

INTRODUCTION

In the research field of drug design two questions about molecular recognition are commonly examined. First, what spatial similarities are there in three-dimensional structure, and associated molecular fields, for drug molecules that exert their action by binding to the same site? Second, where the binding site structure is known, how complementary to each other are the ligand and site? Molecular graphics can aid a visual search for pattern matches of similarity or complementarity, but it would be extraordinarily difficult to carry out exhaustive searches by graphical methods. An optimized search, together with an indication of the quality of any match discovered, for pattern matches is desirable.

Point sets of atom positions can be examined by the distance geometry techniques pioneered by Crippen.[1] These methods have the advantage that they are rotationally invariant, although this property can cause difficulties with stereoisomers. However, if two molecules, each containing n atoms, are to be compared, then $(n^2 - n)/2$ distances have to be calculated and there are $n!$ possible orderings for the distances. Therefore, even with an optimized branch-and-bound recursive-descent tree-search only a small number of atom matches can reasonably be searched.[2]

A different approach to molecular structure comparison is needed if surfaces are to be matched. Rank correlation analysis can be used to assess similarity or complementarity in pattern between surfaces. If the two surfaces are monotonically related, then the method of matching is scale invariant.[3] This type of analysis can be applied successfully to molecular structural surfaces, such as the accessible surface, or to other parameters, for example, the molecular electrostatic potential computed on a three-dimensional surface. Semiregular arrays of points on a surface can be generated by gnomonic projection;[4] this projection reduces the statistical distortions that can arise in sampling points distributed on curved surfaces.

Our paper tackles the problem of searching for orientations of two molecules, A and B, that produce measurable similarities in surface properties. Thus we take a patch, of predetermined shape, fixed on the surface of molecule A and characterize that surface. Then a search is made through a window of that patch projected onto molecule B by rotating B so that the surface characteristics between A and B exhibit an optimal match in the windowed region. The procedure outlined presupposes that we know the patch on A that is to be matched by rotating B around its body axes in 3-space. This is a common case in medicinal chemistry where the chemist has an inkling of the active face he wants to compare. A more complicated situation arises if the patch on A is not defined so that a blind search is needed. Both molecules A and B then have to be rotated against a common window to determine the best matched patches; this is a six-dimensional problem. Our procedure handles surface pattern searching in rotational 3-space, but it could be extended to include 6-space.

Brute-force calculations, where an incremental change is introduced to each of the rotation axes in turn, would give gross approximations for pattern matching. Narrow trajectories in rotational space could be missed because the step-size of the search grid may be too large. A

 0263-7855/87/030152-07 $03.00

better procedure would be to optimize the search by using an algorithm that decides on a suitable search direction and rotates B toward the nearest orientation where the pattern match is greatest. Numerous algorithms are available in standard program packages to solve this minimization problem. By judicious choice of a search strategy further optimization can be achieved. In comparing two molecules we search for the minimum dissimilarity between faces as molecule B is rotated.

In the total search space there may be more than one minimum. The positions of other minima are often desired in drug design to find all orientations of a ligand that fit the geometry of the site. In the accompanying paper a method is described for identifying significant local minima in rotational 3-space.[5]

An optimized search procedure is explicitly given here and tested on the two dissimilar neurotoxins saxitoxin (STX) and tetrodotoxin (TTX). These drug molecules have been used in previous studies of molecular pattern matching, and comparison between the different methods can easily be made.

METHODS

Conditions for optimization

The optimization problem of finding a pattern match in rotational n-space can be expressed

$$\text{minimize } f(x);\ x \in \mathbb{R}^n \tag{1}$$

where $f(x)$ is the objective function containing the independent variables x. The function must be continuous and differentiable. Differentiability is important since this property defines the conditions for a minimum. The gradient vector of f, denoted by $g(x)$ is

$$g(x) = \nabla f(x) = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \cdot \\ \cdot \\ \frac{\partial f}{\partial x_n} \end{pmatrix} \tag{2}$$

and the Hessian matrix of $f(x)$, $\mathbf{G}(x)$, is

$$\mathbf{G}(x) = \nabla^2 f(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix} \tag{3}$$

A minimizing algorithm seeks to find by iteration a point x^* that satisfies the conditions

$$\|g(x^*)\| = 0 \tag{4}$$

and

$\mathbf{G}(x^*)$ is a positive semidefinite matrix.

Both conditions define a local minimum. For a local minimum the eigenvalues of $\mathbf{G}(x^*)$ are all non-negative; these can be checked numerically. For a strong local minimum the eigenvalues of $\mathbf{G}(x^*)$ are all positive; the matrix is then said to be positive definite.

The search for the minimum proceeds as follows. Let the objective function be $f(x)$; at each iteration the value of f will decrease so that the pathway is continuously descending to the minimum x^*, therefore for k iterations $f_{k+1} < f_k$. Let the current estimate of x^* be x_k.

Step 1. Determine whether the function has converged at this point; if the function has converged $x^* = x_k$ then exit; otherwise continue.
Step 2. Compute the direction of search; this is a unit vector $\mathbf{p}_k$.
Step 3. Decide on a step length by computing a scalar α_k so that $f(x_k + \alpha_k \mathbf{p}_k) < f(x_k)$.
Step 4. Find the new value $x_{k+1} \leftarrow x_k + \alpha_k \mathbf{p}_k$, reset $k \leftarrow k + 1$ and return to Step 1.

The optimization algorithm used in this paper, the NAG routine E04JBF,[6] is based on quasi-Newton methods whereby derivatives are computed by calls from the routine; an analytical formula for the derivatives of $f(x)$ does not have to be supplied. Quasi-Newton procedures gradually build up curvature information as the descent progresses by tracking the behavior of past and current iterations. The Hessian matrix $\mathbf{G}$ is not explicitly computed at each iteration but is approximated by a positive semidefinite matrix $\mathbf{H}$, which is updated during each iteration. This optimization procedure is fully explained by Fletcher[7] and Gill, Murray and Wright.[8] The routine E04JBF is comprehensive and allows the user to monitor and fine-tune the minimization as the descent occurs; one useful feature is that the user can limit the number of function evaluations allowed so that feasible regions for a minimum may be discovered in the neighborhood of x^* without having to go to the length of computing x^* exactly each time.

The objective function

The need for the objective function to be continuous and differentiable has an important consequence for pattern matching. In previous studies[3,4] rank correlation analysis has been used to define the quality of a pattern match. In its simplest form, Spearman's rank correlation coefficient, R_{rank}, is given by

$$R_{\text{rank}} = 1 - 6 \sum_{k=1}^{n} d_k^2/(n^3 - n) \tag{5}$$

where d_k is the difference in ranks between the corresponding pairs of surface parameters and n is the number of pairs in the data.

Since the pattern match is only calculated over a region of n data points, there are a finite number of values that R_{rank} can assume. Furthermore, small shifts in orientation may not alter the rank, thus the number of stationary points would be large and would be found in all rotational regions where the rank was unchanged. Therefore, an objective function based on R_{rank} would be discontinuous and undifferentiable.

An alternative procedure for constructing an objective function would be to take the error function for the two surfaces. Suppose that a surface parameter $P_{i\text{A}}$, can be computed at a point i for molecule A together with a corresponding parameter $P_{i\text{B}}$ for molecule B. Then if the surface encompasses n points, the residual, e_i, at each point is given by

$$e_i = P_{i\text{A}} - P_{i\text{B}} \tag{6}$$

and the error function for the surface can be expressed in sums of squares form as

$$f(x) = \sum_{i=1}^{n} e_i^2 \qquad (7)$$

This expression has the advantage that the summation is composed of positive terms, and f is a direct measure of the dissimilarity of the surface. If f has the value of zero, both surfaces are exactly similar. The arrangement of points, i must be exactly comparable and well distributed. Gnomonic projection of an icosahedral tessellation provides a constant semiregular array of points on a molecular surface.[4]

Molecular rotations

Rotation of molecule B can be achieved without degeneracy by rotation through the conventional Euler angles ψ about the z-axis, θ about the x-axis and φ about the new z-axis. The rotation matrix **R** is

$$\mathbf{R} = Z\,\varphi X\theta Z\psi$$

The function parameters x_1, x_2 and x_3 are assigned the angles φ, ψ and θ respectively. The matrix **R** has the elements

$$\begin{aligned}
r_{1,1} &= \cos x_1 \cos x_2 - \sin x_1 \cos x_3 \sin x_2 \\
r_{1,2} &= -\cos x_1 \sin x_2 - \sin x_1 \cos x_3 \cos x_2 \\
r_{1,3} &= \sin x_1 \sin x_3 \\
r_{2,1} &= \sin x_1 \cos x_2 + \cos x_1 \cos x_3 \sin x_2 \\
r_{2,2} &= -\sin x_1 \sin x_2 + \cos x_1 \cos x_3 \cos x_2 \\
r_{2,3} &= -\cos x_1 \sin x_3 \\
r_{3,1} &= \sin x_3 \sin x_2 \\
r_{3,2} &= \sin x_3 \cos x_2 \\
r_{3,3} &= \cos x_3
\end{aligned}$$

If x_1, x_2, x_3 are the Euler angles for rotation around the three-body axes of molecule B, then the objective function to be minimized is $f(x_1,x_2,x_3)$. This function is continuous and differentiable. The minimization problem is unconstrained. At each function call, the pierce-points of the gnomonic projection through the molecular surface are computed together with the surface parameter P_{iB}. The initial positions for x_1,x_2,x_3 lie between $0 \rightarrow 2\pi$ for x_1, x_2 and $0 \rightarrow \pi$ for x_3. No bounds are set on the minimization since the search trajectory may involve a number of revolutions around any of the three-body axes and may proceed in a negative direction. Thus any end-point rotation will need rescaling if it lies outside the initial bounds.

Improving the optimization

Since the purpose of the minimization is to find the global minimum together with other discrete local minima, a large number of randomly distributed starting orientations is needed to increase the probability of finding these minima. Unless the behavior of the objective function is well understood, it is not possible to specify in advance how many starting positions are needed to be sure of encountering the global minimum.

Minimization from a large number of starting positions to complete the search to each nearest local minimum could be very expensive in computer time. In order to be economic, a method has been developed to carry out the search in two stages; the method is described in detail in the accompanying paper[5]. The first stage is to start from random positions and allow a partial minimization to occur by an exit from the routine at step 4 after a preset number of function calls have been made (40). The effect of this is to lead to a concentration of end-points loosely clustered in the region of each feasible minimum. The rotational 3-space is then searched by cluster analysis to identify regions of space in which local minima might be expected.

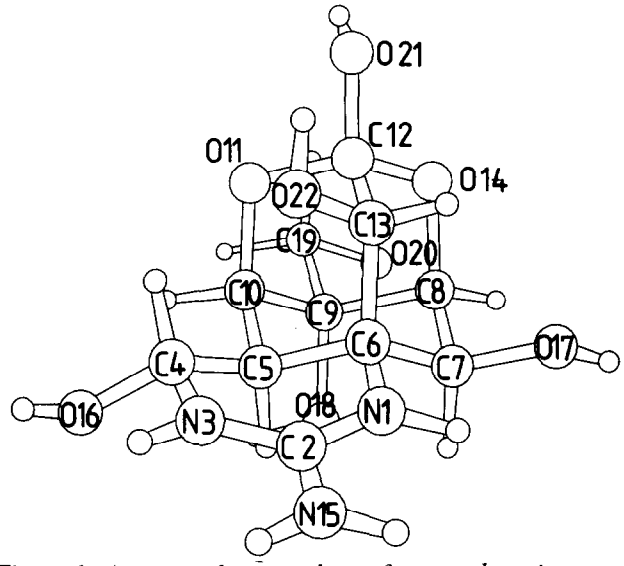

Figure 1. Atom numbering scheme for tetrodotoxin

This clustering problem is not straightforward. First, the clusters are not discrete; they are fuzzy; membership of a cluster has to be decided on location together with the value of the residual. This is analogous to monothetic division. Secondly, the number of clusters present is not known at the start of the analysis. The minimum residual position of each cluster, found by optimization at level 1, is used as the starting point for the second stage, where minimization is allowed to proceed until the local minimum is found (level 2). Each run in this step may need a large number of function calls (2000) for the minimization to be complete. Thus unnecessary amounts of repeated calculations are removed by filtering out all but one point in each cluster and points that are not identified with a strong cluster with low residuals are discarded. This optimization improvement is based on the assumption that in a cluster region there is a smooth surface with only one local minimum embedded therein.

Figures 1 and 2 provide the key to atom identification in each molecule of tetrodotoxin and saxitoxin; initial starting coordinates are presented in Tables 1 and 2. The molecular surfaces examined by the search program were the accessible surface match and the match between the electrostatic potentials projected onto the accessible surface. The graphical display method has been described previously;[4] molecular skeletons are drawn by the program of Beppu.[9] Two point densities, 25 and 90 points, were investigated to assess the matching procedures. Furthermore, as a test of the search algorithm, tetrodotoxin held in a fixed orientation was compared with random starting orientations of another tetrodotoxin molecule. The search window has been defined as a hemisphere for the gnomonic projection with positive z-coordinates. Thus, where the molecules have the origin of their coordinate system located at their center of mass, approximately half of tetrodotoxin is used for comparison.

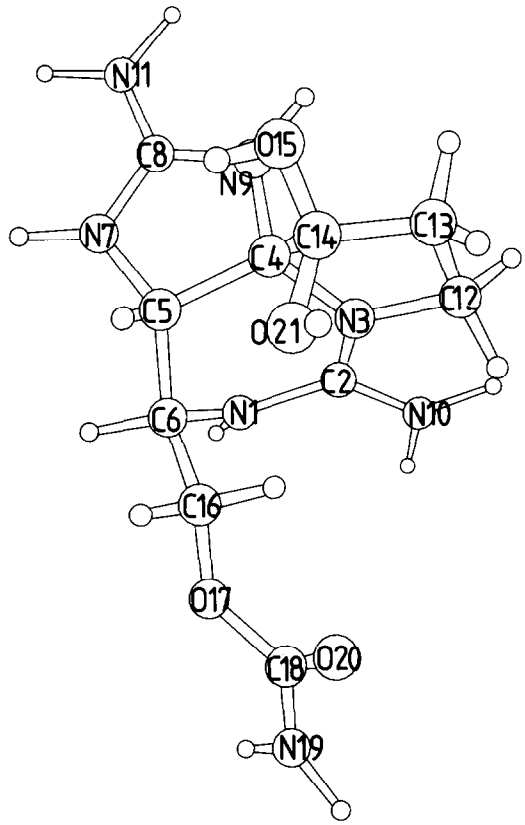

Figure 2. Atom numbering scheme for saxitoxin

Table 1. Atomic Cartesian coordinates for the heavy atoms of tetrodotoxin

Atom	x	y	z
N1	−0.6329	2.1563	1.0209
C2	−0.9620	3.0665	0.1209
N3	−0.8642	2.8260	−1.2135
C4	−0.5405	1.4974	−1.7528
C5	−0.7531	0.4586	−0.7060
C6	−0.1581	0.7994	0.6573
C7	−0.6247	−0.2216	1.6792
C8	−0.0567	−1.5971	1.2498
C9	−0.6652	−2.0521	−0.1602
C10	−0.1831	−0.9443	−1.1122
O11	1.2399	−0.9272	−1.1837
C12	1.8182	−0.5986	0.1190
C13	1.3940	0.7767	0.5478
O14	1.3734	−1.5701	1.0736
N15	−1.4434	4.2924	0.5462
O16	−1.4110	1.2324	−2.8726
O17	−0.0973	0.1168	2.9990
O18	−2.0633	−1.9927	−0.0607
C19	−0.1352	−3.3953	−0.6315
O20	−0.6163	−4.4168	0.2457
O21	3.1624	−0.7381	−0.0453
O22	1.9315	1.7405	−0.3629

RESULTS

Accessible surface matching

Rotational positions for the best matches between saxitoxin and tetrodotoxin are given in Table 3; these

Table 2. Atomic Cartesian coordinates for the heavy atoms of saxitoxin

Atom	x	y	z
N1	−1.2095	−0.2206	1.2760
C2	−1.5122	−0.4184	−0.0145
N3	−0.6200	−0.0102	−0.9199
C4	0.3585	1.0335	−0.6174
C5	0.8498	1.0269	0.8518
C6	0.1950	−0.1030	1.6389
N7	0.4586	2.3408	1.3142
C8	−0.1992	3.0079	0.4157
N9	−0.2360	2.3492	−0.7432
N10	−2.5833	−1.0920	−0.3917
N11	−0.7450	4.2154	0.5854
C12	−0.6259	−0.4356	−2.3431
C13	0.5602	0.3334	−2.9151
C14	1.4158	0.7557	−1.7448
O15	2.1396	1.8812	−2.0924
C16	0.9043	−1.4245	1.4770
O17	0.1644	−2.3915	2.2036
C18	0.0878	−3.6202	1.6760
N19	−0.6913	−4.4034	2.3865
O20	0.7659	−3.9695	0.7397
O21	2.1808	−0.2737	−1.1910

show the local minima in dissimilarity between the parameter surfaces. With the accessible surfaces there are marked differences between the calculations with n = 25 or n = 90 in equation 7; matched positions do not show 1:1 correspondence. Computations using 25 points give very different positions for the first three minima. R_{rank} is not correlated with the value of the residual. For example, the global minimum does not have the largest R_{rank}. If 90 points are used to calculate the residual, the first four minima are closely located and have very similar residuals. These four minima are located near to the second minimum found with n = 25. Once again the order of R_{rank} does not follow the order of the residuals. Position 5 of the 90 point data is near to position 3 with n = 25. Both global minimum positions are distinctly different. Color Plate 1 shows the accessible surface of TTX (white) with STX superimposed (red) drawn in stereo. Each drawing in the color plates used 424 plotted points to provide a graphical impression of the whole surface. It can be seen that the searching program has matched the large protuberances together with the clefts on the accessible surfaces.

Electrostatic potential matching

Matched positions in the electrostatic potential for n = 25 and n = 90 are listed in Table 3. The large residual reflects the fact that the difference in electrostatic potential between the surfaces is large, about 200 kJ mol^{-1}; this difference is due to tetrodotoxin being a monocation whereas saxitoxin is a dication. For both point densities the correlation between R_{rank} and the residual is poor. A rank correlation coefficient of zero is found for the fourth position with n = 25. Where the calculations were performed using 90 points, the first four minima are close together and near to the global minimum found with n = 25. A comparison of the best match between the potentials for TTX and STX is shown in Color Plates

Table 3. Rotational matches for the five lowest minima given as the rotational coordinates x_1, x_2, x_3 of saxitoxin, with tetrodotoxin kept in a fixed orientation

Minimum number	Euler angles			Residual Å^2	Spearman rank correlation coefficient
	x_1	x_2	x_3		
Accessible surface (n = 25)					
1	4.85	4.83	1.36	8.11	0.6669
2	4.46	1.85	0.91	9.01	0.6769
3	3.67	2.57	0.37	9.73	0.6731
4	0.88	1.20	1.78	12.50	0.5800
5	1.02	1.46	2.08	13.59	0.6023
Accessible surface (n = 90)					
1	4.32	1.93	0.56	34.08	0.5551
2	4.33	1.97	0.78	34.45	0.6310
3	4.34	1.95	0.75	34.48	0.6426
4	4.16	2.06	0.49	34.64	0.6563
5	3.89	2.34	0.42	34.85	0.6411
Electrostatic potentials (n = 25)					
1	4.07	0.54	1.96	61407	0.5373
2	6.28	0.31	1.99	64978	0.1077
3	4.19	1.29	2.35	69143	0.4662
4	0.96	6.23	1.73	69819	−0.1954
5	5.24	1.47	1.48	69877	0.0000
Electrostatic potentials (n = 90)					
1	4.04	0.49	1.96	222261	0.4850
2	3.52	0.39	2.07	223749	0.4302
3	3.49	0.36	2.05	224068	0.4451
4	4.09	0.61	2.03	227240	0.4703
5	2.91	0.27	2.04	231064	0.0701

Table 4. Matches obtained by rotating tetrodotoxin against itself

Minumum number	Euler angles			Residual Å^2	Spearman rank correlation coefficeint
	x_1	x_2	x_3		
Accessible surface (n = 25)					
1	0.0	0.0	0.0	0.0	1.0000
2	1.48	1.51	1.59	9.14	0.7454
3	3.51	3.74	3.00	9.18	0.6731
4	0.12	0.28	2.91	9.46	0.6231
5	0.09	0.25	3.06	10.02	0.6669
Electrostatic potential (n = 25)					
1	0.0	0.0	0.0	0.0	1.0000
2	5.09	2.68	2.72	1151.19	0.5392

2 and 3. It has to be remembered in this computation that the accessible surfaces are not being matched; only the potentials gnomonically projected are being used to minimize the residuals between the two molecules. Regions of low potential show correspondences at the bottom of each picture together with a region of high potential to the upper left. A region of high potential on the right is spatially mismatched.

Rotations of tetrodotoxin against itself

Molecule A of TTX was kept in a fixed orientation, and a second molecule B of TTX was assigned random starting positions with n = 25. Table 4 displays the results for the accessible surface and electrostatic potential matches. With the 20 final positions using the accessible surface comparison, 15 successfully reached the global minimum. At this position the molecules are exactly superimposed. However, other discrete minima were encountered. Color Plate 4 illustrates the accessible surfaces for the match nearest to the global minimum.

Where the electrostatic potentials were used in the calculations, 19 positions from the 20 clusters minimized to the global position and the potentials were exactly superimposed. Color Plate 5 shows the second local minimum and should be compared with Color Plate 2. Regions of high potential match approximately in the upper left and right of each picture; areas of low potential are found at the bottom of the plots.

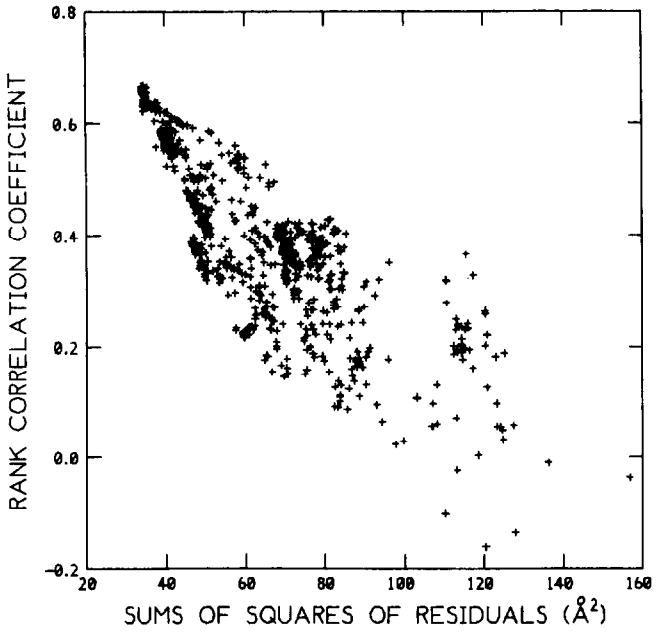

Figure 3. Scattergram of R_{rank} against residual for the accessible surface minimization between tetrodotoxin and STX saxitoxin at level 1 optimization

Relationship between residual and R_{rank}

A scattergram for R_{rank} plotted against residual between the accessible surfaces for different orientations of STX compared with a fixed TTX is shown in Figure 3; each point was taken from level 1 optimization (n = 90). The scatter has a triangular distribution. At low values of R_{rank} the range of scatter in the distribution is large; as R_{rank} increases, the scatter in the residuals decreases. This difference between the methods for matching surface parameters is reflected in the calculation of each coefficient. The sums of squares calculation is dependent only on the difference between paired points and independent of the order in which the differences are summated (equation 7). The rank correlation, on the other hand, summates the differences in ranks between corresponding members of the two sets of points and is not affected by the actual values of the points (equation 5).

DISCUSSION

The method, outlined in this paper, of searching for matches between surface properties of dissimilar molecules is optimized at two levels. This enables one to attempt a comprehensive search procedure with substantial savings in computer time. But the optimization still takes 35 min cpu on an IBM 3084Q to compute 20 cluster minima for n = 90 in the accessible surface search; even then, not all minima converge satisfactorily. The bottleneck in these calculations was tracked down using a Fortran timing procedure; approximately 80% of cpu time is spent doing an unavoidable square-root in the algorithm to calculate the pierce-points. Further square-root calls are made in the computation of the molecular electrostatic potential. Nevertheless, despite this restriction the optimization procedure leads to a 50-fold reduction in computing time compared with a full-scale single minimization step.

The surface minimization between two identical TTX molecules with different starting orientations, provides a good test of the search procedure; the true global minimum is found in a high proportion of the final positions. Other matched, but not global minimum, positions exist; they are probably due to the approximate symmetry introduced in TTX by the adamantane-like portion. These findings illustrate the effectiveness of the matching procedure and the ability of the method to find other discrete positions for matches. However, where the global minimum position is unknown, as is the case with dissimilar molecules, one cannot be absolutely certain whether the true global minimum has been generated. With a large number of random starting positions, the probability of finding the true global minimum is increased.

Color Plate 1 illustrates how striking the similarity of the accessible surface match can be between TTX and STX; both surfaces look to be well superimposed. The low residuals for the global minimum and the other near minima give an average difference between the accessible surface of only 0.62Å. A surface fit of this degree of closeness between dissimilar molecules is encouraging. Further exploration with other molecular structures is needed to provide a more exhaustive testing of the search capabilities of the algorithm.

Searches for the match between the electrostatic potentials give a pattern corresponding to the disposition of gross features between the test molecules. However, the overall outlines shown in Color Plates 2, 3 and 5 would suggest a poor structural superposition. This is entirely due to the fact that the matching parameter is limited to the difference between the two potentials at each pair of points and does not include a term for the accessible surface values. It may be possible to combine more than one parameter in the summation step by normalizing each parameter value within its own distribution to standardize the scales; summation would then be through a set of squared-difference terms. This suggestion would need careful evaluation to assess whether preferential weighting of particular terms should be included to make chemical sense.

The purpose of using different point densities was to discover whether the positions of the local minima, produced by each value of n, were approximately equivalent. In future usage, care will have to be taken in selecting an appropriate point density for the matching problem. The use of computing resources is proportional to n. Further work is needed to establish sensible values for n with each type of surface parameter. The values obtained for R_{rank} with electrostatic potential matching, using n = 25 and n = 90, are 0.537 and 0.485. In the previous paper[4] a comparison between two fixed orientations of TTX and STX, but using 521 points, gave R_{rank} = 0.794. The results are not comparable for two reasons: (1) the minimization procedure used here operates on the residual value, which is not equivalent to R_{rank}; (2) the number of points, n, affects the calculation of R_{rank}.

In general, drug-receptor interaction is a reversible reaction; relatively weak non-bonding forces hold the complex together. These forces may be crudely partitioned into hydrogen bonds, electrostatic forces, van der Waals forces and hydrophobic interactions. Thus, the overall interaction is governed by the molecular shape at the interacting face and the spatial disposition of these forces. Atom connectivity patterns in the ligand are likely to be less important for the binding step if the three-dimensional pattern of forces can be produced by alterna-

tive connectivity arrangements. Therefore, in studies of QSAR it is necessary for research to move away from the restrictions imposed by the limitations of a congeneric series. Minimization of the dissimilarity in surface parameters generates matched orientations for molecular comparisons; this development provides a useful addition to the armory of QSAR and circumvents the need to perform the study within a framework of structural resemblances. In practice, all that is needed is a description of the active face of one of the molecules; the program then searches the second molecule for the corresponding face with a good fit. Other discrete matching orientations are also generated and these may provide the medicinal chemist with alternative orientations to be included in the QSAR.

ACKNOWLEDGEMENT

P.M.D. wishes to thank the Wellcome trust for continued financial support through the Senior Lectureship Scheme.

REFERENCES

1 Crippen, G. M. *Distance geometry and conformational calculations*. Research Studies Press, Letchworth, UK (1981)
2 Danziger, D. J. and Dean, P. M. The search for functional correspondences in molecular structure between dissimilar molecules. *J. Theor. Biol.* 1985, **116,** 215–224
3 Namasivayam, S. and Dean, P.M. Statistical method for surface pattern matching between dissimilar molecules: electrostatic potentials and accessible surfaces. *J. Mol. Graph.* 1986, **4,** 46–50
4 Chau, P.-L. and Dean, P. M. Molecular recognition 3D surface structure comparison by gnomonic projection. *J. Mol. Graph.* 1987, **5,** 97–100
5 Dean, P. M. and Callow, P. Molecular recognition: identification of local minima for matching in rotational 3-space by cluster analysis. *J. Mol. Graph.* 1987, **5,** 159–164
6 E04JBF: a comprehensive quasi-Newton algorithm for finding an unconstrained minimum of several variables. Numerical Algorithms Library routine
7 Fletcher, R. *Practical methods for optimization: Volume 1: Unconstrained optimization*. John Wiley and Sons, London (1980)
8 Gill, P. E., Murray, W. and Wright, M. H. *Practical optimization*. Academic Press, London (1984)
9 Beppu, Y. NAMOD: a computer program for drawing perspective diagrams of molecules. Quantum Chem. Program Exchange Vol 11 program number 370 (1978)