Molecular recognition: blind-searching for regions of strong structural match on the surfaces of two dissimilar molecules

P. M. Dean
University of Cambridge, Department of Pharmacology, Hills Road, Cambridge CB2 2QD, UK

P. Callow
Computer Laboratory, Corn Exchange Street, Cambridge CB2 3QG, UK

P.-L. Chau
Department of Physics, Cavendish Laboratory, Madingley Road, Cambridge CB3 0HE, UK

This paper illustrates a method of searching for matched accessible surfaces on two dissimilar molecules without specifying a particular face to be matched. Both molecules are allowed to rotate and their residuals are minimized to obtain matches. To identify matched orientations, 6-dimensional cluster analysis is used.

Keywords: blind-searching, pattern-matching, accessible surfaces, tetrodotoxin, saxitoxin

Received 27 October 1987
Accepted 18 November 1987

INTRODUCTION

One of the most common questions asked in QSAR studies is: What parts of the molecules match? In other words, are there faces on each molecule that would fit a complementary receptor structure? In the past, attempts to answer this question have resorted to a comparison of the structural formulas by overlaying visually matched, or analogous, moieties. This method fails if a large degree of structural dissimilarity exists between the molecules. One of the difficulties of using methods based on comparing structural formulas is that too much emphasis is paid to bonding topologies. In drug-receptor interaction, the forces holding the complex together are more related to pairwise interactions between atoms at the site and specific ligand atoms at the interface;[1] local bonding networks are of secondary importance.

Several methods have been developed specifically to tackle this important problem. If the molecular interaction between a number of ligands and their binding site occurs through having certain site-points in common, such as hydrogen-bonding donors and acceptors, then it is possible to search for positionally matched sets of ligand-points by tree-searching through a difference distance matrix.[2] However, this can be effectively used only for small numbers of ligand-points on each molecule because of a combinatorial explosion in the searching procedure. This method generates a number of matched atom subsets that can be geometrically superimposed. Other adjacent parts of the molecule, not included in the putative set of ligand-points, then have to be incorporated in a further stage of structure matching and the whole face assessed statistically.[3,4]

More recently, researchers have presented methods for actively searching the surface of one molecule for matches with a fixed face of a second molecule.[5-7] These methods are applicable to the problem of searching for any chemical parameter mapped onto a molecular surface and have been used to obtain matches between molecular accessible surfaces and between molecular electrostatic potentials mapped onto molecular surfaces. The basic method used is to minimize the residuals between two faces when one molecule is fixed and the other is allowed to rotate. The minimization is optimized by splitting the process into two levels. A preliminary incomplete minimization is performed from random starting orientations of the first molecule so that it rotates toward a matched orientation. After a short time, the minimization is artificially truncated and all the new positions are subjected to cluster analysis to find spatial regions showing similar rotational coordinates. This part of the optimization is called level 1. The point with the lowest residual from each cluster is then used as input for prolonged minimization so that a number of matched positions can be ascertained (level 2). The output from this procedure provides a number of faces on the first molecule that match a single fixed face on the second one. This method is useful if the biologically active face of one molecule is known.

A more complex situation arises if the ligand designer is trying to compare two molecules that are believed to bind to the same site, but the binding face of neither molecule has been determined. In this final paper of the series, we present a computational method that outlines a solution to this problem. Since both molecules are free to rotate, the only fixed reference frame is the size and shape of the window through which the molecules are to be compared. The problem is one of blind-searching in rotational 6-space, and we have solved

it by minimizing the residuals within the windowed region and by 6-dimensional clustering. Several discrete matched orientations are generated. These can be compared by visual superimposition, by the values of the residuals between the surfaces, and nonparametrically by rank correlation. We developed our method for the general case and tested it on two dissimilar molecules, the neurotoxins saxitoxin (STX) and tetrodotoxin (TTX); we used these molecules throughout this series of papers as a constant test system.

METHODS

The method of blind-searching for pattern matches between two dissimilar molecular surfaces is very similar to that described earlier for comparing one whole molecule with a fixed face of another.[6,7] Searching is split into two levels. Level 1 allows a short minimization (40 function calls) from many random starting positions for both molecules; this is followed by cluster analysis. Points with the smallest residual from each cluster are input to prolonged minimization in level 2. There are, however, important differences in detail between searching in rotational 3-space and searching in 6-space.

(1) In our previous work in 3-space, we kept one molecule in a fixed orientation and gave the other molecule random initial orientations. In this paper, we have given both molecules random initial orientations with Euler angles, x_n, in the range x_1, x_2, x_4, $x_5 = 0 - 2\pi$; x_3, $x_6 = 0 - \pi$. (x_1, x_2, x_3 are angles for molecule 1; x_4, x_5, x_6 are angles for molecule 2.)

(2) We have modified the squared Euclidean distance metric so that the angles x_r, where $r = 1, 2, 4, 5$, are subject to the condition: if $x_r > \pi$, then $x_r \leftarrow 2\pi - x_r$. This ensures that the clustering algorithm does not arbitrarily split the clusters at the boundary value of 0 and 2π.

(3) We have used an agglomerative single-linkage hierarchical clustering procedure as before, but with minor modifications. Previously, the separation of clusters was checked by visual inspection of 3-dimensional (3D) scattergrams; with 6 dimensions, this is not possible. Icicle plots and dendrograms provide a method for inspecting the cluster behavior and guide the user to select an appropriate number of clusters to be used. This procedure is necessary, since trial runs showed that chaining can be a significant problem with 6-dimensional data.

(4) At level 2 optimization, we have increased the number of function calls to 2,640 in line with the increase in dimensionality of the minimization problem.

The method of blind-searching is outlined in the following steps.

Step 1. The atomic coordinates of each molecule are fixed in an arbitrary orientation, for reference purposes only, for which the Euler angles x_n ($n = 1, 2 \ldots 6$) are set to 0. This rotation has the TTX and STX molecules aligned with the atomic coordinates given in Tables 1 and 2 of Dean and Chau;[6] it corresponds to the 8-atom match of Danziger and Dean.[2] All further rotations are made relative to this orientation.

Step 2. A set of ray vectors, used to calculate pierce points with the molecular surfaces, is computed and stored to be used throughout the minimization. These vectors have been chosen to issue from the surface of an imaginary sphere using an icosahedral tessellation to provide a semiregular array of points.[4] Any patch on this spherical surface can be used as a window for comparing the two molecules. In this paper the patch chosen is a hemisphere with a density of 90 points.

Step 3. Both molecules are given a random orientation around each of their Euler axes, and level-1 optimization is begun on the parameter to be minimized; minimization is performed by the NAG routine E04JBF. In this paper the matching process is limited to accessible surface fitting within the window; the difference in distance between respective points on the accessible surface is calculated exactly as described before.[6] Minimization is stopped after 40 function calls and the end points (Euler angles) stored for further processing.

Step 4. The output from Step 3 is used in hierarchical single-linkage agglomerative cluster analysis by the SPSSx package. Inspection of the frequency plot of residual values gives an indication of an appropriate cutoff value to filter out large residuals. In searching for matches between saxitoxin and tetrodotoxin, a cutoff value, R_p, of < 30 Å^2 is used; for tetrodotoxin rotating against itself to find nonglobal matches, R_p is in the range 10–25 Å^2. Since the single-linkage clustering procedure will produce the same result if the distance matrix is reduced to ranks, it is not necessary to compute the lengths of the arcs in rotational 6-space; the chords give the same ranks. Furthermore, using the squares of chord lengths saves computing time by avoiding repeated calls to the square-root function. This approach is preferable to computing directly the squared Euclidean distances from the angles. If the latter are employed, distortion takes place wherever there is a rotational difference exceeding π between two points; a particularly serious boundary effect occurs, whereby points for which one of the Euler angles is close to 0 are prevented from clustering with other points close to 2π. An appropriate cluster number is obtained from the clustering algorithms by inspection of the icicle plot and the dendrogram. The cluster number, g, has been set to 20; this value appears to be adequate to over-determine the number of clusters necessary for further minimization.

Step 5. From each cluster, the orientation associated with the minimum residual value is taken as input to level-2 optimization. Prolonged minimization is performed for 2,640 function calls, or less if a satisfactory minimum has been achieved.

Step 6. The matches between the accessible surfaces of the molecules, found by blind-searching, have been drawn as before by a modified version of the program of Beppu.[8]

RESULTS

Efficiency of the minimization

In searching for the matches between TTX and STX, we used 1,053 starting positions at the beginning of level-1 optimization. This density of starting positions would produce a maximum average distance of 0.0073 rad. After 40 function calls, the mean chord length trajectory

traveled by the minimization routine was 1.85 rad. Thus, the trajectory traveled is about 250 times the average distance between starting points. This calculation indicates that the relative orientations between the molecules are shifted substantially and well beyond the average starting grid length during the first minimization step. Table 1 shows the decrease in residual during the first six iterations with MAXCAL (an argument to the NAG algorithm E04JBF) set at 40 function evaluations. The mean value of the residual falls off exponentially and after six iterations has decreased to 45% of the starting value. In level-2 minimization, the mean chord distance traveled was 1.38 rad and the mean residual decreased to 38.36 Å^2; that is, to 31% of the original starting value. Therefore, despite the prolonged minimization of a further 2,640 function calls, the distance traveled is smaller and the residual is decreased less than in the initial step. This observation is to be expected for a well-behaved minimization problem proceeding along a smooth trajectory and moving shorter and shorter distances per function call as the local minimum is approached.

Cluster analysis

A direct visual assessment of cluster data in 6 dimensions is not feasible. However, icicle plots and dendrograms can be used to determine the behavior of the clustering algorithm to separate the data into an appropriate number of clusters.

Figure 1 shows an example of an icicle plot for TTX and STX rotated against each other at level-1 minimization; the data from the minimization step has been filtered[7] to retain for clustering only the 60 points with residuals $R_p < 30$ Å^2. The ordinate shows the case number and the abscissa the cluster number as the data is successively agglomerated. At the left, all the data falls into 2 clusters; on the right, 59 clusters are present. The user has to assess an appropriate number of clusters, g, into which the points are to be divided. The figure gives an impression of horizontal icicles (some of them joined together) "hanging" from the left-hand side. The objective is to select a value for g corresponding to a level in the icicle plot at which any well-defined clusters remain discrete, while keeping to a minimum the number of unattached points and weak clusters; for each g there is a corresponding threshold distance, d_0, which defines

Table 1. The decrease in mean residual as minimization proceeds through six iterations, with the maximum number of function calls set to 40. The residual is the square of the distance between the accessible surfaces of TTX and STX summated over 90 semiregularly spaced points

Iteration number	Mean residual (Å^2)	n
0	120.77	58
1	89.47	58
2	75.04	58
3	65.74	58
4	59.87	58
5	56.46	58
6	55.32	49

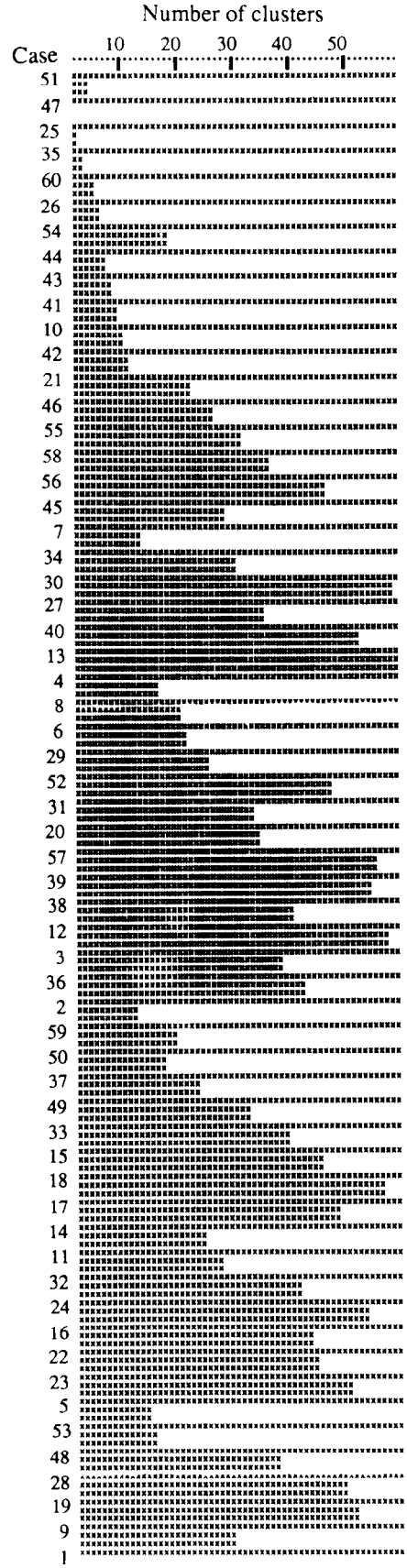

Figure 1. An icicle plot from the cluster analysis of rotational matches between TTX and STX. Only cases with residuals $R_p < 30$ Å^2 have been included

the minimum permissible separation between clusters that remain unfused.

The choice of g (and d_o) on the basis of the shape of the icicles is not a precise procedure. However, whereas in biological taxonomic studies, say, the choice of a partition may have important implications, in this instance the partition is intended merely to provide a basis for further iteration. Setting g too high, and d_o too low, at worst permits the existence of clusters that will converge in level-2 optimization, while retaining some potentially interesting outliers. In Figure 1 there appear to be five or six large icicles that would correspond to that number of major clusters. However, if 5 clusters are specified, this whole group would form only one cluster; single cases *51, 47, 25,* and *35* form the other 4 clusters. (Cases are denoted in italic type, clusters in bold type.) If we specify 20 clusters, there will be 5 large clusters and 15 clusters composed of single cases. The 5 multicase clusters shown in Table 2 contain 75% of the data points from the sample. Statistical analysis of the spread of points within each cluster for all 6 dimensions can provide an indication of whether the cluster is small and compact, large and diffuse or chained in any dimension. Table 3 shows that clusters **1, 3** and **5** have the lowest overall standard deviation and form compact clusters; clusters **2** and **4** are about twice as diffuse, while the smallest deviation in any dimension is found for cluster **4** in x_6, where the range is 0 to π and for cluster **1** in x_2 and x_5, where the range is 0 to 2π.

Table 2. Cumulative frequencies of clustered cases where the residual R_p < 30 Å^2 for the accessible surface match between TTX and STX. Cluster number is in ascending order of residual for the best case in each cluster

Cluster number	Frequency	Cumulative percentage
1	5	8.3
2	13	30.0
3	6	40.0
4	14	63.3
5	7	75.0

The other 15 clusters contain one case each

The dendrogram of Figure 2 illustrates the progress of the allocation of points to each cluster starting from the lowest level, where each case is a separate cluster, and progressing until all clusters have been hierarchically agglomerated. During single-linkage clustering on the distance matrix $\mathbf{D}_g$, the distance, d_{ij}, between the two clusters **i** and **j** is defined as the distance between their nearest neighbors.[9] At any stage, the next link formed will be that corresponding to the smallest element of $\mathbf{D}_g$. To construct the new matrix with i and j combined as a single cluster **(i, j)**, the **i** and **j** rows and columns are simply purged and rearranged: If **k** is another cluster, then $d_{(ij)k}$ is taken as $\min\{d_{ik}, d_{jk}\}$. If fusion is stopped at threshold d_o, the clusters are at least d_o apart measured in terms of the distance between nearest points.[10] Each successive node in the dendrogram records a fusion event; its height on the scale indicates the separation between the two elements (whether grouped or individual cases) making up the new cluster.

Cluster formation for TTX and STX can be traced as follows: The first cluster forms between cases *4* and *13*; a second separate cluster forms between *27* and *30*; the third by fusion of *3* and *12*; and so on. The first triplet is formed between the pair (*39, 57*) and *38*. Thus, we can follow the agglomeration until cases *47* and *51* are finally linked to the rest to form one comprehensive cluster. On the dendrogram, the 5 major clusters are found to be linked for a d_o value of about 1.3; their individual nodes are located above case numbers (*19, 6, 27, 15, 55*) on the dendrogram. With d_o < 1.3, these 5 clusters can be considered as separate groups; they also contain the smallest residuals produced at level **1**

Table 3. Statistical analysis of cluster numbers 1 to 5 for each rotational dimension $x_1 \ldots x_6$ (rad). The number of cases in each cluster is given in Table 2, and the match is between TTX and STX

	Mean Euler angles with standard deviation (σ)											
Cluster number	x_1	σ	x_2	σ	x_3	σ	x_4	σ	x_5	σ	x_6	σ
1	4.10	0.95	5.75	0.29	1.62	0.38	1.24	1.10	0.45	0.29	0.97	0.46
2	2.51	1.20	4.29	1.75	1.76	0.30	2.81	2.01	5.25	1.50	1.34	0.52
3	1.34	0.98	1.38	0.30	1.51	0.51	4.65	0.81	2.05	0.54	2.19	0.40
4	3.09	1.57	1.75	1.46	1.70	0.70	3.01	1.70	1.35	0.93	0.89	0.25
5	4.67	1.13	1.29	0.36	1.29	0.27	1.72	1.03	2.13	0.57	2.42	0.28

Sums of deviations for all 6 dimensions

Cluster number	Σσ
1	3.47
2	7.28
3	3.54
4	6.61
5	3.64

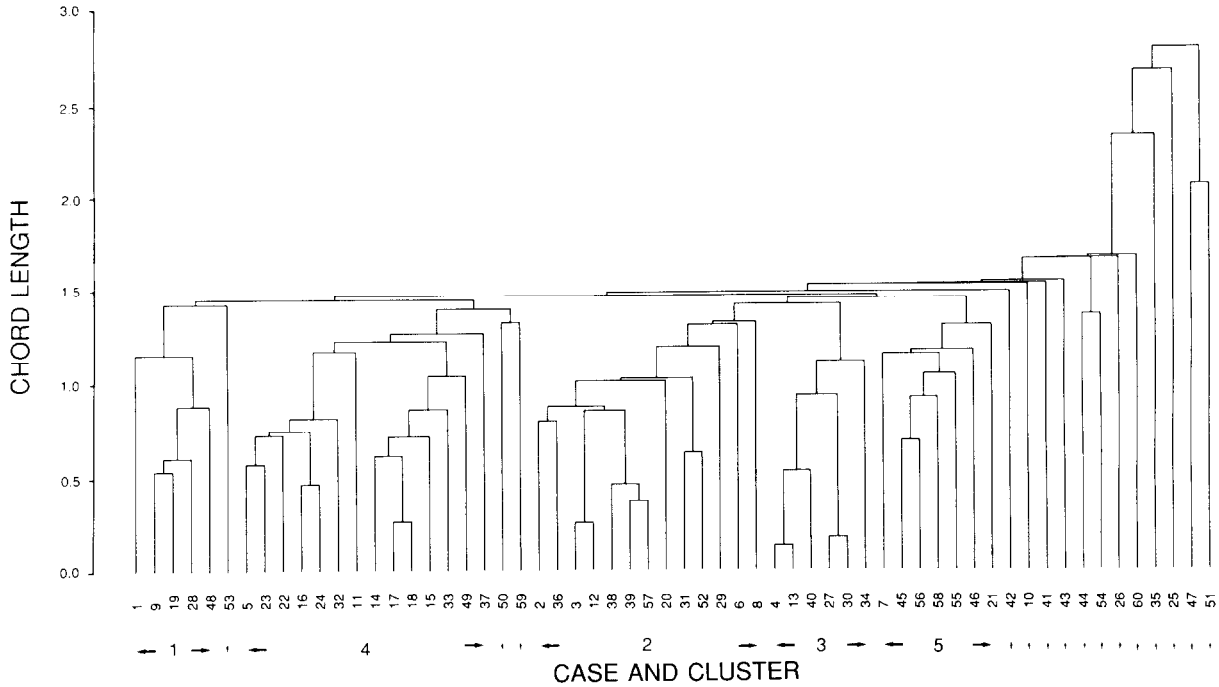

Figure 2. A dendrogram showing the allocation of cases to clusters as the clustering algorithm proceeds. The data are taken from the rotational matches between TTX and STX, shown in Figure 2. The five multicase clusters have been numbered to correspond to those found in Table 2

and are therefore of primary interest for level-2 optimization.

The two graphical methods are complementary and supply the user with a procedure for assessing whether the clusters are discrete, even though there is no direct visual representation of the data.

Revealed molecular matches

STX and TTX

After level-2 minimization, the orientations of STX and TTX are given in Table 4 together with the values of residuals and R_{rank}; the orientations are presented in ascending order of residual. Note that extensive minimization affects the ordering of the match; the cluster numbers are taken in ascending order of residual from level 1. Cluster number **18**, which contained only a single point, has the smallest final residual and $R_{rank} = 0.83$. Clusters **1** to **5** are found in a different order in the best eight clusters. Clusters **18** and **2** show very similar values for x_2, x_3, x_5, and x_6; also, clusters **18** and **1** are similar for x_1, x_2, x_3, x_5 and x_6 — remember that both values for x_5 are close to equivalent rotations of 0 and 2π. Clusters **1, 3, 5, 12** and **10** minimized to stable discrete local minima; the other cluster positions shown in Table 4 had not reached a true local minimum. Nevertheless, the positions obtained have low residuals.

The relative shapes of the accessible surfaces in the two best-matched orientations can be seen in the stereo molecular graphics shown in Color Plates 1 and 2. This illustrates the best matched position for clusters **18** and **2**. Surface protrusions on both hemispherical projections follow each other closely. The r.m.s. separation between the surfaces is 0.452 Å.

TTX and TTX

Two identical molecules of tetrodotoxin were used in a blind search in an analogous way to STX and TTX. If the search algorithm is efficient, many matches should be found that are globally matched with zero values for the residuals and identical pairs of Euler angles, although these angles may differ from one position to another. It has been found that at a cutoff value of $R_p < 10$ Å^2, all clusters at level 1 lead to a global match at level 2. If R_p is in the range 10–25 Å^2, then not all positions are globally matched; 13 of the 20 clusters examined minimized to positions with significant residuals (Table 5). Only a few points reached a stable local minimum. The best nonglobally matched positions for two TTX molecules are shown in Color Plates 3 and 4. The molecular orientation corresponding to cluster **5** has a mean distance between the surfaces of 0.397 Å; the R_{rank} is 0.78 compared with 1.0 for the globally matched positions.

DISCUSSION

Structural comparisons between diverse molecules are widespread in sciences employing molecular recognition; they are frequently used in studies of biomolecular interactions, particularly in QSAR, which is the principal method of research in the pesticide and pharmaceutical industries. Comparisons between molecular skeletons work well only if the structures contain the same pattern of connectivity. Where there is substantial structural dissimilarity, other methods have to be developed to compare similarity in shape. The philosophical assumption behind the new science of molecular recognition is that the ligands bind to their receptor in the same way that pieces fit together in a jigsaw. In general, it is believed that shape and the pattern of other molecular para-

Table 4. Positions of the cluster minima for rotations of TTX against STX after level-2 optimization. The positions of both molecules are expressed in Euler angles (rad) with R_p < 30 Å². The cluster number is that assigned to it at level-1 optimization

	TTX			STX				
Cluster number	x_1	x_2	x_3	x_4	x_5	x_6	Residual (Å²)	R_{rank}
18	5.94	6.12	2.16	3.50	5.87	0.57	18.40	0.83
2	1.46	6.12	1.72	5.05	6.19	0.99	20.10	0.73
1*	5.06	5.64	2.11	2.20	0.34	0.38	20.25	0.81
3*	0.53	1.40	1.83	3.88	1.94	1.95	21.12	0.70
4	2.12	0.57	2.48	5.95	2.24	1.06	21.37	0.72
9	4.77	0.10	1.57	2.11	6.05	1.27	21.60	0.71
6	0.32	2.93	1.64	4.63	0.10	0.66	22.13	0.78
5*	3.95	1.00	1.84	1.30	2.28	1.82	22.71	0.63
17	3.59	0.30	2.47	4.33	0.99	1.05	23.97	0.76
12*	1.48	0.08	0.91	5.13	0.83	1.84	25.67	0.60
7	1.22	6.10	1.00	4.82	0.28	1.65	25.84	0.64
8	1.70	2.95	1.94	0.43	5.73	0.75	25.88	0.76
13	5.59	0.43	1.99	3.39	5.37	1.15	26.53	0.66
19	3.88	5.38	0.48	5.72	3.47	0.06	26.84	0.73
15	3.88	5.38	0.48	5.72	3.47	0.06	26.84	0.73
10*	4.69	6.19	1.75	6.22	0.07	1.01	27.39	0.65
11	0.62	5.44	0.56	1.49	4.49	0.06	27.41	0.76
16	5.15	4.92	0.64	6.01	4.04	0.28	27.86	0.76
14	4.25	5.54	0.56	4.97	4.73	0.10	28.48	0.72
20	4.95	3.16	0.31	2.59	1.10	1.63	28.78	0.60

*Denotes that the algorithm reached a discrete stable minimum. Note that clusters 19 and 15 converged to the same position from different starting positions

Table 5. Positions of the minima for nonglobally matched rotations of TTX against TTX. R_p in range 10–25 Å²

	TTX (a)			TTX (b)				
Cluster number	x_1	x_2	x_3	x_4	x_5	x_6	Residual ($Å_2$)	R_{rank}
5	2.59	0.81	0.78	4.98	2.33	0.85	14.27	0.78
7	2.81	2.21	1.35	5.94	1.11	1.91	18.08	0.77
8	1.22	1.63	2.42	4.37	1.71	0.91	19.48	0.71
16	5.62	0.63	1.46	3.67	5.96	2.35	19.57	0.74
9	0.75	1.14	1.14	4.03	2.28	2.10	19.91	0.72
18*	2.51	6.38	1.02	0.85	0.39	1.86	20.41	0.63
10	6.25	6.01	2.35	1.90	0.58	1.44	20.69	0.73
13	1.57	5.99	1.86	3.54	0.18	1.61	21.55	0.67
14*	4.39	0.61	1.32	2.64	6.20	2.39	21.67	0.71
15	4.30	1.42	2.87	1.07	2.02	0.45	22.10	0.72
17	0.80	0.69	2.55	5.12	2.50	2.32	22.96	0.72
19	5.41	2.49	2.38	1.23	0.86	2.49	23.18	0.72
20	0.52	0.47	2.83	4.01	2.51	0.44	24.47	0.73

*Denotes that the algorithm reached a stable minimum. The 7 other clusters minimized to the global minimum for exact matching

meters, such as the molecular electrostatic potential, are principal determining factors in the recognition mechanism.[11] Therefore, shape-matching and pattern-matching are crucially important research areas in molecular recognition.

We have addressed this problem: Given two molecules, find faces that produce a high degree of surface matching between them. Any matches that the algorithm produces are of course dependent on the size and perimeter shape of the patch acting as a window through which the searching is performed. Here we have limited the window to a hemisphere with its origin at the molecular center of mass. This simulates the problem of searching for surface shapes that could nestle into a cleft. Other semiregular shapes, such as Coon's patches, could be used easily for a window. Even more complex topographies might be more appropriate for problems where numerous separate patches are needed as a window; for example, this might be important if we were to investigate molecules believed to bridge

across a cleft. You would simply have to select which pierce-point vectors should be used for the matching procedure.

In blind-searching, so named because the target orientations are unspecified, you find a number of matches with low residuals and high-rank correlation coefficients. Visual inspection of the match by molecular graphics suggests that the matches obtained look sensible ones. The accessible surfaces of TTX and STX at the best orientations show a good match with a high value for R_{rank} of 0.83. Remember that this coefficient is obtained for a hemispherical projection of molecules of very dissimilar structure. The similarity in surface match for these two faces is quite remarkable and may indicate the face bound to the sodium channel. The r.m.s. differences between the surfaces are small, about 0.4 Å. This small difference is good, considering that the two molecules are rotated around their own centers of mass and that the mean radii of gyration of the molecules, using atoms at the contact surface, differs by 0.1 Å. You can make a more precise match by including a translational element to the coordinate system in the matching routines so that the molecular faces are shifted together for a tighter fit. For TTX rotated against itself, the algorithm behaved in a predictable manner; we obtained many globally matched orientations. Where the matches show significant residuals, it appears that a degree of symmetry about the adamantane cage can give rise to matched positions by realignment around a pseudorotation axis.

The value of a comprehensive method of searching for matched surfaces is that you can identify analogous faces in molecules without any preconceived notion about linking a particular face of the ligand with biological activity. If we had a group of dissimilar molecules that showed all the binding characteristics of being attached to the same site, then it should be possible to identify the most likely orientation for binding by scanning the best 20 orientations of each pair of ligands and finding the common orientation that has an acceptable low residual. This idea is worth giving a trial using a number of ligands having biological actions at the same site. Our method of blind-searching promises to be a useful practical addition to quantitative methods in molecular recognition.

ACKNOWLEDGEMENT

P. M. Dean 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, Chichester, UK, 1981
2 Danziger, D. J., and Dean, P. M. The search for functional correspondences in molecular structure between two 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: 3-d surface structure comparison by gnomonic projection. *J. Mol. Graph.*, 1987, **5,** 97–100
5 Dean, P. M., and Chau, P.-L. Searching for patterns on dissimilar drug molecules. *Br. J. Pharmac.* (in press)
6 Dean, P. M., and Chau, P.-L. Molecular recognition: optimized searching through rotational 3-space for pattern matches on molecular surfaces. *J. Mol. Graph.*, 1987, **5,** 152–158
7 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
8 Beppu, Y. NAMOD: a computer program for drawing perspective diagrams of molecules. *Quantum Chem. Program Exchange,* 1978, **11,** number 370
9 Everitt, B. *Cluster Analysis,* 2nd Ed. Heinemann Educational Books, London, 1980
10 Mardia, K. V., Kent, J. T., and Bibby, J. M. *Multivariate Analysis.* Academic Press, London, 1979
11 Dean, P. M. *Molecular Foundations of Drug-receptor Interaction.* Cambridge University Press, Cambridge, UK 1987