Journal of Computer-Aided Molecular Design, **14:** 573–591, 2000.
KLUWER/ESCOM

A novel method of aligning molecules by local surface shape similarity

D.A. Cosgrove[a,b,*], D.M. Bayada[b,†] & A.P. Johnson[b]
[a]*AstraZeneca, Alderley Park, Macclesfield, Cheshire SK10 4TG, U.K.;* [b]*School of Chemistry, University of Leeds, Leeds LS2 9JT, U.K.*

Received 7 September 1999; Accepted 18 February 2000

Summary

A novel shape-based method has been developed for overlaying a series of molecule surfaces into a common reference frame. The surfaces are represented by a set of circular patches of approximately constant curvature. Two molecules are overlaid using a clique-detection algorithm to find a set of patches in the two surfaces that correspond, and overlaying the molecules so that the similar patches on the two surfaces are coincident. The method is thus able to detect areas of local, rather than global, similarity. A consensus overlay for a group of molecules is performed by examining the scores of all pairwise overlays and performing a set of overlays with the highest scores. The utility of the method has been examined by comparing the overlaid and experimental configurations of 4 sets of molecules for which there are X-ray crystal structures of the molecules bound to a protein active site. Results for the overlays are generally encouraging. Of particular note is the correct prediction of the 'reverse orientation' for ligands binding to human rhinovirus coat protein HRV14.

Introduction

The importance of shape in the molecular recognition process is well established. A large volume of X-ray structural data of ligands bound to proteins has established the fact that tightly binding ligands have a high degree of shape complementarity with their receptors. A strong implication of this observation is that different molecules that bind to the same receptor are capable of adopting conformations with a common shape motif. As a consequence, methods for quantifying the shape similarities of two or more molecules, and aligning molecules accordingly, have been the subject of much study. Good and Richards have recently performed a thorough review of this area [1].

Two principal approaches are possible in the study of molecular shape comparison (MSC). One, which might be termed 'volumetric', involves the computation of the overlap of volume of some function of the molecules, be they hard spheres [2–5] or Gaussian functions [6–9] based on atomic coordinates, grid-based molecular fields [10] or electron density [11, 12].

The second approach, of particular interest to the present work, involves methods that align molecules based on the shapes of the surfaces of molecules. This is a particularly difficult problem because the mathematical description of a surface as an infinitesimally thin object does not lend itself to gradient based optimization. Nevertheless, a means of overlaying according to surface shape is an attractive proposition for several reasons. Ligand-receptor docking is essentially a surface-surface phenomenon [13] – the atomic makeups of the two molecules are largely immaterial except insofar as they define the surface shapes and properties. The surface contact between ligand and receptor is also generally only partial. Some area of both will not be in contact, so it is not always appropriate to compare their total shapes, as volumetric methods generally do. Indeed, some receptors may bind with ligands of markedly different sizes, such as turkey ovomucoid inhibitor (TOMI) [14], a 56 residue peptide, and difluoroketone inhibitor (DFKi) [15], a molecule of less than 40 heavy atoms, which both bind to and inhibit the action of the enzyme elastase.

*To whom correspondence should be addressed. E-mail: david.cosgrove@alderley.zeneca.com
†Current address: Dept. of Molecular Design and Informatics, N.V. Organon, 5340 BH Oss, the Netherlands

Volumetric matching methods cannot work in such cases, as they can do no more than embed the smaller molecule within the larger. Methods based on surface shape similarity are in principle capable of solving these problems.

Methods that can identify and align the common areas of surface shape similarity in two or more molecules are potentially useful in two ways. It is reasonable to infer from such alignments that the portions of the surface that are in common define the shape of some part of the active site. This knowledge can assist in the design of further molecules that will bind at the active site. Conversely, those parts of the surfaces that are not in common may be assumed not to be in contact with the active site. These areas may be amenable to changes that do not affect the binding affinity but may alter some other physical property of the molecule, such as its solubility or metabolic profile to make it more attractive as a drug candidate. Surface shape overlays and analysis can help in both parts of this problem in drug design.

An early solution to surface based overlays was due to Hermann and Herron [16]. Dots were generated on the hard-sphere surfaces of the two molecules to be compared. The user defined those atoms to be used in the scoring, and a systematic search was performed moving one molecule with respect to the other, counting for each orientation the number of surface dots on the nominated atoms on the two surfaces that came within a given tolerance. Masek et al. [17] have described a 'molecular skin', which is the volume difference between two hard-sphere surfaces based on a molecule, the inner surface built with one set of atomic radii, and the outer with the radii all increased by 0.4 Å. This has the effect of giving the surface a volume, and thus turns the problem into a volumetric one for which conventional optimization techniques are appropriate. Overlay of two molecules is achieved by maximising the common 'skin' volume. Perkins et al. [18] have described a similar approach employing a grid-based description of the surface. A grid is laid over each molecule independently, and each grid point marked as either interior, exterior or on the surface with respect to the molecule. One grid is then held stationary whilst simulated annealing is used to optimise the number of surface grid-points that are near each other by moving the other molecule. More recently, Grant and Pickup have reported an extension of their Gaussian volume overlay method that allows a process that they refer to as 'internal docking' [19]. This involves the simultaneous optimisation of the overlaps of both the volume and surface area of the Gaussian representations of the two molecules. Gnomonic projection is the projection of a property at, for instance, a point on a molecule's surface onto a sphere, and Dean et al. have used such projections for surface overlays in the following manner [20–22]. A surface is created for a molecule, and a sphere enscribed around the surface, concentric with the molecule. Rays are projected from the centre of the sphere in fixed directions, and where the ray passess through the surface, the so-called pierce point, a relevant property such as electrostatic potential, or the distance from the centre, is calculated. This property is assigned to the point on the sphere where the ray cuts it. Two molecules so described may be overlaid by rotating one sphere (along with its associated molecule) with respect to the other so that the similarity on all points on the two spheres is maximised. Poirrette et al. [23] have used a genetic algorithm to optimise the overlay of two surfaces described by taking points on a surface and placing them in a regular grid of cubes. An overlay was scored by counting the number of points in close proximity on the two surfaces, using the cubes to speed the calculation. Robinson et al. have recently reported a method of aligning 3D structures using 2D projection that might also be considered a surface shape method, and is in some respects similar to a form of gnomonic projection. A hard-sphere surface is systematically rotated and a large number of 2D projections produced. The shape of each projection is characterised by measuring the distance from the centre of the molecule to the edge of the surface projection at fixed intervals around the molecule. To find the overlay of two molecules, all 2D projections are compared, and the orientations of the pair with the most similar projection shapes define the overlay. Just like gnomonic projection, whilst able to detect local surface shape similarities, the approach involves rotations about the two molecule centres, and so is only appropriate for similar sized molecules. Rosen et al. have used points at the face-centres of molecular (Connolly) surfaces combined with a geometric hashing technique derived from computer-vision research to align protein structures on the basis of shape [24]. Recently [25] Norel et al. have used a similar technique for docking protein-protein complexes and have found shape complementarity to be very much the most important criterion in the binding of such complexes.

In the current work, a method is described that allows surfaces derived by some means to be overlaid by the identification and superposition of areas on the

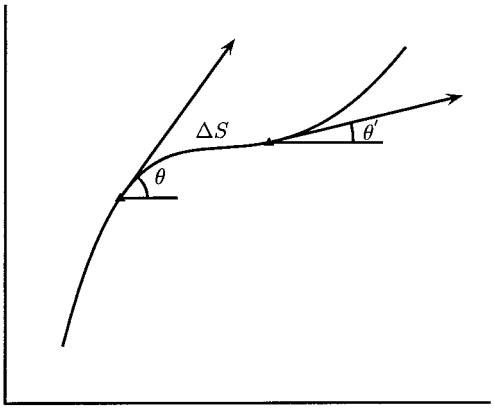

Figure 1. Definition of curvature as the rate of change of θ with respect to *S*, the distance along the curve.

two surfaces that are of similar curvature. Curvature is used as it is a fundamental property of a surface. The curvature at a point on the surface describes in some manner the shape of the surface at that point. It thus provides a direct method of comparing the shapes of two surfaces. The method is general in that the means of surface generation is not important, so long as it is possible to identify points that lie on the surface, and exhaustive in that all possible matches are identified and scored. The relative sizes of the two surfaces are not important in the comparison, and the method is capable of finding local areas of similarity rather than purely global ones. Also explored is a method allowing the combination of a number of pairwise similarity overlays into a consensus overlay for a set of molecules. To this end, the program SPAt (**S**urface **P**atch **A**lignmen**t**) has been written to create the surfaces of a set of molecules, find the areas of similarity on the surfaces, and perform the overlays.

Methods

The basic element used by the present work to describe the surface of a molecule is a 'patch' enclosing on the surface an area of similar regional curvature. The surface of the molecule is divided up into a set of such patches, each patch being characterised by a centre and the average maximum curvature of the underlying surface. It is these patches that are compared between two molecular surfaces to facilitate the superimposition of one on the other. The calculation of the patches is described in the following subsections.

The definition of surface curvature

One definition of curvature is the rate of change of direction of the curve. More formally, it is the change of direction of the tangent to the curve as one moves along the curve, as shown in Figure 1. This may be written as

$$\kappa = \lim_{\Delta S \to 0} \frac{\Delta\theta}{\Delta S}$$

where κ is the curvature, θ is the angle the tangent makes with the x axis, and S is the arc length or distance along the curve. For a function $f(x)$ for which $f'(x)$ and $f''(x)$ are continuous it may be shown that

$$\kappa = \frac{f''(x)}{\{1 + [f'(x)]^2\}^{3/2}}.$$

For any point P on such a curve, it is possible to define an osculating circle which has the properties:

1. It touches the curve at P.
2. It has the same slope as the curve at P.
3. It has the same second derivative as the curve at P.

Such a circle has a radius that is the reciprocal of its curvature.

This definition of curvature can be extended to a surface in three-dimensional space, and an osculating circle may be similarly described for a general point P on the surface, with the added condition that the circle lies in a plane containing the normal to the surface at P. There is a difficulty, however, in that there are an infinite number of such planes, formed by rotating about an axis defined by the normal at P. Euler has described two principal curvatures, κ_1 and κ_2, at P, for which κ_1 is the maximum curvature and κ_2 is the minimum and hence the radii of the osculating circles are a minimum and a maximum. Walker et al. [26] have derived formulae for calculating the principal curvatures of a contoured surface of Gaussian functions representing a molecular charge distribution. Using the curvatures, points on the surface where characterised as being convex, concave or saddle. These descriptions were used to allow the comparison of the shapes of two molecular charge distributions, but not in a manner that allows the overlay of two molecules.

The curvature κ described above is the local or pointwise curvature at a point. This is not necessarily the most revealing description of the curvature of a surface, however. For instance, if the surface is based on a hard-sphere representation, such as the solvent-accessible surface of Lee and Richards [27], the curvature at any point on the surface is merely the reciprocal of the radius of the sphere defining the

surface at that point. Since molecules of biological interest are made up of only a small number of different elements, all with similar atomic radii, this can result in a large number of points with identical local curvatures. A more useful definition of curvature would take into account the point and the points around it on the surface giving rise to a general curvature for the region around the point, which we term the regional curvature. By varying the size of the region around a point that is considered, one can smooth out local irregularities in the surface and capture a more general curvature for that part of the surface. For instance, when comparing the surfaces of two small molecules it is appropriate to consider the variations in the surface shape at an atomic scale, whereas when comparing the shapes of protein surfaces, a less detailed description may be more useful, using larger regions around each point to define the curvature at the point. It is such regional curvatures that are the subject of this paper, the creation of which are described in the next subsection.

Calculating the regional curvature of a surface

In order to calculate the regional curvature at a point P on a surface, we extend the concept of an osculating circle to a more general circle that passes through P and two other points on the surface R and R', which we term the reference points for P. The radius of this circle, which one might term a 'pseudo osculating circle', is the reciprocal of κ_r, the regional curvature of the surface at P. The algorithm for finding the reference points and hence regional curvature is as follows:

1. Find a set of points on the surface along with the normals to the surface at those points. The particular way in which this is done in the current work is described in a later sub-section. Other ways might also be valid, as the overlay procedure is in principle independent of the surface creation method. Any program for generating points defining a molecular (Connolly) surface, for instance, would be equally relevant. Different surface creation methods may, however, give rise to different shaped surfaces. This will effect the overlay results, although the process remains the same.
2. For each point P on the surface, find a set of near-neighbours. The method of establishing these will depend on the surface creation algorithm. In the current work, using contoured surfaces, the contouring algorithm results in a set of triangles, with the surface points as corners, that may be used to draw the surface. The corners of the triangles about point P are used as its near-neighbours. For other surfaces, such as those produced by a dot-based molecular surface program, a search procedure might be required to find all points within a threshold distance of P that would be the near-neighbours for P. Neither of these methods necessarily result in each point having the same number of near-neighbours. The contouring method used here is guaranteed to find a minimum of 3 near-neighbours for each point.
3. For each P, find the set of n_{ref} points that are nearest to P, but are at least *min_ref_dist* from it *along the surface.* These are the reference points for that point.
4. For each P
   (a) Take a pair of reference points of P.
   (b) Find the circle that passes through P and the two reference points in the current pair.
   (c) This circle is not necessarily a valid pseudo osculating circle, as it may not be in the plane of the normal at P. In order to normalise it, its radius is projected onto the normal at P by finding the vector from the centre of the circle to P and taking the dot product of this vector and the normal at P.
   (d) The reciprocal of this normalised radius is the local curvature of the surface at that point as defined by this reference point pair.
   (e) Repeat this for all other reference point pairs of P.
5. Note the largest and smallest curvatures from the above process, as these are an approximation of the principal regional curvatures at P.

It is important that the reference points are found by measuring the distance along the surface rather than through space (a normal Cartesian distance calculation) since it is possible to envisage the situation in a convoluted surface such as in Figure 2. Point R is just beyond *min_ref_dist* of point T when the distance is measured through space and so might be considered as a reference point. However, it is on a completely different part of the surface from T and its use as a reference point is therefore meaningless. Point R' is correctly selected as a suitable reference point. In the current work, the distance along the surface between two points is approximated by finding the shortest path from near-neighbour to near-neighbour using a modified form of Dijkstra's graph traversal algorithm [28]. The problem of finding reference points in this way is mathematically identical to the problem of finding the

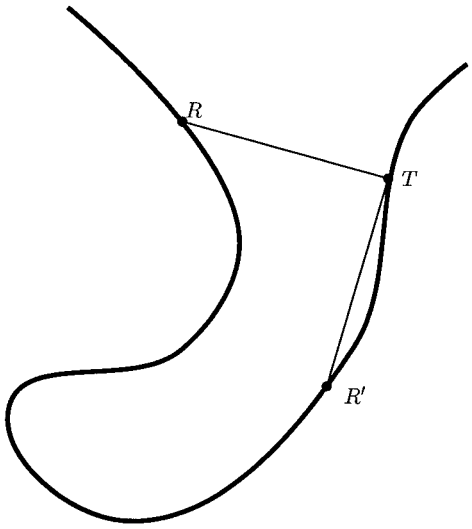

Figure 2. A portion of a surface showing the point T and two possible reference points R and R'.

Table 1. Rules for classifying surface points by maximum and minimum curvature values

Shape class	Max. curvature value	Min. curvature value
CONVEX	$> cutoff$	$> cutoff$
CONCAVE	$< -cutoff$	$< -cutoff$
SADDLE	$> cutoff$	$< -cutoff$
CYLINDER	$> \|cutoff\|$	$< \|cutoff\|$
CYLINDER	$< \|cutoff\|$	$> \|cutoff\|$
FLAT	$< \|cutoff\|$	$< \|cutoff\|$

shortest road distance between two towns via intervening towns, as distinct from the straight-line or aerial distance. Dijkstra's algorithm provides one efficient solution to this problem.

The value of *min_ref_dist* has an important effect on the curvature calculation. With small values, such as the 2.5 Å used in this work, the curvature is calculated on an atomic scale. It is thus capable of picking up subtle differences in curvature due to changes in atomic radius and so is appropriate for comparing small molecules. A larger value of *min_ref_dist* (5 Å, for instance) smoothes out the variations at the atomic level, and so is more appropriate for comparing larger features such as those on protein surfaces.

Formation of regions of similar curvature

Having calculated the maximum and minimum curvatures of the surface at each point, the points are placed in one of five classes: CONVEX, CONCAVE, SADDLE, CYLINDER or FLAT according to the values of their maximum and minimum curvatures, using the rules given in Table 1. The value of *cutoff* depends on the value used for *min_ref_dist* when finding the reference points; a value of 0.1 has been found appropriate when *min_ref_dist* is 2.5 Å, and this combination was used in the current work. This value was arrived at empirically by colouring the surfaces of a small diverse set of molecules according to the curvature class found using a variety of values for *cutoff* and inspecting the classifications to see if they appeared valid.

Once all the points have been placed in one of the five classes, they are collected together into contiguous regions of points of the same class. This is done in two stages. A point is selected that has not yet been placed in a region and its near-neighbours examined. If more than half of them are in the same class as the original point, the new point is placed in the region, and each of its near-neighbours are treated similarly in turn in a recursive process. The region will thus comprise all points that have at least half their neighbours in the same class. Once the region has been created in this way, the patches of similar curvature are created, as described below. When all possible patches have been extracted from the region, it is enlarged by continuing the recursive building process, but this time including points in the region so long as they have at least one neighbour that is in the same class. More patches are extracted from the enlarged region, if possible, and the patch building process moves on to the next region that has yet to be done. The final shapes and sizes of the regions have a slight dependence on the order in which the first point of a region is selected. If a point P has, for example, 4 CONCAVE and 4 CONVEX neighbours, which region it is placed in will depend on whether CONCAVE or CONVEX regions are created first. This problem will only arise at the fringes of the region and we do not consider it to be of major concern.

Building patches from regions of similar curvature

Once the region has been created, as described above, it is divided into circular patches of similar curvature. However, before this can be done, the region must be examined and corrected if necessary. The arbitrary nature of the classification cutoff frequently leads to

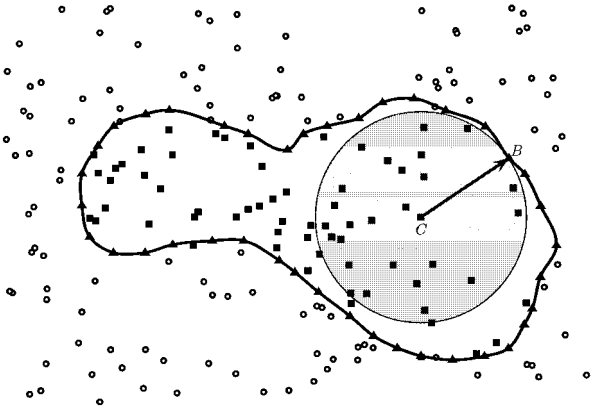

Figure 3. Producing a patch (shaded circle) from a region (enclosed by line). BORDER points are shown as triangles, INTERNAL points as squares. Points not in the region are circles. *C* is the central point of the patch, and *B* is the BORDER point nearest to *C*.

points that are completely surrounded by a region of a different class. This occurs when the curvatures of the point in question lie just on the opposite side of the cutoff from its neighbours. This can lead, for instance, to one or two CYLINDER points, with principal curvatures of, say 0.11 and 0.09 in the middle of a region of CONVEX points, with curvatures of, for example, 0.11 and 0.11. Such invaginated points can seriously effect the production of patches, so they must be found and placed within the region. This is done by examining the near-neighbours N each point P in the region. If one of the N's has more than three-quarters of *its* near-neighbours within the region, N is placed within the region.

The region is then divided into a number of circular patches which are selected so as to be as large as possible. This is done by finding the largest patch (in terms of radius) that will fit within the region, and then the next largest that does not overlap with the first, and so on until it is not possible to make any more patches from the region. So as not to finish the process with a large number tiny patches, a minimum patch size of 10 points is imposed, which is of the order of 0.5–1% of the points in the surfaces of the small molecules considered here. Figure 3 shows the result of applying the algorithm below to a hypothetical region.

1. Assign each point in the region to one of the two classes BORDER or INTERNAL. A point is a BORDER point if less than half its near-neighbours on the surface are in the region, otherwise it is an INTERNAL point.
2. Find the most central point in the region. This is done by calculating the distance from each INTERNAL point to each BORDER point. The most central point is the one whose minimum distance to a BORDER point is maximum.
3. Find the BORDER point that is nearest to the central point. This will mark the furthest extent of this patch, and ensures that the patch will be the largest possible that can fit within the irregular shape of the region.
4. Include in the patch all points on the surface (whether in the region or not) that are nearer to the central point than the nearest BORDER point.
5. Calculate the mean principal curvatures of all the points that are in the patch. These means will be used as the principal curvatures of the patch when patches are compared.
6. Calculate the origin of the patch. Since the maximum curvature of the patch is equal to the reciprocal of the radius of a circle that approximately fits the surface at the central point of the patch, the origin of the patch can be found by projecting the surface normal into the surface for a distance of the recriprocal of the mean maximum curvatures of the points in the patch. This distance becomes the 'radius' of the patch.
7. If the patch is CONVEX or CONCAVE, remove from the patch all points that are further from the patch origin than 1.1 times the radius of the patch. The motivation for this trimming procedure is given below.
8. Remove the points within the most recent patch from the region and repeat the process until no further patches can be generated.

The procedure just described results in the surface being described by a number of patches that are characterised by their origin, central point and radius. A further property is also calculated, which is the angle between the normal of the central point and the vector joining the patch origin to the point in the patch that is furthest from the central point.

Figure 4 shows a section through the surface, with a CONVEX region delineated by a bold line. Before trimming, the whole of the bold region is inside a single patch, with origin O and central point CP, shown by the circle. The BORDER point BP is poorly represented by this patch, and the trimming process rectifies this by removing it and other similarly remote points from the patch, such that they can be placed in other patches, perhaps with central points CP' and CP''. The trimming process does not necessarily result in CONCAVE and CONVEX patches of roughly equal angles. The trimmed patch, origin O, shown in Fig-

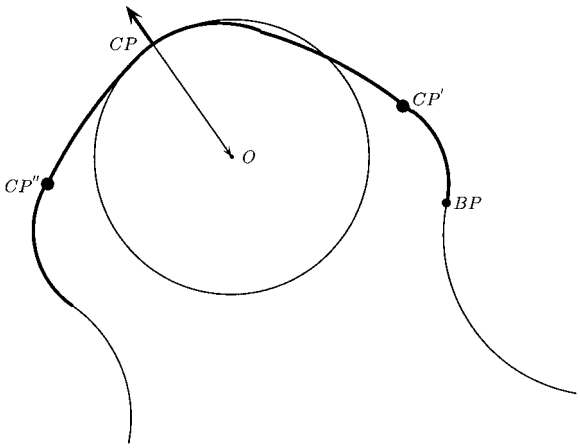

Figure 4. Erroneously large patch, which will be trimmed. O is the origin of the patch, CP is the central point, and BP is a border point that is poorly represented by the patch, and will be trimmed. CP' and CP'' show where the central points of two new patches might be found after the trimming process.

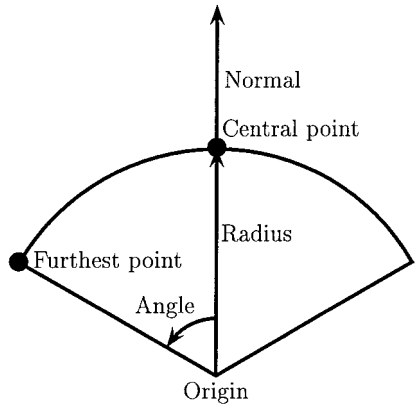

Figure 5. The descriptors that characterise a patch. If the patch is CONCAVE, the origin will be outside the surface, if CONVEX, inside. If FLAT, CYLINDER or SADDLE it may be on either side depending on the curvature values.

ure 4 would probably have a characteristic angle of about 90°. If the surface in the region of CP was more spherical, then the angle could be larger as more points would fall within the threshold distance of 1.1 times the patch radius. In fact if the surface was exactly spherical, one patch could describe it completely, with a characteristic angle of 360° even after trimming. It is, however, perhaps worth noting that the algorithm described above would not find such a patch, as all the surface points would be in one region which would not have any BORDER points. Luckily, completely spherical molecules are rare in medicinal chemistry.

The characteristic properties of a patch are shown in Figure 5. Typically, a surface containing around 1500 points can be described in approximately 50 patches, which contain all the salient features of the surface and thus greatly simplify its description.

Using the patches to match two surfaces

When the patches have been created for two surfaces, some means is required for comparing them and overlaying one on the other. This is done as follows:

1. Create a list of all the patches in the two surfaces that match. Two patches are deemed matching if they are in the same surface class (CONVEX, CONCAVE, etc.) and their two principal curvatures are equal within a tolerance. A value of 0.05 was used for the tolerance, a figure that was arrived at empirically.
2. Create a list of pairwise matches between all possible pairs of patches in the two surfaces. A pairwise match occurs between two pairs of patches if $patch_1$ in $surface_1$ matches $patch_1$ in $surface_2$ and $patch_2$ in $surface_1$ matches $patch_2$ in $surface_2$, the centres of the two pairs are an equal distance apart within a tolerance (1.0 Å in the current work), and their internal angles are such that it is possible that the two $patch_1$'s could overlap, at the same time as the two $patch_2$'s.
3. A correspondence table is created for all possible pairwise matches, with a 1 in the table for those pairwise combinations that match, and with a 0 otherwise.
4. A clique detection algorithm is used to find all cliques in the correspondence table with a minimum clique size of 3.
5. All cliques are taken and scored (described below) and the highest score corresponds to the clique that gives the best shape match between the two surfaces.
6. The two surfaces, with their associated molecules, are overlaid according to the highest scoring clique.

A clique is a maximal complete subgraph of a graph. Maximal means that it is not contained in any other complete subgraph and complete means that all nodes in the subgraph are connected to all other nodes. In the present context, the clique gives a set of patches in $surface_1$ that correspond to a set in $surface_2$. The clique detection algorithm used is that due to Bron and Kerbosch [29], which Brint and Willett have found effective for similar overlay purposes [30].

The cliques are scored by overlaying the surfaces according to the corresponding patches. This is done

by finding the transformation that gives the best RMS fit of the patch origins for $surface_1$ and the corresponding patch origins in $surface_2$ according to the clique. The minimum clique size is 3 because this is the smallest number of points that can give an unambigious overlay. The rigid-body transformation is found using the method of Diamond [31], with $surface_1$ being held stationary and points that comprise $surface_2$ transformed. The score is calculated by counting the number of points on $surface_2$ (the moving surface) that fall within 1 Å of a point on $surface_1$. The score is then 10 000 times the fraction of $surface_2$ points that are counted. This multiplication is arbitrary, and performed in order to make tables of scores more human-readable, and slightly more compact. In the case of a tie in the score, the mean distance between the corresponding points is used as a tiebreaker, with the overlay with the smaller mean distance, and hence with the surfaces closer together, being preferred. Because of the scoring method chosen, the overlay of molecule A on molecule B does not necessarily result in the same score as for the overlay of B onto A. For instance if A is much larger than B, the fraction of the area of A that is in close proximity to B will obviously be less than the fraction of B's surface that is close to A. The overlay itself, in terms of the relative positions of atoms in A and B, will be the same, however.

Finding a consensus overlay for a set of molecules

There frequently arise in the area of computer-aided molecular design situations where it is necessary to overlay a number of molecules into a common reference frame. When carrying out, for instance, CoMFA analyses [32], the relative orientations of the molecules within the set can have a dramatic effect on the final QSAR produced [33]. Most workers choose one molecule, possibly because it is the most rigid or the most active, and perform pairwise overlays onto that molecule by some means. However, the overlays of molecule B onto molecule A and C onto A may not produce the best overlay of B onto C. This is generally the case for SPAt overlays, and examples can be seen in Tables 2 and 3. For instance, overlaying all the molecules in that set onto 4tmn results in acceptable overlays for 1thl, 1tmn, 5tmn and 6tmn, but poor overlays for 1tlp, 3tmn and 5tln. Overlaying onto 1thl gives good results for 1tlp and 3tmn, but not for some of the molecules overlaid well onto 4tmn.

Several workers have previously addressed the problem of multi-molecule alignments. Kearsley [34] and Diamond [35] used similar methods based on minimising RMS values between pre-selected corresponding atoms, and Mestres et al. [36] have extended their MIMIC methodology to optimise the average molecular field similarity of all molecules in an input set.

We present here a novel, if somewhat time-consuming, method of resolving the multi-molecule overlay problem. All combinations of overlays are performed, resulting in a square matrix of results. A correspondence graph is formed in which the nodes represent the molecules and the (directed) edges the overlays required to achieve the best consensus overlay. Two molecules A and B are connected by an edge in the graph in the direction A→B if the highest scoring overlay for molecule A is onto B (B stationary, A moving). When the graph has been fully constructed, the overlays are carried out in an arbitrary order, ensuring that if a molecule is moved in a particular overlay, all molecules connected to it have the same transformation applied to them. Thus, for the rhinovirus inhibitors discussed below, whose overlay graph is shown in Figure 6, when 2rs5 is overlaid onto 1ruc, 2r04 and 1hrv are moved using the same transformation. In this way it does not matter whether, for instance, 1hrv is overlaid onto 2r04 (another required overlay) before or after the overlay of 2rs5. Occasionally it is seen that the highest score for molecule A is for molecule B and vice versa. In these cases, the overlay with the higher score is retained in the graph. This is in fact a special case of a more general problem that might arise, namely, that cycles are found in the graph. We have not yet observed this more general phenomenon, but one way of resolving the problem would be to remove from the cycle the lowest-scoring edge. This is effectively what has been done in the two molecule special case.

If the correspondence graph has no disconnections, the overlay is relatively straightforward, and the best result, according to the overlay scores, can be found by performing the overlays in the directions of the arrows. None of the results described in this paper falls into this category, however, and we have only seen this situation in small datasets of a few molecules. Instead, two or more disconnected sub-graphs are seen. This is not necessarily a disadvantage, however, as it indicates that the molecules have been found to have more than one common shape. We believe that it is advantageous that a set of molecules be split in this way if there is no large commonality of shape between them. Rather this, than try and force all the mole-

Table 2. RMS difference, overlaid to experimental for inhibitors of thermolysin

Molecules	1thl	1tlp	1tmn	3tmn	4tmn	5tln	5tmn	6tmn
1thl	n/a	0.78	0.35	0.48	0.40	6.77	1.08	5.31
1tlp	0.78	n/a	0.94	0.44	9.69	6.24	1.37	9.25
1tmn	0.46	0.97	n/a	1.06	0.88	7.34	0.89	0.68
3tmn	0.49	0.48	1.57	n/a	7.71	7.15	9.23	5.70
4tmn	0.30	7.31	0.57	5.96	n/a	8.24	1.97	2.06
5tln	6.33	6.01	8.52	5.84	10.81	n/a	9.73	9.93
5tmn	0.77	1.54	0.75	9.75	1.85	10.56	n/a	0.12
6tmn	5.66	9.01	0.54	9.93	2.22	6.32	0.12	n/a

Table 3. Raw overlay scores for thermolysin inhibitors

Molecules	1thl	1tlp	1tmn	3tmn	4tmn	5tln	5tmn	6tmn
1thl	n/a	8740	9331	6395	7537	5381	6077	4763
1tlp	8306	n/a	8183	6120	5937	5573	6312	4828
1tmn	9399	8281	n/a	5957	7836	4658	6438	6444
3tmn	9646	9720	8898	n/a	7203	6554	6669	7055
4tmn	6725	5562	6907	4158	n/a	4128	6863	6421
5tln	7319	7688	6132	6014	6493	n/a	5479	6682
5tmn	6112	6968	6308	4601	7322	3669	n/a	10000
6tmn	4494	4882	6416	4504	7012	4957	10000	n/a

cules into one set even if that means that the shape overlap is small. Notwithstanding this, the overlay procedure does propose the best way of uniting the sub-graphs, by finding the highest scoring inter-group overlay for all the groups. These overlays are shown by dotted lines in the correspondence graphs and provide a means for finding one consensus overlay for all the molecules.

Whilst this procedure is demonstrated here in the context of SPAt overlays, it is more generally applicable so long as the scoring function used is comparable from pair to pair, i.e., that if the score for the overlay of molecule A onto molecule B is greater than for the overlay of A onto C, it is fair to assume that the first overlay is better than the second.

Generating the surfaces

The method described in the preceding subsections is general, and can be used on any surface for which it is possible to find points on the surface with their associated normals. In the present method, the surface has been created by contouring the density of atom-based Gaussian functions. The particular Gaussians used are those proposed by Grant et al. in [9], and the density was computed using the formula

$$\rho^g = 1 - \prod_{i}^{n \text{ atoms}} (1 - \rho_i^g)$$

where ρ^g is the value of the Gaussian density at a point in space, and ρ_i^g is the ith atom's contribution to that density. The densities are calculated on a three-dimensional grid extending 4 Å beyond the molecule in each direction, with a spacing of 0.5 Å. The surface is then created by finding the three-dimensional contour at a density of 0.1 using the Marching TetraCubes algorithm of Chan and Purisima [37]. An advantage of using a contoured Gaussian surface is that the gradients of the function, and hence the normals to the surface, can be calculated analytically at very little extra cost.

A note about timings

On a modern workstation, such as an R5000 SGI O2, generation of the surface and patches for one of the molecules used in the tests described below takes of the order of 1 to 2 min. The time taken for the compar-

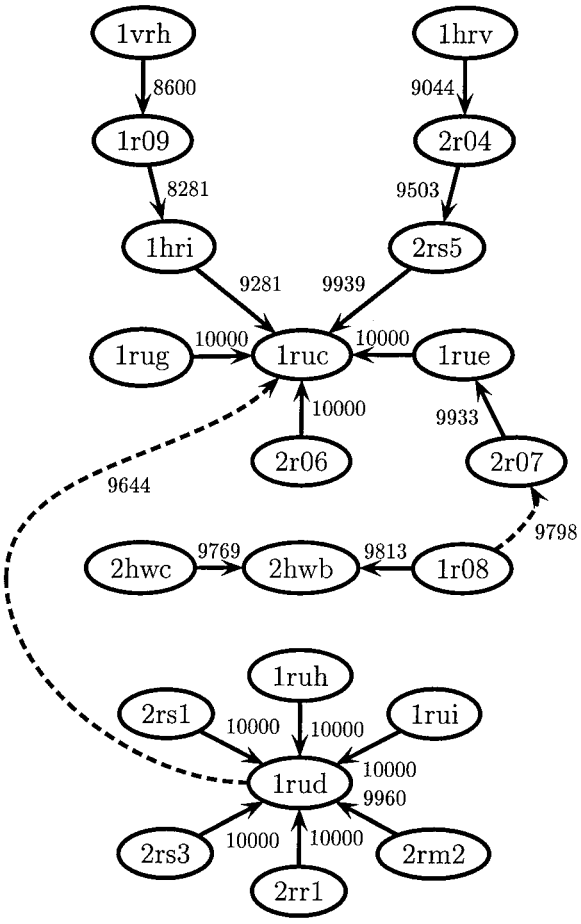

Figure 6. Overlay graph for inhibitors of rhinovirus HRV14.

ison of two surfaces depends very much on the number of cliques found. The rate-limiting step is the calculation of the score. Times can be from less than a minute if the surfaces are very dissimilar, to ten minutes or more if the surfaces show many points of similarity. The test cases reported below generally took several hours each to complete. This is clearly not competitive with other methods of molecular overlay. The program RigFit [7], for instance, performs flexible overlays on drug-sized molecules on average in less than 1 min. Although direct comparisons are difficult, given the improvements of CPU speed over the intervening years, the times for SPAt compare favourably with the times reported by Masek et al. for Skinny [38] which are of the order of 30 min for 1 pairwise overlay. This comparison is more relevant as Skinny also uses a surface-shape overlay, rather than the atom- and volume-based approach of RigFit.

Results

Initial tests have concentrated on using the surface shape overlay to predict the relative orientations of ligands bound to the same protein active site, since these are the only systems for which there is strong experimental data on which to judge the results of the overlay process. To this end, several series of ligands were extracted from the Brookhaven database, where, in each series, the ligands were bound to the same or a similar protein. Within a series, the protein structures containing the ligands were aligned by performing an RMS fit on the backbone atoms of matching residues, and the ligands separated from the protein structure. The result of this was a set of ligands aligned as they have been found to bind in the appropriate active site. The test was then to see whether SPAt could reproduce the relative orientations, using the experimental conformations for the molecules. We have not, as yet, explored the possibilities and difficulties of molecular flexibility. In principle, this could be tackled by generating populations of low-energy conformations for each molecule and using the rigid overlays described herein on each conformation. We are currently investigating this area.

In the subsequent discussions the RMS errors reported are all the RMS difference in the atomic coordinates for the molecule in its overlaid orientation as found by SPAt and the original experimental coordinates derived from the aligned protein structures. Because of the way the overlay procedure works, with all the molecules ultimately being overlaid onto one molecule (directly or indirectly), one molecule within each overlaid group will have an RMS error of 0.0 – it has not been moved from its experimental (input) position.

Inhibitors of a coat protein from human rhinovirus HRV14

The Brookhaven database has 21 structures of inhibitors bound to the same site in protein HRV14. Twelve of these are complexed with the native form, and 9 with various mutants. Of the 9 ligands complexed to mutant proteins, there are only 5 distinct compounds, with 2 compounds appearing in 3 structures each. All 21 compounds were used, since the duplicate structures showed small conformational differences, and thus had slightly different shapes. There are strong similarities in the structures of the inhibitors, some of which are shown in Figure 7. The

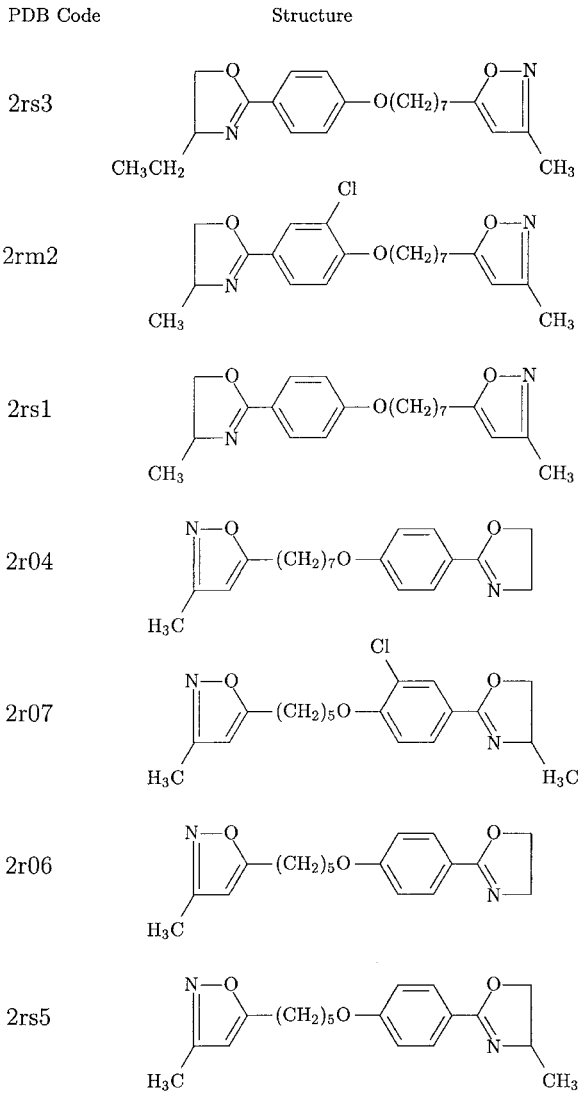

Figure 7. Some inhibitors of rhinovirus coat protein HRV14 drawn in the relative orientations found in the protein active site.

molecules are drawn in the relative orientations they are found within the active site. The remarkable thing about these inhibitors is that extremely similar molecules, such as 2rm2 and 2r07 (which differ only by the number of methylene groups in the alkyl chain), bind with their 'similar ends' at opposite ends of the protein active site. To our knowledge, no one has yet predicted the correct relative orientations of all 21 of these inhibitor structures, although a number of groups have reported success with some subsets.

Table 4 shows the RMS differences between the overlaid position of each molecule obtained using SPAt and its experimentally determined coordinates,

Table 4. RMS differences and maximum distances between corresponding atoms in the overlaid and experimental conformations for the SPAt overlays of rhinovirus HRV14 inhibitors

Molecule	RMS	Max. dist.	Group
1hri	0.38	0.73	1
1r09	4.71	4.77	1
1vrh	1.46	2.30	1
1ruc	0.00	0.00	1
1rue	0.37	0.57	1
2r07	0.49	0.68	1
1rug	0.02	0.03	1
2r06	0.04	0.06	1
2rs5	0.65	0.75	1
2r04	0.74	1.50	1
1hrv	0.93	1.17	1
1r08	0.50	0.80	2
2hwb	0.00	0.00	2
2hwc	0.36	0.73	2
1rud	0.00	0.00	3
1ruh	0.35	0.56	3
1rui	0.05	0.06	3
2rm2	0.41	0.52	3
2rr1	0.20	0.31	3
2rs1	0.04	0.07	3
2rs3	0.32	0.61	3

Figure 8. Experimental (red) and SPAt-overlaid (green) configurations of HRV14 inhibitor 1r09. Experimental configuration of 1hri in blue.

and the maximum distance of an atom from its corresponding atom. The worst RMS value is 4.71 Å for 1r09. It can be seen from Figure 8, which shows the experimental (red) and SPAt-overlaid (green) configurations for 1r09, that the orientation of the molecule is correct, and the error is due in most part to a translation along the long axis of the molecule. The error has come about because 1r09 is overlaid onto 1hri (blue in Figure 8), by aligning the two benzene rings on the two molecules, which is not seen in the experimental result. The second worst overlay is that of 1vrh, with an RMS of 1.46 Å. As can be seen from Figure 6, these two molecules also had the lowest overlay scores. The dashed lines in Figure 6 show the best-scoring over-

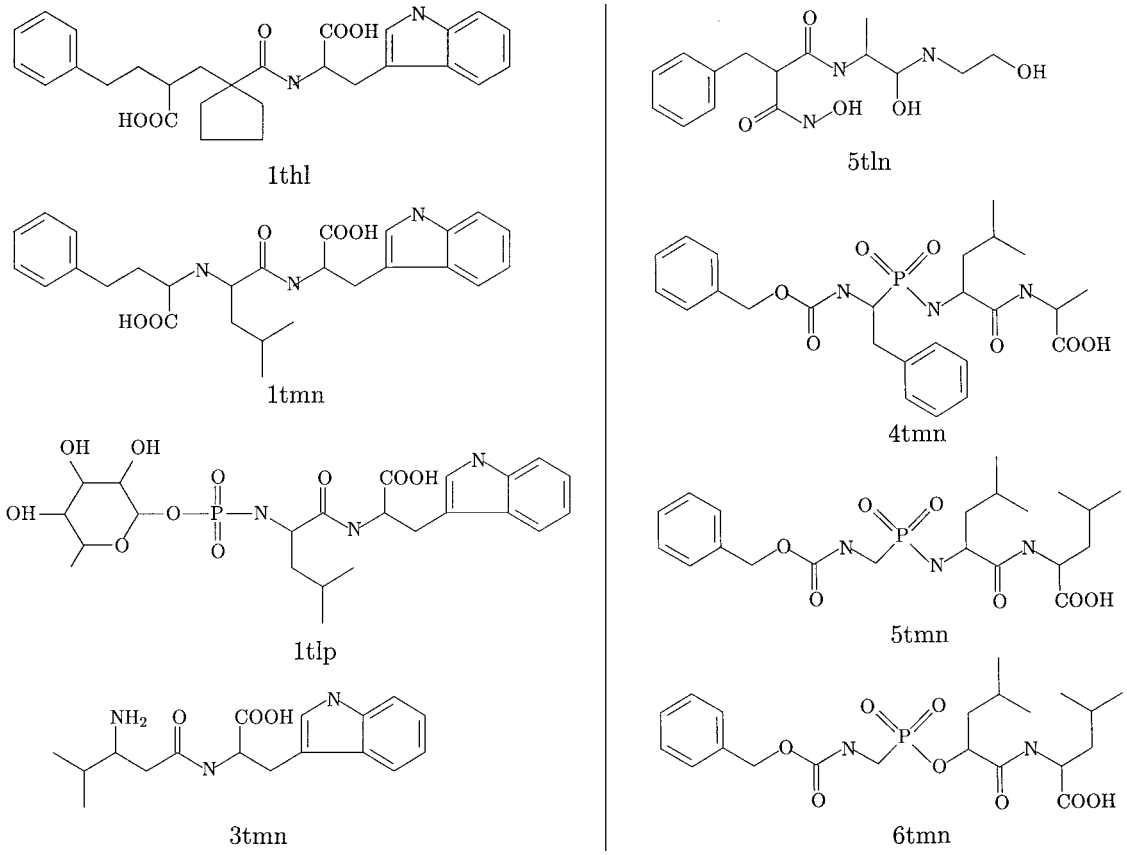

Figure 9. 2D structures of thermolysin inhibitors.

lays that might be used to combine the groups into one overlay. The scores for these overlays are on a par with the scores for the intra-group overlays, and better than the two worst intra-group overlays. On this basis, one might expect that the final overlay of all 21 molecules would be successful, and this is borne out by the results, as given in Table 5. The RMS values for the molecules moved in the inter-group overlay (groups 2 and 3) have deteriorated slightly, but are still very good.

Böhm and Klebe have discussed the binding of two of these inhibitors [39] (2rs1 and 2r04), their molecules 14 and 15 respectively, and failed to provide an explanation for the fact that they bind in opposite modes despite their very close similarities, which extend to their volumes and electrostatic potentials. The SPAt program has successfully predicted the binding of these two inhibitors solely on the basis of surface shape. The point is worth making, however, that the resolutions for these structures are between 2.9 Å and 3.0 Å. It would therefore seem unwise to draw too strong a conclusion from the SPAt results based on these data.

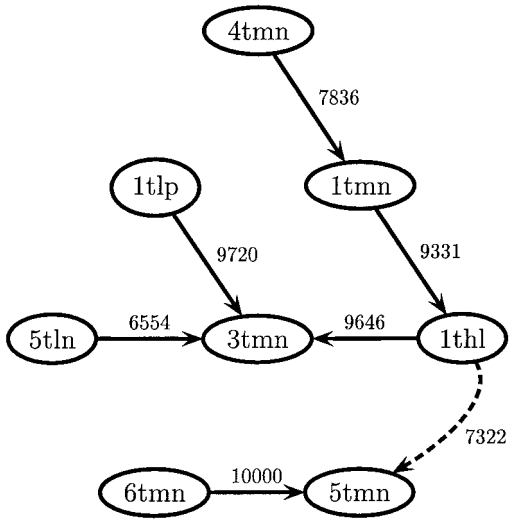

Figure 10. Overlay graph for inhibitors of thermolysin.

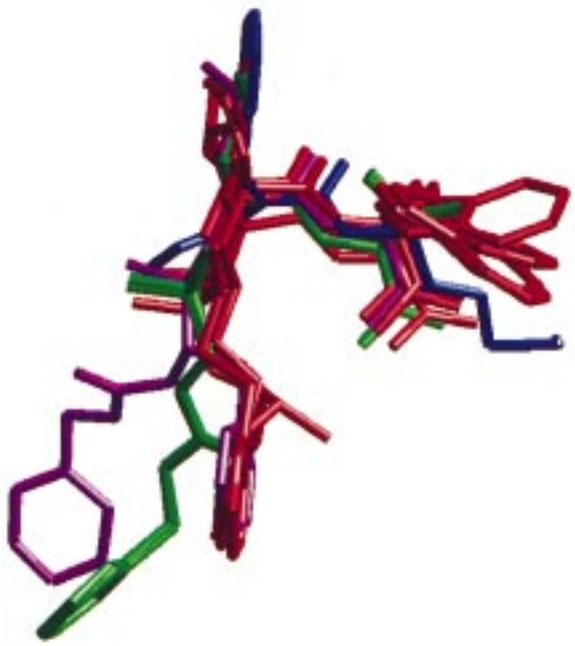

Figure 11. Experimental alignments of all 8 thermolysin inhibitors. Group 2 inhibitors in green, group 1 in red apart from 4tmn (purple) and 5tln (blue).

Figure 12. Experimental and overlaid alignments of 8 thermolysin inhibitors – experimental results in red (5tln blue), overlaid results in green (5tln purple).

Inhibitors of thermolysin

Eight thermolysin inhibitors were found in the Brookhaven database, and the ligands extracted from the protein active sites. Their chemical structures are shown in Figure 9, and the SPAt overlay graph is shown in Figure 10. This graph is also disconnected, consisting of 2 sub-graphs of 6 and 2 molecules, respectively. The RMS values for the overlaid to experimental positions are shown in Table 6, for overlays within the two sub-graphs. With the exception of 5tln, the results are again highly encouraging. As shown in Figure 9, the overlay score for 5tln onto 3tmn, 5tln's best scoring overlay, is substantially lower than for the other overlays in its group.

Figure 11 shows the experimental alignments of all 8 molecules, with the group 1 molecules in red, blue and purple, and the group 2 molecules in green. It can be seen that the group 2 molecules are sufficiently different that the splitting into two groups is reasonable. The purple molecule is 4tmn. At first sight, it appears to be different from the other group 1 molecules, in that it has a phenyl group on a chain out to the left that the other molecules do not have. However, close examination shows that this group is 'extra' in that it also has the phenyl group in the middle of the figure possessed by the other group 1 molecules. Figure 11 also shows that 4tmn does not possess the indole moiety at the right of the figure that the other group 1 molecules have. It is all the more pleasing then, that 4tmn is overlaid so well by SPAt (RMS error of 1.12 Å). The blue molecule in Figure 11 is 5tln. It is also somewhat different from the other molecules, being rather smaller. It is perhaps disappointing that it has not been placed in group 2, since there seems to be a significant overlap of the chains.

The best-scoring inter-group overlay, as shown in Figure 10, is to overlay 1thl onto 5tmn. The RMS results after this overlay are given in Table 7. Again, they are not as good as the intra-group figures, but, with the exception of 5tln, still quite promising. The final overlays are shown in Figure 12, with the experimental structures in red (5tln in blue) and the overlaid structures in green (5tln in purple).

Inhibitors of elastase

Thirteen elastase inhibitors were extracted from the PDB database. They were all high-quality (resolutions < 2 Å) structures of porcine serine protease with inhibitors containing 25 or more heavy atoms. Figure 13 shows the 2D structures of the inhibitors, and Figure 14 the SPAt overlay graph. Table 8 contains the RMS values for the intra-group overlays, along with the group details. There are three groups: group 1, the largest, has 9 molecules in it, of which 5 have been well overlaid, and 4 rather less successfully. This set is extremely challenging, as can be seen from Figure 15, which shows all 13 in the same active site reference frame. There is very little commonality discernible between the 13 molecules, which occupy a relatively large volume. Inspection of the individual structures

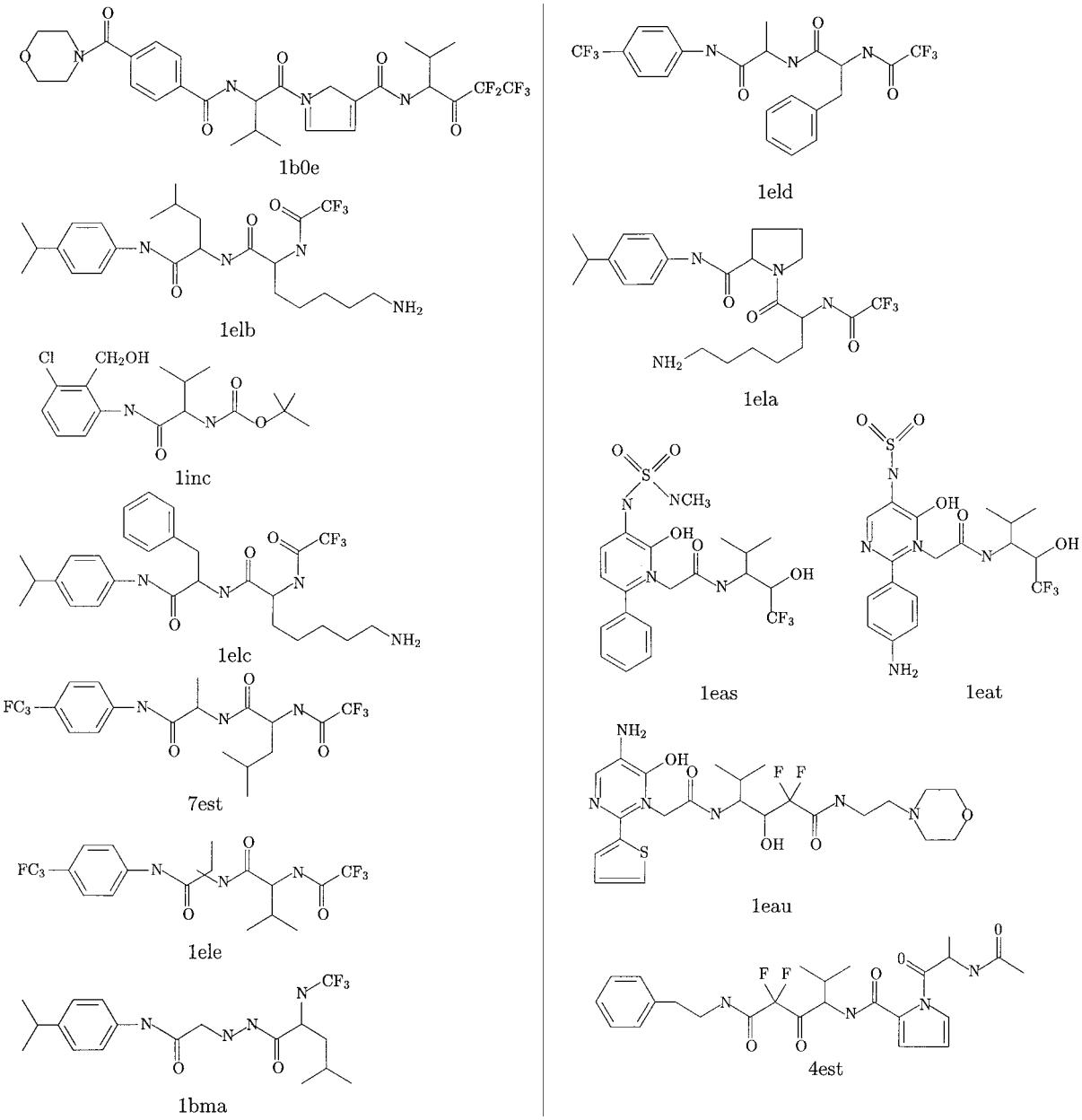

Figure 13. 2D structures of elastase inhibitors.

shows that the molecules project in a number of different directions. The group 1 overlays are shown in Figure 16, with the 5 well-overlaid molecules in red, and the other 4 in green. The experimental orientation of 1ela (the most poorly overlaid of the 'good 5') is shown in blue. The other two groups, of two molecules each, have reasonable overlays. It is notable, as seen in Figure 14, that in group 1 the four poor overlays have substantially lower scores than the 5 good ones. The inter-group overlays involve first overlaying 1eau onto 1eas, and then 1eau onto 1inc. The first of these produces a good overlay, but the second a very poor one, such that the 4 molecules in groups 2 and 3 end up with RMS values of between 6.7 and 7.5 Å. Although the score for the overlay of 1eau onto 1inc has a score of 7735, it is by no means the lowest scoring overlay in Figure 14. It would be difficult to argue convincingly that one would have predicted in advance the poorness

Table 5. RMS differences and maximum distances between corresponding atoms in the overlaid and experimental conformations for the final SPAt overlays of rhinovirus HRV14 inhibitors

Molecule	RMS	Max. dist.
1hri	0.38	0.73
1r09	4.71	4.77
1vrh	1.46	2.30
1ruc	0.00	0.00
1rue	0.37	0.57
2r07	0.49	0.68
1rug	0.02	0.03
2r06	0.04	0.06
2rs5	0.65	0.75
2r04	0.74	1.50
1hrv	0.93	1.17
1r08	0.97	1.48
2hwb	0.80	1.65
2hwc	0.62	1.04
1rud	0.79	1.47
1ruh	1.04	1.88
1rui	0.79	1.47
2rm2	1.01	1.45
2rr1	0.80	1.41
2rs1	0.80	1.46
2rs3	0.86	1.34

Table 6. RMS differences and maximum distances between corresponding atoms in the overlaid and experimental conformations for the intra-group overlays of thermolysin inhibitors

Molecule	RMS	Max. dist.	Group
1thl	0.46	0.68	1
1tmn	0.00	0.00	1
1tlp	0.94	1.34	1
5tln	6.24	9.67	1
4tmn	0.88	1.72	1
3tmn	0.55	0.86	1
5tmn	0.12	0.16	2
6tmn	0.00	0.00	2

Table 7. RMS differences and maximum distances between corresponding atoms in the overlaid and experimental conformations for the final overlays of thermolysin inhibitors

Molecule	RMS	Max. dist.
1thl	0.83	1.43
1tmn	0.77	1.50
1tlp	0.54	0.75
5tln	7.33	11.14
4tmn	1.85	3.65
3tmn	0.79	1.19
5tmn	0.00	0.00
6tmn	0.12	0.16

Table 8. RMS differences and maximum distances between corresponding atoms in the overlaid and experimental conformations for the intra-group overlays of elastase inhibitors

Molecule	RMS	Max. dist.	Group
1b0e	5.14	9.04	1
1elb	4.66	7.81	1
1inc	6.75	10.01	1
1elc	5.51	9.41	1
7est	0.28	0.47	1
1ele	0.00	0.00	1
1bma	0.31	0.41	1
1eld	0.52	0.97	1
1ela	0.72	1.32	1
1eas	0.33	0.43	2
1eat	0.00	0.00	2
1eau	0.81	1.64	3
4est	0.00	0.00	3

of the final overlay, as one might, perhaps, have done with the four bad intra-group overlays in group 1.

Kinase inhibitors

The Brookhaven database contains 13 kinase proteins with ligands bound to the active site. They are not all the same protein, but the ATP binding site is well conserved, and the alignments were performed using the ATP binding residues only. Figure 17 shows the 2D structures of the ligands. The overlay graph is displayed as Figure 18 and Table 9 contains the RMS results for the overlays within the three groups

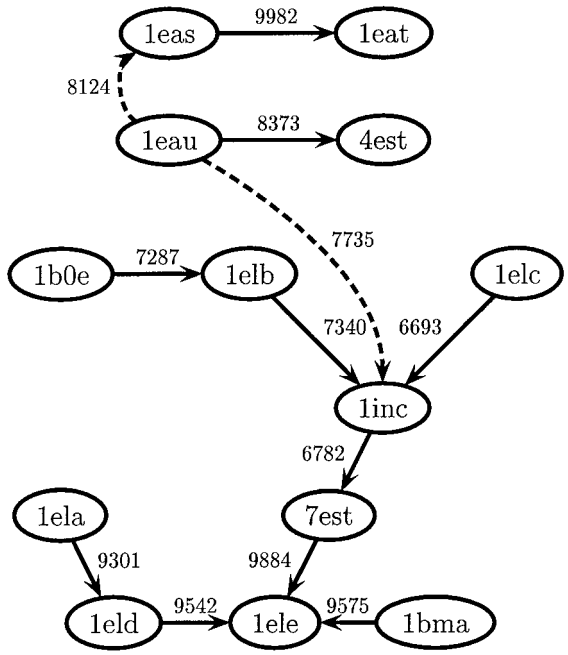

Figure 14. Overlay graph for inhibitors of elastase.

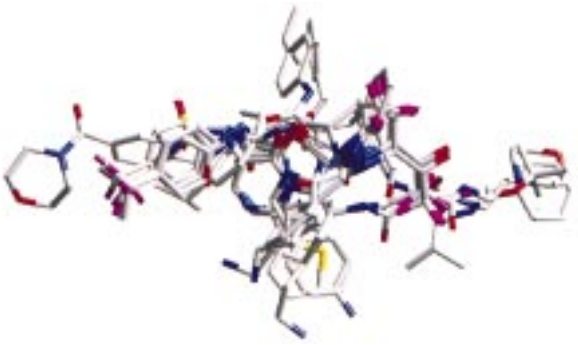

Figure 15. Experimental configurations of all 13 elastase inhibitors.

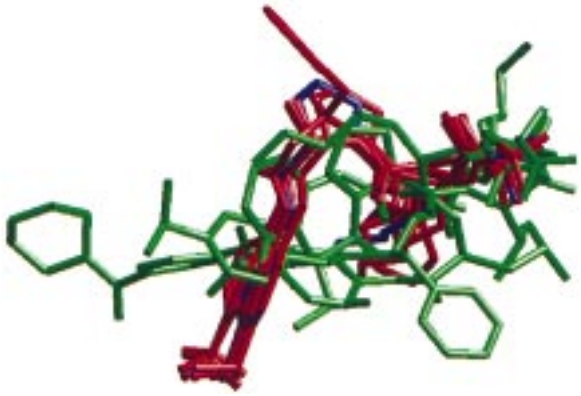

Figure 16. Overlays of group 1 elastase inhibitors, with well overlaid structures in red, 4 poorly overlaid in green. Experimental configuration of 1ela in blue.

Table 9. RMS differences and maximum distances between corresponding atoms in the overlaid and experimental conformations for the intra-group overlays of kinase inhibitors

Molecule	RMS	Max. dist.	Group
1agw	3.34	4.74	1
1fgi	3.68	6.20	1
1bkx	4.85	8.67	1
1atp	9.25	16.51	1
2csn	0.00	0.00	1
1ian	9.72	18.70	1
1aq1	1.13	1.96	2
1stc	0.00	0.00	2
1ckp	5.81	9.93	3
1ydr	0.43	0.72	3
1yds	0.00	0.00	3
2hck	4.93	8.60	3
1ydt	1.04	2.18	3

found. The results show that in this case SPAt has been largely unsuccessful. The results after the inter-group overlay was performed were even worse and are not presented. Two possible reasons for the bad performance in this case, other than the differences in the proteins themselves, are that the ligands are quite widely diverse compared with the other 3 test cases, and the kinase active site is large and loosely defined. There are 4 recognised pockets within it, and none of the published inhibitors occupy all four [40]. There is thus very little commonality of shape with the 13 structures, which is amply demonstrated by the poor results.

Discussion and conclusions

The surface overlay method described herein has performed more than adequately in three of the four test cases investigated. Of particular note is the ability to identify correctly the binding mode of 20 of the 21 HRV14 rhinovirus inhibitors and find the correct relative orientation of the remaining one. To our knowledge, this has not been achieved before and is all the more noteworthy in that it has been achieved on the basis of shape alone – there is no chemical component to the molecular description at all aside from that implicit in the shape engendered by the makeup of the molecules. Whilst the shape-based overlays are clearly successful in many of the pairwise overlays, we be-

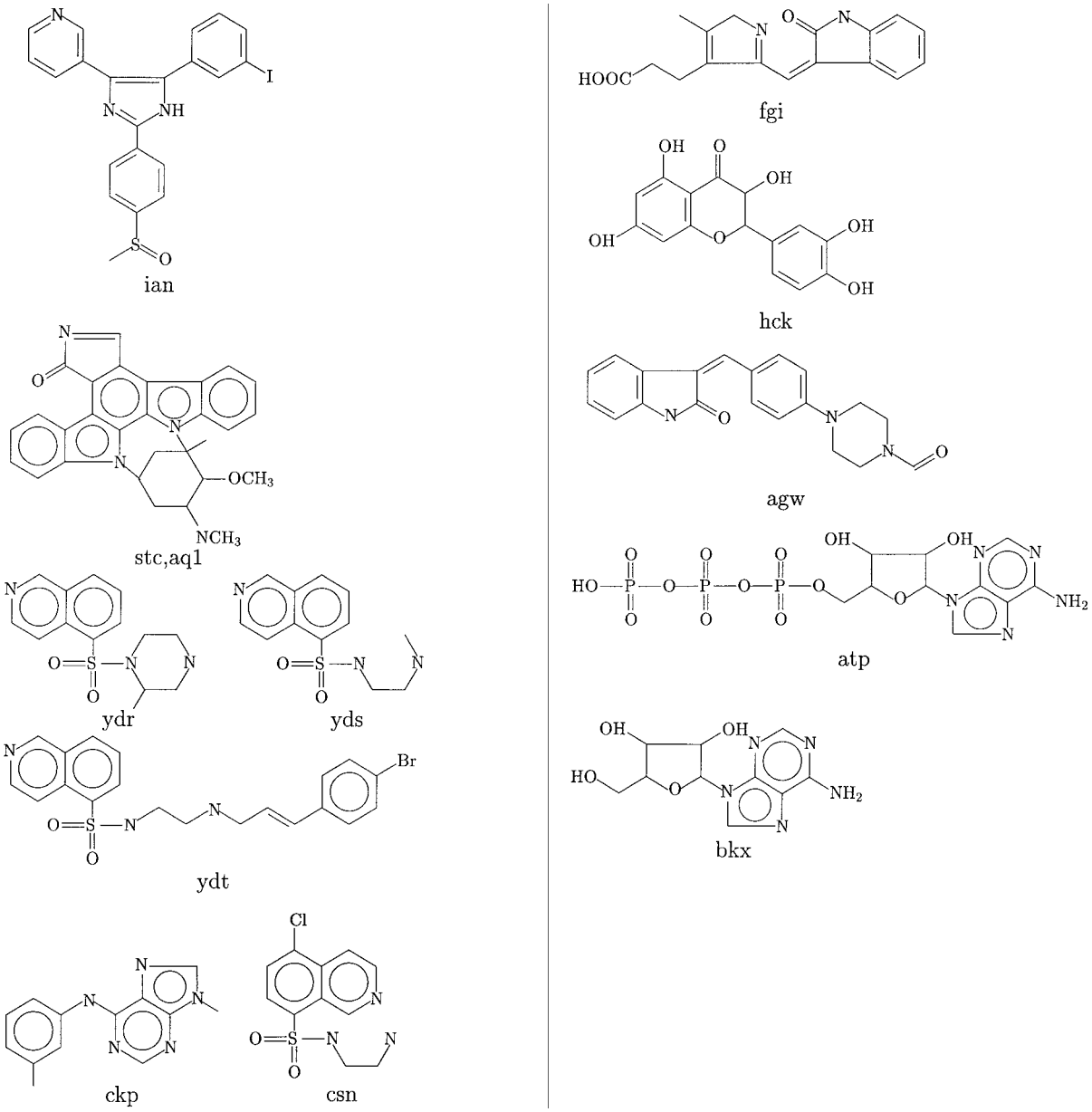

Figure 17. 2D structures of kinase inhibitors.

lieve that the overall success for these overlays is due in at least as much part to the consensus overlay. There is no molecule in the set for which all other molecules are overlaid onto it correctly. It is only by selecting a particular set of pairwise overlays prioritised by the overlay scores that these results are achieved. It is not impossible that other overlay techniques used in conjunction with the consensus procedure we have developed would be equally successful on this test set.

In the elastase and thermolysin examples, SPAt divides the molecules into several groups, based on the shape. Examination of these groups shows that the separations are reasonable. We believe that this behaviour is desirable in that SPAt will not attempt to force an overlay that is not warranted. However, a possible resolution of the groups into one overlay is available, and the user has the opportunity to decide whether to accept it or not. The final two examples demonstrate that the shape overlay procedure does not

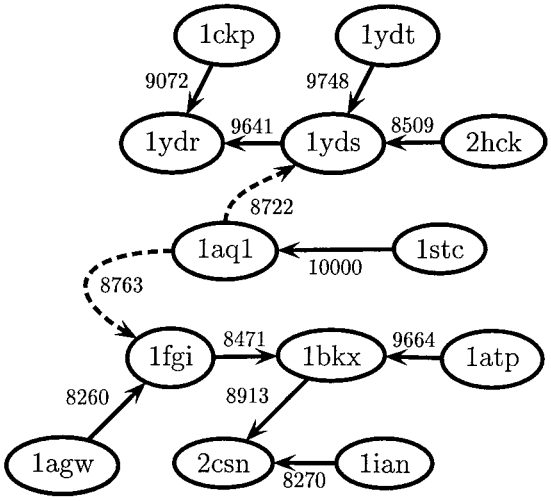

Figure 18. Overlay graph for inhibitors of kinase.

function well with a set of compounds that are diverse in size and binding mode, but is able to extract a set of similar shapes of roughly homologous molecules from a diverse set. They also demonstrate that the SPAt procedure is capable of making successful alignments on the basis of partial, rather than global, shape similarity. It is generally, although not always, the case that the poor overlays can be identified by markedly lower overlay scores in the correspondence graph.

Clearly, the tests described in this paper are the simplest imaginable, since the crystallographically determined conformations of the molecules were used. Whilst providing a demonstration of the effectiveness of the SPAt method for performing shape-based overlays, they are of limited utility in themselves. Future work will address the issue of conformational flexibility. Another area for future work is the addition of chemical information onto the surface. Whilst shape is an important factor in determining ligand binding at a receptor, clearly the chemical composition is also critical, an observation that is borne out by the kinase results. The incorporation of additional information into the surface description could be achieved by labelling the surface patches according to the nature of the underlying atoms and building the labels into the procedures for establishing patch equivalences. We are currently investigating these possibilities.

Acknowledgements

D.A.C. acknowledges the generous support in funding and time of Zeneca Pharmaceuticals and latterly AstraZeneca. Philip Jewsbury has provided valuable discussions on kinase inhibitors. Tony Slater and Dave Timms were extremely helpful in assembling the rhinovirus and elastase datasets. The authors are grateful to Andrew Grant for his meticulous reading of an early draft and his many thoughtful suggestions.

References

1. Good, A.C. and Richards, W.G., Perspect. Drug Discov. Design, 9–11 (1998) 321.
2. Bohacek, R.S. and Guida, W.C., J. Mol. Graph., 7 (1989) 113.
3. Stouch, T.R. and Jurs, P.C., J. Chem. Inf. Comput. Sci., 26 (1986) 4.
4. Marsili, M., Floersheim, P. and Dreiding, A.S., Comput. Chem., 7 (1983) 175.
5. Masek, B.B., Merchant, A. and Matthew, J.B., J. Med. Chem., 36 (1993) 1230.
6. Good, A.C., Hodgkin, E.E. and Richards, W.G., J. Chem. Inf. Comput. Sci., 32 (1992) 188.
7. Lemmen, C., Hiller, C. and Lengauer, T., J. Comput.-Aided Mol. Design, 12 (1998) 491.
8. Kearsley, S.K. and Smith, G.M., Tetrahedron Comput. Methodol., 3 (1990) 615.
9. Grant, J.A., Gallardo, M.A. and Pickup, B.T., J. Comput. Chem., 17 (1996) 1653.
10. Mestres, J., Rohrer, D.C. and Maggiora, G.M., J. Comput. Chem., 18 (1997) 934.
11. Carbo, R., Leyda, L. and Arnau, M., Int. J. Quantum Chem., 17 (1980) 1185.
12. Nissink, J.W.M., Verdonk, M.L., Kroon, J., Mietzner, T. and Klebe, G., J. Comput. Chem., 18 (1997) 638.
13. Klebe, G., In Kubinyi, H. (Ed.), 3D QSAR in Drug Design: Theory, Methods and Applications, ESCOM, Leiden, 1993, pp. 173–199.
14. Bode, W., Wei, A., Huber, R., Meyer, E., Travis, J. and Neumann, S., EMBO J., 5 (1986) 2453.
15. Takahashi, L.H., Radhakrishnan, R. and Rosenfield Jr., R.E., J. Am. Chem. Soc., 111 (1989) 3368.
16. Hermann, R.B. and Herron, D.K., J. Comput.-Aided Mol. Design, 5 (1991) 511.
17. Masek, B.B., Merchant, A. and Matthew, J.B., Proteins, 17 (1993) 193.
18. Perkins, T.D.J., Mills, J.E.J. and Dean, P.M., J. Comput.-Aided Mol. Design, 9 (1995) 479.
19. Grant, J.A. and Pickup, B.T., In van Gunsteren, W.G., Weiner, P.K. and Wilkinson, A.J. (Eds), Computer Simulation of Biomolecular Systems. Theoretical and Experimental Applications, Vol. 3, Ch. 5, Kluwer/ESCOM, Dordrecht, 1997, pp. 150–176.
20. Dean, P.M. and Chau, P.-L., J. Mol. Graph., 5 (1987) 97.
21. Dean, P.M. and Chau, P.-L., J. Mol. Graph., 5 (1987) 153.
22. Dean, P.M. and Callow, P., J. Mol. Graph., 5 (1987) 159.
23. Poirrette, A.R., Artymiuk, P.J., Rice, D.W. and Willett, P., J. Comput.-Aided Mol. Design, 11 (1997) 557.
24. Rosen, M., Lin, S.L., Wolfson, H.J. and Nussinov, R., Protein Eng., 11 (1998) 263.
25. Norel, R., Petrey, D., Wolfson, H.J. and Nussinov, R., Proteins, 36 (1999) 307.
26. Walker, P.D., Ateca, G.A. and Mezey, P.G., J. Comput. Chem., 12 (1991) 220.

27. Lee, B. and Richards, F.M., J. Mol. Biol., 55 (1971) 379.
28. McHugh, J., Algorithmic Graph Theory, Prentice-Hall International Editions, New York, NY, 1990, pp. 90–114.
29. Kerbosch, J. and Bron, C., Commun. ACM, 16 (1973) 575.
30. Brint, A.T. and Willett, P., J. Chem. Inf. Comput. Sci., 27 (1987) 152.
31. Diamond, R., Acta. Crystallogr., A44 (1988) 211.
32. Cramer III, R.D., Patterson, D.E. and Bunce, J.D., J. Am. Chem. Soc., 110 (1993) 5959.
33. Cramer III, R.D., DePriest, S.A., Patterson, D.E. and Hecht, P., In Kubinyi, H. (Ed.), 3D QSAR in Drug Design. Theory, Methods and Applications, ESCOM, Leiden, 1993, pp. 443–485.
34. Kearsley, S.K., J. Comput. Chem., 11 (1990) 1187.
35. Diamond, R., Protein Sci., 1 (1992) 1279.
36. Mestres, J., Maggiora, G.M. and Rohrer, D.C., J. Mol. Graph., 15 (1997) 114.
37. Purisima, E.O. and Chan, S.L., Comput. Graph., 22 (1998) 83.
38. Masek, B.B., Merchant, A. and Matthew, J.B., J. Med. Chem., 36 (1993) 1230.
39. Böhm, H.-J. and Klebe, G., Angew. Chem. Int. Ed. Engl., 35 (1996) 2588.
40. Jewbury, P.J., Private communication.