"Shrink-Wrap" Surfaces: A New Method for Incorporating Shape into Pharmacophoric 3D Database Searching

John H. Van Drie

Pharmacia & Upjohn, Inc., Kalamazoo, Michigan 49008

Received June 5, 1996[⊗]

A new method for representing a molecular surface is described, the "shrink-wrap surface", which allows a simple and rapid computation of steric overlap. We describe how this surface may be used graphically, how it may be used in a 3D database search, and how such surfaces may be developed for a pharmacophoric 3D database search query based on the conformers of a set of active molecules.

INTRODUCTION

Two general categories of 3D database searching have been developed: (1) shape-based methods such as DOCK[1] which uses as a database query a description of the shape of a protein pocket as determined by X-ray crystallography, and (2) pharmacophore-based methods such as ALADDIN[2] which use as a database query a set of geometric and topological constraints. These methods are distinct, in that the former treats the shape constraints as primary, and only recently have they been extended to include chemical constraints, while the latter treats the chemical constraints as primary and allows shape constraints to be included in a primitive way. In ALADDIN, two methods were described for including shape constraints: "dummy" atoms, and the CLASH routine. With dummy atoms (otherwise called "outrigger" atoms[3]), one defines a location in space relative to 3 or 4 features of the pharmacophore and defines an exclusion sphere around that point into which no atoms of a candidate molecule may protrude. With CLASH, candidate molecules were translated/rotated into a common reference frame and tested for their overlap with a series of point on a 3D grid which defined the excluded volume. Both of these methods have their own set of difficulties in practical applications of pharmacophoric 3D database searching; indeed, the fact that they are both rarely used at all is a testament to their impracticality. The need exists to extend the methodology of pharmacophoric 3D database searching to allow the description of shape, which (a) allows steric clashes to be quickly determined, (b) allows the magnitude of the clash to be computed, (c) allows subtle definition of the shape, (d) allows robust discrimination by this shape to be performed, and (e) may be readily computed. To this end, we describe a new method for representing surfaces, the "shrink-wrap representation",[4] and describe its application to pharmacophoric 3D database searching.

THE SHRINK-WRAP REPRESENTATION OF A SURFACE

Begin with a standard definition of a molecular surface, either the van der Waals surface or the Connolly−Richards surface.[5] The result of either such computation is a set of Cartesian coordinates $\{x_i, y_i, z_i\}_{i=1}^{N}$. Compute the geometric center of this surface

$$\bar{x} = \frac{1}{N}\sum x_i, \quad \bar{y} = \frac{1}{N}\sum x_i, \quad \bar{z} = \frac{1}{N}\sum z_i$$

and convert the coordinates of the N points to a spherical polar representation, using

$$(\bar{x},\bar{y},\bar{z})$$

as the origin. From this set of points in spherical coordinates $\{r_k, \theta_k, \phi_k\}_{k=1}^{N}$ compute an array of radii

$$\{r(\Theta_i,\Phi_j)\}_{i,j=1}^{\sigma}$$

as follows: Divide the range of ϕ, $[0,2\pi]$, into σ equal steps. Similarly divide the range of θ, $[0,\pi]$, into σ equal steps. σ is a user-adjustable resolution parameter. For each point (r_k, θ_k, ϕ_k) determine which increment of solid angle is nearest, i.e., find i and j such that

$$\Theta_i \approx \theta_k \quad \Phi_j \approx \phi_k$$

If no value of $r(\Theta_i,\Phi_j)$ has been assigned, assign $r(\Theta_i,\Phi_j) = r_k$. If $r(\Theta_i,\Phi_j)$ has been previously assigned, and $r_k > r(\Theta_i,\Phi_j)$, assign $r(\Theta_i,\Phi_j) = r_k$, else skip this point. If, after processing all N points, "holes" remain (i.e., values of i and j for which $r(\Theta_i,\Phi_j)$ is undefined), these holes are "patched" using the following logic:

Define $r(\Theta_{i+1},\Phi_j)$ as *above*
Define $r(\Theta_{i-1},\Phi_j)$ as *below*
Define $r(\Theta_i,\Phi_{j+1})$ as *left*
Define $r(\Theta_i,\Phi_{j-1})$ as *right*
Define as *neighbors* the set *above*, *below*, *left*, *right*.
If(*above* is defined AND *below* is defined AND
  left is defined AND *right* is defined)
  $r(\Theta_i,\Phi_j) = 1/4$(*above* + *below* + *left* + *right*)
If (only 3 of *neighbors* are defined)
  $r(\Theta_i,\Phi_j) = 1/3$(sum of 3 defined *neighbors*)
If(only 2 of *neighbors* are defined)
  $r(\Theta_i,\Phi_j) = 1/2$(sum of 2 defined *neighbors*)

This logic is applied until all $r(\Theta_i,\Phi_j)$ are defined. This logic may need to be applied multiple times, but in practice that is rare for reasonable values of σ and high densities of Cartesian points of the original surface. The net result has been to convert a set of N Cartesian coordinates to an array of σ^2 radii. Typical values of σ are 36. Higher values allow the surface to be represented with more refinement; lower values afford a coarser surface, but one for which computa-

[⊗] Abstract published in *Advance ACS Abstracts,* January 1, 1997.

S0095-2338(96)00464-7 CCC: $14.00

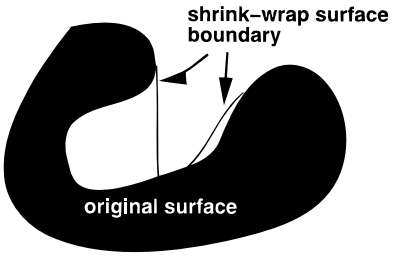

Figure 1. Display of how shrink-wrap surface may be not coincident with original molecular surface. In solid black is the original surface. The thin lines indicate where the shrink-wrap surface is not coincident with the original surface.

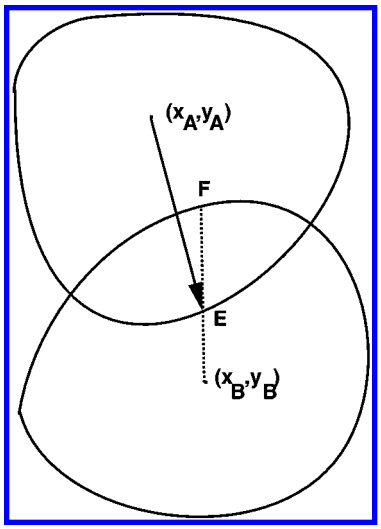

Figure 2. Detection of interpenetration of two shrink-wrap surfaces. Point E is a point of the shrink-wrap surface centered at (x_A, y_A). Referring to the shrink-wrap surface centered at (x_B, y_B), the point of that surface which has the same Θ_j is point F. Since the distance of E from (x_B, y_B) is less than $r_B(\Theta_j)$, the point E penetrates the surface B.

tions may be performed more rapidly. Notice that some information is lost during this conversion. Figure 1 shows the net effect of converting an original shape as depicted into its shrink-wrap form. The loss of definition of some invaginations is reminiscent of that seen in the plastic packaging of odd shapes which is called shrink-wrapping; hence the name.

GRAPHIC DISPLAY OF SHRINK-WRAP SURFACE

Each set of four points $r(\Theta_i,\Phi_j)$, $r(\Theta_{i+1},\Phi_j)$, $r(\Theta_i,\Phi_{j+1})$, $r(\Theta_{i+1},\Phi_{j+1})$ defines the four points of a polygon (where $i+1$ and $j+1$ are understood to be mod σ). Normals are taken in the direction from the origin to each polygon point. This set of polygons defines a surface which is topologically equivalent to a sphere. For graphic display purposes, to view the molecules inside this surface, one can either display the surface transparently or "create a window" by displaying only over a limited range of Θ_i.

CALCULATING STERIC OVERLAP BETWEEN TWO SHRINK-WRAP SURFACES

Let us first consider this problem in two dimensions, as shown in Figure 2. Consider a shrink-wrap surface A centered at (x_A, y_A), described by a set of radii $r_A(\Theta_i)$, and another shrink-wrap surface B centered at (x_B, y_B) described by a set of radii $r_B(\Theta_i)$. The A surface intersects the B surface if any surface point of A (such as point E in Figure 2) is interior to the B surface. Point E is interior to B if it is closer to the center of B than point F, the spherical projection of E onto the B surface. Each surface point of A is tested by computing its Cartesian coordinates

$$x = r_A \cos(\Theta_A) + x_A$$

$$y = r_A \sin(\Theta_A) + y_A$$

and converting these to polar coordinates *relative to the center of B*:

$$r_B = \sqrt{(x - x_B)^2 + (y - y_B)^2}$$

$$\Theta_B = \arctan\frac{y - y_B}{x - x_B}$$

This is compared with the projection of E onto the B surface, point F. Point F is determined by selecting that point on the B surface having a value of Θ closest to Θ_B. In three dimensions, the principle is the same; here one uses the following equations:

$$x = r_A \cos(\Theta_A) \sin(\Phi_A) + x_A$$

$$y = r_A \sin(\Theta_A) \sin(\Phi_A) + y_A$$

$$z = r_A \cos(\Phi_A) + z_A$$

$$r_B = \sqrt{(x - x_B)^2 + (y - y_B)^2 + (z - z_B)^2}$$

$$\Theta_B = \arctan\frac{y - y_B}{x - x_B}$$

$$\Phi_B = \arctan\frac{\sqrt{(x - x_B)^2 + (y - y_B)^2}}{z - z_B}$$

where the center of A is (x_A, y_A, z_A) and that of B is (x_B, y_B, z_B). To simply detect the existence of steric overlap, it suffices to use the equations above. However, it is frequently useful to quantitate the amount of overlap; during a database search, one can use an adjustable parameter to indicate how much overlap must be detected before a candidate is rejected. To compute the amount of overlap, let us once again consider the two-dimensional case, depicted in Figure 3. Construct points E and F as before, computed during step i of the computation, and construct points G and H at step $i+1$. The area of the quadrilaterial EFGH can be computed straightforwardly. Stepping through all points of A which penetrate B, we can effectively partition the overlap area into a set of quadrilaterals; the overlap area is the sum of the areas of each of these quadrilaterals. In three dimensions, the principle is the same. However, one caveat is that it is critically important how one computes the volume of the three-dimensional equivalent of a quadrilateral, a hexahedron (a solid with six faces and eight vertices, with no edges necessarily the same length). It is important to use a volume computation method which converges as σ gets very large. The following volume computation method works well (Figure 4): take an arbitrary hexahedron EFGHMNOP and partition it into two irregular prisms EFGHMN and GHMNOP. Take each of these irregular prisms and further partition them each into three irregular tetrahedra. The

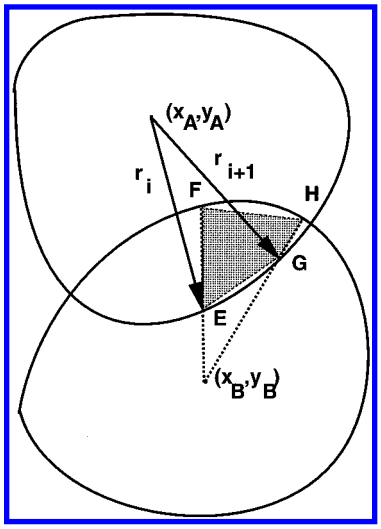

Figure 3. Computation of interpenetration area of two shrink-wrap surfaces. Points E and F were computed as in Figure 2. Points G and H were computed similarly, at the next step in the shrink-wrap array for surface A. The interpenetration area can be partitioned into a series of parallelograms like EFGH.

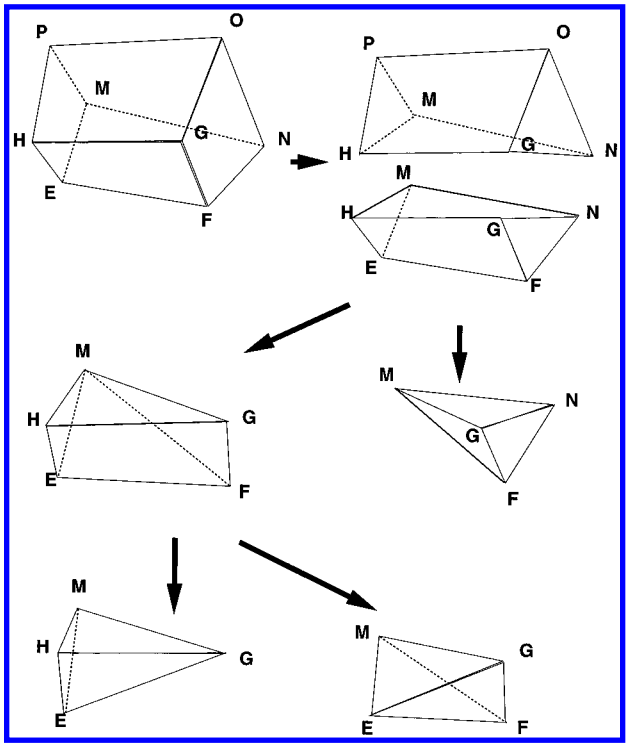

Figure 4. Decomposition of an arbitrary hexahedron EFGHMNOP into six irregular tetrahedra. The first step is to decompose EFGHMNOP into two irregular prisms EFGHMN and MNOPGH. Prism EFGHMN can be decomposed into an irregular tetrahedron MFGN and an irregular pyramid EFGHM. The irregular pyramid can be decomposed into two irregular tetrahedra MHEG and MEGF. Hence, given the formula for the volume of an irregular tetrahedron, one can compute the volume of an arbitrary hexahedron.

volume of an irregular tetrahedron EFGM is simply one-sixth the determinant of the three vectors EF, EG, and EM (i.e., this is the 3 × 3 determinant composed of the three components of the three vectors[6]):

$$\text{vol (EGFM)} = \frac{1}{6}|\overline{\text{EF}} \wedge \overline{\text{EG}} \wedge \overline{\text{EM}}|$$

Note also that this volume computation method does not require that one actually computes the length of any of the sides of the hexahedron (which requires a square root calculation).

USE OF A SHRINK-WRAP SURFACE IN A 3D DATABASE SEARCH

Each candidate molecule/conformer which meets all the usual pharmacophoric constraints must be transformed to the absolute reference frame in which the shrink-wrap surface is defined. The standard method for performing such a transformation is to find that rotation and translation which minimizes the rms (root-mean-square deviation) between points on the candidate and a set of target points.[7] The coordinates of these target points may be computed by applying the query to a reference molecule/conformer. For each candidate molecule/conformer, a shrink-wrap surface is computed based on its van der Waals surface, using the same origin for the computation as the origin of the steric exclusion shrink-wrap surface. Two possibilities are available for using the surface in a 3D database search: one can immediately reject the candidate upon detection of any steric clash. This method is fast, but somewhat inaccurate, and can show a high degree of sensitivity to details of the pharmacophore query. The slower, but more robust method, is to perform a complete computation of steric overlap volume. If this volume exceeds a threshold (e.g., 10.0 Å^3), it can be rejected. Of course, one can detect for this condition as the volume is being computed (as the overlap region is being partitioned into hexahedra), and the computation can stop once the threshold is exceeded. Employing such a threshold criterion allows the steric test to be performed in a 3D database search in less than 0.1 s/compound for a drug-like database on an 150 MHz R4400 processor.

INFERRING A SHRINK-WRAP SURFACE FROM A SET OF FLEXIBLE MOLECULES

Due to conformational flexibility, this is a more subtle problem than one might initially anticipate. Those conformers of the active molecules which match the pharmacophore query are transformed to a common coordinate system, with the center of mass of a reference molecule/conformer as the origin, and the position of the pharmacophore features on that reference molecule as the target points for applying the minimum-rms transformation. We then employ the following algorithm:

```
initialize the inferred-shrink-wrap surface by storing in it the shrink-wrap
surface computed from the van der Waals surface of the reference
molecule/conformer

 for all active molecules {
     for all conformers of this molecule which meet the pharmacophore query {
                 compute a shrink-wrap surface from the van der Waals surface
                 of this molecule/conformer
                 compute the volume difference between this shrink-wrap
                 and the inferred-shrink-wrap-surface
      }
      extend the inferred-shrink-wrap-surface by performing a union of the
        inferred-shrink-wrap-surface and that shrink-wrap surface from the
        previous loop which has  the smallest volume difference
      }
```

The union W of two shrink-wrap surfaces $\{S(\Theta_i,\Phi_j)\}_{i,j=1}\sigma$, and $\{T(\Theta_i,\Phi_j)\}_{i,j=1}\sigma$, is defined by

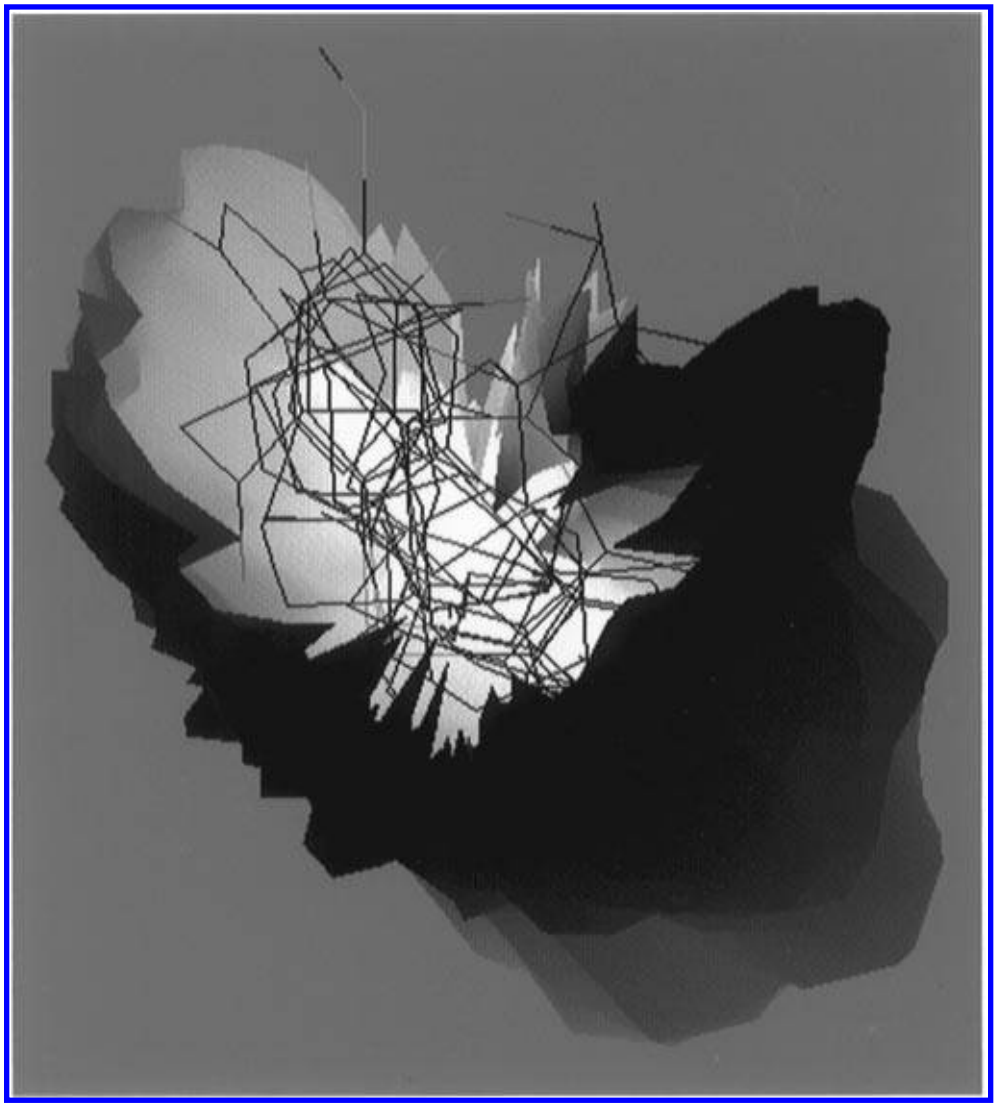

Figure 5. Cutaway view of the shrink-wrap surface inferred from a set of 12 D_2 antagonists taken from Seeman et al.[9] Conformational analysis was performed using Catalyst.[8]

$$\forall i,j\ W(\Theta_i,\Phi_j) = \max\,(S(\Theta_i,\Phi_j), T(\Theta_i,\Phi_j))$$

This algorithm ensures that for each succeeding molecule, a conformer is chosen which minimizes the amount that the volume of the inferred-shrink-wrap surface grows. This algorithm has clear deficiencies, in that its result depends on the arbitrary choice of a reference molecule, and that its result depends on the *order* of the molecules in the list of active molecules. That first deficiency does not have a significant impact, when the pharmacophore query is tightly defined (i.e., has very small tolerances, e.g., < 0.2 Å). To deal with the second deficiency, we apply a canonical ordering to the list of active molecules: they are ordered by their degree of conformational flexibility. This is easily done when one uses the conformational analysis capabilities provided by the commercial software Catalyst,[8] which prunes the list of conformers for each molecule to ensure maximum dissimilarity among the conformers; in this case, the list of active molecules is sorted simply by the number of conformers for each molecule. Figure 5 shows an example of an inferred shrink-wrap surface for a series of 10 D2 antagonists taken from Seeman *et al.*[9] Computing such a surface required 9.0 CPU-s of a 150 MHz R4400 processor. Larger datasets (e.g., hundreds of molecules) can still be done interactively, typically requiring a few minutes.

SUMMARY, CONCLUSIONS

A novel method for representing a molecular surface has been presented, and its use in both graphic display and in detecting steric clashes in a 3D database search was described. This method has the primary advantage that the steric overlap volume can be quickly and accurately computed. This method may be more suitable for the description of shape constraints in pharmacophoric 3D database searching than methods hitherto described. Furthermore, an algorithm was presented for inferring such a surface from a series of active molecules, suitable for use in a 3D database search for molecules of similar activity.

ACKNOWLEDGMENT

The idea for developing the shrink-wrap representation arose originally out of a colloboration with J. W. Erickson, in the development of the PROPACK software for solving the crystallographic translation problem by finding that translation which minimizes total interpenetration volume of a protein with all of its symmetric images.[4] The idea of a shrink-wrap representation was developed independently

and essentially simultaneously, by S. D. Kahn (then at the University of Illinois, Urbana).

REFERENCES AND NOTES

(1) Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. A Geometric Approach to Macromolecule-Ligand Interactions. *J. Mol. Biol.* **1982**, *161*, 269−288.

(2) Van Drie, J. H.; Weininger, D.; Martin, Y. C. ALADDIN: An integrated tool for computer-assisted molecular design and pharmacophore recognition from geometric, steric, and substructural searching of three-dimensional molecular structures. *J. Comp.-Aided Mol. Design* **1989**, *3*, 255−251.

(3) Sheridan, R. P.; Nilakantan, R.; Rusinko, A.; Bauman, N.; Haraki, K. S.; Venkataraghavan, R. 3DSEARCH: A system for three-dimensional substructure searching. *J. Chem. Inf. Comput. Sci.*,**1989**, *29*, 255−260.

(4) Van Drie, J. H.; Erickson, J. W., disclosed at the 1990 Mid-Atlantic Crystallography Conference but never published; and independently but not published, S. D. Kahn (personal communication).

(5) Richards, F. Areas, Volumes, Packing and Protein Structure. *Ann. Rev. Biophys. Bioeng.* **1977**, *6*, 151−176. Connolly, M. Solvent-Accessible Surfaces of Proteins and Nucleic Acids. *Science* **1983**, *221*, 709−713.

(6) See, for example: Bronshtein, I. N.; Semendyayev, K. A. *Handbook of Mathematics*; Frankfurt: Deutsch, 1985; p 207.

(7) Kabsch, W. A discussion of thee solution for the best rotation to relate two sets of vectors. *Acta Crystallogr. A* **1978**, *34*, 827−828.

(8) Smellie, A. S.; Teig, S. L.; Towbin, P. Poling: Promoting Conformational Variation. *J. Comput. Chem.* **1995**, *16*, 171−187. Catalyst is commercial software available from Molecular Simulations, Inc., San Diego, CA.

(9) Seeman, P.; Watanabe, M.; Grigoriadis, D.; Tedesco, J. L.; George, S. R.; Svensson, U.; Nilsson, J. L.; Neumeyer, J. L. Tetrahedral Model of D_2 Receptor. *Mol. Pharm*. **1985**, *28*, 391−399.

CI960464+