Flexible 3D Searching: The Directed Tweak Technique

Tad Hurst

TRIPOS Associates, Inc., 1699 South Hanley, Suite 303, St. Louis, Missouri 63144

Received July 30, 1993*

Recently 3D searching systems have begun to address the flexibility of structures. A common technique is to store or produce a small set of conformations for each structure to be examined for matches to the query. These methods do not adequately address all of conformational space. Directed tweak is a new technique which allows 3D flexible searching on an interactive time scale and finds more structures which match the 3D query than any other method described to date. Van der Waals interactions and ring flexibility can also be addressed.

INTRODUCTION

Many techniques are now available to assist researchers in the quest for new biologically active compounds. One technique which has been of great interest in the past 5 years involves searching large databases of structures for those which can present a particular pharmacophore. This technique is often called pharmacophore searching or simply 3D searching.[1-5]

In 3D searching, the query is defined as a set of distance constraints or angle constraints between molecular features. The molecular features can be atom centers or more generalized features such as centroids and least-squares lines. When atom–atom distances are the only constraints considered, the molecular system to be examined can be described as a fully connected graph in which each atom is a node attributed by atom-type and the edge between any two nodes is attributed with the distance between the two node-atoms. The search of a pharmacophore is thus a subgraph detection problem in which the desired result is one which maps the atom/nodes of the query to those of the structure (many such mappings are possible) and for which the constraints/edges are satisfied by the structure distances/edges. This process is efficiently addressed by the Ullman algorithm[6] or its variants. Modifications of this algorithm can be used to address non-atom-centered types of queries, even when the system allows "on-the-fly" definition of such features, thus making it impossible to enumerate the nodes and edges of the graph to be searched ahead of time.

The problem with such systems is that the conformational flexibility of the structures is not adequately addressed. The majority of pharmaceutical and agricultural chemicals are flexible. It is common to have 6–7 rotatable bonds per structure in databases created at large pharmaceutical and agrochemical sites.[7] If only one conformation for each structure stored in the database is available for investigation, it is common that such 3D searches find fewer than 20% of the structures which could achieve the query geometry in some reasonable conformation.

Traditional approaches to this problem involve generation of multiple conformers for each structure. A set of conformers is chosen to represent all of the conformational space available to each structure. In some systems, these multiple conformations are stored in the database with the structure. Davies *et al.*[8] reported a method of using multiple conformations to generate multiconformational condensed keys. The conformations themselves are then discarded. The conformations are regenerated using the same technique after the hits are found.

The problem with these approaches is that the number of conformations required to span conformational space is extremely large. In active analog studies using the systematic search technique, it is often found that millions of conformations are required to span conformational space.[9] In 3D searching, the same will be true unless the query has very large tolerances.

If in addition, only conformations which are energy minimized are generated, another problem is seen. Often, the bound conformation is not an energy minimal conformation. In fact, the receptor often activates the structure by placing it in a more energetic conformation.[10] Thus, using only a handful of energy-minimal conformations for 3D searching will bypass many relevant conformations.

A better approach is to investigate the flexibility of the molecule at search-time, using the query to direct the conformation changes needed to find a match to the query. We call this "query-directed conformational exploration". Many techniques have been investigated. These techniques include systematic search (from SYBYL[11]), random search, distance geometry, genetic algorithm, and directed tweak. The results of these investigations are reported by Willett and Clark.[12] This paper reports the details of the directed tweak technique, which is the method of choice from among those investigated.

Directed tweak is a torsional space minimizer, in which the rotatable bonds of the structures are adjusted at search time to produce a conformation which matches the 3D query as closely as possible. This involves the use of analytical derivatives and, because of this, is a very fast technique. This method is applicable when the queries contain distance constraints only. The flexibility of the ring systems can also be investigated with this technique.

The conformations of the structures that match the query must not be extremely high in energy. To this end, the directed tweak technique also can address the van der Waals interactions between the atoms of the database structure, thus producing energy-accessible conformers which present the query geometry.

HISTORY

The directed tweak method for 3D searching is based on the random tweak method used for protein loop searching reported by Levinthal *et al.*[13] and uses torsional space

* Abstract published in *Advance ACS Abstracts*, January 15, 1994.

0095-2338/94/1634-0190$04.50/0

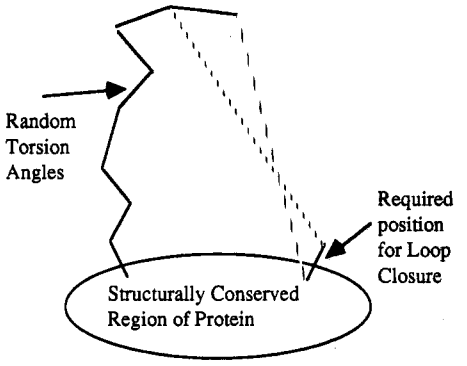


Figure 1. Random tweak technique for protein loop generation.

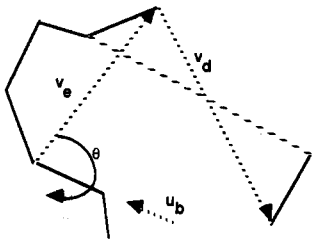


Figure 2. Vectors used in random tweak derivative calculations.

minimization as described for proteins by Levitt.[14] The random tweak technique finds suitable and varied conformations of loops which close two ends of a structurally conserved region of a protein. The structurally conserved regions are determined by sequence homology.

In the random tweak method, the loop residues are attached at one end to the structurally conserved residues, and the torsion angles of the loop residues are set to random values. The sequence to be tweaked contains one more residue than the actual loop, and the final residue must be made to overlap with the residue in the structurally conserved region of the protein to which it corresponds. Because of the initially random torsion angles, the final residue of the loop region may start quite far from the desired positions (Figure 1).

The discrepancy vectors are defined from the randomly set positions of the α-carbon and the nitrogen of the final residue and the desired (structurally conserved) positions of these two points. A minimization technique is then used to adjust (tweak) the torsion angles of the loop region in order to minimize the discrepancy vectors. This brings the loop down into the desired position. By using a number of different random starting conformations, several examples of loop closures can be produced.

The only parameters adjusted are the torsion angles of the loop residues. With van der Waals (VDW) interactions ignored, the pseudoenergy space is very well behaved, and the minimization does not seem often to get stuck in local minima.

This technique is extremely fast because the derivatives of the pseudoenergy are analytically determined, and are simple. The partial derivative of each discrepancy vector "**i**" with respect to each rotatable bond "k" in the loop region of the protein is given by

$$\delta d_i/\delta\theta_k = 2(\mathbf{u}_b \times \mathbf{v}_e) \cdot \mathbf{v}_d/d_i \quad (1)$$

where d_i is the length of the **i**th discrepancy vector, ($\|\mathbf{v}_d\|$), θ_k is the kth rotatable bond angle (Figure 2), $\mathbf{u}_b$ is a unit vector along the direction of the rotatable bond, $\mathbf{v}_e$ is the vector from the end of the rotatable bond to the atom in the final residue of the protein loop to be closed, and $\mathbf{v}_d$ is the discrepancy vector between the desired position of the atom in the final residue of the protein loop and its current position.

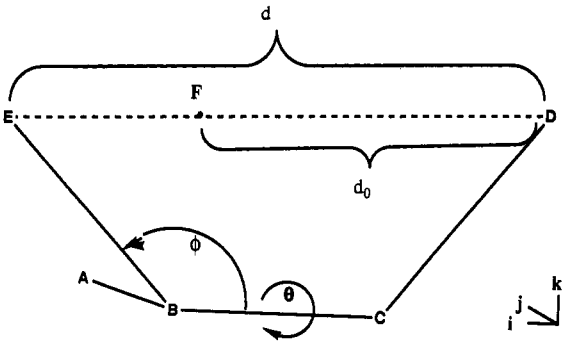


Figure 3. Vectors used in directed tweak derivative calculations.

THE DIRECTED TWEAK TECHNIQUE

The extension of the random tweak technique to 3D searching for distance constraints is straightforward. Rather than using a random starting configuration, the initial 3D coordinates for the structure which are stored in the database can be used. The function to be minimized will involve the sum of the squares of the deviations from the distances found in the structure to the distance values expressed in the constraints of the 3D query. This differs from the protein loop application in that the distances are "tweaked" to given positive values rather than to zero.

The derivatives for distance constraints are developed as follows: First consider all points in the coordinate reference frame **i**, **j**, **k** defined by the rotatable bond θ (points *A–C*, Figure 3)

$$\mathbf{i} = (\mathbf{V}_{BC}/\|\mathbf{V}_{BC}\|) \quad (2)$$

$$\mathbf{k} = (\mathbf{i} \times \mathbf{V}_{AB})/\|(\mathbf{i} \times \mathbf{V}_{AB})\| \quad (3)$$

$$\mathbf{j} = \mathbf{k} \times \mathbf{i} \quad (4)$$

Thus, the unit vector along the rotatable bond has coordinates (1,0,0) in this frame of reference. The vector $\mathbf{V}_{EB}$ = E–B then has the coordinates

$$\mathbf{V}_{EB} = (d_{EB}\cos(\phi), d_{EB}\sin(\phi)\cos(\theta), d_{EB}\sin(\phi)\sin(\theta)) \quad (5)$$

$$= (d_{EB})(\cos(\phi), \sin(\phi)\cos(\theta), \sin(\phi)\sin(\theta)) \quad (6)$$

where d_{EB} is the distance from *E* to *B*. Now, the derivative of point *E* and of vector $\mathbf{V}_{EB}$ is given by

$$\delta\mathbf{V}_{EB}/\delta\theta = \delta E/\delta\theta = (d_{EB})(0, -[\sin(\phi)]\sin(\theta), \sin(\phi)\cos(\theta)) \quad (7)$$

$$= (d_{EB})\sin(\phi)(0, -[\sin(\theta)], \cos(\theta)) \quad (8)$$

and the derivative of the distance constraint vector $\mathbf{V}_{ED}$ = *E–D* is also

$$\delta\mathbf{V}_{ED}/\delta\theta = \delta E/\delta\theta = (d_{EB})\sin(\phi)(0, -[\sin(\theta)], \cos(\theta)) \quad (9)$$

Now, the cross product of $\mathbf{V}_{EB}$ with **i** gives

$$\mathbf{i} \times \mathbf{V}_{EB} = (1, 0, 0) \times ((d_{EB})(\cos(\phi), \sin(\phi)\cos(\theta), \sin(\phi)\sin(\theta))) \quad (10)$$

$$= ((d_{EB})(0, -[\sin(\phi)]\sin(\theta), \sin(\phi)\cos(\theta))) \quad (11)$$

$$= ((d_{EB})(\sin(\phi))(0, -[\sin(\theta)], \cos(\theta))) \quad (12)$$

giving the equivalence

$$(0, -[\sin(\theta)], \cos(\theta)) = (\mathbf{i} \times \mathbf{V}_{EB})/d_{EB}/\sin(\phi) \quad (13)$$

Thus from eqs 9 and 13 the derivative for the distance constraint vector is

$$\delta\mathbf{V}_{ED}/\delta\theta = (d_{EB})/\sin(\theta)(\mathbf{i} \times \mathbf{V}_{EB})/d_{EB}/\sin(\phi) \quad (14)$$

$$= \mathbf{i} \times \mathbf{V}_{EB} \quad (15)$$

$$= \mathbf{V}_{BC} \times \mathbf{V}_{EB}/\|\mathbf{V}_{BC}\| \quad (16)$$

$$= \mathbf{u}_{BC} \times \mathbf{V}_{EB} \quad (17)$$

where $\mathbf{u}_{BC}$ is the unit vector along the rotatable bond. Now the length of vector $\mathbf{V}_{ED}$ is given by

$$d = \|\mathbf{V}_{ED}\| = (\mathbf{V}_{ED} \cdot \mathbf{V}_{ED})^{1/2} \quad (18)$$

and the derivative with respect to the rotatable bond is

$$\delta d/\delta\theta = (\delta/\delta\theta)((\mathbf{V}_{ED} \cdot \mathbf{V}_{ED})^{1/2}) \quad (19)$$

$$= {}^1/_2((\mathbf{V}_{ED} \cdot \mathbf{V}_{ED})^{-1/2}) \cdot 2(\delta/\delta\theta(\mathbf{V}_{ED}) \cdot \mathbf{V}_{ED}) \quad (20)$$

$$= (\delta/\delta\theta(\mathbf{V}_{ED}) \cdot \mathbf{V}_{ED})/\|\mathbf{V}_{ED}\| \quad (21)$$

$$= (\mathbf{u}_{BC} \times \mathbf{V}_{EB}) \cdot (\mathbf{V}_{ED}/\|\mathbf{V}_{ED}\|) \quad (22)$$

$$= (\mathbf{u}_{BC} \times \mathbf{V}_{EB}) \cdot \mathbf{u}_{ED} \quad (23)$$

where $\mathbf{u}_{ED}$ is the unit vector along the line from E to D.

Extension of the minimization technique reported by Levinthal *et al.* requires the computation of a derivative matrix $\mathbf{M}$, where $\mathbf{M}_{ik} = \delta d_i/\delta\theta_k$, calculated using eq 23. The equation to be solved is therefore

$$\mathbf{Mr} = \mathbf{d} \quad (24)$$

where $\mathbf{r}$ is the vector of rotatable bond increments and $\mathbf{d}$ is the vector of deviations from the ideal (query) distances. In the random tweak method for protein loop searching, this system of simultaneous equations is always underdetermined, since there are always at least four rotatable bonds and exactly four constraints. Thus if we consider a vector $\mathbf{v}$ such that $\mathbf{r} = \mathbf{M}^T\mathbf{v}$, it can be seen that any solution to eq 25for $\mathbf{v}$ also generates

$$\mathbf{MM}^T\mathbf{v} = \mathbf{d} \quad (25)$$

a solution $\mathbf{r}$ for eq 24. Now $\mathbf{MM}^T$ is a square matrix whose order is the number of constraints, and eq 18 can be solved directly for a solution $\mathbf{v}$. Multiplying $\mathbf{v}$ by $\mathbf{M}^T$ generates the vector of desired rotatable bond increments $\mathbf{r}$, which is an exact solution to the linear approximation. These approximations are taken iteratively until the desired solution is found.

This analysis is valid as long as the number of constraints is not larger than the number of adjustable values (rotatable bonds). When there are fewer rotatable bonds than distance constraints, the matrix product $\mathbf{MM}^T$ is guaranteed to be singular. In this case, it is desirable to find the rotatable bond vector which best approximates the desired distance constraints.

Left-multiplying both sides of eq 24 by $\mathbf{M}^T$, a system of equations is obtained in which one can solve for the rotatable bond increments to be applied (eq 26). Here $\mathbf{M}^T\mathbf{M}$ is a square

$$\mathbf{M}^T\mathbf{Mr} = \mathbf{M}^T\mathbf{d} \quad (26)$$

matrix whose order is the number of rotatable bonds. This matrix equation represents a vector-extended expression of the nonlinear least-squares technique derived more rigorously by Wentworth,[15] and successive applications of the rotatable bond increments results in a least-squares solution.

So, eqs 25 and 26 would appear to be a set which could be used in any situation, depending on the number of rotatable bonds and distance constraints. Equation 25 would be used when the number of distance constraints is not greater than the number of rotatable bonds, and eq 26 would be used when the distance constraints outnumber the rotatable bonds.

Two problems are encountered using these equations. First, eq 25 was derived with the assumption that an exact solution is available. Use of this equation is very fast when a solution is available, but does not behave well when no solution is to be found. Iterations using the results of eq 25 are often seen to neither converge nor diverge for queries which cannot be matched by the structure, but tend to wander aimlessly through conformational space.

The second problem arises from the difficulty in determining which method to use. The set of rotatable bonds of a chemical structure can be degenerate. For example, consider structure **1** (Chart 1). Since all four rotatable bonds are collinear, they all affect the distances in exactly the same manner. Since there are four rotatable bonds and only two distance constraints, eq 25 would be indicated. The $\mathbf{MM}^T$ matrix produced is singular however. It is also possible for the distance constraints commonly found in a 3D query to be degenerate.

A better approach is to use minimization techniques such as steepest-descent or BFGS,[16] which are more common in molecular mechanics treatments, to minimize a pseudoenergy term. In this case the function to be minimized is the sum of the squares of the differentials between the query distances and the actual differences

$$Q = \sum (d - d_0)^2 \quad \text{take over all constraints} \quad (27)$$

Thus the derivatives of Q with respect to the rotatable bond angles to be adjusted are

$$\delta Q/\delta\theta_i = \sum \delta/\delta\theta_i (d_k - d_{0(k)})^2 \quad (28)$$

For a single distance constraint k, this gives

$$\delta/\delta\theta_i (d - d_0)^2 = 2(d - d_0) \cdot (\delta/\delta\theta_i)(d - d_0) \quad (29)$$

$$= 2(d - d_0) \cdot (\delta/\delta\theta_i)(d) \quad (30)$$

Substitution of eq 22 gives

$$\delta/\delta\theta_i (d - d_0)^2 = 2(d - d_0)(\mathbf{u}_{BC} \times \mathbf{V}_{EB}) \cdot (\mathbf{V}_{ED}/\|\mathbf{V}_{ED}\|) \quad (31)$$

$$= 2(d - d_0)/d \cdot (\mathbf{u}_{BC} \times \mathbf{V}_{EB}) \cdot \mathbf{V}_{ED} \quad (32)$$

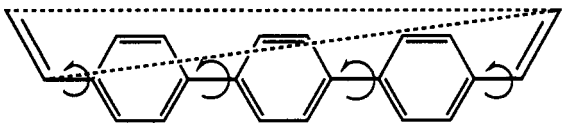
The equation has a singularity when $d = 0$, a circumstance which should not occur often. When it does occur, the derivative reduces to

$$\delta/\delta\theta_i (d - d_0)^2 = -2d_0 \|\mathbf{u}_{BC} \times \mathbf{V}_{EB}\| \quad (33)$$

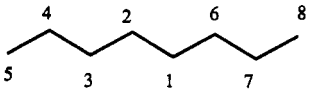
and remains well behaved.

We used a form of steepest-descent minimization. In this, the direction of steepest descent is found by summing over the constraints the derivatives for each rotatable bond. If a step in this direction does not produce an improved evaluation, the maximum allowed step size is reduced. This continues until the maximum allowed step is smaller than a cutoff value (1×10^{-8}), the derivative sum for all rotatable bonds is less than a cutoff value (0.01), or the query constraints are all met.

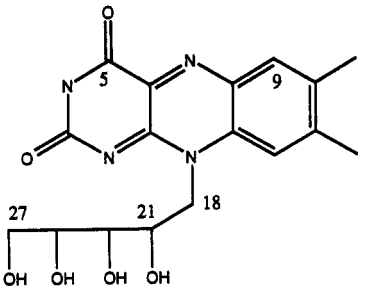
Chart 1



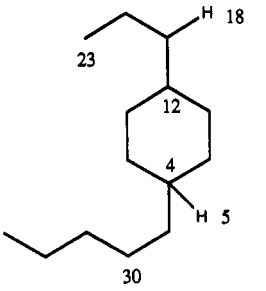
1



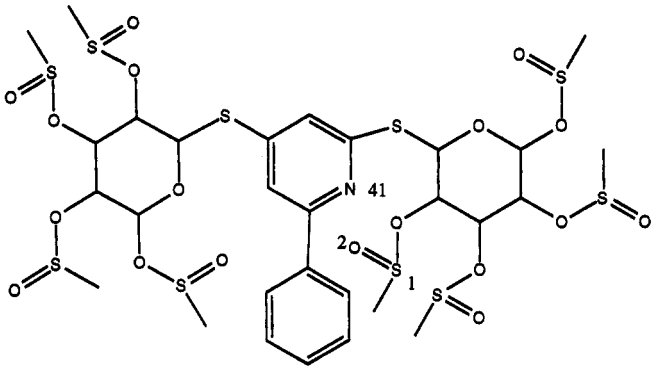
4



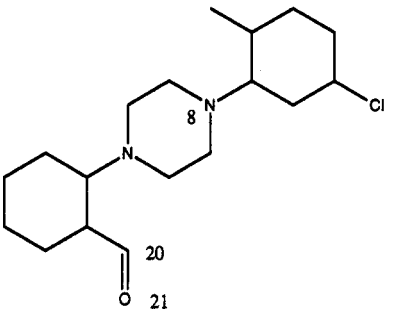
2



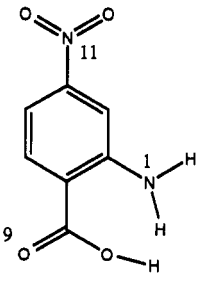
3



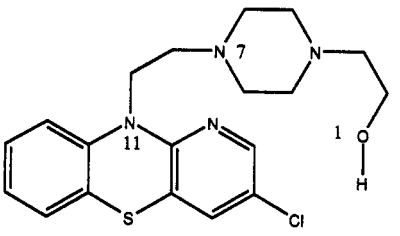
5



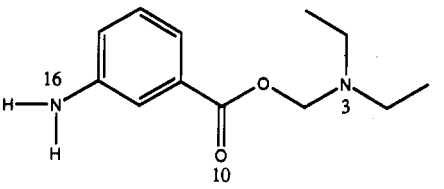
6



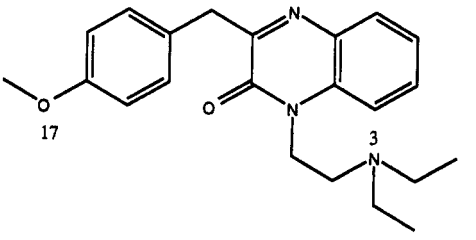
7



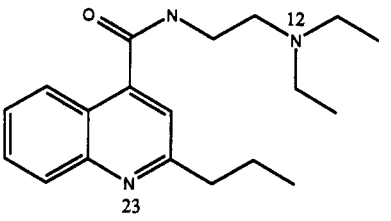
8



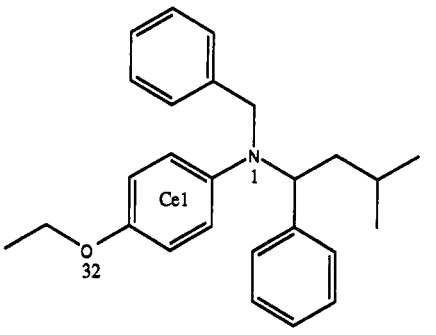
9



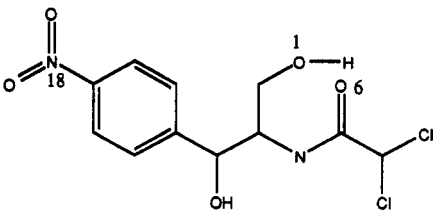
10



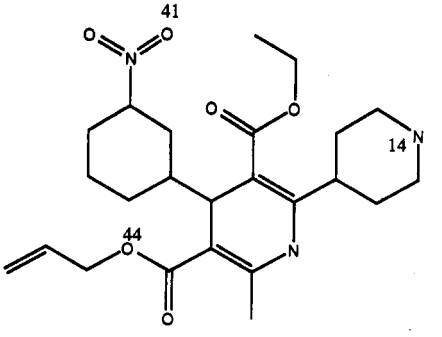
11



12



13



14

RING FLEXIBILITY AND BUMP CHECKING

The flexibility of ring systems in the database structures to be investigated can also be addressed by the directed tweak method. To do this, rotatable bonds in rings are successively deleted until no more remain. In place of each deleted ring closure, a distance constraint is added that has the original bond length as its target distance, and a tolerance of 0.2 Å. Minimization then proceeds normally. A more complete approach would use additional constraints to maintain the bond angles in the ring. This was not implemented for the work reported in this paper.

The VDW interactions are addressed by adding more constraints. A full Lennard-Jones 6–12 treatment could be used, but for our purposes the attractive nature of the VDW interaction is not important. We are interested in structures which satisfy the 3D query and do not have close contacts. Thus, we add to the pseudoenergy a term proportional to $1/r^2$ for each applicable pair of atoms. The VDW interactions between atoms which are separated by 1, 2, or 3 bonds are ignored, as are those between atoms which do not span any rotatable bond (the distance between them does not vary). There is also no contribution to the gradient when the two atoms are not in VDW contact. Thus, the contribution to the pseudoenergy is

$$Q_{vdw} = c/r^2 \quad \text{for } r < \text{vdwSum} \tag{34}$$

$$Q_{vdw} = c/\text{vdwSum}^2 \quad \text{for } r \geq \text{vdwSum} \tag{35}$$

where vdwSum is the sum of the VDW radii for the two atoms. Typically, c is chosen as 0.25. If the VDW constraints are included, it will often be true that the minimization path will encounter a local minimum. This is apparently not the case when all VDW interactions are ignored. For this reason, we first minimize the structure without the VDW interactions to see if a conformation can be obtained which matches the query geometry. Then, if bump-checking is desired, the VDW interactions are turned on and the structure is "relaxed" to a reasonable conformation that is still constrained to match the 3D query. The structures which are then found to match will be free of VDW bumps. These structures will not be extremely high in energy but can still be fairly energetic. The hits found can be further prioritized by postsearch treatment using a full molecular mechanics treatment or other techniques.

RANDOM CONFORMATION AND MULTICONFORMATION GENERATION

As a comparison, we have investigated 3D searching of databases in which are stored random conformations and multiple conformations for each structure. The random conformations were generated with a special script in SYBYL (RANDCONF). This procedure perturbs the rotatable bonds of each structure and then uses the SCAN command of SYBYL to relieve any close contact problems. RANDCONF takes a SYBYL multi-mol2 file as input and generates another SYBYL multi-mol2 file that contains one random conformation for each structure in the original hit list.

Two methods of multiple conformation generation were also written using SYBYL command scripts. The first method, RANDCONFMULT, generates unique energy-minimized conformations for each structure. This is accomplished by first perturbing the rotatable bonds to random values, followed by a full molecular mechanics minimization. Any conformation which is within 0.2 rms (rms = root mean square) of a previously encountered conformation of the same structure is discarded, as is any conformation which is more the 70 kcal/mol above the lowest energy conformation found so far.

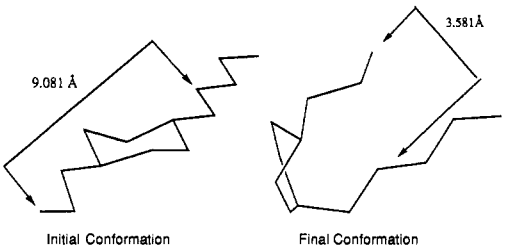


Figure 4. Initial and final conformations for test 8, Table 1.

The second method, RANDCONFMULTSCAN,[17] is similar to the first in that the rotatable bonds are perturbed randomly. A full minimization is not performed, however. The bad bumps are relieved using the SYBYL SCAN function, which is much faster. No conformations are discarded.

Searches of databases where performed using the UNITY[18] program from TRIPOS Associates, Inc. UNITY provides the ability to perform static 3D searches over one or more conformations of structures stored in a database. The directed tweak method of conformational exploration is also incorporated into UNITY.

RESULTS

Tests of the directed tweak method were performed using several structures and distance constraints. With the VDW turned off, and ring bonds not adjusted, the minimization times were quite fast (Table 1). With ring flexibility considered, the times vary, usually requiring more time. Occasionally, ring-flexible searching is faster, probably because the additional degrees of freedom allow a solution to be found (Table 1, test 5).

Ring searching allows conformations of rings to vary significantly. The starting conformation of structure 3 presents the cyclohexane ring in a chair conformation (Figure 4). The results of test 8, in which a short distance constraint is used, transforms the ring into a twisted boat conformation. When queries (Chart 2) are applied to large databases, ring-flexible searching often finds 10–20% more structures than searches in which rings are not perturbed (Table 2). In all cases tested, all structures which matched the ring-rigid search are found in the ring-perturbed search, as expected.

VDW searching is often much slower than searches in which there is no bump checking. When queries are applied to large databases, VDW checking often eliminates 10–15% of the structures found when no bump-checking is done. In most of these cases, the eliminated structures cannot achieve the query geometry without having close contacts and thus should be eliminated from the hit list.

The flexible searching was tested for completeness by using the RANDCONF procedure described above. In these tests, a query is applied to the CAS-30k database, using the static (single conformation) 3D searching of UNITY. The results are then written out in SYBYL mol2 format, and the RANDCONF procedure is used in SYBYL to randomize the conformations. The structures with their randomized conformations are then loaded into a new small database. Each structure in this database is known to have at least one conformation which matches the original query, but the stored conformation may not match the query. When static 3D searching is used on this randomized database with the original query, a good portion of the database structures are not found as expected (Table 3).

To compare these results with multiconformation searching, the procedures RANDCONFMULT and RANDCONFMULTSCAN were used. These procedures were applied to

Table 1. Minimization Times for Directed Tweak[a]

no.	structure	constraints atom–atom, dist ± tol	no Ring no Bump	Rings no Bump	bumps no Ring	rings bumps
1	**2**	5–18, 5.5 ± 0.1	0.0006	0.0181	0.0006	0.0178
			not found	not found	not found	not found
2	**2**	5–27, 5.5 ± 0.1	0.0021	0.0163	4.290	1.133
			found	found	found	found
3	**2**	5–21, 5.0 ± 0.1	0.0009	0.0143	0.1268	3.694
			found	found	found	found
4	**2**	5–21, 4.0 ± 0.1	0.0054	0.1762	0.0054	0.1753
			not found	not found	not found	not found
5	**2**	5–21, 5.5 ± 0.1	0.0076	0.0030	0.0074	1.217
		9–21, 6.0 ± 0.1	not found	found	not found	found
6	**2**	5–21, 5.5 ± 0.3	0.0056	0.0044	0.0056	0.1972
		9–21, 5.0 ± 0.3	not found	found	not found	found
7	**3**	5–18, 5.5 ± 0.1	0.0032	0.0022	0.0030	0.0936
			not found	found	not found	found
8	**3**	5–18, 3.5 ± 0.1	0.0049	0.0021	0.0049	0.0908
			not found	found	not found	found
9	**3**	23–30, 3.5 ● 0.1	0.0212	0.0612	0.0202	0.7899
			not found	found	not found	found
10	**3**	23–30, 7.2 ± 0.1	0.0013	0.0026	0.0842	0.1582
			found	found	found	found
11	**3**	4–12, 5.5 ± 0.1	0.0006	0.0115	0.0006	0.0114
			not found	not found	not found	not found
12	**4**	5–8, 6.0 ± 0.01	0.0071	0.0070	0.0110	0.0110
		1–8, 3.0 ± 0.01	found	found	found	found
13	**4**	5–8, 6.0 ± 0.01	0.0304	0.0302	0.0401	0.0404
		1–8, 3.75 ± 0.01	found	found	found	found
		3–7, 4.185 ± 0.05				
14	**4**	2–6, 6.514 ± 0.05	0.0002	0.0002	0.0002	0.0002
		1–8, 3.75 ± 0.01	not found	not found	not found	not found
15	**5**	2–41, 4.8 ± 0.3	0.0055	0.0087	1.0901	1.2908
		1–41, 4.0 ± 0.3	found	found	found	found
16	**6**	8–21, 4.8 ± 0.3	0.0007	0.0075	0.0329	0.3936
		8–20, 4.0 ± 0.3	found	found	found	found
17	**7**	1–9, 3.996 ± 0.146	0.0005	0.0004	0.0004	0.0005
		1–11, 4.875 ± 0.010	not found	not found	not found	not found
		9–11, 6.295 ± 0.395				
18	**8**	1–7, 6.614 ± 0.714	0.0180	0.0104	8.4045	3.7274
		1–11, 4.930 ± 1.080	found	found	not found	found
		7–11, 4.444 ± 0.474				
19	**9**	3–10, 5.906 ± 0.006	0.0065	0.0076	3.0763	3.0265
		3–16, 5.040 ● 1.070	found	found	found	found
		10–16, 4.920 ± 1.070				
20	**10**	3–17, 7.430 ± 1.530	0.0021	0.0079	0.2843	0.4463
		3–21, 5.04 ± 1.070	found	found	found	found
		17–21, 4.930 ± 1.080				
21	**11**	8–12, 5.938 ± 0.037	0.0014	0.0015	0.0851	0.0833
		8–23, 4.481 ± 0.631	found	found	found	found
		12–23, 5.340 ± 1.070				
22	**12**	1–32, 7.0 ± 0.50,	0.0015	0.0018	0.1111	0.1121
		1-Cent1, 5.15 ± 0.5	found	found	found	found
23	**13**	1–6, 3.0 ± 0.1	0.9850	1.0908	1.0274	1.0874
		1–18, 8.0 ± 0.1	not found	not found	not found	not found
		6–18, 6.0 ± 0.1				
24	**14**	14–41, 7.08 ± 0.56	0.0344	0.0089	0.0346	4.2724
		14–44, 8.62 ± 0.58	not found	found	not found	found
		41–44, 3.35 ± 0.65				

[a] Distances and tolerances are in angstroms. Times are cpu seconds per minimization, averaged over 100 minimizations from random starting conformations. All results (found or not found) are as expected for these structures and queries. Tests were conducted on an SGI Personal Iris 4D/35.

the same mol2 files described above for RANDCONF testing. When RANDCONFMULT was applied to the CAS-30k results from query 1, the average number of conformations/structure produced was 15.1. When a small UNITY database of these conformations was searched (static mode) for the original query (query 1), 109 of 117 structures were found.

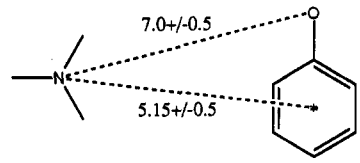
When RANDCONFMULTSCAN was applied to the same set of 117 structures, 20 conformations/structure were produced. When a small database of these conformations was loaded and searched (static mode), 115 of 117 were found. RANDCONFMULTSCAN produced a better representation of conformational space than did RANDCONFMULT and was much faster. RANDCONFMULTSCAN was therefore used in further testing.

Six queries were used to test multiconformational searching and directed tweak searching, as implemented in the UNITY system. In some cases (Table 3), multiconformational searching produces most of the expected hits. When directed tweak is applied, nearly all structures are found. In these tests, only one structure was not found (Table 3, query 4). This structure was investigated and found to be in a local minimum. When several random starting conformations were used, this structure was found 42% of the time.

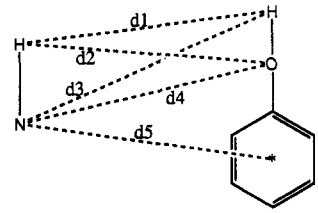
CONCLUSIONS

Directed tweak is an excellent method of exploring conformational space when looking for a match to a 3D query. It is fast and finds nearly all qualified hits.

Chart 2

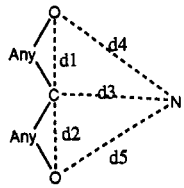


Query 1



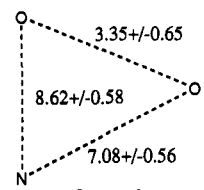
d1: 6.88+/-0.5
d2: 7.21+/-0.5
d3: 6.57+/-0.5
d4: 6.87+/-0.5
d5: 5.05+/-0.5

Query 2

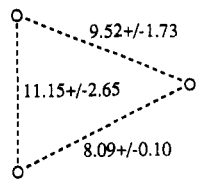


d1: 2.3+/-0.1
d2: 2.5+/-0.1
d3: 5.3+/-0.1
d4: 3.2+/-0.1
d5: 5.0+/-0.1

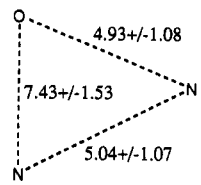
Query 3



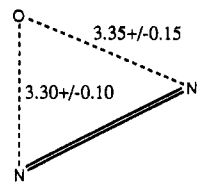
Query 4



Query 5



Query 6



Query 7

Table 2. Results of Searching in Large Databases[a]

query	static search	flexible no ring, noVDW	flexible ring, noVDW	flexible no ring, VDW	flexible ring, VDW
1	117	511	602	450	534
2	2	116	135	103	114
3	0	61	112	28	41

[a] Searches were performed against the CAS-30k database, provided by Chemical Abstracts Service. 3D structures in this database were generated with CONCORD.[19]

FUTURE DEVELOPMENTS

Development is currently under way which will extend the directed tweak method to include angle constraints and excluded volumes. More complete treatment of ring closures is also underway. In addition, this effort will allow the use of any arbitrary evaluation function, such as CoMFA QSAR prediction or receptor binding affinity.

ACKNOWLEDGMENT

The author thanks Prof. Peter Willett and Dr. Dave Clark (University of Sheffield) for contributions in testing and analysis, Prof. Robert Pearlman (University of Texas) and Yvonne Martin (Abbott Laboratories) for many valuable discussions, and Chemical Abstracts Service for supplying the test databases used in this work. Also, the author thanks the UNITY consortium companies for their input, testing, support, and funding of this work.

Table 3. Comparison of Directed Tweak with Single-Conformational and Multiconformational 3D Searching[a]

query	CAS30k static	random	mult20	directed tweak
2	117	68	115	117
3	2	0	2	2
4	186	81	178	185
5	74	24	52	74
6	50[b]	39	50	50
7	43	29	42	43

[a] The random column reflects the number of hits in a single-random conformational database produced with the RANDCONF procedure. Mult20 refers to the number of hits in a multiconformational database (20 conformers/structure) produced with RANDCONFMULTSCAN. The results shown in the directed tweak column were produced using the same single, random conformation database used in the random column, but by applying the directed tweak method to explore conformational space for a match to the query. [b] The original search was truncated when 50 hits were found.

REFERENCES AND NOTES

(1) Martin, Y. C.; Danaher, E. B.; May, C. S.; Weininger, D. MENTHOR, A Database System for the Storage and Retrieval of Three-Dimensional Molecular Structures and Associated Data Searchable by Substructural, Biological, Physical, or Geometric Properties. *J. Comput.-Aided Mol. Des.* **1988**, *2*, 15–29.

(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. Comput.-Aided Mol. Des.* **1989**, *3*, 225–251.

(3) Sheridan, R. P., Nilakantan, R.; Rusinko, A., III; 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) Jakes, S. E.; Watts, N.; Willett, P.; Bawden, D.; Fischer, J. D. Pharmacophore Pattern Matching in Files of 3-D Chemical Structures: Evaluation of Search Performance. *J. Mol. Graphics* **1987**, *5*, 41–48.

(5) Moock, T. Conformational Searching in ISIS 3D Databases. Presented at the Third International Conference on Chemical Structures, Leewenhorst Congress Center, Noordwijkerhout, The Netherlands, June 6–10, 1993.

(6) Ullmann, J. R. An Algorithm for Subgraph Isomorphism. *J. Assoc. Comput. Mach.* **1976**, *16*, 31–42.

(7) Unpublished statistics from UNITY consortium company databases.

(8) Murrall, W.; Davies, E. K. Conformational Freedom in 3-D Databases. *Chemical Structures II*; Springer-Verlag: Berlin, 1993.

(9) Mayer, D.; Naylor, C. B.; Motoc, G. R., Marshall, G. R. A Unique Geometry of the Active Site of Angiotensin-Converting Enzyme Consistent with Structure-Activity Studies. *J. Comput.-Aided Mol. Des.* **1987**, *1*, 3–16.

(10) Jorgenson, W. L. Rusting of the Lock and Key Model for Protein-Ligand Binding. *Science* **1991**, *254*, 954.

(11) SYBYL is a trademark of TRIPOS Associates, Inc., St. Louis, MO.

(12) Clark, D. E.; Jones, G.; Willett, P.; Kinney, P. W.; Glen, R. C. Pharmacophore Pattern Matching in Files of Three-Dimensional Chemical Structures: Comparison Of Conformational-Searching Algorithms for Flexible Searching. *J. Chem. Inf. Comput. Sci.*, following paper in this issue.

(13) Shenkin, P. F.; Yarmush, D. L.; Fine, R. M.; Wang, H.; Levinthal, C. Predicting Antibody Hypervariable Loop Conformation. I. Ensembles of Random Conformations for Ringlike Structures. *Biopolymer* **1987**, *26*, 2053–2085.

(14) Levitt, M. Protein Folding by Minimization. *J. Mol. Biol.* **1983**, *170*, 723–764.

(15) Wentworth, W. E. Rigorous Least Squares Adjustment. *J. Chem. Educ.* **1965**, *2*, 96–103.

(16) Fletcher, R. *Practical Methods of Optimization*, 2nd ed.; Wiley: New York, 1987.

(17) The RANDCONF, RANDCONFMULT, and RANDCONFMULTSCAN SYBYL command procedures are available on request.

(18) Unity is a trademark of TRIPOS Associates, Inc., St. Louis, MO.

(19) Rusenko, A., III; Skell, J. M.; Balducci, R.; McGarty, C. M.; Pearlman, R. S. CONCORD, A Program for the Rapid Generation of High Quality Approximate 3-Dimensional Molecular Structures; TRIPOS Associates: St. Louis, MO, 1988.