Analysis of Conformational Coverage. 1. Validation and Estimation of Coverage

Andrew Smellie,* Scott D. Kahn, and Steven L. Teig

Molecular Simulations Inc., 555 Oakmead Parkway, Sunnyvale, California 94086

Received October 12, 1994®

Given a topology of a molecule, there are many ways to estimate the size and scope of the conformational space.[1−4] This paper studies how well a collection of conformations covers conformational space. A modified version of a systematic search was developed to exhaustively search the conformational spaces of small molecules, and a heuristic was applied to extract a small set of conformations that covers the space of the exhaustive set. It demonstrates in principle that a small number of conformations is sufficient to represent the low-energy conformational spaces of small- to medium-sized molecules. These representative conformations are used to show that a common empirical measure of flexibility, the so-called 3^N estimator, is ineffective at estimating the number of conformations required to cover low-energy conformational spaces.

INTRODUCTION

Advances in the understanding of conformational preferences of molecules have played an important role in the development of chemistry.[5] While initially concerned with the understanding of mostly rigid molecular assemblies, chemists have been increasingly motivated to explore the conformational preferences in larger and more flexible systems. It has become common practice to seek an understanding of time-averaged structures through the enumeration of all local minima and the location of a global minimum.[6−9]

Ongoing advances in conformational analysis of organic molecules have enabled the more efficient exploration of conformational space, but the goal of conformational analysis (*i.e.*, minima enumeration) has remained unchanged. Time-averaged structures have been validated with the wealth of data on isolated organic molecules in the gas phase and in crystals, but the lack of good structural data in less homogeneous media continues to limit the understanding of conformational preferences in condensed phase environments (*e.g.*, in solution or at an interface).

Unlike conformational preferences in the gas phase, the conformations that are possible in the condensed phase reflect a balance of inter- and intramolecular forces.[10] Intermolecular interactions can easily perturb the preferred (gas phase) geometry of a conformer, and thus it becomes much less relevant to enumerate minima, especially when evaluated *in vacuo*, as these minima will depend largely on the local environment of any particular compound. The induced fit of a ligand into a protein receptor,[11] and allosteric effects[12] within protein systems are two well-studied examples of this assertion.

We would like to introduce a new term to describe the accessible conformations (*i.e.*, rather than local minima) a molecule can adopt within an energy threshold. Such a "conformational model" is not necessarily limited to a simple collection of conformations but can be thought of as an abstract description of the accessible conformational space. One version (used in this paper) of a conformational model is a simple collection of conformations. Alternatively, the bounded distance matrix from distance geometry[7] is a conformational model.

Conformational models of small organic molecules generally consist of a collection of conformations consisting of single conformers,[13] multiple conformers,[14] or poled conformers.[15] The resolution of a conformational model can be defined in terms of the expected RMS distance of an arbitrary low-energy conformer from a conformation in the model. The resolution necessary for a given application is largely dependent upon the nature of the application.[16,17] It is crucial that the resolution of the conformational model, in terms of how well the model covers the low-energy regions of accessible conformational space, be consistent with the precision demanded by the application.

This paper investigates the relationship between the size of a conformational model (*i.e.*, the number of conformers) and its resolution. An analogy for this analysis is the dot density of a halftoned image and the perceived spatial and intensity resolution of the image.[18] In the case of conformational models, comparisons have been made between a quasi-exhaustive set of conformers (*i.e.*, a complete image) and smaller sets of conformers (*i.e.*, a halftoned image) that were chosen to maximize the coverage of the larger set.

The paper addresses whether, in principle, a small collection of conformers can represent a quasi-exhaustive set of conformers in any meaningful manner for a collection of small- to medium-sized drug molecules. That a small collection of conformations is sufficient to cover a quasi-exhaustive set of energy minimized conformations is demonstrated using a selection heuristic that extracts a small set of conformations subject to the requirement that the small set adequately covers the conformational space of the larger set. An empirical relationship between the number of conformations extracted and the resolution of the conformational coverage is developed. A novel metric, the hole-size metric, is used to measure relative conformational coverage.

® Abstract published in *Advance ACS Abstracts,* March 1, 1995.

0095-2338/95/1635-0285$09.00/0

Table 1. Torsion Grid Points Used in the Study

torsion symmetry	grid point values explored
sp^3–sp^3	0, 60, 120, 180, 240, 300
sp^2–sp^2	0, 90, 180, 270
sp^3–sp^2	0, 60, 120, 180, 240, 300

GENERATING CONFORMATIONS TO COVER CONFORMATIONAL SPACE

The experiment consists of generating a quasi-exhaustive conformational model for each test molecule and applying a heuristic to select a subset of conformers from this conformational model such that the subset covers the space spanned by the exhaustive set, subject to a user-specified tolerance on the coverage. The quasi-exhaustive search is carried out in torsion space by representing the molecule as a rooted tree of rotatable bonds. The term "quasi-exhaustive" is used to reflect the fact that no systematic search is truly exhaustive, because the search is confined to a predetermined set of search values for the torsion angles. There is a directed edge in the tree for every rotatable bond, and there is a node in the tree that contains every atom whose position is completely defined by its parent edge. For this study, rings are assumed to be rigid and remain fixed throughout the search. This is the familiar torsion tree used previously in systematic search methods.[19,20]

The total number of states available to the molecule in the torsion tree representation is traditionally given by 3^N, but even this daunting number was thought to constitute a crude search for this study. In fact, the number of theoretical states is denoted by M^N, where M represents the number of grid points (i.e., torsion values) explored for a given torsion and N is the number of rotatable bonds. M varies from 6 (for sp_3 vs sp_3 torsions) to 4 (for sp_2 vs sp_2 torsions). Statistics are gathered at each stage to show how the number of conformers used to represent the conformational space differs from the theoretical M^N maximum. Each stage of the experiment is described more fully below, but first some terms must be defined:

M = total number of positions explored for each rotatable bond

N = total number of rotatable bonds

C_T = total number of conformers obtained from the M^N estimator

V_{vdw} = VDW collision radius

C_{vdw} = total number of conformers after rejection of any conformer with an interatomic distance $< V_{vdw}$

E_{energy} = maximum permitted energy above the estimated global minimum

C_{energy} = total number of conformers after rejection of conformers with energies more than E_{energy} above the estimated global minimum

The quasi-exhaustive model is obtained by an implementation of systematic generation of all conformations by exhaustive exploration about rotatable bonds. Enhancements were made to a traditional systematic search to address some of the fundamental problems of a grid-based search. These will be described later. An exhaustive search is performed in torsion space, using the torsion values shown in Table 1. As is common in most systematic search techniques, scaling factors were applied to the VDW terms to partially compensate for the coarse granularity of the torsion search grid, though the scaling is quite conservative because of a novel method of searching. The VDW radius of atom pairs was scaled empirically by 0.7 (for 1−4 interactions) and 0.9 (for all other interactions). At this stage, conformations are rejected if there is any VDW violation $< V_{vdw}$.

To keep the run times tractable, the systematic search was performed using a relatively coarse grid, but it was found that this often led to very few conformers generated. This was due to a VDW clash that was detected early in the exploration of the coarse grid that resulted in the search tree being pruned at an early stage, even if a small displacement might have resolved the VDW clash. One unsatisfactory solution to this problem is to scale down the VDW radii of the atoms to reduce the changes of a clash. Large reductions in VDW radii[21] can lead to many conformers being accepted (i.e., not pruned) with VDW clashes that cannot ultimately be resolved by energy minimization.

The robustness of the quasi-exhaustive search was improved by the selective use of a one-dimensional minimization during combinatorial exploration of rotatable bond values. The minimization is "one-dimensional" in the sense that only one variable is being changed during the minimization. Any implementation of a systematic search is confounded by a fundamental dilemma: it is desirable to search a finer torsion grid to improve the conformational coverage obtained at the expense of exponentially increasing run times, whilst searching a coarser grid decreases run times at the expense of poorer coverage. Coarse grids have the further limitation that it is possible that a small change in the current torsion value being explored (i.e., a small deviation from the grid) might result in a VDW clash-free (partial) conformer being found. It is primarily for this reason that VDW radii are scaled down to reduce this problem.

An alternative solution considers the grid point (i.e., a specific value of a torsion angle during the combinatorial search) as a permitted range rather than a single value. This "fuzzy grid" is illustrated in Figure 1. Here, a torsion T, defined by atoms a-b-c-d, is being explored at some arbitrary place in the systematic search. Atoms X have already been placed (by exploration of torsions earlier in the search), and atoms Y are contained in the local rigid group (defined as the group of atoms whose positions are wholly defined by T). The current value of torsion a-b-c-d can be thought of as belonging to a segment delimited by S. The arc of the segment depends upon the granularity of the search. In Figure 1, a torsion value of 180° in a segment of arc ± 60° is shown.

When, for example, $T = 180°$ (i.e., a particular point of the grid) is explored and a VDW clash of any atom in set Y is detected with any atom of set X, the conventional solution is to abandon this grid point and move to the next. This is acceptable for a fine grid but might miss parts of accessible conformational space for a coarse grid. In this experiment a 1D minimization is performed with T as the variable, minimizing the penalty function given in eq 1 below.

$$F = \sum_{i}^{N_X} \sum_{j}^{N_Y} \frac{1}{d_{ij}^2} \tag{1}$$

where

N_X = number of atoms in the partial conformer (X's in Figure 1)

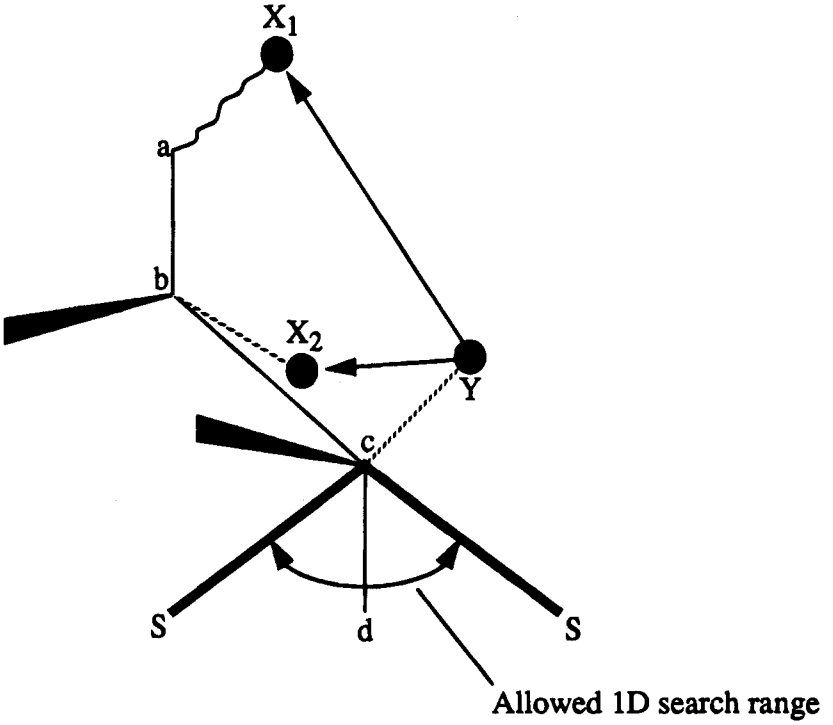


Figure 1. The range and penalty term of a 1D minimization.

N_Y = atoms in the local rigid group for the current dihedral (Y's in Figure 1)

d_{ij} = distance between atom i (of the X set) and atom j (of the Y set)

The penalty function in eq 1 is a simple function that attempts to force the atoms in sets X and Y away from each other. The minimization terminates when (a) all VDW clashes have been resolved, or (b) torsion T leaves its appropriate segment, or (c) a local minimum is found. Termination condition (a) means that the new torsion value is accepted and the search continues. Termination conditions (b) or (c) cause the trial grid point to be rejected. The use of a 1D minimization eliminates the grid effect. In principle, all torsion states are accessible during the search—even those states not directly on the grid.

The effectiveness of the 1D minimization is summarized in Table 2, where it can be seen that the 1D minimization typically resolves VDW clashes about 30% of the time. These results are studied in greater detail later in the paper. The total number of conformations obtained at this stage is given by C_{vdw}. Duplicates are not rejected at this stage, but some effort is made to ensure that they do not arise by a novel implementation of energy minimization, described below.

Refinement of the quasi-exhaustive model involves a *segmented energy minimization* in an attempt to minimize the conformational energy. A standard conjugate gradient minimizer[22] was used with a version of the CHARMm force field,[23] and the energy was minimized until the RMS gradient was less than 0.1 kJ/Å or the energy changed by less than 0.01 kJ over five iterations. An estimate of the globally minimum energy is accumulated on the fly, and a conformer is rejected if it is more than E_{energy} kcal above the estimated global minimum found. The total number of conformations obtained at this stage is given by C_{energy}.

Table 2. Number of Grid Points Resolved by 1D Minimization

mol. no.	tot. no. of grid pts explored (A)	tot. no. of failed grid pts (B)	ratio B:A	tot. no. of grid points resolved (C) by 1D minimztn	ratio C:B
1	826	446	0.54	161	0.36
2	49596	29213	0.59	8528	0.29
3	737836	442897	0.60	138914	0.31
4	11386	7572	0.67	2269	0.30
5	576544	289015	0.50	109203	0.38
6	1988488	1623630	0.82	269243	0.09
7	473436	296377	0.63	22388	0.08
8	53754	7289	0.14	2981	0.41
9	138200	9968	0.07	6814	0.68
10	1924	897	0.47	121	0.13
11	1014698	579375	0.57	49434	0.09
12	27828	22723	0.82	4144	0.18

The minimization is segmented in the sense that torsion constraints are added to every rotatable bond to prevent the torsion from leaving its current quadrant when viewed as a Newman projection. This is illustrated in Figure 2. The purpose of the segmented minimization is to prevent conformers from coalescing during minimization by constraining each torsion to the local well that is explored during the systematic search. Remember, we are interested in generating a diverse set of conformations that span the conformational space within some energy threshold of the estimated global minimum. By constraining the minimization we permit the conformers to span the space and not fall

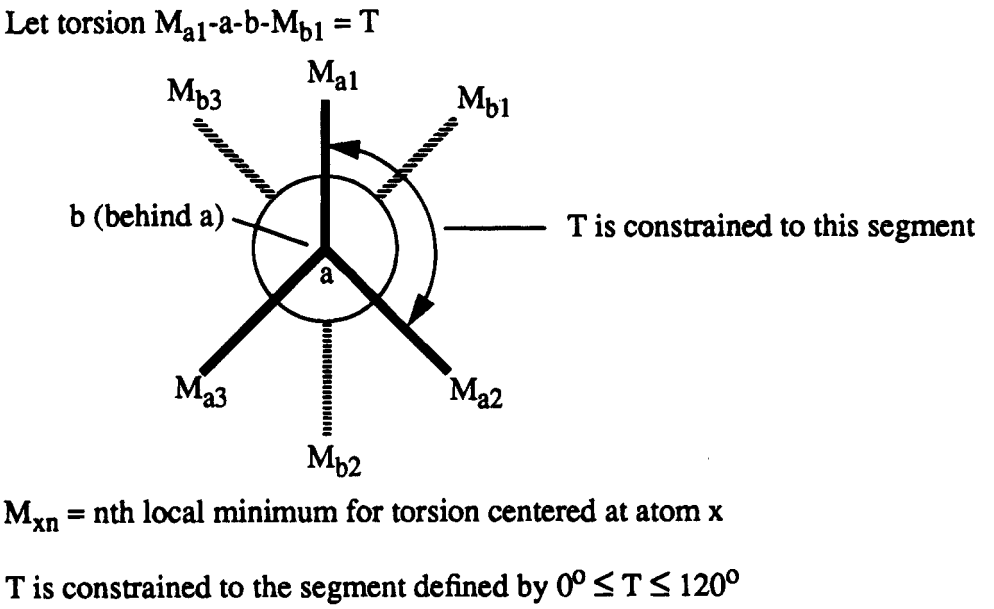


Figure 2. Torsion constraints imposed during energy minization.

into local minimum wells upon minimization. These local minimum wells will still be found by the search because there is at least one grid point in the search that encompasses the wells. The segmented minimization uses a novel torsion constraint based upon the hyperbolic tangent and is described in Appendix 1. The constraint function has the property that it is smooth and continuously differentiable, is virtually zero when the constraint is satisfied, and rises very sharply when the constraint is violated.

Experimental results are shown for the collection of molecules shown in Figure 3. The molecules selected were all flexible molecules typical of drug-sized molecules, with four or more rotatable bonds per molecule. We first consider the efficacy of 1D minimization during the systematic search.

Efficacy of 1D Minimization during Systematic Search. Table 2 demonstrates the efficacy of the 1D minimization introduced into the systematic search to locate VDW clash-free conformers. The second column shows the number of grid points explored, denoted by A. This is *not* the number of complete conformations explored but is rather the total number of times a single torsion was incremented (and checked for VDW clashes) during the systematic expansion of all the torsions.

The second column of Table 2 shows how many of these grid points resulted in a VDW clash between atoms, denoted by B. In conventional systematic searching, the detection of such a clash would result in backtracking up the search tree, and this grid point would be ignored. The fourth column shows the ratio of B:A and shows what proportion of grid points fail the test for VDW clashes.

The final columns of Table 2 show how many grid points with VDW clashes were resolved by a 1D minimization, denoted by *C*. "Resolved" in this case means that after 1D minimization, there were no VDW clashes remaining. The ratio of grid points resolved to grid points that displayed a clash are shown in column 6. It can be seen that the efficacy of 1D minimization varies widely from molecule to molecule. In the most effective case, minimization resolved 68% of "bad" grid points (molecule 9) and in the least effective case resolved 8% of grid points (molecule 7).

Why might 1D minimization fail to help to resolve clashes? Certainly the minimization performed here was necessarily (for speed concerns) crude, involving at most five steps of steepest descent minimization[27] and did not permit the torsion to leave its local well. The use of a more robust technique such as conjugate gradients[22] or quasi-Newton minimization[28] would probably improve the ratio of VDW clashes resolved. Also the fact that this is a *1D* minimization means that it is possible to get into situations that cannot be resolved by varying just one torsion at a time. There might be *no* position of the torsion being minimized that would resolve all the VDW clashes. Often it is necessary to vary more than one torsion to resolve such clashes. Despite these caveates, the 1D minimization is typically seen to resolve about one-third of the "bad" torsion grid points.

It is not the scope of this paper to compare the efficacy of systematic search with 1D minimization to other forms of conformational analysis. It is clear that the search method used here is at *least* as effective as systematic search on a regular grid of the resolution shown in Table 1. It is also clear that the addition of 1D minimization results in more conformations found, because, on average, 30% of grid points explored are resolved. However, the thrust of this paper is to demonstrate how compact these quasi-exhaustive models are, as demonstrated in the next section.

COVERING CONFORMATIONAL SPACE WITH CONFORMATIONS

At this stage, conformational coverage, and a technique for assessing such coverage, must be defined and justified. These topics have been previously addressed,[15,24] and a summary is given in Appendix 2. Put briefly, conformational coverage can be measured using a two-set measure that uses two conformational models of the same molecule named A and B. A is a quasi-exhaustive model, and a quantity is defined that is the largest "hole" conformer model A finds in model B. The hole size of B is calculated by looking at pairwise interconformer distances, where distance is any metric that quantifies the difference between two conformers, and finding MAX($H(A_i, B)$)—where $H(A_i, B)$ is the minimum of the distance from conformer i of A to each conformer j

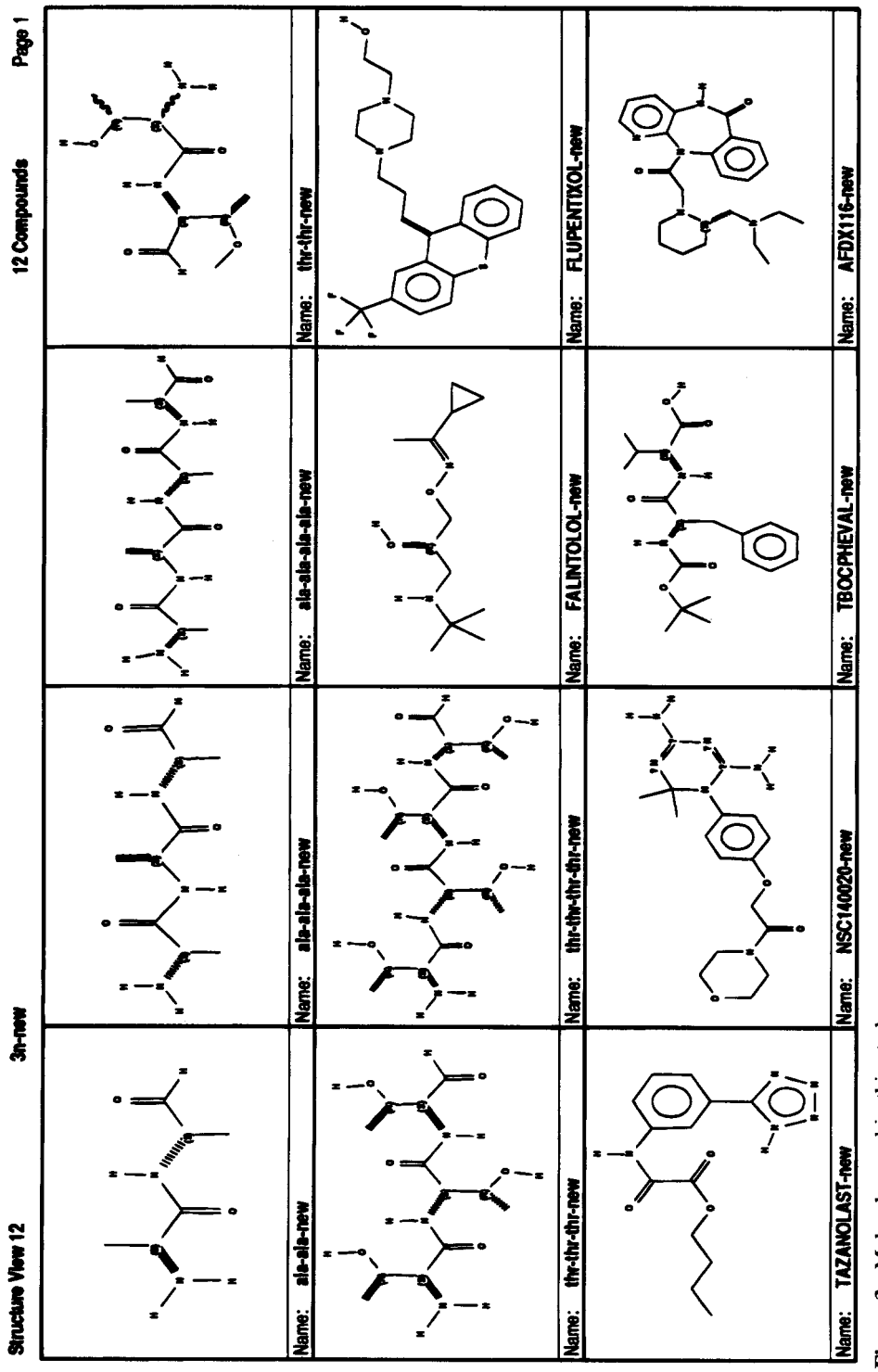


Figure 3. Molecules used in this study.

Table 3. Conformational Models Generated and Extracted

mol. no.	no. rot. bonds	M^N	C_{vdw}[a]	C_{energy}[b]	C_{subset}[c]	C_{subset}[d]	C_{subset}[e]
1	4	864	393	138	47	7	3
2	7	124416	19916	2205	421	46	8
3	10	17915904	300000	12331	1725	155	19
4	6	31104	4147	1335	310	50	10
5	10	26873856	300000	37956	4932	306	30
6	14	2.322e10	300000	32377	4176	411	54
7	8	1119744	120531	39294	15171	426	23
8	7	186624	39648	16196	2124	146	18
9	8	221184	100731	4140	847	83	14
10	5	2304	680	188	152	45	11
11	11	161243136	300000	12766	3674	443	55
12	7	186624	4610	1852	354	88	23

[a] The systematic search was terminated if the number of VDW clash free conformers was greater than 300 000. [b] Conformers rejected if more than 10 kCal above the estimated global minimum. [c] Largest hole the conformer set C_{energy} finds in set C_{subset} = 0.5 Å. [d] Largest hole the conformer set C_{energy} finds in set C_{subset} = 1.0 Å. [e] Largest hole the conformer set C_{energy} finds in set C_{subset} = 1.5 Å.

of B. For this experiment, the RMS difference in heavy atom positions, computed using rigid body superposition,[25] is used as the interconformer distance metric.

A representative conformer set C_{subset} was generated from the quasi-exhaustive set C_{energy} using a simple heuristic, outlined in Algorithm 1. The conformer subset is generated until the maximum hole size found by set C_{energy} in set C_{subset} is less than a user-supplied value. These maximum hole size values were chosen to be on the same scale as reasonable database searching tolerances for distances.[24] This is explored further in a subsequent paper,[26] when these extracted conformational models are used for hypothesis generation and database searching.

Algorithm 1: Selecting a subset of conformers for a given coverage.

user supplies:
holeThreshold = Maximum two-set hole threshold permitted.
select a random conformer C from set C_{energy}
C_{subset} <-- *C*
do {
find conformer C in set C_{energy} *most distant from any conformer in* C_{subset}.
C_{subset} <-- C_{subset} *U C*
measure the largest hole, H_{max} *that set* C_{energy} *finds in set* C_{subset}.
} while (H_{max} < *holeThreshold)*

Conformational Coverage with Point Conformations. Table 3 shows the numbers of conformers generated and extracted by the heuristic outlined in Algorithm 1. The third column shows the theoretical number of conformational states, the M^N number. The fourth column shows the number of conformers that were found by the systematic search, with 1D minimization, that were VDW clash free. These conformer sets are denoted by C_{vdw}. Some molecules were found to be very flexible, and the systematic search was terminated at 300 000 VDW clash-free conformations. Column 5 shows the number of conformers that were retained after energy minimization, denoted by C_{energy}. Here, conformers were retained that were within 10 kcal/mol of the estimated global minimum for the whole set. Columns 6–8 show conformational models extracted from the energy set (C_{energy}) subject to hole size requirements of 0.5, 1.0, and 1.5 Å, respectively. These models are denoted by C_{subset}. An alternative way to describe the conformational models extracted is that there is no conformer from the original set that is further than the specified hole size from any conformer in the extracted set. Table 4 assesses the efficacy of M^N as an estimator of conformational coverage by computing ratios of M^N to the number of conformations in different conformational sets. Molecules were entered in this table if the systematic search was completed successfully (i.e., before generating 300 000 conformers). Table 5 assesses the efficacy of C_{vdw} as an estimator of conformational coverage by computing the ratio of C_{vdw} and the number of conformers in different conformational sets. Molecules were entered in this table if they exceeded the 300 000 conformer limit for VDW clash-free conformers (where the systematic search was terminated).

Table 4. M^N as an Estimator of Conformational Coverage

mol. no.	no. rot bonds	M^N	$\frac{M^N}{C_{vdw}}$	$\frac{M^N}{C_{energy}}$	$\frac{M^{N\,(a)}}{C_{energy}}$	$\frac{M^{N\,(b)}}{C_{subset}}$	$\frac{M^{N\,(c)}}{C_{subset}}$
1	4	864	2.2	6.3	18.4	123.4	288.0
2	7	124416	6.2	56.4	295.5	2704.6	15552.0
4	6	31104	7.5	23.3	100.3	622.0	3110.4
7	8	1119744	9.3	28.5	73.8	2628.5	48684.5
8	7	186624	4.7	11.5	87.9	1278.2	10368.0
9	8	221184	2.2	53.4	261.1	2664.9	15798.9
10	5	2304	3.4	12.3	15.2	51.2	209.5
12	7	186624	40.5	100.8	527.2	2120.7	8206.3

[a] Largest hole the conformer set C_{energy} finds in set C_{subset} = 0.5 Å. [b] Largest hole the conformer set C_{energy} finds in set C_{subset} = 1.0 Å. [c] Largest hole the conformer set C_{energy} finds in set C_{subset} = 1.5 Å.a

Table 5. C_{vdw} as an Estimator of Conformational Coverage

mol. no.	rot. bonds	C_{vdw}	$\frac{C_{vdw}}{C_{energy}}$	$\frac{C_{vdw}^{(a)}}{C_{subset}}$	$\frac{C_{vdw}^{(b)}}{C_{subset}}$	$\frac{C_{vdw}^{(c)}}{C_{subset}}$
3	10	300 000	24.3	173.9	1935.5	15789.5
5	10	300 000	7.9	60.8	980.4	10000.0
6	14	300 000	9.3	71.8	730.0	5555.6
11	11	300 000	23.5	81.7	677.2	5454.5

[a] Largest hole the conformer set C_{energy} finds in set C_{subset} = 0.5 Å. [b] Largest hole the conformer set C_{energy} finds in set C_{subset} = 1.0 Å. [c] Largest hole the conformer set C_{energy} finds in set C_{subset} = 1.5 Å.

There are several important observations to be made from Tables 3–5.

(a) The number of VDW clash free conformations (C_{vdw}) is a small fraction of the total number of theoretical states (M^N), even with the 1D minimization rescuing some of these states that would otherwise have been rejected by a conventional systematic search. At best M^N estimates the number of accessible conformers to within a factor of 2 (molecule 1) but more often errs by a factor of 5–10.

(b) Considering energy limits the size of the conformational space: Now our M^N estimator errs by a factor of 10–100 in estimating the number of conformers found subject to an energy threshold. Clearly, the conformational space bounded by energy is much smaller than the conformational space bounded by the condition that the conformer be VDW clash free. This is an important result. It suggests that simply resolving VDW clashes is *insufficient to produce a conformer that is acceptable on steric and energetic grounds.* Therefore, it should be noted that simply checking for "bump-relief" in molecules during 3D database searching is not sufficient to guarantee that a low-energy conformer is returned by the search. Even cursory inspection of the C_{vdw} and C_{energy} columns of Table 3 shows that a VDW clash free conformer has about a 1-in-10 chance of being energetically reasonable (to within 10 kcal of the estimated global minimum energy).

(c) When the concept of coverage is introduced, and an attempt is made to cover conformational space with a small collection of diverse conformers, the M^N estimator breaks down completely. It overestimates the number of conformers required by a factor of 10−100 (for a hole size of 0.5 Å) to 100−10 000 (for a hole size of 1.5 Å). It is apparent that the M^N estimator in Table 4 (or the C_{vdw} estimator in Table 5) performs poorly at estimating the number of conformers required to cover space at a resolution of 0.5−1.5 Å.

(d) The typical numbers of conformers required to cover conformational space is surprisingly low. Contrary to the frequently-quoted assertion that "millions of conformations are required to cover conformational space",[29] it can be seen that for hole sizes of 1 Å, the number of conformers required ranges from the low tens to a few hundred. Cursory inspection of the last three columns of Table 3 suggests that as the hole size requirement decreases, an exponentially decreasing number of conformations are required to cover the space. As a reminder, "hole size requirement" means that in extracting a small conformational set from a larger one, extraction of conformers continues until the original conformer set finds a hole no larger than a specified value in the set being extracted.

(e) The exponentially decreasing number of conformations required to cover conformational space (as a function of hole size requirement) is demonstrated further in Figure 4. In this graph, the number of conformers extracted from the full energy minimized set (C_{energy}) is plotted as a function of hole size. This shows, for each molecule and hole size, how many conformers are required to represent the conformational space such that there is no other conformer further away than the specified hole size. It can be seen that to cover the space to a hole size of 1 Å, between 10 and 450 conformers are required, depending on the molecule. Conversely, if one is permitted to spend 200 conformers covering the space, hole size ranging from less than 0.6 to 1.2 Å can be attained.

CONCLUSIONS

This paper has analyzed conformational coverage as an alternative approach to conformational analysis. Previous approaches have depended on the elucidation of the global minimum energy conformation and other low-energy local minima. We explored how well low energy conformational spaces of small- to medium-sized drug molecules can be covered by small collections of conformations and compared the number of conformations required with an oft-quoted empirical estimator of flexibility: the M^N estimator.

Systematic search was used to explore the low-energy conformational spaces of the test molecules thoroughly using a novel variation that performs 1D minimizations during the systematic expansion of the search tree of rotatable bonds. It was found that the usefulness of 1D minimization varied widely from molecule to molecule, resolving bad VDW clashes from 8% to 68% of the time, as shown in Table 2.

It was found that the number of VDW clash free conformers found by the systematic search was significantly lower than that estimated by M^N. When energy is considered, and only conformers within 10 kcal energy threshold of the estimated global minimum are retained, the number of conformers remaining is smaller still. It can be seen from Table 4 that the M^N estimator overestimates the number of conformers required to cover low-energy conformational spaces by a factor of 6−100.

The hole size metric, a concept of conformational coverage introduced in,[15] permits the measurement of conformational coverage by a small number of conformations—the hole size metric. Using this metric, conformer subsets were extracted from the set of all low-energy conformations subject to a particular hole size requirement. It was found that surprisingly few conformations are required to cover the space to a hole size H (i.e., there exists no other low-energy conformation that differs by more than HÅ RMS from any conformer in the extracted conformer set). It was found that an average of the following occurred:

*For 2828 conformers covered all holes of size 0.5 Å, and M^N overestimates by a factor of 10−500.

*For 184 conformers covered all holes of size 1.0 Å, and M^N overestimates by a factor of 50−3000.

*For 22 conformers covered all holes of size 1.5 Å, and M^N overestimates by a factor of 200−40 000.

The relationship between hole size and the number of conformations required to cover space to that particular hole size was plotted in Figure 4. It demonstrates the exponential nature of conformational coverage—that an exponentially increasing number of conformers is required to cover space to finer tolerances. However, it can be seen that good coverage (to about 1 Å RMS) can be obtained with 10−450 conformers, depending on the molecule.

It may be concluded from this study that low-energy conformational spaces of small- to medium-sized molecules, typified by the molecules in Figure 3, can be represented by 10−500 conformations for a hole size of 1 Å.

ACKNOWLEDGMENT

The authors would like to thank Jonathan Greene and Peter Towbin for many valuable discussions and insights that contributed to this work.

APPENDIX 1: TORSION CONSTRAINTS USING THE HYPERBOLIC TANGENT FUNCTION

What is meant by a requirement that a torsion angle between any four noncollinear points lies between a user-specified minimum and maximum value? We contend that the user means that the torsion angle should be free to adopt any value, without prejudice, as long as it remains in the specified range, and should be effectively disallowed when the range is violated.

One functional form proposed to model a torsion subject to a constraint is shown in eq 2.

$$f_i = \sum_i W|(\text{sign}((\cos t_0)^2, t_0) - (\text{sign}((\cos t)^2, t)))| \quad [30] \tag{2}$$

sign(a,b) = standard FORTRAN function computing the value of a * sign of b

i = loop over all torsion constraints

t_0 = center of allowed torsion range

t = current torsion value

This does not fully model the intention of a torsion constraint, nor is it smooth and continuously differentiable for the purposes of local minimzation. Equation 2 is not continuously differentiable and is nonzero if the torsion constraint is satisfied but not exactly at $t = t_0$.

We propose a new functional form for torsion constraints based on the hyperbolic tangent (tanh) and a sine function

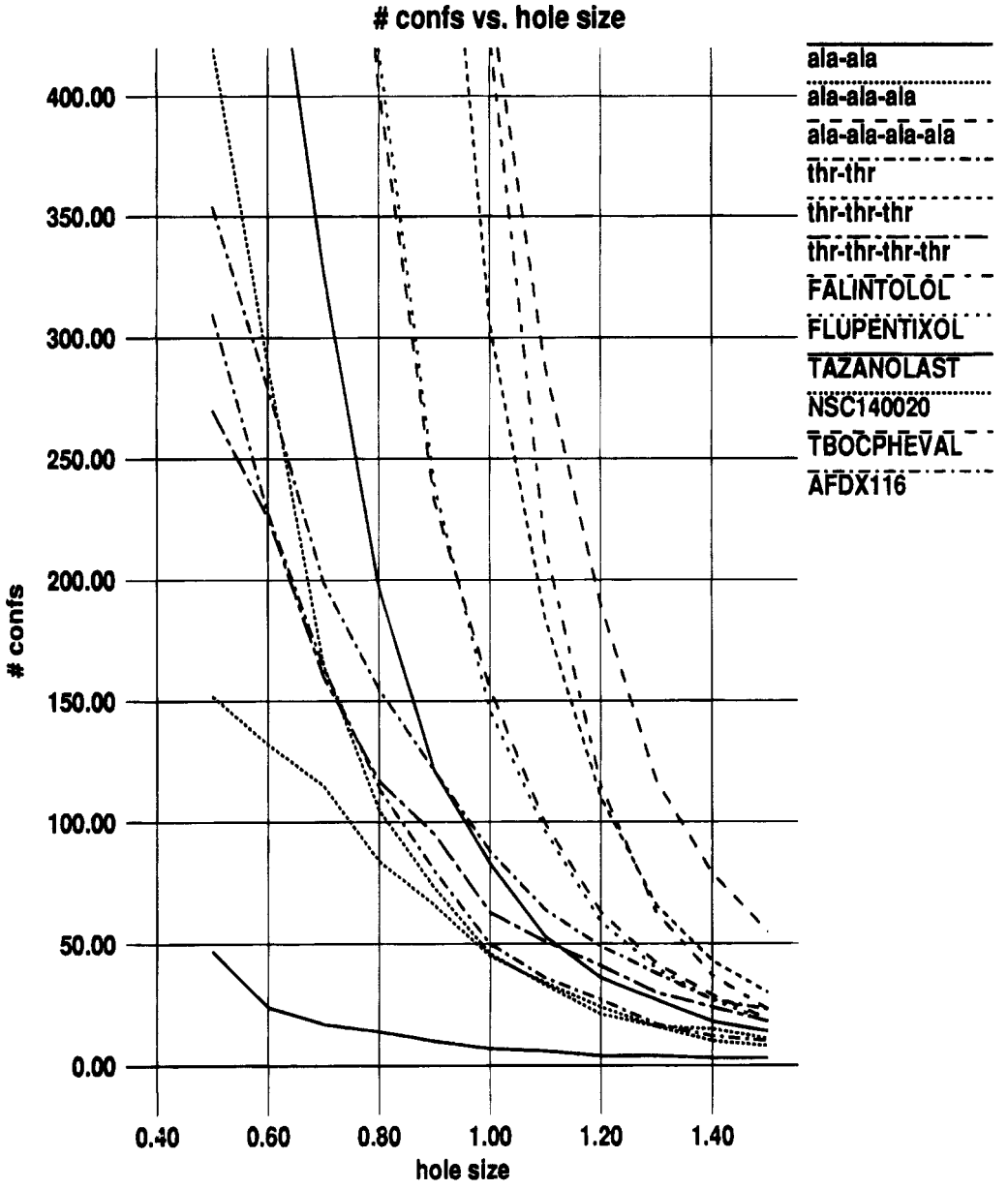


Figure 4. Plot of the number conformers required to cover space to a given hole size.

that has a unique minimum at $t = t_0$ for a torsion constraint with tolerance tol, which we call tanh scaling. The tanh scaling functions are described by

$$F = f^n(\tanh(w_1 f) + 1) \tag{3}$$

where

$$f = w_2\left(\sin^2\left(\frac{1}{2}(t - t_0)\right) - C\right) \tag{4}$$

and

$$C = \left(\sin\left(\frac{1}{2}(\text{tol})\right)\right)^2 \tag{5}$$

To ensure that F is an even function, n must be even, and we have used $n = 2$ in this study. Equation 4 is a simple sine function with a unique minimum, over the range $0-2\Pi$, at $t = t_0$. A constant C is used to translate the curve parallel to the Y-axis so that f is negative when the torsion constraint is satisfied and positive elsewhere. This is illustrated in Figure 5, where it can be seen that eq 4 has many of the properties desirable in a torsion constraint function: it is smooth and continuously differentiable and has a unique minimum at the center of the allowed torsion range, t_0. The final problem is that of scaling this equation to have a negligible effect within the allowed torsion range, which is achieved by the use of the hyperbolic tangent function.

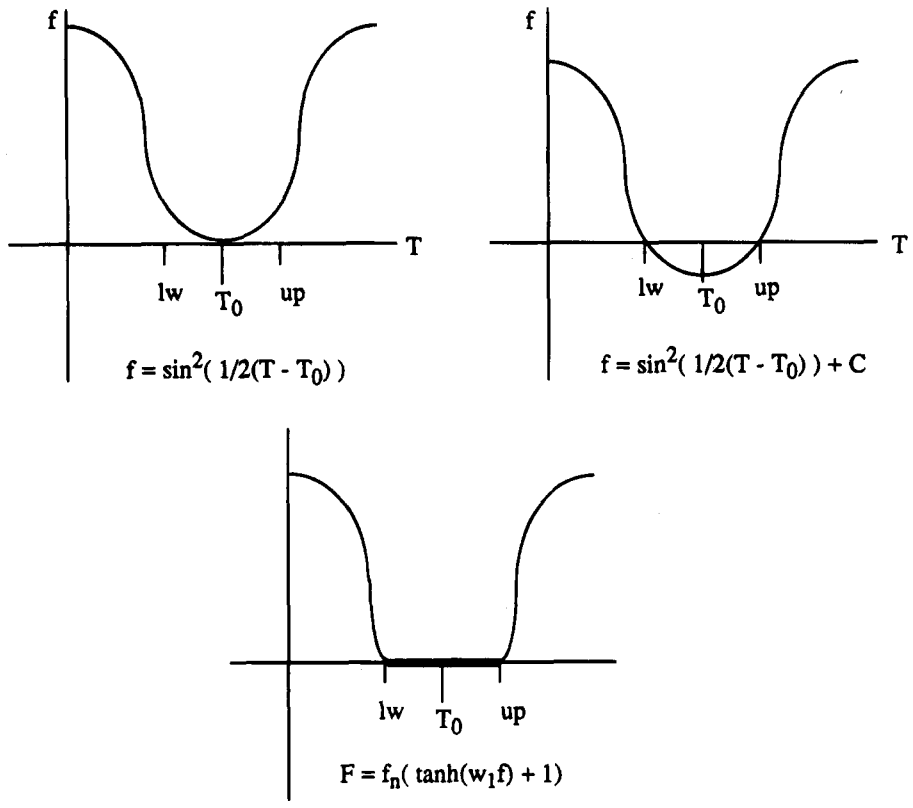
If eq 4 is used as an argument to eq 5, this has the effect of scaling the function smoothly and quickly approaches zero anywhere within the allowed torsion range and rises rapidly everywhere else. All that remains is to compute the constants w_1 and w_2.

The constant w_2 controls how much the torsion is penalized when it deviates from the allowed range. It is prudent to scale the penalty term to dominate even for small deviations, and w_2 can be obtained by considering the maximum value of f and F.

From Eq 4

$$f_{\max} = w_2(1 - C) \tag{6}$$

and from eq 3



lw, up = Lower and upper bounds of torsion T.

T_0 = Midpoint between lower and upper bounds.

C = A constant used to offset the torsion constraint.

F = Tanh scaling function for torsion constraints

Figure 5. Modeling a torsion constraint with a sine function.

$$F_{max} = 2f_{max}^n \tag{7}$$

so

$$F_{max} = 2w_2^n(1 - C)^n \tag{8}$$

and

$$w_2 = n\sqrt{\frac{F_{max}}{2(1 - C)^n}} \tag{9}$$

To penalize the torsion violation adequately, a large value for F_{max} is required: we use a value of 10 000 kJ.

The constant w_1 controls how rapidly the function if "switched on" from (almost) zero within the allowed torsion range to a large nonzero value when the constraint is violated. A desirable property of the function is that it rapidly falls to a near-zero value when the constraint is satisfied. This is stated more formally in eq 10.

Find

$$w_1 \text{ such that } F = f_s^n K \text{ at } t = t_0 \pm St_{tol} \; (K \ll 1, 0 \le S < 1) \tag{10}$$

where t_{tol} = tolerance of torsion constraint

At $t_s = t_0 \pm St_{tol}$, from eq 4

$$f_s = w_2\left(\sin^2\left(\frac{1}{2}(t_s - t_0)\right) - C\right) = w_2\left(\sin^2\left(\frac{1}{2}(St_{tol})\right) - C\right)$$

but, from eq 3 at $t = t_s$

$$F_s = f_s^n(\tanh(w_1 f_s) + 1) = f_s^n K$$

so

$$K = (\tanh(w_1 f_s) + 1) \tag{11}$$

which gives

$$w_1 = \frac{a\tanh(K - 1)}{f_s} \tag{12}$$

We scaled the function using $K = 10^{-6}$ and $S = 0.95$. This means that at 95% of the allowed torsion range (i.e., near the extrema of the range), the torsion penalty function, given by eq 4 is scaled by K (i.e., scaled to insignificant compared to the molecule's energy).

APPENDIX 2: MEASURING CONFORMATIONAL COVERAGE WITH A HOLE SIZE METRIC

The hole-size measure is illustrated in Figure 6. A and B are conformational models (i.e., collections of conformations) of molecule M from conformational analysis methods A and B. Let $d(A_i,B_j)$ be the "distance" between conformers i (of

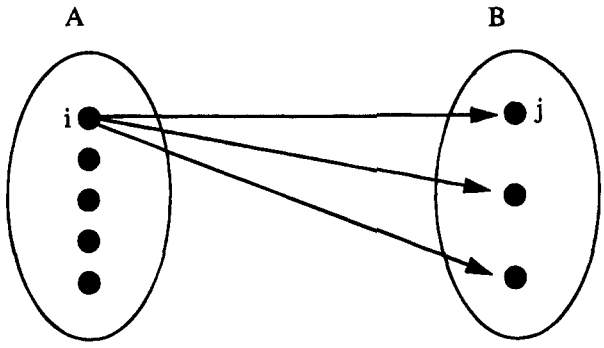


Figure 6. The two-set hole size measure.

A) and j (of B) and let $H(A_i, B) = MIN_i(d(-A_i,B))$. The quantity $H(A_i,B)$ is measuring the hole that the *i*th conformer of A finds in the entire conformational model B. It is the minimum of distances from A_i to all conformers j in B. Thus a lare value of $H(A_i,B)$ means that the *i*th conformer from A has found a large hole in the conformational model of B. Conversely, a large value for $H(B_i,A)$ means that the *i*th conformer from B has found a large hole in the conformational model A.

One final quantity remains to be defined, namely the largest hole that the conformational model A finds in the conformational model B for a given molecule *M*, denoted by $H(A,B)$. This is simply the maximum of all $H(A_i,B)$, for every conformer i from conformational model A and represents the largest hole that (any conformer of) A finds in the space of B.

REFERENCES AND NOTES

(1) Clustering: Romesburg, H. C. *Cluster Analysis for Researchers*; Lifetime Learning Pub., Wadsworth, Inc.
(2) Univariate topological indices: Mihalic, Z.; Nikolić, S.; Trinajstić, N.; Comparative Study of Molecular Descriptors Derived From The Distance Matrix, *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 28−37.
(3) Univariate topological indices: Trinajstić, N. *Chemical Graph Theory*, 2nd ed.; CRC: Boca Raton, FL, 1992.
(4) Bivariate topological indices: Petitjean, M. Applications of the Radius-Diameter Diagram to the Classification of Topological and Geometrical Shapes of Chemical Compounds. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 28−37.
(5) For a review, see: Leach, A. R. *Reviews in Computational Chemistry*; Lipkowitz, K., Boyd, D. B., Eds.; VCH Publishers, Inc.: 1991; Vol 2, Chapter 1, pp 1−47.
(6) Saunders, M. Stochastic Exploration of Molecular Mechanics Energy: Hunting for the Glocal Minimum. *J. Am. Chem. Soc.* **1987**, *109*, 3151−3154.
(7) Crippen, G. M.; Havel, T. F. *Distance Geometry and Molecular Conformations*; Research Studies Press, Wiley Publishers: 1988.
(8) Lipton, M.; Still, W. C. The Multiple Minimum Problem in Molecular Modelling. Tree Searching Internal Coordinate Space. *J. Comput. Chem.* **1989**, *9*, 343−355.
(9) Dammkoehler, R. A.; Karasek, S. F.; Berkley-Shands, E. F.; Marshall, G. A. Constrained Search of Conformational Space. *J. Comp. Aided Mol. Des.* **1989**, *3*, 3−21.
(10) Rigby, M.; Smith, E. B.; Wakeham, W. A.; Maitland, G. C. *The Forces Between Molecules*; Clarendon Press: Oxford, 1986.
(11) Ferhst, A. *Enzyme Structure and Mechanism*, 2nd ed.; Freeman: New York, 1985; pp 331−339, 342−343.
(12) Ferhst, A. *Enzyme Structure and Mechanism*, 2nd ed.; Freeman: New York, 1985; pp 263−292.
(13) CONCORD, Tripos Associates, St. Louis, MO 63144.
(14) See, for example: Saunders, M.; Jiminez-Vasquez, H. A. Stochastic Search for Lactone and Cycloalkene Conformers. *J. Comput. Chem.* **1993**, *14*, 330−348.
(15) Smellie, A.; Teig, S. L.; Towbin, P. Poling: Promoting Conformational Variation. *J. Comput. Chem.*, in press.
(16) Hypothesis in Catalyst, Technical White Paper, BioCAD Corp., Sunnyvale, CA 94086.
(17) Bures, M. G.; Danaher, E.; DeLazzer, J.; Martin, Y. C. New Molecular Modeling Tools Using Three-Dimensional Chemical Substructures. *J. Chem. Inf. Comput. Sci.* **1992**, *34*, 218−223.
(18) Eastman Kodak *Halftone Methods for the Graphic Arts*; Data Book Q-3, Eastman Kodak Company: Rochester, NY, 1968.
(19) Lipton, M.; Still, W. C. The Multiple Minimum Problem in Molecular Modelling. Tree Searching Internal Coordinate Space. *J. Comput. Chem.* **1989**, *9*, 343−355.
(20) Dammkoehler, R. A.; Karasek, S. F.; Berkley-Shands, E. F.; Marshall, G. A. Constrained Search of Conformational Space. *J. Comp. Aided Mol. Des.* **1989**, *3*, 3−21.
(21) Leach, A. R. An Algorithm to Directly Identify A Molecule's "Most Different" Conformations. *J. Chem. Inf. Comput. Sci.*, in press.
(22) Hestenes, Seifel, Methods of Conjugate Gradients for Solving Linear Systems. *J. Res. Nat. Bur. Standards* **1952**, *49*, 409−436.
(23) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaninathan, S.; Karplus, M. CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations, *J. Comput. Chem.* **1983**, *4*, 187−217.
(24) Greene, J.; Kahn, S.; Savoj, H.; Sprague, P.; Teig, S. L. Chemical Function Queries for 3D Database Search. *J. Chem. Inf. Comput. Sci.*, submitted for publication.
(25) Ferro, D. R.; Hermans, J. A Different Best Rigid-Body Molecular Fit Routine. *Acta Crystallogr.* **1977**, *A33*, 345−347.
(26) Smellie, A.; Kahn, S. D.; Teig, S. L. Analysis of Conformational Coverage 2. Applications of Conformational Models. xxx.
(27) Bunday, B. Basic Optimisation Methods. E. Arnold Pub.: Victoria, Australia, 1984.
(28) Gill, P. E.; Murray, W.; Wright, M. H. *Practical Optimization*; Acad. Press Inc.: San Diego, CA, 1981.
(29) Hurst, T. Flexible 3D Searching: The Directed Tweak Technique. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 190−196.
(30) Henry, D. Molecular Design, Ltd., private communication, March 1994.

CI940114S