Journal of Computer-Aided Molecular Design, 3 (1989) 3–21
ESCOM

J-CAMD 061

Constrained search of conformational hyperspace

Richard A. Dammkoehler[a,*], Steven F. Karasek[a], E.F. Berkley Shands[a] and Garland R. Marshall[b]

[a]*Department of Computer Science, Washington University, St. Louis, MO 63130 and* [b]*Department of Pharmacology, Washington University School of Medicine, St. Louis, MO 63110, U.S.A.*

Received 12 January 1989
Accepted 6 February 1989

Key words: Pharmacophore; Systematic search; Active analog approach; Angiotensin converting enzyme; Active site geometry

SUMMARY

We introduce a new method for determining pharmacophore or active site geometries by analysis of the structures of a series of active compounds. The method, constrained search, and the key concepts on which it is based, is described and illustrated by its application to 28 potent inhibitors of angiotensin-converting enzyme (ACE). The data set is one utilized by Mayer et al. [J. Comput.-Aided Mol. Design, 1 (1987) 3–16] to determine a unique geometry for the active site. Our experiment validated the previously reported results, obtained by a systematic search, while reducing the computer time requirement by more than two orders of magnitude. The experiment also identified a previously unrecognized alternative active site geometry for the ACE series.

INTRODUCTION

The common problem facing most medicinal chemists is the need to design compounds to interact with a therapeutic target whose 3D structure is unknown. Attempts to infer properties of the receptor by varying the structure have been hampered by the inherent flexibility of most drugs and the lack of an obvious basis for comparison. The process of varying the structure and observing the results on biological activity has led to the concept of the pharmacophore, i.e., the idea that certain functional groups play an important role in recognition and activation and that the 3D arrangement of such groups may encode some common pattern explaining the activity of diverse chemical structures at a given receptor [1–3]. Variation of the chemical structure of active compounds leads quickly to the conclusion that some parts of the molecule are critical for activity,

*To whom correspondence should be addressed.

whereas others can be changed, causing only minor variations in affinity. These observations led to the concept of the pharmacophore at the turn of the century. The methodology described here provides a means for determining the 3D topography of pharmacophoric groups.

Constrained search is a new procedure for determining the existence of common, 3D orientations of specified functional groups in a series of compounds, a critical aspect of the Active Analog Approach [4]. If the chemical structures of the molecules contained in the series are sufficiently diverse, a unique orientation or geometry can be identified. In the applications in medicinal chemistry, the compounds considered are often those which are active at a particular receptor, and the desired geometry is that of the pharmacophore or that of the binding groups of the receptor. Alternatively, the analysis may indicate the absence of commonality, or a set of equally probable geometries whose uniqueness must be evaluated by analysis of the structures of models of additional compounds or by 3D QSAR determinations [5].

In contrast to previously described systematic conformation search procedures from this group [6,7], constrained search utilizes information derived from the analysis of the substructure of each molecule in the series to restrict the current search and is similar in concept to the ensemble approach described by Sheridan et al. [8]. Results from each molecule in turn are used to restrict searches for subsequent molecules of the series. Effectively, this is a restriction of the search to those regions of conformational hyperspace in which sterically allowed conformations of all previously analyzed molecules are known to exist. In other words, the assumption of the existence of a common geometry is used as an active filter to truncate the combinatorial nature of the search problem. As a result, the computational requirement for the analysis of the structures of a series of molecules is (in general) significantly reduced.

We have applied constrained search to a data set, 28 potent inhibitors of ACE, utilized by Mayer et al. [1], to determine a unique geometry for the active site. Our experiment validated the previously reported results obtained with systematic search [9], while reducing the computational requirement by more than two orders of magnitude. In the sections that follow, we describe the concepts underlying constrained search, the performance of constrained search versus systematic search, and an experiment which has identified a previously unrecognized alternative active site geometry for the ACE series. The detailed evaluation of that alternative receptor site geometry is currently underway in our laboratory.

FUNDAMENTAL CONCEPTS

The pharmacophore model

For our purposes, a pharmacophore is defined in terms of the spatial relationship between functional groups whose orientation at the receptor site is essential for recognition and binding. Given a series of active analogs and the correspondence between functional groups contained in molecules in the series, search procedures can be used to determine the existence of common 3D orientations of the functional groups. A specific orientation is called a pharmacophore geometry. Typically, one atom in each functional group is selected as a reference and is called a pharmacophore reference point. One can also augment each molecule by postulated receptor groups in correct geometrical chemical association with the functional group. This is the receptor site hypothesis used in the ACE analysis. Except for the introduction of additional variables to represent the vir-

tual bonds between the functional groups and receptor points, the two applications are identical. For simplicity, we shall discuss the methodology in terms of the pharmacophore.

A pharmacophore model describes the distance relationships between reference points associated with k functional groups in each molecule. Here, k, the number of functional groups, is the dimensionality of the model. We define a regular model as one in which each of the $k(k-1)/2$ pairwise distances is a variable, and refer to models in which one or more of the distances are fixed or constant as restricted models. A pharmacophore geometry is represented by a point in S, a $k(k-1)/2$ dimensional distance space, whose orthogonal basis vectors are the distances between pairs of reference points. A partial geometry is defined as a subset of the pairwise distances and corresponds to a point in an $n(n-1)/2$ dimensional subspace of S, $(0<n<k)$. Estimates, or a priori knowledge of the minimum and maximum values of the allowed distances, define a closed partition of this continuous distance space. In constrained search, we approximate the enclosed volume by further subdividing the partition into hypercubes of edge length e.

Geometric similarity

In constrained search, two geometries represented by points in the distance space S are assumed to be similar if they are both contained in the same hypercube partition. Clearly, this assumption is valid when the hypercube edge is infinitesimal, and the geometries are, in that case, identical. As the value of the edge length increases, the probability of a classification error increases and uncertainties regarding the validity of the search results are introduced. A pragmatic approach, made feasible by the performance of constrained search, is to iterate the search, decreasing the edge length on each successive trial. This process terminates when it is no longer possible to identify any common geometries in the series.

Conformational hyperspace

The topography of a molecular conformation is completely determined by A, a set of Cartesian coordinates, where $|A| = n$ is the number of constituent atoms, and $a_i \varepsilon A$ is the coordinate triple defining the equilibrium position of the atomic nucleus of the ith atom. For a conformationally flexible molecule, there is an infinite number of coordinate sets, each specifying the topography of a particular conformation. Within the fixed valence geometry approximation, we represent the infinite set of conformations of a molecule with m rotational degrees of freedom in a continuous space in which each dimension corresponds to a variable torsion angle, ω_j, $(1 \leq j \leq m)$*.

Let P be a point in the continuous hyperspace. The coordinates of P, $(p_1, p_2 \ldots, p_m)$, $0 \leq p_j < 360$, are the values of the torsion angles $(\omega_1, \omega_2 \ldots, \omega_m)$ defining a particular conformation of the molecule. The point P represents a sterically allowed conformation if, for all nonbonded atom pairs, the inequality below is satisfied.

$$d_{ij} - c_{ij} \geq 0, \quad (1 \leq i < j \leq n) \tag{1}$$

*We note, as did Brucceleri and Karplus [10], that the choice of ω_j for a general torsion angle conflicts with the more common use of ω as the peptide torsion angle. However, consistency with the notation used by Go and Scheraga [11] is maintained, and we continue their usage.

Here, d_{ij} is the Euclidian distance between the nonbonded atoms i and j, and c_{ij} is the sum of their van der Waals radii.

We approximate the continuous hyperspace by a discrete topological space of the same dimensionality. A generic point in that space is defined by $P(s \cdot x_1, s \cdot x_2 \ldots s \cdot x_m)$, $(0 \leq x_j < 360/s)$. We further restrict the values of x_j to the integers $\{0, 1, 2 \ldots r\}$, where r is the largest integer less than 360/s, forming a uniform grid over all dimensions of the continuous hyperspace. Numbers in the set $\{0 \cdot s, 1 \cdot s \ldots r \cdot s\}$ are used in both systematic and constrained search as trial torsion angle values. The variable s is called the scan parameter, and its value determines the number of trial torsion angle values and the number of points, NP, in the discrete m-dimensional space. If s is an integer, $(0 < s \leq 360)$, then the number of points (i.e., the number of conformations to be evaluated) is given by the formula below:

$$NP = r^m \tag{2}$$

where $r = 360/s$.

Consider a molecule in which there are six torsional rotations. If we select a scan parameter of 3.6°, then there are 100 trial torsional angle values to apply on each of the six rotations. Using the formula above, we find that a trillion conformations must be generated and evaluated for consistency with the van der Waals constraints in order to systematically search the discrete conformational hyperspace available to that molecule. This formidable computational task can be reduced to manageable proportions by the approach described in the next section. There we consider a molecule as a collection of molecular fragments called aggregates, and simulate the process of assembling sterically allowed conformations from those building blocks.

Composition of molecular fragments

We represent the topology of the molecular structure as a set of points representing the atomic nuclei and a set of point pairs or lines representing covalent bonds. If the molecule is cyclic, one bond is deleted from each distinct ring and distance constraints (discussed in the next section) are introduced in the subsequent search in order to preserve the ring geometry. From this acyclic description, we decompose the structure into maximal subsets (aggregates) of atoms whose interatomic distances are invariant with respect to a torsional rotation around an axis defined by the three-space coordinates of any two atomic nuclei in the subset. Two aggregates are said to be adjacent if they share a pair of atoms connected by a covalent bond. Adjacent aggregates can be combined if there is no rotational freedom associated with the shared bond. Once identified, the aggregates are assigned sequential numbers which determine the order in which they will be combined in the process of building a complete molecule from its constituent parts. A detailed presentation of the mathematical basis for these operations is given in Refs. 6 and 7.

In general, there are m torsional degrees of freedom and m+1 aggregates associated with a molecule. However, the positions of the pharmacophore reference points and the distances between them are usually determined by a subset of the torsional rotations. This case is illustrated in Fig. 1, where the structure of a molecule with seven torsional degrees of freedom is described by a graph model. The seven edges in the graph are labeled to indicate their correspondence to the torsional rotations, and the eight vertices represent the set of aggregates. By convention, the ver-

tex labeled A_0 contains the first pharmacophore reference point, and, in this example, the second and third reference points defining a 3D pharmacophore model are contained in A_2 and A_5. Evidently, the distance between a point in A_0 and a point in A_2 is determined by a subset of the rotations $\{\omega_1, \omega_2\}$. Similarly, the distance between a point in A_0 and A_5 is determined by the subset $\{\omega_3, \omega_4, \omega_5\}$. The third distance is determined by the subset which contains all elements of those two subsets. The remaining rotations, ω_6 and ω_7, determine the position of atoms contained in aggregates A_6 and A_7, but do not influence the pharmacophore geometry. In subsequent sections of this paper, we will refer to the atoms contained in A_0 and in aggregates corresponding to rotations which determine the pharmacophore distances as the *pharmacophoric core* of the molecule (i.e., the subset of atoms contained in A_0, A_1, ... A_5 of Fig. 1).

Starting with the first pair of aggregates A_0 and A_1, the search program, using a predetermined list of trial values, finds all values of the first torsion angle which will produce sterically allowed combinations of A_0 and A_1. We refer to each sterically allowed combination of j aggregates ($1 \leq j \leq m$) as a *partial* conformation, and define a *complete* conformation as a sterically allowed combination of all $m+1$ aggregates. For each partial conformation of A_0 and A_1, we determine the values of the second torsional angle which produce partial conformations of A_0, A_1 and A_2. The process in which larger combinations are assembled by including additional aggregates continues until a complete conformation is found and recorded, or in attempting to include the next aggregate, the program determines that there are no trial values which produce sterically allowed

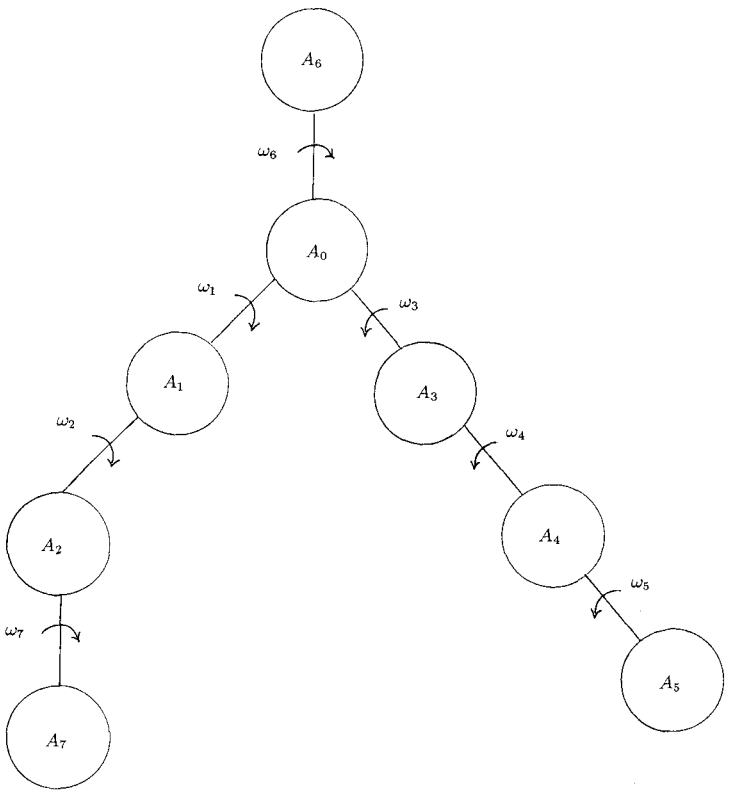

Fig. 1. Illustrating the relationship between aggregates and rotations.

combinations. In either case, the orientation of the last aggregate in the current conformation is changed (i.e., the atoms in the last aggregate are rotated creating a new partial conformation), or the last aggregate is deleted and the orientation of the last aggregate in the reduced conformation is changed to create a new partial conformation. The process continues, adding and deleting aggregates in a systematic manner, until all possible complete conformations have been identified and recorded.

The procedure as described is called systematic search and is technically a depth first search with chronological backtracking [12]. The term backtracking references deleting an aggregate from the current conformation and returning to consideration of a smaller partial conformation. The advantage of a depth first search is its logical simplicity and ease of implementation in hardware or software. However, the time complexity of systematic search, dominated by the cost of verifying that each partial or complete conformation satisfies the van der Waals distance constraints, is substantial. Clearly, as we demonstrate with constrained search, other procedures can significantly reduce the time required for a conformation search.

Distance constraints

In Refs. 6 and 7, we present the derivation of an equation with scalar coefficients which describes the variable distance between two atoms as a function of a single torsional variable. If a_i represents the nucleus of atom i and a_j represents the nucleus of atom j, then, as defined by a_s and a_r, atom j rotates around an axis as illustrated in Fig. 2, the square of the interatomic distance is given by

$$d_{ij}^2(\omega) = d_1 + d_2 \cdot \cos\omega + d_3 \cdot \sin\omega \tag{3}$$

By resolving the vector $\boldsymbol{v}$ into three orthogonal components, using the direction cosine vector $\boldsymbol{u}$,

$$\boldsymbol{v}_3 = \boldsymbol{v} \times \boldsymbol{u}, \quad \boldsymbol{v}_2 = \boldsymbol{u} \times \boldsymbol{v}_3, \quad \boldsymbol{v}_1 = \boldsymbol{v} - \boldsymbol{v}_2, \tag{4}$$

the coefficients of Eq. 3 are obtained by the operations shown below:

$$d_1 = |\boldsymbol{s}|^2 + |\boldsymbol{v}|^2 - 2(\boldsymbol{s} \cdot \boldsymbol{v}_1) \tag{5}$$
$$d_2 = -2(\boldsymbol{s} \cdot \boldsymbol{v}_2) \tag{6}$$
$$d_3 = -2(\boldsymbol{s} \cdot \boldsymbol{v}_3) \tag{7}$$

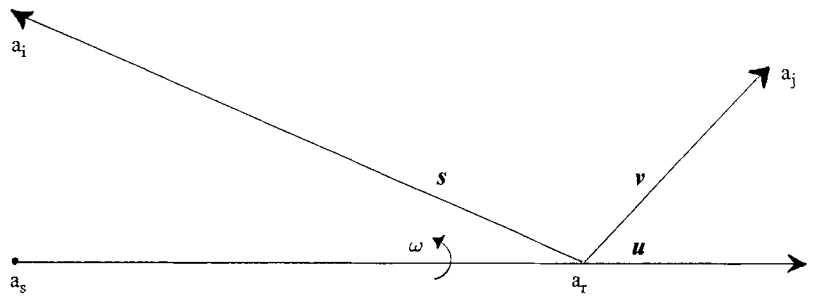

Fig. 2. Illustrating the differential distance function.

If, for any pair of atoms, the minimum and maximum distances are known, then prior to the last torsional rotation on the path from the aggregate containing atom i to the aggregate containing atom j, the equation above can be used to eliminate trial values of ω which would produce conformations at variance with the known distance range(s). Similarly, if i and j are pharmacophore reference points and the minimum and maximum distances define the length of an edge of the closed partition of distance space (i.e., the range of allowed distances between pharmacophore reference points), then values of ω which produce a partial or complete conformation outside the enclosed partition can be eliminated.

The distance Eq. 3 can be converted to quadratic form by a substitution of variables, $x = \sin\omega/(1+\cos\omega) = \tan(\omega/2)$, then $\sin\omega = 2\times/(1+x^2)$ and $\cos\omega = (1-x^2)/(1+x^2)$, and

$$d_{ij}^2(x) = (ax^2 + bx + c)/(1 + x^2) \tag{8}$$

where $a = d_1 - d_2$, $b = 2d_3$, and $c = d_1 + d_2$.
The values of x which minimize or maximize $d_{ij}^2(x)$ are given by

$$x = \pm \frac{\sqrt{d_2^2 + d_3^2 \mp d_2}}{d_3} \tag{9}$$

Those values may be substituted into Eq. 8 to find the minimum and maximum distances between atoms i and j as a function of the torsional variable.

Equation 10 is called the differential distance function. Observe that the value of the function

$$\delta_{ij}(\omega) = d_{ij}^2(\omega) - c_{ij}^2 \tag{10}$$

is positive when $d_{ij}^2(\omega) \geq c_{ij}^2$. Again, c_{ij} is the sum of the van der Waals radii of two nonbonded atoms i and j.

By converting Eq. 10 to quadratic form, we can evaluate the resulting discriminant,

$$D = b^2 - 4(a - c_{ij}^2)(c - c_{ij}^2) \tag{11}$$

to determine whether there are any values of $\omega = 2 \arctan(x)$ for which $\delta_{ij}(\omega) = 0$. If $D > 0$, then δ_{ij} has real roots, indicating that the van der Waals constraint will be violated for some range of values of ω. If $D \leq 0$, the Eq. 10 has complex or real double degenerate roots indicating that $\delta_{ij}(\omega)$ is strictly negative or respectively positive for all ω.

The differential distance function is used to minimize the number of pairwise distances which must be evaluated in order to verify that each partial (or complete) conformation satisfies the van der Waals distance constraints. If the function has only imaginary roots then there are no values of ω which would produce a contact between atoms a_i and a_j, and no distance calculations involving such pairs are required. If for any atom pair, $\delta_{ij}(\omega)$ is strictly negative, then all values of ω will produce a contact and there are no sterically allowed compositions of the partial conformation containing a_i and the aggregate containing a_j.

CONSTRAINED SEARCH

A pseudocode description of the constrained search procedure is shown in Fig. 3. The data ar-

```
Procedure Search (m, Hcubes): recursive

    BEGIN size : = m

        Update   (AG[m])
        Screen   (LC[m], LP[m], IM[1], Angles)
        Validate (AG[m], LP[m], CD[m], Angles)

        IF size = all aggregates

        THEN Record (LP[m], IM[1], IM[2], Hcubes)
             Backtrack (core size, m)

        ELSE WHILE    (Angles > 0)     AND
                      (Hcubes > 0)     AND
                      (size = m)

                      Rotate (AG[m], LP[m], Angles)
                      m: = m + 1
                      Search (m, Hcubes)

             END WHILE

             m: = m - 1

        END IF

    END
```

Fig. 3. Pseudocode description of constrained search.

ray containing trial torsion angle values is globally accessible. The data arrays CD[m] hold the van der Waals constraint distances. The data arrays AG[m] contain the atomic coordinates of atoms in each aggregate. The arrays LC[m] are lists of distance constraints applicable to a composition of the first m aggregates. The parameter m corresponds to the number of aggregates contained in the current partial conformation. Search is called initially by an external procedure with the parameter value m set to one. The parameter Hcubes specifies the number of candidate geometries at the start of each search. Search is called recursively each time an aggregate is to be added to the current partial conformation. A return from Search corresponds to the deletion of an aggregate from the current conformation. The subroutine Update calculates the coordinates of the atoms contained in aggregate m to reflect the current values of rotation variables on the path from A_0 to A_m. If any distance constraints terminate in aggregate m, the subroutine Screen, using Eq. 3 and the upper and lower bound of each constraint, determines which values contained in LT will produce partial or complete conformations in which those constraints are satisfied. If aggregate m contains a pharmacophore reference point, subroutine Screen determines which of the trial values will produce conformations containing a candidate pharmacophore geometry. Here, Eq. 3 is also used to calculate the pharmacophore distance(s). The distance value(s) are compared with the set of candidate pharmacophore geometries, and torsional angle values which do not produce

partial or complete conformations of interest are discarded. The remaining trial values are collected in the data array LP[m]. This feature will be discussed in greater detail in the following section. Subroutine Validate calculates the interatomic distances between atoms contained in aggregate m and those contained in the current partial conformation using each of the values in LP[m] and van der Waals minimum distances contained in CD[m]. Trial values which do not produce sterically allowed partial or complete conformations are removed from LP[m].

Subroutine Record is called when sterically allowed complete conformations have been generated. In Record, the minimum and maximum values of distances between pharmacophore reference points are updated as well. On completion of a constrained search, these values define the closed partition of distance space for the next molecule in the series.

The inner loop of Search is executed once for each value in LP[m] which will produce a sterically allowed partial conformation. Subroutine Rotate calculates the coordinates of the atoms contained in aggregate m using each torsion angle value, and returns the number of torsional values remaining in LP[m]. A globally accessible array containing the current values of the torsion angles is also updated by the subroutine Rotate. This information is also used in subroutine Record when the coordinates of all sterically allowed conformations are requested. Then Search is recursively called with parameter m + 1 to continue the composition process.

Constrained search maintains two special data structures called the information maps. The primary map, IM[1], contains encoded descriptions of the hypercubes containing pharmacophore geometries common to all previously analyzed analogs. These are referred to as the candidate geometries. Logically, the primary map contains all possible hypercubes and candidate geometries prior to the processing of the first molecule in a series. During the analysis of each molecule, the secondary map, IM[2], contains descriptions of pharmacophore geometries which were contained in the primary map when the analysis began and subsequently found in a sterically allowed conformation of the current molecule. The primary map is referenced by subroutine Screen to select as trial values of the torsion angle only those which produce conformations containing partial or complete pharmacophore geometries common to all previously analyzed molecules. The subroutine Validate determines whether or not such conformations are sterically allowed and subroutine Record updates the secondary map when each complete conformation is generated. In Record, the hypercube containing the pharmacophore geometry is also removed from the primary map, preventing the redundant generation and evaluation of any other conformation of the current molecule containing the same geometry. Record returns the number of hypercubes remaining in the primary map.

The look-ahead capability implemented in subroutines Screen and Record accounts for 88.2% of the observed performance improvement in constrained search. In effect, the number of partial or complete conformations which must be evaluated for consistency with the van der Waals distance constraints is minimized by this approach*.

Performance is also enhanced by the capability to terminate a constrained search when the existence of all candidate pharmacophore geometries in a particular molecule has been verified. In the analysis of the ACE series, this situation was encountered 14 times. Termination occurs when

---

*The computational complexity of each invocation of Validate is of order $0(v_m \cdot n_m^2)$ where v_m is the number of torsion angle values constrained in LP[m] and n_m is the number of atoms in a conformation consisting of the first m aggregates.

all hypercubes containing candidate pharmacophores have been removed from the primary information map. Continuing the search when all relevant information has been extracted from a particular structure clearly produces no additional information regarding the validity of the initial pharmacophore hypothesis. This feature accounts for 11.7% of the performance improvement.

A very small performance gain is obtained by eliminating the exhaustive evaluation of all conformations which are produced by varying the position of atoms contained in the *pharmacophoric core* of the molecule. Identification of the first combination of all angles which results in a sterically allowed complete conformation is sufficient to verify the existence of any particular pharmacophore geometry. Unnecessary calculations are avoided by backtracking to the smallest partial conformation containing all pharmacophore reference points. The subroutine Backtrack sets the value of m to that of the last aggregate containing a pharmacophore reference point. The orientation of the atoms in the last aggregate in that partial conformation is then modified, and the composition process is continued.

VERIFICATION AND PERFORMANCE

In order to test our implementation of constrained search, we repeated the experiment performed by Mayer et al. [1] using the data set containing models of 28 potent inhibitors of ACE. A closed partition of the distance space defined by the four-point restricted active site model (Fig. 4) was formed using a range of 2–12 Å on each of the five variable distances. The enclosed volume was subdivided into hypercubes with an edge length of 0.25Å.

Each of the 28 compounds shown in Fig. 5 were analyzed by our constrained search program, maintaining the order of analysis employed by Mayer et al. [1]. A scan parameter of 10° was used to duplicate the conditions of the earlier experiment. The number of torsional variables, as determined by Mayer et al. [1], varies across the series. In those terms, the smallest molecule(s) have three torsional degrees of freedom and the largest has nine. These data are summarized in Table 1.

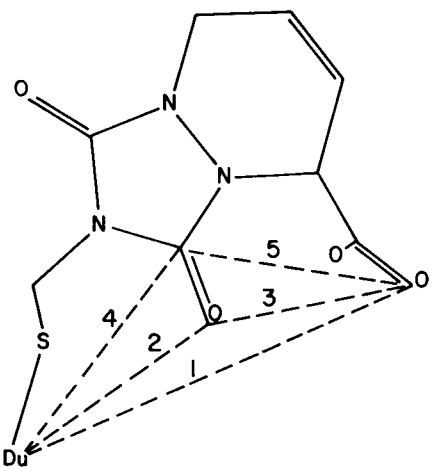

Fig. 4. Defining distances between the most important groups of the active site hypothesis derived from the structure–activity data. [From Mayer et al., J. Comput.-Aided Mol. Design, 1 (1987) 8.]

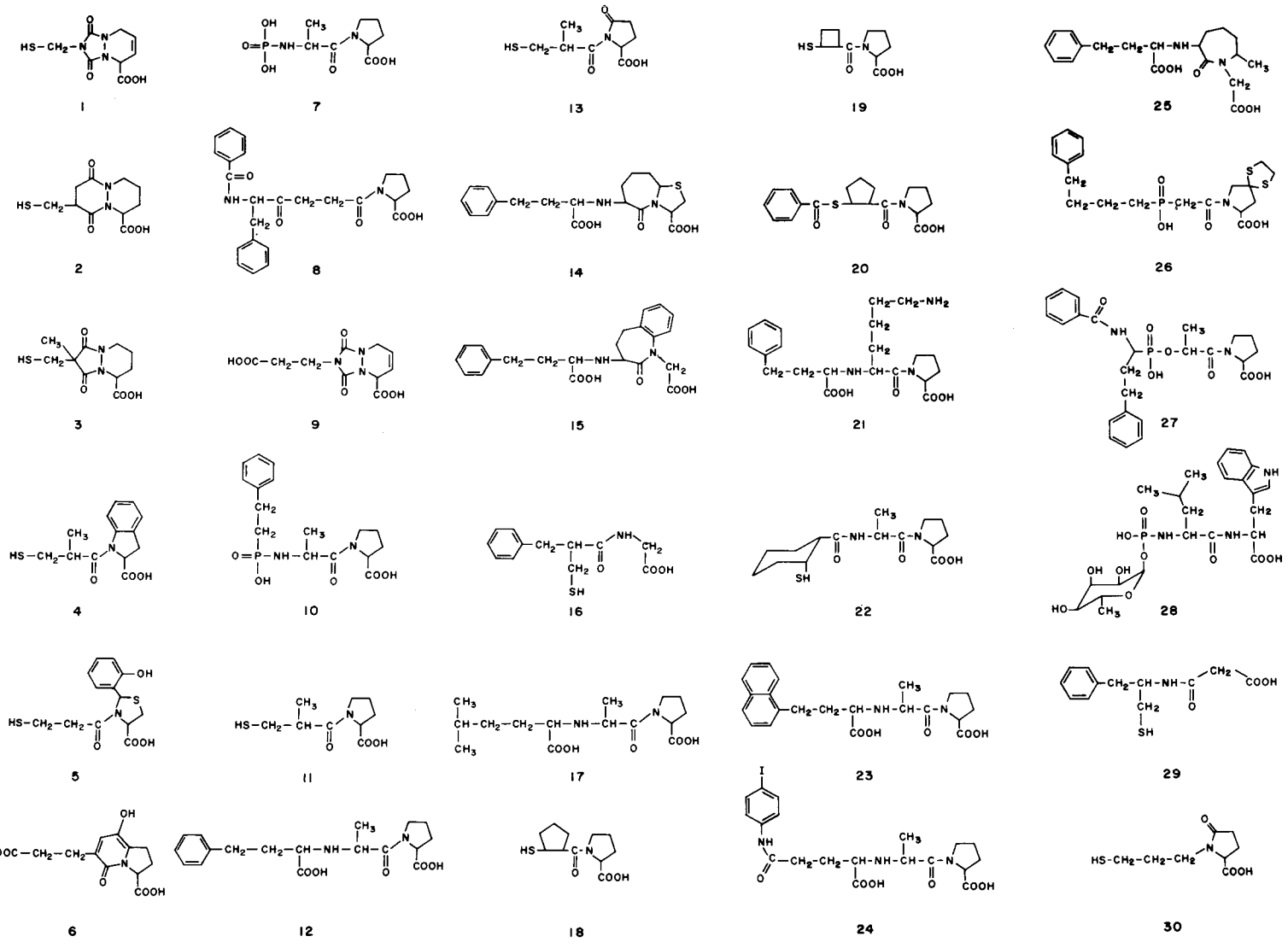

Fig. 5. Twenty-eight analyzed potent ACE inhibitors (1–28) and two inactive molecules. [From Mayer et al., J. Comput.-Aided Mol. Design, 1 (1987) 5.]

TABLE 1
RESULTS OF CONSTRAINED SEARCH ON THE ACE SERIES

Inhibitor no.	Rotatable bonds	No. of conformations found[a]	No. of Hypercubes	CPU time (s)
01	3	1316	1070	11.86
02	3	297	228	3.13
03	3	167	93	2.98
04	3	83	57	0.71
05	5	70	51	4.34
06	5	60	49	49.04
07	6	71	42	17.30
08	6	52	42	57.89
09	5	57	42	68.80
10	9	58	41	11.29
11	5	63	41	7.57
12	8	54	41	254.10
13	5	65	35	2.30
14	8	33	31	47.91
15	8	43	30	146.27
16	7	46	30	4.71
17	8	41	30	164.67
18	3	17	13	0.18
19	3	15	11	0.26
20	5	2	2	0.29
21	8	2	2	107.26
22	6	2	2	6.32
23	7	3	2	57.07
24	8	2	2	0.18
25	7	2	2	15.78
26	5	2	2	2.79
27	7	2	2	1.50
28	7	3	2	0.42
			Total	1046.92

[a]The number of complete conformations evaluated.

On completion of the constrained search, we found that only two hypercubes contained geometries common to each molecule in the series. As shown in Table 2, each hypercube contains one of the two common geometries reported previously, and, therefore, we conclude that the experiments produced equivalent results.

To further test the robustness of the constrained search implementation, we experimented with a variety of reorderings of the molecules. We found that orderings in which the number of torsio-

TABLE 2
DISTANCE VALUES OBTAINED BY CONSTRAINED AND SYSTEMATIC SEARCH

Systematic search	Published results [1] – 0.2 Å resolution				
	8.5432	5.0620	3.8082	5.2785	4.0953
	8.6946	4.9381	3.8082	5.3442	4.0953
Constrained search	0.25 Å resolution				
	8.5000	5.0000	3.7500	5.2500	4.0000
	8.5000	4.7500	3.7500	5.2500	4.0000
Constrained search	Alternate site – 0.25 Å resolution				
	6.0000	2.7500	3.7500	3.2500	4.0000
	5.7500	2.7500	3.7500	3.5000	4.0000

nal rotations increased monotonically reduced the time requirement by an additional factor of four. The results obtained with one such ordering are shown in Table 3. We also repeated Mayer's systematic search experiment under controlled conditions to obtain reliable performance data. Systematic search [9], running in single user mode on a μVAX II with nine Mbytes of main memory, was used to re-analyze the data set. The results are summarized in Table 4.

Direct comparison of the total time required for each search indicates that constrained search is 300 times faster than systematic search. The reduction in run time from more than 87 h to 17.4 min is attributable to the look-ahead feature, the ability to terminate when all candidate geometries have been found, and the ability to eliminate the time cost associated with torsional variations which do not affect the pharmacophore distances. In order to determine the relative importance of these three factors, we repeated the constrained search enabling various combinations of these program capabilities. The results are shown in Table 5. Note that these data were obtained on a μVAX III with 16 Mbytes of primary storage, running in single user mode using the ordering shown in Table 3. The μVAX III is approximately four times the speed of a μVax II.

SENSITIVITY ANALYSIS

The results of a search experiment clearly depend on a multiplicity of factors; the structural diversity of the molecules in the series and the values chosen for the scan parameter, hypercube edge length and van der Waals constraint distances. We have qualitatively explored the relationship between two of those factors using the ACE series and the active site model defined previously. Selection of the value of the hypercube edge length is one of the most important decisions which must be considered when designing a constrained search experiment. For constant values of the scan parameter, we repeated the search using hypercube edge lengths of 0.125 Å and 0.50 Å, respectively. No common geometries were found in the first case and two disjoint clusters of hypercubes were identified in the second case. One of the two clusters contained both the geometries identified when the edge length was 0.25 Å. The distance values are shown in Table 6.

TABLE 3
RESULTS OF CONSTRAINED SEARCH ON THE ACE SERIES

Inhibitor no.	Rotatable bonds	No. of conformations found[a]	No. of hyper-cubes	CPU time (s)
18	3	1102	940	3.83
01	3	322	263	6.96
02	3	36	32	1.24
19	3	33	26	0.37
03	3	32	21	2.08
04	3	25	21	0.48
20	3	8	8	0.62
05	5	7	6	1.45
13	5	5	5	0.94
06	5	8	5	12.69
09	5	5	5	0.91
11	5	5	5	1.67
26	5	5	5	7.55
07	6	5	2	10.32
08	6	3	2	0.59
22	6	2	2	6.35
16	7	2	2	0.60
23	7	3	2	57.07
25	7	2	2	15.76
27	7	2	2	1.52
28	7	3	2	0.40
12	8	3	2	1.06
14	8	2	2	0.39
15	8	2	2	1.66
17	8	3	2	68.04
21	8	2	2	107.35
24	8	2	2	0.18
10	9	2	2	1.40
			Total	313.48

[a]The number of complete conformations evaluated.

We then identified the first (and only) molecule in the series which had not produced a geometry corresponding to a hypercube in the second cluster. The scan parameter for one rotation in that molecule was reduced to 5° and the search was repeated with the hypercube edge length set to 0.25 Å. In this experiment, the same two distinct and equally probable common geometries were again identified. They are shown in Fig. 6, and the distance values are tabulated in Table 2.

As a final step in the analysis, we performed searches on each of the 28 ACE inhibitors to deter-

TABLE 4
RESULTS OF SYSTEMATIC SEARCH ON THE ACE SERIES

Inhibitor no.	Rotatable bonds	No. of conformations found[a]	No. of OMAP points	CPU time (s)
01	3	9.9E3	1779	52.16
02	3	9.1E2	203	32.18
03	3	7.7E2	77	11.59
04	3	1.5E2	24	3.66
05	5	3.2E3	23	43.82
06	5	4.2E3	23	163.04
07	6	1.5E4	23	239.67
08	6	2.4E4	23	13804.53
09	5	8.9E3	23	1725.90
10	9	2.2E4	23	26073.35
11	5	7.0E3	23	320.72
12	8	2.6E5	22	107684.68
13	5	4.0E2	19	45.61
14	8	9.2E3	17	1876.29
15	8	1.2E4	17	88279.39
16	7	8.6E3	17	6674.25
17	8	9.7E4	17	7016.68
18	3	1.3E1	7	6.82
19	3	3.0E1	5	2.76
20	5	5.0E0	1	7.75
21	8	5.2E3	1	9008.02
22	6	3.1E2	1	572.78
23	7	2.1E3	1	2913.67
24	8	2.7E4	1	27611.93
25	7	1.8E2	1	14911.28
26	5	1. 5E3	1	636.05
27	7	4.8E2	1	406.39
28	7	4.5E2	1	3168.61
			Total	313293.58[b]

[a]The number of complete conformations evaluated.
[b]Approximately 87 h of CPU time.

mine the correspondence between sample points in angle space and geometries contained in the hypercubes identified by constrained search. These results, summarized in Table 7, show for each molecule and the alternative geometries, the number of sterically allowed conformations contained in the sample. Here, we include only those conformations presenting either geometry at the receptor site discounting the multiplicity obtainable by varying the position of atoms outside the

TABLE 5
IMPORTANCE OF FACTORS CONTRIBUTING TO IMPROVED PERFORMANCE

Run no.		CPU (s)[a]	% Gain
1	No optimization	10652	–
2	Look-ahead only	1318	88.2
3	Look-ahead and early termination	79	99.9
4	All optimizations	78	100.0

[a]Obtained on a μ VaxIII system.

TABLE 6
DISTANCE VALUES OBTAINED BY CONSTRAINED SEARCH (0.5 Å RESOLUTION)

Group I	5.5000	3.0000	3.5000	3.5000	4.0000
	5.5000	2.5000	3.5000	3.0000	4.0000
	6.0000	3.5000	3.5000	3.5000	4.0000
	6.0000	3.0000	3.5000	3.5000	4.0000
	6.5000	3.5000	3.5000	4.0000	4.0000
Group II	8.0000	5.0000	3.5000	5.0000	4.0000
	8.0000	4.5000	3.5000	5.0000	4.0000
	8.5000	5.0000	3.5000	5.0000	4.0000
	8.5000	4.5000	3.5000	5.0000	4.0000

pharmacophoric core. The ranges of distances used to define the closed partitions of distance space for each search are given in Table 8.

DISCUSSION AND CONCLUSIONS

Systematic search procedures [4,13] have been characterized as incremental approaches [8]. In contrast, constrained search and Sheridan's ensemble distance geometry method [8] use the structural information derived from a set of compounds to identify common pharmacophore or active site geometries. Distance constraints of several types are used as active filters to restrict the search to regions of conformational hyperspace containing sterically allowed conformations common to the set of compounds. These filters correspond to the rules and filters employed by Moult and James [14], and provide a means for directly utilizing NMR distance measurements [15,16].

Our most recent implementation of constrained search also supports multiple lists of trial torsion angle values. As a result, well-defined angles range for main-chain, single bond, torsional variables may be utilized. Where appropriate, the atoms (aggregates) comprising side chains or lying outside the pharmacophoric core can be restricted to staggered conformations using the multiple list mechanism.

Constrained search preserves the completeness property of a systematic search while substantially reducing the computer time requirement when the set of compounds is structurally diverse. The uncertainties of a Monte Carlo sampling of conformational hyperspace are avoided, al-

TABLE 7
NUMBER OF CONFORMATIONS CONTAINING THE ACTIVE SITE GEOMETRY

Inhibitor no.	I	II
	Published [1] site geometry	Alternate site geometry
01	29	2
02	10	3
03	23	1
04	12	1
05	18	2
06	353	1
07	66	1
08	981	2
09	1002	2
10	20	2
11	36	2
12	4006	2
13	10	1
14	142	2
15	189	2
16	487	2
17	7141	2
18	4	1
19	10	2
20	3	1
21	5940	3
22	137	2
23	5679	3
24	6033	2
25	565	2
26	29	2
27	258	2
28	297	3

TABLE 8
DISTANCE CONSTRAINTS APPLIED IN CONSTRAINED SEARCH

Published [1] site geometry					
8.5000	4.7500	3.7500	5.2500	4.0000	(lower bound)
8.7500	5.2500	4.0000	5.5000	4.2500	(upper bound)
Alternate site geometry					
5.7500	2.7500	3.7500	3.2500	4.0000	(lower bound)
6.2500	3.0000	4.0000	3.7500	4.2500	(upper bound)

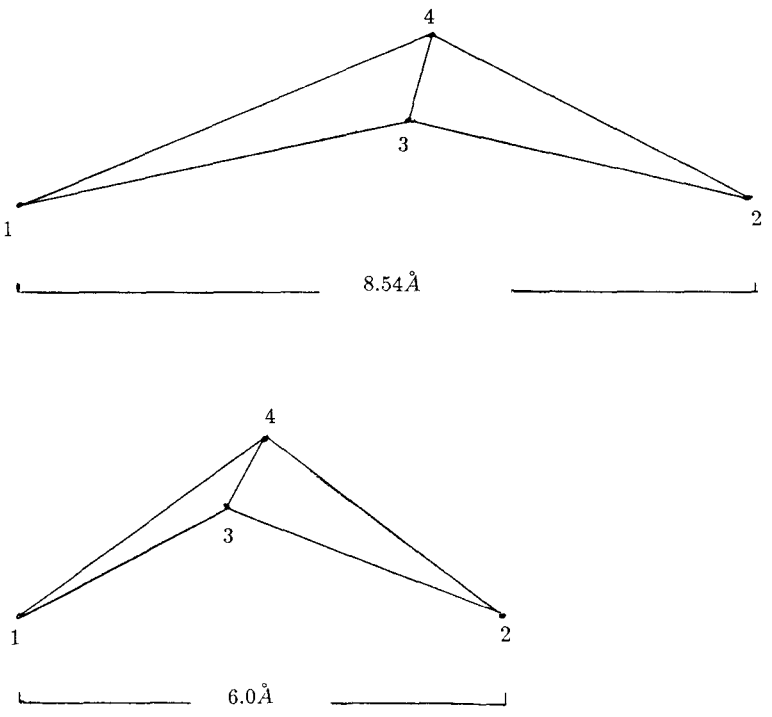

Fig. 6. Alternative active site geometries.

though some potentially important geometries may be missed if the scan parameter is too large. In our experiments with the ACE series, we were able to iteratively explore conformation space systematically varying the sampling parameters and sets of conformations to be evaluated. The efficiency of constrained search and the reordering of the data set allow us to reproduce the results of Mayer's experiment [1] in ca. 312 s. This represents a performance improvement of more than three orders of magnitude. Use of a faster processor provides the means for obtaining those same results in ca. 77 s and effectively provides an interactive environment for pharmacophore and active-site hypothesis testing.

ACKNOWLEDGEMENTS

Constrained search is implemented in standard Fortran-77 and is operational on DEC μVaxII, μVaxIII, SUN 4/280, and Convex-C1 systems in our laboratory. Input and output files are compatible with SYBYL 5.1, licensed from Tripos Associates, Inc. (St. Louis, MO). We wish to thank Dr. Dorica Mayer for providing the data set containing the ACE inhibitors and Dr. Chris Naylor for assistance in performing the systematic search experiment as well as Dr. Denise Beusen for helpful discussion of the manuscript.

This work was supported by a grant from the National Institute of General Medical Sciences (GM 24483).

REFERENCES

1 Mayer, D., Naylor, C.B., Motoc, I. and Marshall, G.R., J. Comput.-Aided Mol. Design, 1 (1987) 3–16.
2 Marshall, G.R., Annu. Rev. Pharmacol. Toxicol., 27 (1987) 193–213.
3 Sheridan, R.P. and Venkataraghavan, R., J. Comput.-Aided Mol. Design, 1 (1987) 243–256.
4 Marshall, G.R., Barry, C.D., Bosshard, H.E., Dammkoehler, R.A. and Dunn, D.A., In Olson, E.C. and Christoffersen, R.E. (Eds.) Computer-Assisted Drug Design (ACS Symposium Series, No. 112), American Chemical Society, Washington, DC, 1979, pp. 205–226.
5 Marshall, G.R. and Cramer, III, R.D., Trends Pharmacol. Sci., 9 (1988) 285–289.
6 Motoc, I., Dammkoehler, R.A. and Marshall, G.R., In Trinajstic, N. (Ed.) Mathematics and Computational Concepts in Chemistry, Ellis Horwood, Chichester, 1986, pp. 222–251.
7 Motoc, I., Dammkoehler, R.A., Mayer, D. and Labanowski, J., Quant. Struct.-Act. Relat., 5 (1986) 99–105.
8 Sheridan, R.P., Ramaswamy, N., Dixson, J.S. and Venkataraghavan, R., J. Med. Chem., 29 (1986) 899–906.
9 SYBYL, Molecular Modeling System, Tripos Associates, Inc. St. Louis, MO.
10 Brucceleri, R.E. and Karplus, M., Macromolecules, 18 (1985) 2767–2773.
11 Go, N. and Scheraga, H. A., Macromolecules, 3 (1970) 178–187.
12 Winston, P.H., Artificial Intelligence, 2nd edition, Addison-Wesley, Reading, 1984, p. 82.
13 Schulman, J.M., Sabio, M.L. and Disch, R.L., J. Med. Chem., 26 (1983) 817–823.
14 Moult, J. and James, M.N.G., Proteins: Structure, Function and Genetics, 1 (1986) 146–163.
15 Tonge, A.P., Murray-Rust, P., Gibbons, W.A. and McLachlan, L.K., J. Comput. Chem., 9 (1988) 522–538.
16 Havel, T. and Wuthrich, K., Bull. Math. Biol., 46 (1984) 673–698.