Computational Geometry: Theory and Applications, Vol. 10, No. 4, p. 263–272, 1998

RAPID: Randomized Pharmacophore Identification for Drug Design

P. W. Finn [b] L. E. Kavraki [c] J.-C. Latombe [d] R. Motwani [d] C. Shelton [d]
S. Venkatasubramanian [d] A. Yao [a]

[a] *Department of Computer Science, Princeton University, USA*

[b] *Pfizer Central Research, Sandwich, U.K.*

[c] *Department of Computer Science, Rice University, USA*

[d] *Department of Computer Science, Stanford University, USA*

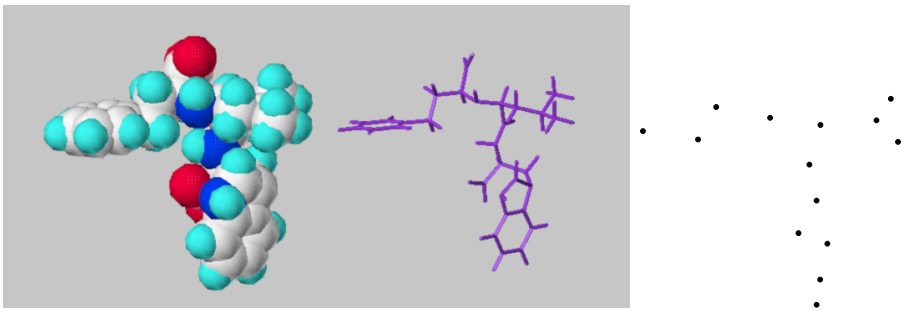
Abstract

This paper describes a randomized approach for finding invariants in a set of flexible and chemically distinct ligands (drug molecules) that underlies an integrated software system called RAPID currently under development. An invariant is a collection of features embedded in $\Re^3$ which is present in one or more of the possible low-energy conformations of each ligand. Such invariants are called pharmacophores and contain the parts of the ligand that are primarily responsible for its binding with a receptor. The identification of pharmacophores is crucial in drug design since frequently the structure of targeted receptor is unknown but a number of molecules that interact with it have been discovered by experiments. In these cases the pharmacophore is used as a template for building more effective drugs. It is expected that our techniques and results will prove useful in other applications such as molecular database screening and comparative molecular field analysis.

1 Introduction

Computational chemists working in the area of structure-based drug design consider both chemical and geometric properties of the interacting molecules when developing new pharmaceutical drugs [4]. The underlying assumption is that drug activity, or pharmacophoric activity, is obtained through the molecular recognition and binding of one molecule (ligand) to a pocket of another, usually larger, molecule (receptor). This assumption is supported by experimental results showing molecules with geometric and chemical complementarity in their binding conformations [6].

When the three-dimensional structure of the receptor is known, docking methods [4] exploit both the geometric and the chemical information available. However, the geometric structures of relatively few molecules have been obtained via X-ray crystallography or NMR techniques. In an effort to develop pharmaceutical drugs for receptors whose structure is unknown, chemists start with



(a) (b) (c)

Fig. 1. (a) space-filling, (b) stick, and (c) set-of-features models of 1TMN

a collection of ligands that have been experimentally discovered to interact with the considered receptor [5,19]. By examining the chemical properties and the possible shapes of these ligands, they try to identify a set of features embedded in $\Re^3$ that is contained in some active conformation of each (or most) of the ligands. This is called the *pharmacophore* and it is considered responsible for the observed drug activity. The features of the pharmacophore interact with features of the receptor, while the rest of the ligand acts as a scaffold. Once a pharmacophore has been isolated, it can be used to further improve the activity of a pharmaceutical drug [4].

We consider the following problem: Given a set of ligands that interact with the same receptor, find geometric invariants of these ligands, i.e., a set of features embedded in $\Re^3$ that is present in one or more valid conformations of each of the ligands. We refer to this problem as the pharmacophore identification problem. Its solution requires dealing efficiently with large amounts of spatial data and shape information. Ligand molecules are very flexible and can assume many distinct potentially valid conformations. A valid conformation is a rigid spatial realization of the atoms of a molecule whose energy is below a predefined threshold [6]. Besides providing templates for drug design, geometric invariant identification is useful in formulating database queries for retrieving functionally equivalent, but structurally novel, molecules from molecular databases [4] and in suggesting alignments of molecules for input to CoMFA (Comparative Molecular Field Analysis) and other 3D QSAR (Quantitative Structure-Activity Relationship) methods [5].

In this paper we describe our efforts to prototype an integrated software system, called RAPID (RAndomized Pharmacophore Identification for Drug design) for addressing the pharmacophore identification problem. We present briefly the overall structure of RAPID and outline related work in Section 2. The two main modules of RAPID, conformational search and identification of invariants, are described in Sections 3 and 4. In Section 5 we report preliminary experimental results and in Section 6, we conclude with a discussion of some open questions that merit further consideration.

2 Overview of RAPID and Related Work

RAPID tries to identify geometric invariants among a collection of small ligands like the molecule shown in Figure 1. This molecule is called 1TMN and it is an inhibitor of thermolysin. Figure 1(a) shows the space filling model of 1TMN, by drawing a Van der Walls sphere [4] around each atom center. Figure 1(b) shows the corresponding stick model in which only chemical bonds are drawn.

The degrees of freedom of ligands include bond lengths, bond angles (angles between two consecutive bonds), and dihedral or torsional angles (angles formed by the first and third of three consecutive bonds, viewed along the axis of the second bond). In practice, only the torsional degrees of freedom are considered since these are the ones that exhibit large variations in their values. Figure 1(c) shows the conformation of 1TMN of Figure 1(a) as a set of points in $\Re^3$. These points may represent atom centers or groups of atoms aggregated to one point endowed with a feature common to all these atoms (e.g., a rigid benzene ring) [19]. We assume that once a conformation is given, one can automatically transform it to a unique collection of points.

In RAPID, the identification of geometric invariants in a collection of flexible ligands denoted by $M = \{M_1, M_2, \ldots, M_N\}$ is treated as a two-stage process addressing the two following problems:

Problem 1 (Conformational Search) *Given a collection of ligands $M = \{M_1, M_2, \ldots, M_N\}$, the degrees of freedom for each of them, and an energy function E, find for each M_i, a set of conformations $\mathcal{C}(M_i) = \{C_{i1}, C_{i2}, \ldots, C_{ik_i}\}$, such that $E(C_{ij}) \leq$ THRESHOLD and $d(C_{ij}, C_{il}) >$ TOLERANCE for $l \neq j$ and $1 \leq j, l \leq k_i$, where THRESHOLD and TOLERANCE are pre-specified values and $d(\cdot, \cdot)$ is a distance function.*

Problem 2 (Invariant Identification) *Given a collection of ligands $M = \{M_1, M_2, \ldots, M_N\}$ where each M_i has a set of conformations $\mathcal{C}(M_i) = \{C_{i1}, C_{i2}, \ldots, C_{ik_i}\}$, determine a set of labeled points S in $\Re^3$ with the property that for all $i \in \{1, \ldots, N\}$, there exists some $C_{ij} \in \mathcal{C}(M_i)$ such that S is congruent to some subset of C_{ij}. A solution S, if it exists, is called an* invariant *of M.*

In practice, the input may contain ligands that do not contain the pharmacophore. This requires us to consider a relaxation of Problem 2 above, where a geometric invariant need only be present in conformations of some K of the N molecules. Although at this stage the two modules of RAPID work independently, we plan to support their interaction as the system develops. A third module of RAPID, currently under development, involves the computation of molecular surfaces [9,13].

Related Work We offer below a brief overview of related work. The interested reader can find an extensive survey in [10,16]. As far as conformational search is concerned, both systematic and randomized techniques are being investigated [18]. Randomized methods obtain conformations by applying random increments to the torsional DOF of the molecule starting from a user-specified initial conformation [11] or from a previously found low-energy conformation [8]. Recent articles, which attempt to compare different methods, emphasize the superior quality of the results obtained with randomized techniques [11].

Invariant identification is related to the well-studied problem in geometric optimization of finding common point sets [1,17]. Determining the congruence of two point-sets in $\Re^3$ is tractable [1] in the absence of complications such as noise. However, invariant identification is more closely related to the problem of identifying the *largest common point set* (LCP). Unfortunately, the LCP problem turns out to exceedingly difficult; in fact, even for m collections of n points on the real line, the LCP cannot be approximated to within an n^ϵ factor unless $P = NP$, and only weak positive results are known [17]. In computational chemistry, the most popular algorithms for invariant identification are based on clique-detection. For instance, DISCO [19] initially considers a pair of conformations belonging to different molecules and constructs a graph whose cliques correspond to

candidate pharmacophores. Although maximum clique detection is NP-hard, the algorithm seems to work well in practice [19]. The approach can be generalized to n conformations by choosing a reference conformation and comparing it with the other $n-1$ conformations, but it can lead to a combinatorial explosion in the operations performed [5]. Other techniques for pharmacophore identification include expansion of small invariants, hashing techniques, and genetic algorithms (see [16] for a survey).

3 Conformational Search

Our algorithm proceeds as follows. Initially a large number of conformations are generated at random. In contrast with previous randomized search methods, we obtain a random conformation by selecting each degree of freedom from its allowed range according to a user-specified distribution. This distribution is frequently the uniform distribution. However, if some a priori information is available about the preferred values of a particular degree of freedom, then the corresponding values are selected according to a distribution that reflects the a priori information (e.g., Gaussian). An efficient minimizer [7] is then used to obtain conformations at local energy minima. Minimization is the most time-consuming step, so we have carefully optimized this procedure.

To obtain a representative set of conformations from our sample, we partition it into sets that reflect geometric similarity as captured by the distance measure DRMS. We define DRMS(C_i,C_j) as the square root of the mean of the squared distances of the corresponding atoms of C_i and C_j, after C_i is transformed to C_j. This transformation is computed using a basis of three predefined atoms a_1, a_2, and a_3 [3]. The clustering algorithm used is described in Gonzalez [12]. It is an approximation algorithm that runs in time $O(nk)$, where n are the conformations to be clustered and k is the number of clusters, and guarantees a solution within twice the optimal value. The centers of the clusters are returned as representatives of the possible conformations of the molecule.

Our experience with randomized techniques for searching high-dimensional spaces has shown that randomized exploration is superior to systematic exploration when the shape of the underlying space is irregular [15]. The same observation holds for conformational search: a systematic procedure has a higher chance of missing the irregularly shaped basins of attraction of the energy landscape of the molecule (see also [11]). This has been our main motivation for the development of the randomized conformational search procedure described above.

4 Identification of Invariants

The set of cluster centers, denoted by $\mathcal{C}(M) = \mathcal{C}(M_1) \cup \ldots \cup \mathcal{C}(M_N)$ is the input for the invariant identification module. Each conformation in $\mathcal{C}(M)$ is now represented as a set of labeled points in $\Re^3$ (see Section 2). We wish to determine a structure S that is congruent to a substructure of some conformation in every molecule. The congruence relation is with respect to 3-D rotations and translations that ensure equality of labels. Our formulation of the invariant identification problem assumes noise-free data, specifically that all point positions are known exactly. In practice, atom positions are fuzzy and it may not be possible to align them exactly. Therefore, we adopt the

convention that two points p_1 and p_2 are said to *match* when $|p_1 - p_2| \leq \epsilon$, where ϵ is the *point location error*. Similarly, two triangles are said to be congruent if each point in the first triangle is within distance ϵ of the corresponding point in the second.

As mentioned before, the invariant identification problem is a variant of the *largest common point set problem (LCP)* in d dimensions: Given s point sets $P_1, P_2, \ldots, P_s$ in $\Re^3$, determine the point set of maximum cardinality congruent to some subset of each point set. For convenience, we assume that each point set P_i has cardinality exactly n. For arbitrary s and d, LCP is hard to approximate within a factor of n^ϵ, for some $\epsilon > 0$. In the sequel, we consider the following variant of LCP, called LCP-α: determine a point set S of size $|S| \geq \alpha n$ congruent to some subset of each $P_i, 1 \leq i \leq s$. The motivation for focusing on this subproblem is that it more accurately captures our primary application, where pharmacophores are desired to have a certain minimum size.

4.1 Phase 1: Pairwise Matching

In this section, we focus on the invariant identification problem for two point sets, denoted by MATCH. This problem has been studied extensively in the literature [20]. For general α, the best known algorithms were obtained in [1]. These have a worst-case running time of $O(n^{4.6})$ for unknown α, and $O(n^{2.6}/\alpha^2)$ (randomized) when α is known for 3D. In 2D, the corresponding bounds obtained are $O(n^{3.2})$ and $O(n^{2.2}/\alpha)$. However, these bounds apply only to the noise-free model of point sets. The noisy version of the problem was considered in [2] yielding an $O(n^8)$ algorithm in 2D. (Refer to [14] for recent results in the noisy model.)

We now describe two random-sampling schemes for solving LCP-α on noisy data. Our analysis (presented in [10]) assumes that the data is exact. We use the notation $g(n) = \tilde{O}(f(n))$, where f and g are functions, to indicate that $g(n) = O(f(n) \log n)$. Also note that in three dimensions, a unique transformation T (upto reflection) between two point sets P_1 and P_2 is determined by matching three points p, q, r in P_1 with three points s, t, u in P_2.

BASIC-SAMPLE: For some constant c, perform $(c \log n)/\alpha^3$ iterations of the following process: sample a triplet of points $< p_1, p_2, p_3 >$ randomly from P_1; determine three points in P_2 congruent to this set; compute the resulting induced transformation and determine the number of points in P_1 matching corresponding points in P_2; and, if this number exceeds αn, declare SUCCESS.

Theorem 1 *Given a common subset S of size $|S| \geq \alpha n$, the probability that BASIC-SAMPLE fails to declare SUCCESS is $O(1/n)$.*

Theorem 2 *BASIC-SAMPLE runs in time $\tilde{O}(n^{2.8}/\alpha^3)$ using space $O(n^2)$.*

Run-time profiling revealed that BASIC-SAMPLE examines many spurious triples, i.e. tuples that do not yield a large invariant. We propose the following modification of the random sampling procedure to handle this problem.

PARTITION-SAMPLE: For some constant c, perform $c \log n$ iterations of the following process: randomly select two subsets A and B of size $1/\alpha$ from P_1; also select a subset C of size $1/\alpha$ from P_2; store all distances $d(p, q)$, for all $p \in C$ and $q \in P_2 - C$, in a hash table; for every triangle (a, b, p) with $a \in A, b \in B$, and $p \in P1 - (A \cup B)$, probe for $d(p, a)$ and $d(p, b)$ in the hash table to

determine all matching triplets (c, p_1, p_2) with $c \in C$ and $p_1, p_2 \in P_2 - C$; finally, as before, if the resulting transformation induces a match of more than αn points, declare SUCCESS.

Theorem 3 *Given a common subset S of size $|S| \geq \alpha n$, the probability that PARTITION-SAMPLE fails to declare SUCCESS is $O(1/n)$.*

Theorem 4 *PARTITION-SAMPLE runs in time $\tilde{O}(n^{3.4}/\alpha^3)$ using space $O(n/\alpha^2)$.*

Although the asymptotic running time of PARTITION-SAMPLE is worse than that of BASIC-SAMPLE, experiments (see Section 5) reveal that PARTITION-SAMPLE consistently outperforms BASIC-SAMPLE, generating far fewer spurious triples with an improved degree of success. Additionally, experimental results suggest that both predicted running times are overly pessimistic.

There are two issues which deserve further discussion. The first concerns the elimination of redundant solutions, that is solutions satisfying a containment relationship with respect to each other. Given invariants S_1 and S_2, we need to check if $S_1 \subseteq S_2$ by invoking MATCH on these two sets with $\alpha = 1$. The second issue concerns the inaccuracies present when computing the transformation that overlaps 3 points in the two point-sets. A source of problems is the fact that our point locations are noisy and this transformation may not be unique. We employ a number of heuristics to compensate for this. For example, we determine a *seed* transformation T and then sample three random pairs from the set of correspondences that T induces and use these pairs to construct a new transformation. Clearly, in a perfect world, we will obtain T again. However, given the inaccuracies in point location, it turns out that some choices of triplets may yield more correspondences than before.

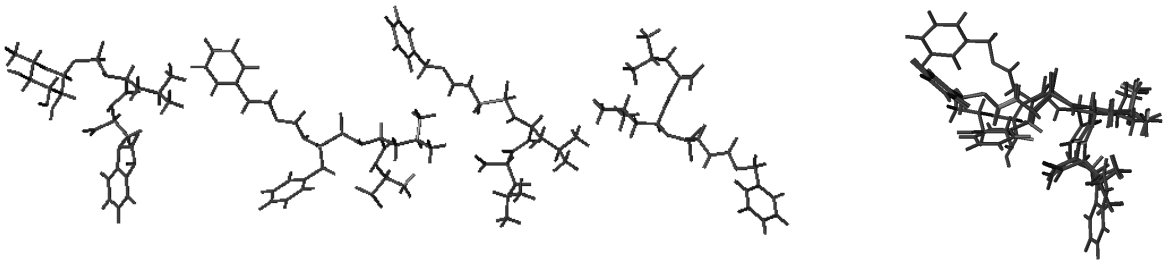
4.2 Phase 2: Multiple Matching

Candidate solutions obtained from Phase 1 are tested against the remaining molecules to determine the invariant. Each MATCH call operates on two conformations. Since each molecule is represented by many conformations, we extend MATCH to two molecules by doing all pair-wise matches between the sets of conformations. Note that comparing a candidate solution against a new conformation may result in $0, 1$, or many solutions, since the solution may decompose into smaller pieces on comparison.

There are various strategies one could use to process multiple molecules. A simple strategy that we implement performs a linear merge. We take each solution and compare it with the next molecule. We do this for all current solutions, concatenate and prune the results, and repeat with a new molecule. In addition to this, we may wish to find an invariant that does not exist in *all* the molecules, but in some fixed number of them. We use a *marking* scheme (described in [10]) to keep track of the number of times an invariant fails to match against a molecule, and reject those invariants which exceed the maximum allowed number of failures.

5 Experimental Results

This section reports experimental results for the algorithms described above. All reported timings are on an SGI Indigo2 with a 175 MHz MIPS R8000 processor and 384MB RAM. Code was written



(a) (b)

Fig. 2. (a) 1TLP, 4TMN, 5TMN, and 6TMN are inhibitors of thermolysin, (b) the molecules overlaped in their active conformation

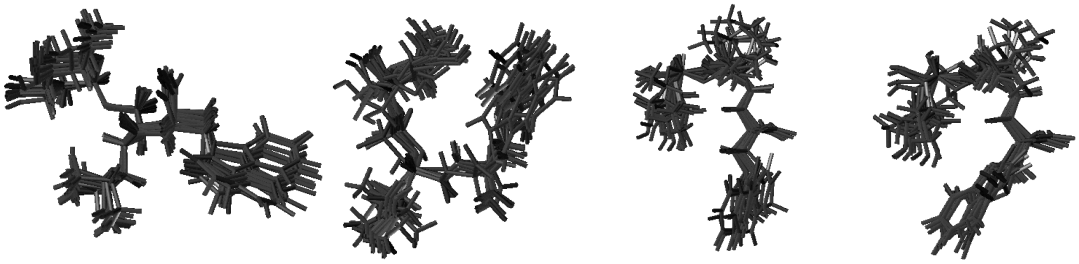


Fig. 3. Different clusters of 1TLP

in C/C++, and compiled using SGI CC. In Figure 2, we show four different inhibitors of the protease thermolysin. These molecules fit into the same cavity of thermolysin and by their presence inhibit the activity associated with that cavity. This example was chosen because all the inhibitors have been crystallized with thermolysin and their active conformations are known and recorded in the PDB database [4]. Note that 1TLP has 69 atoms and 10 torsional degrees of freedom, 4TMN has 68 atoms and 15 degrees of freedom, 5TMN has 64 atoms and 13 degrees of freedom, and 6TMN has 63 atoms and 12 degrees of freedom.

Conformational Search. Each of the molecules in Figure 2(a) was run through our conformational search software. A cutoff value of 20 Kcals/mol was used for the energy of the valid conformations. The clustering algorithm terminated when the average distance of each of the conformations from the center of its assigned cluster dropped below a THRESHOLD of 1.2 $\AA$. It took 9.4h, 34.1h, 10.4h, and 9.2h to create 10,000 conformations of 1TLP, 4TMN, 5TMN, 6TMN correspondingly. The running times for clustering were 5.2m, 18.2m, 15.2m, and 14.1m, producing 128, 253, 241, and 219 clusters. A few clusters of 1TLP are shown in Figure 3. The conformations in a single cluster are overlaid to illustrate that they are close to each other and this justifies using the center of a cluster as its representative. As mentioned in Section 3, an important problem in conformational search is to decide how many conformations to produce. At this stage, this number is determined experimentally: we stop producing new conformations when these do not increase the overall number of clusters significantly.

Identification of Invariants. An experimental comparison of BASIC-SAMPLE and PARTITION-SAMPLE reveals that PARTITION-SAMPLE works significantly better than BASIC-SAMPLE, and a second suite of tests provides some explanation for this. For our input molecules, the "solu-

	Number of Conformations	Running Time	Histogram of Solution Sizes										
			4	5	6	7	8	9	10	11	12	13	total
BASIC SAMPLE	1	7.5		2		1							3
	11	789.35	44	20	10	5	2	1			1		82
	21	13909.64	224	84	38	17	10	18	4	1			395
PARTITION SAMPLE	1	2.8		2						1			3
	11	182.16	15	6	10	6	5	1					43
	21	3530.19	39	35	30	22	20	6	6	2	1	1	162

Fig. 4. Comparison of BASIC-SAMPLE and PARTITION-SAMPLE

tion" consists of the overlapping portions of the molecules when aligned as shown in Figure 2(b). This is the lower right handside T-shaped portion of this diagram. The entire invariant consists of roughly 7 atoms and an additional 7 atoms of "scaffolding," or connecting atoms with no pharmacophore functionality. In all cases, we required the invariant to be present in all of the four molecules.

We run our search procedure on sets consisting of 1, 11, and 21 conformations including the active conformation. The search values for ϵ, δ, and α were set experimentally to 1.3, 0.5, and 0.3. the corresponding prune values are 3.5, 1, and 1. We present in Figure 4 the results of this test. In all cases, the quality of solutions (in terms of the largest solution found) is comparable, and PARTITION-SAMPLE consistently runs significantly faster than BASIC-SAMPLE. When the number of conformations increases, more invariants are produced because some of the added conformations have additional "scaffolding" which also could be matched.

As observed earlier, a formal analysis of the algorithms does not explain the marked difference in performance. Our second suite of experiments attempts to investigate this discrepancy. For this set of tests, we used the four molecules from the above examples, but with only one conformation each (so as to maximize the influence of the basic sampling algorithm). We vary α between 10 and 35 (in intervals of 5) and fix $(\epsilon, \delta) = (1.3, 0.5)$. The parameters that control the running time of the sampling algorithm are (a) the number of candidate triangle-triangle pairs examined (Δ_c), (b) the number of valid transformations produced (Δ_i), and (c) the number of valid transformations that yield solutions above the minimum required size (Δ_m).

An experimental evaluation (shown in [10]), reveals a strong correlation between the running time of the algorithm and Δ_c, and a much weaker correlation between the running time and Δ_m. Our most intersting experiment is the plot of Δ_c against α for the two algorithms in Figure 5. Notice that BASIC-SAMPLE examines a significantly larger number of such pairs than PARTITION-SAMPLE (Δ_c for PARTITION-SAMPLE is scaled by 10 on the graph for ease of reading). However, most of the work that BASIC-SAMPLE performs is wasted effort, as seen in Figure 5, which explains the experimentally good performance of PARTITION-SAMPLE.

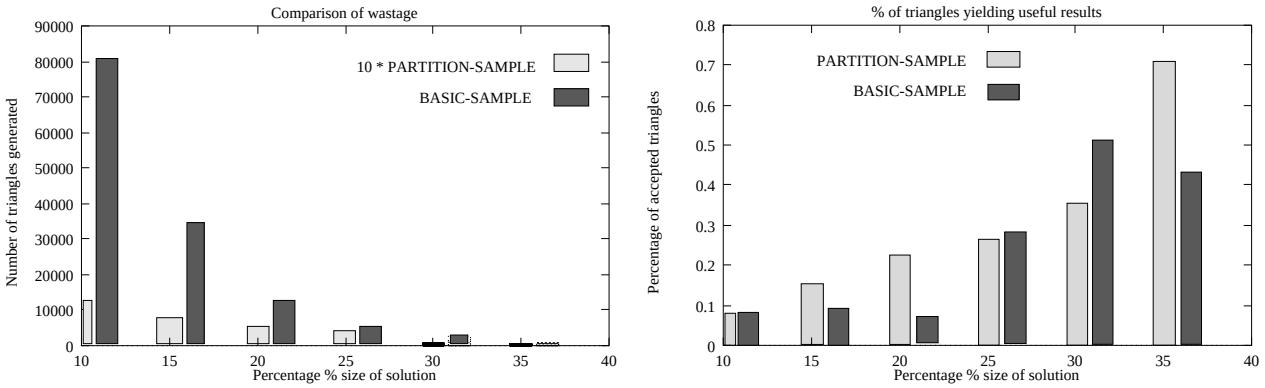


Fig. 5. Number of Matches vs α

6 Discussion

Our goal is to optimize the modules of RAPID to perform experiments that involve 5-20 ligands and a large number of conformations per ligand. To improve conformational search we are investigating better energy minimization algorithms and better incremental clustering techniques. For invariant indentification we are working towards improving our MATCH procedure and understanding the combinatorics of matching noisy point sets in 2D and 3D.

Acknowledgments This research is supported by a grant from Pfizer Central Research. L. Kavraki is partially supported by startup funds from Rice University. R. Motwani is partially supported by an Alfred P. Sloan Research Fellowship, an IBM Faculty Partnership Award, an ARO MURI Grant DAAH04-96-1-0007, and NSF Young Investigator Award CCR-9357849, with matching funds from IBM, Mitsubishi, Schlumberger Foundation, Shell Foundation, and Xerox Corporation. The authors would like to thank Piotr Indyk and Dan Halperin for many helpful discussions. A preliminary version of this paper was published as part of a conference proceedings [10].

References

[1] T. Akutsu, H. Tamaki, and T. Tokuyama. Distribution of distances and triangles in a point set and algorithms for computing the largest common point set. In *Proceeding of the ACM Symposium on Computational Geometry*, pages 314–322, Nice, France, 1997.

[2] H. Alt, K. Mehlhorn, H. Wagener, and E. Welzl. Congruence, similarity and symmetries of geometric objects. *Discrete Comput. Geom.*, 3:237–256, 1988.

[3] K. S. Arun, T. Huang, and S. D. Blostein. Least-squares fitting of two 3-d point sets. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(5):698–700, 1987.

[4] L. Balbes, S. Mascarella, and D. Boyd. A perspective of modern methods in computer-aided drug design. In K. Lipkowitz and D. B. Boyd, editors, *Reviews in Computational Chemistry*, volume 5, pages 337–370. VCH Publishers, 1994.

[5] D. Barnum, J. Greene, A. Smellie, and P. Sprague. Identification of common functional components among molecules. *Journal of Chemical Information and Computer Science*, 36:653–571, 1997.

[6] D. B. Boyd. Aspects of molecular modeling. In K. Lipkowitz and D. B. Boyd, editors, *Reviews in Computational Chemistry*, volume 1, pages 321–351. VCH Publishers, 1990.

[7] R. Brent. *Algorithms for finding zeros and extrema of functions without calculating derivatives.* PhD thesis, Stanford University, 1971.

[8] G. Chang, W. Guida, and W. Still. An internal coordinate monte-carlo method for searching conformational space. *Journal of the American Chemical Society*, 111:4379–4386, 1989.

[9] P. Finn, D. Halperin, L. Kavraki, J.-C. Latombe, R. Motwani, C. Shelton, and S. Venkatasubramanian. Geometric manipulation of flexible ligands. In M. Lin and D. Manocha, editors, *LNCS Series - 1996 ACM Workshop on Applied Computational Geometry*, pages 67–78. Springer-Verlag, 1996.

[10] P. Finn, L. Kavraki, J.-C. Latombe, R. Motwani, C. Shelton, S. Venkatasubramanian, and A. Yao. Rapid: Randomized pharmacophore identification. In *Proceedings of the International Symposium on Computational Geometry*, pages 324–333, Nice, France, 1997.

[11] A. Ghose, J. Kowalczyk, M. Peterson, and A. Treasurywala. Conformational searching methods for small molecules: I. study of the sybyl search method. *Journal of Computational Chemistry*, 14(9):1050–1065, 1993.

[12] T. F. Gonzalez. Clustering to minimize the maximum intercluster distance. *Theoretical Computer Science*, 38:293–306, 1985.

[13] D. Halperin and C. Shelton. A perturbation scheme for spherical arrangements with application to molecular modeling. In *ACM Conference on Computational Geometry*, pages 183–192, Nice, France, 1997.

[14] P. Indyk, R. Motwani, and S. Venkatasubramanian. Geometric matching under noise: Combinatorial bounds and algorithms. Manuscript.

[15] L. Kavraki. *Random Networks in Configuration Space for Fast Path Planning.* PhD thesis, Stanford University, 1995.

[16] L. E. Kavraki. Geometry and the discovery of new ligands. In J.-P. Laumond and M. Overmars, editors, *Algorithms for Robotic Motion and Manipulation*, pages 435–448. A. K. Peters, 1997.

[17] S. Khanna, R. Motwani, and F. F. Yao. Approximation algorithms for the largest common set. Technical Report STAN-CS-95-1545, Stanford University, 1995.

[18] A. Leach. A survey of methods for searching the conformational space of small and medium sized molecules. In K. Lipkowitz and D. Boyd, editors, *Reviews in Computational Chemistry*, volume 2, pages 1–47. VCH Publishers, 1991.

[19] Y. Martin, M. Bures, E. Danaher, J. DeLazzer, and I. Lico. A fast new approach to pharmacophore mapping and its application to dopaminergic and benzodiazepine agonists. In *Journal of Computer-Aided Molecular Design*, volume 7, pages 83–102, 1993.

[20] P. Raghavan and S. Irani. Combinatorial and experimental results for randomized point algorithms. In *Proc. 12th. ACM. SCG*, pages 68–77, 1996.