From Computer Vision to Protein Structure and Association

Haim J. Wolfson [a] and Ruth Nussinov [b]

[a] *Computer Science Department, School of Mathematical Sciences, Tel Aviv University, Tel Aviv 69978, Israel. Telefax : +972-3-640-6476, e-mail: wolfson@math.tau.ac.il.*

[b] *Sackler Inst. of Molecular Medicine, Faculty of Medicine, Tel Aviv University, Tel Aviv 69978, Israel. Fax : +972-3-641-4245, e-mail : ruthnu@post.tau.ac.il. and Laboratory of Experimental and Computational Biology, SAIC, NCI-FCRDC, Bldg 469, rm 151, Frederick, MD 21702, USA. Tel.: (301) 846 - 5579, Fax : (301) 846 - 5598, e-mail: ruthn@fcrfv1.ncifcrf.gov.*

1 Introduction

In this chapter we discuss the application of novel pattern matching methods to the biomolecular recognition and docking problem.

The problem of receptor-ligand recognition and binding is encountered in a very large number of biological processes. For example, it is a prerequisite in cell-cell recognition and in enzyme catalysis and inhibition. The type of molecules involved is very diverse as well: proteins, nucleic acids (DNA and RNA), carbohydrates and lipids. Whereas the receptor molecules are usually large (often proteins), the ligands can be fairly large (e.g. proteins) or small, such as drugs or cofactors. Efficient docking algorithms of drug molecules to target receptor proteins is one of the major ingredients in a computer assisted drug design scheme.

The docking task usually requires shape complementarity of the participating molecules. This complementarity reflects the effect of the van der Waals interactions, which are very sharp at short distances. Thus, there are two simultaneous geometric criteria for a good docking solution. First, the docked molecules should not penetrate each other, and, second, one expects a fairly large interface of tightly matched molecular surfaces. Our task is to detect appropriate conformations of both molecules, so that there is a three dimensional rotation and translation of one of them which brings it into close fit with the other molecule according to the above mentioned criteria.

Preprint submitted to Elsevier Science *15 December 1999*

Enumerating over all the degrees of freedom of the participating molecules is a formidable task, unless the molecules are very small. Consequently, the first generation of automatic docking methods assumed that the molecules behave as rigid objects, thus, freezing all the degrees of freedom except for the three rotational and three translational degrees of freedom which are necessary to move one molecule into close fit with the other. This is the, so called, "key-in-lock" principle. For proteins, it has been argued [1,2] that the rigid approximation is justified by the similarity of the crystallographic structures between the bound and the unbound proteins [1] while many other cases exist where more substantial conformational changes have been observed. In the last decade, a number of docking methods have been developed which tackle the rigid docking problem with considerable success, especially, in the "bound" experiments. Kuntz et al. ([3,4]) developed the DOCK program, Connolly ([5]) docked molecular surfaces by fitting quadruples of topographical extrema, Jiang and Kim ([6]) suggested the "soft docking" approach, Katchalski-Katzir et al. ([7]) exploited the Fourier transform technique for fast enumeration of all possible 3-D translations of one molecule with regard to the other. Our group suggested two main methods for the rigid docking task. The first is the Computer Vision based Geometric Hashing method ([8,9]), which can efficiently dock ligands to moderate size receptors, without any ***a-priori*** knowledge of the binding site location. The second method ([10,11]) was especially geared towards the docking of large protein-protein complexes and is based on matching pairs of sparsely sampled extremal points of the molecular surface. These points are similar to the ones used by Connolly ([5]). We shall present these two methods in detail in the following sections.

In the recent few years, a considerable effort was done to develop techniques for the docking of conformationally flexible ligands to rigid receptors. One of the first attempts was to introduce limited flexibility into the DOCK technique ([12,13]). Lengauer's group developed the FLEXX method ([14]), which is especially geared towards docking of flexible drug molecules into previously known binding sites. Jones et al. ([15]) tackle the flexible docking task by genetic algorithms. Our group ([16,17]) developed an efficient algorithm for docking of either flexible ligands or flexible receptors possessing a small number of internal rotational bonds. Lengauer and Rarey ([18]) have compiled a very good survey of the recent docking methods.

In this chapter we focus on the rigid docking techniques that we have developed. We shall discuss only the geometric aspect of shape complementarity. Our computational approach is based on the ***Geometric Hashing*** paradigm, originally developed for object recognition tasks in Computer Vision ([19,20]). Although coming from an entirely different application field the underlying

[1] The structure of a single protein in solution is referred here as "unbound", while the structure of the same protein in the docked complex, is called "bound".

geometrical and computational principles bear a similar flavor. Let us sketch briefly the model based object recognition problem in Computer Vision. There one is presented with a data base of shapes (models) and the task is to detect if any of these shapes appear in an observed scene, even if they are partially occluded by other objects. In addition, we do not know ***a-priori*** in what locations of the scene the desired shape might appear. In case a shape does appear in the scene, one has to compute the transformation that has to be applied to its model in the data-base, so that it matches the shape's appearance in the scene. Thus, one has to match shape characteristics of the models to the shape characteristics of the scene. Let us now consider the docking task in an analogous framework. Assume that we are given a database of ligands, and a receptor. The task is to decide whether there is a transformation of a ligand, which results in sufficient shape complementarity with the receptor without penetrating it. In the general case, we do not have ***a-priori*** knowledge on the location of the binding site of the receptor. Already these two formulations alone make the analog quite clear. The ligand database is analogous to the model database. The receptor is analogous to the scene, and the partial occlusion and lack of knowledge of the desired model location of the scene are analogous to the facts that not all of the ligand's molecular surface is complementary to the receptor surface, and that we do not know in advance the location of the binding site [2]. One should bear in mind, though, that while in object recognition one has to match shape characteristics seeking for their similarity, in docking we seek shape complementarity. In this sense the analog between object recognition in Computer Vision and structural protein comparison is even more direct([21]).

We shall present two methods for the rigid docking task. In section 2 we formulate the precise mathematical task that we are going to tackle. Section 3 discusses the molecular surface representation and the shape primitives, that we are going to use in our docking algorithms. In section 4 we reformulate the molecular surface matching problem as a point+normal set matching problem. Section 5 represents the sparse feature large protein-protein docking algorithm, while section 6 presents the Geometric Hashing algorithm based docking algorithm. Finally, in section 7 we discuss further developments and future tasks.

2 Problem Formulation

In this section we describe the mathematical model of the docking problem for which our algorithms are developed. We restrict our discussion to the docking of rigid molecules. One of them, usually, the larger one is called the receptor,

[2] In case we do know this site, it can be readily incorporated in the algorithms to reduce the amount of computation.

and the other is called the ligand. Each molecule is regarded as a collection of spherical atoms. Thus, our input is a set of atomic centers together with their van-der-Walls radii.

The biomolecular recognition or docking problem is defined as follows:

Problem 1 *Given the coordinates of the atomic centers of the receptor and the ligand molecules together with their atomic radii, find those rotations and translations of the ligand, which result in a large contact area between the molecular surfaces of the ligand and the receptor, without causing intermolecular penetration.*

There are two major aspects to the docking problem - geometric and chemical. Geometrically the two molecules should interlock like two pieces of a 3-D jigsaw puzzle. From the chemical standpoint, the candidate docking configuration should be energetically favorable, so that the predicted complex would be stable. This presentation deals only with the geometric aspect of the problem which can be directly stated as a pattern recognition problem. Since detailed energy evaluation calculations are time consuming, our goal is to develop geometric matching algorithms which are fast, result in few tens or hundreds of candidate solutions (with the correct one among them), thus enabling a further chemical verification step to filter these remaining candidates, which could further be submitted to human inspection.

3 Molecular Surface Representation and Interest Feature Extraction

Since molecules dock at their surfaces, we have to define the geometric model of the molecular surface. We have also to define how do we compute complementarity of molecular surfaces, which might be of very intricate form. Our solution is to represent the surfaces by sets of discrete *interest points* with their associated normals, and then seek for transformations which best superimpose complementary sets of these interest features. In this section we discuss several methods to compute such *interest points*.

As was mentioned in the previous section, a molecule is modeled by a set of balls, centered at the atomic coordinates of the molecule. The radii of these balls are their van der Waals radii. Richards ([22]) defined the molecular surface as the surface being created by virtually rolling a probe ball over the van der Waals spheres of the molecule's atoms. Richards' definition has been implemented by Connolly in his MS-DOT program ([23]), which computes the analytic representation of the molecular surface. The probe mimics the size of a water molecule and is usually chosen to be of 1.4 angstrom radius.

This procedure fills narrow crevices and defines the surface which is actually accessible to atoms of the size of the probe ball. A probe ball can touch the van der Waal's surface only in one, two or three points resulting in only three types of molecular surface faces - convex, toroidal or concave (see [24]). A convex face is a connected patch of the atom's van der Waal's sphere touched by the probe ball (one contact point). A toroidal face is traced by the probe ball when it rolls along the groove between two atoms (two contact points). When the probe ball is suspended by three atoms, the part of this ball surface mounting the space among the three atoms becomes a concave face (three contact points).

Connolly's program itself returns a discrete set of points with associated normals which samples the molecular surface at the density requested by the user. One could use this set of points as the ***interest features*** of the docking algorithm. However, a coarse sampling may give an unreliable representation of the molecular surface, while a dense sampling may result in thousands of points to represent the surface of a relatively small molecule, making the computational burden not feasible.

In our research we have used two alternative ***interest feature*** definitions, which we shortly describe below.

3.1 Caps, Pits, and Belts.

In [25] we extract ***interest points*** and their associated normals only at key locations of the surface, covering the crests of the exposed atoms and the seams and dents between them. Specifically, we compute one such point at the center of each molecular face and label them as 'cap' if the face is convex, 'pit' if the face is concave, and 'belt' if the face is toroidal. Normals associated with these points are computed as well.

By examining the interfaces of known receptor-ligand complexes we have found hundreds of pairs of our ***interest points*** to be at excellent complementarity, in both the point location and the normal orientation The abundance and quality of these pairs provides the desired shape representation for our ***Geometric Hashing*** based matching algorithm (see [9,26] for details).

The points obtained are considerably more sparse than the original Connolly's MS representation, yet they reliably represent the overall shape of the molecular surface. However, this representation is still somewhat too dense. In particular, the distances between pairs of points should be larger than the error thresholds allowed in the matching. These error thresholds should be liberal ones, due to the inexact fit of docked molecules. Thus, it is necessary to prune the obtained surface points. We have adopted several pruning mechanisms, such as : (1) selecting points by label; (2) thinning by fusing close points; (3) eliminating points representing faces with too small or too large areas. The output of the pruning operation results in a less accurate surface representation, yet dozens of ***interest point*** pairs still remain in a typical binding site supplying enough information for the matching phase.

This molecular surface representation appeared both in rigid docking ([9]), flexible docking ([17]), and molecular surface superposition ([27]) algorithms.

3.2 Knobs and Holes

The cap, pit and belt interest features provide a reliable molecular shape representation, without requiring a dense sampling of the Connolly surface, yet for large molecules their number is still very big. If we know that the binding interface is expected to be relatively large by itself it would be enough to sample the surface at sparse interest points, which still preserve some salient features of the molecular shape. The properties of such points should be also invariant to the rotation and translation of the molecules.

In a docking algorithm developed by Connolly ([5]) he suggested to use the convexity/concavity extrema of the molecular shape as interest points, which he nicknamed as knobs and holes.

We compute these points as follows. First, the molecular surfaces of the receptor and ligand molecules are approximated by densely sampled dots, using the previously mentioned MS algorithm ([23]). Then, at each MS dot, a sphere of a fixed radius is constructed (we use $6\mathring{A}$ which is an approximation of the radius of an amino acid). A shape function at such a point is defined as the volume of the intersection of the molecule with the sphere. This shape function measures the local convexity/concavity of the surface. When the shape function value is small, the surface is convex (knob); when the value of the shape function is high, the surface is concave (hole). Using a non-extrema suppression technique we choose a relatively small set of critical knobs and holes, which are local minima and maxima of this shape function.

For each critical point we also compute the normal vector. At the end of the process, for each molecule, we have lists of knobs and holes along with their associated normal vectors. Note that the definition of these points is invariant to rotation and translation, since they are dependent only on the intrinsic shape of the molecule.

4 Interest Point Correspondence

In section 3 we have shown that the molecular surface can be represented by sets of interest points coupled with their associated normals. One of the suggested representations is reliable, representing all the major geometric characteristics of the surface, while the other representation was a sparse one, representing only the most salient geometric features of the surface. An important property of the selected points is that their existence and intrinsic location on the molecular surface is not affected by translations and rotations that the molecule may undergo. The need for these interest point based representations stems from our desire to reduce the molecular surface matching task to a point matching task. The reason for this reduction is our ability to develop relatively efficient algorithms for the point matching task which provide the solutions for the original molecular surface matching.

Let us formulate this new task for the ligand-receptor docking. Assume that both molecules are represented by sets of characteristic(interest) points with their associated normals.

Problem 2 *Given the coordinates of the molecular surface interest points+normals of the receptor and the ligand, find those rotations and translations of the ligand, which result in a large set of the ligand interest points superimposed on the receptor interest points with their normals in opposite directions. Only transformations which do not cause penetration of the ligand into the receptor are acceptable.*

One should note several important facts. First, since the molecules interlock only at the binding site, not all the interest points of either the ligand or the receptor will have a matched counterpart, even in an ideal situation. Second, since we do not assume *a-priori* knowledge of the binding site, we cannot know in advance which of the interest points might have a matching counterpart. Third, due to the inherent "inexact match" of molecular surfaces, as well as inaccuracies in the computation of interest points and normals, superimposition does not mean an exact match. It means that in a vicinity of a transformed ligand interest point (sometimes nicknamed - "error region") there is a receptor interest point. Since the computation of normal directions is even more noisy than the computation of interest point locations, quite a liberal error threshold is assigned for opposing normal directions. Fourth, it is important to understand that one cannot restrict himself to the examination of only one transformation giving the maximal superimposed set of ligand-receptor points. Due to all the fuzziness inherent in our computations, the maximal superimposed set might not be the correct one, while a somewhat smaller set, defined by a different transformation may hit the target. This can be even examined on real complexes, where quite often the "correct" solution, appearing in the

PDB does not score at the top of the list, although, it almost always scores quite high. Thus one has to examine all the transformations which suggest a "large enough" superimposition.

It is quite common to divide point set matching tasks into two subtasks - the point correspondence problem, and the corresponding point superposition problem. The correspondence problem is the tough one. Once a correspondence hypothesis(match) among the points is established, there are well known methods to find the rigid transformation which superimposes these points with minimal least squares distance among the point sets (see, e.g. [Schwartz and Sharir, 1987]).

Let us examine the point set correspondence problem more closely. Assume that we have n interest points on the receptor surface and m interest points on the ligand surface and we have to establish a correspondence among them. Without loss of generality we may assume that $m \leq n$. If there are no mutual constraints among the points, the first ligand point can match n receptor points, the second can match $n - 1$ receptor points and so far. Thus, there are $n(n-1)(n-2)...(n-m+2)(n-m+1)$ candidate solutions. This is an enormous enumeration space, e.g. for $n = 200$ and $m = 100$ the number of hypotheses is of the order of 10^{217}. Luckily enough, our problem is much more constrained, since we are dealing with rigid body motion, so when some of the points are rotated and translated, the whole ensemble of points performs the same transformation. Let us examine, what is the minimal number of point correspondences required to fully determine the transformation of the ligand vis-a-vis the receptor.

Kuntz et al. ([3]) and Connolly ([5]) have used four point set correspondences to define an unambiguous rotation and translation of the ligand. However, it is well know that to determine unambiguously a rotation and translation that superimposes one rigid body onto another one needs a minimum of three non-collinear ordered matching point-pairs. The only constraint is that the pairwise distances among the matching points should be identical. Namely, for each pair of two non-degenerate **congruent** triangles $\triangle ABC$ on the receptor and $\triangle abc$ on the ligand, one can define a unique rotation and translation. First, translate the ligand so the the point a is superimposed on A. Then, rotate it till b is superimposed on B. Finally, rotate it again around the ab axis till c touches C. This construction is unique. Now, the number of candidate solutions is at most $\boldsymbol{n(n-1)(n-2) \times m(m-1)(m-2)}$, which is the number of possible triangle correspondences. In the previous example of $\boldsymbol{n = 200}$ and $\boldsymbol{m = 100}$, the number of hypotheses is at most $\mathbf{7.6 \times 10^{12}}$. The number of congruent matching triangles is expected to be much smaller.

Still, one would like to reduce the size of the minimal matching set for two major reasons. First, it will reduce the size of the solution space, and second, it will enable recovery of solutions where we do not have a triplet of matching interest points at the interface. It is easy to see that if we exploit only point information, a pair of matching points in the ligand ab and the receptor AB are not enough to recover the transformation, since we are left with the freedom of rotation around the ab axis (see explanation above). However, in the molecular surface representations that we have seen in section 3 we extract not only points, but also their associated normals. Thus, if the length of the segment ab on the ligand surface is equal to the length of AB on the receptor surface, we can superimpose the segments, and the only remaining degree of rotational freedom is fixed by the requirement that the normal at a should be an opposite direction to the normal at A. The same requirement for the normals at the points b, B supplies an additional constraint. Now, there are at most $n(n-1) \times m(m-1)$ candidate solutions, which is about 3.4×10^8 in our previous example.

5 Large Protein-Protein Docking

5.1 The Algorithm

The technique we describe below was especially designed to handle large protein-protein complexes. In retrospective it performed very well for most of the standard protein-ligand cases, where the ligand is a small molecule, as well. Especially encouraging was its relatively good performance on "unbound" examples, which are the real-life tests of docking techniques.

Following are the major steps of this method :

(i) Computation of the molecular surface of the ligand and the receptor using Connolly's MS algorithm.
(ii) Representation of these molecular surfaces by sparse "interest point + normal" sets of knobs and holes.
(iii) Matching of pairs of ligand interest points+normals to compatible pairs on the receptor.
(iv) Clustering close transformations.
(v) Rejection of candidate transformation clusters, which cause significant penetration of the ligand into the receptor.
(vi) Scoring of the remaining transformations by contact surface area favoring (optionally) contacts of hydrophobic residues and solutions which result in spatially connected binding patches on the molecular surfaces.

For details the reader is referred to [10,11]. Here we briefly sketch some of the above mentioned steps.

Knob and hole extraction. Knobs and holes are extracted by the technique described in section 3. The criterion for a knob is that the value of the shape function is below $\frac{1}{3}\boldsymbol{V}$, where $\boldsymbol{V}$ is the volume of the intersecting ball, while a hole's shape function should be above $\frac{2}{3}\boldsymbol{V}$. Non extrema suppression on 12 nearest neighbors is applied.

Interest point matching. For each pair of critical points+normals from the ligand compute the transformation with each compatible pair from the receptor as described in section 4. Note that compatibility requires that the distances between the points be almost equal (up to an error threshold), that a knob should be paired only with a hole (and vice versa), and that the appropriate normals should be in opposite directions. Since protein-protein interactions are localized, and generally do not span distant regions of the proteins without covering close-range ones, we restrict ourselves to critical point pairs which are closer than $\boldsymbol{20\mathring{A}}$ to each other. To enable fast pruning of incompatible critical points, we assign to each pair of points on the ligand/receptor a signature, which has a symbolic component (knob/hole, hydrophobicity) and a geometric component, which includes the distance between the pair of points, the angles that are formed by the line segment connecting the points with each of the normals, and and the torsion angle between the two planes formed by this line segment and the respective normals. Only pairs with compatible signatures are considered as a match. Compatibility is defined by appropriate error thresholds (see [10,11] for the details).

Clustering of transformations. Since the same (or close by) transformation can be achieved using different matching critical-point/normal pairs, the candidate transformations are clustered to reduce their number to the effective one. This reduces the workload of the next verification and scoring phase.

Verification of penetration and scoring. For each resulting candidate transformation the ligand is transformed to fit the receptor. If the ligand atom centers penetrate inside the receptor, the transformation is discarded. Otherwise, the transformation is scored favorably for ligand surface dots, which are in close vicinity of the receptor molecular surface and penalized for surface dots penetrating into the receptor. Optional scoring techniques take into account only the dots belonging to large connected interface patches and give an additional weight to dots belonging to hydrophobic residues.

Table 1
Bound examples.

	pdb	receptor name	ligand name	res. in $\mathring{A}$
1	1cho	alpha-chymotrypsin (E)	turkey 2 ovomucoid third domain (I)	1.8
2	2hfl	IG*G1 fab fragment (LH)	lysozyme (Y)	2.5
3	2mhb	hemoglobin α chain (A)	β chain (B)	2.0
4	2sni	subtilisin novo (E)	chymotrypsin inhibitor (I)	2.1
5	3hfm	IG*G1 fab fragment (LH)	lysozyme (Y)	3.0
6	4sgb	serine proteinase (E)	potato inhibitor pci-1 (I)	2.1
7	4tpi	trypsinogen (Z)	pancreatic trypsin inhibitor (I)	2.2

5.2 Experimental Results

We have done extensive experimentation with our algorithm both for bound and for the much more difficult unbound cases. In [11] 26 examples of complex prediction are given. The data was taken from the PDB and includes mainly proteinases, immunoglobulins, protein-inhibitor and chain-chain docking examples. It is important to note that all these cases have been solved by one set of program parameters. A sample of the above mentioned examples is given in Table 1, where the PDB name of the complex, receptor and ligand are given as well as the resolution of the crystal structures.

These complexes contain relatively large ligands as can be seen from the information in Table 2, where the number of amino acids, the number of interior and exterior (molecular surface) atoms are given as well as the number of the computed knobs and holes.

Since the tool developed here is a predictive one, in order to evaluate its efficacy we present in Table 3 the best RMS distance from the 'correct' (PDB complex) solution among the highest ranking 10, 50 and 100 solution clusters respectively.

Columns 3,4,5 of the table show the lowest RMS for the solutions without the application of a criterion favoring a binding site with a small number of connected components (connectivity filter), while columns 6,7,8 present the results which have been subjected to this criterion. One can see that already among the 10 top candidate solutions in these complexes 4 out of 7 have an RMS less then 2.0 Å, while among the 100 top candidate solutions all but one have an RMS less or equal 2.0 Å. The run time of the procedure on an 66

Table 2
Size of Bound Examples.

	pdb	aa	int at	ext at	holes	knobs	aa	int at	ext at	holes	knobs
				receptor					ligand		
1	1cho	146	655	392	57	144	53	469	232	31	108
2	2hfl	424	1606	1621	220	270	129	517	483	60	93
3	2mhb	141	581	487	84	118	146	618	515	87	119
4	2sni	275	832	1105	112	155	63	306	206	30	65
5	3hfm	429	1552	1741	214	262	129	516	484	59	110
6	4sgb	227	603	706	60	97	51	247	132	17	56
7	4tpi	230	836	792	108	141	58	279	176	23	60

Table 3
Best RMS among top 10,50 and 100 clusters of the bound examples.

	pdb	fast scoring			with connectivity		
		top 10	top 50	top 100	top 10	top 50	top 100
1	1cho	0.53	0.53	0.53	0.53	0.53	0.53
2	2hfl	48.71	14.43	2.07	14.43	2.07	2.07
3	2mhb	0.69	0.69	0.69	0.69	0.69	0.69
4	2sni	1.16	1.16	1.16	1.16	1.16	1.16
5	3hfm	36.44	18.09	0.94	36.44	0.94	0.94
6	4sgb	13.28	4.50	4.50	3.59	3.59	3.59
7	4tpi	1.40	1.40	1.40	1.40	1.40	1.40

MHz Intel486 PC clone are given in Table 4.

Our success with the reconstruction of docked configurations where the molecules have been crystallized together, has encouraged us to turn to the far more difficult but realistic case, that of the 'unbound' proteins. Here the receptor and the ligand have been determined separately. When the molecules dock, their surfaces adjust to each other, so using our (and other rigid docking) techniques one should expect molecular penetrations. This requires relaxation of

Table 4
CPU time on a 66MHz Intel486 clone.

			cpu time in minutes	
	pdb	match	scoring with connectivity	scoring without connectivity
1	1cho	3.3	40.6	12.6
2	2hfl	20.8	283.6	114.2
3	2mhb	14.3	227.1	97.2
4	2sni	4.4	82.3	26.8
5	3hfm	21.3	332.2	137.4
6	4sgb	1.8	26.5	7.7
7	4tpi	4.1	54.2	17.3

the rejection criteria in the penetration test, resulting in a potentially large increase in the number of docked configurations. To test our method, we have docked 19 examples of unbound proteins, in which the 'correct' complex is also known. This gave us a possibility to compare our results with the real-life solution. Table 5 lists the names of the docked proteins as well as their crystal resolution.

The examples were chosen as large proteins (see Table 6), that have been crystallized both in complex and as separate molecules (unbound proteins).

The lowest RMS solutions among the top ranking 10, 50 and 100 solution clusters are given in Table 7. When looking for the top 100 clusters notice that 3 solutions have an RMS less than 2 Å, and 4 additional solutions are with an RMS below 6.5 Å, however, among the 50 top ranking solutions only one is below 2 Å. This is an unparalleled performance for the 'unbound' case which was by now almost untackled by a purely geometric approach. Introduction of potential energy calculations on our top ranking examples should lead to a significant improvement in the fit. Since energy calculations are very time consuming, the reduction of the transformation space to less than 100 candidates gives a solid starting point for such an approach.

Table 5
Unbound Examples.

	pdb	receptor name	res. in Å	ligand name	res. in Å
1	1hfm-1lym(A)	IG*G1 fv fragment	model	lysozyme (A)	2.5
2	1tgn-4pti	trypsinogen	1.6	trypsin inhibitor	1.5
3	1tld-4pti	beta-trypsin	1.5	trypsin inhibitor	1.5
4	2hfl-6lyz	IG*G1 fab fragment	2.5	lysozyme	2.0
5	2pka-4pti	kallikrein a	2.0	trypsin inhibitor	1.5
6	2ptn-4pti	trypsin	1.5	trypsin inhibitor	1.5
7	2sbt-2ci2	subtilisin novo	2.8	chymotrypsin inhibitor	2.0
8	5cha(A)-2ovo	alpha-chymotrypsin (A)	1.7	ovomucoid third domain	1.5

Table 6
Size of the Unbound Examples.

	pdb	aa	int at	ext at	holes	knobs	aa	int at	ext at	holes	knobs
		receptor					ligand				
1	1hfm-1lym(A)	113	884	830	176	161	129	459	542	109	198
2	1tgn-4pti	218	817	804	139	240	58	182	271	59	22
3	1tld-4pti	220	851	778	145	202	58	182	271	59	22
4	2hfl-6lyz	213	1593	1627	270	223	129	479	522	109	72
5	2pka-4pti	223	915	884	131	153	58	182	271	59	22
6	2ptn-4pti	220	857	772	140	249	58	182	271	59	22
7	2sbt-2ci2	275	1061	873	160	129	65	205	316	64	27
8	5cha(A)-2ovo	237	894	841	135	102	56	136	282	69	22

6 The Geometric Hashing based Docking Method

6.1 The Algorithm

In this section we outline our Geometric Hashing surface matching algorithm which is based on ideas originated in Computer Vision ([19]). Here we use the

Table 7
Best RMS among top 10,50 and 100 clusters of the unbound examples.

	pdb	fast scoring			with connectivity		
		top 10	top 50	top 100	top 10	top 50	top 100
1	1hfm-1lym(A)	27.32	15.29	11.75	34.84	13.67	11.75
2	1tgn-4pti	8.24	8.24	2.75	8.24	8.24	1.85
3	1tld-4pti	10.03	6.26	6.26	13.20	6.26	6.26
4	2hfl-6lyz	28.02	24.83	23.33	28.65	9.53	1.21
5	2pka-4pti	8.11	5.65	5.65	9.90	5.65	5.53
6	2ptn-4pti	10.01	6.14	5.85	6.14	6.14	3.53
7	2sbt-2ci2	10.38	10.11	10.11	16.40	10.11	5.64
8	5cha(A)-2ovo	6.45	1.49	1.49	8.81	1.49	1.49

cap-pit-belt shape representation which is described in section 3.1. Given this surface representation our task is the same as in the previous section, namely, to find a rotation and translation of the ligand, so that a large portion of its surface 'cap' interest points are in the vicinity of corresponding receptor surface 'pit' interest points [3] having normals which are almost in the opposite direction. The resulting transformation should not cause inter-penetration of the ligand and receptor atoms. Since the number of cap/pit interest points is significantly larger than the knob/hole points, we applied a computationally more efficient matching algorithm for this case.

Below, we briefly overview the underlying principles of the Geometric Hashing technique with special emphasis on the salient features of the docking application. For additional details on the Geometric Hashing and its applications to structural comparison of molecules the reader is referred to [28,21,29,30].

Assume that we are given a ligand with m caps and a receptor with n pits. In the method described in section 5 there are $n(n-1) \times m(m-1)$ transformations which are candidate solutions. Some of these candidates are immediately rejected since the signatures of the compared point pairs are incompatible. However, for the remaining candidates a time consuming verification stage, which includes penetration check and surface contact scoring is initiated. Our

[3] Of course, the roles of 'pits' and 'caps' can be reversed.

goal is twofold. First, to reduce the number of candidate solutions which pass to the verification stage and, second, to develop a more efficient matching procedure to detect those candidates. To achieve this we use an associative memory type approach. We learn the geometric attributes of the ligand interest points and store the points in a memory (hash-table) which is indexed by these attributes Then, given the receptor interest point set, we compute their geometric attributes and retrieve from the memory those ligand's points having similar properties. The idea behind this approach in Computer Vision was triggered by the human visual performance. Assume that somebody is observing a pile of objects in an attempt to recognize them. If he spots a sharp corner in this scene, he is most likely not going to retrieve smooth spherical objects from his memory in order to check if they appear at this location of the scene. His memory will be accessed directly at all the familiar objects having sharp corners.

The implementation of the above mentioned idea for molecular surface point matching is easier said than done. If the molecular surface points were colored, each by a different color, we could index the ligand points into a table according to their color, and, given the receptor interest point set, check the color of each of its points, access the memory at this color and retrieve the matching ligand point. However, no such symbolic color, which is representative enough is available in general. (It should be noted though, that the (knob,hole) is a set of two symbolic colors, in this sense.) Hence, we shall look for a 'geometric color' which describes the geometric attributes of our points. A natural quantitative geometric attribute of a point is its coordinate set. This could be our color. Maybe we can compute the coordinates of the ligand caps and store their identity in a table which is accessed by these coordinates. Then, we could compute the coordinates of the receptor pits, and for each such coordinate access the table and check whether there is a ligand cap matching it. However, as we well know, coordinates are defined relative to some reference frame. How can we ensure that the ligand and receptor interest points are computed in the same reference frame ? Moreover, we do know that the ligand geometric representation has to undergo a rotation and translation to bind to the receptor.

Thus, before we adopt the associative memory technique, we must address the following issues :

(i) Rotational and translational invariance.
(ii) Choice of the same reference frame both for the ligand and for the receptor.
(iii) Handling of the fact that only part of the ligand caps will have counterpart receptor pits and vice versa.

These problems are addressed by redundantly representing each molecule in many reference frames, which are naturally attached to this molecule. Now the coordinates of a point will represent not only the point itself, but the point in a given reference frame. We start by unambiguously defining a local 3-D reference frame for a pair of points with their associated normals.

Reference frame definition. Let $(\boldsymbol{p_1}, \boldsymbol{n_1})$ and $(\boldsymbol{p_2}, \boldsymbol{n_2})$ be two points with their associated normals. A Cartesian reference frame is unambiguously defined by two points $\boldsymbol{a}$,$\boldsymbol{b}$ and a direction $\boldsymbol{d}$, which is different from the $\boldsymbol{ab}$ direction. It can be defined uniquely by picking $\boldsymbol{a}$ as the origin, the $\boldsymbol{ab}$ direction as the $\boldsymbol{x}$-axis, the cross-product of $\boldsymbol{ab}$ and $\boldsymbol{d}$ as the $\boldsymbol{y}$-axis, and the cross-product of the new $\boldsymbol{x}$ and $\boldsymbol{y}$ axes as the $\boldsymbol{z}$-axis. Our reference set is based on two critical points $\boldsymbol{a} = \boldsymbol{p_1}$ and $\boldsymbol{b} = \boldsymbol{p_2}$, and the $\boldsymbol{d}$ direction can be taken as the average of $\boldsymbol{n_1}$ and $\boldsymbol{n_2}$. Note that each such reference frame has a shape signature similar to that defined in section 5, which includes the inter-point distance, the angles between the normals and the inter-point segment, and the resulting torsion angle.

Now, we can compute the coordinates of all the other interest points with their associated normals in this frame. Rotation and translation of the molecule does not change these coordinates. However, representing the molecule in a single reference frame is insufficient, since we cannot ensure ***a-priori*** that the reference interest point pair participates in the binding. Hence, we represent the molecule in all possible interest point pair reference frames. Consequently, each interest point is redundantly represented in many different reference frames.

Below we describe our method for matching of one ligand to one receptor, although a data-base of ligands could be explored simultaneously. Likewise, a ligand can be docked to a database of receptors.

There are two major phases in the Geometric Hashing algorithm. The first is the Preprocessing phase, where the ligand geometric information is memorized in a look-up table or hash-table and the second is the, so called, Recognition phase, where the receptor points are efficiently matched to the ligand points via the pre-computed hash-table. One should note that once such a table is built, it can be stored and should not be computed at each run of the algorithm. Also, one should note that the actual runtime of the algorithm is the sum of the times of both phases and not their product. If the ligand table is pre-computed in a previous run, the actual execution time is only the time of the Recognition phase.

Ligand Preprocessing

– Reference sets based on a pair of ligand critical points with their associated normals are chosen.

– For each ligand reference set $\boldsymbol{rs(l_1, l_2)}$, the coordinates of the other ligand critical points (in a restricted vicinity as explained below) and their normals are computed in the Cartesian frame associated with the reference set. These coordinates serve as an index to a hash table (memory), where the reference set and the point information are stored along with additional geometric constraints on this triplet of points+normals.

Recognition

For each pair of receptor critical points do the following :

– Build a reference set based on the pair of receptor critical points (and their associated normals).
– The coordinates of the other receptor critical points and their normals are computed in the Cartesian frame associated with this reference set. For each such coordinate the hash table bins with values close to the coordinate are accessed. The records from the table are retrieved and those ligand reference sets which satisfy the geometric constraints of the receptor reference set receive a vote as matching candidates of this set.

After the voting process is completed for the current receptor reference set, if a ligand reference set receives many votes, the transformation between this set and the receptor reference set is computed, the molecules are superimposed and additional transformation refinement steps are executed. In particular, additional matching points are recovered and the best transformation (in the least squares sense) for the extended set is computed. Transformations causing inter-molecule penetration are discarded.

The flow-chart of the algorithm is given in Figure 1.

One should note that the complexity of he Recognition phase is of the order of $\boldsymbol{n^3}$, where $\boldsymbol{n}$ is the number of receptor interest points.

Following the general description of the algorithm we outline several implementation details which are specific to our application.

6.1.1 Implementation Details

We outline the constraints that we have used in our application. To get the reader a feeling of the (in)accuracies involved, we also mention the numerical values of the thresholds used to derive the experimental results presented below. Obviously, these thresholds can be changed according to the application.

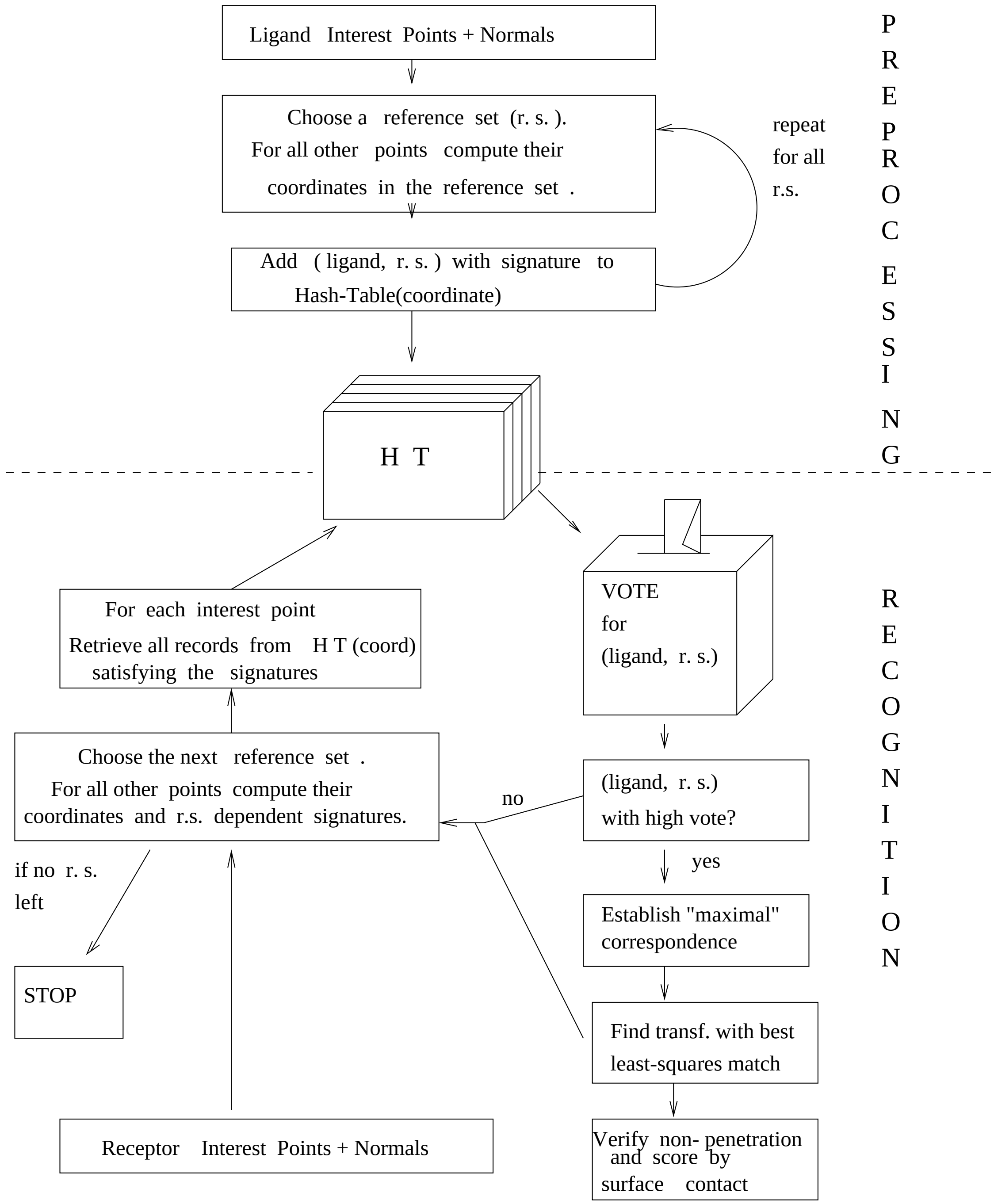

Fig. 1. The flow chart of the Geometric Hashing based docking algorithm.

Choice of reference sets. We do not use all pairs of points to build reference sets. Some distance and directional constraints are applied to a pair of points $\boldsymbol{p_1}$,$\boldsymbol{p_2}$ and their normals $\boldsymbol{n_1}$ and $\boldsymbol{n_2}$ in order to build a reference set based on them:

Distance constraint. The distance between points $\boldsymbol{p_1}$ and $\boldsymbol{p_2}$ should be within a given range:

$$d_{min} < dist(p_1, p_2) < d_{max}.$$

In our experiments we used $d_{min} = 2.5\mathring{A}$ and $d_{max} = 9.0\mathring{A}$, for the non-immunoglobulins. For the immunoglobulins we used $d_{min} = 6.0\mathring{A}$ and $d_{max} = 13.0\mathring{A}$.

Directional constraints. The orientation of the normals n_1 and n_2 should satisfy:

Angle between normals. The angle formed between n_1 and n_2 should be below a given threshold:

$$acos(n_1 \cdot n_2) < 1.2 \text{ radians.}$$

Torsion angle. The torsion angle formed between n_1,p_1, p_2 and n_2 should be smaller than a given threshold. This torsion angle is the angle between two planes. The first is defined by the line segment connecting p_1 and p_2 and by the normal n_1 at point p_1. The second plane is defined by the same line segment and the normal n_2 at point p_2.

$$Torsion(n_1, p_1, p_2, n_2) < 1.4 \text{ radians.}$$

Angle between each normal and the line segment p_1, p_2. Let x be the unit vector in the direction of $p_2 - p_1$. Then the normals should satisfy:

$$0.87 \text{ radians} < acos(n_1 \cdot x), acos(n_2 \cdot x) < 2.4 \text{ radians}$$

and

$$|acos(n_1 \cdot x) - acos(n_2 \cdot x)| < 1.2 \text{ radians.}$$

Choice of points belonging to a reference set. Given a reference set $rs(p_1, p_2)$, built on the critical points p_1 and p_2 (and the average of their normals) we can represent any point in 3D by its coordinates in $rs(p_1, p_2)$. However, we do not take all the points into consideration. We restrict ourselves to those points p which satisfy the constraint :

$$dist(p_1, p), dist(p_2, p) < d_{pmax}.$$

The rationale of this proximity constraint is the fact that we are looking for a binding site which is spatially localized, thus far away points are probably not relevant for the given reference set. By applying this constraint we gain in execution time and reduce the hash table occupancy.

Here we used $d_{pmax} = 15.0\mathring{A}$ for the non-immunoglobulins. For the immunoglobulins we used $d_{pmax} = 18.0\mathring{A}$.

Voting constraints. Given a receptor reference set $rs(r_1, r_2)$ and a receptor point r which satisfy all the above constraints, the coordinates of r in $rs(r_1, r_2)$ are computed. Let us denote these coordinates as the triplet (x, y, z).

The Hash Table is accessed to extract all the records stored at addresses of the form (a, b, c), where (a, b, c) is in the range $(x \pm \epsilon, y \pm \epsilon, z \pm \epsilon)$. We used a value of $\epsilon = 1.5$. Each of these hash table triplets (a, b, c) represents the coordinates of a ligand point in a given ligand reference set. Consider a record stored at an address with index (a, b, c). This record was computed for a ligand point l based on a ligand reference set $rs(l_1, l_2)$. It contributes one vote to the match between the current receptor reference set $rs(r_1, r_2)$ and the ligand reference set $rs(l_1, l_2)$, since the coordinates of r in $rs(r_1, r_2)$ are similar to those of l in $rs(l_1, l_2)$.

Before a vote is tallied for such a match we verify that appropriate distance and angular constraints among the participating triplets are met. For details see [9].

Aligning the receptor with the ligand. After the completion of the voting for a given receptor reference set $rs(r_1, r_2)$, all the ligand reference sets $rs(l_1, l_2)$ which have received a number of votes above a certain threshold are retrieved. In our examples this vote threshold is between 7 to 13. For each such $rs(l_1, l_2)$ with a high vote, we compute the rotation and translation of the ligand which corresponds to this vote. In the voting process all the pairs of matching points (r, l) were recorded, and thus, these pairs can be used along with the reference set points (r_1, l_1) and (r_2, l_2) in the computation of the transformation giving the minimal least squares error ([31]).

Improvement of the transformation. Given a list of matching pairs of atoms and the transformation that best superimposes the matching ligand and receptor points, we verify and "optimize" this match hypothesis. This is done by computing the contact score in a way which is similar to the one described in section 5. Also, the set of corresponding points is verified to satisfy somewhat tighter distance and angular constraints (see [9] for additional details).

Verification of penetration and contact scoring. Finally, the solutions are checked for inter penetration, and for those passing that filter the contact score is computed. These steps are in essence similar to the same steps described in Section 5. The final score indicates the geometric quality of the contact interface among the molecules.

Table 8
The molecules and their interest points.

			receptor		ligand	
			no. surf.	no. crit.	no. surf.	no. crit.
Complex	receptor	ligand	atoms	points	atoms	points
1cpk	Protein Kinase	Prot. Kin. inhibitor	939	544/74	115	119
3dfr	Dihydrofolate Reductase	Methotrexate	576	596/60	214	43
4cpa	Carboxypeptidase A$\boldsymbol{\alpha}$	Inhibitor	761	436/28	166	148
4sgb	Serine Protease B	Potato Inhibitor PCI-1	477	296/68	214	163
1cho	$\boldsymbol{\alpha}$-Chymotripsin	Ovomucoid 3rd domain	620	449/69	226	167
2mbh	Hemoglobin $\boldsymbol{\beta}$-sub.	Hemoglobin $\boldsymbol{\alpha}$-sub.	469	273/58	419	249
1fdl	IgG1 Fab frag.	Lysozyme	646	53	421	439
unb.	5choe	2ovo	636	449/69	227	167
unb.	2hfl	1lyz	616	62	408	452

6.2 Experimental Results

We have carried out dozens of docking experiments (see [9] for listings of results). Some of the results we present in Table 8. The table also lists the number of surface atoms and the number of interest points of the receptor and the ligand. The number of the receptor active site interest points appear next to the receptor's interest points.

As in the results presented in Section 5 our experiments are divided into two classes - bound and unbound. For the first class of bound examples we have examined ligand-receptor complexes appearing in the PDB. We have randomly rotated and translated the ligand, and applied our algorithm to the transformed molecules. The performance of our method is assessed by comparing the proposed docking orientations with the original orientation. To measure the accuracy of a solution, we compute the RMS (root-mean-square) distance between the ligand atomic coordinates in their original orientation and the coordinates of the same atoms in the orientation suggested by our algorithm. Commonly an RMS below 2 Å is considered to be a successful docking orientation.

The second class of experiments involves docking examples of unbound molecules, where the receptor and ligand were independently crystallized. In these experiments, the correct docking orientation is not known. However, the unbound

Table 9
Results of the docking algorithm for the full ligand and receptor surfaces.

receptor	ligand	best rms	no. of sols.	best rank	time full
1cpke	1cpki	1.20	37	1	19.6
3dfr	3dfrm	0.65	1140	7	10.5
4cpa	4cpai	1.51	3249	147	16.3
4sgbe	4sgbi	1.09	827	6	52.0
1choe	1choi	0.80	1721	6	27.3
2mhbb	2mhba	0.79	32	1	4.6
1fdll	1fdly	1.97	285	22	9.8
unbound					
5choe	2ovo	1.32	8490	214	29.7
2hfll	1lyz	1.28	8957	1497	55.8

examples presented here use unbound ligands and unbound receptors that are similar to other ligands and receptors that have also been crystallized as complexes. These complexes serve as a reference for the expected docking orientation. To assess the performance of our docking method, we compute the RMS distance of the unbound ligand in the proposed solution to the ligand in the reference orientation. Here again, RMS below 2 Å is considered a correct solution.

The first 7 examples in the tables belong to the first class, while the last two examples are of 'unbound' docking. In general, one is interested in docking a ligand onto a receptor for which neither a complex exists nor a reference complex can be found. In these cases, the RMS criterion cannot be used to select the correct solution.

Table 9 lists the results obtained in the docking experiments. The RMS of the highest ranking solution, which is still below 1.5 Å (2 Å for immunoglobulins) is listed in the first column. The second column lists the number of solutions which scored above the minimal number of votes in the Geometric Hashing stage and have no significant penetration of the ligand into the receptor. These are subsequently ranked by the surface contact score. The third column shows the rank of the solution whose RMS is listed in the first column. Table 9 demonstrates that the docked solution that is closest to the crystallographically determined one ranks at the very top for the complexes. The RMSD in all cases is below 2.0Å. The times required to complete the docking runs on an SGI Indigo 2 workstation are listed in the next column. The experiments were run using the full receptor surface. One should note that in

Table 10
Results of the docking algorithm for the full ligand and active site receptor surface.

receptor	ligand	best rms	no. of sols.	best rank	time
1cpke	1cpki	1.20	37	1	2.2
3dfr	3dfrm	0.65	202	3	0.7
4cpa	4cpai	1.51	730	28	2.1
4sgbe	4sgbi	1.09	645	2	4.2
1choe	1choi	0.80	389	4	4.2
2mhbb	2mhba	0.79	17	1	1.1
1fdll	1fdly	1.97	67	22	1.6
unbound					
5choe	2ovo	1.32	1490	31	3.4
2hfll	1lyz	1.28	2635	84	3.8

this experiment we have not applied a clustering procedure on the resulting solutions. Since many candidate solutions tend to cluster, application of such a procedure would have improved the ranking of the "best" solutions.

The docking experiments were repeated using only the critical points of the receptor which belong to the interface (see Table 8). This is motivated by the fact that sometimes we can estimate the location of the active site in advance. Table 10 shows the results for this easier case.

7 Further Developments and Future Tasks

We have discussed two methods for rigid molecule docking which are based only on geometric complementarity search. We have shown that rigid docking of relatively large complexes can be accomplished in short time, and that solutions close to the native score high enough to allow their detection. We have discussed major issues in geometric pattern matching, such as the molecular surface representation, extraction of interest features and efficient matching algorithms. It should be clear that these are not independent issues. Especially the interest features extracted and the matching algorithms used may be tightly coupled in the sense that there are certain features sets which can trigger the development of one class of algorithms, while for another algorithm class, suitable features may be detected and developed. We have explained the analogy between object recognition tasks in Computer Vision and prediction of biomolecular docking. We believe that this analogy can help to better understand the geometric shape complementarity issues involved in both disciplines.

In our examples the performance on bound cases was truly satisfactory. Although we achieved reasonable results for the unbound case as well, it is obvious that much work has to be done before the performance on these cases is satisfactory. This is mainly due to the fact that in many cases the rigid approximation of the molecule is not accurate enough for docking purposes. Even if the molecules do not undergo major conformational changes, their surfaces often adjust to each other in the docking process thus departing from their rigid shape.

Another major challenge is dealing with major conformational changes in the ligand or receptor. Several efforts for such flexible docking algorithms are currently in progress and the major ones have been mentioned in the Introduction. The challenge there is to develop efficient algorithms, which can effectively search the much bigger parameter space without increasing significantly the complexity with regard to the rigid docking algorithms.

Acknowledgements

We thank our students and colleagues D. Fischer, S.L. Lin, and R. Norel, who have contributed to this research effort and implemented the algorithms discussed in this chapter. The research of R. Nussinov has been sponsored by the National Cancer Institute, DHHS, under Contract No. 1-CO-74102 with SAIC. The contents of this publication do not necessarily reflect the views or policies of the DHHS, nor does mention of trade names, commercial products, or organization imply of endorsement by the U.S. Government. The research of H.J. Wolfson and R. Nussinov in Israel has been supported in part by grant No. 95-0028 from the U.S.-Israel Binational Science Foundation (BSF), Jerusalem, Israel, by the Rekanati Foundation, by the Basic Research Fund of Tel Aviv University, and by consortium "Daat" of the Israeli Ministry of Trade and Industry. H.J. Wolfson acknowledges the support of the Hermann Minkowski – Minerva Center for Geometry at Tel Aviv University.

References

[1] Janin, J. and Chothia, C. (1990) *J. Mol. Biol.* 265, 16027–16030.

[2] Cherfils, J. and Janin, J. (1993) *Current Opinion in Structural Biology* 3, 265–269.

[3] Kuntz, I.D., Blaney, J.M., Oatley, S.J., Langridge, R., and Ferrin, T.E. (1982) *J. Mol. Biol.* 161, 269–288.

[4] Shoichet, B.K., Bodian, D.L., and Kuntz, I.D. (1992) *J. Comp. Chem.* 13, 380–397.

[5] Connolly, M.L. (1986) *Biopolymers* 25, 1229–1247.

[6] Jiang, F. and Kim, S.H. (1991) *J. Mol. Biol.* 219, 79–102.

[7] Katchalski-Katzir, E., Shariv, I., Eisenstein, M., Friesem, A.A., Aflalo, C., and Vakser, I.A. (1992) *Proc. Natl. Acad. Sci. USA* 89, 2195–2199.

[8] Norel, R., Fischer, D., Wolfson, H.J., and Nussinov, R. (1994) *Protein Engineering* 7, 39–46.

[9] Fischer, D., Lin, S. L., Wolfson, H.J., and Nussinov, R. (1995) *J. Mol. Biol.* 248, 459–477.

[10] Norel, R., Lin, S. L., Wolfson, H.J., and Nussinov, R. (1994) *Biopolymers* 34, 933–940.

[11] Norel, R., Lin, S. L., Wolfson, H.J., and Nussinov, R. (1995) *J. Mol. Biol.* 252, 263–273.

[12] DesJarlais, R.L., Sheridan, R.P., Dixon, J.S., Kuntz, I.D., and Venkataraghavan, R. (1986) *J. Med. Chem.* 29, 2149–2153.

[13] Leach, A.R. and Kuntz, I.D. (1992) *J. Comp. Chem.* 13(6), 730–748.

[14] Rarey, M., Kramer, B., Lengauer, T., and Klebe, G. (1996) *J. Mol. Biol.* 261, 470–489.

[15] Jones, G., Willet, P., and Glen, R. (1995) *J. Mol. Biol.* 245, 43–53.

[16] Sandak, B., Nussinov, R., and Wolfson, H.J. (1995) *Computer Applications in the Biosciences (CABIOS)* 11, 87–99.

[17] Sandak, B., Nussinov, R., and Wolfson, H.J. (1996) Docking of conformationally flexible molecules In *Seventh Symposium on Combinatorial Pattern Matching, CPM 96* pp. 271–287. Springer Verlag Lecture Notes in Computer Science 1075.

[18] Lengauer, T. and Rarey, M. (1996) *Current Opinion in Structural Biology* 6, 402–406.

[19] Lamdan, Y. and Wolfson, H.J. (1988) Geometric Hashing: A General and Efficient Model-Based Recognition Scheme In *Proceedings of the IEEE Int. Conf. on Computer Vision* pp. 238–249 Tampa, Florida.

[20] Lamdan, Y., Schwartz, J. T., and Wolfson, H.J. (1990) *IEEE Trans. on Robotics and Automation* 6(5), 578–589.

[21] Nussinov, R. and Wolfson, H.J. (1991) *Proc. Natl. Acad. Sci. USA* 88, 10495–10499.

[22] Richards, F.M. (1977) *Ann. Rev. Biophys. Bioeng.* 6, 151–176.

[23] Connolly, M.L. (1983) *Science* 221, 709–713.

[24] Connolly, M.L. (1983) *J. Appl. Cryst.* 16, 548–558.

[25] Lin, S. L., Nussinov, R., Fischer, D., and Wolfson, H.J. (1994) *PROTEINS: Structure, Function and Genetics* 18, 94–101.

[26] Lin, S. L. and Nussinov, R. (1996) *J. of Molecular Graphics* 14, 78–90.

[27] Lin, S. L., Xu, D., Li, A., Rosen, M., Wolfson, H.J., and Nussinov, R. (1997) *J. Mol. Biol.* 271(5), 838–8452.

[28] Wolfson, H.J. (1990) Model Based Object Recognition by 'Geometric Hashing' In *Proc. of the European Conf. on Computer Vision* pp. 526–536 Lecture Notes in Computer Science 427, Springer.

[29] Bachar, O., Fischer, D., Nussinov, R., and Wolfson, H.J. (1993) *Protein Engineering* 6(3), 279–288.

[30] Fischer, D., Norel, R., Wolfson, H.J., and Nussinov, R. (1993) *PROTEINS: Structure, Function and Genetics* 16, 278–292.

[31] Schwartz, J.T. and Sharir, M. (1987) *The Int. J. of Robotics Research* 6(2), 29–44.