Small Molecule Recognition: Solid Angles Surface Representation and Molecular Shape Complementarity

R. Norel[1]*, H. J. Wolfson[1] and R. Nussinov[2,3†]

June, 1999

[1] *Computer Science Department*

School of Mathematical Sciences

Tel Aviv University, Tel Aviv 69978, Israel

[2] *Laboratory of Experimental and Computational Biology*

SAIC

NCI-FCRDC

Bldg 469, rm 151

Frederick, MD 21702, USA

[3] *Sackler Inst. of Molecular Medicine*

Faculty of Medicine

Tel Aviv University, Tel Aviv 69978, Israel

*current address: Dept. of Biochemistry & Molecular Biophysics, Columbia University, 630 West 168th St., New York, NY 10032, U.S.A.

†correspondence should be addressed to R. Nussinov at NCI-FCRF Bldg 469, rm 151, Frederick, MD 21702. TEL: 301-846-5579, FAX: 301-846-5598, email: ruthn@ncifcrf.gov.

Abstract

Here we examine the recognition of small molecules by their protein and DNA receptors. We focus on two questions: First, how well does the solid angle molecular surface representation perform in fitting together the surfaces of small ligands, such as drugs and cofactors to their corresponding receptors; And second, in particular, to what extent does the shape complementarity play a role in the matching (recognition) process of such small molecules. Both questions have been investigated in protein–protein binding: "Critical Points" based on solid angle calculations have been shown to perform well in the matching of large protein molecules. They are robust, may be few in numbers, and capture satisfactorily the molecular shape. Shape complementarity has been shown to be a critical factor in protein-protein recognition, but has not examined in drug–receptor recognition.

To probe these questions, here we dock 185 receptor–small ligand molecule pairs. We find that such a representation performs adequately for the smaller ligands too, and that shape complementarity is also observed. These issues are important, given the large databases of drugs that routinely have to be scanned to find candidate, lead compounds.

We have been able to carry out such large scale docking experiments owing to our efficient, computer-vision based docking algorithms. Its fast CPU matching times, on the order of minutes on a PC, allows such large scale docking experiments.

Keywords: docking; small ligand–protein recognition; drugs; molecular surface complementarity; solid angles; surface normals

Running title: Small molecule–receptor recognition

Introduction

Here we examine the recognition of small molecule ligands by their considerably larger receptors. We focus on two inter-related issues. The first probes the representation of the molecular surfaces; The second examines the shape complementarity, and the extent of its determination of their recognition. While these issues have been investigated in large molecule recognition, they have not been specifically addressed in small molecule binding.

Geometrical fitting of two molecules requires adequate description of their molecular surfaces. A dot description is particularly convenient, since it allows matching corresponding dots on the two molecular surfaces. Nevertheless, a dot surface description and matching entails some potential difficulties. On the one hand, if the description of the surface is performed by covering the molecule with too many dots, the combinatorial complexity of the matching will be high, with very long matching times. Additionally, if the dots are too close to each other, they will fall within the error threshold of the matching. Hence, the goal is to describe molecular surfaces by as few dots as possible, while still preserving good surface complementarity, and not missing out on good matching solutions. Ideally, one would want a coarse dot representation. However, it should still preserve the important features of the surface such as the peaks of the hills and the bottoms of the valleys, which would enable recognition of the receptor-ligand molecule pair. Already over a decade ago, Connolly has developed such a scheme [1]. Starting from his initial molecular surface dot (MS dot) description [2,3], Connolly has proceeded to pick only "critical points". Such points describe local knobs and holes, representing, respectively, local minima and maxima of the shape function. The values are calculated by constructing spheres with a given radius, with their origin at each of the original MS dots. The areas of the spheres which are contained within the proteins are measured, and divided by the square of the radius. To qualify as a "critical point", the *solid angle* has to be above (below), certain thresholds. In the matching procedure, a knob pairs with a hole. The solid angles have proven to be robust 4-7]. It suffices for only few (we require two) knob-hole pairs to match to produce good matching solutions. With this surface representation, and our efficient algorithm, we obtain docked solutions in short matching times [4,5]. The solid angle surface representation enables docking into all potential binding sites, even though they are *a priori* unknown. This surface description differs from that of the original DOCK [8] or from Gold [9]. The former focuses on the largest cavity, which is detected by the largest number of intersecting spheres. Hence, it might conceivably encounter difficulties if binding takes place at a site which does not constitute such a hole. On the other hand, GOLD is based on a hydrogen-bonding scheme. For ligands which are largely hydrophobic, there could conceivably be only few hydrogen bonds across the binding interface. In this sense, the solid angle approach is more generally applicable. Thus, recently, it has been incorporated into the DOCK scheme as well [10].

Hence, despite the sparse surface representation, acceptable solutions for matching of two relatively large molecules, has still been obtained [4,5]. However, the question of how well such a molecular surface representation performs in reproducing the binding of small molecules onto their cognate receptors has not been investigated. This question is relevant, as drugs are small molecules. Thus, the description of their molecular surfaces cannot straightforwardly be expected to be as sparse as that of large molecules, and still achieve correct docked solutions. On the other hand, too detailed a description of their molecular surfaces is undesirable, if large databases of drugs are to be scanned against relatively large (protein, DNA) receptors. Speed and efficiency are prime considerations under such circumstances.

The docking procedure is purely geometrical. While the structure of the receptor is available, the location of the binding site is assumed to be unknown, leading to a large number of obtained docked solutions. So long as there is no inter-penetration of the two (receptor–ligand) molecule pair, and the minimum docked quality is achieved, it is accepted as a potential matching solution. At the end of the procedure, each solution is scored and ranked. Scoring and ranking is based largely on rewarding surface contact, and penalizing (minor) surface penetration. (Solutions entailing serious inter-penetration of the two molecules are rejected). This type of geometrical surface scoring ranks the optimally matching molecules, having in principle the more desirable shape complementarity, at the top. Thus, by inspecting the ranked solutions as compared to the actual, experimental crystal complexes, we can address the second, inter-related issue, namely, the extent of shape complementarity in the actual bound molecules, as compared to all other hypothetical matchings. We have already observed that shape complementarity is a major determinant in large-molecule recognition [6]. We have investigated shape complementarity both in the docking of protein–protein *bound* cases, where the structures have been determined when the molecules are bound together in the complex, and in particular, in *unbound* cases, where the structures of the molecules have been determined separately.

Here we examine how well a docking and ranking procedure based on shape complementarity performs for small molecule ligands, most of which are drugs. We compare the performance of our docking which utilizes strictly shape complementarity, to the docking of small drugs where the geometry and the energetics of hydrogen bonds have been the prime consideration. To do that, we carry out systematic examination of shape complementarity. About 160 cases of receptor-drugs are docked. We compare our results to the extensive H-bonds based dockings for drugs [9,11]. To perform these docking experiments, we employ our efficient computer-vision based docking routines [4-6]. The performance of the algorithm using strictly shape complementarity, indicates that for practically all cases, shape complementarity is a prime determinant. In addition, we test the applicability of shape complementarity considerations to protein–DNA and DNA–drug cases. As the backbone of DNA is negatively charged, with the molecule having an appreciable extent of accessible polar surface area, it has been proposed that H-bonds considerations may be a dominant factor in the docking of these molecules as well. These results are consistent with what we have observed in the docking of protein–protein bound, and unbound cases. Previously, for 26 complexed "bound" protein–protein cases, our RMSDs ranged from $0.5\AA$ to $2.5\AA$; for the 19 protein–protein cases whose structures have been determined separately, the RMSDs ranged between $1.2\AA$ to $5.2\AA$. Here, for the 160 tested drugs the range is between $0.4\AA$ to $4.6\AA$. However, when a smaller probe size is used in the generation of the "critical points", with the probe ball able to roll into the smaller pockets and crevices, the worst drug case achieves an RMSD of $2.5\AA$.

Extensive analyses of protein–protein interfaces have shown that even for crystal oligomers the variability of hydrogen bonds geometry may be relatively large, both with respect to angular and to distance distributions [12]. Additionally, superimposing an unbound receptor on a bound one, has illustrated conformational changes in charged residues upon binding. When a given receptor binds to different ligands, there is an extensive variability in the geometry of the hydrogen bonds. The question of shape complementarity is particularly relevant for drugs, where the molecules are small, appear to be more charged and to form more hydrogen bonds with respect to their interface area than the larger protein–protein complexes, and to be more flexible.

Taken together, our previous and current results illustrate that shape complementarity is important for a broad range of ligands. Shape and hydrophobicity appear to play a major role not only in protein–protein binding, but in the binding of drugs and DNA as well.

Nevertheless, in carrying out such studies, and considering their implications, one should bear in mind the implicit assumptions which are inherent to the approach and methodology: First, the ligand and the receptor have fixed geometries; Second, in the description of the molecular surfaces certain radii are utilized [1]; Third, the approach is strictly geometry-based. There is no consideration of the energetics nor of solvent effects. We also note that not all solutions are examined. Only the geometrically top ranking ones.

Methods

Critical Points

We select the critical points from the Molecular Surface dots which are generated by the MS program [2,3]. The MS points describe the molecular surface of the molecules. The selection of critical points for the receptor (the larger of the two molecules) is similar to the way already presented by us [4,5]. For the larger receptor molecule, the procedure is identical. For the smaller drug molecules, some modification has been implemented. Briefly, for each of the drug molecules the following steps are carried out. An atom of the molecule is designated as exterior (E) if an MS dot is generated for it. Otherwise it is an interior atom (I). The molecule is mapped onto a 3D grid (with a resolution of $0.25\mathring{A}$). Interior atoms are mapped first. They are viewed as balls with a radius equal to the atom's van der Waals plus the probe radius. Exterior atoms are mapped next. For these atoms, we use the atom's van der Waals radius. The next step involves computation of the shape function for each of the MS dots. The shape function corresponds to the intersection between the molecule and a sphere of a given radius centered at the MS dot. This shape function measures the local convexity of the surface. The critical points constitute a subset of the local extrema of the function. Dots with high shape function represent a *hole*; dots with low value in the shape function represent a *knob*. Since our goal is to search for deep holes or high knobs, only dots with shape function $> \frac{2}{3}V$ or $< \frac{1}{3}V$ are retained as candidates for a critical point. A candidate dot is selected to be a critical point among its 12 closest neighbors, if its shape function is the local maxima for holes, or local minima for knobs. The surface normal orientation is computed as the segment from the center of mass of the intersection volume, and the dot at which the ball is centered. Further details are given in [4,5]. Previously, for the large protein–protein docking, both knobs and holes were calculated for each molecule. We sought matched knob–hole pairs from the two molecules, making no reference as to which of the molecules contributes the hole, and which the knob. For the drug docking here, we tailor the function to dock small ligands onto a receptor. As usually the small ligand binds into a pocket of the receptor, here we compute holes in the larger receptor, and knobs in the smaller ligand. The shape function of the receptor's holes is computed as previously, that is, the intersection of a ball with a radius of $6\mathring{A}$. However, in the computation of the shape function of the knobs of the ligand, the radius of the intersecting ball has been reduced to $1\mathring{A}$. The rationale for using a ball of radius $6\mathring{A}$ for the protein molecular surface description has been to choose a probe of a radius similar to that of a side chain. In the case of small ligands, like drugs or cofactors, using such a ball is simply too large. Many of the small ligands are smaller than a ball of radius $6\mathring{A}$.

In addition, for the smaller ligand the van der Waals radii are augmented by 30% as compared to the values utilized in the protein–protein docking. By inflating the atoms of small molecules we achieve two goals. First, since the docking algorithm seeks superimposing two matching surface points (one from each molecule) and bringing them together, the resulting ligand orientation is often closer to the receptor than in the native conformation. However, the larger van der Waals radii result in distancing them further apart. Second, the contact area between small (or medium sized) ligands and their receptors is often limited as compared to the cases of protein–protein docking. The fit may not be snug either. Thus the enlarged radii increase the contact area, favoring the scoring scheme. Figure (**1**) illustrates the effect of the different radii sizes. When superimposing two balls with different radii at the same contact points, the center of the smaller ball, having a "normal vdw" (labeled C_n) is closer to the receptor than the center of the larger ball having an inflated van der Waals (labeled C_i). The latter is moved further away from the receptor's atoms' centers.

Docking

In the matching step we use the same routine with the same parameters as described previously [4,5] for the protein–protein case, to the protein–small molecule docking. Here we repeat only the essential features of the procedure in a concise form. 3 non-collinear points from each molecule are needed to compute a rigid transformation which is required in order to superimpose one body onto the other. However, it may happen that there are no 3 independent matching critical points pairs between the receptor and the ligand. Hence, we have proceeded to adopt a slightly altered strategy. This strategy enables both to find the matched surfaces, and, concomitantly, also reduces the complexity of the process. We simply utilize only two critical points. The necessary additional points are provided by the tips of their respective surface normals. This is unlike the original work of Connolly [1] which has inspired our approach. Connolly used 4 independent critical points; however, 4 such points are not always present in the binding interface.

For each pair of points from each of the molecules (we have used any two critical points for the protein–protein docking in our previous procedure; we use two holes for the receptor and two knobs for the ligand in the protein–drug cases here) we compute a "signature". The information required for the calculation of the signature includes: the distance between the dots (d); the two angles formed by the line between the critical points and their respective normals (α_1, α_2), and the torsion angle between the two normals (ω). We then compute the best rigid transformation [13] between the two pairs of points only if the signatures are compatible. We define the signatures as compatible if: a knob in one molecule (the drug) matches a hole in the other (the receptor); the difference between the two distances (d) is less than $2\mathring{A}$; the difference between the two corresponding $\alpha's$ is up to 0.6 radians and the difference between the $\omega's$ is up to 0.7 radians. To impose a stronger constraint on the alignment of the normals we apply two additional requirements: the sum of the differences of the two $\alpha's$ should be less than 0.75 radians, and the sum of the differences of the three angles should be less than 1.2 radians. With these conditions, if one of the normals is not aligned properly, the other must be reasonably well aligned in order for the pair of dots to still be considered a potentially acceptable docked solution. There are several advantages in utilizing the surface normals in the signature, and not only the surface dots. First, we utilize only two critical points in the interface area. Second, the combinatorics of finding the correct match is lower (we use pairs of points, rather than triplets or quartets). And third, the normal orientation is convenient for fast pruning of many wrong solutions.

Geometric Scoring and Overlap Test

The matching is computed locally. Thus, it is necessary to check if the solution is acceptable for the whole protein. That is, it could be that when bringing the ligand onto the receptor there is overlap elsewhere between the two molecules. We proceed then to compute a scoring function, based on geometric features. We award surface contact, penalize overlaps and reject serious overlaps. Since we allow some intermolecular penetrations, we implicitly take into account a certain extent of conformational flexibility. The solutions which are discarded are those in which atoms belonging to the ligand fall into the interior, or "core" of the receptor protein. Ligand atom centers which invade the outer shell of the molecular representation are retained.

For the protein–protein docking the scoring and overlap test is as described previously [5] Here, for the protein–small molecule cases, we simply invert the order of the mapping of the receptor's atoms and points, to obtain a larger interior atoms voxel-core. This reduces the number of solutions, and improves the ranking of the survivors. For the protein–small molecule scoring, the receptor is mapped onto a 3D grid (with a step size of $0.5\mathring{A}$). For each of the MS dots generated for the receptor, a ball of radius $1\mathring{A}$ is mapped onto the grid voxels. These are defined as 's' voxels. The interior atoms (i.e., those for which no MS dots are generated) are mapped as balls of radii equal to their corresponding van der Waals radii. These are defined as 'i' voxels. Lastly, the exterior atoms (i.e., those for which MS dots are generated) are mapped onto the grid voxels as balls with the corresponding van der Waals radii, *overwriting* any previous voxel definition. These are labeled 'e' voxels. With such mapping, the receptor is represented as a 'core' of interior voxels (i), a wide layer of intermediate voxels (e) and a thin layer of surface (s) voxels. After the receptor is mapped, the ligand atoms are transformed using the best rigid transformation matrix obtained in the matching step. If any of the ligand atoms fall into a voxel designated 'i' or 'e' the solution is discarded. If the solution passes the overlap test, MS dots of the ligand (computed using a density of 1 dot per $\mathring{A}^2$) are also transformed and used to compute a scoring function. Three counters are kept, one for ligand dots that fall in 'i' voxels (I), one for MS dots that fall into 'e' voxels (E) and one for MS ligand dots that fall into 's' voxels (S). For the protein–protein and protein–DNA cases the score is computed as: S - 4E - 10I. For the protein–drugs and DNA–drugs the score is 5S + E - I. Surface contact score increases the score, and overlap of the ligand surface dots into the receptor's interior reduces the score. This scoring scheme is used for ranking the solutions.

Results

Our method is entirely automated, with no manual modification of the files. The complex is extracted from the protein data bank, PDB [14], and the receptor and the ligand are separated and docked back together by the docking program. The quality of the docked solutions is computed by measuring the RMSD between the native orientation of the ligand in the PDB crystal complex and the ligand as oriented by the docking scheme. The only information utilized is the atomic coordinates of both molecules (that is, the standard PDB file). Only heavy atoms are used. No hydrogens are included. As we have done previously in our work [4-7] here too, a single set of docking parameters has been employed in all of the examples, with the matching parameter set identical to the one used previously. Thus, a single matching program and parameter set is employed for a wide range of molecular sizes, both for the large protein–protein cases, including the immunoglobulins, and for the small molecule drugs shown here. The only difference relates to the probe size for the drugs. As detailed above, a smaller probe will enable accessing the pockets and crevices on the receptor surface yielding better results. The larger probe size which works well in protein–protein docking, will skip pockets which are too small for it to access.

To shrink the number of docked solutions and, concomitantly improve the ranking of the much smaller drug molecules, we have simply reversed the order of the mapping of the solutions onto the grid in the filtering/ranking step. In the docking of protein–protein examples, we initially map the interior atoms, labeling the respective grid *voxels* where they fall as *interior*. We next proceed to overwrite the *surface* atoms/voxels over these. On the other hand, when we dock the smaller drugs, we start by mapping and labeling the surface voxels. We next overwrite these with the *interior* atoms/voxels, where applicable. Since a penetration of a ligand atom into an *interior* receptor voxel is disallowed, in practice this has the effect of rejecting more solutions. Consequently, the ranking is better for the remaining ones. Hence, by reversing the order of the steps in the program for the ligand, we essentially reduce the receptor shell into which ligand atoms are allowed to penetrate.

Protein–drug cases

For the drug docking, we have examined 141 cases from 101 PDB files [14], chosen by Jones *et al.,* [9] on the basis of their pharmacological interest and 16 cases from 14 PDB files, chosen by Rarey *et al.*, [11]. 21 out of the 141 examples were employed for fine tuning the parameters of the overlap and scoring procedures. We have examined a wide range of molecular sizes both for the receptor (630 to 7400 atoms per molecule) and for the ligand (5 to 92 atoms). As in the protein–protein cases, the quality of the results does not depend on the molecular sizes, neither the receptor's nor the ligand's. Tables 1a,b list the PDB file names, and the CPU times, in minutes, used for the matching and scoring processes. The table also presents the solution with the lowest RMSD which has been obtained. The last column in the table shows the best ranking solution with RMSD $< 5\AA$ which we have obtained. 109 out of 157 cases achieve an RMSD $\leq 1\AA$; 36 cases obtain an RMSD between 1 and 2 $\AA$; that is, 145 out of 157 cases with RMSD $< 2\AA$. More than 92% of the cases are successfully docked. The results are summarized in the histograms in Figure (**2**).

As Tables 1a,b and the histograms show, in 7 cases our optimally docked solutions have an RMSD $> 4\AA$. It is interesting to analyze these separately to see why we failed there. Our worst example is 1baf (Fab fragment of murine monoclonal antibody complexed with its hapten), with an RMSD equal to $4.59\AA$. Visual inspection of the native complex illustrates that the pocket where the ligand binds to the receptor is too small for the $6\AA$ probe in the calculation of the critical points describing the molecular surface (see Methods). Using a $6\AA$ radius probe for the detection of holes, we skip those receptor invaginations where the holes are smaller than that probe size. Hence, we have reduced the probe radius in the calculation of the critical points of the receptor to $3\AA$. Inspection of Table 2 shows the striking effect of this probe size reduction on the RMSD. For the 1baf case, the RMSD has dropped to $1.74\AA$. For the ligand the probe size has been kept as before, at $1\AA$ radius, hence using the previously generated critical points. Table 2 also shows an increase in the CPU time and a lower ranking of the near native solution. This is expected, since a smaller probe generates more critical points, implying more shape signatures to match, and more docked configurations to test. Figures (**3a,b**) depict the 1baf with the 6Å probe as compared to the solution obtained with a probe of 3Å. Figures (**3c,d**) focus on the binding site of 1baf. Figures (**4a,b**) depict 1slt1 (S-lectin complexed with N-acetyllactosamine), with probe sizes 6Å and 3Å respectively. Figures (**4c,d**) show the binding site of 1slt1 with the solutions presented in 4a,b respectively. In this case the ranking rose from 49 to 2, and the RMSDs dropped from 4.57Å to 1.20Å (see Table 2).

DNA cases

27 DNA cases have been docked. Of these, 13 are protein–DNA and 14 are DNA–drug. The protein–DNA cases include 1cgp, 1dnk, 1eri, 1fos, 1fos2, 1hcr, 1lat, 1latB, 2bop, 2drp, 2drp2, 2rve and 2rve2. The CPU times for the matching vary between 0.4 min (for 1hcr) to 18.0 min (for 1eri). The scoring times are appreciably longer, varying between 1.4 min (again for 1hcr) to 75 min (for 1eri). The RMSDs are not as good for these cases. They range between 1.25Å (1dnk) to 3.5Å, with three cases obtaining larger RMSDs: 1cgp (6.08Å), 1fos2 (3.70Å) and 2rve (4.52Å). Figure (**5**) depicts the solution obtained for one example, 1fos, with an RMSD of 3.18Å (CPU times are 1.8 and 12.8 min for the matching and scoring, respectively). Table 3 lists the results obtained in the docking of the drugs to the DNA. The worst case here is that of 131d, with an RMSD of 1.30Å and, a ranking of 4220. This illustrates the large number of ways in which this symmetric, elongated, zig-zagging molecule can fit onto the DNA.

Discussion

Here we have docked small molecule ligands onto their large receptors, using our automated docking routine. The location of the binding site is assumed to be unknown. Hence, the procedure is based on the entire molecular surfaces of both molecules, and their geometric descriptions.

In this work, we examine small molecule drugs, where the extent of the interacting surfaces are relatively small. Furthermore, in protein–protein recognition the bound surfaces are usually, though not always, largely hydrophobic. However, in drugs and in DNA, the surfaces are more polar/charged. Hence, the nature of the docking process may conceivably differ. The goals of this work have been first, to probe the question of the shape representation, given this large difference in size. Specifically, we wished to determine whether a sparsely described molecular surface will still enable obtaining good docked configurations; And, second, given the difference in the chemical nature of the surfaces, in addition to the difference in the interface size, we wished to examine if the shape complementarity still plays an important role in the recognition process. These issues are important, considering the large sizes of the databases of drugs that one frequently needs to go through when searching for optimally bound drugs. Under such circumstances, especially if the active sites are unknown, or if more than one binding site is suspected to exist, the entire molecular surfaces need to be considered. Only after potential lead drugs are detected, the next essential stage is a detailed examination of the energetics. These energy calculations are much more time consuming. Thus, realistically one is able to study only a rather small number of candidates.

Here we have examined a very large number of cases. A total of about 185 cases have been docked, resulting in the most thorough examination of these issues that we are aware of. We find that shape complementarity is an essential feature to consider in the docking, whether of large protein–protein or of small molecules, such as drugs. In addition, we find that sparse molecular surface representation for drugs works well in the matching as well. However, we also find that in order to be able to dock small molecules, it is preferable to use a smaller probe size when describing the molecular surface of the drugs. Too large a probe in the calculation of the critical points may result in skipping the smaller clefts and invaginations into which the smaller molecule may enter. It is interesting to note that even though to a large extent drugs are charged/polar, forming hydrogen bonds and other electrostatic interactions with their respective receptors, still there is a considerable extent of shape complementarity. Shape complementarity largely implicitly reflects hydrophobic and van der Waals interactions.

Over the last few years, a large number of docking routines have been developed and employed [15-31]. Techniques vary, according to the goals at hand. In some, the binding sites are assumed to be known, and the ligands are positioned directly in the active sites. In others, global searches are instituted. Techniques and parameters sometimes vary with the molecule size. Our computer-vision based technique is general, and is highly efficient. Its utility has been shown both for large and for small molecules. Nevertheless, while the matching is attractive, the ranking still needs improvement. This is a general problem, observed in inter-molecular binding, and in folding. Folding and binding are similar processes.

The extensive and systematic tests we carry out utilizing our docking algorithms, and the quality of the results we obtain are encouraging. Here, for the drugs, the program is able to obtain solutions with RMSDs $< 2\mathring{A}$ from the native conformations for 145 out of 157 tested crystal cases. This, along with the fast CPU times, are encouraging for a computer assisted drug design process.

Acknowledgment

We thank Dr. J. V. Maizel for discussions, and for encouragement. We also thank Dr. B. Honig for encouragement. The research of R. Nussinov has been sponsored by the National Cancer Institute, DHHS, under Contract No. 1-CO-74102 with SAIC. The research of R. Nussinov and H. Wolfson in Israel has been supported in part by grant No. 95-00208 from the BSF, Israel, by a grant from the *Israel Science Foundation* administered by the *Israel Academy of Sciences*, by the Magnet fund, administered by the Israel Ministry of Industry, and by a basic science grant administered by Tel Aviv University and the Adams Brain Center grant. 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 endorsement by the U.S. Government. The research of H. Wolfson is partially supported by the Hermann Minkowski-Minerva Center for Geometry at Tel Aviv University.

References

1. Connolly, M. Shape complementarity at the hemoglobin $\alpha_1\beta_1$ subunit interface. *Biopolymers*, **1986**, 25:1229–1247.

2. Connolly, M. Analytical molecular surface calculation. *J. Appl. Cryst.*, **1983**, 16:548–558.

3. Connolly, M. Solvent-accessible surfaces of proteins and nucleic acids. *Science*, **1983**, 221:709–713.

4. Norel, R.; Lin, S. L.; Wolfson, H.; Nussinov, R. Shape complementarity at protein-protein interfaces. *Biopolymers*, **1994**, 34:933–940.

5. Norel, R.; Lin, S. L.; Wolfson, H.; Nussinov, R. Molecular surface complementarity at protein-protein interfaces: the critical role played by surface normals at well placed, sparse points in docking. *J. Mol. Biol.*, **1995**, 252:263–273.

6. Norel, R.; Petrey, D.; Wolfson, H.; Nussinov, R. Examination of Shape Complementarity in Docking of *Unbound* Proteins. *PROTEINS: Structure, Function and Genetics*, **1999**. *in press.*

7. Norel R.; Lin S.L.; Xu D.; Wolfson H.; Nussinov R. Molecular Surface Variability and Induced Conformational Changes upon Protein-Protein Association. *Structure, Motion, Interaction and Expression of Biological Macromolecules*, Ed. Sarma, R.H. & Sarma, M.H., Eds. Adenine Press, Albany, N.Y., **1998**, 33–51.

8. Kuntz, I.; Blaney, J.; Oatley, S.; Langridge, R.; Ferrin, T. A geometric approach to macromolecule-ligand interactions. *J. Mol. Biol.*, **1982**, 161:269–288.

9. Jones, G.; Willet, P.; Glen, R.; Leach, A.; Taylor, R. Development and validation of a genetic algorithm for flexible docking . *J. Mol. Biol.*, **1997**, 267:727–748.

10. Hendrix, D.K.; Klein, T.E.; Kuntz, I.D. Macromolecular docking of a three-body system: The recognition of human growth hormone by its receptor. *Prot. Sci.*, **1999**, 8:1010–1022.

11. Rarey, M.; Wefing, S.; Lengauer, T. Placement of medium-sized molecular fragments into active sites of proteins. *J. Comp.-Aided Mol. Design*, **1996**, 10:41–54.

12. Xu, D.; Lin, S.; Nussinov, R. Protein binding versus protein folding: The role of hydrophilic bridges in protein associations. *J. Mol. Biol.*, **1997**, 265:68–84.

13. Besl, P. J.; McKay, N. D. A method for registration of 3-D shapes. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, **1992**, 14(2):239–256.

14. Bernstein, F.; Koetzle, T.; Williams, G.; Meyer, E.; Brice, M.; Rodgers, J., Kennard, O.; Shimanouchi, T.; Tasumi, M. The Protein Data Bank: A computer-based archival file for macromolecular structures *J. Mol. Biol.*,**1977**, 112:535–542.

15. Goodsell, D.; Olson, A. Automated docking of substrates to proteins by simulated annealing. *PROTEINS: Structure, Function and Genetics*,**1995**, 8:195–202.

16. Cherfils, J.; Duquerroy, S.; Janin, J. Protein-protein recognition analyzed by docking simulations. *PROTEINS: Structure, Function and Genetics*, **1991**, 11:271–280.

17. Jiang, F.; Kim, S. Soft docking : Matching of molecular surface cubes. *J. Mol. Biol.*, **1995**, 219:79–102.

18. Shoichet, B.; Kuntz, I. Protein docking and complementarity. *J. Mol. Biol.*, **1991**, 221:327–346.

19. Wang, H. Grid-search molecular accessible algorithm for solving the protein docking problem. *J. Comp. Chem.*, **1991**, 12:746–750.

20. Katchalski-Katzir, E.; Shariv, I.; Eisenstein, M.; Friesem, A.; Aflalo, C.; Vakser, I. Molecular surface recognition: determination of geometric fit between protein and their ligands by correlation techniques. *Proc. Natl. Acad. Sci. USA*, **1992**, 89:2195–2199.

21. Bacon, D.; Moult, J. Docking by least-square fitting of molecular surface patterns. *J. Mol. Biol.*, **1992**, 225:849–858.

22. Kasinos, N.; Lilley, G.; Subbarao, N.; Haneef, I. A Robust and efficient automated docking algorithm for molecular recognition. *Protein Engineering*, **1992**, 5(1):69–75.

23. Cherfils, J.; Janin, J. Protein docking algorithms: simulating molecular recognition. *Curr. Opin. Struct. Biol.*, **1993**, 3:265–269.

24. Helmer-Citterich, M.; Tramonato, A. Puzzle: a new method for automated protein docking based on surface shape complementarity. *J. Mol. Biol.*, **1994**, 235:1021–1031.

25. Fischer, D.; Lin, S. L.; Wolfson, H.; Nussinov, R. A geometry-based suite of molecular docking processes. *J. Mol. Biol.*, **1995**, 248:459–477.

26. Lengauer, T.; Rarey, M. Computational methods for biomolecular docking. *Curr. Opin. Struct. Biol.*, **1996**, 6:402–406.

27. Walqvist, A.; Covell, D. Docking enzyme-inhibitor complexes using a preference-based free-energy surface. *PROTEINS: Structure, Function and Genetics*, **1996**, 25:403–419.

28. Peters, K. P.; Fauck, J.; Frommel, C. The automatic search for ligand binding sites in proteins in proteins of known three-dimensional structure using only geometric criteria. *J. Mol. Biol.*, **1996**, 256:201–213.

29. Laskowski, R. A.; Luscombe, N. M.; Swindells, M. B.; Thornton, J. M. Protein clefts in molecular recognition and function. *Prot. Sci.*, **1996**, 5:2438–2452.

30. Gabb, H. A.; Jackson, R. M.; and Sternberg; M. J. E. Modelling protein docking using shape complementarity, electrostatics and biochemical information. *J. Mol. Biol.*, **1997**, 272:106–120.

31. Jackson, R.; Gabb, H.; Sternberg, M. Rapid refinement of protein interfaces incorporating solvation: application to the docking problem. *J. Mol. Biol.*, **1998**, 276:265–285.

Figure Legends

Figure 1. The effect of inflating the van der Waals radii of the ligand atoms. Two balls with different van der Waals radii are in contact with the receptor. The center of the larger ball (atom), labeled C_i is further away from the receptor than the center of the "real" vdW, C_n, of the smaller ball.

Figure 2. A summary of the results obtained in the docking of the drugs to their respective protein receptors: The RMSD of the match; The highest ranking solution with an RMSD $< 5\AA$; and the CPU, in minutes on a PC, for the matching (i.e., docking) and for the scoring.

Figure 3. An example of the effect of the different probe size in the docking of a drug. (a) 1baf docked with a $6\AA$ probe ball radius, obtaining an RMSD of $4.59\AA$. (b) 1baf docked with a $3\AA$ probe ball. Here the RMSD is $1.74\AA$. (c) and (d) show the binding site of 1baf with the solutions presented in (a) and (b) respectively. The ranking drops from 969 in (a) to 349 in (b), which still needs further improvement. The PDB complex is in green and blue. Our solution is in red. The graphics program used is RasMol V 2.5, by Roger Sayle, Biocomputing Research Unit, University of Edinburgh.

Figure 4. 1slt, another example illustrating the effect of the probe ball radius on the quality of the docking. (a) The probe ball size is $6\AA$, obtaining an RMSD of $4.57\AA$. (b) The probe ball radius is $3\AA$, obtaining an RMSD of $1.2\AA$. (c) and (d) show the binding site with the solutions presented in (a) and (b) respectively. The ranking increases from 49 in (a) to 2nd in (b). The PDB complex is in green and blue. Our solution is in red. The graphics program is as in Figure 3.

Figure 5. 1fos, a DNA–protein example. The RMSD is $3.18\AA$. Our solution is in red. The graphics program is as in Figure 3.

Table Legends

Table 1. The results obtained in the docking of the drugs. We list the PDB file name and the CPU time in minutes used in the matching step of the algorithm (running on a 586 PC clone, at 133MHz). The next columns list the CPU time used for the scoring step of the algorithm; the lowest RMSD obtained by a solution that passed the overlap test, and the best (highest) rank obtained by a solution that passed the overlap test and that obtained an RMSD $< 5\AA$. (a) Examples taken from Jones *et al.*, 1997 [8]; (b) Examples taken from Rarey *et al.*, 1996 [7]. Cases marked with '*' appear also in Table (a), as they have been employed by both groups of authors.

Table 2. Comparison of the results using the $6\AA$ and the $3\AA$ probe size in the critical points calculation for the receptor. The examples shown here are the ones with highest (worse) RMSDs in Table 1.

Table 3. The results obtained in the docking of 14 DNA–drug cases. The PDB code names, the CPU times, in minutes of the matching and the scoring, the RMSDs which are obtained and the ranking are listed in this order.