Graph-Theoretic Techniques for Macromolecular Docking

Eleanor J. Gardiner* and Peter Willett

Department of Information Studies, Sheffield University, Western Bank, Sheffield S10 2TN, United Kingdom

Peter J. Artymiuk

Department of Molecular Biology and Biotechnology, Krebs Institute, Sheffield University, Sheffield S10 2TN, United Kingdom

Received July 30, 1999

We propose a solution to the problem of docking two macromolecules. We represent each of two proteins as a set of potential hydrogen bond donors and acceptors and use a clique-detection algorithm to find maximally complementary sets of donor/acceptor pairs. Preliminary results are presented which demonstrate the feasibility of the method.

INTRODUCTION

The protein docking problem can be summarized as follows: Given two proteins, we wish to know whether they interact to form a stable complex and, if so, how?[1] This problem is fundamental to all aspects of biological function since proteins must recognize their ligands, antibodies their antigens, etc. Equally importantly, molecules must be able to discriminate between the molecules to which they should bind and all the others. The problem of whether two proteins interact can be solved experimentally, and the mode of interaction can, in favorable cases, be elucidated experimentally by structural techniques such as crystallography. However, while there are over 1000 distinct protein structures deposited in the Protein Data Bank (PDB),[2] and the number is increasing rapidly, it is more difficult to determine the structures of protein−protein complexes, and only about 100 such structures are currently in the PDB.[3] The development of reliable theoretical protein docking techniques is therefore an important goal.

The chief difficulty associated with protein docking is the complexity of the problem. Even assuming that both macromolecules are rigid, there are still six degrees of freedom (three rotational and three translational) in the orientation of one molecule relative to the other. If either (or both) of the molecules is allowed to be flexible, then the number of ways in which one could interact with another becomes rapidly, and probably infeasibly, large. To cope with this, docking routines have to make simplifying assumptions. The most common of these is that one or both molecules are assumed to be rigid. To allow for limited movement of the proteins, most algorithms allow the component molecules to be "soft" in some way, which varies from algorithm to algorithm. For example, a degree of overlap of the two proteins may be tolerated, which could not happen in a real situation. Also, it is very common for some knowledge of the receptor active site to be needed[4,5] so that the entire surface of only one protein need be considered. The second big problem is that most docking methods generate very many putative solutions. A second stage of docking, which usually consists of clustering and/or ranking these solutions, is then needed. The third problem (which is a consequence of the first two) is that docking programs can run for a very long time. The development of ever faster computers and more sophisticated algorithms has helped to reduce the impact of this problem somewhat.[6,7]

Existing docking procedures are based on the idea of some sort of "complementarity" between the two interacting molecules. Shape-based methods essentially try to find maximal regions of surface complementarity between the two molecules, while energy-based methods seek to minimize the interaction energy of the docked complex.[7,8]

Shape-based docking algorithms are posited on the "lock and key" principle in which the ligand fits into the binding site of its receptor as a key fits into a lock. This is scaled up for protein−protein interactions, to imply that "knobs" on the surface of one protein will fit into "holes" on the other, and vice versa.[1] As simple surface complementarity produces too many possible solutions, most methods either incorporate some form of other necessary conditions (e.g., matching patterns of hydrogen donors and acceptors) into the docking algorithm (as in the procedure described in this paper) or use them afterward to score the solutions. Many methods incorporate energy calculations into their scoring procedures. Most, if not all, shape-based algorithms use the molecular surface calculations of Connolly[9] or Lee and Richards,[10]and a wide range of such algorithms have been described in the literature[11−15] as have comparative evaluations using test complexes.[16,17]

In this paper we introduce a shape-based docking method which uses techniques from graph theory to find maximum sets of hydrogen bonds in each putative complex and which then screens these complexes to eliminate steric clashes. Graph theory is widely used for the representation and searching of both 2D and 3D small molecules in chemical information systems, and work at Sheffield has shown that such methods are also applicable to the processing of 3D protein structures.[18−21] Here we focus upon algorithms for

* To whom correspondence should be addressed. Phone: 0114 2222674. E-mail: e.gardiner@sheffield.ac.uk.

10.1021/ci990262o CCC: $19.00
Published on Web 12/17/1999

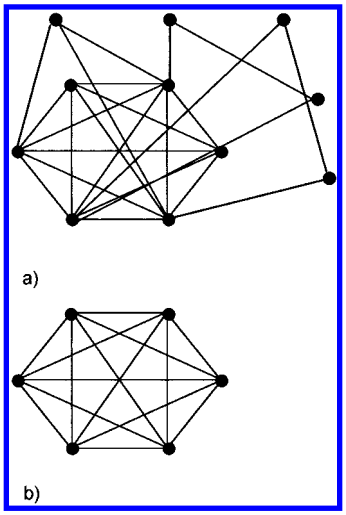

Figure 1. A graph with a maximum clique of size 6. Dots represent vertexes, and lines represent edges. (a) is a graph with 11 vertexes and 24 edges which contains a maximum clique of size 6 as shown in (b).

clique-detection. A clique of a graph, G, is a subgraph of G in which every vertex is connected to every other vertex and which is not contained in any larger subgraph with this property (see Figure 1). Brint and Willett[22] showed that of the clique-detection algorithms then available, the most efficient for finding the cliques which encode the similarity between pairs of small molecules was that due to Bron and Kerbosch.[23] Later work (for example, by Grindley et al.[19]) showed that this was also true for macromolecules, and we have used this algorithm for the work reported here.

METHOD

Our docking procedure is based upon two main aspects of protein−protein interaction, viz., shape-based complementarity and hydrogen-bonding complementarity. We have previously used clique-detection to find regions of maximum common structural similarity between two proteins even in the absence of any sequence similarity.[24,25] It is also possible to detect similarities at a local level of structure, e.g., in comparing similar groups of amino acid residues[20] or similar surfaces.[26] However, it is also clear that these methods could be adapted to match structures that are not the same but rather are complementary to each other. If attention is restricted to the surface of the proteins, and no distinction is made between the interior and exterior of the proteins, then clearly their complementary surfaces can also be regarded as having a similar shape. So, for example, if both surfaces are characterized by a series of dots, then, for those areas where the surfaces are complementary, the patterns of dots are similar. These complementary surfaces can be used to dock the proteins (see Figure 2).

Because complementary surfaces have a similar shape, it should be possible to use clique-detection to find areas of maximum common complementarity between two proteins (in a manner analogous to that used to find 3D structural similarities between pairs of proteins). However, surface similarity will find not only areas of complementarity between two proteins, but also areas of simple surface similarity. The latter can be eliminated as dockings because, if used to dock the proteins, the volumes of the proteins will

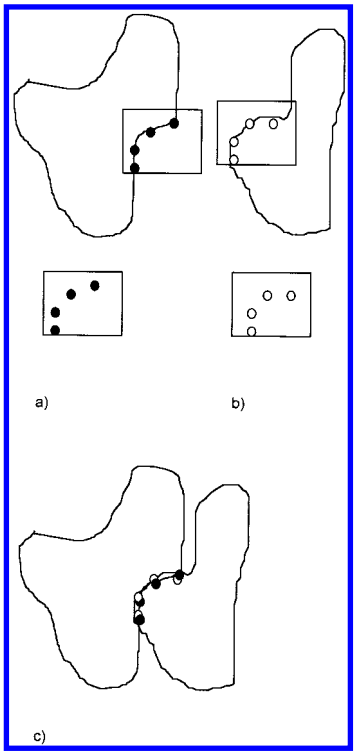

Figure 2. Complementary surfaces are similar in shape. For the two proteins (a b) the surfaces inside the squares have a similar but complementary shape as demonstrated by the pattern of dots. The proteins can be docked by pairing the dots as shown in (c).

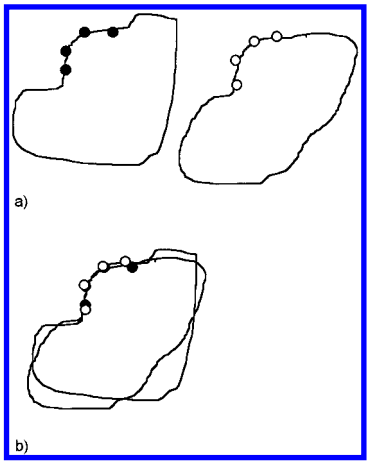

Figure 3. Docking similar surfaces. (a) These two proteins have similar shaped surfaces in the area shown by the dots. (b) When this is used to "dock" the proteins, their volumes overlap unacceptably.

overlap unacceptably, as shown in the example in Figure 3. The dockings which remain should represent areas where the surfaces of the proteins are locally complementary in shape.

Using just geometric complementarity is unlikely to produce only "correct" solutions, and thus there is also a requirement to take account of chemical complementarity: thus far we have used the hydrogen-bonding nature of solvent-accessible atoms to model this type of complementarity. The method outlined here has been described in detail by Gardiner.[27]

In proteins, nitrogen atoms and the oxygen atoms of hydroxyl groups are potential hydrogen bond donors, while all oxygen atoms and some nitrogen atoms are potential

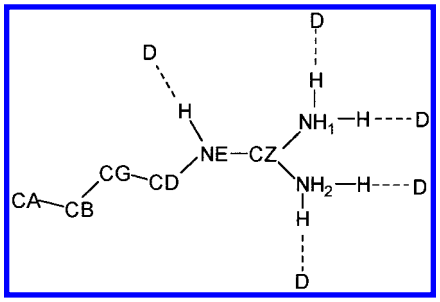

Figure 4. Example of pseudo-"D" atoms. Five pseudo-D atoms are added to the arginine side chain. Each is 2 Å from a side chain H along the line of an N−H bond.

hydrogen bond acceptors. If a hydrogen bond is to be formed then a hydrogen bond acceptor must be positioned approximately 2 Å from the hydrogen atom of a hydrogen bond donor such that the donor/ hydrogen atom/ acceptor are approximately collinear.[28] To model this using graph theory, the following procedure was adopted.

(1) Identify the solvent-accessible atoms in the two proteins that are to be docked. This is done using the program AREAIMOL.[29]

(2) Attach a pseudodonor atom 2 Å from each hydrogen of a potential hydrogen bond donor. In the case of nitrogen donors, the pseudodonors are positioned along the line of the N−H bond. Figure 4 shows the pseudodonor atoms added to the side chain of an arginine residue. As the position of the H of the hydroxyl group is variable, pseudodonor atoms are added in each possible position to the hydroxyl groups of serine, threonine, and tyrosine residues. For each protein, create a file containing the 3D coordinates of the pseudodonor and the (actual) acceptor atoms. For convenience, we can consider the atoms in the first protein's file to be labeled with capital D or A (according to whether they are donors or acceptors) and those in the second with lowercase d or a.

(3) Form the docking graph whose vertexes are the ordered pairs (D,a) and (A,d) of pseudodonors/acceptors or acceptors/pseudodonors (where the first element is from the first protein and the second from the other protein). Vertexes are then joined in the docking graph if the distance between the elements in the first protein is the same as that between the elements in the second protein to within some user-defined tolerance (for which we typically use a value of 1.2 Å). The maximal cliques of the docking graph then correspond to maximal sets of possible hydrogen bonds. Any movement of the proteins on docking is accounted for only in the size of the tolerance permitted in the distance matching.

Many of the sets of hydrogen bonds thus found will be sparsely distributed over the whole surface of the proteins, whereas the area of interaction between two proteins is usually reasonably compact. For example, Janin and Chothia[30] determined the interface area of a number of protein−protein complexes to be in the range 1200−1600 Å^2. The area in each protein is therefore on the order of 600−800 Å^2 which gives a value for the diameter on the order of 20−30 Å (assuming an approximately circular interface). Hence, an additional, user-entered, distance constraint, the *diameter*, is also imposed, and we require that the maximum distance between any two members of a set of hydrogen bond donors/acceptors be less than this diameter.

The formation of a docking graph is illustrated in Figure 5.

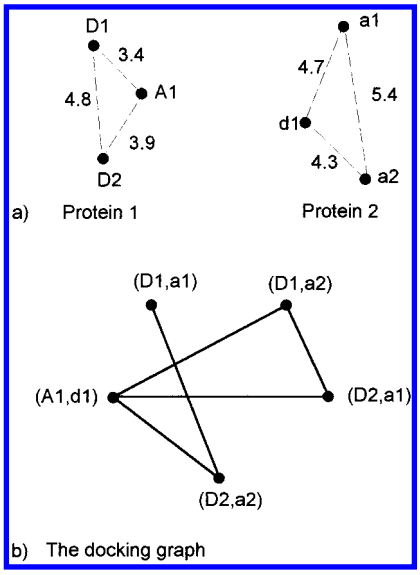

Figure 5. Forming a docking graph. (a) shows two imaginary proteins: the first with pseudodonor atoms D1 and D2 and acceptor A1 and the second with pseudodonor atom d1 and acceptors a1 and a2. The interatomic distances are marked on the dashed lines. (b) The vertexes of the docking graph are (D1, a1), (D1, a2), (D2, a1), (D2, a2), and (A1, d1). If a tolerance of 1.2 Å is used, then these vertexes will be joined as shown and a maximum clique of size 3 will be found. This clique represents a possible docking of the two proteins which is realized by superposing A1 upon d1, D1 upon a2, and D2 upon a1.

(4) A potential docking is then obtained by superposing the two proteins according to a predicted set of hydrogen bond pairs and calculating the RMSD of the result. These dockings are then screened to remove those with high RMSD and also those with interprotein contacts which are too close (typically closer than 1.5 Å).

The method is illustrated in the flowchart shown in Figure 6.

RESULTS AND ANALYSIS

To develop the program, we initially concentrated on just one example, the complementarity determining regions (CDRs) of the Fab fragment Hy/Hel-5 and its antigen hen egg white lysozyme (code 3hfl[31]). We carried out many hundreds of runs using a large number of different parameter combinations. This enabled us to obtain a working parameter set which we used as a basis for the other examples described in this paper. For the test runs, we determined the nine hydrogen-bonding pairs of 3hfl using the program DISTANG[29] and compared them with those predicted by the docking procedure. For our initial tests, "good" dockings were defined as those which predicted a majority of the correct hydrogen bonds (i.e., those identified by DISTANG). These tests demonstrated the viability of the method and also allowed appropriate values for the user-entered clique-detection parameters, tolerance and diameter, and screening parameters, RMSD and closest permitted contact, to be chosen. Table 1 shows the hydrogen bonds of the crystallized complex and also the six hydrogen bonds predicted by a good docking. Five of the predicted hydrogen bond pairs are correct. The incorrect bond is shown in the shaded row. Our program predicted that atom Thr 43 O of the lysozyme hydrogen bonds to Asn 58 ND2 of the Fab H chain, whereas

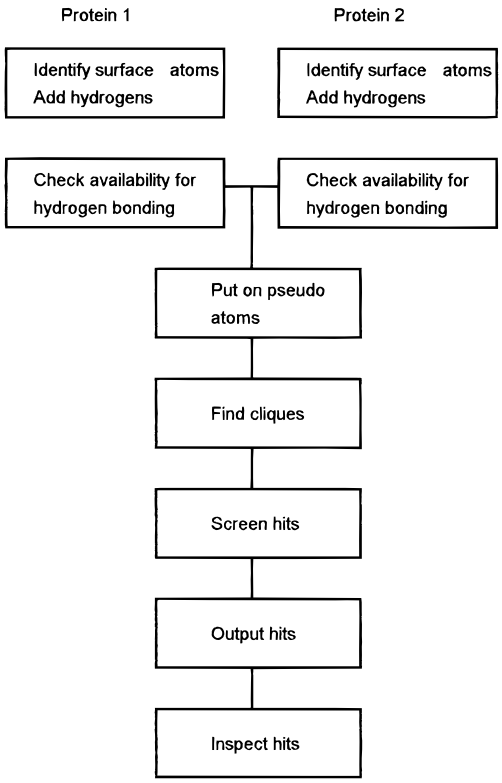

Figure 6. The docking procedure.

Table 1. Comparison of the Hydrogen Bonds of the Crystallized 3hfl Complex and Those of a Good Predicted Docking[a]

Hydrogen bonds of crystallized complex					
Lysozyme Residue	Atom	Chain	Fab Residue	Atom	Chain
Gln 41	O	Y	Ser 56	OG	H
Thr 43	OG1	Y	Asn 58	ND2	H
Tyr 53	OH	Y	Trp 33	NE1	H
Arg 45	NH1	Y	Glu 50	OE2	H
Gly 67	O	Y	Tyr 97	N	H
Asn 44	OD1	Y	Arg 93	NH2	L
Arg 45	NE	Y	Gly 92	O	L
Arg 45	NH2	Y	Gly 92	O	L
Arg 45	NH2	Y	Trp 91	O	L
Hydrogen bonds of a predicted complex					
Lysozyme Residue	Atom	Chain	Fab Residue	Atom	Chain
Gln 41	O	Y	Ser 56	OG	H
Thr 43	**O**	**Y**	**Asn 58**	**ND2**	**H**
Asn 44	OD1	Y	Arg 93	NH2	L
Arg 45	NE	Y	Gly 92	O	L
Arg 45	NH1	Y	Glu 50	OE2	H
Arg 45	NH2	Y	Trp 91	O	L

[a] The predicted docking gave an RMSD of 1.3 Å on superposition of pseudoatoms and no interprotein contact closer than 1.5 Å. The boldfaced row shows the incorrect bond predicted by the docking procedure.

Thr 43 OG1 is the atom involved in the bond in the actual complex.

The number of potential solutions found increases greatly with any relaxation in the tolerance allowed on distance matching. Our initial tests on 3hfl demonstrated that, for this complex, a tolerance of 1.2 Å allowed correct dockings to be found while the total number of predicted dockings (i.e., cliques in the docking graph) was kept within manageable limits (less than 100 000 in this case), which took 15 CPU hours running the program on a Silicon Graphics Origin 200 workstation using an 180 MHz R10000 processor.

We also experimented with different values for the diameter and found that a diameter of 18 Å was sufficiently large to contain a correct set of matching hydrogen bonds, while allowing larger values produced very many more cliques with no increase in the number (or size) of correct solutions found.

At present our work focuses on known examples to evaluate both the parameter values and the methodology. A hit produced by the clique-detection procedure is one of three types: it may be implausible (due, for example, to steric clash), it may be plausible but incorrect (i.e., the predicted complex does not resemble the crystallized complex—a false positive), or it may indeed predict at least some correct hydrogen bonds and thus position the ligand relatively correctly with respect to the protein. Clearly, it is only possible to distinguish between the last two if the answer is known. The aim of a screening procedure is to remove all of the implausible hits and as many of the false positives as possible while not rejecting correct hits. More sophisticated screening methods may then be used to rank the remaining hits.

Our present screening method is very unsophisticated. We simply remove all hits with RMSD upon superposition of hydrogen bonds which is greater than a user-entered RMSD value, and also all hits with contacts closer than a user-entered closest contact value. Tests on the 3hfl complex showed that, for this complex, hits with RMSD of more than 1.7 Å or closest contacts nearer than 1.4 Å were unlikely to be correct dockings. Longer close contacts can be tolerated since the molecules are flexible.

A set of potential dockings of the 3hfl complex is illustrated in Figure 7 which shows the Cα trace of the lysozyme molecule in yellow with a cluster of predicted orientations of the Fab and also two other isolated orientations (in magenta). The cluster of hits are close to the actual position of the Fab in the complex (which is shown in cyan), while the isolated predictions are incorrect. Most of our successful tests with 3hfl produced similar results, with multiple predictions of correct hits. The clusters of hits are produced for two reasons: first, the diameter of 18 Å means that hydrogen bond donors/acceptors further apart than this appear in different cliques; second, hydrogen bond acceptors may accept two hydrogen bonds, and these are not permitted to be in the same clique and therefore appear in two separate cliques.

The method was then further tested and developed on four more antibody/antigen complexes and also two serine protease/inhibitor complexes: in each of these cases the test proteins were obtained by separating the PDB complex into the two molecules and reorienting them. The proteins used are detailed in Table 2 and the results summarized in Table 3. In all cases, the docking graphs generated were too complex when the complete solvent-accessible surfaces of both proteins were considered. Thus, for each of the antibody/antigen complexes, only the part of the Fab containing the CDRs was docked with the antigen. In the case of the serine proteases, a region of 17 Å in radius centered on the catalytic triad was docked with the inhibitor. When the docking is performed, it is necessary to specify the minimum hit size.

Table 2. The Test Complexes[a]

complex	protein	ligand	N_p	N_l	citation
1fdl	antibody D1.3	lysozyme	46	97	Fischmann et al.[32]
1vfb	antibody D1.3	lysozyme	44	82	Bhat et al.[33]
3hfm	HyHEL-10	lysozyme	54	93	Padlan et al.[34]
1mlc	antibody D44.1	lysozyme	50	81	Braden et al.[35]
1cho	bovine α-chymotrypsin	turkey ovomucoid third domain	46	45	McPhalen and James[36]
2sni	subtilisin Novo	chymotrypsin inhibitor 2	49	51	Fujinaga et al. [37]

[a] "Complex" is the PDB code of the complex, "protein" is the larger of the two proteins, "ligand" is the smaller, and N_p and N_l are the number of protein residues and number of ligand residues used in the docking. The 1fdl and 1vfb complexes are composed of the same proteins, but 1vfb is crystallized to higher resolution.

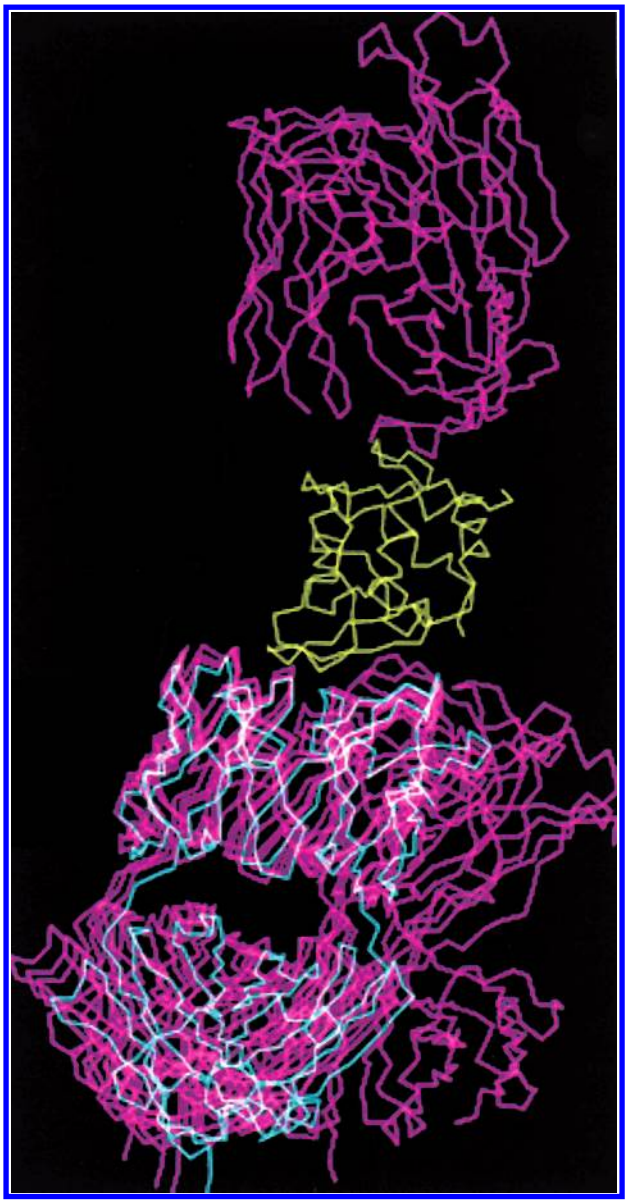

Figure 7. Docking the 3hfl complex. Predicted dockings of the lysozyme with the Fab superposed onto the actual complex structure (3hfl). The Cα trace of the crystallized 3hfl complex is shown with the lysozyme molecule in yellow and the Fab in cyan. The Cα traces of the predicted Fabs are magenta. Three can be seen to be close to the correct orientation.

This is used to control the volume of output as, typically, reducing the minimum hit size by one results in a 100-fold increase in the number of hits. We therefore start the docking with the minimum hit size set large (usually set at 10) and then repeat the dockings, decreasing the value by one each time until hits are found which pass all the screening tests.

Table 3. Results of Docking Bound Complexes[a]

complex	V	T	$N_{cliques}$	ω	hits	good	HB
1fdl	36 451	151	3 829 119	13	52	12	8
1vfb	28 101	47	104 674	11	2	1	8
3hfm	40 017	23	86 587	11	4	2	8
1mlc	34 001	56	802 176	13	19	0	
1cho	11 942	4	887 271	10	7	6	6
2sni	14 477	7	20 157	9	8	5	8

[a] V is the number of vertices of the docking graph, T is the time taken for clique-detection (CPU hours), $N_{cliques}$ is the number of cliques found of minimum size (i.e., the largest clique size at which correct hits were found, except for 1mlc where the minimum of 6 was the smallest clique size that was feasible to find), ω is the size of a maximum clique, "hits" is the number of potential dockings which pass the closest contact test, "good" is the number of these of which the majority of the predicted hydrogen bonds are found in the crystallized complex, and HB is the total number of hydrogen bonds predicted by a good hit.

The timings given are for a C program run on a Silicon Graphics Origin 200 workstation using an 180 MHz R10000 processor.

One of the aims of these further tests was to establish whether the parameters used in the 3hfl tests could be satisfactorily transferred to the docking of other complexes. Thus, in docking each of these complexes, we began by using the same parameter values (distance tolerance 1.2 Å, diameter 18 Å, RMSD < 1.7 Å, and no contact closer than 1.4 Å), but if no hits were found (as proved to be the case for 1vfb) or several hits with no clusters were found (as in the first dockings of 1fdl), then the parameters were varied until satisfactory sets of hits were found. In the case of 1fdl, the diameter was increased to 20 Å, while in the case of 1vfb the closest permitted contact distance was reduced to 1.3 Å and the maximum RMSD was increased to 1.8 Å. For both complexes the changes required are small (on the order of 10% of the parameter value used when 3hfl is docked).

As Table 3 shows, the docking process was generally successful using coordinates taken from the complex, although some problems were found. It proved impossible to dock the antibody/antigen complex, 1mlc, unless attention was restricted to the known binding residues of the antibody. This was due to the comparatively few hydrogen bonds present in this complex, which meant that the correct set of predicted hydrogen bonds was lost among the many incorrect ones. The run times for some of the complexes were very long (up to one week), which is verging on the unacceptable, and for the 1fdl complex over three million cliques were found that contained more than eight vertexes.

A fundamentally more difficult problem is that of docking native proteins. The reason for this is that there may be

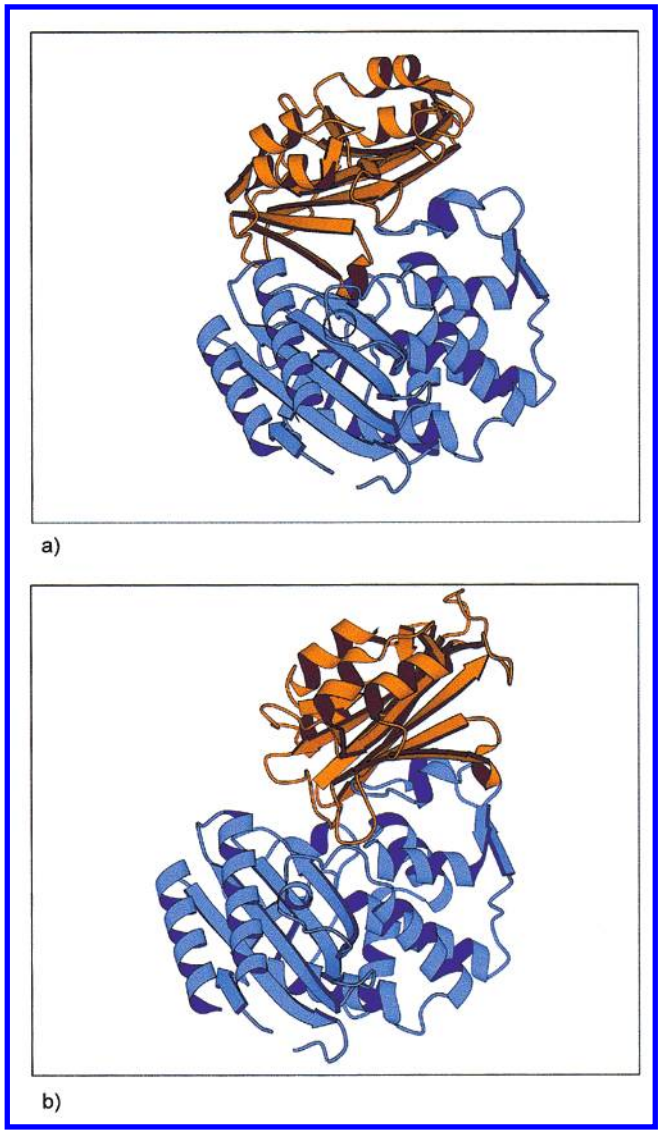

Figure 8. Ribbon trace of predicted TEM−1/BLIP complexes. TEM-1 is shown in blue and BLIP in orange. (a) Although contacts are plausible, this orientation does not match the correct complex. (b) Although the contacts are less plausible (some closer than 1 Å), this orientation is an approximation to the crystallized complex.

unforeseen movements in side chains, or even a major conformational change, relative to the native structures. In addition, it is also possible that the native structures themselves, especially side chain positions, may be perturbed by crystal contacts[38] and therefore deviate from the solution structure. We attempted to dock the native β-lactamase TEM-1[39] with the native β-lactamase inhibitory protein BLIP.[40] These are two proteins which were featured in a docking challenge, and their correct mode of interaction was predicted by all six challenge entrants.[16] TEM-1 and BLIP are two relatively large proteins. It was therefore necessary to use only part of one of the proteins for the docking procedure to be computationally feasible. It was decided to assume that BLIP would inhibit TEM-1 by obstructing the known TEM-1 active site. All residues within a radius of 15 Å of the TEM-1 nucleophile Ser 70 OG[39] were selected, and only these were docked against the inhibitor.

The only predicted docking with RMSD < 2 Å and no contacts closer than 1.2 Å yielded the apparently plausible solution shown in Figure 8a: however, this solution does not match the actual complex. No correct hits were found by the docking procedure until the docking was restricted to known binding residues of both molecules. When docking was performed using only these residues, the orientation shown in Figure 8b was found. This complex resembles the crystallized complex, despite the presence of some contacts which are closer than 1 Å. The main reasons for the failure of this docking are the conformational changes that take place in both proteins on binding. These changes, although small in global terms (the RMSD between the native and complexed main-chain atoms of TEM-1 was 0.35 Å, while for BLIP it was 0.9 Å[41]), mean that some individual side chains which are strongly involved in binding move a good deal. Thus, for TEM-1, the ring of Tyr 105 (which forms part of four hydrogen bonds in the complex) moves "considerably" [41] as does the side chain of Asp 49 of BLIP (which is also involved in four hydrogen bonds in the complex). Our docking program attempts to deal with movement of the protein by using tolerances on distance matches. The problem then is that, for movements of this size, the tolerances are so high that many millions of other matches are also allowed. In fact it is computationally infeasible to use tolerances which are high enough when more than just the binding residues of the proteins are docked. Some docking challenge entrants succeeded[16] where we failed because their methods were based largely on more gross shape matching procedures that were less affected by individual atom movements.[8,42,43]

A second problem is that our assumption that the proteins are rigid means that proposed enzyme−inhibitor contacts are disallowed because they are too close. Thus, in fact, the movement of the Tyr 105 (of TEM-1) allows BLIP residues to be positioned so that Asp 49 can make its hydrogen bonds,[41] but any such docking we find is rejected on the grounds that the enzyme and inhibitor make contacts which are too close. To relax the close contact criterion means that too many incorrect dockings survive our screening process.

SUMMARY

We have developed a novel graph-theoretic technique for protein docking. The preliminary results presented here demonstrate the feasibility of the method, although some severe limitations are also apparent, most notably its failure to dock native proteins. We expect that future developments of the program will ameliorate some of these limitations. In particular, we are experimenting with "colouring" the hydrogen bond donor and acceptor atoms by labeling each with characteristics of its local environment such as local shape and charge. In forming the docking graph, we then do not create vertexes where the donor/acceptor pair have similar colors. Initial tests show that this approach reduces the size of the docking graph by approximately 50%, thus allowing higher tolerances to be used.

There is also a requirement for more sophisticated screening of potential hits. We are currently investigating electrostatic and buried surface area screens.

ACKNOWLEDGMENT

We thank Professor M. G. James for supplying us with the coordinates of TEM-1 and BLIP, the reviewers for their comments on the first version of this paper, the Engineering and Physical Sciences Research Council and the Biotech-

nology and Biological Sciences Research Council for funding this research, and the Biotechnology and Biological Sciences Research Council and Tripos Inc. for computing resources. The Krebs Institute is a designated Biotechnology and Biological Sciences Research Council Biomolecular Sciences Centre.

REFERENCES AND NOTES

(1) Connolly, M. L. Shape complementarity at the hemoglobin $\alpha 1\beta 1$ subunit interface. *Biopolymers* **1986**, *25*, 1229−1247.
(2) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F. J.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, M.; Tasumi, M. The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* **1977**, *112*, 535−542.
(3) Gabb, H. A.; Jackson, R. M.; Sternberg, M. J. E. Modelling protein docking using shape complementarity, electrostatics and biochemical information. *J. Mol. Biol.* **1997**, *272*, 106−120.
(4) Bacon, D. J.; Moult, J. Docking by least-squares fitting of molecular-surface patterns. *J. Mol. Biol.* **1992**, *225*, 849−858.
(5) Kasinos, N.; Lilley, G. A.; Subbarao, N.; Haneef, I. A robust and efficient automated docking algorithm for molecular recognition. *Protein Eng.* **1992**, *5*, 69−75.
(6) Shoichet, B. K.; Kuntz, I. D. Molecular docking using shape descriptors. *J. Comput. Chem.* **1992**, *13*, 380−397.
(7) Hart, T. N.; Read, R. J. In *The protein folding problem and tertiary structure prediction*; Merz, K., Jr., Le Grand, S., Eds.; Birkhauser: Boston, 1994; pp 77−108.
(8) Janin, J. Protein−protein recognition. *Prog. Biophys. Mol. Biol.* **1996**, *64*, 145−166.
(9) Connolly, M. L. Solvent-accessible surfaces of proteins and nucleic acids. *Science* **1983**, *221*, 708−713.
(10) Lee, B.; Richards, F. M. The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* **1971**, *55*, 379−400.
(11) Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Perrin, T. E. A geometric approach to macromolecule-ligand interactions. *J. Mol. Biol.* **1982**, *161*, 269−288.
(12) Shoichet, B. K.; Kuntz, I. D. Protein docking and complementarity. *J. Mol. Biol.* **1991**, *221*, 327−346.
(13) Jiang, F.; Kim, S. H. Soft docking−matching of molecular-surface cubes. *J. Mol. Biol.* **1991**, *219*, 79−102.
(14) Katchalski-Katzir, E.; Shariv, I.; Eisenstein, M.; Friesem, A. A.; Aflalo, C.; Vakser, I. A. Molecular-surface recognition−determination of geometric fit between proteins and their ligands by correlation techniques. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 2195−2199.
(15) Jackson, R. M.; Gabb, H. A.; Sternberg, M. J. E. Rapid refinement of protein interfaces incorporating solvation: Application to the docking problem. *J. Mol. Biol.* **1998**, *276*, 265−285.
(16) Strynadka, N. C. J.; Eisenstein, M.; KatchalskiKatzir, E.; Shoichet, B. K.; Kuntz, I. D.; Abagyan, R.; Totrov, M.; Janin, J.; Cherfils, J.; Zimmerman, F.; Olson, A.; Duncan, B.; Rao, M.; Jackson, R.; Sternberg, M.; James, M. N. G. Molecular docking programs successfully predict the binding of a beta-lactamase inhibitory protein to TEM-1 beta-lactamase. *Nat. Struct. Biol.* **1996**, *3*, 233−239.
(17) Dixon, J. S. Evaluation of the CASP2 docking section. *Proteins: Struct., Funct., Genet.* **1997**, 198−204.
(18) Mitchell, E. M.; Artymiuk, P. J.; Rice, D. W.; Willett, P. Use of techniques derived from graph-theory to compare secondary structure motifs in proteins. *J. Mol. Biol.* **1990**, *212*, 151−166.
(19) Grindley, H. M.; Artymiuk, P. J.; Rice, D. W.; Willett, P. Identification of tertiary structure resemblance in proteins using a maximal common subgraph isomorphism algorithm. *J. Mol. Biol.* **1993**, *229*, 707−721.
(20) Artymiuk, P. J.; Poirrette, A. R.; Grindley, H. M.; Rice, D. W.; Willett, P. A graph-theoretic approach to the identification of 3-dimensional patterns of amino acid side-chains in protein structures. *J. Mol. Biol.* **1994**, *243*, 327−344.
(21) Artymiuk, P. J.; Grindley, H. M.; Poirrette, A. R.; Rice, D. W.; Ujah, E. C.; Willett, P. Identification of beta-sheet motifs, of psi-loops, and of patterns of amino acid-residues in 3-dimensional protein structures using a subgraph-isomorphism algorithm. *J. Chem. Inf. Computer Sci.* **1994**, *34*, 54−62.
(22) Brint, A. T.; Willett, P. Algorithms for the identification of 3-dimensional maximal common substructures. *J. Chem. Inf. Comput. Sci.* **1987**, *27*, 152−158.
(23) Bron, C.; Kerbosch, J. Finding all cliques of an undirected graph. *Commun. ACM* **1973**, *16*, 575−577.
(24) Artymiuk, P. J.; Grindley, H. M.; Park, J. E.; Rice, D. W.; Willett, P. 3-Dimensional structural resemblance between leucine aminopeptidase and carboxypeptidase-A revealed by graph-theoretical techniques. *FEBS Lett.* **1992**, *303*, 48−52.
(25) Artymiuk, P. J.; Poirrette, A. R.; Rice, D. W.; Willett, P. A polymerase I palm in adenylyl cyclase? *Nature* **1997**, *388*, 33−34.
(26) Poirrette, A. R.; Artymiuk, P. J.; Rice, D. W.; Willett, P. Comparison of protein surfaces using a genetic algorithm. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 557−569.
(27) Gardiner, E. J. Computational analysis of protein binding sites and surfaces: A computational analysis of biological and chemical graphs. *Ph.D. Thesis*, University of Sheffield, Sheffield, U.K., 1999.
(28) Baker, E. N.; Hubbard, R. E. Hydrogen bonding in globular proteins. *Prog. Phys. Mol. Biol.* **1984**, *44*, 97−179.
(29) Collaborative Computational Project Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr., Sect. D* **1994**, *50*, 760−763.
(30) Janin, J.; Chothia, C. The structure of protein−protein recognition sites. *J. Biol. Chem.* **1990**, *265*, 16027−16030.
(31) Cohen, G. H.; Sheriff, S.; Davies, D. R. Refined structure of the monoclonal antibody HyHEL-5 with its antigen hen egg-white lysozyme. *Acta Crystallogr., Sect. D* **1996**, *52*, 315−326.
(32) Fischmann, T. O.; Bentley, G. A.; Bhat, T. N.; Boulot, G.; Mariuzza, R. A.; Phillips, S. E. V.; Tello, D.; Poljak, R. J. Crystallographic refinement of the 3-dimensional structure of the Fab d1.3-lysozyme complex at 2.5 Angstrom resolution. *J. Biol. Chem.* **1991**, *266*, 12915−12920.
(33) Bhat, T. N.; Bentley, G. A.; Boulot, G.; Greene, M. I.; Tello, D.; Dallacqua, W.; Souchon, H.; Schwarz, F. P.; Mariuzza, R. A.; Poljak, R. J. Bound water-molecules and conformational stabilization help mediate an antigen−antibody association. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 1089−1093.
(34) Padlan, E. A.; Silverton, E. W.; Sheriff, S.; Cohen, G. H.; Smithgill, S. J.; Davies, D. R. Structure of an antibody antigen complex−crystal-structure of the Hyhel-10 Fab-lysozyme complex. *Proc. Natl. Acad. Sci. US.A.* **1989**, *86*, 5938−5942.
(35) Braden, B. C.; Souchon, H.; Eisele, J. L.; Bentley, G. A.; Bhat, T. N.; Navaza, J.; Poljak, R. J. 3-Dimensional structures of the free and the antigen-complexed Fab from monoclonal antilysozyme antibody-D44.1. *J. Mol. Biol.* **1994**, *243*, 767−781.
(36) McPhalen, C. A.; James, M. N. G. Structural comparison of 2 serine proteinase protein inhibitor complexes−eglin-c-subtilisin Carlsberg and ci-2-subtilisin Novo. *Biochemistry* **1988**, *27*, 6582−6598.
(37) Fujinaga, M.; Sielecki, A. R.; Read, R. J.; Ardelt, W.; Laskowski, M.; James, M. N. G. Crystal and molecular-structures of the complex of alpha-chymotrypsin with its inhibitor turkey ovomucoid 3rd domain at 1.8 Angstrom resolution. *J. Mol. Biol.* **1987**, *195*, 397−418.
(38) Kossiakoff, A. A.; Randal, M.; Guenot, J.; Eigenbrot, C. Variability of conformations at crystal contacts in bpti represent true low-energy structures−correspondence among lattice packing and molecular-dynamics structures *Proteins: Struct., Funct., Genet.* **1992**, *14*, 65−74.
(39) Strynadka, N. C. J.; Adachi, H.; Jensen, S. E.; Johns, K.; Sielecki, A.; Betzel, C.; Sutoh, K.; James, M. N. G. Molecular-structure of the acyl-enzyme intermediate in beta-lactam hydrolysis at 1.7 Angstrom resolution. *Nature* **1992**, *359*, 700−705.
(40) Strynadka, N. C. J.; Jensen, S. E.; Johns, K.; Blanchard, H.; Page, M.; Matagne, A.; Frere, J. M.; James, M. N. G. Structural and kinetic characterization of a beta-lactamase-inhibitor protein. *Nature* **1994**, *368*, 657−660.
(41) Strynadka, N. C. J.; Jensen, S. E.; Alzari, P. M.; James, M. N. G. A potent new mode of beta-lactamase inhibition revealed by the 1.7 Å X-ray crystallographic structure of the TEM-1-BLIP complex. *Nat. Struct. Biol.* **1996**, *3*, 290−297.
(42) Shoichet, B. K.; Kuntz, I. D. Predicting the structure of protein complexes: A step in the right direction. *Chem. Biol.* **1996**, *3*, 151−156.
(43) Sternberg, M. J. E.; Gabb, H. A.; Jackson, R. M. Predictive docking of protein−protein and protein-DNA complexes. *Curr. Opin. Struct. Biol.* **1998**, *8*, 250−256.

CI990262O