PROTEINS: Structure, Function, and Genetics 20:320–329 (1994)

Hydrophobic Docking: A Proposed Enhancement to Molecular Recognition Techniques

Ilya A. Vakser[1] and Claude Aflalo[2]

[1]Department of Membrane Research and Biophysics, Weizmann Institute of Science, Rehovot 76100, Israel and [2]Department of Life Sciences, Ben Gurion University, Beer Sheva 84105, Israel

ABSTRACT **In the classical procedures for predicting the structure of protein complexes two molecules are brought in contact at multiple relative positions, the extent of complementarity (geometric and/or energy) at the surface of contact is assessed at each position, and the best fits are retrieved. In view of the higher occurrence of hydrophobic groups at *contact sites*, their contribution results in more intermolecular atom–atom contacts per unit area for correct matches than for false positive fits. The hydrophobic groups are also potentially less flexible at the surface. Thus, from a practical point of view, a partial representation of the molecules based on hydrophobic groups should improve the quality of the results in finding molecular recognition sites, as compared to full representation. We tested this proposal by applying the idea to an existing geometric fit procedure and compared the results obtained with full vs. hydrophobic representations of molecules in known molecular complexes. The hydrophobic docking yielded distinctly higher signal-to-noise ratio so that the correct match is discriminated better from false positive fits. It appears that nonhydrophobic groups contribute more to false matches. The results are discussed in terms of their relevance to molecular recognition techniques as compared to energy calculations.**

Key words: protein recognition, hydrophobicity, macromolecular surface complementarity, docking algorithm

INTRODUCTION

The concepts of structural homology and complementarity between proteins are well established now among researchers in different fields of molecular biology. Various experimental strategies are being designed and implemented to identify or demonstrate interaction between macromolecules.[1] The basis for such associations is, in principle, reducible to physicochemical rules. This can be validated by considering the atomic structures of macromolecular complexes available in databanks. However, in view of the limited number of such resolved complexes, obtained mostly after cocrystallization of the components, predictive tools are needed to evaluate the odds for interaction between given molecular species suspected to interact, and furthermore to determine the sites at which they interact. Thus, computer modeling approaches to molecular recognition represent a rapidly developing area in recent years.

The basic formalism to the molecular recognition problem may be shortly stated as follows. Two molecules with known 3-D structures (including the identity of all atoms and their coordinates) are given. In the most general case, no additional knowledge is available. The problem is to bring these molecules in contact in order to get maximal compatibility at their respective interacting surfaces. This requires an exhaustive search in a multidimensional space upon which a large number of relative positions (rotation and translation), and/or conformations of the molecules are tested for fit at the contact area. Existing algorithms for molecular recognition may roughly be separated into procedures based on minimization of the energy of interaction,[2–6] and those based on a search for geometric surface complementarity.[7–13] Both approaches usually consider the molecules as rigid bodies, which seems justified for most cases of known complexes involving large molecules.[14,15] However, a recent trend is emerging in molecular recognition procedures,[16,17] which explicitly takes into account multiple conformations of smaller ligands (e.g., oligopeptides).[18,19]

Geometric surface complementarity has been recently found useful as a simplifying approach to efficiently scan and retrieve potential sites for interaction between molecules. It is based essentially on maximizing intermolecular surface contact while avoiding overlap. The major difficulties usually encountered in this approach are as follows:

1. The procedures generate a large number of relative positions which represent quantitatively adequate solutions to the search for geometric fit. This is equivalent to the "multiminima problem" in en-

Received March 28, 1994; revision accepted August 16, 1994.

Address reprint requests to Dr. Ilya A. Vakser at his present address: Center for Molecular Design, Washington University, Lopata Hall, Box 1099, One Brookings Drive, St. Louis, MO 63130.

ergy calculations. Thus, the correct match appears among many different "false" solutions of comparable competency, even as assessed by different methods.[20]

2. The approach is based on the assumption of rigid bodies, and is expected to be sensitive to local conformational changes which may occur upon complex formation. This is diagnosed by the failure in matching two molecules using the structures of the free components, while the correct match is more readily found when the respective structures in a dissociated complex are used.[3,13,20]

3. In many algorithms, the computation time is strongly dependent on the number of atoms (or residues) in the molecules. It is increased when their physicochemical properties, used for proper representation of the molecules, are taken into account for the calculation of the fit itself. Most recognition procedures become inapplicable in the case of large macromolecular entities, because of the excessive amount of information to be processed.

Ideally, any procedure should take into account the minimal amount of information relevant to recognition, leaving out less important, or less defined details which may generate "noise" in the quantitation of the fit between molecules. The detailed atomic structure of macromolecules contains a huge amount of information, which not only puts a heavy load on computation, but may also be detrimental to the specific task of finding recognition sites. Indeed, the position of atom groups at the surface of macromolecules may vary according to their environment. For example, the relative location of a charged residue in water may be different than that in a crystal lattice, or at the interface of a bimolecular complex. Moreover, the microenvironment of such a relocated group will in turn affect the position of neighboring atom groups. Thus, this mutual relation between the environment and the position of residues or atoms at the surface of a macromolecule makes the "rigid body" assumption untenable when the positional information, available from a single source (e.g., data bank entry), is used strictly.[19] This difficulty is prominent in the general case where the structures of the native components are well defined under one set of conditions, but little is known of their interaction in different circumstances.

It appears that due to potential flexibility of atom groups at the surface of macromolecules, the relevance of their exact position in a crystal to the calculations designed to identify recognition sites may be questionable.[14] Attempts to circumvent this problem have been implemented and tested by allowing some degree of overlapping at the contact site,[14] a "soft" docking,[10] or some "tolerance"[13] in the calculation of geometric fit. However, in many cases the global conformation of macromolecules does not change significantly upon complex formation. Thus, the changes at the surface must be local and attributable to only some atom groups which must rearrange at the interface. One must identify these flexible groups and discriminate them from stiffer groups at the surface by assigning less weight to their contribution in the geometric or energy calculations. This would confer some statistical "fuzziness" to the surfaces of macromolecules in their representation.[21] The position and potential for flexibility of an atom group are determined by all interactions between this group and its environment. The physicochemical characteristics of intermolecular interactions (van der Waals, electrostatic, hydrogen bonding, hydrophobic effects, etc.) have been the subject of intensive investigation. The two major elements traditionally used in molecular recognition approaches are steric and electrostatic properties. A third element, hydrophobicity, has recently been explicitly introduced in procedures for protein folding[22,23] and the estimation of 3-D quantitative structure–activity relationships (QSAR).[24,25] The relative contribution of each kind of interaction to the spatial and energetic relation between molecules in a complex is still a matter of controversy. In this paper we have no intention to take part in this discussion, but rather explore a basic approach for improving the performance of general molecular recognition techniques. The results of a geometric procedure based on collision checks and complementarity of full represented molecular surfaces will be compared to these obtained with the same procedure, but using a partial description of the molecules to be matched.

RATIONALE

Our purpose is to match two molecules of known 3-D structure in a crystal and predict their relative position upon their association as solvated species in an aqueous solution. In the case of macromolecules, the surface density of hydrophobic atoms should generally be higher at the binding site than elsewhere on the surface. This thermodynamically sound assumption is supported by the results of a systematic survey of known complexes,[26] indicating that on the average (1) a substantial complementarity between hydrophobic groups is manifest at the area of contact, and (2) the hydrophobicity of the interface is higher than that elsewhere on the surface.

Based on these premises, we propose to exploit the hydrophobicity of surface groups on the molecules to be matched. We separate atom groups into hydrophobic (aromatic C and CH; aliphatic CH, CH_2, and CH_3; and sulfur) and "nonhydrophobic" (C^{α} and carbonyl C, all N, all O). Nonhydrophobic atoms in both polar and nonpolar residues are essentially excluded from consideration in the process of matching surfaces. We look for geometric complementarity be-

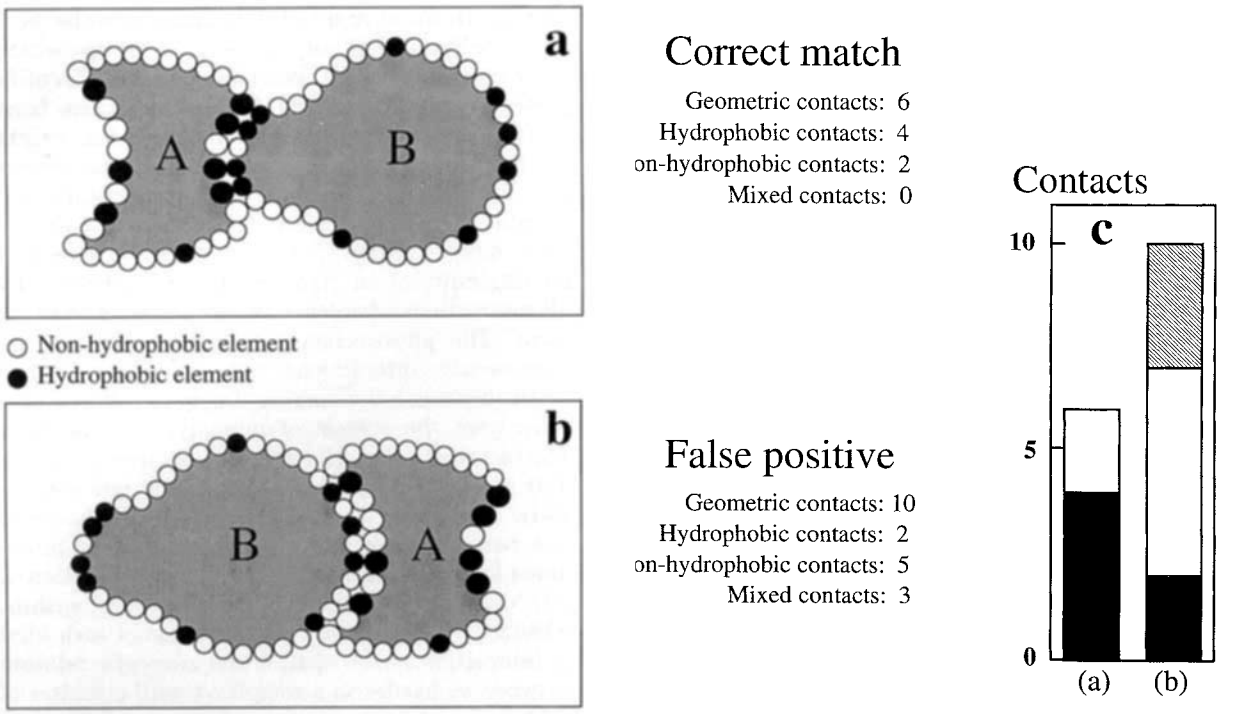

Fig. 1. Schematic illustration of two hypothetical association modes in a bimolecular complex. In the "correct match" **(a)**, hydrophobic groups are statistically more represented at the surface of contact. The "false positive" configuration **(b)** has lower density of hydrophobic groups and mixed pairing at the interface. The number and quality of the contacts in each case are depicted as a stacked histogram **(c)**, where the closed, open, and hatched bars represent hydrophobic, nonhydrophobic, and mixed contact, respectively.

tween molecules whose surfaces are represented by the hydrophobic atom groups only.

The rationale of this approach is illustrated in Figure 1. In the correct match (Fig. 1a) a high complementarity between the surfaces in contact applies, which involves both geometric fit and qualitative match of paired surface elements. However, when no distinction is made between surface elements (as in most geometric matching procedures), some orientations may yield contact scores of similar values or even higher than the correct match (see refs. 12 and 20). These false positive fits (Fig. 1b) are due to the contribution of mismatches (unaccounted as such) at the surface of contact. The selective elimination of one class of elements results on the one hand in a general reduction of the number of contacts, but on the other hand, in a more specific evaluation of complementarity (see Fig. 1c). Instead of accounting for the chemical features of each atom,[27] we chose to eliminate radically the nonhydrophobic elements from protein surfaces. In solution, these elements are assumed to be in contact with the aqueous solvent mainly, and thus are relatively free to move in a restricted space at the interface. On the other hand, hydrophobic groups are relatively "anchored" at the surface since they interact with the water-excluded core, which represents the rigid skeleton of the molecule. The elimination of any atom from the molecular surface would leave a gap which provides more freedom to accommodate local conformational changes at the surface of contact between the molecules. Thus, the space originally occupied by hydrophilic atom groups at the surface of one molecule in a crystal would be available for rearrangement of interacting hydrophilic atom groups from both of the molecules in the process of complex formation. Since hydrophobic groups are statistically more represented in contact sites (see above), they should also contribute relatively more intermolecular atom–atom contacts per unit area for the correct matches than for the false positive fits.

Such an approach should partially relieve the three major difficulties of geometric molecular recognition mentioned above as stated below.

1. Problem of false solutions. The elimination of nonhydrophobic patches in the representation of the molecular surface, while decreasing the quality of the match, should affect more the false positive fits than the correct match (see Fig. 1c). Therefore, considering the score for a good geometric fit as a signal, the proposed change must improve the "signal-to-noise ratio."

2. Problem of conformational changes. Assuming

TABLE I. Values Assigned to Points in the Grid Representation of the Molecules*

	Geometric surface recognition				Hydrophobic surface recognition			
	Mol. A			Mol. B	Mol. A			Mol. B
Grid point	Core	Surf 1	Surf 2	All	Core	Surf 1	Surf 2	All
Hydrophobic†	−10	1	0.5	1	−5	1	0–0.5	1
Nonhydrophobic‡	−10	1	0.5	1	−5	0	0	0

*Mol., molecule; Surf 1, first surface layer (closest to the core); Surf 2, second surface layer.
†Grid points included in the space (defined in the text) occupied by the following atom groups: aromatic C and CH, aliphatic CH, CH_2, and CH_3, and sulfur.
‡Grid points included in the space belonging to the molecule which does not overlap with hydrophobic atom groups.

that nonhydrophobic groups at the surface are in general more exposed to the aqueous solvent, one may consider them as more flexible. As such, they are potentially subject to small conformational changes upon their interaction with another molecule. Thus, taking into account only the more rigid hydrophobic groups, leaving some room at the surface, would contribute to a better tolerance for local conformational changes involving more the hydrophilic groups.

3. Problem of atoms number. The number of atoms is reduced by eliminating nonhydrophobic groups. Therefore, the number of operations performed by a recognition procedure should be reduced accordingly.

Similar "hydrophobic/nonhydrophobic" representations of molecules by either atoms or residues have been used and applied to the problem of peptide–receptor interaction,[28] or to protein folding,[23] respectively. In light of the above, this approach may be considered as a potential improvement for many molecular recognition procedures. We tested it on the algorithm for geometric docking developed in our group[13] with appropriate modifications to implement the hydrophobic docking approach.

METHODS

Geometric Surface Recognition Algorithm

The algorithm developed previously[13] applies pattern recognition techniques. Briefly, the atoms in the molecules, represented as spheres of given radii, are first projected on a 3-D grid so that each grid point included in a sphere is assigned a nonzero value (see below). One or two surface layers for the first molecule (molecule A) are then generated by sequential projections of the atoms with radii expanded by one or two grid steps, respectively. These layers represent a volume available for interaction with the second molecule (molecule B). The interior of molecule A actually represents a volume which is not available to molecule B. The surface and interior of molecule A are differentiated by assigning different values to the grid points in each region, but these in molecule B are not (see Table I). Next, the relative orientation of molecule B is systematically varied by rotation around three axes, and projected again. A three-dimensional correlation function is calculated at each orientation, using Fourier transformation and multiplication of the 3-D arrays representing the molecules. The values in the inverse Fourier transform of the resulting 3-D array assess the degree of molecular surface contact (positive contribution) and penetration (negative contribution) for each relative shift of the molecules in the three translational directions (for details see ref. 13). Thus, each of these scores, representing the algebraic sum of contributions, is associated with six numbers describing the relative translation and rotation of the molecules. The algorithm provides a sorted list of values (scores) indicating the extent of geometric match between the surfaces of the molecules at different relative positions. The procedure is equivalent to a systematic search of six degrees of freedom, but faster by design.

Hydrophobic Surface Recognition Algorithm

The geometric surface recognition procedure was minimally modified in order to eliminate nonhydrophobic atom groups in the 3-D grid representation of the molecules. While molecule B is represented by its hydrophobic atoms only (see Fig. 2b), we chose to exclude the nonhydrophobic atom groups selectively from the surface layers of molecule A, preserving a full representation of the interior (see Fig. 2a), to allow for a partial assessment of collisions. Before the projection on the grid, atoms are sorted so that the hydrophobic ones are projected last, and thus confer to the grid points upon which they are projected a "hydrophobic character," even though the same points might also be associated with other nonhydrophobic atoms. This step ensures that a grid point of molecule B or at the surface of molecule A associated with at least one hydrophobic atom group retains its identity as such. In summary, only hydrophobic atoms of both molecules will contribute to molecular surface contact, while collisions will be checked between hydrophobic atoms of molecule B and all atoms of molecule A. This situation is intermediate between the strict assessment of collisions with all atoms, as in the geometric algorithm above,

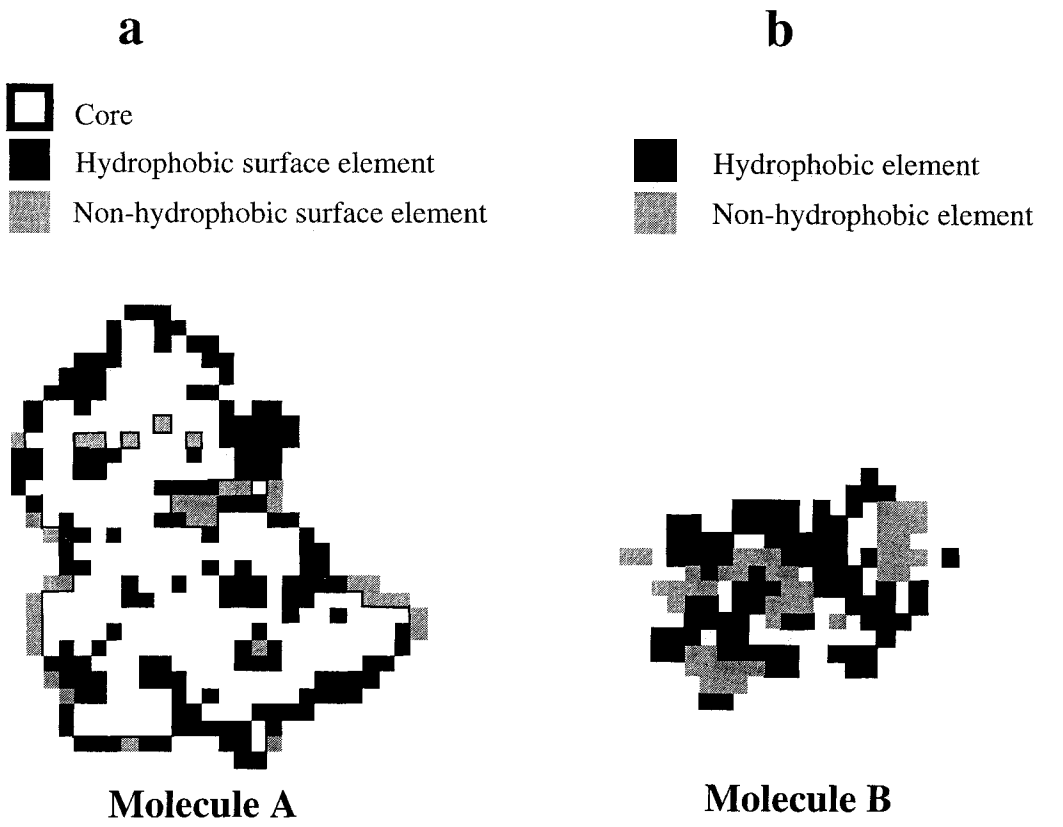

Fig. 2. Typical cross section through the 3-D grid representations of the molecules. Molecule A is trypsin, molecule B is trypsin inhibitor. The relative positions of the molecules were arbitrarily chosen. The gray areas represent nonhydrophobic surface patches which are eliminated in the new representation of the molecules in hydrophobic docking.

and a more tolerant collision check involving only hydrophobic atoms in both molecules. Except for the reduced representation of the molecules, the computation remains identical to that described for geometric docking.

Implementation of the Algorithm

The size of the 3-D grid was chosen as $64 \times 64 \times 64$ giving a grid step in the range 1.5–1.7 Å, which represents a coarser resolution than that previously used ($90 \times 90 \times 90$).[13] This considerably speeds up the computation procedure. For the rotation of molecule B, we used 2340 different orientations separated by an angular step of 20°. At each orientation, the three highest-score relative positions (denoted as peaks) are stored and saved for further sorting.

In order to preserve the quality of the results at the new (lower) resolution, we had to choose a new set of parameters, optimized for these conditions. The relative extent of the surface in geometric docking was increased by using two layers of grid points (about 3 Å instead of 2 Å wide at the higher resolution) around the core of molecule A. This increased tolerance was somewhat compensated by assigning lower values to the external surface layer. The values assigned to grid points in molecule B and to internal surface grid points in molecule A (see Table I) are set to 1 at any resolution. Since the molecules are represented by a smaller number of grid points at the lower resolution, the absolute values for the scores are expected to be lower. Moreover, one has to decrease the relative contribution of the penalty for penetration. Thus a value of -10 (instead of -15 at the higher resolution) was assigned to the points in the interior of molecule A.

A similar argumentation pertains for setting the parameters in the hydrophobic docking. Although the molecules are projected upon 3-D grids at the same resolution as in the geometric docking, the number of atoms in the new representation is reduced. Thus, to lower the relative contribution of penetration, a value of -5 was taken for the interior of molecule A. Moreover, since the molecules are represented by a smaller number of elements, thereby increasing the tolerance for mismatch, we found it necessary to give up the external surface layer of molecule A. However, in cases of matching two native molecules, in which a significant change in conformation is known or suspected to occur upon complex formation, we did use a second surface

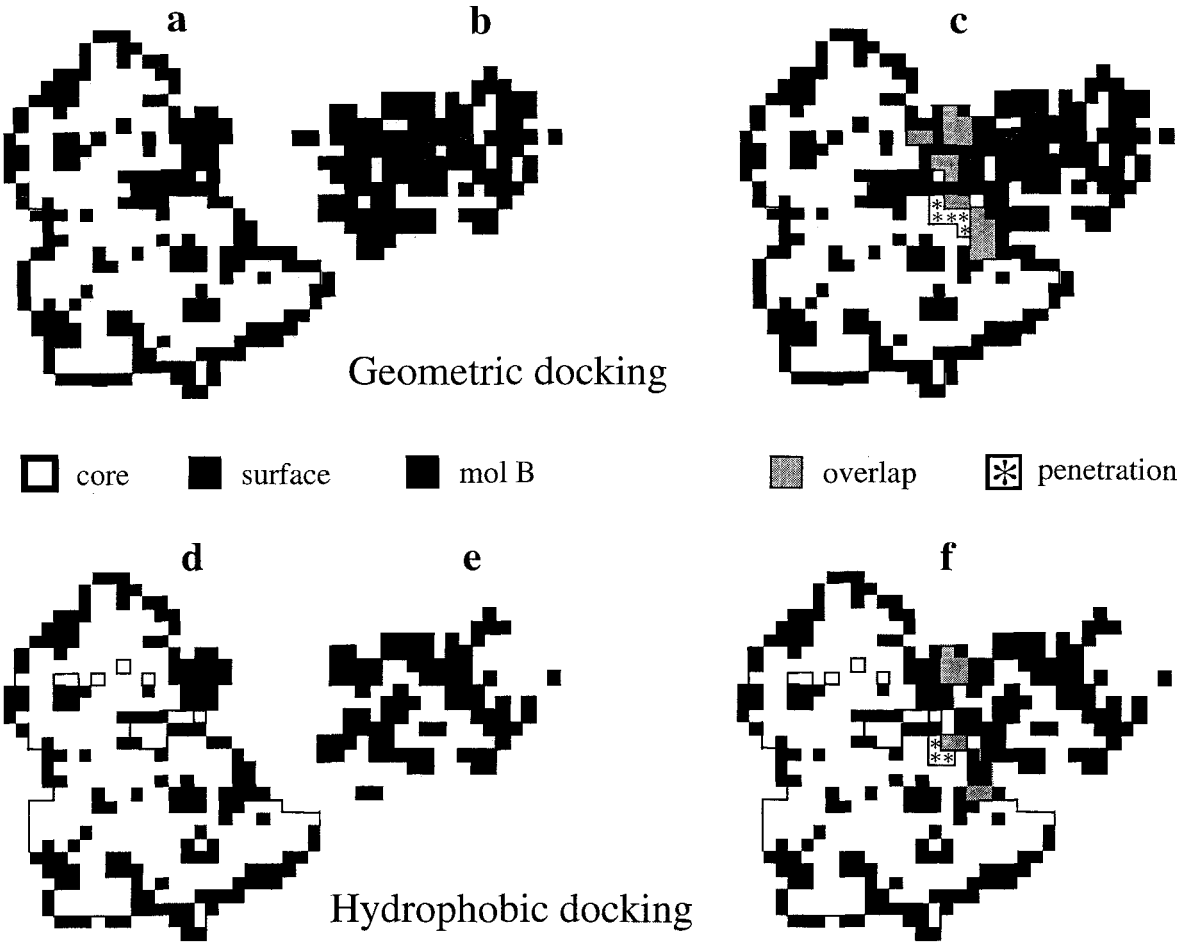

Fig. 3. Cross sections through the 3-D grid representations of the molecules. The upper and lower panels represent results from geometric and hydrophobic matching, respectively. Molecule A **(a, d)** and molecule B **(b, e)** are in the same orientations as in Figure 2. The relative positions of the molecules are these of a complex, in which molecule B was shifted to the right by 10 grid steps. With both geometric and hydrophobic docking, the molecules are represented at identical relative positions for comparative purposes. In **(c)** and **(f)**, the structure of the complex is shown. The core and the first surface layer in molecule A (the second surface layer is not shown) are distinguished by the patterns indicated in the figure. The molecular contacts between the molecules (overlap and penetration) are depicted by a different set of patterns.

layer. This effects a softer approach in the matching process.

The above-mentioned parameters have been tested in several cases for consistency of the results in geometric docking with these obtained at the higher resolution. They are considered quasioptimal, since in most cases the highest score was associated with the correct complex configuration. A discrimination stage at higher resolution, used in our previous report,[13] was not implemented in the present work. With both the geometric and hydrophobic docking, different sets of parameters were tested independently. Changes in the parameters invariably resulted in a negative effect on the quality of the results, as assessed from the relative position of the correct match in the sorted list of scores.

In Figure 3 the representations of molecule A (Fig. 3a,d), molecule B (Fig. 3b,e), and their complex (Fig. 3c,f) are shown for geometric (upper) and hydrophobic (lower) docking by appropriate two-dimensional sections of the 3D grids. The molecules and their relative position were chosen for illustrative purposes, so that two types of contact in the complex are shown. These are overlaps between elements of molecule B and surface elements of molecule A (positive contribution to the score) or their penetration in the core of molecule A (large negative contribution to the score).

RESULTS AND DISCUSSION

We tested the competence of the hydrophobic docking procedure in predicting the correct configuration for bimolecular complexes. We applied the algorithm on various known protein complexes taken from the Brookhaven Protein Data Bank,[29] and considered as representative cases. These are α1–β1

subunits of human deoxyhemoglobin (2HHB),[30] α1–β1 subunits of horse methemoglobin (2MHB),[31] acid proteinase–peptide inhibitor (3APR),[32] and trypsin–trypsin inhibitor (2PTC).[33] For comparative purposes, we also applied the geometric docking procedure on the same pairs.

Representation of the Molecules

Our distinction between hydrophobic and nonhydrophobic atoms in molecules yields typically about 70% nonhydrophobic atoms in total. However, the number of grid elements in the hydrophobic representation of molecule B was reduced by about 40% only (see Fig. 2b). In the projection of molecule A onto a 3-D grid, we used all atoms to generate the core. Indeed, the number of corresponding grid elements does not vary between the geometric and hydrophobic representations of the core. However, the number of points representing the first (main) surface layer of molecule A was reduced by about 20% in the hydrophobic—as compared to the geometric—image (see Fig. 2a). This reduction was almost abolished for the second surface layer (not shown), due to a higher volume occupancy by the expanded hydrophobic atoms.

In order to test the general validity of the approach, based on assigning a priori more weight to hydrophobic atom groups, we did also represent the molecules by nonhydrophobic atoms. This was done simply by swapping the identity of hydrophobic/nonhydrophobic atoms in the algorithm. Thus, all molecule B and the surface of molecule A have been represented using nonhydrophobic atoms only. With either the full or hydrophobic representation of the molecules, the correct match was always found among the few first peaks (see Fig. 4). However, applying the geometric complementarity algorithm in matching the exclusively nonhydrophobic surfaces resulted in a failure in retrieving the correct complex configuration at any significant position in the sorted list of peaks (rank of less than 1000).

Complex Prediction

The prediction results for four complexes are summarized in Figure 4. It shows histograms of the sorted 15 highest-score complex configurations for each pair of molecules. The upper part in each panel gives the results obtained with geometric docking, when the molecules are represented by all atoms. The lower part shows the results of hydrophobic docking, when only hydrophobic elements are represented in the areas to be matched. Finally, the quality of the results obtained by both procedures is compared in Figure 4e. This is done by calculating a "signal-to-noise ratio," in which the signal was taken as the score for the correct match, and the noise arbitrarily defined as the average of the scores associated with the first 20 false positive configurations of the corresponding pairs.

For 2HHB (Fig. 4a), 3APR (Fig. 4c), and 2PTC (Fig. 4d), a significant improvement in the quality of prediction for the correct match is evident. In these cases, an increase in the signal-to-noise ratio is observed (Fig. 4e). However, with 2MHB (Fig. 4b), the new approach yields results similar to these of geometric docking.

In the 3APR complex, the bound oligopeptide inhibitor is immersed in a deep hydrophobic pocket in the acid proteinase molecule. This pair is representative of complexes characterized by predominantly hydrophobic intermolecular contact areas. The scores obtained in this case with both approaches are relatively low, reflecting the small size of the ligand. The quality of the results with geometric docking is high, and it is substantially improved with hydrophobic docking. The other molecular pairs represent complexes with larger contact interfaces of a lesser hydrophobic character.

With the hemoglobin complexes, the correct configuration was retrieved in both procedures as the first peak in the sorted list. In the case of human deoxyhemoglobin (2HHB, Fig. 4a), the hydrophobic docking yielded better results than with geometric docking since the correct match is distinguished well above the other matches, as shown by the improved relative score (see also Fig. 4e). Korn and Burnett[26] have referred to interfaces in protein complexes as predominantly hydrophobic relative to noncontact surfaces. The interface between α1 and β1 subunits of human deoxyhemoglobin is peculiar in this respect since the hydrophobicity at the interface was found lower than at noncontact surfaces. Examination of the structure of the complex indeed reveals strong cross-linking through salt bridges often involving water molecules at the interface. It is interesting to note that nevertheless, a distinctly positive effect is observed in our results after the exclusion of nonhydrophobic atoms.

In the case of horse methemoglobin (2MHB), the correct match is already distinctly higher with geometric docking (Fig. 4b), as assessed by the signal-to-noise ratio (Fig. 4e). This may be the reason why hydrophobic docking does not improve the results further.

With the complex trypsin–trypsin inhibitor, often considered as a model for electrostatic interactions between proteins,[5] the correct configuration was not retrieved as the highest score match in either procedure. The results previously obtained with geometric docking at higher resolution (grids of $90 \times 90 \times 90$)[13] indicate that the correct match was found in position 3. In the present results, it was found at position 13 of the sorted list in geometric docking. The apparent difference with the present result (geometric docking) reflects the strong dependence of the prediction quality on the resolution, as previously shown.[13] However, in hydrophobic docking the rank for the correct match went up to posi-

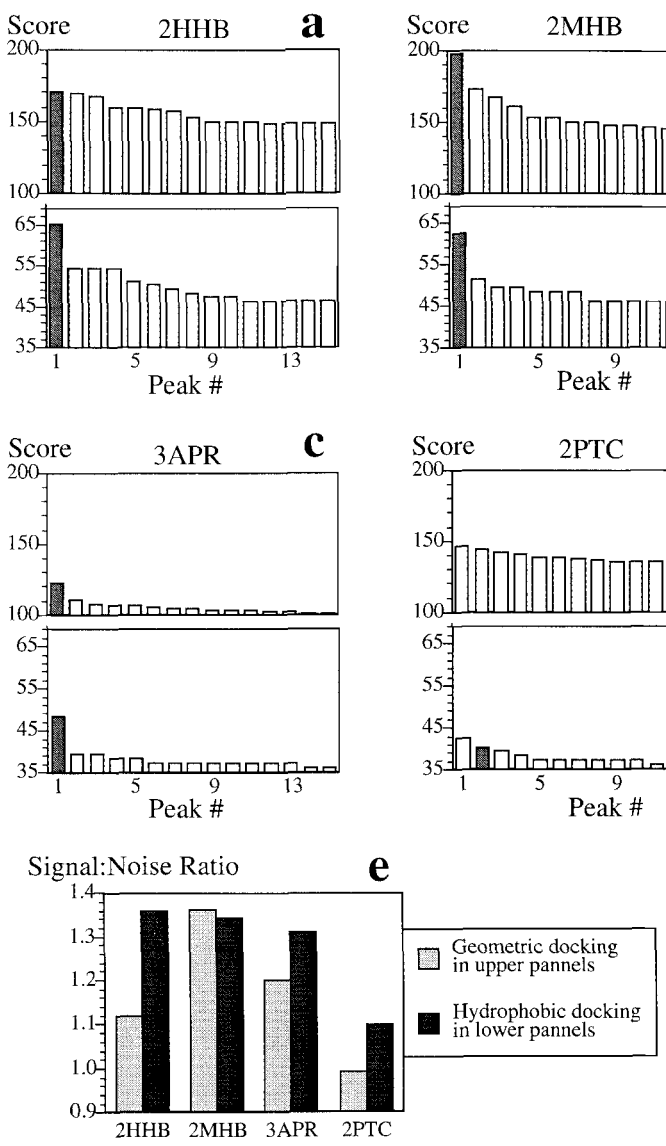

Fig. 4. Scores of predicted complex configurations for different pairs of molecules **(a–d)**. The shaded bars represent the correct configuration. In **(e)** a signal-to-noise ratio for each set of results was calculated as described in the text.

tion 2 (see Fig. 4d), indicating a significant improvement of the prediction as assessed also by the increase in signal-to-noise ratio (see Fig. 4e).

We used both variants of the docking algorithm to match free trypsin inhibitor (4PTI)[34] with the trypsin molecule from the 2PTC complex. Some relevant differences in structure between free trypsin inhibitor and that bound to trypsin are mainly due to substantial changes in the orientation of the side-chains from Lys-15, Arg-17, and Arg-39. These relatively large conformational changes create major difficulties in the prediction of the correct match when using the structures of the free inhibitor.[10,12,14] The geometric docking procedure failed in finding the relative position corresponding to the correct match among the first 7000 highest-score peaks stored in the final sorted list. *Computer* graphics analysis of the molecules brought together in the same relative position as in the complex 2PTC indicate substantial penetration of atoms from the Lys-15 side-chain from the inhibitor, especially the ε-amino group. With hydrophobicity docking, the position corresponding to the correct configuration was found with a rank of 5394 (score of 53 compared to 104 for the first peak). In this case, the nonhydrophobic atom groups being omitted from the representation of the molecules, less collisions arising from conformational changes in hydrophilic surface

residues are detected. In addition, the other factor characteristic to hydrophobic docking, a higher density of hydrophobic groups at the binding site, may also contribute to the improved performance. However, these are more difficult to identify in the results. Nevertheless, the moderate increase in performance strengthens the positive trend demonstrated in the case of molecules in the correct conformation (i.e., in a complex). This example emphasizes the difficulty in matching molecules in which a conformational change occurs upon complex formation. It illustrates the concept of superfluous and detrimental information, and the need for more refined approaches in molecular recognition to take into account local flexibility of atom groups at the surface.

CONCLUSION AND PERSPECTIVES

The approach presented here is not designed to assess the role or the importance of hydrophobic or other forces in intermolecular interactions. Its purpose is rather to improve molecular recognition techniques, designed to assess the potential for association between molecules and predict their relative position in a complex formed in solution, based on the previous knowledge of their 3-D structure in a crystal. We show that the geometric surface complementarity procedure works more efficiently with a hydrophobic representation of molecules as compared to that using their full representation. Indeed, hydrophobic docking yielded improved results than the previously established geometric docking procedure. When assessing the extent of this improvement, it is important to keep in mind that the hydrophobic docking results were obtained with the molecules represented by only one-third of their surface heavy atoms, as compared to the full representation of molecules in the geometric docking. We would suggest that even if the results for cocrystallized structures (Fig. 4) were not improved at all, the factors of increased tolerance to conformational changes and sharply reduced number of atoms would still be of significant importance for docking studies. The tests conducted using "nonhydrophobic" representation appear to yield results with much less predictive power than either the former representations. Taken together, the results strongly suggest that hydrophobic groups at the surface substantially contribute to geometric surface *molecular recognition*. Both the approach itself and the results illustrate the basic difference between intermolecular energy calculations and recognition techniques. In energy calculations, especially in the case of strong electrostatic interactions, the necessity of taking into account the contribution of charged groups is obvious. However, in recognition procedures a more faithful representation of molecules may be detrimental in predicting the correct match by increasing the relative score of false positive matches.

We plan to validate the approach further by testing more molecular pairs, whether in complex or resolved separately. The latter cases will serve as more realistic tests for the ability of the algorithm to tolerate small conformational changes. We also intend to test the merit of the distinction between hydrophobic and non-hydrophobic elements (atoms, atom groups, residues, etc.) in the representation of the molecules in other recognition algorithms using different approaches. It is anticipated that partial representations using elements relevant to recognition not only reduce the computation time but should also improve the signal-to-noise ratio in the corresponding procedures. Finally, different methods for molecules representation and complementarity quantitation may help to determine whether the critical factor relevant to recognition is hydrophobicity by itself or it is only (cor)related in a broader sense to a more complex ensemble of physicochemical characteristics, reflecting for instance the flexibility of structural elements.

ACKNOWLEDGMENTS

We wish to thank Prof. Ephraim Katchalski-Katzir for his support in developing the concept of "hydrophobic complementarity." We are grateful to the molecular recognition group at WIS (Drs. M. Eisenstein and I. Shariv, Profs. I. Steinberg and A. A. Friesem) for extensive discussion of the topics. The authors acknowledge the support of Foundation RASCHI.

REFERENCES

1. Guarente, L. Strategies for the identification of interacting proteins. Proc. Natl. Acad. Sci. U.S.A. 90:1639–1641, 1993.
2. Goodford, P.J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. J. Med. Chem. 28:849–857, 1985.
3. Warwicker, J. Investigating protein-protein interaction surfaces using a reduced stereochemical and electrostatic model. J. Mol. Biol. 206:381–395, 1989.
4. Goodsell, D.S., Olson, A.J. Automated docking of substrates to proteins by simulated annealing. Proteins 8:195–202, 1990.
5. Sharp, K.A., Honing, B. Electrostatic interactions in macromolecules: Theory and applications. Annu. Rev. Biophys. Biophys. Chem. 19:301–332, 1990.
6. Yue, S.-Y. Distance-constrained molecular docking by simulated annealing. Protein Eng. 4:177–184, 1990.
7. Kuntz, I.D., Blaney, J.M., Oatley, S.J., Langridge, R., Ferrin, T.E. A geometric approach to macromolecule-ligand interactions. J. Mol. Biol. 161:269–288, 1982.
8. Connolly, M.L. Shape complementarity at the hemoglobin alpha 1-beta 1 subunit interface. Biopolymers 25:1229–1247, 1986.
9. DesJarlais, R.L., Sheridan, R.P., Seibel, G.L., Dixon, J.S., Kuntz, I.D., Venkataraghavan, R. Using shape complementarity as an initial screen in designing ligands for a receptor binding site of known three-dimensional structure. J. Med. Chem. 31:722–729, 1988.
10. Jiang, F., Kim, S.H. "Soft docking": Matching of molecular surface cubes. J. Mol. Biol. 219:79–102, 1991.
11. Nussinov, R., Wolfson, H.J. Efficient detection of three-dimensional structural motifs in biological macromolecules by computer vision techniques. Proc. Natl. Acad. Sci. U.S.A. 88:10495–10499, 1991.

12. Shoichet, B.K., Kuntz, I.D. Protein docking and complementarity. J. Mol. Biol. 221:327–346, 1991.
13. 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. 89:2195–2199, 1992.
14. Wodak, S.J., Janin, J. Computer analysis of protein-protein interaction. J. Mol. Biol. 124:323–342, 1978.
15. Janin, J., Chothia, C. The structure of protein-protein recognition sites. J. Biol. Chem. 265:16027–16030, 1990.
16. Leach, A.R., Kuntz, I.D. Conformational analysis of flexible ligands in macromolecular receptor sites. J. Comput. Chem. 13:730–748, 1992.
17. Calfisch, A., Niederer, P., Anliker, M. Monte Carlo docking of oligopeptides to proteins. Proteins 13:223–230, 1992.
18. Kuntz, I.D. Structure-based strategies for drug design and discovery. Science 257:1078–1082, 1992.
19. Marshall, G.R. 3D structure of peptide-protein complexes: Implications for recognition. Curr. Opin. Struct. Biol. 2:904–919, 1992.
20. Cherfils, J., Janin, J. Protein docking algorithms: Simulating molecular recognition. Curr. Opin. Struct. Biol. 3:265–269, 1993.
21. Agishtein, M.E. Fuzzy molecular surfaces. J. Biomol. Struct. Dyn. 9:759–68, 1992.
22. Casari, G., Sippl, M.J. Structure-derived hydrophobic potential: Hydrophobic potential derived from X-ray structures of globular proteins is able to identify native folds. J. Mol. Biol. 224:725–732, 1992.
23. Chan, H.S., Dill, K.A. Polymer principles in protein structure and stability. Annu. Rev. Biophys. Biophys. Chem. 20:447–490, 1991.
24. Kellogg, G.E., Semus, S.F., Abraham, D.J. HINT: a new method of empirical hydrophobic field calculation for CoMFA. J. Computer-Aided Mol. Des. 5:545–552, 1991.
25. Kim, K.H. 3D-quantitative structure-activity relationships: Describing hydrophobic interactions directly from 3D structures using a comparative molecular field analysis (CoMFA) approach. Quant. Struct.-Act. Relat. 12:232–238, 1993.
26. Korn, A.P., Burnett, R.M. Distribution and complementarity of hydropathy in multisubunit proteins. Proteins 9:37–55, 1991.
27. Shoichet, B.K., Kuntz, I.D. Matching chemistry and shape in molecular docking. Protein Eng. 6:723–732, 1993.
28. Galaktionov, S.G., Tseitin, V.M., Vakser, I.A., Prokhorchik, E.V. Amphiphilic properties of angiotensin and its fragments. Biophysics 33:595–598, 1988.
29. Abola, E.E., Bernsein, F.C., Bryant, S.H., Koetzle, T.L., Weng, J. Protein Databank. In "Crystallographic Databases—Information Content, Software Systems, Scientific Applications." Allen, F.H., Bergerhoff, G., Sievers, R. (eds.). Bonn: Data Commission of the International Union of Crystallography, 1987: 107–132.
30. Fermi, G., Perutz, M.F., Shaanan, B., Fourme, R. The crystal structure of human deoxyhaemoglobin at 1.74 Å resolution. J. Mol. Biol. 175:159–174, 1984.
31. Ladner, R.C., Heidner, E.G., Perutz, M.F. The structure of horse methaemoglobin at 2.0 angstroms resolution. J. Mol. Biol. 114:385–414, 1977.
32. Suguna, K., Bott, R.R., Padlan, E.A., Subramanian, E., Sheriff, S., Cohen, G.H., Davies, D.R. Structure and refinement at 1.8 Å resolution of the aspartic proteinase from Rhizopus chinensis. J. Mol. Biol. 196:877–900, 1987.
33. Marquart, M., Walter, J., Deisenhofer, J., Bode, W., Huber, R. The geometry of the reactive site and of the peptide groups in trypsin, trypsinogen and its complexes with inhibitors. Acta Crystallog., Sect. B 39:480–490, 1983.
34. Wlodawer, A., Deisenhofer, J., Huber, R. Comparison of two highly refined structures of bovine pancreatic trypsin. J. Mol. Biol. 193:145–156, 1987.