Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 2195–2199, March 1992
Biophysics

Molecular surface recognition: Determination of geometric fit between proteins and their ligands by correlation techniques

(protein–protein interaction/surface complementarity/macromolecular complex prediction/molecular docking)

EPHRAIM KATCHALSKI-KATZIR[†‡], ISAAC SHARIV[§], MIRIAM EISENSTEIN[¶], ASHER A. FRIESEM[§], CLAUDE AFLALO[‖], AND ILYA A. VAKSER[†]

Departments of [†]Membrane Research and Biophysics, [§]Electronics, [¶]Structural Biology, and [‖]Biochemistry, Weizmann Institute of Science, Rehovot 76100, Israel

Contributed by Ephraim Katchalski-Katzir, October 24, 1991

ABSTRACT **A geometric recognition algorithm was developed to identify molecular surface complementarity. It is based on a purely geometric approach and takes advantage of techniques applied in the field of pattern recognition. The algorithm involves an automated procedure including (*i*) a digital representation of the molecules (derived from atomic coordinates) by three-dimensional discrete functions that distinguishes between the surface and the interior; (*ii*) the calculation, using Fourier transformation, of a correlation function that assesses the degree of molecular surface overlap and penetration upon relative shifts of the molecules in three dimensions; and (*iii*) a scan of the relative orientations of the molecules in three dimensions. The algorithm provides a list of correlation values indicating the extent of geometric match between the surfaces of the molecules; each of these values is associated with six numbers describing the relative position (translation and rotation) of the molecules. The procedure is thus equivalent to a six-dimensional search but much faster by design, and the computation time is only moderately dependent on molecular size. The procedure was tested and validated by using five known complexes for which the correct relative position of the molecules in the respective adducts was successfully predicted. The molecular pairs were deoxyhemoglobin and methemoglobin, tRNA synthetase–tyrosinyl adenylate, aspartic proteinase–peptide inhibitor, and trypsin–trypsin inhibitor. A more realistic test was performed with the last two pairs by using the structures of uncomplexed aspartic proteinase and trypsin inhibitor, respectively. The results are indicative of the extent of conformational changes in the molecules tolerated by the algorithm.**

The association of proteins with their ligands involves intricate inter- and intramolecular interactions, solvation effects, and conformational changes. In view of such complexity, a comprehensive and efficient approach for predicting the formation of protein–ligand complexes from the structure of their free components is not yet available. However, with some assumptions, such predictions become feasible, and several attempts based on energy minimization have been partially successful (1–6). Another simplifying approach that could alleviate some of these difficulties is based on geometric considerations.

The three-dimensional (3D) structures of most protein complexes reveal a close geometric match between those parts of the respective surfaces of the protein and the ligand that are in contact. Indeed, the shape and other physical characteristics of the surfaces largely determine the nature of the specific molecular interactions in the complex. Furthermore, in many cases the 3D structure of the components in the complex closely resembles that of the molecules in their free, native state. Geometric matching thus seems to play an important role in determining the structure of a complex.

Several investigators have exploited a geometric approach to find shape complementarity between a given protein and its ligand (7–19). They considered geometric match between molecular surfaces as a fundamental condition for the formation of a specific complex and pointed out the advantages of the geometric approach (13). In this approach, which treats proteins as rigid bodies, the complementarity between surfaces is estimated. Furthermore, the geometric analysis could serve as the foundation for a more complete approach including energy considerations. However, the methods heretofore developed for analyzing geometric matching do not seem to simultaneously fulfill the requirements for generality, accuracy, reliability, and reasonable computation time.

In this paper, we present a geometry-based algorithm for predicting the structure of a possible complex between molecules of known structures. This relatively simple and straightforward algorithm relies on the well-established correlation and Fourier transformation techniques used in the field of pattern recognition. The algorithm requires only that the 3D structure of the molecules under consideration be known. Moreover, it provides quantitative data related to the quality of the contact between the molecules. The algorithm was tested and validated in the analysis of the following complexes, whose structures are known: the α–β hemoglobin dimer, tRNA synthetase–tyrosinyl adenylate, aspartic proteinase–peptide inhibitor, and trypsin–trypsin inhibitor. The correct relative position of the molecules within these complexes were successfully predicted.

METHOD

Geometric Recognition Algorithm. We begin with a geometric description of the protein and the ligand molecules, derived from their known atomic coordinates. The two molecules denoted by **a** and **b**, are projected onto a three dimensional grid of $N \times N \times N$ points, where they are represented by the discrete functions

$$a_{l,m,n} = \begin{cases} 1 & \text{inside the molecule} \\ 0 & \text{outside the molecule,} \end{cases} \quad \text{[1a]}$$

and

$$b_{l,m,n} = \begin{cases} 1 & \text{inside the molecule} \\ 0 & \text{outside the molecule,} \end{cases} \quad \text{[1b]}$$

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: 3D, three dimensional; DFT, discrete Fourier transform; IFT, inverse Fourier transform.

[‡]To whom reprint requests should be addressed.

where l, m, and n are the indices of the 3D grid ($l, m, n = \{1 \ldots N\}$). Any grid point is considered inside the molecule if there is at least one atom nucleus within a distance r from it, where r is of the order of van der Waals atomic radii. Examples for two-dimensional cross sections of these functions are presented in Fig. 1 *a* and *b*.

Next, to distinguish between the surface and the interior of each molecule, we retain the value of 1 for the grid points along a thin surface layer only and assign other values to the internal grid points. The resulting functions thus become

$$\bar{a}_{l,m,n} = \begin{cases} 1 & \text{on the surface of the molecule} \\ \rho & \text{inside the molecule} \\ 0 & \text{outside the molecule,} \end{cases} \quad [2a]$$

and

$$\bar{b}_{l,m,n} = \begin{cases} 1 & \text{on the surface of the molecule} \\ \delta & \text{inside the molecule} \\ 0 & \text{outside the molecule,} \end{cases} \quad [2b]$$

where the surface is defined here as a boundary layer of finite width between the inside and the outside of the molecule. The parameters ρ and δ describe the value of the points inside the molecules, and all points outside are set to zero. Two-dimensional cross sections of these functions are shown in Figs. 1 *c* and *d*.

In our method, matching of surfaces is accomplished by calculating correlation functions. The correlation between the discrete functions $\bar{a}$ and $\bar{b}$ is defined as

$$\bar{c}_{\alpha,\beta,\gamma} = \sum_{l=1}^{N} \sum_{m=1}^{N} \sum_{n=1}^{N} \bar{a}_{l,m,n} \cdot \bar{b}_{l+\alpha,m+\beta,n+\gamma}, \quad [3]$$

where α, β, and γ are the number of grid steps by which molecule **b** is shifted with respect to molecule **a** in each dimension. If the shift vector $\{\alpha,\beta,\gamma\}$ is such that there is no contact between the two molecules (see Fig. 2*a*), the correlation value is zero. If there is a contact between the surfaces

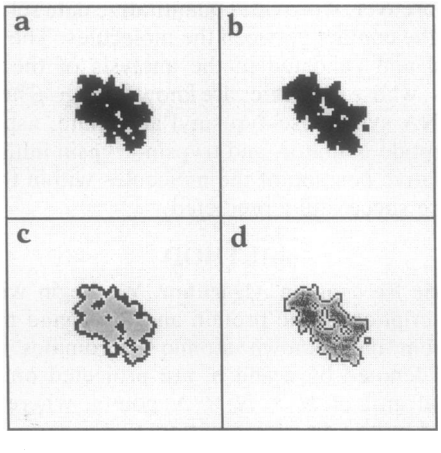

FIG. 1. Typical cross sections through the 3D grid representations of the molecules. (*a*) Cross section (at $l = 46$) through the function $a_{l,m,n}$, derived by projecting the α subunit of hemoglobin (from 2HHB; see text) onto a 3D grid ($N = 90$). The values 0 and 1 are represented in white and black, respectively. (*b*) The cross section $b_{46,m,n}$ was similarly derived for the β subunit (from 2HHB). Other details are as in *a*. (*c*) The cross section (at $l = 46$) through the function $\bar{a}_{l,m,n}$, which was obtained by distinguishing the surface layer from the interior of the molecule in the function $a_{l,m,n}$. The large negative value for ρ is represented in gray. (*d*) Cross section $\bar{b}_{46,m,n}$, similarly derived from $b_{l,m,n}$. The small positive value for δ is represented in a different shade of gray. The values for r and η were 1.8 Å and 1.2 Å, respectively.

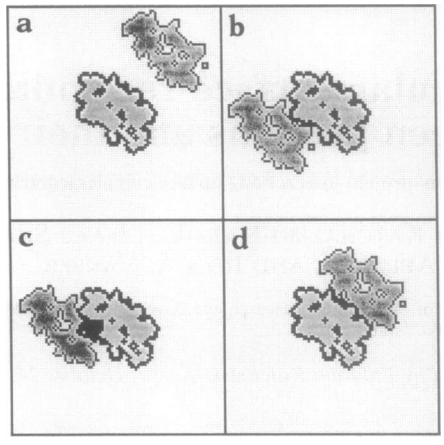

FIG. 2. Different relative positions of molecules **a** and **b**, illustrated by the cross sections $\bar{a}_{46,m,n}$ and $\bar{b}_{46,m,n}$ from Fig. 1. The relative orientation of the molecules is as in the known α–β dimer. (*a*) No contact. (*b*) Limited contact. (*c*) Penetration. The penetrated area is represented in black. (*d*) Good geometric match, as indicated by the extensive overlap of complementary surface layers.

(Fig. 2*b*), the contribution to the correlation value is positive. Nonzero correlation values could also be obtained when one molecule penetrates into the other (Fig. 2*c*). Since such penetration is physically forbidden, a distinction between surface contact and penetration must be clearly formulated. To do so, we assign large negative values to ρ in $\bar{a}$ and small nonnegative values to δ in $\bar{b}$. Thus, when the shift vector $\{\alpha,\beta,\gamma\}$ is such that molecule **b** penetrates molecule **a**, the multiplication of the negative numbers (ρ) in $\bar{a}$ by the positive numbers (1 or δ) in $\bar{b}$ results in a negative contribution to the overall correlation value. Consequently, the correlation value for each displacement is simply the score for overlapping surfaces corrected by the penalty for penetration.

Positive correlation values are obtained when the contribution from surface contact outweighs that from penetration. Thus, a good geometric match (such as in Fig. 2*d*) is represented by a high positive peak, and low values reflect a poor match between the molecules. A cross section of a typical correlation function for a good match is presented in Fig. 3. The coordinates of the prominent peak denote the relative shift of molecule **b** yielding a good match with molecule **a**. The location of the recognition sites on the surface of each molecule can readily be determined from these coordinates. In addition, the width of the peak provides a measure for the relative displacement allowed before matching is lost.

A direct calculation of the correlation between the two functions (see Eq. **3**) is rather lengthy, since it involves N^3 multiplications and additions for each of the N^3 possible relative shifts $\{\alpha,\beta,\gamma\}$, resulting in an order of N^6 computing steps. Therefore, we chose to take advantage of Fourier transformation that allowed us to calculate the correlation function much more rapidly. The discrete Fourier transform (20) (DFT) of a function $x_{l,m,n}$ is defined as

$$X_{o,p,q} = \sum_{l=1}^{N} \sum_{m=1}^{N} \sum_{n=1}^{N} \exp[-2\pi i(ol + pm + qn)/N] \cdot x_{l,m,n}, \quad [4]$$

where $o, p, q = \{1 \ldots N\}$ and $i = \sqrt{-1}$. The application of this transformation to both sides of Eq. **3** yields (21)

$$C_{o,p,q} = A^*_{o,p,q} \cdot B_{o,p,q}, \quad [5]$$

where C and B are the DFT of the functions $\bar{c}$ and $\bar{b}$, respectively, and A^* is the complex conjugate of the DFT of

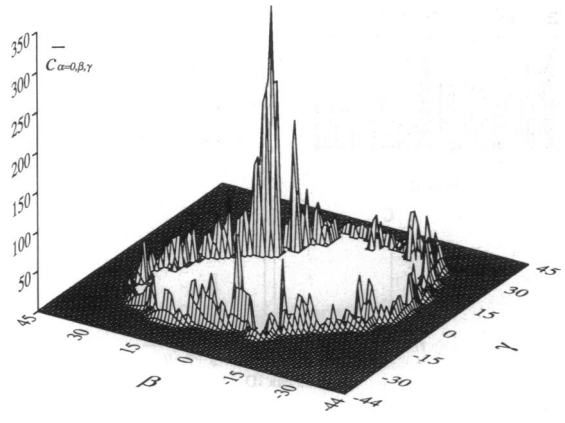

FIG. 3. Cross section (at $\alpha = 0$) through a 3D correlation function $\bar{c}_{\alpha,\beta,\gamma}$. The correlation function shown was calculated for the α and β subunits of hemoglobin, oriented as in the dimer (from 2HHB, see Figs. 1 *c* and *d*). The correlation value at each shift vector $\{0,\beta,\gamma\}$ is represented by the height of the graph. The prominent peak at $\{\alpha = 0, \beta = 14, \gamma = 17\}$ corresponds to the correct match between the molecules (see Fig. 2*d*). Other intermolecular surface contacts (such as in Fig. 2*b*) give rise to the low positive correlation values around the center of the graph. The negative correlation values caused by penetration (see Fig. 2*c*) are omitted, leaving the empty area at the center.

$\bar{a}$. Eq. **5** indicates that the transformed correlation function C is obtained by a simple multiplication of the two functions A^* and B. The inverse Fourier transform (20) (IFT), defined as

$$\bar{c}_{\alpha,\beta,\gamma} = \frac{1}{N^3}\sum_{o=1}^{N}\sum_{p=1}^{N}\sum_{q=1}^{N}\exp[2\pi i(o\alpha + p\beta + q\gamma)/N]\cdot C_{o,p,q}, \quad [6]$$

is used to obtain the desired correlation between the two original functions $\bar{a}$ and $\bar{b}$. The Fourier transformations can be performed with the fast Fourier transform algorithm (20), which requires less than the order of $N^3 \ln(N^3)$ steps for transforming a 3D function of $N \times N \times N$ values. Thus, the overall procedure leading to Eq. **6** is significantly faster than the direct calculation of $\bar{c}$ according to Eq. **3**.

Finally, to complete a general search for a match between the surfaces of molecules **a** and **b**, the correlation function $\bar{c}$ has to be calculated for all relative orientations of the molecules. In practice, molecule **a** is fixed, whereas the three Euler angles defining the orientation of molecule **b** (*xyz* convention in ref. 22) are varied at fixed intervals of Δ degrees. This results in a complete scan of $360 \times 360 \times 180/\Delta^3$ orientations for which the correlation function $\bar{c}$ must be calculated.

The entire procedure described above can be summarized by the following steps:

(*i*) derive $\bar{a}$ from atomic coordinates of molecule **a** (Eq. **2**),
(*ii*) $A^* = [\mathrm{DFT}(\bar{a})]^*$ (Eq. **4**),
(*iii*) derive $\bar{b}$ from atomic coordinates of molecule **b** (Eq. **2**),
(*iv*) $B = \mathrm{DFT}(\bar{b})$ (Eq. **4**),
(*v*) $C = A^* \cdot B$ (Eq. **5**),
(*vi*) $\bar{c} = \mathrm{IFT}(C)$ (Eq. **6**),
(*vii*) look for a sharp positive peak of $\bar{c}$,
(*viii*) rotate molecule **b** to a new orientation,
(*ix*) repeat steps *iii*–*viii* and end when the orientations scan is completed, and
(*x*) sort all of the peaks by their height.

Each high and sharp peak found by this procedure indicates geometric match and thus represents a potential complex. The relative position and orientation of the molecules within each such complex can readily be derived from the coordinates of the correlation peak, and from the three Euler angles at which the peak was found.

Implementation of the Algorithm. To implement our algorithm, it is necessary to assign specific values to the various parameters involved—i.e., the surface layer thickness, r, Δ, ρ, δ, N, and the grid step size denoted by η. The choice of these values is based on a number of considerations, outlined in this section.

We begin by noting that the match between the functions $\bar{a}$ and $\bar{b}$ is not perfect. One reason is that the structure of known complexes reveals small gaps between the molecules, which are also reflected in their mathematical representation. Furthermore, the functions $\bar{a}$ and $\bar{b}$ are derived from atomic coordinates sets that do not include hydrogen atoms. This, in addition to the limited accuracy of the coordinates, may affect the quality of the match. Finally, minor conformational changes may occur at the surface of molecules upon complex formation (locally induced fit). Such changes are not incorporated in the functions $\bar{a}$ and $\bar{b}$ when they represent native molecules that are assumed to be rigid. Therefore, penetration and small gaps occur along the contact area. To ensure that the correct match between molecules is not missed, our algorithm must be able to tolerate these imperfections. This is achieved by assigning more than one layer of grid points to the surface in $\bar{a}$ so that the surface thickness for molecule **a** is 1.5–2.5 Å (see Fig. 1*c*). Consequently, penetrations and gaps that are smaller than these values are tolerated. It should be noted that an inherent drawback in the choice of a thicker surface layer is the concomitant increase in the number of faulty matches.

The thickness of the surface layer also influences the angular tolerance. This tolerance is defined as the maximal deviation from the correct match orientation that would still result in a distinct correlation peak. Typically, a surface layer thickness of 2 Å yielded an angular tolerance of about $\pm 10°$. Thus, the angular step Δ was set to 20°, resulting in 2916 different orientations of molecule **b** at each of which the correlation function had to be evaluated.

The parameter r, used to derive the functions $a_{l,m,n}$ and $b_{l,m,n}$ (see Eq. **1**), was set to 1.8 Å, which is larger by about 0.2 Å than the average van der Waals radius for carbon, nitrogen, and oxygen. This compensated for the fact that hydrogen atoms, missing in the coordinates sets, are not projected on our grids.

The parameters ρ and δ, representing the interior of the molecules, were set to -15 and 1, respectively. This ensures that the correlation value is substantially reduced in case of penetration. Several other choices for ρ and δ, in the ranges $\rho << -1$ and $0 \leq \delta \leq 1$, did not significantly affect the performance of the algorithm.

Another important parameter of the algorithm is the grid step size, η. Optimal results were obtained when η was set to 0.7–0.8 Å, corresponding to half of the carbon–carbon bond length. Yet, since the product $\eta \cdot N$ should be larger than the size of any potential complex, a finer grid requires a larger number of points N. This leads in turn to excessive computation time. Therefore, we performed an initial scan of the angular orientations with larger grid steps ($\eta \approx 1.0$–1.2 Å); thus, computations that would take days with the finer grid were performed in hours. However, with such large grid steps, spurious correlation peaks, which may even be higher than the correct peak, appear. Hence, the scan stage was followed by a discrimination stage, in which the correlation functions were recalculated with a finer grid ($\eta \approx 0.7$–0.8 Å), but only for those orientations that yielded the highest peaks in the scan stage. This discrimination stage enhanced the correct correlation peak and suppressed spurious peaks.

A FORTRAN program was developed for implementing the algorithm. The parameters of the program, in accordance

with the arguments given above, were assigned the following values: r = 1.8 Å, Δ = 20°, ρ = −15, δ = 1, N = 90 ($\eta \approx$ 1.0–1.2 Å) for the scan stage, and N = 128 ($\eta \approx$ 0.7–0.8 Å) for the discrimination stage. The program was run on a Convex C-220 computer with the Veclib fast Fourier transform subroutine. The computation time for each iteration (steps *iii–viii* in the summarized algorithm) in the scan stage was 9 sec. The total computation time for matching two molecules in the range of 1100 atoms each, including both the initial scan and the discrimination stage, was typically 7.5 hr.

RESULTS

Our algorithm was applied to several known complexes, whose coordinates are given in the Brookhaven Protein Data Bank (Brookhaven National Laboratory, Upton, NJ) to test its ability to predict correct structures of protein complexes. We chose complexes that represent a wide variety of relative sizes for molecules **a** and **b** (30–2500 atoms). These are two hemoglobin variants: human deoxyhemoglobin (23) (designated 2HHB) and horse methemoglobin (24) (designated 2MHB), representing naturally occurring heterodimers; and three complexes: tRNA synthetase–tyrosinyl adenylate (25) (designated 3TS1), aspartic proteinase–peptide inhibitor (26) (designated 3APR), and trypsin–trypsin inhibitor (27) (designated 2PTC). In these tests, the component molecules were treated as separate entities by using their respective atomic coordinates within the complex. Additional tests were performed with native aspartic proteinase (28) and its peptide inhibitor (designated 2APR) and with trypsin and native trypsin inhibitor (29) (designated 4PTI). The relative position of the molecules yielding the best geometric fit in a complex, as determined by the algorithm, was finally compared with the corresponding known complex.

The results are summarized in Fig. 4. It shows histograms of 10 correlation peaks for each pair of molecules. The left side of each panel presents the highest 10 peaks obtained at the scan stage, whereas the right side shows the peaks reevaluated for the same 10 orientations in the discrimination stage. As evident from the figure, the correlation peak for the known complex (shaded) is not necessarily the highest in the scan stage. However, the highest peak that was obtained after discrimination represents the right orientation and position of molecule **b** with respect to **a**, and it is significantly higher than the other peaks.

Application of the algorithm to the α and β subunits of human hemoglobin (2HHB in Fig. 4*a*) revealed that the highest peak at the scan stage (score 312), corresponds to the well-known α–β dimer. In the horse methemoglobin variant, however (2MHB in Fig. 4*b*), the correct position for the dimer is represented by the third peak (score 290) in the sorted histogram for the scan stage. Nevertheless, both these peaks became predominant at the discrimination stage (scores 302 and 347 for 2HHB and 2MHB, respectively). The hemoglobin molecules contain two α–β dimers symmetrically arranged so that each α subunit is in contact with two β subunits. The algorithm should thus yield, in principle, two major correlation peaks for the interaction between α and β subunits. The first, mentioned above, corresponds to the tight contact between the subunits of the α–β dimer, and the other corresponds to the looser contact between the α subunit of one dimer with the β subunit of the other. This second expected peak (not shown) was rather low (scores 190 and 178 for 2HHB and 2MHB, respectively), so it was not included among the 10 peaks in the scan stage. However, it was enhanced upon recalculation with the finer grid (scores 260 and 185, respectively), in contrast with the spurious peaks, which were all reduced. The relation between the extent of geometric fit in these two associations may reflect the well-known higher stability for the interdimer association.

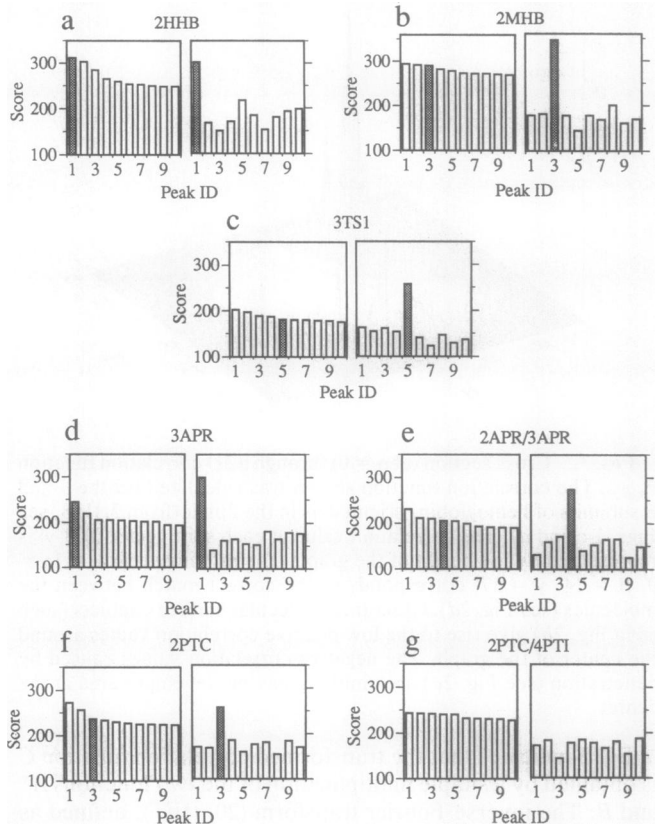

FIG. 4. Correlation results for different pairs of molecules. The pairs are identified by their respective codes (see text). In each panel, the histogram on the left shows the 10 highest correlation peaks obtained in the scan stage (η = 1.0–1.2 Å), sorted by their score. Each of these peaks was obtained at a different relative orientation of the molecules and corresponds to a potential geometric match. The shaded peak in each histogram corresponds to the known complex between the molecules considered. The histogram on the right side of each panel shows the scores obtained at the discrimination stage (η = 0.7–0.8 Å), for the 10 orientations singled out in the scan stage. Note that in the discrimination stage the spurious peaks (plain) are suppressed, whereas the correct peak (shaded) becomes prominent.

Next, we applied the algorithm to the tRNA synthetase-tyrosinyl adenylate pair (3TS1 in Fig. 4*c*), which served as an example for a complex between a high molecular weight protein and a small ligand. In this case the correlation peak, which corresponds to the correct position of the ligand in the complex, was not the highest one at the scan stage. However, discrimination yielded the expected result—i.e., the correct orientation was associated with a peak distinctly higher than the other peaks.

Further assessment of the procedure was carried out by analyzing the complex between aspartic proteinase and its peptide inhibitor (3APR in Fig. 4). This system illustrates a case in which the structure of the protein in the complex closely resembles that of the native protein (26, 28). It is thus possible to look for the best match between the structure of the complexed peptide and the protein, either in its complexed (3APR) or native (2APR) structure. With the complexed protein, the correct relative position of the ligand yielded the highest-peak already at the scan stage (Fig. 4*d*), whereas with the native protein, the peak describing the correct position was only the fourth in the sorted list (Fig. 4*e*). However, the hierarchy of the peaks changed markedly in the discrimination stage, where the highest correlation peak indicated a structure closely resembling that of the

known complex. When the native protein is used, the correlation peaks at both stages are somewhat lower than the corresponding ones for the protein in the complex, indicating a slightly poorer fit.

Analysis of the complex trypsin–trypsin inhibitor (2PTC in Fig. 4) was chosen because the native structure of one of the components, the inhibitor, differs from that in the complex. Specifically, conformational changes involving the side chains of three amino acids, located in the binding site of the inhibitor, occur upon complex formation (27, 29). When the structure of the inhibitor in the complex was used (Fig. 4*f*), the highest peak after discrimination corresponded to the correct position of the inhibitor in the complex. However, when the native structure of the inhibitor (4PTI) was used (Fig. 4*g*), the algorithm did not yield a distinct correlation peak neither in the scan stage nor in the discrimination stage. This result indicates that the extent of the conformational change occurring at the surface of the inhibitor upon binding to trypsin exceeds that tolerated by the algorithm.

CONCLUSION

Our geometry-based algorithm predicts the structure of complexes formed between the two constituent molecules by using their atomic coordinates, without any prior information as to their binding sites. The molecular surfaces need not undergo transformation except a simple 3D digitization; thus, all the surface geometric features are fully preserved within the accuracy of the grid step size. The values chosen for the parameters of the algorithm are general and do not have to be readjusted for each molecular pair. Our algorithm exploits Fourier transformation and correlation techniques, so that all possible associations between the molecules are evaluated much more rapidly than the equivalent exhaustive search in six dimensions. Another important feature of the algorithm is that the computation time is approximately proportional to $k\ln(k)$, where k is the number of atoms in the complex. Consequently, the increase in computation time with larger molecules is moderate.

We tested our algorithm on five known complexes, for which the correct structure of the complex was predicted from the atomic coordinates of the component molecules within the complex. A test carried out using the coordinates of native aspartic proteinase (see Fig. 4*e*) also resulted in the prediction of the correct known complex structure. However, when the algorithm was applied to trypsin and its native inhibitor, no distinct match was found (see Fig. 4*g*). This is most likely due to the known conformational change in the trypsin inhibitor binding site upon complex formation (27, 29) (see also refs. 4, 18, and 19). The results of our tests indicate that as long as the conformational changes are small, the algorithm may be used successfully to predict the structure of hitherto unknown complexes from the structure of two known components. Further enhancements of the algorithm are presently being developed to introduce some physical features to the molecular interface, such as surface charges and degrees of hydrophobicity.

We thank I. Steinberg for helpful discussions and A. Heimrath and D. Revacha for technical assistance. M.E. acknowledges support from the Kimmelman Center for biomolecular structure and assembly; C.A. and I.A.V. thank the Ministry of Absorption and "Fondation RASCHI" for partial financial support; and I.S. thanks the Ministry of Science and Technology for support.

1. Wodak, S. J. & Janin, J. (1978) *J. Mol. Biol.* **124,** 323–342.
2. Goodford, P. J. (1985) *J. Med. Chem.* **28,** 849–857.
3. Billeter, M., Havel, T. F. & Kuntz, I. D. (1987) *Biopolymers* **26,** 777–793.
4. Warwicker, J. (1989) *J. Mol. Biol.* **206,** 381–395.
5. Goodsell, D. S. & Olson, A. J. (1990) *Proteins* **8,** 195–202.
6. Yue, S.-Y. (1990) *Protein Eng.* **4,** 177–184.
7. Greer, J. & Bush, B. L. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 303–307.
8. Kuntz, I. D., Blaney, J. M., Oatley, S. J., Langridge, R. & Ferrin, T. E. (1982) *J. Mol. Biol.* **161,** 269–288.
9. Zielenkiewicz, P. & Rabczenko, A. (1984) *J. Theor. Biol.* **111,** 17–30.
10. Zielenkiewicz, P. & Rabczenko, A. (1985) *J. Theor. Biol.* **116,** 607–612.
11. Fanning, D. W., Smith, J. A. & Rose, G. D. (1986) *Biopolymers* **25,** 863–883.
12. Novotny, J., Handschumacher, M., Haber, E., Bruccoleri, R. E., Carlson, W. B., Fanning, D. W., Smith, J. A. & Rose, G. D. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 226–230.
13. Connolly, M. L. (1986) *Biopolymers* **25,** 1229–1247.
14. DesJarlais, R. L., Sheridan, R. P., Seibel, G. L., Dixon, J. S., Kuntz, I. D. & Venkataraghavan, R. (1988) *J. Med. Chem.* **31,** 722–729.
15. Chirgadze, Y., Kurochkina, N. & Nikonov, S. (1989) *Protein Eng.* **3,** 105–110.
16. Lewis, R. A. & Dean, P. M. (1989) *Proc. R. Soc. London Ser. B* **236,** 141–162.
17. Wang, H. (1991) *J. Comput. Chem.* **12,** 746–750.
18. Jiang, F. & Kim, S. H. (1991) *J. Mol. Biol.* **219,** 79–102.
19. Schoichet, B. K. & Kuntz, I. D. (1991) *J. Mol. Biol.* **221,** 327–346.
20. Elliott, D. F. & Rao, K. R. (1982) in *Fast Transforms: Algorithms, Analyses, Applications* (Academic, Orlando, FL), pp. 58–90.
21. Papoulis, A. (1962) in *The Fourier Integral and Its Applications* (MacGraw-Hill, New York), pp. 244–245.
22. Goldstein, H. (1980) in *Classical Mechanics* (Addison-Wesley, Reading, MA), p. 608.
23. Fermi, G., Perutz, M. F., Shaanan, B. & Fourme, R. (1984) *J. Mol. Biol.* **175,** 159–174.
24. Ladner, R. C., Heidner, E. G. & Perutz, M. F. (1977) *J. Mol. Biol.* **114,** 385–414.
25. Brick, P., Bhat, T. N. & Blow, D. M. (1989) *J. Mol. Biol.* **208,** 83–98.
26. Suguna, K., Padlan, E. A., Smith, C. W., Carlson, W. D. & Davies, D. R. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 7009–7013.
27. Marquart, M., Walter, J., Deisenhofer, J., Bode, W. & Huber, R. (1983) *Acta Crystallogr. Sect. B* **39,** 480–490.
28. Suguna, K., Bott, R. R., Padlan, E. A., Subramanian, E., Sheriff, S., Cohen, G. H. & Davies, D. R. (1987) *J. Mol. Biol.* **196,** 877–900.
29. Wlodawer, A., Deisenhofer, J. & Huber, R. (1987) *J. Mol. Biol.* **193,** 145–156.