A fast computer algorithm for finding an optimum geometrical interaction of two macromolecules

Martin Šantavý and Jaroslav Kypr*

Faculty of Nuclear Sciences and Physical Engineering, Technical University, Prague, Czechoslovakia

*Institute of Biophysics, Czechoslovak Academy of Sciences, Brno, Czechoslovakia

A computer algorithm that finds an optimum interaction of two rigid bodies made up of many hard spheres is reported. The interaction is considered optimal if the bodies have a maximum interface area. The algorithm is very fast and depends only slightly on the size of the interacting bodies so that it can be used, for example, to scan the acceptability of 10^5–10^8 relative positions of very large biomacromolecules.

Keywords: biomacromolecular interaction, biomacromolecule surface representation, interacting biomacromolecules interface, geometric approach

received 30 January 1984, revised 19 April 1984

Finding a position of equilibrium of molecules in various types of complexes is a problem of general interest. The problem is particularly significant in biology, where molecular interactions play crucial roles in elementary processes in the cell. One can mention, for example, proteins like repressors that bind to DNA and regulate gene activity.

The biomacromolecular interactions are governed by mechanisms about which there has been little experimental information for a long time. However, recent successes in crystal structure determination of some DNA binding proteins strongly suggest principles of DNA-protein interactions[1]. One of these principles is employed here to formulate a fast computer algorithm which searches for an optimum interaction between two rigid bodies composed of many hard spheres.

FORMULATION OF THE PROBLEM

Consider two rigid bodies composed of many hard spheres, which may, for example, represent a DNA fragment and a repressor, and search for their optimum position in the complex. For this purpose an object function should be defined, saying which position is better and which is worse. Our notions about the acceptability of the complex are derived from the experiments with the DNA binding proteins[1]. They suggest a substantial geometric complementarity of surfaces of the participating biomacromolecules.

We are going to apply the algorithm to scan a very large number of relative positions of molecules that have 10^3–10^4 atoms. Consequently, characterization of the interaction by means of atom pair potentials including all atoms in both molecules is unthinkable since the position acceptability should be evaluated very fast. In addition, the time required for the evaluation should be almost independent of the number of atoms, N, in the molecules. These requirements are fulfilled by introducing the interface of interacting molecules.

Molecular surface and the interface

Before the interface of interacting molecules can be made up one should define the surface of a single molecule. Several types of surfaces have been introduced[2]. A test sphere is usually adopted (its radius often equals the van der Waals radius of a water molecule) that is rolled on the van der Waals surface of the molecule. The centre of the test sphere moves on a surface that is called the accessible surface of the molecule. Its generation is relatively straightforward but it characterizes inappropriately the molecule properties that we need in the present study. We should work with the molecular surface, which is the sheet on which the test sphere rolls.

Direct generation of the molecular surface is difficult and time-consuming. Hence, we derived an equation which transforms the accessible surface into the molecular. For this purpose a space-fixed reference system XYZ was introduced to localize the molecule. Consider a part, S_A, of the accessible surface and a function f_A that maps the X coordinates of the points in S_A onto the YZ plane (Figure 1). Thus for any point $(y, z) \in \Omega$ the point $(f_A(y, z), y, z)$ falls in S_A, where Ω is the universal set. Similarly a function, f_M, maps points in the molecular surface, S_M, onto the YZ plane. There is a relation between f_A and f_M which can be shown to be:

 0263-7855/84/020047-03 $03.00

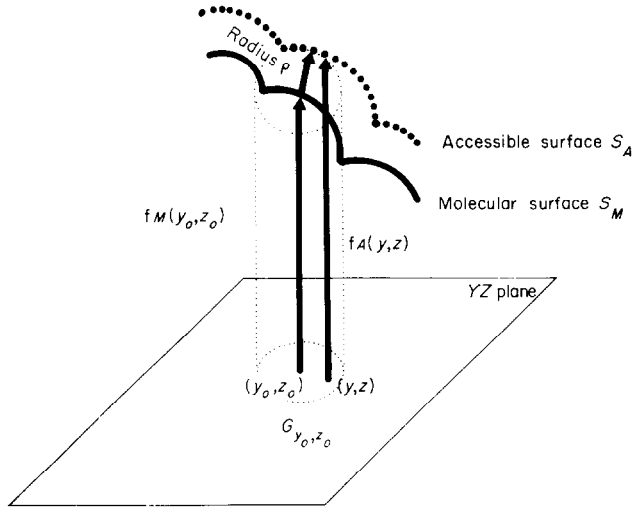

Figure 1. A diagram indicating the quantities used to transform the accessible surface S_A *into the molecular* S_M

$$f_M(y_o, z_o) = \min_{(y,z) \in G_{y_o,z_o}} (f_A(y,z) - (\rho^2 - (y - y_o)^2 - (z - z_o)^2)^{1/2})$$

where G_{y_o,z_o} is the neighbourhood set of points in the circular area with radius ρ, in the YZ plane, around the point (y_o, z_o). The relation holds for all (y_o, z_o) such that $G_{y_o,z_o} \in \Omega$ and was used in the present study to generate the molecular surfaces.

Take two nonpenetrating molecules with the molecular surfaces S_1 and S_2. There is a gap between them that can be filled with spheres which simultaneously touch both surfaces. The centres of these spheres create what we call the interface. Centres of spheres that touch both molecules simultaneously and have radii less than ϵ make up domains C_ϵ in the interface. It is convenient to define the total area $\mu(C_\epsilon)$ of the domains as a value of the object function characterizing the relative orientation of the molecules. Let $g(y, z)$ be a function such that $C_\epsilon = \{(g(y, z), y, z) \mid (y, z) \in \Gamma\}$, Γ is a projection of $C\epsilon$ into the YZ plane. Then the object function:

$$\mu(C_\epsilon) = \iint_\Gamma [1 + (\partial g/\partial y)^2 + (\partial g/\partial z)^2]^{1/2} dy\, dz$$

NUMERICAL REPRESENTATION OF MOLECULAR SURFACES AND CALCULATION OF THE OBJECT FUNCTION

Consider two molecules called B and C located in the reference system XYZ. Naturally, the object function is expected to have the largest values in positions when the molecules touch. Consequently, computations were only performed for such cases. In order to get the molecules to touch, a rectangular equidistant grid with a grid point separation s is introduced in the YZ plane. Those parts of the molecular surfaces of B and C that seem to participate in the interaction are represented by matrices B_{ij} and C_{ij}. They contain X coordinates of those points in the circumscribed regions of the molecular surfaces whose projections into the YZ plane coincide with the grid points. Having B_{ij} and C_{ij} determined, a matrix D_{ij} can be constructed:

$$D_{ij} = (B_{ij} - C_{ij} - m)/2, \quad \text{where } m = \min_{i,j} (B_{ij} - C_{ij}).$$

Thus, by the subtraction of m, the molecule C is shifted in the X direction so that in D_{ij} the two molecules are in contact at their closest point (Figure 2). The value of the object function is in this discrete approximation computed using the formula:

$$\mu \approx s^2 \sum_{(y_i,z_j) \in \Gamma} [1 + (\partial g(y_i, z_j)/\partial y)^2 + (\partial g(y_i, z_j)/\partial z)^2]^{1/2},$$

where (y_i, z_j) are the grid points in the YZ plane and $\Gamma = \{(y_i, z_j) \mid D_{ij} < \epsilon\}$. In a preliminary calculation the derivatives $\partial g(y_i, z_j)/\partial y$ and $\partial g(y_i, z_j)/\partial z$ can be neglected. If the calculations are to be more accurate, then we put:

$$\partial g(y_i, z_j)/\partial y = (b_{i+1,j} + c_{i+1,j} - b_{i-1,j} - c_{i-1,j})/4s,$$

$$\partial g(y_i, z_j)/\partial z = (b_{i,j+1} + c_{i,j+1} - b_{i,j-1} - c_{i,j-1})/4s,$$

where $b_{i,j}$ and $c_{i,j}$ are elements of the matrices B_{ij} and C_{ij}.

The grid representation enables a simple transfer of the molecules to touch and a fast numerical computation of the value of the object function. But it has other advantages. First, having generated the matrices B_{ij} and C_{ij}, the object function can be evaluated not only in one orientation but also in many others which we call related orientations. They arise when the molecules are shifted in the directions perpendicular to the X axis by integer multiples of s or by relative rotations by $\pi/2$ around the X axis. Surface matrices in the original and related orientations are simply interconnected. This makes the algorithm very fast and largely independent of the number N. Second, when searching for m only those values of D_{ij} are pushed down in the memory structure 'STACK' which may contribute to the value of the object function.

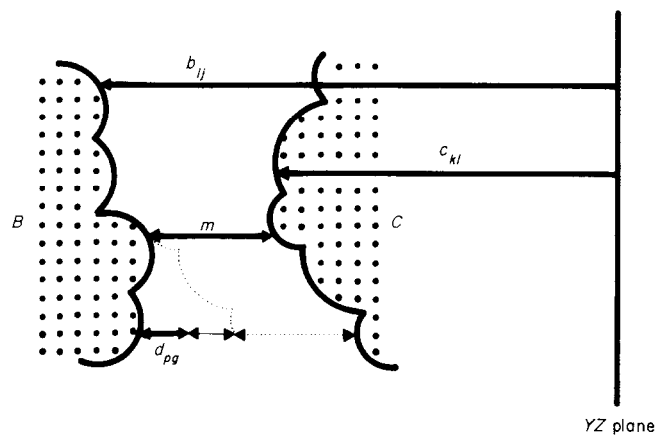

Figure 2. A diagram showing the quantities used to evaluate the matrix D_{ij}

ORGANIZATION OF THE COMPUTATION

The computation is organized according to the following schedule:

1 Read the atomic coordinates of the molecules B and C and declare for what orientations you wish to know the value of the object function.
2 Take consecutive orientations of the molecule B in the reference system XYZ.
  2 I Generate B_{ij}.
  2 II Take consecutive orientations of the molecule C.
    2 II i Generate C_{ij}.
    2 II ii For given ϵ's, compute values of the object function in the original and related orientations.
    2 II iii Put out the results.
    2 II iv If there is an orientation of C to be evaluated, go to 2 II. If not, go to 3.
3 If there is an orientation of B to be evaluated, go to 2. If not, go to 4.
4 End.

The schedule ensures that the value of the object function is evaluated in all relative positions of the molecules that one requires. Construction of the algorithm permits the simultaneous performance of most operations. The use of a multiprocessor computer thus would speed up the computation substantially. The average time of computing a value of the object function is also substantially shorter if the computation is simultaneously carried out for several ϵ's and many related orientations.

Test example

We have written a computer program GIFALG (geometrical interaction — fast algorithm) in the standard Fortran IV which made testing of the reported algorithm possible on an ICL 4-72. The program is available upon request. We tested interactions of two rigid bodies composed of several hundred spheres whose surfaces can be characterized as an ellipsoid and ellipsoidal hollow in a plane. It is clear that the object function of this system is maximal when the ellipsoid is in the hollow and minimal when the ellipsoid only touches an edge of the hollow.

The computation confirmed our views and, furthermore, showed that $\mu(C_{\epsilon''}) > \mu(C_{\epsilon'})$ if $\epsilon'' > \epsilon'$, which is a desired property of the object function. When ϵ was small there was substantial noise in the dependence of the object function on the orientation of the molecules. In this case the object function is too sensitive to minute unevenness of the molecular surfaces. On the other hand, when ϵ was too large the object function was not sensitive enough, and lost the property of characterizing the interaction quantitatively. We found that ϵ was optimal when it was comparable to the radii of the molecule atoms. By way of illustration, we depict the dependence of the object function on a shift of the tested ellipsoidal bodies in the YZ plane when other parameters of the interaction were kept constant (Figure 3). The region where the values of the object function are maximal corresponds to the situation when the ellipsoid is in the ellipsoidal hollow.

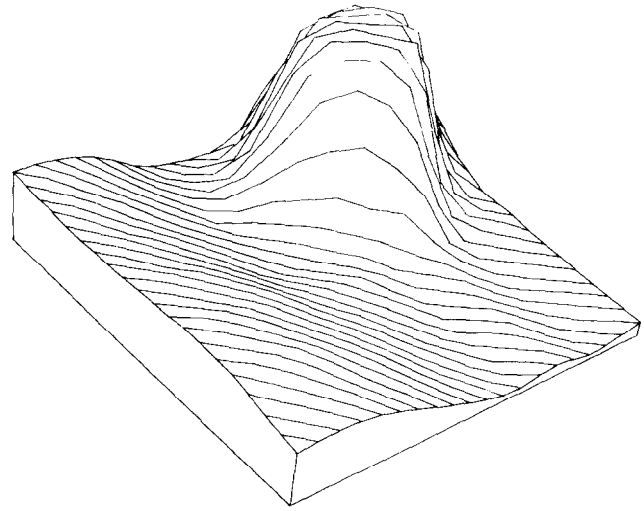

Figure 3. Dependence of the object function on a shift of the tested ellipsoidal bodies in the YZ *plane*

CONCLUSIONS

The algorithm presented here allows those positions of two macromolecules in the complex to be found where their geometrical interaction is optimal. It has two main advantages in comparison to a similar proposed algorithm[3]. First, it is much faster; the computer time to calculate dependences shown in Figure 1, where we worked with 60 × 60 molecular surface matrices and when 20 × 20 related orientations of the molecules were under consideration, never exceeded a few seconds. Second, the algorithm works fast even when very large molecules are studied. These properties make it convenient for scanning the whole conformational space because 10^5–10^8 positions of the two molecules can be examined in a reasonable time. Consequently, it can be used to find starting positions for energy computations of stable molecular complexes or as an auxiliary tool to monitor properties of molecular complexes in computer graphics studies. The algorithm will readily accommodate further principles observed experimentally in macromolecular complexes. However, it may also provide valuable biological information used on its own.

ACKNOWLEDGEMENTS

We are grateful to Dr A J Morffew and the referees for helpful suggestions.

REFERENCES

1 **Takeda, Y, Ohlendorf, D H, Anderson, W F and Matthews, B W** *Science* Vol 221 (1983) pp 1020–1026.
2 **Richards, F M** *Ann. Rev. Biophys. Bioeng.* Vol 6 (1977) pp 151–176
3 **Kuntz, I D, Blaney, J M, Oatley, S J, Langridge, R and Ferrin, T E** *J. Mol. Biol.* Vol 161 (1982) pp 269–288