TOM: a FRODO subpackage for protein-ligand fitting with interactive energy minimization

Christian Cambillau* and Eduardo Horjales

Swedish University of Agricultural Sciences, Department of Molecular Biology, Biomedical Center, Box 590, S-75124, Uppsala, Sweden

We have implemented the docking subpackage TOM into FRODO[1] for studying protein-ligand interactions with interactive energy-minimization procedures. When using TOM, models of protein-ligand complexes are first created in a graphics display unit, followed by interactive energy-minimization treating both ligand and parts of the receptor as flexible units. The potential energy function includes Coulomb and van der Waals interactions. At present, two versions of TOM are available, running on a DEC VAX computer, an Evans and Sutherland PS300 display unit, or a Silicon Graphics IRIS workstation.

Keywords: energy minimization, ligand fitting

Received 22 July 1987
Accepted 26 August 1987

One of the most challenging goals in pharmaceutical research is the capability to perform rational drug design. This involves mainly comparisons between a biologically active molecule with a known three-dimensional (3D) structure and a model one. Molecular graphics and quantum mechanics are the current tools. The increasing number of protein structures solved by X-ray crystallography or built from related structures[2] has led biologists to use this information in cases where the known protein is the key of a biological problem. This includes, for instance, enzyme/ligand interaction studies.[3-5] Moreover, a profitable use of site-directed mutagenesis[6,7] allowing the specific replacement of a residue requires knowledge of the protein structure.

For techniques such as energy minimization, reliable potential energy functions describing the atom-atom interactions in proteins are essential.[8-11] Yet the potential energy functions used in molecular dynamics provide a sufficient basis for the development of fast energy-minimization docking programs coupled with model-building systems for studying protein-ligand interactions.

Structural flexibility is relevant for the process of binding small molecules to proteins. Methods for introducing this flexibility into the model building are required for studying protein-ligand interactions. Energy-minimization procedures can be used to evaluate part of these flexibilities.

This paper describes TOM, a FRODO subpackage that includes an energy-minimization procedure called FIT that considers the translational, rotational and internal torsional degrees of freedom of the ligand as variables and the torsional angles of neighboring protein side chains. The potential energy function used includes Coulomb and van der Waals interactions. Some programs that perform ligand fitting have already been described: Busetta *et al.*[12] developed the program DOCKER for studying the interactions of a protein with a bound peptide. Bush[13] developed extensions of FRODO, which have been used to study thermolysin substrate and inhibitor binding.[14] Finally, Muller[15] has reported the use of a search for low-energy conformers of an isolated ligand when designing inhibitors of dihydrofolate reductase.

PROGRAM DESCRIPTION

We have chosen FRODO as the basis for our fitting subpackage TOM. FRODO is a flexible program that has been implemented on many types of displays. It has all the facilities needed to manipulate a protein model and any small molecule as a ligand. To incorporate a ligand into a model, TOM needs a description of the connectivity and the geometry of the molecule. Both are stored in a special file organized as the Z-matrix, usually encountered in quantum mechanical studies (see, for example, GAUSSIAN 70, Hehre *et al.*[16]). The file also specifies which torsional angles will be considered as flexible ones and includes atomic partial charges and atom types. The geometry is described through a series of bond distances, bond angles and torsion angles. These data can be included directly in the Z-file or calculated by TOM, provided a set of coordinates for the ligand is known. TOM does not include at present any optimization of the internal bond lengths or bond angles of the ligand; this can be considered a limitation of the program.

The choice of the geometric variables to be optimized in the energy minimization is critical to obtain an interactive program using minimum CPU time without intro-

This paper is taken in part from the thesis of E. Horjales, Uppsala, 1985

*To whom correspondence should be sent. Present address: CRMC2-CNRS, Campus de Luminy, case 913, 13288 Marseille-Cedex 09, France

 0263-7855/87/040174-04 $03.00

ducing unreliable simplifications. Some of the arguments used to make the choice described above are the following. Torsion angles of side chains present low-energy barriers and large movements, as seen in the usually larger observed crystallographic temperature factors.[17] Furthermore, neighboring side chains may be involved in concerted motions with very low-energy barriers. Energy minimization of flexible side chains naturally takes into account these structural fluctuations, which are very difficult to simulate and evaluate with manual model-building techniques.

Main chains are often part of secondary structures, allowing them less flexibility and often requiring higher energies to displace them from their equilibrium positions. Furthermore, when main chains are considered to be fixed, the kinematics of the system are highly simplified, because the movement of each side chain does not depend on the variables that determine the position of the other side chains. This speeds up the calculation of the energy and its derivatives.

The procedure FIT uses a conjugate gradient procedure[18] for the energy minimization. Two different energy functions are available. Function 1 can be used to relieve short contacts present in the initial model, idealizing in this way the initial geometry. It is the sum of a Lennard–Jones term for each pair of nonbonded atoms and a cosine function for the atom pairs related through a torsion angle. The Lennard–Jones term has a minimum at a distance of 3.5 Å for all atom pairs. When bad contacts are present in the initial model, the effect of this function is to relieve these contacts, thus idealizing the initial contacts. If no bad contacts are present, the attractive terms of the function dominate. The function can be applied to crystallographic data sets without including the charged hydrogen atoms.

Function 2 is a potential energy function that includes terms for nonbonded and torsional interactions used in the Groningen molecular dynamics system (GROMOS).[8] This is the sum of the Coulomb interaction and a Lennard–Jones potential with van der Waals radii depending on the atom type. Furthermore, different repulsive terms are used for nonpolar and polar interactions. The parameters used are those published by Hermans *et al.*[8] A shorter van der Waals radius is used for 1 to 4 atom pairs of each torsional interaction.[19] To use this potential, it is necessary to incorporate the partially charged hydrogen atoms into the coordinate data set, which can be done with special commands included in TOM.

Macromolecular model-building programs divide the structure into units, be they protein residues, cofactors, or ligands. The structure is generated as a sequence of units of definite type. In TOM, as in FRODO, all standard unit types are listed in a dictionary, including the names of the atoms and the ideal geometry for each unit type.

For each unit type, the dictionary contains one line for each atom in the unit. The first datum in each line is the atom name. This is followed by the connectivity indices (the same included in the Z-file for the ligand), the partial charge and the atom type.[8] The coefficients for defining the van der Waals interaction for each atom type and a matrix indicating whether the interaction between two atom types is polar or nonpolar are listed at the end of the dictionary.[8]

In TOM the ligand is not included in the dictionary, but all the corresponding information is taken from the Z-file. In this way, changes in the chemical formula or the geometry of the ligand can be introduced easily.

FIT uses a central unit, which is usually the ligand, to define a flexible and a fixed zone of atoms. These are selected as the first step when running FIT. These zones are defined using two distances, $R1$ and $R2$. The flexible zone includes atoms of the central unit, as well as all protein side chains containing some atom at a distance less than $R1$ from any atom in the central unit.

If the ligand is not the central unit and contains some atom at a distance less than $R1$ from any atom in the central unit, it is also included in the flexible zone. When a protein residue is selected as the central unit instead of the ligand, only the side chain is considered as flexible, and rotations or translations are disabled.

The present version of TOM also defines an additional file, the ZONE-file. This file is an alternate way to define the list of flexible and fixed residues. In this way, the calculations will always include the same interacting pairs, and thus the energy values resulting from different runs can be compared.

The value $R2$, which is used to select the fixed zone, is also considered as an energy cutoff: The interaction between atom pairs whose distance is larger than $R2$ is taken as zero. Furthermore, those pairs whose distance is larger than ($R2$-1Å) are smoothed using an exponential coefficient to avoid oscillations appearing with the inclusion or exclusion of interactions.

The potential mentioned above does not properly describe the covalent bonds often found between the protein and the ligand. To overcome this limitation, it is possible to assign special interaction potentials to some of the atomic pairs. Harmonic or Lennard–Jones functions can be chosen as special interactions. They are defined by giving the distance at which the function has a minimum and a constant defining the magnitude of the potential.

SOME APPLICATIONS

Model-building and energy-minimization procedures with TOM have been used to study protein-ligand interactions in alcohol[20,21] and sorbitol[22] dehydrogenases, uteroglobin,[23] serine proteases,[24] and retinol binding protein.[25]

- Docking results from a complex between alcohol dehydrogenase, LADH, and a pyrazole derivative have been compared with crystallographic results for this complex. Different strategies for using computer graphics with interactive energy minimization were tested and discussed.[20] A productive substrate binding mode in liver alcohol dehydrogenase for secondary alcohols was determined. These docking results have been compared to some of the extensive amount of kinetic data available for this enzyme that has been organized by others in terms of a diamond lattice description. Very good agreement between kinetic and model-building results was obtained.[21] A model for the tertiary structure of the sorbitol dehydrogenase subunit was built using computer graphics from its homology with LADH. Some important differences in the models of both enzymes account for the different observed stereospecificities.[22]

- A uteroglobin[26]-like protein has been found in rat lung, which binds progesteron (the uteroglobin physiological ligand) and polychlorinated biphenyl bis (methylsulfonyl) derivatives. TOM was used to fit progesteron and the biphenyl derivatives into the 3D structure of uteroglobin (Color Plates 1–3). The results show that the biphenyl derivatives bind more strongly uteroglobin than progesteron. A similar pattern has been found with rat-lung protein.[27]

HOW TO USE TOM: SOME USEFUL CRITERIA

TOM can be used, in studying a protein-ligand binding mode, to search for a complex conformation coherent with the available structural and biochemical information. The program takes into account some structural fluctuations in the system and produces information on the stability of a manually built conformation. However, the same criteria used for accepting or rejecting a model from normal model-building studies should also be applied to the results from TOM.

Some of the criteria applied when using TOM are the following:

The large number of degrees of freedom involved, even in simple problems, demands a careful search of all the possible binding conformations. It may be difficult to obtain good results when docking small or very flexible ligands that have much freedom to move in the active site and very low energy barriers. Different protein flexibility modes can be involved in the calculations, raising difficulties in the selection of the best fit. By contrast, bulky rigid ligands are highly restrained, and it is easier to find one or a few minima with high energy barriers.

The difficulties mentioned above can, in principle, be present for very flexible protein side chains. However, it is possible to overcome them if the original model of the protein is obtained from crystallographic studies. Thus, we do not need to search for the possible conformations, but starting from the crystallographic model we investigate the perturbations produced by the ligand in the crystallographic conformation of the protein. Under these conditions, TOM is very efficient in evaluating possible side-chain displacements related to the flexibility of the structure.

The GROMOS potential energy function used in FIT is a coherent set of parameters. The groups of the ligand included in the description of this potential[8] should include the charges and van der Waals radii described there. However, in docking studies it is common to find some interactions that are not defined by this potential. In this case, there are two possible solutions: (a) define the charges of the system and the van der Waals radii in order to reproduce the structural data available (as in References 20–22 for the Zn atom); or (b) replace the interaction with a special function. The choice of parameters needs to be based on some structural information.

The energy-minimization procedure FIT calculates the local minimum of the energy function, presumably the one closest to the initial conformation. The results may depend on the variables used in the minimization. Many minima obtained when assuming the enzyme to be rigid are not minima if flexible side chains are taken into account. On the contrary, if the initial model contains main-chain contacts and side chains are assumed to be flexible, the side chains can be displaced to high-energy conformations during the minimization procedure. This can lead to local minima with higher energy than those found by first optimizing the flexible ligand in a rigid enzyme and then allowing side-chain flexibility in a second run. It is therefore desirable to run all the minimizations using this two-step procedure.

FIT converges to some minimum in each run, and the problem of choosing those that are relevant to the solution of the problem in question requires some experience with the program. Models from FIT are generally better with respect to contacts than those built manually. The criteria to accept or discard the models are, as in model building, the coherence of a sufficient amount of structural and biochemical data. In addition, the value of the interaction energy can be used as another criterion showing qualitative differences between two models if the same interaction pairs are included in the calculations. This can be achieved using the special option ZONE-file for choosing the flexible and the fixed residues.

Even with the difficulties listed above, it is possible to obtain qualitative evaluations of the possible modes of binding of a proposed ligand, especially when comparing similar binding geometries. When kinetic data are available for many docked substrates, it is possible to obtain information on the geometries of productive binding modes. Energy minimizations provide information on structural flexibility and can be used to reject some of the manually built models if they are not stable. These qualitative conclusions can be very useful in guiding experimental work in areas such as drug design.

ACKNOWLEDGEMENTS

We gratefully acknowledge Professor C. I. Branden for suggesting this work to us and for constant support. We wish to express friendly gratitude to Dr. Alwyn Jones, who helped us find a way in FRODO land, saving us often from Trolls (related to Bugs, but bigger). Professor F. Quiocho kindly provided us with the PS300 version of FRODO. We are also indebted to Dr. J. P. Mornon for sending us the refined coordinates of the uteroglobin.

REFERENCES

1aJones, T. A. *J. Applied Crystallography,* 1978, **11,** 268–272
1bJones, T. A. in *Computational Crystallography* (Sayre, D., ed.), 1982, Clarendon Press, Oxford, pp. 303–317
2 Blundell, T., Sybanda, B. L., and Pearl, L. *Nature,* 1983, **304,** 273–275
3 Wipff, G. *et al. J. Am. Chem. Soc.,* 1983, **105,** 997
4 Blaney, J. M. *et al. J. Am. Chem. Soc.,* 1982, **109,** 6424–6434
5 Pincus, M. R., and Scheraga, H. A. *Accounts in Chemical Research,* 1981, **14,** 299–306
6 Winter, C., and Feersht, A. R. *Trends Biotechnol.,* 1984, **2,** 115–119
7 Villafranca, J. E. *et al. Science,* 1983, **222,** 782–788
8 Hermans, J. *et al. Biopolymers,* 1984, **23,** 1513–1518

9 Levitt, M. *J. Mol. Biol.*, 1974, **82,** 353–420
10 Momany, F. A. *et al. J. Phys. Chem.*, 1975, **79,** 2361–2381
11 Karplus, M., and McCammon, J. A. *Annual Review in Biochemistry*, 1983, **15,** 263–300
12 Busetta, B., Tickle, I. J., and Blundell, T. L. *J. Applied Crystallography*, 1983, **16,** 432
13 Bush, B. L. *Computational Chemistry*, 1984, **8,** 1
14 Hangauer, D. G., Monzingo, A. F., and Matthews, B. W. *Biochemistry*, 1984, **23,** 5730–5741
15 Müller, K. *J. Mol. Graphics*, 1984, **2,** 58
16 Hehre *et al. Quantum Chemical Program Exchange* No. 26, Chem. Dept., University of Indiana, Bloomington, IN
17 Artymiuk, P. J. *et al. Nature*, 1979, **280,** 563–568
18 Fletcher, R., and Reeves, C. M. *The Computer Journal*, 1964, **7,** 149–154
19 Aqvist, J. *et al. J. Mol. Biol.*, 1985, **183**
20 Horjales, E. Thesis, Uppsala, 1985
21 Horjales, E., and Branden, C. I. *J. Biol. Chem.*, 1985, **260,** 15445–15451
22 Eklund, H. *et al. Biochemistry*, 1985, **24,** 8005–8012
23 Cambillau, C., and Gillner, M. Unpublished results
24 Kranlis, P. Unpublished results
25 Jones, T. A. Unpublished results
26 Morize, I. *et al. J. Mol. Biol.*, 1987, **194,** 725–739
27 Lund, J. *et al. Mol. Pharmacology*, 1985, **27,** 314–323