KARLE, J. & HAUPTMAN, H. (1953). *Acta Cryst.* **6**, 131-135.
KENDALL, M. G. & STUART, A. (1977). *The Advanced Theory of Statistics*, Vol. 1, 4th ed. London: Griffin.
KLUYVER, J. C. (1907). *Proc. K. Ned. Akad. Wet.* **8**, 341-350.
MYLLER-LEBEDEFF, W. (1907). *Math. Ann.* **64**, 388-416.
PRINCE, E. (1982). *Mathematical Techniques in Crystallography and Materials Science.* New York: Springer.
ROGERS, D. & WILSON, A. J. C. (1953). *Acta Cryst.* **6**, 439-449.
SHMUELI, U., WEISS, G. H., KIEFER, J. E. & WILSON, A. J. C. (1984). *Acta Cryst.* A **40**, 651-660.
SHMUELI, U. & WILSON, A. J. C. (1981). *Acta Cryst.* A **37**, 342-353.
SPIEGEL, M. R. (1974). *Fourier Analysis with Application to Boundary Value Problems. Schaum's Outline Series.* New York: McGraw-Hill.
TITCHMARSH, E. C. (1937). *Introduction to the Theory of Fourier Integrals.* Oxford: Clarendon Press.
WEISS, G. H. & KIEFER, J. E. (1983). *J. Phys. A*, **16**, 489-495.
WEISS, G. H., SHMUELI, U., KIEFER, J. E. & WILSON, A. J. C. (1985). In *Structure and Statistics in Crystallography*, edited by A. J. C. WILSON, pp. 23-42. Guilderland, NY: Adenine Press.

Acta Cryst. (1986). A**42**, 83-85

Molecular Speleology: The Exploration of Crevices in Proteins for Prediction of Binding Sites, Design of Drugs and Analysis of Surface Recognition

BY ARTHUR M. LESK*

MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, England

(*Received* 24 *June* 1985; *accepted* 8 *August* 1985)

Abstract

A method is described for analyzing molecular-surface complementarity, including the binding of ligands to proteins or the interaction of elements of secondary structure in protein interiors. A computer program can identify and model molecules that satisfy general criteria for good binding affinity. Computational tests are presented. This approach is likely to have useful application in the analysis of surface recognition in proteins, including the identification of binding sites, and in the design of drugs for specific targets, by (i) suggesting potential pharmacophores to the medicinal chemist for further computational analysis or laboratory testing, (ii) suggestion of derivatives of a known ligand to enhance its affinity, or (iii) searching a data base of known drugs for a match to the predicted ligand.

Introduction

Emil Fischer first proposed the 'lock and key' model of enzyme-substrate interactions. We now recognize the importance of surface complementarity not only for ligand binding, but for the interactions of packed α-helices and β-sheets in protein interiors which are crucial in stabilizing native conformations (Lesk, 1981; Chothia, 1984). Important applications of computational methods for analyzing molecular complementarity include:

(1) Analysis of the packing in protein interiors: What will be the effect of a mutation on the conformation of a protein (Lesk & Chothia, 1980)? What freedom do packed secondary structures have to facilitate and transmit conformational changes (Chothia, Lesk, Dodson & Hodgkin, 1983)?

(2) Prediction of ligands complementary to specific clefts in proteins. Can we thereby design drugs of high affinity and specificity (Tickle, Sibanda, Pearl, Hemmings & Blundell, 1984; Beddell, 1984)? Can we rationalize the specificities of antibodies? With the application of protein-engineering techniques to antibodies, it will be useful to analyze changes in the antigen-binding site (Neuberger, 1983).

Given a protein structure that contains a cleft, how can one identify a ligand that has a structure complementary to the cleft? Analyses of protein-ligand interactions suggest that loss of solvent-accessible surface area, and complementarity in shape and charge distribution are the major determinants of affinity and specificity (Janin & Chothia, 1978; Chothia, 1984; Kollman, 1984). Studies of complementarity have used physical models [including making casts, using known protein structures as molds (Blow & Smith, 1975)], empirical parameters characterizing hydrophobicity (Smith, Hansch, Kim, Omiya, Fukumura, Selassie, Jow, Blaney & Langridge, 1982) and interactive computer graphics (Langridge, Ferrin, Kuntz & Connolly, 1981; Busetta, Tickle & Blundell, 1983).

We describe here a computational technique to explore clefts in proteins and suggest candidate ligands. It does not require the facilities of interactive-graphics packages, but could easily and profitably be integrated with them. [This problem should be distinguished from a related one: determining the

* Permanent address: Fairleigh Dickinson University, Teaneck-Hackensack Campus, 1000 River Road, Teaneck, NJ 07666, USA.

0108-7673/86/020083-03$01.50

optimal geometry of interaction of a protein and a *prespecified* ligand, which has come to be known as the 'docking' problem (Pattabiraman, Levitt, Ferrin & Langridge, 1984; Cambillan, Horjales & Jones, 1984). The two problems are complementary in that ligands suggested by the current work could be submitted to a 'docking' program.] Our goals are in some respects similar to those of Kuntz, Blaney, Oatley, Langridge & Ferrin (1982) and Goodford (1985).

Methods

The fundamental idea is to define, within the target cleft, a potential-energy field for possible ligand atoms. The negative of this distribution has maxima (*i.e.* peaks) at which individual ligand atoms would find themselves at positions of low energy. The problem of generating a molecule by connecting peaks into a stereochemically acceptable bonded constellation has been solved in another context: to interpret electron-density maps in X-ray crystallography (Koch, 1974; Main & Hull, 1978).

Thus, the method comprises three tasks, each of which has known computational solutions:

(1) *Establishing a potential-energy field*: Given a protein structure, place a probe atom at various points in its vicinity, and compute the interaction energy of the single atom with the protein. To model a hydrophobic group in a nonpolar cleft, a function as simple as a hard-sphere potential with an attractive r^{-6} term at higher radii may suffice (see next section). To model charged, polar, or hydrogen-bonding interactions, the potential must be correspondingly more realistic (Levitt, 1983*a*, *b*).

The result is a function $f(x, y, z)$ that is positive and very large in regions of space occupied by protein atoms, and negative in regions where a single probe atom would be energetically comfortable. To cast the next step into familiar form, note that peaks in $-f(x, y, z)$ correspond to sites of favorable interaction.

(2) *Peak-picking*: Searching a three-dimensional array [in this case, the values of $-f(x, y, z)$] and interpolating to find the positions and heights of local maxima is a standard computational task in crystallography (Main & Hull, 1978).

(3) *Peak parsing*: *extracting a stereochemically reasonable constellation of atoms from a set of peaks*: Crystallographers have solved this problem as well, in order to facilitate interpretation of electron-density maps (Koch, 1974; Main & Hull, 1978). In this work we associate with each peak a variable $x_i = 0$ or 1 ($x_i = 1$ implies that peak i is included in a structure, and $x_i = 0$ that peak i is not included), and a scalar weight w_i (*e.g.* the peak height). The quantity $\sum_i w_i x_i$ will be large for a structure that includes many strong peaks. The constraints of excluded volume, connectivity, and valence are expressible in terms of the distance matrix of the peak positions, as *linear* inequalities in the x_i.

Each of the following problems is computationally feasible:

3(*a*) Find the maximal subset of peaks obeying the stereochemical constraints (maximize $\sum_i w_i x_i$) (Geoffrion, 1969).

3(*b*) Assign a threshold T and find all subsets of peaks obeying the stereochemical constraints for which $\sum_i w_i x_i \geq T$ (Hammer & Rudeanu, 1968; Lesk, 1973).

3(*c*) Given a known ligand, the affinity of which we should like to enhance, introduce its atoms at their known positions, calculate f in the presence of the protein and ligand, determine the peaks in $-f$, and find subsets of peaks that constitute chemically feasible modifications of the given ligand.

Results

As an example we examined the specificity pocket of bovine chymotrypsin. Fig. 1 shows a substrate analog, formyl-L-trytophan, determined crystallographically by Steitz, Henderson & Blow (1969). We used coordinates of α-chymotrypsin from the Protein Data Bank (2CHA) (Bernstein, Koetzle, Williams, Meyer, Brice, Rodgers, Kennard, Shimanouchi & Tasumi, 1977), deleting the tosyl group, and a simple hard-sphere plus van der Waals-like potential: $f(r) = -r_i^{-6}$ for $r_i \geq 3{\cdot}5$ Å, $f(r) = \infty$ for $r < 3{\cdot}5$ Å, where r_i is the distance from the probe position to atom i of the protein; the summation extended over all residues within 6 Å of the pocket. The program determined the positions and heights of the peaks of $-f$ and assembled a subset into a putative molecule, shown in Fig. 2. Fig. 3 compares the experimental ligand and the predicted one. (The coordinates were *not* rotated or translated relative to each other to improve the fit; they have been rotated together, by the same amount, to a viewpoint perpendicular to the tryptophan ring.)

The program has suggested a ligand similar in nature and position to part of the experimental substrate analog and to the side chains of peptides for which the enzyme is specific. Because no polar interactions or hydrogen bonding were included in the

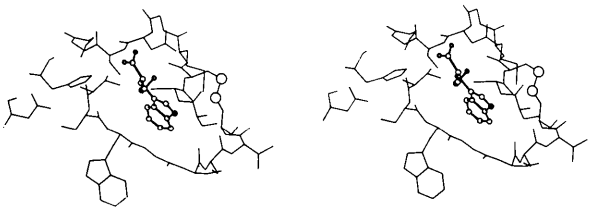

Fig. 1. Binding of the substrate analog formyl-L-tryptophan to bovine chymotrypsin; as determined crystallographically by Steitz, Henderson & Blow (1969).

potential function, none appears in the result. [Incidentally, the carboxyl group of the substrate analog was not visible in the difference map (Steitz, Henderson & Blow, 1969).]

It would be possible to compare the set of predicted peaks with the side chains of each of the twenty natural amino acids to see which ones it resembles (Lesk, 1979). In pharmacological applications, a data bank of known drugs could be searched for entries similar to a predicted ligand.

Foci of future developments include: (1) Examination of more complex potential functions, including polar interactions and hydrogen bonding. These might make it possible to identify individual atom types in the predicted ligand, or at least to assign them a charge or electronegativity. (2) Analysis of surface recognition in proteins, with the ultimate goal of predicting the geometries of interaction of α-helices and β-sheets. (3) Collaboration with medicinal chemists on applications to drug design.

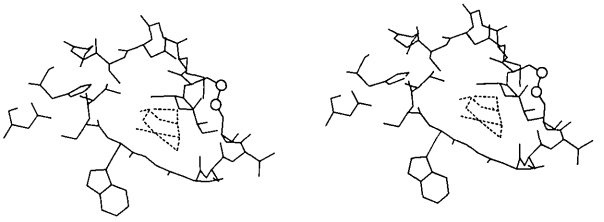

Fig. 2. Ligand fragment predicted from peaks in a hard-sphere plus van der Waals-like attractive potential. The potential function contained no polar interactions or hydrogen-bonding terms.

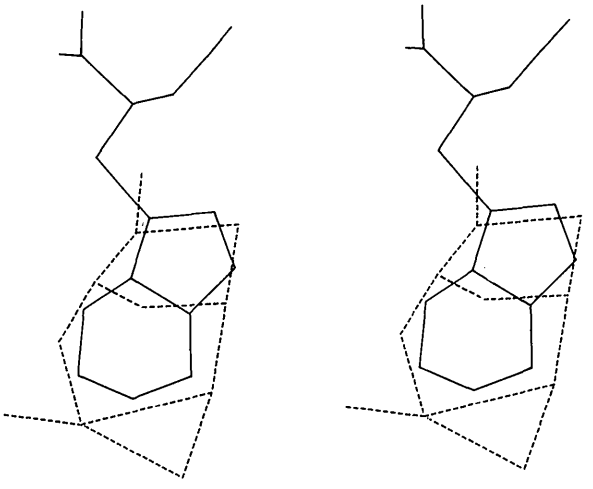

Fig. 3. Comparison of the experimental and predicted ligands. They are shown in the positions they occupy relative to the protein (not in an optimal superposition).

Work supported by US National Science Foundation research grant PCM 83-20171. I thank A. M. Geoffrion for the use of his program *RIP30C* and P. Main for the opportunity to study his computer programs.

References

BEDDELL, C. R. (1984). *Chem. Soc. Rev.* **13**, 279-319.

BERNSTEIN, F. C., KOETZLE, T. F., WILLIAMS, G. J. B., MEYER, E. F. JR, BRICE, M. D., RODGERS, J. R., KENNARD, O., SHIMANOUCHI, T. & TASUMI, M. (1977). *J. Mol. Biol.* **112**, 535-542.

BLOW, D. M. & SMITH, J. M. (1975). *Philos. Trans. R. Soc. London Ser. B*, **272**, 87-97.

BUSETTA, B., TICKLE, I. J. & BLUNDELL, T. (1983). *J. Appl. Cryst.* **16**, 432-437.

CAMBILLAN, C., HORJALES, E. & JONES, T. A. (1984). *J. Mol. Graphics*, **2**, 53-54.

CHOTHIA, C. (1984). *Annu. Rev. Biochem.* **53**, 537-572.

CHOTHIA, C., LESK, A. M., DODSON, G. G. & HODGKIN, D. C. (1983). *Nature (London)*, **302**, 500-505.

GEOFFRION, A. M. (1969). *Oper. Res.* **17**, 437-454.

GOODFORD, P. J. (1985). In preparation.

HAMMER, P. L. & RUDEANU, S. (1968). *Boolean Methods in Operations Research and Related Areas.* New York: Springer-Verlag.

JANIN, J. & CHOTHIA, C. (1978). *Biochemistry*, **17**, 2943-2948.

KOCH, M. H. J. (1974). *Acta Cryst.* A**30**, 67-70.

KOLLMAN, P. A. (1984). *X-ray Crystallography and Drug Action*, edited by A. S. HORN & C. J. DE RANTER, pp. 63-82. Oxford: Clarendon Press.

KUNTZ, I. D., BLANEY, J. M., OATLEY, S. J., LANGRIDGE, R. & FERRIN, T. E. (1982). *J. Mol. Biol.* **161**, 269-288.

LANGRIDGE, R., FERRIN, T. E., KUNTZ, I. D. & CONNOLLY, M. L. (1981). *Science*, **211**, 661-666.

LESK, A. M. (1973). *J. Comput. Phys.* **12**, 150-152.

LESK, A. M. (1979). *Commun. ACM*, **22**, 219-224.

LESK, A. M. (1981). *Introduction to Physical Chemistry*, Englewood Cliffs, NJ: Prentice-Hall.

LESK, A. M. & CHOTHIA, C. (1980). *J. Mol. Biol.* **136**, 225-270.

LEVITT, M. (1983*a*). *J. Mol. Biol.* **168**, 595-620.

LEVITT, M. (1983*b*). *J. Mol. Biol.* **170**, 723-764.

MAIN, P. & HULL, S. E. (1978). *Acta Cryst.* A**34**, 353-361.

NEUBERGER, M. S. (1983). *EMBO J.* **2**, 1373-1378.

PATTABIRAMAN, N., LEVITT, M., FERRIN, T. E. & LANGRIDGE, R. (1984). *J. Mol. Graphics*, **2**, 58-59.

SMITH, R. HANSCH, C., KIM, K. H., OMIYA, B., FUKUMURA, G., SELASSIE, C. D., JOW, P. Y. C., BLANEY, J. M. & LANGRIDGE, R. (1982). *Arch. Biochem. Biophys.* **215**, 319-328.

STEITZ, T. A., HENDERSON, R. & BLOW, D. M. (1969). *J. Mol. Biol.* **46**, 337-348.

TICKLE, I. J., SIBANDA, B. L., PEARL, L. H., HEMMINGS, A. M. & BLUNDELL, T. L. (1984). *X-ray Crystallography and Drug Action*, edited by A. S. HORN & C. J. DE RANTER, pp. 427-440. Oxford: Clarendon Press.