Journal of Medicinal Chemistry

Volume 27, Number 5 May 1984

Perspective

Drug Design by the Method of Receptor Fit

Peter J. Goodford

Laboratory of Molecular Biophysics, Oxford, England. *Received October 17, 1983*

The functions of enzymes have been studied since yeast was first used to make bread and beer, but it is only in the last 25 years that the structures of any biological macromolecules have been observed with reasonable accuracy. The molecular shape of sperm whale myoglobin was described[1] in 1958 by using X-ray crystallography, which is a method for determining the regularly recurring positions of atoms in a crystalline lattice.[2] By the early 1960's the structure of myoglobin crystals had been resolved in greater detail,[3-5] and the outline shape of horse hemoglobin had been published.[6] However, these were not enzymes, and several more years passed before the structure of hen egg-white lysozyme appeared in the literature,[7] thereby dispelling a widespread but erroneous belief that the globins had represented a special class of proteins that were uniquely amenable to X-ray methodology.

Of course an important factor that influenced the initial choice of proteins for X-ray studies was the availability of material that could be purified and crystallized. Neither sperm whale myoglobin, horse hemoglobin, nor hen lysozyme was particularly relevant to drug design or any other practical application. On this 25th anniversary we can appreciate what a great achievement it was for mankind to be able to observe the shape of myoglobin at all, and yet today it is already feasible to select a specific macromolecular target and work toward a chosen research objective like the design of a new drug. The method is advancing rapidly. Less than 10 years ago the pioneering drug-design studies in this field were directed toward hemoglobin,[8,9] because it was the most appropriate *available* protein structure, or toward angiotensin converting enzyme (ACE), whose structure was not available but could perhaps be inferred[10] by analogy with carboxypeptidase. However, progress was already being made on the X-ray crystallography of dihydrofolate reductase (DHFR), which had been *selected* as a target enzyme by Kraut and his colleagues for the rational design of specific drugs.[11,12]

Those workers at the University of California at San Diego were in liason with industrial scientists from Merck, and they chose bacterial DHFR because it was already known to be a target enzyme for effective chemotherapeutic agents.[13] Moreover, another industrial group at the Wellcome Research Laboratories was working on a very similar DHFR project[14] that led from the observed enzyme structure to the design of novel high-affinity inhibitors.[15] In fact, the DHFR study developed into a major international effort over 10 years, and many of the typical problems that can beset X-ray crystallographers were encountered and had to be overcome. It was not easy to obtain DHFR crystals at first, but the work finally showed that the method is no longer limited to particularly favorable proteins, and the list of targets today already includes the hemagglutinin glycoprotein of influenza virus,[16] carbonic anhydrase,[17] renin, phospholipase-A2,[18] and the

(1) Kendrew, J. C.; Bodo, G.; Dintzis, H. M.; Parrish, R. G.; Wyckoff, H.; Phillps, D. C. *Nature (London)* **1958**, *181*, 662-666.
(2) Blundell, T. L.; Johnson, L. N. "Protein Crystallography"; Academic Press: London, 1976, p 565.
(3) Kendrew, J. C.; Dickerson, R. E.; Strandberg, B. E.; Hart, R. G.; Davies, D. R.; Phillips, D. C.; Shore, V. C. *Nature (London)* **1960**, *185*, 422-427.
(4) Edmundson, A. B.; Hirs, C. H. W. *Nature (London)* **1961**, *190*, 663-665.
(5) Kendrew, J. C.; Watson, H. C.; Strandberg, B. E.; Dickerson, R. E.; Phillips, D. C.; Shore, V. C. *Nature (London)* **1961**, *190*, 666-670.
(6) Cullis, A. F.; Muirhead, H.; Perutz, M. F.; Rossman, M. G.; North, A. C. T. *Proc. R. Soc. London, Ser. A* **1961**, *265*, 161-187.
(7) Blake, C. C. F.; Koenig, D. G.; Mair, G. A.; North, A. C. T.; Phillips, D. C.; Sarma, V. R. *Nature (London)* **1965**, *206*, 757-761.
(8) Beddell, C. R.; Goodford, P. J.; Norrington, F. E.; Wilkinson, S.; Wootton, R. *Br. J. Pharmacol.* **1976**, *57*, 201-209.
(9) Goodford, P. J. In "Drug Action at the Molceular Level"; Roberts, G. C. K., Ed.; MacMillan: London, 1977, p 291.
(10) Cushman, D. W.; Chueng, H. S.; Sabo, E. F.; Ondetti, M. A. 1977 *Biochemistry* **1977**, *16*, 5484-5491.
(11) Matthews, D. A.; Alden, R. A.; Bolin, J. T.; Freer, S. T.; Hamlin, R.; Xuong, N.; Kraut, J.; Poe, M.; Williams, M.; Hoogsteen, K. *Science* **1977**, *197*, 452-455.
(12) Matthews, D. A.; Alden, R. A.; Bolin, J. T.; Filman, D. J.; Freer, S. T.; Hamlin, R.; Hol, W. G. J.; Kisliuk, R. L.; Pastore, E. J.; Plante, L. T.; Xuong, N.; Kraut, J. *J. Biol. Chem.* **1978**, *253*, 6946-6954.
(13) Burchall, J. J.; Hitchings, G. H. *Mol. Pharmacol.* **1965**, *1*, 126-136.
(14) Baker, D. J.; Beddell, C. R.; Champness, J. N.; Goodford, P. J.; Norrington, F. E. A.; Smith, D. R.; Stammers, D. K. *FEBS Lett.* **1981**, *126*, 49-52.
(15) Kuyper, L. F.; Roth, B.; Baccanari, D. P.; Ferone, R.; Beddell, C. R.; Champness, J. N.; Stammers, D. K.; Dann, J. G.; Norrington, F. E. A.; Baker, D.; Goodford, P. J. *J. Med. Chem.* **1982**, *25*, 1120-1122.
(16) Wilson, I. A.; Skehel, J. J.; Wiley, D. C. *Nature (London)* **1981**, *289*, 366-373.

0022-2623/84/1827-0557$01.50/0

β-lactamases,[19] as well as hemoglobin, DHFR, and ACE.

X-ray Methods

A short review does not permit full discussion of all the structural methods, but certain aspects must be mentioned before assessing the current role and future prospects for "drug design by the method of receptor fit". In X-ray crystallography the primary observations are used to calculate the electron density of a molecule in a crystal. It is the density that defines the atomic positions, and developments over the last 25 years have led to progressively more consistent results when the approach is applied to macromolecules. Methods have been developed[20a] for isolating purer proteins in larger amounts than was formerly possible and for growing crystals from smaller quantities of material. Modern apparatus allow more observations to be made with more accuracy and less crystal damage than before, so that the electron density itself is better defined. Large modern computers permit the results to be calculated with more accuracy and less subjectivity than was ever possible on earlier machines, which could never store all the data needed for such computations. The agreement of the protein structure with the primary observations is now assessed quantitatively, and an iterative process is used to improve the fit. Additional information, such as the characteristic values of bond lengths and bond angles and the amino acid sequence of the protein, is also used to give a final structure that accurately reflects the conformation of the macromolecule in the crystal (Figure 1).

It is necessary to sound cautions. A protein may have been most skillfully observed and the results most thoroughly calculated, but scientists who have trained in another discipline may still have to deal with certain well-known problems before they interpret the results. Perhaps the atoms or groups of a flexible molecule can occupy one or other of two alternative positions in the crystal, with half strength electron density occurring at each. Some parts may be able to take up several different positions, and these can appear as blurs in the electron-density map. This is particularly common at the periphery of macromolecules, where long flexible side chains, such as the $CH_2CH_2CH_2CH_2NH_3^+$ of lysine, may be floating in solvent molecules that have been trapped in the spaces between the macromolecules themselves. Even worse, peripheral amino acid residues may be somewhat distorted out of their natural positions by interactions with neighboring macromolecules in the crystal and can therefore occupy misleading positions in the electron-density map. Quite often the same macromolecules may occur in two nonequivalent orientations liable to different distortions of this type. For instance, the molecules of *Escherichia coli* DHFR with methotrexate[11] have two such positions in their crystals, although it was not actually possible to establish any definite differences between them when the observations were first published. Detailed interpretation[21,22] has now established small differences that are, for example, caused by the presence of hydrated calcium ion in the crystal, but the overall similarity of these and other DHFR structures shows that crystal-packing forces are not sufficient to distort the macromolecular structure very significantly.

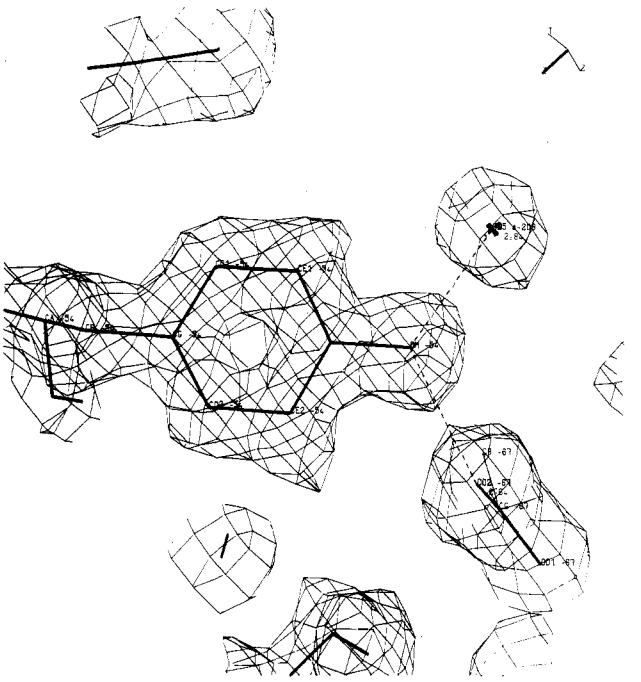


Figure 1. The X-ray observations and fitted structure of the tyrosine-54 residue and nearby groups in human lysozyme, by kind of permission of Artymiuk and Blake.[20b] The "hole" in the middle of the tyrosine ring is clearly defined, and so is the water molecule 203. This figure was objectively produced by computer, including the dashed lines, which indicate close contacts between the phenolic hydroxy group of the tyrosine and two neighboring oxygen atoms. There are no other close contacts in the figure, but these indicate one hydrogen bond with a 2.84-Å length to water molecule 203 and another of the same length to a side-chain carboxyl oxygen (OD2) of aspartic acid-67. The phenolic oxygen atom donates the hydrogen for the latter bond, since the carboxyl oxygen is not normally a donor in proteins, and the phenolic oxygen therefore accepts the hydrogen bond from the water molecule 203.

All these cautions may seem disconcerting to the scientist who has trained in another discipline, but it is actually a strength of the X-ray approach that potential artifacts and errors can be so explicitly defined. The observations themselves often show what is happening. For instance, one may observe that a group actually has partial occupancy in each of the alternative positions that it can occupy, and the thermal displacements of individual atoms can actually be determined as a guide to their local mobility in the crystal. Furthermore, the observed similarities between related molecules confirm the overall correctness of the findings, and in several cases (the *c*-type cytochromes, hemoglobin, serine proteases, acid proteases, or lysozyme as well as DHFR) there have been confirmatory comparisons between several different crystal forms or structures.

In the early days of protein crystallography it was not easy to make such comparisons. There were fewer results, and the available observations were less readily exchanged between different laboratories because of computer incompatibilities and because some scientists were rightly cautious about providing results from an untried technique

(17) Gill, E. W.; Goodall, B. M.; Hancock, P. J. B. In "Quantitative Approaches to Drug Design"; Dearden, J., Ed.; Elsevier: Amsterdam, 1983; pp 121–122.
(18) Dijkstra, B. W.; Kalk, K. H.; Hol, W. G. J.; Drenth, J. *J. Mol. Biol.* **1981**, *147*, 97–123.
(19) Dideberg, O.; Charlier, P.; Dive, G.; Joris, B.; Frere, J. M.; Ghuysen, J. M. *Nature (London)* **1982**, *299*, 469–470.
(20) (a) McPherson, A. *Methods Biochem. Anal.* **1976**, *23*, 249–345. (b) Artymiuk, P.; Blake, C. C. F. *J. Mol. Biol.* **1981**, *152*, 737–762.
(21) Bolin, J. T.; Filman, D. J.; Matthews, D. A.; Hamlin, R. C.; Kraut, J. *J. Biol. Chem.* **1982**, *257*, 13 650–13 662.
(22) Filman, D. J.; Bolin, J. T.; Matthews, D. A.; Kraut, J. *J. Biol. Chem.* **1982**, *257*, 13 663–13 672.

until everything had been most thoroughly checked. However, several developments have now promoted the ready exchange of such information. An essential requirement was the publication in 1970 of standard IUPAC–IUB abbreviations and symbols for the description of the conformation of polypeptide chains,[23] replacing an earlier (1966) convention. This enabled a standard data bank of observations to be set up[24] at Brookhaven with results tabulated in an internationally agreed format. Moreover, the protein crystallographic community started to buy compatible computers, together with compatible computer graphics systems, for displaying macromolecular structures, and similar programs are now being used on these machines in different laboratories world-wide. This standardization, coupled with a growing confidence in the results of protein crystallography, has facilitated the exchange of figures, drawings, programs, and arrays of numbers between different research groups on a reliable day to day basis.

Having said that, it is important to emphasize a significant difference between the data and the programs. While a serious effort is always made to check data before it is disseminated from The Protein Data Bank, no such overall claim can be made for computer programs. Of course the user must be on his guard for inconsistencies in the data, and more recent information may often be obtained by approaching the original workers directly. However, data checks are thoroughly and conscientiously carried out. On the other hand, programs must always be suspect because there are still differences between one computer and another, and each program should be tested by each user for each application on every computer installation.

Related Techniques

It often happens that the same protein has been observed by X-ray crystallography and other techniques. For instance, neutron diffraction[25] can detect hydrogen atoms that may have been missed by X-ray studies, although it should be emphasized that the protonation state of ionizable groups and the formation of salt bridges and hydrogen bonds (Figure 1) in proteins can often by deduced by X-ray crystallography alone.[26] NMR and other spectral techniques are particularly valuable, but caution may be needed when these findings are compared with X-ray results because the experimental conditions can be very different. When these differences are born in mind it is perhaps the compatibilities that are surprising, and spectral methods provide much additional information about structure. For instance, the pK_a of many ionizable groups[27] and detailed information on the modes of ligand binding[27-29] and on protein dynamics[30] can be determined directly by NMR.

Another important method is amino acid sequencing,[4] and the *E. coli* DHFR studies demonstrate[11] the value of a correctly determined protein sequence when interpreting X-ray observations. Moreover, protein sequences may be used to infer the three-dimensional structure of a protein by comparison with related conformations already determined for similar macromolecules. For example, renin is a proteinase that catalyzes the conversion of angiotensinogen to angiotensin I, but no renin structure has been published. However, renin sequences have been observed and compared with those of other proteinases of known structure in order to deduce a plausible renin conformation.[31] Attempts have also been made to predict the tertiary structure of proteins directly from their sequences[32,33] and to use semiempirical functions in order to compute the static structure and picosecond dynamics of protein–ligand interactions.[34-36] Furthermore, one must mention the use of ligand structures by themselves in order to deduce information about the structure of ligand-binding sites.[37-39]

The accuracy and usefulness of these computational approaches will remain open to question until enough theoretical results have been critically compared with experimental findings. However, the crystallographic and spectroscopic methods themselves are now established tools for the precise and reproducible study of macromolecular structures. Equipment, computers, programs, and data on a growing range of proteins are readily available, and it is worth considering almost any macromolecule as a potential target for drug design. The internal consistency of the observations is no longer in doubt. Comparisons between different proteins and between different methods confirm that each biological macromolecule is a unique organic chemical whose conformation and dynamics can be determined with ever-increasing accuracy. Meaningful structures of substrates, inhibitors, solvent and even water molecules[40] can be observed while they are interacting with the proteins themselves. It is too much to claim at this stage that the observed structures always explain macromolecular functions, but good structural determinations seem to be compatible with function and go a long way toward providing functional interpretations.

Observations of Ligand Binding

The relevance of enzyme structure to drug design will be considered under three general headings, beginning with structural observations on the binding of known compounds to macromolecules. The early studies of myoglobin[1] showed heme as a precisely oriented ligand in a cleft of the protein, and the traditional physiology of oxygen transport in vivo has been largely interpreted on the basis of subtle changes in this heme–globin interaction. Many

(23) IUPAC–IUB Commission on Biochemical Nomenclature *J. Mol. Biol.* **1970**, *52*, 1–17.
(24) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. *J. Mol. Biol.* **1977**, *112*, 535–542.
(25) Hanson, J. C.; Schoenborn, B. P. *J. Mol. Biol.* **1981**, *153*, 117–146.
(26) Sawyer, L.; James, M. N. G. *Nature (London)* **1982**, *295*, 79–80.
(27) Ferrige, A. G.; Lindon, J. C.; Patterson, R. A. *J. Chem. Soc., Faraday Trans. 1* **1979**, *75*, 2851–2864.
(28) Matthews, D. A. *Biochemistry* **1979**, *18*, 1602–1610.
(29) Cayley, P. J.; Albrand, J. P.; Feeney, J.; Roberts, G. C. K.; Piper, E. A.; Burgen, A. S. V. *Biochemistry* **1979**, *18*, 3886–3895.
(30) Roberts, G. C. K. *Nature (London)* **1975**, *258*, 112.
(31) Blundell, T.; Sibanda, B. L.; Pearl, L. *Nature (London)* **1983**, *304*, 273–275.
(32) Sternberg, M. J. E.; Cohen, F. E.; Taylor, W. R. *Biochem. Soc. Trans.* **1982**, *10*, 299–301.
(33) van Gunsteren, W. F.; Berendsen, H. J. C. *Mol. Phys.* **1982**, *45*, 637–647.
(34) Naylor, C. B., Cuthbertbon, A.; Richards, W. G. *J. Chem. Soc., Chem. Commun.*, in press.
(35) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Computat. Chem.* **1983**, *4*, 187–217.
(36) Blaney, J. M.; Weiner, P. K.; Dearing, A.; Kollmann, P. A.; Jorgensen, E. C.; Oatley, S. J.; Burridge, J. M.; Blake, C. C. F. *J. Am. Chem. Soc.* **1982**, *104*, 6424–6434.
(37) Gund, P.; Shen, T. Y. *J. Med. Chem.* **1977**, *20*, 1146–1152.
(38) Beddell, C. R.; Goodford, P. J. *Prostaglandins* **1977**, *13*, 493–502.
(39) Humblet, C.; Marshall, G. R. *Drug Dev. Res.* **1981**, *1*, 409.
(40) Blake, C. C. F.; Pulford, W. C. A.; Artymiuk, P. *J. Mol. Biol.* **1983**, *167*, 693–723.

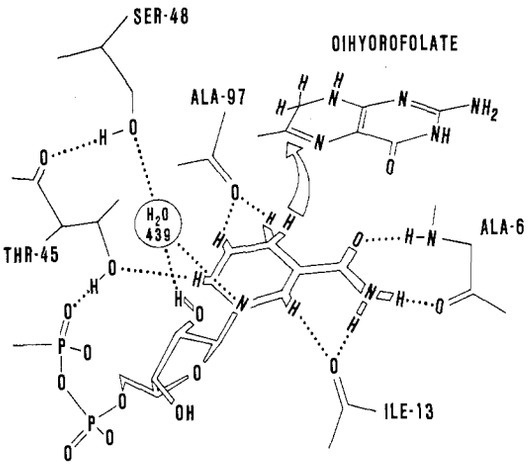


Figure 2. Schematic representation of the proposed[22] complex between *L. casei* dihydrofolate reductase, its substrate dihydrofolate, and its cofactor NADPH. The drawing is based on detailed observations like those in Figure 1, but polar interactions between the enzyme and cofactor are emphasized. The path of hydrogen transfer during catalysis and the crucial role of water molecule 439 are shown. This figure demonstrates the detailed enzyme mechanisms that can be unravelled by X-ray crystallography as a prelude to drug design.

other known compounds have been studied at their binding sites since then, and the general characteristics of binding have been established.

The ligand is often held in position by many interactions with different groups of the protein. These groups may be widely dispersed throughout the amino acid sequence, and an important function of the total macromolecular structure is apparently to hold them in the correct relative orientation. This orientation defines the structure of the cleft, which usually constitutes a binding site. The individual contributions of each group to the total binding energy may be weak, but together they can hold a ligand very tightly. Seventeen hydrophobic interactions between *Lactobacillus casei* DHFR and its cofactor NADPH have been observed[41] and 13 hydrogen bonds. The binding groups are scattered along the protein chain from the amino acid Trp-5 residue up to the Thr-126 residue of the sequence. Of these 30 interactions that contribute to the binding energy, 18 come from conserved groups that occur in the amino acid sequence of all known dihydrofolate reductases or from conserved contacts with the carbonyl and amido groups of the protein backbone. The other 12 contacts are not conserved and presumably vary from one DHFR to another, making different contributions to the total interaction energy. Furthermore, the NADPH can also make direct contact with the inhibitor methotrexate, which binds in a neighboring enzyme cleft or with the substrate during catalysis (Figure 2).

The complexity of this protein–ligand relationship is not atypical, and detailed study is always needed in order to provide a convincing interpretation of ligand binding. The contacts between heme and globin[42] or between the thyroid hormones (T3 and T4) and prealbumin[43] provide similar examples. Moreover, some ligands may have more than one binding mode, and in other cases an apparently trivial change in the structure of a ligand can cause drastic and quite unexpected alterations in the geometry of binding. For example, the inhibitor methotrexate was designed many years ago by considering the structure of natural folates, which are enzyme substrates, and the underlying assumption in this approach must have been that substrate and inhibitor would bind similarly because their structures were similar. Yet recent evidence suggests that the heterocyclic ring of methotrexate may be turned upside down at the receptor cleft of DHFR when compared with substrate,[21] and the original concepts may have been misconceived because insufficient attention was given to the electronic properties of the ligands.

The study of protein–ligand systems has been worthwhile because it has thrown new light on our understanding of enzyme mechanisms. The development of life itself would have been severely restricted if every biochemical reaction had to follow ideal Michaelis–Menten kinetics and if every carrier protein really bound its ligand according to the simplest mass-action laws. Concepts such as cooperativity and allosteric change have provided convincing explanations of the observed deviations from such ideal behavior, and it is no coincidence that these concepts have been developed over the last 20 years in parallel with structural studies of enzyme–ligand interactions. In fact, it is the ligand that often initiates allosteric or cooperative behavior and thereby controls the overall biochemical process. Drugs sometimes work in the same way, and an understanding of such mechanisms can be of direct help in drug design.

The macromolecules in control systems are frequently multisubunit assemblies in which several proteins aggregate together. These proteins are sometimes arranged symmetrically around an axis that passes close to the ligand molecule itself, and the structure of the ligand or drug may need some symmetry if it is to bind effectively.[44] Such tendencies toward a symmetrical interaction may be seen when the allosteric effector 2,3-diphosphoglycerate (DPG) binds to hemoglobin[45] or when the thyroid hormones bind to prealbumin.[43] Because some interactions occur in pairs on each side of the symmetry axis, there may be an effective halving in the number of protein–ligand relationships that have to be studied, and this sometimes simplifies the structural interpretation and subsequent drug design. Similar considerations can apply when ligands intercalate between the chains of DNA.[46]

Another simplifying factor must also be mentioned. As well as weak hydrophobic relationships, there are more obvious interactions between protein and ligand. These naturally catch the eye where some of the weaker effects may be missed at a first inspection. Examples are the charge–charge interactions between cationic arginine side chains of DHFR[41] and anionic phosphate groups of its ligand NADPH. These make a significant contribution to the overall energy of binding, while ligands such as DPG on hemoglobin seem to rely vary largely on charge–charge interactions. The covalent bond between a protein and a ligand, like pyridoxal, provides a different example in which an obvious interaction makes an appreciable binding contribution. Such contacts are very easily detected and can be helpful when one is trying to understand the geometry of binding, but their energetic importance should not be exaggerated. The many weak interactions may provide more free energy of binding to the system as a

(41) Matthews, D. A.; Alden, R. A.; Freer, S. T.; Xuong, N.; Kraut, J. *J. Biol. Chem.* **1979**, *254*, 4144–4151.
(42) Fermi, G.; Perutz, M. F. "Atlas of Molecular Structures in Biology"; Phillips, D. C.; Richards, F. M., Eds.; Clarenden Press: Oxford; 1981; Vol. 2, pp 104.
(43) Oatley, S. J.; Burridge, J. M.; Blake, C. C. F. "Hormone Antagonists"; Agarwal, M. K., Ed.; Gruyter: Berlin, 1982; pp 705–715.
(44) Goodford, P. J. *Proc. Int. Symp. Med. Chem., 6th*, 255–263.
(45) Arnone, A. *Nature (London)* **1972**, *237*, 769–775.
(46) Wakelin, L. P. G.; Waring, M. J. *Biochem. J.* **1976**, *157*, 721–740.

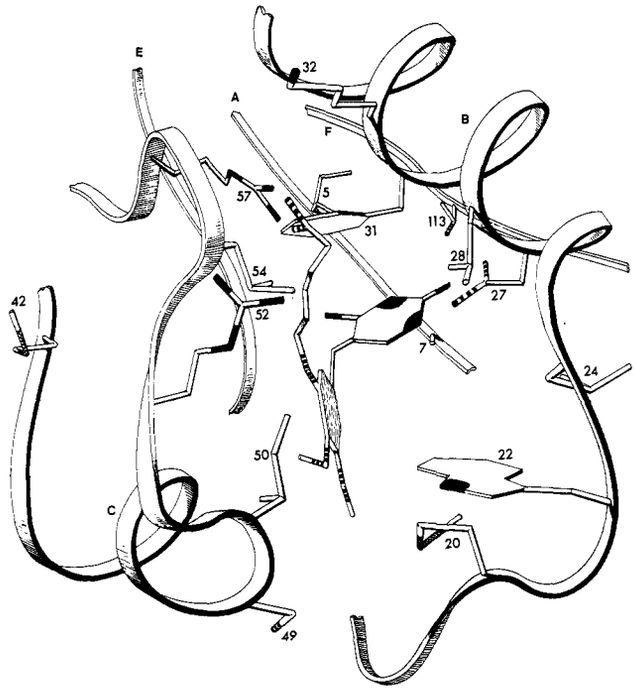


Figure 3. A representation[15] of the substrate binding site in bacterial dihydrofolate reductase. Nitrogen atoms are shown black, oxygen atoms are shown striped, and sulfur atoms are shown crosshatched. The heterocyclic ring of inhibitor II is shown making contact with the aspartic acid 27 residue, and the basic residue arginine-57 lies a few angstroms above the inhibitor. In this view the acidic side chain of compound **2** is shown with its carboxy group and the guanidinium group of the arginine interacting. The compound was a more potent DHFR inhibitor than trimethoprim (1), and X-ray studies showed that it bound as predicted.

whole, and it could be misleading to study the obvious interactions alone.

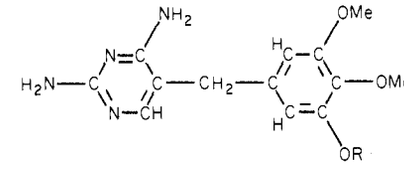
The Design of Ligands

We now progress from observations on the binding of known compounds to more predictive studies. *N,N',N'',N'''*-Tetraacetylchitotetraose is an oligosaccharide that binds to lysozyme, and its δ-lactone (TACL) has been synthesized as an analogue of the predicted transition state of the known hexasaccharide substrate of the enzyme. As expected for a transition-state analogue, TACL bound some 32 times more strongly than *N,N',N'',N'''*-tetraacetylchitotetraose itself, and X-ray studies[47] went a long way toward showing that the geometry of bonding was actually as predicted. In fact this is one of the few cases in which the binding mode of a designed ligand has been directly confirmed, and the findings had a clear influence on the developing concept of transition states in enzymology.

More recently, computer graphics have been used[48] to model a variety of thyroxine analogues into a cleft of the protein prealbumin. It was already known[43] that the natural hormone thyroxine bound at this site, and an unused pocket was identified in the side of the cleft. The relative binding affinities of some previously untested analogues were predicted, and measured affinities were largely compatible with these predictions. The overall approach relied on the geometrical matching of ligand and receptor, and a general method of optimizing this fit has recently been published.[49]

Of course computer graphics are by no means essential for ligand design. A physical model of carbonic anhydrase was used[17] to interpret the binding of acetazolamide, which owes its effectiveness as a diuretic to the inhibition of this enzyme. The presence of hydrophobic residues in the binding cleft was then exploited to design an analogue twice as potent as acetazolamide itself, and more potent inhibitors may be expected from this work in due course. Some difficulties were encountered with the chemical synthesis, and this is not uncommon when novel ligands are being designed to fit a protein. On the other hand, this investigation demonstrates that a macromolecular structure can be of value when choices are being made between alternative synthetically feasible compounds, if it shows which small molecules are most likely to bind at the target site.

These studies were all based on the established structure of a known ligand–protein complex, and the final compounds were more or less close analogues of the starting molecule. Similar methods were used for the DHFR work, in which basic amino acid residues close to the known binding cleft were exploited (Figure 3). The presence of these residues suggested that analogues of the known inhibitor trimethoprim (**1**), containing a carboxylic acid



1, R = CH_3
2, R = $(CH_2)_5COOH$

substituent joined appropriately onto the molecule, might have increased affinity for the enzyme. Experiments confirmed[15] that the designed compound (**2**) was a stronger inhibitor than trimethoprim (**1**) itself and that it bound as predicted (Figure 3).

A rather different approach was adopted for some of the hemoglobin work. That also began from a known compound binding at a known site, but a deliberate attempt was made to design novel ligands that had a *completely different* chemical structure. The starting compound was an acidic molecule (DPG) that is structurally related to the sugars and appears to bind mainly by charge–charge or hydrogen-bond interactions. However, the designed ligands were aromatic dialdehydes that exploited the ability of amino groups on the protein to form covalent bonds.[8] Subsequent work[50] supported the predicted mode of binding for these compounds (Figure 4) and showed that their relative affinities for different hemoglobin variants could be predicted. It also drew attention to entropy as an important factor that may influence the relative binding affinities of different ligands.

From such work one may tentatively conclude that it is feasible to design novel ligands, beginning from the structure of a known ligand–protein complex, and making modest changes to the ligand so that the final compound binds in a similar way. New parts of the receptor cleft can be explored, as shown by the prealbumin studies. Compounds with higher affinity can be designed, as with DHFR. Compounds of different function can be discov-

(47) Ford, L. O.; Johnson, L. N.; Machin, P. A.; Phillips. D. C.; Tjian, R. *J. Mol. Biol.* **1974**, *88*, 349–371.
(48) Blaney, J. M.; Jorgenson, E. C.; Connolly, M. L.; Ferrin, T. E.; Langridge, R.; Oatley, S. J.; Burridge, J. M.; Blake, C. C. F. *J. Med. Chem.* **1982**, *25*, 785–790.
(49) Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. *J. Mol. Biol.* **1982**, *161*, 269–288.
(50) Beddell, C. R.; Goodford, P. J.; Stammers, D. K.; Wootton, R. *Br. J. Pharmacol.* **1979**, *65*, 535–543.

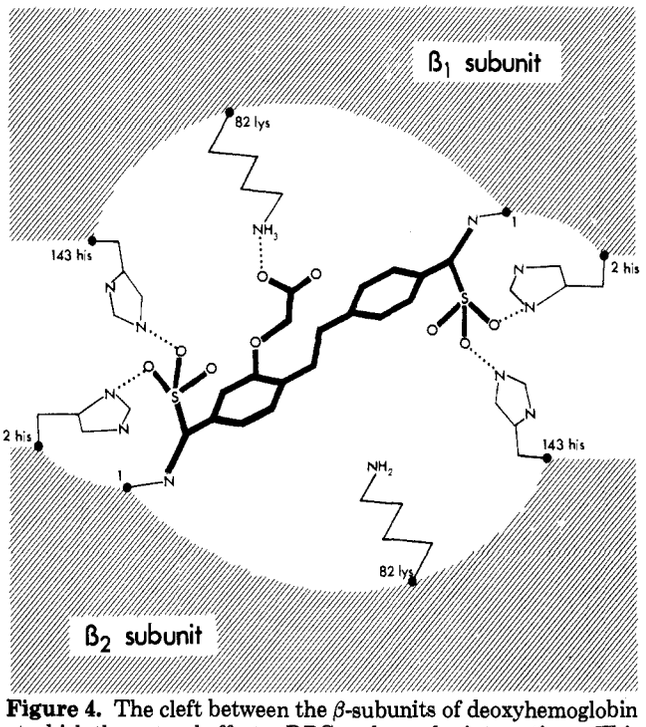


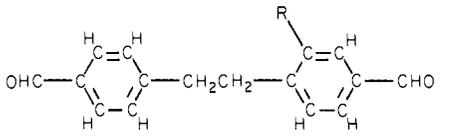
Figure 4. The cleft between the β-subunits of deoxyhemoglobin at which the natural effector DPG makes polar interactions. This site is symmetrical, and a symmetrical ligand (3) was designed[8] to exploit the Schiff base reaction of aldehyde groups with NH_2 groups of the protein. The solubility of the compound was then increased by adding a polar side chain (OCH_2COOH) and by forming a bisulfate addition complex. If the ligands fit as predicted, both these changes could also enhance ligand binding as shown.

ered, such as the inhibitor TACL, starting from previous ideas about substrates binding to lysozyme. On the other hand, the design of compounds with a completely new structure, unrelated to known ligands, may be less straightforward, except in particularly favorable cases like hemoglobin.

The Design of Drugs

Drugs must cure sick people, and this specification calls for compounds with more properties than ligand binding alone. The use of enzyme structures for drug design should be assessed in this wider context, and up to now the successes have been few, if any. The aromatic dialdehydes influenced hemoglobin in solution but were less effective in blood. They have not been tried in man, and neither have the acetazolamide analogues. The DHFR inhibitors were less active than trimethoprim as broad-spectrum antibacterials, although they were better DHFR inhibitors. Perhaps they did not penetrate bacterial cell membranes so well as their parent compound. No therapeutic role has been identified for lysozyme inhibitors, or for thyroid hormone analogues which might bind better to prealbumin than the natural hormone itself.

Such findings suggest two straightforward conclusions. First, the chosen macromolecule must be a therapeutically relevant target. The hemoglobin ligands, for example, were intended to promote the transfer of oxygen from blood to the tissues of the body. Second, the designed compounds must have appropriate physicochemical properties, and it may be surprising that the method of drug design by receptor fit can be of direct relevance here. In the hemoglobin work, for example, straightforward aromatic dialdehydes, such as compound **3**, were too insoluble, and various alterations to the ligand structure were therefore considered in order to increase solubility. A study of the



3, R = H
4, R = OCH_2COOH

receptor cleft (Figure 4) showed that bisulfate addition complexes should not only have the desired solubility characteristics but should also give an improved fit to the receptor as well. This might lead to increased potency so that two birds could be killed with one stone, and the inclusion of an appropriate side chain to give compound **4** might still further increase solubility and also provide an extra carboxylic acid group to interact with basic residue lysine-82 of hemoglobin.[8] Thus, knowledge of the protein cleft was exploited in order to decide how and where a ligand structure should be modified to improve both its binding and its physicochemical properties together. Indeed, this work suggests that the presence of a benzene ring may always be beneficial when designing novel ligands, because the principles of Hammett[51] and Hansch[52] can then be used to predict physicochemical properties.

ACE catalyzes the formation of angiotensin II, which can cause high blood pressure.[53] It is a zinc-containing exopeptidase that cleaves dipeptides from the carboxyl terminal of peptides. The structure of the substrate and the biochemistry of the reaction suggested[10] that the active site of ACE might be similar to the corresponding site of carboxypeptidase, whose structure and function were already understood in detail.[54,55] It was even possible to deduce plausible differences between the two enzymes, and on this basis an outline plan of the site in ACE could be constructed as a template for drug design. Compounds were prepared,[10] and a potent inhibitor (Captopril) was discovered. It became a valued therapeutic agent, and the work constitutes a supreme example of rational drug design. Particular attention must be focused on the elegant biochemical logic.[53] Each step of the angiotensin system was considered in order to choose a therapeutically relevant target, and this analysis led to the inexorable conclusion that certain types of high blood pressure should be controlled by inhibiting ACE. However, one cannot easily claim that the resulting discovery of Captopril was an unambiguous demonstration of the method of receptor fit, because the structure of the binding site in ACE is still unknown. Furthermore, many biochemists would argue that the compounds synthesized during the search for Captopril could have been designed on the basis of general peptide biochemistry, once the substrates for ACE and for carboxypeptidase had been extablished.

The method of receptor fit is not restricted to sites like that in ACE, which are used by natural ligands. Any suitable cleft in the chosen macromolecule can be tried, and the identification of an appropriate "unnatural" site, followed by the successful design of an effective novel drug to fit it, might be a particularly clear demonstration of the overall approach. In fact, an attempt has been made to design such a compound for the treatment of sickle cell

(51) Hammett, L. P. "Physical Organic Chemistry, Reaction Rates, Equilibria and Mechanisms"; 2nd ed.; McGraw-Hill: New York, 1940.
(52) Hansch, C.; Fujita, T. *J. Am. Chem. Soc.* **1964**, *86*, 1616–1626.
(53) Ondetti, M. A.; Cushman, D. W. *J. Med. Chem.* **1981**, *24*, 355–361.
(54) Quiocho, F. A.; Lipscomb, W. N. *Adv. Protein Chem.* **1971**, *25*, 1–78.
(55) Lipscomb, W. N. *Chem. Soc. Rev.* **1972**, *1*, 319–336.

anemia. Patients with this disease have inherited a hemoglobin that polymerizes if their blood loses too much oxygen to the tissues, and the natural effector DPG actually encourages this pathological process when it binds between the β-subunits of deoxyhemoglobin. A biochemical and physiological analysis of the system suggested that a ligand reacting between the α-subunits of oxyhemoglobin might oppose DPG and prevent sickling, but no natural effector binding at this site has been detected in red blood cells. The oxyhemoglobin structure was therefore used in order to design a suitable ligand, and this apparently does bind to oxyhemoglobin and does produce the desired effect[56] on blood in vitro. It also seems to have appropriate physicochemical properties and is absorbed after oral administration to animals, but one cannot yet say if it will become a therapeutic agent in man or, indeed, whether it works by the intended macromolecular mechanism at all.

The Future

At this time it appears that no novel drugs have been unambiguously designed by the method of receptor fit, although the design of ligands has been successful. There are several cases where the structure of a site has been observed with one ligand bound, and related ligands have been designed by making conservative changes to the structure of the first.[15,17,47,48] The compound with highest affinity in a homologous series can be predicted[15] in this situation, and it is also possible to predict the selectivity of related ligands at subtly different binding sites in series of similar proteins.[50] Ligands can be designed to initiate a chosen functional effect, if the function of the macromolecular target is adequately understood, and novel effects can be produced that have no physiological counterpart.[56] One can even produce different effects on the same protein by aiming ligands at different target sites.[8,56]

The approach is inherently less reliable if the structure of the target macromolecule has not been observed experimentally, but inferred by analogy with other proteins or computed indirectly. However, inferred structures have often been of more relevance to therapy than observed receptor sites, and this partly compensates for any diminution in the reliability of the method. The discovery of Captopril demonstrates what a great help it can be if one just thinks about drug receptors as being macromolecules with a definite atomic structure and function. Although the structure of ACE may still be unknown, the great heuristic value of a hypothetical active-site model in the design of inhibitors has been well emphasized in that work.[10]

It may not be easy to decide whether any particular macromolecule should be regarded as a possible receptor. Twenty years ago, for example, nobody would have thought that ligands might be designed to react with hemoglobin and thereby influence the way in which it transports oxygen. However, the picture changed when the functional role of DPG was discovered, and it then became necessary to revise the concept that hemoglobin was a simple transport protein. In the opinion of this author, one needs to have an understanding of function at the molecular level, before trying to use the structure of a macromolecule for effective drug design.

Most attempts to predict ligand affinities up to now have been aimed at predicting the enthalpy of binding, but entropy effects can also be of major importance. Not only the rotation and translation of the ligand, but alternations in the flexibility of both ligand and site, as well as changes in the solvent and solute molecules, all make contributions to the entropy and to binding affinities. Indeed, it may be surprising that any predictions have worked at all, unless the entropy component remains fairly constant when one ligand is compared with another similar molecule. This could account for some of the successful predictions that have been made and might also explain some of the failures. When ligands of widely different size, shape, flexibility, and hydration are compared, it may still be feasible to predict relative enthalpies of binding, but realistic attempts to predict the role of entropy are only just beginning.[57,58]

Real therapeutic agents must be properly absorbed; they must reach the desired site of action in the body; their metabolic breakdown and excretion must be appropriate; they and their metabolites must not be toxic and must not have unacceptable side effects; and they must be sufficiently stable, must satisfy analytical standards, and must indeed comply with many other criteria that are more or less well established. If therapeutic agents are to be designed by the method of receptor fit, they must not only be aimed successfully at an appropriate receptor but must also satisfy these other requirements. Up to now it seems that nobody has explicitly practiced an integrated approach to drug design on this scale, although it is implicit in the accumulated skills of every medicinal chemist. He can use knowledge of the receptor to decide which parts of the trial compound are needed for binding and which can be modified in order to turn a potent ligand into a potential drug.

What, then, does the method of receptor fit offer for a future in which the structure and function of macromolecules will be understood and where doctors may have direct access to the full genetic code of every patient? First, one should be able to design novel drugs of very high affinity for known target sites. This may increase ligand specificity, and if the drug produces undesirable side effects by binding to another macromolecule, it may be possible to diminish affinity for that competing site and thus achieve a better therapeutic index. Systems analysis of a chosen biochemical pathway will enable the most appropriate target site to be identified, so that drugs can be designed to react there and produce more subtle effects than existing therapeutic agents. Furthermore, new sites of action will be identified at which completely new types of therapy can be initiated, as with the compounds designed to react between the α-subunits of hemoglobin and thereby influence sickling. More consistent biological responses may also be obtainable by using conserved protein residues for ligand binding, such as the 18 conserved contacts between DHFR and NADPH, while residues like the remaining 12 contacts will be avoided, since they are liable to vary. On the other hand, sequence variation may be exploited to improve specificity, because systematic differences of protein sequence can often be detected near ligand binding sites. In the past the working assumption was that all receptors in a class were the same, until pharmacological evidence showed that they were different. The progressive subdivision of catacholamine receptors into subclasses provides an example of this philosophy, but in future the a priori assumption should perhaps be that all receptors are different until proved identical. Two receptors may seem to initiate the same response in different tissues, different cell types, different species, or at different times, but it is not reasonable to

(56) Beddell, C. R.; Goodford, P. J.; Kneen, G.; White, R. D.; Wilkinson, S.; Wootton, R. *Br. J. Pharmacol.*, in press.

(57) Dean, P. M. *Br. J. Pharmacol.* **1983**, *79*, 344P.

(58) Jencks, W. P. *Adv. Enzymol.* **1975**, *43*, 219–410

conclude that they are the same on this evidence alone. Structural evidence based on physical and chemical methods is also needed.

These tentative forecasts have two common factors. On the one hand they point toward a new generation of more potent, specific, effective therapeutic agents with less toxicity, reduced side effects, and fewer aberrant responses, which is what people and society at large are seeking. On the other hand, they also point toward more costly research, which is the price that must be paid. Apart from that, one last conclusion seems very probable. Mountaineers climb because the mountains are there and offer them a worthwhile challenge, and scientists will try to design drugs to fit receptors for similar reasons.

Acknowledgment. I thank the Wellcome Foundation at Beckenham, the Laboratory of Molecular Biophysics at Oxford, and all my colleagues for their advice, encouragement, and support. My debt to them is very large, and I am also grateful to the authors who have allowed me to reproduce their figures, and the copyright holders for their permission to do so.

Articles

Biomimetic Approach to Potential Benzodiazepine Receptor Agonists and Antagonists

Filadelfo Guzman,† Michael Cain,† Paul Larscheid,† Tim Hagen,† James M. Cook,*,† Margaret Schweri,‡ Phil Skolnick,‡ and Steven M. Paul§

Department of Chemistry, University of Wisconsin-Milwaukee, Milwaukee, Wisconsin 53201, Laboratory of Bioorganic Chemistry, National Institute of Arthritis, Diabetes, and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20205, and Clinical Neuroscience Branch, National Institute of Mental Health, Bethesda, Maryland 20205. *Received March 4, 1983*

Several β-carbolines, isoquinolines, imidazopyridines, and canthin-6-ones prepared in biomimetic fashion were tested for their ability to bind to the benzodiazepine receptor. Methyl isoquinoline-3-carboxylate (**3a**), methyl 6,7-dimethoxyisoquinoline-3-carboxylate (**3b**), 1-phenyl-3-carbomethoxyimidazopyridine (**6b**), and canthin-6-one (**13a**) bound with moderate affinities, while 2-carbomethoxycanthin-6-one (**13b**) bound to benzodiazepine receptors with an affinity comparable to several pharmacologically active benzodiazepines. The potency of **13b** suggests that the benzodiazepine receptor(s) can tolerate substitution at positions 1 and 9 of a β-carboline without loss of activity if the substituents are trigonal and maintain a planar topography. Moreover, displacement of the carbonyl group by two atoms (**17**) from the aromatic ring (C) of the β-carboline skeleton caused a marked decrease in binding to the benzodiazepine receptor. This observation supports the hypothesis that maximum binding affinity of β-carbolines is achieved when the carbonyl group at position 3 is attached directly to the aromatic pyridine ring.

The demonstration of high-affinity, saturable and stereospecific binding sites for benzodiazepines (viz., receptors) in the CNS[1,2] has radically altered concepts of the molecular mechanisms of benzodiazepine action. Benzodiazepines, the most widely prescribed psychoactive drugs in current therapeutic use,[3] exhibit four principle pharmacological actions: anticonvulsant, anxiolytic, muscle relaxant, and sedative effects.[4,5] Evidence for the multiplicity of benzodiazepine receptors has been reported,[6] and by analogy to the progression of events in the opiate–receptor–enkephalin area,[7,8] the first reports that benzodiazepines have selective and specific high-affinity binding sites in the CNS stimulated a search for an "endogenous ligand" that physiologically acts at these receptors.[9,12] Ethyl β-carboline-3-carboxylate (β-CCE) was first identified in human urine[12a] and found to inhibit potently ($K_i \approx 1$ nM) the in vitro binding of [^{3}H]diazepam to brain benzodiazepine receptors. The potent biological activity of this β-carboline led to the hypothesis that β-CCE or a related compound could be an endogenous ligand of the benzodiazepine receptor. However, subsequent reports suggested that β-CCE was formed during the extraction and isolation procedure employed, since identical treatment of many proteins resulted in the formation of

† University of Wisconsin-Milwaukee.
‡ National Institutes of Health.
§ National Institute of Mental Health.

(1) Squires, R. F.; Braestrup, C. *Nature (London)* **1977**, *266*, 732.
(2) Mohler, H.; Okada, T. *Science* **1977**, *198*, 849.
(3) Greenblatt, D. J.; Shader, R. I. "Benzodiazepines in Clinical Practice"; Raven Press: New York, 1974. Williams, M. *J. Med. Chem.* **1983**, *26*, 619.
(4) Randall, L. O.; Kappel, B. In "The Benzodiazepines"; Garattini, S.; Mussini, E.; Randall, L. O. Eds.; Raven Press, New York, 1973; pp 27–51.
(5) Mohler, H.; Okada, T.; Heitz, H. P.; Ulrich, J. *Life Sci.* **1978**, *22*, 985.
(6) Nielsen, M.; Schou, H.; Braestrup, C. *J. Neurochem.* **1981**, *36*, 276.
(7) Pert, C.; Snyder, S. *Science* **1973**, *179*, 1011.
(8) Hughes, J.; Smith, T. W.; Kosterlitz, H. W.; Fothergill, L. A.; Morgan, B. A.; Morris, H. R. *Nature (London)* **1975**, 258, 577.
(9) Mohler, H.; Polc, P.; Cumin, R.; Pieri, L.; Kettler, R. *Nature (London)* **1979**, *278*, 563. Collelo, C. D.; Hockenberry, D. M.; Bosmann, D. M.; Fuchs, H. B.; Folkers, K. *Proc. Natl. Acad. Sci. U.S.A.* **1978**, *75*, 6319. Skolnick, P.; Marangos, P. J.; Goodwin, F.; Edwards, M.; Paul, S. *Life Sci.* **1978**, *23*, 1473. Ally, A.; Manku, M. S.; Horrobin, D. F.; Karmali, R. A.; Karmazyn, M.; Morgan, R. O. *Neurosci. Lett.* **1978**, *7*, 31. Asano, T.; Spector, S. *Proc. Natl. Acad. Sci. U.S.A.* **1979**, *76*, 977. Davis, L. G.; Cohen, R. K. *Biochem. Biophys. Res. Commun.* **1980**, *92*, 141. Davis, L. G.; McIntosh, H.; Reker, D. *Pharmacol. Biochem. Behav.* **1981**, *14*, 839.

0022-2623/84/1827-0564$01.50/0