Designing Drugs to Fit a Macromolecular Receptor *

By C. R. Beddell
WELLCOME RESEARCH LABORATORIES, LANGLEY COURT, BECKENHAM, KENT BR3 3BS

1 Introduction

In the search for new and better drugs, many different approaches have been used. The present review is concerned with just one of these; the selection of chemical structures by a relatively new process involving matching of chemical structure to a macromolecular target of known architecture. The relationship of this approach to other approaches in medicinal chemistry is also considered.

The development of a series of drugs may start with a compound, available from plant, animal, or synthetic source, which possesses some element of the desired characteristics of a drug. When there does exist a very well defined starting point for series development (*e.g.* a natural regulator molecule) the options open to a medicinal chemist are usually huge and different approaches towards identifying a suitable drug will meet with different levels of success, given the restrictions on time and effort which can be expended. Sometimes there may be no molecule regarded as intrinsically suitable as a starting point, perhaps because no known compound has the appropriate potency, or specificity, or because potential compounds pose problems of synthesis.

The need to reduce the size of the problem is thus obvious and one of the methods employed to achieve this reduction is that of establishing principles specific to the problem in hand which help rationally to direct attention to certain new chemical entities. For example, there may be established relationships between the biological activity of the compounds and certain physicochemical properties, which in turn may indicate compounds with more appropriate physicochemical (and thus biological) properties. The medicinal chemist's problem is to optimize within a current series of compounds and to generate new series. To do the former, trends are sought relating the physicochemical properties of the molecules to the required effects, for use as constraints in the choice of further compounds for synthesis. To do the latter, the chemist has often had to rely upon chance observations of chemicals possessing interesting biological activities.

The search for trends is a problem in itself, as succinctly described by Ganellin.[1]

> 'This is the dilemma for medicinal chemists; a simple structural change in a drug molecule has multiple consequences. There is usually no clear indication

* This review is based on a paper presented on 20th January 1983 to the Perkin Division and Fine Chemicals and Medicinal Substances Group of the Royal Society of Chemistry, the Fine Chemicals Group of the Society of Chemical Industry, and the Pharmaceutical Society.

[1] C. R. Ganellin, in 'Quantitative Approaches to Drug Design', ed. J. C. Dearden, Elsevier Science Publishers B.V., Amsterdam, 1983, p. 239.

of which chemical properties need to be measured or whether they will have importance in determining the biological activity of a compound. Much effort may be expended in seeking for possible property–activity relationships with no apparent successful outcome. This complexity, which is seemingly unanalysable, drives chemists to empiricism'.*

Suppose a small chemical change is made in a drug. This one change may alter certain attributes of the drug, such as its intrinsic stability, its susceptibility to modification by an enzyme, its ability to partition between different body phases, its tendency to bind to one or more receptors, and its ability to produce a response from each receptor. In each process in which the modified drug behaves differently, it may be changes in some specific physicochemical property or combination of properties of the drug which are most influential (for example, molecular orbital energies for stability, steric properties for binding, lipophilicity for partition). Now suppose the overall measured effect (E) from all these processes were a linear function of each property (P_i) in the drug.

$$E = c_1P_1 + c_2P_2 \cdots + c_nP_n \tag{1}$$

In principle it would be possible to solve for the n coefficients c, given a knowledge of the n P values and the value of E, provided that there were enough (at least n) independent equations and hence E values. Since many properties vary from position to position in the drug, and some are intrinsically vectorial rather than scalar, a great many parameters and hence equations would be needed. In practice, given also the error usually associated with biological and physicochemical property estimates, a relationship only becomes apparent when some property comes to dominate in producing the effect in some region of this multi-dimensional property space. This itself implies a non-linear relationship, since the partial derivatives (equation 2) are unchanging, and thus the relative importance of property alterations would be invariant, in the linear case.

$$\frac{\partial E}{\partial P_i} = c_i \tag{2}$$

A further problem which arises is that of congruence ambiguity. In the use of a set of equations, one equation per molecule, there is an assumption of comparability between the property value P_{1i} for one molecule and a corresponding value P_{2i} for the next, and so on, but the matching of values of position-dependent properties between molecules presumes a spatial superposition of molecules, achieving a congruence between property values, which is relevant to the functionally important processes. Clearly, congruence in one process is established by additional positioning parameters for each drug, and congruence may be different in different processes.

Many approaches have been used to search for physicochemical property–activity relationships (P.A.R.). It is not within the scope of this review to detail

* Reproduced by permission of the author and Elsevier Science Publishers B.V. from 'Quantitative Approaches to Drug Design', 1983, p. 239.

[2] (*a*) 'Strategy in Drug Research', ed. J. A. K. Buisman, Elsevier Scientific Publishing Co., Amsterdam, 1982. (*b*) 'Quantitative Approaches to Drug Design', ed. J. C. Dearden, Elsevier Science Publishers B.V., Amsterdam, 1983.

them, but examples of many are provided in recent conference proceedings.[2] The present review is concerned with one such approach, dealing with relationships between the physicochemical properties of the drug and the complementary properties of the receptor. The receptor in this context is any macromolecule with which a series of compounds could potentially interact and, by virtue of this interaction, influence pharmacokinetic or pharmacodynamic aspects of the biological actions of the compounds. Consideration is given to the extent to which, in such an approach, the basic problems in P.A.R. of finding local and global optima are overcome, and to what extent the approach is sensitive to some of the factors which can confound classical P.A.R. approaches, namely the multiplicity and interdependence of molecular properties, congruence ambiguity, and the multiplicity of potential functional forms for the relationship. No review of this length can cover the already substantial literature in this area. Examples of work are chosen with the object of providing a balanced rather than exhaustive view of the field; references are likewise limited, but should provide the reader with entry points to the literature.

2 Fitting to the Macromolecule

Beginning with the establishment of the three-dimensional structure of DNA [3] and myoglobin,[4] the *X*-ray crystallographic technique has provided detailed information on the architecture of many macromolecules, including structural and binding proteins, enzymes, and t-RNAs. There is much known at a qualitative level of the forces which maintain the active conformations of these molecules (van der Waals repulsion and attraction, Coulombic interaction between charged centres, hydrophobic forces in an aqueous environment, hydrogen bonding, for example) and these same types of force, in varying degree, are involved in the interaction between macromolecule and ligand. Indeed, Phillips and his colleagues [5] showed that in the interaction of a chain of sugar residues constituting an oligosaccharide with a chain of binding sites on hen lysozyme, there was a rough correspondence, in the absence of strain, between the unitary free energy for binding of each sugar residue and the total number of contacts less than 4Å between the residue and the protein. With the possibility that so simple a concept might be more widely applicable, several groups have, during the last ten years, pitted their wits against macromolecules by attempting to design small molecules to fit to them.

A. Non-enzymic Proteins.—The following three examples of macromolecule fitting involve non-enzymic proteins, functional through binding with either a small ligand or with a receptor. In each case, the design of chemicals to bind to the macromolecule required fitting to sites in the protein composed of parts of more than one subunit. Compounds were thus designed to bind at intersubunit sites near interfaces between subunits.

[3] J. D. Watson and F. H. C. Crick, *Nature* (*London*), 1953, **171**, 737.
[4] J. C. Kendrew, G. Bodo, H. M. Dintzis, R. G. Parrish, H. Wyckoff, and D. C. Phillips, *Nature* (*London*), *1958*, **181**, 662.
[5] L. N. Johnson, D. C. Phillips, and J. A. Rupley, *Brookhaven Symp. Biol.*, 1968, **21**, 120.

(i) *Haemoglobin.* Haemoglobin contains two types of peptide chain (α and β) each enclosing haem-bound iron and mutually interacting non-covalently in an $\alpha_2\beta_2$ tetramer. The reversible oxygen binding to the iron can be described by a graph showing saturation with oxygen against oxygen partial pressure. The sigmoid relationship observed (Figure 1) is the result of a transition between two conformational states of the tetramer, one with low oxygen affinity (the 'tense' T-state) and the other with high oxygen affinity (the 'relaxed' R-state)[6] (Figure 2). Either state alone would generate a rectangular hyperbola such as shown in Figure 1, the curve to the right relating to the state with lower oxygen affinity. In general a right-shift of the sigmoid curve denotes an overall lowering of affinity for oxygen and this is achieved by a natural regulator molecule, 2,3-diphospho-D-glycerate (DPG) (1) (this and several subsequent compounds are illustrated as the neutral form but exist predominantly as charged species at pH values near to those occurring physiologically) which binds between the β-chains with selectivity for the T-state compared with the R-state. Figure 3 shows how DPG with its

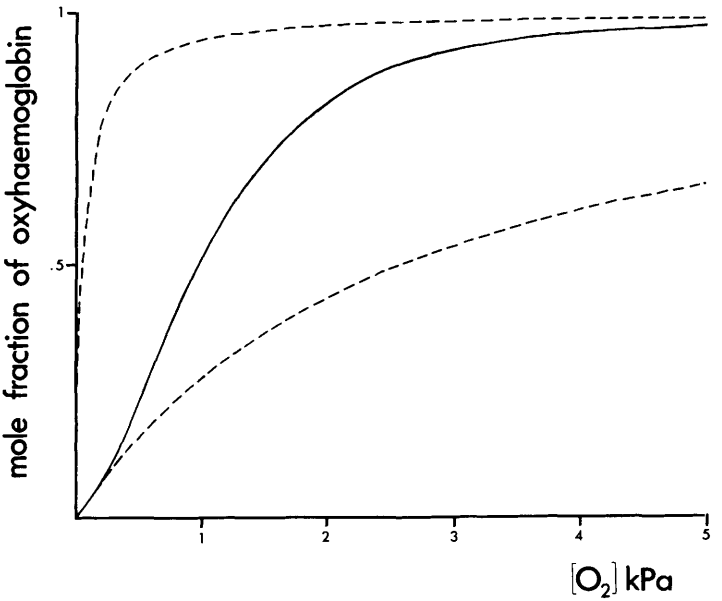


Figure 1 *Relationship between fractional saturation of haemoglobin with oxygen and oxygen partial pressure. The typical sigmoid binding curve (solid) arises from the transition between two conformational states of the haemoglobin, which individually, in the absence of transition are calculated to bind oxygen according to a rectangular hyperbola such as shown to the left (R-state) and right (T-state) of the sigmoid curve*

[6] J. Monod, J. Wyman, and J.-P. Changeux, *J. Mol. Biol.*, 1965, **12**, 88.

anionic phosphate and carboxylate groups is complementary to cationic imidazolyl and α-amino-groups in the protein and binds to the T-state by charge-assisted hydrogen bonding.[7] DPG interacts less favourably with the R-state, owing to substantial geometric differences between states.

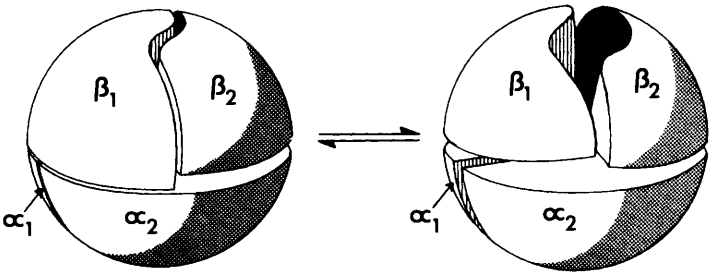
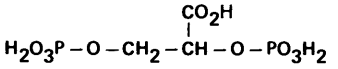
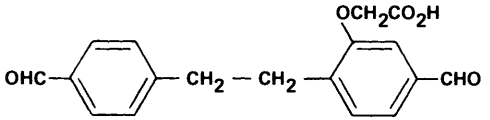


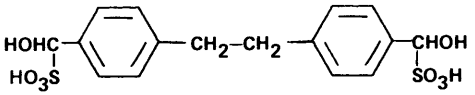
Figure 2 *Diagrammatic representation of the* $\alpha_2\beta_2$ *haemoglobin tetramer, showing interconversion between the R-state (left) and T-state. The cleft between the* β*-chains in the T-state in which* (1) *binds is also illustrated*



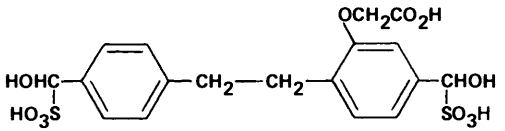
(1)



(2)



(3)



(4)

Compounds Designed to Lower Oxygen Affinity. With the help of physical models constructed from co-ordinates provided by Dr. M. F. Perutz[8,9] and representing

[7] A. Arnone, *Nature (London)*, 1972, **237**, 146.
[8] M. F. Perutz, H. Muirhead, J. M. Cox, and L. C. G. Goaman, *Nature (London)*, 1968, **219**, 131.
[9] W. Bolton and M. F. Perutz, *Nature (London)*, 1970, **228**, 551.

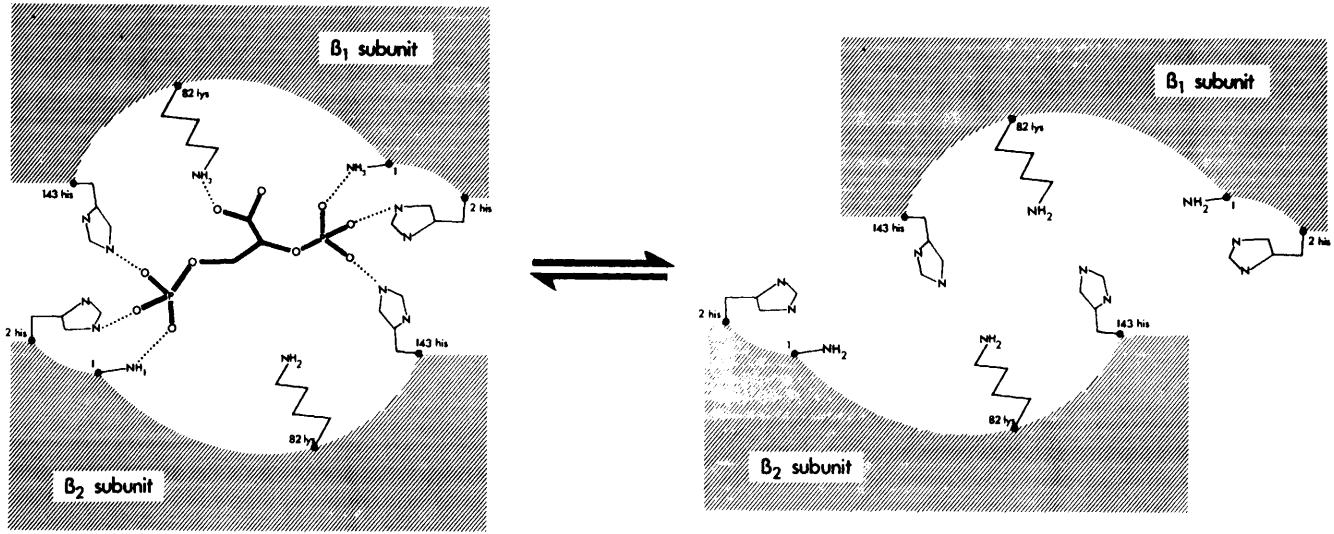


Figure 3 *Diagrammatic representation of the region between the haemoglobin β-chains in the T-state (left) and R-state. The complementary nature of the cationic-group-containing cavity in the T-state and the anionic* (1) *can be seen*

regions between the β-chains, compounds (2)—(4) were designed to interact selectively with some of the T-state DPG binding-site residues.[10] In contrast to DPG, which interacts only non-covalently, compounds (2)—(4), having aldehyde groups (or the bisulphite addition complex of aldehyde), can in principle form reversible covalent bonds with the *N*-terminal amino-groups. In addition, since these compounds could position sulphonate and/or carboxylate groups near to cationic groups in the protein they may also interact with the protein by charge-assisted hydrogen bonding, *i.e.* ionically. Two particular types of interaction between compound and protein were thus considered, namely ionic and covalent, and it was postulated by molecular modelling that (4) would make most (two covalent, five ionic) and (2) fewest (two covalent, one ionic) interactions (see Figure 4). As shown in Figure 5 each compound produced a right-shift of the oxygen saturation curve and a simplistic count of interactions between compound and protein correlated with the size of the right-shift.

N.m.r. studies[11] and crystallographic studies[12] have supported the postulate that such compounds interact with the DPG binding-site, and further evidence has been obtained from the following biochemical experiment. The postulated relationship between compound–protein interactions and binding can in principle be explored by varying the structure of the compound and/or the binding site so that more or fewer interactions are made. The oxygen saturation curves for compounds (1), (2), and (4) each binding to one of six different haemoglobins (Table 1) were used to establish the free energy of binding for each compound to

Table 1 *Residues in the DPG binding-site of selected haemoglobins*

	β-*chain residues*		
Haemoglobin	**1**	**2**	**143**
A_I (human)	valine	histidine	histidine
A_{Ic} (human)	valine-glucose adduct	histidine	histidine
Raleigh (human)	*N*-acetylalanine	histidine	histidine
F_I (human)	*N*-acetylvaline	histidine	serine
F_{II} (human)	valine	histidine	serine
Horse	valine	glutamine	histidine

each haemoglobin (3 × 6 = 18 combinations) and separately, the number of covalent and ionic interactions in each pairing were assessed (Figure 6) by the same simple modelling considerations[13] as above. For example, human haemoglobin

[10] C. R. Beddell, P. J. Goodford, F. E. Norrington, S. Wilkinson, and R. Wootton, *Br. J. Pharmacol.*, 1976, **57**, 201.
[11] F. F. Brown and P. J. Goodford, *Br. J. Pharmacol.*, 1977, **60**, 337.
[12] A. J. Geddes, personal communication.
[13] C. R. Beddell, P. J. Goodford, D. K. Stammers, and R. Wootton, *Br. J. Pharmacol.*, 1979, **65**, 535.

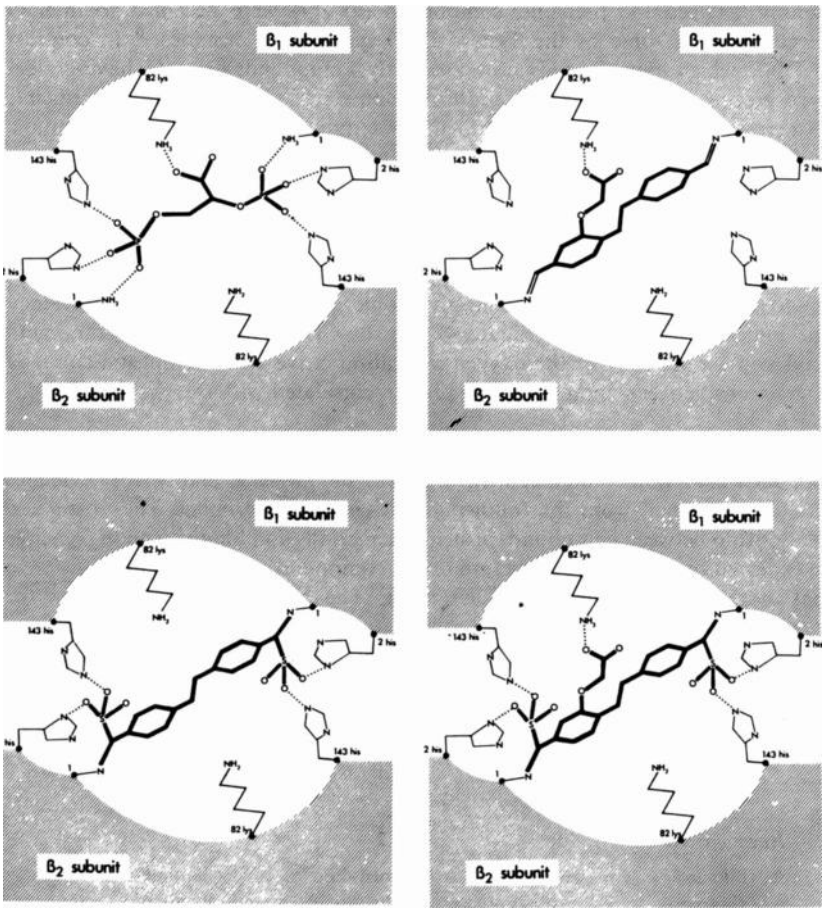


Figure 4 *Diagrammatic representation of the region between T-state haemoglobin β-chains, showing the primary interactions between protein and respectively* (1) (*top left*), (2) (*top right*), (3) (*bottom left*), *and* (4). *The manner of binding for* (1) *has been observed crystallographically, that for the other compounds is postulated. Covalent interactions with residues in each β-chain are shown by solid bonds, charge-assisted hydrogen bonds* (*ionic*) *are represented by dotted lines*
(Reproduced by permission of Macmillan Press Ltd. from *Br. J. Pharmacol.*, 1976, **57**, 201)

has histidine at position β-2 whereas horse haemoglobin has glutamine. DPG paired with human haemoglobin interacts ionically at this position, but with horse, this interaction is absent as glutamine is uncharged. With the inclusion of replicate determinations there were 29 estimates of binding energy and a least-squares analysis revealed that the relationship between binding energy (ΔG) and the respective number of covalent (n_C) and ionic (n_I) interactions could be expressed linearly.

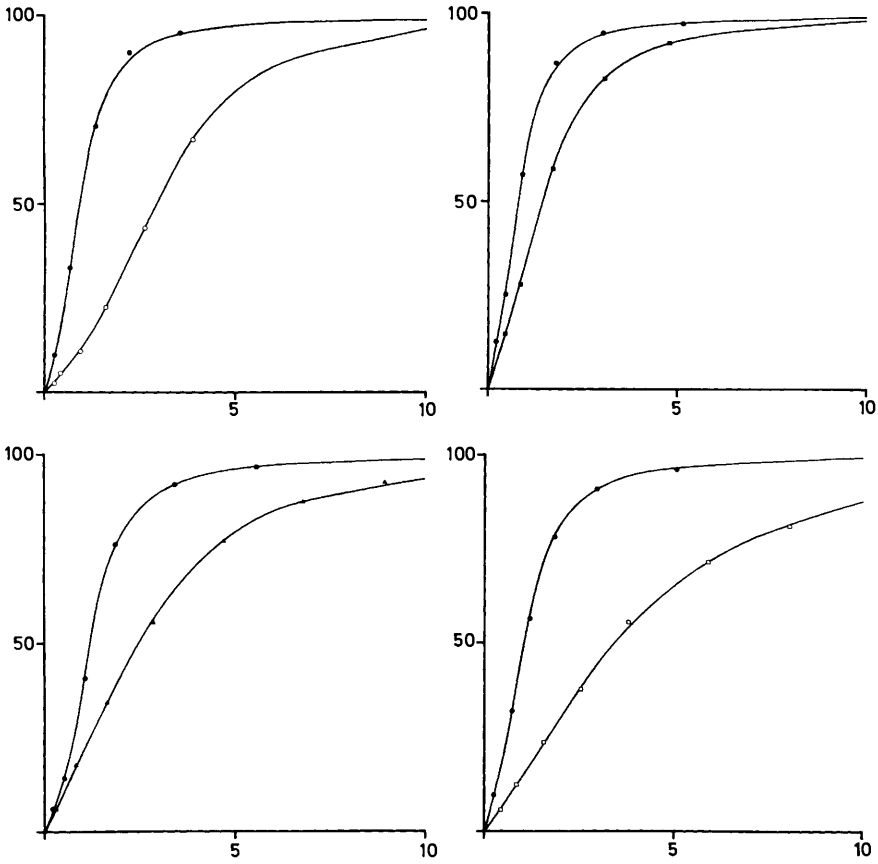


Figure 5 *Each diagram shows two oxygen saturation curves: The one to the left is for haemoglobin whilst the one to the right is for haemoglobin in the presence of one of the compounds* (1)—(4) *respectively at* 2.5mM *concentration, compound assignments being as laid out in Figure* 4. *Note that* (2) *produced a right-shift smaller than that of* (1) *whilst the largest right-shift is produced by* (4), *correlating with the number of interactions counted in Figure* 4. *Abscissae, oxygen pressure* (kPa). *Ordinates, per cent oxyhaemoglobin*
(Reproduced by permission of Macmillan Press Ltd., from *Br. J. Pharmacol.*, 1976, 57, 201)

$$\Delta G = -3.14n_{\mathrm{I}} - 6.78n_{\mathrm{C}} - 8.29 \qquad (3)$$
$$(+0.30)\ (\pm 0.68)\ \ (\pm 1.37)$$
$$n = 29;\ r = 0.928;\ F_{2,26} = 81.15$$

The relationship developed shows that the strength of the covalent bond ($-$ 6.78 kJ mol^{-1}) is about twice that (-3.14 kJ mol^{-1}) of the ionic bond. The equation is significant at the 0.1% level and 86% of the variation in the data is explained. Figure 7 shows the approximately linear relationship between the values of binding free-energy estimated from oxygen saturation curves and values calculated by equation (3) for the values of n_{I} and n_{C} shown in Figure 6.

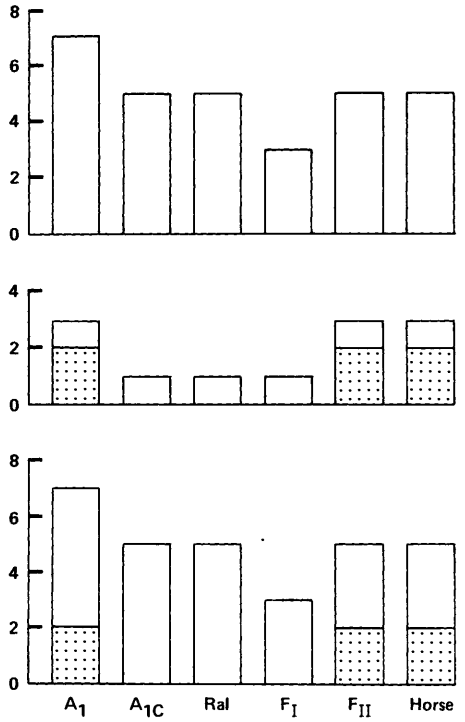


Figure 6 *Histogram of the postulated number (n) of interactions between various haemoglobins and compounds* (1) *(top),* (2) *(middle), and* (4) *respectively as postulated from study of molecular models. Bar height represents the total number of interactions (covalent and ionic), and shading denotes covalent interactions. The eighteen combinations of haemoglobin and compound provide a wide variation in total contacts (between* 1 *and* 7) *and also in the ratio of ionic to covalent interactions (between* 7:0 *and* 1:2)

The various studies with these right-shifting compounds indicate the following points:

(1) It is feasible by modelling studies which highlight qualitatively specific types of intermolecular interaction, to design novel ligands (*a*) which interact with a protein of defined amino-acid sequence and architecture, and (*b*) which selectively differentiate between different conformations of the same protein to produce thereby the expected allosteric effect.

(2) Binding free-energy can be regarded as a linear function of the numbers of each type of postulated interaction to a first approximation.

Compounds Designed to Raise Oxygen Affinity. There exists between the α-chains of haemoglobin a site comparable to that between the β-chains (since α and β are homologous) but different in detail owing to differences between the amino-acid sequences of the two types of peptide chain. Molecular models of this

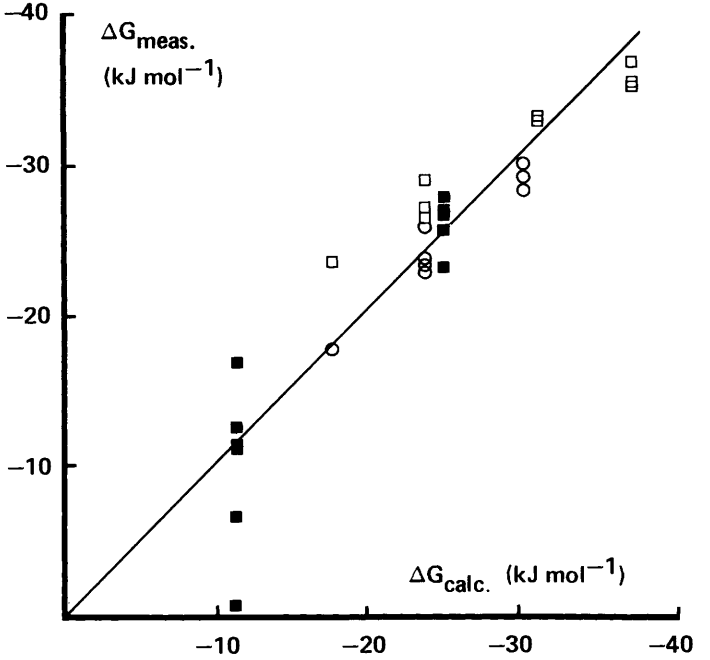
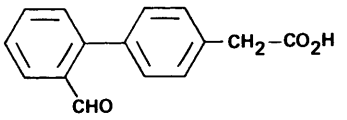
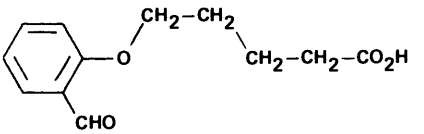


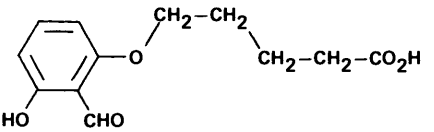
Figure 7 *Relationship between free energy of binding for compound to T-state haemoglobin as estimated from oxygen saturation curves (ΔG_{meas}), and as calculated from equation* (3) *(ΔG_{calc}) with the values of n_C and n_l shown in Figure* 6. *Compounds are denoted, respectively,* (1) *by open circles,* (2) *by solid squares, and* (4) *by open squares*



(5)



(6)



(7)

region in the R and T states were again constructed from co-ordinates provided by Dr. M. F. Perutz[8,9] and molecules were tailored to interact at this site with a selectivity for the R state.[14] Such selectivity should shift the oxygen saturation

[14] (*a*) C. R. Beddell, P. J. Goodford, G. Kneen, R. D. White, S. Wilkinson, and R. Wootton, *Br. J. Pharmacol.*, 1984, **82**, 397; (*b*) G. Kneen and R. D. White, *Br. J. Pharmacol.*, 1981, **74**, 965P.

curve to the left. Molecules (5)—(7) were designed to interact covalently (by virtue of the aldehyde function) with one *N*-terminal group in the protein and to interact ionically *via* the carboxylate group with the other *N*-terminal group (Figure 8). Such binding would also position the molecules between the hydrophobic side-chains of two proline residues. Thus the central part of each molecule was designed to be non-polar in character to promote hydrophobic interaction.

It was found, as predicted, that molecules (5)—(7) could shift the oxygen saturation curve to the left. Furthermore a greater left-shift was obtainable in the presence of the right-shifting compound L-*myo*-inositol hexaphosphate (8) which binds at the DPG binding-site (saturation curves are shown in Figure 9). Compound (5) showed a small left-shifting effect. It was anticipated that a hydroxy-

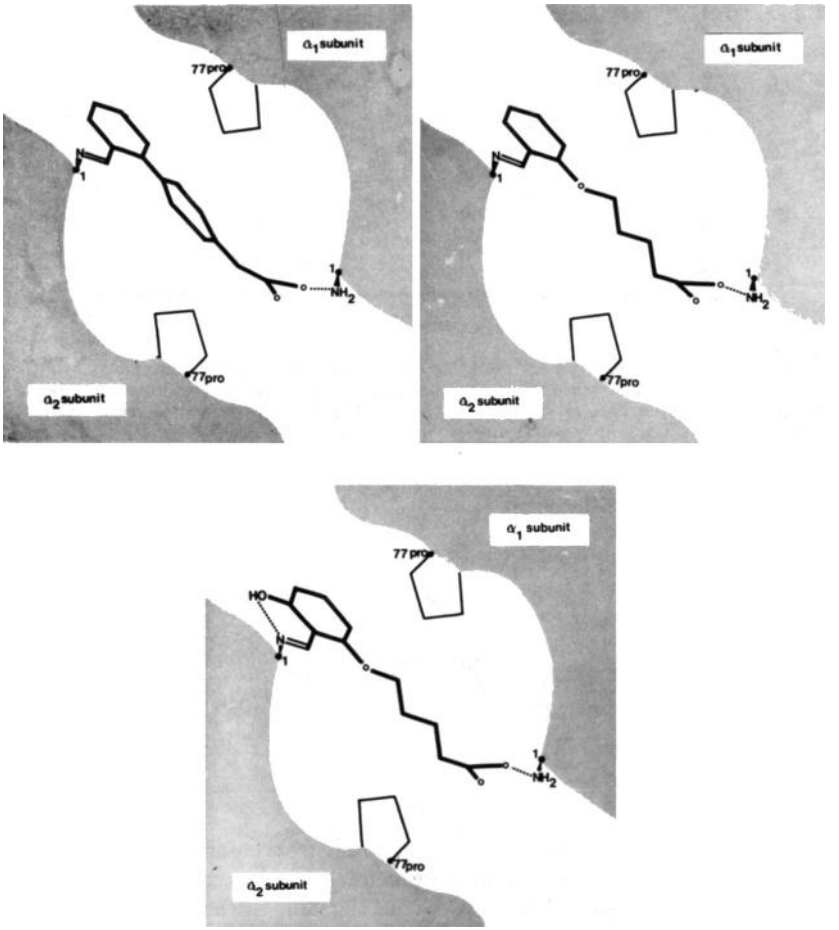
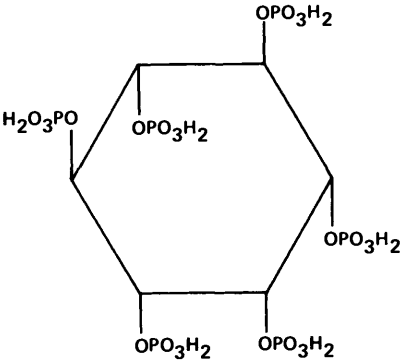


Figure 8 *Diagrammatic representation of the region in R-state haemoglobin between the α-chains, with the postulated manner of binding for compounds* (5) (*top left*), (6) (*top right*), *and* (7)

group adjacent to the aldehyde group would promote Schiff's base formation and hence binding, but owing to initial synthetic difficulties a more flexible analogue (6) was made. This was slightly less effective than (5) in these particular experiments, a finding perhaps attributable to its increased flexibility and less suitable hydrophobic moiety. However, addition of a hydroxy moiety produced compound (7), which was substantially more potent.



(8)

Thus in this study it again proved possible, by consideration of the macromolecule structure, to design selective chemicals to produce a desired effect. The relative potencies of the various compounds can be partially rationalized in terms of the postulated interactions, although it is known that the binding properties of such compounds are more complex than the simple modelling considerations had suggested.[14a] One of these compounds is now under consideration as a potential therapeutic agent for sickle cell disease.[14b] In this disease, red blood cells contain S haemoglobin. At low levels of oxygen in the capillaries, the T state polymerizes and distorts the cells to a sickle shape; these distorted cells do not function normally and clinical symptoms arise. Inhibition of the sickling of such cells would be expected of a compound designed to favour the R-state and thereby to lower the level of T-state, and some such compounds have indeed been shown to inhibit and to reverse *in vitro* sickling.

(ii) *Insulin.* Insulin contains two types of peptide chain, denoted A and B, typically of 21 and 30 residues respectively. These are linked covalently through two disulphide bonds. The hormone circulates in blood primarily at the monomer level, but it is stored intracellularly as a hexamer.[15] Crystals of the hexamer, grown in the presence of zinc ions, have been intensively studied by *X*-ray crystallography.[16] By molecular modelling to the insulin hexamer, Andrews and co-workers[17] have designed small organic molecules to bind in a cavity along the central axis of this

[15] T. Blundell, G. Dodson, D. Hodgkin, and D. Mercola, *Adv. Protein. Chem.*, 1972, **26**, 279.
[16] T. Blundell, J. F. Cutfield, S. M. Cutfield, E. J. Dodson, G. G. Dodson, D. C. Hodgkin, D. A. Mercola, and M. Vijayan, *Nature (London)*, 1971, **231**, 506.
[17] D. T. Manallak, E. F. Woods, and P. R. Andrews, personal communication.

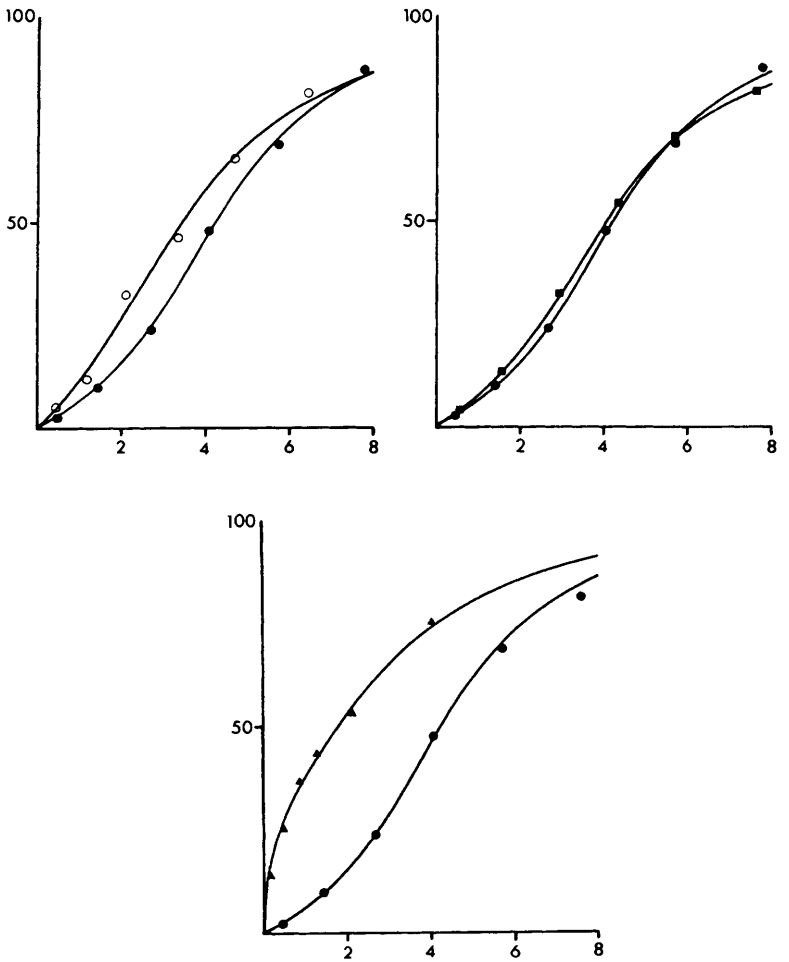
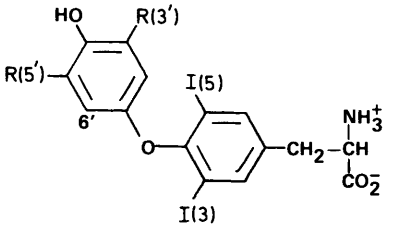
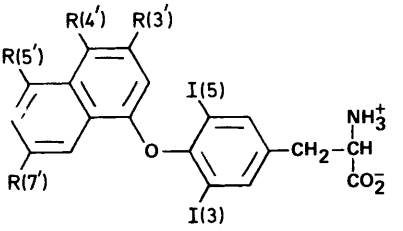


Figure 9 *Each diagram shows oxygen saturation curves for haemoglobin in the presence of* 100 μM *compound* (8) *(solid circles) and for haemoglobin in the presence of both* 100 μM *compound* (8) *and* 50 μM *compound* (5), (6), *or* (7), *with compound assignment as laid out in Figure 8. Relative to the control curve, each compound shifts the oxygen saturation curve to the left, and* (7) *produces the largest shift. Abscissae, oxygen pressure* (kPa), *Ordinate, per cent oxyhaemoglobin*

hexamer and to stabilize the hexamer against dissociation by interaction between subunits. One of the molecules designed was found to influence *in vitro* the distribution of monomer, dimer, tetramer, and hexamer species in the direction of higher molecular weight; the net weight-average molecular weight determined by sedimentation analysis in a solution of insulin was increased by 50%.

(iii) *Prealbumin.* The hormone L-thyroxine (9) interacts with various proteins, including a cell nucleus receptor and in plasma, a globulin and prealbumin. The crystal structures of human serum prealbumin alone[18,19] and in complex with (9)[20] have been used[21] as a basis for macromolecular fitting of thyroxine analogues (10)—(16). Prealbumin is a tetramer of identical subunits which are oriented to form a central channel containing two (9) binding sites. The 222 symmetry of the tetramer requires that not only are the two sites identical, but also that each site itself has on average two-fold symmetry. The binding of (9) to one site presumably alters the other in some way, since the binding of the second molecule is weakened, but statistical averaging in the crystal obscures the finer details. The hormone binds along a two-fold axis with its phenolic hydroxy-group buried deep within the binding channel and its carboxy- and amino-groups ion-paired with side chains of residues lysine-15 and glutamate-54 respectively at the mouth of the binding channel. Figure 10 is a schematic drawing of some of the compound (9)–prealbumin interactions, and shows six pockets capable of binding an iodine substituent, of which four are occupied by the iodines of (9). Of the remaining pockets one is occupied by a crystallographically well-defined water molecule, leaving one pocket empty. Owing to the two-fold symmetry, these six pockets are symmetrically disposed in pairs, denoted in Figure 10 as P(1) and P(1′), P(2) and P(2′), P(3) and P(3′). The inner ring iodines, I(3) and I(5), bind to identical hydrophobic pockets, P(1) and P(1′), lined with the methyl groups of leucine-17, threonine-106, alanine-108, and valine-121 and the polymethylene side-chain of lysine-15. The iodine atoms, 3′ and 5′, of the phenolic ring fit into pockets unrelated by symmetry. The 3′-iodine atom, proximal to the inner ring, binds to pocket P(2), composed of the carbonyl oxygen of lysine-15A, the side chain of leucine-17A, the methyl group of alanine-108A, and the peptide backbone of alanine-109A (the individual protein subunits being denoted, A, B, C, D respectively). The 5′-iodine atom, distal to the inner ring, fits into pocket P(3′) formed by the methyl and





(9) R(3′) = R(5′) = I
(10) R(3′) = R(5′) = Br
(11) R(3′) = Br; R(5′) = H
(12) R(3′) = R(5′) = H

(13) R(3′) = Br; R(4′) = OH; R(5′) = R(7′) = H
(14) R(3′) = R(5′) = R(7′) = H; R(4′) = OH
(15) R(3′) = R(4′) = R(7′) = H; R(5′) = OH
(16) R(3′) = R(4′) = R(5′) = H; R(7′) = OH

[18] C. C. F. Blake and S. J. Oatley, *Nature (London)*, 1977, **268**, 115.
[19] C. C. F. Blake, M. J. Geisow, S. J. Oatley, B. Rérat, and C. Rérat, *J. Mol. Biol.*, 1978, **121**, 339.
[20] C. C. F. Blake, *Proc. R. Soc. Lond., Ser. B.*, 1981, **211**, 413.
[21] J. M. Blaney, E. C. Jorgensen, M. L. Connolly, T. E. Ferrin, R. Langridge, S. J. Oatley, J. M. Burridge, and C. C. F. Blake, *J. Med. Chem.*, 1982, **25**, 785.

carbonyl groups of alanine-108C, the backbone nitrogen and carbonyl group of alanine-109C, the backbone nitrogen and the side chain of leucine-110C, and the hydroxy-groups of serine-117C and threonine-119C. The water molecule is held in pocket P(3) by hydrogen bonds with the hydroxy-groups of serine-117A and threonine-119A and possibly with the phenolic hydroxy-group of (9). A possible close contact (< 3.2 Å) between I(3′) and the carbonyl oxygen of alanine-109A may reflect the polarizability and charge-transfer ability of iodine and may contribute significantly to binding.

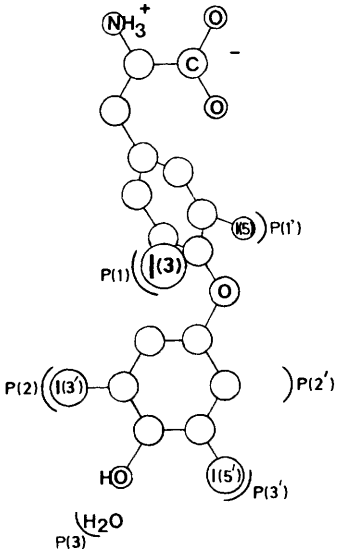


Figure 10 *A schematic illustration of thyroxine* (9) *in its complex with prealbumin. Some pockets* [P(1)—P(3)] *in the protein and symmetry mates* [P(1′)—P(3′)] *in the neighbourhood of* (9)*, accommodate iodine atoms;* P(3) *accommodates water near the phenolic hydroxy-group and* P(2′) *is unoccupied*
(Reproduced by permission of the authors[21] and the American Chemical Society from *J. Med. Chem.*, 1982, **25**, 785)

Molecular modelling of a variety of thyroid hormone analogues by Blaney *et al.*[21] led these workers to conclude that analogues with high binding-affinity would occupy at least three of the four outer-ring pockets P(2), P(2′), P(3), and P(3′); the less empty space in the binding site, the more tightly would the analogue bind. This was thought to be due to the stronger van der Waals attraction associated with increased surface complementarity between ligand and protein. During modelling work it was observed that an empty pocket P(2′) could be filled by the presence of an appropriate substituent at the 6′-position in thyroid hormone analogues. Compounds (10)—(16) were therefore modelled to the protein. The α-napthyl analogues were suitable in part, because of their relative rigidity, thus constraining fitting options during molecular modelling. Based on the complementarity of the

molecular-surface models for the fitting of analogue to protein, the authors predicted that (13) would bind more tightly than either (10) or (11) since (13) is the only analogue which can simultaneously fit all available pockets in the binding site. Similarly (14) was expected to bind more tightly than (12). Comparing (13)—(16), (13) matched the surface of the binding site most closely, with no bad contacts (*i.e.* contacts shorter than van der Waals). Compounds (14)—(16) all lack the bromine substituent, so that one pocket is empty. The phenolic hydroxy-groups on (14) and (15) appeared to be equally well accommodated by the binding site but the hydroxy-group on (16) would 'collide' with the surface of the binding site and (16) might thus bind in a relatively strained manner or in an alternative orientation. Therefore the predicted order of binding affinity for the α-napthyl analogues was K(13) > K(14) ≃ K(15) > K(16).

Table 2 *Binding to prealbumin of some thyroxine analogues: comparison between modelling predictions and binding constants estimated by competition-binding assay (data from ref.* 21)

Binding Constant Ratio	*Predicted Binding Constant Ratio*	*Observed Binding Constant Ratio*
K(13)/K(10)	>1	3.3
K(13)/K(11)	>1	17.9
K(14)/K(12)	>1	5.3
K(13)/K(14)	>1	17.9
K(14)/K(15)	≃1	0.25
K(15)/K(16)	>1	9.0

The match between predicted and measured relative affinity is shown in Table 2. The predicted relative affinities obtained by molecular modelling studies refer to L-enantiomorphs. Likewise the measured affinities also refer to L-enantiomorphs, except for compounds (13), (14), and (15) which were racemic. The authors found that the affinity of racemic (9) is 0.59 that of L-(9), which, if true also for other compounds, would provide the factor necessary to enable the affinity of the L-enantiomorph to be calculated from that obtained with racemic material. The measured relative affinities shown in Table 2 which involve these three compounds incorporate this correction factor. The experimentally determined binding affinities are in the predicted order except for the difference between (14) and (15), which difference was rationalized by more detailed investigation of the fit. It was observed that the hydroxy-group of (15) can interact more strongly with the site than that of (14). The hydroxy-group of (15) in pocket P(3′), comes within hydrogen-bonding distance of the hydroxy-groups of serine-117C and threonine-119C. That of (14) is not deep enough to make such interactions directly, though it might do so through an intermediate water molecule.

In another paper[22] the estimation of binding enthalpy by molecular mechanics computation for D-(9), L-(9), and the deamino (17) and decarboxy (18) analogues was reported. It was concluded that the rank order of binding constants for the four compounds $\{K(17) > K[\text{L-}(9)] > K[\text{D-}(9)] > K(18)\}$ can be matched

[22] J. M. Blaney, P. K. Weiner, A. Dearing, P. A. Kollman, E. C. Jorgensen, S. J. Oatley, J. M. Burridge, and C. C. F. Blake, *J. Am. Chem. Soc.*, 1982, **104**, 6424.

by the calculations, provided in addition an allowance is made for the differences in the free energies of hydration for the complexed and uncomplexed species. Inspection of the modelling work reveals that the analogues are not modelled in a precisely congruent manner. The deamino-compound (17) is modelled with the carboxy-group between symmetrically paired amino-groups of the protein and interacting with both [whereas the carboxy-group of (9) is modelled asymmetrically]. The amino-group of the decarboxy-compound (18) is unable to form good interactions if symmetrically placed and, however placed, makes fewer close interactions than does the carboxy-group of the deamino-compound. Hence it is expected that (17) will bind more tightly than (18) and this was found to be the case. However, it is not immediately obvious why (17) binds more tightly than (9), when (9) has an extra group with which to interact with the protein. The anomaly was attributed to different degrees of desolvation accompanying binding. For (9), with a carboxy- and amino-group, both of which would be desolvated to some extent upon binding, the energetic penalty is greater than for the deamino (or decarboxy) analogues with one less polar group. This work is of particular importance in that it attempts to incorporate the influences upon binding of hydration. Hydration is ubiquitous and hydration energies are substantial and it might seem surprising that so many other studies should have been successful, despite neglect of hydration. Such success might be hoped for in circumstances in which the hydration changes which accompany the interaction of a series of compounds with the macromolecular receptor are 'well-behaved' in the sense that they can be considered to be the sum of a series of components, one for each chemical part of the compounds, and that therefore the hydration-change component for any individual chemical group is constant. This circumstance may well apply when each chemical group, present in a number of different compounds, experiences the same molecular environment regardless of which compound carries it into the binding site in the macromolecule. However, if, for various compounds in the series, the binding site varies in shape, or the compounds bind in different conformations or orientations, it may be essential to consider hydration explicitly before correct predictions can be made of the relative strengths of interaction between compounds and receptor.

B. Enzymic Proteins.—The following examples of studies with enzymes all involve modelling to active sites within a single subunit. The object of each study was the production of compounds which would bind at the active site and thereby inhibit enzymic action.

(i) *Angiotensin-converting Enzyme.* Angiotensin-converting enzyme (ACE) cleaves the inactive decapeptide angiotensin I (19) in blood between the eighth and ninth residues to form an active octapeptide angiotensin II (20) (Figure 11). The octapeptide can induce vasoconstriction, antidiuresis, and antinatriuresis.[23] Inhibitors of the enzyme may thus produce therapeutically useful effects, including reduction in blood pressure. A compatible influence might in principle arise

[23] D. W. Cushman, M. A. Ondetti, H. S. Cheung, E. F. Sabo, M. J. Antonaccio, and B. Rubin, in 'Enzyme Inhibitors as Drugs', ed. M. Sandler, Macmillan Press Ltd., London, 1977, p. 231.

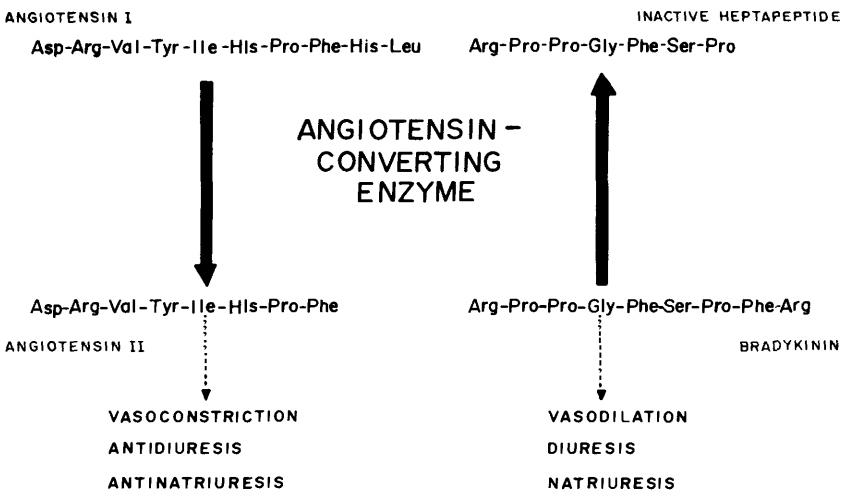
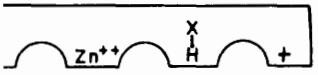
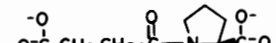
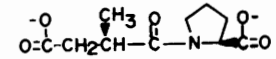
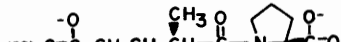
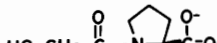
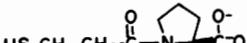
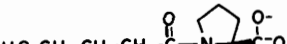
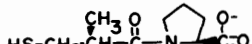


Figure 11 *Reactions effected by ACE. Angiotensin I* (19) *is converted into active hormone angiotensin II* (20), *and bradykinin* (21) *is converted into an inactive heptapeptide* (22) (Reproduced by permission of the authors[23] and Macmillan Press Ltd. from 'Enzyme Inhibitors as Drugs', 1980, p. 231)

as well from the concomitant protection of bradykinin (21) against hydrolysis by ACE to an inactive heptapeptide (22) with a resultant potentiation of bradykinin-induced vasodilation, diuresis, and natriuresis. Many peptide substrates and inhibitors of the enzyme have been found,[23] but of particular relevance in the present context are the molecules shown in Table 3. Biochemical work on the enzyme had revealed that it contains essential zinc and Cushman *et al.*[23] postulated that there might be a relationship between the structure of the active site of this zinc-containing carboxydipeptidase and that known to exist in the zinc-containing bovine pancreatic carboxypeptidase A. This latter enzyme hydrolyses the carboxy-terminal residue from a polypeptide in the manner shown diagramatically in Figure 12. Cushman *et al.* constructed a hypothetical model for ACE in which the hydrolysed peptide group interacted with the zinc atom, and the terminal carboxylate interacted ionically with a cationic centre in the protein. The model for ACE differed from that of carboxypeptidase in that the separation between these interaction sites was increased to change the catalytic activity from carboxypeptidase to carboxydipeptidase (Figure 13). The model of the enzyme also incorporated pockets to accommodate amino-acid side-chains. Byers and Wolfenden[24] suggested that benzylsuccinic acid (23) (Figure 12) is a 'bi-product analogue' in its manner of binding to carboxypeptidase A and one way in which it might bind, due to Cushman *et al.*[23] is shown in Figure 12. Furthermore several peptides with *C*-terminal proline are inhibitors of ACE, and these two considerations led Cushman *et al.*[23] to consider succinylproline (24) as a potential inhibitor of ACE, binding as

[24] L. D. Byers, and R. Wolfenden, *J. Biol. Chem.*, 1972, **247**, 606.

Table 3 *Activities* in vitro *of carboxyalkanoyl and mercaptoalkanoyl amino-acid inhibitors of angiotensin-converting enzyme.* I_{50} *is the concentration of a compound producing a* 50 *per cent inhibition of the activity of rabbit-lung converting enzyme, or a* 50 *per cent inhibition of the contractile activity of angiotensin* I *on guinea-pig ileum strip;* A_{50} *is the concentration producing a* 50 *per cent augmentation of the contractile action of bradykinin*
(Reproduced by permission of the authors[23] and Macmillan Press Ltd. from 'Enzyme Inhibitors as Drugs', 1980, p. 231)

		RABBIT-LUNG CONVERTING ENZYME		GUINEA-PIG ILEUM	
				AI	BK
		I_{50}(μM)	K_I(μM)	I_{50}(μM)	A_{50}(μM)
(24)		330	-	440	37
(25)		22	2.5	18	0.87
(26)		4.9	0.8	23	4.9
(27)		1.1	-	0.19	0.037
(28)		0.20	0.012	0.30	0.025
(29)		9.7	-	12	5.1
(30)		0.023	0.0017	0.023	0.0032

shown in Figure 13. As seen in Table 3, (24) is inhibitory, albeit weakly. Furthermore, the model correctly predicted the enhanced activity seen for analogues of (24) in which a methyl group is attached to the succinyl residue with a configuration which would match that of the corresponding L-residue [at R(2)] in substrate and product. Thus D-2-methylsuccinyl-L-proline (25) is an analogue of Ala-Pro, the terminal dipeptide in a number of peptide inhibitors derived from snake venom. The model is compatible with the high activity also shown by compounds with a chain length slightly longer than succinyl [*e.g.* the glutaryl derivative (26)] but derivatives longer than glutaryl are, as the authors expected, less active.

A very important advance, made with the aid of the rough hypothetical model of the enzyme, arose when the putative carboxylate zinc ligand was replaced by other functional groups with suitable chemical and positional characteristics for tight binding to zinc. Thus (Table 3) a sulphydryl-containing analogue of (24), *i.e.* (28), is about a thousand-fold more potent than (24). Shortening or elongation of (28) lowered potency. Captopril (30) is even more potent, and also about a thousand-fold more potent than the corresponding analogue, (25).

This study shows how fitting to even a rough model of the macromolecule can be

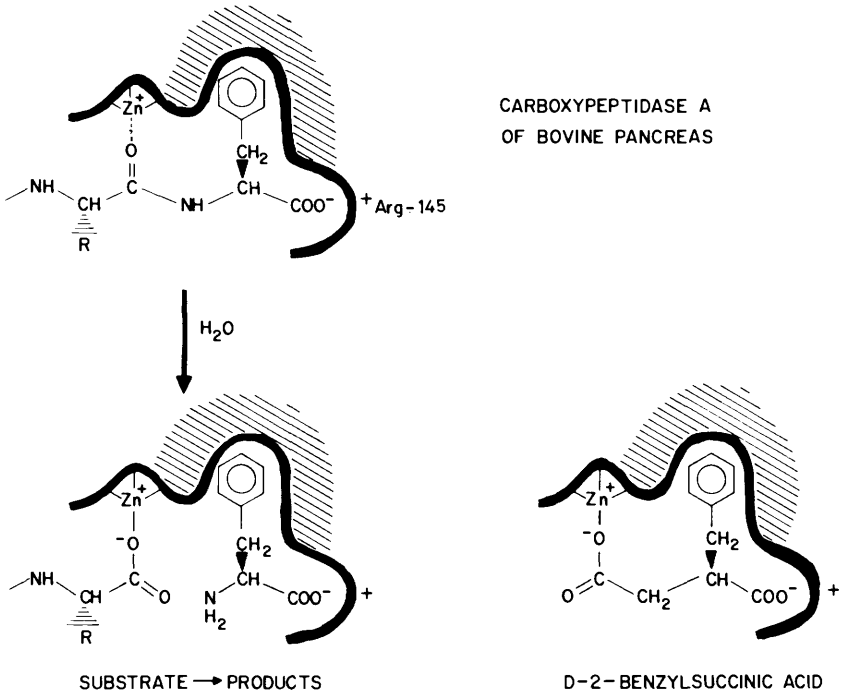


Figure 12 *Diagrammatic model of the binding of peptide substrate (top) and products to the active site of carboxypeptidase. A potential binding mode for the inhibitor* D*-2-benzylsuccinic acid* (23) *is also shown*
(Reproduced by permission of the authors[23] and Macmillan Press Ltd. from 'Enzyme Inhibitors as Drugs', 1980, p. 231)

successfully combined with other sources of knowledge in the design of new drugs. Clearly, the possibility of dramatic improvements in potency through the strengthening of a single putative interaction (with zinc) is a most important factor, but other interactions, such as those made by the methyl substituent, contribute significantly to binding.

(ii) *Carbonic Anhydrase C.* Carbonic anhydrase is inhibited by primary sulphonamides and the utility of acetazolamide (31), both as a diuretic and in the treatment of glaucoma has been attributed to such inhibition. The structure of the C isozyme in the presence and absence of (31) has been determined crystallographically.[25] The enzyme active-site contains a hydrophilic region, in which a zinc atom is co-ordinated to the imidazoles of three histidines; nearby is a hydrophobic region, provided by residues valine-121, isoleucine-91, and phenyl-

[25] K. K. Kannan, I. Vaara, B. Notstrand, S. Lövgren, A. Borell, K. Fridborg, and M. Petef, in 'Drug Action at the Molecular Level', ed. G. C. K. Roberts, Macmillan Press Ltd., London, 1977, p. 73.

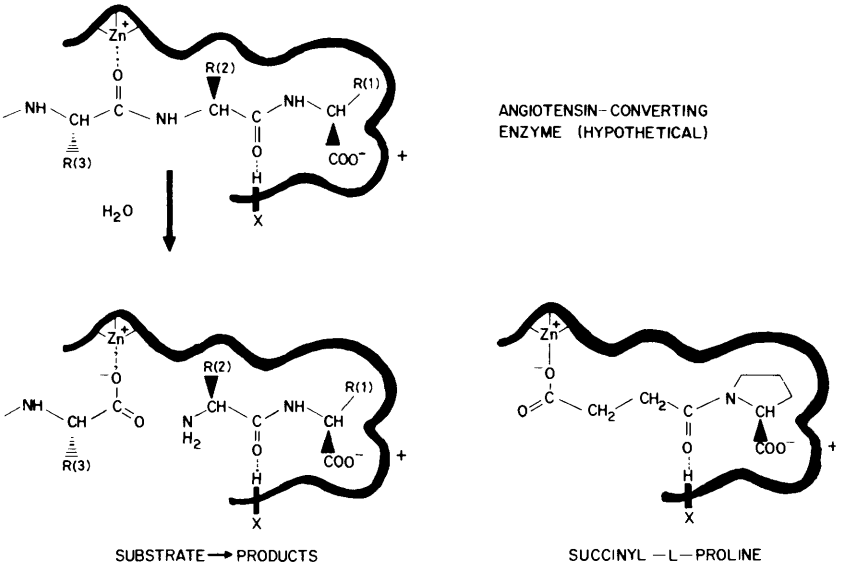
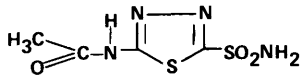


Figure 13 *Diagrammatic representation of a hypothetical model of the active site of ACE with bound substrate (top) and product and with the competitive inhibitor succinyl-L-proline* (24). *X–H represents a group in the protein hydrogen-bonding with the terminal peptide carbonyl group, and residue side chains* R^1 *and* R^2 *are accommodated in pockets. The zinc provides an important interaction with the carbonyl group of the peptide to be hydrolysed (substrate), or with a carboxyl group in product or inhibitor*
(Reproduced by permission of the authors[23] and Macmillan Press Ltd. from 'Enzyme Inhibitors as Drugs', 1980, p. 231)

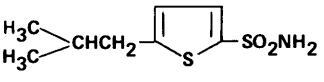
alanine-131. Compound (31) has been described by Gill *et al.*[26] to be interacting as shown in Figure 14, with co-ordination of the sulphonamide nitrogen and of one oxygen to the zinc. The thiadiazole moiety is placed with the sulphur in contact with the hydrophobic side-chain of valine-121, and the nitrogen atoms may be hydrogen bonded with water molecules; the water molecules in turn could be hydrogen bonded to main-chain carbonyl groups of residues threonine-200 and proline-201. The acetamido-chain carbonyl could accept a hydrogen bond from the side-chain nitrogen of glutamine-92. Thus the interactions between ligand and protein are mainly of polar character. Gill and his colleagues, by inspection of a physical model of the active site and by modelling, devised some alternative sulphonamides with groups which might exploit the hydrophobic pocket. The isobutyl thiophene analogue (32) was modelled with the alkyl side-chain in the hydrophobic pocket, the sulphur atom in contact with valine-121, and the sulphonamide group binding to zinc. It was reported to be twice as potent as (31) at inhibiting the enzyme.

(iii) *Dihydrofolate Reductase.* Dihydrofolate reductase (DHFR) is a widely distributed enzyme which regenerates tetrahydrofolate (33) from folate (34) and

[26] E. W. Gill, B. M. Goodall, and P. J. B. Hancock, ref. 1, p. 121.



(31)



(32)

from dihydrofolate (35). Compound (34) arises from spontaneous oxidation of reduced derivatives, especially in foods. The product from DHFR, (33), acquires a methylene group from serine (Figure 15) to become 5,10-methylenetetrahydrofolate (36). This methylene is transferred from (36) to deoxyuridylate and the thymidylate so formed is incorporated into DNA. Concomitant with this transfer is the oxidation of the 5–6 bond, thus generating the DHFR substrate (35) and completing the cycle. Without the active cycle of reactions in which DHFR plays a part, cells will be unable to synthesize DNA, unless they can acquire thymidylate by salvage routes. Thus inhibition of DHFR will stop cell division.

By comparing the inhibitory properties of various diamino-derivatives, Burchall and Hitchings[27] and others demonstrated that their potencies varied with DHFR origin (vertebrate, trypanosomal, plasmodial, and bacterial) and this variation has now been shown to be related to structural differences between the DHFRs. Indeed, we now know the amino-acid sequence of several DHFRs, the three-dimensional architecture of DHFR from mouse, chicken, *E.coli*, and *L.casei*, in

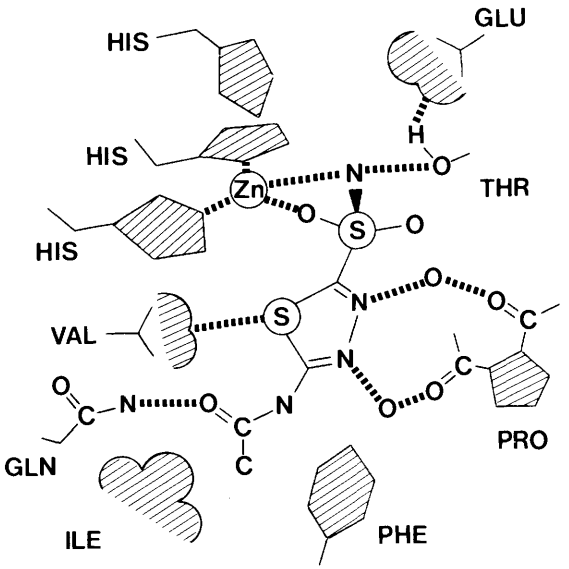
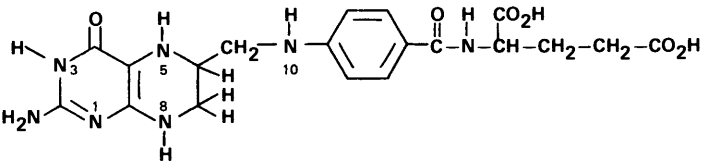
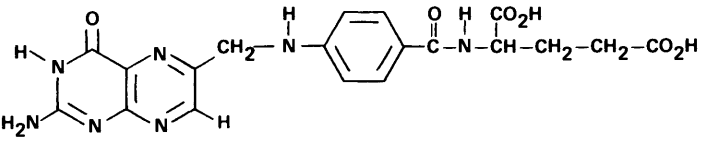


Figure 14 *Schematic illustration of the active site of carbonic anhydrase to show the binding of acetazolamide* (31)
(Reproduced by permission of the authors[26] and Elsevier Science Publishers B.V. from 'Quantitative Approaches to Drug Design', 1983, p. 121)

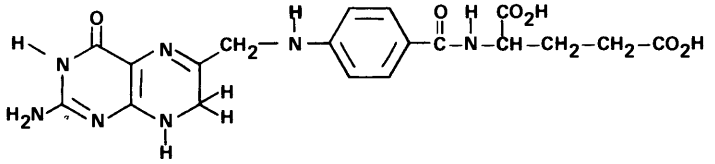
[27] J. J. Burchall, and G. H. Hitchings, *Mol. Pharmacol.*, 1965, **1**, 126.



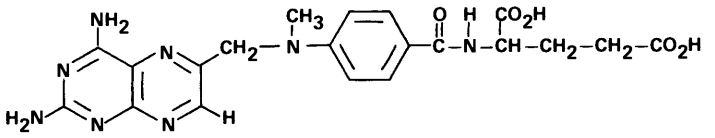
(33)



(34)



(35)



(37)

complexes with inhibitors and/or cofactors, and something of the dynamics and multiplicity of conformational states which can occur (see reviews[28-31]).

DHFR is a target for several drugs. Methotrexate (37) inhibits DHFR from most sources and is commonly used in cancer therapy. Pyrimethamine (38), a specific inhibitor of plasmodium DHFR, is used as an antimalarial. Trimethoprim (39), a selective inhibitor of DHFR in many bacteria, is widely used as an ntibacterial, whilst diaveridine (40) has found some veterinary use as an anticoccidial. As a result of widespread crystallographic, binding, and kinetic studies, DHFR is now perceived to be a useful protein for testing fitting methods. Much work is in progress, and one published study sheds some light on the

[28] G. H. Hitchings, and S. L. Smith, *Adv. Enzyme Regul.*, 1980, **18**, 349.
[29] G. H. Hitchings, and B. Roth, in 'Enzyme Inhibitors as Drugs', ed. M. Sandler, Macmillan Press Ltd., London, 1980, p. 263.
[30] C. R. Beddell, in 'X-Ray Crystallography and Drug Action', ed. A. S. Horn and C. J. De Ranter, Oxford University Press, 1984, p. 169.
[31] G.C.K. Roberts, in 'Chemistry and Biology of Pteridines', ed. J. A. Blair, Walter de Gruyter, Berlin, in press.

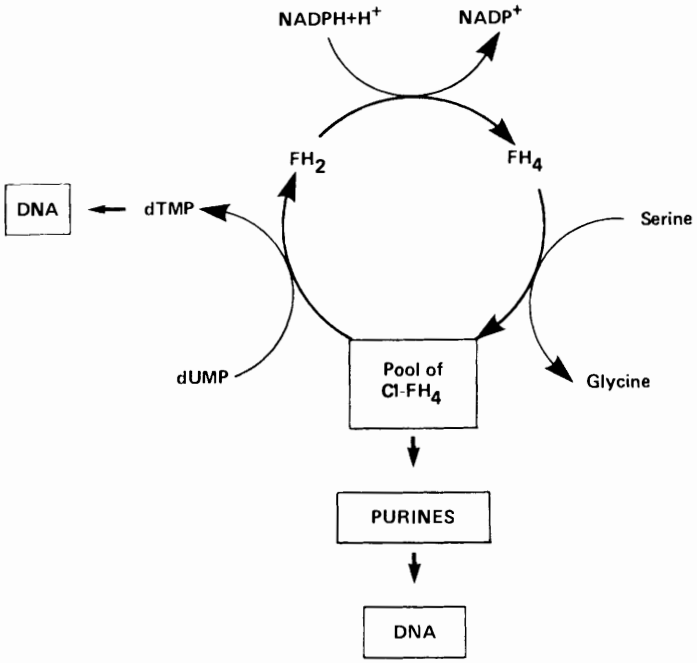
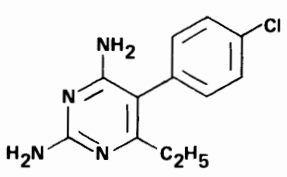
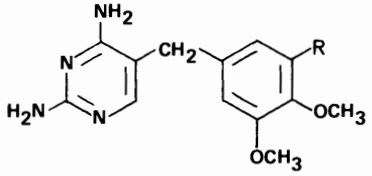


Figure 15 *Some metabolic interrelationships of reduced folate cofactors*

accuracy of fitting predictions. Following the elucidation of the binding mode for (37) to DHFR[32-35] (Figure 16), (39) was modelled to the enzyme.[36] The conformation selected for the compound was compatible with the site geometry, the bound position of (37), and with activity data of 6-substituted compounds, in



(38)



(39) R = OMe
(40) R = H

[32] D. A. Matthews, R. A. Alden, J. T. Bolin, S. T. Freer, R. Hamlin, N. H. Xuong, J. Kraut, M. Poe, M. Williams, and K. Hoogsteen, *Science*, 1977, **197**, 452.
[33] D. A. Matthews, R. A. Alden, J. T. Bolin, D. J. Filman, S. T. Freer, R. Hamlin, W. G. J. Hol, R. L. Kisliuk, E. K. Pastore, L. T. Plante, N. H. Xuong, and J. Kraut, *J. Biol. Chem.*, 1978, **253**, 6946.
[34] D. A. Matthews, R. A. Alden, S. T. Freer, N. H. Xuong, and J. Kraut, *J. Biol. Chem.*, 1979, **254**, 4144.
[35] J. T. Bolin, D. J. Filman, D. A. Matthews, R. C. Hamlin, and J. Kraut, *J. Biol. Chem.*, 1982, **257**, 13 650.
[36] L. F. Kuyper, B. Roth, D. P. Baccanari, R. Ferone, C. R. Beddell, J. N. Champness, D. K. Stammers, J. G. Dann, F. E. A. Norrington, D. J. Baker, and P. J. Goodford, *J. Med. Chem.*, 1982, **25**, 1120.

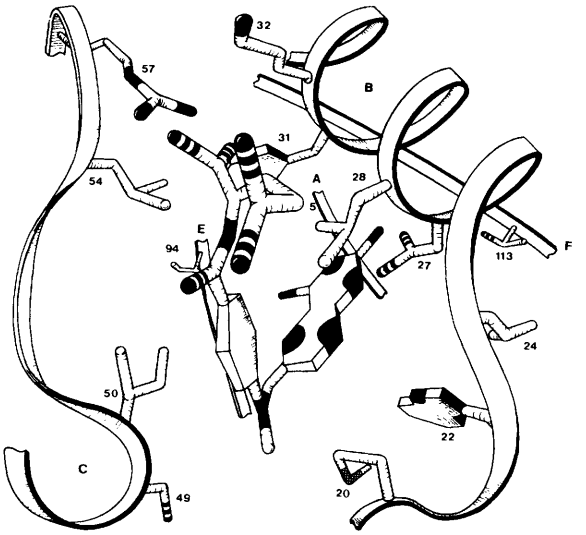


Figure 16 *Schematic illustration of the active site of* E.coli *dihydrofolate reductase with bound methotrexate. Selected atoms are highlighted, oxygen by stripes, nitrogen in black, and sulphur by hatching. The proximity of* N(1) *and the 2-amino-group of methotrexate to the carboxy-group of residue-27 and of the α-carboxy-group to the guanidinium group of residue-57 are examples of charge-assisted hydrogen-bonded interactions. The benzene ring in methotrexate is in hydrophobic contact with the side chains of residue-50 (isoleucine) and -28 (leucine)* (Reproduced by permission of the authors[38] and Walter de Gruyter from 'Chemistry and Biology of Pteridines', 1983, p. 545)

which steric hindrance has been inferred.[37] This conformation was shown shortly thereafter to be essentially correct[38] (Figure 17). In the molecular model of the enzyme with (39) bound there is an arginine side-chain close by. This interacts with the glutamate α-carboxy-group of methotrexate when this binds, by charge-assisted hydrogen bonding. Molecular modelling considerations led to the replacement of one meta methoxy-group of (39) by a carboxyalkoxy-group, and a chain length (five methylene groups) was selected to optimise interaction between the carboxylate and the guanidinium group of this arginine. This compound (45) and analogues [(41)—(44) and (46)] with shorter or longer chains were made (Table 4). Most were more effective than (39) in binding to and inhibiting the enzyme, but the first selected compound (45) was the most potent. Esters (47)—(51) were less potent inhibitors than the respective acid, demonstrating the important contribution to binding made by the free carboxylate group. Furthermore, a crystallographic study[36] of (45) confirmed the mode of binding (Figure 18), whilst study of (42) revealed (Figure 18) that lower potency was associated with loss of some hydrogen

[37] B. Roth, E. Aig, K. Lane, and B. S. Rauckman, *J. Med. Chem.*, 1980, **23**, 535.

[38] (*a*) D. J. Baker, C. R. Beddell, J. N. Champness, P. J. Goodford, F. E. A. Norrington, D. R. Smith, and D. K. Stammers, *FEBS Lett.*, 1981, **126**, 49. (*b*) D. J. Baker, C. R. Beddell, J. N. Champness, P. J. Goodford, F. E. Norrington, B. Roth, and D. K. Stammers, in 'Chemistry and Biology of Pteridines', ed. J. A. Blair, Walter de Gruyter, Berlin, 1983, 545.

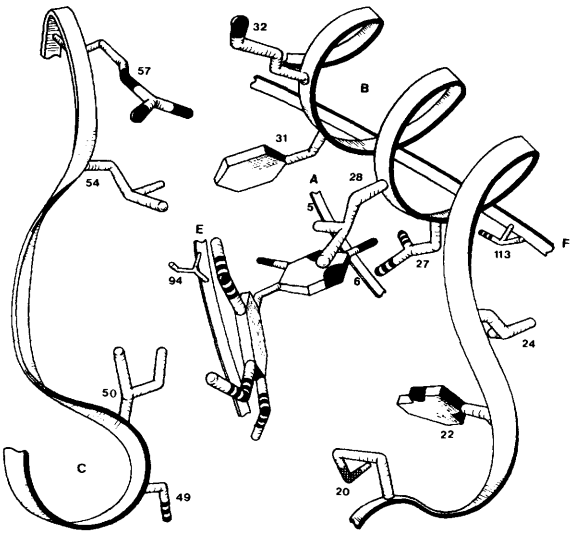


Figure 17 *Schematic illustration of the active site of* E.coli *dihydrofolate reductase with bound trimethoprim. Like methotrexate,* N(1) *and the 2-amino-group interact with aspartate-27. As with the benzene ring in methotrexate, the benzene ring in trimethoprim makes hydrophobic interaction with isoleucine-50 and leucine-28*
(Reproduced by permission of the authors[38] and Walter de Gruyter from 'Chemistry and Biology of Pteridines', 1983, p. 545)

bonding, as expected for a compound with a short carboxyalkoxy-group.

It has been established that (39) adopts a conformation in the complex with vertebrate DHFR[39] different from that which it assumes in bacterial DHFR. Although the architectures of the bacterial and vertebrate DHFRs are rather similar, especially at the active site, (39) is clearly sensitive to the differences that exist. It will be interesting to see whether fitting studies simulate correctly this sensitivity.

In conclusion, trimethoprim and many analogues are able to discriminate in their binding between homologous proteins of similar (though not identical) conformation but of different amino-acid sequence, and the increased binding of certain analogues is associated with the acquisition of additional attractive interactions between compound and protein.

(iv) *Serine Proteases.* Crystallographic studies on the homologous serine proteases—chymotrypsin,[40] trypsin,[41] and elastase[42]—revealed that these endopeptidases have a closely similar main-chain fold and a cleft in which various

[39] D. A. Matthews and K. Volz, in 'Molecular Structure and Biological Activity' ed. J. F. Griffin and W. L. Duax, Elsevier Science Publishing Co., New York, 1982, p. 13.
[40] D. M. Blow, *Acc. Chem. Res.*, 1976, **9**, 145.
[41] R. Huber and W. Bode, *Acc. Chem. Res.*, 1978, **11**, 114.
[42] (*a*) H. C. Watson, D. M. Shotton, J. M. Cox, and H. Muirhead, *Nature* (*London*), 1970, **225**, 806. (*b*) D. M. Shotton and H. C. Watson, *Nature* (*London*), 1970, **225**, 811.

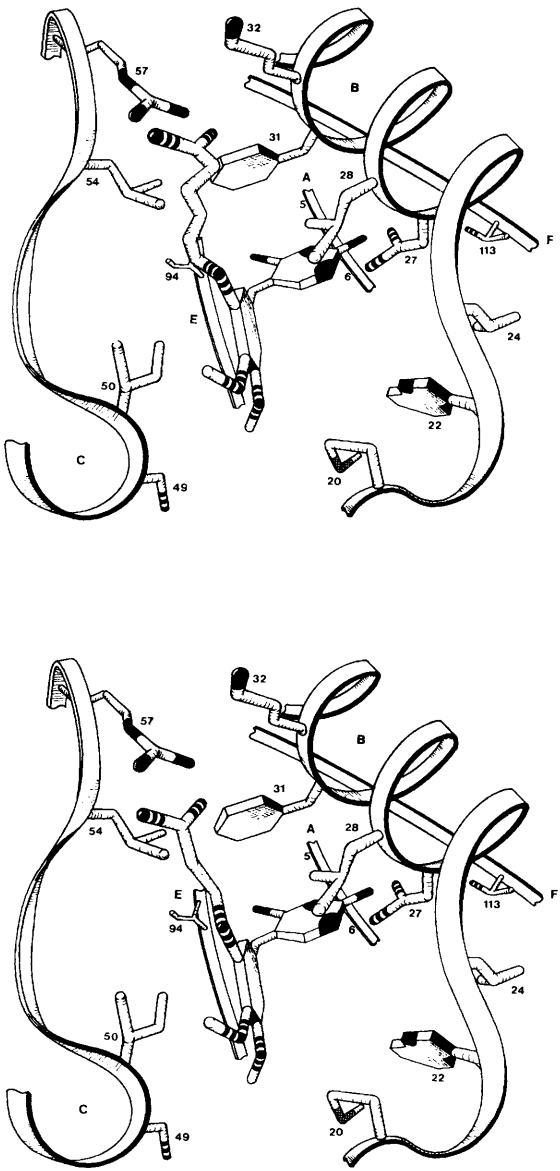


Figure 18 *Schematic illustration of the active site of* E.coli *dihydrofolate reductase with two analogues of trimethoprim. Each bears a carboxylate group, but the longer linkage between this group and the benzene ring in* (45) *(top) permits a closer interaction with the guanidinium group of arginine-57 for this compound than for* (42) *(bottom)*
(Courtesy of D. Baker)

Table 4 *The relative binding to* E. Coli *DHFR of analogues of trimethoprim in the absence (binary complex) and presence (ternary complex) of reduced cofactor (data from ref.* 36)

	R	*Binary complex*	*Ternary complex*
(39)	OCH_3 (trimethoprim)	1.0	1.0
(41)	OCH_2CO_2H	0.8	0.5
(42)	$O(CH_2)_2CO_2H$	3.4	3.5
(43)	$O(CH_2)_3CO_2H$	6.7	37
(44)	$O(CH_2)_4CO_2H$	7.7	20
(45)	$O(CH_2)_5CO_2H$	16	54
(46)	$O(CH_2)_6CO_2H$	7.7	26
(47)	$OCH_2CO_2CH_3$		0.1
(48)	$O(CH_2)_3CO_2CH_2CH_3$		2.8
(49)	$O(CH_2)_4CO_2CH_3$		1.7
(50)	$O(CH_2)_5CO_2CH_3$		1.5
(51)	$O(CH_2)_6CO_2CH_3$		0.7

peptide ligands bind. There exists a bovine pancreatic polypeptide of 58 residues which inhibits trypsin and chymotrypsin and the architecture of this inhibitor is also known.[43] The inhibitor binds to, but is not hydrolysed by, the enzyme. However, if the disulphide bridge cystine-14–cystine-38 in the inhibitor is reduced, trypsin[44] and chymotrypsin[45] hydrolyse the inhibitor between residues lysine-15 and alanine-16. Blow and his colleagues[46] used computer methods to fit the inhibitor to the enzyme, with the susceptible bond in the region of the conserved residues (serine-195 and histidine-57, chymotrypsin numbering) which had been indicated from various studies to be directly concerned with hydrolysis. A unique fit was derived for the complex between inhibitor and chymotrypsin[46] and trypsin[47] respectively. These predictions were confirmed and proved to be accurate when the crystal structure of the trypsin–inhibitor complex[48,49] was subsequently determined. The only apparent problem with the computed fit was the unduly close approach of the side chain of serine-195 to the main chain of the inhibitor in the region of lysine-15 and alanine-16. These contacts were compatible with those expected to precede the formation of a tetrahedral adduct between the serine-195 hydroxy-group and the carbonyl carbon of lysine-15, and distortion of this region was reported for the complex. This study illustrates the ability of skilled workers correctly to predict fit between macromolecule and ligand, by optimization of steric fit and complementarity.

[43] J. Deisenhofer, and W. Steigemann, *Acta Crystallogr., Sect. B.*, 1975, **31**, 238.
[44] L. F. Kress, and M. Laskowski, *J. Biol. Chem.*, 1967, **242**, 4925.
[45] M. Rigbi, in 'Proc. Inst. Res. Conf. on Proteinase Inhibitors', Walter de Gruyter, Berlin, 1971, p. 74.
[46] D. M. Blow, C. S. Wright, D. Kukla, A. Rühlmann, W. Steigemann, and R. Huber, *J. Mol. Biol.*, 1972, **69**, 137.
[47] R. Huber, D. Kukla, A. Rühlmann, and W. Steigemann, *Cold Spring Harbor Symp. Quant. Biol.*, 1971, **XXXVI**, 141.
[48] R. Huber, D. Kukla, W. Bode, P. Schwager, K. Bartels, J. Deisenhofer, and W. Steigemann, *J. Mol. Biol.*, 1974, **89**, 73.
[49] A. Rühlmann, D. Kukla, P. Schwager, K. Bartels, and R. Huber, *J. Mol. Biol.*, 1973, **77**, 417.

Hassall and his colleagues used the results of crystallographic studies on serine proteases to design[50] elastase inhibitors. The motive for the work was the possible use of such inhibitors to treat diseases where tissue degradation by elastase (*e.g.* pancreatitis, arthritis, emphysema) might be occurring. The study started from *N*-acylated derivatives of $(Ala)_3$ or Ala-Pro-Ala. The working hypothesis assumed that these bound with the *C*-terminal residue in site S(1) adjacent to the active-site serine. By examining a physical model of the enzyme, these workers concluded that the *C*-terminal residue could be replaced by a carboxylate-free moiety. Substitution by *N*-cycloalkyl groups improved inhibitory potency, estimated as K_i, by up to nearly two orders of magnitude. However, the authors also indicate that the presumed binding sites S(1), S(2), S(3) for the respective residues, deduced by analogy with those known for trypsin and chymotrypsin, in fact differ from those observed when peptides are crystallographically observed complexed with elastase. The binding of two inhibitors from this study differs from that of the above peptides. Thus, in the case of elastase, which is a peptidase with a shallow cleft, the binding mode is evidently susceptible to modest structural changes in the ligand and direct observation of binding is a much-needed check on the validity of the binding-mode hypothesis which is being used to assist the choice of new structures for synthesis.

C. Non-protein Macromolecules.—Nucleic acids are the only non-protein macromolecules reviewed here. Work on, for example, the guest–host complex with crown ethers and cyclodextrins and studies on macromolecular carbohydrates, though adding greatly to our understanding of complementarity, falls outside the scope of the present review.

(i) *Nucleic Acid.* The Watson–Crick double-stranded right-handed B-helix model[3] for DNA, which is supported by the DNA fibre-diffraction data of Wilkins, Franklin, and co-workers, has hydrogen-bonding between the nucleotide bases, adenine with thymine and guanine with cytosine. It is considered to represent an important form of intracellular DNA, although other right-handed forms (*e.g.* A), left-handed forms (*e.g.* Z) and unpaired forms may also be functionally important. Some compounds bind to DNA by interaction with the backbone or by hydrogen-bonding with bases accessed *via* the minor or major groove. Most compounds which bind tightly to DNA possess a large polarizable aromatic polycyclic system which might in principle interpose in the stack of base pairs, pushing them apart along the helix axis by about 3.4 Å, the van der Waals thickness of the polycyclic system. Such an intercalation model was proposed over twenty years ago by Lerman[51] (Figure 19), and since then there have been many attempts better to define the extent of DNA elongation, unwinding, base-sequence specific binding, and neighbourhood effects associated with intercalation and also the precise nucleotide geometries at and near the intercalation site.

[50] C. H. Hassall, W. H. Johnson, and N. A. Roberts, *Bioorg. Chem.*, 1979, **8**, 299.
[51] L. S. Lerman, *J. Mol. Biol.*, 1961, **3**, 18.

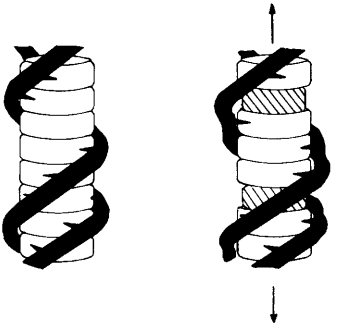


Figure 19 *Schematic illustration of the DNA duplex (left) and of the intercalation of a compound (shaded) according to the Lerman model. The DNA sugar–phosphate backbone is shown in black with the attached nucleotide bases unshaded discs*
(Courtesy of S. Neidle)

In a recent review of intercalation Neidle[52] concludes:

> 'It is now apparent that drug intercalation is a much more complex family of phenomena than was hitherto imagined. Structural studies have started to reveal the extent of these subtleties; many more are needed to clarify questions of, in particular, sequence specificity and intercalation into long sequences of oligonucleotide. Such analyses will undoubtedly enable rational drug design to be conducted to higher levels of sophistication than at the present time'.*

Many drugs act at the DNA level (*e.g.* reviews[53,54]) and it is perhaps fortunate that the complexities have emerged only recently, as workers in this field might otherwise have been deterred from any attempts at designing drugs to fit DNA. Indeed, in a quite recent review of DNA-directed drug design, Henry[55] concluded that whilst little knowledge exists to support the design of tumour DNA-specific anti-tumour drugs, nevertheless active and novel DNA-directed compounds have been produced which might not have been considered in the absence of receptor concepts. For example, Kundu *et al.*[56] attempted to design a non-intercalating DNA binding compound. It was anticipated that a functionalized cyclopenta[*f*]isoquinolin-3(2*H*)-one (52) would hydrogen-bond to a cytosine–guanine base-pair, as shown in Figure 20. This compound showed a low level of cytotoxicity and did bind weakly to calf DNA, with some apparent selectivity for cytosine–guanine base-pairs.

* Reproduced by permission of the author[52] and Oxford University Press, from '*X*-Ray Crystallography and Drug Action', 1984, p. 129.

[52] S. Neidle, in 'X-Ray Crystallography and Drug Action', ed. A. S. Horn, and C. J. De Ranter, Oxford University Press, 1984, p. 129.

[53] S. Neidle, in 'Progress in Medicinal Chemistry', ed. G. P. Ellis and G. B. West, Vol. 16, Elsevier/North Holland Biomedical Press, Amsterdam, 1979, p. 151.

[54] 'Molecular Aspects of Anti-cancer Drug Action', ed. S. Neidle and M. J. Waring, Macmillan Press Ltd., London, 1983.

[55] D. W. Henry, in 'New Approaches to the Design of Antineoplastic Agents' ed. T. K. Bardos and T. I. Kalman, Elsevier Science Publishing Co., Amsterdam, 1982, p. 5.

[56] (*a*) N. G. Kundu, J. A. Wright, K. L. Perlman, W. Hallett, and C. Heidelberger, *J. Med. Chem.*, 1975, **18**, 395. (*b*) N. G. Kundu, W. Hallett, and C. Heidelberger, *J. Med. Chem.*, 1975, **18**, 399.

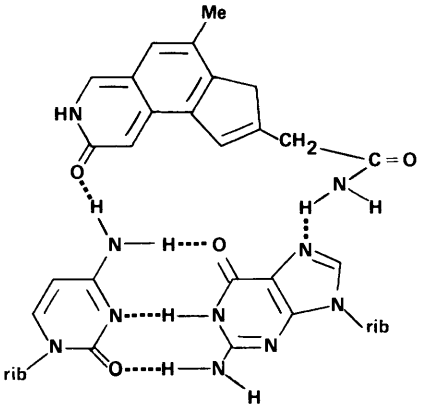
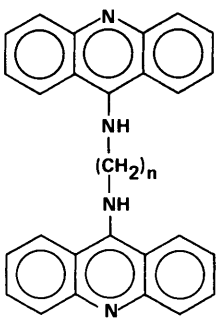
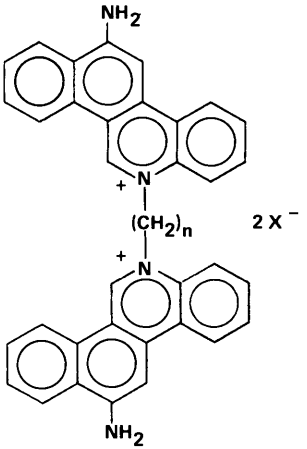


Figure 20 *Postulated hydrogen bonding between* (52) *and DNA nucleotide bases*

The intercalation concept has been incorporated into design, especially in regard to the development of bifunctional intercalators. These are molecules with two planar polycyclic systems, linked with a spacer sufficiently long that the ring systems can both intercalate, at different sites in the DNA helix. Chen *et al.*[57] prepared bis-9-aminoacridines (53), the rings linked with chains of various lengths. Compounds with a chain length, *n*, of 6 and greater could double intercalate and showed cytotoxic and/or *in vivo* anti-tumour activity.[58,59] However, the lack of correlation



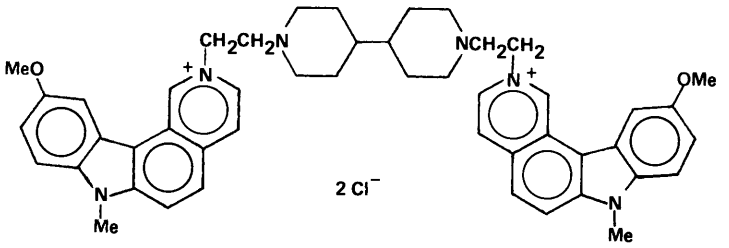
(53)



(54)

[57] T. K. Chen, R. Fico, and E. S. Canellakis, *J. Med. Chem.*, 1978, **21**, 868.
[58] R. G. McR. Wright, L. P. G. Wakelin, A. Fieldes, R. M. Acheson, and M. J. Waring, *Biochemistry*, 1980, **19**, 5825.
[59] L. P. G. Wakelin, M. Romanos, T. K. Chen, D. Glaubiger, E. S. Canellakis, and M. J. Waring, *Biochemistry*, 1978, **17**, 5057.

between cytotoxic and anti-tumour effects led to investigation of the anti-tumour mechanism[57,60] and revealed that this involved action at the cell surface.

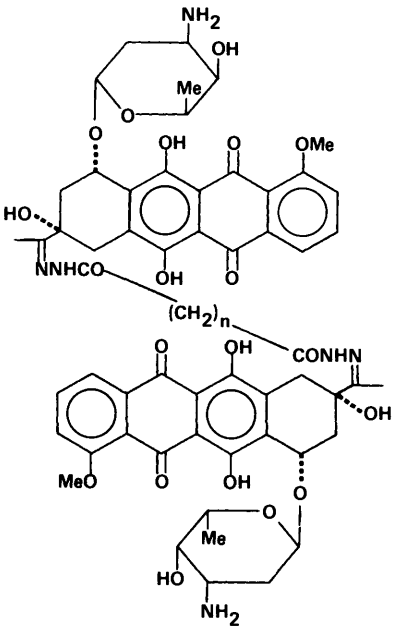


(55)

Henry and his colleagues[61] designed the 12-aminobenz[*i*]phenanthridine (54) to provide substantial charge delocalization, on the grounds that such a property was a general feature of intercalating nitrogen heterocycles. This system also appeared from modelling studies to intercalate well and some mono-functional molecules were found to have significant anti-tumour activity, although bifunctional molecules were the most potent and efficacious. Chain lengths of 3 and 10 were less effective than those in the range 7—9. Le Pecq and co-workers[62] prepared the double intercalator (55) which cured a substantial percentage of animals in the L1210 murine leukaemia experimental anti-tumour system. Small changes in the structure often abolished activity; it may be important for activity in potential bis-intercalators for the structure to resist self-stacking of the aromatic parts. Daunorubicin and adriamycin are intercalative DNA binders used clinically as anti-cancer agents. Several detailed potential binding modes have been reviewed[53] and one such[63] was used for further development. In this model, the molecule intercalated. In the wide groove of the DNA double-helix three successive phosphate groups of one DNA chain interacted with, respectively, the 9-hydroxy-, the protonated 3′-amino- and the 4′-hydroxy-groups. The 9-acetyl group of daunorubicin was not involved in interactions, and thus it was felt that coupling of two daunorubicin molecules through the 9-position might generate an active bifunctional molecule. One such series of molecules is (56). Some molecules were indeed bifunctional intercalators and proved to have substantial anti-tumour activity. The crystallographic determination[64] of the structure of a complex between daunorubicin and a self-complementary hexanucleotide has now provided a detailed picture of one way in which such molecules might intercalate in DNA.

[60] R. M. Fico, T. K. Chen, and E. S. Canellakis, *Science*, 1977, **198**, 53.
[61] W. Fleming, M. Lerom, P. Sturm, C. Mosher, W. W. Lee, D. Taylor, and D. W. Henry, in preparation.
[62] D. Pelaprat, A. Delbarre, I. Le Guen, B. P. Roques, and J.-B. Le Pecq., *J. Med. Chem.*, 1980, **23**, 1336.
[63] D. W. Henry, in 'Cancer Chemotherapy', ed. A. C. Sartorelli, ACS Symposia Series No. 30, 1976, p. 15.
[64] G. J. Quigley, A. H.-J. Wang, G. Ughetto, G. van der Marel, J. H. van Boom, and A. Rich, *Proc. Natl. Acad. Sci., USA*, 1980, **77**, 7204.

Other cases of bis-intercalator design are known (see refs. 65—69, for example, and references therein) but will not be reviewed here. Although the detailed models used in these various studies of the intercalation complex may in some instances be wrong in detail, it appears that even gross features inherent in such a complex are sufficient to point the way towards new and useful molecules.



(56)

3 Receptor Mapping

The present review is concerned with 'receptor fitting', defined and exemplified above. It is the author's experience that such fitting has commonly been confused with 'receptor mapping' and even with 'receptor structuring'. To assist those following the literature a brief comparison of the terms is provided here (see also Gund [70]).

Receptor mapping is the process of deducing the structure and other properties of the receptor by studying the properties of small ligand 'probes', often themselves drugs, which interact to various degrees with the receptor. Naturally some parts of any one ligand may not interact appreciably with the receptor. Those parts of the

[65] M. M. Becker, and P. B. Dervan, *J. Am. Chem. Soc.*, 1979, **101**, 3664.
[66] K. F. Kuhlmann, N. J. Charbeneau, and C. W. Mosher, *Nucleic Acids Res.*, 1978, **5**, 2629.
[67] B. K. Sinha, R. M. Philen, R. Sato, and R. L. Cysyk, *J. Med. Chem.*, 1977, **20**, 1528.
[68] B. F. Cain, B. C. Baguley, and W. A. Denny, *J. Med. Chem.*, 1978, **21**, 658.
[69] J. W. Lown, B. C. Gunn, K. C. Majumdar, and E. McGoran, *Can. J. Chem.*, 1979, **57**, 2305.
[70] P. Gund, *Trends Pharmacol. Sci.*, 1982, **3**, 56.

several ligands which do interact will provide good structural clues and are collectively used to define the 'pharmacophore'. Thus, part of the receptor-mapping approach involves establishing which parts of ligands comprise the pharmacophore. It is then necessary to deduce the conformation of the receptor-bound ligands and their mutual orientations. The ligands are then considered in superposition and receptor groups are postulated in the space around the ligands in a way which would give rise to interactions between groups in the receptor model and groups in the ligands compatible with the experimental observations of the interaction (as manifest in, for example, binding, agonism, or antagonism). The approach is well illustrated by recent work employing graphics as an aid to three-dimensional representation.[71-73]

Receptor structuring is the direct determination of receptor structure with wave/particle probes (*e.g. X*-rays, neutrons, and electrons in crystallography, radiofrequency electromagnetic radiation in n.m.r., electrons in electron microscopy). *X*-Ray crystallography has been especially valuable in providing a detailed and comprehensive picture of macromolecular architecture.

Thus receptor mapping and receptor structuring are, respectively, indirect and direct ways of establishing receptor structure. Receptor mapping is also related in a complementary way to receptor fitting and perhaps it is this complementarity which causes confusion. The receptor-fitting process predicts how a ligand of known chemical structure would interact with a receptor of known architecture; receptor mapping predicts how such ligands would interact with postulated models of the receptor. In receptor fitting, therefore, it is primarily the small molecule which is adjusted to fit the large; in receptor mapping there is much adjustment of the receptor model to fit the ligands. Naturally, considerable skill is needed in both procedures, and in the latter it is particularly difficult to establish internally the accuracy and uniqueness of the model. There may be numerous potential models for the receptor that would lead to predicted interactions with ligands which are in accord with the measurements of the interactions. The predictions will also be sensitive to the assignment of the pharmacophore, conformation, and relative orientation for each ligand. In principle these could all be established by a direct determination of the architecture of the complex between each ligand and receptor. Then ligand conformation and disposition are observed directly and interaction between parts of the ligand and receptor are inferred by proximity and used to define the pharmacophore. From this it would then be possible to map the receptor and compare the map with the directly observed structure of the receptor. Indeed there are already some studies with DHFR[74-76] and with prealbumin,[77] 'receptors' of known structure, which are leading towards the production of receptor models by receptor mapping. Similar work, if genuinely performed

[71] G. R. Marshall, in ref. 1, p. 129.
[72] W. E. Klunk, B. L. Kalman, J. A. Ferrendelli, and D. F. Covey, *Mol. Pharmacol.*, 1983, **23**, 511.
[73] J. R. Sufrin, D. A. Dunn, and G. R. Marshall, *Mol. Pharmacol.*, 1981, **19**, 307.
[74] G. M. Crippen, *J. Med. Chem.*, 1980, **23**, 599.
[75] A. K. Ghose, and G. M. Crippen, *J. Med. Chem.*, 1983, **26**, 996.
[76] A. K. Ghose, and G. M. Crippen, in ref. 1, p. 99.
[77] G. M. Crippen, *J. Med. Chem.*, 1981, **24**, 198.

without recourse to crystallographic information on such macromolecules, could be used to evaluate different receptor-mapping methods through comparison with the rapidly increasing structural information.

Some idea of the difficulty of meaningful receptor mapping may be gained from the studies of Hopfinger,[78-79] in which protein *X*-ray crystallographic and n.m.r. information was explicitly ignored to enable the reliability of indirect methods to be ultimately established. From partially hindered analogues of trimethoprim, taken together with enzyme inhibition data, a conformation for trimethoprim in its complex with *E.coli* DHFR was inferred and used to calculate shape and potential energy parameters for other analogues, which in turn were tested for correlation with bovine DHFR inhibition data. However, the deduced conformation for trimethoprim in the *E.coli* DHFR, as judged by the torsion angles quoted, was not the same as that deduced crystallographically. Similarly n.m.r.[80] data for trimethoprim in bacterial or in vertebrate DHFR failed to confirm the calculated conformation. The difficulty of assigning correctly the bound conformation by indirect methods is indeed substantial for most drugs, since these are commonly flexible molecules, and suitable rigid or semi-rigid analogues of most are not available. Establishment of mutual frames of reference may also not be straightforward. The substrate for DHFR differs from such 4-amino-containing inhibitors as aminopterin and (37) by virtue of its 4-oxo-group. There is now good evidence[35,81,82] that this small difference causes a 180° difference in orientation of the pteridine ring in the DHFR complex. Thus it is clearly difficult to establish indirectly the receptor-bound conformations of individual ligands and to establish their mutual orientations, and the receptor-mapping technique, being crucially dependent on such knowledge, is liable to proceed in error.

4 Characteristics of the Receptor-Fit Approach

A. Form of Relationship.—In the receptor-fit approach the derivation of a relationship between properties of the compound and the biological effect occurs in two steps. First, the properties of groups in the compound in conjunction with corresponding properties for the receptor are used in the estimation of the interaction between compound and receptor. To do this, physical or computer-based models of the molecules are manipulated, the interaction being calculated by means of an assumed relationship between interaction strength and molecular properties. The relationship used varies from guesswork, through simple additivity rules, to molecular mechanical and quantum mechanical formulations. The second step relates the calculated interaction to the biological/biochemical effect and will

[78] A. J. Hopfinger, *J. Med. Chem.*, 1981, **24**, 818.
[79] A. J. Hopfinger, *J. Med. Chem.*, 1983, **26**, 990.
[80] B. Birdsall, G. C. K. Roberts, J. Feeney, J. G. Dann, and A. S. V. Burgen, *Biochemistry*, 1983, **22**, 5597.
[81] P. A. Charlton, D. W. Young, B. Birdsall, J. Feeney, and G. C. K. Roberts, *J. Chem. Soc., Chem. Commun.*, 1979, 922.
[82] J. C. Fontecilla-Camps, C. E. Bugg, C. Temple, J. D. Rose, J. A. Montgomery, and R. L. Kisliuk, *J. Am. Chem. Soc.*, 1979, **101**, 6114.

of course depend upon the nature of the effect and upon the mechanisms operating in the macromolecular system. Thus for DHFR, inhibitory potency has been related to binding constant (and hence to binding free-energy) by various simple relationships.[83] For haemoglobin, the two-state mechanism and refinements generate rather more complex equations; nevertheless, from these, a measure of effect, such as a change in the partial pressure of oxygen at a selected degree of saturation, or fractional liberation of oxygen at fixed partial pressure, can be related to the individual affinities of the compound to each of two conformational states.[84] There is then, with adequate knowledge of the mechanism by which the macromolecule contributes to effects, *'a priori'* reason to anticipate the *form* of relationship. Such *'a priori'* reasoning is of course also encountered in situations besides molecular fitting. If, for example, partitioning of a compound through lipid is important, a linear free-energy relationship involving partition coefficient may be apparent, or if species of different charge partition and pK_a is important, a linear free energy relationship involving Hammett's σ might be sought. However, in most instances the form of relationship cannot be anticipated *'a priori'*, as the mechanisms by which the drugs interact in the biological system are not understood in detail.

B. Multiplicity of Properties.—Many types of physicochemical property might conceivably be involved in producing effects in biological systems, and even more parameters are needed to describe properties which vary as a function of position (position-dependent scalars) and also of direction (position-dependent vectors). Only a very small proportion is ever considered in practice. The primary hope in classical P.A.R. studies is that one or two properties only will be of importance in at least some closely defined region of property–activity space. In the receptor-fitting studies described above, a number of qualitatively different properties, *e.g.* flexibility, covalent bond geometry, steric repulsion, hydrophobicity, hydrogen-bond potential, and ionic properties, are considered concurrently in a manner which makes allowance for positional dependence and vectorial nature. A drawback is that this has largely been done through interaction with physical models and therefore is subject to problems of incompleteness, some subjectivity, and imprecision. However, this situation is now changing with the introduction of a combination of graphics-docking procedures to generate trial fits and subsequent molecular mechanics energy minimization procedures for the estimation of binding enthalpy. With the aid of modern computer hardware and software, it is feasible to attempt some estimate of the contribution to binding of several different properties, each assigned at numerous different positions in the small molecule. The molecular mechanics approach employed by Kollman and co-workers has been mentioned. This approach has been much used by Scheraga,[85] Warshel,[86]

[83] D. P. Baccanari, S. Daluge, and R. W. King, *Biochemistry*, 1982, **21**, 5068.
[84] P. J. Goodford, J. St-Louis, and R. Wootton, *Br. J. Pharmacol.*, 1980, **68**, 741.
[85] M. R. Pincus, and H. A. Scheraga, *Acc. Chem. Res.*, 1981, **14**, 299.
[86] A. Warshel, *Acc. Chem. Res.*, 1981, **14**, 284.

North[87,88] and co-workers to study protein–ligand interactions and recently also by Neidle[52] and Kollman[89] to study drug–nucleotide complexes. The explicit neglect of entropy remains a potential problem, although where correlations between free energy and interaction enthalpy have been found, entropy neglect has presumably not been a confounding factor. The nature of the algorithms and parameters used in the molecular mechanics approach at present are simple. Sophisticated treatments, taking into account in a more precise way, for example, the influence of solvent, and also property changes associated with polarization, are under study. In some methods of a statistical (Monte-Carlo) or dynamical nature, the estimation of entropy becomes approachable.

C. Superposition.—The mutual alignment of different ligands emerges naturally from the process of fitting to the macromolecule. Thus, if several potential fits of a given compound are generated interactively and that which is calculated to bind with minimum binding-enthalpy selected as the most relevant, then the position and orientation of each compound in turn is established in a single frame of reference based upon the macromolecule. This consequently defines the superposition of the various compounds as a by-product. In principle, the alignment problem is solved in this procedure, although it is clear that, owing to the limited radius of convergence of minimization methods, and their inability to overcome large energy barriers in the manner in which they have so far been used in this type of work, it is unlikely that the true fit would be discovered if one fairly close to it in nature were not generated in the interactive docking process. For success therefore, the method requires substantial skill. As mentioned previously, it has been established that the roughly isoelectronic and isosteric substitution of amino for oxo at the 4-position of pteridine derivative ligands of DHFR inverts the orientation of the bound pteridine. Hence one must be alert to the possibility of substantial differences in orientation arising from local changes in the structure of the ligand.

Although superposition is a byproduct, it is of course true that the receptor itself may adopt different conformations for each compound, and compounds may bind at different sites. The whole concept of superposition underlying classical P.A.R. may be misleading when one attempts to superimpose drugs in order to define properties in common, if the drugs bind in different molecular environments. In these circumstances receptor fitting, being independent of assumptions concerning congruence amongst molecules, may be more robust than P.A.R. methods based on superposition. This fundamental difference of course arises because classical P.A.R. methods are essentially correlation methods, dealing with *what* the biological system does in response to different drugs. Receptor fit is one of several mechanistic approaches that use physicochemical knowledge of the system, *i.e. how* it operates, to predict response.

[87] S. B. Brown, A. A. Chabot, E. A. Enderby, and A. C. T. North, *Nature (London)*, 1981, **289**, 93.
[88] E. A. Potterton, in 'Computer Modelling of DHFR-Ligand Interactions', PhD Thesis, Leeds University, 1983.
[89] A. Dearing, P. Weiner, and P. A. Kollman, *Nucleic Acids Res.*, 1981, **9**, 1483.

D. Finding the Local Best Compound.—Consideration of the receptor structure focuses attention on drug action through certain molecularly defined mechanisms. Binding of the drug to the receptor is a prerequisite and binding may be sterically competitive with a natural substrate or regulator, or may selectively favour allosterically a pre-existing or induced state of the receptor. The binding process involves complementarity between ligand and receptor and thus at each position in space in the vicinity of the receptor there is, in principle, an assignable property which must be possessed by the drug for complementarity to exist. The problem then is to optimise complementarity by changing the compound structure, and this is currently done by intelligently devising new structures with locally improved properties. In principle this structural search could be done by computer-automated structure generation with matching to a property field, but the computational problem is not simple in the general case and has yet to be accomplished. However, the illustrative studies described above show that, as with the classical P.A.R. approach, it is not too difficult to select compounds with improved properties, once it is known which properties matter. Furthermore, when structures are generated and fitted, the enthalpy calculation can now provide a relative index of fit and thus a numerical prediction of effectiveness. It is too early to show how reliable this measure is, but preliminary indications are that even the very simple approach based on the totting up of interactions can be useful, so the more detailed calculations of modern molecular mechanics should not fall short of this.

E. Generating Structures in New Areas of Property Space.—In exploring new types of structure, the receptor-fit approach provides guidance and potential binding sites can be explored to which there is no known ligand. The fitting of compounds to such sites requires skill and imagination and will be quite open ended if a great many different types of structure would in fact interact favourably at the site. The medicinal chemist does not need to try to explore all the possibilities however, since a level of fit can be established for a structure which, once achieved by modelling, would establish that structure as the starting point of a new series. The criteria might be based on certain minimum numbers of covalent, ionic, hydrogen-bond, and hydrophobic interactions between compound and protein, or might be based on a calculated binding enthalpy. Potential structures can be quickly eliminated if, for example, during the attempted fit, two groups of like charge or any two atoms are obligatorily so close, that a large repulsive term would be generated. Those not so eliminated might be ranked on, say, estimated binding enthalpy or estimated ease of synthesis. Thus receptor fit differs from classical P.A.R. in that it is well able to generate starting points for optimization in new regions of property space.

F. Limitations.—A major limitation of receptor fit at the present time is its scope and lead time. Of the dozens of macromolecules that have now been structured, few have been seen as having potential relevance in therapy. To embark on the *'de novo'* structuring of a macromolecule demands faith that it is a relevant and important molecule for study. Facilities for the following tasks would be essential:

preparation of substantial quantities of the pure material; determination of residue sequence by direct analysis of protein or nucleic acid; preparation of crystals; structural determination, in general by the preparation of isomorphously prepared crystals containing specifically located heavy atoms. The biochemical procedures for the isolation of protein present in small amounts in organs and tissues have benefitted from selection procedures that induce larger levels of material in cell lines or micro-organisms, and affinity chromatography is now of immense value in the fractionation process. Amplified production by genetic engineering is likely to have substantial impact upon the production of macromolecules by the end of this century. In a few instances, there may be time-saving alternative procedures. In favourable cases the amino-acid sequence can be largely determined from the electron density map determined crystallographically, although there will be numerous instances in which residues will be assigned ambiguously—valine and threonine, for example, are isosteric, and could also resemble aspartic acid and asparagine if there were disorder. Conversely, if the amino-acid sequence of target protein and the crystal structure of a homologous protein are known, a model of the target protein may be built by intelligent superposition upon the known structure.

Another limitation of receptor fit is that it directs attention intrinsically to one facet of the overall process, but it is well known that a compound effective at a receptor *in vitro* may be less effective *in vivo*. The compound may be unfavourable in respect of absorption, distribution, metabolism, or elimination in the more complex bioassays involving cells, micro-organisms, organ systems, and animals. Hence there must in general be a complementary effort, involving P.A.R. at the various levels of biochemical and biological assay, which identifies properties in the compound that come to dominate in the less purified systems.

5 Concluding Remarks

Targeted approaches to drug generation have been considered conceptually for many years. Ehrlich, in his search for synthetic drugs early in the present century, came to adopt the belief that specific receptors must exist. His belief was in part founded on Langley's research on alkaloids.[90] Advances in our knowledge of function, structure, and mechanism for biological macromolecules arising in part from the advent of new techniques in biochemistry and biophysics have progressively made targeted approaches more realistic. However it is clear that in the past the untargeted approach has been the main influence in the generation of new drugs.[91,92] The synthesis of diverse analogues of natural biochemicals, and the study of these in diverse biological assays, have been the source of many of today's drugs. Productive research has often depended upon skilled choice and use of biological assays which allow the *significant* unexpected event to be spotted, assessed, and exploited, in a manner reminiscent of that associated with the

[90] J. Parascandola, *Trends Pharmacol. Sci.*, 1980, **1**, 189.
[91] T. A. Krenitsky and G. B. Elion, 'Strategy in Drug Research', ed. J. A. K. Buisman, Elsevier Scientific Publishing Co., Amsterdam, 1982, 65.
[92] R. Maxwell in 'Drug Development Research', Alan R. Liss, Inc., New York, in press.

identification of penicillin by Fleming, Florey, and Chain. The tailoring of drugs to cellular targets can of course be approached by correlative methods such as classical P.A.R., but it is now possible to augment this by the receptor-fit method, which provides P.A.R. information at the molecular level through knowledge of the physical chemistry of molecular processes.

With such an historical background, conversion to new and relatively untried technology, such as that described in the present review (which deals with only certain aspects of the overall problem) involves commercial risk. However, to ignore new technologies is to give up the hope for a more predictable approach. In practice, new methods are explored tentatively to see what they have to offer and only much later can the main attractions and defects of a new approach to the generation of drugs be assessed. For the macromolecular-fitting approach, it would seem premature to attempt to make such an assessment, as the methodology is little developed. The structural information which provides its foundations must be interpreted with care.[93] Even so, it is clear that in some studies this approach has already been of use. Such successes encourage those in the field to continue to develop and apply the technique and with the advanced computer graphics and energy computation methods now under development, rapid progress can be expected.

Acknowledgements. I appreciate the interest shown by Professor C. R. Ganellin and by the editorial board which led to this review. I am indebted to Drs. J. N. Champness, L. Kuyper, B. Roth, D. K. Stammers, and R. Wootton for their helpful criticism of the manuscript and to Professors P. R. Andrews, D. M. Blow, F. R. S., and G. R. Marshall, and Drs. J. Blaney, P. J. Goodford, D. W. Henry, R. Maxwell, S. Neidle, S. J. Oatley, G. C. K. Roberts, and L. Sawyer for their advice and information.

[93] P. J. Goodford, *J. Med. Chem.*, 1984, **27**, 557.