Steroids, steroid-binding proteins and hydrophobic binding sites

D. M. Blow, Alice Vrielink* and Peter Brick

Blackett Laboratory, Imperial College of Science, Technology and Medicine, London SW7 2BZ, UK
*Present address: McIntyre Medical Sciences Building, McGill University, 3655 Drummond St, Montréal H3G 1YG, Québec, Canada

Dorothy Crowfoot's first research assignment was to work on steroid structure, a controversial subject when she started her research. She completed this assignment in 1945 with the first three-dimensional structure of a steroid molecule, cholesteryl iodide. More recently, the structures of proteins which bind steroids have been determined. Some of them have a large closed cavity, big enough for the steroid nucleus, filled with up to 13 molecules of ordered water which must be displaced when a steroid is bound. Even more extensive arrays of ordered water molecules are observed in the binding cavity of some fatty acid-binding proteins. It is suggested that the function of these arrays of ordered water is to enhance the free energy released when a hydrophobic substrate is bound to them. This creates a favourable equilibrium which allows the enzymes to draw in these substrate molecules from their existing hydrophobic environment in micelles or membranes.

In October 1928, Dorothy Crowfoot went up to Somerville College, Oxford to study Chemistry (Figure 1). During her first term in Oxford, two Nobel Prizes for Chemistry were announced, which were to have a close relevance to her subsequent career. No Nobel Prize for Chemistry had been announced in 1927, and in 1928 a prize was awarded for the previous year in addition to the 1928 prize.

The 1927 Nobel Prize for Chemistry went to Heinrich Wieland for 'extracting from bile the saturated acid that is the underlying substance of all bile acids and accurately ascertaining its structure', and the 1928 Prize to Adolf Windaus for his studies on cholesterol (from which he derived cholanic acid) and on ergosterol obtained from milk, which he showed to assume the properties of vitamin D after ultraviolet irradiation.

In Wieland's Nobel lecture, whose readership was extended by republication in *Angewandte Chemie*[1], he summarized the current state of his work and presented a structural formula (Figure 2) which was also in full agreement with Windaus' results. Perhaps he was challenged by the wording of the citation. In Fieser and Fieser's classic book on *Steroids*[2], they write that he 'summarized the conclusions reached with a tentative formula, even though this was incomplete and uncertain. Republication of his Nobel lecture thus gave undue prominence to a curious formula... that represented merely the working hypothesis of the time'. Although research soon showed it could not be totally correct, this structure became the accepted formula for the steroid nucleus. The central carbon atom participates in three rings, which would give the molecule a rather globular shape.

In 1932, Dorothy Crowfoot, having completed her undergraduate course, moved to Cambridge to begin research under Desmond Bernal. He had been crystallizing several steroids, and earlier that year he triggered a serious controversy when he published the unit cell dimensions of his crystals[3]. He pointed out that these dimensions, especially a 7 Å dimension found in many crystal forms, were 'difficult to reconcile with the usually accepted sterol formula'. Bernal was in touch with the chemists Rosenheim and King at the National Institute for Medical Research in London, who published

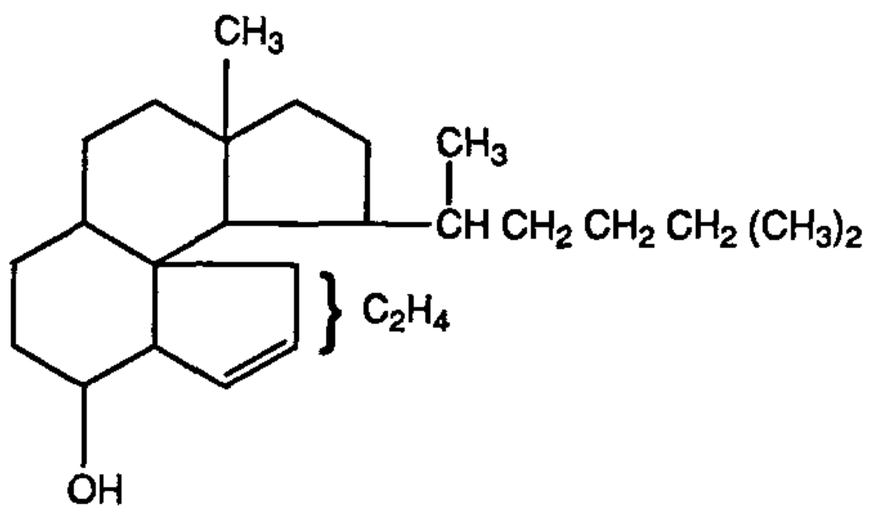
Figure 1. Dorothy Crowfoot aged about 20.

new studies suggesting 'that the ring system of the sterols . . . is that of chrysene, and consists of four six-membered rings'[4], an interpretation which Bernal recognized 'to explain most of the general properties of the crystals whose structures have been studied'[5]. Less than three months later, Wieland's assistant Butenandt, who was to win a Nobel prize himself in 1939 for his work on steroid sex hormones, published a similar suggestion[6]. He proposed 'the existence of four rings' as in 'the new formula recently suggested by Wieland and Windaus' citing 'several papers in press'. He published for the first time the correct formula for the steroid hydrocarbon nucleus without explaining the reason for making ring D five-membered. He infuriated Rosenheim and King by not citing their work, and in almost the next issue of *Nature* they published a stinging letter[7] which pointed out that it was 'not necessary to refer, for "similarities" to the proposed ring system . . . to papers . . . which are still "in the press"', since their own publication already presented the proposals. They also stated that their own observation of chrysene formation 'can obviously take place whether the attached ring [D] is six- or five-membered' even though they had specifically suggested four six-membered rings three months earlier. Wieland and Dane's paper[8], which had been submitted before this acerbic exchange had taken place, was published the following month. It presented experiments which gave clear evidence that ring D is five-membered, but was generous in acknowledging the contribution of Rosenheim and King's experiments, which suggested the chrysene-like arrangement of the rings. This evidently calmed the storm, and two months later Rosenheim and King published their own paper[9], reaching the same conclusion, and giving full recognition to the prior work of Wieland and Dane.

The rumbles of this storm extended widely, and Dorothy must have been very aware of them. There is a kind of archaeological evidence of the depth of the controversy in text books written decades later. Fieser and Fieser[2] summarized it by a series of dated molecular structure diagrams. Witzmann[10] presents a series of almost identical dated diagrams, which have one tiny difference, giving priority to the German group rather than the British, and in fact this may well be justified. But it seems plausible that Dorothy Crowfoot's early experience of scientific controversy over priority, and bickering over who said exactly what and when, may have been an important factor in her wonderful record of constructive international collaboration and generosity towards competitors.

Three years later Bernal and Crowfoot[11] published crystallographic evidence that the proposed steroid structure was correct. They showed that a substance known as Diel's hydrocarbon, derived by reducing steroids, was identical to a synthetic hydrocarbon, corresponding to the structure now accepted. This work was not easy, since the presence of impurities in their samples had led to crystals of different shapes and different melting points, but the unit cells were identical, and the crystal properties became more similar after further purification. It is interesting that there is no mention whether the intensities of the different orders of diffraction were the same, but we are still many years away from the possibility of 'solving' a molecular structure of this complexity by X-ray diffraction, and all intensity observations would have been made on a Bernal oscillation camera which is difficult to use in a reproducible way.

(a)



(b)

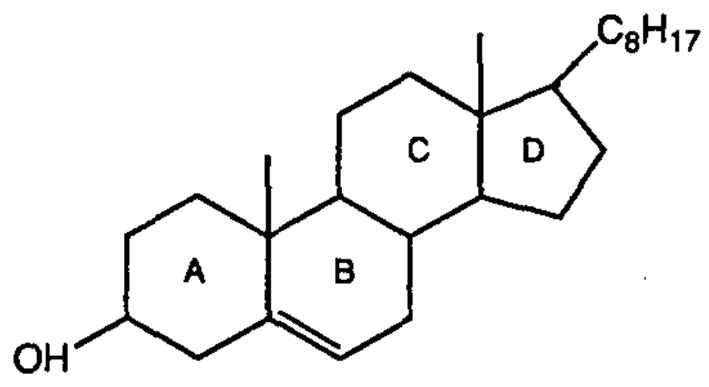


Figure 2. *a*, The structural formula presented by Wieland in his Nobel lecture for the hydrocarbon skeleton of cholanic acid and ergosterol; *b*, The accepted formula of cholesterol.

Detailed structural work on steroids

Dorothy soon undertook ambitious work towards the crystal structure of a steroid, at a time when other known crystal structures were far less complicated. Even those who understand what was involved in calculating a Fourier transform before digital computer programs were available, probably tend to forget just how laborious it was. With Bernal and Fankuchen[12], she showed that the bromine and iodine derivatives of cholesterol are isomorphous, and a two-dimensional Patterson function of the iodide showed a clear iodine–iodine vector. Today one might think of isomorphous replacement but that was not necessary. Carlisle and Crowfoot[13] published their famous paper on the structure of cholesteryl iodide, using what we now call the heavy atom method of phasing. As we read it today, the paper is written in

a matter-of-fact style, and gives little indication how far it extended outside the limits of current knowledge. The three-dimensional Fourier transforms used 301 reflections: every point in the electron density map required the summation of 301 trigonometric terms. It does not mention that this is the first application of the heavy-atom method to molecules in a non-centrosymmetric space group, but it does explain how (in the monoclinic system), a single iodine site, forming a centrosymmetric array, generates an electron density map showing the structure and its mirror image simultaneously.

The illustrations to the paper strongly evoke visits to Dorothy's laboratory. There are the vertical wires with cork markers which she always used. In one diagram, the atomic positions of both the true structure and the mirror image are displayed. The paper describes how the molecular structure was separated from the pair of images, using a cardboard triangle, cut to the angle of 108.5°, with the adjacent sides at a length representing 1.54 Å, to identify triplets of carbon atoms belonging to the same image – a simple but powerful tool typical of Dorothy. Carlisle and Crowfoot's structure finally and utterly vindicated the molecular structures which had been defined with such pain in 1932. The double bond in ring B of cholesterol leads to a rather flat conformation for the molecule, which fits easily into the small 7 Å cell dimension which had drawn Bernal's attention to inconsistency with the 1928 proposals.

Steroid-binding proteins

The first structures of steroid-binding proteins were obtained in Mornon's group on the steroid transport protein, uteroglobin[14–16] (Table 1). Insufficient credit has been given to this work, probably because it was instantly clear that the structure had little or nothing to say about specific steroid recognition. The binding site is obvious, a totally enclosed cavity lying on the two-fold axis of the dimer. Since all steroids are a long way from having two-fold symmetry, it was clear that specific steroid binding must use a different strategy. But, in retrospect, this protein has very interesting features, discussed below.

Regrettably, little can be said about the structure of Δ^5-3-ketosteroid isomerase, because it has only been published at 6 Å resolution[17].

Our work on the structure of a cholesterol oxidase from a *Pseudomonas* species produced a structure with several unexpected features[18]. A strong structural relationship with other flavin-binding enzymes was noticeable in the immediate vicinity of flavin, but the rest of the molecule seemed to have no relation to other known structures. The subsequent publication of structure of glucose oxidase[19] revealed a molecule with a close structural relationship to cholesterol oxidase, even though the only detectable sequence homology is in the flavin-binding domain. Sequence comparisons had identified a family of flavoenzymes including the glucose, methanol and choline oxidoreductases (GMC oxidoreductases)[20]. The structural similarities with glucose oxidase thus revealed cholesterol oxidase as a member of the GMC oxidoreductase family[21].

Cholesterol oxidase contains a large closed cavity, suitably shaped to hold the cholesterol ring-system. The cavity surface is largely non-polar, but adjacent to the flavin at one end of the cavity, five polar residues (glutamate, histidine and three aspargines) form a highly polar cap to it, which obviously defines the position of cholesterol's one hydroxyl group (at C3 in ring A).

Table 1. Crystal structures of steroid-binding proteins

Protein	Ligand	Resolution (Å)	Reference
Uteroglobin $C222_1$	–	1.3	14, 15
$P2_1$	–	1.6	16
Δ^5-Ketosteroid isomerase	–	6	17
Cholesterol oxidase	–	1.8	18
	Dehydroisoandrosterone	1.8	21
Fab' DB3 antibody	–	2.7	22
	Progesterone	2.7	22
	Five other steroids	2.7	23
3α,20β Hydroxysteroid dehydrogenase	–	2.6	24
	Carbenoxalone	2.6	24
3α Hydroxysteroid dehydrogenase	–	3.0	25
17β Hydroxysteroid dehydrogenase	–	2.2	26
Cholesterol esterase	–	1.9	27
	Cholesteryl linoleate	1.9	27

These residues are surrounded by highly-ordered water molecules (Figure 3 *a*), which form part of a network of other ordered water molecules, 13 in all, which fill the cavity.

Access to the cavity seems to be provided by three mobile loops which have relatively high temperature factors. These may swing aside when the enzyme docks to a cholesterol-containing membrane, to provide an entrance to the active site with a highly hydrophobic surface, allowing a cholesterol molecule to displace the water from the cavity. This hypothesis was strengthened when the structure of cholesterol oxidase : steroid complex was available. One of the three loops had moved to a slightly different position, and represented the largest movement of the peptide chain on steroid binding[21].

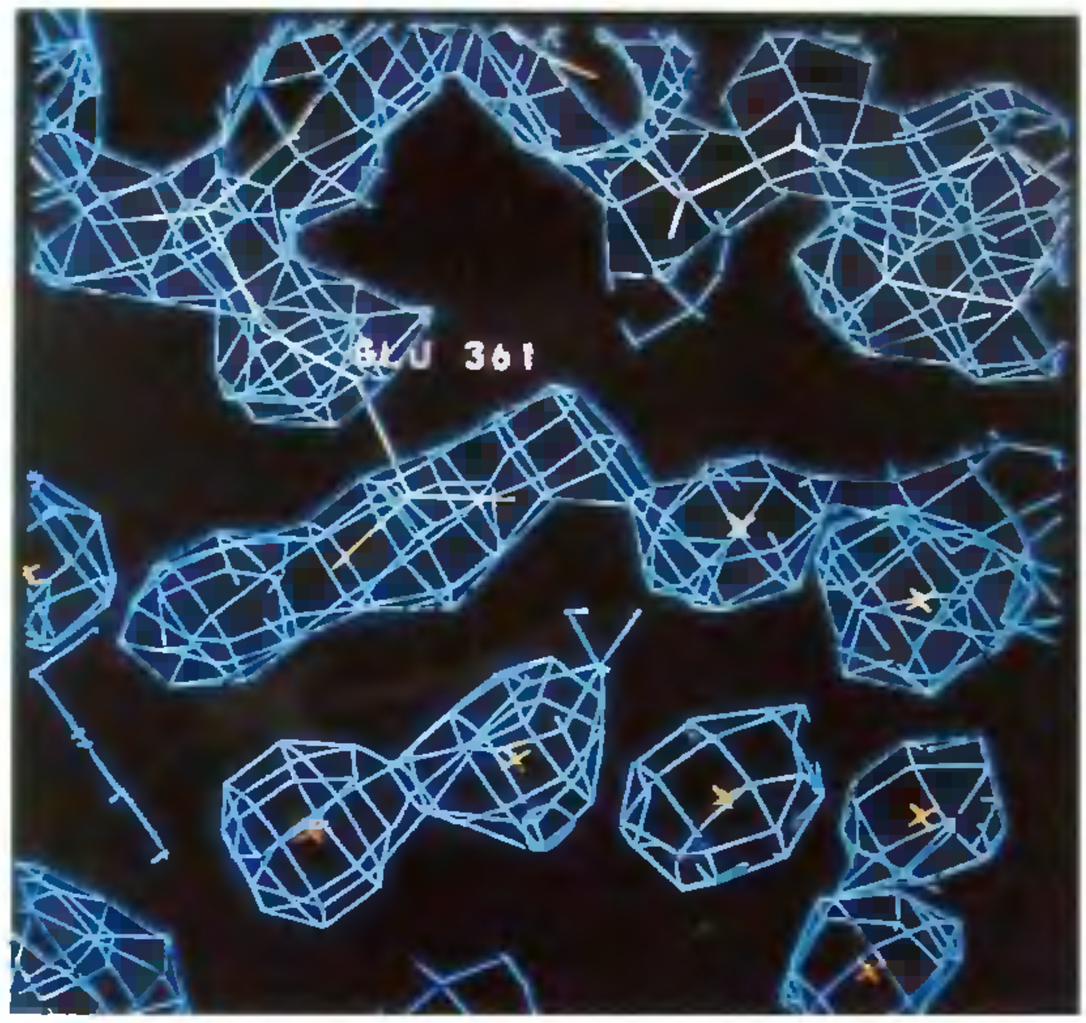
Crystallization of an enzyme–substrate complex raised extreme difficulty, because cholesterol is highly insoluble in water. Another difficulty was that the long aliphatic 'tail' of cholesterol could not be accommodated in the active site, but must hang outside the enzyme, in a way probably not compatible with the crystal packing. In order to achieve a satisfactory level of steroid binding, three effects were exploited. (i) The steroids chosen for study, dehydrisoandrosterone and pregnenalone, have a nucleus identical to that of cholesterol, but lack the long 'tail'. (ii) Detergent micelles were used to solubilize the steroid, and alcohol was introduced into the supernatant, enhancing the steroid's solubility and thus improving the rate of transport from micelles to enzyme sites. (iii) The enzyme was disabled from participating in a full cycle of its action. This was achieved by removal of oxygen from the supernatant, preventing the oxidation of FADH to FAD which normally follows the oxidation of cholesterol. Enzyme–substrate complex appears to have been trapped by this strategy.

The steroids were bound with partial occupancy in the crystals, sufficient to show clearly recognizable electron density for the steroid (Figure 3 *b*). The protein accommodates the cholesterol nucleus with only small movements of a few adjacent side chains, largest in the loop already referred to. All except one of the 13 ordered water molecules is displaced. This last water molecule is at the end of the cavity close to the active site. It is hydrogen-bonded to O3 of the steroid and to two of the residues (His-447 and Asn-485), which form the polar cap adjacent to FAD.

An array of ordered waters in a pocket had already been reported in the two uteroglobin structures[15,16]. The closed pocket lies at the centre of the dimeric molecule, and is similar in size to that observed in cholesterol oxidase. In the two uteroglobin structures, 14 and 6 ordered water molecules respectively can be identified. In $C222_1$ uteroglobin, the 14 water molecules make only four polar interactions with the protein, with threonine and tyrosine side-chains.

The structures of several other steroid-binding proteins are known (Table 1). Some of these are antibodies, whose binding sites bear little resemblance to the struc-

a



b

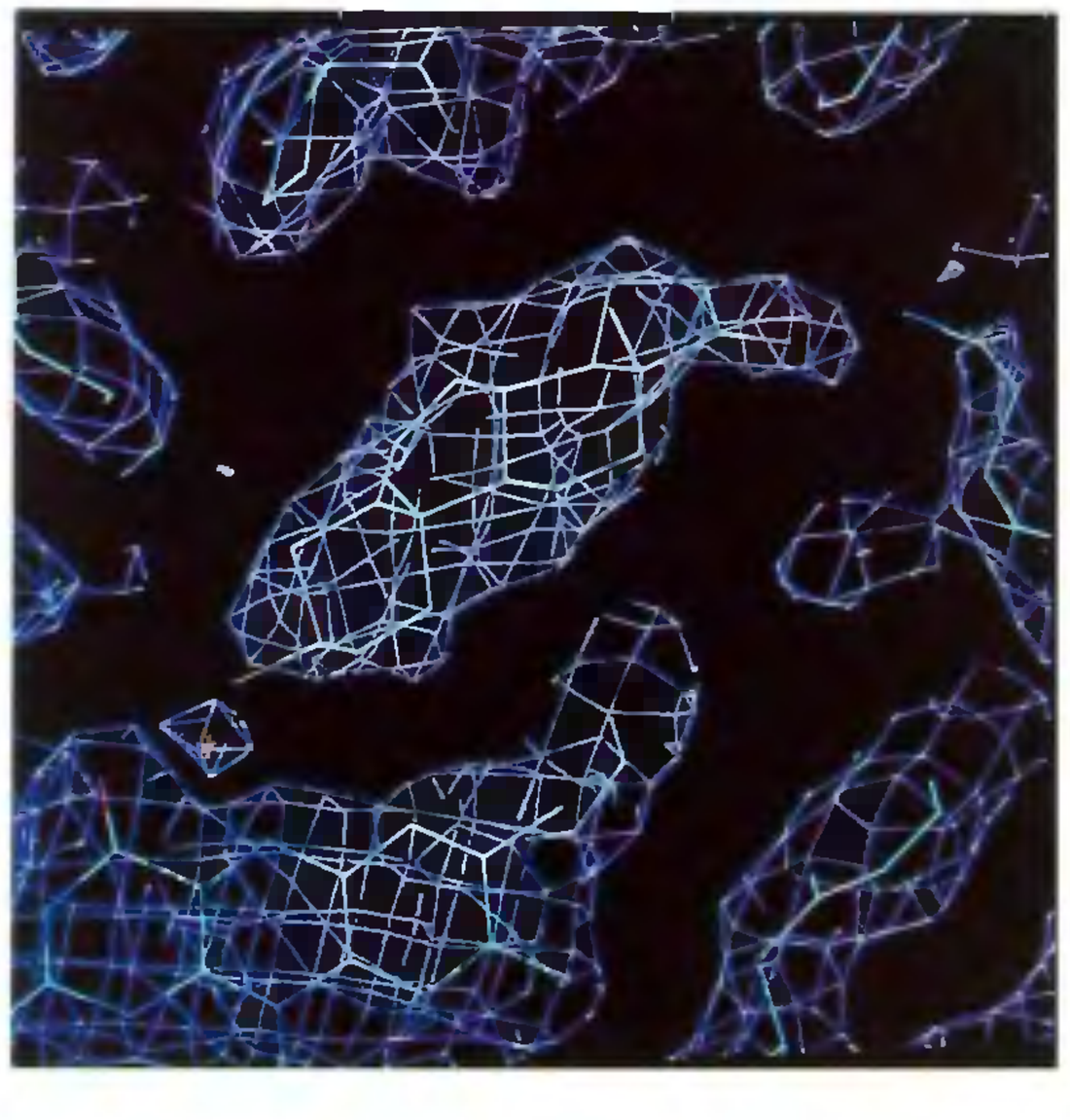


Figure 3. *a*, Part of the electron density in the steroid-binding pocket of cholesterol oxidase, showing 8 of the 13 water molecules. Four water molecules (one at the left-hand edge of this picture) are tightly packed around Glu-361. *b*, Electron density map of the complex with dehydroisoandrosterone, showing the steroid substrate bound in the cavity. The 3-hydroxyl group of the steroid makes a tight polar interaction with Glu-361. It is also tightly bound to the one remaining water molecule in the binding cavity, which appears at the bottom left, adjacent to the flavin. This water is also hydrogen-bonded to His-447 and Asn-485, and possibly to N^5 of the flavin.

tures already discussed[22,23]. In 3α, 20β-hydroxysteroid dehydrogenase, the steroid-binding site is 'deep cleft filled with solvent density' adjacent to a Rossmann fold[24]. Two other hydroxysteroid dehydrogenases have a cavity 11 Å deep at the end of an α/β barrel[25,26]. The latest is cholesterol esterase, a molecule homologous with the lipases, having a spherical hydrophobic cavity which contains some water molecules[27]. None of these publications indicates an extensive network of highly ordered water molecules, though in the structures with resolution worse than about 2.7 Å such a network might be hard to observe.

Other big cavities in proteins which bind hydrophobic substrates

Another class of proteins with large cavities, which bind substrates with predominantly hydrophobic surfaces, is the family which includes retinol-binding proteins, fatty acid-binding proteins and lipid-binding proteins. In at least two of these, both fatty acid-binding proteins, extensive arrays of ordered water have been observed[28-31], especially when substrate is absent (Table 2). So uteroglobin and cholesterol oxidase are not unique in this respect, and the ordered water may have more general significance as a mechanism to increase the affinity of a protein molecules for its hydrophobic target molecules[32].

A wide range of disagreement surrounds the subject of so-called hydrophobic forces[33]. The key problem raised by steroid- and fatty acid-binding proteins is 'How can a soluble protein withdraw a hydrophobic substrate from an environment which already provides a satisfactory hydrophobic-binding surface?'.

It might be thought that the free energy of binding substrate to cholesterol oxidase could easily be found from K_M measurements. This is not so, because satisfactory quantities of cholesterol for assay require its presence under conditions where micelles are formed.

In cholesterol oxidase, the array of ordered waters is maintained by seven interactions with polar groups which are not part of the cavity's hydrophilic cap, which has already been described. They involve a side-chain atom of glutamine and of tyrosine, and four main-chain atoms. When cholesterol is bound, seven hydrogen bonds to

Table 2. Large protein cavities containing water

Protein	Species	Ligand	Bound water molecules	Reference
Steroid-binding proteins				
Uteroglobin	Rabbit C222	–	14	15
	Rabbit P2	–	6	16
Cholesterol oxidase	*Pseudomonas*	–	13	18
		Dehydroisoandrosterone	1	21
Fatty-acid binding proteins (FABP)				
Intestinal FABP	Rat	–	24	28
		14-oate	8	29
		16- and 18-oate	Up to 8	30
Muscle FABP	Human	18-oate	13	31

Table 3. Free energy released on binding steroid

Effect	Assumed unit value	Total energy release
1. Buried ~ 500 Å^2 hydrophobic area *But this is similar to the energy required to remove steroid from a hydrophobic environment in a micelle or a membrane, so in many cases it should be ignored.*	+ 100 J/mol Å^2	+ 50 kJ/mol steroid
2. Lost 7 hydrogen bonds *But several of these hydrogen bonds may be relatively weak because the polar groups on the protein are also making other hydrogen bonds.*	– 6 kJ/mol H-bond	Up to – 42 kJ/mol steroid
3. Disordered 12 water molecules *Depending on how tightly ordered.*	Up to + 10 kJ/mol water	Up to + 120 kJ/mol steroid
4. Ordered one cholesterol molecule *But the cholesterol may also be ordered in the hydrophobic medium from which it is transferred.*	Up to – 180 J/K-mol	Up to – 30 kJ/mol steroid

water are lost. On the other hand, twelve ordered waters become disordered, and a hydrophobic surface of several hundred Å^2 is buried. This disordering of water releases a free energy which may be as large as the latent heat of melting ice, depending on the original degree of order. Using very rough figures, an energy balance sheet can be drawn up (Table 3).

The estimates shown in Table 3 are very rough, and each of them is a high estimate which probably needs adjustment for the reasons given. The energy change on burying hydrophobic area was obtained by Hermann[34] from data on partition coefficients between hydrophobic and aqueous solvents, and corresponds well to the figure established independently by Chothia[35]. The change of free energy on disordering water is obtained from the latent heat of fusion of ice. There is a school of thought which considers that the free energy of burying a hydrophobic surface arises mostly from the water disordering effect. Some may say that items 1 and 3 are different aspects of the same effect, while others consider that there are two separate effects. This emphasizes the vagueness and uncertainty which surrounds the figures.

Table 3 shows two important points. (i) There is ample free energy available to create very tight substrate binding. (ii) The largest term arises from the disordering of water. Even if it is halved to allow for water being partly disordered in the site, it remains the largest term, and it outweighs a high estimate for the compensating energy gain in burying polar groups. It is much larger than in the usual case, because of the highly ordered nature of the water displaced when substrate is bound.

This leads to the following suggestion:

> When a highly insoluble molecule such as a steroid binds to a soluble protein, a hydrophobic site in the protein can do little more than balance the energy gain in detaching the steroid from its normal hydrophobic environment. A further loss of energy can be achieved if ordered water is removed from the protein site and transferred to bulk water. Provided the water is sufficiently ordered, this loss is well able to compensate for the abolition of the few hydrogen bonds which are needed to maintain ordered water in the protein site.

Sadly, we were never able to discuss this work in detail with Dorothy Hodgkin. Not only would she have been excited to see cavities shaped to enclose the cholesterol structure which she had discovered. She would undoubtedly have had imaginative and insightful suggestions for further advancement of this work.

1. Wieland, H., *Angew. Chem.*, 1929, **42**, 421–424.
2. Fieser, L. F. and Fieser, M., *Steroids*, Reinhold, New York, 1959.
3. Bernal, J. D., *Nature*, 1932, **129**, 277.
4. Rosenheim, O. and King, H., *Chem. Ind.*, 1932, **51**, 464–466.
5. Bernal, J. D., *Chem. Ind.*, 1932, **51**, 466.
6. Butenandt, A., *Nature*, 1932, **130**, 238.
7. Rosenheim, O. and King, H., *Nature*, 1932, **130**, 315.
8. Wieland, H. and Dane, E., *Hoppe-S. Z. Physiol.*, 1932, **210**, 268–281.
9. Rosenheim, O. and King, H., *Chem. Ind.*, 1932, **51**, 954–956.
10. Wittmann, R. F., *Schlüssel Leben*, 1977, Verlag Fritz Molden (Trans. Peter, R., *Steroids: Keys to Life*, van Nostrand, New York, 1981).
11. Bernal, J. D. and Crowfoot, D. M., *J. Chem. Soc.*, 1935, 93–100.
12. Bernal, J. D., Crowfoot, D. M. and Fankuchen, I., *Trans. R. Soc.*, 1940, **A239**, 135–182.
13. Carlisle, H. C. and Crowfoot, D. M., *Proc. R. Soc. London*, 1945, **A184**, 64–83.
14. Mornon, J. P., Fridlansky, F., Bally, R. and Milgrom, E., *J. Mol. Biol.*, 1984, **137**, 415–429.
15. Morize, I., Surcouf, E., Vaney, M. C. *et al.*, *J. Mol. Biol.*, 1987, **194**, 725–739.
16. Bally, R. and Delettré, J., *J. Mol. Biol.*, 1989, **206**, 153–170.
17. Westbrook, E. M., Piro, O. E. and Sigler, P. B., *J. Biol. Chem.*, 1984, **259**, 9096–9103.
18. Vrielink, A., Lloyd, L. F. and Blow, D. M., *J. Mol. Biol.*, 1991, **219**, 533–554.
19. Hecht, H. J., Kalisz, H. M., Hendle, J., Schmid, R. D. and Schomburg, D., *J. Mol. Biol.*, 1993, **229**, 153–172.
20. Cavener, D. R., *J. Mol. Biol.*, 1992, **223**, 811–814.
21. Li, J.-Y., Vrielink, A., Brick, P. and Blow, D. M., *Biochemistry*, 1993, **32**, 11507–11515.
22. Arevalo, J. H., Stura, E. A., Taussig, M. J. and Wilson, I. A., *J. Mol. Biol.*, 1993, **231**, 103–118.
23. Arevalo, J. H., Hassig, C. A., Stura, E. A., Sims, M. J., Taussig, M. J. and Wilson, I. A., *J. Mol. Biol*, 1994, **241**, 663–690.
24. Ghosh, D., Weeks, C. M., Grochulski, P., Duax, W. L., Erman, M., Rimsay, R. L. and Orr, J. C., *Proc. Nat. Acad., Sci. USA*, 1991, **88**, 10064–10068.
25. Hoog, S. S., Pawlowski, J. E., Alzari, P. M., Penning, T. M. and Lewis, M., *Proc. Nat. Acad. Sci. USA*, 1994, **91**, 2517–2521.
26. Ghosh, D., Wawrzak, Z., Pletnev, V. *et al.*, *Structure*, 1995, **3**, 279–288.
27. Ghosh, D., Pletnev, V. Z., Zhu, D. W. *et al.*, *Structure*, 1995, **3**, 503–513.
28. Scapin, G., Gordon, J. I. and Sacchettini, J. C., *J. Biol. Chem.*, 1992, **267**, 4253–4269.
29. Eads, J., Sacchettini, J. C., Kromminga, A. and Gordon, J. I., *J. Biol. Chem.*, 1993, **268**, 26375–26385.
30. Sacchettini, J. C., Scapin, G., Gopaul, D. and Gordon, J. I., *J. Biol. Chem.*, 1992, **267**, 23534–23545.
31. Young, A. C. M., Scapin, G., Kromminga, A., Patel, S. B., Veerkamp, J. H. and Sacchettini, J. C., *Structure*, 1994, **2**, 523–534.
32. Scapin, G., Young, A. C., Kromminga, A., Veerkamp, J. H., Gordon, J. I. and Sacchettini, J. C., *Mol. Cell. Biochem.*, 1993, **123**, 3–13.
33. Tanford, C., *The Hydrophobic Effect: Formation of Micelles and Biological Membranes*, Wiley Interscience, New York, 1977.
34. Hermann, R. B., *J. Phys. Chem.*, 1972, **276**, 2754–2759.
35. Chothia, C., *Nature*, 1974, **248**, 338–339.

ACKNOWLEDGEMENTS. We thank Dr Luke Hodgkin for allowing us to reproduce Figure 1. We are grateful to Prof. B. W. Matthews for bringing our attention to the large cavities in the fatty acid-binding proteins, and to Prof. N. P. Franks for his comments on hydrophobic binding.