ELSEVIER

Pharmaceutica Acta Helvetiae 69 (1995) 185–192

PHARMACEUTICA ACTA HELVETIAE

Protein three-dimensional structure and molecular recognition: a story of soft locks and keys

R. Sowdhamini, N. Srinivasan, K. Guruprasad, S. Rufino, V. Dhanaraj, S.P. Wood, J. Emsley, H.E. White, Tom Blundell *

ICRF Unit of Structural Molecular Biology, Department of Crystallography, Birkbeck College, Malet Street, London WC1E 7HX, UK

Abstract

One hundred years ago Emil Fischer proposed a descriptive but provocative analogy for molecular recognition: the lock and key hypothesis. At a time when little was known of the molecular structures of even the relatively simple substrates of enzymes, let alone the complex structures of proteins, this gave an extraordinarily useful visual image of enzyme action. Similar recognition processes, such as antigen–antibody, hormone or growth factor–receptor, lectin–sugar, repressor–DNA and so on, have since been identified in other classes of proteins. Can the Fischer hypothesis be applied to these systems? Has the hypothesis stood the test of time? In this paper, we examine the crystal structures of proteins complexed with their ligand molecules: the pentraxins bound to carbohydrate, several aspartic proteinases complexed with inhibitors, the SH3 domains bound to proline-rich peptide motifs, the periplasmic binding proteins and growth factor systems bound to cell surface receptors. We discuss the modes of binding in terms of surface rigidity, charge and shape complementarity. Such recognition processes are often accompanied by distinct conformational changes at the binding site. The ligand selectivity demonstrated in these systems supports a "soft" lock-and-key hypothesis.

Keywords: Ligand binding; Ligand specificity; Lock-and-key hypothesis; Protein–ligand complexes; Specificity pockets; Surface complementarity

1. Introduction

In the past 35 years the three-dimensional structures of many ligand–protein complexes have been defined and compared to structures of the uncomplexed components both in crystals by X-ray analysis and now increasingly in solution by nuclear magnetic resonance. A few of these complexes are of relatively rigid ligands; but even these generally show some conformational accommodation of the protein during recognition/binding. In many complexes there is evidence of major conformational changes, or even hinge bending in multidomain proteins, so that the ligand is tightly bound to the substrate and often removed entirely from its aqueous environment. Nevertheless, even when the protein and ligand are flexible, the complex most often involves a well-defined three-dimensional complementarity of the bound components; this is not dissimilar to the Fischer "lock and key" concept.

In this paper we describe some structures of complexes of ligands with proteins defined by X-ray analysis in our own laboratory. We examine the nature of the complementarity of ligand and protein in terms of shape, charge and hydrogen bonding. By such analyses we may hope to advance our understanding of molecular recognition and even begin to predict conformational changes that might occur during the process of recognition.

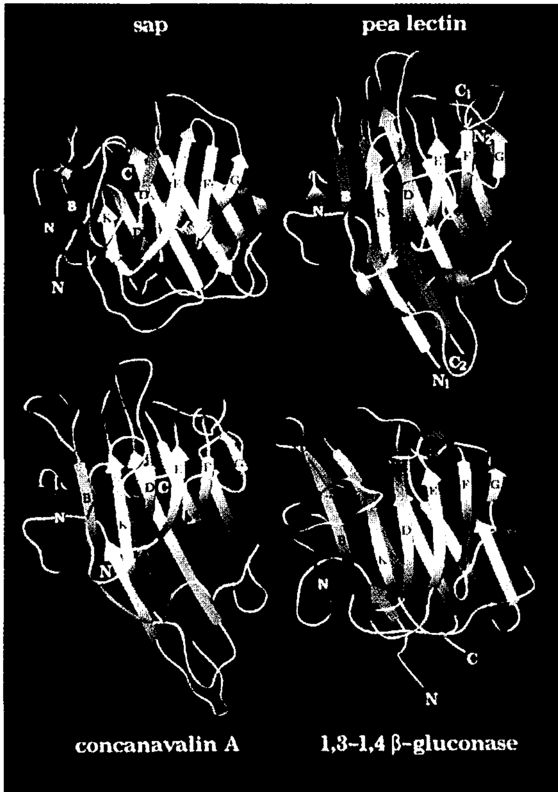
2. Recognition and binding

Fischer's observations were on enzymes, but the simplest systems, and perhaps those closest to the

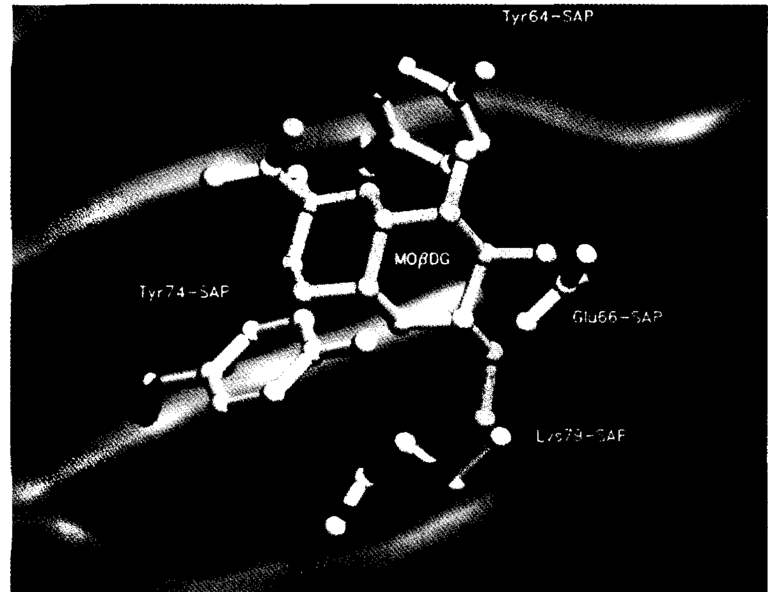
* Corresponding author. Tel.: +44-171-631-6284. Fax: +44-171-631-6805.

0031-6865/95/$09.50 © 1995 Elsevier Science B.V. All rights reserved
SSDI 0031-6865(95)00002-X

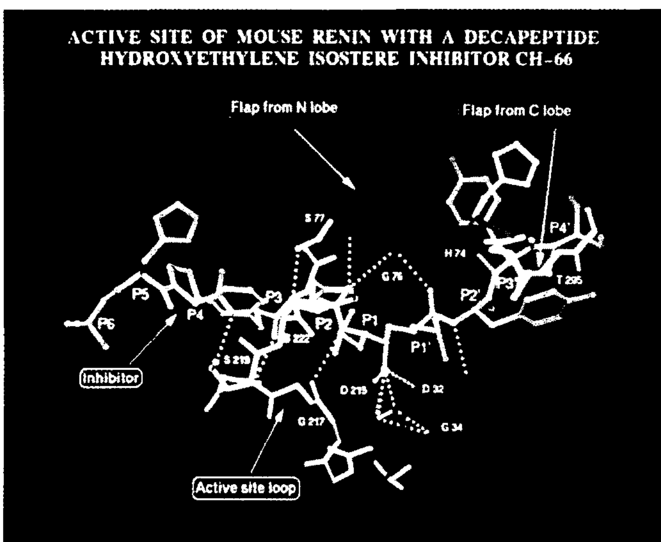
"lock and key" model, are binding modules or domains that are used for recognition in multimeric assemblies that mediate cell–cell interactions, intracellular signal transduction and so on. In these systems modules act as adaptors that bring together enzymes, gated-channels, and other active components. Examples are the



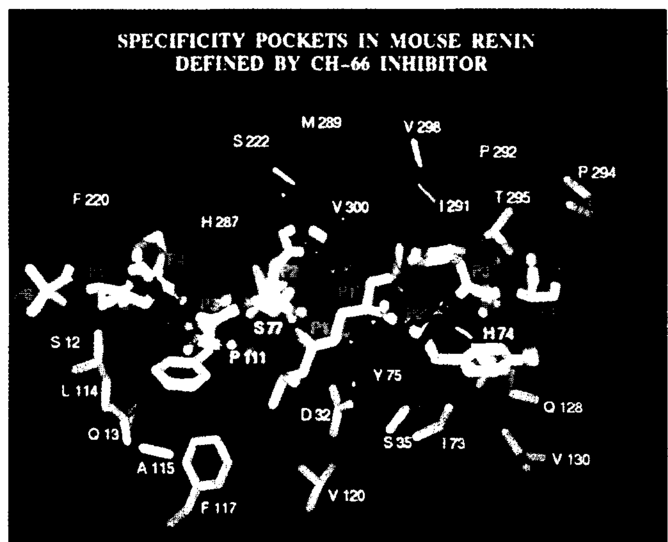
1



2



3a



3c

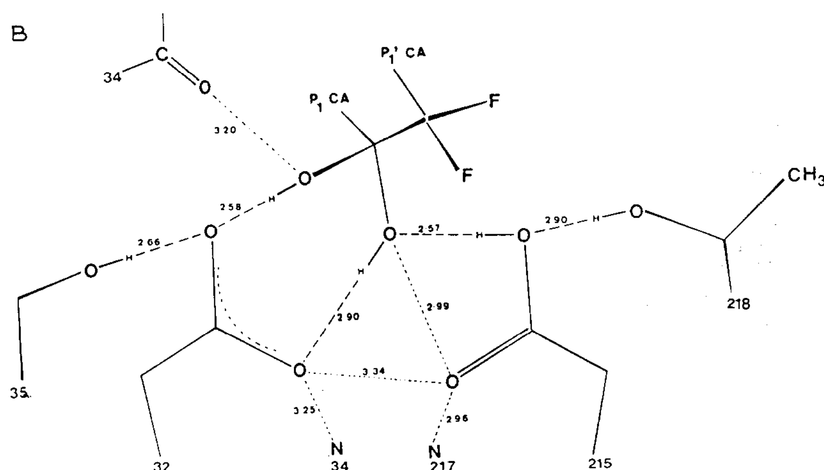


Fig. 3 (continued).

multi-subunit lectins which link cells through carbohydrate recognition and intracellular adaptor proteins such as Grb2, which involve SH2 and SH3 domains, linking phosphorylated peptides of one component, such as a receptor tyrosyl kinase, with proline-rich components further along the pathway of signal transduction.

Let us first consider lectins. Most surprisingly it has recently been found that, although there are no significant sequence similarities, legume lectins such as concanavalin and mammalian lectins such as S-lectins and the pentraxins belong to a common superfamily, i.e. they share a common fold (Lobsanov et al., 1993; Emsley et al., 1994; Liao et al., 1994). Fig. 1 compares the structures of several lectins, viewed from the same orientation. Each comprises an elaborated β-jelly roll fold; in legume lectins and pentraxins identical hydrogen bonding patterns interconnect identical numbers of antiparallel β-strands, although the loops are much longer in the legume lectins. The resultant β-structure is a sandwich, in which one β-sheet is relatively flat and the other strongly concave. The sugar binding in all cases is in the concavity; the β-jelly roll structure makes a relatively rigid framework for binding the carbohydrate ligand.

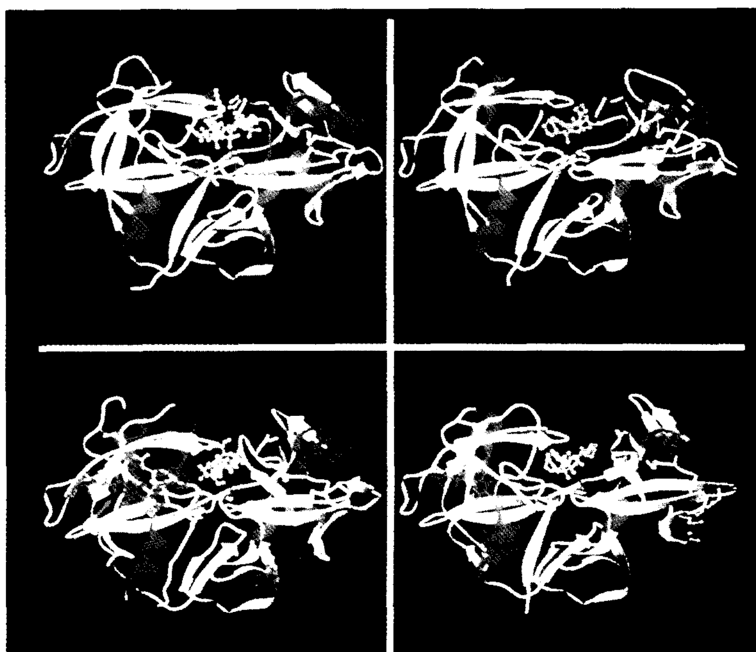
In the legume lectins and the pentraxins the sugar binding is mediated through metal ions, in a way similar to that of C-lectins. In the case of pentraxins two calciums are bound mainly by aspartates and glutamates, some of which are bidentate and bridge between them. Two amide functions link between the metal ion and the carbohydrate hydroxyls; the metal ions have the function of fixing the orientation of the amide groups which are otherwise very flexible. This

Fig. 1. Comparison of the structures of several lectins, viewed from the same orientation (sap: serum amyloid P component). Strands are shown in pink and helices in blue. Despite the remarkable similarity in overall fold, there is very little sequence identity; the positions of N- and C-termini in the proteins are different due to circular permutation.

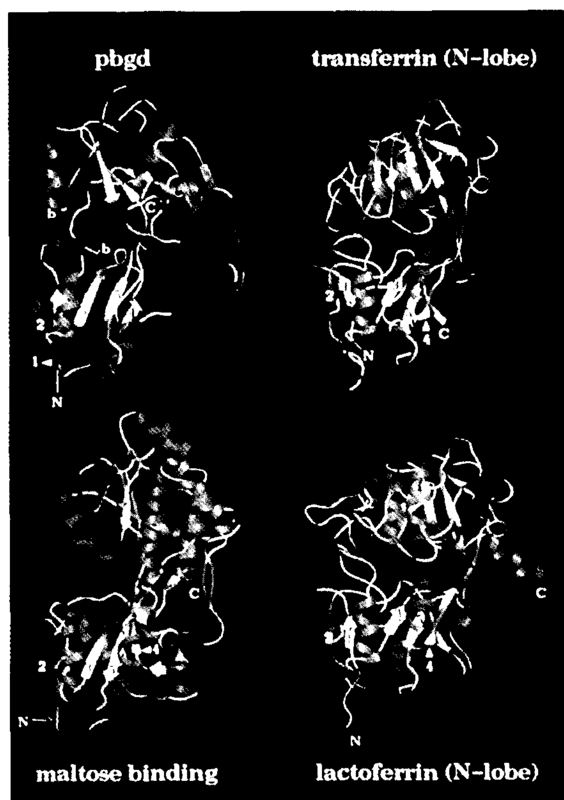
Fig. 2. Mode of protein-ligand interaction illustrated using the crystal structure of the complex of SAP (serum amyloid component) and MoßDG. Atoms of the ligand molecule are shown in yellow. Only selected residues of the protein are shown in grey. Sidechains of equivalent residues in a related protein, CRP (C-reactive protein) are shown in blue. Calcium ions are shown in red.

Fig. 3. (a) The active site of mouse renin, complexed with a decapeptide inhibitor isostere (CH-66) of the transition state. The hydrogen bonding between the substrate and the enzyme is shown: side-chains of residues of the substrate important in ligand recognition are alone shown; inhibitor molecule is in yellow. (b) Hydrogen bonding between the transition state isostere of yeast-proteinase enzyme complexed with a difluorostatine-containing inhibitor, in which there is a tetrahedral gem-diol derived from hydration of the difluorostatine. (c) The active site of mouse renin complexed with CH-66 inhibitor. The peptide inhibitor is shown in white along with van der Waals surface shown in blue.

relatively rigid recognition centre is augmented in pentraxins, unlike legume lectins and S-lectins, by direct interaction of a negatively charged functional group with the metal ions. This is illustrated for the binding of methyloxy-β-D-galactose in Fig. 2. Interestingly, although legume lectins use the equivalent concave sur-



4a



5

6

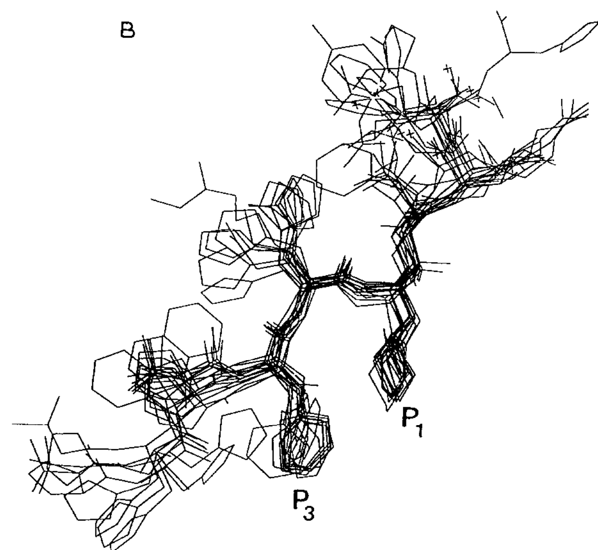


Fig. 4 (continued).

face to bind sugars, the ligands for the manganese and calcium are on different strands of the equivalent topologies. In each case the adaptor function is achieved through oligomerisation: concanavalin A forms tetramers and pentraxins pentamers.

Adaptor molecules require only simple recognition sites; allosteric changes are not needed to convey information to other sites in the molecule. Thus the rigid structure achieved by a β-barrel or sandwich with extensive, inter-strand hydrogen bonds provides an excellent base to form a recognition site. It is interesting that several adaptor domains, involved in signal transduction or cytoskeletal structures, are also largely antiparallel β-proteins. These include SH3 and PH domains, both of which are comprised of antiparallel β-barrels/sandwiches with up-down arrangements or adjacent strands in the barrel that are contiguous in the sequence. This structure also provides a rigid framework. In these cases the molecules are linked covalently rather than as separate subunits in an oligomer.

3. Enzyme recognition of a flexible ligand

For the problem of enzyme recognition of a flexible polymer we take the example of proteolysis, in which the transition state for polypeptide hydrolysis is recognised. A polypeptide is a flexible ligand with a very large number of low energy conformers energetically possible. Of these only those that allow access to both sides of the scissile peptide are useful and this precludes the α-helix as a conformation for substrates in proteolysis.

Most proteinases – serine proteinases, aspartic proteinases and matrix metalloproteinases, for example – involve a distorted β-strand structure involving parallel and/or antiparallel sheet interactions with the enzyme. Fig. 3 shows the active site of the aspartic proteinase, renin, complexed with a decapeptide inhibitor isostere of the transition state (Dhanaraj et al., 1992; Dealwis et al., 1994). Fig. 3a shows the hydrogen bonding between the substrate analogue and the enzyme. Hydrogen bonding with main-chain functions occurs on both sides of the scissile bond. On the N-terminal side (the non-prime side) the main chain H-bonding is replaced by a threonine gamma-hydroxyl to the NH of P_3. These interactions allow a generic mode of binding for a variety of polypeptide sequences, but a very precise orientation of the scissile bond with respect to the catalytic aspartates.

Fig. 3b shows a closer view of the hydrogen bonding between the transition state isostere of a related enzyme complex, in which there is a tetrahedral gem-diol derived form hydration of difluorostatine (Veerapandian et al., 1992). In this case the transition state appears to stabilize the gem-diol rather than an anion as in the serine proteinases. This is probably because many aspartic proteinases are catalytically active at acid pH where it is easier to stabilize the negative charge on a highly hydrogen-bonded carboxylate. In this case the carboxylate of Asp 32 has four hydrogen bond donors, two to each carboxylate oxygen, in a way that is reminiscent of many buried aspartates, that are often conserved in protein structures (Blundell et al.,

Fig. 4. (a) Equivalent views of complexes of aspartic proteinases (human renin, mouse renin, yeast proteinase A and cathepsin D) with inhibitors. Atoms in the inhibitor and disulphide bonds are in yellow; saccharides are in blue. (b) Superposition of 20 different aspartic protease inhibitors complexed with endothiapepsin. The specificity subsites are indicated as P1 and P3.

Fig. 5. The structures of a periplasmic binding protein (maltose-binding protein), porphobilinogen deaminase (pbgd), transferrin and lactoferrin which have a similar fold. Helices are shown in blue and the extended strands are shown in pink and are numbered.

Fig. 6. The interactions between porphobilinogen deaminase and the dipyrrole primer. Atoms in the cofactor are in yellow and side chains of interacting residues of the protein are shown (blue: basic; red: acidic; magenta: hydrophobic).

1988; Overington et al., 1990). This interaction involves "charged hydrogen-bonds".

The specificity is achieved by complementarity between side chains and pockets in the enzyme active site as shown in Fig. 3c. These involve mainly van der Waals interactions, many with aliphatic or aromatic carbons, although occasionally with oxygens or nitrogens of enzyme side chains. The residues vary between members of the homologous family.

The greater specificity of renin, when compared to the more broadly digestive pepsins, is achieved by loops that close over the active site once the substrate is bound. One of these loops, the so called flap, which includes residues 73 to 83, is common to all aspartic proteinases. The flap contains a tyrosine (Tyr 75 or pepsin) that contributes to the specificity pocket 51 and also provides a weak ionic interaction between its aromatic hydrogen (slightly positively charged) and one of the oxygens of the tetrahedral isotostere. In renin, cathepsin D, yeast proteinase and some other aspartic proteinases there is a further flap involving a polyproline loop (residues 293–298), which interacts with the substrate/inhibitor at P_3', and P_4', so providing greater specificity of interactions in this region of the substrate. Although the hydrogen bonds between the enzyme and inhibitor/substrate, particularly those between main-chain functions at P_3, P_1 and P_1', are common to most complexes, small shifts along the active site, giving rise to slightly different positions of $C\alpha$ carbons, are accommodated (Fig. 4a). This is an efficient way of allowing for small differences in side chains and also for binding inhibitors which differ in the number of carbon atoms in the scissile-bond analogue. Identical sidechains of inhibitors can differ radically in their orientations in order to bind opportunistically with favourable groups of the enzyme; for example, note the various orientations of P_2 His in Fig. 4b. Conformational changes in the inhibitor also reflect the fact that pairs of sidechains, such as P_3 and P_1 or P_2 and P_1', of the inhibitor pack together. Thus, when P_1P_1' is replaced by the dipeptide analogue statine, there is then no sidechain at P_1' and the P_2 His moves to fill the pocket S_1'. In a similar way when P_1 is cyclohexylalanyl, rather than leucyl, then P_3 Phe rotates by 90° around the β-gamma bond, to allow better packing with the larger cyclohexyl group. Such subtle differences underline the need for many experimental studies with different ligands, if predictive success is to be achieved in drug discovery.

There are also subtle differences in the relative positions of two "rigid" lobes, comprised of residues 1 to 188 and 304 to 326 for the N-lobe and 189 to 303 for the C-lobe, using pepsin numbering (Sali et al., 1989,1992). The interface between lobes involves helices and sheets that can repack with respect to each other. These differences probably contribute to the differences of K_{cat}/K_m, first observed by Fruton in pepsin, when substrates are extended at P_3, P_4 and P_5. These further complicate a simple "lock and key" model, and emphasise that in drug discovery flexibility needs to be built into receptor models; conversely, the poor precision of X-ray analysis ($\sim$ 0.3 Å) at active sites of enzyme may not be a serious problem in structure-based drug design.

It is interesting to see how divergent protein systems recognise polypeptides. Where the binding is not specifically for peptides, such as in antibodies, sidechain functions provide the principal interactions. Where families of peptides are to be recognised, such as in proteolytic enzymes, histocompatibility antigens or SH3 domains, main chain functions are used to position a fairly extended polypeptide and selectivity is achieved through recognition pockets for sidechains. Note that in enzymes the β-sheet is used where interactions occur on either side of the bound strand. In class II histocompatibility antigens and SH3 domains a polyproline helix conformation is adopted with receptor sidechains, such as asparagine or tryptophan, providing NH donors of hydrogen bonds to peptide CO functions. The more open sites require more prolines in the peptide to define the conformation; specificity is again achieved through sidechain interactions. NMR studies of SH3-peptide complexes point to the presence of a hydrophobic surface composed of several Trp residues (Gout et al., 1993) which undergo changes in chemical shifts upon peptide binding. Recent crystal structures of two SH3-peptide complexes (Musacchio et al., 1994) reveal that apart from the hydrophobic interactions, the presence of hydrogen bonds stabilises the complex formation. The crystal structures derived using 10-residue peptides still do not explain the specificities observed in SH3-peptide interactions. The presence of charged loops may serve to form additional pockets useful in peptide recognition. In some cases of protein peptide recognition, but probably more rarely, binding is achieved by inducing a helical conformation in a hydrophobic environment and recognition is through sidechains; an example is the binding of peptides by calmodulin (Meador et al., 1993).

4. Recognition and binding through hinge-bending

When a ligand needs to be entirely surrounded by the receptor protein, then conformational changes are

often achieved by hinge-bending whereby the ligand first binds to a site on one domain in the open form and a second domain closes on top (Wodak et al., 1987). This is common in enzymes like dehydrogenases, in which one domain may recognise a cofactor and the second a substrate; in the closed form the two are brought together in an environment which is free from water and suitable for a redox reaction. Simpler binding proteins may also operate through a hinge bending mechanism so that all functions of the ligand are recognised and the binding specific; the periplasmic binding proteins belong to this class. Thus, in sulphate or phosphate-binding proteins the anions are at the positive end of helices and have multiple hydrogen bonds that delocalise the charge and provide tight binding.

Interestingly a fold common to the class II periplasmic binding proteins is found in the polymerase, porphobilinogen deaminase. This enzyme has a primer dipyrrole that is covalently attached via a cysteine to a third domain, not found in the periplasmic binding proteins. The enzyme assembles the primer and then four further pyrrole rings by successively deaminating porphobilinogen molecules. A linear tetrapyrrole is then cleaved off and becomes a haem, chlorophyll or B_{12} precursor.

Fig. 5 compares the structures of a periplasmic binding protein, porphobilinogen deaminase, and transferrin, along with other anion binding proteins, all of which have a similar fold. Each protein has two similar doubly-wound $\alpha\beta$ domains; the first four β-strands are parallel and the fifth inserts itself antiparallel between strands 3 and 4 of the other domain so that the domains are connected by two antiparallel strands. The studies on transferrins have shown that the two domains can open and close, hinging on these two strands. The ligands/substrates are bound in the cleft between the two domains close to the N-termini of the α-helices. In the case of porphobilinogen deaminase the developing haem precursor is bound via a series of mainly ionic interactions between positively charged arginines, lysines and histidines on the enzyme and the sidechain carboxylates of the substituted polypyrrole. The interactions between the enzyme and the dipyrrole are shown in Fig. 6.

In the case of porphobilinogen deaminase the hinge bending allows access to the substrate at each stage in the polymerisation. It also allows the developing polymer to be accommodated in such a way that the same catalytic apparatus can be used at each step in the reaction.

5. Recognition leading to assembly and signalling

In the 1970s it was often assumed that most cell signalling would be edited by ligand binding and allow steric changes regulating enzyme activity at a distant site, possibly even in another subunit. This paradigm was influenced by the structures of haemoglobin and aspartate transcarbamylase that supported the two-state model of Monod et al. (1965). It is now clear that many signalling systems, for example those activated by growth factors and antibodies, involve the formation of multimeric structures and probably transphorylation of subunits; they involve aggregation rather than just allostery.

The structures of the complexes of growth hormone (De Vos et al., 1992) and TNF (Banner et al., 1993) with the extracellular domains of membrane bound receptors show that the receptors assemble round the ligands, rather than ligands docking into pockets or clefts or the receptors. This presumably brings the transmembrane helices, the intracellular domains and other proteins associated with them into close proximity so that covalent modification of one subunit by another can occur. Similar mechanisms almost certainly occur on hormone or growth factor binding of most kinases.

The recognition between such receptors and ligands appears to involve large surfaces of the molecules. This has been observed in other hormones and growth factors such as insulin (Blundell et al., 1972) and nerve growth factor (McDonald et al., 1991), where the tertiary structures are clearly important but no individual amino acid is absolutely essential. The interactions appear to be comprised of a larger number of hydrogen bonds, electrostatic and hydrophobic interactions. There is little convincing evidence that small molecules can compete for binding either as agonists or antagonists. Surprisingly many of the protein hormones and growth factors have no full antagonists, even amongst modified proteins. This may indicate the role of dimers of molecules such as nerve growth factor (NGF); in this case the dimers are very tightly bound and NGF circulates as dimers. The explanation for the lack of full antagonists of insulin is less evident as insulin circulates as a monomer and insulins that cannot form dimers can still be active.

6. Conclusions

What can then be said about the lock and key hypothesis? First very few systems even approximate to

the interactions of two preformed complementary surfaces. The closest to this model are binding proteins such as the lectins which have rigid structural templates, often comprised of highly hydrogen-bonded antiparallel β-structures. But even in these systems loops may be variable in conformation and induced fit may contribute to recognition.

Most proteins use flexibility of their internal structures to provide a complementary surface by induced fit. α-Proteins or $\alpha\beta$-proteins provide more malleable structures where helices can move with respect to other helices or parallel β-sheets; such movements can also mediate allosteric effects. Further to this loops between secondary structures can move to allow access of ligands and subsequently bind over the ligand.

Nevertheless, the ligand protein complexes involve well-defined conformers not unlike the lock and key of Fischer. As such conformational changes are extremely difficult to predict, even with knowledge of the structure of the uncomplexed protein, the "lock and key" hypothesis remains the working model in the design of many ligands of interest to the pharmaceutical, agrochemical and other "biotech" industries. Extensive computer software has been developed to discover molecules that are complementary to the liganded or unliganded forms of the protein. Such structure-based design promises to be a major feature of discovery processes, even for elaboration of a lead compound that has been identified by random screening of large databases of synthetic chemicals or natural products produced from microbial fermentation or plant extracts.

References

Banner, D.W., Darcy, A., Janes, W., Gentz, R., Schoenfeld, H.J., Broger, C., Loetscher, H. and Lesslauer, W. (1993) Crystal structure of the soluble human 55 kD TNF receptor-human TNF-ß complex – Implication for TNF receptor activation. Cell 73, 431–445.

Blundell, T., Dodosn, D., Hodgkin, D. and Mercola, D. (1972) Insulin. The structure in the crystal and its reflection in chemistry and biology. Adv. Prot. Chem. 26, 279–402.

Blundell, T., Carney, D., Gardner, S., Hayes, F., Howlin, B., Hubbard, T., Overington, J., Singh, D.A., Sibanda, B.L. and Sutcliffe, M. (1988) 18th Krebs, Hans lecture – Knowledge-based protein modelling and design. Eur. J. Biochem. 172, 513–520.

Dealwis, C.G., Frazao, C., Badasso, M., Cooper, J.B., Tickle, I.J., Driessen, H., Blundell, T.L., Murakami, K., Miyazaki, H., Sueiras-Diaz, J., Jones, D.M. and Szelke, M. (1994) X-ray analysis at 2.0 Å resolution of mouse submaxillary renin complexed with a decapeptide inhibitor CH-66, based on the 4–16 fragment of rat antigiotensinogen. J. Mol. Biol. 236, 342–360.

De Vos, A.M., Ultsch, M. and Kossaikoff, A.A. (1992) Human growth hormone and extracellular domain of its receptor: Crystal structure of the complex. Science 225, 306–312.

Dhanaraj, V., Dealwis, C.G., Frazao, C., Badasso, M., Sibanda, B.L., Tickle, I.J., Cooper, J.B., Driessen, H.P.C., Newman, M., Aguilar, C., Wood, S.P., Blundell, T.L., Hobart, P.M., Geoghegan, K.F., Ammirati, M.J., Danley, D.E., O'Connor, B.A. and Hoover, D.J. (1992) X-ray analyses of peptide–inhibitor complexes define the structural basis of specificity for human and mouse renins. Nature 357, 466–471.

Emsley, J., White, H.E., O'Hara, B.P., Oliva, G., Srinivasan, N., Tickle, I.J., Blundell, T.L., Pepys, M.B. and Wood, S.P. (1994) Structure of pentameric human serum amyloid P component. Nature 367, 338–345.

Gout, I. et al., (1993) The GTPase dynamin binds to and is activated by a subset of SH3 domains. Cell 75, 25–36.

Liao, D.-I., Kapadia, G., Ahmed, H., Vasta, G.R. and Herzberg, O. (1994) Structure of S-lectin, a developmentally regulated vertebrate ß-galactoside-binding protein. Proc. Natl. Acad. Sci (USA) 91, 1428–1432.

Lobsanov, Y.D., Gitt, M.A., Leffler, H., Baranodes, H. and Rini, J.M. (1993) X-ray crystal structure of the human dimeric S-lac lectin, L-14-II, in complex with lactose at 2.9 Å resolution. J. Biol. Chem. 268, 27034–27038.

McDonald, N.Q., Lapatto, R., Murray-Rust, J., Gunning, J., Wlodawer, A. and Blundell, T.L. (1991) New protein fold revealed by a 2.3 Å resolution crystal structure of nerve growth factor. Nature 354, 411–414.

Meador, W.E., Means, A.R. and Quiocho, F.A. (1993) Modulation of calmodulin plasticity in molecular recognition on the basis of X-ray structures. Science 262, 1718–1721.

Monod, J., Wyman, J. and Changeux, J.P. (1965) On the nature of allosteric transitions: a plausible model. J. Mol. Biol. 12, 88–118.

Musacchio, A., Saraste, M. and Wilmanns, M. (1994) High-resolution structures of tyrosine kinase SH3 domains complexed with proline-rich peptides. Nat. Struc. Biol. 1, 546–551.

Overington, J., Johnson, M.S., Sali, A. and Blundell, T.L. (1990) Tertiary structural constraints on protein evolutionary diversity – Templates, key residues and structure prediction. Proc. R. Soc. (London) Ser. B, Biol. Sci. 241, 132–145.

Sali, A., Veerapandian, B., Cooper, J.B., Foundling, S.I., Hoover, D.J. and Blundell, T.L. (1989) High resolution X-ray diffraction study of the complex between endothiapepsin and an oligopeptide inhibitor : the analysis of inhibitor binding and description of the rigid body shift in the enzyme. EMBO J. 8, 2179–2188.

Sali, A., Veerapandian, B., Cooper, J.B., Moss, D.S., Hofmann, T. and Blundell, T.L. (1992) Domain flexibility in the aspartic proteinases. Proteins 12, 158–170.

Veerapandian, B., Cooper, J.B., Sali, A., Blundell, T.L., Rosati, R.L., Dominy, B.W., Damon, D.B. and Hoover, D.J. (1992) Direct observation by X-ray analysis of the tetrahedral intermediate of aspartic proteinases. Protein Sci. 1, 332–328.

Wodak, S.J., Decrombrugghe, M. and Janin, J. (1987) Computer studies of interactions between macromolecules. Prog. Biophys. Mol. Biol. 49, 29–63.