J. Enzyme Inhibition, 1997, Vol. 11, pp. 223–243
Reprints available directly from the publisher
Photocopying permitted by license only

CHARACTERISING NON-COVALENT INTERACTIONS WITH THE CAMBRIDGE STRUCTURAL DATABASE

JOS P.M. LOMMERSE and ROBIN TAYLOR*

Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge CB2 1EZ, UK

(Received 29 July 1996)

Keywords: Cambridge Structural Database; Brookhaven Protein Data Bank; non-covalent interactions; rational drug design; intermolecular geometry.

INTRODUCTION

In recent years, there has been a rapid rise in the number of enzymes whose three-dimensional (3D) structures have been determined by X-ray crystallography or NMR. A number of factors have contributed to this increase, including the development of better equipment for collecting diffraction data and the use of recombinant DNA techniques for producing large quantities of proteins. It is now estimated that the chief repository of protein crystal structures, the Brookhaven Protein Data Bank (PDB)[1], may contain as many as 30,000 entries by the end of the century. Although many of these entries will be near-homologues (e.g. a wild-type enzyme and a mutant produced by site-directed mutagenesis), the availability of so many 3D structures will naturally lead to a greater emphasis on rational, structure-based inhibitor design. After years of effort, there is an increasing body of literature showing that such structure-based design can be successful. Good examples include the design of inhibitors of HIV protease[2], purine nucleoside phosphorylase[3] and thymidylate synthase.[4]

Most enzyme-inhibitor binding is non-covalent. Hence, an understanding of non-covalent interactions must lie at the heart of any successful structure-based inhibitor design strategy. In particular, information is required about: (1) what types of nonbonded interactions are likely to occur between a small molecule

*Correspondence.

and the residues at an enzyme active site; (2) what geometrical preferences these interactions will exhibit; and (3) what energy the interactions will contribute to protein-ligand binding. Relevant results can be obtained from a number of experimental techniques, including infra-red[5], microwave[6] and NMR spectroscopy.[7] Theoretical methods such as molecular orbital calculations and molecular dynamics simulations can also add insight. However, the most useful source of information is crystallography, which give us the ability to characterise a huge variety of nonbonded interactions at atomic resolution.

There are three main databases of crystal structures: the PDB, the Cambridge Structural Database (CSD)[8], and the Inorganic Crystal Structure Database (ICSD).[9] As already mentioned, the PDB contains protein structures. The CSD contains small-molecule organic and metallo-organic structures and the ICSD contains inorganics. The latter is relatively unimportant in the study of non-ionic interactions, but both the PDB and the CSD are extremely valuable. Each has strengths and weaknesses. The PDB has the obvious advantage that it is the most direct source of information on enzyme-inhibitor interactions. On the other hand, the precision of most PDB structures is comparatively low and there is always the chance of subjective bias due to the use of empirically-parameterised force fields in the refinement process. Also, the number and diversity of small-molecule ligands in the PDB is still small. In contrast, the CSD, while less directly relevant, contains a large number of very precise structures and a high degree of molecular diversity. Thus, the CSD and PDB are to some extent complementary.

In this review, we describe how the CSD may be used to characterise non-covalent interactions. We begin with a short summary of the CSD itself: the data it contains and how it can be searched. We then take four different types of non-covalent interactions, all of potential relevance to enzyme-inhibitor binding, and illustrate how they may be studied with the CSD. In each case, we compare conclusions derived from CSD analysis with those drawn from the more limited, but more directly relevant data in the PDB.

CSD DATA AND SOFTWARE

The CSD currently contains the results of about 150,000 small-molecule crystal-structure determinations. Data are abstracted from the primary literature or, sometimes, deposited directly as private communications. The database is fully comprehensive back to 1930 and about 12,000 new entries are added annually. Each new entry is taken through a series of checks, important amongst which are a test of the internal consistency of the published crystallographic data and a matching of the 3D structure onto a 2D chemical diagram drawn by an abstractor.

. The CSD can be searched with the program QUEST.[10] At its most basic level, this program provides simple text-based searching, e.g. for compound or author names. However, the most common use of QUEST is probably for 2D substructure searching. The software permits the user to define in a flexible way a query substructure, which may include variable bond and element types, specification of bond cyclicity/acyclicity, etc. Molecules in the CSD containing the query substructure are found rapidly by standard chemical connectivity-searching methodology. Hit structures (i.e. structures in the CSD containing the specified substructure) may be visualised in QUEST, both as isolated molecules and as expanded crystallographic unit cells.

QUEST can also perform 3D searches. At the intramolecular level, these permit the user to specify that hit molecules must contain a particular spatial arrangement of atoms (e.g. as in "pharmacophore searching"). In addition, and of particular relevance to this paper, QUEST is unique in providing 3D *intermolecular* searching. This involves the encoding of two (or more) molecular fragments, together with one (or more) distance ranges between these fragments. For example, in searching for intermolecular hydrogen bonds, the user might draw an O-H group and a C=O group, and specify that the intermolecular H...O(carbonyl) distance be in the range 1.8–2.5 Å. The search software will first locate all CSD structures that contain both an O-H and C=O group. By exhaustively applying all crystallographic symmetry operators, QUEST will then examine O-H...O=C contacts *between neighbouring molecules in the lattice* in order to see whether the user-specified contact exists anywhere within the crystal structure. As with 2D substructure searches, hits thus found may be stored and visualised. A particularly effective method of visualisation is to overlay the hits so that one of the fragments is exactly superimposed.[11] This shows the spatial distribution of the second fragment around the first (e.g. O-H groups around C=O). Such a presentation is useful for revealing geometrical preferences (e.g. O-H groups are found to lie mainly along the lone-pair directions of the carbonyl oxygen[12] with roughly linear O-H...O angles).

HYDROGEN BONDS TO PHOSPHATE GROUPS

Phosphate groups frequently occur in biomolecules, notably in the DNA backbone and in common cofactors. They have wide opportunities to form hydrogen bonds, both as donors and, more usually, as acceptors. A (negatively) charged phosphate group will act as a strong acceptor, since a hydrogen bond is predominantly an electrostatic attraction between an electron-rich acceptor and the donor group.[13] Two types of oxygen atoms can be distinguished in phosphate groups (Figure 1). First, there are oxygen atoms that link the central phosphorus to another heavy atom,

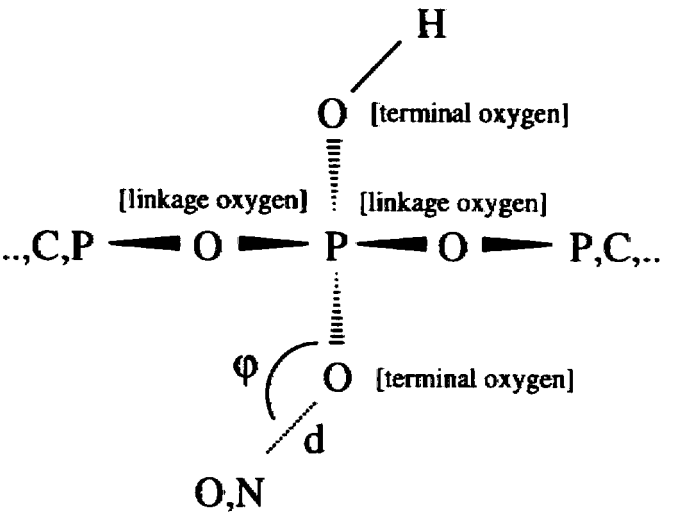


FIGURE 1 Linkage and terminal oxygen atoms in phosphates and definition of hydrogen bond geometrical parameters.

usually a second phosphorus (e.g. in diphosphates) or a carbon atom. These will be denoted as 'linkage-oxygens'. They act as hydrogen-bond acceptors only. The second type of oxygen will be denoted 'terminal-oxygens', as they are on terminal positions of phosphate groups. An oxygen of this type may possibly bear a hydrogen atom, in which case it can serve both as an acceptor and as a donor of hydrogen bonds.

In this section, we illustrate how the CSD may be used to characterise the hydrogen-bonding properties of both types of oxygen. Specifically, we consider, (1) the relative frequencies with which they form hydrogen bonds, and (2) the directionality or geometry of the hydrogen bonds. The results from the CSD are compared with those from the PDB.

Searches of the CSD clearly reveal a large difference in the hydrogen-bonding properties of the two types of phosphate oxygen (Figure 2a,b). There is a direct competition between these oxygens to form hydrogen bonds and it is evident that the terminal oxygens form far more than the linkage oxygens (1680 times [94%] compared with 98 times [6%]). However, one should be cautious in interpreting this. Calculation of the solvent accessible surface, by probing a central phosphate group in, for example, a triphosphate moiety, using a sphere of 2.0 Å radius, shows that approximately 85% of the accessible surface is accounted for by the terminal

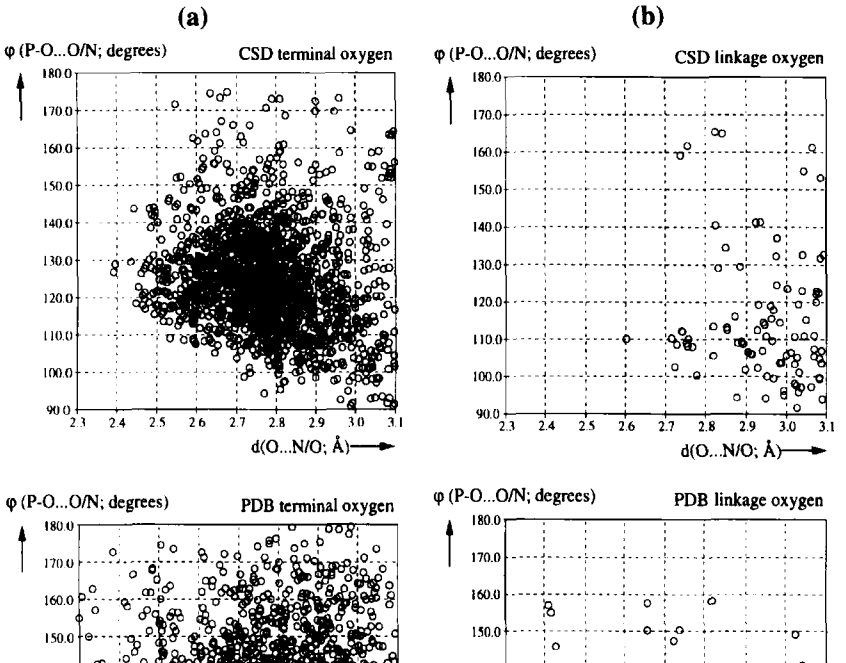


FIGURE 2 Scattergrams representing hydrogen bonding to terminal and linkage oxygens of phosphate groups as derived from CSD and PDB data. All intermolecular van der Waals contacts between phosphate oxygens and any oxygen or nitrogen (d(O...N/O)≤3.1 Å) were stored. The angle of approach of the oxygen or nitrogen to the phosphate oxygen, φ (angle P-O...O/N, degrees) has been plotted as function of the distance d(Å), see also Figure 1. (a) Hydrogen bonds involving the terminal phosphate oxygen atoms in the CSD. (b) Hydrogen bonds involving the linkage phosphate oxygen atoms in the CSD. (c) Hydrogen bonds involving the terminal phosphate oxygen atoms in the PDB. (d) Hydrogen bonds involving the linkage phosphate oxygen atoms in the PDB.

oxygens and only 15% by the linkage oxygens. Therefore, on statistical grounds, one would expect the former to accept more hydrogen bonds. Nevertheless, the hydrogen-bond (O...O or O...N) distances are on average much shorter for H-bonds involving terminal oxygens (2.7–2.9 Å) than for those to linkage oxygens (scattered between 2.7–3.1 Å). This suggests that terminal oxygens do form stronger H-bonds.

The second interesting feature found from the CSD is that there is a marked directional preference for hydrogen bonds with φ between 110–140° (φ = P-O...O/N angle). This preference is very clear for the terminal oxygen acceptors. This means that hardly any hydrogen bonds form in a linear position with respect to the P-O axis of the phosphate ($\varphi = 180°$). One might argue that this position should actually be preferred, because it is sterically the most accessible.

Two major factors may play a role here. First, intrinsic hydrogen-bond properties might lead to a stronger hydrogen bond when formed in the lone pair direction of the terminal oxygen acceptor ($\varphi \cong 55°$). Many other types of hydrogen-bond acceptor have already been found to exhibit such lone-pair directionality, e.g. carbonyl oxygens[12] or amine groups. However, this directionality has never been shown to be very clear-cut. Therefore, the other factor might be more important. Since (charged) phosphate groups are very strong acceptors, they will generally form, if possible, more than one hydrogen bond. In such cases, steric effects will cause the hydrogen bonds to form preferentially in a 'halo' around each terminal oxygen atom.

Summarising, the CSD clearly shows that, (1) terminal oxygens accept hydrogen bonds far more frequently and at shorter distances than linkage oxygens, and (2) the position of the hydrogen-bond donor is mostly in or near the direction of the lone pairs of the oxygen atoms.

Let us now turn to the PDB. Besides inorganic phosphate, there are over 160 examples of crystal structures of proteins which are complexed with ligands containing mono-, di-, tri- and higher oligophosphate groups. As for the CSD we have performed a search for the frequency and orientational preference of hydrogen-bond formation towards the terminal and linkage oxygens of such groups. We have also studied a number of individual PBD entries in detail.

For the investigation of frequencies and orientational preferences, the organic phosphate ligands were defined as carbon-containing compounds that include phosphorus atoms to which four oxygen atoms are bonded, and which are designated as 'HETATM' in the PDB entry. As in the CSD searches, contacts less than 3.1 Å were found between the phosphate oxygen atoms and protein oxygen or nitrogen atoms. These contacts were assumed to be hydrogen bonds, as it is unlikely that two electron-rich atoms can approach each other so closely without a stabilising hydrogen between them.

The results are shown in Figure 2c,d for the terminal and linkage oxygens, respectively. The data points are more scattered in these PDB plots than in those from the CSD (Figure 2a,b), presumably because of the lower experimental precision of the former, but the same general features are present. Many more hydrogen bonds form to terminal oxygens than to linkage oxygens (2580 [94%] compared with 171 [6%]), and the contacts to the terminal oxygens are shorter on average and appear

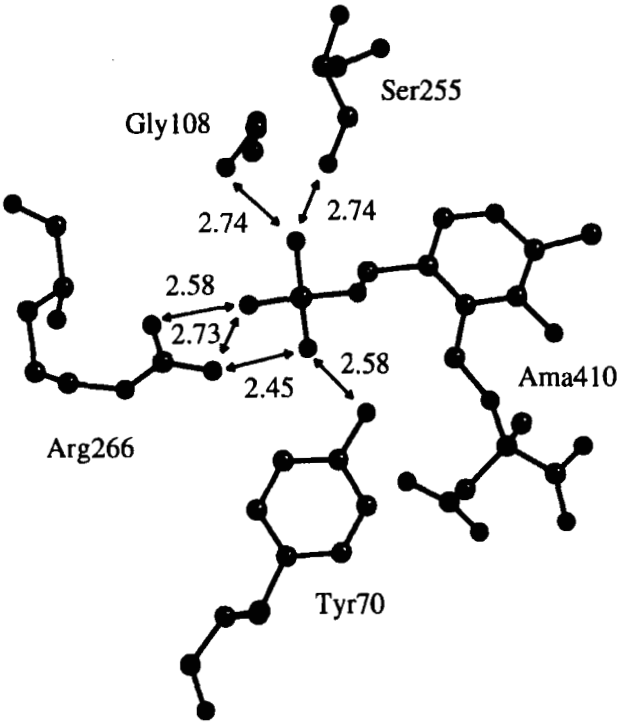


FIGURE 3 Representation of the hydrogen-bond network involving the phosphate group of 2-methyl-aspartyl-pyridoxal-5′-phosphate and the protein active site of aspartate aminotransferase (PDB entry 1ASL).

to have a preferred position in or near to the lone pairs of the phosphate oxygens. This demonstrates an excellent consistency between CSD and PDB data.

In Figure 3, we have represented the hydrogen bonds formed to the monophosphate moiety of 2-methylaspartyl-pyridoxal-5′-phosphate in its complex with apartate aminotransferase[14] (PDB code 1ASL). The terminal phosphate oxygens are involved in a network of hydrogen bonds with the protein and possibly with water molecules too. In contrast, the linkage oxygen of the pyridoxal moiety does not accept any hydrogen bond. Furthermore, the orientation of the hydrogen-bond donors around the phosphate is consistent with the preferred geometry observed in the CSD, i.e. φ angles of less than 150°.

Another typical H-bonding pattern is found in the complex between adenosine triphosphate and ATPase[15] (PDB code 1NGF). Part of the triphosphate moiety is

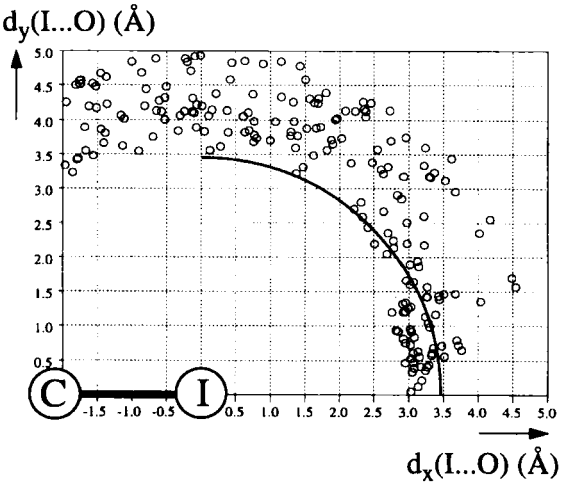
exposed to the solvent. The terminal oxygens form at least six primary hydrogen bonds to crystallographically observed water molecules. Eight other hydrogen bonds are formed, all to terminal oxygens. In seven cases, the phosphate acts as H-bond acceptor, and only once as a donor (to the carbonyl oxygen of Tyr15; O...O 2.7 Å). Again, the orientations of the eight hydrogen bonds agree with the observations made in the CSD: no linear P-O...O/N geometries are present.

However, linkage oxygens do form hydrogen bonds on some occasions. The phosphate moiety of the inhibitor P1,P5-bis-adenosine-5′-pentaphosphate, complexed with adenylate kinase[16] (1AKE), forms fifteen hydrogen bonds to the surrounding amino acid residues of the protein. Three of these can be judged as hydrogen bonds between phosphate linkage oxygens and N-H groups. The O...N distances of these bonds are about 3.0 Å. On the other hand, the twelve hydrogen bonds involving the teminal phosphate oxygen atoms have an average intermolecular O...N/O distance of 2.4 Å. This indicates that the hydrogen bonds to the terminal oxygens are stronger than those to the linkage oxygens, which agrees with the more precise data from the CSD.

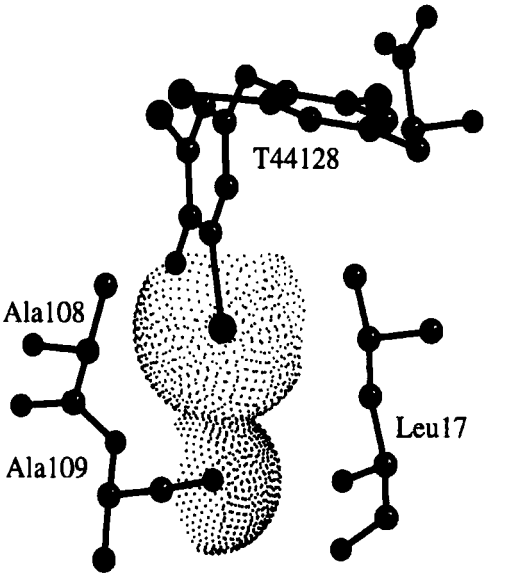
HALOGEN...OXYGEN INTERACTIONS

The large diversity of chemical structures in the CSD makes it an excellent source for finding novel interactions. A type of nonbonded interaction which is relatively unknown is the attraction between halogen atoms and electronegative atoms, notably oxygen or nitrogen.

Previously, we have investigated this type of interaction by searching the CSD and performing *ab initio* based intermolecular calculations.[17] Briefly, the results demonstrate that there can exist a highly directional, attractive interaction between oxygen or nitrogen and the heavier halogens (X: chlorine, bromine or iodine, but not fluorine). Specifically, O or N atoms are observed in the CSD to form short contacts to carbon-bonded halogen atoms along the extension of the C-X axis (the 'head-on' direction; Figure 4a). The X...O/N distances tend to be shorter than the sum of van der Waals radii (i.e. there is penetration of atomic volumes). When allowance is made for geometrical factors, the density of observed contacts in the region of penetrated volumes is much higher than would be expected by random chance. Considering the various halogen atoms and various types of electronegative atoms, the following trends are seen. First, the relative density of contacts in the penetrated part of the halogen atom increases with increasing atomic number (i.e. Cl<Br<I). Secondly, the interaction appears to be stronger for oxygen atoms than nitrogen atoms, and also stronger at lower hybridisation states (sp or sp^2, rather than sp^3).



(a)



(b)

FIGURE 4 (a) Distribution of sp^2-hybridised oxygen atoms around carbon-bonded iodine atoms within a sphere of 5 Å from the iodine; based on structures from the CSD. All data points have been projected in the C-I...O plane. Note that this leads to higher densities of the data points at large d_y values in this 2D plot than is representative for the actual 3D space.[17] (b) The intermolecular I...O interaction between 3,5,3′,5′-tetraiodo-L-thyronine and trans-thyretine (PDB entry 1ETA).

It is generally assumed that unusually short nonbonded contacts are attractive. However, an alternative, geometrical explanation has been proposed for the short X...O/N contacts in the CSD, viz. that halogen atoms are polar-flattened[18] (i.e. have a shorter radius along the extension of the C-X axis than perpendicular to this axis). The crystallographic data cannot be used to distinguish between these two explanations, so *ab initio* based Intermolecular Perturbation Theory calculations were performed to gain extra insight. They showed that the carbon-bonded halogen atom has an anisotropic distribution of electron density around its nucleus, the density being higher perpendicular to the C-X axis than in the head-on direction. This has two consequences. First, the anisotropy causes the halogen atom to be non-spherical, having a longer radius perpendicular to the C-X axis. Therefore, the atom-atom repulsion at a fixed distance from the halogen nucleus is larger perpendicular to the C-X axis than in the head-on direction. This is consistent with the "polar-flattening" argument. Secondly, however, the anisotropy leads to a quadrupole at the halogen nucleus with a relatively low electron density region when looking in the head-on direction. This can result in an electrostatic attraction between an (electronegative) oxygen or nitrogen atom and a halogen atom when the contact is in the head-on direction. These two major effects, the non-spherical shape and quadrupole moment, therefore lead to a directional and attractive interaction. Another consequence is that when halogen atoms are in an electron-withdrawing environment, they will have an even lower electron density in the head-on direction, leading to a stronger interaction. This agrees with the CSD data, which suggest that the strongest attractions occur between the more polarisable, heavier halogens, and the more electronegative atoms.

Although the X...O and X...N contacts should be listed in the category of 'weak nonbonded interactions', it has been shown that the interaction energy can be stronger than –10 kJ/mol, which is about half the strength of an average, neutral hydrogen bond.[17] Thus, it has the potential to be important in the design of enzyme inhibitors. There is tentative evidence for the interaction in the PDB structure of trans-thyretine (prealbumin) complexed with 3,5,3′,5′-tetraiodo-L-thyronine[19,20] (PDB entries 1ETA and a point mutated variant in 1ETB; Figure 4b). The ligand forms a short I...O contact to a backbone carbonyl of the protein. The oxygen penetrates the iodine in a fairly linear direction (C-I...O angle 161°) at a distance of 3.1 Å, approximately 0.4 Å less than the sum of the van der Waals radii. Steinrauf *et al.* argued that the contact is forced by a narrow binding site and should destabilise the complex in comparison with the 1ETB variant, which has a wider binding site and a longer I...O distance (3.2 Å).[20] Nevertheless, in both complexes the penetration of the iodine and oxygen atoms is similar to that seen in CSD structures (Figure 4a), suggesting that it may well be attractive. Iodine is a very 'soft' and polarisable halogen, and, as it is attached to an

aromatic ring, it is in an electron withdrawing environment. This will favour I...O interaction.

HYDROGEN BONDING TO AROMATIC π SYSTEMS

It is now commonly known that π-systems can act as weak hydrogen bond acceptors. The CSD contains many well-determined examples of such hydrogen bonds, which can be used to characterise the preferred geometry of the interaction and to get information on which π-systems might preferentially form it. In this review, we consider aromatic π-ring systems only, but any π-electron system might, in principle, serve as an acceptor.

Generally, as a hydrogen bond is mainly an electrostatic interaction, the π-systems that have higher electron densities will act as stronger acceptors. Therefore, in order to strengthen this weak hydrogen bond, one might include electron donating substituents, such as a hydroxyl group in phenol, or use electron-rich heterocyclic systems such as indole.

In Figure 5a,b, the observed distributions of O-H and N-H groups around benzyl and indole-3-yl ring systems are represented. Both π-systems are potentially important hydrogen-bond acceptors in proteins as side chain constituents of phenylalanine and tryptophan, respectively. In the case of the benzyl acceptor (Figure 5a), a clustering of O-H and N-H vectors occurs above the centre of the ring. They are pointing straight to the centre of the ring. The average distance of this cluster from the ring centre is approximately 2.8 Å. It is interesting to observe that the other O-H and N-H vectors above the plane of the ring are also pointing to the centre of the π-system. Thus, the whole π-system seems to act as a hydrogen bond acceptor. At the edge of the ring, the vectors are randomly distributed.

The indole-3-yl ring acceptor (Figure 5b) shows a striking preference for formation of hydrogen bonds to the centre of the six-membered rather than the five-membered ring. The hydrogen bonds are very well oriented and, with respect to the total number of contacts, they occur frequently, which indicates that this bond is relatively strong compared with those formed by other π-acceptors. The average distance from the cluster of donor groups to the centre of the six-membered ring is approximately 2.7 Å. The five membered hetero-ring π-electrons also accept hydrogen bonds, but fewer. The preference for the six-membered ring can be rationalised by considering the electrostatic potential of indole (Figure 6), which is more negative above the six-ring than above the five-ring.[21]

Let us now consider O-H...π and N-H...π hydrogen bonds in the PDB. In general, it is difficult to find unequivocal evidence for hydrogen bonding to π-systems in the PDB, as crystallographically determined coordinates of hydrogen atoms

FIGURE 5 (a) Top and edge view of a 3D scattergram showing the distributions of N-H and O-H groups around benzyl ring systems. For all contacts shown, the distance between the H atom and at least one benzyl atom is less than v, where v = sum of van der Waals radii plus 1.5 Å (b) Top and edge view of a 3D scattergram showing the distributions of N-H and O-H groups around indole groups. For all contacts shown, the distance between the H atom and at least one indole atom is less than v.

are absent. Nevertheless, they may often be inferred with reasonable confidence. In PDB entry 1MCR, the light chain of λ-immunoglobulin has been complexed with N-acetyl-L-His-D-Pro-OH by diffusion of this peptide into the protein crystal[22] (Figure 7a). The edge of the imidazole ring of the ligand histidine residue faces the plane of the phenol of Tyr93. The position of the N^{δ} atom of the imidazole is roughly above the centre of the phenol ring. Although the coordinates of the hydrogen atoms are absent, one can assume that N^{δ} is protonated, as it would be unlikely that an unprotonated nitrogen would point into the middle of the π-electron cloud of the phenol residue, due to electrostatic repulsion. The aromatic phenol ring is relatively electron-rich and will be a reasonably strong acceptor.

O-glycosyl hydrolase has been complexed with 4-acetyl-amino-3-aminobenzoic acid[23] (PDB entry 1IVE, Figure 7b). The aromatic ring of this inhibitor is thought to act as a hydrogen bond acceptor for Tyr406. Furthermore, close to the other side

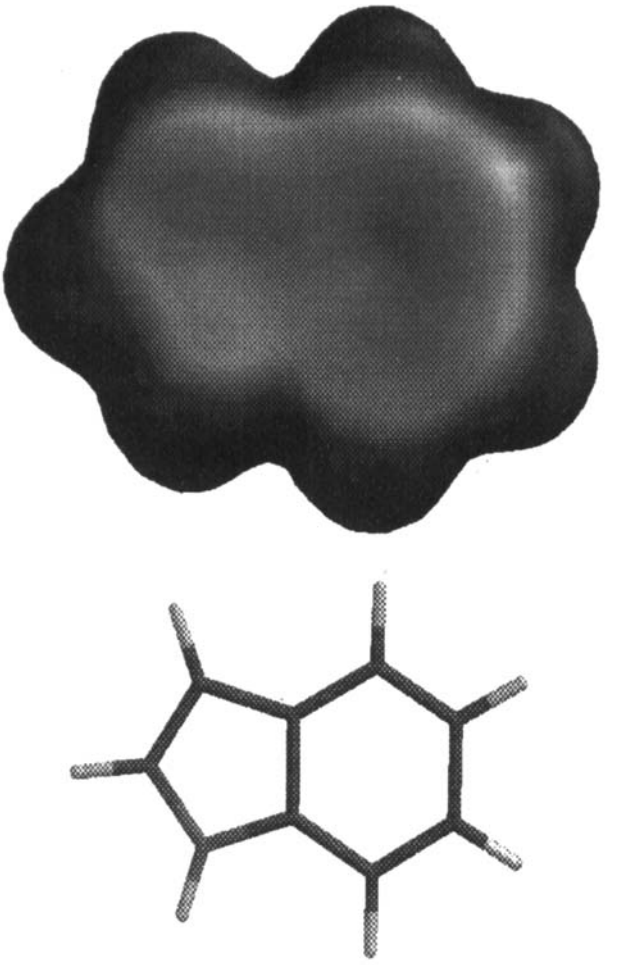


FIGURE 6 Electrostatic potential surface of indole, in the range −25.0 to 25.0 mHartree; calculated with Spartan[21] at the 6–31G** basis-set level, indole geometry fully optimised (see Colour Plate at rear).

of the inhibitor's aromatic plane is the carboxyl group of Asp151, O...C contact distances being between 2.9–3.5 Å. It seems unlikely that a negatively charged carboxylate group would approach a π-electron cloud so closely, so the acid group may be unionised and forming an O-H...π hydrogen bond.

Although hydrogen bonds to π-systems are few in inhibitor...protein complexes, there are many examples amongst protein side-chain...side-chain interactions.

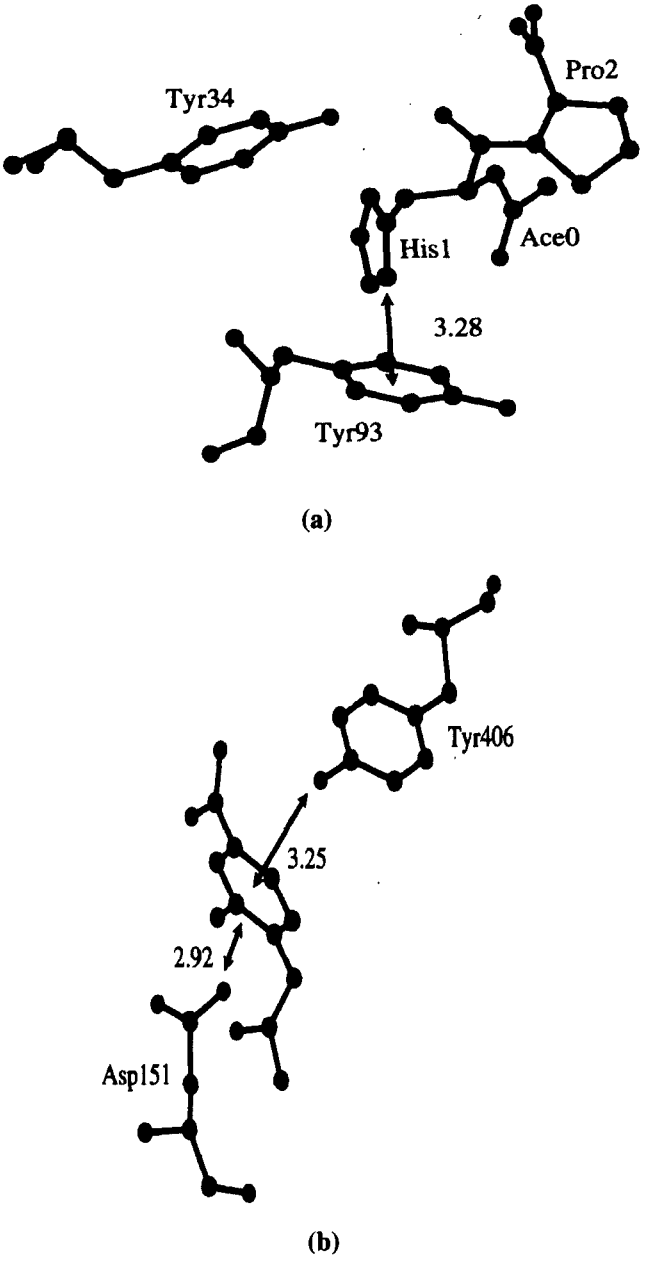


FIGURE 7 (a) Possible hydrogen bonding to a π-system in the complex of N-acetyl-L-His-D-Pro-OH with the light chain of λ-immunoglobulin (PDB entry 1MCR); (b) Possible hydrogen bonding to a π-system in the complex of 4-acetyl-amino-3-aminobenzoic acid with *O*-glycosyl hydrolase (PDB entry 1IVE).

These might have important consequences in ligand binding or in determining protein structures. Such an interaction has been suggested recently, in which the hydrogen bond to the π-system of a tyrosine residue interferes with the binding of glutathione to the active site of rat μ-glutathione S-transferase[24] (PDB entry 1GST[25]). Here, Tyr6 is involved in a (weak) hydrogen bond to the sulphur of glutathione, and acts as a hydrogen bond acceptor for the hydroxyl group of Thr13. This latter O-H...π hydrogen bond appears to be essential in binding glutathione. Mutation of Thr13 to alanine (1HNA[26]) or valine (1GNE[27]) turns out to reduce the binding considerably. The threonine O-H...π hydrogen bond seems to force Tyr6 into a position necessary for the formation of the hydrogen bond to the sulphur of glutathione, rather than forming a hydrogen bond to another, stronger acceptor.

The importance of hydrogen bonds to π-systems may be illustrated by more examples. It has even been proposed that C-H...π interactions are of sufficient importance to influence binding in biomolecular systems![28] Although the C-H...π interaction will be very weak, there are so many present in every protein that the total interaction might become significant.

INDOLE STACKING INTERACTIONS

The non-covalent interactions formed by indole and its derivatives are important because of the occurrence of this ring system in the side chain of tryptophan. As early as 1958, Pullman and Pullman[29] pointed out that indole is a particularly electron-rich aromatic system. Consequently, it might be expected to form attractive intermolecular interactions with electron deficient species. We have already seen that the π-system of indole is actually capable of accepting a hydrogen bond (see above); what other interactions might it form?

A search of the CSD for structures containing both an indole ring (possibly substituted) and a quaternary nitrogen finds several interesting hits. Typical amongst these are 6-bromohypaphorine[30] (CSD code BHYPOR, Figure 8a), the complex between 3-methyladenine and indole-3-acetic acid[31] (HICVEU, Figure 8b), 5-(2-(indole-3-yl-propionyloxy)ethyl)-3,4-dimethylthiazolium iodide[32] (CEDMAZ10) and 1-methyl-3-carbamido-pyridinium N-acetyl-L-tryptophanate[33] (MCPATR). In each case, the molecular moiety containing the positively charged nitrogen is stacked with the indole ring system, these moieties being, respectively, trimethylammonium, adeninium, thiazolium and pyridinium. The results therefore suggest that indole is capable of interacting with a wide variety of electron deficient species. Indeed, this appears to be a strong tendency. In the fourteen CSD structures found in the above search, a stacking interaction between indole and the quaternary nitrogen moiety occurs in all but two (Table I). It may be that the benefits

TABLE I CSD structures containing indole rings and quaternary nitrogen systems

CSD code	*Compound*	*Stacking?*	*Ref.*
BHYPOR	L-6-Bromohypaphorine	Yes	30
BIWKAT10	9-(3-Indol-3-yl-propyl)-1-methyladeninium iodide dihydrate	Yes	34
BOMTUS10	Thiamine-indole-3-propionate perchlorate hydroperchlorate methanol solvate	No	32
CEBZUE10	7-Methyl-9-ethylguaninium indole-3-acetate	Yes	35
CEDMAZ10	5-(2-(Indol-3-yl-propionyloxy)ethyl)-3,4-dimethylthiazolium iodide	Yes	32
DAFBAN	3-Carbamoyl-1-(3-(indol-3-yl)-propyl)-pyridinium sulfate hydrate	Yes	36
DOTXUF10	Tryptaminium 7-methylguanosine 5′-mono-phosphate trihydrate	Yes	35
FEVZUB	1-(3-(3-Indolyl)propyl)pyridinium bromide	No	37
GATSAV	7-Methylguanosinium indole-3-acetate mono-hydrate	Yes	35
HICVEU	3-Methyladenine indole-3-acetic acid pentahydrate	Yes	31
INYECP	1-(2-Indol-3-ylethyl)-3-carbamido-pyridinium chloride monohydrate	Yes	38
MCPATR	1-Methyl-3-carbamido-pyridinium *N*-acetyl-L-tryptophanate monohydrate	Yes	33
MEADIN10	1,9-Dimethyladeninium indole-3-acetate trihydrate	Yes	34
SEKXIP10	Tryptophanylglutamic acid 7-methylguanosine-5′-phosphate trihydrate	Yes	39

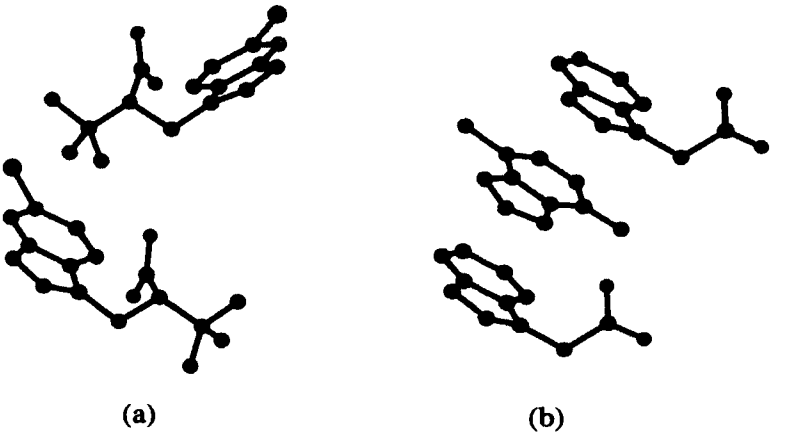
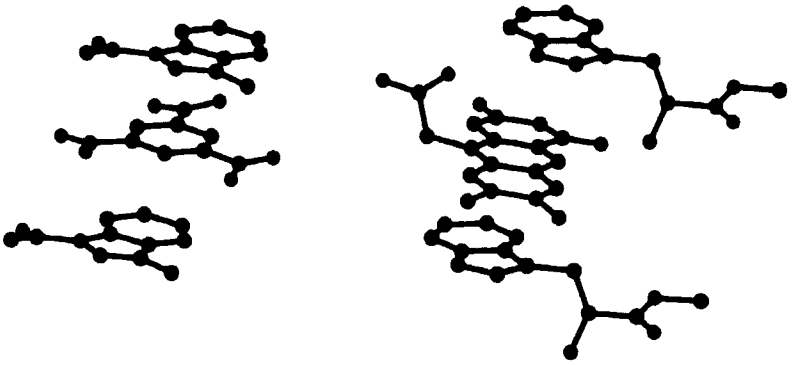


FIGURE 8 The indole ring system interacting with electron-deficient species in the CSD structures (a) 6-bromohypaphorine (CSD code BHYPOR), and (b) the complex between 3-methyladenine and indole-3-acetic acid (HICVEU).



(a) (b)

FIGURE 9 The indole ring system interacting with electron-deficient systems in, (a) the complex between 3-methylindole acetate and trinitrobenzene (ASKNBZ), and (b) the complex between 7,8-dimethylisoalloxazine-10-acetic acid and L-tryptophan methyl ester (CEHDIC).

of π-stacking arrangements are outweighed by strong ionic forces in the two exceptions, which both contain inorganic counterions (perchlorate, bromide).

Further searches of the CSD show that indole rings will also form stacking interactions with groups that, while not formally positively charged, are nonetheless electron deficient. A good example is the complex between 3-methylindole acetate and trinitrobenzene[40] (ASKNBZ, Figure 9a). Perhaps more relevant in biology is the stacking of indole and isoalloxazine in the complex between 7,8-dimethylisoalloxazine-10-acetic acid and L-tryptophan methyl ester[41] (CEHDIC, Figure 9b).

One of the concerns that is always present when studying small-molecule crystal structures is that the observed intermolecular interactions may be influenced by crystal packing forces. This is a particular problem when studying interactions between aromatic systems, since stacked arrangements may arise because they represent an efficient way of close-packing planar molecules.[42] Thus, we must consider the possibility that some of the interactions illustrated in Figures 8 and 9 may be due to crystallographic packing effects rather than reflecting inherent intermolecular preferences. On closer examination, however, this explanation seems unlikely. If it were true, we would be as likely to see indole stacking with itself as with the electron deficient moiety. In fact, indole...indole stacking occurs in none of the structures listed in Table I. Moreover, when the structures of simple indole derivatives such as tryptophan[43] and tryptamine hydrochloride[44] are examined, adjacent ring systems are seen to be in a T-shaped or tilted arrangement

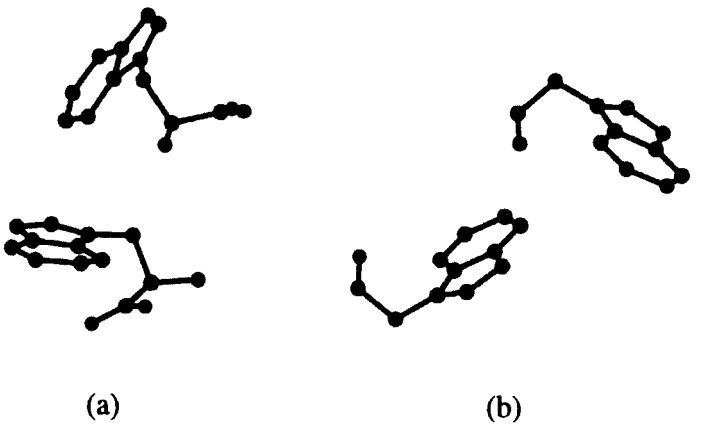


FIGURE 10 Tilted or T-interactions between adjacent indole ring systems in the CSD structures of (a) tryptophan (QQQBTP01) and (b) tryptamine hydrochloride (TRYPTA10).

(QQQBTP01, TRYPTA10, Figure 10). In other words, when no electron deficient moiety is present, indole appears to show a much reduced tendency to form stacking interactions. Interestingly, Lynch *et al.* solved the structures of the complexes between indole and trinitrobenzene, dinitrobenzene and 2-pyridone.[45] Stacking occurs in the first two complexes but not in the third, suggesting that the pyridone ring system is insufficiently electron deficient.

The CSD data imply that tryptophan side chains in proteins should be capable of binding a variety of electropositive groups. This expectation is amply fulfilled when the PDB is examined. Three examples are shown in Figure 11a,b,c. The first of these illustrates the binding of the choline moiety of phosphocholine to a Fab antibody[46] (PDB code 2MCP; cf. CSD structure BHYPOR, Figure 8a). The second shows a dinitrobenzene ring system sandwiched between two tryptophan side chains, again in an antibody complex[47] (1BAF, cf. CSD structure ASKNBZ, Figure 9a). Finally, Figure 11c shows the binding of the aminoacridinium anti-Alzheimers' drug tacrine to the active site of acetylcholinesterase[48] (1ACJ). It is noteworthy that, in each case, the species interacting with the indole ring system is electron deficient (in two cases, formally positively charged) *and that the electropositive charge is highly delocalised.* Delocalisation occurs in 1BAF and 1ACJ because the electron deficient moieties are aromatic. In the case of 2MCP, simple MO calculations are sufficient to show that the formal positive charge of the trimethylammonium group is, in practice, roughly equally distributed over the nine methyl hydrogen atoms. It seems likely that these extensive regions of "soft" electropositive charge are well complemented by the delocalised electron-rich π cloud of the indole ring system.

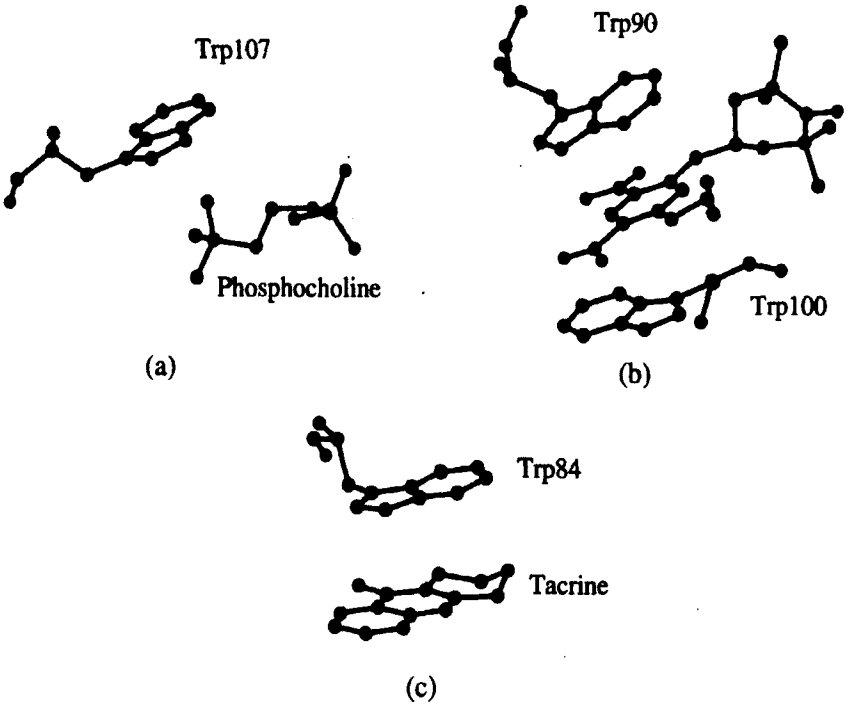


FIGURE 11 The indole ring system of tryptophan interacting with electron deficient ligand groups in three PDB complexes. (a) A phosphocholine – Fab antibody complex (2MCP); (b) a dinitrobenzene-containing molecule bound to an antibody (1BAF); (c) tacrine complexed with acetylcholinesterase (1ACJ).

SUMMARY

This review describes how the CSD can be used to study non-covalent interactions. Several different types of information may be obtained. First, the relative *frequencies* of various interactions can be studied; for example, we have shown that the terminal oxygen atoms of phosphate groups accept hydrogen bonds far more often than the linkage oxygens. Secondly, information can be obtained about the *geometries* of nonbonded contacts; for example, hydrogen bonds to P-O groups rarely form along the extension of the P-O bond, whereas short contacts between oxygen and carbon-bound iodine show a strong preference for linear C-I...O angles. Thirdly, the CSD can be searched for *novel interactions* which may be exploited in inhibitor design; for example, the I...O contacts just

mentioned, and N-H...π hydrogen bonds. Finally, the CSD can suggest *synthetic targets* for medicinal chemistry; for example, molecules containing delocalised electron deficient groups such as trimethylammonium, pyridinium, thiazolium and dinitrophenyl have a good chance of binding to an active-site tryptophan.

Although the CSD contains small-molecule crystal structures, not protein-ligand complexes, there is considerable evidence that the contacts seen in the two types of structures are similar. We have illustrated this a number of times in the present review and additional evidence has been given previously by Klebe.[49] The major advantages of the CSD are its size, diversity and experimental accuracy. For these reasons, it is a useful tool for modellers engaged in rational inhibitor design.

References

[1] Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M. (1977). *J. Mol. Biol.*, **112**, 535.

[2] Lam, P.Y.S., Jadhav, P.K., Eyermann, C.J., Hodge, C.N., Ru, Y., Bacheler, L.T., Meek, J.L., Otto, M.J., Rayner, M.M., Wong, Y.N., Chang, C.-H., Weber, P.C., Jackson, D.A. Sharpe, T.R. and Erickson-Viitanen, S. (1994). *Science*, **263**, 380.

[3] Babu, Y.S., Ealick, S.E., Bugg, C.E., Erion, M.D., Guida, W.C., Montgomery, J.A. and Secrist, J.A.III (1995). *Acta Cryst.*, **D51**, 529.

[4] Appelt, K., Bacquet, R.J., Bartlett, C.A., Booth, C.L.J., Freer, S.T., Fuhry, M.M., Gehring, M.R., Herrmann, S.M., Howland, E.F., Janson, C.A., Jones, T.R., Kan, C., Kathardekar, V., Lewis, K.K., Marzoni, G.P., Matthews, D.A., Mohr, C., Moumaw, E.W., Morse, C.A., Oatley, S.J., Ogden, R.C., Reddy, M.R., Reich, S.H., Schoettlin, W.S., Smith, W.W., Varney, M.D., Villafranca, J.E., Ward, R.W., Webber, S., Webber, S.E., Welsh, K.M. and White, J. (1991). *J. Med. Chem.*, **34**, 1925.

[5] Abraham, M.H., Grellier, P.L., Prior, D.V., Morris, J.J. and Taylor, P.J. (1990). *J. Chem. Soc., Perkin 2*, 521.

[6] Legon, A.C. and Millen, D.J. (1987). *Acc. Chem. Res.*, **20**, 39.

[7] Anelli, P.L., Ashton, P.R., Ballardini, R., Balzani, V., Delgado, M., Gandolfi, M.T., Goodnow, T.T., Kaifer, A.E., Philp, D., Pietraszkiewicz, M., Prodi, L., Reddington, M.V., Slawin, A.M.Z., Spencer, N., Stoddart, J.F., Vicent, C. and Williams, D.J. (1992). *J. Am. Chem. Soc.*, **114**, 193.

[8] Allen, F.H., Davies, J.E., Galloy, J.J., Johnson, O., Kennard, O., Macrae, C.F., Mitchell, E.M., Mitchell, G.F., Smith, J.M. and Watson, D.G. (1991). *J. Chem. Inf. Comput. Sci.*, **31**, 187.

[9] Bergerhoff, G. and Brown, I.D. (1987). In *Crystallographic Databases*, pp. 77-95. Bonn: International Union of Crystallography.

[10] *CSD User Manuals* (1992). Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge, UK.

[11] Taylor, R., Mullaley, A. and Mullier, G.W. (1990). *Pesticide Sci.*, **29**, 197.

[12] Taylor, R., Kennard, O. and Versichel, W. (1983). *J. Am. Chem. Soc.*, **105**, 5761.

[13] Umeyama, H. and Morokuma, K. (1977). *J. Am. Chem. Soc.*, **99**, 1316.

[14] Jaeger, J., Moser, M., Sauder, U. and Jansonius, J.N. to be published.

[15] Flaherty, K.M., Wilbanks, S.M., Deluca-Flaherty, C. and McKay, D.B. (1994). *J. Biol. Chem.*, **269**, 12899.

[16] Müller, C.W. and Schulz, G.E. (1992). *J. Mol. Biol.*, **224**, 159.

[17] Lommerse, J.P.M., Stone, A.J., Taylor, R. and Allen, F.H. (1996). *J. Am. Chem. Soc.*, **118**, 3108.

[18] Nyburg, S.C. and Wong-Ng, W. (1979). *Proc. R. Soc. Lond. A*, **367**, 29.

[19] Hamilton, J.A., Steinrauf, L.K., Braden, B.C., Liepnieks, J.J. and Benson, M.D. (1993). *J. Biol. Chem.*, **268**, 2416.

[20] Steinrauf, L.K., Hamilton, J.A., Braden, B.C., Murrell, J.R. and Benson, M.D. (1993). *J. Biol. Chem.*, **268**, 2425.

[21] Spartan version 4.0, Wavefunction Inc. Irvine, USA.
[22] Edmundsen, A.B., Harris, D.L., Fan, Z.-C., Guddat, L.W., Schley, B.T., Hanson, B.L., Tribbick, G. and Geysen, H.M. (1993). *Proteins, Struct., Funct., Genet.*, **16**, 246.
[23] Jedrzejas, M.J., Singh, S., Brouillette, W.J., Laver, W.G., Air, G.M. and Luo, M. (1995). *Biochemistry*, **34**, 3144.
[24] Xiao, G.Y., Liu, S.X., Ji, X.H., Johnson, W.W., Chen, J.H., Parsons, J.F., Stevens, W.J., Gilliland, G.L and Armstrong, R.N. (1996). *Biochemistry*, **35**, 4753.
[25] Ji, X., Zhang, P., Armstrong, R.N. and Gilliland, G.L. (1992). *Biochemistry*, **31**, 10169.
[26] Raghunathan, S., Chandross, R.J., Kretsinger, R.H., Allison, T.J., Penington, C.J. and Rule, G.S. (1994). *J. Mol. Biol.*, **338**, 815.
[27] Lim, K., Ho, J.X., Keeling, K., Gilliland, G.L., Ji, X., Ruker, F. and Carter, D.C. (1994). *Prot. Sci.*, **3**, 2233.
[28] Kabayashi, K., Asakawa, Y., Kikuchi, Y., Toi, H. and Aoyama, Y. (1993). *J. Am. Chem. Soc.*, **115**, 2648.
[29] Pullman, B. and Pullman, A. (1958). *Proc. Natl. Acad. Sci., USA*, **44**, 1197.
[30] Raverty, W.D., Thomson, R.H. and King, T.J. (1977). *J. Chem. Soc., Perkin Trans. 1*, 1204.
[31] Yamagata, Y., Kato, M. and Fujii, S. (1994). *Chem. Pharm. Bull.*, **42**, 2385.
[32] Ishida, T., Matsui, M., Inoue, M., Hirano, H., Yamashita, M., Sugiyama, K., Sugiura, M. and Tomita, K.-I. (1985). *J. Am. Chem. Soc.*, **107**, 3305.
[33] Ash, R.P., Herriott, J.R. and Deranleau, D.A. (1977). *J. Am. Chem. Soc.*, **99**, 4471.
[34] Ishida, T., Shibata, M., Fujii, K. and Inoue, M. (1983). *Biochemistry*, **22**, 3571.
[35] Ishida, T., Doi, M., Ueda, H., Inoue, M. and Sheldrick, G.M. (1988). *J. Am. Chem. Soc.*, **110**, 2286.
[36] Ishida, T., Ibe, S. and Inoue, M. (1984). *J. Chem. Soc., Perkin Trans. 2*, 297.
[37] Ishida, T., Itoh, M., Doi, M. and Inoue, M. (1986). *Chem. Pharm. Bull*, **34**, 4899.
[38] Herriott, J.R., Camerman, A. and Deranleau, D.A. (1974). *J. Am. Chem. Soc.*, **96**, 1585.
[39] Ishida, T., Iyo, H., Ueda, H., Doi, M., Inoue, M., Nishimura, S. and Kitamura, K. (1991). *J. Chem. Soc., Perkin Trans. 1*, 1847.
[40] Surcouf, E. and Delettre, J. (1978). *Acta Cryst.*, **B34**, 2173.
[41] Inoue, M., Okuda, Y., Ishida, T. and Nakagaki, M. (1983). *Arch. Biochem. Biophys.*, **227**, 52.
[42] Brock, C.P. and Minton, R.P. (1989). *J. Am. Chem. Soc.*, **111**, 4586.
[43] Bakke, O. and Mostad, A. (1980). *Acta Chem. Scand.*, **B34**, 559.
[44] Wakahara, A., Fujiwara, T. and Tomita, K.-I. (1973). *Bull. Chem. Soc. Jpn.*, **46**, 2481.
[45] Lynch, D.E., Smith, G., Byriel, K.A. and Kennard, C.H.L. (1991). *Aust. J. Chem.*, **44**, 809.
[46] Padlan, E.A., Cohen, G.H. and Davies, D.R. (1985). to be published.
[47] Brunger, A.T., Leahy, D.J., Hynes, T.R. and Fox, R.O. (1991). *J. Mol. Biol.*, **221**, 239.
[48] Harel, M., Schalk, I., Ehret-Sabattier, L., Bouet, F., Goeldner, M., Hirth, C., Axelsen, P., Silman, I. and Sussman, J. (1993). *Proc. Natl. Acad. Sci., USA*, **90**, 9031.
[49] Klebe, G. (1994). *J. Mol. Biol.*, **237**, 212.