Acta Cryst. (1995). D**51**, 418–427

Intermolecular Interactions Around Functional Groups in Crystals: Data for Modeling the Binding of Drugs to Biological Macromolecules

BY JENNY P. GLUSKER

The Institute for Cancer Research, The Fox Chase Cancer Center, Philadelphia, PA 19111 *USA*

(*Received* 20 *December* 1994; *accepted* 6 *March* 1995)

Abstract

When a small biologically active molecule enters the body it has the possibility of interacting with one or many of the large variety of macromolecules that are present. This interaction involves a unique complementary fit between the two that depends not only on the shape, but also on the distribution of charges on the surfaces of both [Fischer (1894). *Ber. Dtsch. Chem. Ges.* **27**, 2984–2993]. This recognition between two molecules may elicit a biological response and such processes are the subject of biochemical investigations. The question asked here is: what geometrical information on distances and relative orientations of interacting non-bonded functional groups can be found from X-ray crystallographic investigations?

If X-ray or NMR evidence is not available on how a particular macromolecule interacts with a specific small molecule, it may be necessary to resort to model building in order to obtain this information. A knowledge of the most likely distances between functional groups in different molecules and their most probable relative orientations is essential to a successful outcome of this model building. The necessary geometric data can be extracted from the results of X-ray diffraction studies of crystalline materials. Molecules pack in crystals in a way that is not random, but is governed by the same rules that apply when a molecule binds to its macromolecular receptor. Once numbers are available that describe the appropriate intermolecular forces as a function of distance and angle around a target atom, these data may be extended for use in macromolecular complexes. Energy-minimization techniques may then be used to establish and refine the structural data on probable binding modes of drugs with their macromolecular receptors.

The required metric data on intermolecular interactions are available for compounds containing only C and H atoms (Kitaigorodsky, 1961; Williams, 1970, 1972). The necessary parameters for crystalline hydrocarbons have been derived from experimental data, and have been used to predict and analyze the unit-cell dimensions and possible packing modes of some new compounds (Gavezzotti, 1991). When, however, the molecule contains O or N atoms in addition to C and H atoms, the expression for the energy and the derivation of the required parameters become more complicated, particularly in view of the possibility of hydrogen bonding (Umeyama & Morokuma, 1977; Allinger, 1982; Allinger, Yuh & Lii, 1989; Jeffrey & Saenger, 1991). It is necessary to take all possible interactions around a molecule into consideration when deriving low-energy packing arrangements.

When molecules pack in the crystalline state so that the energy of the total arrangement is minimized, functional groups on these molecules have probably packed in a preferential manner that accommodates their interaction requirements. The directionalities of the binding of functional groups to each other is implicit in the reported atomic coordinates in crystal structure determinations, provided these are preserved as functions of the unit-cell dimensions so that interactions between molecules in different unit cells (or even asymmetric units) can be examined (Glusker, 1985). Since, however, there may be many intermolecular interactions around one molecule in a crystal structure, and some of the interactions may be distorted in order to minimize the total energy of the entire crystal, it is possible that not all measured interactions in a crystal structure have the geometry expected for the most stable interactions. Therefore, it is necessary to analyze the experimentally observed interactions on a statistical basis, taking results from a large number of crystal structure determinations. Such a statistical survey, simplified by the establishment of databases of results of crystal structure determinations, will give an approximate measure of the directional preferences of binding of functional groups. These geometric results on orientational preferences for binding around a selected functional group are the subject of this review.

Intermolecular packing distances and directional preferences of binding of functional groups pertain to biological systems as well as to small-molecule systems. The pharmacophore has been defined by those interested in drug–receptor interactions as the three-dimensional arrangement of functional groups on a drug (small molecule) that is essential for both recognition (by binding) and activation (to give biological activity) of the biological macromolecule. Thus, the pharmacophore is the significant part of a drug involved in binding to the biological receptor in a way that leads to pharmacological activity. The data necessary for a complete

description of a drug–receptor complex are not available unless a three-dimensional determination of the structure of the drug bound to its macromolecular receptor has been reported. On the other hand, a molecule and its interaction probability function can be docked (that is, fit by computer graphics) into the structure of a protein in order to find areas in it that the molecule is most likely to bind (Olson & Christofferson, 1979). The ultimate aim of all such studies is to develop rules that describe the three-dimensional binding characteristics of functional groups. From these rules one may proceed to design drugs, artificial receptors and catalysts.

These geometrical analyses of directional preferences of binding cannot be used to derive the energies of the interactions; these energies need to be derived in some other way (Bürgi & Dunitz, 1988). The geometric data can, on the other hand, be used to derive force fields which can then be used to predict the mode of binding of a drug to a biological macromolecule (Vedani & Dunitz, 1985; Vedani & Huhta, 1990; Klebe & Diederich, 1993; Klebe, 1994). For example, the computer program *GRID* can be used to calculate interactions between a protein molecule (for example) and a chosen probe (the drug) at regularly spaced grid points by use of an appropriately derived potential function (Boobbyer, Goodford, McWhinnie & Wade, 1989; Wade, Clark & Goodford, 1993; Wade & Goodford, 1993). The derived force fields can also be used in attempts to predict crystal structures (Karfunkel & Leusen, 1992; van Eijk, Mooij & Kroon, 1995). Alternatively, the docking of drugs into proteins with pockets of known geometries has been used in various ways for the *de novo* design of protein ligands (Kuntz, Blaney, Oatley, Langridge & Ferrin, 1982; Danziger & Dean, 1989; Moon & Howe, 1991; Bohacek & McMartin, 1992; Böhm, 1992; Klebe, 1994).

Data for such analyses are now readily available. There are several databases of the results of crystal structure determinations that can be used for such statistical analyses of directional preferences of binding. Each of these provide information that is important for modeling specific drug–receptor binding. The Cambridge Structural Database has been developed over the years in Cambridge, England (Allen *et al.*, 1979). It now contains results from over 120 000 crystal structure determinations. The information in the Cambridge Structural Database is divided into numerical data (the unit-cell dimensions and atomic parameters), data representing the atomic connectivity (which atoms are directly bonded to each other), and the bibliography (journal reference). Searches for groups of atoms of specified atomic types and connectivity are extracted from the database by one of the programs in the system (*QUEST*). The results of such a search are stored as a file which can then be subjected, by further input information, to a more specific geometrical analysis. The Inorganic Structural Database (Bergerhoff, Hundt, Sievers & Brown, 1983) contains analogous information on the unit-cell dimensions, space group and atomic positions in crystals of inorganic compounds or complexes. The measured geometries of interactions between groups in proteins and in protein ligands also need to be included in these analyses (Singh & Thornton, 1992). The Protein Data Bank (Bernstein *et al.*, 1977) contains data on protein structure and nucleic acid structures. This includes information on the resolution of the reported structure, its unit-cell dimensions and three-dimensional atomic positions.

The location of an interacting atom of interest (for example, a metal ion) with respect to a chosen group (for example, a carboxylate group) can be compared by superimposing the chosen group from each reported crystal structure and producing a scatterplot of where the interacting atoms lie. Such scatterplots, however, are difficult to compare objectively. They may, therefore, be converted into continuous functions by convoluting each point-like scatterplot entry with a diffuse probability density; this is a 'smearing function' with a Gaussian-like form. The resulting three-dimensional function is contoured in the same way that an electron-density map is contoured (Rosenfield, Swanson, Meyer, Carrell & Murray-Rust, 1984). The development of such pseudo-density contours from three-dimensional scatterplots was necessary in order to view the major locations of scatterplots while, at the same time, eliminating the artifacts that occur because the eye is overly sensitive to outliers. Such pseudo-density maps are very useful for displaying differences between distributions of scatterplots around different functional groups.

Special programs have been designed for analyzing the potential symmetry or other properties of the scatterplots. For example, the scatter of points around a carboxyl group may be reflected across the two planes of symmetry that are found in the carboxylate ion itself. In other words, a metal cation can bind either O atom (or both in some cases) and the differentiation between the two O atoms may have no significance. These symmetries were applied to the scatterplots, when appropriate. Outliers in the probability plots need careful examination because their existence may turn out to be significant when their structures are investigated in detail, and they may invalidate the current interpretation of the data.

The first such analysis of directional preferences of binding around a functional group was carried out in the laboratory of Parthasarathy in 1977 (Rosenfield, Parthasarathy & Dunitz, 1977). As a result of this study it was found possible to classify the manner in which functional groups cluster around a C—S—C group into approximately two main types, as shown in Fig. 1. Electrophiles such as metal cations or hydrogen-bond donors approach in a direction that is nearly perpendicular to the sulfide (C—S—C) plane. Presumably, these electrophiles are interacting with the highest

occupied molecular orbital of the S atom; this is a lone-pair orbital nearly perpendicular to the C—S—C plane, and it can donate electrons to the electrophile. On the other hand, nucleophiles, such as negatively charged groups or ions, approach the S atom approximately in the C—S—C plane, along the extension of one of the C—S bonds, the direction predicted for the lowest unoccupied molecular orbital, which can accept electrons from a nucleophile. Therefore, electrophilic attack of divalent sulfur would be expected to occur in a direction perpendicular to the C—S—C plane, while nucleophilic attack would be expected to occur in the direction of an S—C bond. This work was then extended to S···S contacts, and it was shown that these could generally be considered as electrophile–nucleophile pairs (Guru Row & Parthasarathy, 1981). If one of the S atoms acts as an electrophile, it will approach the other which then acts as a nucleophile.

Since that time there have been several analyses using the Cambridge Structural Database (Britton & Dunitz, 1980; Taylor & Kennard, 1983*a*,*b*, 1984; Taylor, Kennard & Versichel, 1983, 1984*a*,*b*; Bürgi & Dunitz, 1983; Pirard, Baudoux & Durant, 1995). For example, the directions from which hydrogen bonds approach the functional O atoms in ethers, ketones (including cyclic ketones), lactones (cyclic esters), and epoxides have been investigated (Murray-Rust & Glusker, 1984), and the results for each of these types of functional O atoms have been compared with each other. In order to ensure high precision in the information used, only structure determinations that contain no atom heavier than potassium were selected. Crystal structure results were also used only if they contained H-atom positions, had an R value less than 0.10 and contained no atomic disorder. Scatterplots of the locations of hydrogen-bonding groups around a selected functional O atom were made. The number of entries in the database that conformed to the input tests was often high; for example, hydrogen bonding to 286 ketone groups was studied in this way. It was found that the highest concentration of hydrogen-bonded O—H or N—H groups lie in the directions that are conventionally considered those of lone-pair directions of the O atoms, as illustrated in Fig. 2. In the scatterplots of hydrogen-bond locations around the O atom, the maxima for carbonyl O atoms in ketones, enones and esters were found to lie in the plane of the carbonyl group at about 120° to the C=O bond. This result was also found in analyses of protein structures (Baker & Hubbard, 1984). For the C—O—C O atom in epoxides and ethers, two maxima were found in a plane perpendicular to that of the C—O—C group. The locations of these hydrogen-bonding donor groups around these functional O atoms are analogous to those just described for electrophiles around the C—S—C bond. There is a narrower distribution of hydrogen-bond donors to ether O atoms than to the O atoms in ketones or epoxides. The reason for this is still being analyzed. The effects of two, rather than one hydrogen-bonded group approaching the O atom, and also the importance of bulky substituents close to the donor or acceptor of the hydrogen bond are also still being investigated by us.

Of significance in protein interactions is the relative orientation of metal ions to those functional groups

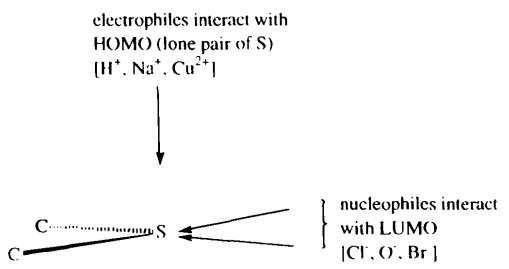


Fig. 1. Directional preferences of non-covalent binding to divalent sulfur in a C—S—C group. Electrophiles approach at an angle of 50–90° from the C—S—C plane, while nucleophiles interact within 30° of the C—S—C plane.

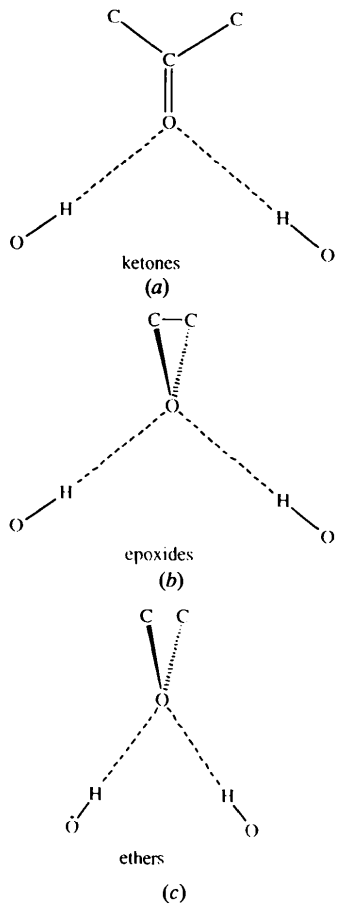


Fig. 2. Surroundings of functional groups in (*a*) ketones, (*b*) epoxides and (*c*) ethers. Note the narrower distribution of hydrogen-bonding positions around ether O atoms.

that bind them, that is, mainly carboxylate, imidazole, and sulfhydryl groups. Carboxylate groups (aspartate or glutamate, for example) each have one delocalized negative charge; each carboxylate O atom has two lone-pair electrons disposed at 120° to the C—$O^{1/2-}$ bond. The geometry of the interactions of a carboxyl group has been investigated by Julius Rebek (Rebek, 1990). He designed and studied several compounds in which he was able to control the directions in which carboxylic acid groups are forced to approach each other. O atoms of the carboxylate groups each contain two lone-pair electrons, designated *syn* and *anti* (*Z* and *E* forms, respectively) and shown in Fig. 3. The proton on another molecule can approach in either of these directions. In the *syn* conformation the proton is on the same side of the C—O bond as the other C—O bond, and in the *anti* conformation it is on the opposite side. For example, the *syn* conformation is found when carboxyl groups dimerize by forming two hydrogen bonds. *Ab initio* quantum chemical studies of formic acid indicate that *syn* lone-pair electrons are more basic than those of the *anti* lone pair (Peterson & Csizmadia, 1979). It was estimated that *syn* protonation is much more favorable than *anti* protonation. The carboxylate group is, therefore, a weaker base when constrained to accept a proton in the *anti* position than when it is in the *syn* position. Thus, carboxylates in the active sites of enzymes generally employ the more basic *syn* lone-pair electrons for metal chelation rather than the less basic *anti* lone-pair electrons (Gandour, 1981).

An analysis of the directions in which metal ions approach a carboxyl group in crystal structures showed that the most likely arrangements of metal cations are *syn*, *anti*, and a third category, 'direct' in which both O atoms of the carboxylate group share the metal ion equally (Carrell, Carrell, Erlebacher & Glusker, 1988). The major finding is that most metal ions bind to a carboxyl group in its plane. Exceptions to this are mainly the alkali metal ions which ionize readily and form strong bases. They have less specific directions of binding, and show extensive out-of-plane interactions with carboxyl groups. Direct bonding, in which the metal ion is in-plane and equidistant from the two O atoms of the carboxylate ion, is preferred in the range $M^{n+}\cdots O$ = 2.3–2.6 Å; otherwise the *syn* lone pair is generally preferred for metal-ion binding. Cations of the appropriate size and, therefore, with a reasonably high percentage of 'direct' binding include divalent calcium, cadmium and mercury, but not magnesium (Einspahr & Bugg, 1981). At $M^{n+}\cdots O$ distances other than the range 2.3–2.6 Å 'direct bonding' is rare. These results are a function of the carboxylate 'bite' size (2.2 Å) and of a need to keep $O\cdots M^{n+}\cdots O$ angles reasonably large (larger than approximately 60°). The overall values for the binding of metal ions to carboxylates were 63% *syn*, 23% *anti* and 14% direct. Exceptions are found at very short and very long $M^{n+}\cdots O$ distances, that is, for very small and very large cations. Such interactions have been analyzed in protein crystal structures (Chakrabarti, 1990*a*). When the carboxyl group is not ionized, it is found, from detailed neutron diffraction studies, that the hydroxyl portion of the carboxyl group, while a powerful hydrogen-bond donor, can only accept one hydrogen bond (Ramanadham, Jakkal & Chidambaram, 1993). This feature was used successfully to establish the ionization states of the active-site carboxyl groups in hen egg-white lysozyme (Ramanadham *et al.*, 1993).

The intermolecular hydrogen bonding of phenolic hydroxyl groups of the type found in tyrosine side chains and in estradiol was investigated by Prout and coworkers (Prout, Fail, Jones, Warner & Emmett, 1988). They found that the donor hydrogen bond does not have to be coplanar with the aromatic ring system but can deviate up to 40° from it. The acceptor hydrogen bond can show a much larger deviation from the plane of the aromatic ring, an effect that is further increased if there is *ortho* substitution in the aromatic ring system.

Imidazole groups in histidyl side chains in proteins also bind to metal ions, and the metal ion nearly always

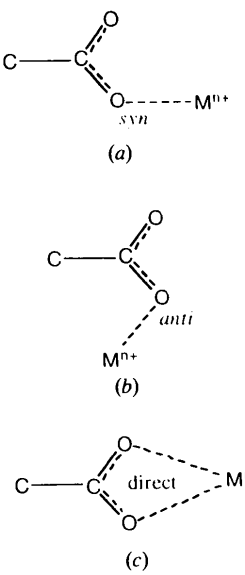


Fig. 3. Binding of metal ions to carboxylate ions. The *syn* and *anti* lone pairs are indicated in (*a*) and (*b*), respectively. When both O atoms of the carboxylate group are shared by the metal ion, the arrangement is described here as 'direct'.

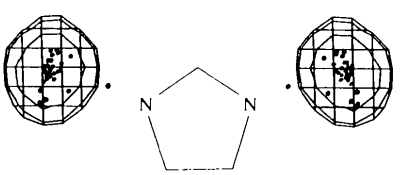
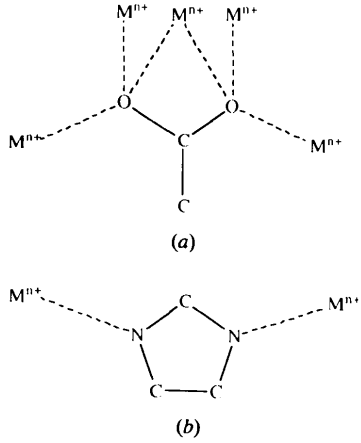


Fig. 4. Scatterplot of sites of binding of metal ions to an imidazole group in small-molecule crystal structures.

lies in or very near the plane of the imidazole group (Fig. 4) (Chakrabarti, 1990*b*; Carrell *et al.*, 1993). One imidazole can bind one or two metal ions, depending on the ionization state and suitability of the metal ion (Zn^{2+} and Cu^{2+} being common binders). Metal ions are found to lie in the plane of the imidazole group even if the oxidation state or coordination number of the metal ion changes. In proteins, the metal ion shows a preference for binding to the N atom furthest from the Cα of the histidine. If the metal binds to the other N atom, the histidine is rotated from the value normally found in proteins. A hydrogen bond to the N atom that is not coordinated by metal serves to stabilize the orientation of the ring. Rotation of the imidazole ring about the $M^{n+}\cdots$N(imidazole) bond is controlled by the steric requirements of the rest of the histidyl residue, and by any hydrogen bonding to other amino-acid side chains such as an adjacent aspartate, as found in D-xylose isomerase, for example. Metal ion–sulfur interactions in proteins have been analyzed in a similar way (Chakrabarti, 1989).

The statistical surveys described so far provide a mechanism for searching for metal-binding positions in proteins. Use is made of the experimental observations that there is a high probability that the metal ion lies in the plane of a neighboring carboxyl or histidyl group. This test is complementary to that suggested by Eisenberg and coworkers (Yamashita, Wesson, Eisenman & Eisenberg, 1990), which probes for the existence of a hydrophilic area in the protein (O, N or S atoms) with a hydrophobic area immediately behind it (C atoms of the carboxylate group). This means that metals bind at sites where the change between hydrophobicity and hydrophilicity is rapid.



(*a*)

(*b*)

Fig. 5. Probes of metal-binding sites in proteins. Shown are (*a*) the surroundings of a carboxylate group, and (*b*) the surroundings of an imidazole group, in-plane in both cases. The coordinates of these probes were docked onto appropriate side chains in the proteins in order to check for sites where several such groups bind to the same metal ion.

In our analysis, probes consisting of a carboxylate or histidine group with the most probable metal-ion positions around them with the geometry shown in Fig. 5 were constructed on a computer graphics system. The carboxylate or histidine groups in the probes were then docked onto these groups on side chains of protein crystal structures (Carrell *et al.*, 1988). This method was successfully used for locating metal-binding sites on the enzyme D-xylose isomerase from *Streptomyces rubiginosus* which binds divalent magnesium, manganese and/or cobalt (Carrell, Rubin, Hurley & Glusker, 1984; Carrell *et al.*, 1989). Two different metal-ion sites, which could not be determined unambiguously from the electron-density map, were successfully located on the enzyme by use of the probes shown in Fig. 5. These metal-ion positions were, by this method using a structural probe, clearly identified on a computer graphics system (Carrell *et al.*, 1988). In the active site of this enzyme, shown in Fig. 6, one metal ion binds three carboxylate groups (one direct, each utilizing *syn* lone-pair electrons) while the other metal ion binds one carboxylate by way of a *syn* lone-pair of electrons and three by way of the *anti* lone pair. In this way the enzyme can control the electronic charge environment in the active site. The metal ion bound to *anti* as well as *syn* lone pairs presumably has a somewhat higher residual positive charge than does the other, for which the charge is more effectively neutralized. The former with *anti* lone pair binding is, indeed, the one that binds substrates.

Specific patterns of groups in metal binding can be identified in crystal structures of small molecules and of biological macromolecules. For example, in the crystal structures of cobalt, magnesium, manganese and other formates, the metal ion binds one O atom of the carboxyl group and a water molecule which forms a hydrogen bond to the other O atom of the carboxyl group, as shown in Fig. 7. This grouping is approximately planar

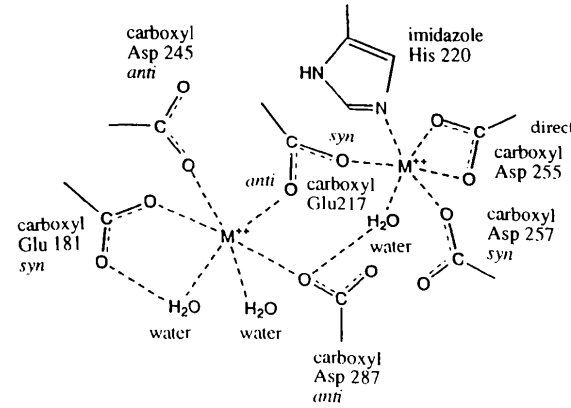


Fig. 6. Carboxylate and imidazole groups and water binding to metal ions in the active site of the enzyme D-xylose isomerase. The lone-pair electrons that bind the metal ion are designated *syn* and *anti*.

(Kaufman, Afshar, Rossi, Zacharias & Glusker, 1993). It also occurs twice in the binding of groups in the active site of D-xylose isomerase. Its role in the mechanism of action is still under investigation. Presumably the carboxyl group may modify any effect of the metal ion on the pK of the bound water molecule.

If a molecule or drug has many functional groups it may be able to bind to biological macromolecules in more than one way. Generally only one of these binding modes is biologically relevant. This cautions us to be judicious in our interpretation of binding data. A possible example of this is the binding of citrate to a positively charged group (a metal ion or the N atom of lysine, arginine or histidine) in the enzyme aconitase (Glusker, 1968). If citrate binds to the enzyme aconitase *via* its α-hydroxycarboxylate group, as shown in Fig. 8, an H atom on one terminal methylene group can be extracted by a functional group on the enzyme. If, however, the citrate binds in the other manner, the functional group is too far from any methylene group to be active as a base. Binding of citrate analogs such as fluorocitrate suggests that citrate can bind in both ways, but it is only enzymatically isomerized in one of these two orientations. Similarly, the substrates of D-xylose isomerase are sugars with many hydroxyl groups. In order to investigate details of the mechanism of action of this enzyme, we studied the modes of binding of a series of substrates, inhibitors,

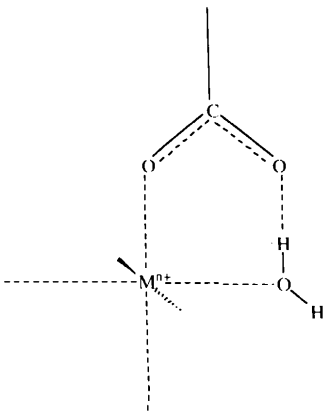


Fig. 7. Motif of carboxylate and water binding to a metal ion. This motif is found in metal salts such as formates, and in proteins (see Fig. 6).

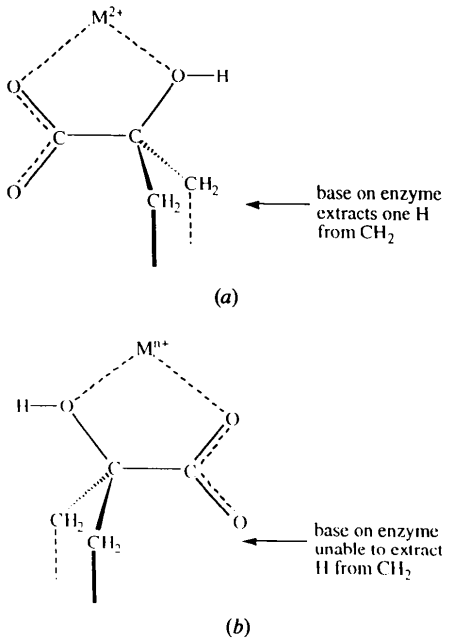


Fig. 8. Two directions of binding of citrate to a cation. (*a*) productive binding and (*b*) non-productive binding. The central α-hydroxycarboxylate group can bind with equal probability in the two ways shown. But in only one of these orientations of binding can an active-site base extract a H atom from a CH_2 group. In the other orientation the base on the enzyme is too far from a H atom to be able to extract it.

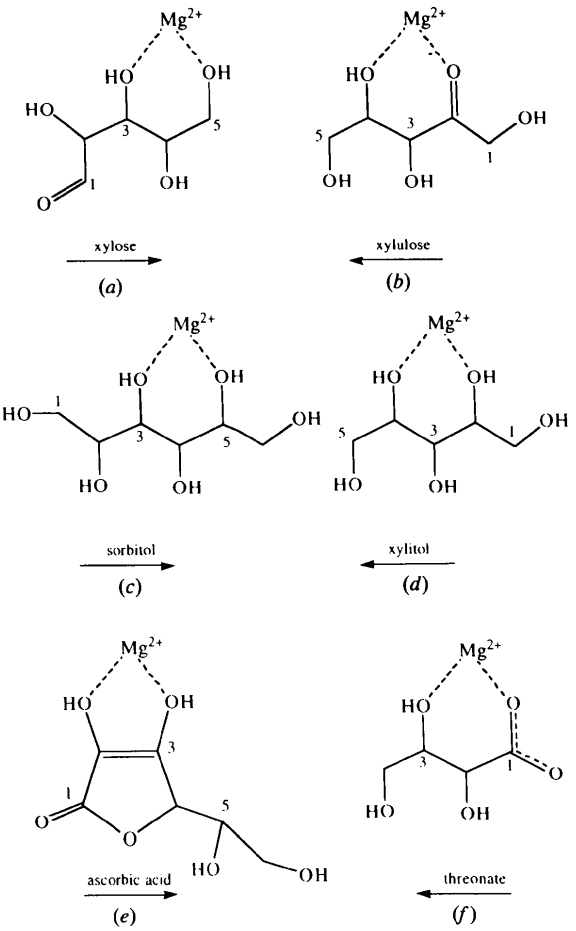


Fig. 9. Two directions of binding of sugars to one of the metal ions in the enzyme D-xylose isomerase. (*a*) Xylose, (*b*) xylulose, (*c*) sorbitol, (*d*) xylitol, (*e*) ascorbic acid and (*f*) threonate. His54 is on the left-hand side of each diagram. The enzyme acts on C1 and C2 of xylose. Note that half of these analogs bind in one direction and half in the opposite direction (indicated by arrows pointing from C1 to C5).

and their analogues (Carrell, Hoier & Glusker, 1994). It appears that some molecules bind to the active site of the enzyme in one direction, while some bind in the alternate direction, as shown in Fig. 9. Thus, if active sites are relatively open and flexible, the substrate may bind in more than one way, although presumably only one binding mode is catalytically competent.

These types of analyses of intergroup orientations suggest that weaker interactions, such as those involving C—H and C—F groups can also be significant in aligning molecules in macromolecular systems (Taylor & Kennard, 1982; Burley & Petsko, 1988; Desiraju, 1989). Such interactions need to be taken into account in molecular modelling and energy minimizations. The interaction between a C—H group and a neighboring O atom can be considered as a very weak hydrogen bond, about one third the energy of an O—H···O hydrogen bond (Sutor, 1963). If, however, there are many such interactions around one molecule, they can have a significant effect on the total energy of an interacting system. For example, the importance of C—H···O interactions in aligning molecules has been shown by the design, by Desiraju and coworkers, of benzylidene ketones that interact by way of C—H···O interactions with 1,3,5-trinitrobenzene (Fig. 10) (Sharma *et al.*, 1994). Use was made of the great acidity of the H atoms in 1,3,5-trinitrobenzene, and the propensity of nitro groups to interact with H atoms bound to C atoms. It was shown by X-ray crystallographic studies that the C—H···O interactions have a significant aligning capability, and that there is a similarity of binding of dibenzylidene acetone, cyclopentanone, and cyclohexanone. These complexes are analogous to the designed complexes containing O—H···O and N—H···O hydrogen bonds, but in the complex described here, only C—H···O interactions could be formed. Investigations of these weak interactions improve our ability to identify all the forces contributing to the binding of a drug to its receptor.

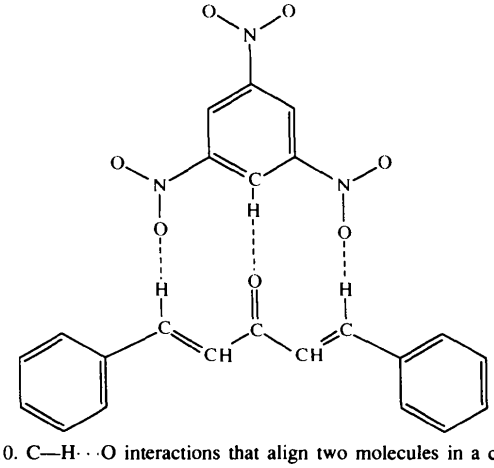


Fig. 10. C—H···O interactions that align two molecules in a complex *via* the nitro and carbonyl groups.

F atoms are often incorporated into drugs and enzyme inhibitors. For this reason intermolecular interactions of a C—F group have also been analyzed in this way (Murray-Rust, Stallings, Monti, Preston & Glusker, 1983). The C—F bond is often considered rather inert, like the C—H bond, a viewpoint that is reinforced by the inertness of Teflon that is used in 'non-stick' frying pans. An analysis of relevant crystal structures in the Cambridge Structural Database shows that F atoms in C—F bonds approach hydroxyl groups with minimum F···H distances of 2.3 Å, much longer than O···H distances of 1.4–1.7 Å in strong O···H—O hydrogen bonds. This suggests that, at best, hydrogen bonds involving C—F groups are very weak. There was no evidence of a preferential angle of approach to C—F groups, unlike the anisotropy observed for C—Cl, C—Br and C—I interactions in a similar type of analysis (Cody & Murray-Rust, 1984; Ramasubbu, Parthasarathy & Murray-Rust, 1986). Electrophiles were found to approach C—Cl, C—Br and C—I bonds at an angle of about 100° (nearly perpendicularly), while nucleophiles approach at about 165° (almost directly).

Fluorine is the most electronegative element and it is often incorporated into drugs and enzyme inhibitors. Why, then, does the C—F bond form hydrogen bonds so rarely, especially since H···F^- forms very short measured hydrogen bonds (H···F^- = 1.15 Å)? The strength of a hydrogen bond (*D*—H···*A*) appears to be a complicated balance of various factors which include the interactions between the fractional charges that have developed on *D*, *A* and H, the polarizability of the electron cloud around the acceptor atom *A* so that it can make its lone pairs available to the proton (the softness of *A*), the transfer of charge from *A* to H (the stabilization as a result of mixing of ground state *A*···H with the excited charge-separated A^-···H^+), the electronic repulsion between *D* and *A*, how readily the hydrogen-bond donor atom *D* will lose its covalently bound H atom as H^+ (related to the electronegativity of *D* and the strength of the *D*—H bond), and how readily the hydrogen-bonded acceptor *A* can accept the H^+ (the electronegativity of *A*). Of these the important attractive forces are electrostatic, polarization effects and charge-transfer interactions. The electrostatic component falls off less rapidly as a function of distance than the others do, and, therefore, at longer distances it is generally the most important. These energy components are balanced by repulsive forces which become important at short distances (Umeyama & Morokuma, 1977; Hunter, 1994). Thus, the H atom in a C—H bond is held firmly by the C atom so that the C—H group does not readily form a hydrogen bond, and when it does it is weak with a relatively long H···*A* distance. The F atom in a C—F group does not, apparently, develop much negative charge, and in addition, it is not readily polarized. If, however, there are strong hydrogen-bond donors available (—OH or —NH, for example),

the C—F bond will form weak interactions with C—H groups, presumably because of slight negative charges on the F atom and slight positive charges on the H atom (Shimoni, Carrell, Glusker & Coombs, 1994). While fluorine is more electronegative than oxygen, oxygen is somewhat more polarizable. Therefore, the probability of O···H—O hydrogen bond formation is significantly higher than that for the formation of a hydrogen bond involving a C—F group. This is exemplified by one of the crystal structures in which the C—F group takes part in hydrogen bonding which involves a three-centered (bifurcated) hydrogen bond, with the H atom shared between an O atom and an F atom, but is closer to the O atom (Fig. 11) (Murray-Rust *et al.*, 1983). Thus, it appears that the C—F group only forms very weak hydrogen bonds (as the hydrogen-bond acceptor). If there are many hydroxyl groups in the structure, the F atom in the C—F can take part in hydrogen bonding only if a bifurcated (three-center) bond is formed. This is because a hydroxyl group can act both as a proton donor or acceptor so that hydrogen bonding can be cooperative, while an F atom attached to a C atom can act only as a proton acceptor. On the other hand, protonated amino groups, which can act only as proton donors, do form N—H···F—C interactions, but only weakly.

Chelating groups that bind metal ions are found in iron-binding hydroxamates, some of which are powerful enough to extract iron from stainless steel. Other examples of metal-chelating groups are provided by α-hydroxycarboxylate groups, HO—CR_2—COO^-, found in many important biochemical compounds such as citrates and malates (Fig. 12). The geometry of their metal binding has been investigated by X-ray crystallographic techniques. If the cation is of a suitable size, the entire chelating group is approximately planar, even though the O···O distance is shorter than the sum of the van der Waals radii for O atoms. It appears that the α-hydroxycarboxylate group requires either a cation or H (atom or ion) chelated to it. A study of one crystal form of potassium citrate revealed, surprisingly, that the α-hydroxycarboxylate group in the citrate ion did not chelate the potassium ion. On the other hand, the citrate hydroxyl H atom formed an internal hydrogen bond within the chelating group (Fig. 12). This was verified by a neutron diffraction study (Carrell *et al.*, 1987). When the central hydroxyl group of citric acid is replaced by fluorine, the α-fluorocarboxylic acid group is still a good chelating group, but no H atom is available to form an internal hydrogen bond. As a result, in the crystal structure of dipotassium 3-fluorodeoxycitrate, a potassium cation is chelated by the α-fluorocarboxylate group. It appears that an α-hydroxycarboxylate group will either chelate a cation or form an internal hydrogen bond *via* the hydroxyl H atom, with the H atom acting as a substitute for the metal ion.

An interaction between a C—F bond and a metal ion (C—F···M^{n+}) has been observed so far by X-ray diffraction studies only when the metal ion is an alkali metal ion (Murray-Rust *et al.*, 1983). In these cases the M^{n+}···F distance is of the same order as the M^{n+}···O distances. The coordination of carbon-bound O and F atoms around the metal ion can be analyzed in terms of the requirement that the sum of the bond valences should be near 1.0 for a monovalent cation (Brown, 1978). This sum is achieved only if the M^{n+}···F—C interaction is taken into account as well as those involving O atoms. The F atom contributes to the local neutralization of the charge of the cation, and a significant member of the first coordination sphere of the monovalent metal ion.

Intermolecular interactions need to be analyzed in such a manner that can be input to any molecular modelling program. This area of study is still in the early stages and much useful information can be expected, as indicated by this preliminary review.

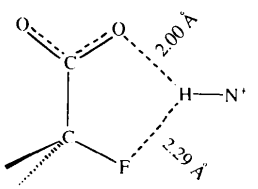


Fig. 11. Sharing of a H atom of a protonated amine between the O and F atoms of an α-fluorocarboxylate. Note that the H···F distance is longer than the H···O distance.

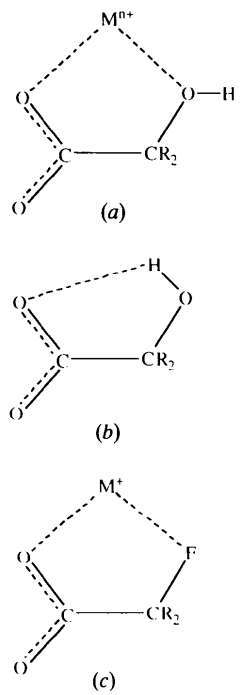


Fig. 12. Binding of an α-hydroxycarboxylate group (*a*) in the presence of a metal ion, (*b*) in the absence of a metal ion (giving an internal hydrogen bond). (*c*) Metal chelation in an α-fluorocarboxylate.

Some rules that can be used for modelling drug–receptor interactions then follow from these types of structural studies of intermolecular interactions.

(*a*) In protein–drug interactions hydrogen bonds, which are generally fairly linear, are important. Main-chain NH groups readily form such hydrogen bonds. The linearity of hydrogen bonds may, however, be perturbed by other interactions such as C—H···O and C—F···H—N, even though these are weak.

(*b*) Main-chain carbonyl groups will form one or two hydrogen bonds at 120° to the C=O axis. This rule also applies to similar functional groups on the drug molecule.

(*c*) Ether and epoxide C—O—C groups tend to form hydrogen bonds in a direction nearly perpendicular to the plane of the group. This means that an epoxide group is a good binding analog of a carbonyl group.

(*d*) C—S—C groups tend to bind positively charged groups (electrophiles) in a direction nearly perpendicular to the plane of this group, while they bind negatively charged groups (nucleophiles) along one of the C—S bonds.

(*e*) Carboxylate groups generally bind metal ions in the plane of the group. They bind to the lone-pair electrons of the carboxylate O atoms which are either *syn* or *anti*. The *syn* lone pairs, if available, seem to be preferred. Cations of a specific size, such as divalent calcium, will share both O atoms of the carboxylate group. Alkali metal cations tend to bind indiscriminately in-plane and out-of-plane.

(*f*) Carboxyl groups show different hydrogen-bonding patterns depending on whether they are ionized or not.

(*g*) Imidazoles such as histidine in proteins bind metal ions in their planes.

(*h*) Phenolic hydroxyl groups such as are found in tyrosine tend to donate hydrogen bonds within 40° of the plane of the aromatic group.

(*i*) Chelating groups such as α-hydroxycarboxylates tend to bind metal ions in their planes.

(*j*) Weak interactions such as C—H···O tend to be formed when the H atom has a slight positive charge, the O atom a slight negative charge and there are no competing stronger hydrogen bonds possible. These interactions serve to align molecules in the more hydrophobic areas of macromolecules.

(*k*) C—F groups also tend to form only weak interactions, a slight negative charge on the F atom interacting with the slight positive charge of the H atom of a hydroxyl or amine group of another molecule. These C—F···H—(O,N) interactions, like C—H···O interactions, compete unfavorably with stronger hydrogen-bonding groups, but, in spite of this, their overall effect may be significant.

This work was supported by grants CA-10925 (to JPG) and CA-06927 (to ICR) from the National Institutes of Health, and by an appropriation from the Commonwealth of Pennsylvania.

References

ALLEN, F. H., BELLARD, S., BRICE, M. D., CARTWRIGHT, B. A., DOUBLEDAY, A., HIGGS, H., HUMMELINK, T., HUMMELINK-PETERS, B. G., KENNARD, O., MOTHERWELL, W. D. S., RODGERS, J. R. & WATSON, D. G. (1979). *Acta Cryst.* **B35**, 2331–2339.

ALLINGER, N. L. (1982). *Molecular Mechanics*. ACS Monograph 177. Washington, DC: American Chemical Society.

ALLINGER, N. L., YUH, Y. H. & LII, J.-H. (1989). *J. Am. Chem. Soc.* **222**, 8551-8566.

BAKER, E. N. & HUBBARD, R. E. (1984). *Prog. Biophys. Mol. Biol.* **44**, 97–179.

BERGERHOFF, G., HUNDT, R., SIEVERS, R. & BROWN, I. D. (1983). *J. Chem. Info.* **23**, 66–69.

BERNSTEIN, F. C., KOETZLE, T. F., WILLIAMS, G. J. B., MEYER, E. F. JR, BRICE, M. D., RODGERS, J. R., KENNARD, O., SHIMANOUCHI, T. & TASUMI, M. (1977). *J. Mol. Biol.* **112**, 535–542.

BOHACEK, R. S. & MCMARTIN, C. (1992). *J. Med. Chem.* **35**, 1671–1684.

BÖHM, H. J. J. (1992). *Comput. Aided Mol. Design*, **6**, 593–606.

BOOBBYER, D. N. A., GOODFORD, P. J., MCWHINNIE, P. M. & WADE, R. C. (1989). *J. Med. Chem.* **32**, 1083–1094.

BRITTON, D. & DUNITZ, J. D. (1980). *Helv. Chim. Acta*, **63**, 1068–1073.

BROWN, I. D. (1978). *Chem. Soc. Rev.* **7**, 359–376.

BÜRGI, H.-B. & DUNITZ, J. D. (1983). *Acc. Chem. Res.* **16**, 153–161.

BÜRGI, H.-B. & DUNITZ, J. D. (1988). *Acta Cryst.* **B44**, 445–448.

BURLEY, S. K. & PETSKO, G. A. (1988). *Adv. Protein Chem.* **39**, 125–189.

CARRELL, C. J., CARRELL, H. L., ERLEBACHER, J. & GLUSKER, J. P. (1988). *J. Am. Chem. Soc.* **110**, 8651–8656.

CARRELL, H. L., GLUSKER, J. P., BURGER, V., MANFRE, F., TRITSCH, D. & BIELLMANN, J.-F. (1989). *Proc. Natl Acad. Sci. USA*, **86**, 4440–4444.

CARRELL, H. L., GLUSKER, J. P., PIERCY, E. A., STALLINGS, W. C., ZACHARIAS, D. E., DAVIS, R. L., ASTBURY, C. & KENNARD, C. H. L. (1987). *J. Am. Chem. Soc.* **109**, 8067–8071.

CARRELL, H. L., HOIER, H. & GLUSKER, J. P. (1994). *Acta Cryst.* **D50**, 113-123.

CARRELL, H. L., RUBIN, B. H., HURLEY, T. J. & GLUSKER, J. P. (1984). *J. Biol. Chem.* **259**, 3230–3236

CARRELL, A. B., SHIMONI, L., CARRELL, C. J., BOCK, C. W., MURRAY-RUST, P. & GLUSKER, J. P. (1993). *Receptor*, **3**, 57–76.

CHAKRABARTI, P. (1989). *Biochemistry*, **28**, 6081–6085.

CHAKRABARTI, P. (1990*a*). *Protein Eng.* **4**, 49–56.

CHAKRABARTI, P. (1990*b*). *Protein Eng.* **4**, 57–63.

CODY, V. & MURRAY-RUST, P. (1984). *J. Mol. Struct.* **112**, 189–199.

DANZIGER, D. G. & DEAN, P. M. (1989). *Proc. R. Soc. London Ser. B*, **236**, 115–124

DESIRAJU, G. R. (1989). *Crystal Engineering. The Design of Organic Solids*. Amsterdam: Elsevier.

VAN EIJCK, B. P., MOOIJ, W. T. M. & KROON, J. (1995). *Acta Cryst.* **B51**, 99–103.

EINSPAHR, H. & BUGG, C. E. (1981). *Acta Cryst.* **B37**, 1044–1052.

FISCHER, E. (1894). *Ber. Dtsch. Chem. Ges.* **27**, 2984–2993.

GANDOUR, R. D. (1981). *Bioorg. Chem.* **10**, 169–176.

GAVEZZOTTI, A. (1991). *J. Am. Chem. Soc.* **113**, 4622–4629.

GLUSKER, J. P. (1968). *J. Mol. Biol.* **38**, 149–162.

GLUSKER, J. P. (1985). *Acc. Chem. Res.* **18**, 95.

GURU ROW, T. N. & PARTHASARATHY, R. (1981). *J. Am. Chem. Soc.* **103**, 477–479.

HUNTER, C. A. (1994). *Chem. Soc. Rev.* pp. 101–109.

JEFFREY, G. A. & SAENGER, W. (1991). *Hydrogen Bonding in Biological Structures*. Berlin/New York: Springer-Verlag.

KARFUNKEL, H. R. & LEUSEN, F. J. J. (1992). *SPEEDUP*, **6**, 43–50.

KAUFMAN, A., AFSHAR, C., ROSSI, M., ZACHARIAS, D. E. & GLUSKER, J. P. (1993). *Struct. Chem.* **4**, 191–198.

KITAIGORODSKY, A. I. (1961). *Organic Chemical Crystallography*. (English translation). (Russian edition, 1955. Academy of Sciences Press: Moscow).

KLEBE, G. (1994). *J. Mol. Biol.* **237**, 212–235.

KLEBE, G. & DIEDERICH, F. (1993). *Philos. Trans. R. Soc. London Ser. A*, **345**, 37–48.
KUNTZ, I. D., BLANEY, J. M., OATLEY, S. J., LANGRIDGE, R. & FERRIN, T. E. (1982). *J. Mol. Biol.* **161**, 269–288.
MOON, J. B. & HOWE, W. J. (1991). *Proteins Struct. Funct. Genet.* **11**, 314–328.
MURRAY-RUST, P. & GLUSKER, J. P. (1984). *J. Am. Chem. Soc.* **106**, 1018–1025.
MURRAY-RUST, P., STALLINGS, W. C., MONTI, C. T., PRESTON, R. M. & GLUSKER, J. P. (1983). *J. Am. Chem. Soc.* **105**, 3026–3214.
OLSON, E. C. & CHRISTOFFERSON, R. E. (1979). Editors. *Computer-Assisted Drug Design*. ACS Symposium Series 112.
PETERSON, M. R. & CSIZMADIA, I. G. (1979). *J. Am. Chem. Soc.* **101**, 1076–1079.
PIRARD, B., BAUDOUX, G. & DURANT, F. (1995). *Acta Cryst.* B**51**, 103–107.
PROUT, K., FAIL, J., JONES, R. M., WARNER, R. E. & EMMETT, J. C. (1988). *J. Chem. Soc. Perkin Trans.* II, pp. 265–284.
RAMANADHAM, M., JAKKAL, V. S. & CHIDAMBARAM, R. (1993). *FEBS Lett.* **323**, 203–206.
RAMASUBBU, N., PARTHASARATHY, R. & MURRAY-RUST, P. (1986). *J. Am. Chem. Soc.* **108**, 4308–4314.
REBEK, J. JR (1990). *Angew. Chem. Int. Ed. Engl.* **29**, 245–255
ROSENFIELD, R. E. JR, PARTHASARATHY, R. & DUNITZ, J. D. (1977). *J. Am. Chem. Soc.* **99**, 4860–4862.
ROSENFIELD, R. E. JR, SWANSON, S. M., MEYER, E. F. JR, CARRELL, H. L. & MURRAY-RUST, P. (1984). *J. Mol. Graphics*, **2**, 43–46.
SHARMA, C. V. K., PANNEERSELVAM, K., SHIMONI, L., KATZ, H., CARRELL, H. L. & DESIRAJU, G. R. (1994). *Chem. Mater.* **6**, 1282–1292.
SHIMONI, L., CARRELL, H. L., GLUSKER, J. P. & COOMBS, M. M. (1994). *J. Am. Chem. Soc.* **116**, 8162–8168.
SINGH, J. R. & THORTON, J. M. (1992). *Atlas of Protein Side-Chain Interactions*. Oxford Univ. Press.
SUTOR, D. J. (1963). *J. Chem. Soc.* pp. 1105–1110.
TAYLOR, R. & KENNARD, O. (1982). *J. Am. Chem. Soc.* **104**, 5063–5070.
TAYLOR, R. & KENNARD, O. (1983*a*). *Acta Cryst.* B**39**, 133–138.
TAYLOR, R. & KENNARD, O. (1983*b*). *Acta Cryst.* B**39**, 517–525.
TAYLOR, R., KENNARD, O. & VERSICHEL, W. (1983). *J. Am. Chem. Soc.* **105**, 5761–5766.
TAYLOR, R., KENNARD, O. & VERSICHEL, W. (1984*a*). *J. Am. Chem. Soc.* **106**, 244–248.
TAYLOR, R., KENNARD, O. & VERSICHEL, W. (1984*b*). *Acta Cryst.* B**40**, 280–288.
UMEYAMA, H. & MOROKUMA, K. (1977). *J. Am. Chem. Soc.* **99**, 1316–1332.
VEDANI, A. & DUNITZ, J. D. (1985). *J. Am. Chem. Soc.* **107**, 7653–7658.
VEDANI, A. & HUHTA, D. W. (1990). *J. Am. Chem. Soc.* **112**, 4759–4767.
WADE, R. C., CLARK, K. J. & GOODFORD, P. J. (1993). *J. Med. Chem.* **36**, 140–147.
WADE, R. C. & GOODFORD, P. J. (1993). *J. Med. Chem.* **36**, 148–156.
WILLIAMS, D. E. (1970). *Trans. Am. Crystallogr. Assoc.* **6**, 21–33.
WILLIAMS, D. E. (1972). *Acta Cryst.* A**28**, 629–635.
YAMASHITA, M., WESSON, L., EISENMAN, G. & EISENBERG, D. (1990). *Proc. Natl Acad. Sci. USA*, **87**, 5648–5652.