Applications of the Cambridge Structural Database in organic chemistry and crystal chemistry

**Frank H. Allen* and
W. D. Samuel Motherwell**

Cambridge Crystallographic Data Centre (CCDC), 12 Union Road, Cambridge CB2 1EZ, England

Correspondence e-mail: allen@ccdc.cam.ac.uk

The Cambridge Structural Database (CSD) and its associated software systems have formed the basis for more than 800 research applications in structural chemistry, crystallography and the life sciences. Relevant references, dating from the mid-1970s, and brief synopses of these papers are collected in a database, DBUse, which is freely available *via* the CCDC website. This database has been used to review research applications of the CSD in organic chemistry, including supramolecular applications, and in organic crystal chemistry. The review concentrates on applications that have been published since 1990 and covers a wide range of topics, including structure correlation, conformational analysis, hydrogen bonding and other intermolecular interactions, studies of crystal packing, extended structural motifs, crystal engineering and polymorphism, and crystal structure prediction. Applications of CSD information in studies of crystal structure precision, the determination of crystal structures from powder diffraction data, together with applications in chemical informatics, are also discussed.

1. Introduction: the CCDC's DBUse database

In 1986 the CCDC began to compile literature references to the scientific applications of the Cambridge Structural Database (CSD: Allen, 2002). The list concentrated on published projects in which CSD information had underpinned experiments in structural knowledge acquisition (structure correlation, chemical informatics). The more straightforward uses of the CSD, *e.g.* to locate references or comparative geometry for use in publications of novel crystal structures, were not included. The list was incorporated into an article (Allen & Kennard, 1987) published in the IUCr monograph *Crystallographic Databases* (Allen, Bergerhoff & Sievers, 1987) and referenced 98 studies. Since that time, the list has been developed into a database, DBUse, which currently contains 856 references to primary journals, review articles and book chapters. Each DBUse entry contains the title of the article, bibliographic details and a short synopsis of the subject matter prepared by CCDC staff. Each article is classified according to type or subject, *e.g. reviews, hydrogen bonding, conformational analysis etc.*, and each article may have several classifications. DBUse is freely available and searchable *via* the CCDC website at http://www.ccdc.cam.ac.uk/dbuse, which also includes a form for notifying the CCDC of papers for potential inclusion in future updates.

The DBUse classification scheme is under detailed review, but some statistics based on current classes, which reveal the very wide range of CSD applications, and on growth over time,

Table 1
Summary statistics for the CSD DBUse database.

(*a*) Number of publications by subject area. Note that many papers have multiple classifications and all classifications are included below.

Subject area	Number
CSD development and applications reviews	86
Methods of data analysis	81
Intramolecular geometry and mean molecular dimensions	147
Conformational analysis and stereochemistry	146
Hydrogen bonding and intermolecular interactions	322
Applications in metallo-organic structural chemistry	107
Molecular modelling and drug design	119
Crystallographic applications	66
Chemical information science applications	44

(*b*) Number of publications by 5-year groupings for the period 1970–2000

Period	Number
1971–1975	5
1976–1980	23
1981–1985	86
1986–1990	139
1991–1995	297
1996–2000	306

are given in Table 1. Two of these areas, applications of the CSD in the life sciences (Taylor, 2002) and applications of the CSD to the structural chemistry of metallo-organic compounds (Orpen, 2002), are surveyed elsewhere in this special issue. The present survey covers the areas of organic chemistry and crystal chemistry, which represented the initial major focus areas for CSD studies and which continue to generate significant fundamental research output.

Any short survey of a large body of published material must necessarily be selective. However, the early research activities did give rise to a number of important reviews and data compilations (*e.g.* Dunitz, 1979; Bürgi & Dunitz, 1983; Allen *et al.*, 1983; Taylor & Kennard, 1984; Allen, Kennard *et al.*, 1987; Orpen *et al.*, 1989; Desiraju, 1989; Glusker *et al.*, 1994), and culminated in the publication of *Structure Correlation* (Bürgi & Dunitz, 1994). In the present paper we cite specific chapters of this book, together with earlier review material when appropriate, to provide leading references. This allows us to concentrate more specifically on work published since 1990. However, even with this limitation, we cannot attempt to do justice to over 300 recent papers and we recommend readers to make use of the DBUse search facilities to locate published applications of the CSD in their own field of interest.

2. Structural systematics and statistical methods

Structural data derived from early crystallographic analyses played a major role in the systematization of structural chemistry, pioneered by Pauling in the 1930s, which generated tables of covalent, van der Waals, ionic and other radii (Pauling, 1939), and tables of standard interatomic distances (Sutton, 1956). The ready availability of increasing volumes of crystal structure data in electronic form, together with appropriate search and analysis software made the derivation

of mean molecular dimensions an obvious application for the developing CSD system (Allen, 2002; Bruno *et al.*, 2002). Two major compilations, derived from a CSD containing some 50 000 structures, provided updated standard interatomic distance tables for organic structures (Allen, Kennard *et al.*, 1987: 682 chemical bond types involving 65 elements) and metallo-organic structures (Orpen *et al.*, 1989: 325 bond types involving the *d*- and *f*-block metals), and these two papers have received more than 3000 citations (Redman *et al.*, 2001). Engh & Huber (1991) have derived mean bond lengths and valence angles for peptidic structures in a CSD of 80 000 entries, basing their classification on 31 atom types that are most appropriate to the protein environment. These data are used extensively in the determination, refinement and validation of novel macromolecular structures, and probably deserve updating from a CSD that is now more than three times larger, as suggested by the original authors. More recently, Clowney *et al.* (1996) and Gelbin *et al.* (1996) have used CSD data to generate structural dictionaries for checking novel structures entering the Nucleic Acid Database (Berman, Westbrook *et al.*, 2002).

In the CSD-based compilations, and in other early studies, we see the emergence of proper statistical treatments of the results, *e.g.*

(i) in the presentation of appropriate descriptive statistics for the distributions of data points that give rise to specific mean values,

(ii) the use of both parametric and non-parametric tests to assess the significance of differences between means and

(iii) the use of covariance, correlation and regression to determine the extent to which two parameters are related. These techniques, together with a variety of multivariate statistical methods, were embodied in the CAMAL subroutine library (Taylor, 1986*b*) and reviewed by Taylor & Allen (1994). The use of multivariate techniques is exemplified below in the context of CSD-based conformational analysis studies.

2.1. Structure correlation

During the late 1970s and 1980s, Dunitz, Bürgi and co-workers enunciated the principle of structure–structure correlation, which underpinned their classic studies of reaction pathways in a series of aminoketones (Dunitz, 1979; Bürgi & Dunitz, 1983, 1994). Essentially, the static distortions exhibited by a specific molecular fragment in a wide variety of crystalline environments are assumed to map the distortions that the fragment would undergo along a reaction pathway, *i.e.* the various static fragments may be considered to form a series of structural snapshots along the pathway. This leads to the principle that '*if a correlation can be found between two or more independent parameters describing the structure of a given structural fragment in a variety of environments, then the correlation function maps a minimum energy path in the corresponding parameter space*' (Murray-Rust *et al.*, 1975), or more succinctly: '*observed structures tend to concentrate in low lying regions of the potential energy surface*' (Bürgi & Dunitz,

1983). This principle is clearly exemplified in many of the chapters of *Structure Correlation* (Bürgi & Dunitz, 1994), with further important application reviews being presented by Klebe (1994) and Ferretti *et al.* (1996).

From the mid-1980s the principle was extended to structure–property correlations, particularly through the work of Kirby and others on structure–reactivity correlations. These crystallographic approaches to transition state structure have been extensively reviewed by Kirby (1994). The approach is exemplified by a study of C—O bond length variations in a series of axial tetrahydropyranyl acetals [(I) in Fig. 1], in which C—OR bond breaking should be highly dependent on the orientation of the O lone-pair electrons (Deslongchamps, 1983): optimum $n\sigma$–σ^*(C—OR) overlap stabilizes both the

ground state axial conformation and the oxocarbocation [(II) in Fig. 1]. It is this orbital–orbital overlap that leads to hydrolysis of these compounds [(I) $\Rightarrow$ (II), Fig. 1], which usually occurs in acid, but can occur spontaneously when the leaving group —OR is good enough. Given the high degree of stereochemical control, a relationship might be expected to exist between the length of the C—OR bond in the ground state and the rate at which it is broken. Chemical evidence had shown that the pK_a of the conjugate acid (ROH) of the leaving group was directly related to reactivity and plots of the *exo* and *endo* C—O bond lengths in a variety of axial tetrahydropyranyl acetals *versus* the corresponding pK_a values (Fig. 1) showed linear correlations that had negative and positive slopes respectively, *i.e.* there was an increasing divergence between the two C—O distances for the better leaving groups, as would be expected from the reaction (I) $\Rightarrow$ (II) (Fig. 1). Kirby (1994) then documents how further targeted crystal structure analyses, together with theoretical calculations (Amos *et al.*, 1992) and further crystal structure correlations (Bürgi & Dubler-Steudler, 1988*a,b*), made it possible to predict a late transition state structure which must be close to that of the oxocarbocation [(II) in Fig. 1].
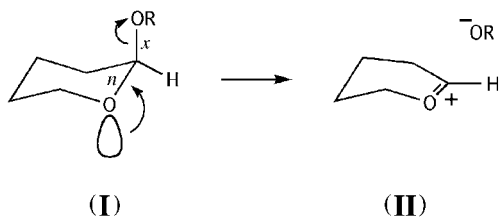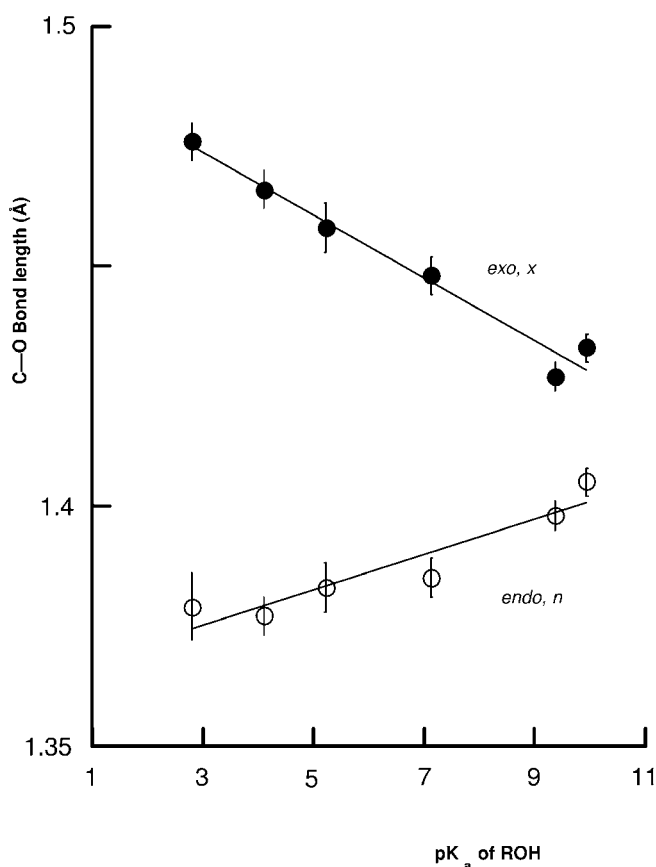
2.2. Conformational analysis

The generation of torsion angle distributions to determine conformational preferences about single rotatable bonds is one of the most common applications of the CSD, particularly in molecular design applications (Taylor, 2002), and in the *ab initio* determination of crystal structures from powder diffraction data (David *et al.*, 1998). One of the fundamental problems in a molecular modelling environment is that CSD distributions are determined in the condensed phase, where conformations can be affected by crystal field effects and may not represent global minima in solution or the geometries adopted when binding to proteins. There are cases where conformations in other phases are known to differ from those observed in crystal structures, *e.g.* in biphenyl (Brock & Minton, 1989), where low rotational barriers are comparable to crystal-field effects. However, a growing body of evidence from structure correlation studies suggested that crystal structure conformations are generally good guides to conformational preferences in solution and this hypothesis was systematically tested (Allen, Harris & Taylor, 1996), and found to hold, by comparing CSD torsional distributions with potential energy curves from high-level *ab initio* calculations for 12 varied molecular fragments. The implications of this study for molecular modelling applications of the CSD are examined by Taylor (2002).

Further evidence supporting the basic tenet of the structure correlation principle, comes from conformational studies of a range of multivariate systems, *i.e.* systems for which two or more torsion angles are required for the conformational definition. For example, Rappoport *et al.* (1990) have studied conformational variations in crystal structures of benzophenone derivatives (Fig. 2) *via* a Ramachandran plot of the two O=C—C(ar)—C(ar) torsion angles TOR1 and TOR2 (taking

Figure 1
Relationship between the lengths (*n,x* in Å) of the endocyclic (*endo*) and exocyclic (*exo*) C—O bonds at the acetal centres of axial tetrahydropyran acetals (I), and the pK_a of the conjugate acid *R*OH of the leaving group (II) (Kirby, 1994).

account of molecular symmetry, as discussed by Taylor & Allen, 1994). In their original paper, Rappoport et al. (1990) show the torsional mapping of Fig. 2 superimposed on a contoured potential energy surface which has energy minima at TOR1, TOR2 = +30, −30° and its symmetry equivalents, connected by low-energy valleys that correspond to the conformational interconversion pathways clearly observed in Fig. 2.

To study the conformational preferences of larger systems defined by $n > 3$ torsion angles, it is necessary to resort to multivariate analytical techniques such as principal component analysis (PCA) and cluster analysis (CA; see e.g. Taylor & Allen, 1994). PCA is a method of dimension reduction which attempts to re-express the original n variables on a set of m mutually orthogonal axes – the principal components or PCs. If m is significantly less than n, it is likely that pairwise scatterplots of PCA scores, i.e. the coordinates of the fragments with respect to the PC axes, will provide conformational mappings that can be interpreted to reveal clusters of preferred conformations, together with any conformational interconversion pathways connecting these clusters. Cluster analysis is a numerical technique which uses dissimilarities between pairs of data points (conformations) to group together those points which are closely similar and to distinguish these clusters from others which are inherently dissimilar.

The value of multivariate techniques for CSD analysis was first illustrated by Murray-Rust & Motherwell (1978) and Taylor (1986a). During the 1990s these techniques were used extensively to map and interpret the conformational spaces of medium rings having 5–8 C (or hetero) atoms (Allen, Doyle &

Auf der Heyde, 1991; Allen et al., 1993; Allen, Howard & Pitchford, 1996), and of macrocyclic ether and thioether ligands (Raithby et al., 1997a,b). These studies have shown that:

(i) PCA results are formally equivalent to the ring puckering analyses of Cremer & Pope (1975), but with a rotation of axes, and

(ii) that the populations of clusters are in qualitative agreement with the relative energies of the conformers identified – the structure correlation principle again.

Thus, for cyclopentane rings PCA reduces a five-dimensional dataset to just two PCs that account for >99% of the variance in the dataset. Fig. 3 shows the symmetry-modified PC1–PC2 mapping of 500 rings retrieved from the CSD and nicely illustrates the well known equi-energetic pseudorotation pathway involving the five envelope and five twist forms and their mirror images, together with their interconversions to and from the planar form at the centre of the plot. PCA has frequently been used in the conformational analysis of coordination compounds, as described by Orpen (1993, 2002). During this period also, some experiments have been conducted into the use of machine learning techniques for conformational classification, as exemplified by the use of conceptual clustering which accurately reproduces the results of some of the analyses cited above (Conklin et al., 1996).

3. Intermolecular interactions

In his Nobel lecture, Lehn (1988) defined supramolecular chemistry as 'the chemistry of the intermolecular bond'. A
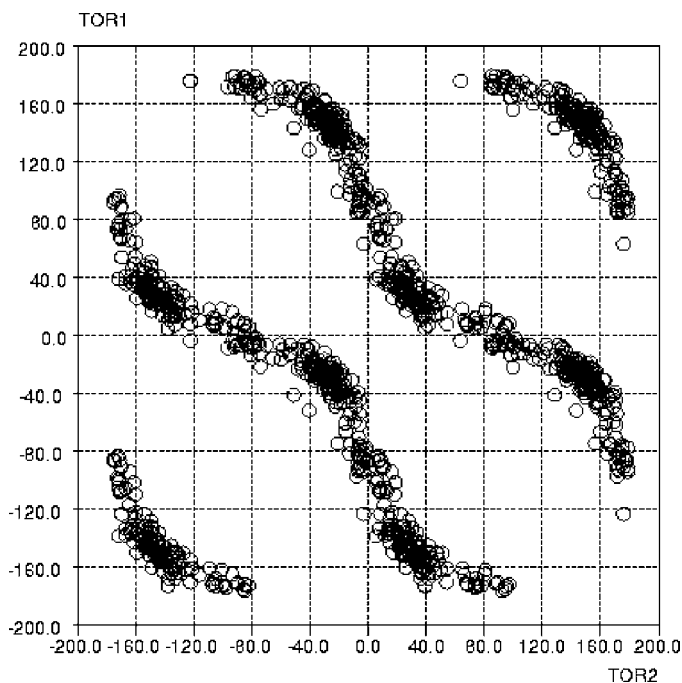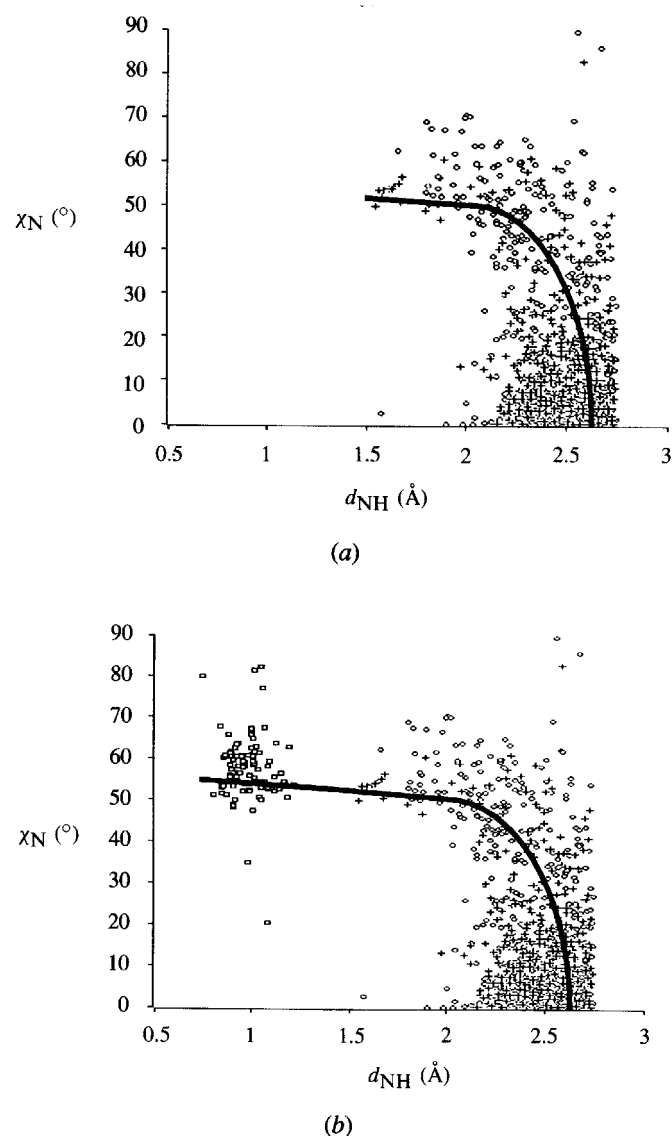

Figure 2
Symmetry-expanded Ramachandran plot of the torsion angles O=C—C(ar)−C(ar) in benzophenone fragments retrieved from the CSD (see Rappoport et al., 1990).
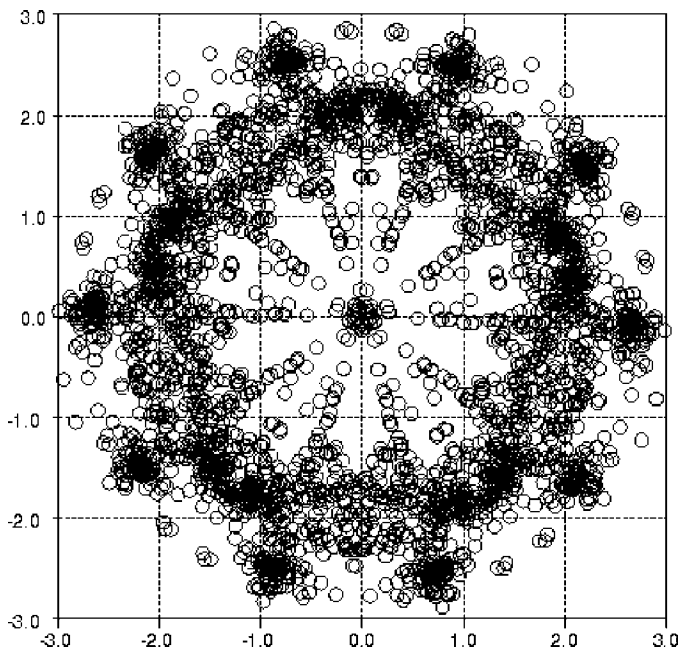
Figure 3
Symmetry-expanded principal component analysis based on the five intra-annular torsion angles for 500 cyclopentane rings retrieved at random from the CSD. Two degenerate PCs account for >99% of the variance in this dataset and the plot axes are PC1 along x and PC2 along y (see Allen, Doyle & Auf der Heyde, 1991).

crystal structure is the archetypal supermolecule and provides direct experimental knowledge concerning the many and varied intermolecular interactions that control molecular aggregation, *i.e.* the types of interaction that occur, their geometrical characteristics, directional preferences and their involvement in common supramolecular motifs. This knowledge is fundamental to our understanding and development of all things supramolecular, including *inter alia*: supramolecular synthesis, crystal structure solution and validation, crystal engineering, protein–ligand docking and crystal structure prediction. The CSD is therefore a primary source of knowledge that can be applied in all of these areas.

Early CSD studies of intermolecular interactions were not simple to perform, relying on a two-stage search process using separate programs:

(a)

(b)

Figure 4
Reaction pathway for the protonation of N in anilines (Allen, Bird *et al.*, 1995). Plots of χ_N (a measure of the pyramidality of nitrogen) *versus* d_{NH} (the N$\cdots$H or N—H distance) (*a*) for nonbonded N$\cdots$H interactions only and (*b*) including bonded N$^+$—H distances.

(i) use of two-dimensional substructure searching to locate CSD entries containing the chemical functional groups of interest, followed by

(ii) three-dimensional geometric searching to locate specific intermolecular interactions.

Data analysis often involved the use of commercial or local code available to individual researchers. In 1992, facilities for molecular (two-dimensional) and supramolecular (three-dimensional) substructure searching were integrated and improved within the framework of the *QUEST3D* program (Cambridge Crystallographic Data Centre, 1994), and this fundamental search philosophy is now available within the CCDC's *ConQuest* software (Bruno *et al.*, 2002). Data analysis facilities are available *via Vista* (Cambridge Crystallographic Data Centre, 1995) and much enhanced structure visualization is now provided within *Mercury*[1] (Taylor & Macrae, 2001; Bruno *et al.*, 2002).

Despite the operational difficulties, many significant CSD studies of hydrogen-bonded systems were published prior to 1992, including a number of important surveys and reviews (Kroon *et al.*, 1975; Jeffrey & Maluszynska, 1982, 1986; Murray-Rust & Glusker, 1984; Taylor & Kennard, 1982, 1984; Desiraju, 1991*a*). Much of the early material is also discussed and referenced in monographs by Jeffrey & Saenger (1991) and by Jeffrey (1997). Given this available background, we concentrate here on CSD-based studies of intermolecular interactions published in the past decade.

3.1. van der Waals radii

The hard-sphere atomic model is central to the study of intermolecular interactions and, to an approximation, atomic van der Waals radii can be regarded as transferable from one structure to another. The van der Waals radii of Bondi (1964) are the most highly cited compilation in modern literature and were assembled from a variety of sources; however, the original paper issues a caution about the general validity of the derived values for assessing limiting contact distances in crystals. Rowland & Taylor (1996) have tested this statement for the common non-metallic elements and find excellent overall agreement between CSD results and the Bondi (1964) values. The only significant discrepancy is for hydrogen, where the Rowland & Taylor (1996) analysis yields a recommended value of 1.1 Å, which is 0.1 Å shorter than that of Bondi (1964). A reduction in v(H) was also suggested by Steiner & Saenger (1991, 1992*a*), who studied non-bonded H$\cdots$H distances in cooperative $\cdots$O—H$\cdots$O—H$\cdots$ systems studied by neutron diffraction. These authors cite a minimum H$\cdots$H contact of 2.05 Å with some polar flattening of the van der Waals sphere for H. They conclude that a global value for v(H) should be closer to 1.0 Å than to the 1.2 Å cited by Bondi (1964).

3.2. Combining CSD-based studies with quantum chemistry calculations

CSD-based studies cannot provide anything better than crude estimates of the relative strengths of hydrogen bonds

and one of the important developments during the 1990s has been to combine CSD studies with calculations of interaction energies using a variety of *ab initio* computational methods. These methods have been applied not only to the stronger hydrogen bonds (see *e.g.* Llamas-Saiz *et al.*, 1992; Allen, Lommerse *et al.*, 1997; Bock *et al.*, 1998), but also to a wide variety of weaker hydrogen-bonded systems (see *e.g.* Rzepa *et al.*, 1994; Steiner *et al.*, 1995; Dunitz & Taylor, 1997; Lommerse & Cole, 1998) and to interactions that are not mediated by hydrogen (see *e.g.* Lommerse *et al.*, 1996; Allen *et al.*, 1998). These and other examples of the combined crystallographic/ computational approach are noted in the sections below.

Of particular importance are quantum-mechanical methods that

(i) take proper account of basis set superposition errors (Stone, 1993), and

(ii) yield separate components of the total interaction energy, E_t, which have distinct physical significance, *i.e.* terms that quantify the electrostatic, exchange–repulsion, polarization, charge transfer and dispersion energy components. Such a separation is effected, for example, using the intermolecular perturbation theory module (IMPT: Hayes & Stone, 1984) of the *CADPAC*6.0 software package (Amos, 1994).

4. Hydrogen bonding

4.1. Systematic studies of stronger hydrogen bonds

Hydrogen bonds are the most important intermolecular interactions that determine molecular recognition and aggregation in biology and chemistry, and are implicated in the structures of many novel materials. Crystal structure data has
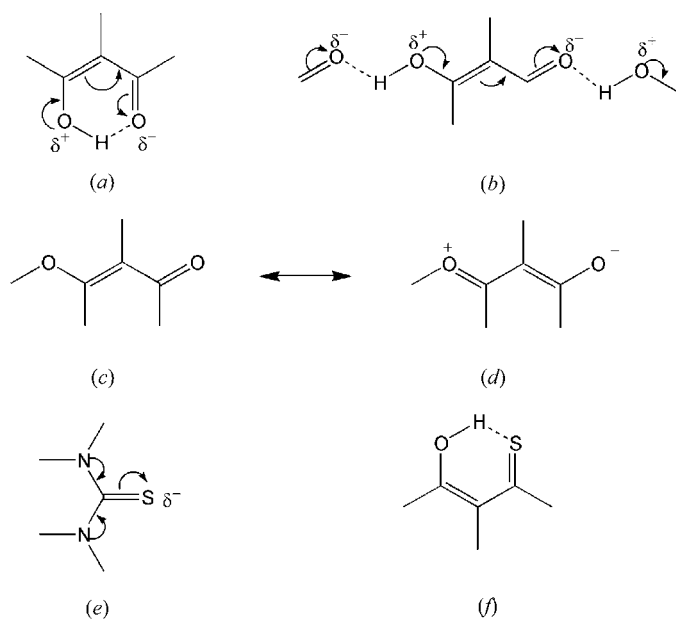
Figure 5
Resonance-assisted and resonance-induced hydrogen bonding in enolones (*a*)–(*d*) and thiones (*e*) and (*f*) (see text).

been the principal source of knowledge about hydrogen bonds for many years, as shown by a scan of the references contained in the DBUse database. Many recent studies have used CSD system software to quantify the metrical and directional properties of the common strong hydrogen bonds involving N—H and O—H donors and a variety of N and O acceptors (see *e.g.* Steiner & Saenger, 1992*b*; Pirard *et al.*, 1995; Chertanova & Pascard, 1996; Grabowski, 1998). Other stronger hydrogen bonds, such as N,O—H···halide bonds, have also been studied in this manner (Steiner, 1998*a*).

However, these basic studies have been extended in a number of ways. Thus, Gavezzotti & Filippini (1994) have used CSD data on N,O—H···N,O systems to calibrate empirical hydrogen-bond potentials. Steiner & Saenger (1994) and Steiner (1995) have studied O—H···O and N—H···N systems, respectively, and have shown that a smooth relationship exists between the lengthening of the O,N—H bond and the shortening of the O···O, N···N separations. Allen, Cole & Howard (1995*a*,*b*) have studied the geometry of N,O—H···NH₂ hydrogen bonds in anilines and have mapped the reaction pathway for protonation of the amino-N (Fig. 4): as the amine group rotates out of the plane of the phenyl ring the conjugative interaction is disengaged, generating a lone pair on the increasingly pyramidal N atom which then acts as a (increasingly strong) hydrogen-bond acceptor.

4.2. Resonance-assisted and resonance-induced hydrogen bonding

Gilli and co-workers (Gilli *et al.*, 1989, 2000; Bertolasi *et al.*, 1993, 1996; Ferretti *et al.*, 1996) have made extensive studies of hydrogen bonding in beta-enolones (Fig. 5), beta-enaminones and related systems. In their original study (Gilli *et al.*, 1989), they noted that the enolone group undergoes greater delocalization when the fragment forms either intramolecular (Fig. 5*a*) or infinite intermolecular chains (Fig. 5*b*) of hydrogen bonds. They report that the proportion of the two resonance forms (Figs. 5*c* and *d*) changes from an 87:13 mixture in non-hydrogen-bonded situations to a 63:37 mixture for intramolecular hydrogen bonding and to a 52:48 mixture in the intermolecular case. These data are backed up by detailed structure correlation studies, combined with the determination of novel structures, and the use of IR and NMR spectroscopy and quantum chemical calculations. These authors (Bertolasi *et al.*, 1996) and others (*e.g.* Jeffrey, 1997) also note the strengthening of hydrogen bonds when the acceptor carries a full charge, a factor that plays an important role in the crystal engineering of metallo-organic structures (Brammer *et al.*, 1995; Aullon *et al.*, 1998; Braga *et al.*, 2000).

In a related study of N,O—H hydrogen bonding to $(R_1,R_2)C=S$ acceptors, Allen, Bird *et al.* (1997) showed that hydrogen-bond formation to the S acceptor is only induced by the electron-donating ability of the C-substituents R_1, R_2. If the R_1,R_2 substituents are not electron-donating, the electronegativity difference between the C and S atoms in the C=S bond is close to zero and hydrogen-bond formation is inhibited. However, with R_1,R_2 = N, as in thiourea and its deriva-

tives, the effective electronegativity of S is significantly increased by resonance effects (Fig. 5e) so that the S atom now becomes an effective acceptor. Steiner (1998b) has extended this study to show that a small number of very short intramolecular $O-H\cdots S=C$ hydrogen bonds exist in monothio-β-diketones (Fig. 5f, in which R_1,R_2 are both C atoms), a further example of resonance induction.

4.3. Hydrogen-bond competition effects

In many cases, there exists competition for available donor-H atoms between different and apparently strong acceptor atoms. For example, it is known (Lommerse et al., 1997) that hydrogen bonds to esters form almost exclusively to the carbonyl-O rather than to the ether-O, whereas both $>C=O$ and $C-O-C$ O atoms are known to be good acceptors in isolated situations. This observation was explained by Lommerse et al. (1997) by the use of detailed IMPT calculations: hydrogen bonds to the carbonyl-O are some $10\ \text{kJ mol}^{-1}$ stronger than those to the ether-O in the ester environment, and the ether-O simply loses out on the basis of energetic competition for available donor-H. These authors have used similar methods to study acceptor competition between N and O in a variety of aromatic heterocycles (Nobeli et al., 1997), and use detailed IMPT calculations to show that O is again disfavoured on energetic grounds. They go on to show that $O-H\cdots O$(furan) bonds form in only 3% of possible cases since these bonds are particularly weak ($ca\ -14\ \text{kJ mol}^{-1}$), even weaker than those to the ether-O in esters. In both furan and esters, the ether-O is bonded to Csp^2 atoms and the relative weakness of Csp^2-O as a hydrogen-bond acceptor has been extensively studied by Böhm et al. (1998). More recently, the CSD has been used in conjunction with quantum chemical calculations to study competition between amino-N and cyano-N (Ziao et al., 2001), in which the latter is shown to be the more potent acceptor, and by Steiner (2001) to study acceptor competition for the strong carboxyl donor. A detailed understanding of acceptor competition and the effects of local acceptor environments are of vital importance in crystal engineering and studies of protein–ligand interactions.

4.4. Weaker hydrogen bonds

Perhaps the most consistent theme running through recent hydrogen bond research has been the identification, study and discussion of weak interactions involving:

(i) weak donors and strong acceptors exemplified by $C-H\cdots O$ and $C-H\cdots N$ bonds,

(ii) strong donors and weak acceptors, such as $O,N-H\cdots F,Cl$ and $O,N-H\cdots\pi$, and

(iii) weak donors and weak acceptors, such as $C-H\cdots\pi$, $C-H\cdots Cl-C$ etc.

The CSD has been vital in these researches, together with ab initio calculations, and correlation of geometry with spectroscopic and other data. Indeed, the existence of this research area owes much to early CSD-based research and particularly to the highly cited paper by Taylor & Kennard (1982), which provided indisputable evidence for the existence of $C-H\cdots O$ hydrogen bonds via a careful statistical analysis of neutron diffraction results for small molecule structures. This paper, which refuted emphatically the embargo, arising from the views of Donohue (1968), on the discussion of short $C-H\cdots O$ contacts as true hydrogen bonds, has currently received more than 1000 citations (Redman et al., 2001). Such has been the level of research activity over the past two decades that the whole area has recently been the subject of an extensive monograph by Desiraju & Steiner (1999). Given the existence of this encyclopaedic work, what follows are some short, and perhaps subjective, notes on a number of key issues, including some material that has appeared only very recently.

4.4.1. Weak donors–strong acceptors. The $C-H\cdots O$ hydrogen bond has continued to be a major focus since the mid-1980s, prompting two reviews (Desiraju, 1991a, 1996) which illustrate their importance in supramolecular design and provoking significant interest in their involvement in protein secondary structure stabilization (see e.g. Derewenda et al., 1995). However, the criteria used for classifying both $C-H\cdots O$ and $C-H\cdots N$ contacts as hydrogen bonds have been discussed by many authors and were called into question again by Cotton et al. (1997), who judged that existing criteria do not distinguish between what is a hydrogen bond and what is 'nothing more than a classical van der Waals interaction'. This point was immediately addressed by Steiner & Desiraju (1998) using $C-H\cdots$acceptor angular data derived from the CSD, and by Mascal (1998) and Thalladi et al. (2000) who used both CSD distance and angle distributions and a number of novel structures to show that there are real precedents for $C-H\cdots N$ bonding.

4.4.2. Strong donors–weak acceptors. Another area to receive attention has been the effectiveness (or not) of fluorine and chlorine as hydrogen-bond acceptors. Two papers (Howard et al., 1996; Dunitz & Taylor, 1997) use the CSD and appropriate ab initio calculations to show that organic fluorine hardly ever accepts hydrogen bonds, although F in anionic situations, e.g. BF_4^-, is an effective acceptor. However, the situation for Cl (Br and I) is completely different. Beginning with work that showed that metal-bound chlorine often accepts hydrogen bonds (Aullon et al., 1998), Brammer et al. (2001) performed a highly detailed study combining CSD geometrical statistics with high-level ab initio calculations. This study showed that $O-H$ and $N-H$ interactions with $C-Cl$ exhibit strong geometrical similarities to their $M-Cl$ counterparts, although they are energetically much weaker. In both cases, the donor-H approaches Cl in a direction which is approximately perpendicular to the $M-Cl$ or $C-Cl$ bond – a fact that will be alluded to again below. Notably, however, Brammer et al. (1999) had already shown that metal-bound fluorine exhibited marked departures from the hydrogen-bond geometries observed for $M-X$ ($X=$ Cl, Br, I) acceptors.

Spectroscopic evidence for the existence of $N,O-H$ hydrogen bonding to π-acceptor density in acetylenic, aromatic, olefinic and pseudo-olefinic (such as cyclopropane) systems is well documented (see e.g. Joris et al., 1968). The first CSD study of these interactions was carried out by Levitt &

research papers

Perutz (1988), prompted by observations in protein structures. Since that time a number of other contributions (Viswamitra *et al.*, 1993; Steiner *et al.*, 1995; Allen, Lommerse *et al.*, 1996) have continued to explore this area, and Ciunik & Desiraju (2001) have discussed the proper statistical treatment of the geometry of hydrogen bonds to multi-atom acceptors.

4.4.3. Weak donors–weak acceptors. Interactions between C−H donors, of varying acidities, and phenyl and acetylenic π-systems are the most widely studied of this class of interactions. There is CSD and *ab initio* computational evidence (Steiner, 1995, 1996; Fan *et al.*, 1996) of hydrogen-bond characteristics for interactions involving the more acidic (*e.g.*
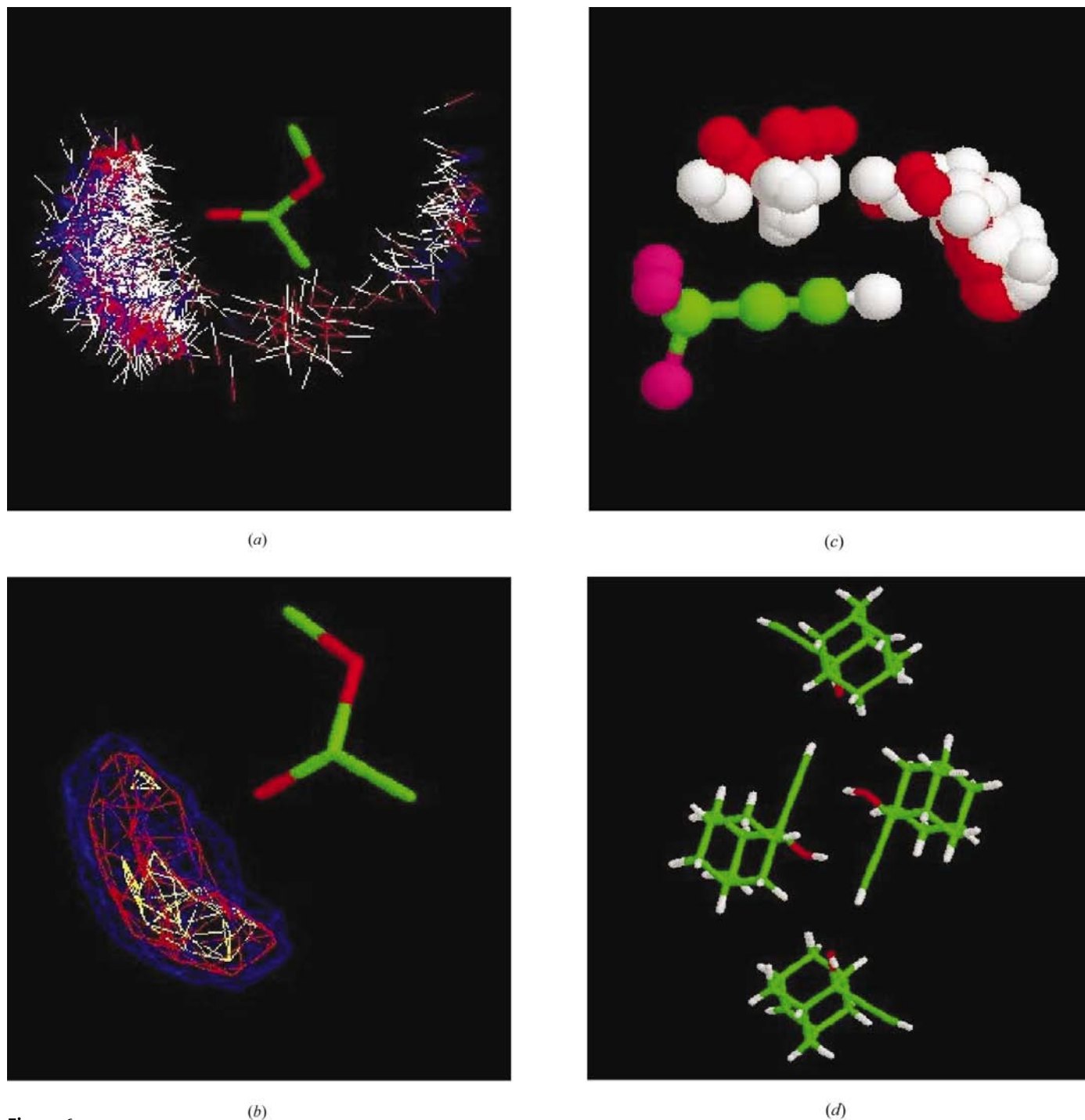
Figure 6
The IsoStar knowledge-base of intermolecular interactions (Bruno *et al.*, 1997). (*a*) Basic scatterplot of N−H and O−H donors (contact atoms) around an ester central group. (*b*) Contoured version of scatterplot (*a*). (*c*) Hyperlinking from IsoStar to the CSD: scatterplot of O−H donors around an acetylenic central group. (*d*) Hyperlinking from IsoStar to the CSD: The structure BETXAZ (Lin *et al.*, 1982) that generates the shortest O−H$\cdots\pi$ (acetylene) interaction.

acetylenic, phenyl and ethylenic) C—H, but interaction energies are low (from *ca* -3 to -7 kJ mol^{-1}). Nevertheless, this is a burgeoning area of investigation and existing evidence from a wide range of sources is collected in a recent monograph (Nishio *et al.*, 1998). Interest has also centred on C—H$\cdots$Cl interactions, where Aäkeroy *et al.* (1999) use CSD statistics to propose the universality of C—H$\cdots$Cl hydrogen bonds, but Thallapally & Nangia (2001) conclude that while C—H$\cdots$Cl$^-$ and C—H$\cdots$Cl—Metal systems exhibit hydrogen-bond character, C—H$\cdots$Cl—C systems are best regarded as van der Waals interactions. Nevertheless, C—H$\cdots$Cl,Br,I—C hydrogen bonds have recently been invoked by Gibb *et al.* (2001) as being among the range of tools available to the supramolecular chemist from their studies of host–guest complexation. Finally, Thalladi *et al.* (1998) have suggested that, while organic fluorine is a poor acceptor of hydrogen bonds from strong donors, the C—H$\cdots$F interaction may play an important role in structural stabilization.

5. Interactions not mediated by hydrogen

In a recent review, Desiraju (1995) illustrates the structural importance of a wide range of attractive non-bonded interactions that are not mediated by hydrogen and notes the value of the CSD in identifying and characterizing such interactions. In this context, the combination of CSD analysis, to indicate preferred interaction geometries, and *ab initio* calculations, to explore the energetics of interactions in these preferred regions, is of considerable value. By quantifying the interaction energies, the combined approach not only helps to predict likely interaction patterns, but also their relative robustness by comparison with *e.g.* hydrogen-bonded analogues.

The strong tendency for the halogens X = Cl, Br, I to form short contacts to other halogens and especially to electronegative O and N atoms (Nyburg & Faerman, 1985) is well known (Price *et al.*, 1994). Recent combined CSD/IMPT studies of C—$X\cdots$O=C< systems (Lommerse *et al.*, 1997) and C—$X\cdots$O(nitro) systems (Allen, Lommerse *et al.*, 1997) showed a marked preference for the $X\cdots$O interaction to form along the extension of the C—X bond, with interaction energies in the range from -7 to -10 kJ mol^{-1}. These energies are comparable with those of C—H$\cdots$O hydrogen bonds and have been used by Desiraju and co-workers (see Desiraju, 1995, for leading references) to engineer a number of novel small-molecule crystal structures. By contrast, C—Cl$\cdots$Cl—C interactions tend to form with C—Cl$\cdots$Cl angles approaching 90° (Price *et al.*, 1994). Taken together, these studies provide a picture of Cl as presenting a quadrupolar electrostatic environment to potential interacting species, with an area of positive density pointing outwards along the extension of the C—Cl bond, thus interacting with O(δ^-) and a region of negative density in directions perpendicular to the C—Cl bond, thus interacting with Cl(δ^+) and H(δ^+), as noted in the M—Cl$\cdots$H bonds discussed above. Similar computational methods have been used to study short S$\cdots$O and S$\cdots$N interactions that occur in thiazole nucleosides (Burling & Goldstein, 1992) and

the general prevalence of such interactions has been documented using the CSD system (Burling & Goldstein, 1993).

The importance of multi-centre interactions between functional groups has been studied by Taylor *et al.* (1990), who constructed the composite crystal-field environments for carbonyl and nitro groups during a search for isosteric replacements in modelling protein–ligand interactions. Building on this work, Allen *et al.* (1998) have shown that antiparallel dipolar interactions involving >C=O groups have interaction energies (*ca* -20 kJ mol^{-1}) that are comparable to those of medium-strength hydrogen bonds and are clearly involved in stabilizing the crystal packing of *e.g.* 3-oxosteroids. Carbonyl–carbonyl interactions have also been shown to be significant in proteins and are responsible for stabilizing certain protein secondary structures (Maccallum *et al.*, 1995a,b), and for stabilizing the partially allowed Ramachandran conformations of asparagine and aspartic acid (Deane *et al.*, 1999).

6. The IsoStar knowledge base

A vast amount of information on intermolecular interactions is now gathered together in a readily accessible form in the CCDC's IsoStar knowledge base (Bruno *et al.*, 1997), which has formed part of the distributed CSD system since 1997. For a given contact between a central group (A) and a contact group (B), CSD search results for an interaction $A\cdots B$ are transformed into an easily visualized form by overlaying the A moieties. This results in a three-dimensional scatterplot showing the experimental distribution of the B moieties around the (static) central group A (see Fig. 6a), which can also be presented in contoured form (Fig. 6b). IsoStar contains data retrieved from the CSD, from protein–ligand complexes stored in the Protein Data Bank (PDB: Berman *et al.*, 2002) and also contains nearly 1000 potential energy minima calculated using distributed multipole analysis (see *e.g.* Beyer *et al.*, 2001) and IMPT calculations. Version 1.4 of IsoStar, released in October 2001, covers 300 central groups, 45 contact groups and contains over 18 000 scatterplots. The user may interact with the basic scatterplots: to generate contoured surfaces, to change the $A\cdots B$ distance limit for data presentation, to control the display style *etc.* Importantly, the scatterplot data are hyperlinked to the master CSD. Thus, Fig. 6(c) shows the scatterplot of O—H contacts to an acetylenic group. By clicking on the shortest O—H$\cdots\pi$ contact, the user is presented with the original crystal structure from which that contact was derived: CSD Refcode BETXAZ (Lin *et al.*, 1982), in which the structure comprises a tetrameric motif connected by —C≡C—H$\cdots$O and O—H$\cdots\pi$ hydrogen bonds. IsoStar therefore contains a vast amount of information of use in crystal engineering and organic crystal chemistry, and also provides ready access to information that is invaluable in rational ligand design (Taylor, 2002).

7. Structural motifs and graph sets

The previous sections have illustrated how the CSD system, often in combination with other experimental and computa-

tional data, has been used to study the local geometry of non-bonded interactions. However, crystallographers have long recognized, classified and discussed the structural patterns or motifs formed by sets of interactions, and their role in controlling molecular aggregation to form extended crystal structures. A series of papers by Etter and co-workers in the late 1980 s and early 1990 s concentrated on the importance of hydrogen-bonded patterns in determining and engineering crystal structures (see *e.g.* Panunto *et al.*, 1987; Etter & Reutzel, 1991; Görbitz & Etter, 1992). Then, building on ideas first proposed by Kuleshova & Zorkii (1980), the classification of hydrogen-bond patterns using graph theoretical descriptors was developed (Etter, 1990; Etter *et al.*, 1990; Bernstein *et al.*, 1990, 1995) into a viable tool for comparing and contrasting extended crystal structures, including those of polymorphs. The main problem inherent in the method was its reliance on human recognition and encoding of the often complex hydrogen-bonding patterns. This problem was finally solved, for first-level graph sets at least, using graph-theoretical algorithms due to Grell *et al.* (1999) and encoded within the CCDC's *RPLUTO* structure visualization program[1] (Motherwell *et al.*, 1999, 2000), which operates on structures retrieved from the CSD, or on structure data in CIF format (Hall *et al.*, 1991; Brown & McMahon, 2002).

8. Crystal engineering

The field of crystal engineering (Desiraju, 1989) developed apace from the mid-1980s and is a major applications area for new knowledge derived from the CSD. Desiraju (1995) has encapsulated the importance of robust and reproducible supramolecular patterns in structural design by describing them as supramolecular synthons. This terminology recognized the synthetic basis of crystal engineering by borrowing the term 'synthon' from the world of molecular organic chemistry (Corey, 1967). In his paper Desiraju (1995) cites the use of the CSD to perform supramolecular retrosynthesis, *i.e.* to use existing structures to recognize and then employ the most frequently occurring (and hence the most robust) non-bonded structure-determining patterns. To bring some automation to the retrosynthetic analysis, Allen *et al.* (1999) modified CSD system software to recognize and generate a probability ranking for bimolecular hydrogen-bonded ring motifs in published organic structures, restricting the analysis to the stronger hydrogen bonds involving N,O—H donors and N,O acceptors. The concept of motifs is also shown to be important in a study of intramolecular hydrogen bonds presented by Bilton *et al.* (2000) and which have significant implications for supramolecular organization.

Although hydrogen bonds are involved in the formation of many robust synthons, there is considerable evidence (see *e.g.* Desiraju, 1995) that synthons involving other non-bonded interactions also occur predictably. An early mention in crystal engineering is the C—Cl· · ·Cl—C synthon leading to the

engineering of a short cell axis of *ca* 4 Å (see Sarma & Desiraju, 1986). Large functional groups such as tetraphenyl-P+ cations have been observed to form multiple phenyl embrace motifs (Dance & Scudder, 1996). Another type of interaction is exemplified by C—halogen· · ·π interactions with aromatic systems, which also appear to play an important role in crystal packing (Prasanna & Row, 2000). A recent review of database research in crystal engineering is presented by Nangia (2002).

9. Polymorphism and pseudo-polymorphism

The CSD release of October 2001 (245 392 entries) contains only 4335 unique compounds with reported polymorphism, less than 2% of entries. This figure almost certainly does not represent the true frequency of occurrence of polymorphic forms for organic molecules under normal laboratory conditions. The most common reason for the use of X-ray analysis is to determine a molecular structure unambiguously, hence most investigators will choose the most suitable crystalline sample available and will not be motivated to search for other polymorphs. However, despite the often quoted opinion that the number of reported polymorphs is proportional to the time spent looking for them (McCrone, 1965), it turns out that intensive investigations of certain pharmaceutical compounds usually reveal only a few polymorphs, although the molecules are flexible and have many hydrogen-bonding possibilities. Furthermore, many common compounds that have been studied for up to 100 years still show just one polymorph, *e.g.* naphthalene, benzoic acid, sucrose *etc.* Crystal structure lattice energy calculations of polymorphs reveal a narrow energy range between forms; the largest current survey, for 204 pairs of CSD polymorphs (Gavezzotti & Filippini, 1995), discusses differences in density, static packing energy, lattice-vibrational entropy and other crystal properties.

The term 'pseudo-polymorphism' has been used to describe cases where a compound exists in a pure crystalline form, and also co-crystallized with solvent molecules. An analysis of hydration in organic crystals (Desiraju, 1991*b*) showed that the occurrence of a hydrated form increases as the hydrogen-bond donor/acceptor ratio increases. A survey of approximately 5000 CSD structures containing common solvent molecules (Nangia & Desiraju, 1999) shows the importance of multipoint recognition of solvents by both strong and weak hydrogen bonds. A comprehensive survey of CSD structures containing any type of solvent molecule (Görbitz & Hersleth, 2000) identifies over 300 solvent types and the relative frequencies of occurrence are given.

10. Crystal and molecular symmetry and packing analysis

A definitive survey of space-group occupancy within the CSD (Wilson, 1988, 1990, 1993) showed the rarity of certain space groups for organic molecules (75 space groups had no examples). Their rarity is attributed to the presence of mirror planes or rotation axes, which give rise to inefficient packing

---

[1] *RPLUTO* and *Mercury* can be downloaded for non-commercial research purposes from http://www.ccdc.cam.ac.uk/prods/.

and conflict with the Kitaigorodskii (1973) principle of closest packing as a determinant of crystal structure. The influence of molecular symmetry on the extended crystal packing and space group has been investigated. An study of 129 pairs of racemic and chiral structures of the same molecule was performed by Brock *et al.* (1991) in the hope that the racemic crystals would generally be more closely packed, *i.e.* have a higher density, but results were inconclusive.

The relative importance of various symmetry elements in the molecule in governing the choice of space groups reported in the CSD has also been discussed (Brock & Dunitz, 1994*a*; Filippini & Gavezzotti, 1992*a*; Belsky *et al.*, 1995). Further work on the prevalence of polar and chiral space groups in the CSD (Brock & Dunitz, 1994*b*) looked for molecular factors which might influence the choice of space groups, *e.g.* space groups containing mirror planes nearly always include mirror-symmetric molecules located on special positions, and crystals with threefold rotation axes usually contain molecules having threefold symmetry. However, the inverse relationship is not generally true and in general molecules with molecular symmetry such as a mirror-plane do not predictably choose space groups with a mirror plane.

Recently a computational methodology has been developed for perceiving molecular symmetry (Cole *et al.*, 2000). A comprehensive calculation has now been performed for approximately 200 000 CSD entries and a database constructed which allows easy searching for combinations of molecular and space-group symmetry elements (Yao *et al.*, 2002). Simple statistics can be easily derived, *e.g.* there were 25 710 entries for molecules with an inversion symmetry element C_i (-1) in their point groups, and 19 817 (77.1%) reside on a Wyckoff position of -1 symmetry. As well as confirming some of the earlier findings of Brock & Dunitz (1994*b*), the database provides joint frequency distributions for specific combinations of molecular symmetry elements and the occupied Wyckoff positions, and *vice versa*.

As the CSD has grown in size the question has arisen as to how many times the wrong space group has been used in structure refinement. In general, the reported CSD space group is correct, but some cases of false assignments have been given (see *e.g.* Baur & Kassner, 1992; Marsh *et al.*, 2002). A discussion of wrongly assigned space groups and suggestions for avoiding such mistakes has been given by Marsh (1995).

The case of structures having more than one molecule in the asymmetric unit has been studied (Steiner, 2000), both with regard to effects on the molecular conformation of the crystallographically independent molecules (Gautham, 1992) and the wider question of pseudo-symmetry between these independent molecules, *e.g.* in $P\bar{1}$ (Desiraju *et al.*, 1991). Another study in $Pca2_1$ and $Pna2_1$ shows local pseudo-centres of symmetry lying mostly near preferred positions *e.g.* $x = 1/8$ and $y = 1/4$ in $Pca2_1$ (Marsh *et al.*, 1998).

Studies on packing patterns often take a wider view than details of synthons and although hydrogen bonds are important they are not the sole determinant of crystal structure formation. For example, the occurrence of hydrophobic

regions in hydrogen-bonded crystals is a recognized pattern (Görbitz & Etter, 1992; Cole *et al.*, 1998). Another generalized pattern is the formation of bilayers in metal carboxylate structures where the influence of hydrophobic substituents is important (Vela & Foxman, 2000). A systematic study of fused ring hydrocarbons (Gavezzotti & Desiraju, 1988) gave some correlations between the packing energy and various molecular indices, and some classification of basic patterns of packing. A more generalized view of packing was given by a study of the distribution of molecular centres (geometric centroids) in the normalized unit cell for common space groups (Motherwell, 1997) and showed strong preferences for certain centroid positions, *e.g.* $P2_1$ shows most molecules lying at $x = 1/4$ or $z = 1/4$. The recognition that there are certain spatial patterns preferred for the molecular coordination sphere (Holden *et al.*, 1993) has formed the basis for a crystal structure prediction program (*MOLPAK*) using these patterns as templates. The question of classifying structures according to the broad similarity of their packing patterns has been hinted at by Desiraju (1995). Sometimes there are networks of molecules common to several families of compounds, when one concentrates not on the chemical details but the broad pattern of intermolecular bonding.

11. Crystal structure prediction

For many years a relatively small group of researchers have been pursuing the goal of *ab initio* prediction of crystal structures given only the chemical diagram for a molecule (for comprehensive reviews see Gavezzotti, 1994, 1998; Verwer & Leusen, 1998). The predominant method of approach is to calculate the lattice energy for a range of space groups with one molecule in the asymmetric unit, usually using a rigid molecular model. Many thousands of possible crystal structures are generated and various mathematical optimization methods are used to seek the global minimum in the energy surface. This approach is showing signs of success and blind tests of the prediction of unpublished small organic structures have recently been organized by the CCDC (Lommerse *et al.*, 2000; Motherwell *et al.*, 2002). However, although structures can sometimes be predicted with reasonable accuracy, no single methodology is reliable. The basic problem is that there are many more minima lying within 5 kJ mol^{-1} of the global minimum than was expected even 10 years ago. This typically leads to a set of about 30 possible polymorphs for a 25-atom molecule, and the final decisions on energy alone at zero temperature are often made on the basis of very small (*ca* 0.5 kJ mol^{-1}) energy differences. Attempts have been made to include entropy due to vibrational motion in the crystal, which sometimes improves prediction (van Eijck *et al.*, 2001). These small differences in energy depended crucially on the force fields used and these force fields have over the years been constructed and validated against the CSD (see *e.g.* Filippini & Gavezzotti, 1992*b*).

The calculation of the electrostatic component of the energy has often shown that a multipole description is more

effective in reproducing CSD structures (Beyer *et al.*, 2001). Other methods have sought to use:

   (i) statistical potentials based directly on the interatomic distances between molecules in the CSD (Hofmann & Lengauer, 1997),

   (ii) direct fitting of distance frequency curves (Motherwell, 2001) or

   (iii) quasi-potentials based on the fitting of interaction scattergrams between chemical groups in the CCDC's IsoStar knowledge base (Lommerse *et al.*, 2000),

but none of these methods were successful in the blind tests. However, since crystallization is a kinetic process, it is accepted that the single lowest free-energy structure predicted may not always be the polymorph observed experimentally. Thus, the future role of the CSD may be to provide larger samples of crystal structures for molecules similar to the target molecule and thus argue, from patterns and analogy between structures, that certain members of a low-energy set of possible structures are more likely to exist under given conditions.

12. Structure determination from powder data

The determination of crystal structures from experimental X-ray powder diffraction data has become routine for small molecules, when a reasonable molecular model can be constructed. The structures of many thousands of trial crystals are generated and the simulated powder pattern fitted against the observed data, with global optimization techniques similar to those used in crystal structure prediction being used to seek the best fit (David *et al.*, 1998; Engel *et al.*, 1999). The CSD provides a valuable check on the geometry of the molecular model and sometimes, *e.g.* in the case of flexible acyclic and cyclic systems with almost equal internal energy, the observed CSD conformations may provide an essential guide to the areas of parameter space that need to be considered. Owing to the loss of data in the one-dimensional powder pattern with overlapping peaks, there is a complexity limit on the maximum number (*ca* 20) of independent parameters that may be fitted unambiguously to the data. So for the case of a single molecule in the asymmetric unit with 6 degrees of freedom it is usually found that up to 14 torsion angles may be treated as parameters. An example where solution was only possible when restricting certain torsion angles to CSD observed ranges is tetracaine hydrochloride (Nowell *et al.*, 2002) with a total of 18 parameters, of which nine are torsion angles and two torsion angles were restricted.

13. Analysis of structural precision

The earliest work which used the CSD data to assess the precision of crystal structures was carried out by Taylor & Kennard (1986). They carried out a statistical analysis of 100 structures that had been determined independently by two different research groups and found that the s.u.'s of non-H atom positional parameters were invariably underestimated, with the positions of heavier atoms being less reliable than for light atoms and that the s.u.'s of cell parameters were grossly underestimated. Allen *et al.* (1995*a,b*) returned to this theme by exploring the relationship between the mean isotropic s.u. for different element types and various experimental and structural factors, *e.g.* the R factor, the atomic number of the element, the atomic number of the heaviest element in the structure, the $\sin\theta/\lambda$ limit for data collection *etc.* and multiple linear regression techniques were used to generate predictive equations for isotropic s.u. values. While the predictions were generally acceptable across a broad range of structures, the findings of Taylor & Kennard (1986) were reinforced.

14. Applications in chemical information science

The CSD has been used both in-house and externally to develop improved, and sometimes novel, methods of searching and comparing two-dimensional and three-dimensional structure representations. For example, two-dimensional substructure searches within *QUEST*3D and *ConQuest* depend for their efficiency on a set of bit-screens, *i.e.* bit-encoded yes/no information about the chemical structure of each molecule or ion in the database. Each bit in the map records the presence or absence of a specific substructural feature, such an atom having a specific connection pattern, a ring of a specific size and constitution, a specific functional group *etc.*, as described by Allen, Davies *et al.* (1991) and Allen (2002). Each substructure query is then analysed using the same software to generate a query bitmap which is compared with the bitmap for each database entry, a very rapid computational process. Only those CSD entries, typically less than 5%, which contain all of the chemical features of the query need to be further analysed using the more computationally intensive atom-by-atom, bond-by-bond substructure matching procedure. Using a similar philosophy, Poirette *et al.* (1991, 1993) have examined the use of valence-angle data and generalized torsion-angle data as a basis for screening systems for specific types of three-dimensional queries posed to databases of three-dimensional molecular structures, and a subset of these screens is implemented within the current CSD system.

   During the 1980s, the concept of two-dimensional molecular similarity was pioneered by Willett and co-workers (Willett, 1986) and has been applied with some success as a method for locating new pharmaceutical lead compounds. Essentially, the two-dimensional screen bitmaps of a molecule of interest are compared pairwise with the bitmaps stored in a database of two-dimensional molecules, and database molecules are ranked and examined according to their similarity to the query molecule using a variety of similarity coefficients. A number of attempts have been made to extend this philosophy to detect three-dimensional molecular similarities in the CSD, either by using bit-encoded representations of three-dimensional structure or by comparing molecular distance matrices (Pepperell & Willett, 1991; Bath *et al.*, 1994, 1995).

15. Conclusions

The aim of this review has been exactly that which underpinned the corresponding review (Allen & Kennard, 1987) in the first collection of papers on crystallographic databases (Allen, Bergerhoff & Sievers, 1987), namely to provide a commentary on the published research applications of the CSD, as recorded in the CCDC's (then embryo) CSDUse database. This commentary is, however, very different since:

(i) the sheer number of publications now makes it impossible to cover more than a selected, and maybe subjective, subset, and

(ii) the scientific breadth covered by the publications has expanded significantly, leading to a greater difficulty in organizing the material.

Indeed, many of the sections and subsections of this survey have themselves become the subject matter for monographs and topical reviews in their own right.

The major difference, though, is in the nature of the research applications themselves. CSD applications have always aimed to generate new knowledge in crystallography and structural chemistry and in 1987 that was perhaps the ultimate aim. Today, however, we see a more directed approach to CSD-based research – the generation of new knowledge for a purpose: e.g.

(i) to determine conformational preferences for use in molecular modelling or in solving new crystal structures from powder diffraction data,

(ii) to recognize, categorize and use information about non-bonded interactions to advance research in the supramolecular arena, and

(iii) to try to unravel the complexities of crystal packing and ultimately, perhaps, to catalyse real improvements in the field of ab initio crystal structure prediction.

It is this pathway from data to knowledge to applications (Allen et al., 1990) that is now clearly visible in the recent research activities catalogued in the DBUse database and, as with all of the crystallographic databases, the CSD is now of significant value in a research environment in which informatics is now an increasingly important component. In the future, links between the CSD and other databases and, importantly, much improved interfaces between CSD information and a wide variety of computational procedures, will be essential for its more complete integration into the informatics environment.

References

Aäkeroy, C. B., Evans, T. A., Seddon, K. R. & Palinko, I. (1999). *New J. Chem.* pp. 145–152.
Allen, F. H. (2002). *Acta Cryst.* B**58**, 380–388.
Allen, F. H., Baalham, C. A., Lommerse, J. P. M. & Raithby, P. R. (1998). *Acta Cryst.* B**54**, 320–329.
Allen, F. H., Bergerhoff, G. & Sievers, R. (1987). Editors. *Crystallographic Databases.* IUCr: Chester.
Allen, F. H., Bird, C. M., Rowland, R. S., Harris, S. E. & Schwalbe, C. H. (1995). *Acta Cryst.* B**51**, 1068–1081.
Allen, F. H., Bird, C. M., Rowland, R. S. & Raithby, P. R. (1997). *Acta Cryst.* B**53**, 696–701.
Allen, F. H., Cole, J. C. & Howard, J. A. K. (1995a). *Acta Cryst.* A**51**, 95–111.
Allen, F. H., Cole, J. C. & Howard, J. A. K. (1995b). *Acta Cryst.* A**51**, 112–121.
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. & Watson, D. G. (1991). *J. Chem. Inf. Comput. Sci.* **31**, 187–204.
Allen, F. H., Doyle, M. J. & Auf der Heyde, T. P. E. (1991). *Acta Cryst.* B**47**, 412–428.
Allen, F. H., Harris, S. E. & Taylor, R. (1996). *J.Comput.-Aided Mol. Des.* **10**, 247–254.
Allen, F. H., Howard, J. A. K. & Pitchford, N. A. (1993). *Acta Cryst.* B**49**, 910–928.
Allen, F. H., Howard, J. A. K. & Pitchford, N. A. (1996). *Acta Cryst.* B**52**, 882–891.
Allen, F. H. & Kennard, O. (1987). *Crystallographic Databases*, edited by F. H.Allen, G. Bergerhoff & R. Sievers, pp. 55–76. IUCr: Chester.
Allen, F. H., Kennard, O. & Taylor, R. (1983). *Acc. Chem. Res.* **16**, 146–153.
Allen, F. H., Kennard, O., Watson, D. G., Brammer, L., Orpen, A. G. & Taylor, R. (1987). *J. Chem. Soc. Perkin Trans. 2*, pp. S1–S19.
Allen, F. H., Lommerse, J. P. M., Hoy, V. J., Howard, J. A. K. & Desiraju, G. R. (1996). *Acta Cryst.* B**52**, 734–745.
Allen, F. H., Lommerse, J. P. M., Hoy, V. J., Howard, J. A. K. & Desiraju, G. R. (1997). *Acta Cryst.* B**53**, 1006–1016.
Allen, F. H., Motherwell, W. D. S., Raithby, P. R., Shields, G. P. & Taylor, R. (1999). *New J. Chem.* pp. 25–34.
Allen, F. H., Rowland, R. S., Fortier, S. & Glasgow, J. I. (1990). *Tetrahedron Comput. Methods* **3**, 757–774.
Amos, R. D. (1994). *The Cambridge Analytical Derivatives Package – CADPAC6.0.* Chemistry Department, Cambridge University, Lensfield Road, Cambridge, England.
Amos, R. D., Handy, N. C., Jones, P. G., Kirby, A. J., Parker, J. K., Percy, J. M. & Su, M. D. (1992). *J. Chem. Soc. Perkin Trans. 2*, pp. 549–558.
Aullon, G., Bellamy, D., Orpen, G. A., Brammer, L. & Bruton, E. A. (1998). *Chem. Commun.* pp. 653–654.
Bath, P. A., Poirette, A. R., Willett, P. & Allen, F. H. (1995). *J. Chem. Inf. Comput. Sci.* **35**, 714–716.
Bath, P. A., Willett, P. & Allen, F. H. (1994). *J. Chem. Inf. Comput. Sci.* **34**, 141–147.
Baur, W. H. & Kassner, D. (1992). *Acta Cryst.* B**48**, 356–369.
Belsky, V. K., Zorkaya, O. N. & Zorky, P. M. (1995). *Acta Cryst.* A**51**, 473–481.
Berman, H. M., Battistuz, T., Bhat, T. N., Bluhm, W. F., Bourne, P. E., Burkhardt, K., Feng, Z., Gilliland, G. L., Iype, L., Jain, S., Fagan, P., Marvin., J., Ravichanran, V., Schneider, B., Thanki, N., Padilla, D., Weissig, H., Westbrook, J. D. & Zardecki, C. (2002). *Acta Cryst.* B**58**, 899–907.
Berman, H., Westbrook, J., Feng, Z., Iype, L., Schneider, B. & Zardecki, C. (2002). *Acta Cryst.* B**58**, 889–898.
Bernstein, J., Davis, R. E., Shimoni, L. & Chang, N.-L. (1995). *Angew. Chem. Int. Ed. Engl.* **34**, 1555–1573.
Bernstein, J., Etter, M. C. & MacDonald, J. M. (1990). *J. Chem. Soc. Perkin Trans. 2*, pp. 695–704.
Bertolasi, V., Ferretti, V., Gilli, P., Issa, Y. M. & Sherif, O. E. (1993). *J. Chem. Soc. Perkin Trans 2*, pp. 2223–2228.
Bertolasi, V., Gilli, P., Ferretti, V. & Gilli, G. (1996). *Chem. Eur. J.* **2**, 925–934.

Beyer, T., Day, G. M. & Price, S. L. (2001). *J. Am. Chem. Soc.* **123**, 5086–5094.

Bilton, C., Howard, J. A. K., Shields, G. P. & Allen, F. H. (2000). *Acta Cryst.* B**56**, 849–856.

Bock, H., Nagel, N. & Nather, C. (1998). *Z. Naturforsch. Teil B*, **53**, 1389–1400.

Böhm, H.-J., Brode, S., Hesse, U. & Klebe, G. (1998). *Chem. Eur. J.* **2**, 1509–1513.

Bondi, A. (1964). *J. Phys. Chem.* **68**, 441–452.

Braga, D., Maini, L., Grepioni, F., De Cian, A., Felix, O., Fischer, J. & Hosseini, M. W. (2000). *New J. Chem.* pp. 547–553.

Brammer, L., Bruton, E. A. & Sherwood, P. (1999). *New J. Chem.* **23**, 965–968.

Brammer, L., Bruton, E. A. & Sherwood, P. (2001). *Cryst. Growth Des.* **1**, 277–290.

Brammer, L., Zhao, D., Ladipo, F. T. & Braddock-Wilking, J. (1995). *Acta Cryst.* B**51**, 632–640.

Brock, C. P. & Dunitz, J. D. (1994a). *Chem. Mater.* **6**, 1118–1127.

Brock, C. P. & Dunitz, J. D. (1994b). *Mol. Cryst. Liq. Cryst. Sci. Technol. Sect. A*, **242**, 61–69.

Brock, C. P. & Minton, R. P. (1989). *J. Am. Chem. Soc.* **111**, 4586–4593.

Brock, C. P., Schweizer, W. B. & Dunitz, J. D. (1991). *J. Am. Chem. Soc.* **113**, 9811–9820.

Brown, I. D. & McMahon, B. (2002). *Acta Cryst.* B**58**, 317–324.

Bruno, I. J., Cole, J. C., Edgington, P. R., Kessler, M., Macrae, C. F., McCabe, P., Pearson, J. & Taylor, R. (2002). *Acta Cryst.* B**58**, 389–397.

Bruno, I. J., Cole, J. C., Lommerse, J. P. M., Rowland, R. S., Taylor, R. & Verdonk, M. L. (1997). *J. Comput.-Aided Mol. Des.* **11**, 525–537.

Bürgi, H.-B. & Dubler-Steudler, K. C. (1988a). *J. Am. Chem. Soc.* **110**, 4953–4957.

Bürgi, H.-B. & Dubler-Steudler, K. C. (1988b). *J. Am. Chem. Soc.* **110**, 7291–7299.

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

Bürgi, H.-B. & Dunitz, J. D. (1994). *Structure Correlation*, edited by H.-B. Bürgi & J. D. Dunitz, pp. 163–204. Weinheim: VCH Publishers.

Burling, F. T. & Goldstein, B. M. (1992). *J. Am. Chem. Soc.* **114**, 2313–2320.

Burling, F. T. & Goldstein, B. M. (1993). *Acta Cryst.* B**49**, 738–744.

Cambridge Crystallographic Data Centre (1994). *Quest3D*. Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge, England.

Cambridge Crystallographic Data Centre (1995). *Vista*. Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge, England.

Chertanova, L. & Pascard, C. (1996). *Acta Cryst.* B**52**, 677–684.

Ciunik, Z. & Desiraju, G. R. (2001). *Chem. Commun.* pp. 703–704.

Clowney, L., Jain, S. C., Srinivasan, A. R., Westbrook, J., Olson, W. K. & Berman, H. M. (1996). *J. Am. Chem. Soc.* **118**, 509–518.

Cole, J. C., Taylor, R. & Verdonk, M. L. (1998). *Acta Cryst.* D**54**, 1183–1193.

Cole, J. C., Yao, J. W., Shields, G. P., Motherwell, W. D. S., Allen, F. H. & Howard, J. A. K. (2000). *Acta Cryst.* B**57**, 88–94.

Conklin, D., Fortier, S., Glasgow, J. I. & Allen, F. H. (1996). *Acta Cryst.* B**52**, 535–549.

Corey, E. J. (1967). *Pure Appl. Chem.* **14**, 19–37.

Cotton, F. A., Daniels, L. M., Jordan IV, G. T. & Murillo, C. A. (1997). *Chem. Commun.* pp. 1673–1674.

Cremer, D. & Pople, J. A. (1975). *J. Am. Chem. Soc.* **97**, 1354–1358.

Dance, I. & Scudder, M. (1996). *Chem. Eur. J.* **2**, 481–486.

David, W. I. F., Shankland, K. & Shankland, N. (1998). *Chem. Commun.* **8**, 931–932.

Deane, C. M., Allen, F. H., Taylor, R. & Blundell, T. L. (1999). *Protein Engng*, **12**, 1025–1028.

Derewenda, Z. S., Lee, L. & Derewenda, U. (1995). *J. Mol. Biol.* **252**, 248–262.

Desiraju, G. R. (1989). *Crystal Engineering: The Design of Organic Solids.* New York: Academic Press.

Desiraju, G. R. (1991a). *Acc. Chem. Res.* **24**, 270–276.

Desiraju, G. R. (1991b). *Chem. Commun.* pp. 426–428.

Desiraju, G. R. (1995). *Angew. Chem. Int. Ed. Engl.* **34**, 2311–2327.

Desiraju, G. R. (1996). *Acc. Chem. Res.* **29**, 441–449.

Desiraju, G. R., Calabrese, J. C. & Harlow, R. L. (1991). *Acta Cryst.* B**47**, 77–86.

Desiraju, G. R. & Steiner, T. (1999). *The Weak Hydrogen Bond in Structural Chemistry and Biology.* Oxford University Press.

Deslongchamps, P. (1983). *Stereoelectronic Effects in Organic Chemistry.* Oxford: Pergamon Press.

Donohue, J. (1968). *Structural Chemistry and Molecular Biology*, edited by W. Davidson & E. Rich, pp. 443–465. San Francisco: W. H. Freeman.

Dunitz, J. D. (1979). *X-ray Analysis and the Structure of Organic Molecules.* Ithaca, New York: Cornell University Press.

Dunitz, J. D. & Taylor, R. (1997). *Chem. Eur. J.* **3**, 83–90.

Eijck, B. P. van, Mooij, W. T. M. & Kroon, J. (2001). *J. Phys. Chem B*, **105**, 10573–10578.

Engel, G. E., Wilke, S., Konig, O., Harris, K. D. M. & Leusen, F. J. J. (1999). *J. Appl. Cryst.* **32**, 1169–1179.

Engh, R. A. & Huber, R. (1991). *Acta Cryst.* A**47**, 392–398.

Etter, M. C. (1990). *Acc. Chem. Res.* **23**, 120–126.

Etter, M. C., MacDonald, J. M. & Bernstein, J. (1990). *Acta Cryst.* B**46**, 256–262.

Etter, M. C. & Reutzel, S. M. (1991). *J. Am. Chem. Soc.* **113**, 2586–2598.

Fan, M.-F., Lin, Z., McGrady, J. E. & Mingos, D. M. P. (1996). *J. Chem. Soc. Perkin Trans. 2*, pp. 563–568.

Ferretti, V., Gilli, P., Bertolasi, V. & Gilli, G. (1996). *Cryst. Rev.* **5**, 3–98.

Filippini, G. & Gavezzotti, A. (1992a). *Acta Cryst.* B**48**, 230–234.

Filippini, G. & Gavezzotti, A. (1992b). *Mol. Cryst. Liq. Cryst.* **238**, 37–41.

Gautham, N. (1992). *Acta Cryst.* B**48**, 337–338.

Gavezzotti, A. (1994). *Acc. Chem. Res.* **27**, 309–314.

Gavezzotti, A. (1998). *Cryst. Rev.* **7**, 5–121.

Gavezzotti, A. & Desiraju, G. R. (1988). *Acta Cryst.* B**44**, 427–434.

Gavezzotti, A. & Filippini, G. (1994). *J. Phys. Chem.* **98**, 4831–4837.

Gavezzotti, A. & Filippini, G. (1995). *J. Am. Chem. Soc.* **117**, 12299–12305.

Gelbin, A., Schneider, B., Clowney, L., Hsieh, S.-H., Olson, W. K. & Berman, H. M. (1996). *J. Am. Chem. Soc.* **118**, 519–528.

Gibb, C. L. D., Stevens, E. D. & Gibb, B. C. (2001). *J. Am. Chem. Soc.* **123**, 5849–5850.

Gilli, G., Bellucci, F., Ferretti, V. & Bertolasi, V. (1989). *J. Am. Chem. Soc.* **111**, 1023–1030.

Gilli, P., Bertolasi, V., Ferretti, V. & Gilli, G. (2000). *J. Am. Chem. Soc.* **122**, 10405–10417.

Glusker, J. P., Lewis, M. & Rossi, M. (1994). *Crystal Structure Analysis for Chemists and Biologists.* Weinheim: VCH Publishers.

Görbitz, C. H. & Etter, M. C. (1992). *Int. J. Pept. Protein Res.* **39**, 93–110.

Görbitz, C. H. & Hersleth, H.-P. (2000). *Acta Cryst.* B**56**, 526–534.

Grabowski, S. J. (1998). *Tetrahedron*, **54**, 10153–10160.

Grell, J., Bernstein, J. & Tinhofer, G. (1999). *Acta Cryst.* B**55**, 1030–1043.

Hall, S. R., Allen, F. H. & Brown, I. D. (1991). *Acta Cryst.* A**47**, 655–685.

Hayes, I. C. & Stone, A. J. (1984). *J. Mol. Phys.* **53**, 83–105.

Hofmann, D. W. M. & Lengauer, T. (1997). *Acta Cryst.* A**53**, 225–235.

Holden, J. R., Du, Z. & Ammon, H. L. (1993). *J. Comput. Chem.* **14**, 422–437.

Howard, J. A. K., Hoy, V. J., O'Hagan, D. & Smith, G. T. (1996). *Tetrahedron*, **52**, 12613–12622.

Jeffrey, G. A. (1997). *An Introduction to Hydrogen Bonding*. Oxford University Press.

Jeffrey, G. A. & Maluszynska, H. (1982). *Int. J. Biol. Macromol.* **4**, 173–185.

Jeffrey, G. A. & Maluszynska, H. (1986). *J. Mol. Struct.* **147**, 127–142.

Jeffrey, G. A. & Saenger, W. (1991). *Hydrogen Bonding in Biological Structures.* Berlin: Springer Verlag.

Joris, L., Schleyer, P. & Gleiter, R. (1968). *J. Am. Chem. Soc.* **90**, 327–336.

Kirby, A. J. (1994). *Adv. Phys. Org. Chem.* **29**, 87–183.

Kitaigorodskii, A. I. (1973). *Molecular Crystals and Molecules.* New York: Academic Press.

Klebe, G. (1994). *J. Mol. Struct.* **308**, 53–89.

Kroon, J., Kanters, J. A., van Duijneveldt–van de Rijdt, J. G. C. M., van Duijneveldt, F. B. & Vliegenthart, J. A. (1975). *J. Mol. Struct.* **24**, 109–129.

Kuleshova, L. N. & Zorkii, P. M. (1980). *Acta Cryst.* B**36**, 2113–2115.

Lehn, J.-M. (1988). *Angew. Chem. Int. Ed. Engl.* **27**, 90–112.

Levitt, M. & Perutz, M. (1988). *J. Mol. Biol.* **201**, 751–754.

Lin, S. Y., Okaya, Y., Chiou, D. M. & Le Noble, W. J. (1982). *Acta Cryst.* B**38**, 1669–1673.

Llamas-Saiz, A. L., Foces-Foces, C., Mo, O., Yanez, M. & Elguero, J. (1992). *Acta Cryst.* B**48**, 700–713.

Lommerse, J. P. M. & Cole, J. C. (1998). *Acta Cryst.* B**54**, 316–319.

Lommerse, J. P. M., Motherwell, W. D. S., Ammon, H. L., Dunitz, J. D., Gavezzotti, A., Hofmann, D. W. M., Leusen, F. J. J., Mooij, W. T. M., Price, S. L., Schweizer, B., Schmidt, M. U., van Eijck, B. P., Verwer, P. & Williams, D. E. (2000). *Acta Cryst.* B**56**, 697–714.

Lommerse, J. P. M., Price, S. L. & Taylor, R. (1997). *J. Comput. Chem.* **18**, 757–780.

Lommerse, J. P. M., Stone, A. J., Taylor, R. & Allen, F. H. (1996). *J. Am. Chem. Soc.* **118**, 3108–3116.

Maccallum, P. H., Poet, R. & Milner-White, E. J. (1995*a*). *J. Mol. Biol.* **248**, 361–373.

Maccallum, P. H., Poet, R. & Milner-White, E. J. (1995*b*). *J. Mol. Biol.* **248**, 374–384.

Marsh, R. E. (1995). *Acta Cryst.* B**51**, 897–907.

Marsh, R. E., Kapon, M., Hu, S. & Herbstein, F. H. (2002). *Acta Cryst.* B**58**, 62–77.

Marsh, R. E., Schomaker, V. & Herbstein, F. H. (1998). *Acta Cryst.* B**54**, 921–924.

Mascal, M. (1998). *Chem. Commun.* pp. 303–304.

McCrone, W. C. (1965). *Physics and Chemistry of the Solid State*, Vol. 2, p. 725. New York: Interscience.

Motherwell, W. D. S. (1997). *Acta Cryst.* B**53**, 726–736.

Motherwell, W. D. S. (2001). *Mol. Cryst. Liq. Cryst.* **356**, 559–567.

Motherwell, W. D. S., Ammon, H. L., Dunitz, J. D., Dzyabchenko, A., Erk, P., Gavezzotti, A., Hofmann, D. W. M., Leusen, F. J. J., Lommerse, J. P. M., Mooij, W. T. M., Price, S. L., Scheraga, H., Schweizer, B., Schmidt, M. U., van Eijck, B. P., Verwer, P. & Williams, D. E. (2002). Accepted for publication.

Motherwell, W. D. S., Shields, G. P. & Allen, F. H. (1999). *Acta Cryst.* B**55**, 1044–1056.

Motherwell, W. D. S., Shields, G. P. & Allen, F. H. (2000). *Acta Cryst.* B**56**, 466–473.

Murray-Rust, P., Bürgi, H.-B. & Dunitz, J. D. (1975). *J. Am. Chem. Soc.* **97**, 921–922.

Murray-Rust, P. & Glusker, J. P. (1984). *J. Am. Chem. Soc.* **106**, 1018–1025.

Murray-Rust, P. & Motherwell, W. D. S. (1978). *Acta Cryst.* B**34**, 2534–2546.

Nangia, A. (2002). *CrystEngComm*, 17.

Nangia, A. & Desiraju, G. R. (1999). *Chem. Commun.* pp. 605–606.

Nishio, M., Hirota, M. & Umezawa, Y. (1998). *The CH/π Interaction.* New York: Wiley-VCH.

Nobeli, I., Price, S. L., Lommerse, J. P. M. & Taylor, R. (1997). *J. Comput. Chem.* **18**, 2060–2074.

Nowell, H., Attfield, J. P., Cole, J. C., Cox, P. J., Shankland, K. & Maginn, S. J. (2002). *New J. Chem.* In the press.

Nyburg, S. C. & Faerman, C. H. (1985). *Acta Cryst.* B**41**, 274–279.

Orpen, A. G. (1993). *Chem. Soc. Rev.* pp. 191–197.

Orpen, A.G. (2002). *Acta Cryst.* B**58**, 398–406.

Orpen, A. G., Brammer, L., Allen, F. H., Kennard, O., Watson, D. G. & Taylor, R. (1989). *J. Chem. Soc. Dalton Trans.* pp. S1–S83.

Panunto, T. W., Urbanczyk-Lipkowska, Z., Johnson, R. & Etter, M. C. (1987). *J. Am. Chem. Soc.* **109**, 7786–7797.

Pauling, L. (1939). *The Nature of the Chemical Bond.* Ithaca, New York: Cornell University Press.

Pepperell, C. A. & Willett, P. (1991). *J. Comput.-Aided Mol. Des.* **5**, 455–474.

Pirard, B., Baudoux, G. & Durant, F. (1995). *Acta Cryst.* B**51**, 103–107.

Poirette, A. R., Willett, P. & Allen, F. H. (1991). *J. Mol. Graphics*, **9**, 203–217.

Poirette, A. R., Willett, P. & Allen, F. H. (1993). *J. Mol. Graphics*, **11**, 2–14.

Prasanna, M. D. & Row, T. N. G. (2000). *Cryst. Engng*, **3**, 135–154.

Price, S. L., Stone, A. J., Lucas, J., Rowland, R. S. & Thornley, A. E. (1994). *J. Am. Chem. Soc.* **116**, 4910–4918.

Raithby, P. R., Shields, G. P. & Allen, F. H. (1997*a*). *Acta Cryst.* B**53**, 476–489.

Raithby, P. R., Shields, G. P. & Allen, F. H. (1997*b*). *Acta Cryst.* B**53**, 241–251.

Rappoport, Z., Biali, S. E. & Kaftory, M. (1990). *J. Am. Chem. Soc.* **112**, 7742–7750.

Redman, J., Willett, P., Allen, F. H. & Taylor, R. (2001). *J. Appl. Cryst.* **34**, 375–389.

Rowland, R. S. & Taylor, R. (1996). *J. Phys. Chem.* **100**, 7384–7391.

Rzepa, H. S., Smith, M. H. & Webb, M. L. (1994). *J. Chem. Soc. Perkin Trans 2*, pp. 703–707.

Sarma, J. A. R. P. & Desiraju, G. R. (1986). *Acc. Chem. Res.* **19**, 222–228.

Steiner, T. (1995). *Chem. Commun.* pp. 1331–1332.

Steiner, T. (1998*a*). *Acta Cryst.* B**54**, 456–463.

Steiner, T. (1998*b*). *Chem. Commun.* pp. 411–412.

Steiner, T. (2000). *Acta Cryst.* B**56**, 673–676.

Steiner, T. (2001). *Acta Cryst.* B**57**, 103–106.

Steiner, T. & Desiraju, G. R. (1998). *Chem. Commun.* pp 891–892.

Steiner, T. & Saenger, W. (1991). *Acta Cryst.* B**47**, 1022–1023.

Steiner, T. & Saenger, W. (1992*a*). *Acta Cryst.* B**48**, 551–552.

Steiner, T. & Saenger, W. (1992*b*). *Acta Cryst.* B**48**, 819–827.

Steiner, T. & Saenger, W. (1994). *Acta Cryst.* B**50**, 348–357.

Steiner, T., Starikov, E. B., Amado, A. M. & Teixeira-Dias, J. J. C. (1995). *J. Chem. Soc. Perkin Trans. 2*, pp. 1312–1326.

Steiner, T., Starikov, E. B. & Tamm, M. (1996). *J. Chem. Soc. Perkin Trans. 2*, pp. 67–71.

Stone, A. J. (1993). *Chem. Phys. Lett.* **211**, 101–109.

Sutton, L. E. (1956). *Tables of Interatomic Distances and Configuration in Molecules and Ions*. London: The Chemical Society, Special Publication No. 11.

Taylor, R. (1986*a*). *J. Mol. Graphics*, **4**, 123–131.

Taylor, R. (1986*b*). *J. Appl. Cryst.* **19**, 90–91.

Taylor, R. (2002). *Acta Cryst.* B**58**, 879–888.

Taylor, R. & Allen, F. H. (1994). *Structure Correlation*, edited by H.-B. Bürgi & J. D. Dunitz, pp. 111–162. Weinheim, Germany: VCH Publishers.

Taylor, R. & Kennard, O. (1982). *J. Am. Chem. Soc.* **104**, 5063–5070.

Taylor, R. & Kennard, O. (1984). *Acc. Chem. Res.* **17**, 320–326.

Taylor, R. & Kennard, O. (1986). *Acta Cryst.* B**42**, 112–120.

Taylor, R. & Macrae, C. F. (2001). *Acta Cryst.* B**57**, 815–827.

Taylor, R., Mullaley, A. & Mullier, G. W. (1990). *Pestic. Sci.* **29**, 197–213.

Thalladi, V. R., Gehrcke, A. & Boese, R. (2000). *New J. Chem.* **24**, 463–470.

Thalladi, V. R., Weiss, H.-C., Blaser, D., Boese, R., Nangia, A. & Desiraju, G. R. (1998). *J. Am. Chem. Soc.* **120**, 8702–8710.

Thallapally, P. K. & Nangia, A. (2001). *CrystEngComm*, 27.

Vela, M. J. & Foxman, B. M. (2000). *Cryst. Engng*, **3**, 11–31.

Verwer, P. & Leusen, F. J. J. (1998). *Rev. Comput. Chem.* **12**, 327–365.

Viswamitra, M. A., Radhakrishnan, R., Bandekar, J. & Desiraju, G. R. (1993). *J. Am. Chem. Soc.* **115**, 4868–4869.

Willett, P. (1986). *Similarity and Clustering in Chemical Information Systems.* Letchworth, England: Research Studies Press.

Wilson, A. J. C. (1988). *Acta Cryst.* A**44**, 715–724.

Wilson, A. J. C. (1990). *Acta Cryst.* A**46**, 742–754.

Wilson, A. J. C. (1993). *Acta Cryst.* A**49**, 795–806.

Yao, J. W., Cole, J. C., Pidcock, E., Allen, F. H., Howard, J. A. K. & Motherwell, W. D. S. (2002). Submitted for publication.

Ziao, N., Graton, J., Laurence, C. & Le Questel, J.-Y. (2001). *Acta Cryst.* B**57**, 850–858.