Supramolecular Chemistry
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713649759

From data to knowledge—Use of the Cambridge Structural Database for studying molecular interactions

Olga Kennard[a]
[a] Cambridge Crystallographic Data Centre, Cambridge, UK

To cite this Article Kennard, Olga(1993) 'From data to knowledge—Use of the Cambridge Structural Database for studying molecular interactions', Supramolecular Chemistry, 1: 3, 277 — 295
To link to this Article: DOI: 10.1080/10610279308035171
URL: http://dx.doi.org/10.1080/10610279308035171

SUPRAMOLECULAR CHEMISTRY, Vol. 1, pp. 277–295
Reprints available directly from the publisher
Photocopying permitted by license only

From data to knowledge—Use of the Cambridge Structural Database for studying molecular interactions

OLGA KENNARD

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

(Received July 28, 1992)

The ordered arrangement of molecules and ions in crystal structures is the result of non-bonded interactions, albeit modulated in the crystalline state by random crystal packing forces. Each individual crystal structure gives but a snapshot of these interactions. It is only by examining a large number of structures containing similar molecules or functional groups that we can gain any generalized knowledge about the rules governing molecular interactions. Indeed if we knew these rules it should be possible to predict crystal structures likely to be formed by specific molecules, whereas at present each new crystal structure requires a *de novo* experimental solution. We would also be able to evaluate the relative contribution of non-bonded and crystal packing forces and extrapolate to the interactive behaviour of molecules in solution and, most importantly, in the environments encountered in biological systems. A key to these studies is the Cambridge Structural Database (CSD) system where experimental results from over 100,000 individual structure determinations are stored in computer readable form. The system also provides the computational tools needed to locate the relevant structures and analyse the numerical data using a variety of statistical techniques. This paper describes the latest Version of the CSD system (Version 5) released Autumn 1992. The use of the system is illustrated by several examples, such as the preferential interaction of certain functional groups, the study of the C—H...O bond and drug-DNA interactions.

INTRODUCTION

Synthetic organic chemistry has, traditionally, been the art of one step transformations from a set of starting materials to a target molecule. It has essentially involved atom by atom, group by group procedures using catalysts and other reagents to make or break covalent bonds and has relied heavily on the incorporation of protecting groups. Currently this art is undergoing a profound conceptual revolution. It is moving away from these traditional methods and aiming instead at the assembly of new molecules from molecular components utilizing spatially directed non-bonded interactions, templates, and other three-dimensional devices. The impetus for this change is based on the insight gained through biochemical, chemical and structural studies of the synthetic pathways used in nature. Synthesis and transformation in biological systems is based on self replication and self assembly, using information encoded in the system itself. Compounds are assembled molecule by molecule from molecular components by extensive use of non-covalent intermolecular interactions. These interactions are often spatially directed and the three-dimensional structure of the components and their relative orientation is a crucial aspect of the molecular assembly. The bonding interactions persist, albeit transformed, in the final product and are the basis for information transfer at a molecular level.

This new direction in chemistry is leading not only to the synthesis of traditional compounds with efficiency resembling that of natural systems, but also the ready and elegant assembly of molecules carrying in their structure further information for self-replication. This type of thinking also plays an important role in the design of novel compounds for the transfer of molecular information for use in molecular electronics, in molecular sensors and in a host of other applications. Molecular recognition is the key concept in these new developments and involves the understanding of non-bonded interactions such as hydrogen bonds, van der Waals forces, charge interactions, solvation effects and thermodynamic concepts of entropy and enthalpy.

Evidently the more we understand the principles underlying molecular recognition the more likely we are to be able to harness these to our own purposes. Much of what we know about molecular recognition at the enzyme level is derived from X-ray diffraction analysis of macromolecules. The many studies of small molecule crystal structures, however, also contain a

wealth of relevant information. Indeed, if you stop to think about it, any crystal, even one composed of the simplest molecules, is an example of a supramolecule. The smallest crystal contains many millions of molecules held together in exquisite order by the very forces we are seeking to understand and exploit for chemial goals. It is a sobering thought, however, that our present knowledge of the intermolecular forces responsible for these molecular organizations is not sufficient to predict the crystal structure of even the simplest compound. There is as much to be learnt from the data currently available and hence the title of this paper 'from data to knowledge'.

THE CAMBRIDGE STRUCTURAL DATABASE (CSD)

The experimental tool used here for studying intermolecular interactions and for improving our understanding of the principles underlying these interactions is the CSD.[1]

Table 1 Statistical overview of the Cambridge Structural Database, October 1992

(*a*) General data

Number of entries	102,589
Number of compounds	91,325
Number of different literature sources	634
Number of entries with 3D-co-ordinates	90,315
Number of error-free 3D co-ordinate sets	88,565
Number of entries with errors corrected by CCDC	11,520
Number of atoms with 3D co-ordinates	4,737,722
Number of X-ray studies	101,803
Number of neutron studies	786
Absolute configuration by X-ray methods	2635
Number of low temperature studies	10,795
Number of entries with perfect connectivity matching	76,537
Number of entries with partial connectivity matching	8793
Number of entries with no connectivity matching	2990
Number of entries for which matching is impossible	14,269

(*b*) Distribution of entries in the different chemical classes

Classes	*Types of Compounds*	*Number of entries*	*% of database*
1–12	Simple aliphatics	5471	5.3
13–23	Monocyclic hydrocarbons	5223	5.1
24–31	Polycyclic hydrocarbons	4337	4.2
32–42	Heterocyclic compounds	16,543	16.1
43–59	Natural products	12,356	12.0
60–61	Molecular complexes, clathrates	2811	2.7
62–70	Main group compounds	11,837	11.5
71–75	Transition metal complexes (sigma, pi)	17,603	17.2
76–86	Transition metal complexes (co-ordination)	26,408	25.7

(*c*) Precision of structural results

Reliability factor, R	*Precision*	*Number of entries*	*% of database*
1–3	Exceptional	6778	6.6
3–4	Very high	16,976	16.5
4–5	High	21,225	20.7
5–7	Good	29,400	28.7
7–9	Average	13,422	13.1
9–10	Fair	3564	3.5
10–15	Poor	6482	6.3
15 and over	Bad	1656	1.6
Not reported	?	3086	3.0

CCDC,

The CSD contains results derived from the X-ray diffraction analysis of organic and organometallic compounds. In 1962 when I first visited Japan I reported, with some pride, that there were approximately 700 known crystal structures and almost 100 new structures published each year.[2] In 1992 we processed the 100,000th entry for the database by a Japanese author. Currently the database increases by over 10,000 new structures annually.

Table 1 gives a statistical overview of the October 1992 release of the database. It contains 102,589 entries, retrieved from 634 different literature sources and is fully retrospective. There are 90,315 entries with three-dimensional co-ordinates. For 11,520 entries, errors were corrected by the staff of the Data Centre. The chemical and crystallographic connectivities have been fully matched for 76,537 entries, as explained later. 8793 entries have been partially matched. The distribution of entries within the various chemical classes highlights the large number of heterocyclic compounds and natural products. There are 2811 molecular complexes and clathrates (Classes 60 and 61) of special interest here. Main group compounds and metal complexes comprise 54.4% of the database. Some 72% of the entries are of good precision or better with reliability factors (R) of less than 7%. A further 17%, with R factors of between 7 and 10% would still be judged adequate for most molecular modelling experiments.

Information categories

We can think of the information contained in the CSD in three categories: 'chemical information' which is encoded in the conventional chemical structural diagram. One such diagram, for morphine, is shown in Figure 1. The chemical information displayed in such a diagram is transformed internally, as part of our software system, into a conventional chemical connection table which defines for each atom in the diagram, the element type, the charge, the number of hydrogen atoms, and the connections via specified bond types to neighbouring atoms. The bond type designations we use are very detailed with seven different bond types, more extensive than those used by most chemical information management tools. To speed up the eventual searches based on chemical information an extensive set of screens is generated from the connectivity table. The screens range from element types to complex structural fragments. The two-dimensional diagram co-ordinates are retained so that the publication quality diagrams can be retrieved at any time from the database.

The 'crystallographic information', which comprises the bulk of the database contents, gives for each

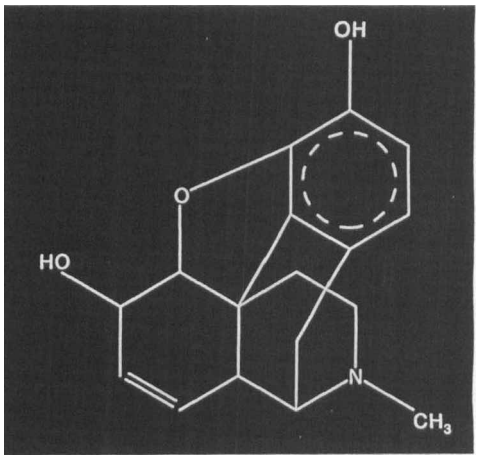


Figure 1 Chemical structural diagram of morphine. Information relating to this diagram is stored in the database as a chemical connectivity Table and the X, Y diagram co-ordinates.

entry the crystallographic constants such as unit cell parameters, symmetry, space group and most importantly, the three-dimensional co-ordinates derived from the crystal structure determination. There is information also about the crystallographic connectivity analogous to the chemical connectivity, except that it does not specify the bond type. The reason for this is that in many crystal structure determinations some or all hydrogen atoms are not located and thus the bond type is not directly determined. Figure 2(*a*) shows the three-dimensional structure of morphine and Figure 2(*b*) the extended crystal structure with all the symmetry-related molecules in a unit cell.

It should be noted that the co-ordinates recorded in the database are not necessarily those reported in the original paper. They may well have been transformed by the staff of the Cambridge Crystallographic Data Centre from the asymmetric unit of the crystallographer to the unique chemical entity, the so called crystal-chemical unit, which can equally be regarded as the repeat unit of the crystal. From the earliest days we have made a deliberate effort to make crystallographic results understandable to chemists with no previous knowledge of symmetry, space group theory and other matters. The co-ordinates in the database can therefore be used directly to derive information about a molecule or ion, without the need for symmetry transformations.

The atomic co-ordinates included in the database are checked in Cambridge for internal consistency against the bond lengths reported in original publications. In the course of this checking process a large number of errors are discovered and resolved either by our own computer programs or by correspondence with the authors (Table 1). There are of course

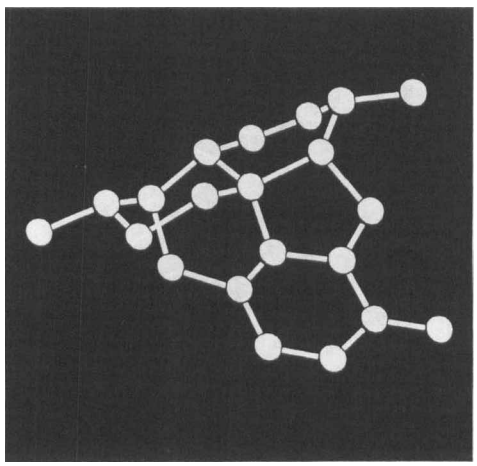


Figure 2(*a*) Three dimensional structure of morphine. The following information specifying atom type, number of attached atoms, charge, bond type and chemical connectivity is relevant to the 3D molecular structure of morphine: element type, three-dimensional atomic co-ordinates (x, y, z), connectivity of atoms within covalent distances which forms the crystal chemical unit.

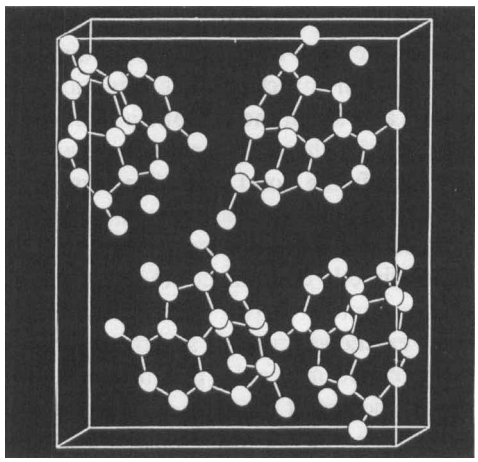


Figure 2(*b*) Extended crystal structure of morphine. The following information relevant to the extended crystal structure of morphine is stored in the database: unit cell dimensions, number of molecules in the repeat unit of the crystal, symmetry relations between molecules (space group).

numerous other checks before an entry is archived in the database.

The third information type is 'bibliographic'. This comprises the unique six letter reference code, which identifies the entry in the database, the name of the compound, a qualifier (e.g. antibiotic), the chemical formula in terms of residues which correlate with the crystal chemical unit, the bibliographic information, a data summary, and remarks. Table 2 shows the bibliographic citation for morphine and the more extensive information, illustrating concepts such as the residue designation for ions and solvents, for naloxone hydrochloride dihydrate.

Matching numbers

The lack of bond type information in the crystallographic connectivity record has for many years been a major obstacle to the ready and user friendly exploitation of the data stored in the database. The problem is illustrated in Figure 3 which shows diagramatically the crystallographic and chemical connectivities of 4-ammoniocyclohex-2-ene-1-carboxylic acid. Although the double bond in the cyclohexene ring can be identified from the bond length as the bond between atoms 5 and 6 of the crystal structure, the chemical concept (double bond) has not been explicitly identified in the X-ray experiment. Similar considerations apply to the C=O double bond. To correlate the two connectivities it is necessary therefore to examine the topology and geometry of the molecule determined crystallographically and use a variety of rules to map the two connectivities onto each other. Atoms 7 and 3 of the cyclohexene ring in the chemical diagram can then be associated with the three-dimensional co-ordinates of atoms 6 and 5 of the X-ray structure and the hydroxy O-5 atom with the co-ordinates of O-10. Once these assignments are made all other atoms in the two connectivities can also be matched. The extended connectivity record is depicted in Figure 4.

With the help of some specially developed computer programs and by reference to the original papers we were able to fully or partially match the connectivities for the majority of entries in the database. Matching numbers are derived for all new entries before they are registered in the database.

Table 2 Bibliographic information for two entries in the Cambridge Structural Database

MORPHM
(−)-Morphine monohydrate
C17 H19 N1 O3, H2 O1
E.Bye
Acta Chem. Scand. Ser. B, (1976) **30**, 549
COOR = 43// *SPAC = P212121// *RFAC = .0450//

NALOXC02
Naloxone hydrochloride dihydrate
opiate antagonistic activity, at 90 deg. K,
full data refinement
C19 H22 N1 O4 1+, Cl1 1−, 2(H2 O1)
C.L. Klein, R.J. Majeste, E.D. Stevens
J. Am. Chem. Soc., (1987) **109**, 6675
*REMA = High-order refinement Rfact = 0.033,
multipole refinement Rfact = 0.027,
monopole refinement Rfact = 0.037
*COOR = 53// *SPAC = P212121// *RFAC = .0330//
*CASN = 51481608

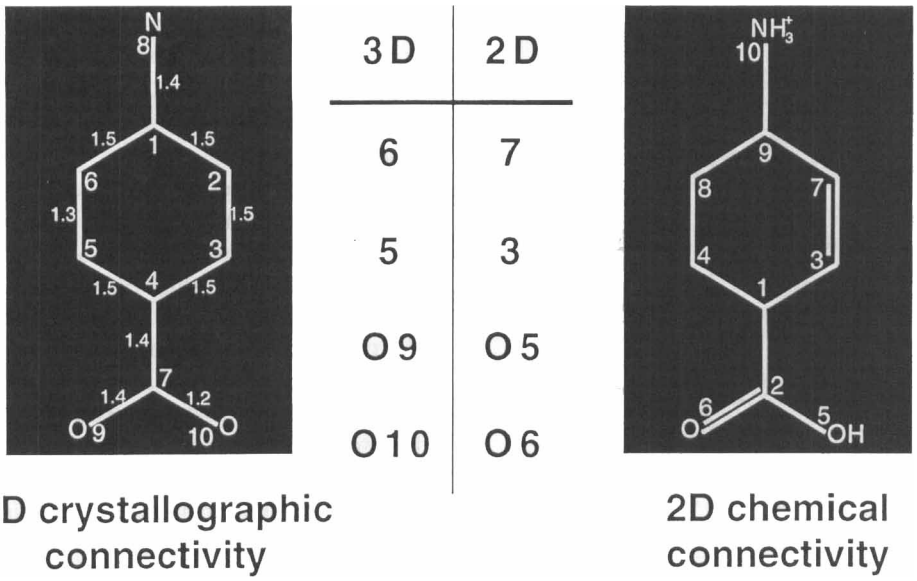


Figure 3 *Link between 2D and 3D structures; the 'matching number'. Two representations of the connectivity of 4-ammoniocyclohex-2-ene-1-carboxylic acid. The 3D crystallographic connectivity depicted on the left lacks information about bond type, but the distances between the atoms can be derived from the atomic co-ordinates determined in the X-ray analysis. The matching numbers are numbers such as those in the centre of the diagram which define the matches between the atoms in the two connectivity representations and identify, for example, the hydroxy O atom of the chemical diagram with the O-10 atom of the crystal structure.*

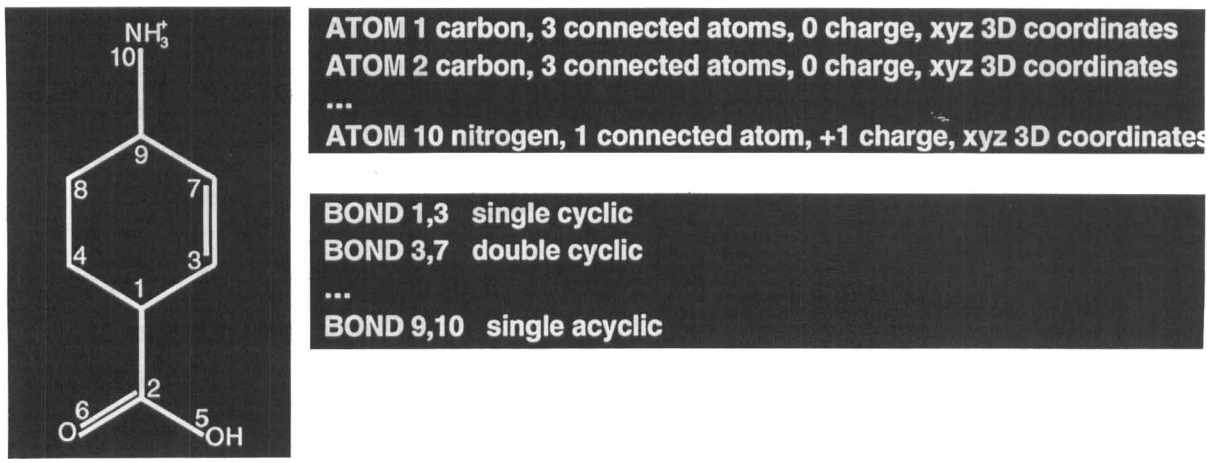


Figure 4 Extended connectivity record. Each atom of the chemical connectivity record is associated with the x, y, z co-ordinates of the matched atom as determined by the X-ray structure analysis.

Software system

The introduction of the matching numbers has made it possible to search the entire database by an interactive graphics program 'Quest 3D'. Other software components of the Version 5 system are a statistical package for the analysis of numeric data and the display program 'Pluto'.

Quest 3D allows the user to specify a variety of search constraints or 'tests'. Results from individual tests can be logically combined to form a complete query. Search constraints can be 'one-dimensional', testing individual numbers or character strings to select, for example, papers from a given period or by a specified author. Two-dimensional constraints are applied in chemical connectivity searches to locate specific molecules or, more importantly, specific substructural fragments. The specification of search constraints may be extended to three dimensions by requiring that geometrical parameters for a molecule or fragment be in pre-defined ranges. Thus, we may retrieve pharmacophoric patterns through use of distance constraints, or specific conformers of a chemical fragment by use of torsion angle constraints.

Figure 5 illustrates how a distance constraint is

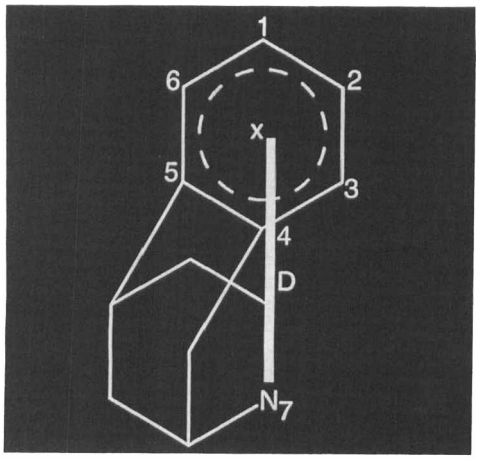


Figure 5 Specification of a distance involving the centroid of a ring and an atom in Version 5. Two commands are used. The SETUP X command followed by the specification of the ring atoms initiates the calculation of the co-ordinates of the centroid. The DEFINE command specifies the distance D as the distance between the centroid X and atom 7. D can then be used as a distance constraint as shown in Figure 6.

introduced. In this case it is desired to search for all compounds in the database that have a morphine-like fragment where the distance D between the centroid of the six-membered ring and the N-7 atom is within a certain range of values. In order to define this distance it is necessary to specify the centroid of the ring using a so-called 'Setup' command. When such a command is followed by the atom numbers either as shown in the diagram or selected directly on the graphics screen then X is automatically specified as the centroid. The 'Define' comand specifies D as the distance between X and atom 7. The D-range can then be used as a distance constraint as shown in Figure 5. Other Setup commands specify vectors, planes, normals, ring pucker and the orientation of vectors to lone pair planes. The Define command can define distances, angles or torsional angles. In Version 5 all these commands can be implemented graphically using chemist-friendly menus.

Figure 6 illustrates the various outputs obtained in a search for the fragment shown in Figure 5 with D in the range of 4.3 to 4.65 Å. The search finds all entries

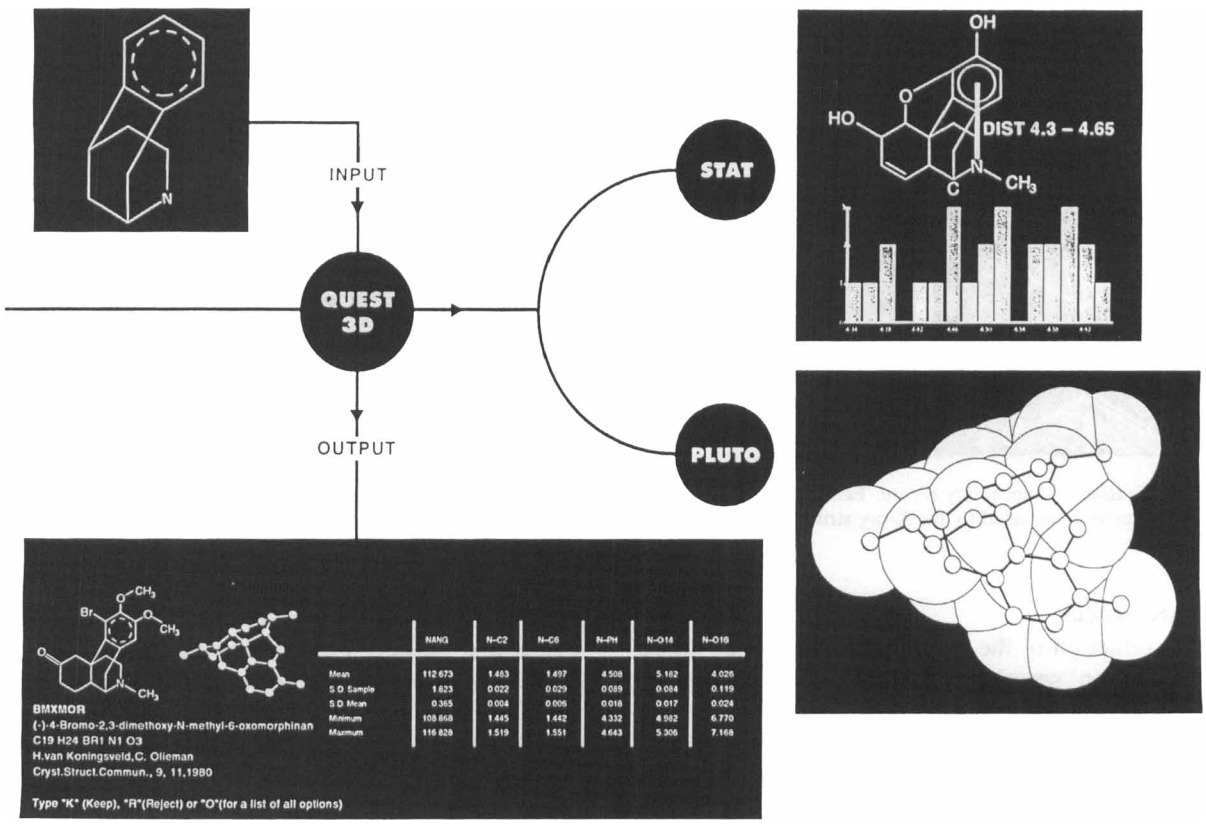


Figure 6 Flowchart for Version 5 of the CSD system showing a search with distance constraint as specified in Figure 5. The output at the bottom of the diagram is the default option for each 'hit' containing the specified morphine fragment. It gives the chemical diagram of the molecule containing the fragment, a three-dimensional diagram using the co-ordinates determined in the X-ray analysis and the bibliographic information. The table on the right gives statistical data for specified angles and distances for all hits located by the search. The histogram on the upper right of the Figure depicts the distribution of the distance D for all fragments located in the search. The diagrams on the left show two different styles from the plot program PLUTO.

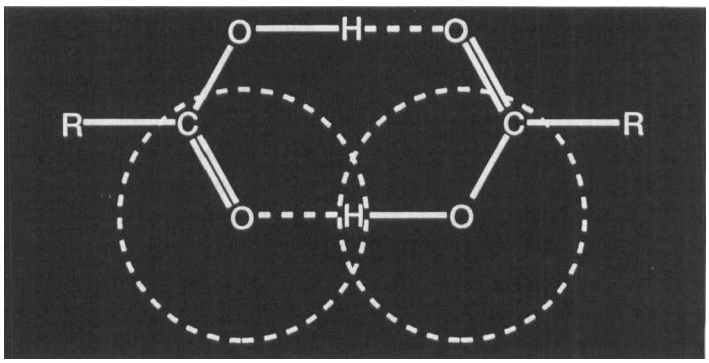
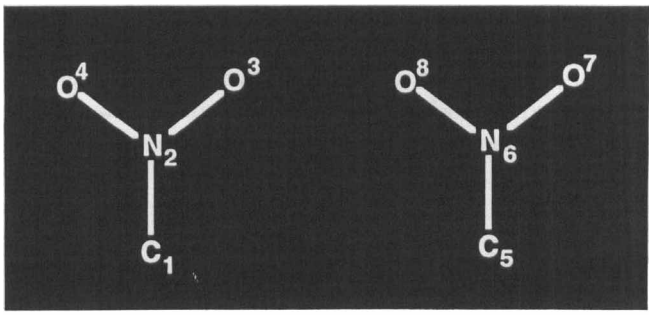


Figure 7 Intermolecular crystallographic connectivity. The use of van der Waals radii to define 'connections' between non-bonded atoms.




```
CONTACT INTER 2 8 2.0 3.1
```

Figure 8 Search for contact atoms between nitro–nitro groups. Command specifies that the search is to be confined to distances between N-2 and O-8 of between 2.0 and 3.1 Å.

which satisfy the chemical substructure and distance criteria. For each 'hit' the output shows the chemical and crystallographic structure and a summary table of all the defined distances, angles and torsion angles requested. The histogram on the top right of Figure 6 is the statistical analysis of the distribution of D for the different 'hits', and the diagram on the bottom right is a superposition of two different Pluto drawings.

Searches for intermolecular interactions

One of the most exciting new developments is that in Version 5 intermolecular and intramolecular constraints can be applied with equal ease. Instead of the intramolecular crystallographic connectivity discussed so far one can define an intermolecular crystallographic connectivity as shown in Figure 7. In this case the van der Waals radii are used to define 'connections' between non-bonded atoms. Once such a motif is defined and located all geometric calculations can be applied using the same Setup and Define commands.

Just one illustration of such an intermolecular search is given; the search for contact distances between atoms of interacting nitro groups. Figure 8 is a direct reproduction from the graphics screen where command 'Contact inter' automatically sets up the connected unit of Figure 7 followed by the specification of a range of distances between two atoms N-2 and O-8 which is to be the constraint of the search. An early example of the knowledge which can be derived from such a search is described in the next section.

When developing the interactive graphics system of Version 5 we tried to use modes of thinking familiar to chemists, particularly chemical diagrams, and to introduce three-dimensional concepts such as the definition of centroids (Fig 5), planes etc., in a way which can be understood almost intuitively.

We hope that the provision of these readily usable tools will lead to new discoveries both about the fundamental rules that govern intermolecular interactions, and also as pointers to new structures that might mimic natural products like vancomycin which exploit these principles for specific purposes involving molecular recognition. Version 5 will be released in October 1992 to academic institutions worldwide and to many companies who subscribe to the system.

EXAMPLES OF USE OF CSD TO STUDY INTERMOLECULAR INTERACTIONS

There are a growing number of studies of intermolecular interactions utilizing the database and the computational tools which were available with earlier versions of the CSD system. Some of these studies are referenced at the end of this paper. The list is kept current and is issued with each update of the database. In this section just three examples of such studies are described. Two involve the use of large data sets derived from the database and the third is a more in-depth analysis of a relatively small group of structures. The reader is advised to turn to the bibliography of selected papers in the Appendix for a more balanced and extensive overview of the types of investigations which have been published and which should prove an inspiration for further studies. The goldmine represented by the database has hardly been touched as yet.

Environment of functional groups

Robin Taylor and his collaborators[3] undertook an investigation to determine the preferred intermolecular environment of nitro and carbonyl groups. They hoped that such a study would help in the design of new pesticides in which certain functional groups are replaced to minimize unwanted features, while the ability of the compound to bind to the receptor site involved in anti-pesticide activity is retained. This pioneering study is a good example of the types of studies that can be carried out, especially since the knowledge derived could well have application to the group-by-group synthesis approach discussed at the beginning of this paper.

Taylor and his colleagues first searched the CSD for all structures containing the nitro group. They created a mini database of all entries for which hits were obtained. They then examined each of these crystal structures in turn to locate atoms which were involved in short intermolecular contacts with atoms of other functional groups. Short contacts were defined as distances of less than the sum of the van der Waals radii of the atoms involved. This was a two-step procedure, but with the release of Version 5 it will be possible to locate such contacts directly with Quest 3D as indicated in Figure 8.

Table 3 shows the distribution of these contact atoms in the 1765 crystallographically independent nitro groups located in the 1988 version of CSD. As indicated in the Table, C, H, N and O contacts are very frequent.

As the third step in this investigation, all the functional groups located were brought into a standard orientation and displayed simultaneously on a 3D graphics colour terminal using a program developed by Taylor and co-workers. Since the functional groups were all in the same orientation it was possible to superimpose them exactly and thus the contact atom displayed represented the crystal field environment of the functional group in the various crystal structures identified by the original search. An analysis of the composite picture, by interactive graphics, gave an indication of the preferred contact arrangements for the particular functional group.

Figure 9 shows the nitro–nitro contacts; Figure 9(*a*) the distribution of oxygen atoms around the nitrogen of the nitro group, and Figure 9(*b*) that of the nitrogen atoms around the oxygen of the nitro group. In both cases two views are presented: one orthogonal, looking down on the plane of the nitro group and one edge on. Note that since the NO_2 group is symmetrical all contact atoms have been reflected into one quadrant.

Although these contact positions have been selected from a large number of structures and were influenced by the crystal packing forces, it is clear that there is a pronounced tendency for the oxygen atoms to cluster above the nitrogen atoms of the reference nitro group, and a smaller cluster of nitrogen atoms above the reference oxygen atoms. The distribution suggests that the preferred packing arrangements for nitro–nitro groups are as shown in Figure 10. It is likely that such arrangements are stabilized by electrostatic attractions between the nitrogen and oxygen atoms.

The same technique can also be used for investigating hydrogen bonding. In this case the distribution of hydrogen bonding H atoms is plotted, as shown in Figure 11, for O—H and N—H hydrogen atoms. This Figure also shows two orthogonal views of the distribution, one looking down on the NO_2 plane (Fig 11*a*) and one edge on (Fig 11*b*). There appears to be a marked preference for the hydrogen bond to occur close to or actually in the plane of the nitro group and largely in the direction of the idealized oxygen sp^2 lone pairs. The most likely hydrogen bonding pattern is depicted in Figure 11(*c*), although another possible arrangement is also discussed in the paper.

The authors have also examined contacts to hydrogen atoms bonded to carbon and found that again there were relatively few contacts above the plane of the nitro group. (Figs 12*a* and 12*b* show a

Table 3 Crystal field environment of the nitro group. Distribution of the different types of atoms at contact distances of less than the sum of the van der Waals radii from atoms of the nitro group

C	*H*	*Br*	*Cl*	*F*	*I*	*K*	*N*	*Na*	*O*	*S*	*Se*	*Te*
907	3153	23	33	16	4	139	341	1	909	23	3	6

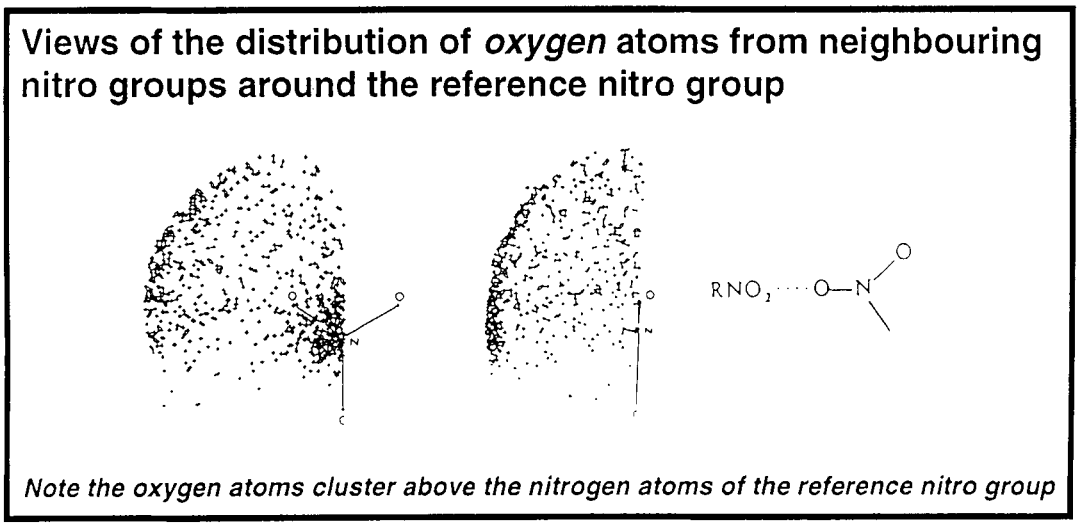


Figure 9(*a*) Nitro–nitro contacts. Views of the distribution of the oxygen atoms from neighbouring nitro groups around the reference nitro group. Orthogonal views perpendicular and edge on to the plane of the nitro group.

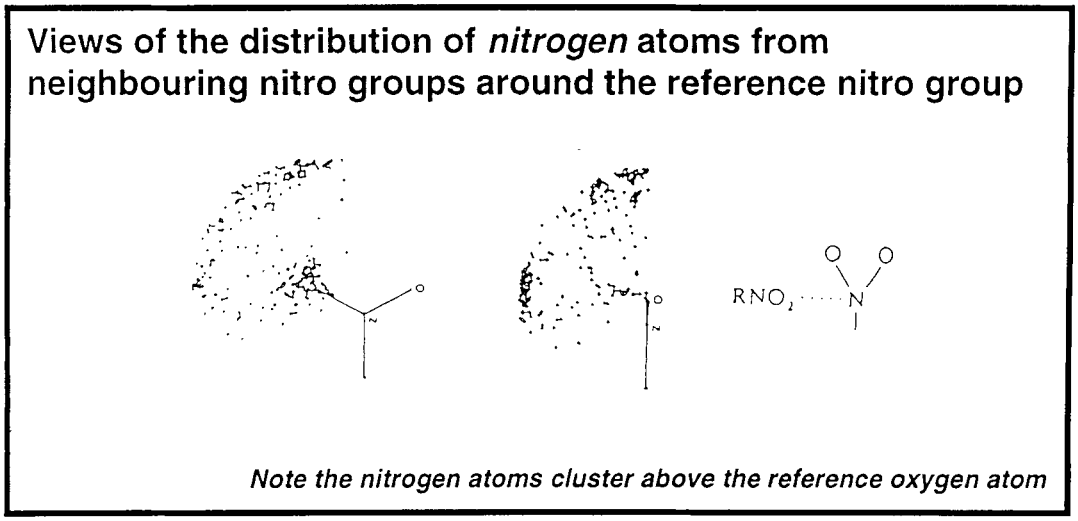


Figure 9(*b*) Nitro–nitro contacts. The distribution of the nitrogen atoms of the nitro group around the reference nitro group. Orthogonal views perpendicular and edge on to the plane of the nitro group.

selection of the various NO_2...H—C contacts). It would appear from this plot that there may be an energetic advantage for the hydrogen atoms to lie close to the nitro group plane and it may possibly be involved in C—H...O hydrogen bonding.

It is necessary to refer to the full paper for detailed discussion of the results but even this brief summary should indicate the relative ease with which it is possible to display the crystal-field environment of functional groups by use of the structural information accumulated in the database. Such investigations will in future be greatly facilitated by the use of Quest 3D described in the first part of this paper.

The knowledge derived from the study was that there are strong systematic features in the composite environments of nitro groups, with a tendency to hydrogen bond in approximately the direction of the oxygen bond pairs, and strong electrostatic interactions in the direction normal to the plane of the group. Evidently, this knowledge is of value not only for the specific purpose of pesticide design but also is of intrinsic chemical interest as well.

The C—H...O bond

The second example I would like to discuss is an earlier study by Taylor and myself[4] which also involved searches for contact distances. We used, and indeed developed this technique to examine the reality or otherwise of the so-called C—H...O bond.

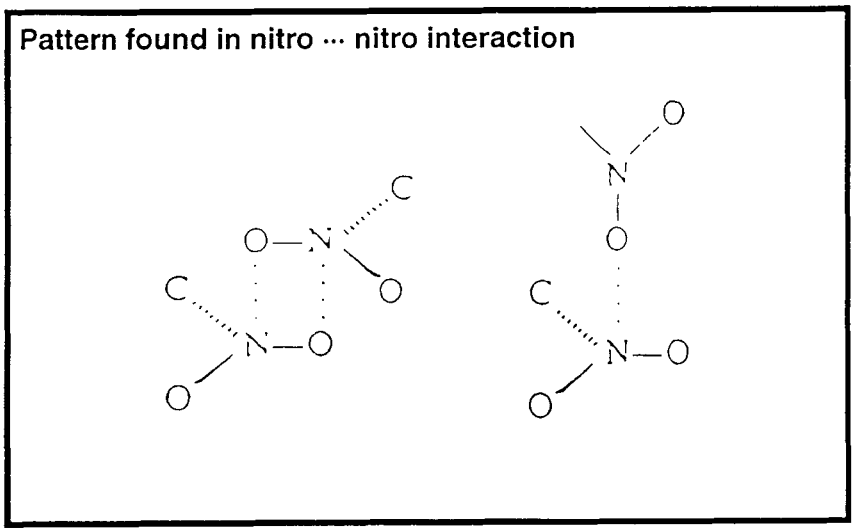


Figure 10 The most favourable packing arrangements of nitro– nitro groups.

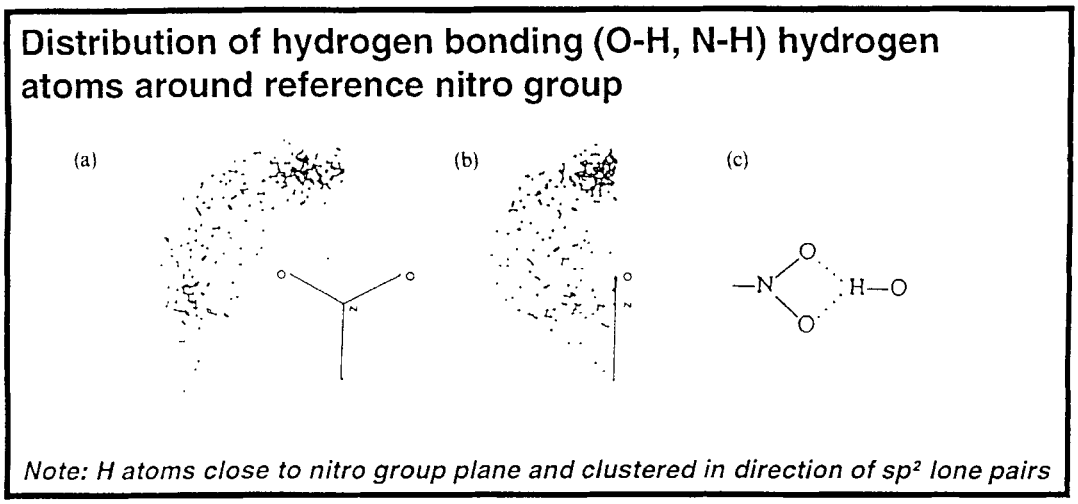


Figure 11 Hydrogen bonding to the nitro group. Distribution of the hydrogen bonds H—O and H—N around the reference nitro group. Orthogonal views normal (*a*), and edge on (*b*), to the plane of the nitro group. (*c*) The most likely hydrogen bonding pattern suggested by the preponderance of hydrogen atoms in the plane of the nitro group and clustering in the direction of the sp^2 lone pair.

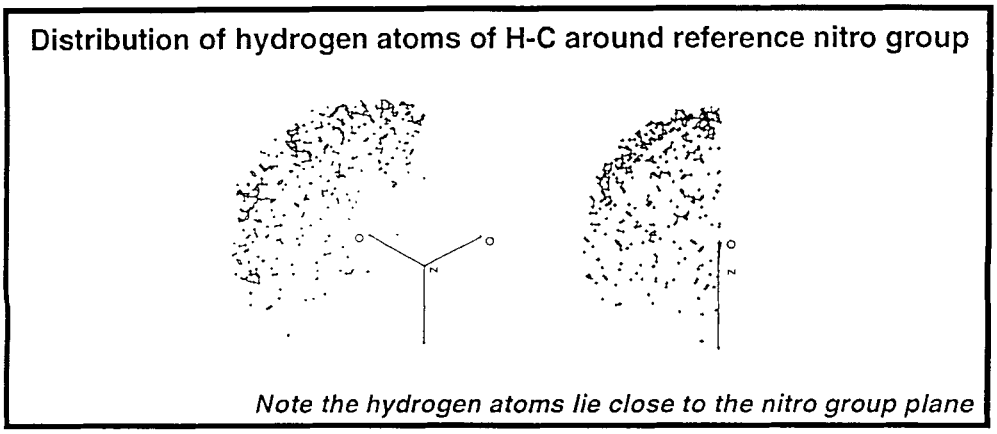


Figure 12 Distribution of the H atoms of H—C around the nitro group.

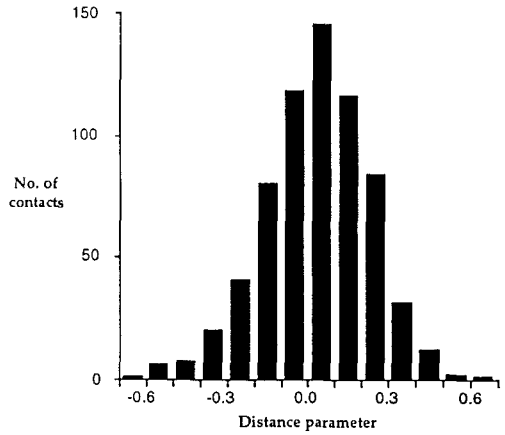


Figure 13 Histogram of the distribution of contact distances for the C—H bond retrieved from 113 neutron diffraction studies.

The existence of the C—H donor in H bonds had been the subject of debate for many years. Strong evidence in favour of the C—H...X bonds comes from spectroscopy and from quantum mechanical calculations. In contrast, the crystallographic evidence has been contentious. The influence of Donohue[5], who dismissed the question 'the C—H...O bond, what is it?' by stating that 'it is not', acted as a strong deterrent. Crystallographers were inhibited from using individual structure determinations, where close C—H...X contacts were found, as evidence of some type of H bonded interactions. Even as late as 1990 Julius Rebek[6] quotes only one structure determination, that of an 'unconventional' hydrogen bond in caffeine,[7] when discussing evidence for the C—H...O bond.

By the 1980s, with the availability of large numbers of precise crystal structure determinations, including neutron studies where actual hydrogen positions were located, it seemed possible to obtain a definitive answer to the question from a statistical analysis of the numerical data accumulated in the structural database. The study was based on 113 neutron diffraction studies which yielded 661 C—H...X contacts. The histogram in Figure 13 shows the distribution of d where $d = v(\mathrm{H}) + v(\mathrm{X}) - r(\mathrm{HX})$, where $v(\mathrm{H})$, $v(\mathrm{X})$ are van der Waals radii and $r(\mathrm{HX})$ is the H...X distance. Positive values of d indicate contacts shorter than the relevant sum of H + X van der Waals radii. They are thus the contact atoms as in the previous study.

The observed distribution is compared in Table 4 with a statistically derived expected distribution based on the stoichiometries of the sample compounds. Observed C—H...O contacts are twice as common as might be expected; when only the shorter contacts are taken into account the proportion of C—H...O rises to more than five times the expected number. These factors, together with the overall similarity of the C—H...O angular distribution with those of O—H...O and N—H...O, implied that the majority of short C—H...O contacts may reasonably be regarded as H bonds. It is notable that nine of the ten shortest C—H...O contacts found in this survey involved donor groups of the type N^+—C—H.

We have not as yet repeated the analysis with the more extensive data now in the database but there seems little need since, at least in crystallographic circles, the C—H...O and the C—H...O bonds are now well established. They have been noted particularly in a number of biological systems where they appear to have a strong stabilizing role intermediate between short hydrogen bonds and van der Waals forces.

DNA drug interactions

The C—H...O contacts have certainly been observed in crystal structures of medium-sized molecules of biological importance, most interestingly in oligonucleotides complexed with drugs. Here some of these studies are discussed very briefly to illustrate the knowledge which can be derived from a comparatively small number of crystal structures of a specific group of compounds, rather than the statistical analyses of specific data items selected from the 100,000 entries in the database.

The deoxyoligonucleotides, small fragments of DNA, are such a group. There are some 130 published structures in the database and co-ordinates are available for approximately 70 of them. The structures determined to date fall into less than 10 different structure types principally because of the difficulty of crystallizing these types of compounds, even with the

Table 4 Statistical analysis of the observed distribution of C—H...O contacts

		Observed distribution		
Atom type	*Predicted distribution*[a]	*All contacts*	*Contacts with $d > 0.0$ Å*	*Contacts with $d > 0.3$ Å*
C	0.30	0.18(122)	0.16(63)	0.04(2)
H	0.47	0.30(197)	0.23(90)	0.02(1)
Br	0.00	0.00(0)	0.00(0)	0.00(0)
Cl	0.01	0.05(31)	0.03(13)	0.02(1)
N	0.04	0.03(11)	0.03(11)	0.00(0)
O	0.18	0.44(289)	0.54(211)	0.91(42)
P	0.00	0.00(0)	0.00(0)	0.00(0)
S	0.00	0.00(4)	0.00(1)	0.00(0)
Total	1.00	1.00(661)	1.00(389)	1.00(46)

[a] Results obtained with an oxygen van der Waals radius of 1.50.

Table 5 DNA complexed with drugs

Complex	*Reference*
d(CGCGAATTCGCG)/cisplatin	9
d(CGCGAATTBrCGCG)/netropsin	10
d(CGCGAATTCGCG)/netropsin	11
d(CGCGAATTCGCG)/distamycin	12
d(CGCGAATTCGCG)/berenil	13
d(CGCGAATTCGCG)/DAPI	14
d(CGCGAATTCGCG)/Hoechst 33258	15–17
d(CGCGAATTCGCG)/Hoechst 33258	18
d(CGTACG)/daunomycin	19
d(CGATCG)/adriamycin	20
d(CGATCG)/daunomycin	21, 20
d(CGTpsACG)11-deoxydaunomycin	22
d(CGATCG)/4′-epiadriamycin	23
d(CGTACG)/triostin A	24
d(CGTACG)/echinomycin	25
d(GCGTACGC)/triostin A	26
d(BrCGBrCG)/proflavin	27
d(CGTpsACG)/nogalamycin	28, 29
d(mCGTpsAmCG)nogalamycin	30
d(TBrUGGCCAA)/chromomycin	31, 32
d(ATGCATATGCAT)/actinomycin	33
d(CGCAAATTTTGCG)/netropsin	34
d(TGATCA)/daunomycin	35
d(TGTACA)/daunomycin	35
d(mCGTpsAmCG)/U 58872	30

ps = Rp-phosphorothioate, mC = 5-methylcytosine, BrC = 5-bromocytosine, BrU = 5-bromouracil.

sophisticated purification techniques now available. (For a recent review see reference 8.)

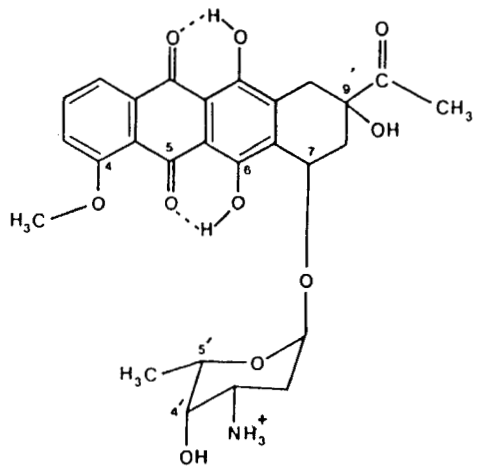
Some 30 structures in this collection relate to DNA fragments complexed with drugs. The specific aim of the studies was to determine the interactions between the drug and the DNA molecules. Certain natural products, particularly antibiotics, appear to bind selectively to specific base sequences and it is hoped that by studying the binding forces it will be possible to design antibiotics and anti-tumour agents with more selective targeting abilities.

The types of drugs which interact with DNA and which have been studied crystallographically fall into three categories: the groove binders which bind to the minor groove of DNA, the monofunctional intercalators which both bind to the minor groove and intercalate between the base pairs, and the bifunctional intercalators where two chromophores intercalate at two different positions into the double helix. The various DNA–drug complexes that have been analysed are listed in Table 5.

Monofunctional and bifunctional intercalators

Monofunctional intercalators, typified by daunomycin (Fig 14), have been extensively studied. The different crystal structures which have been published are listed in Table 5. As can be seen from Table 5 daunomycin was complexed with four different self-complementary DNA hexamers: d(CGATCG), d(CGTACG), d(TGATCA) and d(TGTACA) (for references see Table 5). The same hexamers were also crystallized with molecules closely related to daunomycin such as 14-hydroxydaunomycin (adriamycin), 4′-epiadriamycin and 11-deoxyadriamycin. It was thus possible to examine the same molecule in different crystal environments and different molecules in the same crystal environment. Figure 15 is a stylized diagram of the DNA double helix with two molecules of daunomycin intercalated at the d(TpG) step in the hexamer d(TGATCA).

In Figure 16 the intermolecular contacts between the drug and the DNA double helix are compared in a pair of studies where daunomycin was complexed to two sequences, d(TGATCA) and d(TGTACA). In both structures the principal interactions are via hydrogen bonds, some mediated by water molecules or sodium ions. Close van der Waals contacts also have an important stabilizing role. The result which emerges from these studies is that there is marked mutual adaptation of both the DNA and the drug molecules in such a way as to maximize hydrogen bonding and van der Waals interactions to suit the particular functional groups available for binding. The role of the water molecules is particularly important in mediating these interactions. As shown in Figure 17 the drug molecule adapts to the different DNA sequence environments particularly through changes in the orientation of the sugar moiety in relation to the intercalating planar chromophore. It seems that the best hope of improving base specificity is to reduce the flexibility of the drug molecule, possibly by making



Daunomycin

Figure 14 Molecular formula of daunomycin.

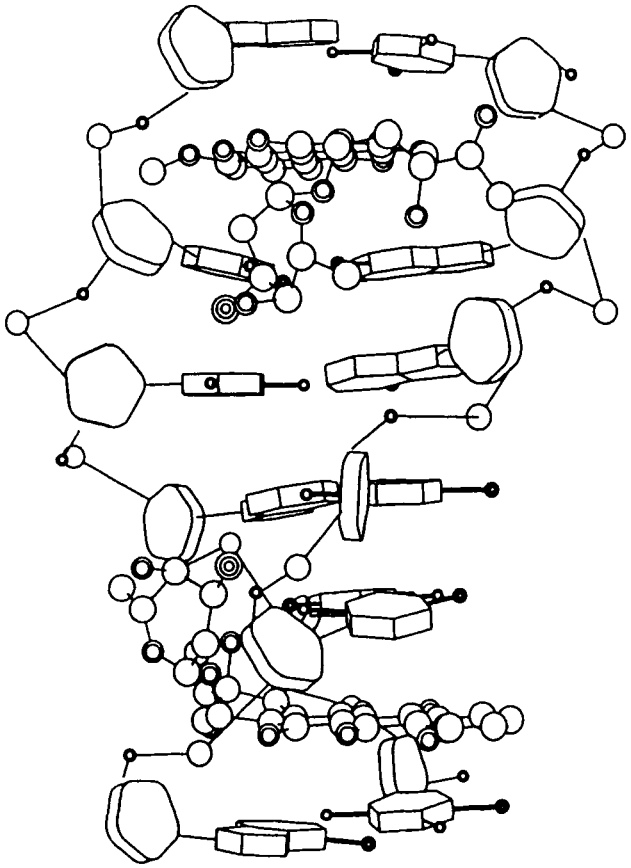


Figure 15 A view from the minor groove side of the complex between daunomycin and a d(TGATCA).

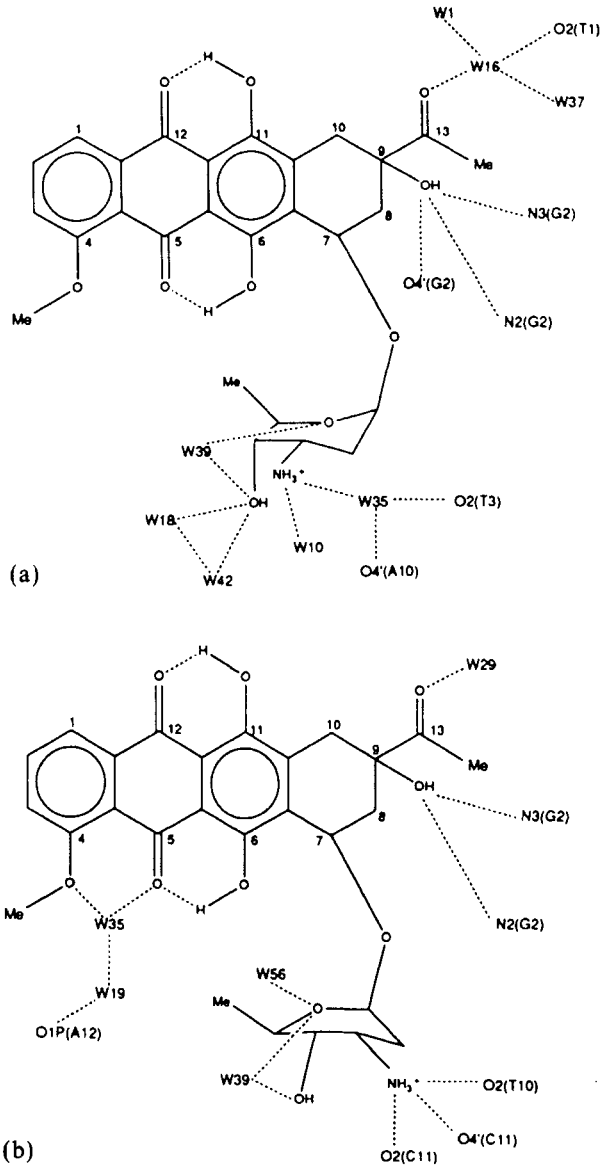


Figure 16 Two-dimensional diagram to show the hydrogen bonding and close contact (<3.3 Å) interactions with complexes between daunomycin and a DNA hexamer. Interaction in the complex daunomycin d(TGTACA) (*a*), and d(TGATCA) (*b*).

the sugar moiety more rigid, a line of thinking which was not anticipated at the start of these X-ray studies.

The last example of the influence of a selected number of X-ray structure determinations on our view of molecular interactions concerns the bifunctional intercalators triostin A and echinomycin (Fig 18). These molecules have two planar quinoxaline groups separated by a fairly rigid cyclic depsipeptide system. The two chromophores are optimally aligned (with the aid of a D-serine) to intercalate and insert two adjacent base pairs from the minor groove side of the B-DNA helix. This group of compounds has been the subject of numerous theoretical studies in an attempt to understand the interaction between the drug and the DNA and some very elegant docking programs were developed as part of these studies. However, it was not until the experimental determination of some of these complexes by Rich and Wang and co-workers (for references see Table 5) that the true importance of the various types of molecular interactions was appreciated. The X-ray analyses uncovered a complex balance involving O—H...O,

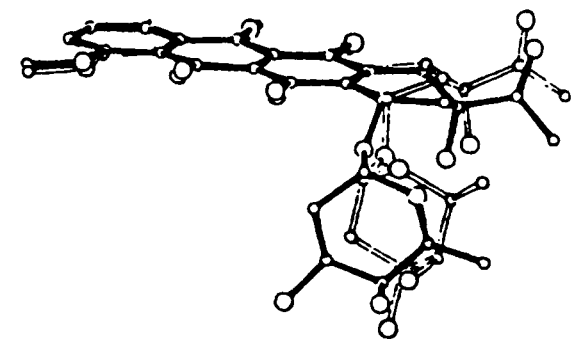


Figure 17 Superposition of the daunomycin molecule as determined in two different crystalline complexes with d(TGATCA) and d(TGTACA).

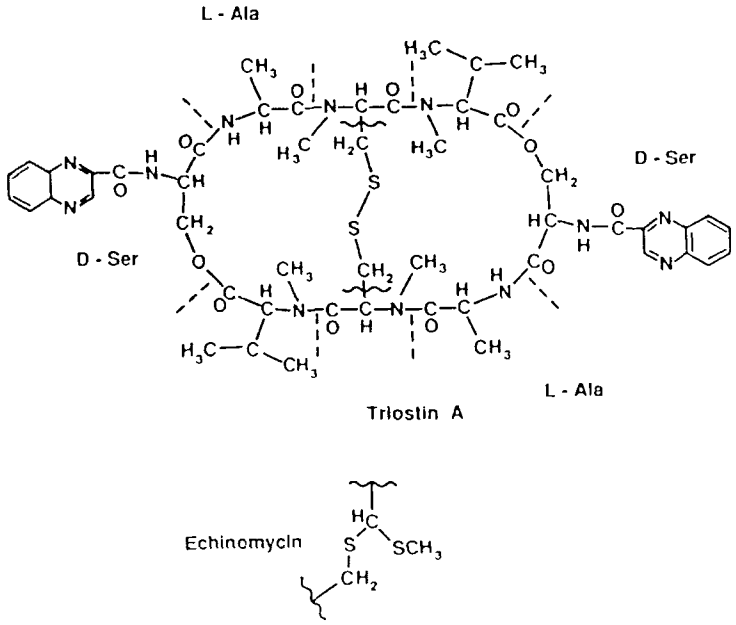


Figure 18 Formulae for the bifunctional intercalators triostin A and echinomycin.

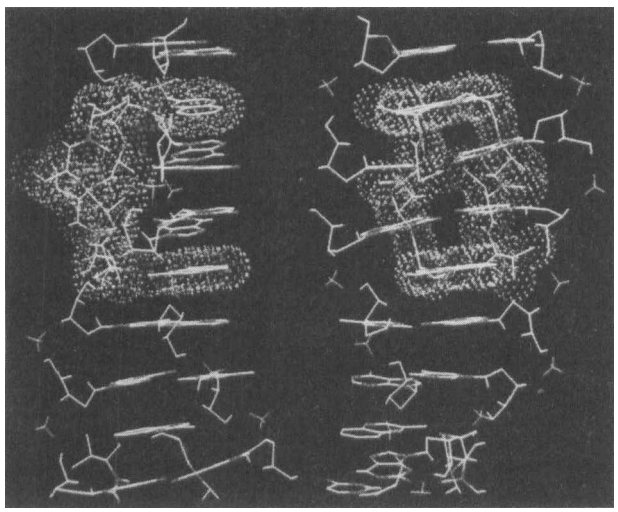


Figure 19 Views of the d(GCGTACGC)–triostin complex characterized by Quigley *et al.*[26] The high degree of distortion of the DNA produces a ladder-like duplex. (*a*) Side view of the drug highlighting the bis-intercalation. (*b*) View from the major groove side of the DNA. In both cases the drug is depicted with a dot surface representing a 50% van der Waals area. (See color plate I at the back of this issue.)

O—H...N and C—H...O hydrogen bonds, van der Waals contacts and stacking interactions. In the resultant stable DNA–drug complex the DNA double helix itself is modified, with some of the Watson–Crick base pairs replaced by quasi-Hoogstein pairs as shown in Figure 19. The disruption of the DNA double helix was certainly not predicted by any of the theoretical studies previously published. The 'take home' lesson seems to be that there is a need for continuous dialogue between experimentalists and molecular modellers with the structural database providing a common source of both theoretical and experimental information.

CONCLUSION

These examples from the DNA field are only an indication of the information that can be derived from limited sets of crystallographic data. For this information to be transformed into 'knowledge', in the sense of generally applicable rules, we need to analyse the very large data sets which are now available for many classes of compounds and held in the structural database. They represent the most extensive experimental data on intermolecular interactions to be found and I hope that by the use of the conceptual and computational tools, which will shortly become available in the next version of the CSD system, this data will make a real contribution to our understanding and application of molecular interactions to the 'new chemistry' which is the subject of this meeting.

ACKNOWLEDGEMENTS

I would like to thank my colleagues, present and past, at the Cambridge Crystallographic Data Centre who contributed to the development of the Cambridge Structural Database system since 1965 and to my

collaborators on the DNA studies, which is supported by the Medical Research Council. I thank Dr Robin Taylor for allowing me to quote extensively from his work and Dr Stephanie Garner with her help in the preparation of this paper. I would also like to acknowledge the contribution of the many authors who have used the database for the various studies listed in the bibliography and who, by their interest and enthusiasm, encouraged us in the continued improvement of the CSD system.

REFERENCES

1 (a) Allen, F.H.; Kennard, O.; Taylor, R.; *Acc. Chem. Res.* **1983**, *16*, 156. (b) 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.; *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 187.
2 Kennard, O.; *Scientific Information in the Fields of Crystallography and Solid State Physics*, Cryst. Soc. Japan, **1962**, p. 62.
3 Taylor, R.; Mullaley, A.; Mullier, G.W.; *Pestic. Sci.* **1990**, *29*, 197.
4 Taylor, R.; Kennard, O.; *J. Amer. Chem. Soc.* **1982**, *104*, 5063–5070.
5 Donohue, J.; *Selected Topics in Hydrogen Bonding* (Rich, A. and Davidson, N., eds.), Structural Chemistry and Molecular Biology, W.H. Freeman, San Francisco, **1968**, p. 443.
6 Rebek, J.N.; *Host-Guest Molecular Interactions from Chemistry to Biology* Ciba Foundation Symposium, John Wiley, Chichester, **1991**, p. 105.
7 Sutor, J.; *Acta Cryst.* **1958**, *11*, 453.
8 Kennard, O.; Hunter, W.N.; *Angew. Chem. Intl. Ed. Engl.* **1991**, *103*, 1254.
9 Wing, R.M.; Pjura, P.; Drew, H.R.; Dickerson, R.E.; *EMBO J.* **1984**, *3*, 1201.
10 Kopka, M.L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R.E.; *Proc. Natl. Acad. Sci. USA* **1985**, *82*, 1376.
11 Coll, M.; Aymami, J.; van der Marel, G.A.; van Boom, J.H.; Rich, A.; Wang, A.H.-J.; *Biochemistry* **1989**, *28*, 310.
12 Coll, M.; Frederick, C.A.; Wang, A.H.-J.; Rich, A.; *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 8385.
13 Brown, D.G.; Sanderson, M.R.; Skelly, J.V.; Jenkins, T.C.; Brown, T.; Garman, E.; Stuart, D.I.; Neidle, S.; *EMBO J.* **1990**, *9*, 1329.
14 Larsen, T.A.; Goodsell, D.S.; Cascio, D.; Grzeskowiak, K.; Dickerson, R.E.; *J. Biomol. Stereodyn.* **1989**, *7*, 477.
15 Teng, M.-K.; Usman, N.; Frederick, C.A.; Wang, A.H.-J.; *Nucleic Acids Res.* **1988**, *16*, 2671.
16 Pjura, P.E.; Grzeskowiak, K.; Dickerson, R.E.; *J. Mol. Biol.* **1987**, *197*, 257.
17 Quintana, J.R.; Lipanov, A.A.; Dickerson, R.E.; *Biochemistry* **1991**, *30*, 10294.
18 Carrondo, M.A.A.F.; Coll, M.; Aymami, J.; Wang, A.H.-J.; van der Marel, G.A.; van Boom, J.H.; Rich, A.; *Biochemistry* **1989**, *28*, 7849.
19 Wang, A.H.-J.; Ughetto, G.; Quigley, G.J.; Rich, A.; *Biochemistry* **1987**, *26*, 1152.
20 Frederick, C.A.; Williams, L.D.; Ughetto, G.; van der Marel, G.A.; van Boom, J.H.; Quigley, G.J.; Wang, A.H.-J.; **1990**, *29*, 2538.
21 Moore, M.H.; Hunter, W.N.; Langlois d'Estaintot, B.; Kennard, O.; *J. Mol. Biol.* **1989**, *206*, 693.
22 Williams, L.D.; Egli, M.; Ughetto, G.; van der Marel, G.A.; van Boom, J.H.; Quigley, G.J.; Wang, A.H.-J.; Rich, A.; Frederick, C.A.; **1990**, *215*, 313.
23 Williams, L.D.; Frederick, C.A.; Ughetto, G.; Rich, A.; *Nucleic Acids Res.* **1990**, *18*, 5533.
24 Wang, A.H.-J.; Ughetto, G.; Quigley, G.J.; Hakoshima, T.; van der Marel, G.A.; van Boom, J.H.; Rich, A.; *Science* **1984**, *225*, 1115.
25 Ughetto, G.; Wang, A.H.-J.; Quigley, G.J.; van der Marel, G.A.; van Boom, J.H.; Rich, A.; *Nucleic Acids Res.* **1985**, *13*, 2305.
26 Quigley, G.J.; Ughetto, G.; van der Marel, G.A.; van Boom, J.H.; Wang, A.H.-J.; Rich, A.; *Science*, **1986**, *232*, 1255.
27 Westhof, E.; Hosur, M.V.; Sundaralingam, M.; *Biochemistry* **1988**, *27*, 5742.
28 Liaw, Y.-C.; Gao, Y.-G.; Robinson, H.; van der Marel, G.A.; van Boom, J.H.; Wang, A.H.-J.; *Biochemistry* **1989**, *28*, 9913.
29 Williams, L.D.; Egli, M.; Gao, Q.; Bash, P.; van der Marel, G.A.; van Boom, J.H.; Rich, A.; Frederick, C.A.; *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 2225.
30 Gao, Y.-G.; Liaw, Y.-C.; Robinson, H.; Wang, A.H.-J.; *Biochemistry* **1990**, *29*, 10307.
31 Ogata, C.M.; Hendrickson, W.; Gao, X.; Patel, D.J.; *Abstr. Amer. Crystallogr. Assoc. Meet.* **1989**, *17*, 53.
32 Ogata, C.M.; Hendrickson, W.; *Abstr. XVth Intl. Congr. Intl. Union Crystallogr.* **1990**, *C-146*, 04.02.03.
33 Jain, S.C.; Sobell, H.M.; *Abstr. Amer. Crystallogr. Assoc. Meet.* **1990**, *18*, 86.
34 Tabernero, L.; Aymami, J.; Subirana, J.A.; *Abstr. XVth Intl. Congr. Intl. Union Crystallogr.* **1990**, *C-147*, 04.02.08.
35 Nunn, C.M.; van Meervelt, L.; Zhang, S.; Moore, M.H.; Kennard, O.; *J. Mol. Biol.* **1991**, *222*, 167.

APPENDIX

Bibliographic references to the scientific utilization of the Cambridge Structural Database

A. Hydrogen bonding

Relationship between symmetry in hydrogen-bonded benzoic acids and the formation of acentric crystal structures. Frankenback, G.M. and Etter, M.C. (1992). *Chem. Mater.* **4**, 272.

Hydrogen bonds to carboxylate groups. The question of three-centre interactions. Gorbitz, C.H. and Etter, M.C. (1991). *J. Chem. Soc., Perkin Trans.* **2**, 131.

Hydrogen bonds to carboxylate groups. Syn/anti distributions and steric effects. Gorbitz, C.H. and Etter, M.C. (1992). *J. Am. Chem. Soc.* **114**, 627.

An *ab initio* study of crystal field effects. Part 3: Solid- and gas-phase geometry of formamide, modelling the changes in a peptide group due to hydrogen bonds. Propelier, P., Lenstra, A.T.H., Van Alsenoy, C. and Geise, H.J. (1991). *Struct. Chem.* **2**, 3.

The C—H...O hydrogen bond in crystals: What is it? Desiraju, G.R. (1991). *Acc. Chem. Res.* **24**, 290.

Iminophosphorane-substituted proton sponges. Part 2. Preparation and crystal structure of four phosphoranylidene ammonionaphthalene derivatives. Llmas-Saiz, A.L., Foces-Foces, C., Elguero, J., Molina, P.,

Alajarin, M. and Vidal, A. (1991). *J. Chem. Soc., Perkin Trans.* **2,** 1667.

Hydration in organic crystals: prediction from molecular structure. Desiraju, G.R. (1991). *J. Chem. Soc., Chem. Commun.* 426.

Hydrogen bonding in the crystal structure of the tetrasaccharide stachyose hydrate: a 1:1 complex of two conformers. Jeffrey, G.A. and Huang, D-Bin (1991). *Carbohydr. Res.* **210,** 89.

Hydrogen bonds in concreto and in computro: The sequel. Stouten, P.F.W., Van Eijck, B.P. and Kroon, J. (1991). *J. Mol. Struct.* **243,** 61.

Hydrogen bonding in alkylammonium chlorometalates of the first transition metals. Emerson, K., Roman, P., Luque, A., Gutierrez-Zorilla, J.M. and Martinez-Ripoll, M. (1991). *Inorg. Chem.* **30,** 1878.

Hydrogen bond directed cocrystallization and molecular recognition properties of acyclic imides. Etter, M.C. and Reutzel, S.M. (1991). *J. Am. Chem. Soc.* **113,** 2586.

Stereochemistry of water molecules in hydrates of small biological molecules. Jeffrey, G.A. and Maluszynska, H. (1990). *Acta Cryst., B (Str. Sci)* **46,** 546.

Encoding and decoding hydrogen bond patterns of organic compounds. Etter, M.C. (1990). *Acc. Chem. Res.* **23,** 120.

Strength and linearity of C—H...O bonds in molecular crystals: a database study of some terminal alkynes. Desiraju, G.R. (1990). *J. Chem. Soc., Chem. Commun.* 454.

N—H...Nsp^2 hydrogen interactions in organic crystals. Llamas-Saiz, A.L. and Foces-Foces, C. (1990). *J. Mol. Struct.* **238,** 367

The role of intramolecular hydrogen bonding as a determinant of the conformational profiles of cGMP and cAMP. Topiol, S., Morgan Junior, T.K., Sabio, M. and Lumma Junior, W.C. (1990). *J. Am. Chem. Soc.* **112,** 1452.

Distance dependence of C—H...O interactions in some chloroalkyl compounds. Desiraju, G.R. (1989). *J. Chem. Soc., Chem. Commun.* 179.

Correlation between crystallographic and spectroscopic properties for C—H...O bonds in terminal acetylenes. Desiraju, G.R. and Murty, B.N. (1987). *Chem. Phys. Lett.* **139,** 360.

Hydrogen bond formation in nitroanilines: The first step in designing acentric materials. Panunto, T.W., Urbanczyk-Lipokowska, Z., Johnson, R. and Etter, M.C. (1987). *J. Am. Chem. Soc.* **109,** 7786.

A survey of the geometry of hydrogen bonds in the crystal structures of barbiturates, purines and pyrimidines. Jeffrey, G.A. and Maluszynska, H. (1986). *J. Mol. Struct.* **147,** 127.

Lone-pair directionality in hydrogen bond potential functions for molecular mechanics calculations: the inhibition of human carbonic anhydrase II by sulfonamides. Vedani, A., Dunitz, J. (1985). *J. Am. Chem. Soc.* **107,** 7653.

Geometry of the N—H...O=C hydrogen bond. 3. Hydrogen bond distances and angles. Taylor, R., Kennard, O. and Versichel, W. (1984). *Acta Cryst., B (Str. sci)* **40,** 280.

Hydrogen bond geometry in organic crystals. Taylor, R. and Kennard O. (1984). *Acc. Chem. Res.* **17,** 320.

Directional hydrogen bonding to sp^2- and sp^3-hybridized oxygen atoms and its relevance to ligand–macromolecule interactions. Murray-Rust, P. and Glusker, J.P. (1984). *J. Am. Chem. Soc.* **106,** 1018.

Geometry of the N—H...O=C hydrogen bond. 2. Three-center (bifurcated) and four-center (trifurcated) bonds. Taylor, R., Kennard, O. and Versichel, W. (1984). *J. Am. Chem. Soc.* **106,** 244.

Hydrogen-bonding patterns in pyranose and pyranoside crystal structures. Jeffrey, G.A. and Mitra, J. (1983). *Acta Cryst., B (Str. Sci)* **39,** 469.

Comparison of X-ray and neutron diffraction results for the N—H...O=C hydrogen bond. Taylor, R. and Kennard, O. (1983). *Acta Cryst., B (Str. Sci)* **39,** 133.

Geometry of the N—H...O=C hydrogen bond. 1. Lone-pair directionality. Taylor, R., Kennard, O. and Versichel, W. (1983). *J. Am. Chem. Soc.* **105,** 5761.

A survey of hydrogen bond geometries in the crystal structures of amino acids. Jeffrey, G.A. and Maluszynska, H. (1982). *Intl. J. Biol. Macromol.* **4,** 173.

Crystallographic evidence for the existence of C—H...O, C—H...N, and C—H...Cl hydrogen

bonds. Taylor, R. and Kennard, O. *J. Am. Chem. Soc.* **104,** 5063.

Bibliography of crystal structures of carbohydrates, nucleosides and nucleotides, 1977–1978. Jeffrey, G.A. and Sundaralingam, M. (1981). *Adv. Carbohydr. Chem. Biochem.* **38,** 417.

B. Intermolecular interactions, reaction pathways, molecular dynamics

Quantitative analysis of the relative importance of symmetry operators in organic molecular crystals. Filippini, G. and Gavezzotti, A. (1992). *Acta Cryst., B (Str. Sci)* **48,** 230.

Comparison of conformations of thiocrown ethers in solid and solution with information on conformational flipping from molecular dynamics calculations. Lockhart, J.C. and Tomkinson, N.P. (1992). *J. Chem. Soc., Perkins Trans.* **2,** 533.

Crystal engineering and solid state chemistry of some beta-nitrostyrenes. Pedireddi, V.R., Sarma, J.A.R.P. and Desiraju, G.R. (1992). *J. Chem. Soc., Perkins Trans.* **2,** 311.

Stereochemical aspects of hydration of carbohydrates in aqueous solutions. 2. Kinetic medium effects. Galema, S.A., Blandamer, M.J. and Engberts, J.B.F.N. (1992). *J. Org. Chem.* **57,** 1995.

Molecular self-assemblies: Monte Carlo predictions for the structure of the one-dimensional translation aggregate. Perlstein, J. (1992). *J. Am. Chem. Soc.* **114,** 1955.

Database analysis of crystal-structure-determining interactions involving sulphur: Implications for the design of organic metals. Desiraju, G.R. and Nalini, V. (1991). *J. Mater. Chem.* **1,** 201.

H...H van der Waals distance in co-operative O—H...O—H...O hydrogen bonds determined from neutron diffraction data. Steiner, T., Saenger, W. (1991). *Acta Cryst., B (Str. Sci)* **47,** 1022.

Linear Si—O—Si fragments: Structure of 1,1,3,3-tetramethyl-1,3-bis[(2-oxo-1-pyrrolidinyl)methy]-disiloxane dihydrochloride and analysis of displacement parameters. Schklover, V.E., Burgi, H.-B., Raselli, A., Armbruster, T. and Hummel, W. (1991). *Acta Cryst., B (Str. Sci)* **47,** 544.

Ab initio and crystal structure analysis of like-charged ion pairs. Gao, Jiali, Boudon, S. and Wipff, G. (1991). *J. Am. Chem. Soc.* **113,** 9610.

Generation of possible crystal structures from the molecular structure for low-polarity organic compounds. Gavezzotti, A. (1991). *J. Am. Chem. Soc.* **113,** 4622.

Comparative study of imidazole hydration: *Ab initio* and electrostatic calculations vs. Cambridge Structural Database analysis. Alagona, G., Ghio, C., Nagy, P., Simon, K. and Naray-Szabo, G. (1990). *J. Computational Chem.* **11,** 1038.

Gavezzotti, A. *Electron Crystallography of Organic Molecules* (1990). p. 77. in Prediction of the crystal packing of organic molecules: Geometry and intermolecular potentials. (eds. J.R. Fryer and D.L. Dorset) Kluwer Academic Publishers.

Use of crystallographic data in searching for isosteric replacements: composite crystal-field environments of nitro and carbonyl groups. Taylor, R., Mullaley, A. and Mullier, G.W. (1990). *Pestic. Sci.* **29,** 197.

Crystal packing of hydrocarbons. Effects of molecular size, shape and stoichiometry. Gavezzotti, A. *Acta Cryst., B (Str. Sci)* **46,** 275.

Packing analysis of organic crystals containing C=O or C≡N groups. Gavezzotti, A. (1990). *J. Phys. Chem.* **94,** 4319.

Tetraarylporphyrin sponges. Composition, structural systematics and applications of a large class of programmable lattice clathrates. Byrn, M.P., Curtis, C.J., Khan, S.I., Sawin, P.A., Tsurumi, R. and Strouse, C.E. (1990). *J. Am. Chem. Soc.* **112,** 1865.

Stereochemistry of phosphate–lewis acid interactions: Implications for nucleic acid structure and recognition. Alexander, R.S., Kanyo, Z.F., Chirlian, L.E. and Christianson, D.W. (1990). *J. Am. Chem. Soc.* **112,** 933.

Ring-opening reactions of mesoionic [1,2,4]triazolo-[1,5-c]quinazolines. Preparation of 1,2,4-triazolo[3,4-b][1,3,4]thiadiazoles by a translocative rearrangement. Molina, P., Arques, A., Alias, M.A., Saiz, A.L.L., de la Concepcion Foces-Foces, L. (1989). *Liebigs Ann. Chem.* 1055.

Face selectivity of the Diels-Alder reaction of thebaine-like morphinandienes, a computational approach, Chemistry of Opium Alkaloids, Part XXXI, Linders, J.T.M. and Maat, L. (Belg. 1989). *Bull. Soc. Chim.* **98,** 265.

The nature of halogen...halogen interactions: are short halogen contacts due to specific attractive forces or to close packing of nonspherical atoms. Desiraju, G.R. and Parthasarathy, R. (1989). *J. Am. Chem. Soc.* **111,** 8725.

Intermolecular triple proton and deuteron transfer in crystalline 3,5-dimethylpyrazole studied by NMR, NQR, and X-ray methods. Smith, J.A.S., Wehrle, B., Aguilar-Parrilla, F., Limbach, H.-H., de la Concepcion Foces-Foces, M., Cano, F.H., Elguero, J., Baldy, A., Pierrot, M., Khurshid, M.M.T. and Larcombe-McDouall, J.B. (1989). *J. Am. Chem. Soc.* **111,** 7304.

Cyano–halogen interactions and their role in the crystal structures of the 4-halobenzonitriles. Desiraju, G.R. and Harlow, R.L. (1989). *J. Am. Chem. Soc.* **111,** 6757.

Statistical analysis of some structural properties of solid hydrocarbons. Gavezzotti, A. (1989). *J. Am. Chem. Soc.* **111,** 1835.

Crystal packing modes of hydrocarbons. Statistics on all the available crystal structures in the Cambridge data files. Gavezzotti, A.Z. (1988). *Kristallogr.* **185,** 717.

A systematic analysis of packing energies and other packing parameters for fused-ring aromatic hydrocarbons. Gavezzotti, A. and Desiraju, G.R. (1988). *Acta Cryst., B (Str. Sci)* **44,** 427.

Empirical potential energy surfaces relating structure and activation energy. 2. Determination of transition-state structure for the spontaneous hydrolysis of axial tetrahydropyranyl acetals. Burgi, H.B. and Dubler-Steudle, K.C. (1988). *J. Am. Chem. Soc.* **110,** 7291.

Empirical potential energy surfaces relating structure and activation energy 1. Metallocyclopentene ring inversion (s-*cis*-eta4-butadiene)metallocene complexes and related compounds. Burgi, H.B. and Dubler-Steudle, K.C. (1988). *J. Am. Chem. Soc.* **110,** 4953.

Dynamic processes in crystals examined through difference displacement parameters cdeltaU: Pseudo-Jahn-Teller distortion in *cis*-Cu(II)N_4O_2 co-ordination octahedra. Stebler, M. and Burgi, H.B. (1987). *J. Am. Chem. Soc.* **109,** 1395.

A revision of van der Waals atomic radii for molecular crystals (II): Hydrogen bonded to carbon. Nyburg, S.C., Faerman, C.H. and Prasad, L. (1986). *Acta Cryst., B (Str. Sci)* **42,** 106.

Role of Cl...Cl and C—H...O interactions in the crystal engineering of 4-A short-axis structures. Sarma, J.A.R.P. and Desiraju, G.R. (1986). *Acc. Chem. Res.* **19,** 222.

Dipole-dipole interactions and the inversion motif in the crystal structures of planar chloro aromatics: The unusual packings of 1,2,3-trichlorobenzene and 1,2,3,7,8,9-hexachlorodibenzo-*p*-dioxin. Desiraju, G.R., Sarma, J.A.R.P. and Krishna, T.S.R. (1986). *Chem. Phys. Lett.* **131,** 124.

Angular preferences of intermolecular forces around halogen centers: preferred directions of approach of electrophiles and nucleophiles around the carbon–halogen bond. Ramasubbu, N., Parthasarathy, R. and Murray-Rust, P. (1986). *J. Am. Chem. Soc.* **108,** 4308.

Stereochemistry of the $R_1(X{=})C(sp^2)\text{-}N(sp^3)R_2R_3$ fragment. Mapping of the *cis–trans* isomerization path by rotation around the C—N bond from crystallographic structural data. Gilli, G., Bertolasi, V., Bellucci, F. and Ferretti, V. (1986). *J. Am. Chem. Soc.* **108,** 2420.

A revision of van der Waals atomic radii for molecular crystals: N, O, F, S, Cl, Se, Br and I bonded to carbon. Nyburg, S.C. and Faerman, C.H. (1985). *Acta Cryst., B (Str. Sci)* **41,** 274.

The chloro-substituent as a steering group: A comparative study of non-bonded interactions and hydrogen bonding in crystalline chloro-aromatics. Sarma, J.A.R.P. and Desiraju, G.R. (1985). *Chem. Phys. Lett.* **117,** 160.

Co-ordinative interactions in chelated complexes of silicon. XII. Mapping the expansion of co-ordination of silicon from 4 to 5. Crystal structure of 1,2,3,4-tetrahydro-1-trimethylsilyl-1,10-phenanthroline. Klebe, G. (1985). *J. Organomet. Chem.* **293,** 147.

Internal molecular motion of triphenylphosphine oxide: Analysis of atomic displacement parameters for orthorhombic and monoclinic crystal modifications at 100 and 150K. Brock, C.P., Schweizer, W.B. and Dunitz, J.D. (1985). *J. Am. Chem. Soc.* **107,** 6964.

Crystal environments and geometries of leucine, isoleucine, valine and phenylalanine provide estimates of minimum nonbonded contact and preferred van der Waals interaction distances. Gould, R.O., Gray, A.M., Taylor, P. and Walkinshaw, M.D. (1985). *J. Am. Chem. Soc.* **107,** 5921.

Mapping the atomic environment of functional groups: turning 3D scatter plots into pseudo-density contours. Rosenfield Junior, R.E., Swanson, S.M., Myer Junior, E.F., Carrell, H.L. and Murray-Rust, P. (1984). *J. Mol. Graphics* **2,** 43.

Reaction pathways from structural data: Dynamic stereochemistry of zinc(II) compounds. Auf der Heyde, T.P.E. and Nassimbeni, L.R. (1984). *Acta Cryst., B (Str. Sci)* **40,** 582.

Dynamic processes in crystals examined through difference vibrational parameters cdeltaU: The low-spin–high-spin transition in tris(dithiocarbamato)-iron(III) complexes. Chandrasekhar, K. and Burgi, H.B. (1984). *Acta Cryst., B (Str. Sci)* **40,** 387.

An investigation into the role of chloro-substituents in hydrogen-bonded crystals: The crystal structures of the dichlorophenols. Thomas, N.W. and Desiraju, G.R. (1984). *Chem. Phys. Lett.* **110,** 99.

Iodine...X(O, N, S) intermolecular contacts: Models of thyroid hormone–protein binding interactions using information from the Cambridge Crystallographic Data Files. Cody, V. and Murray-Rust, P. (1984). *J. Mol. Struct.* **112,** 189.

A six-site intermolecular potential scheme for the azabenzene molecules, derived by crystal structure analysis. Price, S.L. and Stone, A.J. (1984). *Mol. Phys.* **51,** 569.

Interactions between functional groups. Part I. Crystal structure of 2-phenylethynyl-*N*,*N*-dimethylaniline at 98K: A remarkably short C—H...C distance. Wallis, J.D. and Dunitz, J.D. (1984). *Helv. Chim. Acta* **67,** 39.

Reaction pathways from structural data: Dynamic stereochemistry of nickel compounds. Auf der Heyde, T.P.E. and Nassimbeni, L.R. (1984). *Inorg. Chem.* **23,** 4525.

Bond length and reactivity. Variable length of the C—O single bond. Allen, F.H. and Kirby, A.J. (1984). *J. Am. Chem. Soc.* **106,** 6197.

Internal molecular motions in crystals. The estimation of force constants, frequencies and barriers from diffraction data. A feasibility study. Trueblood, K.N. and Dunitz, J.D. (1983). *Acta Cryst., B (Str. Sci)* **39,** 120.

Chemical reaction paths. 9. Conformational interconversions of Wilkinson's catalyst and of related square-planar $XM(PR_3)_3$ compounds as determined from systematic analysis of solid-state structural data. Chandrasekhar, K. and Burgi, H.-B. (1983). *J. Am. Chem. Soc.* **105,** 7081.

Intermolecular interactions of the C—F bond: The crystallographic environment of fluorinated carboxylic acids and related structures. Murray-Rust, P., Stallings, W.C., Monti, C.T., Preston, R.K. and Glusker, J.P. (1983). *J. Am. Chem. Soc.* **105,** 3206.

Directional preferences of ether O-atoms towards alkali and alkaline earth cations. Chakrabarti, P. and Dunitz, J.D. (1982). *Helv. Chim. Acta* **65,** 1482.

Chemical reaction paths. 8. Stereoisomerization path for triphenylphosphine oxide and related molecules: Indirect observation of the structure of the transition state. Bye, E., Schweizer, W.B., Dunitz, J.D. (1982). *J. Am. Chem. Soc.* **104,** 5893.

Analysis of the atomic environment of quaternary ammonium groups in crystal structures using computerized data retrieval and interactive graphics: Modeling acetylcholine–receptor interactions. Rosenfield Junior, R.E. and Murray-Rust, P. (1982). *J. Am. Chem. Soc.* **104,** 5427.

Chemical reaction paths. 7. Pathways for SN2 and SN3 substitution at Sn(IV). Britton, D. and Dunitz, J.D. (1981). *J. Am. Chem. Soc.* **103,** 2971.

Directional preferences of approach of nucleophiles to sulfonium ions. Britton, D. and Dunitz, J.D. (1980). *Helv. Chim. Acta* **63,** 1068.

Computer retrieval and analysis of molecular geometry. 4. Intermolecular interactions. Murray-Rust, P. and Motherwell, W.D.S. (1979). *J. Am. Chem. Soc.* **101,** 4374.

Directional preferences of nonbonded atomic contacts with divalent sulfur. 1. Electrophiles and nuclophiles. Rosenfield Junior, R.E., Parthasarathy, R. and Dunitz, J.D. (1977). *J. Am. Chem. Soc.* **99,** 4860.