Concluding remarks

The pressure-induced structural transformation of CHD involves, apart from other effects, the interchange of donor and acceptor sites by the enolic H atoms in the hydrogen bonds. The structure, despite its close similarities with antiferroelectric crystals, is ordered in the high-pressure phase, also retaining its low-pressure symmetry. It is possible that the crystals have a domain structure in both low- and high-pressure phases. The proposed mechanism of the transformation strongly suggests an important role of electrostatic interactions in the transformation and in the jump of the enolic H atom to its other site in the OH···O bond. This mechanism also affords an explanation of the phase transition observed in CPD and of the exceptional stability of MCPD at high pressures; however, several points still need to be confirmed and further studies are being carried out on the CHD crystals.

The author is grateful to Professor Z. Kałuski for encouragement, to Dr R. J. Nelmes for his invitation to use the high-pressure and X-ray facilities of the Physics Department, University of Edinburgh, to Professor M. C. Etter of the Department of Chemistry, University of Minnesota, for stimulating discussions and providing the 1,3-cyclohexanedione samples crystallized from 2-pentanone, and to Dr J. Koput (Adam Mickiewicz University) for his expertise in the MNDO calculations. This study was partly supported by the British Council and by the Polish Academy of Sciences, Project CPBP, 01.12.

References

ALLMANN, R. (1977). In *Homoatomic Rings, Chains and Macromolecules of Main Group Elements*, edited by A. RHEINGOLD. Amsterdam: Elsevier.
DEWAR, M. J. S. & THIEL, W. (1977). *J. Am. Chem. Soc.* **99**, 4899–4907.
DREMIN, A. N. & BABARE, L. V. (1984). *J. Phys. (Paris) Colloq.* **C8**, 177–186.
DUPONT, L., LAMOTTE, J., CAMPSTEYN, H. & VERMEIRE, M. (1978). *Acta Cryst.* B**34**, 1304–1310.
ETTER, M. C., URBAŃCZYK-LIPKOWSKA, Z., JAHN, D. A. & FRYE, J. S. (1986). *J. Am. Chem. Soc.* **108**, 5871–5876.
FINGER, L. W. & KING, H. E. JR (1978). *Am. Mineral.* **63**, 337–342.
KATRUSIAK, A. (1989). *Acta Cryst.* C**45**, 1897–1899.
KATRUSIAK, A. (1990*a*). *Acta Cryst.* C**46**. In the press.
KATRUSIAK, A. (1990*b*). *J. Appl. Cryst.* **23**. Submitted.
KATRUSIAK, A. (1990*c*). *Acta Cryst.* B**46**. Submitted.
KATRUSIAK, A. (1990*d*). *Acta Cryst.* B**46**. In preparation.
KATRUSIAK, A. & NELMES, R. J. (1986). *J. Phys. C*, **19**, L765–L772.
KING, H. E. JR (1981). *High-Pressure Crystallography with a CAD-4*. Enraf–Nonius, Delft, The Netherlands.
MERRILL, L. & BASSETT, W. A. (1974). *Rev. Sci. Instrum.* **45**, 290–294.
MOTHERWELL, W. D. (1976). *PLUTO*. Program for plotting molecular and crystal structures. Univ. of Cambridge, England.
NICOL, M. & YIN, G. Z. (1984). *J. Phys. (Paris) Colloq.* **C8**, 163–172.
PIERMARINI, G. J., BLOCK, S., BARNETT, J. D. & FORMAN, R. A. (1975). *J. Appl. Phys.* **46**, 2774–2780.
SAMARA, G. A. & SEMMINGSEN, D. (1979). *J. Chem. Phys.* **71**, 1401–1407.
SAPRIEL, J. (1975). *Phys. Rev. B*, **12**, 5128–5139.
SEMMINGSEN, D. (1974). *Acta Chem. Scand. Ser. B*, **28**, 169–174.
SHELDRICK, G. (1976). *SHELX76*. Program for crystal structure determination. Univ. of Cambridge, England.
SINGH, I. & CALVO, C. (1975). *Can. J. Chem.* **53**, 1046–1050.
TUN, Z., NELMES, R. J. & MCINTYRE, G. J. (1987). *J. Phys. C*, **20**, 5667–5675.

Acta Cryst. (1990). B**46**, 256–262

Graph-Set Analysis of Hydrogen-Bond Patterns in Organic Crystals

BY MARGARET C. ETTER* AND JOHN C. MACDONALD

Department of Chemistry, University of Minnesota, Minneapolis, MN 55455, USA

AND JOEL BERNSTEIN

Ben-Gurion University of the Negev, Beer Sheva, Israel

(*Received* 17 *August* 1989; *accepted* 15 *November* 1989)

Abstract

A method is presented based on graph theory for categorizing hydrogen-bond motifs in such a way that complex hydrogen-bond patterns can be disentangled, or decoded, systematically and consistently. This method is based on viewing hydrogen-bond patterns topologically as if they were intertwined nets with molecules as the nodes and hydrogen bonds as the lines. Surprisingly, very few parameters are needed to define the hydrogen-bond motifs comprising these networks. The methods for making these assignments, and examples of their chemical utility are given.

* Alfred P. Sloan Foundation Fellow, 1989–1991.

0108-7681/90/020256-07$03.00

Introduction

Since hydrogen bonds were first proposed, efforts to characterize them, measure them, and understand them have proliferated (Latimer & Rodebush, 1920; Schuster, Zundel & Sandorfy, 1976; Joesten & Schaad, 1974). As it became obvious that these interactions occurred in many kinds of materials, and were important for determining the properties and activities of most biochemicals, hydrogen-bond classifications were developed; the most successful of these relate to their spectroscopic and chemical properties (Taylor & Kennard, 1984; Murray-Rust & Glusker, 1984; Emsley, 1980). Wells (1962) was one of the first to recognize the importance of hydrogen-bond patterns, independent of the geometry or of spectroscopic or thermal properties of the bonds. This perspective provided new insights into the role of hydrogen bonds in controlling the structures of ensembles of molecules. The consequences of repetitive hydrogen-bond interactions throughout a crystal structure or throughout a self-assembled aggregate structure had not previously been addressed. In other words, this topological point of view allowed one to characterize subsets of crystal structures or arrays, namely, the hydrogen-bonded subset.

Wells proposed a classification scheme for describing hydrogen-bond structures in inorganic compounds based on consideration of molecules as single points with hydrogen bonds as lines emanating from these points. Hamilton & Ibers (1968) developed this idea further, characterizing hydrogen-bonded networks with two numbers (N, M), the number of hydrogen bonds per point (N), and the number of molecules to which a point is hydrogen bonded (M). Kuleshova & Zorky (1980) recognized that these early classification schemes were actually an application of graph theory, a mathematical formalism for analyzing graphs and networks (Harary, 1967). This theory has been used for many different chemical applications, such as analysis of stereochemical topology (Walba, 1987), development of synthetic strategies (Fujita, 1988*a*) and coding of reaction pathways (Fujita, 1988*b*). Kuleshova & Zorky developed Wells' idea and applied graph theory to hydrogen-bond patterns in organic crystal structures. They identified finite sets, chains, layers and frameworks of hydrogen-bonded molecules, and they made initial surveys of the crystallographic literature to determine whether certain graph sets occurred more frequently than others. They also studied a set of hydrogen-bonded polymorphs and found that about half of them had the same graph set for both polymorphs (Zorky & Kuleshova, 1980).

The graph-set method presented here uses a molecular version of graph-set representations of hydrogen-bond patterns where functional groups and molecular structure are used explicitly. Graph theory has been adapted here rather freely to preserve a strong chemical component. The derivation from single points to molecules as the basic unit of a graph is important in order to represent hydrogen-bond patterns of organic molecules where different functional groups on a single molecule contribute in different ways to the hydrogen-bond network. Attempts are made to make the method applicable to as many different kinds of systems as possible within the constraints of keeping the notation and the process simple. Arrays that are easiest to handle using the graph-set method involve typical neutral organic molecules with a low density of hydrogen bonds, as opposed to structures like urea or oxalic acid which have high hydrogen-bond densities.

Graph sets, as presented here, are descriptions like the organic chemist's empirical formula or the crystallographer's space group. They tell how many donors and acceptors are used in a hydrogen-bond pattern and what the nature of the pattern is, yet they also highlight common features of molecular aggregates that are not addressed by empirical formulae or by the symmetry considerations of space groups. To analyze a hydrogen-bond pattern, the criteria for identifying hydrogen bonds must be established by the investigator. For the present purpose we find it useful to consider the collective behavior of molecules as they self assemble by interactions between hydrogen-bond donors and hydrogen-bond acceptors, without setting arbitrary limits on hydrogen-bond lengths or energies. The topologies of the patterns are considered rather than their particular geometries.

Operational definitions are given here and the process of assigning graph sets is explained.* Examples of how to apply graph sets to problems of decoding and comparing hydrogen-bond patterns are given at the end of this article.

Graph-set definitions

The set of molecules to be analyzed is called an *array*. Some or all of the molecules in the array are

* Several of the concepts developed here have a direct analogy with graph-theoretical concepts, while others have a more distant relation. To clarify these relations, the equivalencies are given, with the first term being the one used here, and the second italicized term being the one used by graph theoreticians (Harary, 1967): hydrogen bonds/*lines*; molecules/*points*; array/*graph*; motif/*spanning walk*; the superscript character *a* used in the notation/*indegree of a directed graph*; subscript character, *d*/*outdegree of a directed graph*. The most important difference between the two methods is that mathematical graphs involve points and lines having no structural properties. The points used in our method are molecules whose structure is implicit in the graph-set assignments. Likewise, we differentiate between different kinds of hydrogen bonds, but in mathematical graph sets all lines are identical.

associated through hydrogen bonds. This array need not be in a crystal structure, but regularly repeating sets such as those found in crystal structures are particularly suitable for graph-set assignments. The geometry of a hydrogen bond is not critical in these analyses, so even a hand-drawn picture of a hypothetical set of molecules is sufficient for making graph-set assignments.

A *network* is a subset of an array in which each molecule in the network is connected to every other molecule by at least one hydrogen-bonded pathway. A network contains any number of different kinds of hydrogen bonds. Our challenge is to define the morphology of this network.

A *motif* is a special type of network. *It is a hydrogen-bonded set in which only one type of hydrogen bond is present.* A hydrogen-bond type is defined by the chemical nature of the proton donor and acceptor used in the hydrogen bond. Thus, a hydrogen bond between a phenol and a nitro group would be a different type from a hydrogen bond between a phenol and a ketone. A motif is constructed by identifying all occurrences of one of these types of hydrogen bonds throughout the network. The subset of molecules that becomes hydrogen-bonded together by this operation is called a motif. The ability to single out one type of hydrogen-bond pattern at a time is one of the most useful features of this graph-set method.

Graph sets are assigned first to motifs, and then to networks. Frequently only one or a few motifs need to be assigned to answer questions about preferred aggregate patterns. This level of analysis is easy to learn, and to remember. The process of making the assignments is what is important here. It allows one to see clearly how multiple hydrogen-bond patterns are interrelated and to focus on the chemistry of particular sets of molecules.

A graph set is specified using the pattern designator (G), its degree (r), and the number of donors (d) and acceptors (a), as shown:

$$G_d^a(r).$$

G is a descriptor referring to the pattern of hydrogen bonding. It has four different assignments based on whether hydrogen bonds are inter- or intramolecular: S, C, R and D. S (standing for self) denotes an intramolecular hydrogen bond. Different patterns of intramolecular hydrogen bonds are not specified with this notation. A method for specifying particular intramolecular hydrogen bonds in cyclic peptide structures has been developed and may be useful for other types of molecules as well (Karle, 1981). For intermolecular bonds, C refers to hydrogen-bonded chains that are infinite, R refers to rings. A typical ring pattern is a cyclic carboxylic acid dimer. D refers to noncyclic dimers and other finite hydrogen-bonded sets, such as a phenol hydrogen-bonded to acetone.

The parameter r refers to the degree, being either the number of atoms in a ring or the repeat length of a chain. For a ring, in an S or R set, the degree is defined as the number of atoms in the ring, counted by traversing the ring in one direction along the shortest chain of covalent and hydrogen bonds until all the atoms in the ring have been counted once. For a chain, C, the degree is the repeat length of the monomer unit in the chain, *i.e.*, the number of atoms encountered by traversing the shortest pathway from the hydrogen atom of one hydrogen bond to the acceptor atom of the next. For D motifs the degree is the number of atoms in the entire length of the hydrogen-bonded set starting with the proton of the first hydrogen bond, proceeding along the shortest pathway, and ending with the acceptor atom in the last hydrogen bond of the set. If there is only one hydrogen bond in the motif, then the degree of the D pattern is 2. Since this pattern occurs so frequently, $r = 2$ is considered the default degree value for a D pattern and it is not specified. The parameters d and a refer to the number of different kinds of donors (d) and acceptors (a) used in the hydrogen-bond pattern. When all the molecules in a set are the same, d and a will be the number of participating donors and acceptors per molecule. When the set is composed of different kinds of molecules, d and a will be the sum of the participating donors and acceptors from all the different molecules. The default value for d and a is 1.

Assigning graph sets to motifs

First, identify the different types of hydrogen bonds in the array of interest. One motif will be generated for each type of hydrogen bond.

Second, rank the hydrogen bonds by chemical priority. This step is necessary so hydrogen-bond patterns can be reconstructed from graph sets [prioritization can be performed consistently by using the Cahn–Ingold–Prelog rules (IUPAC, 1970) which have been extended here to include hydrogen bonds (see *Appendix*)]. If hydrogen bonds are encountered that are not covered by application of these rules, they should be enumerated arbitrarily before proceeding further.

Third, generate a motif, selecting the highest priority hydrogen bond, H(1), and finding all occurrences of this bond in the array. To identify a motif choose one molecule as the starting point and identify all molecules that are attached to it by H(1). Then proceed from each of these molecules to all others bonded to them by H(1), *etc.* until a molecule is encountered which has no additional attached molecules, or until it is obvious that the set is infinite.

The process is like gluing certain molecules in an array together [with H(1) glue], then pulling one molecule out of the array to see which additional molecules come with it, and whether they form an infinite (*C*) or a finite (*S*, *R* or *D*) set. *This step is the essential decoding step in the procedure.*

Fourth, assign a graph set to the motif. A straightforward example is given as example (I) in Table 1, with graph set *D*. Example (II) is an infinite chain, with degree 7. Its complete graph set is *C*(7) (*A* stands for proton acceptor and *D*H for proton donor). In some cases H(1) occurs in a ring as well as in a chain. In these cases, the ring may be indicated as a subset, given in brackets. For (III), the complete graph set is $C_1^2(8)[R_1^2(4)]$. If the hydrogen atom of a single hydrogen bond occurs in two of the *same* rings at the same time, then the graph set assigned to one of the patterns is doubled. For (IV) (Table 1) the graph set is $2R_1^2(4)$.

Fifth, repeat steps three and four until graph sets have been assigned to all hydrogen-bond types.

Assigning graph sets to first-order networks

A first order network, N_1, is just a sequential listing of the graph sets that correspond to each motif in the network, where M_1 is the motif containing the highest priority hydrogen bond:

$$N_1 = M_j \cdots M_3 M_2 M_1.$$

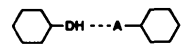
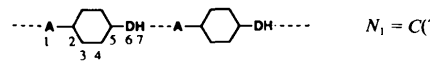
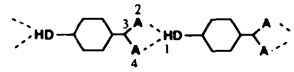
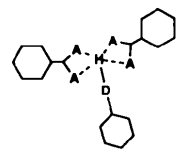
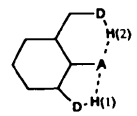
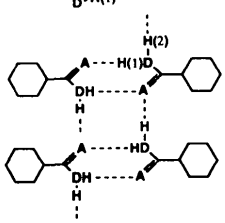
In examples (I)–(III) above, where there was only one kind of hydrogen bond in each example, their first-order networks consist of only one motif, M_1. In example (V), Table 1, there are two different kinds of hydrogen bonds in the pattern, so two motifs are needed to represent N_1. Since the highest priority hydrogen bond is *S*(5) and the second highest priority one is *S*(6), $N_1 = S(6)S(5)$.

Assigning graph sets to higher-order networks

Although all the information about individual hydrogen-bond contributors is given in the motifs, other hydrogen-bond patterns may be present in the structure, and in some cases it may be useful to identify them. These other patterns arise from combinations of two or more different types of hydrogen bonds (recalling that motifs contain only one kind of hydrogen bond). Higher-order networks, N_i, are assigned by adding sequentially one new graph set per network. The lowest degree pattern is given the highest priority, as for motifs, and is assigned first. The lowest degree pattern will usually contain only two different kinds of proton donors, but it could contain more. If there are still two new patterns with the same priority (*e.g.*, two different kinds of chains with $r = 6$) then the ranking is specified by the user.

Table 1. *Examples of graph-set assignments*

The hexagons are meant to represent any organic ligand. This specific form is given so the degrees of the patterns can be assigned in this table. If the organic ligands were other kinds of groups, then the degrees of some of the patterns would change.

(I)		$N_1 = D$
(II)		$N_1 = C(7)$
(III)		$N_1 = C_1^2(8)[R_1^2(4)]$
(IV)		$N_1 = 2R_1^2(4)$
(V)		$N_1 = S(6)S(5)$ $N_2 = S_2^1(9)$
(VI)		$N_1 = C(4)R_2^2(8)$ $N_2 = R_4^2(8)$

There is a simple procedure for finding higher-order networks. Search for that pattern (of any graph-set type) that has the smallest number of atoms in its repeat unit, yet contains two or more different kinds of proton donors. In searching for chains, start with H(1) for example, and proceed along a hydrogen bond and along the continuing atom chain until reaching H(2). Then find a pathway that returns to H(1) in another molecule without including any other H atoms. If this sequence propagates itself then it is accepted as a chain. The procedure is applied to all the pairwise combinations of the individual proton donors, and thereafter to triplet combinations, *etc.*

Rings are readily identified by a similar procedure. A ring of second order or above must contain at least two types of hydrogen bonds. Hence start again with H(1), proceed along its hydrogen bond and along its intramolecular atomic chain to H(2). The smallest ring possible would then include a second H(1), and a second H(2), returning in a cycle to the starting H(1). Again, proceeding through all pairwise combinatorial possibilities higher-order rings can be identified. Rings with more than two types of hydrogen bonds in them are assigned as even higher networks, with ring size (or degree) determining their respective priorities.

Higher-order networks arise by combination of motifs from two different kinds of hydrogen bonds. In example (VI), Table 1, one kind of hydrogen bond gives rise to a ring and another kind of hydrogen bond gives rise to chains. The first-order network for (VI) is $C(4)R_2^2(8)$. When the dimer is repeated by a chain, several new networks are created. The highest priority one is a new eight-membered ring between the original dimers. Thus, $N(2)$ is $R_4^2(8)$. Graph sets can be assigned to even higher-order networks in the same way as demonstrated for N_2.

Troubleshooting

Most, but not all, hydrogen-bond patterns of small molecules can be assigned with the method outlined above. For some compounds, there may be complicating circumstances that make the assignments intractable. For hydrogen-bond patterns that are very difficult to visualize or to decode a useful approach is to assign just part of the pattern to a graph set. In some cases an optional notation using O superscript and O subscript can be used to resolve difficulties:

$$[G_d^a(r)]_o^o$$

An example where optional fields are useful is in comparison of Watson–Crick, Hoogsteen and reverse Hoogsteen hydrogen-bond patterns (Hoogsteen, 1959, 1963; Saenger, 1984). The graph set of the Watson–Crick pattern is characteristic, but the graph sets of the other two types are identical to one another even though their hydrogen-bond patterns are not. An optional field identifying which oxygen atom is used in the hydrogen-bond pattern resolves this ambiguity allowing each of the three hydrogen-bond types to have a unique and useful graph set, as shown in Fig. 1.

Applications of graph sets

The use of graph sets for designing new materials, deconvoluting crystal structure data, and comparing and contrasting sets of molecules has been demonstrated (Etter, 1989; Bernstein, Etter & MacDonald, 1989). This approach views hydrogen-bonded sets of molecules as distinct chemical species that can be studied independently. The molecular sets are conceptually like macromolecules. A particular graph set may be common to many kinds of molecules. Such sets are isographic as well as isoentropic since the number of intermolecular hydrogen-bond connections are the same for all sets of molecules with the same graph set. For example, *N*-methyl-4-nitroaniline (VII) and 4-nitrophenol (VIII) are isographic (Panunto, Urbañczyk-Lipkowska, Johnson & Etter, 1987) even though the proton-donating ability of a phenol ($pK_a = 9{\cdot}9$) is much greater than that of an aniline ($pK_a = 27$), Fig. 2.

Similarly, the graph sets of three polymorphs of iminodiacetic acid (IX) have been analyzed. Despite the extraordinary complexity of the intermeshed hydrogen-bond patterns in these structures, and despite the fact that the molecules are zwitterions (the graph-set method was developed here primarily for use with neutral organic molecules), the similarities and differences in the graph sets of the three polymorphs provides an explicit basis for comparison and insight into the chemical similarities and differences of the polymorphs (Bernstein *et al.*, 1989), Fig. 3.

Since hydrogen bonds are generally weak interactions, most hydrogen-bonded assemblies have been studied in the solid state where they are stable enough for spectroscopic and chemical investigations. The occurrence of certain hydrogen-bond motifs throughout a series of compounds, *e.g.*,

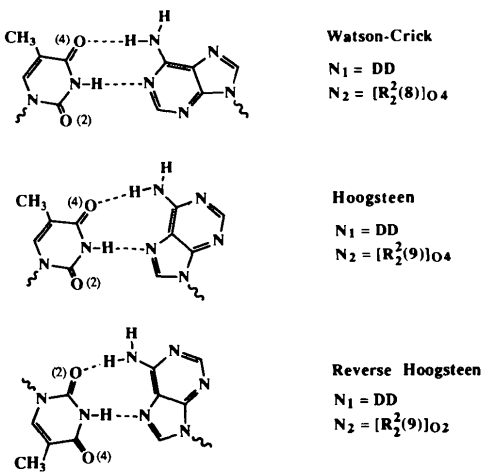
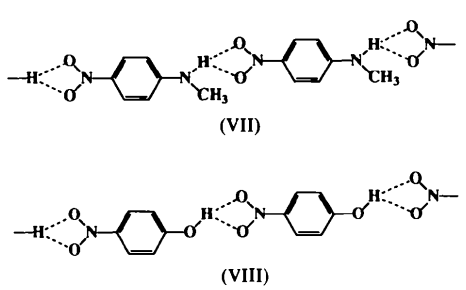


Fig. 1. An example of the use of optional fields (brackets with sub/superscripts) to explicitly define hydrogen-bond patterns. In these examples the ambiguity about O2 or O4 as the hydrogen-bond acceptor has been resolved with the use of optional fields.



(VII)

(VIII)

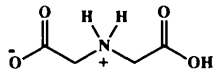
Fig. 2. Isographic hydrogen-bond patterns of a secondary aniline and a phenol. Both sets of molecules have the same first- and second-order graph sets: $N_1 = R_1^2(4)$ and $N_2 = C_1^2(8)$.

$C(4)R_2^2(8)$ for primary amides, or $R_2^2(8)$ for carboxylic acids, suggests that these patterns may be present in solution and may be controlling the structure of crystal nucleation sites. Even though the hydrogen-bond aggregates would be transient in solution, and may even be present in very low concentrations, at the instant when nucleation begins these structures must be forming. It is unlikely that other crystal-packing forces, which vary almost randomly throughout a series of compounds, could by chance direct all primary amides into the same hydrogen-bond pattern. Rather, it is more likely that crystal-packing forces influence the observed geometry of the hydrogen bonds, but do not determine their connectivity patterns. Preferred hydrogen-bond modes derived from crystal structures may be useful for determining the preferred modes of association of the individual functional groups on complex multi-functional molecules like biopolymers. Hopefully, graph-set analysis will aid in transferring information about molecular recognition properties of molecules from the solid state to solution and mobile phases, and will provide a unifying methodology for describing hydrogen-bonded sets of molecules.

APPENDIX

Assigning hydrogen-bond priorities

The Cahn–Ingold–Prelog (CIP) rules deal specifically with priorities of groups attached to a carbon atom. We propose here that the same protocol be used for assigning priorities to hydrogen bonds based first on the priorities of the hydrogen atoms in the hydrogen bonds, secondly on the acceptor atoms, and finally on other hydrogen-bond criteria (IUPAC, 1970). Often the CIP rules can be applied unchanged to hydrogen bonds. For example, a hydrogen atom attached to an oxygen would be higher priority than one attached to a nitrogen since the oxygen has higher atomic number.



Iminodiacetic acid zwitterion (IX)

Polymorph 1:	N_1 = C(5)R_2^2(10)C(8)	$N_2 = R_2^2(14)$
Polymorph 2:	N_1 = C(5)R_2^2(10)C(8)	$N_2 = R_4^2(8)$
Polymorph 3:	N_1 = C(5)C(5)C(8)	

Fig. 3. Graph-set assignments for the three polymorphs of (IX). All three polymorphs have C(5)C(8) in common in their first-order networks N_1. Polymorphs 1 and 2 differ in their second-order networks N_2.

When CIP rules are not directly applicable they can be extended as follows with the highest priority rules listed first:

1. Primary amides – the *cis* hydrogen atom is higher priority than the *trans* hydrogen atom.

2. Primary amines – the hydrogen atom on the highest priority side of the molecule is given highest priority.

3. In a carboxylic acid, the hydrogen atom with *cis* geometry is higher priority than one with *trans* geometry.

4. A hydrogen atom in an intermolecular hydrogen bond has priority over one in an intramolecular hydrogen bond.

5. The CIP priorities of acceptor atoms in hydrogen bonds are used to determine priorities of the hydrogen atoms when two hydrogen atoms are identical by all other criteria above.

6. When the acceptor and donor atoms in two hydrogen bonds are identical, then the following criteria apply:

(*a*) Lone pairs of electrons are treated as acceptors. The highest priority lone pair is on the highest priority side of a molecule.

(*b*) A hydrogen atom in a two-center bond is higher priority than one in a single bond. Likewise, hydrogens in multicenter bonds have even higher priorities.

(*c*) A hydrogen atom in a short hydrogen bond has a precedence over one in a longer bond.

Financial support from NIH (GM 42148-01) and the Petroleum Research Fund of the American Chemical Society to MCE is gratefully acknowledged.

References

BERNSTEIN, J., ETTER, M. C. & MACDONALD, J. C. (1989). *J. Chem. Soc. Perkin Trans.* 2. In the press.

EMSLEY, J. (1980). *Chem. Soc. Rev.* **9**, 91–124.

ETTER, M. C. (1989). *Acc. Chem. Res.* In the press.

FUJITA, S. J. (1988*a*). *J. Chem. Inf. Comput. Sci.* **28**, 128–137.

FUJITA, S. J. (1988*b*). *J. Chem. Inf. Comput. Sci.* **28**, 137–142.

HAMILTON, W. C. & IBERS, J. A. (1968). *Hydrogen Bonding in Solids*, pp. 19–21. New York: W. A. Benjamin.

HARARY, F. (1967). *Graph Theory and Theoretical Physics*. New York: Academic Press.

HOOGSTEEN, K. (1959). *Acta Cryst.* **12**, 822–823.

HOOGSTEEN, K. (1963). *Acta Cryst.* **16**, 907–916.

IUPAC (1970). *J. Org. Chem.* **35**, 2849–2867.

JOESTEN, M. D. & SCHAAD, L. J. (1974). *Hydrogen Bonding*. New York: Marcel Dekker.

KARLE, I. (1981). *The Peptides*, Vol. 4, edited by E. GROSS & J. MEIENHOFER, pp. 1–54. New York: Academic Press.

KULESHOVA, L. N. & ZORKY, P. M. (1980). *Acta Cryst.* **B36**, 2113–2115.

LATIMER, W. M. & RODEBUSH, W. H. (1920). *J. Am. Chem. Soc.* **42**, 1419–1433.

MURRAY-RUST, P. & GLUSKER, J. P. (1984). *J. Am. Chem. Soc.* **106**, 1018–1025.

PANUNTO, T. W., URBAŃCZYK-LIPKOWSKA, Z., JOHNSON, R. & ETTER, M. C. (1987). *J. Am. Chem. Soc.* **109**, 7786–7797.
SAENGER, W. (1984). *Principles of Nucleic Acid Structure*, ch. 6. New York: Springer-Verlag.
SCHUSTER, P., ZUNDEL, G. & SANDORFY, C. (1976). *The Hydrogen Bond*, Vols. I–III. Amsterdam: North-Holland.
TAYLOR, R. & KENNARD, O. (1984). *Acc. Chem. Res.* **17**, 320–326.
WALBA, D. M. (1987). *Graph Theory and Topology in Chemistry, Studies in Physical and Theoretical Chemistry*, Vol. 51, edited by R. B. KING & D. H. ROUVRAY, pp. 23–42. Amsterdam: Elsevier.
WELLS, A. F. (1962). *Structural Inorganic Chemistry*, pp. 294–315. Oxford: Clarendon Press.
ZORKY, P. M. & KULESHOVA, L. N. (1980). *Zh. Strukt. Khim.* **22**, 153–156.

Acta Cryst. (1990). B**46**, 262–266

Structure of a Low-Temperature Polymorph of Chenodeoxycholic Acid, $C_{24}H_{40}O_4$, Determined with Synchrotron Radiation

BY P. J. RIZKALLAH

Department of Chemistry, Liverpool University, PO Box 147, *Liverpool L*69 3*BX, England, and SERC Daresbury Laboratory, Warrington WA*4 4*AD, England*

MARJORIE M. HARDING

Department of Chemistry, Liverpool University, PO Box 147, *Liverpool L*69 3*BX, England*

P. F. LINDLEY

*Department of Crystallography, Birkbeck College, Malet Street, London WC*1*E* 7*HX, England*

AND A. AIGNER AND A. BAUER

Diamalt Aktiengesellschaft, Werk Raubling, Postfach 1151, *D*-8201 *Raubling, Federal Republic of Germany*

(*Received* 21 *July* 1989; *accepted* 17 *October* 1989)

Abstract

The low-temperature polymorph of chenodeoxycholic acid ($3\alpha,7\alpha$-dihydroxy-5β-cholanic acid) forms very fine needle-shaped crystals (30–60 μm cross-section), containing a variable solvent content. The crystal and molecular structure has been solved using synchrotron radiation in conjunction with an area detector. $M_r = 392{\cdot}6$ (excluding solvent), hexagonal, space group $P6_5$, $a = 22{\cdot}250\ (5)$, $c = 10{\cdot}255\ (2)$ Å, $Z = 6$, $V = 4396{\cdot}7$ Å^3, $F(000) = 1296$, $D_m = 1{\cdot}12\ (6)$ g cm^{-3} (flotation in bromobenzene/toluene), $D_x = 0{\cdot}89$ g cm^{-3} for $C_{24}H_{40}O_4$ without solvent of crystallization, D_x (max.) $= 1{\cdot}17$ g cm^{-3} for $C_{24}H_{40}O_4{\cdot}C_6H_5Br$, $\lambda = 1{\cdot}5418$ Å for unit-cell determination, $\lambda = 0{\cdot}895$ Å for intensity measurements, $\mu = 1{\cdot}09$ cm^{-1}, m.p. = 388–390 K, $T = 291\ (1)$ K. The structure refined to $R = 0{\cdot}11$ for 578 reflections with $|F_o| > 2\sigma(F)$. The molecular conformation is similar to that of the high-melting-point (438–439 K) form, but the molecular packing is far more open. All the oxygen atoms are involved in intermolecular hydrogen bonding. The crystal structure contains wide cylindrical channels, parallel to **c**, sufficient to accommodate molecules with van der Waals diameters up to 8·0 Å.

0108-7681/90/020262-05$03.00

Introduction

Chenodeoxycholic acid, $3\alpha,7\alpha$-dihydroxy-5β-cholanic acid (CDCA), $C_{24}H_{40}O_4$, is a bile acid which is an efficacious agent for inducing dissolution of cholesterol gallstones (see, for example, Hofmann & Paumgartner, 1975). The cholanic bile acids form a number of mixed crystals which may be sub-divided into two categories, choleic acids and 'canal' complexes. The choleic acids contain highly stable complexes of host (bile acid) and guest (second organic molecule) which may, indeed, be stable outside the crystalline state. The crystal structures of the canal complexes contain channels or 'canals' which contain the guest molecules. The host–guest interactions in these complexes are usually less specific than for the choleic acids and can have non-integral coordination numbers. The canals can be either hydrophilic whereby they readily accommodate guest molecules with polar substituents, or hydrophobic where the guest molecules tend to be apolar. Although several canal complexes involving deoxycholic acid, DCA, have been reported, very few where the host molecule is another bile acid have been characterized. Polymorphism of CDCA has been reported (Guiseppetti & Paciotti, 1978), and the high-temperature