Acta Crystallographica Section B
Structural
Science
ISSN 0108-7681

On the inclusion of solvent molecules in the crystal structures of organic compounds

Carl Henrik Görbitz* and Hans-Petter Hersleth

Department of Chemistry, University of Oslo, PO Box 1033 Blindern, N-0315 Oslo, Norway

Correspondence e-mail:
c.h.gorbitz@kjemi.uio.no

Received 30 September 1999
Accepted 7 January 2000

The Cambridge Structural Database has been searched for all crystal structures including organic solvent molecules (solvates) and solvent water molecules (hydrates). Well above 300 different solvent molecules were identified and the frequencies with which they occur in crystal structures, as a function of the year of publication, were established. The crystal structures are classified as 'organic' and 'metalloorganic'; it is shown that the relative prevalences of various cocrystallized solvents are different in the two groups. Several frequently used organic solvents are also common ligands for metal ions. Special interest has been focused on the existence of heterosolvates, *i.e.* crystal structures which include more than one type of solvent molecule. Up to five different types of solvent molecules were found in a single crystal structure. It is suggested that the use of solvent mixtures during crystallizations may prove to be a more useful and versatile approach for obtaining crystals of high-molecular-weight organic compounds than has hitherto been recognized.

1. Introduction

The crystal structure determination for an organic molecule often reveals that a cocrystal with one or more solvent molecules has been formed. In such cases the solvent usually receives little attention, or may even be regarded as a nuisance that causes crystals to be unstable or introduces disorder into an otherwise ordered structure. For many compounds, however, it should be kept in mind that the presence of solvent molecules is of vital importance for a successful crystallization outcome. Structures with large empty channels or cavities are generally not stable (Kitaigorodski, 1961) and occasionally crystals may not be obtained if the intrinsically unstable scaffolding of solute molecules is not supported by carefully selected solvent molecules. Previous investigations (Van der Sluis & Kroon, 1989) indicate that the relative number of structures with organic solvent molecules is steadily increasing, suggesting that it becomes increasingly more important to consider the possibilities for solvent inclusion in crystallization experiments.

Solvent molecules in crystals have essentially three different functions (Van der Sluis & Kroon, 1989):

(*a*) as participants in hydrogen-bonding networks (Görbitz & Gundersen, 1996*a*,*b*; Görbitz, 1997, 1999; Görbitz & Torgersen, 1999);

(*b*) as space fillers, with no strong interactions between solvent and solute molecules;

(*c*) as ligands completing the coordination around a metal ion.

A solvent molecule may simultaneously have at least two of these functions. Cocrystallized alcohol molecules thus usually combine functions (*a*) (hydroxyl group) and (*b*) (alkyl group) and in this way form bridges between non-polar and polar regions of the crystal.

Functions (*a*) and (*b*) are almost always shown by solvents being used during the synthesis and/or for the subsequent crystallization, but ligands for metals may be introduced in a deliberate step during synthesis and can be very tightly bound.

In this paper we present a comprehensive discussion of the inclusion of solvents in crystal structures based on a systematic survey of the available experimental material in the Cambridge Structural Database (Allen & Kennard, 1993). Water is by far the most common solvent in organic structures and dominates all the statistics when included. For this reason, structures with only solvent water molecules, called 'hydrates', were treated separately. The term 'solvate' has been reserved for structures incorporating organic solvent molecules. Structures with solvent water and an organic solvent molecule are consequently called 'hydrate solvates'. Throughout, emphasis has been put on solvent molecules with functions (*a*) and (*b*) (free solvent molecules).

2. Methods

There is no clear definition of a solvent, and thus no obvious distinction between a solvent molecule and a solute molecule. We have used a very broad definition of the solvent, demanding initially only that the compound is a liquid at room temperature.

The retrieval of entries from the Cambridge Structural Database (CSD; October 1998 release; Allen & Kennard, 1993) was essentially a five-step process, as follows.

(i) Identification of unique CSD Refcodes in two categories of structures; organic (ORG) and metalloorganic (MORG).

(ii) Initial searches for solvates within these two subsets of the database.

(iii) Final searches for solvates and hydrates, identification of solvate mixtures. Study of ligands for metal ions.

(iv) Extraction of publication-date-dependent information on solvent inclusion.

(v) Study of correlation between molecular weight and solvent inclusion.

2.1. Step (i)

The differentiation between ORG and MORG structures was made on the basis of a bit screen (No. 57) for each CSD entry, which is set to 1 (ORG) if the compound does not contain any transition element, actinide, lanthanide, Al, Ga, In, Tl, Ge, Sn, Pb, Sb, Bi or Po. Other entries belong to the MORG group (bit screen set to 0). This simple definition of a MORG compound may occasionally violate the conception of a true MORG compound by researchers working in the field.

Some compounds have been subject to special interest by crystallographers in the past, reflected by the occurrence of two or more entries in the database, all within the same Refcode family *ABCDEFnn*, where *ABCDEF* is the basic six-letter code and *nn* is a number. An in-house program was used to eliminate duplicates and keep only one Refcode for each compound. For compounds with more than one distinct crystal form, it would have been desirable to keep one Refcode for each, but it is our impression (with no statistics to confirm it) that polymorphism is uncommon among solvates. We believe the possible omittance of a small number of solvate polymorphs is a very minor problem with the statistics presented.

Since the scope of the current study was not to explore fine structural details, no limits on quality indicators, such as the *R* factor, have been used, and no restrictions on experimental data type were implemented.

Two files with unique Refcodes for ORG and MORG structures were generated in this step, and formed the basis for all subsequent searches.

2.2. Step (ii)

There is no specific bit screen or other indication in a CSD entry that the structure is a solvate, and with the variety of possible solvents a specific structural search is not possible. Initially, solvates were identified by searching for the text strings 'solvate' or 'clathrate'. A considerable amount of time was spent manually weeding out those structures which most chemists would not consider as solvates, particularly for the 'clathrate' subset which were often considered to be 1:1 or 1:2 complexes rather than solvates (KEZZUK, VIVZAB).[1]

2.3. Step (iii)

After establishing which solvents were the 50 most common within the ORG and MORG categories, respectively, new searches for structures containing each specific solvent molecule were carried out. Accordingly, structures which include the top 50 solvent molecules, but were not identified as solvents in the title of the CSD entry, were found, giving a separate file of Refcodes. The number of additional structures was typically small, around 2–5% of the number already retrieved in the initial search. One additional file of Refcodes (No. 51) for each category was obtained by searching for the text strings 'solvate' and 'clathrate' with the limitation that none of the first 50 solvates were allowed to be present. Hydrates were retrieved by searching for the keyword 'hydrate' as well as the structure H–O–H. The latter option resulted in about 7% fewer hits, and very few entries that had not already been hit by the first alternative. We believe that the material retrieved by these search procedures is very close to complete for the important solvents used in organic chemistry.

Solvent mixtures were identified by a second in-house computer program that filtered the contents of the 50 files of Refcodes for ORG and MORG structures, giving as output

---

[1] Representative examples of structures of various categories discussed in the text are identified by their CSD Refcodes. A list of associated references is available from the IUCr electronic archives (Reference: OS0038). Services for accessing these data are described at the back of the journal.

Table 1
Solvent inclusion in organic structures.

The column headed '% free' gives the percentage of the total number of free solvates. The column headed '% free + other' gives percentage of free solvates which include at least one additional type of solvent molecule. The column headed '% free + water' gives the percentage of free solvates which also incorporate water molecules. For each solvent tabulated in this paper, the CSD Refcodes for structures that include that solvent (in the form of files that can be used as input to the *QUEST* program) are available from the IUCr electronic archives (Reference: OS0038)

Solvent		Occurrences as free	% free	Hetero-solvates	% free + other	Solvate hydrates	% free + water	Occurrences as ligand	Ligand + solvates
1	Methanol	762	14.20	40	5.2	115	15.1	37	7
2	Dichloromethane	457	8.52	32	7.0	17	3.7	0	0
3	Benzene	444	8.27	15	3.4	11	2.5	6	3
4	Ethanol	398	7.42	13	3.3	80	20.1	18	5
5	Chloroform	395	7.36	27	6.8	25	6.3	0	0
6	Acetonitrile	389	7.25	17	4.4	47	12.1	16	1
7	Acetone	332	6.19	12	3.6	38	11.4	9	1
8	Toluene	279	5.20	15	5.7	8	2.9	3	1
9	Tetrahydrofuran	199	3.71	6	3.0	12	6.0	472	52
10	Ethyl acetate	155	2.89	6	3.9	16	10.3	3	1
11	Diethyl ether	148	2.76	9	6.1	2	1.4	134	3
12	Dioxane	139	2.59	5	3.6	12	8.6	18	3
13	Dimethyl sulfoxide	129	2.40	5	3.9	18	14.0	7	0
14	*N*,*N*-Dimethylformamide	114	2.12	3	2.6	12	10.5	18	2
15	*n*-Hexane	92	1.71	10	10.9	4	4.3	0	0
16	Acetic acid	83	1.55	5	6.0	18	21.7	1	0
17	Pyridine	73	1.36	2	2.7	9	12.3	48	9
18	2-Propanol	61	1.14	1	1.6	9	14.8	2	0
19	Cyclohexane	49	0.91	7	14.3	1	2.0	0	0
20	*p*-Xylene	42	0.78	2	4.8	2	4.8	0	0
21	Tetrachloromethane	41	0.76	2	4.9	1	2.4	0	0
22	Carbon disulfide	34	0.63	0	0.0	0	0.0	0	0
23	Nitrobenzene	30	0.56	9	30.0	2	6.7	0	0
24	Nitromethane	30	0.56	3	10.0	6	20.0	0	0
25	1,2-Dichloroethane	27	0.50	3	11.1	2	7.4	0	0
26	*n*-Pentane	25	0.47	3	12.0	0	0.0	0	0
27	1-Propanol	24	0.45	0	0.0	8	33.3	3	0
28	1,2-Dimethoxyethane	19	0.35	0	0.0	0	0.0	87	3
29	Chlorobenzene	16	0.30	2	12.5	0	0.0	0	0
30	*m*-Xylene	15	0.28	0	0.0	0	0.0	0	0
31	2-Butanone	14	0.26	1	7.1	0	0.0	0	0
32	*n*-Butanol	14	0.26	0	0.0	0	0.0	2	0
33	*o*-Xylene	14	0.26	0	0.0	1	7.1	0	0
34	Ethylenediamine	13	0.24	0	0.0	0	0.0	18	3
35	1,2-Dichlorobenzene	12	0.22	0	0.0	0	0.0	0	0
36	*N*,*N*-Dimethylacetamide	12	0.22	4	33.3	1	8.3	5	0
37	Bromobenzene	11	0.20	0	0.0	1	9.1	0	0
38	Cyclohexanone	11	0.20	0	0.0	1	9.1	0	0
39	Formic acid	11	0.20	1	9.1	1	9.1	0	0
40	*n*-Heptane	11	0.20	0	0.0	1	9.1	0	0
41	Acetophenone	10	0.19	0	0.0	0	0.0	1	0
42	Benzonitrile	9	0.17	0	0.0	0	0.0	3	0
43	Ethylene glycol	9	0.17	0	0.0	2	22.2	0	0
44	1,1,2-Trichloroethane	8	0.15	0	0.0	0	0.0	0	0
45	2-Butanol	8	0.15	0	0.0	1	12.5	0	0
46	*t*-Butanol	8	0.15	0	0.0	0	0.0	5	0
47	Acetylacetone	7	0.13	0	0.0	0	0.0	1	0
48	Diethyl ketone	7	0.13	0	0.0	0	0.0	0	0
49	Propionic acid	7	0.13	0	0.0	0	0.0	0	0
50	Dibromomethane	6	0.11	0	0.0	0	0.0	0	0
Other solvents		296	5.52	4	1.4				
Total		5366†		132†	2.5	462†	9.0	917	94

† The sum of the column entries is slightly larger than the number of structures since heterosolvates are included more than once.

only those Refcodes which were present in more than one file. Mixtures where both solvent molecules had rank >50 were easily identified by manual checking of file 51 for each group. The main problem was associated with mixtures where one solvent molecule had rank $\leq$50 and one had rank >50. After substantial manual checking we believe that most of these structures have been found. Hydrate solvates including one of the top 50 solvents in each group were found by filtering a file with hydrate Refcodes against the 50 files for organic solvent molecules.

When a solvent molecule (or anything else) acts as a ligand for a metal ion, the CSD entry has a metal$\cdots$ligand bond encoded, and this bond needs to be defined in the search fragment with the bond type set as 'any'. Searches were carried out with solvents of the top 50 list for ORG and MORG compounds, excluding non-aromatic molecules without heteroatoms, such as cyclohexane, *n*-pentane *etc.* Separate searches were furthermore carried out to identify structures in which the same compound is present both as free solvent molecules and as a ligand.

2.4. Step (iv)

The time-dependent occurrences of solvates and hydrates were calculated for the 15 most common solvents and 'others' by carrying out a number of searches with limits on the *YEAR qualifier for each entry.

2.5. Step (v)

One of the objects of this investigation was to establish the correlation between the molecular weight (or molecular volume) of the solute and the propensity to include a solvent molecule in the crystal structure. Unfortunately, there is no easy way of retrieving the molecular weight or volume of the solute molecule from the CSD entry. Instead, the information stored in each database entry as *MVOL has been used, which essentially is the sum of the solute and the solvent volumes. This is not what we had in mind originally, but since most of the solvent molecules are small compared with the solute molecules, the resulting statistics should still be of interest.

3. Results and discussion

3.1. General observations

Among 87 185 ORG and 101 457 MORG entries in the CSD, there are 76 966 and 91 146 unique Refcodes, respectively. The number of hydrates is 6233 (8.1%) for the ORG structures and 9615 (10.5%) for the MORG structures, while for solvates the numbers are 5366 (7.0%) and 15 782 (17.3%), respectively. These figures include structures with at least one free solvent molecule, and not structures where the solvent acts only as a ligand for a metal ion.

3.2. Solvates

For ORG structures we have identified (by our broad definition of such) 220 different cocrystallized organic solvent molecules. Additionally, 30 structures were found for which the nature of the solvent molecule was unknown. The 50 most common solvent molecules for the ORG structures are given in Table 1. Methanol is on the top of the list (14.2%) and solvents with rank from 1 to 15 occur in 82.4% of the total number of solvates. Solvents with rank from 16 to 50 constitute 14.4%, while the remaining solvents with rank from 51 to 220 constitute 5.5% of the solvent molecules. The sum of these three figures is 102.4%, showing that some structures contain more than one type of solvent molecule (see below).

The number of different solvent molecules identified in MORG crystal structures is 195, which is smaller than observed for ORG structures, despite the fact that the number of MORG structures is significantly larger. The list of the 50 most common solvents for the MORG structures, Table 2, contains many of the same entries as the list for ORG structures, Table 1, and 13 of the top 15 solvents in the ORG list are also among the top 15 solvents in the MORG list. Nevertheless, 99 solvent molecules have been found exclusively in MORG structures, bringing the total number of solvent molecules identified in this survey up to 319. Dichloromethane is the dominating solvent molecule, occurring in 23.5% of the total number of solvates. The top 15 solvents are observed in 94.5% of the solvates, while solvents with rank from 16 to 50 and 51 to 195 are observed in only 7.7 and 2.0% of the solvates, respectively (sum = 104.1%). This means that for MORG compounds there is less diversity with respect to solvent inclusion than among organic structures in general.

The relative number of structures that include solvent molecules has not remained constant with publication date, but rather shows a steady and significant increase through the years (Fig. 1). For the MORG group, the percentage is as high as 23.1% in 1995–1997. This increase is probably a result of increasingly larger and more complex molecules being included in the database. Fig. 2 shows how the fraction of

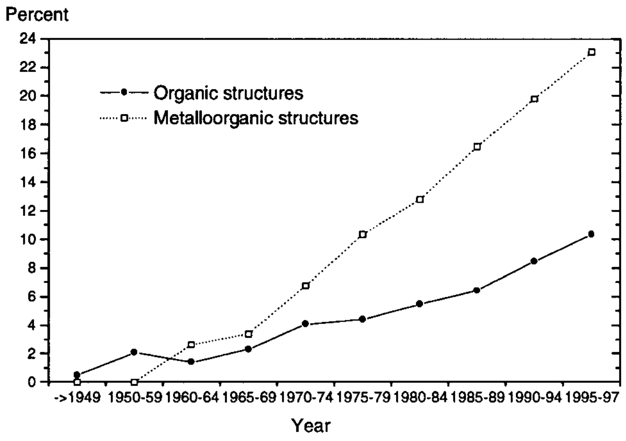

Figure 1
Proportion (in %) of the published crystal structures which contain cocrystallized organic solvent molecules.

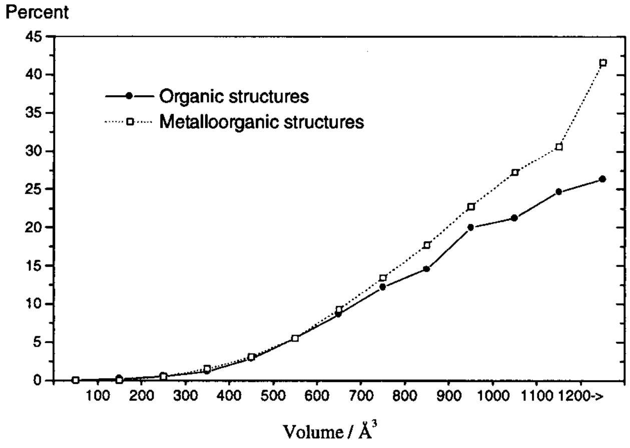

Figure 2
Correlation between the proportion of solvates and the combined molecular volume of the solute and the solvent molecules.

Table 2
Solvent inclusion in metalloorganic structures.

The column headed '% free' gives the percentage of the total number of free solvates. The column headed '% free + other' gives percentage of free solvates which include at least one additional type of solvent molecule. The column headed '% free + water' gives the percentage of free solvates which also incorporate water molecules. For each solvent tabulated in this paper, the CSD Refcodes for structures that include that solvent (in the form of files that can be used as input to the *QUEST* program) are available from the IUCr electronic archives (Reference: OS0038).

Solvent		Occurrences as free	% free	Hetero-solvates	% free + other	Solvate hydrates	% free + water	Occurrences as ligand	Ligand + solvates
1	Dichloromethane	3704	23.47	231	6.2	127	3.4	5	0
2	Acetonitrile	1493	9.46	127	8.5	111	7.4	909	183
3	Benzene	1439	9.12	99	6.9	17	1.2	250	25
4	Toluene	1374	8.71	93	6.8	10	0.7	118	8
5	Tetrahydrofuran	1354	8.58	106	7.8	14	1.0	1556	235
6	Methanol	1203	7.62	145	12.1	211	17.5	492	153
7	Chloroform	976	6.18	79	8.1	38	3.9	0	0
8	Acetone	815	5.16	51	6.3	69	8.5	56	10
9	Diethyl ether	645	4.09	138	21.4	29	4.5	142	10
10	Ethanol	596	3.78	74	12.4	121	20.3	191	36
11	*N*,*N*-Dimethylformamide	454	2.88	29	6.4	54	11.9	247	53
12	*n*-Hexane	377	2.39	69	18.3	8	2.1	0	0
13	Pyridine	248	1.57	12	4.8	20	8.1	1578	186
14	Dimethyl sulfoxide	145	0.92	10	6.9	22	15.2	278	18
15	*n*-Pentane	142	0.90	30	21.1	4	2.8	0	0
16	Nitromethane	130	0.82	18	13.8	11	8.5	1	1
17	1,2-Dichloroethane	128	0.81	10	7.8	3	2.3	3	0
18	Dioxane	89	0.56	5	5.6	1	1.1	47	14
19	Chlorobenzene	83	0.53	5	6.0	4	4.8	1	0
20	Cyclohexane	66	0.42	11	16.7	2	3.0	0	0
21	Carbon tetrachloride	64	0.41	10	15.6	2	3.1	0	0
22	2-Propanol	52	0.33	8	15.4	4	7.7	45	0
23	Ethylenediamine	46	0.29	1	2.2	3	6.5	924	8
24	Ethyl acetate	45	0.29	6	13.3	6	13.3	13	1
25	Acetic acid	44	0.28	3	6.8	10	22.7	25	1
26	1,2-Dimethoxyethane	42	0.27	4	9.5	1	2.4	160	15
27	*n*-Heptane	37	0.23	6	16.2	2	5.4	0	0
28	*p*-Xylene	28	0.18	3	10.7	0	0.0	18	3
29	Carbon disulfide	26	0.16	2	7.7	0	0.0	16	0
30	Aniline	26	0.16	2	7.7	1	3.8	50	2
31	4-Methylpyridine	22	0.14	1	4.5	5	22.7	190	20
32	*n*-Butanol	22	0.14	5	22.7	4	18.2	3	0
33	1,1,2,2-Tetrachloroethane	21	0.13	4	19.0	1	4.8	0	0
34	Benzonitrile	21	0.13	0	0.0	1	4.8	52	2
35	Diglyme	19	0.12	0	0.0	0	0.0	39	1
36	*m*-Xylene	19	0.12	2	10.5	1	5.3	13	1
37	*N*,*N*-Dimethylacetamide	19	0.12	1	5.3	8	42.1	18	0
38	1,2-Dichlorobenzene	17	0.11	3	17.6	1	5.9	0	0
39	Mesitylene	17	0.11	4	23.5	1	5.9	116	4
40	Nitrobenzene	17	0.11	0	0.0	2	11.8	2	0
41	*o*-Xylene	16	0.10	2	12.5	0	0.0	6	0
42	Ethylene glycol	16	0.10	0	0.0	1	6.3	18	2
43	1-Propanol	15	0.10	2	13.3	2	13.3	4	1
44	Propionitrile	13	0.08	3	23.1	1	7.7	23	0
45	Dibromomethane	12	0.08	1	8.3	0	0.0	1	0
46	Cyclopentane	11	0.07	1	9.1	0	0.0	0	0
47	Pyrrole	11	0.07	0	0.0	0	0.0	0	0
48	1,1,2-Trichloroethane	10	0.06	0	0.0	0	0.0	0	0
49	Di-isopropyl ether	10	0.06	2	20.0	1	10.0	0	0
50	Fluorobenzene	10	0.06	0	0.0	1	10.0	4	0
Other solvents		316	2.00	2	0.6				
Total		15 782†		697†	4.4	863†	5.5	7614	993

† The sum of the column entries is slightly larger than the number of structures since heterosolvates are included more than once.

solvates depends on the molecular volume [volume of solvent + solute(s); see above]. The correlation between solvent inclusion and molecular volume (or rather the sum of solute and solvent volumes; see above) is evident both for ORG and MORG compounds, and within the latter group more than 40% of the structures with molecular volume above 1200 Å^3 include organic solvent molecules. Evidently, large molecules can often not pack efficiently without leaving cavities that are occupied by solvent molecules.

The time-dependent occurrence of some of the most popular solvents is shown in Fig. 3. In general, the percentages for methanol, ethanol, acetone and benzene have come down, while the percentages for acetonitrile, toluene and dichloromethane are increasing. The reduction for benzene (inversely related to the increase for toluene) is clearly associated with the potential health hazards involved with direct exposure, but since Fig. 3 shows percentages and the total number of solvates increases, all solvents, including benzene, occur in larger absolute numbers today than before. The increases for *n*-hexane and *n*-pentane for MORG structures are also noteworthy; *n*-pentane was actually observed for the first time as a solvent molecule in this group in 1978 (TMSICR, BPEAFE, TPFEBP) and its use has since increased uniformly to reach a current fraction of 1.2% (Fig. 3).

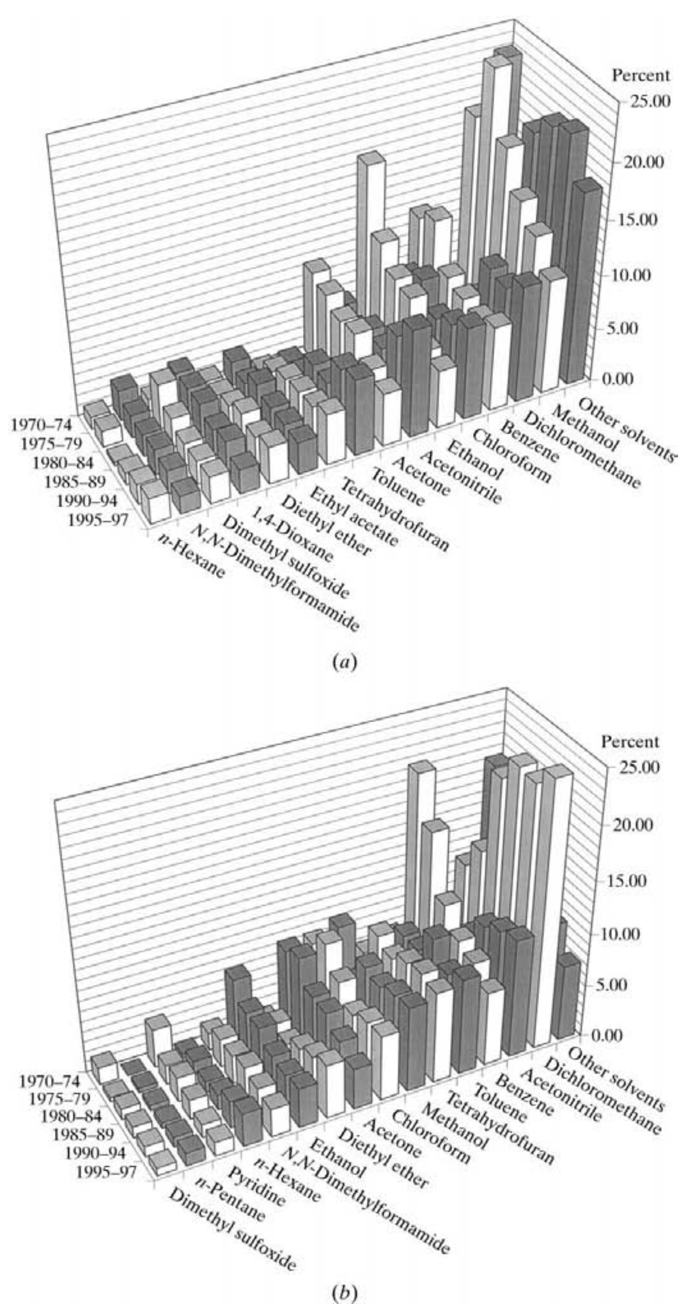

Figure 3
The variation of the relative frequencies of the 15 overall most common solvents in crystal structures with time of publication. (*a*) Organic structures. (*b*) Metalloorganic structures.

Although various solvents undoubtedly have different propensities to be cocrystallized with organic and metalloorganic compounds (see below), we believe that the data presented in Tables 1 and 2 and Fig. 3 mostly indicate which solvents are (and have been) in use by organic chemists for synthesis and crystallization purposes.

3.3. Hydrates

After an initial rise, the fraction of hydrates among ORG structures declined from almost 18% in the late 1950s to stabilize at 7% after 1985 (Fig. 4). The development is rather similar for MORG structures, with a peak at 18% around 1950, at which time a high percentage of the structures would be classified as coordination compounds, and a stabilization at about 10% after 1985.

3.4. Heterosolvates

In the ORG group, we found 131 structures with two different organic solvent molecules and a single structure (KEHDAC) with three solvent molecules. Some statistics for these structures are presented in Table 3. The occurrence of a solvent in a mixture is obviously related to the total frequency with which the solvent occurs in crystal structures, but we are still able to say that some compounds are more prone to be included in a heterosolvate than others. These are *n*-hexane, chloroform, dichloromethane and diethyl ether, and to a smaller extent toluene and methanol. Apart from this it is difficult to find specific trends in the statistical material. Even the most common heterosolvate (methanol–dichloromethane) (KAGWIY, RIQMOT, WETKIP) has been observed only nine times, and just 28 combinations of solvents have been observed more than once in ORG structures.

For MORG compounds, we found 673, 22 and 2 heterosolvates with two, three (GAGRUB, NABGEC, TOJMEK) and four (LALRIZ, VONMUG) different solvent molecules, respectively (Table 4). Overall, more MORG structures than ORG structures contain two or more solvent molecules.

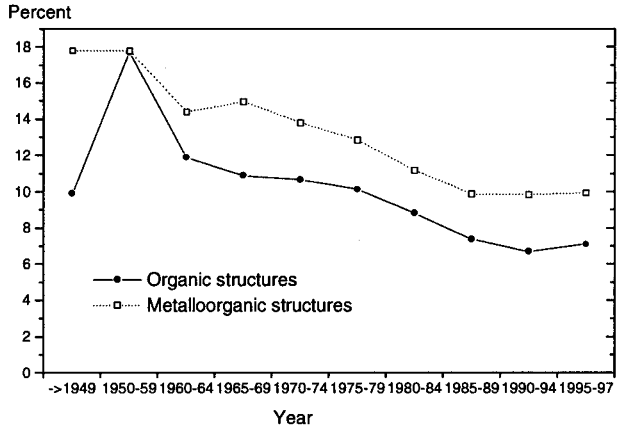

Figure 4
The publication-date-dependent occurrence of water (hydrates) in crystal structures.

Table 3
The most common heterosolvates for organic structures.

Solvents			Number	%
1	Dichloromethane	Methanol	9	6.9
2	Chloroform	Methanol	6	4.6
3	Ethanol	Methanol	6	4.6
4	Acetonitrile	Methanol	5	3.8
5	Chloroform	Dichloromethane	4	3.1
6	Dichloromethane	*n*-Hexane	4	3.1
7	Acetone	Dichloromethane	3	2.3
8	Acetone	Methanol	3	2.3
9	Acetonitrile	Chloroform	3	2.3
10	Acetonitrile	Dichloromethane	3	2.3
11	Benzene	Chloroform	3	2.3
12	Dichloromethane	Diethyl ether	3	2.3
13	Tetrahydrofuran	Toluene	3	2.3
Other combinations of two solvents			76	58.0
Total			131	
Combinations of three solvents			1	

Table 4
The most common heterosolvates for metalloorganic structures.

Solvents			Number	%
1	Dichloromethane	Diethyl ether	44	6.5
2	Dichloromethane	*n*-Hexane	29	4.3
3	Acetonitrile	Diethyl ether	25	3.7
4	Dichloromethane	Methanol	23	3.4
5	Chloroform	Methanol	22	3.3
6	Acetonitrile	Methanol	21	3.1
7	Benzene	Dichloromethane	19	2.8
8	Benzene	Tetrahydrofuran	19	2.8
9	Tetrahydrofuran	Toluene	18	2.7
10	Diethyl ether	Methanol	17	2.5
11	Dichloromethane	Ethanol	16	2.4
12	Acetonitrile	Dichloromethane	15	2.2
13	Chloroform	Dichloromethane	15	2.2
14	Dichloromethane	Toluene	15	2.2
Other combinations of two solvents			375	55.7
Total			673	
Combinations of three solvents			22	
Combinations of four solvents			2	

Among the top 15 solvents, *n*-pentane, diethyl ether, *n*-hexane, ethanol and methanol stand out as likely to participate in a heterosolvate. It is noteworthy that these are all linear molecules without aromatic groups or branching, although further down the list there are high percentages for cyclohexane and CCl_4.

It is very interesting that the proportion of heterosolvates has increased dramatically in recent years (Fig. 5). This is probably in part a result of generally more complex molecules being synthesized, which more easily leave cavities of various sizes when forming a crystal. It is difficult to tell if this development also reflects increased use of solvent mixtures by organic chemists for crystallization purposes. From reading the original literature on the crystallization experiments, it is our impression than heterosolvates usually result from a synthesis which involves different solvents in subsequent steps, but also from liquid diffusion experiments involving two different solvents and from slow evaporation from a solvent mixture.

3.5. Hydrate solvates

The most common organic solvents in hydrate solvates are those which are completely miscible with water, such as methanol, ethanol, acetone, acetonitrile and *N*,*N*-dimethylformamide, but also less polar molecules like dichloromethane and tetrahydrofuran. Indeed, we find an unexpectedly high number of hydrates which incorporate rather hydrophobic, unpolar solvents, like benzene (BOHDUX, CVERBZ10), chloroform (DIGBEA, BUYXUO) and diethyl ether (KIXGIH, SAXFAY). For a molecule like 4-methylpyridine, 23% of the solvates also incorporate solvent water molecules (GADFIA). When screening the methods used for obtaining the crystals used in the diffraction experiments, we find either that water and the organic solvent were used in separate steps during a synthesis, or that an organic solvent saturated with water (*e.g.* 'wet benzene'; ZUDBOP) was used for crystallization.

The hydrate solvates are not confined to structures with one single type of organic solvent molecule. Two different types of solvent molecules in addition to water occur in 22 organic (LAKWAV, GAJFAY, TUFXAT) and 65 MORG (JUCKUN, LAHTUJ, PIRMOS) structures. Four MORG structures (TOJMEK, TOVXOR, VURHAR, YAVFIK) even have three types of organic solvent molecules, and one structure (LALRIZ) appears (there is some uncertainty due to disorder) to have four additional different solvent molecules, giving an unparalelled total number of five different cocrystallized solvent molecules.

3.6. Common solvent molecules as ligands for metal ions

The above discussion has dealt with structures in which the solvent molecules are not tightly associated with metal ions. In this section we focus on structures in which the overall 50 most common free solvent molecules in crystals act as ligands for metal ions. Most of these structures evidently fall into the MORG category, but a fair number of structures in the ORG category also include metal ions, *i.e.* various salts and complexes of the alkali and alkaline-earth metals.

In our statistical material on MORG compounds, four common solvents dominate as ligands for metal ions (Table 2). They are pyridine (20.7%), tetrahydrofuran (20.4%), ethylenediamine (12.1%) and acetonitrile (11.9%). While benzene forms the familiar interaction with the metal ion positioned over the face of the ring [or in a few instances at an edge of the ring (BZCATC10, DEMGEH)], pyridine (except HEVBEP) and its derivatives as well as non-aromatic solvents all interact with the metal ion through a lone pair on an O atom or N atom, usually in a monodentate fashion, but also in a bidentate fashion for a number of ethylenediamine ligands (ADENCO10, AEIACO) and in a bi- or tridentate fashion for diglyme [$O(C_2H_4\text{–}O\text{–}CH_3)_2$; BELKAE, PEDTUN, DIRYUY]. One exception is CS_2, where the lone pairs sit on an S atom. A few examples of Cl–CH_2–Cl acting as a ligand were also found (JAMSAR, LADLOR). There is never a distinct upper limit for a ligand$\cdots$metal distance, and for several structures it is a matter of discussion whether or not an interaction is really present.

Water is observed as a ligand in 5166 MORG structures, which is 5.7% of the total number of structures and more than three times more than for tetrahydrofuran and pyridine.

It is not uncommon that a crystal structure contains two or more crystallographically independent solvent molecules of the same type, with one (or more) acting as a ligand for a metal ion and one (or more) acting as a free solvent molecule. This is observed for about 13% of the MORG structures with common solvent molecules as ligands, and for 10% of the ORG structures. In 31% of the MORG structures where methanol is a ligand, it is also found as a free solvent molecule (BIHCOK, BUFYOQ), while the corresponding figure for a strongly binding ligand like ethylenediamine is just 0.9% (KILBAI).

3.7. The use of solvent mixtures for crystallization purposes

The choice of solvent(s) for crystallization is usually dictated by the intuition of the chemist, taking into consideration the nature of the solute molecule and solvent parameters, such as solubility, reactivity, hydrophobicity/polarity, vapour pressure, boiling point/melting point, hydrogen-bond donating/accepting ability, *etc*. For certain compounds, the presence of special (reactive) functional groups or other aspects of the structure may limit the array of potentially useful solvents dramatically, while for others, obtaining favourable solubility properties [not too high, not too low (Van der Sluis *et al.*, 1989)] may be more important. Under any circumstance, the use of solvent mixtures could prove useful. Firstly, instead of setting up a large number of crystallization experiments in the search for the 'right' solvent for high-quality crystal formation, a smaller number of experiments with solvent mixtures can be used. Secondly, mixed solvent crystals may be obtained, unavailable with pure solvents.

Mixtures can easily be derived from the most common solvents, for instance by adding small percentages of one (or more) other solvent such as chloroform, 1,2-dichloroethane and possibly linear molecules (see above) like *n*-pentane or *n*-hexane to dichloromethane, or replacing methanol with a mixture containing methanol and ethanol and other simple alcohols. Mixtures containing hydrogen-bond donating and accepting solvents as well as more unpolar solvents should be considered, if appropriate, to allow inclusion of solvent into two or more types of cavities with different properties.

An unfortunate liability of solvent inclusion in crystal structures is that solvent molecules tend to be disordered more often than the solute molecules. In the worst case of liquid-like disorder, positive crystallographic identification of the solvent molecule may be impossible (Van der Sluis & Kroon, 1989). Less serious types of disorder make the structure refinement more laborious; data at high 2θ angles are lost and the final R factor may be disappointingly high. Sometimes this is unavoidable, but the choice of solvent(s) should be reconsidered when the results are completely inadequate. A solvent molecule that fits better into the pertinent cavity, by being larger or having a different shape, may be found (Görbitz & Torgersen, 1999; Kraaijkamp *et al.*, 1983). On the other hand, disorder associated with symmetry operators usually does not create serious problems. Toluene is for instance often found on special positions, which may force solvent molecule disorder. This is why structures with toluene have a higher chance of disorder than those containing benzene (Van der Sluis & Kroon, 1989).

As far as mixtures are concerned, one should avoid simultaneously using two molecules with roughly the same shape and volume, since both could be incorporated at the same solvent site and create a disorder which is difficult to resolve (Görbitz & Torgersen, 1999). It should also be pointed out that since the components in a mixture have different vapour pressures, the composition will change (though in a predictable way if the mixture is ideal) during evaporation. In particular, care must be taken to avoid the essential volatile components drifting off before saturation is reached and precipitation can start.

It is finally worth pointing out that water molecules are excellent candidates for filling small cavities in crystal structures (and forming hydrogen bonds), and if the solute molecule is stable in the presence of water, an aqueous mixture, or one that is saturated with water, could be used.

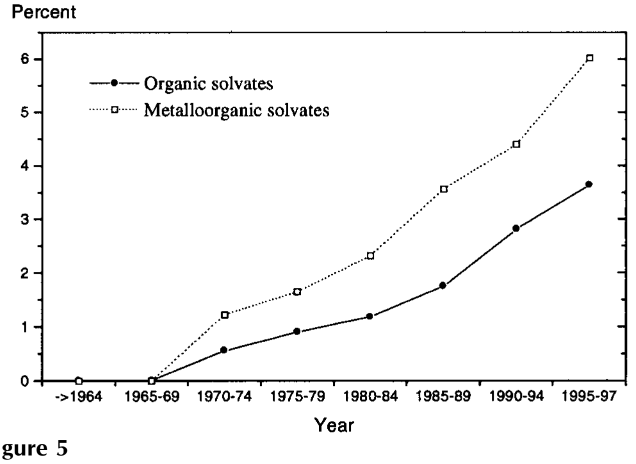

Figure 5
The proportion of solvates in which there are two or more different types of solvent molecules (heterosolvates).

4. Conclusions

The increased proportion of crystal structures which include cocrystallized solvent molecules is related to the increasing complexity of the molecules being synthesized, making a packing arrangement without cavities or channels less easily attainable. Finding the right solvent molecules may be of critical importance for obtaining crystals, and even apparently minor changes, such as the repositioning of a methyl group from 1-propanol to 2-propanol, may be sufficient. Furthermore, many hypothetical crystal structures will contain more than one type of cavity, with different sizes, requiring at least two types of solvent molecules for crystal formation. We therefore advocate the use of solvent mixtures in order to improve the chances of a successful crystallization outcome.

Note added in proof: In a recent article based on a CSD survey of non-metal-containing organic structures, Nangia & Desiraju (1999) have estimated the abilities of several common organic solvents to be included in the crystal. The results are quantified as values for corrected occurrence ($O_{\rm corr}$) with 1.00 as the scaled average value. DMF is on the top of the list with $O_{\rm corr}$ = 5.69 followed by DMSO (4.73), dioxane (4.70) and *p*-xylene (3.81). Methanol and dichloromethane have $O_{\rm corr}$ values 1.17 and 1.38, respectively.

The authors thank Per Groth for writing essential computer programs, and Professor Mats Tilset, University of Oslo, for critically reading this manuscript.

References

Allen, F. H. & Kennard, O. (1993). *Chem. Des. Autom. News*, **8**, 31–37.

Görbitz, C. H. (1997). *Acta Cryst.* C**53**, 736–739.

Görbitz, C. H. (1999). *Acta Cryst.* C**55**, 670–672.

Görbitz, C. H. & Gundersen, E. (1996*a*). *Acta Chem. Scand.* **50**, 537–543.

Görbitz, C. H. & Gundersen, E. (1996*b*). *Acta Cryst.* C**52**, 1764–1767.

Görbitz, C. H. & Torgersen, E. (1999). *Acta Cryst.* B**55**, 104–113.

Kitaigorodski, A. I. (1961). *Organic Chemical Crystallography*. New York: Consultants Bureau.

Kraaijkamp, J. G., van Koten, G., Vrieze, K., Grove, D. M., Klop, E. A., Spek, A. L. & Schmidpeter, A. (1983). *J. Organomet. Chem.* **256**, 375–389.

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

Van der Sluis, P., Hezemans, A. M. F & Kroon, J. (1989). *J. Appl. Cryst.* **22**, 340–344.

Van der Sluis, P. & Kroon, J. (1989). *J. Cryst. Growth*, **97**, 645–656.