International Tables for Crystallography (2006). Vol. F, Figure 24.3.2.2, p. 664.

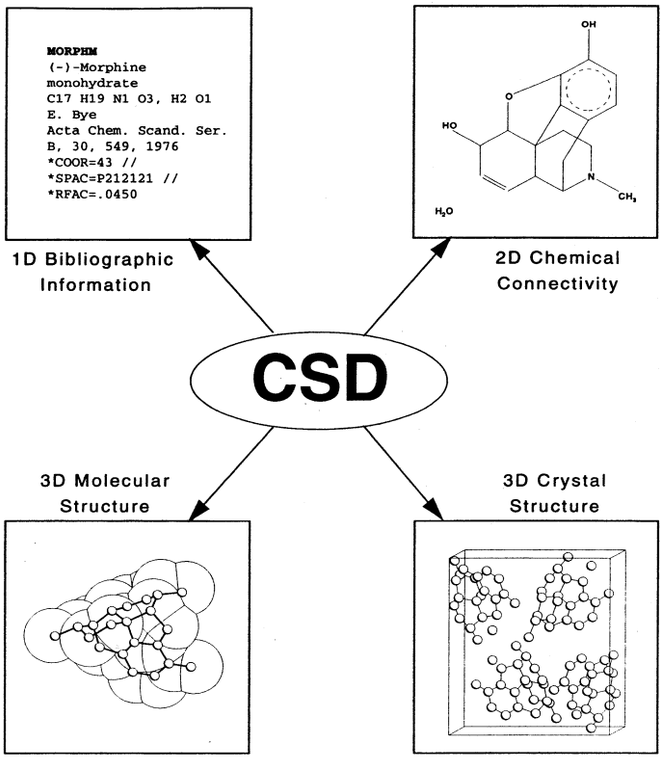


Fig. 24.3.2.1. Information content of the Cambridge Structural Database (CSD).

typical CSD entry is illustrated schematically in Fig. 24.3.2.1. Individual data items can be categorized into three different groupings which are most conveniently described in terms of their dimensionality.

24.3.2.3. *1D bibliographic and chemical data*

The one-dimensional data for each entry comprise chemical and bibliographic text strings, together with certain individual numerical items, *viz* chemical compound name and any common synonym(s), chemical formula, authors names, journal name and literature citation, text comment reflecting any special experimental details (non-room-temperature study, absolute configuration determined, neutron study *etc.*). The cell parameters, crystal data, space group and precision indicators also fall into this category.

24.3.2.4. *2D chemical connectivity data*

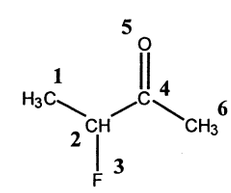
The formal two-dimensional chemical structural diagram for each entry (Fig. 24.3.2.2) is encoded in the form of a compact connection table. Chemical connectivity is recorded in terms of a set of atom and bond properties. The atom properties recorded are: atom number, element type, number of connected non-H atoms, number of terminal H atoms and the formal atomic charge. Bond properties are encoded as a pair of atom numbers and the formal chemical bond type that connects those atoms. Bond types employed in the CSD connectivity descriptions are: single, double, triple, quadruple (metal–metal), aromatic, delocalized double and π bonds. Bond types are (automatically) coded negative if the bond forms part of a cyclic system.

24.3.2.5. *3D crystal structure data*

The three-dimensional data consist of the fractional coordinates and symmetry operators for each entry. This information, together with the cell dimensions, is used to establish a crystallographic connectivity using standard covalent radii. The chemical and crystallographic connectivities are then mapped onto one another, using graph-theoretic algorithms, so that the chemical atom and bond properties are associated with the three-dimensional structure for search purposes. The CSD always records coordinates for complete molecules. Thus, if a molecule adopts a special position in the assigned space group, *i.e.* the asymmetric unit is some fraction of the total number of atoms in the molecule, then the CSD system also records those symmetry-generated atoms that complete the chemical entity. This speeds up the search process and also makes the data more accessible to non-crystallographers.

24.3.2.6. *Derived data and bit-encoded information*

Derived data are calculated directly from the evaluated raw data and stored in the master archive for search purposes. Numerical items such as Z', the number of chemical entities in the asymmetric unit, is a typical (real) numerical data item in this category. However, by far the most useful of the derived data items are a set of 682 individual pieces of yes/no information which are encoded as a bitmap, referred to as the *screen* record. The first 155 of these bits record information about (*a*) the elemental constitution of the compound, (*b*) results of the data-validation procedure and (*c*) summary information about the data content of the entry. These bits can be accessed directly by the user as search keys. The most important parts of the bitmap contain codified yes/no information about the presence/absence of specific features in the complete 2D or 3D structures held in the CSD. When a chemical substructure is entered as a query, its constitution is analysed in the same way to produce a bitmap for the query. Logical comparison of the query bitmap with the bitmap stored for each full CSD entry is computationally rapid, and quickly eliminates those entries that do not contain the requested features. Only those entries that pass this initial screening process need enter the detailed and computationally intensive atom-by-atom, bond-by-bond connectiv-



Atom Properties						
Atom Number	1	2	3	4	5	6
Element Number	C	C	F	C	O	C
No. Connected Non-hydrogen Atoms	1	3	1	3	1	1
No. Terminal Hydrogen Atoms	3	1	0	4	0	3
Net Charge	0	0	0	0	0	0

Bond Properties					
Atom 1 of Bond	2	2	2	4	4
Atom 2 of Bond	1	3	4	5	6
Bond Type	1	1	1	2	1

Fig. 24.3.2.2. 2D chemical connectivity data for a simple organic molecule.

Copyright © 2006 International Union of Crystallography

references