

3. CIF DATA DEFINITION AND CLASSIFICATION

Compatibility with PDB format files (§3.6.8.3.2)

DATABASE_PDB_CAVEAT
 DATABASE_PDB_MATRIX
 DATABASE_PDB_REMARK
 DATABASE_PDB_REV
 DATABASE_PDB_REV_RECORD
 DATABASE_PDB_TVECT

The purpose of entries in the DATABASE category group is to provide pointers that link the mmCIF to all database entries that result from the deposition of the file. For mmCIF, the relevant category is DATABASE_2, which replaces the DATABASE category of the core dictionary.

Note the distinction between the database pointers provided here and those in the STRUCT_REF family of categories. The latter are intended to provide links to external database entries for any aspect of any subset of the structure that the author may wish to record, including previous determinations of the same structure, other structures containing the same ligand or references to the sequence(s) of the macromolecule(s) in sequence databases. In contrast, the links provided in DATABASE_2 refer to the entire contents of the mmCIF and are designed to cover situations in which the entire file is deposited in more than one database (for example, in the PDB and in a database for protein kinases).

3.6.8.3.1. Related database entries

Data items in these categories are as follows:

(a) DATABASE

- *_database.entry_id*
 → *_entry.id*
 _database.code_CAS
 _database.code_CSD
 _database.code_ICSD
 _database.code_MDF
 _database.code_NBS
 _database.code_PDB
 _database.code_PDF
 _database.code_depnum_ccdc_archive
 _database.code_depnum_ccdc_fiz
 _database.code_depnum_ccdc_journal
 _database.CSD_history
 _database.journal_ASTM
 _database.journal_CSD

(b) DATABASE_2

- *_database_2.database_id*
- *_database_2.database_code*

The bullet (•) indicates a category key. Where multiple items within a category are marked with a bullet, they must be taken together to form a compound key. The arrow (→) is a reference to a parent data item. Items in italics have aliases in the core CIF dictionary formed by changing the full stop (.) to an underscore (_).

The DATABASE category is retained in the mmCIF dictionary, but only for consistency with the core dictionary.

The role of the data items in the DATABASE_2 category is to store identifiers assigned by one or more databases to the structure described in the mmCIF. In the data model used in the core CIF dictionary, each database has an individual data item. The data model in mmCIF is more general. It comprises the data items *_database_2.database_id*, which identifies the database, and *_database_2.database_code*, which is the code assigned by the database to the entry. Thus a new database can be referred to without needing to add an additional data item to the dictionary. If a structure has been deposited in more than one database, the values of *_database_2.database_id* and *_database_2.database_code* can be looped.

The institutions and individual databases recognized in the DATABASE_2 category in the current version of the mmCIF dictionary are CAS (Chemical Abstracts Service), CSD (Cam-

bridge Structural Database), ICSD (Inorganic Crystal Structure Database), MDF (Metals Data File), NDB (Nucleic Acid Database), NBS (the Crystal Data database of the National Institute of Standards and Technology, formerly the National Bureau of Standards), PDB (Protein Data Bank), PDF (Powder Diffraction File), RCSB (Research Collaboratory for Structural Bioinformatics) and EBI (European Bioinformatics Institute). It is intended that new databases will be added to this list on an ongoing basis; the purpose of specifying a list of possible databases in the dictionary is to ensure that each database is referenced consistently.

3.6.8.3.2. Compatibility with PDB format files

Data items in these categories are as follows:

(a) DATABASE_PDB_REV

- *_database_PDB_rev.num*
 _database_PDB_rev.author_name
 _database_PDB_rev.date
 _database_PDB_rev.date_original
 _database_PDB_rev.mod_type
 _database_PDB_rev.replaced_by
 _database_PDB_rev.replaces
 _database_PDB_rev.status

(b) DATABASE_PDB_REV_RECORD

- *_database_PDB_rev_record.rev_num*
 → *_database_PDB_rev.num*
- *_database_PDB_rev_record.type*
 _database_PDB_rev_record.details

(c) DATABASE_PDB_MATRIX

- *_database_PDB_matrix.entry_id*
 → *_entry.id*
 _database_PDB_matrix.origx[1][1]
 _database_PDB_matrix.origx[1][2]
 _database_PDB_matrix.origx[1][3]
 _database_PDB_matrix.origx[2][1]
 _database_PDB_matrix.origx[2][2]
 _database_PDB_matrix.origx[2][3]
 _database_PDB_matrix.origx[3][1]
 _database_PDB_matrix.origx[3][2]
 _database_PDB_matrix.origx[3][3]
 _database_PDB_matrix.origx_vector[1]
 _database_PDB_matrix.origx_vector[2]
 _database_PDB_matrix.origx_vector[3]
 _database_PDB_matrix.scale[1][1]
 _database_PDB_matrix.scale[1][2]
 _database_PDB_matrix.scale[1][3]
 _database_PDB_matrix.scale[2][1]
 _database_PDB_matrix.scale[2][2]
 _database_PDB_matrix.scale[2][3]
 _database_PDB_matrix.scale[3][1]
 _database_PDB_matrix.scale[3][2]
 _database_PDB_matrix.scale[3][3]
 _database_PDB_matrix.scale_vector[1]
 _database_PDB_matrix.scale_vector[2]
 _database_PDB_matrix.scale_vector[3]

(d) DATABASE_PDB_TVECT

- *_database_PDB_tvect.id*
 _database_PDB_tvect.details
 _database_PDB_tvect.vector[1]
 _database_PDB_tvect.vector[2]
 _database_PDB_tvect.vector[3]

(e) DATABASE_PDB_CAVEAT

- *_database_PDB_caveat.id*
 _database_PDB_caveat.text

(f) DATABASE_PDB_REMARK

- *_database_PDB_remark.id*
 _database_PDB_remark.text

The bullet (•) indicates a category key. Where multiple items within a category are marked with a bullet, they must be taken together to form a compound key. The arrow (→) is a reference to a parent data item.