Decifer and CSD Editor User Guide

1 Decifer and CSD Editor User Guide
  1.1 Conditions of Use

2 Introduction
  2.1 Installation notes
    2.1.1 Other requirements to use DeCIFer at the command line
  2.2 Licensing DeCIFer

3 DeCIFer
  3.1 Updating an existing in-house database
  3.2 Standardising bonds in a legacy database
  3.3 Adding new structures to/creating a new in-house database
    3.3.1 Adding additional information to the CIF before conversion
    3.3.2 Creating a new database from a set of CIFs
    3.3.3 Adding new CIFs to an existing database
    3.3.4 Using existing chemical diagrams in the conversion
    3.3.5 Adding citation information to the entries during conversion
    3.3.6 Conversion to current CSDS database format
  3.4 Using in-house databases with CSD Portfolio software
    3.4.1 Activating your in-house database for use with ConQuest
    3.4.2 Generating your in-house database for use with Mogul
    3.4.3 Generating your in-house database for use with WebCSD

4 CSD-Editor
  4.1 Opening the CSD-Editor
  4.2 Opening entries in the CSD-Editor
  4.3 Editing entries in the CSD-Editor
    4.3.1 Overview Tab
    4.3.2 Compare tab
    4.3.3 All text tab
4.3.4 Diagram tab
4.3.5 Visualiser tab
4.3.6 All Hits tab

4.4 Saving entries in the CSD-Editor

4.5 CSD-Editor FAQs
4.5.1 How should I deal with polymeric structures?
4.5.2 How should I deal with disordered structures?
4.5.3 How do I select all atoms in a molecule in the 3D structure?
4.5.4 How do I select a group of atoms in the 3D structure?
4.5.5 What does the colour coding in the identifier list signify?
4.5.6 How do I edit data values?
4.5.7 How do I edit the Z and Z’ values?
4.5.8 How do I recalculate the formula using the CCDC Diagram Editor?
4.5.9 How do I fix truncated identifier warnings?
4.5.10 Why are some of my entries marked as existing and skipped when they are unique?
4.5.11 How can I tell where the CSD-Editor is saving edits to my database?
4.5.12 Why do I see an ‘Operation Failed’ message when trying to open my database in CSD-Editor?
4.5.13 Why does my database entry not contain all the information I expected?
4.5.14 What should I do if I get an ‘internal error’ warning?
4.5.15 How do I add a pi-bond to a chemical diagram using ChemDraw?
4.5.16 What is the ‘-journal’ file that is sometimes created alongside my database files?
4.5.17 What should I do if the DeCIFer output gives the warning

5 Appendix
5.1 Automate the Addition of new CIFs to an existing database
5.2 CSD-Editor errors and warnings
5.2.1 No formula given - no comparison with connectivity
5.2.2 Given formula does not agree with chemical connectivity
- Abstracted: A; Calculated: B
5.2.3 formula has unbalanced charges
5.2.4 2D: A x B valency C atoms D
5.2.5 diagram atom A has both aromatic and delocalised bonds
5.2.6 B aromatic atoms(s) that are in rings with poor geometry
5.2.7 no compound name
5.2.8 compound name: mismatched parentheses A
5.2.9 compound name: letter followed by number
5.2.10 atom A has unusual valency B
5.2.11 crystal connectivity not polymeric but formula polymeric
5.2.12 crystal connectivity polymeric but formula not polymeric
5.2.13 crystal connectivity bond: A B C short
5.2.14 crystal connectivity: long bonds may have been added manually
5.2.15 crystal structure: A % voids
5.2.16 Dcalc is too small, no density check
5.2.17 Dcalc differs from Dx by >2%, try adjusting Z value
5.2.18 no R factor for entry category=3
5.2.19 formula: estimated vol A calc vol B differs by C %; check Z
5.2.20 calc fw A differs from publ fw B by C (D %)
5.2.21 Element A in formula but not in compound name
5.2.22 Element A in compound name but not in formula
5.2.23 compound name: poly or catena needed for polymeric formula
5.2.24 Chem & Crystal connectivities not matched
5.2.25 Crystal fully matches Chem connectivities, but not vice-versa
5.2.26 Chem fully matches Crystal connectivities, but not vice-versa
5.2.27 Chem & Crystal matching impossible
5.2.28 published bond: A B given but not found in calc
5.2.29 published bond: A B short
5.2.30 published bond: A B long
5.2.31 spacegroup does not agree with symmetry operators
5.2.32 no Z value
5.3 CIF attributes read by DeCIFer
5.4 Glossary
1 Decifer and CSD Editor User Guide

2024.1 CSD Release

Copyright © 2024 Cambridge Crystallographic Data Centre

Registered Charity No 800579

1.1 Conditions of Use

The Cambridge Structural Database Portfolio (CSD Portfolio) including, but not limited to, the following: ConQuest, CSD-Editor, Decifer, Mercury, Mogul, IsoStar, DASH, CSD Conformer Generator, Hermes, GOLD, SuperStar, the CSD Python API, web accessible CSD tools and services, WebCSD, CSD sketchers, CSD data files, CSD data updates, the CSD database, sub-files derived from the foregoing data files, documentation and command procedures, test versions of any existing or new program, code, tool, data files, sub-files, documentation or command procedures which may be available from time to time (each individually a Component) encompasses database and copyright works belonging to the Cambridge Crystallographic Data Centre (CCDC) and its licensors and all rights are protected.

Any use of a Component of the CSD Portfolio, is permitted solely in accordance with a valid Licence of Access Agreement or Products Licence and Support Agreement and all Components included are proprietary. When a Component is supplied independently of the CSD Portfolio its use is subject to the conditions of the separate licence. All persons accessing the CSD Portfolio or its Components should make themselves aware of the conditions contained in the Licence of Access Agreement or Products Licence and Support Agreement or the relevant licence.
In particular:

- The CSD Portfolio and its Components are licensed subject to a time limit for use by a specified organisation at a specified location.

- The CSD Portfolio and its Components are to be treated as confidential and may NOT be disclosed or re-distributed in any form, in whole or in part, to any third party.

- Software or data derived from or developed using the CSD Portfolio may not be distributed without prior written approval of the CCDC. Such prior approval is also needed for joint projects between academic and for-profit organisations involving use of the CSD Portfolio.

- The CSD Portfolio and its Components may be used for scientific research, including the design of novel compounds. Results may be published in the scientific literature, but each such publication must include an appropriate citation as indicated in the Schedule to the Licence of Access Agreement or Products Licence and Support Agreement and on the CCDC website.

- No representations, warranties, or liabilities are expressed or implied in the supply of the CSD Portfolio or its Components by CCDC, its servants or agents, except where such exclusion or limitation is prohibited, void or unenforceable under governing law.

Licences may be obtained from:

Cambridge Crystallographic Data Centre 12 Union Road Cambridge CB2 1EZ, United Kingdom

Web: http://www.ccdc.cam.ac.uk

Telephone: +44-1223-336408

Email: admin@ccdc.cam.ac.uk
2 Introduction

- DeCIFer and CSD-Editor enable users to build their own private 'in-house' databases of structures from crystallographic data that are then searchable either independently of, or in unison with, the Cambridge Structural Database (CSD).

- DeCIFer is a data conversion program provided as part of the Cambridge Structural Database Portfolio. The primary objective of the program is to create high quality structural data files in a format searchable by ConQuest.

- CSD-Editor is a database editing program that allows users to manually modify entries from their own private databases to maintain consistent and up-to-date information.

2.1 Installation notes

DeCIFer requires:

- a pre-installed current version of the CSD
- a pre-installation of Python (version 3.9)
  - The CSD Portfolio installation includes a version of Python that is suitable.


2.1.1 Other requirements to use DeCIFer at the command line

You will need Python 3.9 installed and available - the CSD Python API installed with the CSD Portfolio software can be used for this purpose, or Python 3.9 can be downloaded from https://www.python.org/downloads/ if you do not have it installed already.
The Python 3.9 installation must be the default Python and the PATH environment variable must be configured to allow it to be run from anywhere.

### 2.2 Licensing DeClIFer

You will need to activate your DeClIFer installation in order to run it. If you have not already done so for other CCDC software, use the CCDC Activation tool and select either online or offline activation, or alternatively connect to a local licence server. More details can be found via the Help option in the activation tool. The CCDC Software Activation tools will be located in the `<INSTALLDIR>\ccdc-utilities\software-activation` folder of your DeClIFer install. To start the graphical user interface, run `ccdc_activator_gui.exe`.

For full details on licensing please consult the "Licensing Overview" section of the CSD Release and Installation Notes.

### 3 DeClIFer

DeClIFer has been developed by the CCDC to facilitate the creation of in-house databases of crystal structures from experimental data (.cif files). DeClIFer can be used directly, from a Command Prompt, or via the associated database editing software known as CSD-Editor.

#### 3.1 Updating an existing in-house database

CSD-Editor uses the `csdsqlx` file format, different to the format used by other CSD Portfolio software such as ConQuest. The flow chart shown in Figure 1 outlines the workflow for the production of an in-house database and the relationships between data formats.
Any legacy in-house database used by PreQuest will be in ASER format and consist of at least three files (.ind, .msk and .tcd). It is recommended the initial conversion of this database to the new database format is carried out using the chemical diagram standardiser tool.

### 3.2 Standardising bonds in a legacy database

In order to gain the maximum benefit from searching your in-house database alongside the CSD it is important that the chemistry of the structures is represented in a consistent manner. If you have existing in-house database structures containing aromatic moieties, these may be represented with either Kekule alternating single-double bonds or with aromatic bond types. To ensure the CSD bond representation is used throughout you can convert the bond types in your existing in-house database (.ind file) using the standardiser tool.

**Windows** (assuming DeCIFer has been installed in the default location):

1. Open a Command Prompt
2. Navigate to the directory in which the database is located
3. Type "C:\Users\USERNAME\CCDC\ccdc-software\csd-editor\chemical_diagram_standardiser.bat" --database [filepath to in-house database .ind file]
4. The generated database can be edited using the CSD-Editor (see Section 3 CSD-Editor)

**Linux** (assuming DeCIFer has been installed in the default location):

1. Open a command line
2. Navigate to the directory in which the database is located
3. Type /home/USERNAME/CCDC/ccdc-software/csd-editor/bin/chemical_diagram_standardiser.sh --database [filepath to in-house database .ind file]
4. The generated database can be edited using the CSD-Editor (see Section 3 CSD-Editor)

This will produce a single file, standard.csdsqlx, which can be edited with CSD-Editor (see Section 3 CSD-Editor). It is envisaged that this file is the ‘master’ database to which new structures are added and from which derivative database formats (for example for use in ConQuest and Mogul) will be produced.

Figure 1: Workflow showing the steps required to produce an in-house database

3.3 Adding new structures to/creating a new in-house database

3.3.1 Adding additional information to the CIF before conversion

DeCIFer will automatically create a database from a folder of CIF files. By default DeCIFer uses the CIF datablock name (beginning data_) to create the database entry identifier. The identifier will also be truncated to a maximum of 8 characters, with all letters capitalized e.g. data_Cmpd1 will be given the database identifier CMPD1. Two entries with datablock identifiers data_ciftitle1 and data_CIFtitle2 would both by default be given the database identifier CIFTITLE, and as such the second entry would not be automatically added to a database, as identical database identifiers are forbidden. To specify the database identifier used by DeCIFer an additional custom CSD data field should be added to each datablock of the CIF database_code_CSD
The same character limit will apply to this identifier.

DeCIFer will automatically include experimental data from standard CIF fields (e.g. crystal colour and habit); for additional information DeCIFer will also pick up data from the custom CSD data field

_ccdc_publ_extra_info

Data from this section is added to the CSD-Editor ‘Remarks’ field.

### 3.3.2 Creating a new database from a set of CIFs

#### 3.3.2.1 Creating a new database with CSD-Editor

It is possible to create a database by uploading CIFs directly into CSD Editor. However, if many CIF files are to be added simultaneously it is suggested that DeCIFer is used via the Command Prompt, as this provides more information in output files and also provides a number of different conversion options (see sections below).

To create a database from CSD-Editor:

1. Upload any entries you already have in CSD-Editor using the **File > Upload** menu option or the icon in the toolbar. If there are existing entries in CSD-Editor you will only be given the options of either adding new entries to that existing database (see [Adding to an existing database with CSD-Editor](#)) or cancelling the import.

2. From CSD-Editor navigate to **File > Open...** or press the ‘Open a new or existing session’ icon
3. From the Open CSD-Editor Session pop-up, navigate to the folder of CIFs (to see the CIF files it may be necessary to change the File type dropdown to **CIF file (*.cif)**. The other option is the **CSDSQLX Database file (*.csdsqlx)** which refers to an existing database).

4. Highlight the CIFs to convert (multiple selections can be made by holding either the 'Shift' or 'Ctrl' key whilst picking files) and click the **Open** button.

5. CSD-Editor will then open a **Master Database** pop-up to prompt for the name of the new database. By default this will be in the same folder location as the CIF files, however a new folder location can be chosen. Enter the name of the database (for example **CSD_database_1**) into the **File name** box and click **Save**.
A dialogue will then be shown while the CIFs are processed. This stage may take some time and so please leave this dialogue open until the process has finished.

If any of the new structures have the same identifier as an existing entry in the chosen database then a dialogue will be displayed. Click **Yes** to replace the entry in the database with the new entry and click **No** to keep the existing entry.

The structures will then be shown in the CSD-Editor for review. The database will have been created in the specified folder location.

### 3.3.2.2 Creating a new database with DeCIFer (Command Prompt)

**Windows** (assuming DeCIFer has been installed in the default location):

1. Add the input CIFs to a folder
2. Open a Command Prompt

3. Navigate to the directory in which the input CIFs are stored

4. Type "C:\Program Files\CCDC\decifer2021.1.0\decifer.bat" [filepath to CIF folder] > output1.out 2>&1

5. The generated database can be edited using the CSD-Editor (see CSD-Editor)

**Linux** (assuming DeCIFer has been installed in the default location):

1. Add the input CIFs to a folder

2. Open a command line

3. Navigate to the directory in which the input CIFs are stored

4. Type /home/USERNAME/decifer_2021.1.0/bin/decifer.sh [filepath to CIF folder] > output1.out 2>&1

5. The generated database can be edited using the CSD-Editor (see CSD-Editor)

### 3.3.3 Adding new CIFs to an existing database

#### 3.3.3.1 Adding to an existing database with CSD-Editor

It is quick and easy to add new entries to an existing .csdsqlx database using CSD-Editor.

To add to an existing database from CSD-Editor:

1. From CSD-Editor navigate to **File > Open...** or press the ‘Open a new or existing session’ icon (note: this can be done with an existing database already open in the Editor, or from a new session when the CSD-Editor windows will be empty).
2. From the Open CSD-Editor Session pop-up, navigate to the folder of CIFs (to see the CIF files it may be necessary to change the File type dropdown to CIF file (*.cif)). The other option is CSDSQLX Database file (*.csdsqlx) which refers to an existing database).

3. Highlight the CIFs to convert (multiple selections can be made by holding either the 'Shift' or 'Ctrl' key whilst picking files) and click the Open button

4. There are two possible alternatives at this point:

   ◦ If starting from a new session: CSD-Editor will then open a Master Database pop-up to prompt for the name of the database. Navigate to the folder location of the existing .csdsqlx database and select this. The new entries will be added.
If an existing database is already open in CSD-Editor: CSD Editor will ask for confirmation to add the CIFs to the existing database. Selecting Yes will add the CIFs to the database and the structures will then be shown alongside the entries open in CSD-Editor for review. Choosing No at this point will cancel the conversion of the new CIF files.

3.3.3.2 Adding to an existing database with DeCIFer (Command Prompt)

You can add new structures to a database but it is always recommended you make a copy of your database first, append the new entries to this database and then when you are happy replace the existing database.

Windows (assuming DeCIFer has been installed in the default location):

1. Add the input CIFs to a folder
2. Open a command line
3. Navigate to the directory in which the input CIFs are stored
4. Type "C:\Users\USERNAME\CCDC\ccdc-software\csd-editor\decifer.bat" [filepath to CIF folder] --target-database [filepath to existing .csdsqlx database] > output1.out 2>&1
5. The generated database can be edited using the CSD-Editor (see CSD-Editor)

Linux (assuming DeCIFer has been installed in the default location):

1. Add the input CIFs to a folder
2. Open a command line
3. Navigate to the directory in which the input CIFs are stored
4. Type /home/USERNAME/CCDC/ccdc-software/csde-editor/bin/decifer.sh [filepath to CIF folder] --target-database [filepath to existing .csdsqlx database] > output1.out 2>&1
5. The generated database can be edited using the CSD-Editor (see CSD-Editor).

It is possible to automate this process by creating a cronjob. See the appendix.

### 3.3.4 Using existing chemical diagrams in the conversion

This option is not available for CIFs added directly using CSD-Editor. Should you wish to use existing chemical diagrams as part of the database you can add diagram files in mol file format to the same directory as the CIFs. Each mol file should have the same name as the matching CIF.

**Windows** (assuming DeCIFER has been installed in the default location):

1. Add the input CIFs and .mol files to a folder
2. Open a command line
3. Navigate to the directory in which the input CIFs and mol files are stored
4. Type "C:\Users\USERNAME\CCDC\ccdc-software\csd-editor\decifer.bat" [filepath to CIF folder] > output1.out 2>&1
5. The generated database can be edited using the CSD-Editor (see CSD-Editor)

**Linux** (assuming DeCIFER has been installed in the default location):

1. Add the input CIFs and .mol files to a folder
2. Open a command line
3. Navigate to the directory in which the input CIFs and mol files are stored
4. Type /home/USERNAME/CCDC/ccdc-software/csdc-editor/bin/decifer.sh [filepath to CIF folder] > output1.out 2>&1
5. The generated database can be edited using the CSD-Editor (see CSD-Editor)

### 3.3.5 Adding citation information to the entries during conversion

This option is not available for CIFs added directly using CSD-Editor. You can add citation information to all the entries during a conversion by using a citation file. An example of the format of this file is available as part of the installation ([file path to directory] \csd-editor\examples\fully_populated.citx). You can add authors and publication information to this file which will then be picked up by the conversion process. The coden refers to the CCDC coden number for individual journal titles.

**Windows** (assuming DeCIFer has been installed in the default location):

1. Add the correct data to the citation file
2. Add the input CIFs and the citation file to a folder
3. Open a Command Prompt
4. Navigate to the directory in which the input CIFs are stored
5. Type
   ```
   C:\Users\USERNAME\CCDC\ccdc-software\csd-editor\decifer.bat [filepath to CIF folder] --citation [filepath to .citx file] > output1.out 2>&1
   ```
6. The generated database can be edited using the CSD-Editor (see CSD-Editor)

**Linux** (assuming DeCIFer has been installed in the default location):

1. Add the correct data to the citation file
2. Add the input CIFs and the citation file to a folder
3. Open a Command Prompt
4. Navigate to the directory in which the input CIFs are stored
5. Type `/home/USERNAME/CCDC/ccdc-software/csd-editor/bin/decifer.sh [filepath to CIF folder] --citation [filepath to .citx file] > output1.out 2>&1

6. The generated database can be edited using the CSD-Editor (see CSD-Editor)

### 3.3.6 Conversion to current CSDS database format

Before your CSD-Editor .csdsqlx in-house database can be used by CSDS software it must be converted into the current CSDS database format. These files can then be used to create in-house databases in ConQuest and Mogul.

**Windows** (assuming DeCIFer has been installed in the default location):

1. Open a command line
2. Navigate to the csd-editor folder (typically C:\Users\USERNAME\CCDC\ccdc-software\csd-editor)
3. Type `ccdc_babel -csdsqlx <path_to_input_database>\<name_of_database.csdsqlx> –csdsql3 <path_to_output_database>\<name_of_database.csdsql>`

**Linux** (assuming DeCIFer has been installed in the default location):

1. Open a command line
2. Navigate to the DeCIFer folder
3. Type `/home/USERNAME/CCDC/ccdc-software/csd-editor/bin/ ccdc_babel -csdsqlx <path_to_input_database>/ <name_of_database.csdsqlx> –csdsql3 <path_to_output_database>/ <name_of_database.csdsql>

### 3.3.6.1 Combining databases

The `ccdc_babel` program used for database conversion can also be used to append pre-existing databases together. The following command will add the entries from `new_database.csdsqlx` to the file `existing_database.csdsqlx`
**Windows** (assuming DeCIFer has been installed in the default location):

1. Open a command line

2. Navigate to the csd-editor folder (typically C:\Users\USERNAME\CCDC\ccdc-software\csd-editor)

3. Type `ccdc_babel -auto <path_to_new_database.csdsqlx>/<name_of_new_database.csdsqlx> –csdsqlx <path_to_existing_database.csdsqlx>/<name_of_existing_database.csdsqlx>`

**Linux** (assuming DeCIFer has been installed in the default location):

1. Open a command line

2. Navigate to the DeCIFer folder

3. Type `/home/USERNAME/CCDC/ccdc-software/csd-editor/bin/ccdc_babel -auto <path_to_new_database.csdsqlx> <name_of_new_database.csdsqlx> –csdsqlx <path_to_existing_database.csdsqlx> <name_of_existing_database.csdsqlx>`

This will result in the entries from first specified database being added to the second. Any entries which have identical refcodes between the two databases will not be appended.

### 3.4 Using in-house databases with CSD Portfolio software

The `.csdsqlx` database produced by DeCIFer is required only for maintaining and updating the database. For all other uses of an in-house database with CSD Portfolio software it is strongly recommended to use the `.csdsql` database, as this provides faster search results for some features.
3.4.1 Activating your in-house database for use with ConQuest

ConQuest can search in-house databases separately or in addition to the main CSD database. After you create or update an in-house database that you wish to search using ConQuest it must first be activated using the procedure described below.

In order to activate your in-house database you must copy (or soft-link: Linux only) the in-house database file (.csdsql) to the same directory as the CSD V5.45 database files. On Linux this is typically:

```
<INSTALLDIR>/CCDC/ccdc-data/csd
```

While on Windows, this is normally:

```
C:\Users\USERNAME\CCDC\ccdc-data\csd
```

You must then run the Activate program.

For Linux, to run the activate database program type:

```
<INSTALLDIR>/ccdc-software/conquest/bin/activate
```

To run the activate database program on Windows select: **Programs -> CCDC -> Activate in-house database** from the Windows Start menu.

Note: Windows users may require escalated administrator privileges to activate databases. To do so, right-click on the **Activate Inhouse Databases** menu item and select **Run as Administrator**.

In all cases a pop-up window like the one shown below will be displayed:
The activation program attempts to find the directory where the database is located by examining your ConQuest defaults file. However, if the correct directory cannot be found then use the Change button to locate the required directory.

All databases in the selected directory which have not been activated will be displayed in the Database to activate pull-down menu. Simply select the databases you wish to activate.

The contents of the Database Name dialogue box will be used by ConQuest to identify which databases you wish to search or view. Choose a name relevant to the database contents.

Enter a version number for the database in the Database Version dialogue box. If you have an earlier version of a database, with the same Database Name in the same directory, it is important to make sure that the most recent version has the largest version number so that it is used in preference to the others.

Enter the date for the database in the Database Date dialogue box.

Use Search/Display Order pull-down menu to select the order in which activated databases will be displayed and searched by ConQuest. Taking the above example, if you select After CSD v5.43, when both databases are selected for searching in ConQuest the main CSD database will be searched before the in-house database.
When you are happy with your selections press the **Activate** button; this will create an `.inf` file for the database, which will then be viewable and searchable the next time you start ConQuest.

To close the activate database program press the **Exit** button.

In ConQuest, navigate to **View Databases > Available Databases...**

This opens an **Available Databases** pane which allows you to add a directory containing your new database, and also manage which databases are searched.

### 3.4.2 Generating your in-house database for use with Mogul

The `mogulbuilder` tool is available as a utility provided in the RP API Add-on to the CSD Python API which can be downloaded from the CCDC User Registration Portal ([https://downloads.ccdc.cam.ac.uk/user_portal/](https://downloads.ccdc.cam.ac.uk/user_portal/)).

To run the `mogulbuilder.py` script (found in utilities > mogulbuilder) you will need to specify the path to the ASER database from which you wish to generate the Mogul library, an output directory and a name for the data library. The latter is required for linking back to the underlying structures in the Mogul GUI.

#### 3.4.2.1 Windows

1. Open a Command Prompt
2. Navigate to the download location of your `mogulbuilder` script (note the RP API Add-on must be unpacked (unzipped) after downloading from the User Portal)
3. Type `mogulbuilder.py [filepath to database in current CSDS format (.csdsql file)] [path to output directory] [name for the data library]`

The `mogulbuilder` tool generates a large number of files and Mogul data libraries; these should be created in a new, separate folder to the original database to avoid potential conflicts with pre-existing files.
The name for the data library can be any alphanumeric name.

3.4.2.2 Linux

1. Open a command line

2. Navigate to the download location of your mogulbuilder script (note the RP API Add-on must be unpacked (unzipped) after downloading from the User Portal)

3. Type `/home/<mogulbuilder file location>/mogulbuilder.py`
   `[filepath to database in current CSDS format (.csdsql file)]`
   `[path to output directory] [name for the data library]`

The mogulbuilder tool generates a large number of files and Mogul data libraries; these should be created in a new, separate folder to the original database to avoid potential conflicts with pre-existing files.

The name for the data library can be any alphanumeric name.

The building of a Mogul data library requires a set of configuration files. These configuration files contain the definitions of the keys used in the Mogul data library. The recommended practice is to make use of the configuration files provided with the CSDS Mogul data library. On Windows mogulbuilder will try to identify the location of these configuration files using the CSDS information stored in the registry. However, if mogulbuilder fails to automatically detect the location of these files then you can specify the location explicitly.

    mogulbuilder.py –m path_to_csds_mogul_data aser_database output_directory name

For more information on how to use mogulbuilder please have a look at the help text available by making use of the -h option.

    mogulbuilder.py –h

Once you have built your Mogul data you then need to configure Mogul to use the new data alongside the CSD Mogul data. To specify the databases to use when running a Mogul analysis open the Mogul Data Library Settings dialogue. This can be found under the Databases menu.
In the dialogue use the Add button to add another Mogul data library. The CSD Mogul data should already be listed (and cannot be deleted).

Note that you can use the check boxes to select which data libraries you want to use in your Mogul analysis. You must always have at least one database selected.

### 3.4.3 Generating your in-house database for use with WebCSD

Your WebCSD server will use the .csdsq1 database. Please refer to your WebCSD User Instructions document for more information on configuration options.
4 CSD-Editor

4.1 Opening the CSD-Editor

Windows (assuming DeCIFer has been installed in the default location):

1. Navigate to the csd-editor folder (typically C:\Users\USERNAME\CCDC\ccdc-software\csd-editor)
2. Open the csd_editor.exe file

Linux (assuming DeCIFer has been installed in the default location):

1. Open a command line
2. Type /home/USERNAME/CCDC/ccdc-software/csd-editor/bin/csd_editor

4.2 Opening entries in the CSD-Editor

Open the CSD-Editor and click the File menu button and select the Open option.

This will open the Open CSD-Editor Session dialogue; from this dialogue check that the CSDSQLX Database File option is selected in the dropdown box in the bottom right hand corner of the
dialogue. Then locate the .cslsq database that was created by DeCIFer/CSD-Editor and add it to the File name box and click Open.

This will open the dialogue named Crystal Structure Selection. The selected database will be listed in the left hand Available Structures column. Click on the database that you want to load entries from and this will expand to show all the entries in this database. Select entries from this list and click the Add Selected Items button to move the structures into the Selected Structures column. Once all the entries that you want to open are in the Selected Structures column you should click OK to open the entries in the CSD-Editor.

If you have a refcode list (a .gcd file) of the entries that you want to open from your selected database you can click the Open identifier list file button, select the refcode list and click Open. You can
create a refcode list using ConQuest or you can manually create one from the CSV file which is created when you run DeClIFer on a set of CIFs. This can be done by selecting the list of Structure names and saving them in a separate text file with a .gcd file extension.

If you want to open a different database then click **Cancel** and use the **Open** option in the **File** menu.

The entries that you have selected will be saved in a local temporary database (usually in \C:\Users\USERNAME\csd_editor). When you edit these structures the edits will be saved in this temporary database every time you move between entries or close the CSD-Editor. Since the edits are saved you can close the CSD-Editor and reopen it which will reopen the edited structures. When you click **Save All** the edits will be saved back into the database from which you selected the structures initially.
4.3 Editing entries in the CSD-Editor

The CSD-Editor has six different tabs. The functionality of these tabs will be described over the next few pages.

4.3.1 Overview Tab

1. **2D chemical diagram.**

2. **3D crystal structure.** Generated from the atomic coordinates.

3. **Crystallographic, numeric and textual data items.** Data items can be edited by double clicking in the relevant box and tab can be used to exit the editing. Some of the boxes are greyed out to indicate that they are not editable.

4. **Top menu bars.** The menu items **File, Edit, View, Insert, Action, Calculate** and **Help** are available. These menus contain a number of different options but all of these options can also be accessed through toolbars and buttons on the interface. Note that there are no options available in the **Insert** menu.

5. **Save All button.** Saves all the structures back into the database from which they were downloaded.
6. **Save the Current Entry button.** Saves just the current structure that is displayed back into the database from which it was downloaded.

7. **Undo and redo buttons.** Allows the last command(s) to be undone or redone.

8. **Undo All button.** Allows all the changes since the entry was loaded (or last saved) to be undone. If you move onto a different entry then you can no longer undo the changes.

9. **ChemDraw Chemical Diagram button.** Enables the 2D chemical diagram to be edited using the ChemDraw program (if this is installed on the computer). The location of the installation of ChemDraw can be stored by clicking on the Help menu item and selecting *Session Info*. The path for ChemDraw can then be added to the *External Diagram Editor* field.

10. **Generate Chemical Diagram from current crystal structure button.** Generates a 2D chemical diagram from the 3D crystal structure. Note that the diagram generated each time the button is used may be different.

11. **DOI button.** If a DOI is included as part of the *Literature Reference* field, clicking this button opens the relevant paper in a browser if access to the article is available.

12. **Generate Formula from Chemical Diagram button.** Opens a dialogue which allows the formula to be calculated. If there are multiple molecules in the chemical diagram then the multipliers for each molecule can be defined.

13. **Generate a Packing Summary for the Crystal Structure.** Displays the Cell contents, the Formula unit and the Z and Z' values for the crystal structure.

14. **Delete Entry button.** Deletes the currently selected entry. The entry can be undeleted by clicking this button again. The deletion is finalised once the *Save All* button is clicked.

15. **Arrows to move between structures.** The left hand button moves to the first entry in the list, the central button to the previous entry and the right hand button to the next entry.
16. **Compare Entry button.** Used in cases where an entry is available to compare. This functionality is unlikely to be useful in most common situations.

17. **List of entries.** The software determines the relevant menu depending on the current entry being viewed. In most cases the display gives a list of all the entries in the current editing session colour coded by matching status. Green indicates that all of the chemical moieties in the 2D diagram match perfectly on to the moieties in the 3D structure. Amber indicates that only some of the chemical moieties in the 2D diagram match moieties in the 3D structure, likely to be due to disordered or unmodeled molecules in the 3D structure. A purple colour indicates moieties present in the 3D structure have not been represented in the 2D diagram, suggesting the 2D diagram may not represent all moieties modelled from the X-ray data. Red indicates that none of the moieties in the 2D diagram match the moieties in the 3D structure. Black indicates that a match is impossible because of the absence of a 2D diagram.

18. **Current entry identifier.** The identifier for the structure currently displayed.

19. **Filter Structure Menu.** A drop-down menu which can be used to filter the list of entries. The filter options allow display of either all the structures in the current editing session, just the **Not matched** (red) entries or just the **Partially matched** (amber) entries.

20. **Arrows to move between structures.** Moves between structures in the displayed list.

21. **Details of Errors, Warnings and Remarks.** This section gives a list of the errors, warnings and remarks related to the structure. Some of the messages are from the DeCIFer analysis and some are calculated from the current structure. A list of the common errors and warnings is given in the Appendix.

22. **Matching Status.** Gives details of the matching status of the entry.
4.3.2 Compare tab

This tab is unlikely to be useful in most common situations.

4.3.3 All text tab

1. **All Text tab.** This is used to view all the text items in the entry including some text items that are not given in the **Overview** tab.

2. **Crystallographic, numeric and textual data items.**

3. **Non-editable data.** Some of the boxes are greyed out to indicate that they are not editable.
1. **Diagram tab.**

2. **Right-click menu.** The diagram can be copied and pasted between entries by right-clicking on the diagram and using the **Copy** and **Paste** options. It can also be used to **Print** the diagram.
4.3.5 Visualiser tab

A lot of the functionality is identical to the functionality in Mercury. For a more detailed description of the functionality in the Visualiser tab see the Mercury documentation, [https://www.ccdc.cam.ac.uk/support-and-resources/documentation-and-resources/](https://www.ccdc.cam.ac.uk/support-and-resources/documentation-and-resources/).

1. **Visualiser tab.**

2. **Display Style.** Select the display style for the 3D structure from the options in the drop-down menu.

3. **Colour Mode.** Select the colouring scheme for the atoms and bonds in the 3D structure. A particularly useful function in some cases is the ability to colour the 3D structure by matching status so that it is possible to see which molecules in the structure are matched to the 2D (green) and which are unmatched (red). Another useful function is the **by Element or Suppression** option which indicates which atoms have been suppressed in the 3D structure.

4. **Picking Mode.** Select the picking mode from options in the dropdown box. This defines what happens when you left-click on items in the 3D structure; whether it selects the atoms or measures a distance, angle, etc.
5. **Normal and lasso picking modes.** Click the icons to change the picking mode from the normal picking mode to the lasso picking mode.

6. **Show Labels options.** Click the **Show Labels** check box to show/hide atom labels. Select options from the two drop-down menus to select the type of atom that is labelled and the type of label that is shown. A useful function is the **Show Labels for Siteless Neighbours** option which will display the labels for only atoms which have siteless atoms (e.g. atoms with hydrogens that have no coordinates).

7. **Display Options.** Click the relevant check box to display certain options in the 3D structure.

8. **More Information button.** Opens the **More Information** dialogue. From this dialogue, it is possible to suppress, unsuppress, set the hydrogen count for or delete atom(s) using a mouse click for a single row, Shift+click for a block and Ctrl+click for separated rows. This dialogue is customisable and can show different attributes about the atoms, including label, occupancy and ADP.

9. **Force Match button.** Used in situations where edits have been made to either the 2D diagram or the 3D structure and there is no longer a perfect or partial match between the two. In such a situation, if the chemistry represented in the 2D diagram is close enough to correctly matching the 3D structure, then the bond types/hydrogen atom assignments etc. in the 3D will automatically be edited to achieve as complete a match as possible.

10. **Suppress button.** Used to suppress atoms in the 3D structures, to help represent disordered molecules. Select all the atoms that you want to suppress and then click the **Suppress** button. The suppressed atoms will be coloured light green and will not bond to any other atoms. See the FAQ ‘**How should I deal with disordered structures?**’ for more information.

11. **Unsuppress button.** Used to unsuppress atoms in the 3D structures. Select all the atoms that you want to unsuppress and then click the **Unsuppress** button.
12. **Apply Auto Edit button.** Useful if DeCIFer has been unable to determine all the chemistry in the structure correctly (often shown by the presence of unknown bond types in the 3D structure). The **Apply Auto Edit** button is a quick one-click option that uses the default settings suitable for most situations. To manually configure the Auto Edit options the **Auto Edit** button should be used instead.

13. **Auto Edit button.** Useful if DeCIFer has been unable to determine all the chemistry in the structure correctly (often shown by the presence of unknown bond types in the 3D structure). The default settings are suitable for most situations and clicking **Apply** produces another pop-up window to show how many (if any) bonds and atoms have been changed, added or removed and presents a choice to **Accept** or **Revert** the changes.

14. **Radii button.** Used to edit the default bonding limits for pairs of elements in the structure and can be a useful way of removing unwanted bonds from the 3D structure. A tick box is provided to either include or exclude chemically-unlikely bonds. For most common cases it is better for this box to be unchecked.

15. **Manual Edit button.** This opens the dialogue shown in the 3D structure and allows editing of the 3D structure including adding atoms and bonds, changing element types and editing atom and bond properties.

16. **Set Bond Type to button.** Used to change the bond type of bonds in the 3D structure.

17. **Suppress Atoms and Bonds button.** Used to suppress atoms and bonds in the 3D structure (the **Unsuppress atoms** button below will unsuppress atoms). This functionality can be useful if you have disorder that has not been correctly resolved.

18. **Make 2D button.** Generates a 2D chemical diagram from the 3D crystal structure.
19. **Projection button.** This produces a 2D diagram in the same orientation as the currently selected view of the 3D structure. This diagram is generally of lower quality than the diagram produced by the **Make 2D** button but it is much faster.

20. **Edit Diagram button.** Enables the 2D chemical diagram to be edited using the CCDC Diagram Editor.

21. **Spacegroup button.** Allows alternative settings of the space group to be chosen.

22. **Packing button.** Allows various options for packing and slicing the structure.

23. **Voids button.** Calculates the void volume as a percentage of the unit cell volume.

24. **Structures button.** Used when there are two 3D structures in the display to colour the structures differently and to move them apart. This will not be useful in most common situations.

### 4.3.6 All Hits tab

This tab is used to compare multiple structures but it is unlikely to be useful in most common situations.

### 4.4 Saving entries in the CSD-Editor

During your editing session entries are automatically saved to a local temporary database after moving between entries and the CSD-Editor can be closed and opened without any edits being lost.

Once you have finished editing entries in your session you should click the **Save All** button which will save your edits back into the database that you downloaded the structures from.
4.5 CSD-Editor FAQs

4.5.1 How should I deal with polymeric structures?

The easiest way to deal with polymeric structures is to reduce the relevant bond distance pairs to the metal until the structure is not polymeric. This can be done by going to the Visualiser tab, pressing the Radii button and reducing the relevant bond distance limits for all bonds which occur to the metal in the structure. Once all the polymeric bonds have gone the Make 2D button can be used to generate a new diagram. In some cases it may be necessary to use the Force Match button to match the structure.

4.5.2 How should I deal with disordered structures?

If a structure contains disorder then the minor occupancy sites must be suppressed to leave a single representative atomic position that will then be matched against the chemical diagram. Suppressed atoms are not deleted; they simply take no further part in the establishment of the crystallographic connectivity and therefore will not be used in ConQuest 3D searches. In the majority of cases, DeCIFer will have dealt with any disorder in the structure. If there are atoms in the 3D structure that have not been dealt with correctly then these atoms can be suppressed and unsuppressed manually using either the Manual Edit dialogue or the More Info dialogue in the Visualiser pane. It is advisable to check that in the top menu, colour is set to by Element or Suppression.

4.5.3 How do I select all atoms in a molecule in the 3D structure?

This can be done by holding the Shift key and clicking an atom in the molecule that you want to select.

4.5.4 How do I select a group of atoms in the 3D structure?

This can be done by holding the Shift+Alt keys and using the mouse to lasso atoms.
4.5.5 What does the colour coding in the identifier list signify?

Green indicates that all of the chemical moieties in the 2D diagram match perfectly onto the moieties in the 3D structure. Amber indicates that only some of the chemical moieties in the 2D diagram match moieties in the 3D structure, likely to be due to disordered or unmodeled molecules in the 3D structure. A purple colour indicates moieties present in the 3D structure have not been represented in the 2D diagram, suggesting the 2D diagram may not represent all moieties modelled from the X-ray data. Red indicates that none of the moieties in the 2D diagram match the moieties in the 3D structure. Black indicates that a match is impossible because of the absence of a 2D diagram.

4.5.6 How do I edit data values?

Data values can be edited by double-clicking in the box. Some of the boxes enable free-format text editing whereas others provide either a pop-up box or drop-down menu with suggested text. For free-format textboxes, Tab can used to exit the textbox.

4.5.7 How do I edit the Z and Z' values?

To edit the Z value you should double click in the Z, Z' field and add the correct Z value; the Z' value will be calculated automatically when you tab out of the field.

4.5.8 How do I recalculate the formula using the CCDC Diagram Editor?

This can be done as follows:

- Open the CCDC Diagram Editor by clicking the Edit chemical diagram for incoming entry button
- Click the TO-SCIED button at the top of the Editor page
- Click GEN-HYDS
· Click **GENERATE-FORMULA** and press Return (or enter the number) for each moiety in the 2D diagram

· Click **RETURN-AND-SAVE** to return to save the formula and return to the CSD-Editor.

### 4.5.9 How do I fix truncated identifier warnings?

A warning about a truncated identifier means an identifier was processed which exceeded the length limit. The normal limit is 8 characters. Ensure identifiers do not exceed this length.

### 4.5.10 Why are some of my entries marked as existing and skipped when they are unique?

The IDs of the entries may exceed the length limit. This will result in truncation to 8 characters. Truncated IDs may no longer be unique, resulting in duplicates that will not be added to the output database.

For more information on a database created or updated via CIFs being added directly through the CSD-Editor it is possible to output a log file that will document the database creation process. This requires the addition of a **-logfile** argument when opening CSD-Editor e.g. for a typical Windows installation typing "C:\Users\USERNAME\CCDC\ccdc-software\csd-editor\csd_editor.exe" -logfile [location and name of log file] will enable logging as part of the database creation. (e.g. typing "C:\Users\USERNAME\CCDC\ccdc-software\csd-editor\csd_editor.exe" -logfile C:\Users\USERNAME\csd_editor\deciferlog1.txt will save a log file called deciferlog1.txt in the C:\Users\USERNAME\csd_editor\ folder.)

### 4.5.11 How can I tell where the CSD-Editor is saving edits to my database?

Any additional structures or updates of existing database entries are only added to your existing database once the CSD-Editor session has been saved, when the changes are applied to the
original source database. CSD-Editor can be closed and reopened without uploading database entries; in this case edits made within CSD-Editor will be saved but not applied to your source database.

The location of the source database can be found from the Help menu under the Session Info option. This will record your database location as the Data Source, with the Session File and Database File locations referring to the files updated during the CSD-Editor session.

4.5.12 Why do I see an ‘Operation Failed’ message when trying to open my database in CSD-Editor?

When making changes to database entries in CSD-Editor the structures are saved as a separate CSD-Editor session file. If you attempt to re-open a database as part of a pre-existing session it may cause an error message to occur (see below).

To resolve the error: First close the CSD-Editor session, and then locate the folder containing the CSD-Editor session files (typically C: \Users\USERNAME\csd_editor) and delete the files present. Once these files are removed CSD-Editor can be reopened and used as normal.
4.5.13 Why does my database entry not contain all the information I expected?

DeCIFer is designed to quickly produce a database from a set of CIF files. In rare, particularly challenging cases, DeCIFer will struggle to fully interpret the information from the CIF; DeCIFer has a preset timeout, meaning some of the curation steps, such as a diagram generation, will be skipped if too much time elapses. This can be seen in CSD-Editor where the phrase ‘DeCIFer is running in TEXT ONLY mode’ will be seen in the lower panel should a timeout have occurred. In this case it may be necessary to manually resolve disorder in the structure, and generate the 2D diagram by using one of the diagram generation options.

4.5.14 What should I do if I get an ‘internal error’ warning?

When deleting and adding entries to a database you may very occasionally get an ‘internal error’ warning message. If you get this message then it is recommended that you close and reopen the CSD-Editor.

4.5.15 How do I add a pi-bond to a chemical diagram using ChemDraw?

Open ChemDraw and draw the atoms for the structure.

Select the Multiple bonds option and change all the pi-bonded bonds to the single and dotted bond (tautomeric bond).

Select all the atoms that are to be pi-bonded to the metal and select the Structure-’Add Multi-Center Attachment’ option. A multi-center attachment point will appear in the diagram. Then select the dotted bond option and draw a bond between the metal atom and the multi-center attachment point.
Close ChemDraw and the diagram in the CSD-Editor will have the pi-bonds displayed.

4.5.16 What is the '-journal' file that is sometimes created alongside my database files?

This is a temporary file that appears alongside the database file and is used as part of the normal database transaction rollback mechanism. This file should not be deleted while the database is in use.

4.5.17 What should I do if the DecIFer output gives the warning

'DecIFer is running in TEXT ONLY mode' for the addition of all new structures?

You should check if the Structure Assigner Reference Database Location is set correctly. To check this you should open up the file located in your APPDATA settings which is usually located at C:\Users\USERNAME\AppData\Roaming\CCDC\Decifer.ini. The StructureAssignerReferenceDatabaseLocation should be set to your latest installation.

If there is any problem then please delete this Decifer.ini file.
5 Appendix

5.1 Automate the Addition of new CIFs to an existing database

On Linux it is possible to automate the process of adding a cif file to database, by creating a cronjob.

You may need to ask your system administrator to setup the cronjob and make sure users have the appropriate permissions to access the folder where cif files are to be added.

The steps are:

1. Create a folder hierarchy.

   `<db_dir>` - full path to base folder
   
   `<db_dir>/cifs` - cif files go here

   Create a new file `<db_dir>/update_database.sh`. Add the following contents (or something similar that fits your own requirements):

   ```sh
   #!/bin/sh
   # change the following two lines
   DB_DIR=<db_dir>
   DECIFER_DIR=<decifer_installation_dir>/csd-editor
   echo "---------------------------------" >> ${DB_DIR}/decifer.log
   2>&1
   date >> ${DB_DIR}/decifer.log 2>&1
   ${DECIFER_DIR}/bin/decifer.sh
   ${DB_DIR}/cifs
   --target-database ${DB_DIR}/organics.csdsqlx
   >> ${DB_DIR}/decifer.log 2>&1
   echo >> ${DB_DIR}/decifer.log 2>&1
   ```

2. Create a cronjob. For example to run the update every Sunday at 8:05 add the following to the cronfile.

   ```
   5 8 * * 0 DISPLAY=:99 <db_dir>/update_database.sh
   ```
Usually this should be created using `crontab -e` and then adding in the line above, however this may vary depending on your system setup.

Users add their cif files to `<db_dir>/database_update/cifs`.

Errors are reported to `<db_dir>/decifer.log` and the database is `<db_dir>/organics.csdsqlx`. We have used `organics.csdsqlx` as the name but any database name can be used.

### 5.2 CSD-Editor errors and warnings

Below is a list of the common errors and warnings that are found in the CSD-Editor along with an explanation of each. Note that A, B, C and D will be replaced by the relevant atom(s) or calculated value in the messages given in the CSD-Editor.

#### 5.2.1 No formula given - no comparison with connectivity

The formula may be missing and can be calculated by using the CCDC Diagram Editor (see CSD-Editor FAQs).

#### 5.2.2 Given formula does not agree with chemical connectivity - Abstracted: A; Calculated: B

The formula does not agree with the 2D diagram and can be recalculated by using the CCDC Diagram Editor (see CSD-Editor FAQs).

#### 5.2.3 formula has unbalanced charges

The formula does not charge balance. The charges can be edited and the formula recalculated using the CCDC Diagram Editor (see CSD-Editor FAQs).
5.2.4 2D: A x B valency C atoms D

The 2D diagram contains atoms which have unusual valencies. There may be H atoms missing from C atoms and these can be added by using the CCDC Diagram Editor (see CSD-Editor FAQs).

5.2.5 diagram atom A has both aromatic and delocalised bonds

An atom in the 2D diagram has unusual aromatic and delocalised bonds. This can be fixed by using the CCDC Diagram Editor.

5.2.6 B aromatic atoms(s) that are in rings with poor geometry

In the 2D diagram an aromatic ring has unusual geometry. This can be fixed by using the CCDC Diagram Editor.

5.2.7 no compound name

The compound name field is empty. This is only important to fix if you want to search for compound names in your database.

5.2.8 compound name: mismatched parentheses A

The compound name does not contain the same number of opening and closing brackets.

5.2.9 compound name: letter followed by number

The compound name contains a letter followed immediately by a number. This is only important to fix if you want to search for compound names in your database.

5.2.10 atom A has unusual valency B

In the 3D structure an atom has an unusual valency. This may be ok but can be changed if required by editing the 3D structure in the Visualiser tab.
5.2.11 crystal connectivity not polymeric but formula polymeric

The formula is polymeric but the 3D structure is not. It is possible that the formula should not be polymeric and can be edited by removing the n multipliers using the CCDC Diagram Editor.

5.2.12 crystal connectivity polymeric but formula not polymeric

The formula is not polymeric but the 3D structure is. It is possible that the formula should be polymeric and can be edited by adding n to the multipliers using the CCDC Diagram Editor.

5.2.13 crystal connectivity bond: A B C short

A bond in the 3D structure is short. This may be due to unresolved disorder.

5.2.14 crystal connectivity: long bonds may have been added manually

This is a warning that the 3D structure contains bonds which are longer than the default limits in the CSD-Editor. These may have been added manually and so there is no change required.

5.2.15 crystal structure: A % voids

This is a warning that the structure contains void space. This may sometimes be due to missing or unmodelled solvent molecules.

5.2.16 Dcalc is too small, no density check

The calculated density is very small. This may be due to an incorrect Z value. To change the Z value you should double click in the Z, Z’ field and add the correct Z value, the Z’ value will be calculated automatically when you tab out of the field.
5.2.17 Dcalc differs from Dx by >2%, try adjusting Z value

The calculated density differs from the density given in the CIF. This may be due to an incorrect Z value. To change the Z value you should double click in the \( Z, Z' \) field and add the correct Z value, the \( Z' \) value will be calculated automatically when you tab out of the field.

5.2.18 no R factor for entry category=3

No R-factor is given. A category 3 entry is a structure which has had a complete refinement.

5.2.19 formula: estimated vol A calc vol B differs by C %; check Z

The cell volume is estimated from the molecular formula using average effective atomic volumes derived by D. Hoffman, but based on the '17Å per non-H atom' idea and is compared with the volume calculated from the cell parameters. If there is a difference then the structure may have an incorrect Z value or may contain void space.

5.2.20 calc fw A differs from publ fw B by C (D %)

The calculated formula weight differs from the formula weight given in the CIF. This may be due to an incorrect formula which can be recalculated using the CCDC Diagram Editor (see CSD-Editor FAQs).

5.2.21 Element A in formula but not in compound name

The calculated formula contains an element that is not given in the compound name. This can be corrected by adding the element name to the compound name. This is only important to fix if you want to search for compound names in your database.
5.2.22 Element A in compound name but not in formula

The compound name contains an element that is not given in the calculated formula. This can be corrected by removing the element name from the compound name. This is only important to fix if you want to search for compound names in your database.

5.2.23 compound name: poly or catena needed for polymeric formula

The formula is polymeric but the compound name does not contain ‘catena’ or ‘poly’. To correct, add one of these names to the compound name. This is only important to fix if you want to search for compound names in your database.

5.2.24 Chem & Crystal connectivities not matched

None of the chemical moieties in the 2D diagram match with any of the moieties in the 3D structure. This may be fixed by editing the 3D structure or the 2D diagram. Using the Force Match button is a way of adding the chemistry in the 2D diagram to the 3D structure.

5.2.25 Crystal fully matches Chem connectivities, but not vice-versa

Some of the chemical moieties in the 2D diagram do not match with some of the moieties in the 3D structure. This may be fixed by editing the 3D structure or the 2D diagram. Using the Force Match button is a way of adding the chemistry in the 2D diagram to the 3D structure.

5.2.26 Chem fully matches Crystal connectivities, but not vice-versa

Some of the chemical moieties present in the 3D structure are not represented in the 2D diagram. Using the Edit > Create Chemical Diagram option will represent all 3D moieties in the 2D diagram;
tools such as the Auto Edit function from the visualiser pane may be required to consistently represent the chemistry shown in the 3D structure.

5.2.27 Chem & Crystal matching impossible

It is not possible to match the moieties in the 3D structure with the moieties in the 2D structure. This is probably due to a missing 2D diagram.

5.2.28 published bond: A B given but not found in calc

A bond given in the CIF does not correspond to a bond given in the calculated 3D structure. This may indicate that there is a bond missing in the 3D structure or that the published bond list contains an incorrect bond. The published bond list can be edited by clicking on More Info in the Visualizer tab and then clicking on Published Bond.

5.2.29 published bond: A B short

A bond given in the CIF is shorter than the default limits in the CSD-Editor. This may indicate that there is a problem with the position of an atom in the 3D structure or that the published bond list is incorrect. The published bond list can be edited by clicking on More Info in the Visualizer tab and then clicking on Published Bond.

5.2.30 published bond: A B long

A bond given in the CIF is longer than the default limits in the CSD-Editor. This may indicate that there is a problem with the position of an atom in the 3D structure or that the published bond list is incorrect. The published bond list can be edited by clicking on More Info in the Visualizer tab and then clicking on Published Bond.
5.2.31 spacegroup does not agree with symmetry operators

This is a message that is added when an entry is created from DeCIFer and indicates that the space group given in the CIF does not agree with the list of symmetry operators given in the CIF. The list of symmetry operators can be viewed by clicking on More Info in the Visualizer tab and then clicking on Symmetry.

5.2.32 no Z value

The Z value is missing from the entry. To add the Z value you should double-click in the Z, Z' field and add the correct Z value, the Z' value will be calculated automatically when you tab out of the field.

5.3 CIF attributes read by DeCIFer

DeCIFer creates a database entry by looking for a number of standard and CCDC-specific CIF attributes; further details on the standard CIF attributes and their expected contents can be found from the IUCr's Core CIF dictionary page https://www.iucr.org/resources/cif/dictionaries/cif_core. The table below gives the mapping between CIF attributes and CCDC database entry fields.

<table>
<thead>
<tr>
<th>Database entry identifier (Refcode)</th>
<th>_database_code_CSD</th>
</tr>
</thead>
<tbody>
<tr>
<td>If not present:</td>
<td>data_&lt;Cif datablock name&gt;</td>
</tr>
</tbody>
</table>

Compound name

<table>
<thead>
<tr>
<th>_chemical_name_systematic, _chemical_name_common</th>
</tr>
</thead>
<tbody>
<tr>
<td>A single name is added to the Compound Name field, with any additional names added as a semicolon delimited list in the Synonym field. If both attributes are present, the _chemical_name_systematic is given as the Compound Name</td>
</tr>
</tbody>
</table>

Unit cell information
Cell dimensions obtained from _cell_length_a, _cell_length_b, 
_cell_length_c, _cell_angle_alpha, _cell_angle_beta, 
_cell_angle_gamma

Space group is first determined from symmetry operators 
_space_group_symop_operation_xyz (or if not present 
_symmetry_equiv_pos_as_xyz)

This is compared to the given value in 
_symmetry_space_group_name_H-M with a warning given if the two 
do not correlate. If no space group is given in the CIF, this is 
calculated using the symmetry operators with a warning 
reported.

Atomic coordinates

_atom_site_label used for label and check of element type.

Coordinates from _atom_site_fract_x, _atom_site_fract_y 
and _atom_site_fract_z 
Atom_site_occupancy, _atom_site_disorder_assembly and 
_atom_site_disorder_group used in assessment of occupancies 
and disorder.

Literature reference

Author names are taken from _publ_author_name

Default values of To Be Published and the current year are 
given for the publication details.

Formula

This is calculated by DeCIFer; the contents of 
_chemical_formula_moiety and _chemical_formula_sum are 
reported in CSD-Editor for reference.

Remarks

Any content of the attribute _ccdc_publ_extra_info; remarks 
may also be generated by warning from other DeCIFer checks.

Melting point

_chemical_melting_point value assumed to be reported in 
Kelvin (as defined by the IUCr).

Colour
Habit

R-factor (%)
The attributes _refine_ls_R_factor_gt, _refine_ls_wR_factor_gt, _refine_ls_wR_factor_all, _refine_ls_wR_factor_obs, _refine_ls_wR_factor_ref, _refine_ls_R_factor_all and _refine_ls_R_factor_obs are checked for values, with the number converted to a percentage. If multiple values are present, the lowest is reported in the database entry.

Goodness Of Fit
The attributes _refine_ls_goodness_of_fit_ref, _refine_ls_goodness_of_fit_obs, _refine_ls_goodness_of_fit_gt and _refine_ls_goodness_of_fit_all are checked for values. If multiple values are present, the first one is reported in the database entry.

Max Theta
The attributes _diffrn_reflns_theta_max and _diffrn_reflns_theta_full are checked for values. If multiple values are present, the first one is reported in the database entry.

Weighted R-factor (%)
The attributes _refine_ls_wR_factor_ref, _refine_ls_wR_factor_obs, _refine_ls_wR_factor_gt and _refine_ls_wR_factor_all are checked for values, with the number converted to a percentage. If multiple values are present, the first one is reported in the database entry.

Refinement Shift
The attributes _refine_ls_shift/su_max and _refine_ls_shift/esd_max are checked for values. If multiple values are present, the first one is reported in the database entry.

Residual Electron Density Max
_refine_diff_density_max

Residual Electron Density Min
_refine_diff_density_min

Number Of Parameters
_refine_ls_number_parameters

Number Of Restraints
Number Of Constraints

Temperature

Density

Radiation probe

Experiment type

5.4 Glossary

ASER

The legacy Cambridge Structural Database CSD Portfolio (CSDP) database format, consisting of three files (.ind, .msk and .tcd formats)

CSD

Cambridge Structural Database; the repository for all published small-molecule organic and metal-organic crystal structures

CSD-Editor

Software for viewing and updating entries in an in-house database

csdsqlx

The database file format used by CSD-Editor

csdsql

The database file format used by CSD Portfolio software
DeCIfer

The data conversion program used to create an in-house database from crystallographic data in a format searchable by CSD Portfolio software.