

5.3. SYNTACTIC UTILITIES FOR CIF

```

CYCLOPS Check List
-----
Dictionary data names = 2244
New data names in text = 4
[1] Dictionary cif_core.dic 2.0.1 data names = 624
[2] Dictionary cif_mm.dic 0.9.0 data names = 1620

Data names NOT in Dictionary          Line Numbers

_blat1 . . . . .                9  11  94  96
                                   181 183 290 296
_blat2 . . . . .                13  15  98  100
                                   185 187 287 293
_dummy_test . . . . .           5   7  90  92
                                   177 179 201
_rubbish_here. . . . .          431

[1] Dictionary cif_core_2.0.1.dic
[2] Dictionary cif_mm.dic

                                   Line Numbers

[2] _atom_site.calc_attached_atom  413
[1] = _atom_site_calc_attached_atom 412
[2] _atom_site.calc_flag . . . . . 410
[1] = _atom_site_calc_flag         409
[2] _atom_site.fract_x . . . . .   38  44  50  390
[1] = _atom_site_fract_x           389
[2] _atom_site.fract_y . . . . .   39  45  51  394
[1] = _atom_site_fract_y           393
[2] _atom_site.fract_z . . . . .   40  46  52  398
[1] = _atom_site_fract_z           397
[2] _atom_site.id . . . . .        37  43  49  386
[1] = _atom_site_label             385
[2] _atom_site.thermal_displace_type 406
[1] = _atom_site_thermal_displace_type 405
[2] _atom_site.type_symbol . . . . 416 420 424 428
                                   434 438 442 450
[1] = _atom_site_type_symbol       415 419 423 427
                                   433 437 441 449

[later in the validation output file, showing the transition to unreferenced data names ... ]

[1] _symmetry_cell_setting . . . . 319
[2] = _symmetry.cell_setting       320
[1] _symmetry_space_group_name_H-M 323
[2] = _symmetry.space_group_name_H-M 324
[1] _symmetry_space_group_name_Hall 327 445
[2] = _symmetry.space_group_name_Hall 328 446

[1] Dictionary cif_core_2.0.1.dic
[2] Dictionary cif_mm.dic

                                   Names Not Referenced

[2] _atom_site.aniso_B[1][1]
[2] _atom_site.aniso_B[1][1]_esd
[2] _atom_site.aniso_B[1][2]
[... portion of output omitted ...]

[2] _atom_site.aniso_U[3][3]_esd
[2] _atom_site.attached_hydrogens
[1] = _atom_site_attached_hydrogens
[2] _atom_site.auth_asym_id
[2] _atom_site.auth_atom_id
[2] _atom_site.auth_comp_id
[2] _atom_site.auth_seq_id
[2] _atom_site.B_equiv_geom_mean
[1] = _atom_site_B_equiv_geom_mean
[2] _atom_site.B_equiv_geom_mean_esd
[2] _atom_site.B_iso_or_equiv
[1] = _atom_site_B_iso_or_equiv
[2] _atom_site.B_iso_or_equiv_esd
[... remainder of output omitted ...]

```

Fig. 5.3.4.1. Sample output from *CYCLOPS*. The output has been edited and reformatted slightly to fit into the present column width.

respectively. The special character hyphen ('-') may also be supplied as an argument to '-i' or '-o' to indicate standard input or standard output.

Finally, if the operating system supports the passing of environment variables to a program, the names of the input file, output file and dictionary file may be passed through the values of \$CYCLOPS_INPUT_TEXT, \$CYCLOPS_VALIDATION_OUT or \$CYCLOPS_CHECK_DICTIONARY, respectively.

5.3.5. File transformation software

This section describes a number of applications that transform an input CIF either to another CIF that contains a subset of the original contents or to other formats suitable for use with general processing tools. (Conversion to other crystallographic data formats is not discussed here.)

5.3.5.1. QUASAR: a data extractor

The oldest CIF manipulation program is *QUASAR* (Hall & Sievers, 1993), which was described as the prototype CIF application in the original standard specification paper (Hall *et al.*, 1991). Much of the functionality of *QUASAR* has now been included in the *cif2cif* program (Section 5.3.5.2). However, it remains useful as an application in its own right, and so is briefly described here.

5.3.5.1.1. Purpose

The program was designed to read a *request list* of data names, to locate the associated data in an input CIF and to output the data in the order of the request list. The output retains local conformance to CIF syntax rules, but the output file may not be strictly CIF conformant. For example, the same data can be requested multiple times and will be reproduced as often as requested in the output stream, a feature forbidden within a legal CIF.

5.3.5.1.2. Mode of operation

Written as a pure Fortran77 application, *QUASAR* requires three data streams: a file containing the request list, an input CIF and an output file. In an operating system such as Unix, it is convenient to attach the request list to the standard input channel; the first two lines of the input stream then take the form *star_arc_infile* and *star_out_outfile*, where *infile* and *outfile* are the file names of the input and output files, respectively.

The assignment of an output file may be replaced by a line containing *star_log*. When this is done, the program will test the syntactic validity of the input CIF and write any error messages to the standard output channel. In this mode the program may be used as a syntactic validator, although it is more tolerant of certain syntactic errors than *vcif* (Section 5.3.2.1).

5.3.5.1.3. The request list

Fig. 5.3.5.1 is an example request list, intended to highlight some of the special features of the way the program operates. Fig. 5.3.5.2 shows an example CIF against which this request list will be tested; Fig. 5.3.5.3 shows the output. Both figures have been modified slightly to fit on the printed page; they are derived from the sample files distributed with the program.

The request list begins with directives specifying the input and output file names (*qtest.cif* and *qtest.out*, respectively). The file may contain comments prefaced by a hash character #; this is a useful feature for annotating a request list. Another use for such comments is seen in the standard request list distributed to authors for papers published in *Acta Crystallographica*. Here, data names that are *not* normally published are hidden within the request list as comments and may be activated if they occur in a *publ_manuscript_incl_extra_item* loop within a CIF (see Section 5.7.2.3).