

5.2. STAR FILE UTILITIES

```

data_Gaussian

loop_
  _basis_set_atomic_name
  _basis_set_atomic_symbol
  _basis_set_atomic_number
  _basis_set_atomic_mass
loop_
  _basis_set_contraction_scheme
  _basis_set_funct_per_contraction
  _basis_set_primary_reference
  _basis_set_source_exponent
  _basis_set_source_coefficient
  _basis_set_atomic_energy
  loop_
  _basis_set_function_exponent
  _basis_set_function_coefficient

  hydrogen  H  1  1.0079
#
(2)->[2]  1:  PKC1.1.1  R44 . -0.485813
  1.3324838E+01  1.0
  2.0152720E-01  1.0 stop_
(2)->[2]  1:  PKC1.2.1  R33 . -0.485813
  1.3326990E+01  1.0
  2.0154600E-01  1.0 stop_
(2)->[1]  2  PKC1.14.1  R24 R24 -0.485813
  1.3324800E-01  2.7440850E-01
  2.0152870E-01  8.2122540E-01 stop_
(3)->[2]  2:1  PKC1.23.1  R75 R75 -0.496979
  4.5018000E+00  1.5628500E-01
  6.8144400E-01  9.0469100E-01
  1.5139800E-01  1.0000000E+01 stop_ stop_

  lithium  Li  3  6.94
#
(4)->[4]  1:  PKC3.1.1  R44 . -7.376895
  3.4856175E+01  1.0
  5.1764114E+00  1.0
  1.0514394E+00  1.0
  4.7192775E-02  1.0 stop_
(9,4)->[3,2]  7:2:1,3:1
  PKC3.9.1  R2  R98 -7.431735
  921.271  0.001367  138.730  0.010425
  31.9415  0.049859  9.35329  0.160701
  3.15789  0.344604  1.15685  0.425197
  0.44462  0.169468  0.44462  -0.222311
  0.076663  1.116477  0.028643  1.0
  1.488  0.038770  0.2667  0.236257
  0.07201  0.830448  0.02370  1.0 stop_
(4,3)->[3,2]  4:2:1,2:1
  PKC3.30.1  R77 R77 -7.419509
  1.09353E+02  1.90277E-02  1.64228E+01  1.30276E-01
  3.59415E+00  4.39082E-01  9.05297E-01  5.57314E-01

  5.40205E-01  -2.63127E-01  1.02255E-01  1.14339E+00
  2.85645E-02  1.00000E+00

  5.40205E-01  1.61546E-01  1.02255E-01  9.15663E-01
  2.85645E-02  1.00000E+00 stop_ stop_

```

Fig. 5.2.2.1. Example quantum chemistry basis set functions in STAR File format.

coefficients). At the innermost loop level, a loop packet is simply a row within a table of exponents and coefficients of the basis set function.

If one were to treat this example file as a database of indeterminate structure and query the values associated with one of the data names, for example, `_basis_set_function_exponent`, one would retrieve a series of strings `1.3324838E+01`, `2.0152720E-01` etc. However, the value strings in themselves are insufficient to allow the reconstitution of any data structure in the file. One also needs an expression of the levels within the nested loop structure at which the values were located, and an indication that they were associated with different packets of information at those various levels. This additional information about the context of each value is sufficient to determine its position within the data structure

```

data_Gaussian
loop_
  loop_
  loop_
  _basis_set_function_exponent
  stop_
  stop_

  1.3324838E+01 2.0152720E-01 stop_
  1.3326990E+01 2.0154600E-01 stop_
  1.3324800E-01 2.0152870E-01 stop_
  4.5018000E+00 6.8144400E-01 1.5139800E-01 stop_
  stop_

  3.4856175E+01 5.1764114E+00 1.0514394E+00
  4.7192775E-02 stop_

  921.271 138.730 31.9415 9.35329 3.15789 1.15685
  0.44462 0.44462 0.076663 0.028643 1.488 0.2667
  0.07201 0.02370 stop_

  1.09353E+02 1.64228E+01 3.59415E+00 9.05297E-01
  5.40205E-01 1.02255E-01 2.85645E-02 5.40205E-01
  1.02255E-01 2.85645E-02 stop_
  stop_

```

Fig. 5.2.2.2. Retrieval from the example file in Fig. 5.2.2.1 of the value of `_basis_set_function_exponent` with associated context.

without any other *a priori* information regarding the data model. The context is most easily expressed by listing the output values in STAR File format.

Fig. 5.2.2.2 is an output listing of the requested values for this example, where the context is expressed as the innermost of three nested loop levels and distinct packets at this level are indicated. It will be seen also that by tracing the disposition of `stop_` words the embedding within higher-level loop packets can also be inferred.

5.2.2.3. Context in data sets

Another indicator of context in the previous example is the data-block header, which was reproduced in the output of Fig. 5.2.2.2.

The STAR File allows data instances in three types of location: in a data block, in a save frame or in a global block.

The usual way to partition a STAR File is by data blocks; each such block represents a data set in which a data name (associated with a single or multiple values) may be declared once only.

Data blocks may include save frames. A save frame is an encapsulated subsidiary data set, effectively insulated from the contents of the surrounding data block, in which data items may occur that have the same names as items in the parent data block. Indeed, 'parent' is potentially a misleading term, since no relationship is implied between the data within a save frame and those in the data block in which the save frame occurs. A reference to a save frame may, however, occur as a data *value* within the data block where the save frame is specified. Recall from Section 2.1.3.6 that save frames within a data block are uniquely identified by the *frame-code* header.

Global blocks may also occur in a STAR File, preceding or interspersed between data blocks. For each data item defined within a global block, that definition is inherited by each succeeding data block that does not contain an internal definition of a data item with the same name. If there is a definition of a data item with the same name within a data block, that internal definition overrides the global definition within that data block. The situation is then re-evaluated in the next data block. If that data block does not contain an internal definition, the global definition holds.