Chapter 9.5

IDEAL

A program to compare one set of atomic coordinates with one or more other similar sets

R.O. GOULD, N. MOULDEN AND P. TAYLOR,
UNIVERSITY OF EDINBURGH
SCOTLAND, UK

9.5.1 General Principle of operation

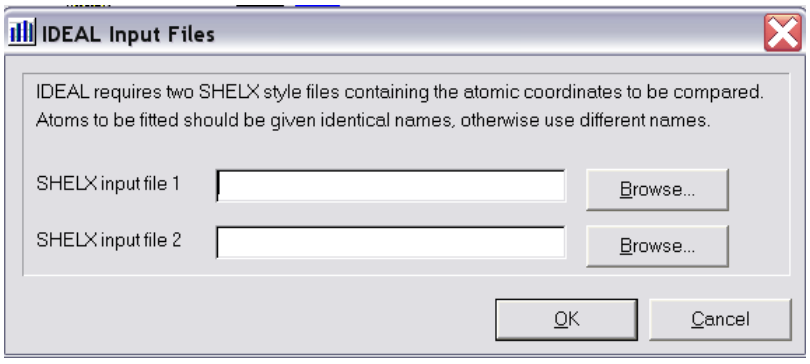
An initial set of atomic coordinates is read in from the previously prepared file IDEAL.INPUT, and stored in orthogonal Angstrom form. As each subsequent set is read in, atom names common to it and the first set are identified, and providing that at least three such atoms are found, an attempt is made to superimpose the two sets. The initial fit is made by:

1. Translating the origin of both sets to the centres of gravity of the common atoms.
2. Defining the positive z-axis to lie in the direction of the atom furthest from the origin in the first set.
3. Defining the x-axis to lie such that the atom lying furthest from the z-axis lies in the xz-plane with a positive x- coordinate.
4. Rotating the atoms in the second set using both the rotations calculated for it and the inverse of the rotations calculated for the first set.

The initial fit is then refined by iterative least squares, rotating the second set about the x-, y- and z-axes to minimise the sum of the squares of the deviations of equivalent atoms from one another until convergence (a change in the r.m.s. deviation of less than 0.1% of its previous value) occurs.

9.5.2 Input file format -IDEAL.INPUT

In the current version of WinGX, a preliminary version of the input file *IDEAL.INPUT* is created automatically, from two SHELX parameter files. Some manual editing of this file will usually be necessary.



The input format is designed to be simple rather than elegant. Each line is read and interpreted according to the first four characters on it. Lines relating to each set of coordinates must be given together, with the CELL line (if any) first, and the END line last. The following are the possible lines:

1. CELL : to be followed by six numbers on the same line in free format, giving a, b, c, alpha, beta and gamma. These may be in direct or reciprocal space ($Å^{-1}$),

with the angles either in degrees or as cosines. If no CELL is given, Angstrom-coordinate input is assumed, so if two sets are given from the same structure, the CELL line must be repeated.

2. ATOM : to be followed by an atom name of up to four characters, and three coordinates, x, y, and z. The word ATOM is optional, but if it is omitted, the atom name must come in columns 1-4, and must, of course, not correspond to another type of card. Anything beyond the z-coordinate will be ignored.
3. END : signals the end of a coordinate set. There must be at least two of these in any run.
4. FINI(SH) : signals the end of the job.
5. INVE(RT) : placed anywhere before an END-card will invert all the atoms in that set to -x,-y,-z. It may not be used in the first set.
6. SHEL(X) : used anwhere in a run indicates that all subsequent CELL and ATOM cards will be in SHELX-format, i.e. the first number after CELL or the atom-name will be ignored. Other SHELX cards must not be present. AFIX, for example, will be treated as an atom!
7. Lines with blanks in columns 1-4 will be ignored.

Below is a simple example input file *IDEAL.INPUT* to compare two sets of atoms. Note that the order of the atoms is unimportant, only the names are compared:

```
ATOM  O1  1.9046  3.4531  4.5047
ATOM  C2  1.1244  2.2765  4.2905
ATOM  O3  0.1562  2.1305  5.3469
ATOM  C4 -0.6103  3.2144  5.6631
ATOM  O4 -1.6185  3.0046  6.2954
ATOM  C5 -0.0995  4.5072  5.2557
ATOM  C6  1.2575  4.6284  4.7334
ATOM  O6  1.8640  5.6560  4.5506
END
CELL  7.544  8.576  12.014  78.01  83.36  71.41
O1   .2370   .2956   .8258
C4   .3679   .5907   .7843
O4   .4656   .6615   .7618
C5   .2929   .5148   .8982
C2   .2001   .4138   .7210
O3   .3369   .5015   .6997
C6   .2617   .3499   .9183
O6   .2582   .2612  1.0104
INVERT
END
FINISH
```

In this case, all eight atoms of set 1 also occur in set 2; this is not necessary; any or all sets may have extra atoms. The INVERT line will cause the enatiomorph of the atoms in set 2 to be fitted to set 1. Additional sets of coordinates may be added before the FINISH line.

9.5.3 Output file - IDEAL.LST

For each CELL, the orthogonalisation matrix used is given. Coordinates are listed in both fractional (XF etc.) and orthogonal (XO) form. The coordinates of the initial and final fits are given relative to the first set, but with the origin at the centre of

gravity. DIFF is the distance in Angstroms between the corresponding atoms in the two sets. Large values of DIFF may indicate wrongly labelled atoms or that the two structures are actually mirror images. Finally, all of the coordinates in the second or subsequent set are listed relative to the cell of the first set.

A sample output file (*IDEAL.LST*) is shown below

```
IDEAL, Version of 19.4.94
 R.O. Gould, N. Moulden and P. Taylor, University of Edinburgh
 Input file name: IDEAL.INS

 Beginning of data for set number 1
 Cell:     8.307  22.468   8.883   90.00   96.53   90.00
 Orthogonalisation matrix:
   8.2529   0.0000   0.0000
   0.0000  22.4680   0.0000
  -0.9448   0.0000   8.8830
 Cell volume =    1647.1

 Name       xf      yf      zf         xo      yo      zo
 Ru1     0.2065  0.3875  0.1996      1.704   8.705   1.578
 Fe1     0.4490  0.3744  0.4327      3.705   8.413   3.419
 Fe2     0.1545  0.3573  0.4883      1.275   8.027   4.191
 C11     0.3345  0.4054  0.0374      2.760   9.109   0.016
 C12    -0.0095  0.3920  0.0917     -0.078   8.808   0.823
 C13     0.2250  0.3019  0.1577      1.857   6.783   1.188
 C14     0.1919  0.4707  0.2590      1.584  10.575   2.119
 C15     0.5984  0.3619  0.5945      4.939   8.131   4.715
 C16     0.5201  0.4476  0.3834      4.292  10.056   2.914
 C17     0.5537  0.3352  0.2980      4.569   7.531   2.124
 C18     0.1357  0.3367  0.6799      1.120   7.565   5.912
 C19    -0.0039  0.4118  0.4757     -0.032   9.252   4.229
 C110    0.0312  0.2964  0.4056      0.258   6.659   3.574
 C111    0.3436  0.2993  0.4774      2.835   6.724   3.916
 C112    0.3021  0.4239  0.5547      2.493   9.523   4.642
 O11     0.4064  0.4138 -0.0591      3.354   9.298  -0.909
 O12    -0.1336  0.3961  0.0278     -1.102   8.900   0.373
 O13     0.2341  0.2544  0.1279      1.932   5.716   0.915
 O14     0.1816  0.5193  0.2870      1.499  11.669   2.378
 O15     0.6950  0.3529  0.6919      5.736   7.929   5.489
 O16     0.5672  0.4934  0.3563      4.681  11.086   2.629
 O17     0.6250  0.3099  0.2167      5.158   6.963   1.334
 O18     0.1160  0.3236  0.8001      0.957   7.270   6.998
 O19    -0.1053  0.4460  0.4731     -0.869  10.020   4.302
 O110   -0.0476  0.2582  0.3588     -0.393   5.801   3.232
 O111    0.3657  0.2492  0.4918      3.018   5.599   4.023
 O112    0.3109  0.4671  0.6245      2.566  10.496   5.253

 Beginning of data for set number 2
 Cell:     9.203  11.903  15.117   81.55   87.29   66.72
 Orthogonalisation matrix:
   8.4531   0.0000   0.0000
   3.6125  11.7738   0.0000
   0.4351   1.7491  15.1170
 Cell volume =    1504.5

 Name       xf      yf      zf         xo      yo      zo
 Ru1     0.0714  0.2157  0.2915      0.604   2.798   4.815
 Fe2     0.3382 -0.0015  0.3240      2.859   1.204   5.042
 Fe1     0.2312  0.0668  0.1621      1.954   1.622   2.668
 P1a    -0.1539  0.3919  0.2376     -1.301   4.058   4.210
 O1a     0.2450  0.2898  0.0561      2.071   4.297   1.462
```

```
O2a    -0.0660  0.0759  0.0903     -0.558   0.655   1.469
O3a     0.4281 -0.0980  0.0401      3.619   0.393   0.621
O4a     0.1879 -0.1540  0.2541      1.588  -1.134   3.654
O5a     0.5447  0.0787  0.1964      4.604   2.894   3.344
O6a     0.4676  0.1245  0.4312      3.953   3.155   6.940
O7a     0.1612 -0.0754  0.4721      1.363  -0.305   7.075
O8a     0.6024 -0.2411  0.3615      5.092  -0.662   5.305
O9a    -0.1347  0.0634  0.3108     -1.139   0.260   4.751
O10a    0.2942  0.3509  0.2439      2.487   5.194   4.429
O11a    0.0406  0.2706  0.4817      0.343   3.333   7.773
C1a     0.2346  0.2069  0.0994      1.983   3.283   1.967
C2a     0.0452  0.0782  0.1196      0.382   1.084   1.964
C3a     0.3513 -0.0344  0.0879      2.970   0.864   1.421
C4a     0.2258 -0.0714  0.2467      1.909  -0.025   3.703
C5a     0.4372  0.0584  0.2210      3.696   2.267   3.633
C6a     0.4143  0.0804  0.3880      3.502   2.443   6.186
C7a     0.2234 -0.0431  0.4135      1.888   0.300   6.273
C8a     0.5028 -0.1479  0.3439      4.250   0.075   5.159
C9a    -0.0558  0.1171  0.3032     -0.472   1.177   4.764
C10a    0.2125  0.3005  0.2606      1.796   4.306   4.558
C11a    0.0480  0.2516  0.4097      0.406   3.136   6.654
C12a   -0.3313  0.4228  0.3069     -2.801   3.781   5.235
C13a   -0.3389  0.3467  0.3829     -2.865   2.858   6.247
C14a   -0.4763  0.3742  0.4320     -4.026   2.685   6.978
C15a   -0.6060  0.4764  0.4062     -5.123   3.420   6.710
C16a   -0.6008  0.5545  0.3315     -5.079   4.358   5.720
C17a   -0.4645  0.5278  0.2812     -3.926   4.536   4.972
C18a   -0.1341  0.5412  0.2263     -1.134   5.888   4.309
C19a   -0.0513  0.5627  0.2923     -0.434   6.440   5.381
C20a   -0.0417  0.6766  0.2893     -0.352   7.816   5.539
C21a   -0.1151  0.7686  0.2208     -0.973   8.634   4.632
C22a   -0.1987  0.7502  0.1558     -1.680   8.115   3.581
C23a   -0.2081  0.6364  0.1585     -1.759   6.741   3.419
C24a   -0.2215  0.3845  0.1280     -1.872   3.727   2.511
C25a   -0.3430  0.3454  0.1211     -2.899   2.828   2.286
C26a   -0.3852  0.3300  0.0380     -3.256   2.494   0.984
C27a   -0.3063  0.3522 -0.0368     -2.589   3.040  -0.074
C28a   -0.1866  0.3908 -0.0304     -1.577   3.927   0.143
C29a   -0.1430  0.4069  0.0519     -1.209   4.274   1.434

Start coordinates of fit for set 2
Name       x1      y1      z1         xn      yn      zn        diff
Ru1     1.704   8.705   1.578      1.692   8.713   1.543       0.038
Fe2     1.275   8.027   4.191      1.279   8.027   4.196       0.005
Fe1     3.705   8.413   3.419      3.714   8.406   3.450       0.032

rms misfit before cycle 1:    0.029    shifts(deg.):  -0.07  -0.26   0.02

rms misfit before cycle 2:    0.028    shifts(deg.):   0.00   0.00   0.00

rms misfit before cycle 3:    0.028    shifts(deg.):   0.00   0.00   0.00

rms misfit before cycle 4:    0.028    shifts(deg.):   0.00   0.00   0.00

rms misfit before cycle 5:    0.028    shifts(deg.):   0.00   0.00   0.00

Final coordinates of fit for set 2
Name       x1      y1      z1         xn      yn      zn        diff
Ru1     1.704   8.705   1.578      1.685   8.712   1.546       0.038
Fe2     1.275   8.027   4.191      1.284   8.026   4.200       0.013
Fe1     3.705   8.413   3.419      3.716   8.408   3.443       0.026

Coordinates for set 2 in cell 1:
Name       xf      yf      zf         xo      yo      zo
Ru1     0.2041  0.3877  0.1957      1.685   8.712   1.546
```

Fe2	0.1556	0.3572	0.4894	1.284	8.026	4.200
Fe1	0.4502	0.3742	0.4355	3.716	8.408	3.443
P1a	0.2953	0.4125	-0.0388	2.437	9.268	-0.624
O1a	0.5631	0.4933	0.3598	4.647	11.082	2.664
O2a	0.6471	0.3014	0.2499	5.341	6.772	1.608
O3a	0.6796	0.3630	0.7071	5.608	8.155	5.639
O4a	0.3689	0.2492	0.4945	3.044	5.599	4.044
O5a	0.3101	0.4650	0.6342	2.559	10.447	5.341
O6a	-0.1111	0.4434	0.4710	-0.917	9.962	4.289
O7a	-0.0450	0.2588	0.3542	-0.371	5.816	3.189
O8a	0.1063	0.3149	0.7919	0.877	7.076	6.934
O9a	0.2695	0.2559	0.1299	2.224	5.750	0.900
O10a	0.1917	0.5188	0.2927	1.582	11.656	2.419
O11a	-0.1505	0.3876	0.0757	-1.242	8.708	0.814
C1a	0.5154	0.4472	0.3829	4.253	10.047	2.914
C2a	0.5672	0.3311	0.3153	4.681	7.440	2.265
C3a	0.5900	0.3670	0.6015	4.869	8.246	4.785
C4a	0.3454	0.2996	0.4778	2.851	6.732	3.918
C5a	0.2980	0.4229	0.5620	2.459	9.501	4.711
C6a	-0.0056	0.4115	0.4739	-0.046	9.245	4.215
C7a	0.0352	0.2970	0.4006	0.291	6.673	3.525
C8a	0.1302	0.3325	0.6775	1.074	7.472	5.895
C9a	0.2464	0.3042	0.1577	2.033	6.835	1.168
C10a	0.1979	0.4704	0.2579	1.633	10.569	2.104
C11a	-0.0159	0.3873	0.1172	-0.131	8.701	1.056
C12a	0.2017	0.3708	-0.2039	1.665	8.331	-2.002
C13a	0.0920	0.3257	-0.1944	0.759	7.318	-1.814
C14a	0.0265	0.2955	-0.3226	0.219	6.640	-2.891
C15a	0.0693	0.3097	-0.4599	0.572	6.958	-4.151
C16a	0.1766	0.3549	-0.4727	1.458	7.974	-4.366
C17a	0.2442	0.3853	-0.3452	2.015	8.657	-3.298
C18a	0.2606	0.4897	-0.1036	2.151	11.003	-1.166
C19a	0.1146	0.5171	-0.0817	0.946	11.619	-0.834
C20a	0.0793	0.5740	-0.1369	0.654	12.896	-1.291
C21a	0.1893	0.6031	-0.2142	1.562	13.550	-2.081
C22a	0.3327	0.5768	-0.2387	2.745	12.959	-2.435
C23a	0.3684	0.5201	-0.1832	3.041	11.685	-1.976
C24a	0.5117	0.4000	-0.0423	4.223	8.988	-0.859
C25a	0.5661	0.3480	-0.1025	4.672	7.818	-1.446
C26a	0.7313	0.3364	-0.0935	6.036	7.559	-1.521
C27a	0.8402	0.3760	-0.0241	6.934	8.448	-1.008
C28a	0.7873	0.4274	0.0354	6.498	9.602	-0.430
C29a	0.6230	0.4397	0.0272	5.141	9.880	-0.347