Fragment Search By Integrated Patterson and Direct Methods (PATSEE)

PATSEE is a public domain program for location of a fragment of known geometry by integrated Patterson, packing, and direct methods; it was written by Ernst Egert* at the Goettingen University Institut für Anorganische Chemie, Tammanstraße 4, D-37077 Goettingen, Germany.

This program attempts to combine the merits of both Patterson and direct methods in order to position a fragment of known geometry in the unit cell. It is valid for all space groups in all settings.

The rotation search can find the orientation of a search model of any size and allows one torsional degree of freedom. (Searching for two independent fragments simultaneously would be very time consuming because a large number of combinations of orientations would have to be considered.) The translation search may locate up to two independent fragments of any size (including single atoms) taking into account known atoms at fixed positions, if any.

PATSEE expects to read an input file "NAME.INP" which is contains TITL ... UNIT (as for XS) etc., followed by further instructions, a search fragment, and possibly the list of E-values and tabulated asymmetric unit of the sharpened Patterson function (created using the PSEE instruction in XS). Alternatively the E-values and Patterson may be read from a separate file (NAME.PAT). The "best" final coordinates are written to a file (NAME.REP) ready for partial structure extension using XS (edit or renamed to NAME.INS for this purpose). Execute PATSEE in dialogue mode by:

    $ **PATSEE name**   <RETURN>

or submit it as a batch job by:

    $ **X PATSEE/BATCH name**   <RETURN>

1 Strategy

A structure solution with PATSEE normally consists of three stages:

1. The data are processed with XS using the PSEE command to calculate the sharpened Patterson function and the largest n E-values. This instruction creates an output file (NAME.INP) to be read by PATSEE. Normally, 200 E-values are sufficient.

2. A fragment search is performed with PATSEE. The normal procedure will be a rotation search (ROTS) followed by a translation search (TRAN) within the same run. The input must contain:

    a. The usual XS sequence TITL ... UNIT.

    b. ROTS and/or TRAN.

    c. Atomic coordinates of the search model(s); at least one atom belonging to FRAG 1 (default) must be present.

     d.   E-magnitudes and Patterson grid values, which must form the end of the input stream (conveniently appended via END n and read from NAME.PAT).

All other input instructions are optional but - like the various parameters in the ROTS and TRAN commands - could prove very useful when applied by an expert.

3.   The fragment undergoes partial structure expansion with another run of XS. PATSEE writes a file with suitable instructions for the expansion of the best solution starting from the first (best) orientation (which, however, may *not* be the fragment position with the highest overall CFOM). Note that the last line (HKLF 4) may have to be changed according to your data type and format. For the tangent expansion and E-Fourier recycling, positions of atoms heavier than Al are automatically kept fixed provided that they come first in the atom list.

Any other solution can also be tested conveniently by specifying SPIN (with the appropriate rotation angles phi1, phi2, and phi3) and MOVE (with the translational shifts X1, Y1, and Z1, and possibly X2, Y2, and Z2) followed by FRAG (if not Cartesian) and the original input atomic coordinates which produced the particular PATSEE solution; suspicious atoms may be omitted, of course.

In order to demonstrate how a crystal structure is solved by Patterson search using a suitable model, the test structure LAC1 (which used to be difficult for direct methods) serves as an excellent example. It is recommended that you study the test output carefully before using PATSEE with your own data.

2 Input Instructions

All input commands are read in free format with parameters separated by one or more blanks. The input stream must begin with the sequence TITL...UNIT and end with END n or READ; all other commands may be given in any order (for exceptions, see below). Continuation lines are allowed for all instructions except SFAC with element symbols and SYMM.

Characters following "=", "/" or four blank spaces in columns 1-4 are treated as comment and thus only printed (but otherwise ignored).
In the following list of instructions, most parameters have default values which are given in brackets ("#" means that it is problem-dependent). All distances are in Ångstroms.

ROTS nt [#] ns [5] lmax [6] dmin [#]

This initiates the rotation search for fragment 1.

|**nt**| random orientations are tested. If |**nt**| < 1, a suitable number (10000-60000 depending on the Laue group) will be calculated. Negative **nt** produces a list of bond lengths and angles for the search model; this is useful for detecting possible errors in the atomic coordinates. The |**ns**| best rotation solutions are kept and transferred to the subsequent translation search and up to |**ns**| + 3 solutions are printed out. Negative **ns** indicates that the search model has a mirror plane or an inversion center, which makes a more extensive equivalence test necessary. For the rotation search, only vectors shorter than |**IMAX**| are used. |**IMAX**| may be increased for very accurate models and should be reduced for unreliable ones. Vectors shorter than 2 Å are only used if **IMAX** < 0; this is recommended only for short vectors with high weights. Two solutions are regarded as similar if all equivalent atoms are closer than |**dmin**|, whose optimum value depends on the size of the model. Negative **dmin** retains low weight vectors (which are normally ignored in order to save time).

If the SPIN command is present, the specified orientation is refined by a restricted rotation search (this could improve chances of the subsequent translation search); in that case, however, ROTS must occur before the first FRAG 1 atom!

The rotation search starts with phi 1 = phi 2 = phi 3 = 0; thus the combination of ROTS -1 with TRAN gives bond lengths and angles followed by a translation search.

RFOM fr [0.3] nv [100]

A fraction |**fr**| of worst-fitting intramolecular vectors (but not more than **nv**) are used for calculating the RFOM figure of merit, which resembles either the sum (|**fr**| = 1) or the minimum function (**nv** =1). The number of vectors incorporated into RFOM should be increased for bad models so that single (strong) vectors get less weight, and may be reduced for models consisting of a few "heavy" atoms. If **fr** < 0, a sorted (according to the weight) list of intramolecular vectors used for the rotation search is printed together with the corresponding lengths, if this list contains vectors with high weights, a comparison with most prominent peaks in the Patterson function (printed by the XS command PSEE) might be useful.

TRAN nt [#] ns [#] dmin1 [1.8] dmin2 [2.4] nE [100] ntpr [100]

This initiates the translation search for fragments 1 and 2 (if present). |**nt**| random positions are refined for each orientation; if |**nt**| < 1, a suitable number will be calculated from the number of search parameters and their range. Negative **nt** produces a sorted list of E-magnitudes linked by strong three-phase structure invariants. If it is found that a few E-values participate in a large number (e.g. more than ten) of triple-phase relations, they should be omitted (see the READ instruction). The |**ns**| best solutions are retained, but only the atomic coordinates for the first are normally printed (give **ns** a negative value in order to get them all).

Only fragment positions without intermolecular contacts shorter than |**dmin1**| (preliminary distance test) and |**dmin2**| (final distance test), respectively, are assumed to be physically reasonable. If |dmin| < 0.1, the preliminary distance test is skipped (this may save time if the fragment occupies only a small portion of the asymmetric unit). If **dmin1** < 0, a list of atoms used for the first test is printed. |**nE**| and |**ntpr**| are the maximum numbers of E-magnitudes and triple-phase relations used for the translation search. If **ntpr** < 0, the TPRSUM rejection criterion is reduced from 0.5 to 0.3; this is recommended for very small or inaccurate models.

The default TPRSUM rejection criterion is reduced by 0.1 for noncentrosymmetric structures.

The random numbers may be influenced by the user. If the first parameter **nt** on the ROTS and TRAN commands is given as "m.n", n determines the random number seed (however "m.5" has the same effect as "m.0"). So when a rotation search is to be repeated with 60,000 instead of 40,000 tries, this should be done with ROTS 60000.n; otherwise a large proportion of the search will just be a repetition of the previous runs.

If **nE** is made negative, 3 cycles of translation refinement are applied instead of the usual 2. If $r_{min} 2 < 0$, a different formula is used for CFOM.

TFOM fr [0.2] nv [250]

The meaning of **fr** and **nv** is similar to that of the parameters in the RFOM instruction, but for TFOM intermolecular vectors are used. If fr < 0, CFOM will be calculated ignoring the TPRSUM value; this gives the Patterson criterion TFOM more weight and is advisable for the location of single heavy atoms. A special option for very large structures (possibly at low resolution) is invoked if **nv** < 1: a very rapid translation search without using the Patterson function (TFOM is arbitrarily set to 1.0 and the final distance test restricted to a small number

of atoms representing the search model). Note that **fr** and **nv** MUST NOT BOTH BE NEGATIVE!

SPIN phi1 [0] phi2 [0] phi3 [0]

The following atoms are rotated by the specified angles (all in radians) about the x, y, and z axis of a Cartesian coordinate system. This instruction defines the orientation of a fragment for a subsequent translation search or a restricted rotation search (see the ROTS command).

Move delx [0] dely [0] delz [0] sign [1]

The coordinates of the following atoms are changed to:

$$x' = delx + sign \cdot x$$
$$y' = dely + sign \cdot y$$
$$z' = delz + sign \cdot z$$

The three shifts are ignored for atoms belonging to FRAG 1 or 2, which may only be inverted (sign = -1).

FRAG code [1] a [1] b [1] c [1] alpha [90] beta [90] gamma [90]

This instruction defines the coordinate system to which all following atomic positions (until the next FRAG instruction) are referred and thus enables the convenient input of search models from published structures (default: Cartesian coordinates). The six cell parameters in Å must be either specified completely or omitted. The default search fragment (**code** = 1) *must* be present. **code** = 2 indicates a fragment with known orientation (defined by SPIN) but unknown position, for which a translation search is to be performed (simultaneously with fragment 1). **code** < 0 indicates a fragment with known orientation and position (from a previous PATSEE run) fixed by SPIN and MOVE. A special option (code = 0) is provided for fragments whose true crystal coordinates are already known (e.g. heavy atoms from a Patterson interpretation via the XS command PATT); for such fragments the cell parameters must *not* be specified.

name sfac x y z sof [11] U (or U11 to U12)

This defines the name, type, coordinates and site occupation factor for an atom belonging to one of the fragments (see the ATOM description in Section XXX); temperature factors are ignored. On reading such an atom instruction, FRAG, SPIN and MOVE are applied in that order. The site occupation factor can be used to reduce the weight of possibly doubtful, but not definitely wrong atoms. Negative **sfac** indicates that the atom may lie on a special position or take part in short intermolecular interactions so that the packing criteria defined by ROTS and TRAN should not be applied to it (see TWIS for the use of dummy atoms). This is much more economic than softer global criteria in the ROTS and TRAN instruction.

TWIS w0 Dw we [w0+355] rtest [2]

The vector between the atoms immediately preceding and following this instruction is treated as a bond (irrespective of its length) connecting two rigid groups with a variable torsion angle. The number and range of torsion angles to be tested is defined by a minimum (w_0) and a maximum value (w_e) and an increment (**Dw**) given in degrees. It is important to note that the input geometry is arbitrarily defined as $w = 0°$. If only w_0 is given, a single search with the corresponding torsion angle will be performed. The distance |**rtest**| serves as a criterion: 1) for establishing the connectivity and thus for deciding which atoms belong to which part of the flexible fragment (if **rtest** < 0 it is assumed that TWIS separates the two parts) and 2) for detecting short contacts upon rotation (ignored if |**rtest**| < 0.2), in which case the next geometry is immediately generated.

TWIS is only allowed for fragment 1. If, as in ferrocene-type complexes, the rotation axis does not correspond to a bond between two atoms, dummy hydrogen atoms with negative sfac and $0 < sof < 0.5$ can be introduced.

CODE lv1 lv2 lv3 lv4 lv5 lv6 lv7

(All seven parameters must be specified in ascending order.)

Code lv7 incr   [lvn = lv7 - (7-n) · incr]

The CODE instruction, which can take the above two forms, over-writes the automatic encoding of Patterson grid values (0 ... 7) according to seven test levels. This is sometimes necessary if the Patterson distribution is not smooth enough; the warning printed by the program (see test output) can often be ignored since the criterion for this is rather strict. Normally the count for the first (negative Patterson grid values) and possibly the last (origin of Patterson function) test level yields considerably larger numbers than for the others. Of course, the distribution among the test levels could be forced to be regular, but then two levels could be too close (which would disturb RFOM and TFOM) and "features" of the data or structure would be hidden.

END n

This is normally the last input line and indicates that subsequent instructions (generated by XS) are to be read from unit **n**; they will not be interpreted or printed until a READ instruction is found. For the VAX version of PATSEE, n must be 7 and NAME.PAT should be "assigned" to FOR007 in the .COM file.

READ nE 2Θ (max)

This must be the last instruction if the whole input (including the Patterson function) resides in one file; in that case, however, the TITL ... UNIT sequence precending READ in the output file generated by XS has to be removed. The characters between the parameters **nE** (number of E-values in the subsequent list) and 2Θ (max) (limit applied to E-values, *not* to Patterson function) are not interpreted. The E-list contains four reflections per line (h, k, l, and E for each). Since negative E-values are ignored, unsuitable reflections may thus be suppressed (see TRAN), but be careful not to change the format!