

CCP4 NEWSLETTER ON PROTEIN CRYSTALLOGRAPHY

Number 32. June 1996

An informal Newsletter associated with the BBSRC Collaborative Computational Project No. 4 on Protein Crystallography.

Contents

CCP4 - Recent changes

M. D. Winn & A. Ralph

A data collection strategy option in MOSFLM

A. G. W. Leslie

Electron density modification with Solomon

J. P. Abrahams

ROTGEN -A Program for Rotation Images Simulation and Analysis

J. W. Campbell

NEWS FROM THE UPPSALA SOFTWARE FACTORY - 6

Gerard J. Kleywegt

Density growing : a method for local improvement of electron density maps

A.G.Urzhumtsev

Tcl/Tk-based programs. I. CONFOR : user-friendly converter for crystallographic data files

L.M.Urzhumtseva and A.G.Urzhumtsev

Relationships between some rotation descriptions for molecular replacement procedure

A.G.Urzhumtsev

Editors: Sue Bailey and Martyn Winn
Daresbury Laboratory, Daresbury,
Warrington, WA4 4AD, UK

CCP4 - Recent changes

Martyn Winn & Adam Ralph

Daresbury Laboratory,
Daresbury,
Warrington
WA4 4AD, U.K.

Since the last Newsletter, versions 2.15, 3.0 and 3.1 of the CCP4 suite have been released. Version 3.0 included a significant addition to the MTZ format, namely the use of Missing Number Flags (MNF) to indicate missing data. Most programs now read and write MNFs, and it is recommended that MNFs are incorporated into existing data files using the jiffy MTZMNF. With a policy for handling missing data, it is now straightforward (and strongly recommended) to work with complete datasets, i.e. to include all reflections up to the high resolution limit whether or not the corresponding datum is present (see the UNIQUE documentation for more details). A Users' Guide to MNFs can be found at <http://www.dl.ac.uk/CCP/CCP4/user.html>

In addition to improvements to existing programs, a number of new programs were introduced. In particular, have a look at:

1. RESTRAIN from the Birkbeck group: refinement program including, amongst many options, TLS refinement.
2. REFMAC from Garib Murshudov: Maximum Likelihood refinement program.
3. SOLOMON from Jan-Pieter Abrahams: density modification program using solvent flipping.
4. XDLDATAMAN, XDLMAPMAN from Gerard Kleywegt: manipulation and reformatting of data and map files.

Version 3.1 is primarily to fix a number of bugs contained in version 3.0, and to do some general tidying. The most significant/visible changes are as follows: Missing data in MTZ files are now listed by MTZDUMP as the string 'MNF'. ROTAPREP can now convert user-defined formats to a multi-record MTZ file (INPU USER). The functionality of the script \$CETC/uniqueify for completing a dataset has been extended, see unique.doc for details. A version for VMS, \$CETC/UNIQUIFY.COM, has been added. There is a patch file from v3.0 to v3.1.

A full list of changes since version 2.14 (announced in the last newsletter) is given below. Details of program fixes since version 3.1 have been placed on a [Problems Page](#). Perennial problems appear on our [FAQ](#). If you still have problems, then email us at ccp4@dl.ac.uk.

Finally, CCP4 will have a booth at the Commercial Exhibit Show of the XVII IUCr Congress in Seattle. Come and visit us!

Changes to the Suite

Building etc.

- configure: fftbig, tracer and amore have individual compile flags for AIX. OSF/1 optimisation reduced to -O1; problem with FFT SCALE card suggests compiler problem. configure tries to avoid messing with the `unsupported' directory if it's not present.
- HPUX: don't set (invalid) amore (non-)optimisation flags. The correct one reported to be needed, though.
- include/environ.def,default.def: reftmac files added.
- include/ccp4.setup-dist: Incorrectly (for any system) set _RLD_LIST, causing chaos with dynamic loading on Irix and similar systems.
- MAKEFILE.COM: typo in linking of reftmac corrected. Can now compile just one designated program with no checks.
- Fix problems building procheck on VMS.
- ccp4/etc/PROCHECK*.COM: hopefully the problem with PRODIR\$ is now sorted.
- XCCPJIFFY: SUNOS_5 added to Imakefile, SUNOS_4 hints in comments.
- Irix6 configuration is provided, your mileage may vary -- it did for me. Feedback still wanted.

Program changes

- General: A number of programs were found to close files on program termination. This may cause problems on some machines. Here is a list of progs. that were changed (not all programs have been checked): abyscale, act, agrovata, almn, contact, coordconv, crossec, extend, f2mtz, fft, fftbig, gensym, geomcalc, hgen, hklplot, icoefl, mlphare, mtz2various, mtztona4, npo, pdbset, peakmax, polarrfn, prolsq, protin, rfcrr, rotaprep, rstats, sortwater, undupl, vectors, volume.
- AMORE: The HKLM keyword has been reinstated. Default NM=25. Fix: all 0kl reflections of table were 0. Keyword END now accepted. Fix possible problems with inverse FFT due to dimensioning error. Fix failure to get FFT prime factors right. Less likely to overflow limits on HKL. SORTFUN can generate a table -- enables use of Es for TRAFUN. Maximum value for TRAFUN NMOL increased to 9. The ROTATE NSR keyword reinstated.
- CAD: Allow longer LABIN/LABOUT lines. Accept LABIN FILE <i> ALL. The scale factor is now applied to the correct columns and not to inappropriate ones like phase.
- CONTACT: The output has changed to: Res type, Res ID, Atom name.
- DM: correct multi-domain averaging bug. DM 1.6 now current version.
- ECALC: Use LABIN and LABOUT rather than FOBS/ISODIFF -- incompatible change. Introduced SCALE card to scale F and F2OR.
- F2MTZ: Deal with SHELX free R.
- FFT/FFTBIG: A new keyword has been added FILLIN which substitutes F2 for F1 if F1 is missing. This only works if the MTZ file has been properly converted to Missing Number Flags.
- FFTBIG: Compiler/runtime problems caused by undefined LNaN fixed.
- FHSCAL: Resolution card added so that data can be truncated in output file. Problem with eof on input corrected. LIST can take a no. > no. of refls. without crashing. Statistics for old style MTZ the same as new.

- FREERFLAG: Default semantics of free R column changed, along with its type: it's picked from a set of integers, allowing full cross-validation. An existing free R set may be completed when reflections are added (only if using missing data conventions). Systematic absences are now not assigned a freeR flag but are flagged as missing. When using the COMPLETE keyword the freeR flag information is taken from the MTZ file and not the keyword input.
- HKLPLOT: Redimensioning, improved output.
- HKLVIEW: minv replaced by minv3 because of routine renaming in modlib.f
- IPDISP: Support MacScience scanner. spdfil.mar_n now image type "bigovf", to indicate overflow table. ipdisp -m option points to spdfil.mar_n (bigend rather than littleend image)
- MAMA2CCP4: problem fixed wrt quotes in call to ccperr.
- MAPMASK: Symmetry expansion speeded up. New SIGMA option for SCALE (useful when contouring).
- MLPHARE: Phasing power was printed as 0.0 with PRINT AVE. SCALE FP option removed. Resolution in Angstroms is added to tables but not plotted on xlog-graphs. SCRIPT keyword will produce either a VMS COM file or unix script with refined heavy atom parameters. Ensure Angstrom coordinates are in PDB format - Orthogonal code 1. Will now produce NaNs for unphased reflections. Compiler problems caused by undefined CCPNUN fixed.
- MTZ2VARIOUS: Remove RESTRAIN input type (never actually necessary since restrain can specify the input format) now restrain reads MTZ. OUTPUT USER will now output DP and SIGDP. For USER the ordering of the columns is taken from the ordering on LABIN not LSPRGI table. If reflection record has NaN, by default reflection not output. MISS keyword will output them.
- MTZDUMP: Missing data identified through MNFs. File statistics table now gives single "Mean" and "Mean abs." which is calculated from all non-missing data in a column. LRES option added to list resolution. More flexible statistics available via STAT keyword. Missing reflections are printed as "MNF". The VALM keyword has been removed.
- MTZUTILS: Fix zeroed columns from hklin2 for reflexions not in hklin1 with UNIQUE option.
- NA4TOMTZ: Fix incorrect call of `qnan'. (Unlikely to have caused problems before missing data changes.)
- NCSMASK: A crude `sphere mask' can be constructed by making a PDB file with only a few atoms and specifying a large radius. Output map grid was always 3A round coordinates -- now 2A+.
- NPO: colours are now consistent with those defined in XCCPJiffy. CHAIN sub-keyword removed. Input handling improved a little. JOIN RADII HBOND now works as it should. Msecst has been reinstated. Can now have maps stacks like 3,5,7 or 5,10,15 etc.. RADII BOND ALL fixed.
- OVERLAPMAP: changed code for EXCLUDE option. Will have non-zero point in mapout if $|grid_point| < 1.0e-06$ in map1.
- PEAKMAX: symmetry related peaks at the edge of the map are not output.
- PDBSET: new keywords added EXCLUDE HEADERS, EXCLUDE HYDROGENS and SEQUENCE [PDB|SINGLE].
- PHISTATS: output improved + documentation.
- PLTDEV: `-pen' option now works (e.g. for colour PostScript). Colours are now consistent with those defined in XCCPJiffy.
- PROLSQ: the keyword RTEST has been reconstructed. Now you can only have RTEST -1|0. This means that you cannot recalculate structure factors i.e. RTEST +1.

- PROCHECK: Updated to v. 3.3.2 plus fixes. Now the first atom in *.new is printed only once. CONVAX has now been added to distribution.
- PROTOP: Didn't find contacts between the same residue on symmetry related molecules; error in packed symmetry number caused some contacts not to be picked up by prolsq; increase some dimensions; error in checking special inter-chain distances; no longer monitors short hydrogen bonds. Can now use atom names which represent ions e.g. Fe+2, U+6 etc. Non-protein residue atoms all assigned as side-chain not as main-chain and side-chain. Error in handling of multiple disordered planar groups fixed? Atoms now must have VDW radii defined. Hence, changes in dictionaries.
- RFCORR: arrays for PARSER redimensioned to avoid problem with TITLE. Polar angles are output correctly when have exact 2-fold axes.
- ROTAPREP: Format for SCALEPACK options changed from F8.1 to F8.0. Also, updated man page. Can now convert user-defined format to multi-record MTZ file (INPUT USER).
- ROTMAP: the table printed after the OUTPUT ... keyword was incorrect, now fixed.
- RSPS: Failed to handle SGs with Patterson symmetry groups 143-145.
- RSTATS: OUTPUT ASIN changed to LABOUT ALLIN. PROCESS FOBS should work. Ranges in F applied to Fobs only. Xloggraph altered and new column added ln(/) v resolution.
- SCALEIT: failed for more than 9 derivatives
- SFALL: Reads formfactors automatically if not specified, but still no way to distinguish Fe Fe+2, Fe+3 except to enter manually. Obsolete keywords removed.
- SIGMAA: `PARTIAL DAMP' didn't work. HLAC ... HLDC now output correctly. Map coefficients can now be constructed for missing reflections.
- SOLOMON: (solomon.c) parameters for rbfro1 were undefined.
- SORTMTZ: now not only does there need to be an equal number of columns in each file but the column types must be equal.
- STEREO: keyword CONSTRAIN could not be used, now working.
- SURFACE: trapped uninitialised variable, important for output.
- TRACER: prog. now works correctly.
- TRUNCATE: MTZ columns were potentially scrambled for data not from agrovata containing anomalous data.
- ZEROED: initialised some variables to stop IOT trap failure.
- Handling of `missing' data in files added to: abscale, absurd, agrovata, almn, amore, ecalc, f2mtz, fft, fftbig, fhscal, freerflag, hklplot, icoefl, mtz2various, mtzdump, mtztona4, mtzutils, na4tomtz, phistats, polarrfn, postref, rotaprep, reindex, rsearch, rstats, scala, scaleit, sigmaa, sfall, sortmtz, tffc, truncate, unique, udupl, wilson.

New programs

- CROSSEC: Interpolate cross sections and compute anomalous scattering factors (thanks to Don Cromer).
- GEOMCALC: Geometry calculations on molecule (some overlap with `angles')
- HGEN: Generates hydrogen atoms for a protein coordinate file with standard geometry.
- MAPROT: Map skewing, interpolating, rotating and averaging program.
- MTZMNF: Attempts to identify missing data in an MTZ file, and replaces the entry with a Missing Number Flag (e.g. NaN).
- REFMAC: Garib's Maximum Likelihood refinement program. Brand new from York.

- RESTRAIN, TLSANL: `restrain' does constrained anisotropic thermal parameter refinement using the translation/libration/screw-rotation model. Its feature list says: refinement with restrained geometry, rigid body, use of amplitude and phase observations, group anisotropic displacement parameters, disordered solvent corrections (thanks to Birkbeck group). `tlsanl' provides TLS analysis. Note restrain's keyword input isn't in the standard CCP4 form.
- SOLOMON: density modification (phase improvement) by solvent flipping.
- SORTWATER: Replaces `watersort'.
- xdlmapman, xdldataman: have been added thanks to Gerard K.

Library

- QNAN from the diskio routines (almost unused) changed from a function to a subroutine to avoid lossage with f2c's calling conventions.
- protin.dic: Last block wasn't read. WAT residue changed to contain more extensive list of waters (oxygen) only. All other atoms (ZN1,CU) have been removed. All atoms have VDW radii defined. Changed OXT and OE for NFORMYL and NACETYL so they are consistant with OT.
- protin_jp.dic: extra terminal atoms (OE, OXT) added to terminal residues similar to protin.dic.
- protin_vl.dic: extra terminal atoms added to terminal residues similar to protin.dic.
- symop.lib: Errors corrected in SGs 43, 112, 115, 160, 210, 228, 230 (a subset of errors reported with the X-PLOR version...)
- library.nw removed after too many complaints -- source is now library.c
- lib/data/atomsf.lib has comment section added; some programs which read it altered as a consequence.
- lib/data/crossec.lib: Data file for `crossec'.
- CCPOPN uses /dev/null and NL: interchangeably
- Fatal errors on diskio files now report the filename in question.
- Packed image routines (for mosflm, ipdisp): New version.
- MODLIB: MINV has been renamed to MINV3. A new routine has been added to calculate the inverses of nth order matrices MINVN.
- MTZLIB: Facilities for dealing with missing data items. [Null attributes for relational tuples? -- surely not!] Writing files with this facility should probably not be done since the programs aren't guaranteed to swallow the results.
- PARSER: Backslash is now a continuation character like `&' and `-'.
- RESTRAIN dictionaries: chiral_pep4.dic dna.dic nochiral_pep5.dic chiral_pep5.dic nochiral_pep4.dic
- RWBROOK: Add deuterium to atom list.
- SYMLIB: new routine HKLRANGE. Maximum no. of symmetry operators increased to 192. Now consistent with MTZLIB. Error in handling P2 C unique corrected.
- symop.lib: correct groups 120, 166, 210, 219, 220, 227, 228. File now subjected to the overdue check for fulfilling group conditions on the operators.
- x-windows/XCCPJIFFY/XCCPJiffy: colours changed to be consistant with those defined in NPO. Also, Xloggraph and xplot84driver pen colours are the same.
- Altered write statements so they do not exceed vms write limit of 132 characters.

Examples

- *.log-dist: all files except completedata and na4tomtz have been removed from the \$CEXAM/toxd area. Some files have been added to the \$CEXAM/unix/runnable area.
- All examples updated, inline with current progs.
- \$CETC/uniqueify: original cell now retained rather than rounded up version. Problem with some refls. not being assigned a FreeR flag overcome. -f switch modified to accept argument. -p switch added.
- \$CETC/UNIQUIFY.COM: added. VMS version of uniqueify
- Remove unnecessary test_brk.brk

Documentation

- Several .doc files corresponding to man .1 files were missing. diskio.doc was out-of-date.
- Various fixes/updates to documents for amore, ccplib, contact, libsymmetry, lsqkab, ncsmask
- FAQ: the plain text version with searching scripts (ccp4faq) are available. Just type 'ccp4faq'.
- HGEN: Included this time.
- MLPHARE: more information about using MAD data.
- MTZMNF: Clarified that columns not specified in LABIN are output unchanged.
- MTZUTILS: RZONE described correctly now.
- SOLOMON: MULSOLV keyword incorrectly referenced in the doc as SLVMUL.
- UNIQUE: Improved description of uniqueify script.
- Out-of-date manual-usletter.ps removed -- the current one should print OK on letter paper anyhow.
- Updates to mtzlib, npo, sfall, sigmaa documentation and the ccp4.1 man page.

Withdrawal

- COMPLETE removed, see unique.doc for replacement procedure.
- FAQ in \$CCP4 removed. It will be added to FAQ on the Web.
- ENVELOPE, FLATMAP, RESLICE, PRMAP moved unsupported status.
- LCF2MTZ, EXCHANGE and LCFLIB removed in the hope that everyone's using MTZ files -- still available from Daresbury if anyone is in need.
- PLUTO removed, should now use NPO.
- WATERSORT was broken, replaced by `sortwater'.

A data collection strategy option in MOSFLM

Andrew G.W. Leslie

MRC Laboratory of Molecular Biology,
Cambridge
CB2 2QH, U.K.
E-mail: andrew@mrc-lmb.cam.ac.uk

The advantages of collecting a dataset with a very high degree of completeness are well recognised. Even the loss of a few percent of the unique data can have a highly detrimental effect on molecular replacement solutions and density modification procedures. Nevertheless, it is not unusual to see a completeness of below 90%, even for the native data, in published structures. In order to simplify the task of collecting a complete dataset a new semi-automatic data collection strategy has been incorporated into the latest version of MOSFLM (5.30 or later).

The procedure assumes that data are being collected on a detector with a single rotation axis such as the Mar, R-Axis or Mac Science detectors, and is based on the OSCGEN/UNIQUE/COMPLETE set of programs originally written 12 years ago which formed part of the CCP4 suite of crystallographic software (although OSCGEN/COMPLETE have been removed from release 3.0 of CCP4). This problem has also been addressed in a general way by Nikonov & Chirgadze (1985) and a very similar approach has been described for the FAST area detector (Vickovic et al., 1994)

The program requires as input all of the parameters that are associated with processing a set of images, including the crystal symmetry and orientation which can be determined by auto-indexing a preliminary image. Given this information, the program will determine the starting phi value and the total phi rotation required to collect a dataset with the highest possible completeness. For crystal symmetries higher than triclinic, data will be lost in the cusp region unless the principle symmetry axis is inclined to the rotation axis and it is therefore an advantage to deliberately "mis-set" the crystal to eliminate the cusp region (the program will issue a warning message if this conditions has not been met).

While the Laue group symmetry determines the total rotation angle required to collect 100% complete data, in practice it is possible to collect a very high degree of completeness (>95%) with a total rotation considerably less than that required for 100% completeness (particularly if the crystal is oriented so that the rotation is not around a principle axis). For example, in an orthorhombic system it is possible to collect 95% complete data with a rotation of 60% or 90% with a rotation of 45° (in both cases using two non-contiguous segments). This can be particularly useful when collecting data at a synchrotron source and time constraints do not allow collection of the full rotation required by the Laue group. Indeed, there is a good argument for collecting data in this way even if sufficient time is available, because it minimises the risk of ending up with a seriously incomplete dataset in the event of equipment failure. Once the two (or more) most important segments have been collected, remaining beamtime can be used to fill in the remainder of the rotation range. To use this approach the user should specify the total

rotation angle (eg 60 degrees) and the number of segments (up to 3) to be used. The program will then automatically determine the phi values of the segments which will give the maximum possible completeness. Normally the segments will have equal sizes (in phi) but these can be specified by the user. Alternatively the user can specify the phi segments to be generated explicitly. In either case the program will give a detailed breakdown of the completeness, and multiplicity, both as a function of resolution and rotation angle, and the percentage of Friedel pairs is also tabulated. Finally, if incomplete data has been collected for one (or more) crystal(s) and data collection is about to start on a second (or subsequent) crystal, the program can be used to determine the optimum phi values to be used for the current data collection to give the most complete final dataset.

The statistics on completeness take no account of the oscillation angle per image used in data collection and in order to achieve this completeness it is crucial to use an oscillation angle that avoids reflections being lost due to spatial overlap. An option is therefore provided which will determine the maximum possible oscillation angle that can be used while allowing a specified percentage of spatially overlapped reflections (which can be zero). (This is an improved version of the TESTGEN option in the OSCGEN program). In a future release it is planned to allow the effect of a specified data collection scheme (phi range and rotation angles per image) to be tested, but this is not possible at present.

The cpu requirements are determined by the unit cell dimensions and the resolution of the data. On an SGI Indy (100MHz R4600 processor) the cpu time is just over 4 minutes to determine the best two 30° segments for an orthorhombic cell 100 x 100 x 100Å and a maximum resolution of 2.5Å. For large cells or high resolution the calculation can become very lengthy. However, as the completeness depends on the volume of reciprocal space swept out during the crystal rotation, the calculation can be speeded up substantially by reducing the size of the unit cell. For example, if all the cell dimensions are halved, the time required will be reduced by a factor of almost 8. Inevitably this will introduce small sampling errors but providing the smallest dimension of the "reduced" cell is 15-20 times the maximum resolution the sampling errors will be very small. The facility to speed up the calculation has been included in the latest version of MOSFLM (5.40). The option to maximise the completeness of the anomalous data (i.e. the maximum number of anomalous pairs) rather than the unique data has also been implemented in this version.

MOSFLM is available by anonymous ftp as follows:

```
ftp ftp.mrc-lmb.cam.ac.uk
logon as user "anonymous" password <anything>
cd pub
get mosflm_news
```

This describes the latest version available and installation details, and gives documentation on use of the STRATEGY and TESTGEN options. The ftp site can also be accessed via the web by pointing your browser [here](#).

References:

- S.V. Nikonov & YU. N. Chirgadze, *Acta Cryst A* **41**, 195-203 (1985).
I. Vickovic, K.H. Kalk, J. Drenth & B.W. Dijkstra, *J. Appl. Cryst* **27**, 791-793 (1994).

Electron density modification with Solomon

J.P. Abrahams

MRC Laboratory of Molecular Biology,
Hills Road,
Cambridge,
CB5 2QH, U.K.
E-mail: jpa@mrc-lmb.cam.ac.uk

The structure determination of F1 ATPase prompted the development of several new procedures for electron density modification [1]. This resulted in a program, called "Solomon", which has allowed the solution of at least 20 different structures over the last 18 months. The main purpose of this short notice is to draw attention to its inclusion in the CCP4 suite, and to give a brief overview of some of the concepts which, until recently, were unique to Solomon.

Density modification - Why and How

After experimentally determining a phase probability distribution for each of the measured structure factors with techniques like SIR, MIR, or MAD, one can use these distributions to calculate the "best" phases and figures of merit. In practice this is done for each structure factor by an angular integration of its phase probability distribution. For example, a centric structure factor can have two possible phases, and if the probability of it having a phase of 0° is 0.75, its "best" phase will be 0° , and its figure of merit will be $(0.75 - 0.25)$. The "best" phases and figures of merit of non-centric structure factors are calculated similarly in the complex plane.

However, usually more information is available. Rather than being information about individual structure factors, as measured from experimental differences in intensity, this information pertains to the way the structure factors interact with one another, and can be formulated as a set of constraints in real-space. Solomon imposes constraints on solvent flatness, non-crystallographic symmetry, and on the density distribution within the ordered parts of the crystal.

In practice, one first calculates a map using the "best" phases and figure-of-merit weighted structure factor amplitudes. Then one modifies the resulting map according to real-space constraints, and from this map calculates new, modified structure factor amplitudes and phases. The extra information contained within these modified structure factors is then combined with the original phase probability distribution to produce better estimates. The process can be iterated. To summarise, the advised procedure in conjunction with Solomon is:

1. Calculate initial map (FFT)
2. Introduce real-space constraints by modifying the map (Solomon)
3. Calculate structure factors of the modified map (SFALL)

4. Combine experimental phase information with phase information from the new structure factors (SIGMAA)
5. Calculate a new map using these updated "best" structure factors (FFT)
6. Stop or go back to step 2

Solomon is a program which modifies electron density maps and as such is responsible for only one step of the cycle outlined above, unlike programs like SQUASH [2] or DM [3], which essentially incorporate all the above steps, although quite different computational procedures may be used. Either approach has advantages and disadvantages. Solomon works in concert with several CCP4 programs (FFT, SFALL & SIGMAA), and it is distributed with a script which will correctly go through the cycle as outlined above. However, more complicated crystallographic problems will require a more sophisticated approach, in which case the user will have to amend the script.

Solvent density modification by Solomon

Solvent flattening is a standard technique for phase improvement and it involves three separate steps: location of the solvent, modification of the solvent, and the combination of the modified map with experimental phase information.

One should realise that the solvent will have a mean electron density which is very similar to that of the protein, and that solvent and protein are mainly distinguished by the relative featurelessness of the former, and the undulating landscape of the latter. Solomon is unique in its way of locating the solvent by determining a new map, in which at every grid point the local standard deviation of the original map is stored. The user will have to specify the radius within which the local standard deviation is determined, and it was found that a radius slightly larger than the maximum resolution of the map is optimal in most cases. After calculating such a map, Solomon suggests a contour level which will show the protein mask, given a certain solvent content, allowing inspection on a graphics workstation. Solomon uses this map to construct a solvent mask by excluding small islands of protein. If requested, this mask can be stored as an old-style "O" [4] or CCP4 mask and manipulated as any other mask. It is also possible to use solvent masks which were generated in other ways, or were edited by the user.

The solvent masks determined from the local standard deviation of the map have a higher resolution than masks determined by the method suggested by Wang [5]. The higher accuracy of the masks was found to be beneficial.

After locating the solvent, it can be modified. In conventional density modification, the density at every grid point of the solvent is replaced by the mean density of the solvent, but Solomon also allows different types of modification. The density within the solvent can be scaled as follows:

$$r_{\text{mod}} = (r_i - r_{\text{mean}}) \cdot k_{\text{flip}} + r_{\text{mean}} + \text{sadd} \dots\dots\dots (1)$$

(Where r_{mod} = modified electron density at grid point "i", r_i = density at grid point "i", r_{mean} = mean solvent density, k_{flip} = solvent multiplier, sadd = constant to be added to all solvent density.)

It is evident that setting the solvent multiplier k_{flip} to zero is equivalent to flattening the solvent. Setting it to a negative value will "flip" features within the solvent, and it turns out that doing so is desirable in many cases. The constant sadd can be used to reconstruct low resolution features of the electron density, and by setting it to a negative value, the

density of the protein can be "lifted" slightly above that of the solvent. This feature is used in conjunction with protein density truncation and structure factor reconstruction (see below).

The main benefit of flipping solvent features becomes apparent upon combining the modified structure factors with the experimental phase probability distribution. The intricacies of the recombination are beyond the scope of this short report, but an attempt will be made to give the reader a flavour of the sort of difficulties associated with this computation. It can be shown that, provided the sources information are independent, the optimal way of combining the information from model structure factor amplitudes with experimentally determined ones, is through sigmaA-weighting [6]. In the case of density modification, the information carried by the modified structure factors is not strictly independent from the experimental data, but with current methodology it is not possible to calculate how dependent the information actually is. In fact, the degree of independency will vary from structure factor to structure factor and will also crucially rely on the restraints imposed in real-space. As a result of treating the sources of information as independent, the recombined data will be biased. By flipping the solvent instead of flattening it, the modified structure factor amplitudes will be made more different from the original ones, and therefore they will appear to be more independent. Any (accidental) improvement of the phases will result in a more featureless solvent, and the next iteration there will be less density to flip. As a result, the solvent does get flatter as the flipping procedure is iterated, not so much because the solvent is biased to be flat, but rather because of other phase improvements. Solvent flattening cannot be iterated in a similar fashion, but it is entirely possible (and desirable) to flatten the solvent on the very last cycle of the solvent flipping procedure: the bias introduced at this point will not be propagated. It was found that the reduction in the R-factor between the experimental structure factor amplitudes and the modified structure factor amplitudes on the very last flattening cycle is an accurate indicator of the overall phase improvement.

There is a relationship between the solvent content of the crystal and the optimal value for kflip. The higher the solvent content, the less negative kflip should be. If the solvent content is about 30%, kflip should be set to -1.8 to -1.6, if the solvent content is about 50%, the optimum value is about -1, and if the solvent content is 70-75%, a kflip of zero seems to be optimal. Also the amount of averaging influences the optimal value of kflip: with two- or threefold averaging, one should set it at 60% to 80% of the value one would choose in the absence of non-crystallographic averaging, sixfold or higher non-crystallographic symmetry averaging is incompatible with solvent flipping and requires a value for kflip of zero.

Protein density modification with Solomon

Solomon allows two types of density modification in the protein region: averaging and truncation of density. Solomon calculates symmetry related density by cubic spline interpolation, and averages using single, rather than double interpolation. As with masks defining the solvent, masks defining non-crystallographic symmetry are old-style "O" masks or CCP4 masks, and can easily be manipulated with existing programs. Determination of the solvent mask is independent of the specification of the masks defining the non-crystallographic symmetry related density, but one can prevent the solvent mask from intruding into a symmetry mask by inclusion of the appropriate keyword in the script. Although symmetry masks are allowed to overlap, Solomon does provide the opportunity to remove the overlap between these masks.

In density truncation, grid points within the protein region which have a density below a certain specified threshold, are assigned a density equal to this threshold [7]. The result of density truncation of the protein region is that features of high density become sharpened relative to features of lower density. As such it is comparable to histogram matching techniques. However, the sharpening resulting from truncation seems to be beneficial even at resolutions at which the modification resulting from histogram matching is virtually non-existent. Because of this, density truncation was preferred over histogram matching. Other protocols for sharpening protein features were explored, but none of them were very successful, with one exception. In some cases it is better to set the density of truncated grid points to the mean density of all truncated grid points, rather than to the threshold density.

As a result of truncating the density within the protein region, the overall variance of electron density of the protein region relative to the solvent region decreases. This means that the density of the solvent region has to be scaled down since it was found desirable to maintain a constant ratio between the two. This is done automatically by Solomon if requested. Another result of truncation is that the mean density within the protein region will increase, relative to the mean density of the solvent region. Since it was found that this is undesirable if one is reconstructing missing structure factors, this can be corrected for by assigning a value to s_{add} in equation (1).

Reconstruction of missing data

It is vital to try to collect complete data sets, but there will always be some reflections which escape detection. The corresponding structure factors will be systematically absent from all the maps calculated, and as such will introduce a bias, since in density modification procedures one relies on the interaction between structure factors. If certain structure factors are left out, structure factors interacting through real-space with the missing data will incorrectly "assume" them to be zero. Therefore, the script distributed with Solomon allows one to replace missing structure factors with the structure factors of the modified map. In this way, the bias is removed, at the expense of creating potentially dangerous sources of error: since the missing structure factors are not restrained, they can take on unrealistic values in order to satisfy incorrect phase sets. However, if one only includes resolution bins with a completeness of 90% or more, it is quite safe to reconstruct missing structure factors, and in many cases it can be beneficial. Very low resolution terms can also be introduced quite safely.

Availability & Acknowledgements

Solomon is now available from CCP4, and if results obtained with Solomon are published, a reference to [1] should be made. I am grateful to Kevin Cowtan for making Solomon compatible with CCP4 masks.

References:

- [1] Abrahams, J.P. & Leslie, A.W.G. (1996) *Acta Cryst.* **D52**, 30 - 42
- [2] Cowtan, K.D. & Main, P. (1993) *Acta Cryst.* **D49**, 148 - 157
- [3] Cowtan, K. (1994) *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography* 31, 34-38
- [4] Jones, T.A., Zou, J.Y., Cowan, S.W. & Kjeldgaard, M. (1991) *Acta Cryst.* **A47** 110 - 119
- [5] Wang, B.C. (1985) *Methods Enzymolog.* **115**, 90 - 112
- [6] Read, R. (1986) *Acta Cryst.* **A42**, 102 - 116
- [7] Schevitz, R.W., Podjarny, A.D., Zwick, M., Hughes, J.J. & Sigler, P. (1981) *Acta Cryst.* **A37**, 669 – 677

ROTGEN -A Program for Rotation Images Simulation and Analysis

John W. Campbell, CCLRC Daresbury Laboratory

INTRODUCTION

ROTGEN is an X-windows based program for carrying out simulations of X-ray diffraction rotation images and analysing the unique data coverage for one or more crystal settings and series of rotations.

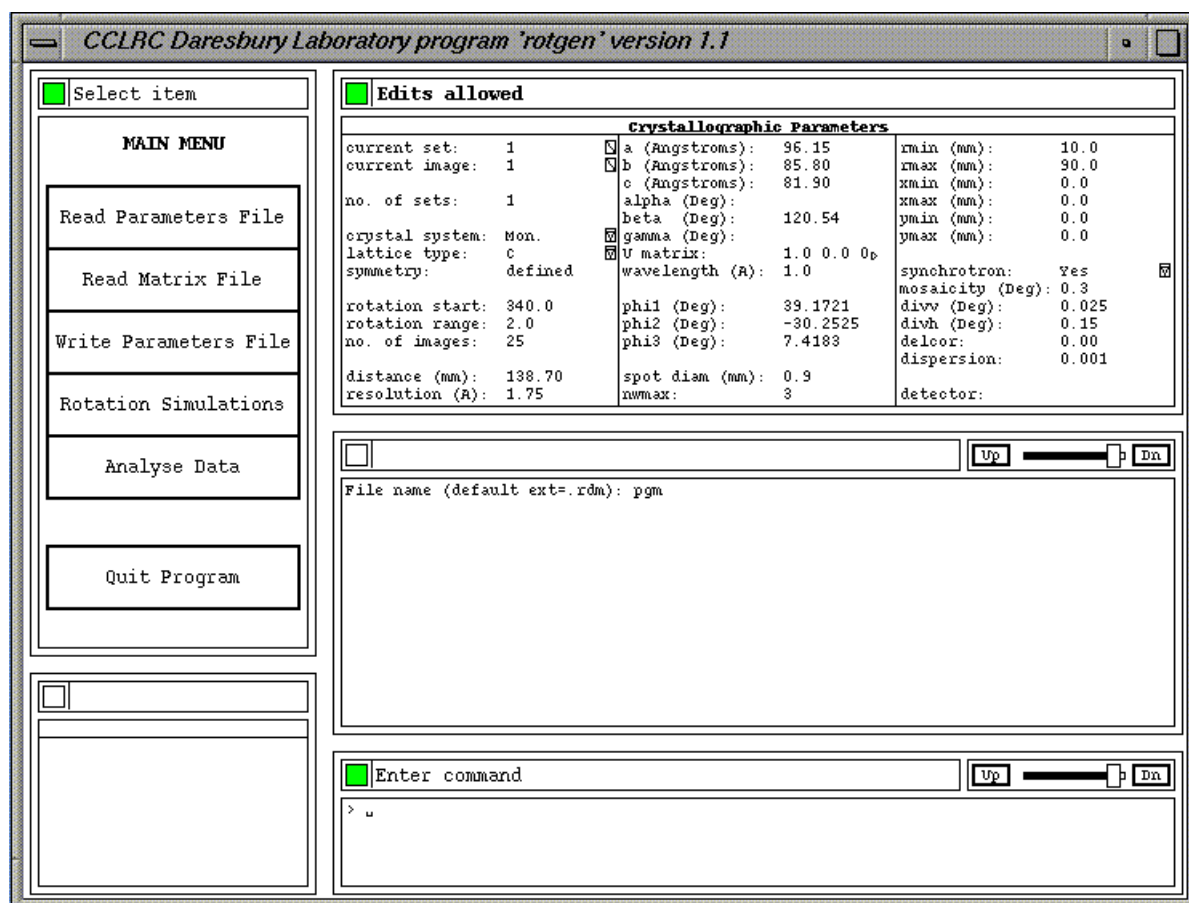


Figure 1. Main Screen of ROTGEN

The current program incorporates functions from the CCP4 (Collaborative Computational Project, Number 4, (1994)) and/or the MOSFLM suite (Wonacott, Dockerill & Brick (1980), Leslie (1992)) programs OSCGEN, UNIQUE and COMPLETE (the function of the latter now incorporated in MTZDUMP). The general treatment of the goniometry is derived from that used in the MADNES program (Messerschmidt & Pflugrath, 1987). The program makes also use of the XDL_VIEW toolkit (Campbell, 1995), library routines from the Daresbury Laboratory Laue Software Suite (Helliwell et al., 1989) and the CCP4 libraries.

The program is a first stage in an on-going development; the main aim in the next stages of development is to provide strategy information for efficient data collection using the

rotation method. If successful procedures are developed, it may well be desirable to integrate the use of the program, perhaps in a re-packaged form, with the PXGEN interface used on PX stations 9.5 and 7.2 at the Daresbury SRS (Kinder, McSweeney & Duke (1996). It is intended however that a stand alone version will also be retained.

The most immediate next steps are to allow the input of previously collected data (from an MTZ file) to be read into the program and included in the data coverage analyses and to provide quicker though slightly approximated analyses.

SCREEN LAYOUT

The basic screen layout for the program is divided into five areas.

Top Left

A menu area used for selecting the options currently available. The options within a menu are, in general arranged such that the normal order of access would be from the top downwards without necessarily using all the options available. The basic menu structure of the program is hierarchical and most menus have an item at the bottom which returns the program to the previous level with a label such as 'Return to Main Menu' or 'Return to Previous Menu'

Top Right

Main parameter table. This is an editable parameter table which contains the main crystal/detector parameters.

Bottom Left

Parameter table 2. This area is used for special purpose parameters relating to the current option selected.

Middle right

This area is the main text input/output window. This is used to output textual information from the program when required and also, at some stages, for prompt and reply sequences e.g. for inputting file names.

Bottom Right

Command window. This window is for inputting program options and parameters via a command line. Keyworded parameter values may be input or inspected. It is anticipated that processing functions will be added at a later stage.

Additional windows e.g. windows for displaying Rotation simulations or graphs are used at various stages of the program. These may be positioned independently of the main program window. Normally they will initially be positioned just to the right of the menu area.

OUTLINE OF THE MAIN OPTIONS

The following are the main options available in ROTGEN selectable via options in the program's main window and menu:

1. Setting up the parameters

The parameters in the main parameter table are assigned default values when the program is started. Individual parameters may be reset by editing the values directly in the parameter table. A new set of parameters may be read in from a keyworded parameters file using the 'Read Parameters File' option.

When the required parameters have been set up, they may be saved in a keyworded parameters file for future use using the 'Write Parameters File' option.

2. Updating the orientation parameters.

The 'Read Matrix File' option may be used to reset the parameters which define the crystal orientation using matrices produced following auto-indexing by programs such as REFIX or DENZO.

3. Rotation simulations

Two basic types of simulations are available, 'colour' simulations and 'interactive' simulations. In each case a window is created with a display area for the simulation, an area to list details of a selected reflection, a control panel and an area used for requesting hard copy (Postscript) output of the simulation. 'Colour' simulations show the show the rotation patterns with the spots colour coded in a number of different ways. 'Interactive' simulations are in black and white and have a slider which allow the user to investigate the effects of changing various parameters such as the oscillation range and mosaicity; spot labelling is also available.

4. Analyse data

This option enables the prediction of the reflections which would be recorded for the defined crystal sets and to analyse the data coverage in terms of the unique data for the space group, cell and resolution. The analysis may be done for either the current crystal set or for all crystal sets within the data set. The results may be presented in the form of histograms describing the data coverage or in terms of a pictorial representation of the reciprocal lattice sections. The analysis data equivalent to that displayed in the histograms is also written to the log file. The analyses automatically written to the log file exclude any spatially overlapped reflections.

THE COORDINATE SYSTEMS

ROTFGEN, in general, makes use of the set of laboratory axes as used by the MOSFLM suite of programs. However this does not mean, for example, that the rotation axis need necessarily be along the direction defined as 'Y' as one of the keyworded parameters enables the direction of the rotation axis to be specified as a vector with respect to the laboratory axes. Similarly, the detector plane need not necessarily be normal to the 'X' axis.

THE PARAMETERS AND PARAMETER FILES

The parameters used by the program form a keyworded data set. They fall essentially into two categories:

1. The crystal parameters such as the crystal system, lattice type, cell parameters, crystal orientation parameters and the resolution limit of the diffraction.
2. The parameters relating to the X-ray and detector system and X-ray source such the wavelength and the positioning of the detector.

The parameters include specifications for one or more crystal orientations and sets of images. Some parameters refer to the dataset as a whole and some may have separate values for each of the crystal sets.

A general Keyword Data Module (KDM) set of functions is used by ROTEGEN to define and handle its keyworded data set. It uses many the same concepts as the Laue Data Module (Campbell, Clifton, Harding & Hao, 1995) but the KDM routines are general whereas the LDM routines were written for a specific set of parameters. The KDM routines enable parameters of various types to be defined e.g. integer, real or character. Additional sets of routines are used to handle the input and output of KDM data and also symmetry data (KWD and KSM routines).

For most parameters a keyword is followed by a single value. The keywords are case insensitive and need not be given in full though a certain minimum number of characters must be given in each case. For a set based parameter the keyword may have a set number appended in square brackets e.g. ROTSTART[2] 50.0; if the set number is omitted, it is assumed that the value following the keyword refers to all sets e.g. RESOLUTION 2.0.

KDM Routines enable the monitoring of changes to KDM parameter values during the execution of a program. KWD/KDM based data files may contain indirect references to data in another file by giving a file name reference of the form '@filename'. These indirect files may currently be nested to a level of up to 20.

THE MATRIX FILES

This option is used to read in crystal orientation data from a matrix file for example after the orientation has been determined via an auto-indexing procedure. Currently two types of matrix file may be read, one as used by the MOSFLM suite and the other as used by the HKL/DENZO suite.

When a MOSFLM suite matrix file is input, the following parameters are updated:

- The cell parameters
- The U-matrix
- The missetting angles ϕ_1 , ϕ_2 , ϕ_3

For a DENZO file input, the following parameters are updated:

- The cell parameters
- The missetting angles ϕ_1 , ϕ_2 , ϕ_3

In contrast to the MOSFLM case, the missetting angles and cell need to be derived from the matrix read in as the required values are not stored in the file. The program derives the missetting angles on the basis of the current value of the U-matrix and this remains unchanged again in contrast with the MOSFLM case.

ROTATION SIMULATIONS

This option allows for the display of Rotation simulations based on the current parameter values for the currently selected crystal set and image. It has two modes of display, a colour mode and an interactive mode. In each mode there is a display area for the simulation, an area to list details of a selected reflection, a control panel area and an area used for requesting hard copy (Postscript) prints of the simulations.

Orientation of the Simulations

The simulations may be displayed in a number of different orientations. By default the orientation is the same as that from an IPDISP program display of a MAR image. An alternative program default may be set either via a command line option or an environment variable or via a parameter table during the running of the program. The orientation options are 'ipdisp', 'denzo' or 'oscgem' indicating compatibility with programs of these names.

Colour Display

'Colour' simulations show the rotation images with the spots colour coded in a number of different ways. The various options are selected via control panel items. Also, details may be listed for selected spots and hard copy Postscript file plots produced if required.

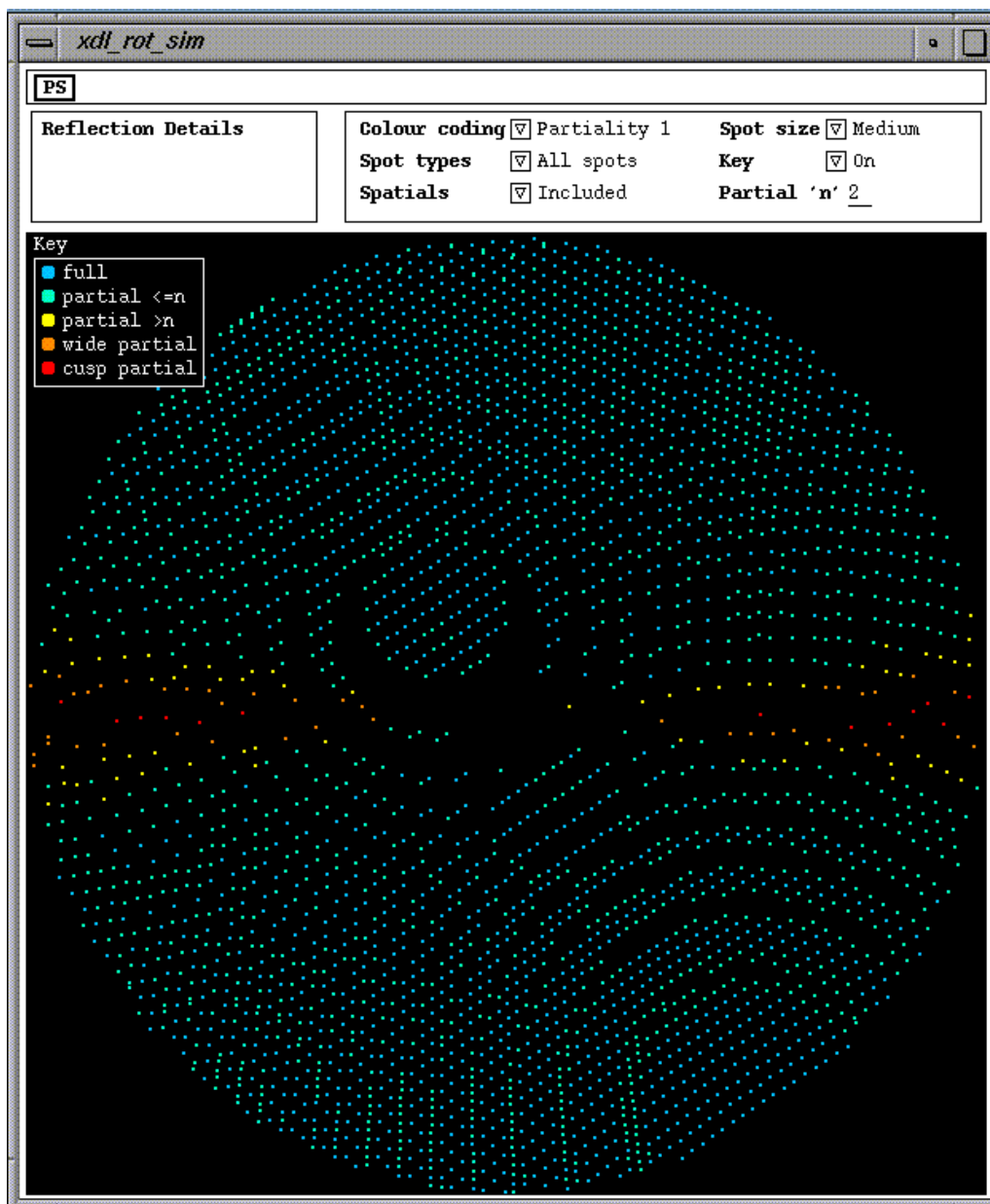


Figure 2. Example of a Colour Rotation Simulation

Control Panel Options

The colour coding depends on the option selected via the Colour coding choice menu on the control panel. The display is qualified by the settings of other control panel items. The options available are:

Fulls/partials

The spots are split into three categories, fulls partials and those which are on the edge of the cusp region and are partly visible. They are colour coded as follows:

```
blue:    fulls
yellow:  partials (up to maximum used in the prediction)
red:     cusp partials
```

Partiality 1

The spots are colour coded by their partiality as follows:

```
blue:    fulls
cyan:    partial <=n as set in the Partial 'n' panel item
yellow:  partial >n and up to the maximum used in the
         prediction
orange:  wide partial (greater than the maximum partiality
         used in the prediction)
red:     cusp partial
```

Partiality 2

The spots are colour coded by their partiality as follows:

```
blue:    fulls
cyan:    partial =2
green:   partial =3
yellow:  partial >3 and up to the maximum used in the
         prediction
orange:  wide partial (greater than the maximum partiality
         used in the prediction)
red:     cusp partial
```

All the same

All the recorded spots are displayed in the same (blue) colour

Overlap status

The spots are colour coded by whether or not they are spatially overlapped. The colour coding is as follows:

```
blue:    not overlapped
yellow:  spatially overlapped spot
```

The 'Spot types' choice menu on the control panel allows the user to choose one of the following three options:

- Display 'all' spots i.e. fulls and partials
- Display fulls only
- Display partials only

The display of spatially overlapped spots may be controlled via the 'Spatial' choice menu on the control panel. There are three options:

- Display all spots including spatially overlapped spots
- Exclude spatially overlapped spots
- Display the spatially overlapped spots only

The user may select one of three spot sizes, small medium or large for the display via the 'Spot size' choice menu on the control panel.

The display, of a key at the top left of the display area showing the colour coding, may be turned on or off via the 'Key' choice menu on the control panel.

The partial 'n' value, used in deciding which category a partial spot is to be included in for the Partiality 1 colour coding option a nodal can be reset via the Partial 'n' value item on the control panel. The value must be in the range of 2 to the maximum value used in the prediction.

Listing Spot Details

When the mouse Button1 is pressed with the cursor on a spot position, details of that spot will be listed in the spot details area. The following information is listed:

The indices

The coordinates

The detector xd, yd coordinates in millimetres from the centre of the pattern.

The partiality

For a fully recorded spot the message 'fully recorded' is given. For a partial the partiality is given and the offset from the current image of the start image for the spot is indicated. 'Wide' or 'Cusp' partials are flagged as such.

The overlap status

The spot will be flagged as a 'non-overlapped' or as an 'overlapped' spot.

The selected spot is marked by a surrounding white circle on the display area. The selection may be removed by clicking Button2 or Button3 of the mouse (or Button1 when the cursor is within the display area but not pointing to a spot). When Button1 is pressed, the nearest spot to the cursor is selected provided that the distance squared (pixels) to the spot is no more than 18.

Hard Copy

To get a hard copy plot in the form of a Postscript file, select the panel button marked PS in the hard copy request area at the top of the view object. A question and answer sequence is then followed using a panel i/o item to the right of the PS button. Invalid replies will give pop-up error notices. The hard copy output may be abandoned by pressing the Escape key when a prompt is displayed.

Interactive Display

When one of the interactive display options is selected, the rotation simulation is displayed as black spots displayed on a white background. Various options may be selected via the control panel items. These include options to vary various parameters via a slider, options to highlight various classes of reflection and options to label selected reflections. Also, details may be listed for selected reflections and hard copy Postscript file plots produced if required.

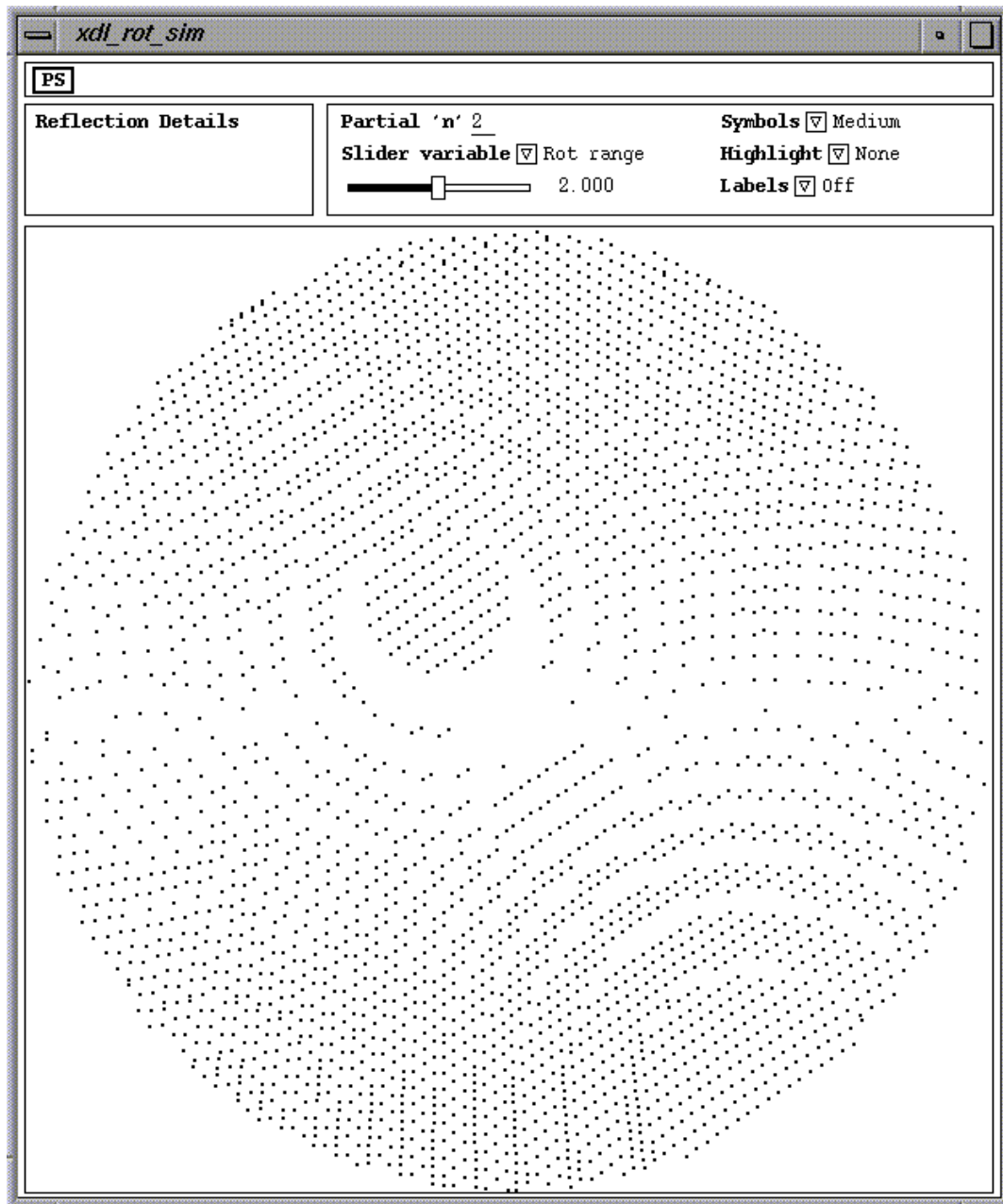


Figure 3. Example of an Interactive Rotation Simulation

Control Panel Options

A 'Slider variable' panel choice item determines which of the available variable parameters can currently be adjusted via the slider. The parameters which may be varied under slider control are:

- The rotation range for an image (in degrees)
- The mosaicity in degrees (Eta)
- The vertical beam divergence in degrees (Divv)
- The horizontal beam divergence in degrees (Divh)
- The wavelength dispersion $\Delta\lambda/\lambda$ times 1000 (Delamb)

The use of the slider was originally designed for use on colour displays which have 'writeable colour maps'. In fact 50 colours are used on the display with the spots being 'colour coded' by the value of the current variable parameter. The 50 colours are then set

to white or black depending on the position of the currently selected slider. If the slider is moved then the display is altered merely by changing the colours in the colour map thus giving a rapid change of pattern as the slider is moved. On displays which do not have writeable colour maps and on monochrome displays, the use of the sliders is less effective as the pattern needs to be redrawn each time a slider is moved.

The current value for the parameter being varied is displayed to the right of the slider. The overall range is defined by the calling program.

The 'Highlight' choice menu on the control panel may be used to enable various classes of spots to be highlighted. The options available are as follows:

No highlighting

This is the default; all spots are shown as black spots.

Highlight fulls

Fully recorded spots are indicated by blue vertical crosses; partials remain as black spots. (on monochrome displays, the corresponding symbols are drawn in black)

Highlight partials

Partials with a partiality \leq the value set via the Partial 'n' panel value item are displayed as blue vertical crosses; partials with a higher partiality are displayed as red crosses and cusp partials are displayed as green stars (or black on a monochrome display).

Note: When spots are highlighted and the soft limit sliders are used, the plot will be redrawn each time a slider is moved. The sliders are best used when no spots are highlighted.

The user may select one of three spot sizes, small medium or large for the display via the 'Symbols' choice menu on the control panel.

The partiality value defining the boundary between the two main classes of partials may be set via the Partial 'n' value item on the control panel. The value must be in the range of 2 to the maximum value used in the prediction.

The 'Labels' choice menu allows a number options for labelling. These are:

Labels off

No labels will be displayed and the labelling facility is currently disabled.

Labels on

Display any labels which have been added to the plot as described below.

If there are labels already displayed when a new labels option is selected, the existing labels will be redrawn as needed.

Note: When labels are displayed the plot may be redrawn if the slider is moved because a spot has disappeared or reappeared. The sliders are best used when no labels are displayed.

Listing Spot Details

When the mouse Button1 is pressed with the cursor on a spot position, details of that spot will be listed in the spot details area. The following information is listed:

The indices

The coordinates

The detector x_d , y_d coordinates in millimetres from the centre of the pattern.

The partiality

For a fully recorded spot the message 'fully recorded' is given. For a partial the partiality is given and the offset from the current image of the start image for the spot is indicated. 'Wide' or 'Cusp' partials are flagged as such.

The selected spot is marked by a surrounding red circle on the display area (black on a monochrome display). The selection may be removed by clicking Button1 when the cursor is within the display area but not pointing to a spot. When Button1 is pressed, the nearest spot to the cursor is selected provided that the distance squared (pixels) to the spot is no more than 18.

The spot details are also listed when a spot is labelled.

Labelling Spots

Providing that one of the labelling options has been selected via the 'Labels' choice menu, spots may be labelled on the plot. The labels are positioned under user control.

Hard Copy

To get a hard copy plot in the form of a Postscript file, select the panel button marked PS in the hard copy request area at the top of the view object. A question and answer sequence is then followed using a panel i/o item to the right of the PS button.

THE ANALYSE DATA OPTION

The 'Analyse Data' option enables the prediction of the reflections which would be recorded for the defined crystal sets and to analyse the data coverage in terms of the unique data for the space group, cell and resolution. The analysis may be done for either the current crystal set or for all crystal sets within the data set. The results may be presented in the form of histograms describing the data coverage or in terms of a pictorial representation of the reciprocal lattice sections. The analysis results equivalent to those displayed in the histograms are also written to the log file. The analyses automatically written to the log file exclude any spatially overlapped reflections.

When the analysis is initiated, the program first calculates a list of the unique reflections for the spacegroup and cell based on the resolution of the first crystal set defined. The program outputs a message for each crystal set for which predictions are being done. If there is more than one image in the set, then, a progress bar will be displayed to indicate the progress of the predictions throughout the images of the set. A cancel button is also displayed in case the user wishes to abandon the analysis before it has been completed.

The results may be presented in the form of histograms describing the data coverage or in terms of a pictorial representation of the reciprocal lattice sections.

Reciprocal Lattice Plot

This option enables the presentation of the results in the form of a pictorial representation of the reciprocal lattice and the extent of coverage by the predicted data. A specially written XDL_VIEW object is used to display, in pictorial form, the coverage of the unique part of the reciprocal lattice by the currently predicted data. The main display area shows a single section of the reciprocal lattice at a time showing which reflections are part of the unique area of the reciprocal lattice and which of these reflections have been predicted or observed. Different categories of reflections e.g. fulls and/or partials may be included in the plot and, where appropriate, merged or unmerged anomalous data pairs may be indicated. Spots predicted as spatial overlaps may be included or excluded as desired. Spots which are classified as unknown full/partial (e.g. derived from a merged dataset) are included in the display as fully recorded (and non-spatially overlapped) spots. If anomalous separated data is displayed then these spots may also be of unknown sign (see below for colour coding). The user may change the section displayed using the 'Up' or 'Down' buttons or the slider provided. Alternatively there is an option to cycle automatically through the sections. In all cases the section axis is along 'l'. On the $l=0$ section the a^* and b^* axes are displayed. Two forms of hard-copy Postscript plots may be produced; one is basically a copy of the currently displayed plot and the other enables plots of all the reciprocal lattice layers (at a selected interval if required) in a more compact format with nine sections being output on each page.

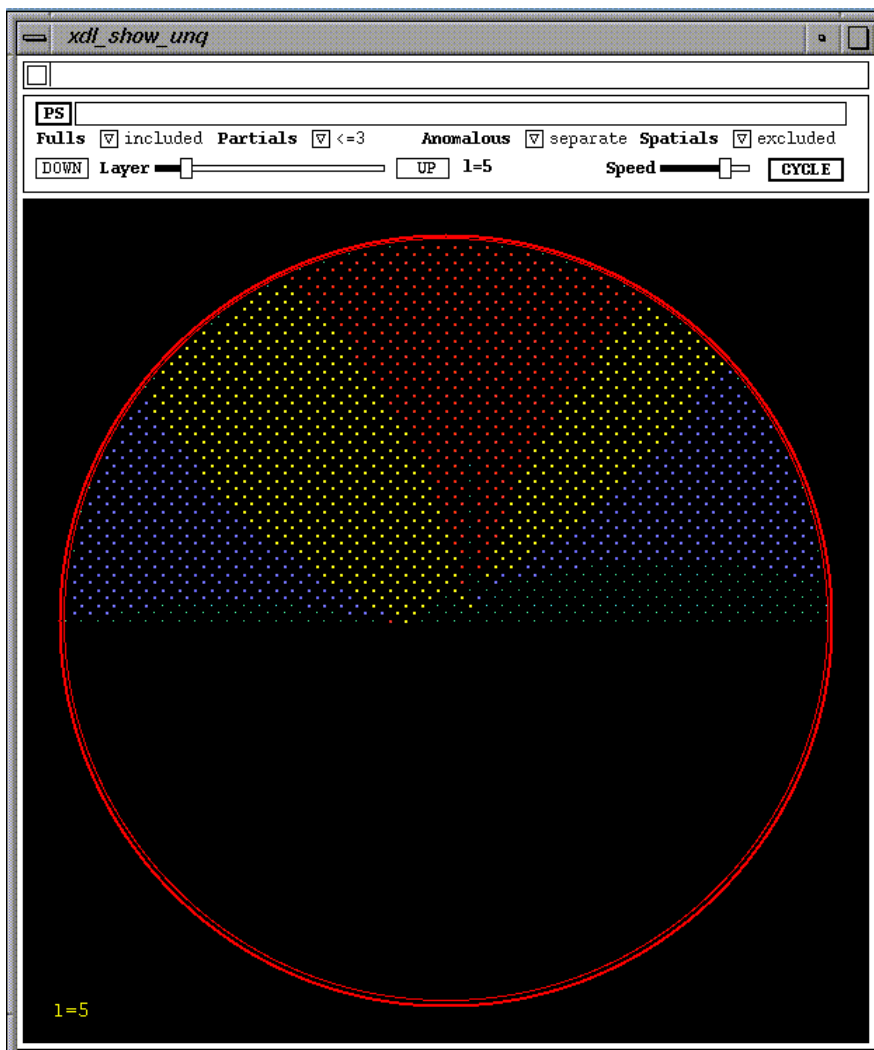


Figure 4. Example of 'Show Unique Coverage' View-object

The points on the unique part of the reciprocal lattice are marked by small greenish dots for those spots which have not been predicted/measured. Where anomalous data are merged, all predicted/measured reflections are represented by yellow dots. Where the anomalous data are to be separated the following colour coding is used.

yellow: Both an I+ and an I- reflection are present.
 red: Only the I+ reflection is present.
 blue: Only the I- reflection is present.
 white : Unknown sign (i.e. both I+ and I- are not present though I+ or I- may be present in addition to measurements flagged as being of unknown sign.

Graphical Plots

Three type of graphical plot may be produced. A 'Partials Toggle' selects which, if any, partial reflections are to be included in the analyses. The following choices are available:

- o Partials with partiality \leq nwmax (the default)
- o None
- o Partials with partiality = 2
- o Partials with partiality \leq 3

The Overlaps Toggle indicates whether spatially overlapped reflections are to be excluded from (the default) or included in the analyses. All the plots currently available are plots against resolution. The resolution range, from infinity to the resolution limit defined for the first crystal set, is divided into ten equal bins of $4 \cdot \sin^2(\theta) / \lambda^2$. The three plots are as follows:

23. Unique Data :This plot shows the percentage of coverage of the unique data within each resolution bin. Where partials are included two lines are drawn, a blue one indicating the coverage for the full reflections plus the included partials and a red one indicating the coverage for the full reflections only. The overall percentage coverage for each of these cases is given at the top of the plot. Where no partials are included a blue line indicates the coverage by the fulls data and the overall coverage by the fulls data is given at the top of the plot.

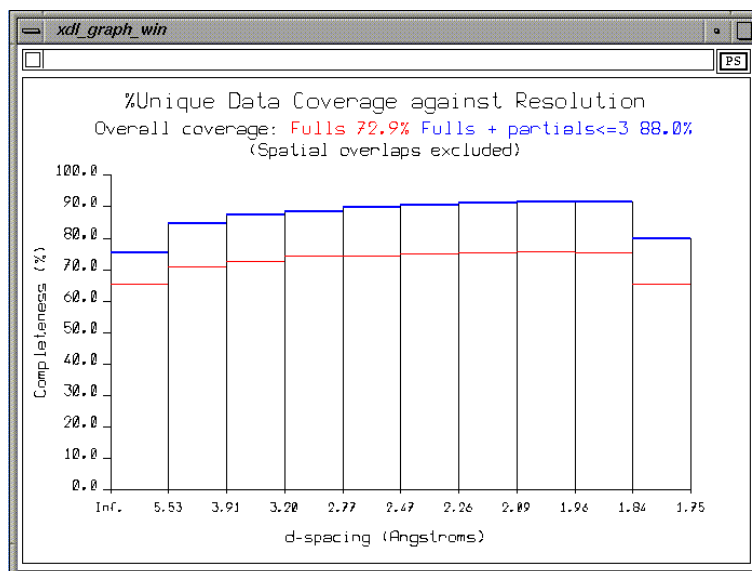


Figure 5. Example of a Unique Data Coverage Graph

24. Acentric Pairs This plot shows the percentage coverage of the unique acentric data within each resolution bin. A reflection must be present for each mate of a Bijvoet pair to be included. Where partials are included in the analysis two lines are drawn, a blue one indicating the coverage for the full reflections plus the included partials and a red one indicating the coverage for the full reflections only. The overall percentage coverage for each of these cases is given at the top of the plot. Where no partials are included a blue line indicates the coverage by the fulls data and the overall coverage by the fulls data is given at the top of the plot.

25. Multiplicity

This plot shows the average reflection measurement multiplicity within each resolution bin. Where partials are included in the analysis two lines are drawn, a blue one indicating the multiplicity for the full reflections plus the included partials and a red one indicating the coverage for the full reflections only. The overall average multiplicity for each of these cases is given at the top of the plot. Where no partials are included a blue line indicates the average multiplicity of the fulls data and the overall average multiplicity of the fulls data is given at the top of the plot.

The analyses automatically written to the log file exclude any spatially overlapped reflections. If the user also wishes to have analyses with the spatially overlapped reflections included, then these must be specifically requested.

THE LOGFILE

When the program ROTGEN is run, a log file is normally written.

When a new parameters file is read in, the values of the keyworded parameters are written to the log file.

When a data analysis has been carried out, then any keyworded parameter values, which have been changed since reading in a new parameters file or since the previous analysis, are written to the log file. Three analyses tables are output in a form which may be accessed by the CCP4 'xloggraph' program if desired. These are for unique data coverage, acentric pairs coverage and reflection multiplicity. All are shown as functions of resolution in 10 resolution bins. These analyses exclude spatially overlapped reflections.

Analogous analyses including spatially overlapped reflections will only be written to the log file if this was specifically requested when examining the analysis results in ROTGEN.

When the program is terminated, the current values of the keyworded parameters are written to the log file.

The use of the 'Rotation Simulations' option does not result in any additional log file output.

SUMMARY OF KEYWORDED PARAMETERS

This section gives a brief summary of the parameters available; a more detailed specification follows in the next section. Keywords followed by a pair of brackets [] may have separate values for each crystal set within the dataset if required.

General, Crystal System and Alignment

| | | |
|------------------|----------|---|
| Title | TITLE | A title for the dataset. |
| No. crystal sets | NUMSETS | Number of crystal sets defined within this dataset. |
| Crystal system | SYSTEM | The crystal system Tri, Mon, Ort, Tet, Hex, Rho or Cub. |
| Lattice type | LATTICE | Lattice type P, A, B, C, I, F or R. |
| Crystal symmetry | SYMMETRY | The space group symmetry followed by the space group number, space group name or symmetry operators (or 'clear' to clear the current symmetry). |

Cell, Resolution, Wavelength and Orientation Parameters

| | | |
|-------------------|---|---|
| Cell | A[], B[], C[], ALPHA[], BETA[], GAMMA[] | The cell parameters in Angstroms and degrees. |
| Resolution | RESOLUTION[] | The resolution limit in Angstroms. |
| Wavelength | WAVELENGTH[] | The wavelength in Angstroms. |
| U-matrix | UMATRIX[] | The basic 'U' setting matrix (9 values). |
| Missetting angles | PHI1[], PHI2[], PHI3[] | The missetting angles in degrees. |

The Rotation Ranges, Mosaicity and Spot Sizes

| | | |
|-------------------|-------------|--|
| Rotation start | ROTSTART[] | Rotation start angle in degrees. |
| Oscillation range | OSCRANGE[] | Oscillation range in degrees (+ve). |
| No. of images | NUMIMG[] | No. of images to collect. |
| Partials limit | NWMAX[] | Only consider partials occurring on up to this number of images. |
| Mosaicity | MOSAICITY[] | The mosaic spread in degrees. |
| Spot size | SPOT_SIZE[] | The spot diameter in mm. |

Detector Parameters

| | | |
|-----------------|--------------------------------|--|
| Distance | DISTANCE[] | The crystal to detector distance in mm. |
| Rmin | RMIN[] | Minimum radius on detector in mm. |
| Rmax | RMAX[] | Maximum radius on detector in mm. |
| Limits | XMIN[], XMAX[], YMIN[], YMAX[] | Detector limits on 'x' and 'y' from pattern centre (mm.). |
| Orientation | DET_ROTATIONS[] | Directions of three orthogonal detector rotation axes. |
| Axes | DET_AXES[] | Detector axis vectors |
| Horizontal axis | IAX_H[] | Horizontal axis number 1-3. |
| vertical axis | IAX_V[] | Vertical axis number 1-3. |
| Beam vector | BEAM_VECTOR[] | Beam vector. |
| Scan axis | SCAN_AXIS[] | Rotation axis vector. |

Other Source Parameters

| | | |
|-------------------|---------------|--|
| Synchrotron | SYNCHROTRON[] | Synchrotron - yes or no. |
| Dispersion | DISPERSION[] | The dispersion $\Delta(\lambda)/\lambda$. |
| Vertical divrg. | DIVV[] | Vertical divergence in degrees. |
| Horizontal divrg. | DIVH[] | Horizontal divergence in degrees. |
| Corr. dispersion | DELCOR[] | Correlated dispersion. |

REFERENCES

Campbell, J.W. "XDL_VIEW, an X-windows-based toolkit for crystallographic and other applications" J. Appl. Cryst. (1995) **28** 236-242

Campbell, J.W., Clifton, I.J., Harding, M.M. and Hao, Q., "The Laue Data Module (LDM) - a software development for Laue X-ray diffraction data processing" J. Appl. Cryst. (1995) **28** 635-640

Collaborative Computational Project, Number 4 "The CCP4 Suite: Programs for Protein Crystallography" Acta Cryst. (1994) **D50** 760-763

Helliwell J.R., Habash J., Cruickshank D.W.J., Harding M.M., Greenhough T.J., Campbell J.W., Clifton I.J., Elder M., Machin P.A., Papiz M.Z. and Zurek S. "The Recording and Analysis of Synchrotron X-radiation Laue Diffraction Photographs" J. Appl. Cryst. (1989) **22** 483-497

Kinder S.H., McSweeney S.M. and Duke E.M.H. "PXGEN: A general purpose graphical user interface for protein crystallography experimental control and data acquisition" J. Appl. Cryst. (1996) submitted for publication.

Leslie A.G.W. In CCP4 and ESF-EACBM Newsletter on Protein Crystallography (1992) No. 26, DRAL Daresbury Laboratory, Warrington, WA4 4AD, England

Messerschmidt A. and Pflugrath J.W. "Crystal orientation and X-ray pattern prediction routines for area detector diffractometer systems in macromolecular crystallography" J. Appl. Cryst. (1987) **20** 306-315

Wonacott A.J., Dockerill S. and Brick P. "MOSFLM program" (1980) Unpublished Notes.

Making the most of your search model

Gerard J. Kleywegt

Department of Molecular Biology

Biomedical Centre, Uppsala University

Uppsala - Sweden

In cases with high sequence homology, Molecular Replacement (MR) is usually a trivial exercise. But, as the structural homology gets lower and lower, solving the structure by means of MR becomes increasingly difficult. Conventional wisdom holds that an RMSD of about 1.5Å between the search model and the actual structure constitutes the limit at which MR can still be used. Fortunately, the MR programs get better, and crystallographers think of new and clever tricks, thereby pushing the limits of the method further and further. On the program end, Axel Brünger has introduced PC-refinement into X-PLOR [1], which is a useful tool to sort the rotation function boys from the rotation function men, and to improve the orientation of the entire molecule or of individual domains prior to the translation function. More recently, he has introduced a direct rotation function [2] which may help in cases where other approaches fail. Liang Tong has developed a locked rotation function [3], which uses knowledge of the rotational relationships between NCS-related molecules to increase the signal-to-noise ratio of the rotation function. More recently, he has also implemented a locked translation function [4]. Jorge Navaza, finally, has carefully redone the mathematics of the fast rotation and translation functions, and his package AMoRe [5] has solved numerous problems which could not be solved by other programs.

In this article we will address some of the tricks that the crystallographer can use in order to produce a more suitable search model in cases where the standard approach is not immediately successful. We shall also discuss some software tools that can be of help in this process.

*** Which parts of the model are likely to be conserved ?**

MR is based on a comparison of the set of intra-molecular vectors (rotation function) and inter-molecular vectors (translation function) calculated from the search model on the one hand, and the Patterson function (or peaks) calculated from the data pertaining to the unknown structure on the other. Therefore, every atom (or rather: atom pair) in the search model that is roughly in the correct position will contribute to the signal, whereas every atom in a wrong position will, at best, contribute to the noise and, at worst, contribute to false signals. If the MR is not immediately successful, a critical appraisal of which parts of the search model are likely to be conserved is in order. The crystallographer can use a large set of heuristics and tools to aid in this process:

- a sequence alignment of the known and unknown protein (and other related proteins) is very valuable. For instance, we know that insertions and deletions in the sequence are most likely to occur in loops, and not in the middle of a regular secondary structure element. Also, in the hydrophobic core of a protein, sidechain torsion angles are often conserved between related residues. For instance, in case of a Phe/Tyr substitution it's often a good guess to assume that the chi1 and chi2

angles will be conserved (in other words, this sidechain need not be cut back to the CB atom).

- assessing the quality of the search model may provide further clues. If it is the result of a low-resolution structure determination, a quick analysis of temperature factors and the Ramachandran plot may reveal regions which are of poor quality. Any such regions are probably best left out of the search model.
- in general, the overall fold and the hydrophobic core are much better conserved than, say, the conformation of loops or of sidechains that point into the solvent.

Parts of the model which are not likely to be conserved at all can be removed; if only the sidechain conformation is uncertain, residues can be cut back to the CB or CG/OG/SG atom. Temperature factors can either be set to a uniform value, or they can be retained. The latter is probably to be preferred in most cases, since it automatically downweights the contribution of regions with high temperature factors. An exception can be made if inspection of the temperature factors shows them to be completely unreliable (there are structures in the PDB which have been refined without temperature-factor restraints at resolutions as low as 3Å [6], the result being amusing to some, but not particularly useful). Another option is to multiply all temperature factors by a constant in order to reproduce the overall B-factor obtained from a Wilson plot, while retaining the pattern of low and high temperature factors within the molecule.

*** When the going gets tough ...**

In the case of non-crystallographic symmetry (NCS), MR becomes more difficult because a monomeric search model constitutes less and less of the scattering matter in the asymmetric unit as there are more and more molecules in it. If one is lucky enough to have a tetrameric search model, and the data indicates that the unknown might form similar tetramers, the MR can of course be carried out with the intact tetramer. If the rotation function fails to give solutions, PC refinement of the individual monomers may be of help. If that fails too, a dimer can be used and only if that fails to give a solution as well should a single monomer be tried. If the search model is monomeric, but the unknown is not, the locked rotation and translation functions may be of use. In the case of multiple-domain structures, MR calculations can be carried out with the intact molecule, with different subsets of the domains, and with individual domains. PC refinement may be necessary in these cases before attempting to solve the translation function.

Often nowadays different crystal forms of a protein are obtained. Clearly, it is worthwhile to try and solve each of these separately. Usually, the one with the best data and the smallest degree of NCS can be solved first. Since the starting phases after MR are often very poor, it is worthwhile to solve one or more of the other crystal forms as well, so that multiple-crystal electron-density averaging can be used to improve the maps.

*** Multiple models**

One "trick" which we have found to work very well in cases where all attempts to solve the structure had failed is to use multiple, superimposed search models (see also [9]). Such models are often available:

- there may be several related protein structures available (we have used this to solve the structure of cellular retinoic-acid-binding protein [7], for example);
- there may be multiple copies of a single protein structure available (for example, when the structure was determined in different labs and/or in different spacegroups,

or when the structure contained NCS, or when different complexes have been solved);

- there may be a "family" of NMR structures, each of which individually may be too poor to crack the problem. We have used this to solve several structures (for instance acetyl-CoA-binding protein, ACBP [8][6]) which could not be solved by using any of the individual models.

This approach works surprisingly often, probably since it implicitly weights the well-conserved parts of the models higher than the more variable (or more poorly determined) parts. Nevertheless, it is often necessary in practice to combine the use of multiple models with the editing out of parts which show a large conformational spread. The approach that has worked for us on several occasions is to superimpose all available models, remove any obvious outliers (in particular for NMR ensembles), to remove regions of large variability, and to cut back all sidechains to CB or CG. Also, we have always used uniform temperature factors since (a) their use would be tautological for multiple models, (b) Bs from different X-ray structures are not well comparable, and (c) NMR models don't have Bs associated with them at all.

When using multiple models, the contrast between correct solutions and incorrect ones is usually very low, so that it is very important to ascertain the correctness of solutions. Also, the correct solutions may be far down the list: for ACBP, rotation function solution 41, and translation function solution 21 turned out to be the correct ones ... It is rather puzzling why it should be so difficult to solve MR problems when a 100% homologous NMR structure is available. In the case of ACBP the problem turned out to be a rather large RMSD of $\sim 2.2\text{\AA}$ on CA atoms. The differences, however, were not random: the helices had all undergone rigid-body translations along their axis compared to the NMR model.

* Tools

Various programs from Uppsala can be used in the process of producing search models:

(1) O [10]:

- to superimpose multiple models
- to delete zones of residues
- to apply conservative substitutions
- to do homology modelling (if one is really desperate; we know of no case where a homology model has been used successfully to solve an MR problem that could not be solved otherwise)

(2) MOLEMAN2 [11], a trivial but useful MOLEcule MANipulating program:

- to do all sorts of temperature-factor manipulation
- to create poly-Gly, poly-Ala or poly-Ser models
- to calculate the distribution of intra-molecular distances (this can be used to find a suitable Patterson integration radius; often $\sim 75\%$ of the largest diameter of the molecule covers 90% of the vectors)

(3) LSQMAN [6], a superpositioning and (NCS-) analysis program:

- to superimpose multiple models
- to find the "central" model and outliers

- to find regions of large variability (e.g., in terms of torsion angles, which is often more informative than the use of RMSD values [6])

(4) SEAMAN [11], a program specifically written for SEArch-model MANipulation which handles multiple models as well as single models:

- to delete loops, turns, general zones, atoms with high temperature factors
- to apply some minimalist substitutions with sidechain-torsion angle conservation
- to change zones to poly-Gly, poly-Ala or poly-Ser

Other Uppsala tools that may be of use for MR work in general include:

- **MOLEMAN2**: to apply rotation and translation function solutions to a model
- **XPLO2D** [11]: to generate plots of 1D and 2D translation functions from X-PLOR
- **XPLO2D**: to generate 2D contour plots from X-PLOR direct rotation function searches [2]
- **MAPMAN** [12]: to convert CCP4 map and X-PLOR "3dmatrix" files from rotation and translation function calculations into O-style maps for viewing on the graphics (can also be used to extract planes or slices from a map for plotting with **O2D** [11])
- **PACMAN** [11]: to do a quick "centre-of-gravity packing" check for possible solutions
- **PACMAN**: to estimate how large a fraction of the cell can possibly accommodate the centre-of-gravity of the molecule, depending on its radius, the spacegroup and the cell dimensions (this can be used to find if spacegroup P213 is more or less likely than P23, for instance)
- **RAVE** [13] [14]: programs for electron-density averaging (single and multiple domain; single and multiple crystal form) plus tools for improving NCS and inter-crystal operators, mask manipulation, etc.
- **auto_amore.csh**: a C-shell script that uses the CCP4 version of AMoRe and PACMAN to do the MR calculations, saving the user a lot of tedious file editing, and making it easy to screen dozens (or hundreds) of rotation and translation function solutions. This script has recently been modified so that the rotation function can be carried out either with Fs or with Es.

* AVAILABILITY

All programs other than O are available to academic users free of charge (from the O ftp server). For more information about these programs, contact GJK

(gerard@xray.bmc.uu.se). For more information about O, contact T. A. Jones

(alwyn@xray.bmc.uu.se).

* REFERENCES

- [1] Brünger, A.T. (1990). *Acta Cryst.* **A46**, 46-57.
- [2] DeLano, W.L. and Brünger, A.T. (1995). *Acta Cryst.* **D51**, 740-748.
- [3] Tong, L., & Rossmann, M.G. (1990). *Acta Cryst.* **A46**, 783-792.
- [4] Tong, L. (1996). *Acta Cryst.* **A52**, in the press.
- [5] Navaza, J. (1994). *Acta Cryst.* **A50**, 157-163.
- [6] Kleywegt, G.J. (1996). *Acta Cryst.* **D52**, in the press.
- [7] Kleywegt, G.J., Bergfors, T., Senn, H., Le Motte, P., Gsell, B., Shudo, K. & Jones, T.A. (1994). *Structure* **2**, 1241-1258.
- [8] Zou, J.Y., Kleywegt, G.J., & Jones, T.A. (1996). To be published.
- [9] Read, R.J. (1990). *Acta Cryst.* **A46**, 900-912.

- [10] Jones, T.A., Zou, J.Y., Cowan, S.W. and Kjeldgaard, M. (1991). *Acta Cryst.* **A47**, 110-119.
- [11] Kleywegt, G.J. (1996). Unpublished programs.
- [12] Kleywegt, G.J. & Jones, T.A. (1996). *Acta Cryst.* **D52**, in the press.
- [13] Jones, T.A. (1992). In "*Molecular Replacement*", pp. 91-105, CCP4.
- [14] Kleywegt, G.J. and Jones, T.A. (1994). In "*From First Map to Final Model*", pp. 59-66, CCP4.

*This file was converted with TextToHTML - (c) 1995 [Logic n.v.](#) - [Kris Coppieters](#)
But edited manually by Gerard himself ;-)*

Density growing : a method for local improvement of electron density maps

A.G.Urzhumtsev

IGBMC du CNRS,
Parc d'Innovation,
BP 163,
67404 Illkirch,
c.u. de Strasbourg, France

and

IMPB of RAS,
Puschino,
Moscow Region,
142292, Russia

1. Introduction. Dummy atoms and phase improvement.

The procedure (Watenpaugh et al. (1973)) of refinement of an atomic model in reciprocal space and phase recalculation from the refined model has been proven as a powerful tool to correct model errors and to build their missing parts. However, in general, the density corresponding to the missing atoms is twice weaker than the one for the atoms included in the phase calculation (Raman, 1959; Main, 1979). Since not-yet-interpreted regions of the map usually correspond to less ordered parts of the molecule and the density in these regions is already weak, the problems of density improvement for such regions is quite acute.

Agarwal and Isaacs (1977) suggested to improve the density by placing dummy atoms in the unit cell with a following «refinement» of such model. Dummy atoms are placed respecting the known interatomic distances in the peaks of the initial map. The phases calculated from the refined dummy-atoms model are used together with the observed amplitudes to obtain new, improved maps.

This idea was further developed by Lunin & Urzhumtsev (1984, 1985) by suggesting new schemes for model building and model modification and by the development of the likelihood approach for phase error estimation in order to calculate properly weighted improved syntheses. Their mixed models contain both «molecular atoms» from the interpreted parts of the map and dummy atoms for uninterpreted ones. Dummy atoms interact with the molecular atoms, however they are really artificial in the sense that they can be at any distance to each other, which is essential in order to improve the map details.

The method was successfully applied in several cases at middle resolution (2.3-3.5Å). A development of this approach for the high-resolution case (Lamzin & Wilson, 1993) resulted in the package ARP distributed through CCP4 (CCP4, 1994). *[Actually, ARP is not at present part of the CCP4 suite -Ed.]*

However, even with all these modifications, the method had three major problems :

- a) incorrect phase error-estimations for the refined models;
- b) impossibility to retrieve missing density;
- c) impossibility of phase extension starting from a resolution lower than approximately 3.5 Å.

The first problem has been recently solved by Lunin & Skovoroda (1995). To solve the second problem, Agard and co-authors (Wilson & Agard, 1993; Baker et al., 1993; Bystroff et al., 1993) used simultaneously the properties of atomicity and connectivity. Their new approach realised through the program PRISM (Wilson & Agard, 1993) allows to build «bridges» between regions with reasonably strong density and to recover missing density in the middle. However, the presence of some initial peaks of density near the poorly defined regions is necessary.

2. Density growing : method description.

The initial idea of the current approach comes from the fact that the map, calculated from the models with omitted parts, does not show these parts at full height. In order to increase the height, an attempt can be done to fill the non-interpreted region by some scatters and to try to optimise their diffracting power automatically.

The density growing is based on the assumption that an electron density map of resolution around 2-3 Å already has been interpreted everywhere except in a small region with poor density. A fine grid (with the step of 1 Å or less) covers this region, and every node is occupied by a dummy atom. The scattering factors for these atoms can be similar to the ones of hydrogens, and their initial parameters should correspond to a weak density (temperature factors may be 10-40 Å² and occupancies are between 0.1-0.5).

As it was shown by Lunin & Urzhumtsev (1984), refinement of dummy models is an important step in the phase improvement. For the density growing, the most important step is the refinement of occupancies of dummy atoms. Usually, a significant part of the dummy atoms does not want to grow, and their occupancies decrease. These dummy atoms (from 50 to 90%, depending on the case) are removed from the model, and the position of other ones is refined without any restraints.

While removing the dummy atoms with very small occupancy changes very little the R-factor, it is important in order to decrease the number of parameters for the following iterations. The usual rule of having a high ratio of Ndata/Nparameters is also applicable in this case. Starting from the occupancy refinement is useful because there is only one parameter per dummy atom at this step, and before position refinement (3 parameters per atom) the number of dummy atoms is already reduced according to the refined occupancy values.

This procedure of refinement of occupancy and atomic positions is automatically repeated varying the low-resolution end. It is important to include all low-resolution data at the initial iterations in order to form the correct shape of the «field» and to remove the dummy atoms from the zones of the bulk solvent. On the contrary, it is useful to finish the iterations with the data of only higher resolution by restricting the low-resolution end to 6-7 Å. Usually, 5-6

iterations are enough, with 20-30 refinement cycles par iteration. Since no geometrical criterion is applied, the procedure is relatively fast.

Structure factor amplitudes calculated from the refined model are used to estimate phase errors in the corresponding calculated phases (Lunin & Skovoroda, 1995). These values are used to prepare the weighted electron density map (Read, 1986; Urzhumtsev et al., 1996). For several experimental data sets the new map showed improved density in the cases of :

- a) missed side chain(s) ;
- b) poor density at the N-terminal of the main chain ;
- c) poor density for an inhibitor in protein-inhibitor complexes.

An analysis of the refined position of dummy atoms did not show any direct way of using them as a guide for the construction of the missing part of the molecule model.

This procedure can also be used instead of simple calculation of the omit-maps (Bhat & Cohen, 1984). The program automatically removes the model atoms from a given region, fills this region with dummy atoms and calculates new structure factors from such mixed model after refinement of the dummy component (which in the case of classical omit-map has zero contribution). Analysis of experimental cases has proven that new maps have an additional information in comparison with the usual omit maps. An intermediate step of the simulated- annealing refinement of the known part of the model (Hodel et al., 1992) before refinement of dummy atoms was also tried. In this case, the new improved map also had more details than the omit one. Retrieval of the density for the well-defined atoms removed from the analysed region can be used as a criterion of the quality of the new map.

The experience showed that «growing» the density (start from small values of occupancy allowing some of them to increase) is more efficient than «suppressing» density (start from large values of occupancy). It is interesting to note that the occupancy of some «grown» dummy atoms of the hydrogen type can be increased to quite large values (2-4).

The same approach is currently being tested for density improvement in the case without an assigned atomic model.

3. Density growing : computer realisation.

The procedure starts from an atomic model and a file of structure factors (FOBS, SIG, TEST values) and produces a map, locally improved in a region of the unit cell. The region is defined by a set of spheres of a prescribed radius around given points. The overlapping parts of spheres contains only one grid corresponding to the sphere «coming first». The procedure consists of 5 steps and is based on X-PLOR (Brünger, 1992). A few parameters at the top of corresponding example files (command files or X-PLOR input files) should be changed to tune the procedure. The main steps are the following.

1. Generation of a dummy-atoms model (program : sphexpl)

Input :

- initial atomic model
- set of points defining the region for improvement (PDB-type lines)

- file with control data

Output :

- atomic model; atoms inside the region are removed (occupancy $q = 0$)
- a set of dummy atoms placed in a grid in the defined region

2. Calculate structure factors FPART from atomic model (X-PLOR)

Input :

- atomic model; atoms inside the region are removed (occupancy $q = 0$)
- structure file for the atomic model (the same used before for the model refinement)
- structure factors file with FOBS, SIG, TEST
- file with unit cell parameters and symmetries (X-PLOR lines)

Output : - structure factors file in the X-PLOR format with atomic structure factors as FPART

3. Dummy-atom model refinement (X-PLOR)

Input :

- model of dummy atoms obtained at step 1
- file of structure factors calculated at step 2
- file with unit cell and symmetries

Output :

- file of structure factors calculated from a mixed model ("real" and dummy atoms)

4. Phase error estimations (rflexpl, Urzhumtsev et al., 1996)

Input :

- file of calculated structure factor at step 3.

Output :

- file of structure factors with estimated fom's

5. Map calculation (X-PLOR)

Input :

- file of structure factors after the step 4

Output :

- a 2Fo-Fc map calculated following Read (1986) with more correctly estimated parameters (Lunin & Skovoroda (1995))

As it was mentioned above, step (1*) of the simulated-annealing refinement (Hodel et al., 1992) of the omit model can be introduced after step (1). This step can be useful in the case of omit-map calculations to check the already built model. Since the omit-model after step (1) has the same list of atoms (omitted atoms are marked by zero occupancy), one can use exactly the same X-PLOR input files for the refinement which were used before.

The programs SPHERXPL and RFLEXPL are written on standard FORTRAN-77 and are available by request from the author. The corresponding e-mail address is sacha@igbmc.u-strasbg.fr.

The author thanks Mr.L.Moulinier for his help in testing the method, Drs. V.Lunin and A.Podjarny for many discussions on the dummy atoms ideas, and Dr. D.Moras for his support of the work. The work was supported by the CNRS through the UPR 9004, by the Institut National de la Santé et de la Recherche Médicale and the Centre Hospitalier Universitaire Régional.

References:

- Agarwal, R.C. & Isaacs, N.W. (1977) Proc.Natl.Acad.Sci.USA, **74**, 2835-2839.
- Baker, D., Bystroff, C., Fletterick, R.J. & Agard, D.A.. (1993) Acta Cryst., **D49**, 429-439.
- Bhat, T.N. & Cohen, G.H. (1984) J.Appl.Cryst., **17**, 244-248.
- Brünger, A.T. (1992) X-PLOR. Version 3.1. A System for X-ray Crystallography and NMR. Yale University, USA.
- Bystroff, C., Baker, D., Fletterick, R.J. & Agard, D.A.. (1993) Acta Cryst., **D49**, 440-448.
- Collaborative Computational Project, Number 4 (1994) **D50**, 760-763.
- Hodel, A., Kim, S.-H. & Brünger, A.T. (1992) Acta Cryst., **A48**, 851-858.
- Lamzin, V.S. & Wilson, K.S. (1993) Acta Cryst., **D49**, 129-147.
- Lunin, V.Yu. & Skovoroda, T.P. (1995) Acta Cryst., **A51**, 880-887.
- Lunin, V.Yu. & Urzhumtsev, A.G. (1984) Acta Cryst., **A40**, 269-277.
- Lunin, V.Yu., Urzhumtsev, A.G., Vernoslova, E.A., Chirgadze, Yu.N., Nevskaya, N.A. & Fomenkova, N.P. (1985) Acta Cryst., **A41**, 166-171.
- Main, P. (1979) Acta Cryst., **A35**, 779-785.
- Read, R. (1986) Acta Cryst., **A42**, 140-149.
- Raman, S. (1959) Acta Cryst., **12**, 964-975.
- Urzhumtsev, A.G., Skovoroda, T.P. & Lunin, V.Yu. (1996) J.Appl.Cryst., sent to editor.
- Watenpaugh, K.D., Sieker, L.C., Herriott, J.R. & Jensen, L.H. (1973) Acta Cryst., **B29**, 943-956
- Wilson, C. & Agard, D.A.. (1993) Acta Cryst., **A49**, 97-104.

Tcl/Tk-based programs. I. CONFOR : user-friendly converter for crystallographic data files

L.M.Urzhumtseva and A.G.Urzhumtsev

IGBMC du CNRS,
Parc d'Innovation,
BP 163,
67404 Illkirch,
c.u. de Strasbourg, France

and

IMPB of RAS,
Puschino,
Moscow Region,
142292, Russia

1. Introduction.

Different crystallographic program packages do not always use the same formats of the crystallographic data files of structure factors, electron density maps or atomic models. However, the data interchange between them is important in order to optimise the process of structure determination, and corresponding service software has been developed in different laboratories for this purpose, for example by Heinemann (1992) for the atomic coordinates and by the Uppsala group (Kleywegt & Jones, 1996) for different data.

A major part of the current crystallographic software is based on the idea of macrolanguage and macrocommands and not on the modern principles of menus. The first reason for this "delay" is the significant computation time for many of the crystallographic programs which makes ineffective their run in interactive mode. The second reason is the variety of different types of computers with different graphic interfaces where the packages are used.

Tcl/Tk (Ousterhout, 1993) is an universal interpreting language in which scripts can be ran with no changes at different types of computers. These scripts allow to run any combination of command and executable files depending on their results and pressed keyboard buttons. The Tcl/Tk is both very powerful, convenient for use and easy to program.

A program to reformat the crystallographic data files can be taken as an example of the software, which could be used at any computer and which is short enough to be ran in the interactive mode. Moreover, during its run a user needs to make several choices : name and type of the files, etc. These conditions allow to include such program in a list of ones for which an effective interface based on the Tcl/Tk language can be developed. Such series of programs is under development in the UPR de Biologie Structurale, and some first results are reported below.

2. Program description.

The program CONFOR allows to change the format for three types of crystallographic data: structure factors, electron density maps and atomic coordinates. The choice of the type of data is defined by a radiobutton on the screen menu (Fig.1), and the following choice of the file is carrying out by clicking in the list of the files like in Microsoft programs.

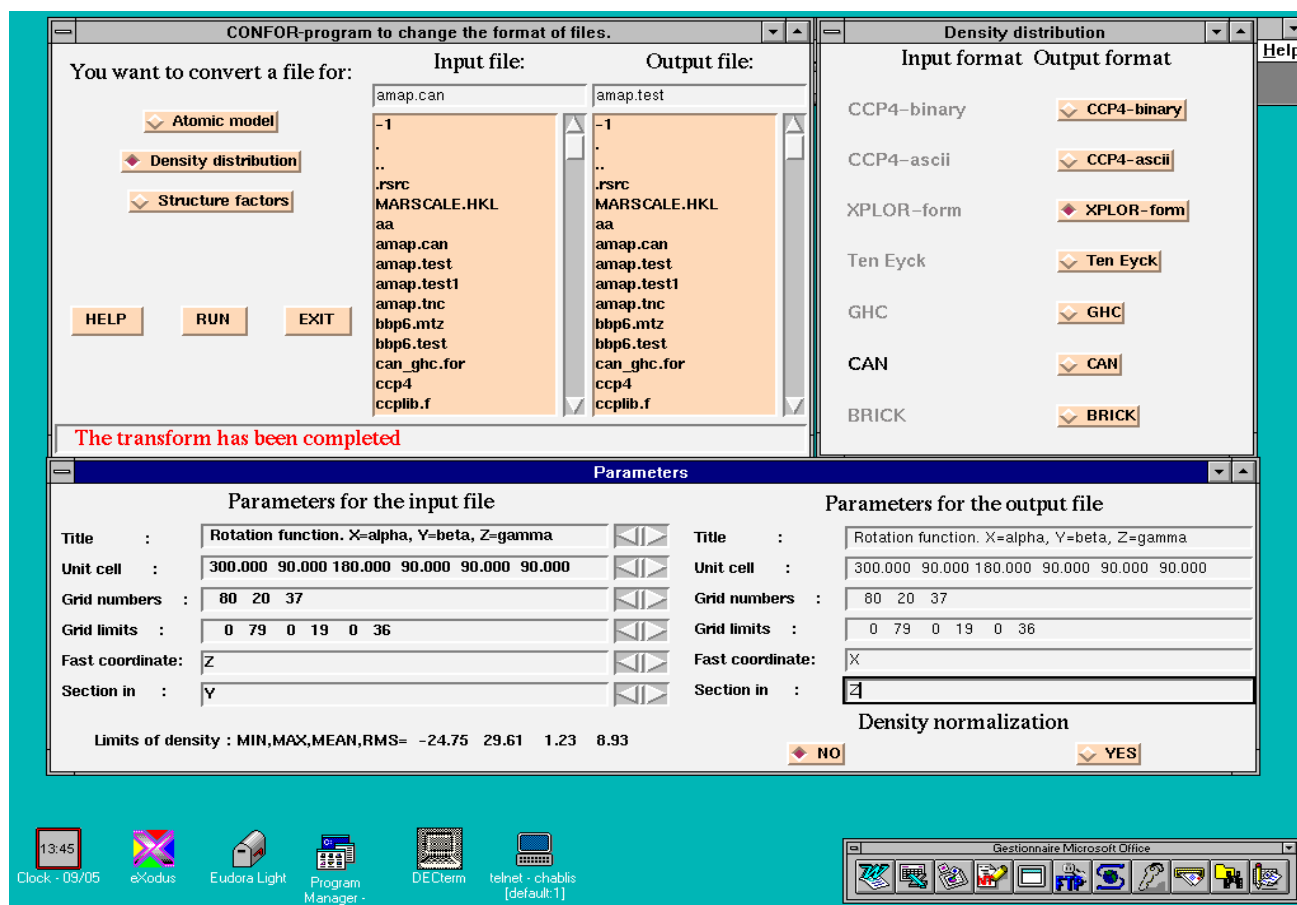


Fig.1. Example of the CONFOR application to change the format of the map file from CAN to X-PLOR with changing the map orientation. [Click on figure for full size image.]

For the map files, the program can normalise the function of the electron density distribution and change the orientation and the limits of the map. For the structure factors, the data values can be reassigned, e.g., for the X-PLOR format (Brünger, 1992) FPART amplitudes can be assigned to FCALC or to FOBS ones. The current version of the program does not yet allow to make any other data transformation, which is supposed to be included in the following versions.

The format of the input file is defined automatically, and the format of the output file and necessary data transformations are chosen from the menus.

The program calculates major statistical information for the data (limits of the model coordinates, minimal and maximal density values, etc).

Any number of files of any type of data can be processed during the same run of the program.

The current version of CONFOR recognises the following input/output formats:

- a)
 - for the structure factors
 - X-PLOR (Brünger, 1992)
 - MARSCALE (Kabsch, 1988)
 - UF (Urzhumtsev et al., 1989, Vernoslova & Lunin, 1993)
- b)
 - for the electron density maps
 - X-PLOR (Brünger, 1992)
 - CCP4-binary (CCP4, 1994)
 - CCP4-ascii (CCP4, 1994)
 - "bricks" (Jones, 1978; Jones et al., 1991)
 - Ten Eyck (1973)
 - GHC (Cohen, personal communication)
 - CAN (Vernoslova & Lunin, 1993)
- c)
 - for the atomic models
 - PDB (Bernstein et al., 1977)
 - PROLSQ (Hendrickson & Konnert, 1980)

The subroutines for other formats, e.g., for the MTZ format of structure factors file, are under development.

The standard Tcl/Tk libraries are necessary to interpret the CONFOR script which calls the basic programs written on standard FORTRAN-77. The CCP4-formatted files are read/written using the CCP4 standard subroutines. The program is available by request from the authors. The corresponding e-mail address is sacha@igbmc.u-strasbg.fr.

The authors thank Dr.J.M.Wurtz for showing a wonder world of the Tcl/Tk language and Drs. D.Moras and A.Podjarny for their interest to the work. We thank the staff of the CCP4 and Dr. T.A.Jones and his collaborators for the permission to use several their basic subroutines. The work was supported by the CNRS through the UPR 9004, by the Institut National de la Santé et de la Recherche Médicale and the Centre Hospitalier Universitaire Régional.

References:

- Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Jr., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977) *J.Mol.Biol.*, **112**, 535-542.
- Brünger, A.T. (1992) X-PLOR. Version 3.1. A System for X-ray Crystallography and NMR. Yale University, USA.
- Collaborative Computational Project, Number 4 (1994) *Acta Cryst.*, **D50**, 760-763
- Jones, T.A. (1978) *J.Appl.Cryst.*, **11**, 268-272.
- Jones, T.A., Zou, J.Y., Cowan, S.W. & Kjeldgaard, M. (1991) *Acta Cryst.*, **A47**, 110-119
- Heinemann, U. (1992) *J.Appl.Cryst.*, **25**, 326-327.
- Hendrickson, W.A. & Konnert, J.H. (1980) In *Biomolecular Structure, Function, Conformation and Evolution*, R.S.Srinivasan, ed., **1**, 43-57. Oxford: Pergamon.
- Kabsch, W. (1988) *J.Appl.Cryst.*, **21**, 916-924.
- Kleywegt, G.J. & Jones, T.A. (1996). *Acta Cryst*, **D52**, in press
- Osterhout, J.K. (1993) *Tcl and the Tk Toolkit*, Addison-Wesley Publishing Company, Inc.
- Ten Eyck, L.F. (1973) *Acta Cryst.*, **A29**, 183-191.
- Urzhumtsev, A.G., Lunin, V.Yu. & Vernoslova, E.A. (1989) *J.Appl.Cryst.*, **22**, 500-506.
- Vernoslova, E.A. & Lunin, V.Yu. (1993) *J.Appl.Cryst.*, **26**, 291-294.

Relationships between some rotation descriptions for molecular replacement procedure

A.G.Urzhumtsev

IGBMC du CNRS,
Parc d'Innovation,
BP 163,
67404 Illkirch,
c.u. de Strasbourg, France

IMPB of RAS,
Puschino,
Moscow Region,
142292, Russia

The final purpose of the molecular replacement method (Rossmann & Blow, 1962; Crowther & Blow, 1967; Rossmann, 1990; Rossmann & Arnold, 1993; and references therein) is to find the position of a given search model in a given crystal. Another meaning of the term "molecular replacement", as a procedure to improve density distribution by averaging through non-crystallographic symmetry, is also widely used, however, to avoid any misleading, is not considered further in these notes. The search of the internal crystal symmetry (self-rotation analysis) as a tool to improve a molecular image is not considered here either.

A search atomic model is defined by a set of three-dimensional coordinates given in some absolute or external Cartesian coordinate system. The result of the molecular replacement search is a transformation which should be applied to these coordinates. The transformation can always be presented as a rotation plus a translation.

There is a number of different descriptions of rotations in three-dimensional space reflecting point of view of different authors. These variants differ by the definition of :

- a) the rotating object (body or coordinate system),
- b) direction of the rotation (clockwise or counter-clockwise, and how "clockwise" is defined),
- c) type of angles (Eulerian or polar),
- d) choice of parameters inside a given type of angles.

This variety is reasonable since different descriptions are convenient in different situations. Unfortunately, this very rich choice does not allow an easy understanding for users of different programs. Any of these definitions is easily used inside one program but their exact understanding is necessary when trying to compare the results of different programs. Actually, what the majority of crystallographers needs is simply to have a rotation matrix, R , which should be applied to the coordinates of the model to place it in the unit cell as the rotation function suggests. However, these matrices are not always derived explicitly while a large variety of options (a)-(d) are used in different programs and articles.

The program CONVROT is developed which calculates the rotation matrix for different types of rotation description and vice versa. Such approach solves two problems. First, it gives directly a rotation matrix which can be applied to the model coordinates in order to obtain a model orientation accordingly to the molecular replacement solution. Secondly, it allows to recalculate the rotational angles from any to any of these systems by use of 2N subroutines and not N*N, where N is the number of different rotation systems : N subroutines to go from rotation angles to the rotation matrix and N - for opposite operations. This also allows easy addition of any new rotation system.

The program can define all symmetry related rotation angles for a given one when the orthogonalisation agreement and the symmetry operations are known.

The program can work with the polar systems introduced by Rossmann & Blow (1962), Dodson et al. (1966), Crowther (1972, 1973) and the eulerian angles systems introduced by Rossmann & Blow (1962), Nordman (1966), Huber (1969), Crowther (1972, 1973), Lattman (1972), Navaza (1994). The system of direction cosines is also used (see, for example, Diamond, 1993). These system are used, in particular, in ALMN (Crowther, 1973; CCP4, 1994), in PROTEIN (Steigemann, 1974), in MERLOT (Fitzgerald, 1988), in X-PLOR (Brünger, 1992) and in AMoRe (Navaza, 1994).

Tables 1-2 give some basic formulae used by the program, the singular cases are not included. [Click on tables to obtain postscript version.] Table 1 uses the following notation:

$$\begin{array}{l}
 \text{RZ}(x) = \begin{vmatrix} \cos x & -\sin x & 0 \\ \sin x & \cos x & 0 \\ 0 & 0 & 1 \end{vmatrix} \\
 \\
 \text{RY}(x) = \begin{vmatrix} \cos x & 0 & \sin x \\ 0 & 1 & 0 \\ -\sin x & 0 & \cos x \end{vmatrix} \\
 \\
 \text{RX}(x) = \begin{vmatrix} 1 & 0 & 0 \\ 0 & \cos x & -\sin x \\ 0 & \sin x & \cos x \end{vmatrix}
 \end{array}$$

Table 1. Rotation angles and corresponding rotations. Matrices $R_X(\xi)$, $R_Y(\xi)$, $R_Z(\xi)$ for the primitive rotations are defined in the text. Most distributed molecular replacement program packages are indicated which use the corresponding rotation system. PROTEIN rotation matrices are inverted X-PLOR ones.

| System | Limits for the triclinic group | Alternative solution | Rotation matrix as a product of matrices for elementary rotations | Program packages |
|--|---|--|--|------------------|
| <i>Poincaré</i> Rossmann, 1962 | $0 \leq \phi < 2\pi$ $0 \leq \psi < \pi$ $0 \leq \kappa < 2\pi$ | $\phi = \phi + \pi$ $\psi = -\psi + \pi$ $\kappa = -\kappa$ | $R_Y(\phi) \cdot R_Z(-\psi) \cdot R_Y(-\kappa) \cdot R_Z(\psi) \cdot R_Y(\phi)$ | X-PLOR |
| <i>Poincaré</i> Dodson et al., 1966 Crowther, 1972 | $0 \leq \phi < 2\pi$ $0 \leq \psi < \pi$ or $0 \leq \psi < 2\pi$ $0 \leq \kappa < 2\pi$ | $\phi = \phi + \pi$ $\psi = -\psi + \pi$ $\kappa = -\kappa$ | $R_Z(\phi) \cdot R_Y(\psi) \cdot R_Z(-\kappa) \cdot R_Y(-\psi) \cdot R_Z(-\phi)$ $R_Z(\phi) \cdot R_Y(\psi) \cdot R_Z(-\kappa) \cdot R_Y(-\psi) \cdot R_Z(-\phi)$ | ALMN AMoRe |
| <i>Direction cosines</i> | $0 \leq \kappa < 2\pi$ $-1 \leq \cos \alpha \leq 1$ $-1 \leq \cos \beta \leq 1$ $-1 \leq \cos \gamma \leq 1$ $\cos^2 \alpha + \cos^2 \beta + \cos^2 \gamma = 1$ | $\kappa = -\kappa$ $\cos \alpha = -\cos \alpha$ $\cos \beta = -\cos \beta$ $\cos \gamma = -\cos \gamma$ | can be written through recalculated angles | AMoRe |
| <i>Eulerian</i> Rossmann, 1962 | $0 \leq \theta_1 < 2\pi$ $0 \leq \theta_2 < 2\pi$ $0 \leq \theta_3 < 2\pi$ | $\theta_1 = \theta_1 + \pi$ $\theta_2 = -\theta_2$ $\theta_3 = \theta_3 + \pi$ | $R_Z(-\theta_3) \cdot R_X(-\theta_2) \cdot R_Z(-\theta_1)$ | X-PLOR |
| <i>Eulerian</i> Lattman, 1966 | $0 \leq \theta_1 < 4\pi$ $0 \leq \theta_2 < 2\pi$ $-2\pi \leq \theta_3 < 2\pi$ | $\theta_1 = \theta_1 + 2\pi$ $\theta_2 = -\theta_2$ $\theta_3 = \theta_3$ | $R_Z[-(\theta_1 - \theta_3)/2] \cdot R_X(-\theta_2) \cdot R_Z[-(\theta_1 + \theta_3)/2]$ | X-PLOR |
| <i>Eulerian</i> Crowther, 1972 | $0 \leq \alpha < 2\pi$ $0 \leq \beta < 2\pi$ $0 \leq \gamma < 2\pi$ | $\alpha = \alpha + \pi$ $\beta = -\beta$ $\gamma = \gamma + \pi$ | $R_Z(-\gamma) \cdot R_Y(-\beta) \cdot R_Z(-\alpha)$ | ALMN MERLOT |
| <i>Eulerian</i> Hoodman, 1966 | $0 \leq \phi < 2\pi$ $0 \leq \theta < 2\pi$ $0 \leq \psi < 2\pi$ | $\phi = \phi + \pi$ $\theta = -\theta$ $\psi = \psi + \pi$ | $R_X(-\phi) \cdot R_Y(-\theta) \cdot R_X(-\psi)$ | |
| <i>Eulerian</i> Huber, 1969 | $0 \leq \psi < 2\pi$ $0 \leq \theta < 2\pi$ $0 \leq \phi < 2\pi$ | $\phi = \phi + \pi$ $\theta = -\theta$ $\psi = \psi + \pi$ | $R_Z(\psi) \cdot R_X(\theta) \cdot R_Y(\phi)$ | PROTEIN |
| <i>Eulerian</i> Navaza, 1994 | $0 \leq \alpha < 2\pi$ $0 \leq \beta < 2\pi$ $0 \leq \gamma < 2\pi$ | $\alpha = \alpha + \pi$ $\beta = -\beta$ $\gamma = \gamma + \pi$ | $R_Z(\alpha) \cdot R_Y(\beta) \cdot R_Z(\gamma)$ | AMoRe |

Table 2. Rotation matrices and angle recalculation for a general case. Following notations are used for brevity : $c\theta = \cos\theta$, $s\theta = \sin\theta$, $c\theta_1 = \cos\theta_1$, $s\theta_1 = \sin\theta_1$, etc; $c^2\theta = \cos\theta \cdot \cos\theta$, $s^2\theta = \sin\theta \cdot \sin\theta$, etc; $R_{jk} = R(j,k)$ is the k -th element of the j -th line of the rotation matrix; $\xi = \text{atan2}(q_1, q_2)$ stands for the angle ξ for with $\sin\xi = k \cdot q_1$ and $\cos\xi = k \cdot q_2$, where k and is a positive normalising factor (definition c corresponds to the inserted FORTRAN function atan2)

| System | Matrix | | | Solution for a general case |
|--|---|---|---|--|
| <i>Polar</i> Reasimann, 1962 | $c\kappa + (1-c\kappa) \cdot \omega^2 \cdot \mu \cdot c^2\theta$ $\omega\mu \cdot [(1-c\kappa) \cdot c\mu \cdot c\theta + \omega \cdot \omega \cdot s\theta]$ $-(1-c\kappa) \cdot \omega^2 \cdot \mu \cdot c\theta \cdot \omega \cdot \omega \cdot s\theta$ | $\omega\mu \cdot [(1-c\kappa) \cdot c\mu \cdot c\theta + \omega \cdot \omega \cdot s\theta]$ $c\kappa + (1-c\kappa) \cdot \omega^2 \cdot \mu$ $-\omega\mu \cdot [(1-c\kappa) \cdot c\mu \cdot \omega \cdot \omega \cdot s\theta + c\theta]$ | $-(1-c\kappa) \cdot \omega^2 \cdot \mu \cdot c\theta \cdot \omega \cdot \omega \cdot s\theta$ $-\omega\mu \cdot [(1-c\kappa) \cdot c\mu \cdot \omega \cdot \omega \cdot s\theta + c\theta]$ $c\kappa + (1-c\kappa) \cdot \omega^2 \cdot \mu \cdot \omega^2\theta$ | $\cos\kappa = (R_{11} + R_{22} + R_{33}) \cdot 1/2$ $\text{tg}\theta = -(R_{12} - R_{21}) / (R_{23} - R_{32})$ $\text{tg}\omega \cdot \cos\theta = -(R_{23} - R_{32}) / (R_{12} - R_{31})$ |
| <i>Polar</i> Crossther, 1972 Dodson, 1966 ($\mu = \omega$) | $c\kappa + (1-c\kappa) \cdot \omega^2 \cdot \omega \cdot c^2\theta$ $(1-c\kappa) \cdot \omega^2 \cdot \omega \cdot c\theta \cdot \omega \cdot \omega \cdot c\theta$ $\omega \cdot [(1-c\kappa) \cdot c\theta \cdot \omega \cdot \omega \cdot s\theta]$ | $(1-c\kappa) \cdot \omega^2 \cdot \omega \cdot c\theta \cdot \omega \cdot \omega \cdot c\theta$ $c\kappa + (1-c\kappa) \cdot \omega^2 \cdot \omega \cdot \omega^2\theta$ $\omega \cdot [(1-c\kappa) \cdot c\theta \cdot \omega \cdot \omega \cdot s\theta + c\theta]$ | $\omega \cdot [(1-c\kappa) \cdot c\theta \cdot \omega \cdot \omega \cdot s\theta]$ $\omega \cdot [(1-c\kappa) \cdot c\theta \cdot \omega \cdot \omega \cdot s\theta + c\theta]$ $c\kappa + (1-c\kappa) \cdot \omega^2 \cdot \omega$ | $\cos\kappa = (R_{11} + R_{22} + R_{33}) \cdot 1/2$ $\text{tg}\theta = -(R_{12} - R_{21}) / (R_{23} - R_{32})$ $\text{tg}\omega \cdot \cos\theta = -(R_{23} - R_{32}) / (R_{12} - R_{31})$ |
| <i>Direction cosines</i> | $c\kappa + (1-c\kappa) \cdot c\kappa^2$ $(1-c\kappa) \cdot c\kappa \cdot c\mu \cdot c\theta$ $(1-c\kappa) \cdot c\kappa \cdot c\mu \cdot c\theta$ | $(1-c\kappa) \cdot c\kappa \cdot c\mu \cdot c\theta$ $c\kappa + (1-c\kappa) \cdot c\mu^2$ $(1-c\kappa) \cdot c\mu \cdot c\mu \cdot c\theta$ | $(1-c\kappa) \cdot c\kappa \cdot c\mu \cdot c\theta$ $(1-c\kappa) \cdot c\mu \cdot c\mu \cdot c\theta$ $c\kappa + (1-c\kappa) \cdot c\mu^2$ | $\cos\kappa = (R_{11} + R_{22} + R_{33}) \cdot 1/2$ $\cos\kappa \cdot \cos\theta = (R_{12} - R_{21}) / (R_{23} - R_{32})$ $\cos\theta / \cos\kappa = -(R_{13} - R_{31}) / (R_{23} - R_{32})$ |
| <i>Eulerian</i> Reasimann, 1962 | $\omega\theta_1 \cdot c\theta_2 \cdot \omega\theta_1^2 + \theta_1 \cdot c\theta_1$ $\omega\theta_1 \cdot c\theta_2 \cdot \omega\theta_1^2 + \theta_1 \cdot \omega\theta_1$ $\omega\theta_1 \cdot \omega\theta_2$ | $c\theta_1 \cdot c\theta_2 \cdot \omega\theta_1^2 + \omega\theta_1 \cdot c\theta_1$ $c\theta_1 \cdot c\theta_2 \cdot \omega\theta_1^2 + \omega\theta_1 \cdot \omega\theta_1$ $-c\theta_1 \cdot \omega\theta_2$ | $\omega\theta_2 \cdot \omega\theta_1$ $\omega\theta_2 \cdot c\theta_1$ $c\theta_2$ | $\theta_1 = \text{atan2}(R_{31}, -R_{32})$ $\cos\theta_2 = R_{33}$ $\theta_3 = \text{atan2}(R_{13}, R_{23})$ |
| <i>Eulerian</i> Lattmann, 1966 | $[(1-c\theta_2) \cdot c\theta_1 + (1+c\theta_2) \cdot \omega\theta_1] \cdot 2$ $[(1-c\theta_2) \cdot \omega\theta_1 \cdot (1+c\theta_2) \cdot \omega\theta_1] \cdot 2$ $\sin[(\theta_4 + \theta_1) / 2] \cdot \omega\theta_2$ | $[(1-c\theta_2) \cdot \omega\theta_1 + (1+c\theta_2) \cdot \omega\theta_1] \cdot 2$ $[(1+c\theta_2) \cdot c\theta_1 \cdot (1-c\theta_2) \cdot \omega\theta_1] \cdot 2$ $-\cos[(\theta_4 + \theta_1) / 2] \cdot \omega\theta_2$ | $\sin[(\theta_4 - \theta_1) / 2] \cdot \omega\theta_2$ $\cos[(\theta_4 - \theta_1) / 2] \cdot \omega\theta_2$ $c\theta_2$ | $\theta_4 = \text{atan2}(R_{31}, -R_{32}) + \text{atan2}(R_{13}, R_{23})$ $\cos\theta_2 = R_{33}$ $\theta_1 = \text{atan2}(R_{31}, -R_{32}) - \text{atan2}(R_{13}, R_{23})$ |
| <i>Eulerian</i> Crossther, 1972 | $c\mu \cdot c\theta + \mu \cdot \omega \cdot \omega \cdot \mu$ $-c\mu \cdot c\theta + \mu \cdot \omega \cdot \omega \cdot \mu$ $c\mu \cdot \omega\theta$ | $\omega \cdot c\theta + \mu \cdot \omega \cdot \omega \cdot \mu$ $-\omega \cdot c\theta + \mu \cdot \omega \cdot \omega \cdot \mu$ $\omega \cdot \omega\theta$ | $-\omega\theta \cdot c\mu$ $\omega\theta \cdot \mu$ $c\theta$ | $\mu = \text{atan2}(R_{32}, R_{31})$ $\cos\theta = R_{33}$ $\theta = \text{atan2}(R_{23}, -R_{13})$ |
| <i>Eulerian</i> Hoodman, 1966 | $c\theta$ $\omega\theta \cdot \omega\theta$ $c\theta \cdot \omega\theta$ | $\omega\theta \cdot \omega\theta$ $-\omega\theta \cdot c\theta + \omega\theta \cdot \omega\theta$ $-c\theta \cdot \omega\theta + \omega\theta \cdot \omega\theta$ | $-\omega\theta \cdot c\mu$ $\omega\theta \cdot c\theta + \omega\theta \cdot \omega\theta$ $c\theta \cdot c\theta + \omega\theta \cdot \omega\theta$ | $\theta = \text{atan2}(R_{21}, R_{31})$ $\cos\theta = R_{11}$ $\mu = \text{atan2}(R_{12}, -R_{13})$ |
| <i>Eulerian</i> Huber, 1969 | $c\mu \cdot c\theta + \mu \cdot \omega \cdot \omega \cdot \mu$ $\omega \cdot c\theta + \mu \cdot \omega \cdot \omega \cdot \mu$ $-c\theta \cdot \omega\theta$ | $-\omega \cdot c\theta$ $c\mu \cdot c\theta$ $\omega\theta$ | $c\mu \cdot \omega\theta + \omega \cdot \omega \cdot \mu \cdot c\theta$ $\omega \cdot c\theta + \mu \cdot \omega \cdot \omega \cdot \mu$ $c\theta \cdot c\theta$ | $\theta = \text{atan2}(R_{32}, -R_{31})$ $\sin\theta = R_{33}$ $\mu = \text{atan2}(R_{12}, R_{22})$ |
| <i>Eulerian</i> Navaza, 1994 | $c\mu \cdot c\theta + \mu \cdot \omega \cdot \omega \cdot \mu$ $\omega \cdot c\theta + \mu \cdot \omega \cdot \omega \cdot \mu$ $\omega\theta \cdot c\mu$ | $-c\mu \cdot c\theta$ $-\omega \cdot c\theta + \mu \cdot \omega \cdot \omega \cdot \mu$ $\omega\theta \cdot \mu$ | $c\mu \cdot \omega\theta$ $\omega \cdot c\theta$ $c\theta$ | $\mu = \text{atan2}(R_{23}, -R_{13})$ $\cos\theta = R_{33}$ $\theta = \text{atan2}(R_{32}, R_{31})$ |

Table 3 gives some simplest relationships between rotation angles.

Table 3a. Basic relationships between polar rotation angles. The polar angles used by Dodson et al. (1966) are the same as the ones defined by Crowther with a small change in notation (ψ instead of ω). AMoRe (Navaza, 1994) uses the polar angles similar to the ones by Crowther only for the self-rotation, therefore the sign of κ is not defined.

| | Rossmann, 1962; $\kappa^z, \delta^z, \psi^z$ | Crowther, 1972; $\kappa^c, \delta^c, \omega^c$ | Direction cosines |
|---|--|---|--|
| Rossmann 1962 $\kappa^z, \delta^z, \psi^z$ | | $\kappa^c = \kappa^z$ $\cos \psi^z = \sin \omega^c \cdot \sin \delta^c$ $\tan \delta^z = -\cos \omega^c / (\sin \omega^c \cdot \cos \delta^c)$ | $\kappa^d = -\kappa^z$ $\cos \psi^z = \cos \psi$ $\tan \delta^z = (-\cos \omega, \cos \omega)$ |
| Crowther 1972 $\kappa^c, \delta^c, \omega^c$ | $\kappa^c = \kappa^z$ $\cos \omega^c = -\sin \psi^z \cdot \sin \delta^z$ $\tan \delta^c = \cos \psi^z / (\sin \psi^z \cdot \cos \delta^z)$ | | $\kappa^d = -\kappa^z$ $\cos \omega^c = \cos \omega$ $\tan \delta^c = (\cos \psi, \cos \psi)$ |
| direction cosines | $\kappa^d = -\kappa^z$ $\cos \omega = \sin \psi^z \cdot \cos \delta^z$ $\cos \psi = \cos \psi^z$ $\cos \omega = -\sin \psi^z \cdot \sin \delta^z$ | $\kappa^d = -\kappa^z$ $\cos \omega = \sin \omega^c \cdot \cos \delta^c$ $\cos \psi = \sin \omega^c \cdot \sin \delta^c$ $\cos \omega = \cos \omega^c$ | |

Table 3b. Basic relationships between some Eulerian rotation angles. Angles defined by Nordman (1966) and Huber (1969) are not included in the table because their recalculation through corresponding rotation matrix elements (Table 2) seems to be more rational.

| | Rossmann, 1962 $\theta_1, \theta_2, \theta_3$ | Lattman, 1966 $\theta_+, \theta_2, \theta_-$ | Crowther, 1972 $\psi^c, \beta^c, \gamma^c$ | Navaza, 1994 $\psi^a, \beta^a, \gamma^a$ |
|---|---|---|--|---|
| Rossmann 1962 $\theta_1, \theta_2, \theta_3$ | | $\theta_1 = (\theta_+ + \theta_-) / 2$ $\theta_2 = \theta_2$ $\theta_3 = (\theta_+ - \theta_-) / 2$ | $\theta_1 = \psi^c + \pi / 2$ $\theta_2 = \beta^c$ $\theta_3 = \gamma^c - \pi / 2$ | $\theta_1 = -\psi^a + \pi / 2$ $\theta_2 = -\beta^a$ $\theta_3 = -\psi^a - \pi / 2$ |
| Lattman 1966 $\theta_+, \theta_2, \theta_-$ | $\theta_+ = \theta_1 + \theta_3$ $\theta_2 = \theta_2$ $\theta_- = \theta_1 - \theta_3$ | | $\theta_+ = \psi^c + \gamma$ $\theta_2 = \beta^c$ $\theta_- = \gamma^c - \psi^c + \pi$ | $\theta_+ = -\psi^a - \gamma^a$ $\theta_2 = -\beta^a$ $\theta_- = -\psi^a + \pi$ |
| Crowther 1972 $\psi^c, \beta^c, \gamma^c$ | $\psi^c = \theta_1 - \pi / 2$ $\beta^c = \theta_2$ $\gamma^c = \theta_3 + \pi / 2$ | $\psi^c = (\theta_+ + \theta_-) / 2 - \pi / 2$ $\beta^c = \theta_2$ $\gamma^c = (\theta_+ - \theta_-) / 2 + \pi / 2$ | | $\psi^c = -\psi^a$ $\beta^c = -\beta^a$ $\gamma^c = -\psi^a$ |
| Navaza 1994 $\psi^a, \beta^a, \gamma^a$ | $\psi^a = -\theta_3 - \pi / 2$ $\beta^a = -\theta_2$ $\gamma^a = -\theta_1 + \pi / 2$ | $\psi^a = -(\theta_+ - \theta_-) / 2 - \pi / 2$ $\beta^a = -\theta_2$ $\gamma^a = -(\theta_+ + \theta_-) / 2 + \pi / 2$ | $\psi^a = -\psi^c$ $\beta^a = -\beta^c$ $\gamma^a = -\psi^c$ | |

The program is written on standard FORTRAN-77. The source is available by request from the author. Corresponding e-mail address is sacha@igbmc.u-strasbg.fr. Any reported comments, found mistakes and formulae improvements are appreciated. The full text of the notes will be published elsewhere. A new user-friendly version of the program is under development.

The author thanks Drs. A.Poterszman and J.C.Thierry for pointing the problem which initiated this work, Drs. J.Navaza, A.Podjarny and B.Rees for their extremely useful discussions about molecular replacement and Dr. D.Moras for his interest to the work. The work was supported by the CNRS through the UPR 9004, by the Institut National de la Santé et de la Recherche Médicale and the Centre Hospitalier Universitaire Régional.

Erratum (9/7/96)

For various reasons, the Tables included in the announcement of the program ROTCONV ("Relationships between rotation agreements"), and in the first release of this article, had some mistypings. Of essential ones, there is a wrong phi value in the Table 1, Line 1 (Rossmann, 1962), column "Alternative solution" ; and a down part of the Table 3 (making this Table even not self-consistent).

One disagreement was indicated by Ulrich Baumann : the PROTEIN rotation matrices are inverted in comparison with the X-PLOR ones and do not coincide with them. Therefore, in Table 1 corresponding matrices for PROTEIN are :

Polar :

$Ry(\phi) Rz(-\psi) Ry(\kappa) Rz(\psi) Ry(-\phi)$

(excluded from line 1 of Table 1)

Eulerian :

$Rz(\theta_1) Rx(\theta_2) Rz(\theta_3)$

(not shown in Table 1; PROTEIN can use this Eulerian system as an alternative to the one by Huber)

The corrected Tables are included in the current version of this article due to kind permission of the CCP4 staff.

The full version of this material is in preparation and is supposed to be sent to J.Appl.Cryst.

References:

- Brünger, A.T. (1992) X-PLOR. Version 3.1. A System for X-ray Crystallography and NMR. Yale University, Connecticut, USA.
Collaborative Computational Project, Number 4 (1994) Acta Cryst., **D50**, 760-763
Crowther, R.A. (1972) In: "The Molecular Replacement Method", M.G.Rossmann, ed., 173-178. New York: Gordon and Breach.
Crowther, R.A. (1973) Manual "The Fast Rotation Function".
Crowther, R.A. & Blow, D.M. (1967) Acta Cryst. **23**, 544-548.
Diamond, R. (1993) In: "International Tables for Crystallography", U.Shmueli, ed., B, 345-373, Dordrecht, Boston, London: Kluwer Academic Publishers.
Dodson, E., Harding, M.M., Hodgkin, D.C. & Rossmann, M.G. (1966) J.Mol.Biol., **16**, 227-241.

Fitzgerald., P. (1988) J.Appl.Cryst., **21**, 273-278
Goldstein, H. (1950) "Classical Mechanics", Cambridge, Massachusetts : Addison-Wesley Press.
Huber, R. (1969) In : "Crystallographic Computing Proc.", F.R.Ahmed, ed., 96-102. Munksgaard: Copenhagen.
Lattman, E.E. (1972) Acta Cryst., **B28**, 1065-1068
Navaza, J. (1994) Acta Cryst., **A51**, 157-163
Nordman, C.E. (1966) Trans.Am.Crystallogr.Assoc., **2**, 29-38
Rossmann, M.G., ed. (1972) "The Molecular Replacement Method. New York: Gordon and Breach.
Rossmann, M.G. (1990) Acta Cryst. **A46**, 73-82.
Rossmann, M.G. & Blow, D.M. (1962) Acta Cryst. **15**, 24-31.
Rossmann, M.G. & Arnold, E. (1993) In : "International Tables for Crystallography", U.Shmueli, ed., B, 230-263, Dordrecht, Boston, London: Kluwer Academic Publishers.
Steigemann, W. (1974) PhD thesis, Technische Univ., München, FRG