

9. SHELXPRO: Protein Interface to SHELX-97

A new program **SHELXPRO** has been added as an interactive user interface between SHELXL and other programs often used by protein crystallographers. It is designed to be self-explanatory so that it can be used without constant reference to a manual. It is started by:

shelxpro name

When started, SHELXPRO creates a log file *name.pro* and a Postscript output file *name.ps*. These may be printed after exiting from SHELXPRO and provide text and graphical summaries of the operations performed. Many options in SHELXPRO expect that the files *name.lst*, *name.fcf*, *name.pdb*, *name.res* etc. have been generated in a SHELXL job using the LIST 6 and WPDB instructions. A menu of possible options is displayed by SHELXPRO; choosing a particular option by typing the appropriate letter (upper or lower case) produces a detailed description of that option, after which the user has the choice of typing <Enter> to continue or N<Enter> to return to the menu. The menu consists of:

[F] New output filename	[V] R(free) files
[A] Anisotropic scaling (Hope & Parkin)	[I] .ins from PDB file
[P] Progress of LS refinement diagram	[L] Luzzati plot
[T] Thermal displacement analysis	[E] Esd analysis
[U] Update .res (and .pdb) to .ins file	[N] NCS analysis
[R] Ramachandran Phi-Psi plot	[K] Kleywegt NCS plot
[M] Map file for O from .fcf	[O] PDB file for O
[H] .hkl file from other data formats	[B] PDB deposition
[D] Convert DENZO/SCALEPACK .sca to .hkl	[C] Color plots (now on)
[X] Write XTALVIEW map coefficients	[W] Write Turbo-Frodo map
[S] Reflection statistics from .fcf	[Z] Least-squares fit
[G] Generate PDB file from .res or .pdb	[Q] Quit

Enter option:

The various options will now be discussed individually. Several of them add Postscript plots to the file *name.ps*. In these plots, the main-chain atoms are often color-coded according to the secondary structure, which the user is prompted for (blue for alpha-helix, green for beta-sheet and red for others). The side-chains are often color-coded according to residue characteristics:

Yellow	= Cys, Met
Green	= Phe, Tyr, Trp, His
Cyan	= Gly, Ala, Leu, Ile, Val, Pro
Red	= Glu, Asp
Blue	= Arg, Lys
Purple	= Gln, Asn
Gray	= Ser, Thr

9.1 Outline of the available features

The options provided by SHELXPRO can be divided into three general groups.

(a) Files and communication with other protein programs

[H] *.hkl* file from other data formats. This provides general interactive reformatting of reflection data files, avoiding the need to write a FORTRAN program or UNIX shell-script each time it is necessary to reformat reflection data.

[D] Convert DENZO/SCALEPACK *.sca* to *.hkl*. This is often the safest and quickest way of generating the *.hkl* reflection data file for SHELXL, SHELXS etc.

[V] R(*free*) files. This adds an R_{free} flag to selected reflections in an *.hkl* file; they may be chosen at random or in thin shells. This is the preferred method of calculating a *free R-factor* using SHELXL, and requires the SHELXL instructions `CGLS n -1` or `L.S. n -1`.

[I] *.ins* from PDB file. This will normally be used when a structure is transferred from another program to SHELXL for the first time. It generates most of the restraints and other extra instructions automatically as well as converting the atoms to fractional coordinates in SHELX format. For editing and updating between SHELXL refinement cycles the following [U] option should be used instead.

[U] Update *.res* (and *.pdb*) to *.ins* file. This should be used to read the *.res* output file from a SHELXL refinement job and update it to create the *.ins* input file for the next job. Alterations such as extra residues or disorder components may be added from a PDB format file written by a graphics program such as O or XtalView.

[G] Generate PDB file from *.res* or *.pdb*. Although SHELXL can write a PDB format file directly, this option provides for more user interaction, e.g. for setting up a PDB format file containing symmetry equivalents or modified temperature factors for use with molecular replacement programs such as AMoRe.

[B] PDB deposition. Collects the information needed for PDB deposition from the *.lst* and *.pdb* files written by SHELXL and creates a file according to the current specifications for deposition with the Brookhaven PDB. The resulting file contains all the compulsory records, but still requires some hand editing e.g. to include information about the data collection..

[F] New output filename. New *.ps* and *.pro* files are started and the previous *.ps* and *.pro* files closed. This enables the Postscript plots to be viewed in another window without leaving SHELXPRO etc.

[C] Color plots (now on). This option toggles color on or off in the Postscript output files. For some journals it may be necessary to produce black and white diagrams rather than color.

[Q] Quit. Terminates SHELXPRO and returns to the command line prompt.

(b) Creation of map (and *pdb*) files for various graphics packages

[M] Map file for O from *.fcf*. This creates a map file that can be read by O and some versions of FRODO. A variety of maps may be created, including Sigma-A maps. SHELXPRO reads the *.fcf* file written by SHELXL (it contains calculated structure factors and phases) and the *.pdb* file (in order to work out the extent of the map).

[W] Write Turbo-Frodo map. Very similar to the corresponding option for O.

[O] PDB file for O. The otherwise exemplary program O is unfortunately not able to read standard PDB format files (as written by e.g. SHELXL) when they contain disordered groups. This option provides a (not very elegant) work-around.

[X] Write XtalView map coefficients. Writes a *.phs* file with coefficients for various types of map including Sigma-A maps for input to XtalView. XtalView should be instructed to calculate an F_o -map whatever type of map is actually required! This produces MUCH better maps than inputting the atoms from SHELXL as a *.pdb* file into XtalView and repeating the structure factor calculation in XtalView (because of various incompatibilities such as the solvent model, anisotropic temperature factors, complex scattering factors as well as approximations made by XtalView in the structure factor calculation).

(c) *Analysis of a structure after refinement with SHELXL*

[P] Progress of LS refinement diagram. Produces a diagram of the *R*-factor as a function of the refinement cycle, with special action for automated water divining (SHELXWAT). The *R*-factors are extracted from the REM instructions in the current *.res* file, which are accumulated there when the U option in SHELXPRO is used to update the *.res* file written by one refinement job to create the *.ins* file for the next.

[T] Thermal displacement analysis. Creates bar-plots to show the variation of B-value (and anisotropy) with residue number for main-chain and side-chain atoms.

[R] Ramachandran Phi-Psi plot. A Ramachandran plot is created and the outliers listed. Reads the *.lst* file that must contain the necessary torsion angles calculated in SHELXL using RTAB instructions.

[K] Kleywegt NCS plot. A Kleywegt plot is a Ramachandran plot with NCS-related residues joined by straight lines. The lines cross the edges of the plot and reappear at the other side if necessary. If the plot is too hairy you may be in trouble..

[N] NCS analysis. Creates bar-plots of differences in B-values and various torsion angles between NCS related monomers. These are read from the *.lst* file so the torsion angles should have been calculated using RTAB instructions in SHELXL.

[S] Reflection statistics from *.fcf*. *R*-factors, data completeness, mean(l/σ) etc. may be calculated for user-specified resolution ranges.

[L] Luzzati plot. Similar to [S] but the resolution ranges are fixed by the program and a Luzzati plot of *R*-factor against resolution is created as well as the statistics.

[E] Esd analysis. Graphical analysis of the esds estimated by a (blocked) full-matrix refinement using SHELXL.

[Z] Least-squares fit. Allows parts of one or more structures to be fitted to each other and r.m.s. deviations calculated. The deviations may be plotted against residue number as bar plots and superimposed structures may be output in suitable format for preparing diagrams with MOLSCRIPT or the XP program in SHELXTL.

[A] Anisotropic scaling (Hope & Parkin). Applies an anisotropic scaling analysis to the *.fcf* file output from SHELXL using LIST 6. It is similar to the action of the HOPE instruction in SHELXL, but is much faster. This instruction may be used as a quick check to see whether the introduction of the HOPE instruction would be justified.

9.2 Communication with other programs

The various options will now be described in more detail. Much of this information is provided by the program when an option is chosen. This section contains useful information on the best ways of using SHELXL for protein refinements.

[H] *.hkl* file from other data formats

The program can read a variety of reflection data file formats and write a *.hkl* file in SHELX *.hkl* format. If the original file contained *F*-values, the *.hkl* file should be read into SHELXL with HKLF 3; if the original file contained intensities, HKLF 4 is appropriate. The input file should contain one reflection per line, but lines may be stripped from the beginning and end, e.g. to process data transferred by email. On reading the file, the first line is displayed. To skip this line and move to the next, hit the <Enter> key. To read *h,k,l*, *F* (or F^2) and $\sigma(F)$ [or $\sigma(F^2)$] from this and subsequent lines in free format, enter the character * followed by <Enter>; to read in fixed format, fill the positions under these quantities with H,K,L,F or S. Thus to read a correctly formatted *.hkl* file, enter the line:

```
HHHHKKKKLLLLFFFFFFFFSSSSSSSS
```

For technical reasons, the following option [D] should always be used instead of [H] to read files produced by SCALEPACK.

[D] Convert DENZO/SCALEPACK *.sca* to *.hkl*

The SCALEPACK *.sca* and SHELXL *.hkl* formats look very similar, but there are some subtle differences. The *.sca* file has three lines of header information but *.hkl* has no header. The *.hkl* file may be terminated by a line with all items zero that is not present in the *.sca* file; however both are also terminated by the end of the file. Unlike *.hkl*, the *.sca* file may contain floating-point numbers in 'l8' format. If the 'anomalous' flag was applied, the *.sca* file may contain reflections h_+ and h_- on the same line, with dummy values if not measured. The [D] option handles these differences and may also be used to extract anomalous ΔF values (with esds) for heavy-atom location using Patterson or direct methods in SHELXS.

[V] *R*(free) files

This command is used to flag say 5 or 10% of the reflections in the *.hkl* file for use as a reference set in calculating free *R*-values (Brünger, 1992). As a rule of thumb, at least 500 reflections or 5% of the total number should be flagged, whichever is larger. It is difficult to obtain statistically meaningful free *R*-values for datasets containing a total of less than 5000 reflections before division into reference and working sets. The flag is applied by making the 'batch number' at the end of each line in the *.hkl* file negative. The unflagged reflections constitute the working set. The *.hkl* file is read into SHELXL in the normal way using HKLF 4 (or 3), and the flags are ignored (i.e. all reflections are used for refinement and no free *R* is calculated) unless the second number on the CGLS (or L.S.) instruction is -1, in which case

only the working set is used for the refinement, and only the reference set is used to calculate the free R -values. It is customary to perform the final refinement using all the data, but not increasing the number of independent parameters or reducing the weights of the restraints. This may be done by simply deleting the second number on the CGLS or L.S. instruction.

The reference set may either be chosen at random or in thin shells. The latter option is strongly recommended if a twinned structure is being refined or if NCS restraints are applied, because otherwise the reference and working sets will not be independent. When the reflections are averaged in SHELXL, they are included in the final reference set only if all contributors have the R_{free} flag set, otherwise they are used in the working set. In such a case it is advisable to use thin shells rather than flagging the reflections at random, otherwise there will not be many reflections left in the reference set after averaging!

Note that if the second CGLS (or L.S.) parameter is negative ($-N$) with N not equal to 1, SHELXL will generate its own reference set consisting of every N 'th reflection (after merging) irrespective of the flags in the *.hkl* file. This possibility is retained for upwards compatibility with SHELXL-93, but is NOT RECOMMENDED, because the reference set may possibly change if a different space group, resolution range, merging procedure or a different version of SHELXL is used, and because it is inappropriate for problems involving NCS or twinning.

[I] *.ins* from PDB file

Usually, when SHELXL is used for a high-resolution refinement, a low-resolution or preliminary refinement will already have been performed with another program, or a model will be available from molecular replacement or map interpretation in the form of a PDB file. SHELXPRO can read PDB files taken from the Brookhaven database as well as files written by X-PLOR and other widely used protein programs. The [I] option incorporates standard Engh & Huber (1991) restraints, and other instructions needed for a refinement job, into the *.ins* file. The program applies some consistency checks and searches for disulfide bridges, generating the necessary restraints automatically. The user may renumber the residues and must specify the residue numbers for N- and C-termini so that appropriate action can be taken. Since SHELXL does not recognize chains, these must be flagged by adding (e.g.) 1000, 2000, ... to the residue numbers (note that the [B] and [G] options in SHELXPRO provide the reverse transformation). It is advisable to ignore hydrogen atoms in the input PDB file because it is better to regenerate and refine them using the riding model in SHELXL.

It is almost inevitable that some hand editing of the resulting *.ins* file will still be necessary. For example, SHELXPRO is not able to define restraints, torsion angles and hydrogen atoms for residues that it doesn't recognize. Bad initial geometry may require the addition of FREE or BIND instructions so that the connectivity array is generated correctly by SHELXL, and chain breaks, ligands or solvent molecules other than water may require special action. The [I] option, followed by any necessary hand editing, should be used once per structure before the first SHELXL refinement. Thereafter it is much more convenient to use the [U] option in SHELXPRO to update the *.res* file from one refinement job to produce the *.ins* file for the next., because special restraints and other instructions are retained, and because there are extra facilities for defining and checking disorder, solvent molecules, etc. The restraints incorporated into the *.ins* file are stored internally in SHELXPRO, so no dictionary file is required (in contrast to the now obsolete program PDBINS supplied with SHELXL-93, which used a dictionary file *shelxl.dic*).

[U] Update .res (and .pdb) to .ins file

This option converts a SHELXL .res file to a new .ins file by including new or changed atoms from PDB format files such as those written by the graphics programs O, Turbo-Frodo and XtalView. All other SHELXL commands are retained unchanged. This option also provides for setting up disorder refinement and updating the list of solvent molecules. The .res file should not contain instructions other than RESI, AFIX, PART and atoms between FVAR and HKLF, and both FVAR and HKLF must be present. Note that although it is possible to set up threefold or multiple disorders in this way, the necessary SUMP restraints must be edited into the .ins file later by hand; no extra editing is needed for twofold disorders. The [U] option may also be used without a .pdb file to update the .res file to .ins and apply various checks. It is recommended that the .res file is always updated to .ins in this way rather than by using an editor, so that the REM records that contain a summary of the course of the refinement are accumulated correctly; if necessary the resulting .ins file can then be edited further with a text editor before rerunning SHELXL.

[G] Generate PDB file from .res or .pdb

The WPDB instruction in SHELXL is normally used to write PDB format files, but the [G] option in SHELXPRO provides additional editing facilities that are particularly useful for the creation of PDB format files for use as molecular replacement search models, and are also sometimes useful before calculating least-squares fits etc. An .ins, .res or PDB format file serves as input. B-values may be reset automatically to typical values, disordered atoms, solvent molecules and H-atoms may be removed, chain ID's (not recognized by SHELXL) may be (re)inserted, and multiple copies of chains may be generated using (non-)crystallographic symmetry. In the resulting PDB file all atoms are isotropic.

[B] PDB deposition

The [B] option reads files .pdb and .lst files written by the 'final' SHELXL refinement job and creates a file with the default extension .ent in PDB format suitable for deposition in Brookhaven. Some of this file is in the form of a template suitable for hand editing, e.g. to include literature references, experimental details, special features of the structure and refinement, etc. The user is prompted for details of chains and possible renumbering of the residues; except for structures consisting of a single chain, chain ID's should be (re)inserted in this way before deposition. The resulting file should contain all the compulsory records, but some of them will need completion by subsequent hand editing. The following notation is used to redefine residue numbers and chains. When prompted by the program, the new chain ID letter (the character '\$' should be used if a blank chain ID is required) is followed by the first and last old residue numbers and the first new residue number. One chain should be specified per input line, and the list of chains is terminated by a blank line. Thus if there were two chains numbered 1001-1189 and 2001-2189, followed by waters with residue numbers 1-111, the following three lines should be entered:

```
A 1001 1189 1  
B 2001 2189 1  
$ 1 111 201
```

For example, residue 1001 in this example would become chain A residue 1. Similarly, residue 2189 becomes chain B residue 189. The solvent water that used to start at residue 1 now starts at residue 201.

For the deposition of reflection data, the CIF format *.fcf* file written by SHELXL may be used directly.

9.3 Creation of map (and pdb) files for various graphics packages

In a computer utopia, interactive graphics packages would all read the CIF format *.fcf* file written by SHELXL directly; this contains all the information necessary for generating maps. For the couple of years before this comes to pass, SHELXPRO provides the necessary generation of maps or (in the case of XtalView) coefficients. For the programs O and Turbo-Frodo, it is also necessary to define the region of space for which the map is calculated; SHELXPRO does this by scanning a PDB file to find the maximum and minimum atomic coordinates in each direction. Furthermore, O is liable to be confused by disordered residues even if these are specified exactly according to the PDB rules (as SHELXL does), so it is also necessary for SHELXPRO (option [O]) to be able to modify the PDB file so that all disorder components are given separate residue numbers. Note that the option [U] provides the reverse procedure, i.e. separate residues obtained using O may be recombined as different disorder components of the same residue for refinement using SHELXL. SHELXPRO does not make the changes that may be required to the *all.dat* connectivity file read by O.

The [M], [O], [W] and [X] options should be self-explanatory. The following questions are asked by the program; usually the answers suggested by the program are suitable, so most of the questions are answered by <Enter>.

Name of *.fcf* file created using SHELXL and LIST 6 [name.fcf]:

Enter name of PDB file [name.pdb]:

Include all waters in the volume covered by map? [Y]:

Number of grid points per cell in x, y and z (the first two MUST be powers of 2, and the last MUST be a multiple of 8) [64 64 88]:

Origin of map along x, y and z (grid points) [-32 -24 24] (must all be multiples of 8):

Extent of map along x, y and z (grid points) [128 136 88] (must all be multiples of 8):

Fourier type (-3= $mF_o - DF_c$ (Sigma-A difference map), -2= $2mF_o - DF_c$ (Sigma-A map), -1= $F_o - F_c$, 0= F_c , 1= F_o , 2= $2F_o - F_c$, n= $nF_o - (n-1)F_c$) [-2]:

Enter reference/working set Sigma-A ratio from SHELXL [0.97]:

Apply sharpening (Y or N) ? [N]:

Enter name of map file [sigmaa.map]:

For XtalView, the questions about the grid are skipped. Note that there is a choice of maps. Thus the input '3' for the Fourier type generates a $3F_o - 2F_c$ map; '4' gives a $4F_o - 3F_c$ map, etc. The sigma-A ratio is calculated in each SHELXL job that uses the free *R*-factor; it is designed

to correct the sigma-A weight for overfitting. For refinement at low resolution this might be about 0.8, for medium resolution 0.9; the default is appropriate for structures with a high ratio of data to parameters. If the free R -factor was not used in the refinement, a estimated value should be input. 'Sharpening' multiplies the coefficients by $\langle F^2 \rangle^{1/4}$, where $\langle F^2 \rangle$ is the mean reflection intensity in the appropriate resolution shell (this factor is used in preference to the almost identical factor $\sqrt{(E/F)}$ because the latter involves a statistical factor for certain reflections that is inappropriate for this application). Finally, the program outputs the maximum and minimum electron density (in sigma units as well as -electrons per cubic Ångström) and electron density histogram.

Note that XtalView MUST be told to do an F_0 synthesis, whatever type of map the coefficients actually represent !

9.4 Analysis of the refined structure

The *.lst* file produced by SHELXL contains a great deal of important information, but for proteins (in contrast to small molecules) it is not very economical to print it out and read it after every job. Many of the following options are designed to summarize the essential information in more digestible form, e.g. as Postscript plots. Usually the *.lst* and/or *.fcf* and sometimes the *.res* or *.pdb* files are required from a SHELXL refinement job in which the LIST 6, FMAP 2 and WPDB instructions were employed.

[P] Progress of LS refinement diagram

At the end of each refinement job, and after each SHELXL stage in the SHELXWAT water divining procedure, SHELXL outputs three lines of remarks to the *.res* file containing current R -values etc. If the *.res* file is edited to the next *.ins* file in such a way as to retain these remarks, they provide a convenient summary of the course of the refinement. The remarks are written after the HKLF instruction so they must be moved ahead of this instruction in order to be preserved; if the [U] option is used to update from *.res* to *.ins* this happens automatically. The [P] option extracts the R -factors from these remarks and prepares a Postscript plot of R -factor against refinement job number. Points that were part of the SHELXWAT water divining procedure are plotted with a smaller horizontal gap between them. This plot provides a convenient summary of the course of refinement; it can be seen at a glance which stage produced the biggest drop in free R -factor, and whether R continues to fall but the free R -factor rises again, indicating over-refinement.

[T] Thermal displacement analysis

This reads a SHELXL *.lst* file from an isotropic or anisotropic refinement and prepares Postscript bar plots of the mean (equivalent) B and (optionally) anisotropy (minimum eigenvalue divided by maximum eigenvalue) against residue number. The refinement should have been performed with FMAP 2, so that the residue diagnostics table is present in the *.lst* file. Unless black and white Postscript output is set, the main-chain plots are color coded according to secondary structure (it is useful to run PROCHECK first to obtain this

information) and the side-chain plots by residue type. The color schemes are defined in the .pro output file.

Alpha-helices and beta-strands are entered one per line with 'A n1 n2' or 'B n1 n2' respectively, where n1 and n2 are the first and last residues of the helix or strand. The letters may be upper or lower case. The list is terminated with a blank line. Thus:

```
a 21 45
b 48 55
a 67 108
```

would define two alpha-helices (residues 21 to 45 and 67 to 108 resp.) and one beta-strand (48 to 55). The alpha-helix regions are colored blue, the beta-strands green, and the rest red. There may be up to four diagrams on one page, starting at the top. Each should be defined by entering three characters: a symbol to label the diagram, then either B (B-values) or A (anisotropy), followed by M (main-chain) or S (side-chain) and then the numbers of the first and last residues. END terminates the list. The program will suggest suitable parameters. A typical sequence, selecting these defaults by <Enter> each time, would be:

```
Next diagram [aBM 1 204]:
Maximum value and step for vertical scale [50 10]:
Next diagram [bAM 1 204]:
Next diagram [cBS 1 204]:
Maximum value and step for vertical scale [60 10]:
Next diagram [dAS 1 204]:
```

Note that no scale needs to be specified for the anisotropy, because the range is always from 0 to 1.

[R] Ramachandran Phi-Psi plot

The [R] option reads the SHELXL .lst output file and extracts the psi and phi torsion angles to make Ramachandran plots. If the main-chain is disordered, only the PART 1 (and of course PART 0) atoms are used. Glycines are included optionally as open squares; prolines are treated as normal residues. A list of outliers appears on the screen and in the .pro file. Residues are color-coded according to residue type unless black and white Postscript has been specified (option [C] in the main menu). The refinement should have been performed with appropriate RTAB instructions for the phi and psi torsion angles and with FMAP 2, so that the residue diagnostics table is present in the .lst file. See Kleywegt & Jones (1996), who kindly provided the distribution table used in SHELXPRO.

[K] Kleywegt NCS plot

This is the same as the normal Ramachandran plot (option [R] above) except that the phi/psi dots for each residue are smaller and residues related by non-crystallographic symmetry (NCS) are joined by lines (Kleywegt, 1996). The lines may cross the edges of the plot and reappear at the other side if this makes the differences between the angles smaller. Ramachandran outliers (as defined by Kleywegt and Jones) are also reported. This plot gives an immediate indication of how well NCS is obeyed for the main-chain atoms, and is also a good indicator of the overall quality of the structure. If the main-chain is disordered, only PART 0 and PART 1 atoms are considered. Glycines are optionally included as open squares; prolines are treated as normal residues. Unless color has been switched off (option [C]) the dots and lines are color-coded according to residue type. The refinement should have been performed with FMAP 2 and the RTAB instructions needed to calculate the phi and psi torsion angles in SHELXL.

[N] NCS analysis

This option provides a detailed analysis of deviations from non-crystallographic symmetry (NCS). The Kleywegt plot [K] can also be used to provide an overall picture of how well NCS is obeyed by the main-chain torsion angles. Before using these options, a SHELXL refinement should be performed in which RTAB is used to calculate the phi, psi, omega and chi1...chi4 torsion angles. The instruction FMAP 2 is also required so that the *.lst* file contains the residue diagnostics table. It is also useful to have secondary structure assignments to hand for color coding of the NCS bar plots; many standard protein programs such as PROCHECK are able to supply this information.

Differences (2 NCS related components) and maximum deviations and r.m.s. deviations (if there are more than two components) are plotted and tabulated as a function of the base residue number (i.e. the residue number minus the offset such as 1000, 2000 ... that SHELXL uses instead of a chain ID). Because of the large number of factors involved this option requires some attention to detail.

Alpha-helices and beta-strands are entered one per line as 'A n1 n2' or 'B n1 n2' where n1 and n2 are the first and last residues of the helix or strand. Base residue numbers should be used and the list is terminated with a blank line. Then the numbers that have to be added to the base residue numbers to generate the NCS related units are defined in answer to a prompt by the program. For fourfold NCS the usual SHELXL convention of numbering equivalent chains 1001..., 2001... etc. would require the input '1000 2000 3000 4000' here. The program then requests the minimum deviations in angles (deg.) and B for output to *.pro* file; 0 would print all and 999 would not print any.:

There may be up to four diagrams on one page, starting at the top. Each should be defined by entering three characters: a symbol to label the diagram, then either D (absolute difference [rms absolute difference from mean if more than 2 components]), M (maximum absolute deviation [from mean if more than 2]) or A (average), followed by the letter H (phi), Y (psi), P (phi and psi), O (omega), C (chi1), T (all chi), M (main-chain B) or S (side-chain B) and then the numbers of the first and last base residues. Note that A is only allowed with S or M and that P or T must be preceded by M. END terminates the list.

The default diagrams are:

aMH (diagram a; maximum absolute deviation of phi angles)
bMY (diagram b; maximum absolute deviation of psi angles)
cMO (diagram c; maximum absolute deviation of omega angles)
dMT (diagram d; maximum absolute deviation of all chi angles)
eMM (diagram e; maximum absolute deviation of main-chain B)
fMS (diagram f; maximum absolute deviation of side-chain B)
gAM (diagram g; average main-chain B)
hAS (diagram h; average side-chain B)

[S] Reflection statistics from .fcf

This option creates reflection statistics from a *.fcf* file written by SHELXL in response to a LIST 6 instruction.. The user must specify the resolution ranges, e.g. to be the same as those used for data reduction. A table of data completeness, *R*-factors etc. is written to the console and to the *.pro* output file.

[L] Luzzati plot

This plots the resolution vs. *R*₁. The *.fcf* file must have been created using LIST 6 in SHELXL. SHELXPRO outputs a Postscript Luzzati (1952) plot, which gives estimates of the average errors in atomic coordinates for an incompletely refined structure assuming perfect data, NOT (as widely assumed by people who have not read this paper which happens to be in French) estimates of the esds in the atomic positions. For small proteins and high resolution data, esds in individual bond lengths and atomic positions may be estimated rigorously using SHELXL (see the [E] option in SHELXPRO described below). Nevertheless, a plot of *R*-factor against resolution is always entertaining.

[E] Esd analysis

This option reads SHELXL *.lst* file and prepares Postscript scatter-plots of esds in atom positions and bond lengths against (equivalent) B values. The refinement should normally have been performed with the SHELXL instructions L.S. 1, DAMP 0 0, BLOC 1 and BOND. If geometrical restraints were used in the refinement the bond length esds will be very low, but high resolution data are required to perform such a refinement without restraints. Similarly the damping has to be switched off because this can also lead to underestimated esds. Disordered atoms, atoms on special positions, and atoms other than C, N and O are not included in the diagrams. Such atoms are recognized by the first letter of their names in the atom coordinate table, so it may be necessary to remove calcium and other atoms that might be mistakenly identified from this table by editing the *.lst* file before running SHELXPRO.

A quadratic may be fitted to the atom radial esds, which enables the results to be compared with the formula suggested by Cruickshank (1996). Note that this formula predicts positional esds in one direction, which should be a factor of $\sqrt{3}$ smaller than the radial esds output by SHELXL.

[Z] Least-squares fit

The [Z] option may be used to perform a least-squares fit of two molecules, taken from the same or different structures. The iterative quaternion method is employed. This option is of necessity rather complex, and it is important to read each request for information by the program carefully because the default action (<Enter>) may well not be suitable and an incorrect answer can lead to complications.

It is necessary first to define the first molecule (called 'current structure'), which is extracted from a PDB format file. The a second molecule ('model') is obtained from another (or possibly the same) PDB file. Both PDB files may be as output by SHELXL or may be taken directly from the PDB databank, so 'chains' may be present. Since the residues may be numbered differently in the two molecules, it is necessary to convert the residue numbers in both molecules to a matching set of residue numbers referred to as SHELXPRO residue numbers. These numbers are also used to annotate the plots etc. The set of residues used for fitting is in general a subset of those used for the plots and calculation of r.m.s. esds.

After performing the fit for specified atoms in each of the specified residues, the program prints the r.m.s. deviation of the atoms fitted and the largest individual deviations (greater than 2σ). Then appears the question:

New current structure (C), new model (M), Repeat fit (R), write PDB file (P), XP file (X), Postscript bar plot of differences (D) or exit (E) [E]:

'R' repeats the fit (possibly using different residues and atoms) of the 'model' (second molecule) to the 'current structure' (first molecule). 'M' replaces the 'model' but keeps the 'current structure'. 'C' starts again with a new 'current structure'. 'P' writes a new PDB format file that contains the two molecules as two separate chains with the SHELXPRO residue numbers; this can be used as input to the program MOLSCRIPT. 'X' writes an orthogonal coordinate file that can be read by the Siemens' SHELXTL program XP and used to make a (stereo) $C\alpha$ -trace of the superposition. 'D' prepares a Postscript bar plot of the differences between the two molecules, using all stored residues, not just those that were fitted.

[A] Anisotropic scaling (Hope & Parkin)

This option reads an *.fcf* file created using the LIST 6 instruction in SHELXL, and writes a NEW *.hkl* file after application of anisotropic scaling by the method of Parkin, Moezzi & Hope (1995). The modification of the observed structure factors in this way is scientifically suspect and is intended for testing purposes only. It is much better to use the HOPE instruction in SHELXL so that parameter correlations are taken into account and the observed data are not modified. The SHELXPRO correction provides a quick test as to whether HOPE in SHELXL will result in a significant improvement; in this case the question about the filename for corrected data should be answered with <Enter>. A 'local' R_{free} test is applied to establish how many parameters [none(!), 12, 18 or 24] may justifiably be fitted. A significant improvement is not to be expected if anisotropic refinement has been performed or if a large number of symmetry equivalents were merged in the data reduction.