ProFit V2.2

Dr. Andrew C.R. Martin,
University of Reading

Document First Written: 25th July, 1996 (University College London)
Last updated: 20th March, 2001

1 Introduction and Methodology

ProFit (pronounced Pro-Fit, not profit!) is designed to be the ultimate program for performing least squares fits of two protein structures. It performs a very simple and basic function, but allows as much flexibility as possible in performing this procedure. Thus one can specify subsets of atoms to be considered, specify zones to be fitted by number, sequence, or by sequence alignment.

ProFit does not try to address the question of sorting out equivalent atoms for you beyond doing a sequence alignment. There are other programs such as SSAP and GAFIT which address that problem. You must specify which residues and atoms you consider to be equivalent although the program supports internal sequence alignment to set the zones automatically.

As of **ProFit** V2.0, iterative updating of fitting zones is now supported. Thus you may give just a small fragment to initiate the fitting process (a minimum of 3 amino acids). Fitting is performed on this region and then all residue pairs within 3Å are included in the fitting zones and the fitting is repeated. This iterates until the Cα RMSd converges to within 0.01Å. This is particularly useful in conjunction with the initial zone specification based on sequence alignment. Convergence typically takes 3–4 cycles.

ProFit V2.0 also introduces multiple structure fitting. The first structure file is used as a reference set for the first fitting stage but the coordinates are averaged after each stage to derive a template used for subsequent fitting. i.e. Given N files to fit, file 2 is fitted to file 1 and an averaged structure, A, is calculated, file 3 is then fitted to A and a new average, A' is calculated. This continues until all N structures have been fitted. The whole procedure iterates until convergence (typically 3 or 4 cycles).

The program will output an RMS deviation and optionally the fitted coordinates. RMS deviations over alternate zones and atoms may also be calculated without performing a new fit. Thus the zones for calculating the RMS deviation can be different from those used for fitting.

While optimised for proteins, non-protein structures may also be fitted if they are stored in the standard Protein Databank (PDB) format.

ProFit is written to be as easily portable between systems as possible and uses a command-driven interface. It is planned to add an optional Expectk graphical interface at some future date.

ProFit uses the McLachlan fitting algorithm, essentially a steepest descents minimisation, as described in McLachlan, A.D. (1982) *Rapid Comparison of Protein Structures*, *Acta Cryst.* **A38**, 871–873. This part of the code is based on an implementation by Dr. Mike Sutcliffe.

In summary, **ProFit** has the following features:

1. Portability between different operating systems
2. Ability to specify atom subsets
3. Ability to specify zones:
   - Numerically
   - By sequence
   - By auto sequence alignment
4. Output RMS deviation over:
   - Fitted reglon
   - Any other reglon
   - Any other atom set
5. Optionally output fitted coordinates in PDB format
6. Integrated help facility
7. Fitting zones derived from sequence alignment
8. Iterative updating of fitting zones
9. Multiple structure fitting

2 Starting the program

The program is started from the command line by typing the command:

```
profit
```

Once the program is started, you may read in structures to be fitted. Alternatively, the PDB files may be specified on the command line:

```
profit reference.pdb mobile.pdb
```

By default, **ProFit** does not read HETATM records from the PDB file. This may be changed from the command line by using the -h flag:

```
profit -h
profit -h reference.pdb mobile.pdb
```

Alternatively, once in the program you may give the `HETATOMS` command before reading in the structures (see Section 3.

If compiled with XMAS[1] file support, the `-x` flag may be used to specify that the files named on the command line are XMAS files instead of PDB files. Note that the program currently will only write PDB format files.

```
profit -x reference.xmas mobile.xmas
```

If `COPT   =` in the Makefile is changed to `COPT   = -DGUNZIP_SUPPORT` then the program can read gzipped PDB files. This will only work on unix-like platforms and assumes that the `gunzip` program is in your path. Note that the uncompressed files will remain in `/tmp` with a name like `readpdb_12345` where 12345 is a process number. You will need to delete these regularly!

Once in the program, you issue commands by typing at the keyboard. These commands may always be abbreviated to the minimum non-ambiguous string. The program is also case insensitive; you may mix upper and lower case at will, though uppercase will be used throughout this documentation.

You exit from the program by typing `QUIT`.

[1] XMAS is an XML-like file format developed at Inpharmatica, Ltd. which is designed for leaf-heavy data such as protein structure data

3 Reading Structures

ProFit reads files in PDB format. If compiled with XMAS support, then XMAS files may also be read, but only PDB files may be written. It uses the concept of a *reference* structure and a *mobile* structure. The reference structure remains static in space and the mobile structure is fitted onto it. When the files are specified on the command line, the reference structure is specified first, the mobile structure second.

Once in the program, you may read the reference structure using the `REFERENCE` command and the mobile structure using the `MOBILE` command. Using these commands causes the equivalent current structure to be deleted from the program's memory first. However, any zone and atom specifications (see Sections 6 and 7) are not deleted. For example, you can read p3hfl.pdb as a new reference structure using the command:

```
REFERENCE p3hfl.pdb
```

and read p3hfm.pdb as a new mobile structure with:

```
MOBILE p3hfm.pdb
```

If compiled with XMAS support, then the XMAS format is specified by placing the keyword `XMAS` after the commands `REFERENCE` or `MOBILE`:

```
REFERENCE XMAS p3hfl.pdb
MOBILE XMAS p3hfm.pdb
```

When you read a structure containing insertions, you will receive a warning message to this effect. This dates from when the program was unable to handle residue specifications containing insertion codes, but is still useful to draw your attention to the fact that they are present.

Note that atoms with coordinates of 9999.00, 9999.00, 9999.00 will be ignored during all calculations allowing atoms with undefined coordinates to be handled.

When fitting multiple structures (new in **ProFit** V2.0), you use the `MULTI` command to read in the structures. See Section 8.

4 Getting Help

To get help within the programs simply type `HELP` and you will be presented with a list of commands which the help facility knows about. The `ProFit>` prompt will also change to `Help>`. You may then type the name of a command to get help on that command. Typing `HELP` once in the help facility will repeat the list of available help topics. Like the main command interface, the help facility will accept upper or lower case and you may abbreviate commands. If your abbreviation is ambiguous (i.e. more than one command starts with the letters you have specified), help will be supplied on all the commands which match[2].

If the help text is longer than 21 lines, you will see a prompt saying

```
More...
```

in which case you should hit the Return (or 'Enter') key to get the next page of help.

Once at the `Help>` prompt, you should simply hit the Return (or 'Enter') key to get back to the main `ProFit>` prompt.

[2]The one exception to this is if the letters you supply are an abbreviation of `HELP`, when the list of help topics will be shown again.

If you know the topic on which you need help, you may type the name of the command after the HELP keyword at the main ProFit> prompt. After the help message is printed, you will be returned directly to the ProFit> prompt. For example, if you want help on the ZONE command, you may type:

```
HELP ZONE
```

Allied to help, is the STATUS command. This tells you the current status of the program: what structures are loaded, fitting zones, atoms and the like.

5 Fitting Structures

Having read in a reference and a mobile structure, you actually fit them by giving the FIT command. When you do this, you will get a message like:

```
Fitting structures...
RMS: 0.366
```

However, this will only work if the two structures are of identical composition i.e. if the sequences are the same and the same atoms are present in both. If there are any mismatches, the first such mismatch will be reported and the RMS deviation will not be calculated.

Since you will frequently need to fit non-identical structures, you may use the ZONE and ATOMS commands to specify which residues should be considered equivalent and which atoms should be considered in the calculation.

If you are using zone or atom specifications, the RMS deviations will be displayed over the atoms and zones specified in those commands.

Normally the fitting procedure will not be completed if there are any mismatched atoms or atoms missing from one of the two structures. The program issues an error message about atoms missing in the mobile structure which are found in the reference structure. The IGNOREMISSSING command causes the program to issue a warning instead of an error and the fitting proceeds ignoring the mismatched atoms. The default behaviour is restored by using the NOIGNOREMISSING command.

6 Specifying Atom Subsets

The ATOMS command is used to specify a subset of atoms to be used in the calculations. It has the syntax:

```
ATOMS atm[,atm]...
```

i.e. you specify the ATOMS keyword followed by one or more atom names separated by commas. A * may be used to specify all atoms and a ~ or ^ may be placed at the beginning of the specification to inverse the selection. For example, to fit only Cα atoms:

```
ATOMS CA
```

to fit N, Cα, C and O atoms:

```
ATOMS N,CA,C,O
```

to fit sidechains only (i.e. everything except N, Cα, C and O atoms):

```
ATOMS ^N,CA,C,O
```

to return to fitting all atoms:

```
ATOMS *
```

The PDB atom name field is 4 characters wide followed by a space. The first two characters are the right-justified element type, so for normal protein and DNA atoms consist of a space followed by a N, C, O, S or P. Thus the atom name field for a Cα contains ' CA '. HETATMs such as calcium will contain the two characters CA in the first two fields. i.e. 'CA '. When you specify an atom type it is matched against the atom name field from the *second character onwards*, unless you preceed it with a `<`. Thus to match a Cα you use `CA`, but to match Calcium, you use `<CA`. For example, as stated above, to match Cα atoms:

```
ATOMS CA
```

while to match calcium atoms

```
ATOMS <CA
```

and to match both Cα and calcium:

```
ATOMS <CA,CA
```

Wildcards are also allowed. A `%` or a `?` may be used to match a single letter at any point in the specification while a `*` may be used to match all remaining characters (thus `C*` is allowed, but `*G` is not). These special characters may be escaped by preceding them with a \. For example to fit all carbons:

```
ATOMS C*
```

or to match all atoms at the γ position:

```
ATOMS ?G*
```

and to match the `C4*` atoms in DNA:

```
ATOMS C4\*
```

If atom names contain spaces (e.g. in heme groups) the whole atom specification must be enclosed in double inverted commas:

```
ATOMS "N A,N B,N C"
```

7 Specifying Zones

The `ZONE` command is used to specify zones in the two structures which are considered equivalent. The complete syntax for the command is:

```
ZONE CLEAR|((*|(X...[,n][/m])|(j-k))[:(*|(X...[,n][/m])|(j-k))])
```

where `X...` is an amino acid sequence, `n` is a number of residues, `m` is the occurrence number, `j` and `k` are residue specifications of the form *[chain]resnum[insert]*. Items in square brackets are optional and alternatives are marked by a `|` and grouped in parentheses.

`ZONE` commands are cummulative. Thus each zone you specify is added to those currently active. To clear all zones (i.e. fit all residues), the `ZONE CLEAR` or `ZONE *` command may be given.

Although it appears complex, the syntax is actually very simple and consists of two identical sections separated by a colon (:). The left half is applied to the reference structure and the right half to the mobile structure. In its simplest form, the right hand half of the expression is absent and the specification is applied to both reference and mobile structures. For example:

```
ZONE 24-34
```

will set the zone to include residues 24–34 in both structures. If you wanted to fit 24–34 in the reference structure with 25–35 in the mobile structure, this simply becomes:

```
ZONE 24-34:25-35
```

You may also specify chain names and insertion code. The chain name is placed before the residue number and the insertion code afterwards. For example:

```
ZONE L25A-L30
```

fits residues 25A–30 in the L chain of both structures.

Simple wildcards may also be used. For example

```
ZONE H*:B*
```

fits the reference H chain with the mobile B chain,

```
ZONE -10:50-59
```

fits from the first residue to residue 10 in the reference structure with 50–59 in the mobile structure.

```
ZONE *:1-100
```

fits all residues in the reference structure with 1-100 in the mobile structure.

If the structure file contains negatively numbered residues and you are using residue numbering, you can escape the minus sign in the residue number using a backslash:

```
ZONE \-4-10:\-1-13
```

will fit residues -4 to 10 in one structure with -1 to 13 in the other.

Alternatively, you may specify the zones to be fitted by giving a sequence fragment. Together with that fragment, you may specify the number of residues to consider starting at that point. If the fragment occurs more than once in the sequence you may specify which occurence you wish to consider. For example:

```
ZONE CAR:VNS
```

fits the first occurence of CAR in the reference set with first occurence of VNS in the mobile set;

```
ZONE CAR,10:VNS,10
```

fits 10 residues starting at the first occurence of CAR in the reference set with 10 residues from the first occurence of VNS in the mobile set;

```
ZONE CAR,5/2
```

fits 5 residues from second occurence of CAR in both structures;

```
ZONE 24-34:EIR,11
```

fits 24-34 in the reference set with 11 residues starting at the first occurence of EIR in the mobile set.

By default, **ProFit** works in 'Residue Number' mode, i.e. the numbers used in zone commands are the numbers seen in the PDB file. The alternative mode is 'Sequential' mode where residues are numbered sequentially throughout the structure (including throughout multiple chains). Any chain names appearing in zone specifications will be ignored in Sequential mode. To switch mode, you use the `NUMBER SEQUENTIAL` or `NUMBER RESIDUE` commands.

7.1 Sequence Alignment

Another way of specifying zones is to let the program do it. **ProFit** does not provide any facilities for calculating structural equivalences, but does allow you to perform a simple Needleman and Wunsch sequence alignment and to apply zones automatically derived from that sequence alignment. This is done by issuing the `ALIGN` command. The sequence alignment is displayed, any currently active fitting zones are cleared and replaced by zones derived from the alignment.

Currently the `ALIGN` command may only be used if the structures contain only one chain.

Additional zones may also be specified in the usual way.

Clearly, it will normally be necessary to use the `ATOMS` command to specify that only backbone or Cα atoms are included in the fitting.

The `GAPPEN` command allows you to specify an integer gap penalty for the sequence alignment performed by the `ALIGN` command. The default value is 5.

7.2 Reading an Alignment

If you have an alignment performed outside **ProFit** you may use this to specify the equivalent zones. Any previously defined fitting zones are automatically cleared first. As with the `ALIGN` command, this can currently only be used with structures having a single chain.

The alignment should be a file in PIR format using – characters to align the sequences. The two sequences are represented by separate entries, i.e. each must have a header of the form:

```
>P1;xxxxxx
title text .......
```

If the PIR file contains multiple chains, it will be rejected. The first sequence will be assumed to be that of the the reference structure and the second is that of the the mobile structure. Any other sequences in the file are ignored.

The `READALIGNMENT` command is used to read in the PIR file.

7.3 Limiting Zones Read From an Alignment

When obtaining fit zones from a sequence alignment, either from ALIGN or from READALIGNMENT, it can be useful to limit the zones of residues used. Normally all aligned residue pairs will be used.

For example, if the alignment were:

```
         1         2         3
123456789012345678901234567890123
ASAHSTGEHNM--PLELLGHISLAM---NPRTY
---HSTADHNLRTPLEVLG--SLAMEDRQPRTY
```

the zones would normally be taken from the following positions in the alignment: 4-11, 14-19, 22-25, 29-33

By using the command:

```
LIMIT 20 28
```

only the zone from 22-25 would be included.

This is particularly useful in conjunction with the ITERATE command (Section 7.4) and when fitting multiple structures (Section 8).

The `LIMIT OFF` command restores the default behaviour of deriving the zones from the whole alignment.

7.4 Iterative Updating of the Fitting Zones

The ITERATE command switches on the iterative updating of fitted zones during subsequent FIT commands. The ITERATE command ma be followed by an optional parameter to specify the cutoff used to include or exclude pairs from the zones. (ITERATE OFF is used to switch it off again.)

Currently the ITERATE command may only be used if the structures contain only one chain.

Note that this immediately does an ATOMS CA since iteration of zones is only performed on Cα atoms. The program gives an informational message to this effect. See notes below if you want to calculate an RMSd over other atoms.

After the initial fit on the specified zones, the zones are updated such that residue pairs with Cα atoms within a specified cutoff (default 3.0Å) are included and those more distant are excluded. The optimum set of equivalences is obtained using a dynamic programming method.

After updating the zones, the structures are refitted and the procedure iterates to convergence of < 0.01Å, (typically 3 or 4 cycles). The RMSd on Cα atoms is shown after each cycle unless the QUIET command is given.

You may specify a minimal initial zone of say 3 amino acids on which to fit first. The zone iteration will expand the zones until as many residues as possible can be equivalenced. Alternatively, this option is particularly useful in conjunction with the ALIGN command. Using ALIGN followed by ITERATE gives a particularly convenient method of fitting two arbitrary structures.

As stated above, the ITERATE command implies ATOMS CA. Having fitted on Cα atoms, you can of course display the RMSd over other atom sets in the usual way using the RATOMS command (e.g. RATOMS N,CA,C,O will display the backbone RMSd).

Should you wish to refit on another atom set using the iterated zones, simply use ITERATE OFF to switch off iteration, select the atom set required using the ATOMS command and use FIT to refit the structures in the usual way. For example, to fit on backbone atoms:

```
ITERATE OFF
ATOMS N,CA,C,O
FIT
```

8 Multiple Structure Fitting

The MULTI command allows a multiple set of structures to be read in for fitting. The filename specified for MULTI is a 'file of files' i.e. it contains a list of filenames which will be read.

MULTI is used in place of REFERENCE and MOBILE to read in a set of structure files. The first structure file is used as a reference set for the first fitting stage, but the coordinates are averaged after each fitting stage to derive an averaged template used for subsequent fitting.

i.e. Given N files to fit, file 2 is fitted to file 1 and an averaged structure, A, is calculated, file 3 is then fitted to A and a new average, A' is calculated. This continues until all N structures have been fitted. The whole procedure iterates until convergence (typically 3 or 4 cycles).

Progress and RMSds are reported at each iteration unless the QUIET command is used.

The resulting fitted files are written with the MWRITE command. Note that there is no "reference" set in the sense used for normal 2-structure fitting; fitted versions of all N files will be written since the reference set is actually an averaged template.

When the MWRITE command is used, the output filenames are the same as the input files, but with the extension replaced by that specified in the MWRITE command. If no extension is specified, then '.fit' will be used. If the input structure files contained no extension, then the extension specified will be appended to the filenames.

Note that since only the extension is changed when writing back the fitted files, you must have permission to write to the directory from which the original files were read.

Note that multiple structure fitting and zone interation can be very slow as these have been added to the earlier pair-wise fitting engine. An increase in speed needs a complete re-design of the code.

8.1 Specifying Zones With Multiple Structure Fitting

Currently, the ZONE command may only be used with multiple structure fitting when the same zone specification may be applied to every structure. i.e. You cannot specify a zone for each structure separating the zones with a colon (:)

Thus, the following are legal zones:

```
ZONE 20-30
ZONE C,3
```

while the following are not:

```
ZONE 24-34:25-35
ZONE CAR:VNS
ZONE 24-34:EIR,11
```

For normal use, it is recommended that the ALIGN and READALIGN commands (possibly in conjunction with the LIMIT command) are used for specifying zones when fitting multiple structures.

9 Calculating the RMSd Over Other Zones and Atoms

Having fitted the structures using the ZONE and ATOMS commands to specify which residues and atoms should be included in the fitting, the RMS deviation may then be calculated over a different region of the structure and/or a different atom set.

This is achieved using the RZONE and RATOMS commands. The syntax of these commands is identical to that of the ZONE and ATOMS commands described in Sections 6 and 7.

As each RZONE or RATOMS command is given, the RMS deviation is reported over the new set of zones or over the new atom set. Don't forget the RZONE commands are cummulative, like the ZONE commands. Note that the RZONE * or RZONE CLEAR behaves slightly differently from ZONE * or ZONE CLEAR since it resets the zones to be the same as those specified for the fitting using ZONE, ALIGN or READALIGNMENT commands.

10 Obtaining Output

10.1 The Fitted Structure

The fitted mobile structure may be written to a file in PDB format using the WRITE command:

```
WRITE fitted.pdb
```

If the first character of the filename is a pipe character (|), then the results will be piped into the specified program. For example:

```
WRITE |less
```

will cause the coordinates to be displayed on the screen using the `less` pager program.

The reference set may also be written:

```
WRITE REFERENCE ref_fitted.pdb
```

(only the three letters 'REF' of the REFERENCE parameter are required). This is only useful if the CENTRE command has been used (see below).

10.2 Centering the Coordinates

By default, the mobile structure is moved to the coordinate frame of the reference set. If the `CENTRE` (or `CENTER`) command is given then the centre of geometry of the fitted coordinates will be located at the origin.

If only two structures are fitted then the `WRITE REFERENCE` command must be used to write the reference set in the origin-centred coordinate frame. If multiple structures are fitted and written using `MWRITE` then the reference set will be written automatically.

10.3 Details of the Fitting

More details about the fitting may be obtained by using the `MATRIX` command. This displays the centres of geometry, the rotation matrix and the translation vector which is the vector between the centres of geometry. Thus to superimpose the mobile structure onto the reference structure using these data, you should translate the mobile set to the origin, apply the rotation matrix, translate back to the original centre of geometry and finally apply the translation vector.

Note that the rotation matrix is not orthogonal and cannot therefore be used to extract Euler angles. This is a result of the fitting method used.

The `NFITTED` command displays the number of atom pairs which were fitted in the last fitting operation. Note that this will not be the number of residues fitted unless you are only fitting one atom type per residue (typically Cα atoms).

10.4 By-residue RMS Deviation

The `RESIDUE` command is used to obtain a by-residue RMS deviation on the currently specified RMS atoms in the currently specified RMS zone. If no `RATOMS` and `RZONE` commands have been used, the atoms and zones used for the fitting will be used.

The `RESIDUE` command may be followed by an optional filename parameter in which case output is directed to the specified file. If the file cannot be opened or a filename is not specified, output appears on the screen. If the first character of the filename is a pipe character (|), then the results will be piped into the specified program. For example:

```
RESIDUE |less
```

will cause the results to be displayed on the screen using the `less` pager program.

11 Modifying the Fit

Normally, no weighting is applied during the fitting i.e. all atoms are weighted equally. The `WEIGHT` command causes the fitting to be weighted by the mean of the B-values in the equivalent atoms. Normally, you wouldn't use this with real B-values, but with some other weight parameter (e.g. SSAP scores).

The `BWEIGHT` command weights the fitting by the inverse of the mean of the B-values in the equivalent atoms. This is useful for genuine weighting by B-values (i.e. the mobile set atoms will be less heavily weighted).

The `NOWEIGHT` command switches off weighting.

Atoms can also be removed from consideration in the fitting and RMS deviation calculations using temperature factors as a cutoff. The `BVALUE` command allows you to specify a B-value cutoff and any atoms with B-values greater than this value will be *ignored completely* in both the fitting and RMS deviation calculations. The B-value may not be higher than this value in either the reference set or the mobile set. For example, if you specify 10, then atoms with B-values greater than 10 will be ignored.

By specifying a negative value for `BVALUE`, you require that any atoms with B-values less than the absolute value you specify will be ignored. For example, if you specify −10, then atoms with B-values less than 10 will be ignored.

The value may be followed by an optional `REF` or `MOB` parameter which restricts checking of B-values to the specified structure.

12 Miscellaneous Commands

The `RMS` command may be used to reprint the RMS deviation over the currently defined set of RMS zones and RMS atoms.

Any operating system command may be run from withing **ProFit** by preceeding it with a `$`. The string following the `$` is passed to the operating system exactly as given and is useful for obtaining directory listings, typing, editing or copying files.

13 Command Summary

$ *command* Passes command to the operating system.

ALIGN Performs Needleman and Wunsch sequence alignment on the sequences of the two structures and derives zones from the equivalent regions in the alignment.

ATOMS *atm[,atm]...* Specifies the atom subset to fit.

BVALUE *cutoff* [REF|MOB] Specify a B-value cutoff. Any atoms with B-values greater than this value will be ignored completely. A negative cutoff specifies that atoms with B-values less than the absolute cutoff should be ignored. The optional `REF` or `MOB` parameter restricts B-value checking to the specified structure.

BWEIGHT Weight the fitting by the inverse of the mean of the B-values in the equivalent atoms.

CENTER [OFF] See `CENTRE`.

CENTRE [OFF] Cause the coordinates to be written (using the `WRITE` or `MWRITE` commands), with the centre of geometry located at the origin instead of in the same coordinate frame as the reference set.

FIT Performs the actual fitting. Returns the RMS deviation over the atoms included in the fit.

GAPPEN *val* Specify an integer gap penalty for the sequence alignment performed by the `ALIGN` command.

HETATOMS Read HETATM records with subsequent `MOBILE` and `REFERENCE` commands.

IGNOREMISSING Ignore any atom mismatches and proceed with the fitting. Such atoms are listed as warnings.

ITERATE [(*limit* | OFF)] Switches on (or off) iterative updating of the zones for fitting. The `ITERATE` command may be followed by an optional distance cutoff (default: 3.0Å) or by the keyword 'OFF' to switch off iterative zone calculation.

LIMIT (*pos1 pos2* | OFF) Limits the range in an alignment (from ALIGN or READALIGNMENT) used to derive zones. `LIMIT OFF` restores the default behaviour.

MATRIX Displays the centres of geometry, rotation matrix and translation vector.

MOBILE [XMAS] *filename* Reads a mobile PDB structure. If compiled with XMAS support, then the XMAS keyword specifies that the input is in XMAS format.

MULTI *filename* Reads a file of files containing a list of structures for multiple fitting.

MWRITE [*ext*] Write the results of multiple structure fitting. The structures are written back using the same filenames with which they were read, but with the extension changed to that specified. If no extension is given, then '.fit' is used. Note therefore, that you must have write permission to the directory from which the input files were read.

NFITTED Reports the number of atom pairs fitted.

NOHETATOMS Do not read HETATM records with subsequent `MOBILE` and `REFERENCE` commands.

NOIGNOREMISSING Restore the default behaviour of issuing an error message for any atom mismatches and halting the fitting proceedure.

NOWEIGHT Normal, non-weighted fitting.

NUMBER (RESIDUE|SEQUENTIAL) Specifies whether zones are based on residue numbers in the PDB file or on sequential numbering (running through all chains).

QUIET [OFF] Switches on (or off) quiet mode. In quiet mode, warning messages are suppressed and progress of iterative zone updating and multiple structure fitting is not reported.

QUIT Exits from the program.

RATOMS *atm[,atm]...* Specifies atoms over which to calculate the RMS deviation. Fitting must already have been performed.

READALIGNMENT *filename* Reads an alignment in PIR sequence file format and sets zones based on that alignment.

REFERENCE [XMAS] *filename* Reads a reference PDB structure. If compiled with XMAS support, then the XMAS keyword specifies that the input is in XMAS format.

RESIDUE [*filename*] Gives a by-residue RMS deviation. If the first character of the (optional) filename is a pipe character (|), then the results will be piped into the specified program. For example:

```
RESIDUE |less
```

will cause the results to be displayed on the screen using the `less` pager program.

RMS Recalculate the RMS deviation over the zones and atoms currently defined with `RZONE` and `RATOMS`.

RZONE *zonespec* Adds a zone specification to the list of zones considered in RMS deviation calculation. `RZONE *` or `RZONE CLEAR` resets the zones for RMSD calculation to be the same as that specified with the `ZONE` command.

STATUS Reports current program status.

WEIGHT Weight the fitting by the mean of the B-values in the equivalent atoms.

WRITE [REFerence] *filename* Writes the fitted structure to a PDB file. If the first character of the filename is a pipe character (|), then the results will be piped into the specified program. For example:

```
WRITE |less
```

will cause the coordinates to be displayed on the screen using the `less` pager program.

If the `REFERENCE` keyword is given (only the letters 'REF' are required), then the reference set will be written. This is used in conjunction with the `CENTRE` command.

ZONE *zonespec* Adds a zone specification to the list of zones considered in fitting. `ZONE *` or `ZONE CLEAR` removes all zone specifications.

14 Copyright

Please note that the program is called **ProFit** — not PROFIT, Profit or profit; this attempts to avoid confusion with the threading program known as PROFIT. **ProFit** was written first and released to the public around the same time.

ProFit was initially written by me while self-employed and trading as SciTech Software. As such it is owned by me alone. Under the conditions of the copyright statement included in the code, enhancements made since at UCL also become my property.

If you did not register the program via the web site, you are requested to send EMail to the author to say that you are using this code so that you may be informed of future updates.

The code may not be made available on other FTP or Web sites without express permission from the author.

The code may be modified as required, but any modifications must be documented so that the person responsible can be identified. If someone else breaks this code, the author doesn't want to be blamed for code that does not work! You may not distribute any modifications, but are encouraged to send them to the author so that they may be incorporated into future versions of the code.

Such modifications become the property of Dr. Andrew C.R. Martin and SciTech Software though their origin will be acknowledged.

While the compiled **ProFit** program may be used by commercial companies, it may not be sold commercially or included as part of a commercial product. The source code or any derivative works may not be sold commercially or used for commercial purposes outside of **ProFit** without prior permission from the author.

This manual was written while at UCL and updated while at Reading as part of consultancy work to Inpharmatica. It is based on the help file included with the program. As such it is joint property of Andrew Martin, UCL and Inpharmatica.

While this software is provided "as is" and free of charge, I do appreciate hearing from people who use it and find it useful. An EMail or a postcard would be nice.

If you find **ProFit** useful, please tell your colleagues about it. Please *do not* pass copies of **ProFit** on to them directly; ask them to obtain it *via* my World Wide Web page (`http://www.bioinf.org.uk/software/profit`)

15 How do I Reference ProFit?

No paper has been published describing **ProFit** itself since it is simply a convenient program (I hope) to let you use a standard fitting algorithm; consequently, it is a little difficult to reference. The exact wording is up to you and dependent on the context, but I suggest something similar to:

> Fitting was performed using the McLachlan algorithm (McLachlan, A.D., 1982 "Rapid Comparison of Protein Structres", Acta Cryst A38, 871-873) as implemented in the program ProFit (Martin, A.C.R., `http://www.bioinf.org.uk/software/profit/`)