

MAID - Copyright David G. Levitt, 2000

Outline:

- I) **Overview.**
- II) **Compiling.**
- III) **Directions for “quick” MAID version.**
- IV) **Screen Version - Pull Down Menus:**
- V) **Use the screen version (“MAID”) to set up the conditions for the batch version and create the "maidfile".**
- VI) **Use the screen version to select the bone region that you want to use for the autofitting routine.**
- VII) **Select the map region that you want to use for autofitting.**
- VIII) **Run the batch version of MAID:**
- IX) **Viewing the output using the screen version.**
- X) **Format of fit file:**

I) Overview.

There are 2 separate programs; 1) Batch Version (“batchmaid”); and 2) Screen Version (“maid”):

1) Batch Version: This version is written in standard C++ and does not require any graphics software. The most recent version is compatible with most Unix and Linux compilers in addition to the SGI C++ compiler. It is the only version that is required for the “quick” MAID run.

2) Screen Version: This only runs on SGI machines and uses OpenGL and Motif. Although the entire fitting can be run from this version, usually the “batch” version is used for the actual run. The screen version is used primarily for debugging the program. It can also be used to set up the initial conditions for the run, to evaluate intermediate aspects of the output and to interactively edit and modify the fits. The use of this screen version provides a convenient way to view and modify the output of the batch version.

II) Compiling:

To compile the two versions:

1) Batch Version: Open the file `makefile_batch` and select the compiler you want. Five compiler options are listed:

```
Cgeneric = CC #SGI compiler - default
#Cgeneric = gcc # Linux compiler
#Cgeneric = g++ # gnu Linux
#Cgeneric = pgCC #Portland compiler
#Cgeneric = cxx #Compact Tru64 alpha
```

Pick the option you want by using the comment symbol (“#”) and save this file. Then copy “`makefile_batch`” to “`makefile`” and enter “`make`” - and this will output the batch version “`maidbatch`”.

1) Screen Version: Copy “`makefile_screen`” to “`makefile`” and enter “`make`” - this will output the screen version “`maid`”.

Depending on the compiler version - you will get a number of “Warnings” during compilation - hopefully these can be ignored.

III) Directions for “quick” MAID run.

A “quick” version has been established to allow the user to run MAID with an absolute minimum of user input. Only a few minutes of user time is required to set up the MAID run which, without any other user intervention, should output a pdb structure. Use the following steps:

A) Make an XPLOR (or CNS) formatted normalized (mean = 0; stddev = 1) map over the region you want MAID to build the fit.

For the typical case where you do not have any information about the location of the molecule, use a default map that covers a region equal to 1 asymmetric unit – with an additional 5 to 10 Angstroms added to each boundary of the asymmetric unit. This will insure that MAID will be able to find at least 1 complete fit. If there is more than 1 molecule per asymmetric unit, MAID will find all of them. This is particularly useful if the map is not averaged, since MAID tries to use all the different fits to improve each fit. MAID also, in the last step, will extend any of the partial fits into complete fits. The only disadvantage of using this large default map is that it will require longer computer time (even large maps for typical proteins should not require more than about 24 hours of computer time). It is important to use a large enough map because, e.g., if any part of a side chain is out of the map region, then this region cannot be fit.

B) Enter: “maidbatch ‘name’ setup” - where name is the name that you will use in subsequent calls to MAID.

This will convert the map file to a special binary file (.den1), create the default skeletonization (bone) file and make the file "**name.maid**" that is required for all subsequent runs. This call will ask you for the name of the XPLOR map. It will then ask if you have a pdb file you want to use. This pdb file is not used in the fitting. It is only used in the graphics version to allow you to compare your MAID generated fits with a tentative pdb. Finally it will ask you for “the skeletonization cut off density value”. This is the only adjustable parameter in MAID. A bone file will be made using this cut off value. This bone file is used to find initial trial starting positions for the sheets and helices. Only bone traces that are 3 residues long will be tested. For good maps, a value of 1.4 or greater should be used, for poorer maps try 1.2. The graphical version of MAID provides a convenient way to look at different values of this cutoff density. In all subsequent runs of MAID, only the binary (.den1) map file is used, so that the XPLOR map can be deleted at this point.

Example: the file “test.maid” created by entering “maidbatch test setup” is shown below (the map name input was endodm24aug98.map):

```
Number of pdbfiles: 1
pdbfile1: endo_18.pdb
Number of denfiles: 1
denfile1: endodm24aug98.map.den1
bonefile1: endodm24aug98.map.den1.bone1.2
spherefile: test.maidispherefile
fitfile: test.maidfit
editbonesfile: test.maideditbones
```

```
pdbsequencefile: test.maidpdbsequence
symrottranfile: test.maidsymrottran
selenometfile: test.maidseleomet
```

The "spherefile" is the name of the file containing a set of spheres to limit the region that is searched for the initial sheets and helices. If "setup" is used, as above, spherefile will just have the entry "-1", indicating that no spheres should be used and the entire map should be searched. The graphical version of MAID provides a convenient way to choose a set of spheres and create the spherefile, or you can just input a set of spheres you have determined by another procedure. A sample sphere file is shown below (the first line is the number of spheres, and next lines are the sphere radius and x, y, z center).

```
3
23.349991 35.844566 48.328846 27.298893
9.250002 18.024784 25.139563 37.160847
16.849995 50.979294 36.292030 51.224781
```

The "fitfile" is the name of the default file that the MAID fits will be written to.. This default file can be overruled and the output fits written to another file as described below.

The symrottranfile contains the set of rotation/translation operators that MAID uses to superimpose the different subunits that it finds.

The selenometfile is not used in this version.

C) Paste the 1 letter amino acid sequence into the file corresponding to "pdbsequencefile" in the maidfile.

This file is required for the "extend" routine of MAID, where the sheets and helices are extended through the loops. It is not required for the "trace" routine that finds the sheets and helices.

Example: for the example maidfile "test.maid" above – the pdbsequence file is test.maidpdbsequence. A sample file is shown below:

```
10
SKSSTASASAK
KIIVKHVTVIGGGLMGAGIAQVAAATGHTVVLVLDQTEDILAKSKKGIEESLRKVAKKKFA
ENPKAGDEFVAKTLSTIATSTDAASVVHSTDLVVEAIVENLKVKNELFKRLDKRAAEHTI
FASNTSSLQITSIANATTRQDRFAGLHFFNPVPSMKLVEVIKTPMTSQKTFESLVDFSKA
LGKHPVSCKDTPGFIVNRLLPYLMEAIRLYERGDASKEDIDTAMKLGAGYPMGPFELLD
YVGLDTTKFIVDGWHEMDAENPLHQPSPLNKLVAENKFGKKTGEGFYKYKAA
LEHHHHH
```

The first line contains just the number (e.g. 10) that corresponds to the first residue listed in line 2. That is, residue #10 is serine, #11 a lysine etc. In most cases this number will just be 1. (It is essential to have this number in the first line).

D) Run MAID using maidbatch:

There several different options that can be used at this point. All of the following examples use the "test.maid" file described above. Since these runs may take 24 hours or more, they are usually set up as batch jobs, with the output redirected to some file.

1) "maidbatch test.maid complete"

This will run the complete MAID fitting routine: "trace" (finds the sheets and helices); "extend" (assigns the amino acid residues and extends through the loops); and "ncs" which expands all the fits using the rotation/translation operators and tries to

group the pdb fits into segments. This command will write the MAID “fits” to the default fitfile listed in "test.maid" (i.e. test.maidfit). If you prefer to write it to some other file, use the command: “maidbatch test.maid complete filename” where “filename” is the name of the file the fits will be written to. When “complete” starts, it starts from nothing, so that the fitfile will be written over, and any previous fits in it will be lost. Some of the other commands below use the previous fits and will add to or modify the fits in the fitfile. In the commands below, “filename” can be omitted and MAID will write to the default fitfile.

- 2) **“maidbatch test.maid trace filename”** where filename can be omitted and the fits will be written to the default fitfile. This command just finds the set of sheets and helices. Since it does not assign the amino acid sequence, the side chains in the fits consist simply of C-C...C atoms.
- 3) **“maidbatch test.maid extend filename”**. This will start with the fits in filename (or in the default fitfile if filename is omitted) that were returned by the "trace" option and assign the amino acids and extend the fits. MAID is written so that it can be restarted any time, automatically continuing on at the point where it stopped. Thus, this command can also restart a previous “extend” run, just continuing on from where the previous run stopped.
- 4) **“maidbatch test.maid ncs filename”**. This will apply the rotation/translation operators to all the partial fits, expanding the fits as far as possible. It also tries to assign the different fits to corresponding pdb segments. It is normally used after the call to “extend” so that it is using fits that have the amino acid residues assigned. **The symmetry operator will be applied to a fit only if at least 50% of the fit lies in density.** Thus, it is important that the map that was originally input into maid is large enough to cover both subunits.
- 5) **“maidbatch test.maid trace_cont filename”**. This command is used if, for some reason, MAID was interrupted (or crashed) during the “trace” run. This command will continue on from the last fit in the fitfile.

E) View the MAID output.

MAID uses its own unique “fitfile” to keep track of the structure being built into the map. This fitfile can be viewed using the MAID graphics version. MAID also converts the fitfile into a corresponding standard pdb file that can be viewed using the standard visualization programs. **There is one important difference between these two structures that it is important to remember.** The advancing ends of the fitfile terminate with a Calpha atom. Thus, it is not a complete residue. When MAID converts the fitfile to a pdb file, it adds a default glycine at each end so that this information is not lost. (When MAID converts a pdb file to a fitfile, these glycines are removed.) Thus, when you look at the output pdb file, there will be extra glycines inserted at the ends so that the sequence will differ from the actual amino acid sequence. Depending on how you use this output, you may want to delete these glycines.

The output of the “complete” run consists of several fitfiles and pdb files:

- 1) The default fitfile (or the file name = "filename" used above) which is the current fit file at the time the program ended. It is the file that should be used if, for example, the program is restarted, or if one runs “extend” after “trace”.

- 2) fitfile_trace (and the corresponding pdb file fitfile_trace.pdb) which are the structures after the helices and sheets have been found. Since no residue assignment has been made at this point, the pdb file consists only of alanines or serines.
- 3) fitfile_extend which is the fitfile after the extend run has been completed.
- 4) fitfile_extend_seg (and the corresponding pdb file fitfile_extend.pdb) which is the fitfile after extension and an attempt to assign the fits to pdb segments.
- 5) fitfile_expand and fitfile_expand_seg (and the corresponding fitfile_expand.pdb) which is the output after all the rotation/translation operators have been applied to the extended fit.

F) Summary.

For most purposes, this is all you need to know to run MAID. The information below provides some brief information about using the graphics version ("maid").

IV) Screen Version - Pull Down Menus:

Enter "maid" to start the graphics version. A cross-eyed stereo image should appear. The directions for rotation, translation, slabbing, picking and zooming are shown on this screen. **Note: some options, such as picking and rotating about specific axis require that the pointer be in the left window.**

There are a large number of pull-down menu options in MAID - most of which are only used for evaluating code. They are somewhat self-explanatory. The options that you will probably find useful are listed below:

PDB Menu:

“Center pick” will center on picked “pdb”, “fit” or “bone”

“Center input” will center on the residue number you input

“Window on next pick” will window the pdb and bones about the picked or input residue.

“Make fit from displayed pdb" will make a fit file that can be input into the autofitting routines.

MAP Menu:

“Map1” options: Turn on/off density contour or bones. . Sets the contour radius, level and color. “Set min density for bones” sets the skeletonization cutoff value that is used for making bones, and will make and display the new bones.

BUILD menu:

“Read fit” reads the default fit file (specified in the X.maid file) to the screen

“Pick Active” activates fit for torsional dynamics, etc.

“Write fit as PDB” converts the fit to XPLOR PDB format. The fit files may have a special “segment” designation that is converted to standard PDB segments **by** this option.

SPHERE Menu:

“Add sphere” will draw a new standard sphere around the picked bone or pdb point.

“Save spheres” will save the sphere data to sphere file specified in X.maid.

“Recall spheres” will display the spheres saved in the spherefile and allow them to be edited

V) Run the screen version ("MAID") to set up the conditions for the batch version and create the "maidfile".

1. Enter ".maid" - this should bring up split screen image.
2. Click on pull down menu "FILE"; choose "NEW" and ENTER NAME: enter name for this data set (e.g. test).
3. ENTER NAME OF PDB FILE: if you have a pdb file, name in standard format pdb file. MAID uses lines that begin with "TER" to indicate different segments. (The segment info in the data line is ignored.)
4. ENTER NAME OF MAP FILE: this must be in XPLOR (or CNS) normalized format (e.g. endodm24aug.map). You must enter either a pdb or map file, but you do not need both. If you have other pdb or map files they can be entered using the ADD PDB and ADD MAP option.
5. When you click on OK, MAID will read the ASCII map file (if you entered a map) and create a binary ".den" file. In the future, when you start MAID, click on "OLD" and enter your data file name (e.g. "endodm") and MAID will use restart. If you entered a PDB file, this pdb will be centered on the screen. If you did not enter a PDB file, the screen will be blank until you make a bone file by pulling down the "MAP" menu, and clicking on MAP1 and BONE. This will make a skeletonization version of your map using the default value of 1.4 std. dev. as the minimum bone density. (To use other minimum densities - use the "SET MIN DENSITY" option in the MAP1 menu).

At this point you will have created and saved a ".maid" file (see, e.g. test.maid in IIIB above).

Once this file is created, the ASCII map file can be deleted. The names listed in the .maid file are default names - they can be changed to whatever names you want.

VI) Select the bone file that you want to use for the autofitting routine.

The bones are used to find the initial test position for the sheets and helices. Since a position will not be tested unless there is a continuous bone at least 3 residues long, it is important to choose the appropriate skeletonization level. The choice of the minimum density used to make the bones is the only adjustable parameter in the entire routine.

Use "Set min density for bones" option in the MAP/MAP1 menu to look at different minimum skeletonization densities. You should look at bones for various minimum densities (1.0 to 1.6) and choose a value to use in the fitting routine. I have found that 1.2 works well for average quality maps, 1.4 for good maps. If the 1.2 bone set does not have reasonable long continuous stretches, then MAID probably will not be very successful. After you have found the appropriate bone level, open the ".maid" file and insert the desired bone name in the bonefile1 position.

EXAMPLE: bonefile1: endodm24aug98.map.den1.bone1.2.
This tells MAID to use the 1.2 std. dev. minimum density bone file for the "Trace" routine.

VII) Select the map region that you want to use for autofitting.

1. Run MAID and bring up a bone file.
2. Click on the "ADD SPHERE" option in the "SPHERE" pull down menu. With the left mouse button (**in the left window, only!!!**) click on a pdb or bone point. A sphere will pop up. Then click on "MOVE AND SIZE" and use the middle mouse button (but2) to translate in the X and Y positions, and the right mouse button to translate in the z position (horizontal mouse movement) and change sphere radius (vertical mouse movement). If the mouse is in the right window, you can use the standard mouse movements to rotate and translate the entire map, then move to the left window and adjust the sphere position and size.
3. Continue adding spheres until you have covered the desired region.
4. Be sure to "SAVESPHERES" when finished. This will write the sphere file that will be used by the trace routine when the batch version is called.

Ideally, one would like to pick a region that covers one complete subunit and not much else. However, this is not usually possible unless one has some detailed prior information about the boundaries of the subunit. The best procedure is to choose a region that may overlap several subunits but is large enough that a complete subunit can be put together from the overlapping regions. MAID will automatically look for symmetry operators (either crystallographic or non-crystallographic) and use these operators to synthesize a complete subunit from the overlapping sections. This routine may get confused if there are more than two regions from the same subunit position - so this places some limits on the size of the chosen region. In general, start with a large region - all it costs is a little computer time!

Picking a region that covers 2 subunits is especially advantageous if the map has not been averaged. MAID will combine the best parts of each of the different subunits to synthesize a complete subunit. For example, if a loop has poor density in one subunit and good density in another, MAID will use the information from the good density region to fill in the poor density loop.

VIII) Run the batch version of MAID as described in section III D above.

IX) View and edit the output of MAID.

As described above in section III E above, MAID uses its own "fit" format to keep track of the structures being built into the map. These fits have begin and end with Calpha atoms. At various steps in the program, these fits are converted to standard PDB format that can be visualized using standard routines. This conversion adds a glycine at each end (see III E). The fits themselves can be viewed using the graphics routine. Under the "BUILD" menu, choose "Read fit". This will read the fit file that is listed in the name.maid file (i.e., for test.maid, the default fitfile is test.maidfit).

X) Refinement of this protein structure.

For now, no special routines are available for the refinement steps. You should use the PDB file output from the last ("ncs") routine as input to CNS, XPLOR or whatever refinement program you prefer. Preliminary tests suggest that it is important to use some sort of phase recombination when making an improved new map to reinput into MAID, otherwise the map will be too heavily biased by the initial MAID fit.

For example, for fit 2 the first side chain (= -12) is a LYS, etc.

If an amino acid sequence has been assigned, the rotomer type is listed in the sixth line (after refinement, this is set = 0).

The next lines list the atom positions.