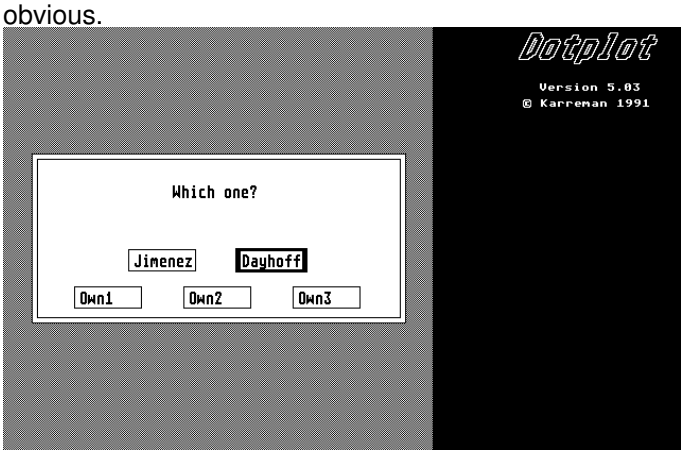
To explain this option I will have to tell you something more about the various methods of comparing proteins and their amino acids.
When DNA files are compared the scoring is fairly simple: for every identical base the score is incremented. This method is also available for proteins, but there are other options as well. Some amino acids are chemically more related than others; Glycine (R-H) is nearer to Alanine ($R\text{-}CH_3$) than to Cysteine ($R\text{-}CH_2\text{-}SH$). This fact can be expressed either as a fraction or as equality within a group.
An other approach is to score for evolutionary relatedness. This means two processes have to be considered and expressed in a number. First, the chance of a certain codon mutating into another has to be calculated, and secondly, the fitness of this mutation has to be assessed. Both the chance and the fitness have to be expressed by a single number.
Both the chemical and the evolutionary method have been incorporated into DOTPLOT.
The chemical scoring table is called ‘‘JIMENEZ’’ after the man who described it first (1). It does not score for individual amino acids but divides them in groups.

The groups are:

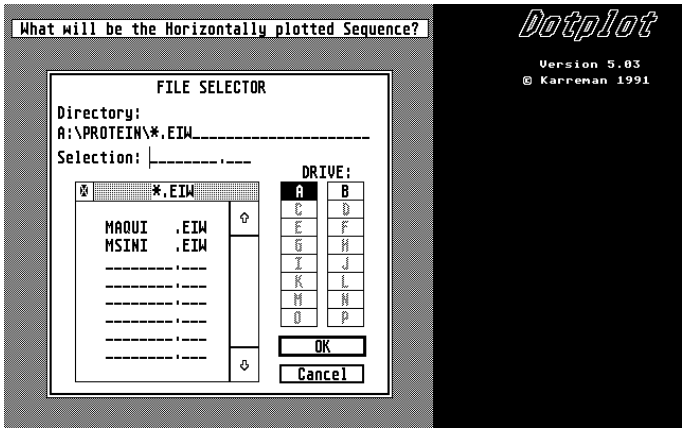
PAGST	: neutral,weakly hydrophobic
QNEDBZ	: hydrophilic, acid amine
HKR	: hydrophilic, basic
VILM	: hydrophobic
FYW	: hydrophobic, aromatic
C	: cross-link forming

All amino acids within the group score equal (=1), between groups they score 0.
The evolutionary approach is represented by ‘‘DAYHOFF’’ again named for an important contributor of this work (2), this is a completely individual scoring table. The relatedness of every amino acid with every other amino acid is expressed as a number between 0 and 2.73.
There are three more tables available in DOTPLOT, the contents of which, as well as their names, defaults and comments can all be changed. So if you feel you have developed an improved scoring system you can change one of these tables to fit, complete with an appropriate name and defaults. If you choose to use a scoring table the next step of DOTPLOT is obvious.



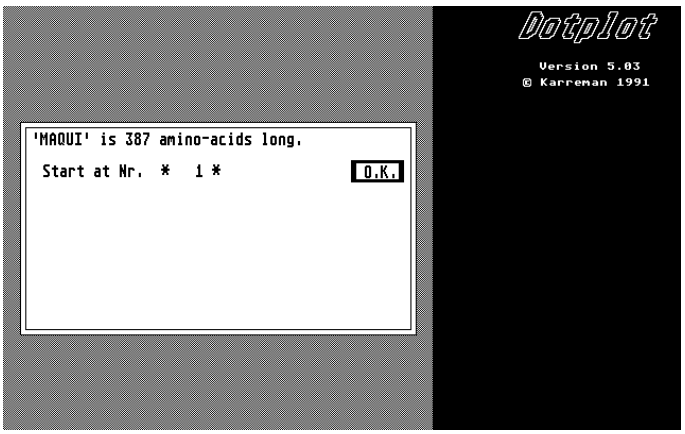
The various scoring-tables.

The program will ask you what table to use. As you see the tables you can change are called Own1 to Own3. This name can be changed by you into any other, so if you get this program as a copy don’t be alarmed if these boxes contain different names. This is also true for a lot of the numbers that are given as default values in the next couple of screens. If they are not the same as described, somebody probably has changed them to fit his/her needs. If you don’t agree with these new values or if you don’t agree with mine you can change them yourself (see page 10).
After you have chosen your table, or if you have selected to use no table and did not see the above screen, you will get the oppertunity to select the sequence that will be plotted horizontally.



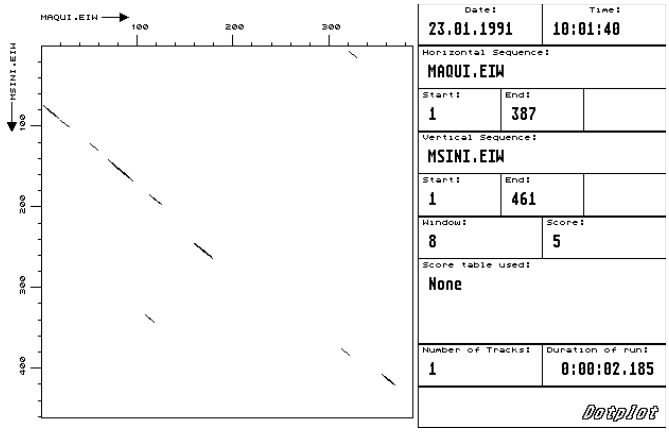
The first sequence input.

And welcome back to those of you that chose to select DNA instead of protein. The picture you get at this stage is a little different but as it is completely analogous you probably will be able to overcome the discrepancies. So for the rest of the story replace "amino acids" by "bases" and it all will be very straightforward. If the differences get to great I will, of course, explain them to a greater extent. The screen, as can be seen above, is filled with a standard selection box and as soon as you have selected a sequence the next screen will appear and look like:



Borders of choosen sequence.

The number between the asterisks is the default value; you can get this value by pressing <RETURN>, or by clicking the left mouse button while putting the arrow in the OK-box. If you want to start at another amino acid just type in the number of the first residue of your selected range and press <RETURN>. The next line will appear at once asking for the last amino acid to be in your selection, the default option (the last residue of the total sequence) is again given between asterisks.
If you are running DNA there will even be a third line asking if you want to use the reverse sequence of the strand on your disk. Another way of doing this is by giving a higher number to start with, than to end with; the program will then automatically reverse the strand. The numbering of the reverse strand will be that of the whole sequence; so if you select 100 bases at the beginning of a sequence of 1000, the numbering will run from 901 to 1000 after reversing.
After the horizontal sequence is selected the whole process is repeated for the vertical sequence. The whole procedure is completely analogous to that of the horizontal sequence, so I doubt if there will be any major problems.
The next input that is asked for is the window size. The question is accompanied by its own default values and can be treated in the same way as the question about the borders of the sequences. The window is that small stretch of one sequence that will be compared to all possible stretches of the other sequence of exactly the same size. The number of residues that are identical in that stretch are a measure of homology. The threshold score for this window is asked next. In proteins the defaults are: window=8 and score =5; so if 5 out of 8 amino acids are identical then there is enough similarity. For scoring amino acids also see page 6



The result of a run.

As soon as this last information has been given to the program, it starts to calculate the degree of similarity and gives the output shown above. Every stretch of similarity is represented by the diagonal in the left boxed area. The right-hand part of the picture is reserved for the statistics of your run; date, time, what sequences and score table you used but also how long it took for DOTPLOT to generate the picture, in this case just over two seconds. There is also a little box that bears the name "tracks", it probably will be one in most cases, only for very long sequences it will increase. This has to do with the fact that DOTPLOT uses a lot of memory, so much in fact that if you study long sequences there is not enough room in your computer for the program to store all its tables. In this case, several tracks will be laid on your screen to build up the complete picture, and these will be counted. You don't have to worry about this process, everything will be calculated by DOTPLOT itself; as a user you will only notice the way the picture was built and nothing of the hard work.
As soon as the picture as shown above is completed, a panel is slid in from the right, you probably will have trouble getting a good look at the original picture before this happens. But don't worry; it is not lost, it can be brought back.
DOTPLOT is now in a mode were you can use the mouse to communicate directly with the program. When positioned over the panel the mouse is represented by an arrow, when over the picture by a cross. In the latter case, when over the picture, two commands are possible: Zoom-in and Show-homology.

Zoom-in. You can zoom in any part of the picture, just place the cross on the left hand top corner of the area you want to enlarge, press the LEFT mouse button and drag the mouse (while you keep the button pressed) downwards and to the right. While doing this you will see that the surrounded area is boxed in. As soon as you release the button the boxed area will be blown up to full size. Note that the panel is affected by this; some or all of the lettering of the rectangle marked "Borders" are now fully black.

Show-homology. If you place the cross directly over a diagonal representing a stretch of homology and press the RIGHT mouse button, the program will show the two corresponding sequences around this stretch. This form of output can be printed directly or saved as an ASCII file. This will enable you to incorporate it into other texts using almost any word processor. You can return to the original screen by either using the "EXIT" button or by pressing the RIGHT mouse button again.

```
MAQUI        3  KKLISLFSGAGGMDIGFMA     21
                  |  | |||| | | |
MSINI       75  PKALSFFSGAMGLDLGIEQ     93
```

Show-homology.

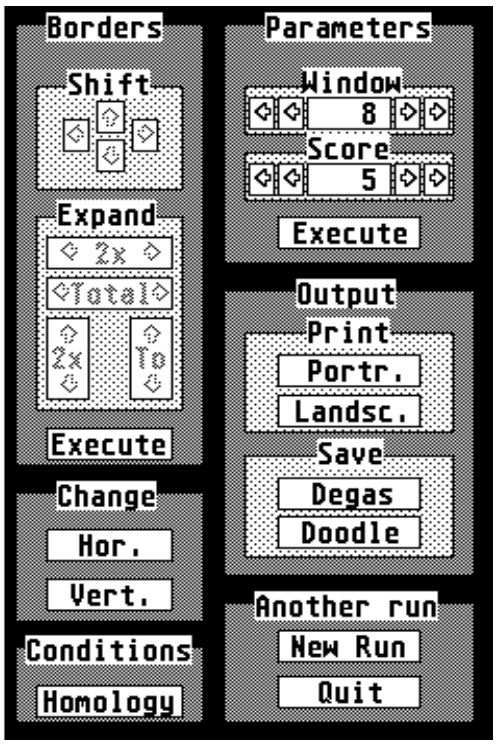
If you are using protein files and a scoring table you will notice that only perfect matches are indicated by vertical lines, and not the related amino acids. Even if not shown the program does use the scoring table you selected, so do not be alarmed about differences between the graphic and the litteral output. DNA users might notice if they occasionally run RNA against DNA sequences that T=U and this couple will be honored by a vertical line.

All other commands can be entered by use of the panel on the right of the screen. The commands are grouped, and these groups are visualized by placing them in the same box.

Borders. These options are all about the borders of that part of the sequences that is shown. The lettering in this box is either black or grey. The button is only active when black, and will not react to you pressing the left mouse button when it is grey. If you did try the Zoom-in option, and selected a small portion of the original picture, all or most will be black.

Shift.The first subgroup of "Borders" is The "shift" option. With the four arrows in this block you can shift the borders of your selection. If you zoomed in and ended up with a picture that runs horizontally from 100 to 200 and now you shift to the right the new borders will be 150 to 250. So the total length remains the same, only the borders shift. The program will NOT do this straight away, but only after you pressed one of the "Execute" buttons. This allows you to press combinations of shift, i.e. to the right and down. It also makes it possible to combine shift and expand.

Expand. As for "Shift", this option is only active when the lettering is in black. Expand is a kind of zoom-out; you can enlarge the area that is represented in the picture by a factor of two -horizontally and/or vertically- or you can take one or both sequences in their total. The program will NOT do this straight away, but only after you pressed one of the "Execute" buttons.

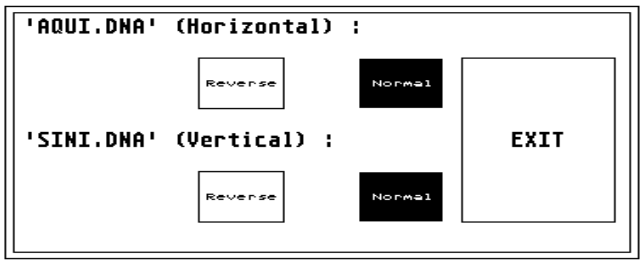


The panel.

Change. The change block has only two options; you can replace the horizontal or the vertical sequence by another one. The way DOTPLOT asks you for this input is exactly the same as before. It is also possible to use only a part of the sequence and, if it is a DNA file, to reverse the sequence.

Conditions. There is only one box, but it depends on whether you have chosen DNA or protein what this option is. In case you choose protein it is "Homology" and it allows you to change or activate a score table. In fact you will be asked if you want to use one, so this is also the way to inactivate a score table. For score tables see also pages 6 and 10.

If you are studying DNA sequences the option will read: "Reverse". Using this box will enable you to reverse one or both sequences. As described on page 7, the numbering will be changed! RNA files will be changed into DNA by the reverse process; since DNA and RNA are completely compatible, as far as DOTPLOT is concerned, you don't have to worry about this.



DNA :Reverse box.

Parameters. The two most crucial parameters of a DOTPLOT run can be changed within this box. This can be done by clicking in the little arrow-boxes. The innermost will either increment (right) or decrement (left) the value of the window or the score by one. Clicking in the outmost will change the values by 10%+1. The program will NOT use the altered parameters, until you have pressed one of the "Execute" buttons.

Output. There are two mayor ways to save the result of your run: 1.) hardcopy on paper or 2.) as a file that can be accessed by other programs.

Print. When you click in one of the two boxes with this heading DOTPLOT will make a hardcopy of the picture and the relevant information about it. It will, in fact, be the picture on page 8. The orientation of the hardcopy can be either in the portrait or the landscape format.

Save. The picture, see top of page 8, can be saved in two formats that are compatible with popular programs. Degas format is not compressed and will use 32034 bytes of memory for every picture. Doodle will take

32000 bytes of memory and so, obviously, is not compressed either. The names of these files can be typed in on the prompt and saved on any disk.

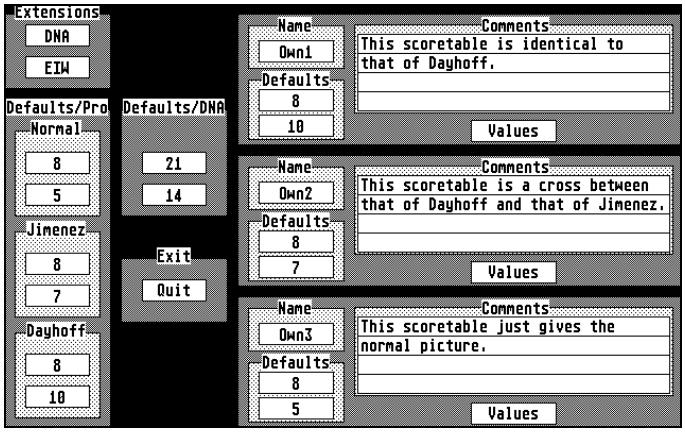
Another run. Here the most extreme commands are grouped. It will take you hours of playful use of DOTPLOT before you will want to use them.

New Run. This one is not so bad. It takes you back to your first real choice: DNA or protein.

Quit. This one is terrible: it does just what the name suggests.

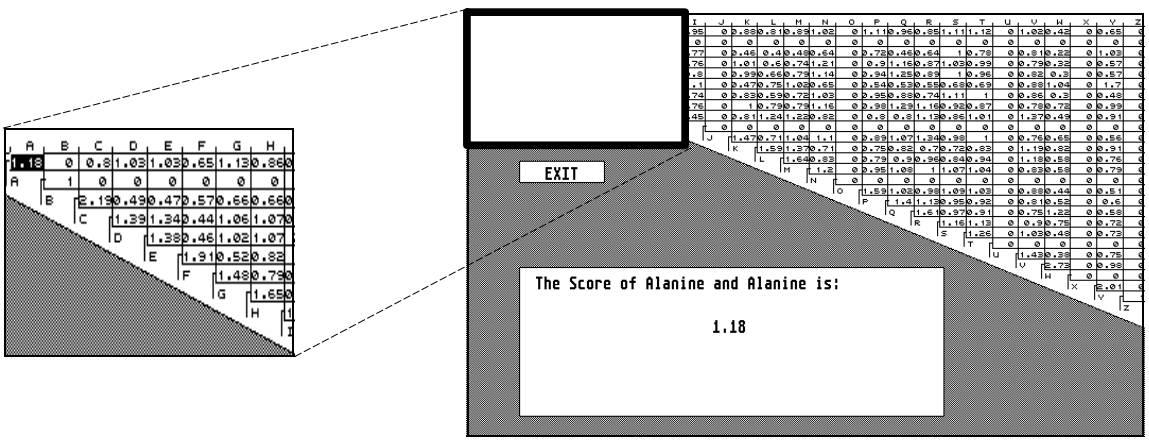
The DOTPLOT Editor.

In the DOTPLOT folder there is a second program called DTPLT_ED.PRG. This program enables you to change the defaults themselves of the DOTPLOT program. On running the editor you will get a screen looking like:



First editor screen.

You will recognize a multi button panel if you see one by now, so I guess this one will not give you too much trouble. The top-left block is for changing the default extensions of DNA and protein files, respectively. If one of these two buttons is clicked upon the program enters the editor mode and you are given the opportunity to edit the old extensions. Likewise you can edit the window- and score-values of the three standard protein score tables, these are all in the elongated box on the left site of the screen. In the middle are two small boxes; one to change the window and score of DNA comparison and the other the Quit option; to stop and leave the program. The whole of the right of the screen is dedicated to the three extra score tables. Of these the names, windows, scores and comments all can be edited in the same way as with the other tables. An additional option is



Second editor screen and partial blow-up.