As you can see it lists the complete set of values of the score table and this will fill almost half of your screen with, all but illegible lettering. You will also find an exit to return to your previous screen and an edit box. When you enter the second screen, the edit box reports on the Alanine-Alanine couple, and if you look at the table above you will see that the little box on the cross of the A-row and the A-column is indeed in reversed video. To change a value; just type in the new one and it will replace the old one. The corresponding box will switch to normal video and the next one will be activated. If you don't want to change all values but only some there are three ways to activate the box of your choice:

Press <RETURN> and keep it pressed until you reach the right box.
Use the arrows on your keyboard.
Simply use the mouse to click in the desired box to activate it.

when you are finished; click in ''Exit'' and then ''Quit''; all changes will be saved and DOTPLOT will be able to run with new sets of defaults and or a new table.
As you might have noticed the table is a 26 by 26 matrix. This means that not only the standard amino acids are represented, but also B(Asx) and Z(Glx). There are four letters that do not stand for any amino acid (J,O,U,X), allthough the X is sometimes used for "any amino acid"; however, this allows you to use the extra letters for such excentrics as selene-coupled amino acids and their likes. It is advisory to give these extremely high auto-score values to highlight their rarety.

Principle of DOTPLOT.

Like all Dotplot programs, this one works with two parameters called the Window (W) and the Score (S). A block of homology is defined as that part of the sequences, with length W, where at least S residues are the same. In case of the DNA comparison of DOTPLOT this means that only if at least 14 out of 21 bases are identical a line is drawn in the picture. For a complete picture all possible stretches, with length W, of one sequence have to be compared with all possible stretches, with length W, of the other sequence. For DNA this means that bases 1 to 21 of the horizontal sequence have to be compared with bases 1 to 21, 2 to 22, 3 to 23 etc. of the vertical sequence. After this first round bases 2 to 22 of the horizontal sequence will again be compared to the vertical sequence: 1 to 21, 2 to 22, 3 to 23 etc. This is the general principle, but DOTPLOT uses an algorithm that has to calculate a lot less than this description suggests at first glance. However, it does give you an idea about the number of calculations needed for a run. It also shows the quadratic nature of DOTPLOT: an increase in length by a factor two will increase the time necessary for a run with a factor of four.
In the case of proteins the definition of the Score has to be revised. In this case it can be defined as the sum of the various, individual, scores. If you run without any score tables this boils down to the same as for DNA, but with other standard values for the Window and Score. With the use of score tables (see also page 6) the new definition comes in really handy; it will explain the Dayhoff defaults where Score is ten out of a Window of only 8. The very high values can be explained by the use of numbers larger than one for some combinations. If you have very good eyes, and can read the numbers in the last figure on page 10, you will see that in this table W(=Trp) scores 2.73 with itself. The table on page 10 is identical to the Dayhoff table and the number expresses the enormous importance of Trp at certain sites in a protein; the chance that an analogous protein will have kept it during evolution is very large indeed.

The files on the disk.

Under the DNA folder there are two files called MAQUI.DNA and MSINI.DNA respectively. Both these DNA's code for a procaryotic DNA-methyltransferase, genes homologous enough to show up on DOTPLOT pictures under default conditions (3,4). The protein translations of these DNA files are also on the disk under the PROTEIN folder. Alltough this means a very straightforward translation in case of M.SinI it is not so for M.AquI; the latter protein is composed out of two polypeptides. For the use of DOTPLOT I have "glued" them together so you can compare them to M.SinI in one simple run. For more information see the files themselves; they are in UWGCG format.

Formats.

The files for DOTPLOT can be off the following formats: Staden, UWGCG, Genbank, EMBL and flat sequence files. All these formats can be used together; two files don't have to be in the same format to be run together.

References.

1. Devereux, J., P. Haeberli and O. Smithies. 1984. A comprehensive set of sequence analysis programs for the VAX. Nucl. Acids Res. 12: 387-395.
2. Schwarz, R. M. and M. O. Dayhoff. 1978. in Atlas of protein sequence and structure 5 sup 3 (M.O.Dayhoff editor) , The national biochemical research foundation, Washington.
3. Karreman, C. and A. de Waard. 1988. Cloning and complete nucleotide sequences of the type II restriction-modification genes of Salmonella infantis. J. Bacteriol. 170:2527-2532.
4. Karreman, C. and A. de Waard. 1990. Agmenellum quadruplicatum M.AquI, a novel modification methylase. J. Bacteriol. 172:266-272.

Christiaan Karreman,
Dept. of Pediatrics gebouw 12,
Academical Hospital Leiden,
P. O. box 9600,
2300 RC Leiden,
The Netherlands.
Phone : 031-70276118

EMail : KARREMAN@RULLF2.LeidenUniv.nl