

Statistical Analysis

The statistical analysis part of MacPattern includes the maximal segment score method, introduced by Karlin and colleagues (Karlin & Altschul, 1990, Karlin et al., 1991, Karlin & Brendel, 1992), and the method of Eguchi and Seto (1992).

MSS Analysis

A variety of different scoring schemes is offered:

Mixed charge segments (Karlin & Altschul, 1990):

2 (D,E,H,K,R), -1 (others)

Basic charge segments (Karlin et al., 1991):

2 (K,R), 0 (H), -2 (D,E), -1 (others)

Acidic charge segments (Karlin & Altschul, 1990):

2 (D,E), -2 (K,R), -1 (others)

Cysteine clusters (Karlin & Altschul, 1990):

5 (C), -1 (others)

Strong hydrophobic segments (Karlin & Altschul, 1990):

1 (A,C,F,I,L,M,V), -1 (G,P,S,T,W,Y), -2 (D,E,H,K,N,Q,R)

Kyte-Doolittle hydropathy index (Karlin & Brendel, 1992) :

3 (I,V,L), 2 (F), 1 (A,C,M), -1 (G), -2 (S,T,W,Y), -3 (P), -4 (D,E,H,N,Q), -5 (K,R)

Transmembrane domains, target frequencies (Karlin & Brendel, 1992):

6 (I), 5 (L), 4 (V), 3 (A,F), 2 (M), 1 (G,W), -1 (C,Y), -2 (S,T), -6 (P), -8 (H), -9 (Q), -10 (N), -16 (D,K,R), -17 (E)

DNA-binding domains, target frequencies (Karlin & Brendel, 1992):

4 (C,R), 3 (K,W), 1 (F,T,I), 0 (E,N,Q,T,V), -1 (A,H,L), -2 (G,M,S), -3 (D), -5 (P)

Other schemes can be added freely, although at present one has to use a resource editor such as ResEdit. Score tables are stored in resources of type 'SCOR'. To add a new table, the brave-hearted can simply create a new resource of this type. A ResEdit template is included in MacPattern and it should be fairly obvious how to fill it in. A future version may include some easier way of adding new scoring systems.

The Maximal Segment Score menu command displays a scrollable list of available scoring

schemes, and one or more can be selected from this list. This dialog also allows one to change the number of high-scoring segments that are displayed in the result window.

At the heart of this analysis is the calculation of the two parameters K and λ (see Karlin & Altschul, 1990). MacPattern uses a slightly modified version of the `karlin()` function taken from the BLAST program suite (Altschul et al., 1990). The time needed for calculation varies considerably depending on the input sequences and the scoring scheme. Under certain circumstances, execution time may become unacceptably long on machines without a mathematical co-processor such as the Macintosh LC. Therefore, MacPattern checks for the presence of a co-processor and issues a warning if it cannot find one and the Maximal Segment Score command is chosen. There is currently no way to interrupt a maximal segment score analysis once it has started!

[Eguchi & Seto Method](#)

When this method is selected, the user can choose between a graphical or textual presentation of results. One can also set the sliding-window size (see Algorithms). Only odd values are allowed for the window size.

In textual representation, the results are displayed in a list. Each row contains four fields: the position of the middle residue of a segment, the segment dissimilarity score, the z-value (i.e., the distance of the score from the mean in units of standard deviation), and the segment sequence. Fields are delimited by tabs so that the list can easily be imported into any spreadsheet program. At the end the mean score and the standard deviation are shown.

If the graphical representation is chosen, the scores are displayed as a graph. The three red horizontal lines indicate the mean value (middle line) and the $\pm 2z$ regions. The picture can be sized using the window's grow box, and can be printed, saved and copied. However, when printed the window will always be sized to fit the page.

If a nucleotide sequence is translated on the fly, the textual representation will give results for all reading frames, whereas in the graphical view only the first reading frame is shown.

Note: The statistical analysis part of MacPattern is still experimental. Although it has been tested to a great extent, it was not given as much attention as the PROSITE or BLOCKS search functions, and there may well be some hidden bugs lurking around.