

## Appendix B: PROSITE Pattern Syntax

The following is extracted from the PROSITE User Manual Rel. 5:

quote start

- The standard IUPAC one-letter codes for the amino acids are used.
- The symbol 'x' is used for a position where any amino acid is accepted.
- Ambiguities are indicated by listing the acceptable amino acids for a given position, between square parentheses '[ ]'. For example: [ALT] stands for Ala or Leu or Thr.
- Ambiguities are also indicated by listing between a pair of curly brackets ' {} ' the amino acids that are not accepted at a given position. For example: AM stands for any amino acid except Ala and Met.
- Each element in a pattern is separated from its neighbor by a '-'.  
- Repetition of an element of the pattern can be indicated by following that element with a numerical value or a numerical range between parenthesis. Examples: x(3) corresponds to x-x-x and [FY](1,2) corresponds to [FY] or [FY]-[FY].
- When a pattern is restricted to either the N- or C-terminal of a sequence, that pattern either starts with a '<' symbol or ends with a '>' symbol.
- A period ends the pattern.

Examples:

PA [AC]-x-V-x(4)-ED.

This pattern can be translated as: [Ala or Cys]-any-Val-any-any-any-any but Glu or Asp

PA <A-x-[ST](2)-x(0,1)-V.

This pattern, which must be in the N-terminal of the sequence ('<'), can be translated as: Ala-any-[Ser or Thr]-[Ser or Thr]-(any or none)-Val

quote end

The index generating software in the MacPattern package checks the input pattern database for any syntax errors.

Note: if you enter a pattern by keyboard, the following rules apply in addition to the ones above:

- You may omit the trailing period.
- You may omit the dashes.
- Characters can be upper or lower case.
- The maximum pattern size is 50 positions. A position is defined by surrounding dashes, whether they are actually used or not. Therefore, a pattern such as M(1,30)-x-K(38,39) [or M(1,30)XK(38,39)] is perfectly fine, since it consists of only three positions, although the matching sequence may be up to 70 residues long.