BioBIKE Pattern Matching (using MATCHES-OF-PATTERN)

Character sets and some special characters:
- \[ \text{abc} \] Set of characters
- \[ \sim \text{abc} \] Set of excluded characters
- \[ \text{a-z} \] Set of characters from first character to last
- \[ \sim \text{a-z} \] Set of excluded characters from first character to last
- \* \hspace{1em} Any character
- \# \hspace{1em} Any digit (equivalent to \([0-9]\))
- \sim \# \hspace{1em} Any non-digit (equivalent to \([-0-9]\))
- \$ \hspace{1em} Any word character (letters and digits) (equivalent to \([0-9a-z]\))
- \sim \$ \hspace{1em} Any non-word character (equivalent to \([-0-9a-z]\))
- ^ or \sim @ \hspace{1em} Any space character (space, tab, and newline)
- @ or \sim ^ \hspace{1em} Any non-space character
- `' \hspace{1em} Either ' or "

Examples:
- \( \text{MATCHES-OF-PATTERN} \ "\ast\ast\ast\ast\ast\ast\ast\ast\ast\ast\ast\" \ \text{IN} \ \text{candidate-gene} \)
  \( \text{Looks for iron-sulfer cofactor binding site in sequence of candidate gene} \)
- \( \text{MATCHES-OF-PATTERN} \ "\#\sim\#\sim\#\sim\#\sim\#\sim\#\sim\" \ \text{IN} \ "\text{LOCUS ANGLNA 2225 bp DNA linear BCT 12-SEP-1993}" \)
  \( \text{Looks for the date within a locus line of a GenBank file} \)
- \( \text{MATCHES-OF-PATTERN} \ "\sim\text{ACGT}\" \ \text{IN} \ (\text{SEQUENCE-OF Cw?0002}) \)
  \( \text{Looks for nonstandard nucleotides within a gene sequence} \)

Repetition symbols
- \{n\} \hspace{1em} Previous element must be present exactly \(n\) number of times
- \{n,\} \hspace{1em} Previous element must be present at least \(n\) number of times
- \{m, n\} \hspace{1em} Previous element may be present anywhere from \(m\) to \(n\) number of times
- ? \hspace{1em} Previous element may be present or absent (equivalent to \{0,1\})
- .. \hspace{1em} Previous element may be present 1 or any number of times
  \hspace{1em} (choose minimum number of times)
- ... \hspace{1em} Previous element may be present 1 or any number of times
  \hspace{1em} (choose maximum number of times) (equivalent to \{1,\})
- ?.. \hspace{1em} Previous element may be absent or present any number of times
  \hspace{1em} (choose minimum length that satisfies the rest of the pattern)
- ?... \hspace{1em} Previous element may be absent or present any number of times
  \hspace{1em} (choose maximum length that satisfies the rest of the pattern)
  \hspace{1em} (equivalent to \{0,\})
Example:

(MATCHES-OF-PATTERN " \[acgt]\..." IN
  " 1021 accacgaagt tgctactggt ggtcagtgcg agctaggctt ccgcttttgg")
  
  Looks for blocks of nucleotides (any length), preceded by a space

(MATCHES-OF-PATTERN "am_{0,5}a $..." IN "I am not a crook (or am I?)")

  Looks “am” within 5 characters of “ a “, followed by a sequence of letters of any
  length

Other special symbols

~ Negation
`
Escape (the character that follows is to be interpreted literally)
`# Pound sign (because # itself is special)
`$ Dollar sign (because $ itself is special)
`* Asterisk (because * itself is special)
`^ Carat (because * itself is special)
(` Group (to be considered a single element in pattern matching)
( ) Remember these elements
| Or