

Rules for Substitution-Cost Function:

1. Indels of any repeating sequences (i.e. "A", "BB", "AAA", "CCCCC") are scored such that only the initial indel is scored according to the substitution cost matrix, and the following indels are scored as 0
2. For a pair of length-3 sequences, if the first and third letters in the sequence match, but the second doesn't, score the mismatch as 0, and the matches as normal, e.g.
 - "ABC" and "ACC" will be scored as "AC" and "AC"
3. Gap-initiation has 3x the cost of regular indels
 - A-- and ABC will score the match (A,A), triple the mismatch (-,B), and the mismatch (-,C)
 - A-- and ABB will score the match (A,A), triple the mismatch (-,B), and ignoring the additional B on the edge as in rule 1
4. Certain matched subsequences will have their scores, as generated by the substitution matrix, doubled:
 - AABBCC
 - ABCCBA
 - CABBAC
5. Certain matched subsequences will have their scores, as generated by the substitution matrix, tripled:
 - ABCABCABC
 - CBACBACBA
 - CABABCAB
6. If a perfectly matching subsequence with no indels has length greater than 30, it's score is quadrupled
7. Matches of length 1 should be ignored, e.g.
 - (-A-,CAB) is scored as (--,CB)

This means a match such as (-A-,BAB) is scored as (--,BB), which is then scored as (-,B) as in rule 1