Sequence Alignment with Motif Database

Susumu Hiraoka
hiraoka@crl.hitachi.co.jp
Central Research Laboratory, Hitachi, Ltd.
1-280 Higashikoigakubo, Kokubunji-shi, Tokyo 185, Japan

1 Introduction

Both sequence alignment and motif scan are effective sequence analysis methods, which give us clews to a function of a newly determined sequence. Sequence alignment tools like FASTA [1] and BLAST [2] search for similar sequences from a sequence database and display the similarity. Motif scan tools search for sequence patterns which are already known to have some kind of functions from a motif database such as PROSITE [3]. Each method has its own limitations. Sequence alignment tools can't find distantly related sequences. Furthermore, they do not display motifs, even if aligned sequences have common motifs. On the other hand, motif scan tools can't find the sequence patterns which are not registered in a motif database.

2 Method

We combined sequence alignment with motif scan and made a new sequence analysis method. The algorithm which we present is an extension of Smith-Waterman sequence comparison [4]. The algorithm temporarily translates amino acid sequences into sequences of motifs. It compares the motif sequences and align them instead of the amino acid sequences. The alignment and the translation which give the maximum score are chosen from all possible alignments and translations.

3 Result

We compared two sequences using conventional Smith-Waterman method and this method. We simplified the result of this method by omitting the comparison of same motifs. The result of this method clearly shows that both of the sequences have leucine zipper pattern which is a periodic repetition of leucine residues at every seventh position. On the other hand, conventional sequence alignment tools do not show motifs nor find leucine zipper pattern.

References

[1] Pearson, W.R., "Searching protein sequence libraries: comparison of the sensitivity and selectivity of the Smith-Waterman and FASTA algorithms," *Genomics*, 11:635–650, 1991.

[2] Altschul, S.F.,et al., "Basic local alignment search tool," *J. Mol. Biol.*, 215:403–410, 1990.

[3] Bairoch, A., et al., "The PROSITE database, its status in 1995," *Nucl. Acids Res.*, 24:189–196, 1995.

[4] Smith, T.F. and Waterman, M.S., "Identification of common molecular subsequences," *J. Mol. Biol.*, 147:195–197, 1981.

```
           130        140        150        160        170        180
TVHUF1 GRRGKVEQLSPEEEEKRRIRRERNKMAAAKCRNRRRELTDTLQAETDQLEDEKSALQTEI
       ... : .  .....:..  ::..: ..    :.::.::. .. :  ::. . ... ..
TVHUM  SNNRKCTSPRSSDTEENVKRRTHNVLE----RQRRNELKRSFFALRDQIPELENNEKAPK
        340        350        360            370        380        390

           190        200        210        220        230        240
TVHUF1 ANLLKEKEKLEFILAAHRPACKIPDDLGFPEEMSVASLDLTGGLPEVATPESEEAFTLPL
       . .::. .  ..::...
TVHUM  VVILKKAT--AYILSVQAEEQKLISEEDLLRKRREQLKHKLEQLRNSCA
            400          410        420        430
```

Fig. 1. Alignment using conventional Smith-Waterman method.

```
              140        150        160        170        180        190
TVHUF1  PEEEEKRRIRRERNKMAAAKCRNRRRELTDTLQAETDQLEDEKSALQTEIANLLKEKEKL
           ..     . .: .     ..      ..::: ..Leucine_zipper_pattern
TVHUM   FALRDQIPELENNEKAPKVVILKKATAYILSVQAEEQKLISEEDLLRKRREQLKHKLEQL
             380        390        400        410        420        430
```

Fig. 2. Alignment using this method.