Extended Sequence Alignment Method for Protein Secondary Structure Prediction

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

1 Introduction

Sequence alignment methods are very effective for secondary structure prediction. However, they are only applicable when the similarity of the sequences is high enough. We previously reported that the extended sequence alignment method, which uses not only amino acid letters but also strings of amino acid letters representing motifs as comparing units, enabled us to find common motifs even among the sequences whose similarities are low [1]. In this paper, we show that the method also enables us to find common secondary structure between the sequences, and applicable for secondary structure prediction.

2 Method

The extended sequence alignment method is the extension of the Smith-Waterman dynamic programming sequence alignment method [2]. The dynamic programming sequence alignment method puts two sequences on a lattice in much the same way as in dot matrix methods. For each point in the lattice the alignment score and the transition which is the last step of alignment path to that point are calculated. Fig. 1 shows the transitions used in the conventional method and in this method. The conventional method considers only three transitions corresponding to gaps in both sequences and amino acid letter comparison. Our method considers additional transitions corresponding to motif comparison which means amino acid letters representing the same motifs. Motif comparisons are done for all motifs in the motif database. In the present study, we made the motif database. It consists of alpha-helix, beta-sheet, turn, and coil motifs that are calculated using Chou-Fasman structure index.

3 Results and Discussions

We compared rtp sequence with histone sequence using this method. Two sequences have the common secondary structure which consists of alpha-helixes and beta-sheets. The alignment by this method (Fig. 3) shows alpha-helix and beta-sheet that are part of common structure between the sequences, though conventional method (Fig. 2) showed no similarity.

References

[1] Hiraoka, S., Sequence alignment with motif database, *Genome Informatics 1997*, Universal Academy Press, 280-281, 1997.

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

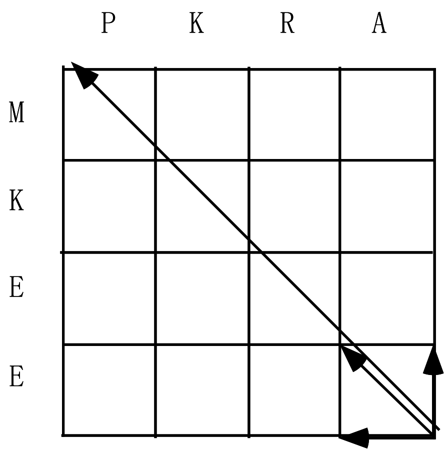

Figure 1: Transitions used in the conventional method (short arrow) and an additional transition used in this method (long arrow).

```
           20        30        40        50        60        70
sp:RTP QRAFLKLYMITMTEQERLYGLKLLEVLRSEFKEIGFKPNHTEVYRSLHELLDDGILKQIK
                                 ..:..: .: .... ..  ... X:..:.:::.X
sp:H5_ YSEMIAAAIRADKSRGGSSRQSIQKYVKSHYK-VG-QHADLQIKLAIRRLLTTGVLKQTK
       30        40        50        60          70        80
```

Figure 2: Alignment by the conventional method.

```
                       10        20                    30        40
sp:RTP_BAC MKEE----KRSSTGFLVKQRAFLKLYMIT------------MTEQERLYGLKLLEVLRSE
           ====    &&&&========:====   .            . ----========. ===
sp:H5_ANSA PKRARAPRKPASHPTYSEMIAAAIRADKSRGGSSRQSIQKYVKSHYKVGQHADLQIKLAI
                20        30        40        50        60        70
```

Figure 3: Alignment by this method. Helices and sheets are represented by '=' and '-' respectively.