2012 International Conference on Environment Science and Engieering
IPCBEE vol.32(2012) © (2012)IACSIT Press, Singapoore

In-Depth Assessment of Local Sequence Alignment

Atoosa Ghahremani and Mahmood A. Mahdavi[†]

Department of Chemical Engineering, Ferdowsi University of Mashhad, Azadi Square, Pardis Campus, 91779-48944, Mashhad, Iran

Abstract. The overall procedure of searching local similarities among amino acid sequences using local alignment tool has been described in detail. Different phases of the alignment and use of available scoring matrices are explained. The update of scoring matrices based on new sequence information regularly take place over time based on the overall procedure assembled here in a single document. This document assists young scientists who seek simple references to understand step-by-step local alignment.

Keywords: sequence alignment, local alignment, dynamic programming, modified local alignment, BLAST.

1. Introduction

Sequence alignment is a way for comparing two or more sequences in order to search a series of specific characters or character patterns with same order. It causes to identify positions and regions in protein sequences that provide insights into the function or structure of an uncharacterized sequence by predicting similarities to a protein that have been studied experimentally and ultimately infer homology between two real related sequences (share a common evolutionary ancestor). Computational approaches to sequence alignment generally ordered into two categories: global alignment (Needleman-Wunsch, 1970) and local alignment (Smith- Waterman, 1981). Two rapid alternative local alignment algorithms compared to the Smith-Waterman algorithm are FASTA (Pearson & Lipman, 1988) and BLAST (Basic Local Alignment Search Tool) (Altschul et al., 1990).

The biological reliable measure that discriminates homologs from unrelated sequences and indicates the degree of functional or structural similarity is score or expectation value. These values are computed from the alignment and give biologists a measure of the relatedness of two sequences. For protein sequence alignment, the 20 × 20 matrices are used to evaluate all possible combinations of amino acid pairs to calculate an overall score for the alignment of two sequences. Several types of scoring matrices have been proposed for protein sequences. Series of PAM (point accepted mutation) (Dayhoff et al., 1978) and BLOSUM (block substitution matrix) (Henikoff & Henikoff, 1992) matrices are the most popular and initial matrices for scoring all possible amino acid substitutions in pair-wise alignment during evolution.

Performance of sequence alignment tools is heavily upon the procedure scoring matrix is constructed. This procedure has been updated over time depending on the number of sequenced proteins and completed proteoms available in the literature. There is, however, an overall process of data gathering and calculations that need to be followed in any update of scoring matrices such as PAM. The overall procedure has not yet been fully assembled in a single document and different references have cited and explained partials of the procedure as required.

In this article, the basic concepts of local alignment algorithm and modified Smith & Waterman algorithm, which is applied for searching similarity in databases, are described and step by step calculation

[†] Corresponding author. Tel.: + 985118805008.
E-mail address: mahdavi@ferdowsi.um.ac.ir