

Abstract

Multiple sequence alignment is the most fundamental problem in bioinformatics research field because it is the first step to analyze organism phylogenetic tree, secondary and tertiary structure prediction of protein and RNA, etc. various of methods and approaches have been published for over the last 30 years. But there is no method that can solve the problem of multiple sequence alignment efficiently and optimally.

Dynamic programming method has been proven to handle pairwise sequence alignment effective and efficiently at both global and local alignment. However, when developed to handle multiple sequences alignment, it often to fail because the requirement of very large resource.

Therefore, in this final, the divide-and-conquer method adapted to make the memory used by dynamic programming to solve multiple sequence alignment problems more efficient.

By applying divide-and-conquer method, the time and space complexity to perform multiple sequence alignment can be reduced from $O(n_m^k)$, where n_m is the maximum length of initial sequence and k is the number of sequences that want to be aligned, to $O(\hat{n}_m^k)$, where \hat{n}_m is the limit of allowed sequence length. However, due to the use of divide-and-conquer, the resulting alignment becomes unoptimal (approximate). To improve the resulting alignment back to optimum, iterative refinement adapted. The space complexity than become $O(n_m^k)$, where L is the limit used.

Keywords: Bioinformatics, multiple sequence alignment, global sequence alignment, memory efficiency, optimization, exact method, dynamic programming, divide-and-conquer, iterative refinement.