BMC Systems Biology

BioMed Central

Poster presentation

Open Access

A parallel algorithm for de novo peptide sequencing

Elisa Mori*[1], Sara Brunetti[1], Sonia Campa[2] and Elena Lodi[1]

Address: [1]Dipartimento di Scienze Matematiche e Informatiche, Università degli studi di Siena, Siena, Italy and [2]Dipartimento di Scienze Informatiche, Università degli studi di Pisa, Pisa, Italy

Email: Elisa Mori* - morie@unisi.it

* Corresponding author

from BioSysBio 2007: Systems Biology, Bioinformatics and Synthetic Biology
Manchester, UK. 11–13 January 2007

Published: 8 May 2007

BMC Systems Biology 2007, **1**(Suppl 1):P61 doi:10.1186/1752-0509-1-S1-P61

This abstract is available from: http://www.biomedcentral.com/1752-0509/1?issue=S1

Introduction

Protein identification is a main problem in proteomics, the large-scale analysis of proteins. Tandem mass spectrometry (MS/MS) provides an important tool to handle protein identification problem. Indeed the spectrometer is capable of ionizing a mixture of peptides, essentially several copies of the same unknown peptide, dissociating every molecule into two fragments called complementary ions, and measuring the mass/charge ratios of the peptides and of their fragments. These measures are visualized as mass peaks in a mass spectrum.

There are two fundamental approaches to interpret the spectra. The first approach is to search in a database to find the peptides that match the MS/MS spectra. This database search approach is effective for known proteins, but does not permit to detect novel proteins. This second task can be dealt with the de novo sequencing that computes the amino acid sequence of the peptides directly from their MS/MS spectra.

In the de novo sequencing problem one knows the peptide mass m_P, and a subset of the masses of its ions $m_1,...,m_n$, and the task is to determine a sequence Q of masses of residues such that subsets of its prefixes and suffixes give the masses in input. The solution is, in general, not unique.

Methods

We reformulate the problem in terms of searching paths in a graph. To this goal, let M_P denote the set of ion masses m_i in input increased with: their complementary masses $m_P - m_i + 2$, the mass of the hydrogen, 1, and of its complementary mass $m_P - 17$. By abuse of notation, $M_P = \{m_1,...,m_n\}$, where $m_i < m_j$ if $i < j$.

We build a directed acyclic graph $G_P = (V, E)$ as follows. Let a node v_i associate to a member m_i of M_P, and an edge from v_i to v_j if $m_j - m_i$ equals the sum of residue masses.

The de novo sequencing problem consists in determining any path from v_1 to v_n in the graph G_P.

Although there is a unique original protein, the de novo sequencing may have in general more solutions (or none). In order to choose one sequence among the possible solutions, researchers have introduced any scoring function [1-3] depending on the masses of the fragments in the spectra. Our algorithm can determine either the solution of maximum score according to any given function or that of maximum length.

We use 3 algorithms:

◦ the first algorithm consists in building the graph;

◦ the second algorithm permits to distinguish the feasible paths that start in v_1 and terminate in v_n among the others;

◦ finally, the third algorithm retrieves the solution of maximum score.

Results and conclusion

The literature offers a wide range of sequential de novo sequencing algorithm, all taking $O(n\ logn)$ time, at least [4,5]. Aiming at lowering such time barrier, we proposed a work-efficient CREW (concurrent-read exclusive write) PRAM [6] algorithm for the de novo peptide sequencing that determines the maximum length sequence in $O(n)$ time by using $log\ n$ processors. Such theoretical result showed that our algorithm is clearly scalable and reaches the theoretical, ideal efficiency.

The next step we are working on is the implementation of the proposed algorithm on a parallel machine to verify such theoretical results and scalability features.

References

1. Bafna V, Edwards N: **On de novo interpretation of tandem mass spectra for peptide identification.** In *Proceeding of ICCMB ACM Press*; 2003.
2. Dancik V, Addona TA, Clauser KR, Vath JE: **De novo peptide sequencing via tandem mass spectrometry.** *Journal of computational biology* 1999, **6:**.
3. Ma B, Zhang K, Hendrie C, Liang C, Li M, Doherty-Kirby A, Lajoie G: **PEAKS: powerful Software for Peptide De Novo Sequencing by MS/MS.** *Rapid Communications in Mass Spect* 2003.
4. Brunetti S, Dutta D, Liberatori S, Mori E, Varrazzo D: **An efficient algorithm for de novo Peptide Sequencing.** In *Proceeding of the ICANNGA Springer Verlag*; 2005.
5. Pandurangan G, Ramesh H: **The Restriction Mapping Problem Revisited.** *Journal of Computer and System Sciences* 2002.
6. Jàjà J: *An introduction to parallel algorithms Addison-Wesley*; 1992.