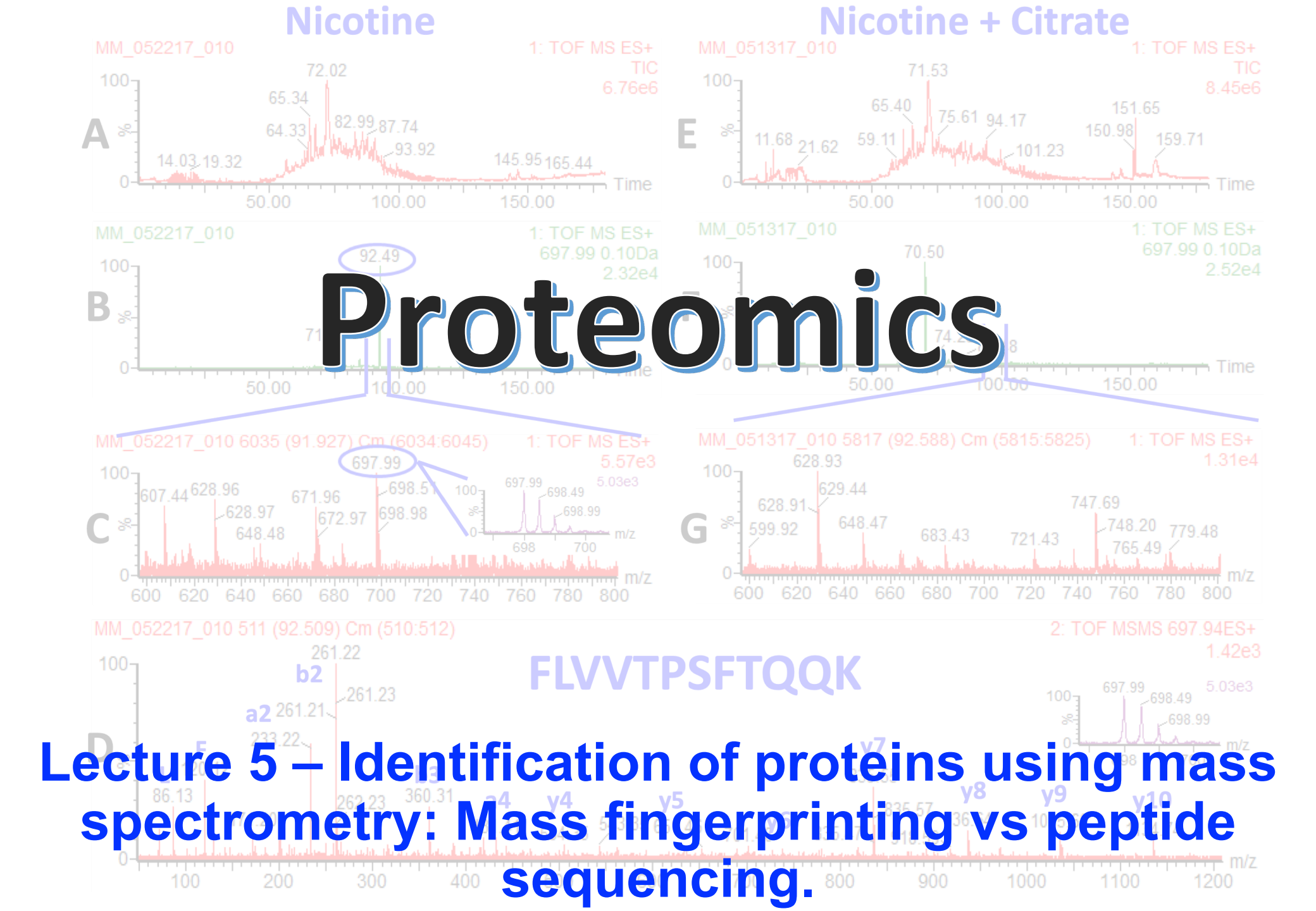
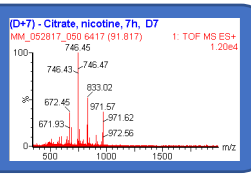
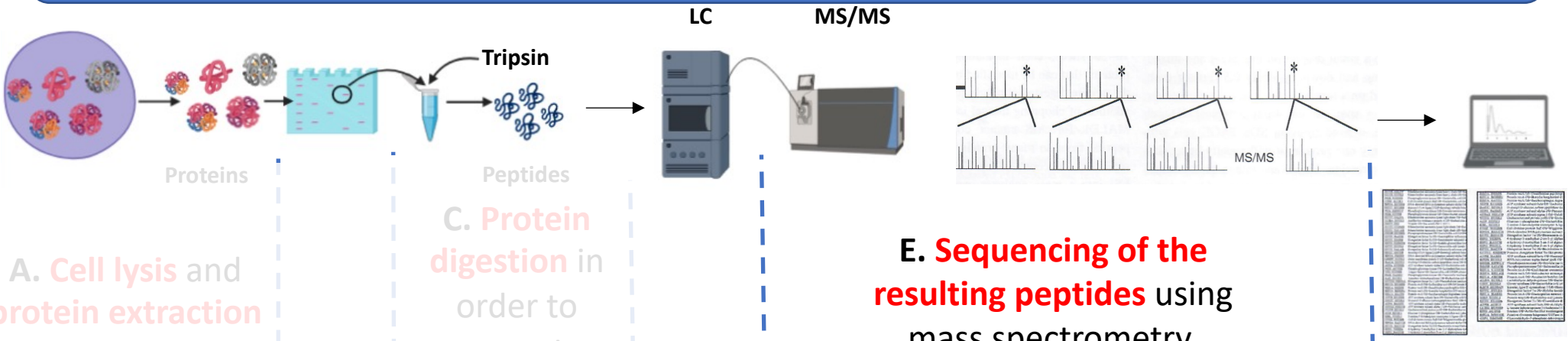


Lecture 5 – Identification of proteins using mass spectrometry: Mass fingerprinting vs peptide sequencing.

II. 2. Identification of proteins using mass spectrometry



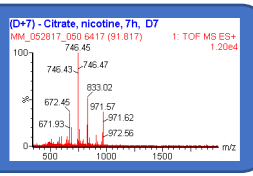


ll proteomic methods involve fractionating proteins from the sample and digesting them to generate a mixture of peptides. Depending on the type of mass spectrometry instrument used, the information obtained from peptide analysis can be used for protein identification. Two distinct approaches for protein identification have been described:

mandatory step).

A. **Mass fingerprinting** - Methods that use mass spectrometers operating in full scan mode to identify the m/z of peptides in the mixture. The obtained values are then used to identify the proteins from which these peptides originate.
B. **Peptide sequencing** - Methods that use tandem mass spectrometers, which identify the m/z of peptides in the mixture and then fragment them to identify the amino acid sequence of each peptide. The amino acid sequences are then used to identify the proteins from the initial mixture.

II. 2. Identification of proteins using mass spectrometry



A. Peptide mass fingerprinting, PMF

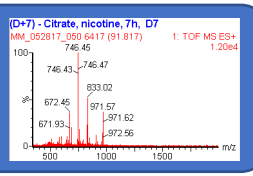
This identification method is based on the presumtion that that, given the complete genome of an organism, it could theoretically be possible to generate a list of all the proteins in that organism by performing in silico translation using computers. Knowing how proteases work (trypsin hydrolyzes the peptide bond at a K or R residue, if not followed by a P), it would also be possible, again using computers, to generate a list of all the peptides that could potentially result from the digestion of each protein constituting the proteome of that species. For the theoretical peptides generated, their origin (the protein from which they originate), length, sequence, and thus their mass could be calculated. In this way, it would be possible to create **extensive lists with the theoretical masses of peptides obtained through virtual digestions, each peptide being associated with the protein from which it originates**. Some of these **peptides**, especially those with more than 6 amino acids**, have unique masses, which means that each protein is associated with a variable number of unique mass values** - a **mass fingerprint of the protein**, created based on the masses of the peptides.

Through laboratory processing of a sample obtained from the organism of interest, the proteins present in the sample are typically hydrolyzed, and the resulting peptides are most commonly analyzed using a **MALDI-TOF** mass spectrometer (Matrix-Assisted Laser Desorption/Ionization coupled with Time-of-Flight mass analyzer). The instrument identifies the **mass-to-charge ratio (m/z) values for each peptide in the sample**. If the **measured mass is unique and matches one of the theoretical masses in the extensive list of theoretical peptide masses obtained through virtual digestions, then it is highly likely that the real peptide matches the theoretical one, indicating the presence of the parent protein in the sample**. The higher the number of real peptides identified with corresponding masses to the virtual peptides belonging to the same protein, the higher the level of confidence that the protein is present in the sample.

Peptide Mass list

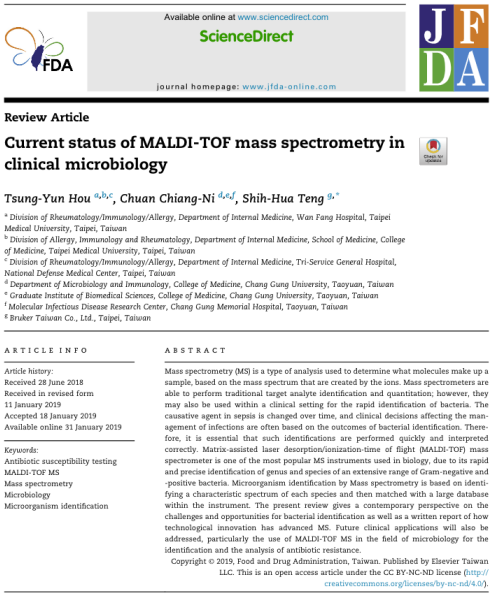
.
.
.
1528.7685
1528.7989
1529.0002
1529.2001
1529.2454
1529.5006
1529.6997
1529.7348 ⇐ ***1529.7348 measured mass***
1529.9978
1530.2332
1530.4567
1531.0003
1531.0107
1531.3004
1531.5656
.
.
.

II. 2. Identification of proteins using mass spectrometry



By matching the masses of peptides measured using a mass spectrometer with the theoretical masses of peptides that could potentially exist in the respective organism, the proteins present in an unknown sample can be identified. The success of peptide mass fingerprinting in protein identification depends on two key factors:

- The **accuracy** of the measurements: It is crucial for the mass spectrometer to provide precise measurements of peptide masses. Any error in determining the real mass of the peptide can lead to incorrect protein identification**.** This is why MALDI-TOFs are prefered for this approach;
- The **uniqueness** of the identified peptides: Peptides with unique masses that correspond to specific proteins increase the confidence in protein identification. If multiple peptides with matching theoretical masses are identified from the measured spectra, it strengthens the certainty of the protein's presence in the sample.
- The **completeness** of the theoretical peptide database: The quality of the database used for matching the measured peptide masses with theoretical masses greatly influences the success of protein identification. A comprehensive and accurate database covering a wide range of organisms improves the chances of finding matches for the measured peptides.



JOURNAL OF FOOD AND DRUG ANALYSIS 27 (2019) 404–414

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.jfda-online.com

Review Article

Current status of MALDI-TOF mass spectrometry in clinical microbiology

Tsung-Yun Hou [a,b,c], Chuan Chiang-Ni [d,e,f], Shih-Hua Teng [g,*]

[a] Division of Rheumatology/Immunology/Allergy, Department of Internal Medicine, Wan Fang Hospital, Taipei Medical University, Taipei, Taiwan
[b] Division of Allergy, Immunology and Rheumatology, Department of Internal Medicine, School of Medicine, College of Medicine, Taipei Medical University, Taipei, Taiwan
[c] Division of Rheumatology/Immunology/Allergy, Department of Internal Medicine, Tri-Service General Hospital, National Defense Medical Center, Taipei, Taiwan
[d] Department of Microbiology and Immunology, College of Medicine, Chang Gung University, Taoyuan, Taiwan
[e] Graduate Institute of Biomedical Sciences, College of Medicine, Chang Gung University, Taoyuan, Taiwan
[f] Molecular Infectious Disease Research Center, Chang Gung Memorial Hospital, Taoyuan, Taiwan
[g] Bruker Taiwan Co., Ltd., Taipei, Taiwan

ARTICLE INFO

Article history:
Received 28 June 2018
Received in revised form
11 January 2019
Accepted 18 January 2019
Available online 31 January 2019

Keywords:
Antibiotic susceptibility testing
MALDI-TOF MS
Mass spectrometry
Microbiology
Microorganism identification

ABSTRACT

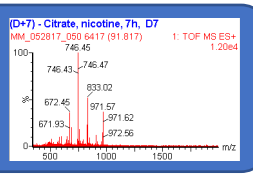
Mass spectrometry (MS) is a type of analysis used to determine what molecules make up a sample, based on the mass spectrum that are created by the ions. Mass spectrometers are able to perform traditional target analyte identification and quantitation; however, they may also be used within a clinical setting for the rapid identification of bacteria. The causative agent in sepsis is changed over time, and clinical decisions affecting the management of infections are often based on the outcomes of bacterial identification. Therefore, it is essential that such identifications are performed quickly and interpreted correctly. Matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) mass spectrometer is one of the most popular MS instruments used in biology, due to its rapid and precise identification of genus and species of an extensive range of Gram-negative and -positive bacteria. Microorganism identification by Mass spectrometry is based on identifying a characteristic spectrum of each species and then matched with a large database within the instrument. The present review gives a contemporary perspective on the challenges and opportunities for bacterial identification as well as a written report of how technological innovation has advanced MS. Future clinical applications will also be addressed, particularly the use of MALDI-TOF MS in the field of microbiology for the identification and the analysis of antibiotic resistance.

Copyright © 2019, Food and Drug Administration, Taiwan. Published by Elsevier Taiwan LLC. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

The main limitations in the application of mass fingerprinting methods relate to:

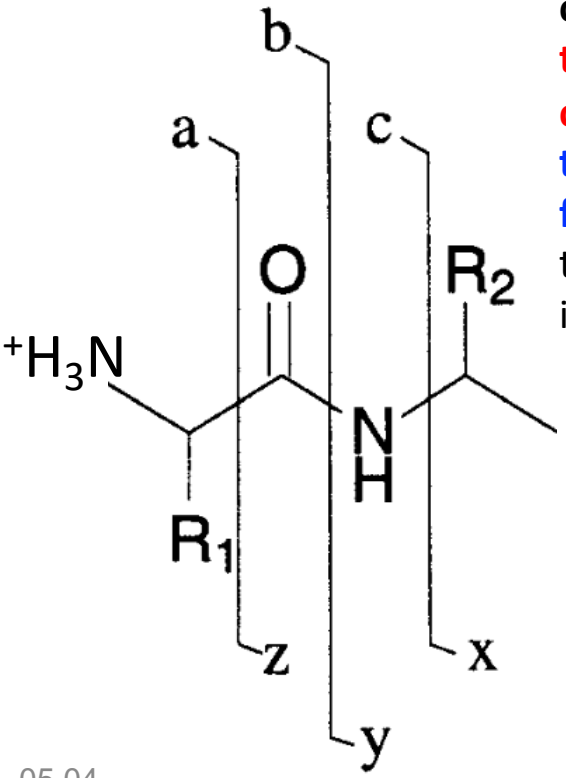
1. While sequence databases are extensive, they rarely contain complete and annotated genomes (gene positions are described, start and stop codons are identified) of a species;
2. In higher organisms, there are numerous homologous proteins, which are very similar in sequence, and their mass fingerprints are also very similar or even identical.
3. Sequence databases rarely contain information about post-translational modifications (PTMs) of proteins. In the case of a protein undergoing PTM, the theoretical mass fingerprint is significantly different from the real one and cannot be identified using this method.

II. 2. Identification of proteins using mass spectrometry



B. **Peptide sequencing** by tandem mass spectrometry (MS/MS) seeks to overcome the limitations of PMF **by fragmenting peptides in a collision cell, analyzing the resulting fragments (fragment ions/daughter ions) to identify the amino acid sequence**. The calculated amino acid sequence is then used to identify the proteins present in the sample.

Peptide sequencing by MS/MS is based on the observation that the interaction between the inert gas and peptides in the collision cell leads to breaking of covalent bonds and generation of peptides fragments. The key point is that **the fragmentation of peptides** (ie. breakingg of covalent bonds) **is rather predicatable and thus**, **a limited set of ions are formed**:



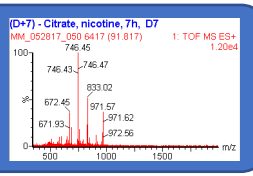
1. **b ions and y ions** - are formed by **breaking the peptide bond between the carbonyl carbon and the amidic nitrogen**. **The b ion carries the amino group from the N-terminal end of the precursor parent peptide and a positive charge on the carbonyl carbon due to the peptide bond cleavage**. **The y ion contains the C-terminal end of the precursor parent peptide and a positive charge on the NH_2 group resulting from the peptide bond cleavage.** If the peptide was obtained by trypsin hydrolysis, the C-terminal end contains the amino acids Arg or Lys - di-aminic amino acids, so y ions may carry an additional positive charge.

COO^-

2. **A ions and z ions** – formed by breaking the bond between the C-carbonyl and Ca atoms. **The a ion carries the amino group from the N-terminal end of the precursor peptide, while the ion z contains the C-terminal end of the precursor peptide**.

3. **c ions and x ions** – formed by breaking the bond between the amidic N and Ca atoms. **The c ion carries the amino group from the N-terminal end of the precursor peptide, while the x ion contains the C-terminal end of the precursor peptide**.

II. 2. Identification of proteins using mass spectrometry



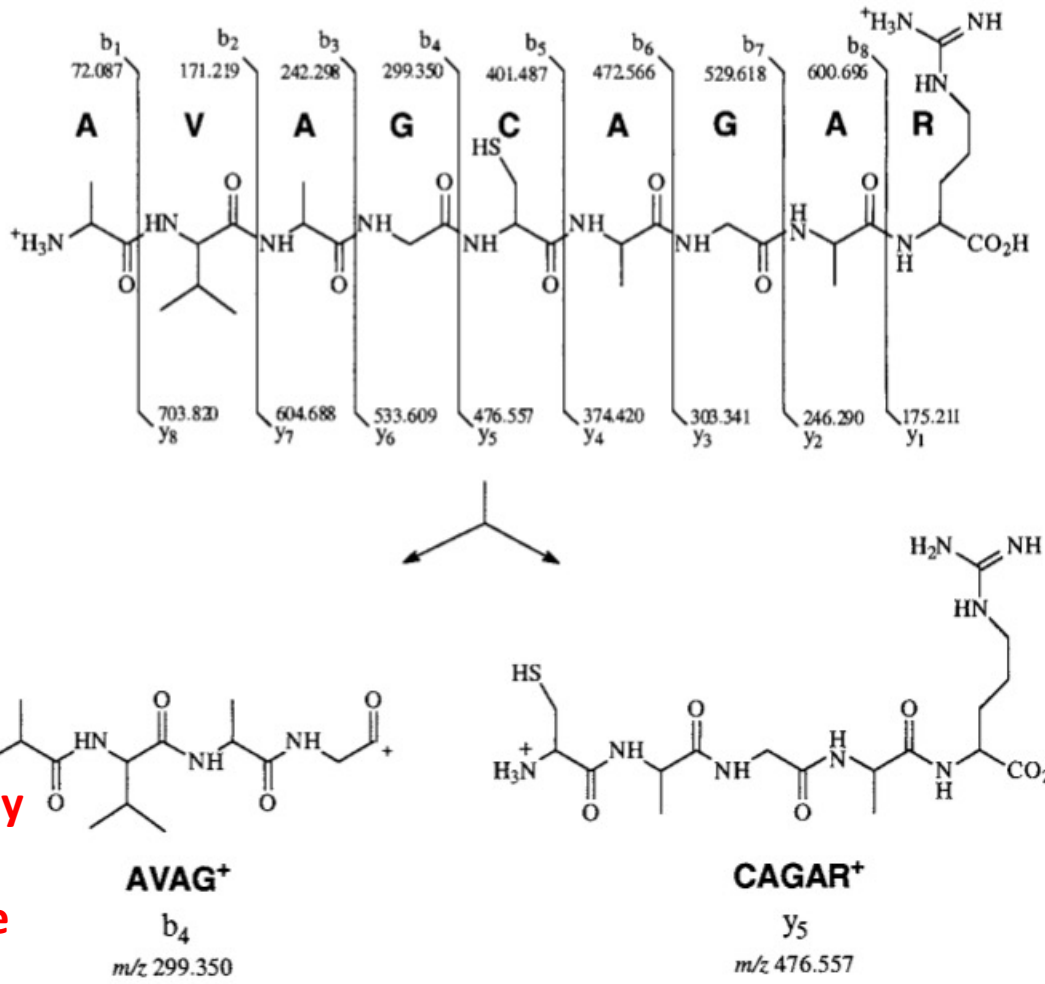
In the collision cells of MS/MS spectrometers used in proteomics - triple-quads, ion traps, or Q-TOF instruments - **peptide bonds are most commonly broken, leading primarily to the formation of y and b ions**. However, there is no preference regarding which peptide bonds or amino acids will be cleaved during this process.

For example, in the case of a peptide with the sequence AVAGCAGAR, any of the peptide bonds can be cleaved, resulting in the formation of a **series of y and b ions** as represented in the figure.

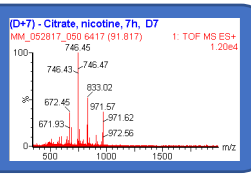
The numbers of the ions within a series correspond to the position of the peptide bond that is cleaved, and consequently, to the number of amino acids contained in that fragment.

The mass of each y or b ion within a series is the sum of the masses of the amino acid residues it contains. This includes the molecular weight of the amino acids (minus 18 Da for the loss of water for the peptide bond formation) plus 1 Da for b ions (from the NH2 terminal hydrogen) or plus 19 Da for y ions (from the OH at the COOH terminal plus 2H+ at the N amide).

The difference in mass between two consecutive y or b ions represents the mass of the amino acid residue by which they differ. This difference is the parameter used for protein sequencing.



II. 2. Identification of proteins using mass spectrometry



The MS/MS spectrum of the peptide AVAGCAGAR with highlighting of the signals for the y and b ion series.

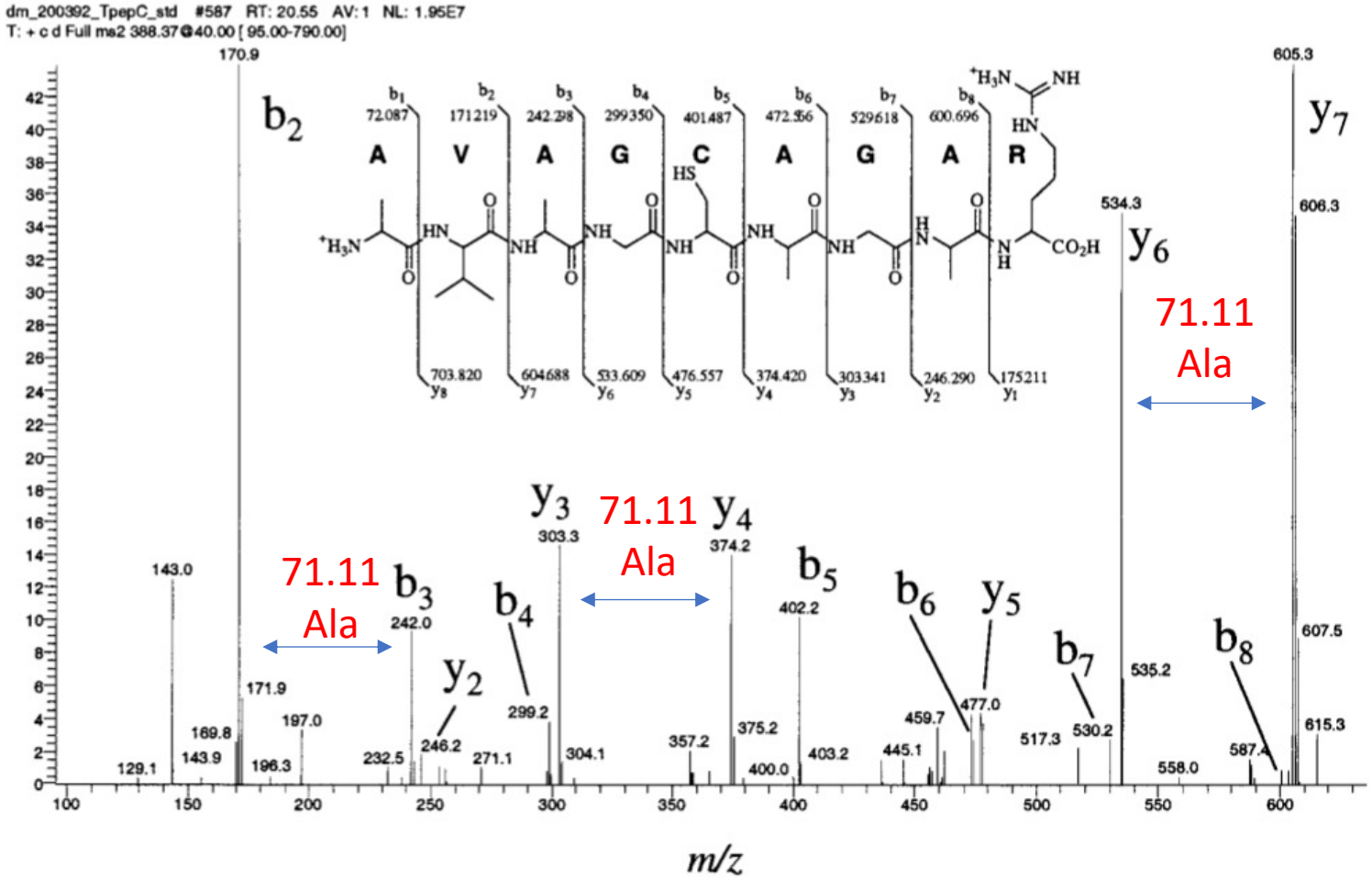
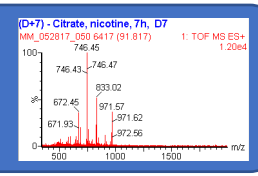


Fig. 4. Annotated MS-MS spectrum of the $[M+2H]^{2+}$ ion of AVAGCAGAR showing b- and y-ions.

II. 2. Identificarea proteinelor prin spectrometrie de masă

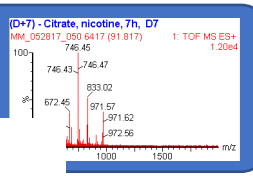


Peptide sequencing by tandem mass spectrometry (MS/MS) **involves recording the mass spectrum resulting from the controlled fragmentation of each peptide through collision-induced dissociation (CID),** followed by the **identification of the y and b ion series based on the dominant signals. The differences in m/z between ions of the same ion series indicate the contained amino acids, while the order of the ions reflects the sequence in which these ions are bound in the peptide structure.** Generally, identifying one ion series (y or b) is sufficient to establish the sequence of a peptide, but it is preferable to establish both series for verification (the y series complements the b series).

Manual interpretation of an MS/MS spectrum of a peptide is called *de novo* sequencing and, depending on the spectrum's complexity and the user's expertise, it can take from 15 minutes to several hours. Therefore, **computer based interpretation of MS/MS spectra is preferred**, **which involves comparing real MS/MS spectra with theoretical MS/MS spectra**. This process involves the following steps:

1. Generating a database with **theoretical peptides** using the known amino acid or nucleotide sequences from the target organism;
2. Fragmenting these **theoretical peptides** and establishing **all possible virtual mass spectra** to be generated and recorded; creating a new database with possible m/z values to be recorded for all theoretical ion series for these **theoretical** peptides;
3. Running the sample and recording the real mass spectrum of a real peptide in the sample and identifying the predominant m/z values;
4. Performing **spectral matching** – identifying which recorded mass spectra are found in the theoretical spectra database. The search is performed by identifying which theoretical ion series best corresponds to the real ion series.

II. 2. Identificarea proteinelor prin spectrometrie de masă



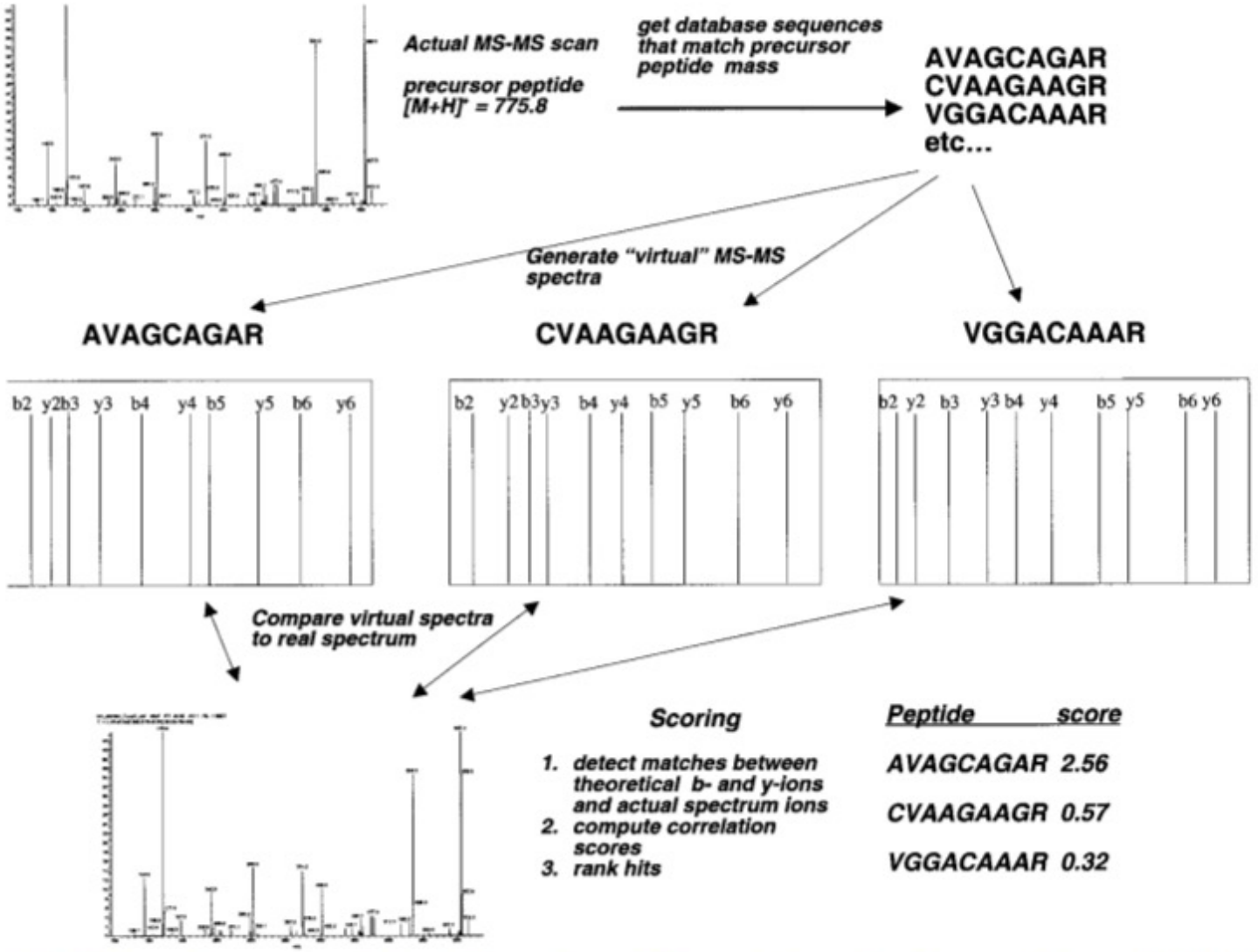


Fig. 1. Schematic representation of operation of Sequest algorithm for correlation of MS-MS spectra with peptide sequences from databases.