Methods xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Methods

journal homepage: www.elsevier.com/locate/ymeth

Towards integrative structural mass spectrometry: Benefits from hybrid approaches

Julien Marcoux *, Sarah Cianférani *

Laboratoire de Spectrométrie de Masse Bio-Organique (LSMBO), IPHC, Université de Strasbourg, 25 rue Becquerel, 67087 Strasbourg, France
IPHC, CNRS, UMR 7178, 67087 Strasbourg, France

ARTICLE INFO

Article history:
Received 18 December 2014
Received in revised form 6 May 2015
Accepted 25 May 2015
Available online xxxx

Keywords:
Structural proteomics
Mass spectrometry
Structural biology

ABSTRACT

Structural mass spectrometry encompasses an increasing range of methods aimed at collecting as much structural information as possible on a biomolecule or its related complexes. Originally limited to the analysis of the primary structures of proteins, mass spectrometry has evolved over the past 20 years to provide information on the secondary, tertiary and even quaternary structure of proteins. Furthermore, the systems investigated with these methods have become more and more complex, as many developments have progressively overcome the main challenges of the size, heterogeneity, and/or solubility of protein complexes. A decade ago, most of these techniques were still the playground of a handful of specialists. However, the potential of these methods and their complementarity to other classical biophysical methods have driven an increasing number of users to develop new techniques and, perhaps more crucially, manufacturers have developed improved instruments and solutions/kits that are now commercially available. Today, more and more groups are combining structural proteomics techniques in order to gain additional information, as we will see in this review. This article will particularly focus on the analysis of peptides and protein complexes. First, the main methods of structural proteomics will be described. Then different possible combinations will be described, including how complementary they are, what synergistic information can be obtained from them, and what their current limitations are.

1. Introduction: main structural MS approaches, what can I get?

Structural MS is a quite recent research area that consists of deriving structural information from MS based data and usually focuses on purified proteins or protein complexes, usually of recombinant origin.

1.1. Primary structure information: the emergence of top-down proteomics

Classical bottom-up LC–MS/MS analysis provides reliable information about the amino acid sequence, including the purity, heterogeneity of the protein(s), and/or post-translational modifications (PTMs) [1,2]. However, analyzing a mixture of protein isoforms (proteoforms) with different combinations of PTMs by bottom-up proteomics is usually not successful, as most of the information is lost upon trypsin digestion (Fig. 1). In contrast, in a top-down analysis, each proteoform is analyzed and fragmented individually, providing additional information on the primary structure of the isoforms. Although early top-down analyses used collision induced dissociation (CID), alternative dissociation techniques based on electron transfer dissociation (ETD) or electron capture dissociation (ECD), are now providing complementary fragmentation profiles, significantly increasing protein sequence coverage [3,4]. The recent study of 150 kDa monoclonal antibodies has significantly increased the size limit of proteins studied by top-down, typically between 10 and 50 kDa [5,6], although this technique was used only for primary sequence confirmation, not for higher-order structural characterization. Top-down proteomics, initially limited to very challenging instruments with superconducting magnets (FT-ICR), has now become more accessible thanks to Orbitrap analyzers. Furthermore, ETD and ECD, unlike CID, do not lead to the loss of certain labile PTMs such as phosphorylation [7], explaining their more and more common use in proteomics. Although sometimes characterized as "opposing" techniques, bottom-up and top-down proteomics are actually complementary or even overlapping, as evidenced by the increasing number of middle-down studies, whereby large polypeptides are generated in solution prior to their gas-phase fragmentation [8–10]. Ten

* Address: IPBS, CNRS, UMR 5089, 205 Route de Narbonne, 31077 Toulouse, France (J. Marcoux). Laboratoire de Spectrométrie de Masse Bio-Organique (LSMBO), IPHC, Université de Strasbourg, 25 rue Becquerel, 67087 Strasbourg, France (S. Cianférani).

http://dx.doi.org/10.1016/j.ymeth.2015.05.024
1046-2023/

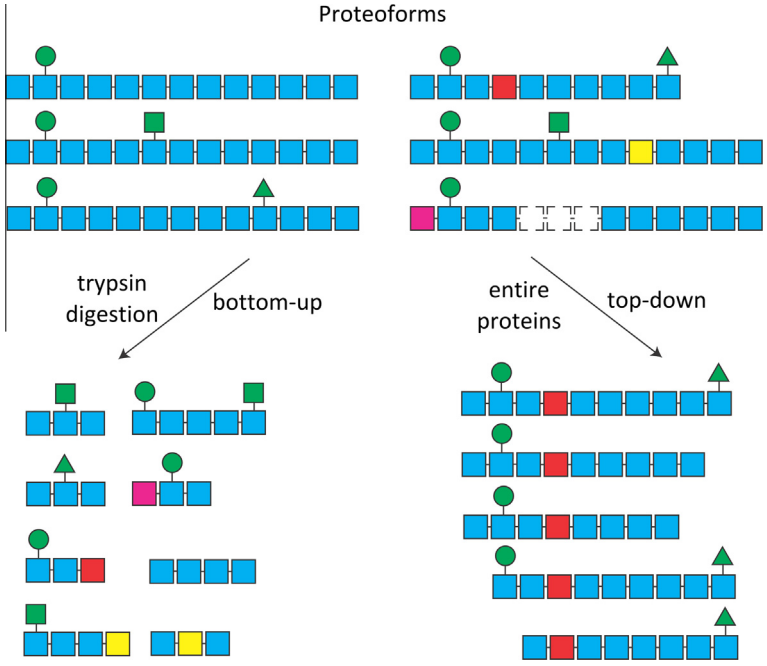

Fig. 1. The combination of post-translational modifications (green symbols), mutations (colored squares) and alternative splicing (dotted squares), can sometimes produce numerous protein isoforms or "proteoforms". After digestion (bottom-up, left), it is impossible to know the origin of the tryptic peptides. In the top-down approach (right), each proteoform is selected and fragmented individually.

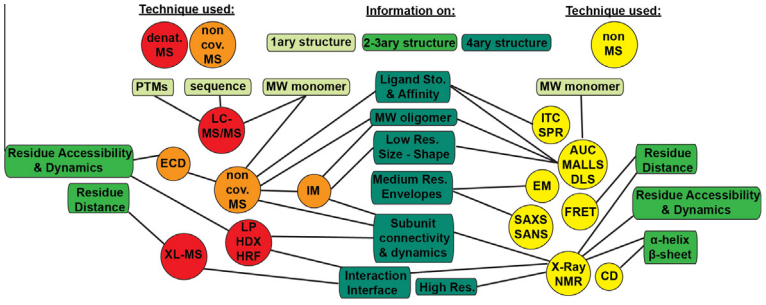

Fig. 2. The main methods coupled to mass spectrometry under denaturing (red) and non-covalent (orange) conditions are complementary to the classical biophysical techniques used for protein characterization (yellow). They provide information ranging from the primary to the quaternary structure of proteins. XL: cross-linking, HDX, hydrogen deuterium exchange, HRF, hydroxyl radical footprinting, IM: ion mobility, ECD: electron capture dissociation, LP: limited proteolysis, ITC: Isothermal Titration Calorimetry, SPR: Surface Plasmon Resonance, AUC: Analytical Ultra Centrifugation, MALLS: Multi Angle Laser Light Scattering, EM: Electron Microscopy, SAXS: Small Angle X-ray Scattering, SANS: Small Angle Neutron Scattering, FRET: Förster Resonance Energy Transfer, CD: Circular Dichroism.

years after the release of the Human Genome Project, the first drafts of the human proteome have just been published [11,12]. Given the increasing interest in epigenetics and PTMs, the identification of proteoforms by top-down approaches will undoubtedly constitute the next characterization level of entire organisms.

1.2. Higher order structure

The two gold standards of structural biology, namely Nuclear Magnetic Resonance (NMR) and X-ray diffraction (XRD), are unfortunately sometimes unable to solve the structure of a particular system for many different reasons, including size, flexibility, heterogeneity of the sample, or the amount of protein needed. All these parameters are, in theory, not a major limitation to most of the MS methods used for secondary and tertiary structure determination. Even though MS-based approaches are rather low-resolution techniques, they represent a potential alternative to get structural information that is complementary to classical biophysical methods (Fig. 2 and Table 1).

1.2.1. Hydrogen/deuterium exchange and covalent labeling

The idea of assessing protein topology by monitoring the exchange kinetics of amide hydrogen atoms dates back from the 50s [13]. Rates of exchange were initially determined by following the decrease of tritium scintillation or by NMR. Pepsin digestion was introduced in the late 70s to increase the resolution of the method [14] and MS was used for the first time to monitor the uptake of deuterium on proteolytic peptides in 1993 [15]. Although NMR has higher resolution, MS had the clear advantage in terms of sensitivity, sample and time consumption, and the size limit of the proteins studied. At the same period, covalent labeling coupled to LC–MS started to emerge as a promising technique to probe protein surface topology [16–20]. A variety of chemical modifications (diethylpyrocarbonate, acetylation, carboxylation, oxidation) (reviewed in [21]) can be used to probe the surface