Methods 59 (2013) 270–277

Contents lists available at SciVerse ScienceDirect

Methods

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

Exposing the subunit diversity within protein complexes: A mass spectrometry approach

Shelly Rozen [a], Alessandra Tieri [a], Gabriela Ridner [a], Ann-Kathrin Stark [a], Tilo Schmaler [b], Gili Ben-Nissan [a], Wolfgang Dubiel [b], Michal Sharon [a,*]

[a] Department of Biological Chemistry, Weizmann Institute of Science, Rehovot 76100, Israel
[b] Department of Surgery, Charité – Universitätsmedizin, Berlin 10117, Germany

ARTICLE INFO

Article history:
Available online 4 January 2013
Communicated by Peter Schuck

Keywords:
Mass spectrometry/methods
Chromatography/methods
Protein complexes
Post-translational modifications
Protein isoforms
COP9 signalosome

ABSTRACT

Identifying the list of subunits that make up protein complexes constitutes an important step towards understanding their biological functions. However, such knowledge alone does not reveal the full complexity of protein assemblies, as each subunit can take on multiple forms. Proteins can be post-translationally modified or cleaved, multiple products of alternative splicing can exist, and a single subunit may be encoded by more than one gene. Thus, for a complete description of a protein complex, it is necessary to expose the diversity of its subunits. Adding this layer of information is an important step towards understanding the mechanisms that regulate the activity of protein assemblies. Here, we describe a mass spectrometry-based approach that exposes the array of protein variants that comprise protein complexes. Our method relies on denaturing the protein complex, and separating its constituent subunits on a monolithic column prepared in-house. Following the subunit elution from the column, the flow is split into two fractions, using a Triversa NanoMate robot. One fraction is directed straight into an on-line ESI-QToF mass spectrometer for intact protein mass measurements, while the rest of the flow is fractionated into a 96-well plate for subsequent proteomic analysis. The heterogeneity of subunit composition is then exposed by correlating the subunit sequence identity with the accurate mass. Below, we describe in detail the methodological setting of this approach, its application on the endogenous human COP9 signalosome complex, and the significance of the method for structural mass spectrometry analysis of intact protein complexes.

1. Introduction

Production of ATP, DNA replication, transcription, protein synthesis and degradation, are only a small list of the biological processes that are carried out by protein complexes. In fact, the entire cell can be viewed as a factory in which its many diverse functions are carried out by the orchestrated activities of such complexes, each of which requires the coordinated action of multiple subunits that assemble into a functional unit of distinct composition and structure.

In recent years, experimental advances on several fronts have expanded our ability to study the structural and functional properties of such large protein complexes [1]. Techniques to isolate and purify multi-subunit complexes have been developed, and technological innovations have enabled the complete sequencing of several eukaryotic genomes [2]. This information, coupled with worldwide proteomics initiatives, have yielded detailed lists of the subunits that comprise protein complexes [3]. Determining the accurate composition of protein assemblies is an important step towards understanding their function. However, such knowledge might not be sufficient for unraveling the full complexity of protein complexes' mode of action, and the regulatory mechanisms that underlie their activities.

Protein complexes are likely not uniform in their structure and function. This is due to the fact that a single subunit may be encoded by more than one gene [4], a single gene might produce several alternative splice forms [5–7], protein subunits may be post-translationally modified (PTM) [8,9] or cleaved [10], and single nucleotide polymorphisms (SNPs) may be present [11]. These multiple forms of a single subunit may be integrated within a protein complex. As a consequence, the presence of a specific protein complex within a cell may actually be represented by a diverse group of functionally distinct entities. Moreover, protein complexes are most likely dynamic assemblies whose composition is altered according to the tissue type, the nature of the cell itself (e.g., normal vs. diseased), and the intracellular localization. Such variability in subunit composition is expected to create unique functional properties, enabling the complex to respond and adapt

* Corresponding author. Fax: +972 8 934 6010.
E-mail address: michal.sharon@weizmann.ac.il (M. Sharon).

1046-2023/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ymeth.2012.12.013

to varying cellular conditions [12]. These abilities may be highly developed, and subject to sophisticated modes of regulation [13,14].

Clarifying the heterogeneity of subunit composition is essential for uncovering the multiple functional forms of protein complexes, and their regulatory pathways. However, relatively little attention has been given to this topic. In the past, only a few studies have sought to determine the variability of protein subunits (for examples, see [15–18]). This is mainly due to the fact that large-scale proteomic analysis, which involves the digestion of proteins into peptides, is limited in its ability to describe subunit diversity. This approach, while providing a high number of identifications, often yields low sequence coverage, hindering the detection of endogenous protein cleavages, PTMs and isoforms, all of which are likely to have high sequence homology. Notably, even when the full array of PTMs is probed and their exact localization within the protein sequence is defined, information regarding their composition on a single subunit, and the number of co-existing combinations, is lost. Thus, of necessity, specialized methods had to be developed, in order to identify the full repertoire of subunit compositions.

Two-dimensional gel electrophoresis followed by liquid chromatography mass spectrometry (LC–MS) analysis constitutes a known strategy for characterizing the heterogeneity of protein species [17,19]. However, the method relies on the extraction of protein variants from polyacrylamide gels, which often yields low recovery. The 2D gel electrophoresis approach was recently advanced by combining it with intact protein separation and mass measurements [15]. "Top down" approaches based on mass spectrometric analyses of intact proteins [20] have also been applied for mapping protein variants [16–18]. However, these "top down" approaches present a challenge, due to the difficulty in predicting fragmentation pathways, compared to tandem MS of peptides [21]. Nevertheless, an elegant study recently used this approach for large-scale intact protein analysis [16], and ~3000 protein species were identified. However, since cell extracts were used, information regarding the ability of each subunit variant to be integrated within a functional complex was lost.

Here, we describe our approach for determining the diversity of protein subunit composition. The underlying principle of the method entails correlation between the accurate mass of the protein subunit, and its sequence identity. Our experimental set-up couples capillary-LC separation, a NanoMate robotic system, and a QToF mass spectrometer (Fig. 1). Initially, the protein complex is loaded onto a monolithic column prepared in-house, where it is decomposed into its "building blocks," by a gradual increase in organic solvent concentration. The individual subunits are then separated from one another, based on their size and chemical properties. Subsequently, the eluted flow is split into two, using the NanoMate robot. One fraction is sprayed directly into the mass spectrometer to accurately determine the mass of the individual subunits, whereas the second fraction is collected into a 96-well plate for sequence identification by tryptic digestion, LC–MS/MS proteomic analysis, and a database search. Our approach is automated and, as a result, easily reproducible, with low sample consumption. Furthermore, this strategy improves upon a method we previously applied, which required two LC separations of the protein subunits [22]: one for exact mass determination, and one for sequence identification. With the introduction of the NanoMate robot, the two applications are integrated in a single LC run. Not only is our new method faster, but it also reduces sample consumption, a critical factor when only minute amounts of the protein complex are available.

2. Rationale

In the course of investigating the structural properties of protein complexes by MS, we realized that there is little agreement between the actual measured mass of a protein subunit, and that predicted by protein sequence databases. Consequently, it was often impossible to determine which protein subunits are present within a particular complex, especially when investigating less established multi-protein assemblies of unknown composition and stoichiometry. Thus, a method enabling correlation of accurate mass with the sequence identity of protein subunits is required. Such an approach would enable us to identify all proteins present in a multi-subunit complex, together with their corresponding intact masses. Only then could we successfully utilize structural MS to determine the non-covalent interactions between subunits, and the overall architecture of the complex [23–25].

3. Materials and methods

3.1. Chemicals and reagents

For column synthesis, fused-silica capillary tubings, polyimide-coated with 200 μm I.D. and 375 μm O.D, were purchased from Polymicro Technologies. The silanization agent: 3-(trimethoxysilyl)propyl methacrylate 98% (3-TMPM); inhibitor: 2,2′-diphenyl-1-picrylhydrazyl (DPPH); monomer: hexyl methacrylate 99% (HEMA); crosslinker: ethylene glycol dimethacrylate 98% (EDMA); porogenic solvents: 1-propanol 99.8%, 1,4-butanediol 99%; and radical initiator: azobisisobutyronitrile, 98.0% (AIBN); were all purchased from Sigma.

For LC separation, methanol and acetonitrile (ULC/MS grade) and trifluoroacetic acid (TFA) (HPLC grade) were purchased from Bio-Lab. Formic acid (FA) >98% was purchased from Fisher Scientific. Ammonium acetate buffer (7.5 M), dithiothreitol (DTT), iodoacetamide (IAA), trypsin and ammonium bicarbonate were purchased from Sigma. Ultrapure water (Milli Q water) was obtained from a Direct-Q 3 Ultrapure Water System (Millipore). Commercially available proteins were used as standards (see Table 1) following purification by size exclusion chromatography, using a Superdex 200 column (GE Healthcare), with 200 mM ammonium acetate.

3.2. Monolithic column synthesis

The fused-silica capillary was pretreated by etching with 1 M NaOH at 100 °C for 3 h, followed by 1 M HCl at 70 °C for 3 h. The capillary was then washed with Milli Q water and methanol until a neutral pH was achieved. Silanization was carried out by treating the inner wall of the capillary with a solution of 3-(trimethoxysilyl)propyl methacrylate (3-TMPM) in neat toluene [30/70 (v/v)] containing 0.005% of the inhibitor 2,2′-diphenyl-1-picrylhydrazyl (DPPH). The silanized capillary was washed with toluene and methanol, and dried by a nitrogen stream.

For preparation of the HEMA monolithic column, a polymerization mixture was prepared by mixing 1-propanol (48.5% (v/v)), 1,4-butanediol [41.6% (v/v)], Milli Q water [9.9% (v/v)] with HEMA [40% (w/w), EDMA (60% (w/w)] and AIBN (1% w/w to the monomers)giving a ratio of monomers to porogenic solvents of 60/40. The mixture was degassed by a stream of nitrogen before it was loaded into an 18-cm pretreated capillary using a syringe pump (Harvard Apparatus) at a flow rate of 10 μl/min. The capillary was sealed at both ends with two silicon rubber caps, and incubated at 50 °C for 24 h. The monolithic capillary column thus obtained was then connected to a nanoAcquity UPLC system (nUPLC) (Waters Corp.), and