Mass Spectrometry and Fluorescence Analysis of Snap-Nappa Arrays Expressed Using E. coli Cell_Free Expression System

Claudio Nicolini1,2,*, Rosanna Spera1,2, Fernanda Festa1, Luca Belmonte1,2, Shaorong Chong1, Eugenia Pechkova1,2 and Joshua Labaer3

1Laboratories of Biophysics and Nanobiotechnology, Department Experimental Medicine, University of Genova, Italy
2Nanoworld Institute Fondazione EL.BA. Nicolini, Largo Piero Redaelli 7, Pradalunga (Bg), Italy
3Virginia G. Piper Center for Personalized Diagnostics, Biodesign Institute, Arizona State University, Tempe, Arizona, USA
4New England Biolabs, Inc, 240 County Road, Ipswich, UK

Abstract

We present the analysis of an innovative kind of self-assembling protein microarray, the “Nucleic Acid Programmable Protein Array” (NAPPA), express with the SNAP tag in E.coli coupled self free expression system. The goal is to develop a standardize procedure to analyze the protein protein interaction occurred on NAPPA array combining label free Mass Spectrometry (MS) and fluorescence technology for protein microarray. We employ in the process “Protein synthesis Using Recombinant Elements” (PURE) system. For the first time an improved version of NAPPA, that allows for functional proteins to be synthesized in situ - with a SNAP tag - directly from printed cDNAs just in time for assay, has been expressed with a novel cell-free transcription/translation system reconstituted from the purified components necessary for Escherichia coli translation - the PURE system – and analyzed both in fluorescence and in a label free manner by four different mass spectrometers, namely three Matrix Assisted Laser Desorption Ionization Time-of-Flight (MALDI-TOF), a Voyager, a Bruker Autoflex and a Bruker Ultraflex, and Liquid Chromatography-Electrospray Ionization Mass Spectrometry (LC-ESI-MS/MS). Due to the high complexity of the system, an ad hoc bioinformatic tool has been needed to develop for their successful analysis. The contemporary fluorescence analysis of NAPPA, expressed by means of PURE system, has been performed to confirm the improved characterization of this new NAPPA-SNAP system.

Keywords: Mass spectrometry; Label-free analysis; Nucleic acid programmable protein array; SNAP tag; E. coli cell-free expression system; PURE express system

Introduction

One of the main challenges in the field of proteomics is to study large number of proteins, identifying their interactions and function. In order to study the complexity of proteome, various proteomic techniques, including protein microarrays, have emerged during the past few years. A protein microarray provides a multiplex platform for high throughput (HT) studies [1,2].

Protein microarrays detection is essentially based on two main strategies, label-based and label-free methods. The label-based technique requires labelling of query molecules with labels such as fluorescent dyes, radioisotopes, epitope tags, etc. [3]; the labeled molecule is then detected via a fluorescence microscope, flow cytometer or some other fluorescence reading instrument [4]. On the contrary label-free analysis do not require the use of reporter elements (fluorescent, luminescent, radiometric, or colorimetric) to facilitate measurements, it can provide direct information on analyte binding to array molecules typically in the form of mass addition or depletion from the array surface [5-7].

Label-based technique presents limitations particularly in the context of measuring naturally occurring ligands, such as in clinical studies where it is not possible to produce fusion proteins: namely the need to obtain a capture antibody for each analyzed protein, and the concern that a labeling molecule may alter the properties of the query protein. A label-free method for the analysis of a microarray would then represent a major advancement [5].

Many label-free techniques such as SPR, quartz crystal microbalance (QCM), carbon nanotubes (CNTs) and nanowires, nanohole arrays, atomic force microscopy (AFM), etc., have been successfully integrated with protein microarrays and are emerging rapidly as a potential complement to labeling methods [7,8]. Among all these techniques, that usually cannot reveal the identity of interaction proteins, MS is powerful to provide chemical and structural information that is difficult to obtain through other means.

Combining MS and other surface techniques could offer new dimensions in protein analysis [5,8]. The integration of microarrays with MS has generated a powerful new tool to deal with the problems in protein analysis and identification area. The most successful example is the ProteinChip® System of Ciphergen Biosystems Inc. The design of the ProteinChip® array was originally derived from chromatography and is divided into two groups according to its surface characteristics: the array reader is a SELDI-TOF-MS instrument equipped with a pulsed UV nitrogen laser source [8].

Also Nedelkov et al. [9] coupled BIACORE with MS to demonstrate its feasibility in detecting multiple protein-protein interactions.In our research we employed two different mass spectrometry (MS) techniques, the Matrix Assisted Laser Desorption Ionization Time-of-Flight (MALDI-TOF) MS and Liquid Chromatography-Electrospray Ionization MS (LC-ESI-MS) (Figure 1).

In a previous research we carried out a feasibility study of MALDI-
with benzylguanine (BG) derivatives, leading to irreversible covalent alkylguanine-DNA alkyltransferase that reacts specifically and rapidly (NAPPA) [5]. The NAPPA method allows for functional proteins to be TOF MS analysis of Nucleic Acid Programmable Protein Array (NAPPA) [5]. The NAPPA method allows for functional proteins to be synthesized in situ directly from printed cDNAs just in time for assay [10,11]. The use of purified proteins was substituted with the use of cDNAs encoding the target proteins for the microarray.

The design of NAPPA has been directed to overcome the limitations of traditional protein microarray technologies, including minimal manipulation of the proteins and protein repertoire – since cDNA is used as a template for the protein expression the availability of comprehensive cDNA libraries makes it possible to use virtually any cDNA sequence on the array. Moreover the protein stability is preserved since proteins are produced just in time for the assay; once the array is activated for analysis, all reactions occur in solution and in real time so stability is not an issue [2,7,11]. On the base of the results obtained [5] many improvements have been done both to NAPPA technology and expression process and to bioinformatic analysis of the data.

Here we present the results obtained analyzing an improved version of NAPPA [10]; in this improved version the proteins were synthesized with the addition of a SNAP tag – therefore hereafter we name SNAP_NAPPA this kind of array [12,13] and translated using a reconstituted Escherichia coli coupled cell-free expression system. The addition of a SNAP tag to each protein enabled its capture to the array through the connection between the liquid chromatography and mass spectrometry (LC-MS) to analyze SNAP-NAPPA. LC-ESI MS, in fact, allows to analyze protein samples co-crystallized with the matrix on a conductive surface; for this reason NAPPA were produced on a standard microscope glass covered with a thin layer of gold.

After the NAPPA expression, the proteins immobilized on the array surface were trypsin digested and immediately after analyzed by MALDI-TOF mass spectrometer for NAPPA analysis. MALDI technique, in fact, allows to analyze protein samples co-crystallized with the matrix on the array surface at the end of the digestion.

SNAP tag is a 20 kDa mutant of the DNA repair protein O6-alkylguanine-DNA alkyltransferase that reacts specifically and rapidly with benzylguanine (BG) derivatives, leading to irreversible covalent labeling of the SNAP tag [12,13,15]. SNAP tag has a number of features that make it ideal for a variety of applications in protein labeling; in particular its substrates are chemically inert towards other proteins, avoiding nonspecific labeling in cellular applications [15]. Moreover also the chemistry and the printing of the NAPPA have been improved.

The last goal of our research is to develop a standardized analysis procedure, able to analyze the protein-protein interactions occurred on NAPPA array in a label free manner. To this aim we employed a MALDI-TOF mass spectrometer for NAPPA analysis. MALDI technique, in fact, allows to analyze protein samples co-crystallized with the matrix on a conductive surface; for this reason NAPPA were produced on a standard microscope glass covered with a thin layer of gold.

After the NAPPA expression, the proteins immobilized on the array surface were trypsin digested and immediately after analyzed by MALDI-TOF MS, without needing to be removed from the array surface. On the base of the previous results [5], however, we decided, in the present research, to add a further investigation technique, coupling mass spectrometry (electrospray ionization, ESI) with liquid chromatography (LC-MS) to analyze SNAP-NAPPA. LC-ESI MS, because of the connection between the liquid chromatography and mass spectrometer, requires removing the trypsin digested solution from the array surface at the end of the digestion.

Figure 1: Fluorescence analysis of SNAP-NAPPA a) Proteins were synthesized by two different IVTT systems, 1-Step Human Coupled IVT (HCIVT) and E. coli IVTT. Slide images were obtained with Power Scanner and the signal intensity was quantified using the Array-ProAnalyzer 6.3. The median intensity across the quadruplicates was measured and the background was corrected through the subtraction of the median value of the negative control with a matching SNAP concentration. b) Proteins yield for different SNAP concentrations, for HCIVT and E. coli IVTT systems. c) The master mix box (spotted with all the reagents of the regular NAPPA spotting mix, except DNA) was the negative control and reference box.
Several improvements have been done also to reduce the sample complexity (i.e. the amount of biological material due to NAPPA chemistry and to the expression system); in particular the in vitro translation-transcription (IVTT) system we used was from E. coli and no more from rabbit reticulocyte lysate (RRL). In the previous experiments, in fact, the proteins were synthesized with a C-terminal glutathione S-transferase (GST) tag (M_w 26 kDa) and translated using a T7-coupled RRL IVTT system [5].

The protein translation reaction, one of the most important regulators of cell behavior, involves the interactions of a large number of components, and has been studied extensively due to its importance in the cell [16]. Two approaches have guided efforts to achieve cell-free translation. One approach, developed over the past decade, is based on crude cell extract, often derived from Escherichia coli, rabbit reticulocytes, or wheat germ [17]. The second approach attempts to reconstitute protein synthesis from purified components of the translation machinery. More than 100 molecules participate in prokaryotic and eukaryotic translation, many of which have been individually purified for biochemical studies of their functions and structures [18].

Shimizu and coworkers firstly developed a protein-synthesizing system reconstituted from recombinant tagged protein factors purified to homogeneity. The system was able to produce protein at a rate of about 160 µg/ml/h in a batch mode without the need for any supplementary apparatus. Moreover, omission of a release factor allowed efficient incorporation of an unnatural amino acid using suppressor transfer RNA (tRNA). The system was termed the “protein synthesis using recombinant elements” (PURE) system [18,19].

The reconstruction of an E. coli-based in vitro translation system using protein components, highly purified on an individual basis, showed that 36 enzymes and ribosomes are sufficient to carry out protein translation [18]. These minimal protein components include the ribosomal proteins; initiation, elongation, and release factors; aminoacyl-tRNA synthetases; and enzymes involved in energy regeneration. In addition, many studies have characterized the properties of such individual proteins in detail, for example, by kinetic analysis and three dimensional structural determination, and to quantify the interactions among the components constituting the system [16].

The drawback of any extract-based systems is that they often contain nonspecific nucleases and proteases that adversely affect protein synthesis. In addition, the cell extract is like a “black box” in which numerous uncharacterized activities may modify or interfere with the downstream assays [20]. Except for the ribosomes and tRNAs, which are highly purified from E. coli, the PURE system reconstitutes the E. coli translation machinery with fully recombinant proteins. These include 10 translation factors (IF1, IF2, IF3, EF-Tu, EF-Ts, EF-G, RF1, RF2, RF3, RRF), 20 aminoacyl-tRNA synthetases and several enzymes for energy regeneration (Table 1). In addition, recombinant T7 RNA polymerase is used to couple transcription to translation.

The PURE system represents an important step towards a totally defined in vitro transcription/translation system, thus avoiding the “black box” nature of the cell extract. The immediate advantage is the significantly reduced level of all contaminating activities. The PURE system has the capacity for a yield of more than 100 µg/ml is today exclusively licensed to New England Biolabs (Ipswich, MA, USA) under the trade-name “PURExpress” [20]. Moreover the E. coli IVTT lysate, respect the RRL one, is totally characterized, which could be an advantage for the subsequent analysis of the results.
| Recombinant proteins | Protein name | Gene symbol | Protein name | Gene symbol |
|----------------------|--------------|-------------|--------------|-------------|
| IF1                  | infA         | GlnRS       |             |             |
| IF2                  | infB         | TrpRS       |             |             |
| IF3                  | infC         | TyrRS       |             |             |
| Methionyl-tRNA formyltransferase | Fmt           | HisRS       |             |             |
| EF-Tu                | TufA         | ProRS       |             |             |
| EF-Ts                | Tif          | ThrRS       |             |             |
| EF-G                 | fusA         | SerRS       |             |             |
| RF1                  | prfA         | LysRS       |             |             |
| RF2                  | prfB         | AsnRS       |             |             |
| RF3                  | prfC         | AspS        |             |             |
| RRF                  | Frr          | AlaRS       |             |             |
| ArgRS                | argS         | GlyRS       |             |             |
| CysRS                | cysS         | PheRS :2-2  | pheS pheT   |             |
| IleRS                | ileS         | Creatine kinase | Ndk         |             |
| LeuRS                | leuS         | Nucleotide diphosphate. Kinase | Ndk         |             |
| MetRS                | metG         | Myokinase   |             |             |
| ValRS                | valS         | Inorganic pyrophosphatase | Ppa         |             |
| GluRS                | gltX         | T7 RNA polymerase |             |             |

| Ribosomal proteins | Protein name | Protein name | Protein name | Protein name |
|--------------------|--------------|--------------|--------------|--------------|
| 30S ribosomal subunit protein S1 | 50S ribosomal subunit protein L17 | 50S ribosomal subunit protein L5 | 30S ribosomal subunit protein S17 | 30S ribosomal subunit protein S2 |
| 30S ribosomal subunit protein S2 | 50S ribosomal subunit protein L18 | 50S ribosomal subunit protein L6 | 30S ribosomal subunit protein S18 | 30S ribosomal subunit protein S2 |
| 30S ribosomal subunit protein S3 | 50S ribosomal subunit protein L19 | 50S ribosomal subunit protein L7/L12 | 30S ribosomal subunit protein S19 | 30S ribosomal subunit protein S2 |
| 30S ribosomal subunit protein S4 | 50S ribosomal subunit protein L20 | 50S ribosomal subunit protein L9 | 30S ribosomal subunit protein S20 | 30S ribosomal subunit protein S2 |
| 30S ribosomal subunit protein S5 | 50S ribosomal subunit protein L21 | 50S ribosomal subunit protein L10 | 30S ribosomal subunit protein S21 | 30S ribosomal subunit protein S2 |
| 30S ribosomal subunit protein S6 | 50S ribosomal subunit protein L22 | 50S ribosomal subunit protein L11 | 30S ribosomal subunit protein S22 | 30S ribosomal subunit protein S2 |
| 30S ribosomal subunit protein S7 | 50S ribosomal subunit protein L23 | 50S ribosomal subunit protein L13 | 30S ribosomal subunit protein L1 | 30S ribosomal subunit protein S2 |
| 30S ribosomal subunit protein S8 | 50S ribosomal subunit protein L24 | 50S ribosomal subunit protein L14 | 30S ribosomal subunit protein L2 | 30S ribosomal subunit protein S2 |
| 30S ribosomal subunit protein S9 | 50S ribosomal subunit protein L25 | 50S ribosomal subunit protein L15 | 30S ribosomal subunit protein L3 | 30S ribosomal subunit protein S2 |
| 30S ribosomal subunit protein S10 | 50S ribosomal subunit protein L26 | 50S ribosomal subunit protein L16 | 30S ribosomal subunit protein L4 | 30S ribosomal subunit protein S2 |
| 30S ribosomal subunit protein S11 | 50S ribosomal subunit protein L27 | 50S ribosomal subunit protein L17 | 30S ribosomal subunit protein L5 | 30S ribosomal subunit protein S2 |
| 30S ribosomal subunit protein S12 | 50S ribosomal subunit protein L28 | 50S ribosomal subunit protein L18 | 30S ribosomal subunit protein S2 | 30S ribosomal subunit protein S2 |
| 30S ribosomal subunit protein S13 | 50S ribosomal subunit protein L29 | 50S ribosomal subunit protein L19 | 30S ribosomal subunit protein S2 | 30S ribosomal subunit protein S2 |
| 30S ribosomal subunit protein S14 | 50S ribosomal subunit protein L30 | 50S ribosomal subunit protein L2 | 30S ribosomal subunit protein S2 | 30S ribosomal subunit protein S2 |
| 30S ribosomal subunit protein S15 | 50S ribosomal subunit protein L31 | 50S ribosomal subunit protein S16 | 30S ribosomal subunit protein L3 | 30S ribosomal subunit protein S2 |
| 30S ribosomal subunit protein L36 | 50S ribosomal subunit protein L37 | 50S ribosomal subunit protein L38 | 30S ribosomal subunit protein S14 | 30S ribosomal subunit protein S2 |

| Ribosomal RNAs | Protein name | Protein name | Protein name |
|----------------|--------------|--------------|--------------|
| 23S rRNA        | 5S rRNA      | 16S rRNA     |              |

| bulk tRNAs | Protein name | Protein name | Protein name |
|------------|--------------|--------------|--------------|
| tRNAaA      | tRNAaA      | tRNAaA      | tRNAaA      |
| tRNAaU      | tRNAaU      | tRNAaU      | tRNAaU      |
| tRNAaV      | tRNAaV      | tRNAaV      | tRNAaV      |
| tRNAaW      | tRNAaW      | tRNAaW      | tRNAaW      |
| tRNAaX      | tRNAaX      | tRNAaX      | tRNAaX      |
| tRNAaG      | tRNAaG      | tRNAaG      | tRNAaG      |
| tRNAaP      | tRNAaP      | tRNAaP      | tRNAaP      |
| tRNAaX      | tRNAaX      | tRNAaX      | tRNAaX      |
| tRNAaY      | tRNAaY      | tRNAaY      | tRNAaY      |
| tRNAaZ      | tRNAaZ      | tRNAaZ      | tRNAaZ      |
| tRNAaS      | tRNAaS      | tRNAaS      | tRNAaS      |
| tRNAaV      | tRNAaV      | tRNAaV      | tRNAaV      |
| tRNAaU      | tRNAaU      | tRNAaU      | tRNAaU      |
| tRNAaX      | tRNAaX      | tRNAaX      | tRNAaX      |
The presence of "background" molecules, in fact, represents the main obstacle to the data interpretation and bioinformatic tools are necessary to improve them. For this reason new matching software have been implemented. SpADS [21] an R [22] implementation of preprocessing algorithms for data reduction and noise suppression was used in order to filter results from background noise i.e. master mix MS spectrum. Moreover, this latter was used coupled to and R implementation of the K Means clustering [23,24].

The MS samples were realized printing SNAP-NAPPA spots on gold coated glass slides in a special geometry, getting proteins with higher density, in order to obtain an amount of protein appropriate for MS analysis. The spots of 300 microns were printed in 12 boxes. The spots in a box were of the same gene and the sample genes immobilized - all genes with a central role in cell signalling [25-29] - were: p53_Human, Cellular tumor antigen p53; CDK2_Human, Cyclin-dependent kinase 2; and Src_Human-SH2, the SH2 domain of Proto-oncogene tyrosine-protein kinase; PTPN11_Human-SH2, the SH2 domain of Tyrosine-protein phosphatase non-receptor type 11.

The spots in a box were of the same gene, and in particular one box apiece was reserved to the sample genes, two boxes were negative controls and reference samples, and six boxes, were printed with the sample genes in an order blinded to the researcher who made MS analysis (Figure 2). The presence of background molecules in fact represents sample genes in an order blinded to the researcher who made MS controls and reference samples, and six boxes, labelled with the letters from A to F, were printed with the sample genes in an order blinded to the researcher who made MS analysis (Figure 2).

Materials and Methods

Production and expression of NAPPA

The full length cDNA for p53 (Mw 43.80 kDa), CDK2 (Mw 33.93 kDa) - both purchased from Open Biosystem, Thermo Scientific- were amplified and cloned into Ndel and Xhol sites in pcOATexp SNAP1 vector [30], a derivative of pcOATexp and pSNAPI (New England Biolabs, Ipswich, MA, USA), Plasmid DNA was purified with NucleoBond® Xtra Maxi (Macherey-Nagel Inc., Bethlehem, PA, USA) and re suspended in water.

Printing mix was prepared with 0.66 μg/ul DNA, capture reagent (BG-PEG-NH2 ranging from 80 to 800 ng/ul; New England Biolabs, Ipswich, MA, USA), followed by secondary antibodies labelled with cy3 (Jackson ImmunoResearch Laboratories, Inc.; West Grove, PA, USA). All antibodies incubations were performed in a 1:300 dilution in PBSTM at RT, with agitation for one hour.

NAPPA slides quantification and data analysis

Slide images were obtained with PowerScanner (Tecan Group Ltd., Männedorf, Switzerland) and the signal intensity was quantified using the Array-ProAnalyzer 6.3 (Tecan Group Ltd., Männedorf, Switzerland), using the default settings. The median intensity across the quadruplicates was measured and the background was corrected through the subtraction of the median value of the negative control with a matching SNAP concentration (Figure 1).

MS analysis

The MS samples were realized printing SNAP-NAPPA spots of 300 microns in 12 boxes of 7x7 or 10x10 spots per box (spaned of 350 microns, centre to centre). The spots in a box were of the same gene, and in particular one box apiece was reserved to the sample genes (p53, CDK2, Src-SH2 and PTPNII-SH2), two boxes were printed with MM as negative control and reference samples, and six boxes, labelled with the letters from A to F, were printed with the sample genes in an order blinded to the researcher who made MS analysis (Figure 2).

This configuration allowed us to identify the samples named A, B, C, D, E and F (named "blinded samples") matching their experimental mass lists with those of the known samples (p53, CDK2, Src-SH2 and PTPNII-SH2) and then to proceed with the identification by peptide mass fingerprint (through data bank search) to further confirm the results.

The analyses were performed using two different MALDI-TOF mass spectrometers, a Voyager-DE STR (Applied Biosystems, Framingham, MA, USA) and an Ultraflex III (Bruker Daltonics, Leipzig, Germany).
For MS analysis, after the incubation, the slides were washed with PBS NaCl (1X PBS with 500 mM NaCl) three times and dried with nitrogen. The proteins synthesized on the NAPPA were trypsin digested: each box (of 16 spots) was overlaid with 5 μl of 0.1% mg/ml trypsin (Trypsin Gold, Mass Spectrometry Grade, Promega, Madison, WI, USA) in 25 mM ammonium bicarbonate (pH 7.5) and incubated in a humid chamber at 37°C for 4 hours [32,33]. At the end of the digestion the trypptic digested solutions were collected and stored in Eppendorf tubes at 4°C for the LC-ESI and Voyager MALDI-TOF MS analysis or the solvent was let evaporating at RT and the slides were stored at 4°C for Ultraflex III MALDI-TOF MS analysis.

For LC-ESI MS analysis, peptide mixtures were analyzed by nanoflow reversed-phase liquid chromatography tandem mass spectrometry (RP-LC-MS/MS) using an HPLC Ultimate 3000 (DIONEX, Sunnyvale, CA U.S.A) connected on line with a linear Ion Trap (LTQ, ThermoElectron, San Jose, CA). Peptides have been desalted in a trap column (Acclaim PepMap 100 C18, LC Packings, DIONEX) and then separated in a reverse phase column, a 10 cm long fused silica capillary (Silia Tips FS 360-75-8, New Objective, Woburn, MA, USA), slurry-packed in-house with 5 μm, 200 Å pore size C18 resin (Michrom BioResources, CA).

Peptides were eluted using a linear gradient from 96% A (H2O with 5% acetonitrile and 0.1% formic acid) to 60% B (ACN with 5% H2O and 0.1% formic acid) in 40 min, at 300 nL/min flow rate. Analyses were performed in positive ion mode and the HV Potential was set up around 1.7-1.8 kV. Full MS spectra ranging from m/z 400 to 2000 Da were acquired in the LTQ mass spectrometer operating in a data-dependent mode in which each full MS scan was followed by five MS/MS scans where the five most abundant molecular ions were dynamically selected and fragmented by collision-induced dissociation (CID) using a normalized collision energy of 35%. Target ions already fragmented were dynamically excluded for 30 s.

Tandem mass spectra were matched against Swiss Prot database and through SEQUEST algorithm [34] incorporated in Bioworks software (version 3.3, Thermo Electron) using fully trypptic cleavage constraints with the possibility to have one miss cleavage permitted, static carbamidomethylation on cysteine residues and methionine oxidation as variable modification. Data were searched with 1.5 Da and static carbamidomethylation on cysteine residues and methionine oxidation as variable modification, carbamidomethylation of cysteine as fixed modifications, peptide tolerance of 50 ppm, all taxa.

For Ultraflex III MALDI-TOF MS each box was overlaid with 2.5 μl of HCCA matrix solution and let it dry. To calibrate the spectra we spotted on the array surface 1 μl of peptide calibration standard solution (Bruker Daltonics, Leipzig, Germany) in HCCA matrix solution.

The MALDI-TOF measures were performed in reflectron mode; the resulting mass accuracy for protein was <50 ppm. MALDI-TOF mass spectra were acquired with a pulsed nitrogen laser (337 nm) in positive ion mode. The algorithm used for spectrum annotation was “Sophisticated Numerical Annotation Procedure” (SNAP). This process used the following detailed metrics: Peak detection algorithm, SNAP; Signal to noise threshold, 10; Relative intensity threshold, 10%; Greatest number of peaks, 100; Quality factor threshold, 100; SNAP average composition, Averaging. Peaks in the mass range of m/z 600-3000 were used for the peptide mass fingerprint.

For MASCOT data bank search we utilized Biotools software v2.2 (Bruker Daltonics, Leipzig, Germany), that allowed automated protein identification via library search with fully integrated MASCOT software v2.2.06 (Matrix Sciences, Ltd., London, U.K.) that searches against the Swiss-Prot/ TrEMBL database. The following parameters were used for the search: Homo sapiens or Bacteria; tryptic digest with a maximum of 2 missed cleavage; eventual methionine oxidation and a mass tolerance of 50 ppm. Identification was accepted based on significant MASCOT Mowse scores (p<0.05).

In order to identify the blinded protein panel (A, B, C, D, E and F), we used SPADs an R package for MS data preprocessing coupled to and R implementation of the K Means clustering SpADs and K Means clustering application on two specimen of 23 and 56 sample respectively was performed, the former composed of only known proteins (p53, CDK2, Src-SH2 and PTPN11-SH2) spectra while the latter composed of all spectra (the same specimen plus A,B,C,D,E,F spectra).

After a first manual selection, in which out layers were deleted from this specimen, a SpADs pre-processing was applied. Pre-processing consist of different operation on the whole spectra or in a selected region, in this case a peak extraction, with a binning window selection was performed. Selected regions of interest (ROI) were selected between 1000/2000 and on 1000/1200 on mZ axis. After pre-processing clustering was performed. Binning windows were selected dependently on this latter ROI, in the former case a binning window of 1000 was used and in the latter a binning window of 500 was used in order to preserve data consistency from flattening.

The experimental mass lists of the blinded panel were matched with those of the known samples (p53, CDK2, SH2-Src and SH2-PTPN11). The same algorithm was used to subtract MM peaks to the other spectra in order to obtain a mass list containing only the peaks obtained from protein digestion.

Results and Discussion

Fluorescence analysis

To verify the proper protein expression and capture on SNAP-NAPPA a preliminary test has been leaded by fluorescence analysis. The same SNAP-NAPPA samples employed for MS analysis (p53, CDK2, SH2-Src and SH2-PTPN11) were spotted on microscope glass in a 2×2 spots per box configuration using increasing SNAP concentrations (Figure 2). As negative control on the gold slides was printed a box only with master mix (Figures 1 and 2), while the positive controls mouse IgG and rabbit IgG were added in a printing mix instead of DNA.
The gain respect RRL is, therefore, more than twenty times higher respect HCIVT, considering the higher SNAP concentration. That expression and capture on the array surface but, moreover, demonstrated antibody. Antibody or anti-p53 antibody followed by a cy3-labeled secondary displayed on the array was measured using respectively anti-SNAP expressed with HCIVT and a third with E. coli after proteins expression fluorescence acquired. Two slides were obtained with this innovative cell free IVTT system. The protein yields obtained through PURE system has then been matched to that obtained increased levels of protein on the array [14,35]. The protein yields form 2 to 8 times E. coli IVTT systems ensured a protein yield from 2 to 8 times higher than RRL. Moreover, HCIVT performs better than RRL IVTT. The yield of protein synthesized in HCIVT is more than 10 times higher than RRL. Moreover, HCIVT showed a robust lot-to-lot reproducibility. In immune assays, the signals of many antigens were detected only in HCIVT-expressed arrays, mainly due to the reduction in the background signal and the increased levels of protein on the array [14,35]. The protein yields obtained through PURE system has then been matched to that obtained with this innovative cell free IVTT system.

In Figure 1 are reported the images of three SNAP-NAPPA slides after proteins expression fluorescence acquired. Two slides were expressed with HCIVT and a third with E. coli IVTT; the level of protein displayed on the array was measured using respectively anti-SNAP antibody or anti-p53 antibody followed by a cy3-labeled secondary antibody.

The results obtained not only confirmed the proper protein expression and capture on the array surface but, moreover, demonstrated that E. coli IVTT systems ensured a protein yield form 2 to 8 times higher respect HCIVT, considering the higher SNAP concentration. The gain respect RRL is, therefore, more than twenty times.

**MALDI-TOF mass spectrometry**

We analyzed by MALDI-TOF MS four copies of SNAP-NAPPA slides with 7x7 spots per box (two by Voyager-DE STR and two by Ultraflex III), and four copies of slides with 10x10 spots per box (two by Voyager-DE STR and two by Ultraflex III). The results were extremely reproducible both with respect to 7x7 spots/box and 10x10 spots/box that with respect to the different spectrometers and no significant difference was appreciable (Figures 2-6).

We conducted two parallel identifications, the first through the matching algorithm comparing blinded and known samples experimental mass lists, and the second submitting experimental mass lists to databank search.

We submitted the experimental mass list obtained for the known samples (p53, CDK2, Src-SH2 and PTPN11-SH2) to MASCOT data bank search. The MASCOT searching engine uses the Mowse scoring algorithm [36] to determine the significance of the peptide fingerprint result. Protein score is -10*log(P), where P is the probability that the observed match is a random event. Protein scores greater than 64 are significant (p<0.05).

In Table 2 are summarized the results obtained with significant scores; in all the samples has been detected with a significant score also the BSA (belonging to MM), that has not been reported in Table 2 for simplicity.

We analyzed by LC-ESI two copies of 10x10 spots/box slides. The data obtained resulted very reproducible, too. The matching of the results in human database allowed us to identify with a good score albumin (ALBU_HUMAN Serum albumin), presumably due to some peptides that are common also to BSA. No other human proteins were identified. We preformed a search against bacterial database; for all the samples we identified approximately the same proteins (essentially from bacterial lysate); the results are reported in Table 3.

**MALDI-TOF data analysis**

Both MALDI-TOF and LC-ESI data identified essentially proteins from SNAP-NAPPA chemistry and from bacterial lysate. These results were not surprising considering the high complexity of the samples analyzed and considering that the concentration of the proteins expressed and captured on the array is, at least in solution; hundred times lower than those of E. coli lyase components.

From Shimizu and co-workers results [18] we know that the PURE system components concentrations are in the range 1.5-40 µg/µl while...
the proteins are expressed in a concentration of about 0.10 µg/µl. Evidently, even if after proteins synthesis and capture the slides have been carefully washed, some lysate proteins remained specifically bound on the slide gold surface. The presence of these “background” molecules represents the main obstacle to the samples identification and we experienced different paths to overcome this obstacle.

Exploiting the data obtained from MM samples analysis we subtracted - the MM experimental mass list to those of known samples and performed a further MASCOT data bank search. Again no significant identifications were obtained.

One of the main advantages of PURE system, that prompted us to use it, is that its components are all recombinant proteins, so all known and well characterized. Thanking advantage of this aspect we built a data bank of all the theoretical mass lists belonging to PURE system recombinant proteins and subtracted them from the experimental mass lists of known samples. The samples protein identification was not possible again.

A further aspect to take in account when analyzing MS data is that the proteins immobilized on the SNAP-NAPPA were synthesized with a SNAP tag and a FLAG tag that could also contribute to the difficulty in matching spectra with databases that are based on tryptic digests of natural proteins. It was then useful to consider strategies that compensate for this; we modified the sequence of sample proteins present in the reference database, adding the tags.

We used these modified sequences to perform a new fingerprint: the theoretical mass lists of the chimeras after in silico trypsin digestion were obtained by means of the software Sequence Editor included into the Biotools package. We matched the experimental mass lists with these theoretical mass lists. The peaks identified are reported in Table 4 together with the chimeras sequences (underlined the fragment identified). The peptides of SNAP tag are in italic and those of FLAG tag in bold italic. The sequence coverage was calculated as the ratio between the number of residues matched and the total number of protein residues (Table 4).

The results obtained allow us to identify CDK2 sample with a percentage of sequence coverage of 22% and sample p53 with a percentage of sequence coverage of 6% for p53 while for -SRC-SH2 and PTPN11- SH2 samples no fragments were identified.

The results obtained from SNAP-NAPPA analysis seem worse if compared with those relative to the NAPPA presented in our previous study [5]. In the previous research the MASCOT databases search also turned out a difficult task, but considering the chimeras sequences we obtained percentages of coverage between 20% and 40%.
The following:

Selection noise subtraction of the mastermix + lysate spectrum was performed on the whole spectra. After region binning window of 10, 100 and 1000 m/z values. The same conditions were applied for both spectra preprocessed with and without ROI extraction this means that each spectrum was preprocessed with a threshold of 0.1 m/z, over the Intensity axes, was applied to every protein spectrum. Finally, in order to overcome noise troubles a threshold of 0.05 m/z was applied to every protein spectrum. SpADS is able to provide results, of the so far discussed preprocessing functions, in an ASCII file. The found peaks were submitted to MASCOTT and results are showed in Figures.

Some tests were performed in order to recognize protein spectra, in particular two main tests were performed for each protein: the former a homologous result was found, indeed CSK2, a casein kinase previously described for two protein spectra i.e. CDK2 and p53. For the latter was performed on the whole spectra. After region selection noise subtraction of the mastermix + lysate spectrum was applied too. A specimen of four different proteins was used for these tests, as in submitted to MASCOTT and results are showed in figures.

Results of the matching of CDK2 and p53 samples experimental mass lists with the theoretical mass list obtained from the trypsin digestion of the sequence of native protein in red and those of FLAG tag in blue. The sequence coverage is calculated as the ratio between the number of residues matched and the total number of protein residues.

### Matching algorithm

In parallel to known samples identification trough MASCOT databank search we developed a matching algorithm to match known and unknown samples.

In order to evaluate the goodness of SpADS preprocessing on SNAP/MS spectra, single spectrum routines of SpADS were used to preprocess data and view results of their application on SNAP/MS protein Spectra. Tests were conducted applying different binning windows for peak extraction this means that each spectrum was preprocessed with a binning window of 10, 100 and 1000 m/z values. The same conditions were applied for both spectra preprocessed with and without ROI selection. Finally, in order to overcome noise troubles a threshold of 0.05 m/z was applied to every protein spectrum. SpADS is able to provide results, of the so far discussed preprocessing functions, in an ASCII file. The found peaks were submitted to MASCOTT and results are showed in figures.

In a ROI between 1000 and 1200 Da was selected in order to highlight differences in spectra. These tests were performed as in submitted to MASCOTT and results are showed in figures.

In a ROI between 1000 and 1200 Da was selected in order to highlight differences in spectra. These tests were performed as in submitted to MASCOTT and results are showed in figures.

### Clustering

Clustering proposed solutions are showed in Figures 8 and 9 for.

**Table 4:** Results of the matching of CDK2 and p53 samples experimental mass lists with the theoretical mass list obtained from the trypsin digestion of the sequence of native protein plus SNAP tag and FLAG tag. After matching results there are reported the sequences of the chimera proteins: the peptides of SNAP tag are in black, those of FLAG tag in blue. The sequence coverage is calculated as the ratio between the number of residues matched and the total number of protein residues.

| Range | Theoretical | Experimental | DELTA | Sequence |
|-------|-------------|--------------|-------|----------|
| [491-504] | 1467.768 | 1467.817 | 0.049 | ALPNNTSSSPKPKK |
| [542-554] | 1361.680 | 1361.786 | 0.106 | EPGSARHSSHLK |
| [11-20] | 1086.603 | 1086.569 | 0.034 | RTTLDPLGK |

**Match to:** SNAP-CDK2 & human-FLAG

**Number of mass values matched:** 65

**Number of mass values matched:** 62

**Sequence Coverage:** 22.4%

**Percentage of experimental masses matched (with background peaks):** 12.3%

**Identified Fragments**

| Range | Theoretical | Experimental | DELTA | Sequence |
|-------|-------------|--------------|-------|----------|
| [11-20] | 1086.603 | 1086.569 | 0.034 | RTTLDPLGK |

**Match to:** SNAP- p53 – FLAG

**Number of mass values matched:** 3

**Sequence Coverage:** 6.3%

**Percentage of experimental masses matched:** 4.8%

**Sequence Coverage:** 22.4%

**Number of mass values matched:** 8

**Number of mass values searched:** 65

**Percentage of experimental masses matched:** 12.3%

**Table 4:** Results of the matching of CDK2 and p53 samples experimental mass lists with the theoretical mass list obtained from the trypsin digestion of the sequence of native protein plus SNAP tag and FLAG tag. After matching results there are reported the sequences of the chimera proteins: the peptides of SNAP tag are in black, those of FLAG tag in blue. The sequence coverage is calculated as the ratio between the number of residues matched and the total number of protein residues.

- **p53**

  Tests were conducted applying different binning windows for peak extraction this means that each spectrum was preprocessed with a binning window of 10, 100 and 1000 m/z values. The same conditions were applied for both spectra preprocessed with and without ROI selection. Finally, in order to overcome noise troubles a threshold of 0.05 m/z was applied to every protein spectrum.
the “only protein” specimen with a ROI selection of 1000/2000 and 1000/1200 respectively. The same results are shown in related Tables 5 and 6. While results for the ROI 1000/2000 are cluster overlapped and hard to investigate using a restricted ROI of 1000/1200 and a more precise sampling approach clusters are suitable and understandable without any further software intervention. In order to compare results, the same processing was then performed for the second specimen, composed of the 56 spectra with known and unknown proteins. In order
Figure 5: Samples mass spectra acquired by Ultraflex III MS. Each one of this spectrum is the sum of 100 single shot spectrum a) full range; b) 1.1 - 2.4 kDa range.

Figure 6: UltraflexIII samples mass spectra summation. The arrows point at the theoretical peak position.

Table 5: Cluster assignment for each known protein sample on a specimen of 23 samples in the ROI 1000/2000 with a binning window of 1000 m/Z. Statistics are based on the SpADS results coupled with K Means clustering given in Figure 8.

| Cluster Assignment |
|-------------------|
| Cluster Number 1 2 3 4 |
| P53 (p) 2 2 1 - |
| PTPN11-SH2 (pt) - 2 2 2 |
| Src-SH2 (s) - 3 - 1 1 |
| CDK2 (c) 3 - 2 - |

Table 6: Cluster assignment for each known protein sample on a specimen of 23 spectra in the ROI 1000/1200 with a binning window of 500 m/Z. Statistics are based on the SpADS results coupled with K Means clustering given in Figure 9.

| Cluster Assignment |
|-------------------|
| Cluster Number 1 2 3 4 |
| P53 (p) 2 3 1 - |
| PTPN11-SH2 (pt) 3 - - 3 |
| Src-SH2 (s) 3 - 1 1 |
| CDK2 (c) - - 2 3 |
to couple these unknown samples with the right protein spectra, preprocessing and clustering algorithms were then run. As in the previous test results cluster for ROI 1000/2000 results in overlapped ensembles hard to evaluate, for this reason this latter is not shown while results are shown for clustering of specimen in region 1000/1200, Figure 8 and related Table 7 (Tables 7 and 8).

The results obtained on the unknown assignment (Table 8) through the bioinformatic processing, appears striking without any further human intervention.

Conclusions

We have here presented our analysis of SNAP-NAPPA, an improved version of NAPPA with a SNAP tag, expressed with a novel cell-free transcription/translation system reconstituted from the purified components necessary for E. coli translation, the PURE system [20], and analyzed by fluorescent label and by label-free Mass Spectrometry.

The fluorescence analysis carried out demonstrated not only the proper SNAP-NAPPA behaviour but also that E. coli IVTT systems ensured a protein yield about 20 times higher respects RRL (Figure 1).

The Mass Spectrometry coupled with ad hoc implemented bioinformatics, as it was expected due to the high complexity of the NAPPA-SNAP system, gave quite encouraging results improving earlier findings with MS without SNAP (5) were very complex and a bioinformatics tool has been developed ad hoc for their analysis [21]. The MS samples were realized printing SNAP-NAPPA spots on gold coated glass slides in a special geometry in order to obtain an amount of protein appropriate for MS analysis.

The samples were printed in 12 boxes of 7×7 spots per box. One box apiece was reserved to the sample genes (p53, CDK2, SH2-Src and SH2-PTPN11), two boxes were negative controls (MM) and reference samples, and six boxes, were printed with the sample genes in an order blinded to the researcher who made MS analysis. We conducted two parallel identifications, the first through the matching algorithm comparing blinded and known samples experimental mass lists, and the second submitting experimental mass lists to databank search.

The databank search of samples experimental mass lists obtained by MALDI-TOF or LC-ESI-MS provided the identification, with significative scores, of molecules of MM or E. coli lysate (Figure 3). Then different strategies have been addressed to overcome the presence of these “background” molecules that represented the main obstacle to the samples identification. Experimental master mix plus E. coli lysate mass lists have been subtracted to samples experimental mass lists and the results have been submitted to MASCOT databank search. Unfortunately this strategy did not give statistically significative results.
Table 7: Cluster assignment for each known protein sample on a specimen of E. Coli lysate (in process).

Table 8: Comparison between the actual protein deposition in the NAPPA array and the assignment made by cluster analysis as explained in the text and in Table 7.

Table: Cluster Assignment

| Cluster Number | 1 | 2 | 3 | 4 |
|----------------|---|---|---|---|
| P53 (p)        | 1 | 1 | 4 | 66% |
| CDK2 (c)       | 3 | 50% | 1 | 20% |
| A = Src – SH2  | 1 | 2 | 3 | 50% |
| B = cdk2       | 2 | 40% | 2 | 1 |
| C = p53        | 4 | 50% | 1 | 20% |
| D = PTPN11-SH2 | 1 | 2 | 20% | 2 |
| E = PTPN11-SH2 | 4 | 100% | - | - |
| F = PTPN11-SH2 | 2 | 4 | 50% | 1 |

Table: Unknown Assignment

| Unknown Samples | Actual Festa Deposition | Via Cluster analysis |
|-----------------|-------------------------|----------------------|
| A               | Src – SH2               | Src – SH2            |
| B               | cdk2                    | cdk2                 |
| C               | p53                     | p53                  |
| D               | PTPN11-SH2              | PTPN11-SH2           |
| E               | – SH2                   | Src                  |
| F               | p53                     | PTPN11-SH2           |

References

1. Rs S, Mehta G, Srivastava S (2010) Label-free detection techniques for protein microarrays: prospects, merits and challenges. Proteomics 10: 731-48.

2. LaBaer J, Ramachandran N (2005) Protein microarrays as tools for functional proteomics. Curr Opin Chem Biol 9: 14-19.

3. Espina V, Woodhouse EC, Wulfkuhle J, Asmussen HD, Petricoin EF 3rd, et al. (2004) Protein microarray detection strategies: focus on direct detection technologies. J Immunol Methods 290: 121-133.

4. Nagl S, Michael S, Wolfbeis OS (2005) Fluorescence analysis in microarray technology. Microchim Acta 151: 1-21.

5. Spera R, La Baer J, Nicolini C (2011) MALDI-TOF characterization of NAPPA-generated proteins. J Mass Spectrom 46: 960-965.

6. Spera R, Terenico TBC, Nicolini C (2013) C NAPPA based nanogranimetric biosensor: preliminary characterization, Sensors and Actuators B: Chemical 182: 682-688.

7. Spera R, Bragazzi NL, Pechkova E, LaBaer J, Nicolini C (2013) Conductometric Monitoring of Protein-Protein interactions. Journal of Proteome Research.

8. Yu X, Xu D, Cheng Q (2006) Label-free detection methods for protein microarrays. Proteomics 6: 5493-5503.

9. Nedelkov D, Nelson RW (2003) Delineating protein-protein interactions via biomolecular interaction analysis-mass spectrometry. J Mol Recognit 16: 9-14.

10. Ramachandran N, Hainsworth E, Bhullar B, Eisenstein S, Rosen B, et al. (2004) Self-assembling protein microarrays. Science 305: 88-90.

11. Ramachandran N, Raphael JV, Hainsworth E, Demirkan G, Fuentes MG, et al. (2008) Next-generation high-density self-assembling functional protein arrays. Nat Methods 5: 535-538.

12. Juilleraat A, Heinis C, Siefart I, Barnikow J, Jaccard H, et al. (2005) Engineering substrate specificity of O6-alkylguanine-DNA alkyltransferase for specific protein labeling in living cells. ChemBiochem 6: 1263-1269.

13. Keppler A, Gredenzis S, Gromenreter T, Pich H, Vogel H, et al. (2003) A general method for the labeling of fusion proteins with small molecules in vivo. Nat Biotechnol 21: 86-89.

14. Festa F, Rollins SM, Vattem K, Hathaway M, Lorenz P, et al. (2013) Robust microarray production of freshly expressed proteins in a human milieu. Proteomics Clin Appl 7: 372-377.

15. Jongsma MA, Litjens RH (2006) Self-assembling protein arrays on DNA chips by auto-labeling fusion proteins with a single DNA address. Proteomics 6: 2650-2655.

16. Matsuura T, Kazuta Y, Aita T, Achi J, Yomo T (2009) Quantifying epistatic interactions among the components constituting the protein translation system. Mol Syst Biol 5: 297.

17. Sieg W, Erdmann VA (1985) The potentials of the in vitro protein biosynthesis system. J Biotechnol 41: 81-90.

18. Shimizu Y, Inoue A, Tomari Y, Suzulki T, Yokogawa T, et al. (2001) Cell-free translation reconstituted with purified components. Nat Biotechnol 19: 751-755.

19. Shimizu Y, Kanamori T, Ueda T (2005) Protein synthesis by pure translation systems. Methods 36: 299-304.

20. Pechkova E, Chong R, Tripathi S, Nicolini C (2010) Cell free expression and APA for NAPPA and protein crystalllography, in Functional Proteomics and Nanotechnology-based Microarrays, Pan Stanford 121-147.

21. Belmonte L, Spera R, Nicolini C (2013) SpAdS: An R Package for Mass Spectrometry Data Preprocessing before Data Mining. Journal of Computer Science and System Biology (in press).

22. R Language Platform StatSoft.

23. Peeples, Matthew A (2011) R Script for K-Means Cluster Analysis.

24. MacQueen JB (1967) Some Methods for classification and Analysis of Multivariate Observations. Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, USA, University of California Press 1: 281-297.

25. Tartaglia M, Meiler EL, Goldberg R, Zampino G, Brunner HG, et al. (2001) Mutations in PTPN11, encoding the protein tyrosine phosphatase SHP-2, cause Noonan syndrome. Nat Genet 29: 465-468.

26. Liu BA, Jablonskowi K, Raina M, Arcé M, Pawson T, et al. (2006) The human and mouse complement of SH2 domain proteins-establishing the boundaries of phosphotyrosine signaling. Mol Cell 22: 851-868.

27. Zheng H, Li S, Hsu P, Qu CK (2013) Induction of a Tumor-Associated Activating Mutation in Protein Tyrosine Phosphatase Ptpn11 (Shp2) Enhances Mitochondrial Metabolism, Leading to Oxidative Stress and Senescence. J Biol Chem 288: 25727-25738.
28. Harris CC (1966) Structure and function of the p53 tumor suppressor gene: clues for rational cancer therapeutic strategies. J Natl Cancer Inst 88: 1442-1455.

29. Hughes BT, Sidorova J, Swanger J, Monnat RJ Jr, Clurman BE (2013) Essential role for Cdk2 inhibitory phosphorylation during replication stress revealed by a human Cdk2 knockin mutation. Proc Natl Acad Sci U S A 110: 8954-8959.

30. Zhou Y, Asahara H, Gaucher EA, Chong S (2012) Reconstitution of translation from Thermus thermophilus reveals a minimal set of components sufficient for protein synthesis at high temperatures and functional conservation of modern and ancient translation components. Nucleic Acids Res 40: 7932-7945.

31. Asahara H, Chong S (2010) In vitro genetic reconstruction of bacterial transcription initiation by coupled synthesis and detection of RNA polymerase holoenzyme. Nucleic Acids Res 38: e141.

32. Koopmann JO, Blackburn J (2003) High affinity capture surface for matrix-assisted laser desorption/ionisation compatible protein microarrays. Rapid Commun Mass Spectrom 17: 455-462.

33. Finnskog D, Ressine A, Laurell T, Marko-Varga G (2004) Integrated protein microchip assay with dual fluorescent- and MALDI read-out. J Proteome Res 3: 988-994.

34. Yates JR 3rd, Eng JK, McCormack AL (1995) Mining genomes: correlating tandem mass spectra of modified and unmodified peptides to sequences in nucleotide databases. Anal Chem 67: 3202-3210.

35. Miersch S, LaBaer J (2011) Nucleic Acid programmable protein arrays: versatile tools for array-based functional protein studies. Curr Protoc Protein Sci Chapter 27: Unit27 2.

36. Pappin DJ, Højrup P, Bleasby AJ (1993) Rapid identification of proteins by peptide-mass fingerprinting. Curr Biol 3: 327-332.