

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

Claudio Nicolini^{1,2,3*}, Rosanna Spera^{1,2}, Fernanda Festa³, Luca Belmonte^{1,2}, Shaorong Chong⁴, Eugenia Pechkova^{1,2} and Joshua Labaer³

¹Laboratories of Biophysics and Nanobiotechnology, Department Experimental Medicine, University of Genova, Italy

²Nanoworld Institute Fondazione EL.B.A. 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-

***Corresponding author:** Claudio Nicolini, Professor, President Nanoworld Institute Fondazione ELBA Nicolini, Italy, E-mail: president@fondazioneelba-nicolini.org

Received August 20, 2013; **Accepted** August 30, 2013; **Published** October 02, 2013

Citation: Nicolini C, et al. (2013) Mass Spectrometry and Florescence Analysis of Snap-Nappa Arrays Expressed Using *E. coli* Cell_Free Expression System. J Nanomed Nanotechnol 4: 181. doi:10.4172/2157-7439.1000181

Copyright: © 2013 Hassan A, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

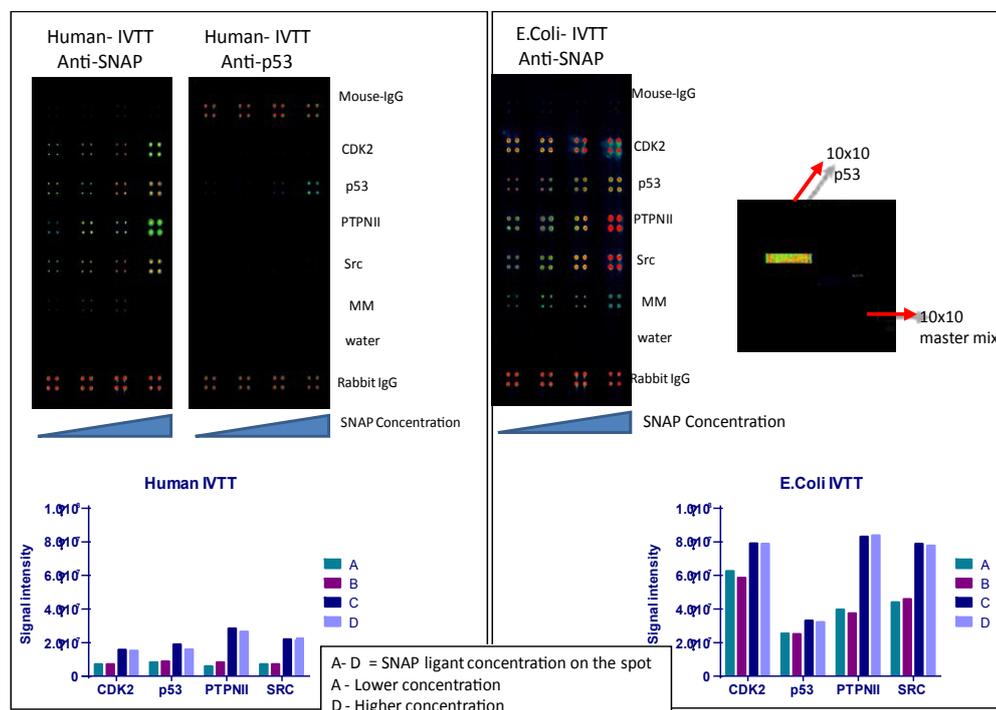


Figure 1: Fluorescence analysis of SNAP-NAPPA a) Proteins were synthesized by two different IVTT systems, 1-Step Human Coupled IVT (HCIVT) and *E. coli* IVTT. Slide images were obtained with Power Scanner and the signal intensity was quantified using the Array-ProAnalyzer 6.3. The median intensity across 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.

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 an anti-SNAP antibody printed simultaneously with the expression plasmid [14].

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.

Anyhow we decided to employ both MALDI-TOF and LC-ESI MS since, as for our previous research one of the main challenges in evaluating the mass spectra obtained from SNAP-NAPPA was the biological material present on the SNAP-NAPPA together with the target proteins, such as the BSA protein. A chromatographic step, before MS analysis, could reduce the complexity of the sample thus providing better results.

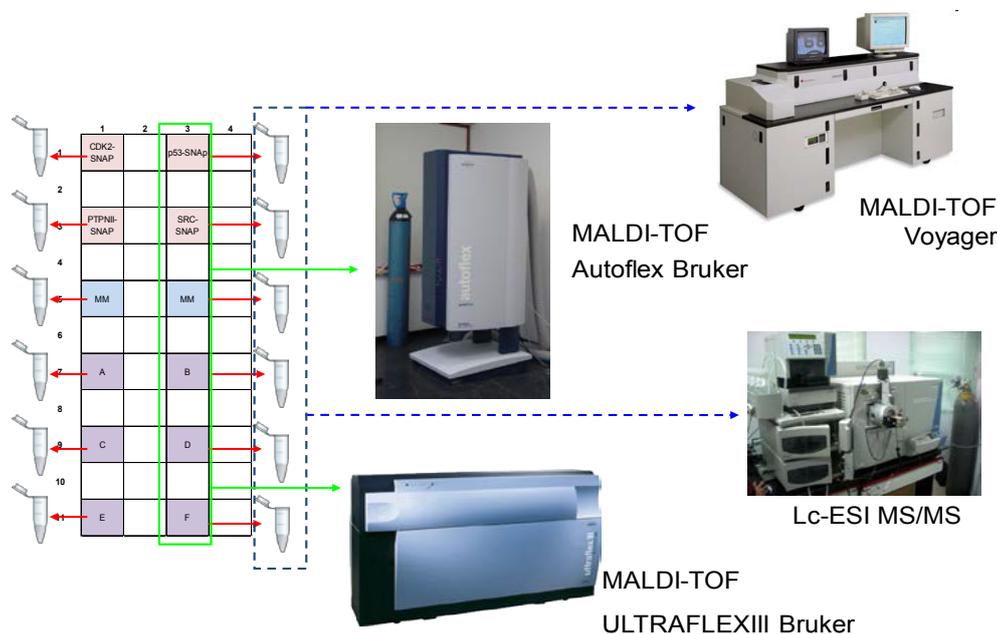


Figure 2: Experimental set-up. Samples were printed on a gold coated glass slides; the array printing was realized in a special geometry for MS analysis. The spots of 300 microns were printed in 12 boxes of 7×7 or 10×10 (spaced of 350 microns, centre to centre). The spots in a box were of the same gene: four boxes were printed with sample genes (p53, CDK2, Src-SH2 and PTPN11-SH2), two boxes were printed with master mix (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. SNAP-NAPPAs were analyzed by LC-ESI and MALDI-TOF MS. We utilized two MALDI-TOF MSs, a Voyager and a Bruker MS. For LC-ESI MS and Voyager MS analysis the sample were collected at the end of trypsin digestion and stored liquid in Eppendorf tubes since the analysis. For Bruker MS analysis the matrix was mixed with the trypsin digested fragment solutions directly on the slides and let to dry before the analysis.

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 because of 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	<i>infA</i>	GlnRS	<i>glnS</i>
IF2	<i>infB</i>	TrpRS	<i>trpS</i>
IF3	<i>infC</i>	TyrRS	<i>tyrS</i>
Methionyl-tRNA formyltransferase	<i>Fmt</i>	HisRS	<i>hisS</i>
EF-Tu	<i>TufA</i>	ProRS	<i>proS</i>
EF-Ts	<i>Tsf</i>	ThrRS	<i>thrS</i>
EF-G	<i>fusA</i>	SerRS	<i>serS</i>
RF1	<i>prfA</i>	LysRS	<i>lysS</i>
RF2	<i>prfB</i>	AsnRS	<i>asnS</i>
RF3	<i>prfC</i>	AspS	<i>aspS1</i>
RRF	<i>Frr</i>	AlaRS	<i>alaS</i>
ArgRS	<i>argS</i>	GlyRS	<i>glyS</i>
CysRS	<i>cysS</i>	PheRS $\square 2 \square 2$	<i>pheS pheT</i>
IleRS	<i>ileS</i>	Creatine kinase	
LeuRS	<i>leuS</i>	Nucleotide diphosphate. Kinase	Ndk
MetRS	<i>metG</i>	Myokinase	
ValRS	<i>valS</i>	Inorganic pyrophosphatase	Ppa
GluRS	<i>gltX</i>	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	50S ribosomal subunit protein L18	50S ribosomal subunit protein L6	30S ribosomal subunit protein S18
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 S4	50S ribosomal subunit protein L20	50S ribosomal subunit protein L9	30S ribosomal subunit protein S20
30S ribosomal subunit protein S5	50S ribosomal subunit protein L21	50S ribosomal subunit protein L10	30S ribosomal subunit protein S21
30S ribosomal subunit protein S6	50S ribosomal subunit protein L22	50S ribosomal subunit protein L11	30S ribosomal subunit protein S22
30S ribosomal subunit protein S7	50S ribosomal subunit protein L23	50S ribosomal subunit protein L13	50S ribosomal subunit protein L1
30S ribosomal subunit protein S8	50S ribosomal subunit protein L24	50S ribosomal subunit protein L14	50S ribosomal subunit protein L2
30S ribosomal subunit protein S9	50S ribosomal subunit protein L25	50S ribosomal subunit protein L15	50S ribosomal subunit protein L3
30S ribosomal subunit protein S10	50S ribosomal subunit protein L27	50S ribosomal subunit protein L16	50S ribosomal subunit protein L4
30S ribosomal subunit protein S11	50S ribosomal subunit protein L28	50S ribosomal subunit protein L32	30S ribosomal subunit protein S13
30S ribosomal subunit protein S12	50S ribosomal subunit protein L29	50S ribosomal subunit protein L33	30S ribosomal subunit protein S14
50S ribosomal subunit protein L35	50S ribosomal subunit protein L30	50S ribosomal subunit protein L34	30S ribosomal subunit protein S15
50S ribosomal subunit protein L36	50S ribosomal subunit protein L31	30S ribosomal subunit protein S16	
Ribosomal RNAs			
Protein name	Protein name	Protein name	
23S rRNA	5S rRNA	16S rRNA	
bulk tRNAs			
Protein name	Protein name	Protein name	Protein name
tRNA ^A alaT	tRNA ^A metT	tRNA ^A glyX	tRNA ^A glnX
tRNA ^A alaU	tRNA ^A metU	tRNA ^A glyY	tRNA ^A gltT
tRNA ^A alaV	tRNA ^A metV	tRNA ^A hisR	tRNA ^A gltU
tRNA ^A alaW	tRNA ^A metW	tRNA ^A ileT	tRNA ^A gltV
tRNA ^A alaX	tRNA ^A metY	tRNA ^A ileU	tRNA ^A gltW
tRNA ^A argQ	tRNA ^A metZ	tRNA ^A ileV	tRNA ^A glyT
tRNA ^A argU	tRNA ^A pheU	tRNA ^A ileX	tRNA ^A tyrU
tRNA ^A argV	tRNA ^A pheV	tRNA ^A ileY	tRNA ^A tyrV
tRNA ^A argW	tRNA ^A proK	tRNA ^A leuP	tRNA ^A valT
tRNA ^A argX	tRNA ^A proL	tRNA ^A leuQ	tRNA ^A valU
tRNA ^A argY	tRNA ^A proM	tRNA ^A leuT	tRNA ^A valV
tRNA ^A argZ	tRNA ^A sec	tRNA ^A leuU	tRNA ^A valW
tRNA ^A asnT	tRNA ^A serT	tRNA ^A leuV	tRNA ^A valY
tRNA ^A asnU	tRNA ^A serU	tRNA ^A leuW	tRNA ^A valZ
tRNA ^A asnV	tRNA ^A serV	tRNA ^A leuX	tRNA ^A glyU
tRNA ^A asnW	tRNA ^A serW	tRNA ^A leuZ	tRNA ^A glyV
tRNA ^A aspT	tRNA ^A serX	tRNA ^A lysQ	tRNA ^A glyW
tRNA ^A aspU	tRNA ^A thrT	tRNA ^A lysT	tRNA ^A glnV
tRNA ^A aspV	tRNA ^A thrU	tRNA ^A lysV	tRNA ^A glnW
tRNA ^A cysT	tRNA ^A thrV	tRNA ^A lysW	tRNA ^A trpT

tRNA ^{Gln} U	tRNA ^{Thr} W	tRNA ^{Lys} Y	tRNA ^{Tyr} T
tRNA ^{Lys} Z	tRNA ^{Val} X		

Table 1: *E. coli* IVTT components.

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;

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. The presence of background molecules in fact represents the main obstacle in the MS data interpretation and bioinformatics tools are mandatory to improve their subtraction with new matching software being recently implemented and optimised [22].

Materials and Methods

Production and expression of NAPPA

The full length cDNA for p53 (M_w 43.80 kDa), CDK2 (M_w 33.93 kDa) - both purchased from DNAsu plasmid repository of the Biodesign Institute, Arizona State University - and the SH2 domain of Src (M_w 25.95 kDa) and PTPN11 (M_w 13.13 kDa) - both purchased from Open Biosystem, Thermo Scientific- were amplified and cloned into NdeI and XhoI sites in pCOATexp SNAPf vector [30], a derivative of pCOATexp and pSNAPf (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/µl; New England Biolabs, Ipswich, MA, USA), protein cross-linker (2.2 mM BS3; Pierce, Rockford, IL, USA) and BSA (3.6 mg/ml; Sigma-Aldrich). As negative controls were prepared in printing mix solution without DNA (hereafter named master mix, MM). Similar to the gene samples, negative controls were prepared with a concentration range from 80 to 800 µg/µl of SNAP capture reagent.

As a positive control (for fluorescence analysis) mouse IgG or rabbit IgG (Pierce, Rockford, IL, USA) were added in a printing mix instead of DNA. Samples were agitated for 90 minutes at 1200 rpm in RT and printed in glass slides (VWR International, Radnor, PA, USA),

which were previously coated for 10 minutes with 2% solution of 3-Aminopropyltriethoxysilane (Pierce, Rockford, IL, USA) in acetone, rinsed in acetone and dried with filtered air.

All samples were printed on a 50 nm gold coated glass slides (Phasis, Switzerland) to allow MALDI-TOF MS analysis, using a Genetix QArray2 with 300 µm solid stealth technology pins (Arrayit Corporation, Sunnyvale, CA, USA). Arrays were stored in an airtight container at room temperature until use.

The printed slides were expressed using a reconstituted *E. coli* coupled cell-free expression system (*E. coli* IVTT) (PURExpress *in vitro* system, New England Biolabs, Ipswich, MA, USA) [14,31]; briefly, slides were blocked in SuperBlock (Thermo Scientific, Rockford, IL, USA) for 1hour at room temperature with constant agitation and dried with filtered air. HybriWells gaskets (Grace Biolabs, Bend, OR, USA) were applied on the top of the slides and 160µl of *E. coli* IVTT, prepared according to the manufacturers' instructions, was added.

Slides were incubated for 90 minutes at 30°C and 30 minutes at 15°C. For fluorescence analysis the slides were incubated for one hour of blocking/washing with PBSTM (1X PBS supplemented with 0.2% Tween 20 and 5% Milk). The levels of protein expression were assayed with anti-SNAP antibody (New England Biolabs, Ipswich, MA, USA) or anti-p53 antibody (Santa Cruz Biotechnology, Inc.; Santa Cruz, CA, 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 (spaced 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 PTPN11-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)

(that represents an updated version of the Bruker Autoflex utilized in our previous research [5], and a LC-ESI MS.

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 μ l of 0.01 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 tryptic 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 (Silica Tips FS 360-75-8, New Objective, Woburn, MA, USA), slurry-packed in-house with 5 μ m, 200 Å pore size C18 resin (Michrom BioResources, CA).

Peptides were eluted using a linear gradient from 96% A (H_2O with 5% acetonitrile and 0.1% formic acid) to 60%B (ACN with 5% H_2O 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 tryptic 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 1 Da tolerance respectively for precursor and fragment ions. A peptide has been considered legitimately identified when it achieved cross correlation scores of 1.8 for $[M+H]^+$, 2.5 for $[M+2H]^{2+}$, 3 for $[M+3H]^{3+}$, and a peptide probability cut-off for randomized identification of $p < 0.001$.

For Voyager MALDI-TOF MS analysis, since the Voyager target is too small to carry a NAPPA slide, 1 μ l of sample (collected from the array surface) was spotted on a standard Voyager target, then 1 μ l of α -cyano-4-hydroxy-cinnamic acid (HCCA, Bruker Daltonics Leipzig, Germany) saturated solution in 0.1% trifluoroacetic acid / acetonitrile (2:1) (matrix solution) was added and finally this solution was let dry. The instrument operated in the delayed extraction mode.

Peptides were measured in the mass range from 750 to 4000 Da; all spectra were internally calibrated using peaks from trypsin autolysis and processed via the Data Explorer software. Proteins were unambiguously identified by searching a comprehensive non-redundant protein database (Swiss Prot) using the program Mascot (www.matrixscience.com). Search settings allowed one missed cleavage with the trypsin enzyme selected, oxidation of methionine

as variable modification, carboamidomethylation 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 μ l of HCCA matrix solution and let it dry. To calibrate the spectra we spotted on the array surface 1 μ l 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 Biotoools 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 1 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 2x2 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.

Sample	Protein identified	Score
p53	EFTU1 (elongation factor) – <i>E. coli</i>	74
CDK2	IF3_SALTI, Translation initiation factor IF-3 - <i>E. coli</i>	73
PTPN11-SH2	IF3_SALTI, Translation initiation factor IF-3 - <i>E. coli</i>	72
SRC-SH2	SYM_STRAW, Methionine--tRNA ligase - <i>E. coli</i>	82

Table 2: MASCOT data-bank search results synthesis (about MALDI-TOF data).

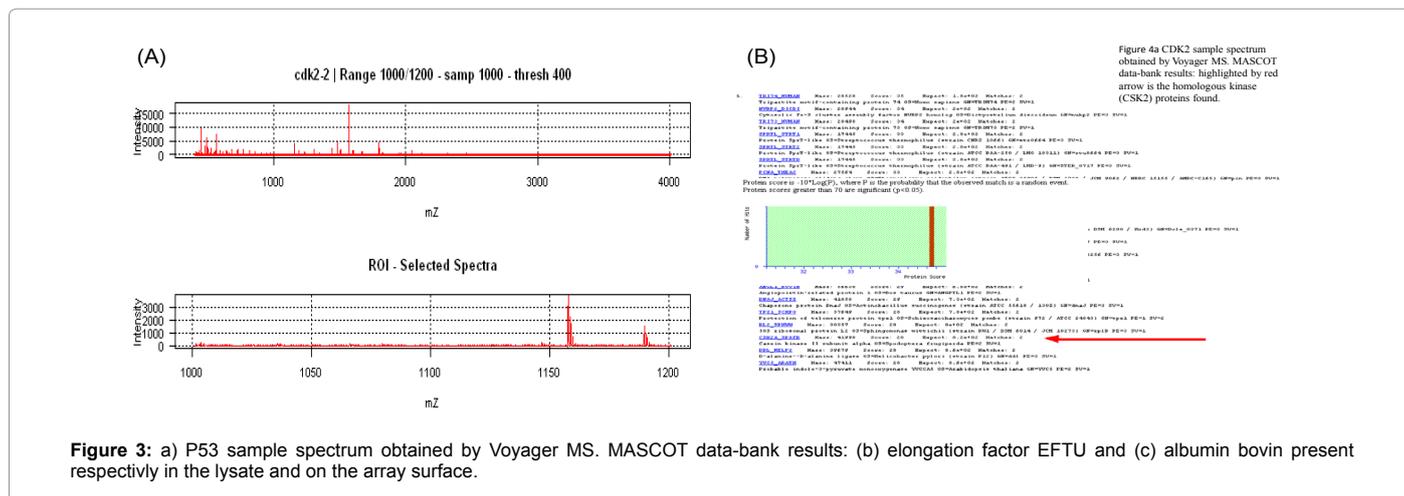


Figure 3: a) P53 sample spectrum obtained by Voyager MS. MASCOT data-bank results: (b) elongation factor EFTU and (c) albumin bovin present respectively in the lysate and on the array surface.

Proteins were synthesized by two different IVTT systems, a new system extracted from human cells (1-Step Human Coupled IVT, HCIVT, Thermo Scientific) and *E. coli* IVTT. It is known [14] that 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 from 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 $-10^3 \log(P)$, 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.

LC-ESI mass spectrometry

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 performed 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* lysate components.

From Shimizu and co-workers results [18] we know that the PURE system components concentrations are in the range 1.5-40 $\mu\text{g}/\mu\text{l}$ while

Protein	Accession	Score
IF2_CITK8 Translation initiation factor IF-2	A8AQ58	300
DNAK_CITK8 Chaperone protein dnaK	A8ALU3	160
IF1_CITK8 Translation initiation factor IF-1	A8AIJ9	153
RL22_AGGAC 50S ribosomal protein L22	P55838	141
EFTS_CITK8 Elongation factor Ts	A8ALC0	136
EFTU_ENTS8 Elongation factor Tu	A7MKI5	132
RPOL_BPT7 DNA-directed RNA polymerase	P00573	122
RPOL_BPT3 DNA-directed RNA polymerase	P07659	114
IOMPA_CITFR Outer membrane protein A (Fragment)	P24016	112
RL16_CITK8 50S ribosomal protein	A8AQK9	111
RPOL_BPK11 DNA-directed RNA polymerase	P18147	108
RL5_CITK8 50S ribosomal protein L5	A8AQK4	107
RL14_CITK8 50S ribosomal protein L14	A8AQK6	103
DNAK1_PHOPR Chaperone protein DnaK 1	Q6LUA7	103
RS2_CITK8 30S ribosomal protein S2	A8ALC1	101
IF2_CITK8 Translation initiation factor IF-2	A8AQ58	100
IF1_NITOC Translation initiation factor IF-1	Q3J7Z5	92
IF1_CITK8 Translation initiation factor IF-1	A8AIJ9	80
RS4_CITK8 30S ribosomal protein S4	A8AQJ1	89
IF2_PECCP Translation initiation factor IF-2	C6DKK3	86
RL4_CITK8 50S ribosomal protein L4	A8AQL8	79
RS9_CITK8 30S ribosomal protein S9	A8AQC0	76
EFTU_BRELN Elongation factor Tu	P42471	75
EFTU1_PHOPR Elongation factor Tu 1	Q6LVC0	70
EFTU_BACFN Elongation factor Tu	Q5L890	70
EFTU_ENTS8 Elongation factor Tu	A7MKI5	70
RL16_CITK8 50S ribosomal protein L16	A8AQK9	70
EFTU_BDEBA Elongation factor Tu	Q6MJ00	67
RS3_CITK8 30S ribosomal protein S3	A8AQL0	60
DNAK1_PHOPR Chaperone protein dnaK 1	Q6LUA7	59
RL32_CROS8 50S ribosomal protein L32	A7MFQ6	49
RS7_CROS8 30S ribosomal protein S7	A7MKJ3	49
EFTU_SOLUE Elongation factor Tu	Q01SX2	46
IF1_PHOPR Translation initiation factor IF-1	Q6LT12	45
RL6_CITK8 50S ribosomal protein L6	A8AQK0	42

Table 3: LC-ESI MS results for bacterial database matching.

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 a specifically bounded 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 form 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 *chimera* proteins sequence (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%.

Match to: SNAP-CDK2_human-FLAG				
Number of mass values searched: 65				
Number of mass values matched: 8				
Sequence Coverage: 22.4%				
Percentage of experimental masses matched (with background peaks): 12.3%				
1 MKNDKDCEMK RTTLDSPGK LELSGCEQGL HRIIFLGKGT SAADAVEVPA PAAVLGGPEP LMQATAWLNA YFHQPEAIEE				
81 FPVPALHHPV FQESFTRQV LWKLLKVVKF GEVISYSHLA ALAGNPAATA AVKTALSGNP VPILIPCHR V QGDLVDVGGY				
161 EGGLAVKEWL LAHEGHRLGK PGLGMENFQK VEKIGEGTYG VVYKARNKLT GEVVALKKIR LDTETEGVPS TAIREISLLK				
241 ELNHPNIVKL LDVIHTENK L YLVFEFLHQD LKKFMDASAL TGIPLPLIKS YLFQLLQGLA FCHSHRVLHR DLKPNQLLIN				
321 TEGAIKLADF GLARAFGVPV RYTHEVVT L WYRAPEILLG CKYYSTAVDI WSLGCIFAEM VTRRALFPD SEIDQLFRIF				
401 RTLGTPEDEV WPGVTSMPDY KPSFPKWARQ DFSKVVPLD EDGRSLLSQM LHYDPNK RIS AKAALAHPPF QDVTKVPVPHL				
481 RLDYKDDDDK				
IDENTIFIED FRAGMENTS				
Range	Theoretical	Experimental	DELTA	Sequence
[290-306]	2019.015	2019.148	0.133	SYLFQLLQGLAFCHSHR
[311-326]	1765.994	1765.922	0.072	DLKPNQLLINTEGAIK
[274-289]	1685.943	1685.995	0.052	FMDASALTGIPLPLIK
[260-272]	1663.897	1663.991	0.094	LYLVFEFLHQDLK
[191-204]	1540.814	1540.742	0.072	VEKIGEGTYGVVYK
[221-234]	1487.747	1487.939	0.192	LDTETEGVPSTAIR
[435-444]	1095.556	1095.571	0.015	VVPLDEDGR
[11- 20]	1086.603	1086.569	0.034	RTTLDSPGK
Match to: SNAP- p53 – FLAG				
Number of mass values searched: 62				
Number of mass values matched: 3				
Sequence Coverage: 6.3%				
Percentage of experimental masses matched: 4.8%				
1 MKNDKDCEMK RTTLDSPGK LELSGCEQGL HRIIFLGKGT SAADAVEVPA PAAVLGGPEP LMQATAWLNA YFHQPEAIEE				
81 FPVPALHHPV FQESFTRQV LWKLLKVVKF GEVISYSHLA ALAGNPAATA AVKTALSGNP VPILIPCHR V QGDLVDVGGY				
161 EGGLAVKEWL LAHEGHRLGK PGLGMEEPQS DPSVEPLSQ ETFSDLWKL PENNVLSPLP SQAMDDLMLS PDDIEQWFTE				
241 DPGPDEAPRM PEAAPRVAPA PAAPTPAAPA PAPSWPLSS VPSQKTYQGS YGFRGLFLHS GTAKSVTCTY SPALNKMFCQ				
321 LAKTCPVQLW VDSTPPPGTR VRAMAIYKQS QHMTEVVRRC PHHERCSDS GLAPPQLIR VEGNLRVEYL DDRNTFRHSV				
401 VVPYEPPEVG SDCTTIHYN MCNSSCMGGM NRRPILIT LEDSSGNLLG RNSFEVRVCA CAGRDRRTEE ENLRKKGEPH				
481 HELPSGSTR ALPNNTSSSP QPKKKPLDGE YFTLQIRGRE RFEMFRELNE ALELKDAQAG KEPGGSRAHS SHLKSCKGQS				
516 TSRHKKLMFK TEGPDSLDY KDDDDK				
IDENTIFIED FRAGMENTS				
Range	Theoretical	Experimental	DELTA	Sequence
[491-504]	1467.768	1467.817	0.049	ALPNNTSSSPQPKK
[542-554]	1361.680	1361.786	0.106	EPGGSRAHSSHLK
[11- 20]	1086.603	1086.663	0.060	RTTLDSPGK

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 proteins 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 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 through 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. Some tests were performed in order to recognize protein spectra, in particular two main tests were performed for each protein: the former is performed on a “region of interest” (ROI) between 1000 and 2000 mZ while 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 the following:

- CDK2
- PTPN11-SH2
- Src-SH2

- 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 400, over the Intensity axes, 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 previously described for two protein spectra i.e. CDK2 and p53. For the former a homologous result was found, indeed CSK2, a casein kinase appears in Figure 5. Similar results are apparent also by processing the p53 protein MS spectra (not shown).

Clustering

Clustering proposed solutions are showed in Figures 8 and 9 for

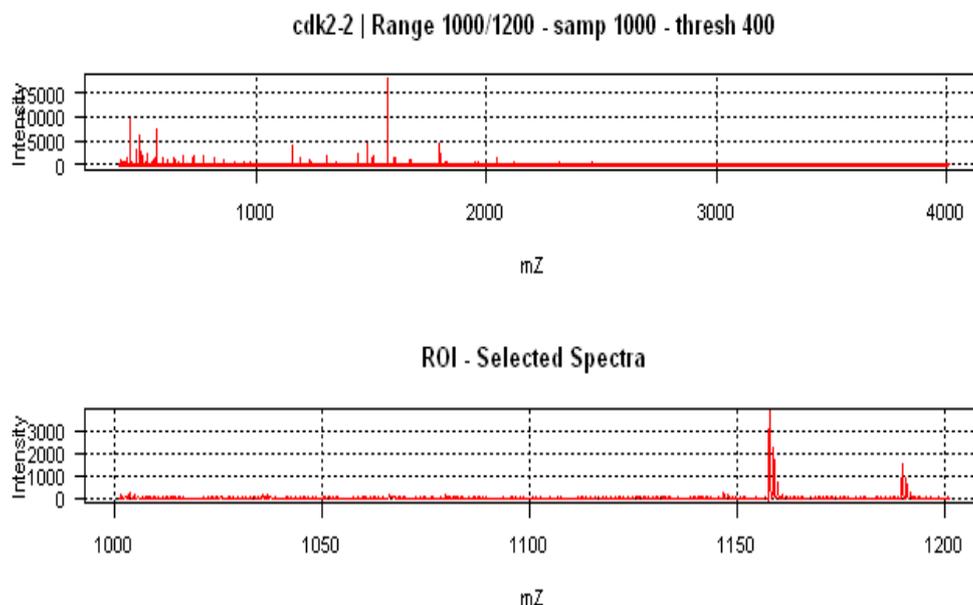
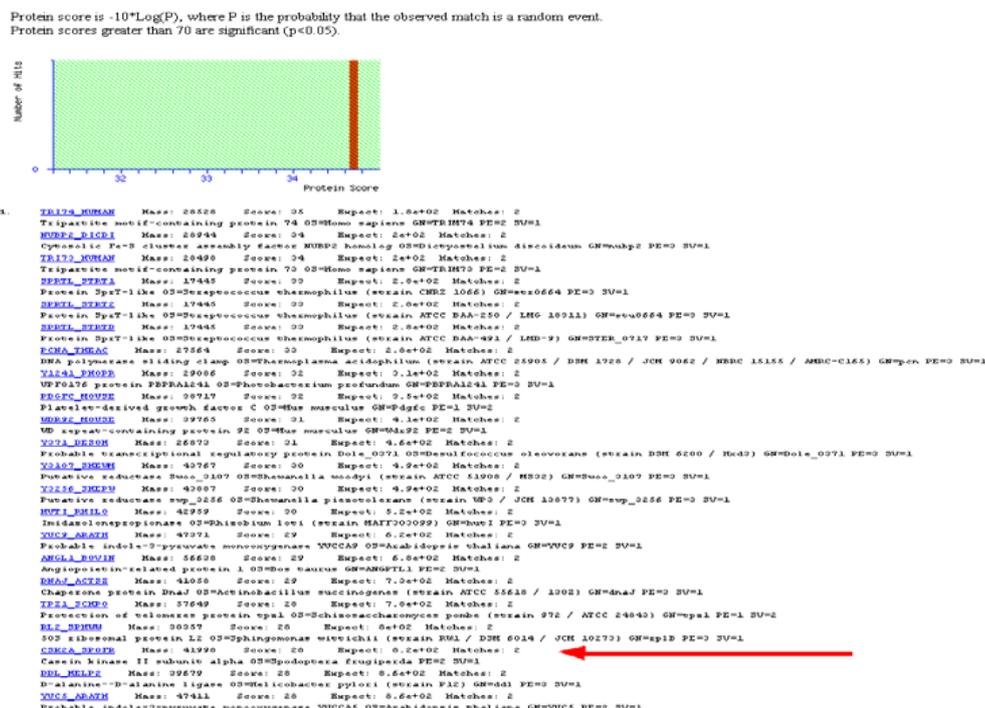


Figure 4a: Experimental mass list of CDK2 (ultraflex data) and experimental mass list [MM+ lysate] (ultraflex data) on the top. ROI selection 1000/1200 of spectra as obtained by Bruker Autoflex MS.



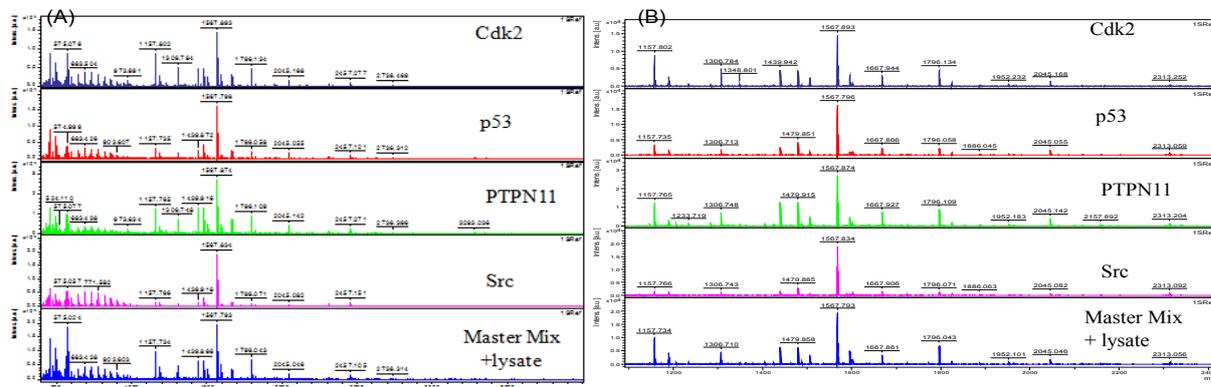


Figure 5: Samples mass spectra acquired by Ultraflex III MS. Each one of this spectrum is the sum of 100 single shot spectrum a) full range; b) 1.1 - 2.4 kDa range.

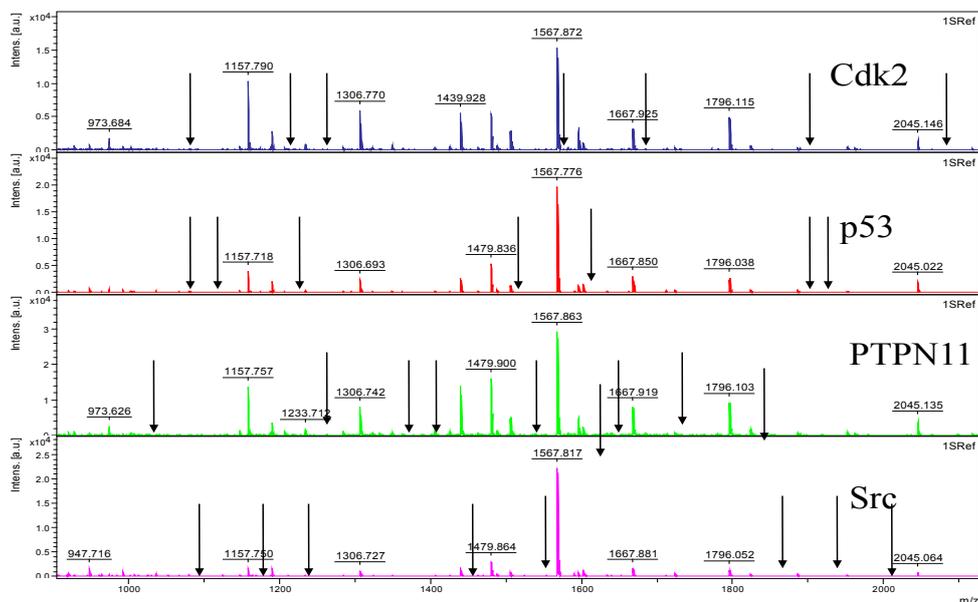


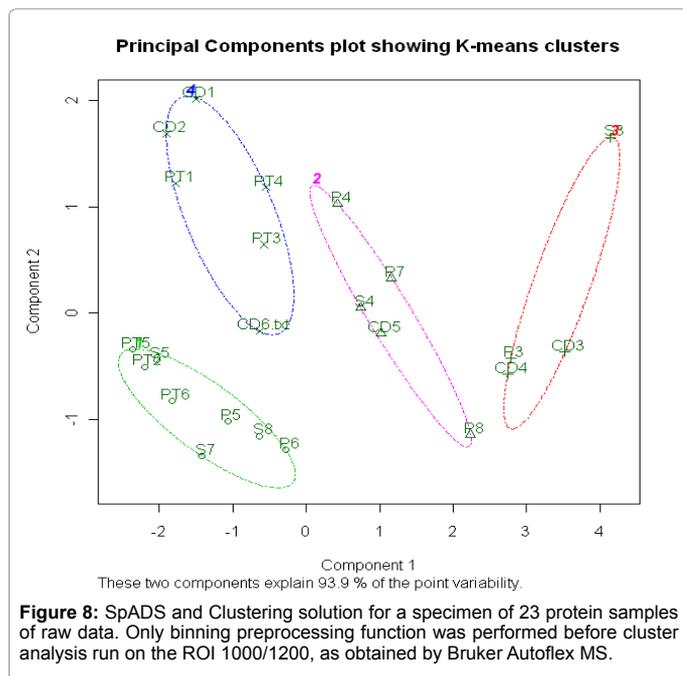
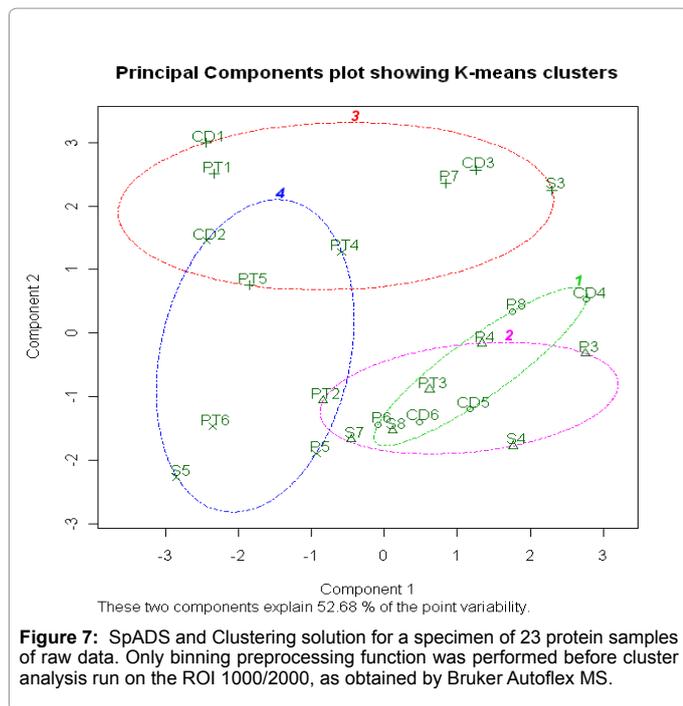
Figure 6: UltraflexIII samples mass spectra summation. The arrows point at the theoretical peak position.

Cluster Assignment				
p53 (p)	2	2	1	1
PTPN11-SH2(pt)	-	2	2	2
Src-SH2 (s)	-	3	1	1
CDK2 (c)	3	-	2	1

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	3	1	-
PTPN11-SH2 (pt)	3	-	-	3
Src-SH2 (s)	3	1	1	-
CDK2 (c)	-	1	2	3

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.



to couple these unknown samples with the right protein spectra, pre-processing 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 10 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

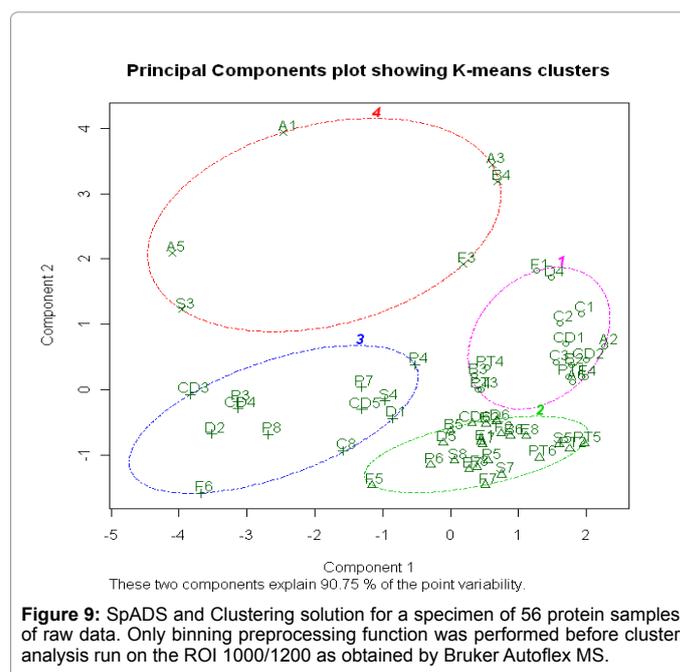
We have here presented our analysis of SNAP-NAPPA, an improved version of NAPPA with a SNAP tag, expressed with a novel cell-free transcription/translation system reconstituted from the purified components necessary for *E. coli* translation, the PURE system [20], and analyzed by fluorescent label and 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 7x7 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



Cluster Assignment				
Cluster Number	1	2	3	4
P53 (p)	1	1	4 (66%)	-
PTPN11-SH2 (pt)	3	3 (50%)	-	-
Src-SH2 (s)	-	3	1	1 (20%)
CDK2 (c)	2 (33%)	1	3	-
A = Src – SH2*	1	2	-	3 (50%)
B = cdk2	2 (40%)	2	-	1
C = p53	4	-	1(20%)	-
D = PTPN11-SH2	1	2 (20%)	2	-
E = PTPN11-SH2	-	4 (100%)	-	-
F = PTPN11-SH2	2	4 (50)	1	1

Table 7: Cluster assignment for each known protein sample on a specimen of 56 spectra. Statistics are based on the SpADS results coupled with K Means clustering given in figure 10. In bold assignment of clusters by human interpretation of cluster results. Highlighted with * striking recognition.

Unknown Assigemnt		
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	PTPN11-SH2
F	p53	PTPN11-SH2

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.

on MS of these SNAP NAPPA array, with the best identification being 22% for CDK2 sample (Figure 3) and poor clustering even on known proteins (Figure 7), apparently worse if compared with those relative to the old MS NAPPA version and presented in Spera et al. [5].

Deciding to postpone now the lengthy subtraction of the theoretical values of all lysate recombinant *E. Coli* components (work still in progress), we pursue then the coupling of our newly developed software SpADS [21] to K Means Cluster algorithm with good results both for known (Figure 8) and unknown (Figure 9) protein identification, up to 67% correct score, quite better than earlier MS without SNAP.

A conservative rule of thumb suggest that with at least hundred times more MS spectra of the unknown protein (a minimum of hundred rather than 1 as was in the limiting worst case and rather than 8 in the best case). The results so far obtained are thereby encouraging even with a quite low number of MS spectra so far acquired and without the subtraction of ab initio known MS spectra of *E. Coli* lysate (in process).

Acknowledgment

Grant Sponsor: MIUR (Ministero dell'Istruzione, Università e Ricerca; Italian Ministry for Research and University)

Grant Contract: Funzionamento (Fondazione El.B.A Nicolini); FIRB Italanonet (RBPR05JH2P) from MIUR to Professor Claudio Nicolini of the University of Genova.

The authors are very grateful to Prof. Marco Crescenzi for his precious cooperation and to Dr. Serena Camerini and Dr. Marialuisa Casella for performing LC-ESI MS and Vojager MALDI-TOF MS measures at the Dept of Hematology, Oncology and Molecular Medicine, Higher Institute of Health, Rome (Italy).

References

1. Ra 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. *Microchimica Acta* 151: 1-21.
5. Spera R, Labaer J, Nicolini C (2011) MALDI-TOF characterization of NAPPA-generated proteins. *J Mass Spectrom* 46: 960-965.
6. Spera R, Terencio TBC, Nicolini C (2013) C NAPPA based nanogravimetric 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 I 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: 86-90.
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. Juillerat A, Heinis C, Sielaff 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, Gendrezig S, Gronemeyer T, Pick H, Vogel H, et al. (2003) A general method for the covalent 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, Adachi J, Yomo T (2009) Quantifying epistatic interactions among the components constituting the protein translation system. *Mol Syst Biol* 5: 297.
17. Stiege W, Erdmann VA (1995) The potentials of the in vitro protein biosynthesis system. *J Biotechnol* 41: 81-90.
18. Shimizu Y, Inoue A, Tomari Y, Suzuki 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 crystallography, 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. Peebles, 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, Mehler 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, Jablonowski 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, Hojrup P, Bleasby AJ (1993) Rapid identification of proteins by peptide-mass fingerprinting. Curr Biol 3: 327-332.