In vitro transcription translation system

Horvatovich, Péter; Vegvari, Akos; Saul, Justin; Park, Jin G; Qiu, Ji; Syring, Micheal; Pirrotte, Patrick; Petritis, Konstantinos; Tegeler, Tony J; Aziz, Meraj

Published in:
Journal of Proteome Research

DOI:
10.1021/acs.jproteome.5b00486

Publication date:
2015

Journal of proteome research

Perspective

pubs.acs.org/jpr

In Vitro Transcription/Translation System: A Versatile Tool in the Search for Missing Proteins

Péter Horvatovich,*[,†] Ákos Végvári,[‡,§] Justin Saul,[∥] Jin G. Park,[∥] Ji Qiu,[∥] Michael Syring,[⊥] Patrick Pirrotte,[⊥] Konstantinos Petritis,[⊥,#] Tony J. Tegeler,[⊥] Meraj Aziz,[⊥] Manuel Fuentes,[▽,○] Paula Diez,[▽,○] Maria Gonzalez-Gonzalez,[▽,○] Nieves Ibarrola,[○] Conrad Droste,[◆] Javier De Las Rivas,[◆] Concha Gil,[¶] Felipe Clemente,[¶] Maria Luisa Hernaez,[¶] Fernando J. Corrales,[□] Carol L. Nilsson,[‡,§] Frode S. Berven,[■,●] Rainer Bischoff,[†] Thomas E. Fehniger,[§,▲,◇] Joshua LaBaer,[∥] and György Marko-Varga[§,▲,△]

[†]Analytical Biochemistry, Department of Pharmacy, University of Groningen, A. Deusinglaan 1, 9713 AV Groningen, The Netherlands

[‡]Department of Pharmacology & Toxicology, The University of Texas Medical Branch, 301 University Boulevard, Galveston, Texas 77555-1074, United States

[§]Clinical Protein Science & Imaging, Biomedical Center, Department of Biomedical Engineering, [▲]Center of Excellence in Biological and Medical Mass Spectrometry, Lund University, 221 84 Lund, Sweden

[∥]Center for Personalized Diagnostics, Biodesign Institute, Arizona State University, Tempe, Arizona 85287, United States

[⊥]Center for Proteomics, Translational Genomics Research Institute, Phoenix, Arizona 85004, United States

[#]Pathology Research, Phoenix Children's Hospital, 1919 East Thomas Road, Phoenix, Arizona 85016, United States

[▽]Department of Medicine and General Cytometry Service-Nucleus, [○]Proteomics Unit, [◆]Bioinformatics and Functional Genomics Research Group, Cancer Research Center (CSIC/USAL/IBSAL), 37007 Salamanca, Spain

[¶]Department of Microbiology & Proteomics Unit, University Complutense, 28040 Madrid, Spain

[□]Center for Applied Medical Research (CIMA), University of Navarra, PRB2-ProteoRed-ISCIII, IDISNA, Ciberhed, 31008 Pamplona, Spain

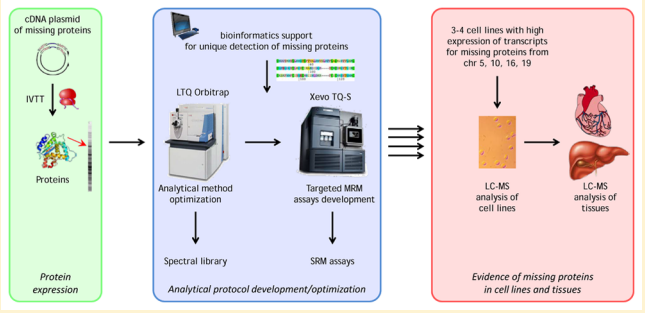
[■]Proteomics Unit (PROBE), Department of Biomedicine, University of Bergen, Postbox 7804, N-5009 Bergen, Norway

[●]The Norwegian Multiple Sclerosis Competence Centre, Department of Neurology, Haukeland University Hospital, Postbox 1400, 5021 Bergen, Norway

[△]First Department of Surgery, Tokyo Medical University, 6-7-1 Nishishinjuku Shinjuku-ku, 160-0023 Tokyo, Japan

Supporting Information

ABSTRACT: Approximately 18% of all human genes purported to encode proteins have not been directly evidenced at the protein level, according to the validation criteria established by neXtProt, and are considered to be "missing" proteins. One of the goals of the Chromosome-Centric Human Proteome Project (C-HPP) is to identify as many of these missing proteins as possible in human samples using mass spectrometry-based methods. To further this goal, a consortium of C-HPP teams (chromosomes 5, 10, 16, and 19) has joined forces to devise new strategies to identify missing proteins by use of a cell-free in vitro transcription/translation system (IVTT). The proposed strategy employs LC−MS/MS data-dependent acquisition (DDA) and targeted selective reaction monitoring (SRM) methods to scrutinize low-complexity samples derived from IVTT. The optimized assays are then applied to identify missing proteins in human cells and tissues. We describe the approach and show proof-of-concept results for development of LC−SRM assays for identification of 18 missing proteins. We believe that the IVTT system, when coupled with downstream mass spectrometric identification, can be applied to identify proteins that have eluded more traditional methods of detection.



KEYWORDS: *Missing proteins, Chromosome-Centric Human Proteome Project, LC−MS, in vitro transcription/translation system, proteomics, bioinformatics*

Special Issue: The Chromosome-Centric Human Proteome Project 2015

Received: May 28, 2015
Published: July 9, 2015

DOI: 10.1021/acs.jproteome.5b00486
J. Proteome Res. 2015, 14, 3441−3451

1. INTRODUCTION

One of the most important goals of the Chromosome-Centric Human Proteome Project (C-HPP) is to identify at least one protein product for each gene within the human genome.[1−3] When the international C-HPP was launched at the Human Proteome Organization (HUPO) congress in Sydney in 2010, 30% of the then known 20 687 protein-coding genes were ambiguously annotated by various platforms of protein measurement lacking experimental evidence of their existence at the protein level (PE1).[4] In an effort to address this discrepancy, HUPO and the C-HPP classified proteins according to the neXtProt definitions (PE2−PE5) that describe the existence of products of protein-coding genes: Protein identification at level PE2 has evidence at the transcript level, such as expressed sequence tags (ESTs) and full-length mRNA, level PE3 has evidence inferred by homology (derived by sequence similarity with known proteins in related species), level PE4 has evidence predicted from gene sequence, and level PE5 refers to uncertain and/or dubious identifications often homologous to known protein sequences.

The latest release of neXtProt (April 28, 2015, available at http://www.nextprot.org) published an updated list of human proteins coded by 20 060 genes.[5,6] The number of proteins classified at levels PE2, PE3, and PE4 together is 2771, whereas an additional number of 597 proteins are classified at level PE5 as questionable existence. These numbers can be compared with 20 055 total, 2948 PE2 + PE3 + PE4, and 616 PE5 protein entries in neXtProt 2014-09-19, the benchmark version of neXtProt for this special issue. Proteins from approximately 18% of all protein-coding genes have never been observed with high confidence at the expression level; thus, they are considered to be "missing" proteins (PE2 + PE3 + PE4) or dubious or uncertain proteins (PE5).

Within HUPO, our chromosome consortia have recently published examples of the discovery of new proteins and proteoforms using the combined resources of genomic, RNA-Seq, and proteomic databases.[7,8] An ENCODE database containing genetic information searchable with proteomic data is essential to define annotations of unknown proteins. To this end, we have translated ENCODE data into a searchable proteomic database (proteoENCODEdb) to improve identifications of novel proteoforms.[8] The databases, in combination with bioinformatic tools, allow comparisons to be made between transcript and protein data.[7,8] These new toolbox concepts, combining proteoENCODEdb searches with experimental mass spectrometric data, enable the detection of alternatively spliced forms at the transcript level and provide direct evidence of their translation to proteins.[7]

There are many factors influencing protein expression, detection, and identification that need to be investigated in order to understand the reasons that impact the discovery of missing proteins (Table 1). It should be noted that there is evidence for proteins in the literature that is not based on detection with MS or antibody approaches but is based on other, such as Edman sequencing or crystallography, data; this evidence is included in current neXtProt annotation. Both biological and experimental factors can significantly influence the success rate of these activities. Our initial studies focused on the identification of missing proteins predicted by transcript evidence (PE2) but not yet identified as expressed proteins in human cells or tissues. In this perspective, the consortia of chromosome 5 (The Netherlands), 10 (USA), 16 (Spain), and 19 (Sweden and USA) present a novel strategy to identify 565 missing proteins that are encoded by these chromosomes.

2. MATERIALS AND METHODS

2.1. In Vitro Transcription/Translation (IVTT) System

Expression vectors (pANT7_cGST) containing an insert corresponding to a protein-coding gene with a glutathione-S-transferase (GST) tag at the carboxy terminus (Supporting Information Figure S1) were acquired from the human expression clone collection of the DNASU plasmid repository at Arizona State University (https://dnasu.org). All clones from

Table 1. Factors Confounding the Discovery of Missing Proteins

interference	error type	examples
technical	MS signal absent	loss during sample preparation and analysis
		low ionization potential
		blocked N/C-terminus
		lack of tryptic cleavage site
	sensitivity to rapid proteolysis	sample degradation
	ionization/detection failure	instrument malfunction
	lack of uniquely identifiable tryptic peptides	incomplete digestion
		no unique peptide with 6−30 amino acid length
biological	low abundance in sample	cytokines, hormones
	silencing of germ cell genes	epigenetic mechanisms (histone conformation, methylation)
	potentially identifiable peptides are modified by PTM	hundreds of potential PTMs blocking sequence
	expression during cell cycle transitions/differentiation/aging	not tested with appropriate form of cell type
	developmentally regulated	embryonic
	transient expression following stimulation or perturbation	absent signal or downstream activators
	ancient genes no longer under transcriptional control in humans	conserved but inactive regions
	DNA elimination	human erythroblasts
informatics	erroneous annotation	limited information on possible translated sequences, e.g., on mitochondrial proteins
	wrongly predicted protein encoding regions	limited reference database
	parsimonious protein assembly of MS/MS identifications	nonexistent mosaic
	large variability in individual isologous sequences per sample	individual antibodies
	expression in rare cell/tissue types	rare cell type not annotated

DOI: 10.1021/acs.jproteome.5b00486
J. Proteome Res. 2015, 14, 3441−3451

the DNASU repository are full-sequence-validated. Proteins were synthesized from plasmid cDNA using the 1-step human in vitro translation kit (Thermo-Pierce) according to the manufacturer's protocol. One microgram of plasmid DNA was added to 25 μL of reaction mix and incubated at 30 °C for 1.5 h.

2.2. Enrichment of Human Recombinant Protein Expressed by the IVTT system

To enrich human recombinant GST-fusion proteins, 8 μL of a slurry (2 μL settled resin) of Pierce glutathione magnetic beads (Thermo-Pierce) was used per sample, equilibrated three times with 25× slurry volumes of phosphate-buffered saline (PBS) and suspended in a final 12.5× slurry volume with PBS at room temperature. A 100 μL aliquot of the equilibrated slurry was added to each well (2 mL). Beads were pulled down by a 96-well plate magnet, and supernatant was removed. Completed IVTT reaction solution was added to the beads, and the bead−protein mixture was shaken at 550 rpm on a plate mixer at 4 °C overnight.

The supernatant was removed, and 100 μL of wash buffer was added to the beads. The beads were mixed by shaking at 550 rpm for 5 min. Six washes were performed in this manner: two with PBS, followed by two with PBS with 863 mM NaCl, and two with 100 μL of PBS with 50 mM ammonium bicarbonate (pH 7.8). After the last wash, the beads were resuspended in 100 μL of PBS with 50 mM ammonium bicarbonate (pH 7.8) and 0.05% PPS silent surfactant (Expedeon, Inc.) and stored at −20 °C until digest. For gel electrophoresis of enriched protein samples, 5 μL of the final resin slurry was added to 2.5 μL of PBS and 7.5 μL of 2× Laemmli sample buffer (Bio-Rad) and run under reducing conditions. The gel was rinsed with water three times for 5 min with rocking and stained in SimplyBlue Coomassie stain (Life Technologies) for 30 min with rocking. The gel was destained in water overnight, with rocking, and imaged with an Alpha Innotech imager system.

2.3. Tryptic Digestion of Human Recombinant IVTT Proteins

Protein samples bound to beads were boiled at 95 °C for 10 min. The samples were then cooled on ice, and urea was added to bring the final concentration to 8 M; the samples were then reduced with 10 mM dithiothreitol (DTT) at 37 °C for 60 min and alkylated with 40 mM iodoacetic acid (IAA) at 37 °C for 30 min in the dark. Proteins were then digested with trypsin (Promega) at a 1:10 trypsin-to-protein ratio at 37 °C for 3 h with shaking. More trypsin was added, again at a 1:10 protein-to-trypsin ratio, and the sample was incubated at 37 °C overnight with shaking. Beads were centrifuged at 500*g* for 2 min, and the supernatant, which contained the digested peptides, was transferred to a new 96-well collection plate. The beads were washed once with 150 μL of 50 mM ammonium bicarbonate (pH 7.8), and the supernatant from this wash was combined with the previous supernatant. The pH was adjusted to <3.0 by 20 μL of formic acid (Promega), and the samples were incubated at 25 °C for 20 min. The digested samples were desalted using a Sep-Pak C18 1 mL Vac Cartridge (Waters) following the manufacturer's protocol with minor modifications. Briefly, the cartridge was conditioned using 1 mL of methanol followed by an aqueous solution of 0.1% formic acid. The samples were then loaded onto the cartridge and washed with 1 mL of 0.1% formic acid. The peptides were eluted from the cartridge with 1 mL of 0.1% TFA in 80% acetonitrile. Peptide samples were dried then resuspended in 50 μL of 0.1% formic acid in water. The peptide samples were stored at −20 °C until they were injected into the mass spectrometer.

2.4. DDA LC−MS/MS Analysis and Data Processing

A Waters nanoAcquity UPLC coupled to a Thermo LTQ-ETD-Orbitrap Velos was used for DDA LC−MS/MS analysis in proof-of-concept studies. The nano-LC system included a Symmetry C18 5 μm particle size, 180 μm × 20 mm trap column and a BEH130 C18 1.7 μm particle size, 100 μm × 100 mm analytical column (Waters). Mobile phases A and B consisted of 0.1% formic acid in water and 0.1% formic acid in acetonitrile, respectively. Samples were loaded onto the trap column for desalting and preconcentration with 99%/1% mobile phase A/B at a flow rate of 5 μL/min for 3 min. Peptides were separated on the analytical column at a flow rate of 500 nL/min using a linear gradient from 3% mobile phase B to 40% mobile phase B in 77 min. The electrospray ion source consisted of a nanospray head (Thermo Fisher Scientific) coupled with a coated PicoTip fused silica spray tip (360 μm o.d., 20 μm i.d., 10 μm diameter emitter orifice, New Objective, Woburn, MA, USA). Samples were analyzed using a positive ionspray voltage and heated capillary temperature of 1.9 kV and 220 °C, respectively. Mass spectrometry data were collected with the instrument operating in DDA MS/MS mode. MS survey scans (m/z 300−2000) were acquired in the Orbitrap analyzer with a resolution of 60 000 at m/z 400 and an accumulation target of 1×10^6. This was followed by the collection of MS/MS scans of the 15 most intense precursor ions with a charge state ≥2 and an intensity threshold above 500 in the LTQ, with the accumulation target of 10 000 and isolation window of 2 Da, normalized collision energy of setting of 35%, and activation time of 30 ms. Dynamic exclusion was used with repeat counts, repeat duration, and exclusion duration of 1, 30, and 60 s, respectively. Peptide and protein identifications were performed using Proteome Discoverer (Thermo Fisher Scientific, San Jose, CA, USA, version 1.2.0.208) by searching against the UniProt human database (version 2014, September 19). A fragment ion mass tolerance of 0.80 Da and a parent ion tolerance of 10 ppm were used. The iodoacetamide derivative of cysteine was specified as a fixed modification, whereas oxidation of methionine was specified as a variable modification.

2.5. SRM Analysis

A Waters XEVO TQ-S triple quadrupole was used during the first proof-of-concept study to analyze 11 IVTT-expressed proteins encoded by chromosome 10. Peptides were separated in a Waters nanoAcquity UPLC system. Digested samples (0.5−1 μg) were injected onto a trap column (Symmetry C18 5 μm 180 μm × 20 mm, Waters Co., Milford, MA, USA) using the same chromatographic conditions as those used with the Orbitrap Velos.

The 18 IVTT-expressed missing protein digests (proteins from chromosome 16 expressed in the second proof-of-concept study) were enriched using GST-agarose beads and were analyzed using an Eksigent nanoflow LC system connected to a 5500 QTRAP. Five microliters of the sample was directly loaded onto a Eksigent trap column (NanoLC Trap Set ChromXP C18 3 μm, 120 Å, 350 μm × 0.5 mm) at 3 μL/min. Following 5 min of loading, the samples were fractionated in a 75 μm × 15 cm ChromXP nano-LC column (Eksigent) at 300 nL/min. Solvent A was water with the addition of 0.1% formic acid. Solvent B was acetonitrile with 0.1% formic acid. The gradient used was 5−65% B in 30 min. The 5500 QTRAP was operated in SRM mode. The SRM transition intensities were used to trigger data-dependent enhanced product ion (EPI) scans.

DOI: 10.1021/acs.jproteome.5b00486
J. Proteome Res. 2015, 14, 3441−3451

Table 2. List of Chromosome 16 Missing Proteins Targeted in SRM Analysis after IVTT[a]

neXtProt ID	protein	Skyline selected peptides	found peptides	total found precursors	total transitions	identified peptides
NX_A5D8T8	CLEC18A. C-type lectin domain family 18 member A	19	14	18	100	9
NX_Q8WV35	LRRC29 Leucine-rich repeat-containing protein 29	9	9	11	53	9
NX_A8K8V0	ZNF785 Zinc finger protein 785	10	7	9	58	7
NX_Q6UXU4	GSG1L Germ cell-specific gene 1-like protein	7	5	8	33	7
NX_Q8N446	ZNF843. Zinc finger protein 843	8	6	9	47	6
NX_Q8IY82	CCDC135. Coiled-coil domain-containing protein lobo homologue	13	6	7	35	5
NX_A6NDY0	PABPN1L. Embryonic polyadenylate-binding protein 2	7	4	5	28	4
NX_Q9UND3	NPIPA1. Nuclear pore complex-interacting protein family member A1	4	3	4	20	3
NX_Q8WTQ4	C16ORF78 Uncharacterized protein C16orf78	10	8	8	40	2
NX_Q8WUS8	SDR42E1. Short-chain dehydrogenase/reductase family 42E member 1	8	4	4	21	2
NX_Q8TDN1	KCNG4 Potassium voltage-gated channel subfamily G member 4	11	3	3	12	2
NX_Q8N0W5	IQCK. IQ domain-containing protein K	4	3	4	19	2
NX_O95947	TBX6. T-box transcription factor TBX6	7	4	5	23	2
NX_Q9ULZ0	TP53TG3. TP53-target gene 3 protein	6	3	5	23	2
NX_P17538	CTRB1. Chymotrypsinogen B	4	4	6	30	1
NX_Q8TAZ6	CKLFSF2 CKLF-like MARVEL transmembrane domain-containing protein 2	6	2	3	12	1
NX_Q6PL45	C16ORF79 BRICHOS domain-containing protein 5	7	5	5	22	0
NX_P17023	ZNF19 Zinc finger protein 19	15	3	6	24	5

[a]Identified peptides are the sum of the 2+ and 3+ precursors spectra using the Paragon algorithm.[38]

Protein identification from EPI data was performed with Paragon software (version 4.5.0.0, build 1654) integrated in Protein Pilot 4.5. The human SwissProt database (20 233 sequences) was specified along with the following search parameters: trypsin protease, carbamidomethyl cysteine as a fixed modification, and methionine oxidation as a variable modification. The peptides identified by search of MS/MS spectra resulting in a confidence interval (CI) $\geq$ 99% were annotated as identified peptides.

SRM transitions for each protein were selected using Skyline software (MacCoss Lab, University of Washington). When possible, peptides were chosen that matched the settings: no trypsin missed cleavage and 7−25 amino acids in length, excluding those containing Met, Trp, and Cys. Transitions between m/z 300 and 1250 for each peptide were selected among the y fragments of 2+ and 3+ precursors. The SRM transition lists for both instruments were generated by default and used without any modification. The raw SRM files were analyzed in Skyline (version 2.6.0.7176) to confirm the peptides and transitions. The .wiff files obtained with the 5500 QTRAP were analyzed in Protein Pilot (AB Sciex, version 4.5) to obtain the group archives used by Skyline to generate the spectral libraries.

SRM methods were optimized with the spectral and transition data. The number of identified peptides for the missing proteins encoded by chromosome 16 and the number of derived SRM transitions are summarized in Table 2. The list of peptides targeted in SRM analysis can be found in Supporting Information Table S1.

3. RESULTS AND DISCUSSION

3.1. Challenges Associated with the Identification of Missing Proteins

Currently, the main goal of the C-HPP is to define standardized protocols of analytical methods and to provide general guidelines to the proteomics community to facilitate identification of all expressed human proteins in complex biological samples.[2,9,10] The C-HPP selected neXtProt, which keeps record of and carefully curates proteomic data submitted to public databases, such as PeptideAtlas, Human Protein Atlas, and SRMAtlas, and regularly releases updates (latest version 2015-04-28) on categories of protein evidence. Bioinformatics plays a vital role in providing accurate protein annotations linked to genomic and transcriptomic data as well as to biological studies reported in the literature. Furthermore, to improve the quality of protein databases, it is recommended that three unique peptides of each protein, generated by bottom-up MS-based proteomics, should be presented with FDR < 1% at the protein level for confident identification. However, many missing proteins have short sequences, resulting in fewer tryptic peptides during proteolysis, and may not fulfill this requirement. Identification of these proteins should be performed based on two (or even single) unique peptide sequences, if high statistical significance from MS and MS/MS spectra are obtained.

In general, there are two main criteria to provide unambiguous MS evidence for the presence of a protein in complex biological samples. First, unique peptides must exist, and, second, these peptides must be detectable and fragmentable by mass spectrometry. In our opinion, the majority of missing proteins are those with low copy numbers that may be easily hidden in highly complex enzymatically digested samples. The first challenge can be approached by use of synthetic proteotypic peptides that are predicted in silico to have a high probabilities to be detected by LC−MS. The recorded MS/MS spectra of the synthetic peptides can also be used to develop SRM assays for more sensitive detection of their endogenous counterparts in biological matrices. Using modern mass spectrometers, a 10-fold gain in sensitivity can be achieved, on average, in SRM mode compared to that with untargeted DDA employing LC−MS/MS instruments equipped with an Orbitrap or time-of-flight mass analyzer.[11] For unambiguous identification, the synthetic

DOI: 10.1021/acs.jproteome.5b00486
J. Proteome Res. 2015, 14, 3441−3451

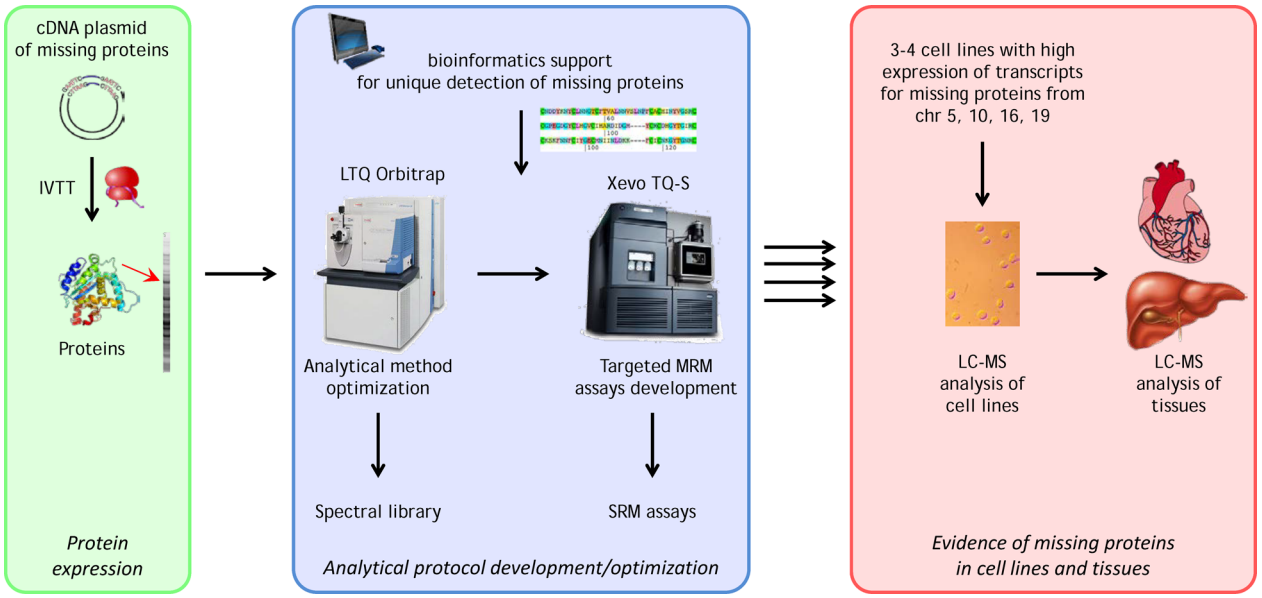


Figure 1. Strategy to identify missing proteins in human cell lines and tissues using IVTT cell-free human protein expression. IVTT-expressed protein extracts were optimized to detect the targeted missing proteins in the most sensitive and specific way. This procedure detects missing proteins while taking genetic variability and post-translational modifications by endoproteases into account. The optimized analytical protocol will be applied to analyze 3 to 4 cell lines enriched in missing proteins to provide evidence of their expression in vitro. Those findings will be further validated by analysis of proteins extracted from human tissue samples. The final step will include functional analysis of missing proteins using subcellular localization data and data obtained from homologous proteins in other species.

peptides must be suitable for ionization and MS detection with high confidence and must also possess unique sequences. To find such tryptic peptides is not a trivial task if we consider only canonical sequences of human proteins collected in databases, such as UniProt. These protein databases typically do not include all human protein sequence variants originating from genetic variability, such as nonsynonymous SNPs, alternatively spliced variants, and RNA editing events. Prediction algorithms to identify proteotypic peptides for missing proteins from primary amino acid sequences of proteins follow a general procedure according to well-established SRM assay development. This procedure applies filters to remove short peptides, to include peptides with a high predicted ionization efficiency during the electrospray process,[12,13] and to exclude peptides with a high predicted probability of having missed cleavages[14] and those with extreme retention times.[15]

3.2. The IVTT-Based Strategy for Identification of Missing Proteins

Here, we propose a novel strategy to allow the detection of missing proteins by a combination of cloning/subcloning the encoding genes into plasmid expression vectors and the use of IVTT human cell-free protein expression[16,17] coupled to LC−SRM and LC−MS/MS analyses. An overview of the approach is presented in Figure 1. In our approach, missing proteins are expressed using the plasmid repository of the human open reading frame (ORF) collection available at DNASU. A profound advantage of having the full-length proteins expressed for identifying missing proteins is the possibility of obtaining experimental data from unique peptides useful in sensitive detection of the proteins by mass spectrometry. In addition, it permits the use of other proteases that can enhance sequence coverage, produce unique peptides different from those obtained with trypsin digestion, and generate peptides with improved detectability and selectivity. The expression of full-length proteins by IVTT readily allows for optimization of instrumental settings such as MS parameters (collision energy and ion transmission) and the use of a complementary fragmentation method (such as electron transfer dissociation (ETD)[18,19] and electron-transfer and higher-energy collision dissociation (EThcD)).[20−22] Therefore, the IVTT approach is independent of the steps upstream of MS detection, such as fractionation with multidimensional chromatography,[23−25] subcellular fractionation,[10,26,27] or enrichment of modified peptides (e.g., phosphorylated,[28−30] glycosylated,[31−33] and lysine acetylated[34]), which increases the sensitivity of detection of endogenous proteins/peptides.

IVTT, in comparison to other recombinant expression systems, enables not only cell-free translation but also post-transcriptional modifications of full-length proteins from plasmid templates in high yields. Human IVTT is optimized to produce properly folded proteins with any size and character in a highly efficient manner. The proteins obtained by IVTT have the same or similar activities as those produced in human cells for the majority of expressed proteins, which is not achievable in bacterial expression systems (e.g., using *Escherichia coli*). In contrast to expression in eukaryotic cells in vivo, the cell-free IVTT system is significantly simpler. The workflow is easy to automate and allows the produced proteins to be harvested in a straightforward manner by eliminating the cell lysis step, leading to improved batch-to-batch repeatability. Additionally, IVTT-produced samples are considerably less complex, containing only the translated protein at a high concentration and the proteins required for translation. The PTMs on proteins produced by IVTT are typically phosphorylations, but disulfide bond formation and glycosylation are also achievable. It has, however, a limited ability to produce membrane and secreted proteins in correctly folded, active form.[16,17,35] Differential expression analysis can be used to quickly identify peptides derived from the translated proteins and optimize the sample preparation protocol.

DOI: 10.1021/acs.jproteome.5b00486
J. Proteome Res. 2015, 14, 3441−3451

lane	gene	GST fusion size (kDa)
1	PARK7	46.9
2	YWHAZ	54.7
3	ALDOA	66.4
4	PLIN3	74.1
5	RUVBL2	78.2
6	ATIC	91.6
7	NCL	103.6
8	EIF3C	132.3
9	PRPS1	61.8
10	G6PD	86.3
11	HNRNPU	117.6

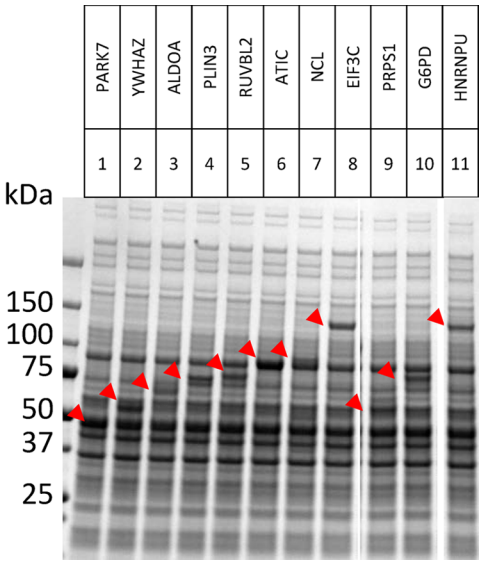


Figure 2. IVTT-based expression of proteins. Left: 11 proteins from chromosome 10 analyzed in the first proof-of-principle study; the molecular weight of the translated protein with a GST tag is shown. Right: SDS-PAGE analysis of IVTT samples stained with Coomassie blue. The red arrows indicate the position of the target proteins. The analyses show that translated proteins are expressed in high quantity in a stable background of proteins necessary for cell-free transcription/translation. These conditions allow optimized analytical assays to be developed for the target proteins.

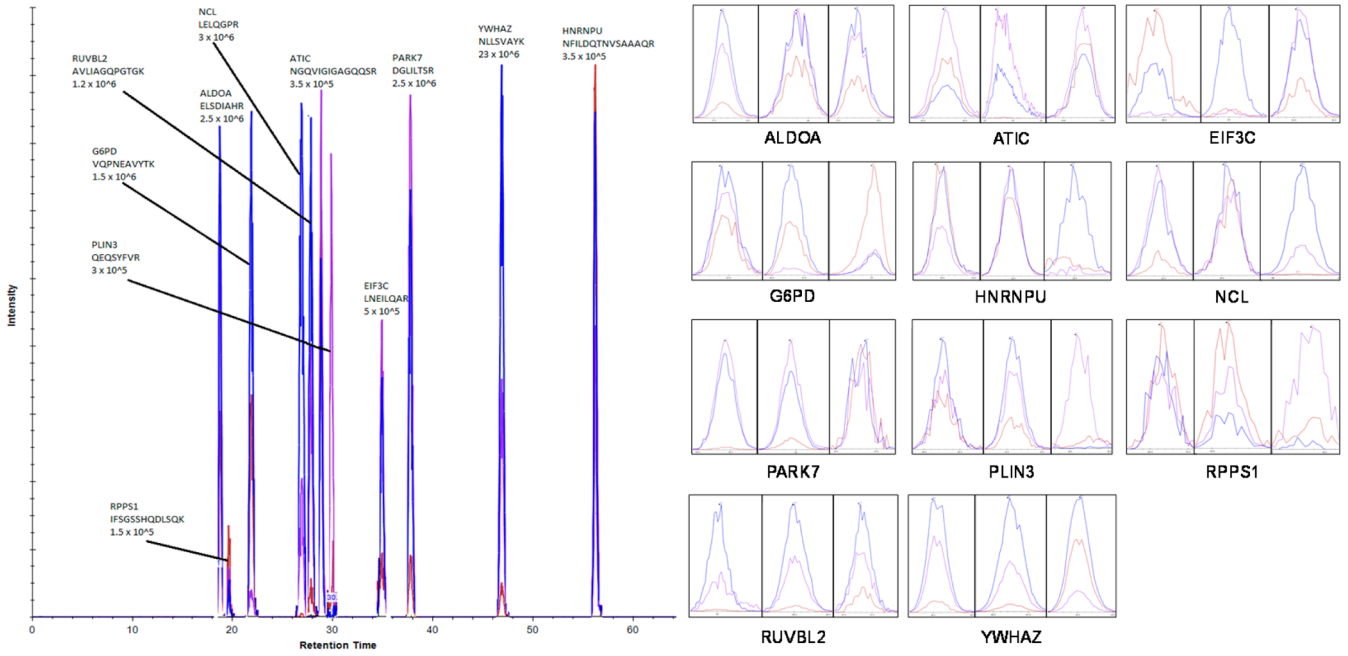


Figure 3. Left: Reconstructed extracted ion chromatograms of the LC−SRM−MS/MS analysis of the PANC-1 cell line (1 μg was injected on-column) depicting the most intense proteotypic peptides for each protein from chromosome 10 expressed in the first-proof-of-principle study. Right: Zoomed-in view of all proteotypic peptides and their corresponding transitions.

In our approach, optimized DDA LC−MS/MS analyses and SRM assays utilize spectra from IVTT-expressed protein to provide the mandatory sensitivity and specificity. We target classes of protein families that have a poor identification level by MS detection, for instance, olfactory and GPCR receptors as well as proteins with high sequence homology. We anticipate that optimized protocols using multiple IVTT-expressed proteins will allow us to distinguish alternative splice variant-derived proteins that have high sequence homology, which represents an analytical challenge.

Ultimately, the power of using IVTT to express missing proteins is that high-quality spectral libraries can be obtained from unique proteotypic peptides, enabling their specific and sensitive detection based on tandem mass spectra in complex biological samples. Spectral library searches consider the intensity of the fragment ions in the peptide spectrum match process; thus, they are more sensitive than database searches, which consider mainly the masses of theoretical fragments derived in silico from peptide and protein sequences expected to be present in the sample.[36,37] One further potential aspect of the

DOI: 10.1021/acs.jproteome.5b00486
J. Proteome Res. 2015, 14, 3441−3451

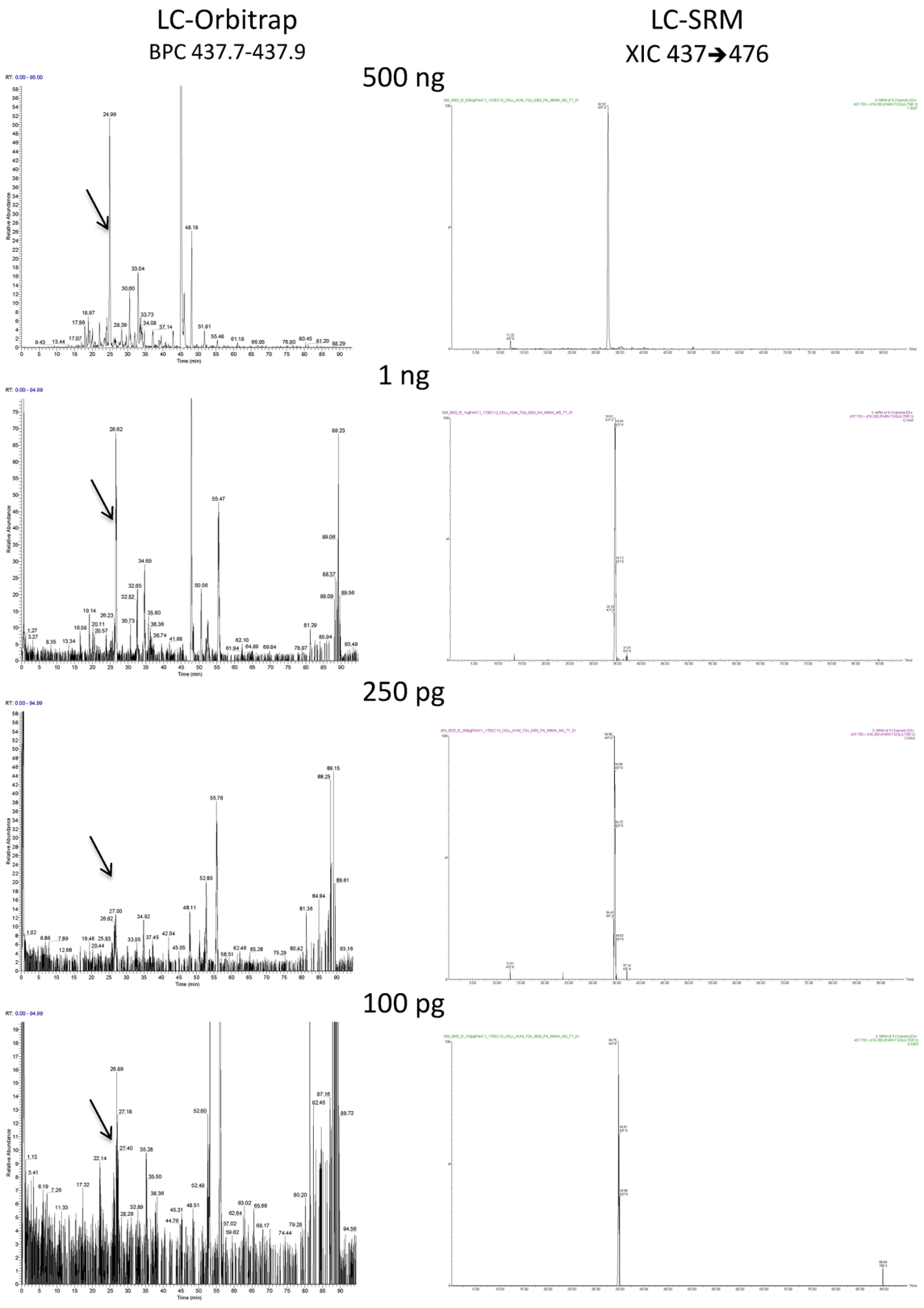


Figure 4. Sensitivity differences of DDA LC−MS/MS (right plots) and LC−SRM (left plots) approaches. Base peak chromatograms (437.7−437.9) and extracted ion chromatograms (437 → 476 m/z) are shown for the proteotypic peptide DGLILTSR derived from the PARK7 protein in a series of chromatograms obtained by injecting decreasing amounts (500, 1, 0.25, and 0.1 ng) of trypsin-digested PANC-1 cell extract. The plots show the higher sensitivity, by a factor of 100, and better selectivity of SRM analysis versus DDA single-stage MS. Arrows in the left plots indicate the peaks corresponding to the target peptide in DDA LC−MS.

IVTT system is its ability to produce proteins labeled with stable isotopes that are considered to be ideal internal standards in both qualitative and quantitative analyses. Improved analytical methods will employ bioinformatics to recognize three or

DOI: 10.1021/acs.jproteome.5b00486
J. Proteome Res. 2015, 14, 3441−3451

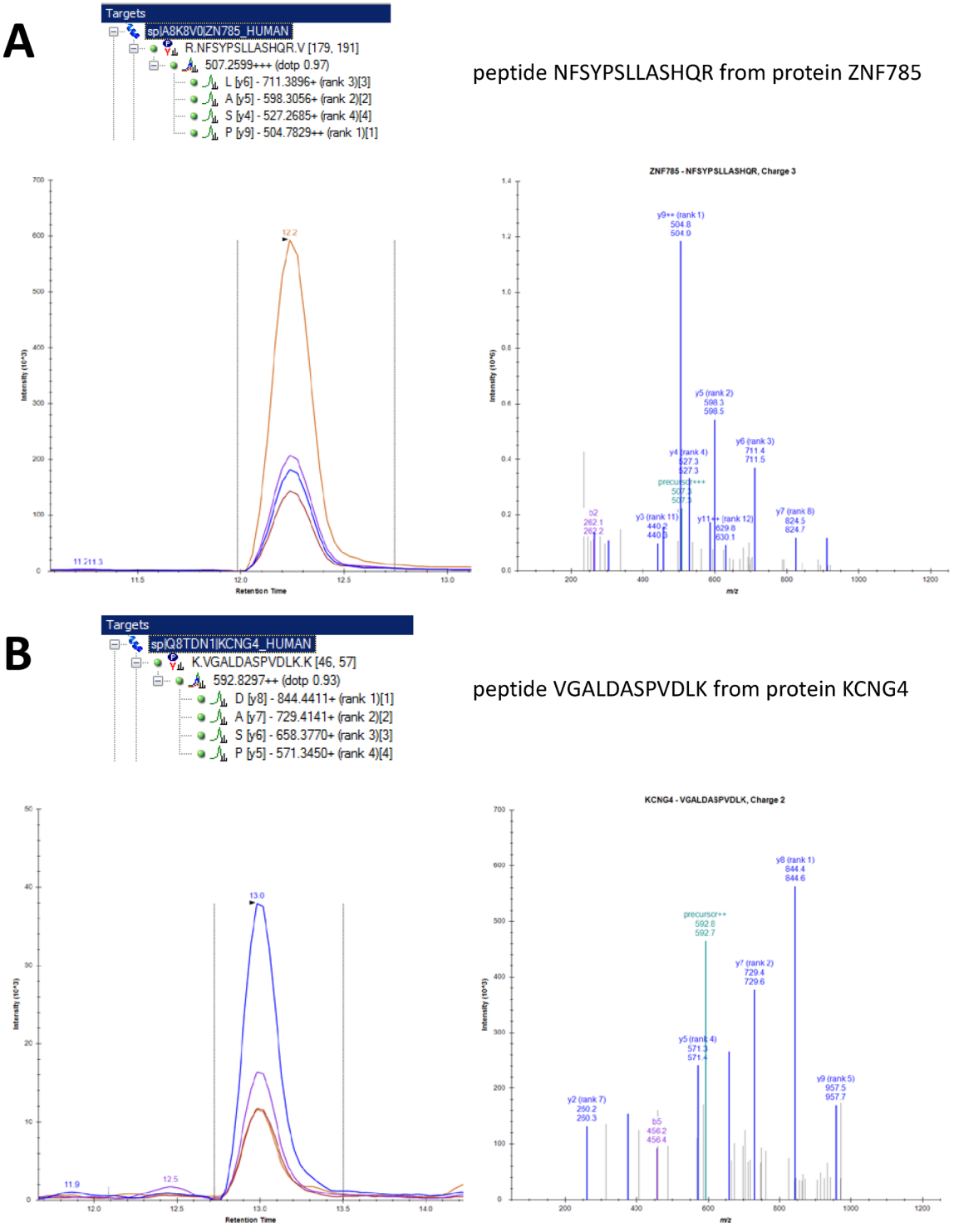


Figure 5. Extracted ion chromatograms and MS/MS spectra of one unique peptide for each of two missing proteins in an IVTT extract. (A) NFSYPSLLASHQR from ZNF785 and (B) VGALDASPVDLK from KCNG4. The *m*/*z* and ranks of transitions are highlighted in the SRM traces and the MS/MS spectra.

more unique peptides generated by different proteases for the identification of human proteins. The uniqueness of these peptides will be assessed based on canonical sequences in UniProt, but this will also take into account genetic variability, such as alternative splice variants, single amino acid variants, and eventual post-translational cleavages by endoproteases. Bioinformatic algorithms can assess fragmentation efficiency and spectral quality of peptides based on various MS/MS settings in data obtained from IVTT extracts and lead to optimization of the MS parameters.

3.3. Choice and Analysis of Biological Samples

To identify missing proteins in subcellular fractions, cell lines with a high predicted probability to contain missing proteins based on a high level of corresponding transcripts will be separated by subcellular fractionation and then analyzed by deep protein sequencing

DOI: 10.1021/acs.jproteome.5b00486
J. Proteome Res. 2015, 14, 3441−3451

methods employing 2D-LC−MS/MS. Enrichment of missing proteins predicted to be expressed at low levels will be performed using specific antibodies and analyzed by SRM or targeted DDA LC−MS/MS.

Biological sample analysis will start with a sensitive SRM assay to obtain preliminary information on the possible presence of missing proteins. This preliminary data will be used to assess if an enrichment strategy is needed to reach the required sensitivity to obtain high-quality MS/MS spectra of unique peptides. The presence of missing proteins will be ultimately confirmed by MS/MS spectra of unique peptides measured by optimized MS settings obtained during optimization of IVTT sample digest analysis. Finally, biological functions of detected proteins will be assigned based on information on homologous protein functions in other species.

A further challenge is to identify the tissues or cell types in which missing proteins have the highest probability of being expressed. Most of the missing proteins (81.8%) have mRNA evidence in chromosomes 5, 10, 16, and 19, levels PE1−PE4, whereas 70.8% of missing proteins have mRNA evidence that includes uncertain, level PE5, proteins as well, which can guide the choice of tissues and cell lines. Furthermore, RNA-Seq data stored in open-source data repositories, such as the Gene Expression Omnibus (GEO) and repositories of the Broad Institute and the Cancer Genome Atlas (TCGA) mRNA catalogues, can be analyzed to identify potential cell and tissue specimens. Cell lines and tissue types are ranked according to the abundance and number of mRNAs associated with missing proteins. Four to six cell types will be selected, which cover most of the missing proteins coded on chromosomes 5, 10, 16, and 19 in sufficient abundances. The identified cell lines will be produced in large amounts using a bioreactor and will be analyzed in six proteomic laboratories: Bergen, Norway (chromosome 19), Lund, Sweden (chromosome 19), Groningen, The Netherlands (chromosome 5), Galveston, TX, USA (chromosomes 10 and 19), and Salamanca and Madrid, Spain (chromosome 16).

3.4. Proof-of-Concept Data from IVTT-Based Targeted Mass Spectrometry Assays for Identification of Missing Proteins

As proof of concept of the approach to develop sensitive SRM assays for missing proteins, we selected 11 non-missing proteins from chromosome 10 expressed in the PANC-1 cell line. We selected proteins of relatively high abundance that could be identified by single-dimension liquid chromatography. The proteins were synthesized by IVTT, enriched using a GST tag, and digested; the proteotypic peptides were identified in data acquired in DDA LC−MS/MS mode, and a SRM method was established as described above. Figure 2 lists the gene names of the proteins that were IVTT-synthesized and shows the SDS-PAGE gels of the IVTT proteins, where the position of the target proteins is marked with a red arrow. Figure 3 shows a reconstructed SRM profile obtained from 1 μg of digested PANC-1 cell line extract where only the time window in which the best three proteotypic peptides per each IVTT protein elute is shown. In order to compare the gain in sensitivity using SRM compared to the profiling (DDA LC−MS/MS) approach, serial dilutions of the PANC-1 digest at 0.1−500 ng on-column were made and injected into a triple quadrupole and Orbitrap instrument operated in SRM and DDA LC−MS/MS mode, respectively. Our results indicate that sensitivity gains of 2−100-fold can be achieved using SRM versus a DDA approach. For example, Figure 4 shows base peak chromatograms (BPC 437.7−437.9) and extracted ion chromatogram (EIC 437 → 476 m/z) for the proteotypic peptide DGLILTSR derived from the PARK7 protein. The figure shows EICs obtained from PANC-1 cells with injected concentrations of 500, 1, 0.25, and 0.1 ng. At all analyzed concentrations, the peptide could be easily identified during the LC−SRM experiments with excellent signal-to-noise ratios even at the lowest analyzed concentration. During DDA experiments, the peptide could no longer be identified by Proteome Discoverer when injecting 1 ng of PANC-1 digest, although it could still be seen at the BPC level.

In a follow-up study, we expressed 18 missing proteins from chromosome 16 using an IVTT expression kit. Table 2 presents the number of in silico identified peptides by Skyline, the number of identified peptides, and the number of derived SRM transitions. Figure 5 shows the extracted ion chromatograms and the MS/MS spectra of one unique peptide for two missing proteins after IVTT expression.

The first study showed that IVTT expression of proteins can be used to efficiently develop sensitive SRM for proteins, resulting in improved sensitivity compared to that with DDA LC−MS/MS analysis. The second study provides proof of principle for the development of SRM assays for missing proteins.

4. CONCLUSIONS

We have developed an approach that addresses many of the analytical and technological challenges related to the identification of missing proteins in complex biological mixtures. Missing human proteins expressed in the IVTT system allow the development, optimization, and assessment of the performance of analytical measurements. The IVTT system produces human proteins with post-translational processing, mimicking the in vivo human environment. Our research will produce important analytical resources in the form of optimized LC−MS/MS methods, LC−SRM assays, and fragment spectral libraries, which can be used to detect human proteins in various types of tissue samples and facilitate their study in the context of multiple diseases. The combination of sensitive analytical methods with bioinformatics and algorithms to ascertain the presence of missing proteins in complex biological samples based on spectral similarity will facilitate the identification of missing proteins and provide an essential tool to develop bioanalytical assays for various protein isoforms occurring due to genetic variability.

■ ASSOCIATED CONTENT

Supporting Information

pANT7_cGST expression vector map (pANT7_cGST.pdf), peptides identified with the Orbitrap LC−MS/MS instrument for 11 non-missing proteins expressed with the IVTT system (IVTTOrbiPeptidespared.xlsx), list of genes with missing proteins and genes with protein evidence in chromosome 5, 10, 16, and 19 (chromosome_5_10_16_19.xlsx), SRM transitions of peptides used for analysis of PANC-1 cell line (IVTTPeptides-proteins-transitions.xlsx), and list of identified peptides after IVTT expression for 30 missing proteins of chromosome 16 (TableS1.xlsx). The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jproteome.5b00486. LC−SRM data of the 18 missing of chromosome 16 measured after IVTT translation is submitted to PASSEL with dataset identifier PASS00712 (http://www.peptideatlas.org/PASS/PASS00712).

DOI: 10.1021/acs.jproteome.5b00486
J. Proteome Res. 2015, 14, 3441−3451

AUTHOR INFORMATION

Corresponding Author

*E-mail: p.l.horvatovich@rug.nl; Tel: +31-50-363-3341; Fax: +31-50-363-7582.

Notes

The authors declare no competing financial interest.

◇We pay tribute to Thomas Edward Fehniger, a friend and recognized scientist, who passed away in July 2015 in Lund.

ACKNOWLEDGMENTS

Support from the Cancer Prevention Research Institute of Texas (RML 1122 to C.L.N.) and the University of Texas Medical Branch (C.L.N.) is gratefully acknowledged. Financial support from the Carlos III Health Institute of Spain (ISCIII, FIS PI11/02114/FIS PI14/01538 and FIS PI12/00624), Fondos FEDER (EU) is gratefully acknowledged. The proteomics units at CIC and UCM belong to ProteoRed, PRB2-ISCIII, supported by grant PT13/0001. P.D. and C.D. are supported by JCYL-EDU/346/2013 Ph.D. scholarships.

ABBREVIATIONS

IVTT, in vitro transcription/translation system; C-HPP, Chromosome-Centric Human Proteome Project; ETD, electron transfer dissociation; EThcD, electron-transfer and higher-energy collision dissociation; PE1, protein evidence (level 1); PE2, evidence at transcript level (level 2); PE3, inferred from homology (level 3); PE4, predicted protein (level 4); PE5, uncertain protein (level 5); SNP, single nucleotide polymorphism

REFERENCES

(1) Paik, Y. K.; Omenn, G. S.; Uhlen, M.; Hanash, S.; Marko-Varga, G.; Aebersold, R.; Bairoch, A.; Yamamoto, T.; Legrain, P.; Lee, H. J.; Na, K.; Jeong, S. K.; He, F.; Binz, P. A.; Nishimura, T.; Keown, P.; Baker, M. S.; Yoo, J. S.; Garin, J.; Archakov, A.; Bergeron, J.; Salekdeh, G. H.; Hancock, W. S. Standard guidelines for the chromosome-centric human proteome project. *J. Proteome Res.* **2012**, *11* (4), 2005−13.

(2) Paik, Y. K.; Jeong, S. K.; Omenn, G. S.; Uhlen, M.; Hanash, S.; Cho, S. Y.; Lee, H. J.; Na, K.; Choi, E. Y.; Yan, F.; Zhang, F.; Zhang, Y.; Snyder, M.; Cheng, Y.; Chen, R.; Marko-Varga, G.; Deutsch, E. W.; Kim, H.; Kwon, J. Y.; Aebersold, R.; Bairoch, A.; Taylor, A. D.; Kim, K. Y.; Lee, E. Y.; Hochstrasser, D.; Legrain, P.; Hancock, W. S. The Chromosome-Centric Human Proteome Project for cataloging proteins encoded in the genome. *Nat. Biotechnol.* **2012**, *30* (3), 221−3.

(3) Legrain, P.; Aebersold, R.; Archakov, A.; Bairoch, A.; Bala, K.; Beretta, L.; Bergeron, J.; Borchers, C. H.; Corthals, G. L.; Costello, C. E.; Deutsch, E. W.; Domon, B.; Hancock, W.; He, F.; Hochstrasser, D.; Marko-Varga, G.; Salekdeh, G. H.; Sechi, S.; Snyder, M.; Srivastava, S.; Uhlen, M.; Wu, C. H.; Yamamoto, T.; Paik, Y. K.; Omenn, G. S. The human proteome project: current state and future direction. *Mol. Cell. Proteomics* **2011**, *10* (7), M111.009993.

(4) ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **2012**, *489* (7414), 57−74.

(5) Lane, L.; Argoud-Puy, G.; Britan, A.; Cusin, I.; Duek, P. D.; Evalet, O.; Gateau, A.; Gaudet, P.; Gleizes, A.; Masselot, A.; Zwahlen, C.; Bairoch, A. neXtProt: a knowledge platform for human proteins. *Nucleic Acids Res.* **2012**, *40* (Database issue), D76−D83.

(6) Gaudet, P.; Argoud-Puy, G.; Cusin, I.; Duek, P.; Evalet, O.; Gateau, A.; Gleizes, A.; Pereira, M.; Zahn-Zabal, M.; Zwahlen, C.; Bairoch, A.; Lane, L. neXtProt: organizing protein knowledge in the context of human proteome projects. *J. Proteome Res.* **2013**, *12* (1), 293−8.

(7) Lichti, C. F.; Liu, H.; Shavkunov, A. S.; Mostovenko, E.; Sulman, E. P.; Ezhilarasan, R.; Wang, Q.; Kroes, R. A.; Moskal, J. C.; Fenyo, D.; Oksuz, B. A.; Conrad, C. A.; Lang, F. F.; Berven, F. S.; Vegvari, A.; Rezeli, M.; Marko-Varga, G.; Hober, S.; Nilsson, C. L. Integrated chromosome 19 transcriptomic and proteomic data sets derived from glioma cancer stem-cell lines. *J. Proteome Res.* **2014**, *13* (1), 191−9.

(8) Nilsson, C. L.; Mostovenko, E.; Lichti, C. F.; Ruggles, K.; Fenyo, D.; Rosenbloom, K. R.; Hancock, W. S.; Paik, Y. K.; Omenn, G. S.; LaBaer, J.; Kroes, R. A.; Uhlen, M.; Hober, S.; Vegvari, A.; Andren, P. E.; Sulman, E. P.; Lang, F. F.; Fuentes, M.; Carlsohn, E.; Emmett, M. R.; Moskal, J. R.; Berven, F. S.; Fehniger, T. E.; Marko-Varga, G. Use of ENCODE Resources to Characterize Novel Proteoforms and Missing Proteins in the Human Proteome. *J. Proteome Res.* **2015**, *14* (2), 603−8.

(9) Smith, L. M.; Kelleher, N. L. Proteoform: a single term describing protein complexity. *Nat. Methods* **2013**, *10* (3), 186−7.

(10) Paulo, J. A.; Gaun, A.; Kadiyala, V.; Ghoulidi, A.; Banks, P. A.; Conwell, D. L.; Steen, H. Subcellular fractionation enhances proteome coverage of pancreatic duct cells. *Biochim. Biophys. Acta, Proteins Proteomics* **2013**, *1834* (4), 791−7.

(11) Canterbury, J. D.; Merrihew, G. E.; MacCoss, M. J.; Goodlett, D. R.; Shaffer, S. A. Comparison of data acquisition strategies on quadrupole ion trap instrumentation for shotgun proteomics. *J. Am. Soc. Mass Spectrom.* **2014**, *25* (12), 2048−59.

(12) Eyers, C. E.; Lawless, C.; Wedge, D. C.; Lau, K. W.; Gaskell, S. J.; Hubbard, S. J. CONSeQuence: prediction of reference peptides for absolute quantitative proteomics using consensus machine learning approaches. *Mol. Cell. Proteomics* **2011**, *10* (11), M110.003384.

(13) Fusaro, V. A.; Mani, D. R.; Mesirov, J. P.; Carr, S. A. Prediction of high-responding peptides for targeted protein assays by mass spectrometry. *Nat. Biotechnol.* **2009**, *27* (2), 190−8.

(14) Lawless, C.; Hubbard, S. J. Prediction of missed proteolytic cleavages for the selection of surrogate peptides for quantitative proteomics. *OMICS* **2012**, *16* (9), 449−56.

(15) Escher, C.; Reiter, L.; MacLean, B.; Ossola, R.; Herzog, F.; Chilton, J.; MacCoss, M. J.; Rinner, O. Using iRT, a normalized retention time for more targeted measurement of peptides. *Proteomics* **2012**, *12* (8), 1111−21.

(16) Festa, F.; Rollins, S. M.; Vattem, K.; Hathaway, M.; Lorenz, P.; Mendoza, E. A.; Yu, X.; Qiu, J.; Kilmer, G.; Jensen, P.; Webb, B.; Ryan, E. T.; LaBaer, J. Robust microarray production of freshly expressed proteins in a human milieu. *Proteomics: Clin. Appl.* **2013**, *7* (5−6), 372−7.

(17) Saul, J.; Petritis, B.; Sau, S.; Rauf, F.; Gaskin, M.; Ober-Reynolds, B.; Mineyev, I.; Magee, M.; Chaput, J.; Qiu, J.; LaBaer, J. Development of a full-length human protein production pipeline. *Protein Sci.* **2014**, *23* (8), 1123−35.

(18) Mikesh, L. M.; Ueberheide, B.; Chi, A.; Coon, J. J.; Syka, J. E.; Shabanowitz, J.; Hunt, D. F. The utility of ETD mass spectrometry in proteomic analysis. *Biochim. Biophys. Acta, Proteins Proteomics* **2006**, *1764* (12), 1811−22.

(19) Syka, J. E.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101* (26), 9528−33.

(20) Frese, C. K.; Altelaar, A. F.; van den Toorn, H.; Nolting, D.; Griep-Raming, J.; Heck, A. J.; Mohammed, S. Toward full peptide sequence coverage by dual fragmentation combining electron-transfer and higher-energy collision dissociation tandem mass spectrometry. *Anal. Chem.* **2012**, *84* (22), 9668−73.

(21) Liu, F.; van Breukelen, B.; Heck, A. J. Facilitating Protein Disulfide Mapping by a Combination of Pepsin Digestion, Electron Transfer Higher Energy Dissociation (EThcD), and a Dedicated Search Algorithm SlinkS. *Mol. Cell. Proteomics* **2014**, *13* (10), 2776−86.

(22) Mommen, G. P.; Frese, C. K.; Meiring, H. D.; van Gaans-van den Brink, J.; de Jong, A. P.; van Els, C. A.; Heck, A. J. Expanding the detectable HLA peptide repertoire using electron-transfer/higher-energy collision dissociation (EThcD). *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (12), 4507−12.

(23) Xie, F.; Smith, R. D.; Shen, Y. Advanced proteomic liquid chromatography. *J. Chromatogr A* **2012**, *1261*, 78−90.

DOI: 10.1021/acs.jproteome.5b00486
J. Proteome Res. 2015, 14, 3441−3451

(24) Di Palma, S.; Hennrich, M. L.; Heck, A. J.; Mohammed, S. Recent advances in peptide separation by multidimensional liquid chromatography for proteome analysis. *J. Proteomics* **2012**, *75* (13), 3791−813.
(25) Horvatovich, P.; Hoekman, B.; Govorukhina, N.; Bischoff, R. Multidimensional chromatography coupled to mass spectrometry in analysing complex proteomics samples. *J. Sep. Sci.* **2010**, *33* (10), 1421−37.
(26) Drissi, R.; Dubois, M. L.; Boisvert, F. M. Proteomics methods for subcellular proteome analysis. *FEBS J.* **2013**, *280* (22), 5626−34.
(27) Fang, Y.; Robinson, D. P.; Foster, L. J. Quantitative analysis of proteome coverage and recovery rates for upstream fractionation methods in proteomics. *J. Proteome Res.* **2010**, *9* (4), 1902−12.
(28) Yang, C.; Zhong, X.; Li, L. Recent advances in enrichment and separation strategies for mass spectrometry-based phosphoproteomics. *Electrophoresis* **2014**, *35* (24), 3418−29.
(29) Beltran, L.; Cutillas, P. R. Advances in phosphopeptide enrichment techniques for phosphoproteomics. *Amino Acids* **2012**, *43* (3), 1009−24.
(30) Fila, J.; Honys, D. Enrichment techniques employed in phosphoproteomics. *Amino Acids* **2012**, *43* (3), 1025−47.
(31) Huang, B. Y.; Yang, C. K.; Liu, C. P.; Liu, C. Y. Stationary phases for the enrichment of glycoproteins and glycopeptides. *Electrophoresis* **2014**, *35* (15), 2091−107.
(32) Ongay, S.; Boichenko, A.; Govorukhina, N.; Bischoff, R. Glycopeptide enrichment and separation for protein glycosylation analysis. *J. Sep Sci.* **2012**, *35* (18), 2341−72.
(33) Ito, S.; Hayama, K.; Hirabayashi, J. Enrichment strategies for glycopeptides. *Methods Mol. Biol.* **2009**, *534*, 195−203.
(34) Zhang, K.; Tian, S.; Fan, E. Protein lysine acetylation analysis: current MS-based proteomic technologies. *Analyst* **2013**, *138* (6), 1628−36.
(35) Chong, S. Overview of cell-free protein synthesis: historic landmarks, commercial systems, and expanding applications. *Curr. Protoc Mol. Biol.* **2014**, *108*, 16.30.1−16.30.11.
(36) Li, H.; Zong, N. C.; Liang, X.; Kim, A. K.; Choi, J. H.; Deng, N.; Zelaya, I.; Lam, M.; Duan, H.; Ping, P. A novel spectral library workflow to enhance protein identifications. *J. Proteomics* **2013**, *81*, 173−84.
(37) Lam, H. Building and searching tandem mass spectral libraries for peptide identification. *Mol. Cell. Proteomics* **2011**, *10* (12), R111.008565.
(38) Shilov, I. V.; Seymour, S. L.; Patel, A. A.; Loboda, A.; Tang, W. H.; Keating, S. P.; Hunter, C. L.; Nuwaysir, L. M.; Schaeffer, D. A. The Paragon Algorithm, a next generation search engine that uses sequence temperature values and feature probabilities to identify peptides from tandem mass spectra. *Mol. Cell. Proteomics* **2007**, *6* (9), 1638−55.

DOI: 10.1021/acs.jproteome.5b00486
J. Proteome Res. 2015, 14, 3441−3451