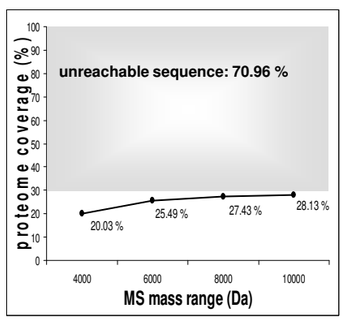
Improved Sequence Coverage During Proteome Analyses with Protein Based Isotopic Labelling Methods (ICPL™)

D. Dosch[1,4], T. Sepp[2], A. Brunner[1], J. Kellermann[1], B. Müller[3], P.C. Wright[4], F. Lottspeich[1],

[1]Max Planck Institut of Biochemistry, Martinsried, Germany, [2]BioPhoretics, Reno, NV, USA, [3]SERVA Electrophoresis, Heidelberg, Germany,
[4] Department of Chemical and Process Engineering, The University of Sheffield, U.K

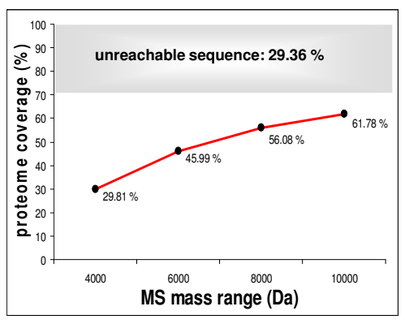
Introduction

Various stable isotope labelling strategies have been developed enabling the quantification of proteins by MS. A general workflow is the differential isotopic labelling of proteins or proteolytic peptides by stable isotopes, mixing the samples and jointly processing them. The relative quantification is done by mass spectrometry, comparing relative ion signal intensities. The sequence coverage of peptide labelling methods is usually very high. A severe drawback is that protein isoforms or modifications often can not be assigned to a specific protein. Protein labelling techniques like ICAT (Cys) or ICPL (Lys) (isotope coded protein label) overcome this weakness as the labelled mixture can be separated and fractionated on the level of intact proteins. Thereby different isoforms as well as modifications can be discriminated. Both technologies are even theoretical not able to cover the full sequence of the human proteome (fig.1).



ICAT Cysteine labeled, cleavage: Arginine/Lysin

- unlabeled proteins: 3.9%
- unreachable sequence: 70.96%



ICPL Lysine labeled, cleavage: Arginine

- unlabeled proteins: <1.37% (n-terminal labeling!)
- unreachable sequence: 29.36%

Fig.1: *Theoretical sequence coverage of the human proteome using the ICAT (a) or the ICPL (b) technology*

Many publications as well as the ABRF 2007 standard proteome experiment show, that only a limited number of proteins can be identified confidently during proteomics experiments.

Main reason for the low identification rate is low sequence coverage due to the size of peptides after proteolysis. Labelling of lysine residues reduces the number of possible cleavage sites for trypsin, usually applied during protein analyses, thereby increasing the average molecular weight of the resulting peptides. A large number of those high molecular weight peptides can not be analyzed in proteomics experiments. They either can not be extracted from the gel during cleavage or can not be detected by MS, depending on type and technology. Especially for low molecular weight proteins the small number of detectable peptides inhibits a confident protein identification.

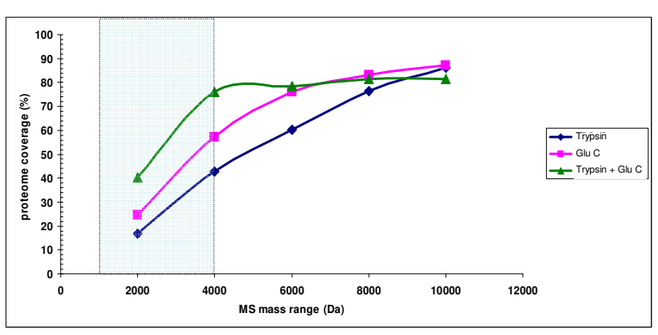


Fig 2: *Theoretical sequence coverage for ICPL labeled samples digested with Trypsin, GluC or the combination of Trypsin and GluC for different mass ranges.* ⬚ *Analysed mass range in typ. MS experiment*

Experimental

The DGPF proteome standard *(SERVA Electrophoresis , Heidelberg, Germany),* containing 8 proteins and a large number of background proteins was taken as "constant proteomic state", identical in amount in each experiment. This standard was spiked with three proteins in different amount in each "proteomic state". The mixtures were differentially labelled using the **ICPL™** kit *(Bruker, Bremen and SERVA,)* after alkylation of the cysteine residues with iodoacetamide. Then the samples were combined and digested either with trypsin *(SERVA),* endoproteinase Glu C *(SERVA)* or a combination of both enzymes. Aliquots of the sample (see Tab.1) were analyzed by LC-MALDI-MS or LC-ESI. The identification rate and sequence coverage of the different cleavage approaches were compared with each other including data of different types of mass spectrometer.

#	Protein Name	MW [kDa]	B/A	C/A	D/A
1	Cytochrome C	11,7	1	1	1
2	Myoglobin	17,8	1	1	1
3	beta Lactoglobulin	18,4	1	1	1
4	Glucose-1-Dehydrogenase	32	1	1	1
5	Lipase	33	1	1	1
6	Catalase	58	1	1	1
7	Albumin	67	1	1	1
8	Glucose Oxidase	77	1,5	0,8	1
9	Ovalbumin	42,7	1	2	5
10	Peroxidase	56,2	0,75	3	0,1

Tab.1: *Standard sample composition and expected ratios for the ICPL quadruplex experiment*

LC-MALDI

The peptide mixture was separated on a *Dionex* Ultimate 3000 nano-LC system using a C-18 column (75µm id. X 30cm) (PepMap100). Peptides were eluted with a 60-min linear gradient of 85% solvent A (0.1% TFA) to 45% solvent B (80% acetonitrile 0.04% TFA). Analysis was done on a proteomics analyzer 4700 *(Applied Biosystems).*

LC-ESI

The experiments were performed on an *Agilent* 1200 nanoflow system connected to an LTQ Orbitrap mass spectrometer *(Thermo Electron, Bremen, Germany)* equipped with a nanoelectrospray ion source *(Proxeon Biosystems, Odense, Denmark).* Binding and chromatographic separation of the peptides took place in a 15-cm fused silica emitter (75-µm inner diameter from *Proxeon Biosystems, Odense, Denmark)* in-house packed with reversed-phase ReproSil-Pur C18-AQ 3 µm resin *(Dr. Maisch GmbH, Ammerbuch-Entringen, Germany).* Peptides were eluted with a 140-min linear gradient of 98% solvent A (0.5% acetic acid *(Fluka)* in H_2O) to 50% solvent B (80% acetonitrile *(Merck)* and 0.5% acetic acid in H_2O).

Results

A theoretical digestion of the standard proteins with trypsin resulted in a sequence coverage (fig. 2) of about 40% within the mass range up to 4 kDa, routinely used during MALDI experiments.

used. A combination of both enzymes increases the number of identified peptides nearly by a factor of two. These theoretical data could be verified during wet lab experiments (Tab. 2).

Sequance coverage (%) (800-4000 Da)	Trypsin		GluC		Trypsin + GluC	
Protein Name	in silico	exp	in silico	exp	in silico	exp
Cytochrome_c	12,5	12,5	48,1	11,5	86,5	50,0
Myoglobin	29,4	9,2	84,1	85,0	80,4	80,4
Beta-lactoglobulin	23,6	23,5	53,9	56,8	51,2	49,4
Glucose1-dehydrogenase)	17,2	9,2	90,4	38,7	85,8	40,6
Lipase	39,2	28,2	9,7	4,1	41,1	40,4
Catalase	60,3	45,2	41,6	13,7	80,4	51,5
Serumalbumin	23,5	22,0	81,2	65,9	79,7	70,7
Glucoseoxidase	38,9	19,5	46,3	19,0	87,3	44,8
Ovalbumin	44,4	26,8	68,8	24,7	80,0	43,1
Peroxidase_C1A	74,7	49,4	38,1	11,0	79,2	54,9

Tab 2: *Comparison of the calculated and observed (Orbitrap) sequence coverage for the different digestion alternatives (FDR=2%)*

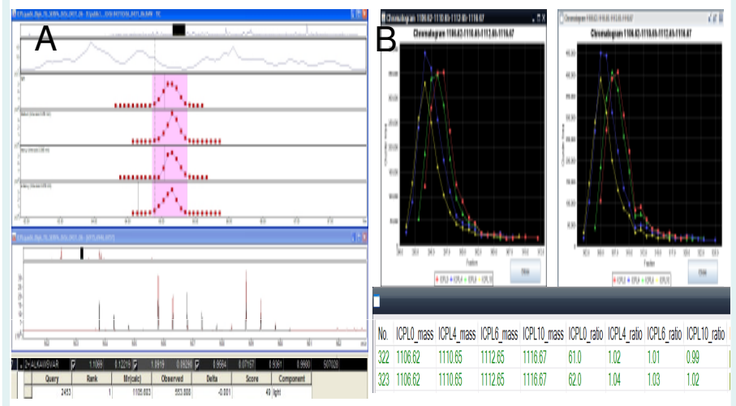


Fig 3: *The quantitation summary for the peptide AL**K**AWSVAR for both the MALDI and ESI workflow. A: Distiller with the quantitation toolbox (Mascot) for the LC-ESI workflow and B: ICPLQuant for LC-MALDI workflow.*

Exp.	#	AccNo	ICPL⁴/ICPL⁰ ratio	ICPL⁴/ICPL⁰ stabw	ICPL⁶/ICPL⁰ ratio	ICPL⁶/ICPL⁰ stabw	ICPL¹⁰/ICPL⁰ ratio	ICPL¹⁰/ICPL⁰ stabw
ESI 1	1	CYC_HORSE	1,06	*1,06*	1,06	*1,13*	0,97	*1,01*
ESI 2	1	CYC_HORSE	1,09	*1,58*	1,01	*1,21*	1,01	*1,64*
MALDI 1	1	CYC_HORSE *	0,82		0,89		0,72	
MALDI 2	1	CYC_HORSE*	0,88		0,94		0,78	
ESI 1	2	MYG_HORSE	1,08	*1,19*	0,99	*1,13*	0,94	*1,27*
ESI 2	2	MYG_HORSE	1,19	*1,16*	1,05	*1,11*	0,9	*1,16*
MALDI 1	2	MYG_HORSE *	1,11		0,45		1	
ESI 1	3	LACB_BOVIN	1,08	*1,04*	1,08	*1,05*	1,07	*1,36*
ESI 2	3	LACB_BOVIN	1,09	*1,2*	1,02	*1,09*	0,96	*1,21*
MALDI 1	3	LACB_BOVIN	0,94	*0,11*	1,08	*0,1*	0,97	*0,06*
MALDI 2	3	LACB_BOVIN	0,96	*0,09*	1,21	*0,13*	0,96	*0,1*
ESI 1	4	DHGA_BACME	1,06	*1,1*	1,04	*1,18*	0,89	*1,12*
ESI 2	4	DHGA_BACME	1,07	*1,14*	0,97	*1,19*	0,95	*1,13*
MALDI 1	4	DHGA_BACME	0,99	*0,09*	0,95	*0,06*	1,03	*0,1*
MALDI 2	4	DHGA_BACME	1	*0,09*	0,94	*0,09*	1,05	*0,16*
ESI 1	5	LIP_BURGL	1,1	*1,25*	1,01	*1,06*	1,04	*1,13*
ESI 2	5	LIP_BURGL *	1,18		1,08		1,08	
MALDI 1	5	LIP_BURGL **	0	*0*	0	*0*	0	*0*
MALDI 2	5	LIP_BURGL **	0	*0*	0	*0*	0	*0*
ESI 1	6	CATA_BOVIN	0,87	*1,16*	0,97	*1,17*	0,79	*1,2*
ESI 2	6	CATA_BOVIN	1,07	*1,15*	1,01	*1,17*	0,96	*1,19*
MALDI 1	6	CATA_BOVIN	1,3	*0,21*	1,05	*0,18*	1,21	*0,27*
MALDI 2	6	CATA_BOVIN	1,18	*0,19*	1,03	*0,18*	1,06	*0,16*
ESI 1	7	ALBU_BOVIN	1,02	*1,12*	1,04	*1,06*	0,96	*1,14*
ESI 2	7	ALBU_BOVIN	1,07	*1,12*	1,05	*1,08*	0,96	*1,12*
MALDI 1	7	ALBU_BOVIN	0,9	*0,15*	0,99	*0,13*	1,02	*0,12*
MALDI 2	7	ALBU_BOVIN	0,94	*0,11*	0,99	*0,08*	0,92	*0,1*
ESI 1	8	GOX_ASPNG	1,48	*1,18*	0,71	*1,22*	1,01	*1,16*
ESI 2	8	GOX_ASPNG	1,42	*1,21*	0,71	*1,27*	0,95	*1,17*
MALDI 1	8	GOX_ASPNG	1,48	*0,39*	0,78	*0,1*	1,04	*0,21*
MALDI 2	8	GOX_ASPNG	1,28	*0,21*	0,8	*0,17*	0,95	*0,09*
ESI 1	9	OVAL_CHICK	1,03	*1,34*	2,22	*1,12*	6,17	*1,21*
ESI 2	9	OVAL_CHICK	1,05	*1,13*	2,29	*1,09*	5,73	*1,3*
MALDI 1	9	OVAL_CHICK	0,91	*0,24*	1,99	*0,21*	4,67	*0,59*
MALDI 2	9	OVAL_CHICK	0,97	*0,17*	2	*0,22*	4,16	*0,88*
ESI 1	#	PER1A_ARMRU	0,81	*1,14*	3,58	*1,1*	0,11	*1,41*
ESI 2	#	PER1A_ARMRU	0,91	*1,21*	3,51	*1,07*	0,1	*1,14*
MALDI 1	#	PER1A_ARMRU	0,88	*0,1*	2,96	*0,99*	0,06	*0,01*
MALDI 2	#	PER1A_ARMRU	0,9	*0,08*	2,88	*0,95*	0,06	*0,01*

Tab. 3: *MALDI and ESI analysis of the combined digest with Trypsin and GluC. Protein quantification was done with ICPLQuant (MALDI) or Mascot Distiller (ESI). Each experiment was repeated twice. (*one peptide ratio **no ICPL quadruplet detected)*

Conclusion

- The ICPL™-technology allows to compare up to 4 different proteomic states in one experiment.
- Excellent quantitative data from both, MALDI (e.g. 4700 Proteomics analyzer) and ESI (e.g. Orbitrap) could be achieved. Both ionisation techniques show highly accurate and reproducible results.
- Experiments using the ICPL-quadruplex technology can readily be analyzed using *ICPLQuant* (MALDI-data) or Mascot Distiller (ESI-data).
- To increase sequence coverage and the number of correct identified proteins a combination of several enzymes has to be used. (Fig. 2).

BIOpHORETICS™ ADVANCED TOOLS FOR BIOEXPLORATION

The University Of Sheffield.

SERVA Electrophoresis