



Protein Identification

By Protein Digestion and UV-LC-MS

Order 12345

John Doe

ABCDE Pharma Inc.

Analysis start date: January 13, 2020

Analysis reporting date: January 20, 2020

Principal Investigator: Alphalyse



Table of Contents

Samples received 3
Objective..... 3
Results..... 5
MS analysis 1: Sample 1 5
MS analysis 1: Sample 2 7

Appendices

Rawdata.zip

SAMPLE



Samples received

The following samples were received at Alphalyse for protein analysis.

Sample 1
Sample 2

Objective

Identification of proteins in solution. The liquid protein sample is digested with trypsin and the resulting peptides analyzed by nanoLC-MS/MS. The MS data are searched against the SwissProt database containing all reviewed public amino acid sequences to identify the protein. The analysis provides the name of the identified protein in the database, the database accession number, the theoretical amino acid sequence and a list of observed peptides.

Introduction

Unknown proteins in liquid samples can be identified by mass spectrometric peptide analysis and database searching. The proteins are cleaved into peptides by specific proteolytic cleavage and the peptide masses and peptide fragment masses determined at high sensitivity by nanoflow HPLC coupled to electrospray mass spectrometry (nanoLC-MS/MS). The database search can find the matching protein if the amino acid sequence is recorded in the database. The protein sequence coverage map shows the observed peptides in the database protein sequence.



Sample preparation

Sample name	Sample 1	Sample 2
Volume sample used for digest	20	10

The samples were after digest resuspended in 50µl 0.1% formic acid and of this only 1µl was injected onto the column

Experimental

DIGESTION

The protein samples were reduced and alkylated with iodoacetamide, i.e. carbamidomethylated, and subsequently digested with trypsin that cleaves after lysine and arginine residues. The resulting peptides were concentrated by Speed Vac lyophilization.

nanoLC-MASS SPECTROMETRY

The peptides were dissolved and injected on a Dionex nano-LC system for MS/MS analysis on a Bruker Maxis Impact QTOF mass spectrometer.

DATABASE SEARCH

The public protein databases are downloaded from SwissProt. The Mascot software finds matching proteins in the database by their peptide masses and peptide fragment masses. The database search considers fixed modification of cysteines by carbamidomethylation, variable oxidation of methionine residues that may be a result of the sample preparation procedure, as well as N-terminal methionine truncation. The database search does not consider other known protein features recorded in the database, such as signal sequences, sequence variants and other post translational modifications. The protein identification is based on a probability-scoring algorithm and the significant best matching protein is shown in the Results. Homologous proteins with a lower score are not included in the report. If the protein from the correct organism is not present in the database, then a significant matching homologous protein from another organism is reported. If several different proteins are identified with a significant score, then several protein identifications are reported for the sample.

The identified database protein sequences are shown together with the obtained sequence coverage maps. The peptides used for the identification are highlighted in the sequence and the matching peptides are listed for comparison of the determined and calculated peptide masses. The same peptide mass can appear in multiple identifications. It is considered a positive identification when at least 2 matching peptides have an Ions Score above 20, or if a protein under 20 kDa has 1 peptide with an Ions Score above 50. The sequence coverage % is not considered for the identification. The total Mascot score provided for each identification is a total sum of all the individual peptide scores.

PROTEIN STRUCTURE AND FUNCTION

The identified protein can be looked up at the UniProt and NCBI websites (National Center for Biotechnology Information) by selecting the UniProt/NCBI Entry hyperlink. UniProt and NCBI contain collected information about the protein function and structure, and links to bioinformatics tools such as Conserved functional domains and Blast homology searches.

SUMMARY TABLE OF IDENTIFIED PROTEINS

The table shows the customer Sample name, the name of the protein identified in the database, the database entry name (accession number), the calculated molecular weight (MW) of the database sequence including eventual signal sequences and fixed modifications used in the database search, the identification Score calculated by the Mascot database search software, and the Sequence Coverage % calculated as the number of amino acids in observed peptides/total number of amino acids in database sequence.



Results

The Mascot protein identifications along with the peptide identifications and search settings are documented in the .xlsx file associated with each sample

MS analysis 1: Sample 1

Id#	Id	Protein name	Coverage	MW	Score
1.1	PRZN_SERMA	Serralysin OS=Serratia marcescens OX=615 PE=1 SV=1	49.7 %	52073	1514
1.2	FLIC_SERMA	Flagellin OS=Serratia marcescens OX=615 GN=fliC PE=3 SV=1	46.7 %	36844	1480
1.3	OMPA_SERMA	Outer membrane protein A OS=Serratia marcescens OX=615 GN=ompA PE=3 SV=1	18.1 %	38516	292
1.4	LPP_ECOLI	Major outer membrane prolipoprotein Lpp OS=Escherichia coli (strain K12) OX=83333 GN=lpp PE=1 SV=1	33.3 %	8375	108
1.5	RL28_ERWT9	50S ribosomal protein L28 OS=Erwinia tasmaniensis (strain DSM 17950 / CIP 109463 / Et1/99) OX=465817 GN=rpmB PE=3 SV=1	12.8 %	9106	53

See Sample1_IDs.xlsx for details.

Protein details

1.1 Protein identification information

Serralysin OS=Serratia marcescens OX=615 PE=1 SV=1

Protein ID: PRZN_SERMA	MW: 52073	pI: 4.6
Mascot score: 1514	Seq. Coverage: 49.7 %	Unique peptides: 17
Uniprot link	Conserved Domain link	Blast link



Peptide map:

1 MQSTKKAIEI TESSLAAATT GYDAVDDLLH YHERGNGIQI **NGKDSFSNEQ** **AGLFITRENQ**
 61 TWNGYKVFQ PVK**LTFSFPD** **YKFSSTNVAG** **DTGLSKFSAE** **QQQAKLSLQ** **SWADVANITF**
 121 **TEVAAGQKAN** ITFGNYSQDR PGHYDYGTQA YAFLPNTIWO QDLDGGQWY NVNQSNVHKP
 181 ATEDYGR**QTF** **THEIGHALGL** **SHPGDYNAGE** **GNPTYNDVTY** **AEDTRQFSLM** SYWSETNTGG
 241 DNGGHYAAAP LLDDIAAIQH LYGANPSTR**T** **GDTVYGFNSN** **TGRDFLSTTS** **NSQKVIFAAW**
 301 **DAGGNDTFDF** **SGYTANQRIN** LNEK**SFSDVG** **GLKGNVSI**AA GVTIENAIGG SGNDVIVGNA
 361 ANNVLK**GGAG** **NDVLFGGGGA** **DELWGGAGKD** **IFVFSAA**SDS **APGASDWIR**D FQKGIDKIDL
 421 **SFFNK**EANSS DFIHFVDHFS GTAGEALLSY NASSNVTDLS VNIGGHQAPD FLVK**IVGQVD**
 481 **VATDFIV**

1.2 Protein identification information

Flagellin OS=Serratia marcescens OX=615 GN=fliC PE=3 SV=1

Protein ID: FLIC_SERMA	MW: 36844	pI: 4.85
Mascot score: 1480	Seq. Coverage: 46.7 %	Unique peptides: 12
Uniprot link	Conserved Domain link	Blast link

Peptide map:

1 **MAQVINTNSL** **SLMAQNNLNK** SQSSLGTAIE RLSSGLRINS **AKDDAAGQAI** **SNRFTANIKG**
 61 LTQASR**NAND** **GISLAQTTEG** **ALNEVNDNLQ** **NIRRLTVQAQ** NGSNSTSDLK **SIQDEITQRL**
 121 SEINR**ISEQT** **DFNGVKVLSS** DQKLTIQVGA NDGETTDIDL KKIDAKQLGM DTFDVTTKSA
 181 KAGAEIATGT KITVDSDATK QADADVTGLA KGQTLVSGTD ADGKSAYFIA TKDDATGDVA
 241 YTKAKVADDG KVTDSGTDAG VKNPLATLDK **ALAQVDGLRS** **SLGAVQNRFD** **SVINNLNSTV**
 301 **NNLSASQSRI** **QDADYATEVS** **NMSRANILQQ** **AGTSVLAQAN** **QSTQNVLSLL** R

1.3 Protein identification information

Outer membrane protein A OS=Serratia marcescens OX=615 GN=ompA PE=3 SV=1

Protein ID: OMPA_SERMA	MW: 38516	pI: 8.27
Mascot score: 292	Seq. Coverage: 18.1 %	Unique peptides: 5
Uniprot link	Conserved Domain link	Blast link

Peptide map:

1 MKKTAIALAV ALAGFATVAQ AAPKDNTWYT GAKLGWSQYH DTGFYGNQYQ NGIGNGP**THK**
 61 DQLGAGAF**LG** YQANQYLGFE LGYDWLGRMP YKGSVN**NGAF** KAQGVQ**LA**AK LSYPIAD**DL**
 121 IYTRLGGMVW RADSKANYGR TGQRLSDHDT GVSPLAAVGV EYALTKNWAT RLDYQFVSNI
 181 GDAGTVGARP DNTMLSLGVS YRFGQDDVVA PAPAPAPAPV VETKRFTLKS **DVLFNFNKST**
 241 **LKAEGQQALD** **QLYTQLSSMD** **PKDGSVVVLG** **YTDVAVGSDQY** **NQKLSEQRAQ** **SVVDYLVSKG**
 301 IP**SDKIS**ARG MGEADAVTGN TCGYKSGRAT KAQIVCLAPD RRVEIEVKGI KDVVTQ**PQG**

Protein Identification

Order 12345



1.4 Protein identification information

**Major outer membrane prolipoprotein Lpp OS=Escherichia coli (strain K12)
OX=83333 GN=lpp PE=1 SV=1**

Protein ID: LPP_ECOLI	MW: 8375	pI: 9.3
Mascot score: 108	Seq. Coverage: 33.3 %	Unique peptides: 2
Uniprot link	Conserved Domain link	Blast link

Peptide map:

1 MKATKLVLGA VILGSTLLAG CSSNAK**IDQL SSDVQTLNAK** VDQLSNDVNA MRSDVQAAKD
61 **DAAR**ANQRLD NMATKYRK

1.5 Protein identification information

**50S ribosomal protein L28 OS=Erwinia tasmaniensis (strain DSM 17950 / CIP
109463 / Et1/99) OX=465817 GN=rpmB PE=3 SV=1**

Protein ID: RL28_ERWT9	MW: 9106	pI: 11.42
Mascot score: 53	Seq. Coverage: 12.8 %	Unique peptides: 1
Uniprot link	Conserved Domain link	Blast link

Peptide map:

1 MSRVCQVTGK RPVTGNNRSH AMNATKRRFL PNLHSHRFWV ESEKRFVTLR VSAKGMRVID
61 KK**GIETVLAD LR**TRGEKY

MS analysis 2: Sample 2

Id#	Id	Protein name	Coverage	MW	Score
2.1	A0A287B5W2_PIG	Trypsinogen isoform X1 OS=Sus scrofa OX=9823 GN=LOC100302368 PE=3 SV=1	 32.5 %	26563	339
2.2	A0A024A2R8_VIBCL	Cholera enterotoxin subunit B OS=Vibrio cholerae H1 OX=1093790 GN=ctx PE=4 SV=1	 47.6 %	14001	219
2.3	A0A023YXN8_ECOLX	Major outer membrane lipoprotein OS=Escherichia coli O145:H28 str. RM12581 OX=1248823 GN=lpp PE=4 SV=1	 48.7 %	8375	150
2.4	A0A2E8K870_9GAMM	BolA family transcriptional regulator OS=Gammaproteobacteria	 7.5 %	11800	64

Protein Identification

Order 12345



bacterium OX=1913989
GN=CMQ20_10180 PE=3
SV=1

See Sample2_IDs.xlsx for details.

Protein details

2.1 Protein identification information

Trypsinogen isoform X1 OS=Sus scrofa OX=9823 GN=LOC100302368 PE=3 SV=1

Protein ID: A0A287B5W2_PIG	MW: 26563	pI: 6.85
Mascot score: 339	Seq. Coverage: 32.5 %	Unique peptides: 6
Uniprot link	Conserved Domain link	Blast link

Peptide map:

1 MNTFVLLALL GAAVAFPTDD DDKIVGGYTC AANSIPYQVS LNSGSHFCGG SLINSQWVVS
61 AAHCYKSRIQ VRLGEHNIDV LEGNEQFINA AKIITHPNFN GNTLDNDIML IKLSSPATLN
121 SRVATVSLPR SCAAAGTECL ISGWGNTKSS GSSYPSLLQC LKAPVLSOSS CKSSYPGQIT
181 GNMICVGFLE GGDSCQGDS GGPVVCNGQL QGIVSWGYG C A QKNKPGVYT KVCNYVNWIQ
241 QTIAAN

2.2 Protein identification information

Cholera enterotoxin subunit B OS=Vibrio cholerae H1 OX=1093790 GN=ctx PE=4 SV=1

Protein ID: A0A024A2R8_VIBCL	MW: 14001	pI: 8.91
Mascot score: 219	Seq. Coverage: 47.6 %	Unique peptides: 6
Uniprot link	Conserved Domain link	Blast link

Peptide map:

1 MIKLFKGVFF TVLLSSAYAN GTPQNITDLC AEYHNTQIHT LNDKIFSYTE SLAGKREMAI
61 ITFKNGATFQ VEVPGSQHID SQKKAIERMK DTLRIAYLTE AKVEKLCVWN NKTPHAI AAI
121 SMAN

2.3 Protein identification information

Major outer membrane lipoprotein OS=Escherichia coli O145:H28 str. RM12581 OX=1248823 GN=Ipp PE=4 SV=1

Protein ID: A0A023YXN8_ECOLX	MW: 8375	pI: 9.3
Mascot score: 150	Seq. Coverage: 48.7 %	Unique peptides: 3

Protein Identification

Order 12345



Uniprot link	Conserved Domain link	Blast link
-------------------------------------	--	-----------------------------------

Peptide map:

1 MKATKLVLGA VILGSTLLAG CSSNAK**IDQL SSDVQTLNAK VDQLSNDVNA MRSDVQAAKD**
61 **DAAR**ANQRLD NMATKYRK

1.4 Protein identification information

**BolA family transcriptional regulator OS=Gammaproteobacteria bacterium
OX=1913989 GN=CMQ20_10180 PE=3 SV=1**

Protein ID: A0A2E8K870_9GAMM	MW: 11800	pI: 6.36
Mascot score: 64	Seq. Coverage: 7.5 %	Unique peptides: 1
Uniprot link	Conserved Domain link	Blast link

Peptide map:

1 MSVKDVIESK **LATVLS**PRHL QVVNESGNHN VPDGSESHFK VVLVSDDFEG ERLLRHHRV
61 NRVLAELEN QIHALALHTY TASEWQSKHG DAPMSPPCKG GSAKTQ

SAMPLE