Proteomics Workflows available at our Campus

Jayashree R.[1] and Dominik Schwudke[2]
[1]C-CAMP; [2]NCBS

Overview of the MS-facility

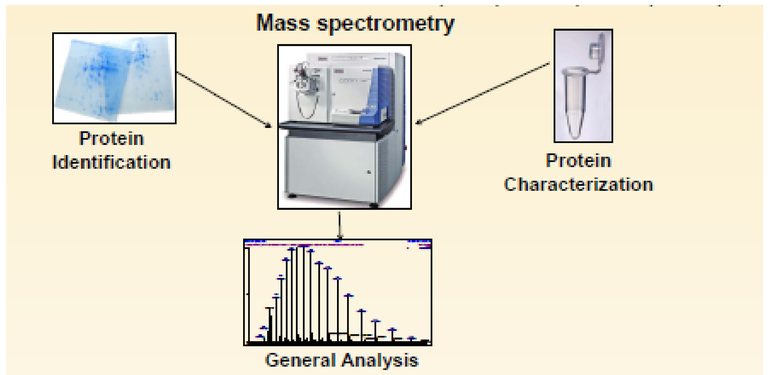


Fig.1 : Instrumentation and overview of services at the MS facility.

The Mass Spectrometry Facility provides scientific and technical support for the research groups at NCBS and associated institutes. The facility comprises of Q-TOF-MS, LTQ Orbitrap Discovery-MS(Thermo), Agilent 1200 Nano-LC and Advion Nanomate system.

Currently, the following services are offered from this facility:
1. Molecular mass measurement of intact proteins and peptides
2. Identification of 1D and 2D gel separated proteins
3. In-solution digestion and characterization of proteins

The facility uses proteomics search engines such as Mascot and Sequest and data analysis/ validation software such as Proteome Discoverer(PD) and Proteome Scaffold(PS).

Facility usage and Service Distribution for the year 2012:

Total of 316 samples from 80 projects and 40 users have been analyzed in the year of 2012. This includes 291 proteomic and 25 general analyses, As against 207 samples from 60 projects of 37 users analyzed in the year of 2011.

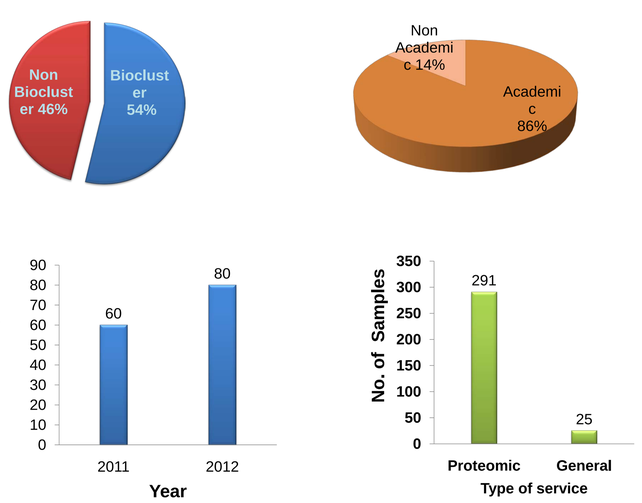
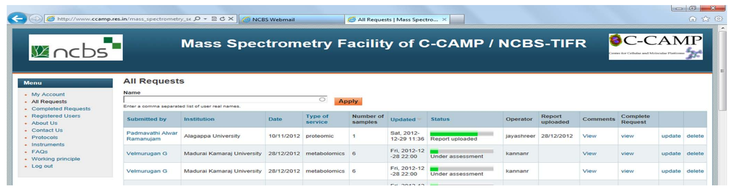


Fig.2 : Usage statistics of the MS-facility for the year of 2012.

Introduction to the Web portal:

http://www.ccamp.res.in/mass_spectrometry_services

The web portal serves as a common platform for both the users and service providers to raise requests, share information and upload /download reports, results etc.



Working principle of C-CAMP/NCBS Mass spectrometry facility

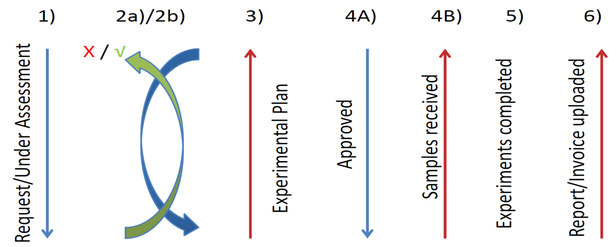


Fig.3 : Screen shot of the Web-portal and Working principle.

Nutshell of services provided:

Protein characterization:
Purified proteins/peptides free from salts and detergents are analyzed on the Q-TOF system. Good quality data are generated with 250-500fmol/µl of sample concentration. The interpreted data along with the report is delivered to the end user

Protein Identification and Quantification :
➢Routine Protein identification: 1D/2D gels/protein samples are subjected to in-gel/in-solution digestion using Proteolytic enzymes such as trypsin. The digested peptides are subjected to Nano LC-MS/MS analysis. Data generated after this data dependent MS/MS is analyzed by database matching using PD software integrated with MASCOT and / or Sequest search engines.
➢Data dependent neutral loss (DDNLMS^3) method for analysis of Phosphorylation PTM in proteins.
➢Homology –driven protein identification, and
➢Label free quantification is performed on the request of the user.

Representative examples:

Routine Protein identification:

In gel digestion and Nano LC MS/MS analysis of Co-IP experiment: for Sarita:
•Proteins were from Drosophila melanogaster. Co-IP of spin using myc-GFP.
•Western blotting showed that the GFP (control) and spin-fusion-GFP migrated at the expected sizes when compared to the protein ladder. GFP (control) is around 27kDa where as the fusion spin-myc-GFP is expected at 100kDa
•Protein was run over a SDS-PAGE and stained with Coomassie.
•3 sets of samples were provided , the gel lanes were sliced into 11 slices each, subjected to in-gel digestion.
•Digested peptides were reconstituted, subjected to 70 minute RPLC-MS/MS analysis.
•Generated data was searched for the identity on MASCOT as search engine using Drome_Uniref90 database.
•Multidimensional Protein Identification Technology (MudPIT) analysis was done to compare the gel lanes.
•High confident peptides with prerequisite of minimum two peptides leading into identification of proteins were selected and list was generated
Results: List of all proteins with peak area based quantification values was provided ,
Protein spinster OS=Drosophila melanogaster GN=spin PE=1 SV=1 - [SPIN_DROME] with accession number Q9GQQ0 was found in the spin-myc-GFP lane.

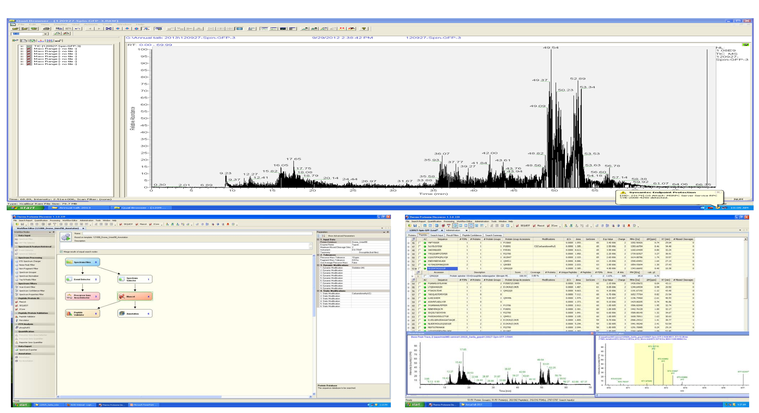


Fig 4: A typical workflow used for the identification, quantification and validation of proteins/peptides.ProteomeDiscoverer 1.3 (PD) is used for peak area based quantification.

Homology–driven protein identification

Proteomics workflow carried out for the sample –Routine workflow

Results from the Error tolerant search -No proteins from Dioscorea Alata were identified. A few proteins from the Plantae family were identified

Suggested way forward:
Homology based proteomics- This analysis will end up providing a list of proteins which are originating from phylogenetically nearest Sequest organism. Reference : Protein Identification Pipeline for the Homology Driven Proteomics, Magno Junqueira et al. Journal of proteomics, 2008. (Pubmed ID: 18639657).

Homology driven proteomics workflow carried out for the sample

Links :
Eagle eye: http://genetics.bwh.harvard.edu/cgi-bin/msfilter/eagleeye.cgi
PepNovo: http://proteomics.ucsd.edu/Software/PepNovo.html
MS Blast: http://genetics.bwh.harvard.edu/msblast/

•The .MGF file generated earlier was subjected to Eagle Eye filtering.
Eagle eye resulted in a zip file with two .MGF files, one with the good spectra and another one with background spectra produced from the submitted raw .MGF file, as well as two tables with filtering report and with filtering settings used.

•The .MGF file with good spectra was subjected to de novo sequencing by PepNovo.

•The PepNovo output was pasted directly into MS BLAST query window and searched against nr database using the *LC-MS/MS Presets option of the MS BLAST web server.*

Accession	Description	Mascot search	MS Blast search	Mass (Da)	Source	Remarks
TrEMBL - B0YMX4/E2ZP01; MS Blast - ABU39826.1;TrEMBL from error tolerant - Q0ZHI0	DinG OS=Pseudomonas aeruginosa/Putative uncharacterized protein OS=Pseudomonas aeruginosa/Putative uncharacterized protein OS=Pseudomonas aeruginosa	✓	✓	40663	Bacteria	
YP_792606.1	hypothetical protein PA14_55410 [Pseudomonas aeruginosa UCBPP-PA14] >gb|ABJ09831.1| hypothetical protein PA14_55410 [Pseudomonas aeruginosa UCBPP-PA14] >gb|EKA41171.1| periplasmic binding protein [Pseudomonas aeruginosa C27]	x	✓	40714.3	Bacteria	Sequence Homologous to DinG of Pseudomonas aeruginosa
YP_001350163.1	periplasmic binding protein [Pseudomonas aeruginosa PA7] >gb|ABR84929.1| periplasmic binding protein [Pseudomonas aeruginosa PA7]	x	✓	40603.2	Bacteria	Sequence Homologous to DinG of Pseudomonas aeruginosa
ZP_10651436.1	ABC-type phosphate transport system, periplasmic component [Pseudomonas sp. GM50] >gb|EJM65282.1| ABC-type phosphate transport system, periplasmic component [Pseudomonas sp. GM50]	x	✓	40076.9	Bacteria	Sequence Homologous to DinG of Pseudomonas aeruginosa
YP_259866.1	alkaline phosphatase [Pseudomonas protegens Pf-5] >gb|AAY92032.1| alkaline phosphatase L [Pseudomonas protegens Pf-5]	x	✓	40493.3	Bacteria	Sequence Homologous to DinG of Pseudomonas aeruginosa
AAW57408.2	DING protein [Hypericum perforatum]	x	✓	36897.08	Plant	Sequence Homologous to DinG of Pseudomonas aeruginosa
NP_001234080	enolase [Solanum lycopersicum]	x	✓	47798.19	Plant	Enolase
ACN50180	enolase [Annona cherimola]	x	✓	47803.21	Plant	Enolase
NP_177543	enolase 1 [Arabidopsis thaliana]	x	✓	51474.34	Plant	Enolase
XP_002877895	enolase [Arabidopsis lyrata subsp. lyrata]	x	✓	47786.13	Plant	Enolase

DinG protein was identified independently in 3 searches – Mascot TrEMBL search, Error tolerant search with TrEMBL database and MS Blast homology based search with nr database.
The DinG protein has a mass of ~ 40KDa.The protein of interest is of 40KDa (from sample user information).Thus this suggests that the protein of interest has a sequence homologous to the DinG protein.
Enolase was also identified in the MS Blast search. This has a mass of about 50KDa.Thus this suggests that Enolase is a second possibility .

Data dependent neutral loss LC-MS^n (DDNLMS^3) method for phosphorylation analysis (Design of a quality control sample)

- Standard phosphopeptide mix(SIGMA) was spiked at 5 different concentrations into Standard BSA digest:
  1. Standard phosphopeptide mix amounts used were at 1,10,25,100 and 250 fmol per peptide (on column).
  2. Standard BSA digest was used at 250fmol (on column).
- LC-MS/MS analysis: The samples were subjected to 70 min standard gradient, using a customized DDNLMS^3 acquisition method for data acquisition on the LTQ-Orbitrap Discovery.
- Data analysis was performed using PD 1.3 applying the workflow template *111229_Sequest_Phospho_dehydration* with a customized phosphopeptide BSA fasta database.

Results:

- Out of the 10 peptides, 5 peptides were identified at 1fmol, 6 peptides were identified at 10fmol concentration, and 7 peptides were identified at 25,100 and 250 fmol concentrations.
- Area linearity was observed with phosphopeptides PP10, PP9 and PP3

S.No	Phosphopeptides	Abbreviation	MW monoisotopic	Relative Signal intensity***	# of Phosphates per amino acid			Sample 1 fmol PP spiked	Sample 10 fmol PP spiked	Sample 25 fmol PP spiked	Sample 100 fmol PP spiked	Sample 250 fmol PP spiked	PD Retention time(min)	Qual browser Retention time(min)
					S	T	Y							
1.1	VLHSGpSR	PP1	835.3822	Weak	1									
1.2	RSpYpSRSR	PP2	1071.4132	Weak	1		1							
1.3	RDSLGpTYSSR	PP3	1221.526	Medium		1		✓	✓	✓	✓	✓	10.8	10.9
1.4	pTKLpTQLRDAK	PP4	1446.7117	Strong		2		✓	✓	✓	✓	✓		20.96
1.5	EVQAEQPSSpSSPR	PP5	1481.6268	Medium	1			✓	✓	✓	✓	✓	9.83	9.96
1.6	ADEPpSSEESDLEIDK	PP6	1743.6844	Strong	1			✓	✓	✓	✓	✓	16.73	16.84
1.7	ADEPSpSEEpSDLEIDK	PP7	1823.6508	Medium	2									
1.8	FEDEGAGFEESpSETGDYEEK	PP8	2334.8446	Strong	1					✓	✓	✓	22.43	23
1.9	ELSNpSPLRENSFGpSPLEFR	PP9	2339.0105	Medium	2				✓	✓	✓	✓	37.67	37.52
1.10	SPTEYHEPVpYANPFYRPTpTPQR	PP10	2810.2011	Strong		1	1	✓	✓	✓	✓	✓	31.49	31.67

Standard phosphopeptide mix(SIGMA) at 25 fmol/peptide spiked in 250fmol BSA digest can be used as a standard to check the method performance for samples with Phosphorylation PTM analysis, since this is the lowest concentration at which 7 peptides are identified.