

Some Recent Enhancements to Protein Prospector

Peter R. Baker¹, Robert J. Chalkley¹,
Jonathan Trinidad¹,
and Alma L. Burlingame¹

¹ Mass Spectrometry Facility, Dept. of Pharmaceutical Chemistry, University of
California, San Francisco, USA

Multi-Sample Data

The ABI 4700 TOFTOF instrument allows each spot on a sample plate to contain a single digest sample with both MS and MSMS data available. As the MSMS data isn't acquired for every peak the MS data can be used to identify post-translation modifications, peptides formed by non-specific enzymatic cleavages or to distinguish proteins with similar sequences.

The MS data can either be searched against the proteins identified by MSMS or the whole database.

The MSMS hits can be used to calibrate the MS scan.

Display of Multi-Sample Results

The Batch-Tag and Search Compare programs can produce separate protein and peptide reports for each spot on a sample plate. Both the MS and MSMS matches for the proteins identified can be displayed.

1 MPSQEVSVNR **VIVHPLVLLS** VVDHNRMGK IGNQKRVGV LIGCVRSKGV LDVSNFAVP FDEDDKRSV WFLDHDYLEN
 81 NYGHFQVNA REPVGWHYH GKLNQNDIA ENLIVRYCP NSVLVIDAK PDLGLPTA YISVEVHDD GSPTSSTEN
 161 **VPSIGAREE** REVGVEHLR DEDITVCSL SQTITNQLG LKLNAGLSD IKQYLQVCD SKPINSQIV YLQDIFNLL
 241 PDITNDQFTG THYVKTNDQ LVVYLASHVR SIALNHLIN NKLANDARE GSDSRAKE KMDSDSDDN KETDQDCKK
 321 ABEKADKGD EGGKSRK

Protein Score 1	m/z	Error (ppm)	Peptide	F	S	R	#	Score	Score Diff	Disc Score
342.2	2677.4878	66	TFEHVPSIGAREEAEVGVVEHLR	1	23	7	3	109.7	73.5	4.67
	1534.8872	35	LHONDIANELVR	1	23	3	2	120.2	71.6	4.61
	1957.2274	45	VIVHPLVLLSVVDHFNRR	1	23	3	4	66.8	31.5	3.41
	1033.5969	24	ITNQLMGLK	1	23	6	4	45.5	7.3	2.68

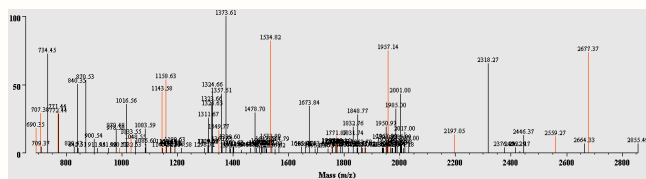
MSMS Coverage

1 MPSQEVSVNR **VIVHPLVLLS** VVDHNRMGK IGNQKRVGV LIGCVRSKGV LDVSNFAVP FDEDDKRSV WFLDHDYLEN
 81 NYGHFQVNA REPVGWHYH GKLNQNDIA ENLIVRYCP NSVLVIDAK PDLGLPTA YISVEVHDD GSPTSSTEN
 161 **VPSIGAREE** REVGVEHLR DEDITVCSL SQTITNQLG LKLNAGLSD IKQYLQVCD SKPINSQIV YLQDIFNLL
 241 PDITNDQFTG THYVKTNDQ LVVYLASHVR SIALNHLIN NKLANDARE GSDSRAKE KMDSDSDDN KETDQDCKK
 321 ABEKADKGD EGGKSRK

MSMS Hits

m/z	MH*	Intensity	Delta ppm	Modifications	Start	End	Missed Cleavages	Sequence
690.35	690.34	69.3	13		328	334	1	(K)IKDRDQK (G)
690.35	690.36	69.3	-9.0	pyroGha	213	217	0	(K)QYLQR(V)
690.35	690.38	69.3	-4.0		315	320	2	(K)IKGQK (A)
707.38	707.38	109.5	-10		213	217	0	(K)QYLQR (V)
771.44	771.45	119.7	-13		203	209	0	(K)ILNAQLR (D)
990.53	990.50	12.6	32		328	337	2	(K)IKDRDQKSR (K)
1033.55	1033.57	49.0	-18	1Met-ox	194	202	0	(K)ITNQLMGLK(G)
1143.38	1143.60	169.5	-13		94	103	0	(R)YGVYHTQPK (L)
1158.63	1158.65	204.4	-10		37	46	0	(R)YGVYHTQPK (S)
1176.38	1176.60	15.0	-17	pyroGha	213	222	1	(K)QYLQVGVGDSK(M)
1349.77	1349.79	64.2	-13		271	282	0	(K)IHALNHLNKK (L)
1534.82	1534.83	311.7	-7.9		104	116	0	(K)LHONDIANELVR (R)
1755.87	1755.88	24.0	-5.7	1Met-ox	256	270	0	(K)TNDQMLVYVYLAQMVR(S)
1771.87	1771.87	46.5	-2.0	2Met-ox	256	270	0	(K)TNDQMLVYVYLAQMVR(S)
1873.02	1873.04	15.5	-6.6		117	132	1	(R)YVPSVIVDQKPK (D)
1953.99	1953.91	37.1	40		49	66	0	(K)YVLSVNSFAVFFDEDDK (D)
1957.14	1957.14	285.2	2.8		11	27	0	(K)VIVHPLVLLSVVDHFNRR (M)
2197.05	2197.03	52.0	10		49	68	1	(K)YVLSVNSFAVFFDEDDK (S)
2559.27	2559.21	45.7	24		133	156	0	(K)IQLPTEAYVSEVVDQSETSK (T)
2677.37	2677.31	280.2	24		157	180	0	(K)TFEHVPSIGAREEAEVGVVEHLR (D)

MS Coverage

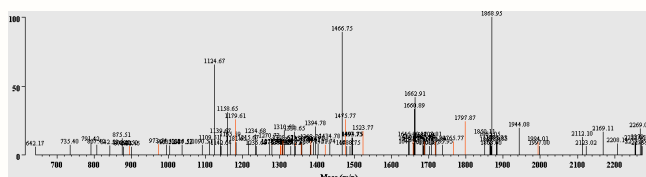


MS Spectrum

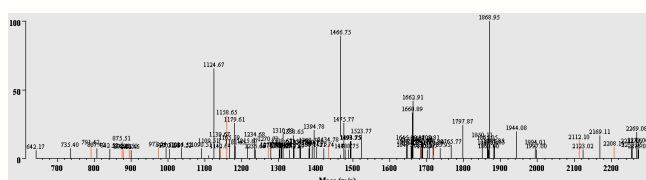
MS Hits

Better Matches Using MS-Homology

Swiss Prot Matches

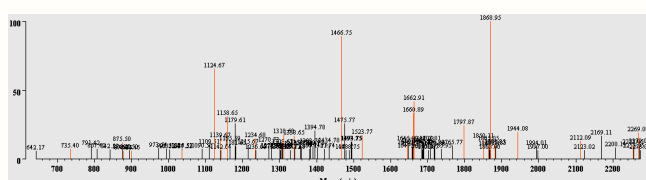


HUMAN Keratin, type II cytoskeletal 1



DICI 26S protease subunit 6B homolog

NCBI nr Match identified by MS-Homology

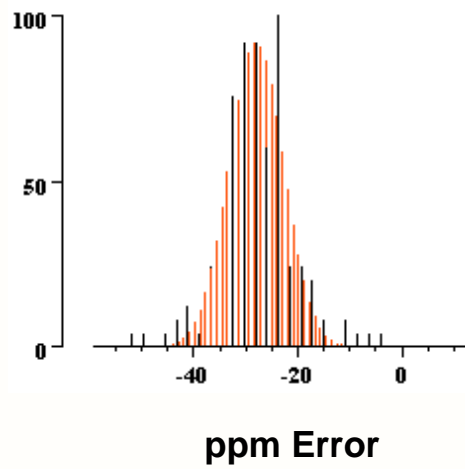


DROME 26S proteasome regulatory complex subunit p50

Identified by using the MSMS hit sequences in an MS-Homology search.

MSMS Hits: VD[IIL][IIL]DPA[IIL][IIL]R, GV[IIL][IIL]YGPPGTGK, YEE[IIL][K][Q][IIL]TAGR

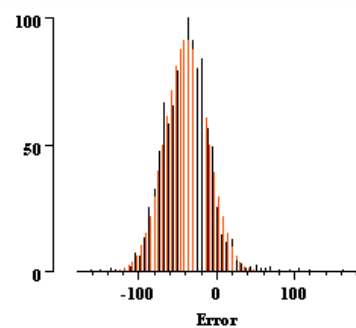
Recalibration of Data Based on Results



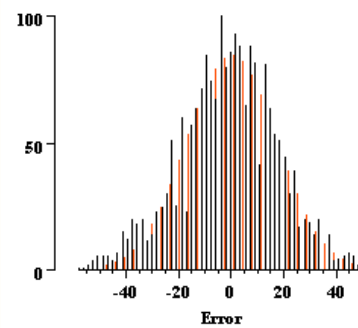
A histogram of the MSMS parent errors of the correctly identified peptides shows a normal distribution with a systematic error of -27ppm . The data can be resubmitted with an error tolerance of 35 ppm rather than the original 100 ppm .

Recalibration of Multi-Fraction Data

Each fraction in a multi-fraction data set can have a separate calibration.



Before Calibration

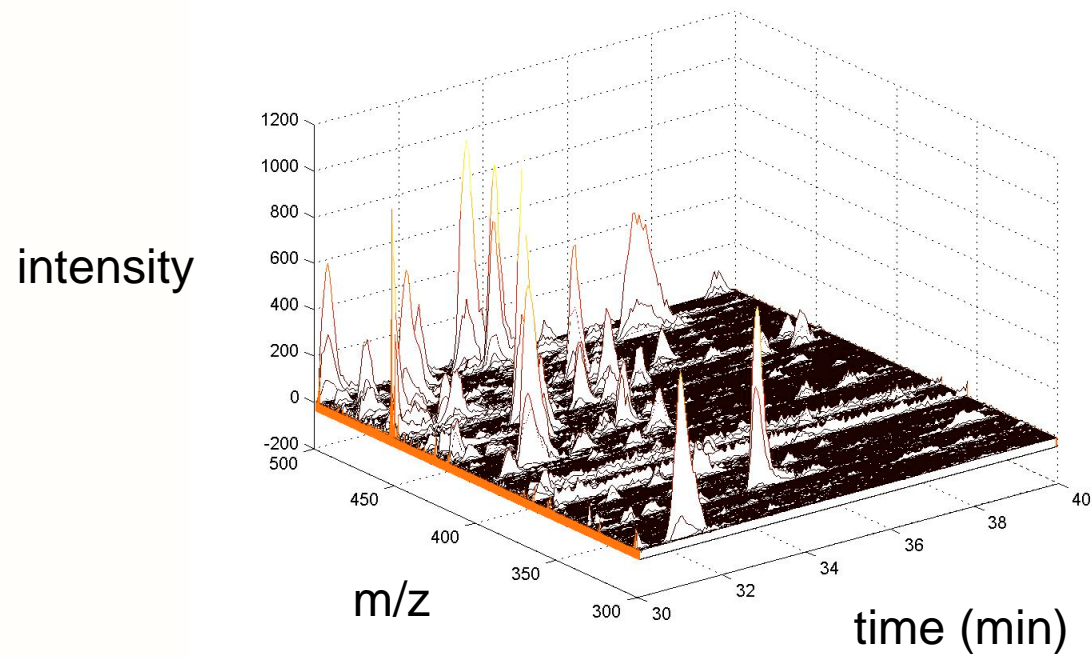


After Calibration

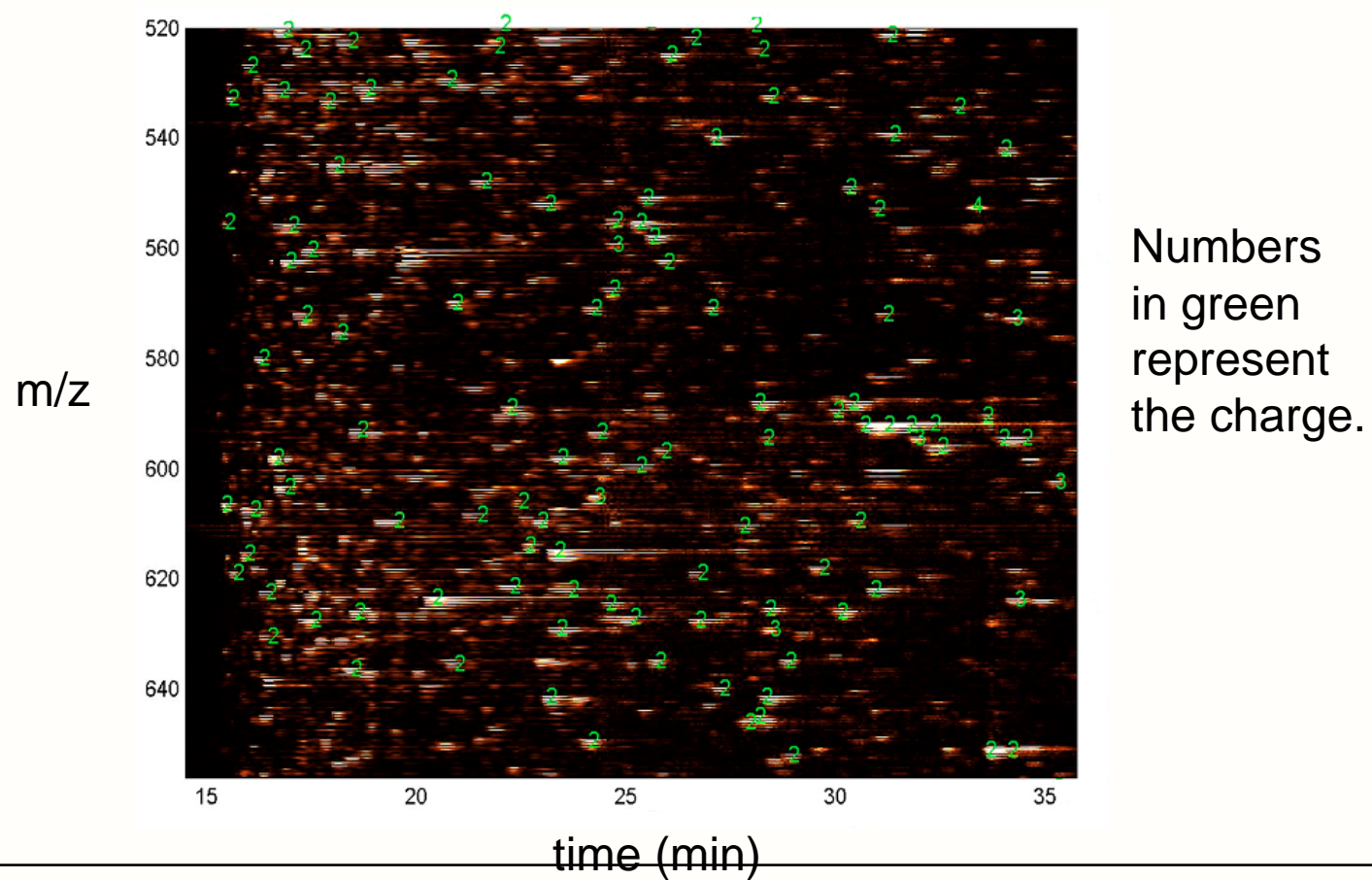
Distribution of Errors in a 6 Fraction Data Set

The calibration enabled the data to be re-searched with a tolerance of $\pm 50\text{ ppm}$ rather than the original $\pm 120\text{ ppm}$.

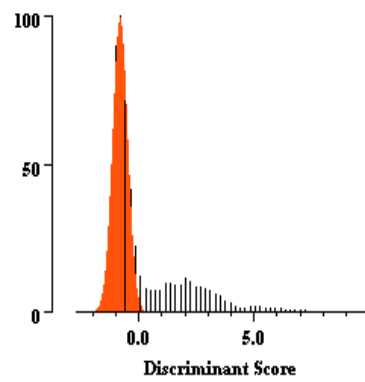
LC MSMS Waterfall Plots



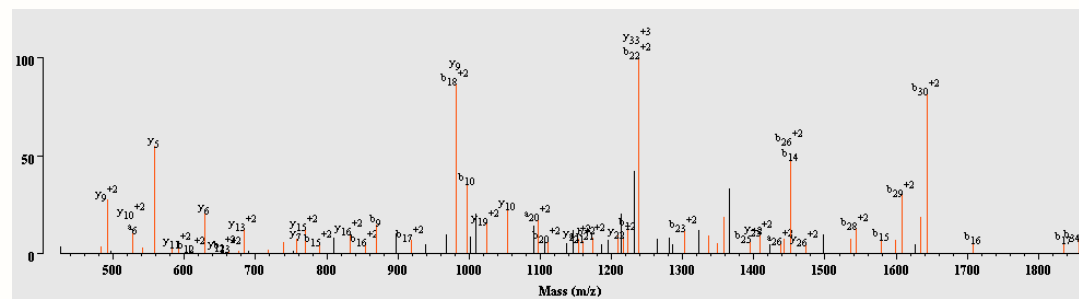
Visualization of Peaks Chosen for MSMS



Support for LTQ Data



Discriminant score graph from data set with 116075 spectra from 82 fractions.



New ion types supported:

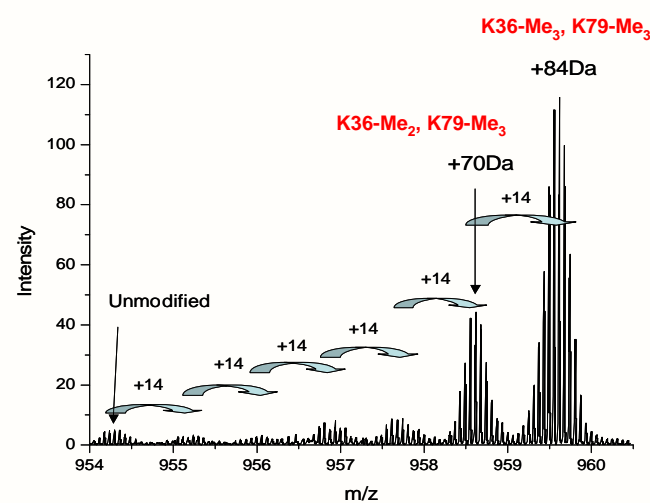
b⁺², b-NH₃⁺², b-H₂O⁺², b-SOCH₄⁺², b-H₃PO₄⁺²
 y⁺², y-NH₃⁺², y-H₂O⁺², y-SOCH₄⁺², y-H₃PO₄⁺²

MS-Tag of Protein ECD

# Unmatched Ions	Sequence	m/z Submitted	MH ⁺ Calculated (Da)	MH ⁺ Error (Da)	Species	Protein Name
11	ARTRQTARKSTGGKAPRKQLASKAARKSAPSTGGVK(TriMeK)KPHRYK PGTVALREIRRFQKSTELLIRKLPFQRLVREIAQDFK(TriMeK)IDLRFQ SSAIGALQESVEAYLVSLFEDTNLAAIHAKRVTIQKKDIKLARRLRGERS	697.7216 ⁺²²	15328.7098	0.0011	YEAST	Histone H3 K4R
14	ARTRQTARKSTGGKAPRKQLASKAARKSAPSTGGVK(DiMeK)KPHRYK PGTVALREIRRFQKSTELLIRKLPFQRLVREIAQDFK(TriMeK)IDLRFQ SSAIGALQESVEAYLVSLFEDTNLAAIHAKRVTIQKKDIKLARRLRGERS	697.7216 ⁺²²	15314.6941	14	YEAST	Histone H3 K4R

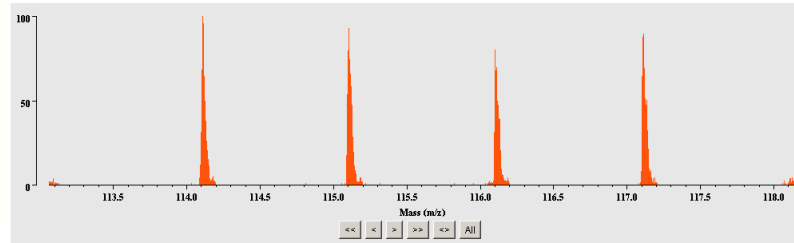
The ECD spectrum with 198 fragment ions contains two major modification states of Histone H3.

MS-Tag correctly identifies the modification sites of these two components.



iTRAQ Quantitation

Link to raw
quantitation data



Acc. #: [P16858](#) Species: MOUSE Name: Glyceraldehyde-3-phosphate dehydrogenase (GAPDH)
pI of Protein: 8.4
Protein.MW: 35810.3

```

1      11      21      31      41      51      61      71
MVKVGVNGFC RIGRLVTRAA ICSGKVEIVA INDPFDLNY HVYMFQYDST HGKFNQTVKA ENGLKLVNGK PIIIFQERDP
81      91      101     111     121     131     141     151
TNIKWGEAGA EYVVESTGVF TTEKAGAGHL RCGAKRVIIS APSADAPHEV HGVNHEKYDN SLKIVSNASC TTNCLAPLAK
161     171     181     191     201     211     221     231
VIHDFGIVE GLMTTVHAIT ATQKTVDGFS GKLWRDGRGA DENLIPASTG ARKAVGRVIP ELNGKLTGMR FRVPTPNVSV
241     251     261     271     281     291     301     311
VDLTCRLKPK AKYDDIKGVV RQASEGPLK ILCVTFDQVV SCDPNSNSHS STFDAGACIA LNDNFVKLLS WVDNEYGYSN
321     331
RVVDLMAYMGR SKE
    
```

Int Pk 1	Int Pk 2	Int Pk 3	Int Pk 4	Area Pk 1	Area Pk 2	Area Pk 3	Area Pk 4
1840	1710	1480	1650	1182	1167	9850	1143

Num Unique	m/z	Error (ppm)	Peptide	F	S	R	#	Int Pk 1	Int Pk 2	Int Pk 3	Int Pk 4	Score	Score Diff	Disc Score
13	860.4192	-6.3	WGEAGAEYVVESTGVFTTMEK	4	79.26	2	1	1840	1710	1480	1650	52.3	38.0	5.53
	865.7486	-8.9	WGEAGAEYVVESTGVFTTmEK	3	74.67	1	1	2390	2390	2000	2320	49.5	33.2	5.00
	862.4451	-9.1	LSWYVDNEYGYSNR	2	64.91	2	1	841.0	752.0	573.0	705.0	48.4	31.6	4.82
	829.4695	-10	GAAQNIPASTGAAK	14	46.72	2	1	636.0	555.0	473.0	544.0	45.5	31.5	4.81
	758.4183	8.5	VVDLMAYMASK	11	76.19	2	1	482.0	443.0	338.0	374.0	44.6	30.2	4.67
	703.3749	0.14	IVSNASCTTNCLAPLAK	11	57.05	2	1	497.0	527.0	443.0	510.0	38.7	26.3	4.24
	822.9295	-4.4	VVDLMAYMASKE	9	79.28	2	1	649.0	615.0	498.0	638.0	39.6	24.8	4.08
	470.2616	-5.9	LTGMAFR	9	51.68	2	1	789.0	729.0	625.0	701.0	36.6	18.6	3.39
	559.6275	3	VVDLMAYmASKE	8	51.32	1	1	518.0	345.0	285.0	379.0	25.3	12.7	2.75
	516.6098	5.3	VVDLMAYmASK	12	54.58	2	1	576.0	572.0	524.0	513.0	23.0	12.3	2.70
	554.2877	-3.6	VVDLMAYmASKE	9	61.59	1	1	817.0	737.0	650.0	691.0	25.3	5.2	1.92
	511.2686	-13	VVDLMAYmASK	9	59.86	2	1	724.0	636.0	550.0	605.0	27.2	5.1	1.91

Acknowledgements

- Funding: NIH NCRR grant 01614 and the Vincent Coates Foundation.