Database Searching of Combined ETD and CID Data Using Protein Prospector

Peter R. Baker, Robert J. Chalkley , Aenoch J. Lynn, Shenheng Guan and Alma L. Burlingame

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

Introduction

- CID and ETD produce complementary fragmentation data.
  - Performing both types of fragmentation sequentially on the same precursor should allow identification of a higher percentage of precursor ions.
- The two types of data produce different fragment ion types, so must be searched with different parameters.
  - It would be useful to be able to correlate/combine the results from the two fragmentation types.

Experimental

In-house peak-picking software separated the ETD and CID peaklists which were then searched by the Protein Prospector Batch-Tag program. The data set, from a recently published study of the murine postsynaptic density pseudoorganelle[1], consisted of 32500 spectra from each fragmentation type. The database search algorithm has been modified to take account of additional fragmentation sometimes found in ETD spectra.

After database searching the results can be displayed in a single report using the Protein Prospector Search Compare program.

Results

	CID		ETD		Merged	
Proteins	441		418		514	
Decoy Proteins	0		5		5	
	Unique Peptides	**Total Peptides**	**Unique Peptides**	**Total Peptides**	**Unique Peptides**	**Total Peptides**
Spectrin alpha chain, brain	119	597	85	406	129	1003
Spectrin beta chain, brain 1	72	411	56	265	80	676
Neurofilament medium polypeptide	33	173	29	139	35	312
Actin, cytoplasmic 2	27	271	23	231	29	502
Tubulin beta-5 chain	26	236	22	161	27	434
Tubulin alpha-1A chain	25	260	20	195	26	455
2',3'-cyclic-nucleotide 3'-phosphodiesterase	23	129	21	86	29	215
ADP/ATP translocase 2	22	154	20	129	24	283
Calcium/calmodulin-dependent protein kinase type II alpha chain	20	242	19	174	24	416
Creatine kinase, ubiquitous mitochondrial precursor	17	168	14	136	19	304

Separated and Merged CID/ETD Protein Report

CID ETD

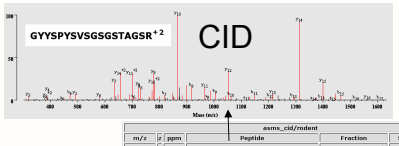
Rank	Uniq Pep	Acc #	asms_cid/rodent				asms_etd/rodent				Protein MW	Species	Protein Name
			Num Unique	Peptide Count	% Cov	Best Expect Val	Num Unique	Peptide Count	% Cov	Best Expect Val			
1		P16546	119	597	51.3	8.3e-11	85	405	40.0	6.6e-9	284599.7	MOUSE	Spectrin alpha chain, brain
2		Q62261	72	411	36.2	1.4e-11	56	265	28.3	1.3e-9	274225.4	MOUSE	Spectrin beta chain, brain 1
2-1	22	Q62696	24	103	12.4	7.2e-8	15	66	8.4	8.0e-8	271066.5	RAT	Spectrin beta chain, brain 2
2-2	5	P15508	9	25	4.3	3.6e-9	4	5	2.1	2.3e-6	245252.3	MOUSE	Spectrin beta chain, erythrocyte
3		P08553	33	173	39.3	2.1e-9	29	139	37.6	1.1e-8	95916.9	MOUSE	Neurofilament medium polypeptide
3-1	24	P46660	31	198	59.1	2.0e-10	19	103	40.7	1.1e-7	55742.3	MOUSE	Alpha-internexin
4		P63260	27	271	69.9	1.5e-11	23	231	68.5	2.4e-9	41793.2	MOUSE	Actin, cytoplasmic 2
4-1	1	P60710	26	263	69.9	1.5e-11	22	229	68.5	2.4e-9	41737.1	MOUSE	Actin, cytoplasmic 1
5		Q7TMM9	25	248	67.4	1.1e-8	23	186	62.0	5.5e-8	49907.4	MOUSE	Tubulin beta-2A chain
6-1	11	Q5DEF9	24	180	66.9	4.4e-9	19	141	51.6	5.5e-8	49586.2	MOUSE	Tubulin beta-4 chain
6		P68369	25	260	62.5	1.5e-11	20	195	58.5	6.5e-10	50136.1	MOUSE	Tubulin alpha-1A chain
6-1	3	P05213	24	240	62.3	1.5e-11	19	169	58.3	6.5e-10	50152.1	MOUSE	Tubulin alpha-1B chain
7		P16330	23	129	61.4	3.6e-11	21	96	48.6	2.5e-7	47123.7	MOUSE	2',3'-cyclic-nucleotide 3'-phosphodiesterase
8		P51881	22	154	62.1	1.7e-11	20	129	55.0	4.2e-8	32931.6	MOUSE	ADP/ATP translocase 2
8-1	10	P48962	21	201	56.0	1.7e-11	16	142	46.6	2.6e-8	32904.6	MOUSE	ADP/ATP translocase 1

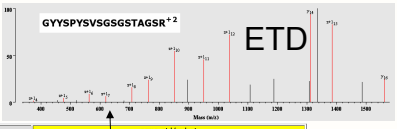
There is now a preferred species option in Search Compare for selecting hits from particular species or taxomonies in favour of others if the hits are equivalent. Here the preference is MOUSE in a RODENT search.

Merged CID/ETD

Rank	Uniq Pep	Acc #	Merged				Protein MW	Species	Protein Name
			Num Unique	Peptide Count	% Cov	Best Expect Val			
1		P16546	129	1003	54.2	8.3e-11	284599.7	MOUSE	Spectrin alpha chain, brain
2		Q62261	80	676	40.4	1.4e-11	274225.4	MOUSE	Spectrin beta chain, brain 1
2-1	22	Q62696	29	169	14.9	7.2e-8	271066.5	RAT	Spectrin beta chain, brain 2
2-2	5	P15508	9	30	4.3	3.6e-9	245252.3	MOUSE	Spectrin beta chain, erythrocyte
3		P08553	35	312	40.2	2.1e-9	95916.9	MOUSE	Neurofilament medium polypeptide
3-1	24	P46660	31	301	59.1	2.0e-10	55742.3	MOUSE	Alpha-internexin
4		P63260	29	502	71.7	1.5e-11	41793.2	MOUSE	Actin, cytoplasmic 2
4-1	1	P60710	28	491	71.7	1.5e-11	41737.1	MOUSE	Actin, cytoplasmic 1
5		Q7TMM9	27	434	67.4	1.1e-8	49907.4	MOUSE	Tubulin beta-2A chain
6-1	11	Q5DEF9	25	321	66.9	4.4e-9	49586.2	MOUSE	Tubulin beta-4 chain
6		P68369	26	455	62.5	1.5e-11	50136.1	MOUSE	Tubulin alpha-1A chain
6-1	3	P05213	25	409	62.3	1.5e-11	50152.1	MOUSE	Tubulin alpha-1B chain
7		P16330	29	215	66.9	3.6e-11	47123.7	MOUSE	2',3'-cyclic-nucleotide 3'-phosphodiesterase
8		P51881	24	283	62.1	1.7e-11	32931.6	MOUSE	ADP/ATP translocase 2
8-1	10	P48962	22	343	59.7	1.7e-11	32904.6	MOUSE	ADP/ATP translocase 1

Separated and Merged CID/ETD Peptide Report





asms_cid/rodent

m/z	z	ppm	Peptide	Fraction	S	Score	Expect
891.8995	2	-0.66	GYYSPYSVSGGSTAGSR	T8102008_ITMSms2cid	58.218	99.7	1.7e-9
907.9499	2	-3.7	FLEGTSC(Carbamidomethyl)IAGVFVDAIK	T8102009_ITMSms2cid	68.542	90.0	9.3e-7
1050.0829	2	-4.4	IQGIN-::pyro-Glu)YINAIKDYELQLITYK	T8102016_ITMSms2cid	78.095	90.0	1.4e-6
872.4671	2	0.27	YLYGTGAVAGVYLPGSR	T8102008_ITMSms2cid	69.831	93.7	2.1e-6
898.9476	2	-2.0	YLQDLLAWVEENQR	T8102008_ITMSms2cid	79.595	94.7	1.2e-5
769.7162	3	-5.4	GFFDPNTHENLTYLQLIER	T8102012_ITMSms2cid	73.058	94.2	1.3e-5
643.8444	2	-2.8	WQAVLAQTDVR	T8102008_ITMSms2cid	47.173	85.2	1.8e-5
1053.8402	2	-0.011	FLEVQYLTQGLIEPDTPGR	T8102007_ITMSms2cid	73.521	97.0	3.5e-5
986.1959	3	-0.23	DLLQPEVAVALLEAQAGTGHIIDPATSAR	T8102016_ITMSms2cid	80.924	89.9	4.6e-5
457.1681	3	3.3	KDWLVYEAQGR	T8102008_ITMSms2cid	51.587	91.1	9.6e-5
621.7969	2	-3.0	GWLVYEAQGR	T8102008_ITMSms2cid	51.587	91.1	9.6e-5
466.5847	3	-2.5	NLELHYQAFLR	T8102009_ITMSms2cid	67.116	85.8	1.7e-4
930.3819	2	-5.9	NTLYQTLEIQR	T8102007_ITMSms2cid	57.429	99.7	1.9e-4
528.2807	2	-4.1	AHAFVYQGR	T8102008_ITMSms2cid	25.962	16.8	3.4e-4
441.9133	3	-1.0	IVITVVEEHER	T8102016_ITMSms2etd	41.443	31.0	8.6e-4
466.2698	2	-0.57	KWRLVYFR	T8102009_ITMSms2cid	61.872	17.9	0.0012
692.2916	2	-2.6	LFNAIHR	T8102009_ITMSms2cid	44.332	17.4	0.0021
458.2732	2	-2.1	LLLWSQR	T8102008_ITMSms2cid	15.109	16.9	0.0080
401.2626	2	-1.7	IIRLVIR	T8102017_ITMSms2etd	48.687	21.5	0.0084

asms_etd/rodent

m/z	z	ppm	Peptide	Fraction	S	Score	Expect
891.8995	2	-0.66	GYYSPYSVSGGSTAGSR	T8102008_ITMSms2etd	58.218	99.4	0.0014
514.9521	3	-0.63	LGFHLPLEVAYQR	T8102010_ITMSms2etd	64.790	40.3	9.7e-7
872.4671	2	0.27	YLYGTGAVAGVYLPGSR	T8102008_ITMSms2etd	69.831	93.3	4.8e-4
898.9476	2	0.9	YLQDLLAWVEENQR	T8102010_ITMSms2etd	80.898	97.0	0.0067
769.7162	3	-5.4	GFFDPNTHENLTYLQLIER	T8102012_ITMSms2etd	73.058	94.2	1.3e-5
457.1681	3	-1.0	KDWLVYEAQGR	T8102008_ITMSms2etd	51.204	44.6	7.8e-5
621.7969	2	-3.0	GWLVYEAQGR	T8102008_ITMSms2etd	51.204	93.1	0.0070
466.5847	3	-2.5	NLELHYQAFLR	T8102009_ITMSms2etd	67.116	85.8	1.7e-4
441.9133	3	-1.0	IVITVVEEHER	T8102016_ITMSms2etd	41.443	31.0	8.6e-4
466.2698	2	-0.57	KWRLVYFR	T8102009_ITMSms2etd	61.872	17.9	0.0012
692.2916	2	-2.6	LFNAIHR	T8102009_ITMSms2etd	44.332	17.4	0.0021
401.2626	2	-1.7	IIRLVIR	T8102017_ITMSms2etd	48.687	21.5	0.0084

Merged

m/z	z	ppm	Peptide	Fraction	S	Score	Expect
891.8995	2	-0.66	GYYSPYSVSGGSTAGSR	T8102008_ITMSms2cid	58.218	99.7	1.7e-9
907.9499	2	-3.7	FLEGTSC(Carbamidomethyl)IAGVFVDAIK	T8102009_ITMSms2cid	68.542	90.0	9.3e-7
514.9521	3	-0.63	LGFHLPLEVAYQR	T8102010_ITMSms2etd	64.790	40.3	9.7e-7
1050.0829	2	-4.4	IQGIN-::pyro-Glu)YINAIKDYELQLITYK	T8102016_ITMSms2cid	78.095	90.0	1.4e-6
872.4671	2	0.27	YLYGTGAVAGVYLPGSR	T8102008_ITMSms2cid	69.831	93.7	2.1e-6
898.9476	2	-2.0	YLQDLLAWVEENQR	T8102008_ITMSms2cid	79.595	94.7	1.2e-5
769.7162	3	-5.4	GFFDPNTHENLTYLQLIER	T8102012_ITMSms2etd	73.058	94.2	1.3e-5
643.8444	2	-2.8	WQAVLAQTDVR	T8102008_ITMSms2cid	47.173	85.2	1.8e-5
1053.8402	2	-0.011	FLEVQYLTQGLIEPDTPGR	T8102007_ITMSms2cid	73.521	97.0	3.5e-5
986.1959	3	-0.23	DLLQPEVAVALLEAQAGTGHIIDPATSAR	T8102016_ITMSms2cid	80.924	89.9	4.6e-5
457.1681	3	3.3	KDWLVYEAQGR	T8102008_ITMSms2etd	45.204	44.6	7.8e-5
621.7969	2	-3.0	GWLVYEAQGR	T8102008_ITMSms2cid	51.587	91.1	9.6e-5
466.5847	3	-2.5	NLELHYQAFLR	T8102009_ITMSms2etd	67.116	85.8	1.7e-4
930.3819	2	-5.9	NTLYQTLEIQR	T8102007_ITMSms2cid	57.429	99.7	1.9e-4
528.2807	2	-4.1	AHAFVYQGR	T8102008_ITMSms2cid	25.962	16.8	3.4e-4
441.9133	3	-1.0	IVITVVEEHER	T8102016_ITMSms2etd	41.443	31.0	8.6e-4
466.2698	2	-0.57	KWRLVYFR	T8102009_ITMSms2cid	61.872	17.9	0.0012
692.2916	2	-2.6	LFNAIHR	T8102009_ITMSms2cid	44.332	17.4	0.0021
458.2732	2	-2.1	LLLWSQR	T8102008_ITMSms2cid	15.109	16.9	0.0080
401.2626	2	-1.7	IIRLVIR	T8102017_ITMSms2etd	48.687	21.5	0.0084

The merged report can either show the best peptide (CID or ETD) or all the peptides identified.

Peptide Difference Results

asms_cid/rodent								asms_etd/rodent						
m/z	z	ppm	Peptide	Fraction	S	Score	Expect	m/z	z	ppm	Peptide	Fraction	S	Score	Expect
807.9499	2	-3.7	FLEGTSC(Carbamidomethyl)IAGVFVDATK	T8102009_1TMSms2cid	68.542	30.0	9.3e-7								
								614.9522	3	-0.63	LGFHLPLEVAYQR	T8102010_1TMSms2etd	84.798	40.3	9.7e-7
1050.0520	2	-4.4	Q(Gln->pyro-Glu)YINAIKDYELQLITYK	T8102016_1TMSms2cid	78.095	30.0	1.4e-6								
643.8444	2	-2.8	WQAVLAQTDVR	T8102008_1TMSms2cid	47.173	25.2	1.8e-5								
1053.0493	2	-0.011	FLEVQYLTGGLIEPDTPGR	T8102007_1TMSms2cid	73.521	27.0	3.5e-5								
986.1959	3	-0.23	DLLQPEVAVALLEAQAGTGHIIDPATSAR	T8102016_1TMSms2cid	80.924	28.9	4.6e-5								
650.3819	2	-5.0	VTLVQTLEIQR	T8102007_1TMSms2cid	57.428	28.7	1.9e-4								
528.2887	2	-4.1	AHAFVVQQR	T8102008_1TMSms2cid	25.963	16.8	3.4e-4								
								441.9133	3	-1.0	EVITYYEEHER	T8102016_1TMSms2etd	41.443	31.0	8.6e-4
498.2688	2	-0.57	RWSLVTFR	T8102009_1TMSms2cid	61.872	17.9	0.0012								
492.2916	2	-2.6	LFNAIIHR	T8102009_1TMSms2cid	44.552	17.4	0.0021								
458.2732	2	-2.1	LLLWSQR	T8102008_1TMSms2cid	55.193	16.9	0.0080								
								401.2626	2	-1.7	IISLVIR	T8102017_1TMSms2etd	48.687	21.5	0.0084

The peptide difference report only shows peptides that are unique to a particular seach.

This type of report makes it possible to see which amino acids are favoured by CID or ETD.

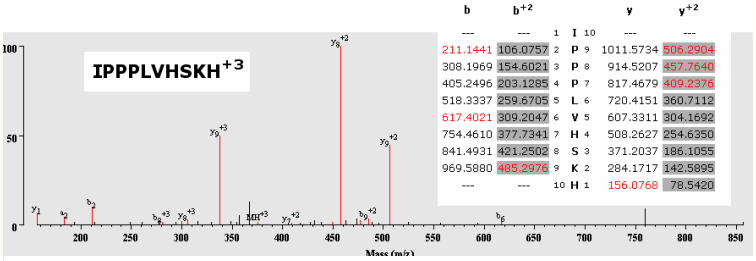
K	H	R	M	W	S	Y	F	V	I	G	P	L	A	T	D	N	E	Q	C
0.60	0.66	0.69	0.75	0.81	0.85	0.90	0.93	1.00	1.01	1.05	1.08	1.08	1.09	1.11	1.28	1.32	1.34	1.52	2.13

Preferred by ETD

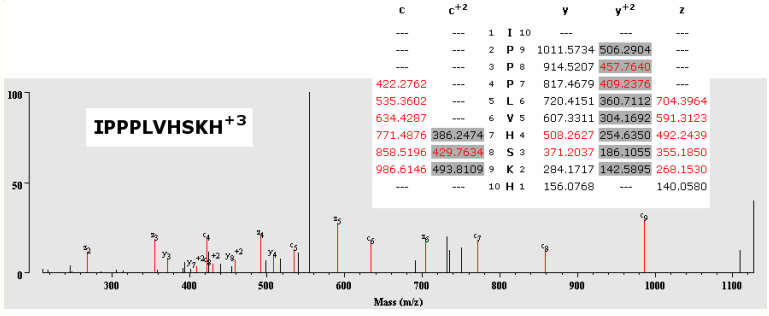
Preferred by CID

Complementary Fragmentation in Proline Rich Peptides

CID



ETD

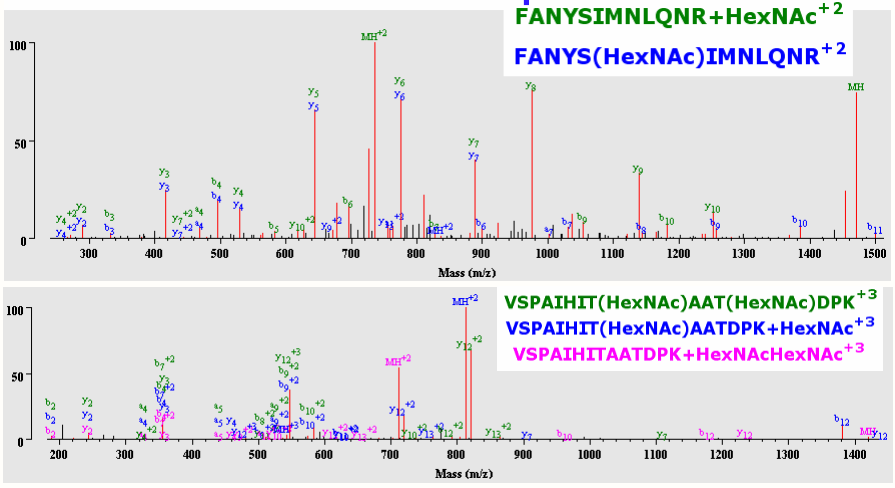


O-GlcNAc Modifications

O-GlcNAc modifications can be searched for in Protein Prospector by specifying HexNAc as a variable modification. This can either be as a modification to Serine or Threonine or as a single/multiple neutral loss from the precursor ion.

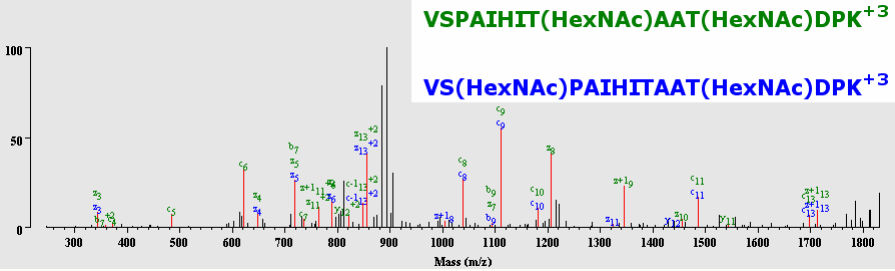
58 exact modification sites were found in addition to several more where it wasn't possible to locate the site unambiguously. 28 of the sites were found on the protein bassoon (O88737).

Fully Annotated O-GlcNAc Modified Spectra



MS-Product can now annotate multiple spectra simultaneously thus allowing O-GlcNAc modified peptides to be fully labelled.

Comparison of Alternative Modification Sites



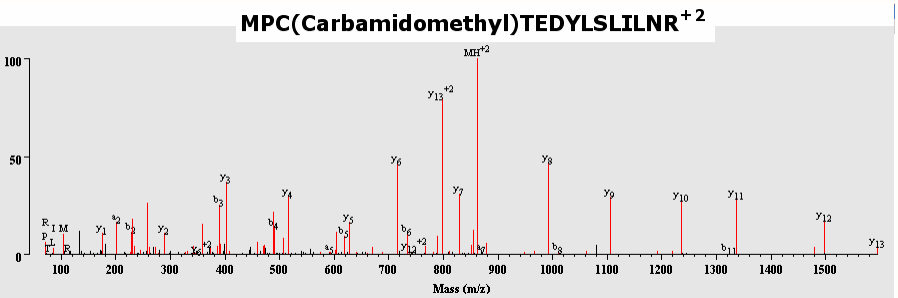
Annotating multiple peptides simultaneously can also be used to compare matches from different modification sites. In this case the green sequence is preferred because of the c_4, c_5, c_6, c_7, z_7, z_8 and z_9 ions. The CID spectrum for this peptide is shown on the previous slide.

Other Potential Applications

Assuming sufficient computer memory there is no limit to the number of searches that can be merged. Thus the following applications are possible:

- Compilation of spectral libraries
- Large scale studies of spectral fragmentation
- Identification of sections of proteins that are not detected in MSMS experiments – maybe because they are modified
- Stable isotope and label free quantitation

Compilation of Spectral Libraries



Best spectrum of a particular peptide from several hundred searches.

Undetectable Parts of Proteins

```
1   MKWVTFISLL LLFSSAYSRG VFRRDTHKSE IAHRFKDLGE EHFKGLVLIA FSQYLQQCPF DEHVKLVNEL TEFAKTCVAD
81  ESHAGCEKSL HTLFGDELCK VASLRETYGD MADCCEKQEP ERNECFLSHK DDSPDLPKLK PDPNTLCDEF KADEKKFWGK
161 YLYEIARRHP YFYAPELLYY ANKYNGVFQE CCQAEDKGAC LLPKIETMRE KVLASSARQR LRCASIQKFG ERALKAWSVA
241 RLSQKFPKAE FVEVTKLVTD LTKVHKECCH GDLLECADDR ADLAKYICDN QDTISSKLKE CCDKPLLEKS HCIAEVEKDA
321 IPENLPPLTA DFAEDKDVCK NYQEAKDAFL GSFLYEYSRR HPEYAVSVLL RLAKEYEATL EECCAKDDPH ACYSTVFDKL
401 KHLVDEPQNL IKQNCDQFEK LGEYGFQNAL IVRYTRKVPQ VSTPTLVEVS RSLGKVGTRC CTKPESERMP CTEDYLSLIL
481 NRLCVLHEKT PVSEKVTKCC TESLVNRRPC FSALTPDETY VPKAFDEKLF TFHADICTLP DTEKQIKKQT ALVELLKHKP
561 KATEEQLKTV MENFVAFVDK CCAADDKEAC FAVEGPKLVV STQTALA
```

Coverage map for BOVINE Serum Albumin for 90 combined searches. The coverage is 91% compared with a maximum coverage of 80% for any individual search. The undetectable regions may because of the signal and propeptide regions of the protein and the prevalance of tryptic cleavage amino acids in some regions.

Conclusions

- Combined CID/ETD search identified around 15% more proteins
- Peptides containing basic amino acids often fragment better ETD
- Peptides containing cysteine often fragment better with CID
- Sequential CID/ETD fragmentation is particularly useful for locating O-GlcNAc modifications
- Protein Prospector can often facilitate O-GlcNAc site assignment
- The ability to merge search results is also useful for other applications such as compiling spectral libraries

Acknowledgements

This work was supported by NIH NCRR grant RR001614, SIG RR019934 and the Biotechnology and Biological Sciences Research Council of the UK. We also thank the Vincent Coates Foundation for support.

References

1) Chalkley, R. J et al, 2009 Identification of Protein O-GlcNAcylation Sites Using Electron Transfer Dissociation Mass Spectrometry On Native Peptides. *PNAS,* Currently Published Online