

Targeted
Metabolomics &
Proteomics
Laboratory

High sensitivity and high specificity analysis of lens peptides

Stephen Barnes, PhD

MCLM 452; 4-7117

sbarnes@uab.edu

BMG744 1-30-12

Barnes Lens Research Group



Stephen Barnes, PhD



Om Srivastava, PhD



Matt Renfrow, PhD



Kevin Schey, PhD,
Vanderbilt



David Stella, PhD
ADFS



Kyle Floyd



Landon Wilson

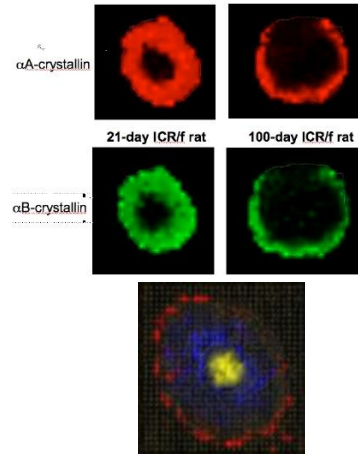
Studying the role of diet and bioactives on lens proteins
 Funding support: P50 AT00477, S10 RR17261, S10 RR19231,
 R21 EY020963, EyeSight Foundation of Alabama

BMG744 1-30-12



Challenges in lens research

- Mass spectrometry imaging has revealed that individual proteins such as α A-crystallin are found in a full-length form in the epithelial cell monolayer, but in many truncated forms throughout the rest of the lens
- We need to develop quantitative peptide assays to identify when these truncated peptides are formed during development



Red, aa1-173; blue, aa1-157;
yellow, aa1-53 – all α A-crystallin

BMG744 1-30-12

Synopsis

- Need for quantitative analysis of specific peptides in proteomics
- Principle of reaction ion monitoring
- Selection of peptides for analysis
- The problem of the complexity of mass space
- Advantages of, indeed need for, a high resolution, high mass accuracy analyzer

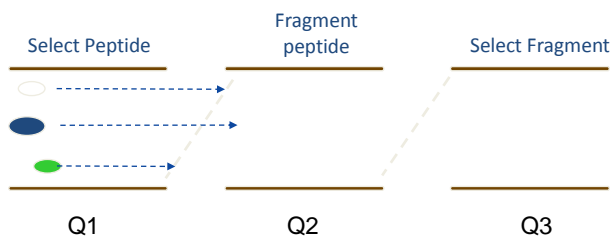
BMG744 1-30-12

Quantitative proteomics

- Proteomics is rapidly moving out of just being about *discovery*
- Investigators want to measure changes occurring in protein levels in whole networks
 - 2-fold changes are not sufficient
 - Critical proteins have important changes in amounts that are 20% or less
- Some proteins are only theoretical
 - A result of truncations, mRNA splicing

BMG744 1-30-12

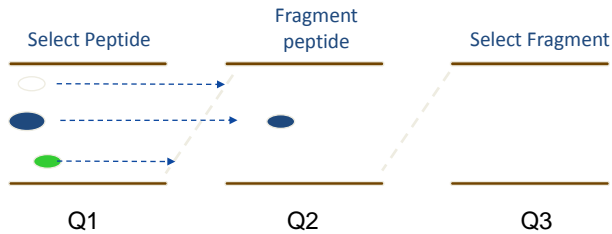
Multiple Reaction Monitoring (MRM)



- Selects one specific peptide ion in the first quadrupole

BMG744 1-30-12

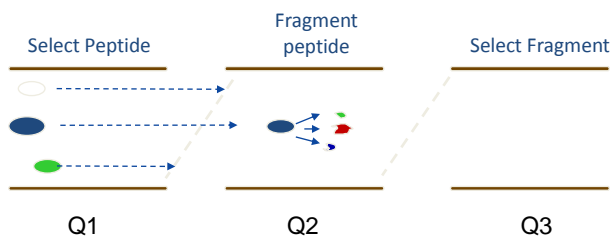
Multiple Reaction Monitoring (MRM)



- Isolated, selected peptide ion enters Q2

BMG744 1-30-12

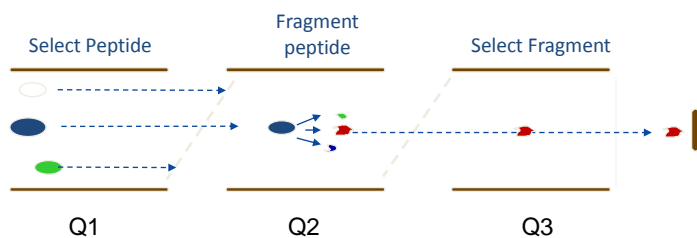
Multiple Reaction Monitoring (MRM)



- Isolated peptide precursor ion collides with gas and breaks into fragment (product) ions

BMG744 1-30-12

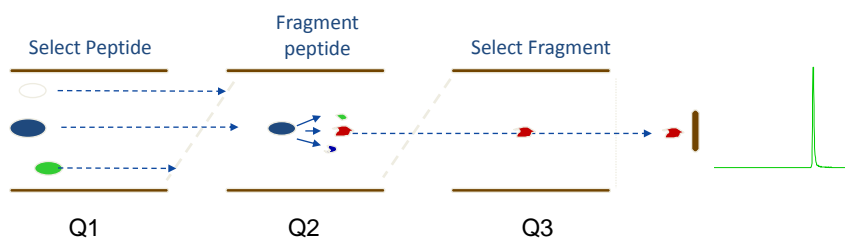
Multiple Reaction Monitoring (MRM)



- Selected fragment ion (based on the sequence of the peptide precursor) isolated in Q3 and hits the detector

BMG744 1-30-12

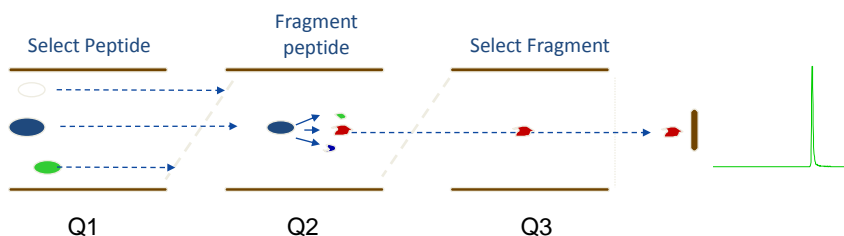
Multiple Reaction Monitoring (MRM) on a triple quad



- Precursor ion-product ion signal is monitored for 20-50 msec

BMG744 1-30-12

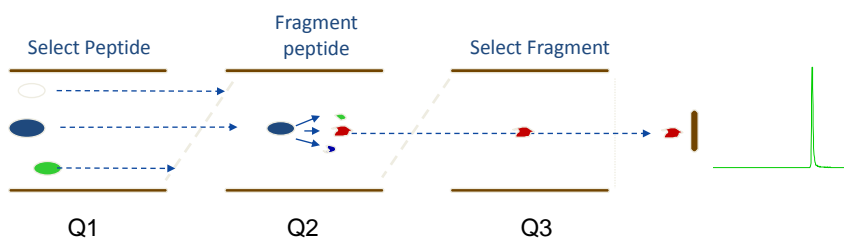
Multiple Reaction Monitoring (MRM) on a triple quad



- Precursor ion-product ion signal is monitored for 20-50 msec
- Then a second pair of precursor-product ions is monitored

BMG744 1-30-12

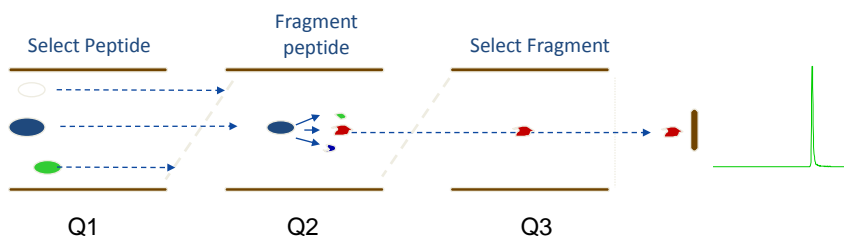
Multiple Reaction Monitoring (MRM) on a triple quad



- Precursor ion-product ion signal is monitored for 20-50 msec
- Then a second pair of precursor-product ions is monitored
- For a 50 msec data collection, 20 different channels of information can be acquired per second

BMG744 1-30-12

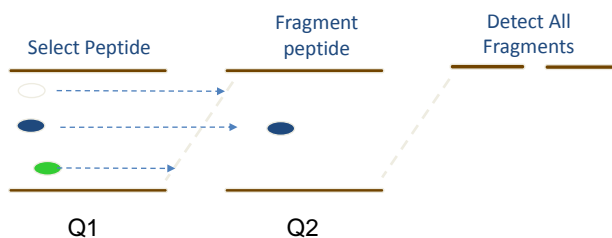
Multiple Reaction Monitoring (MRM) on a triple quad



- Precursor ion-product ion signal is monitored for 20-50 msec
- Then a second pair of precursor-product ions is monitored
- For a 50 msec data collection, 20 different channels of information can be acquired per second
- In each of the following seconds, this sequence of actions is repeated, leading to the generation of ion chromatograms

BMG744 1-30-12

Pseudo MRM Analysis

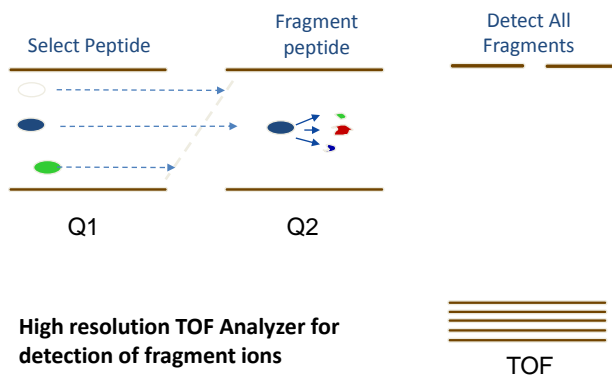


- High resolution TOF Analyzer for detection of fragment ions

TOF

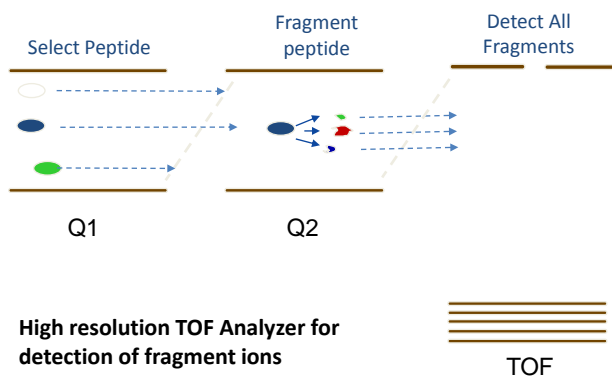
BMG744 1-30-12

Pseudo MRM Analysis



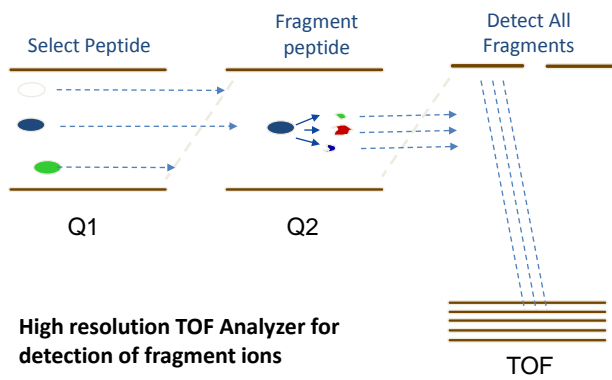
BMG744 1-30-12

Pseudo MRM Analysis



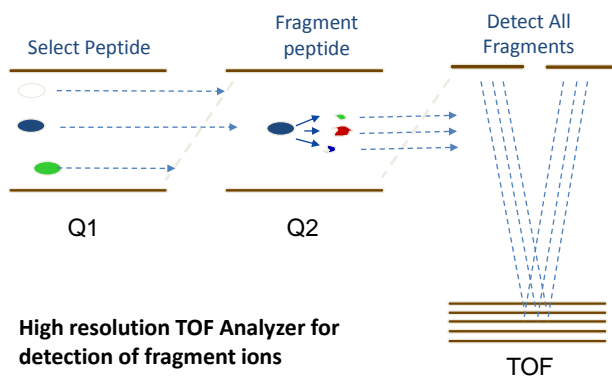
BMG744 1-30-12

Pseudo MRM Analysis



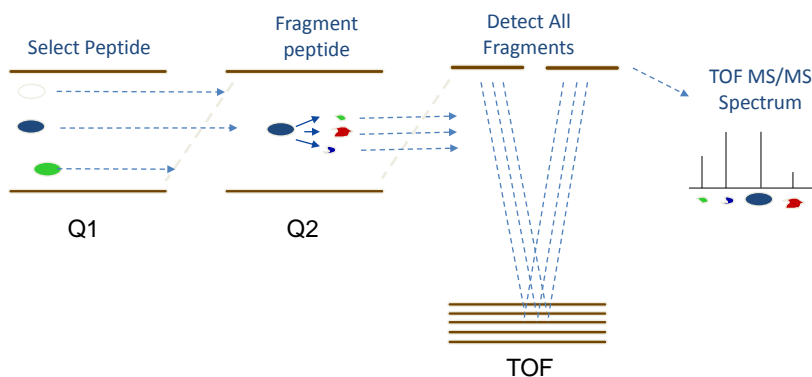
BMG744 1-30-12

Pseudo MRM Analysis



BMG744 1-30-12

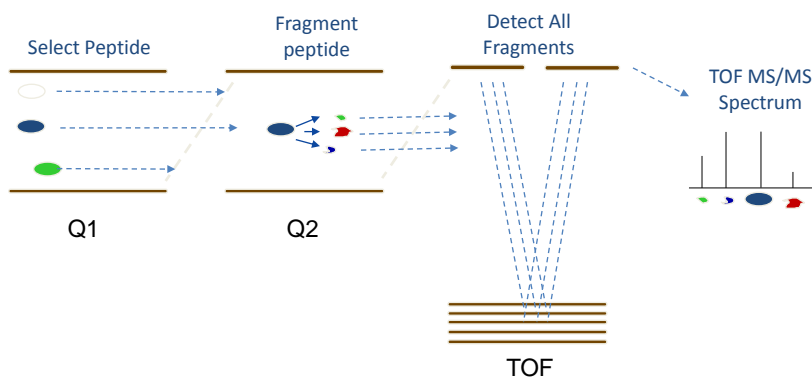
Pseudo MRM Analysis



- The key difference between the TripleTOF and the triple quad is that the entire MSMS spectrum is collected by the TripleTOF in a single 50 sec (or shorter) data acquisition – the selection of product ion to follow is made post-data acquisition

BMG744 1-30-12

Pseudo MRM Analysis



- The key difference between the TripleTOF and the triple quad is that the entire MSMS spectrum is collected by the TripleTOF in a single 50 sec (or shorter) data acquisition – the selection of product ions is made post-data acquisition
- The mass accuracy of the product ions is 3-5 ppm

BMG744 1-30-12

Selecting a peptide for quantitative analysis

- There are databases of peptides from a proteome
 - These have tools to indicate the best peptides for analysis
 - <http://www.srmatlas.org/mrmassays.php>
 - MRMPilot – AB Sciex
 - Skyline 1.1 - <https://skyline.gs.washington.edu/>

BMG744 1-30-12

Searching pathways - MRMPPath

- Proteins rarely operate all on their own, but rather in pathways or groups
- MRMPPath is web-based software that was developed to facilitate recovery of information about suitable proteotypic peptides
- It's based on data mining of the KEGG (Kyoto Encyclopedia of Genes and Genomes) databases

BMG744 1-30-12

Use of MRMPath

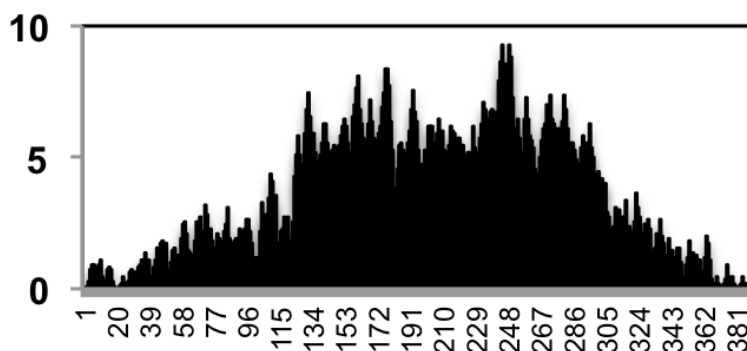
<http://templ.uab.edu/MRMPath/>

Users select the protease they want to use, the pathway of interest and the species in which the research was carried out

- An image of the pathway is presented to them and they can select either a specific protein by clicking on it, or all the proteins in the pathway
- The software does an *in silico* digestion of each protein and filters the peptides to produce those with 7-25 amino acid residues
 - It also removes peptides containing Cys or Met residues
- The user can BLAST each peptide one at a time, or all at once

BMG744 1-30-12

Are there canonical protein sequences??



This plot shows the weighted average of the number of mutations per amino acid residue for p53. There are 1361 described mutations for 393 amino acid residues.

BMG744 1-30-12

Pragmatic selection of a peptide

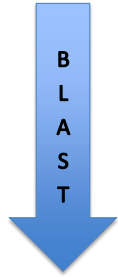
DLG4_HUMAN Mass: 80788 Score: 388 Queries matched: 18 emPAI: 0.68
 Disks large homolog 4 OS=Homo sapiens GN=DLG4 PE=1 SV=3

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
840	404.71	807.40	807.42	-0.02	1	19	0.61	1	K.AFDRATK.L
854	406.23	810.45	810.45	-0.00	1	1	24	3	K.RGFYIR.A
1135	436.21	870.40	870.41	-0.01	0	36	0.0049	1	R.ALFDYDK.T
1173	441.20	880.39	880.39	-0.00	0	40	0.0019	1	R.EYEIDGR.D
1730	519.25	1036.48	1036.49	-0.01	1	20	0.31	1	K.REYEIDGR.D
2022	557.79	1113.57	1113.57	-0.00	0	45	0.0013	1	K.NTYDVVYLK.V
2042	562.24	1122.47	1122.47	-0.00	0	44	0.00038	1	K.DWGSSSGSQGR.E
2048	563.30	1124.58	1124.59	-0.01	0	59	4.8e-005	1	K.IIPGGAAAQDGR.L 2049
2125	578.79	1155.57	1155.58	-0.01	0	50	0.00032	1	K.DLLGEEDIPR.E
2349	418.22	1251.64	1251.66	-0.01	0	41	0.0026	1	R.NASHEQAAALK.N
2357	418.89	1253.65	1253.66	-0.02	0	38	0.0055	1	R.EVTHSAAVEALK.E
2484	438.91	1313.72	1313.73	-0.01	1	27	0.073	1	R.SLENVLEINKR.I
2558	452.23	1353.67	1353.68	-0.01	0	63	1.6e-005	1	K.HCILDVSAVAVR.R
2563	682.32	1362.62	1362.63	-0.01	0	95	6.9e-009	1	R.ANDLLSEFPDK.F
2601	462.90	1385.67	1385.69	-0.01	0	11	2.3	1	K.FGSCVPHTRPK.R
2715	505.28	1512.81	1512.83	-0.01	1	62	2e-005	1	R.KGQILSVNGVDLR.N
2737	513.59	1537.76	1537.77	-0.01	1	32	0.021	1	K.DLLGEEDIPRPR.R

MASCOT PROTEIN SUMMARY REPORT

BMG744 1-30-12

IIPGGAAAQDGR



Accession	Description	Max score	Total score	Query coverage	E-value
gi1957526 NP_001356.1	disks large homolog 4 isoform 1 [Homo sapiens] >q 3318653 qb AAC	28.8	61.9	100%	5e-07
gi5918874 AAD56172.1	post-synaptic density 95 [Homo sapiens]	28.8	61.9	100%	5e-07
gi119610659 RAW90232.1	discs, large homolog 4 (Drosophila), isoform CRA_c [Homo sapiens] >	28.8	61.9	100%	5e-07
gi17658825 P29352.3	ReName: Full=Disks large homolog 4; AltName: Full=Postsynaptic de	28.8	61.9	100%	5e-07
gi192447426 NP_001122289.1	disks large homolog 4 isoform 2 [Homo sapiens] >q 119610661 qb E	28.8	61.9	100%	5e-07
gi221041202 Bdmi12328.1	unnamed protein product [Homo sapiens]	28.8	61.9	100%	5e-07
gi221041202 Bdmi12558.1	unnamed protein product [Homo sapiens]	28.8	61.9	100%	5e-07
gi119610658 RAW90232.1	discs, large homolog 4 (Drosophila), isoform CRA_b [Homo sapiens] >	28.8	61.9	100%	5e-07
gi119610657 RAW90231.1	discs, large homolog 4 (Drosophila), isoform CRA_a [Homo sapiens]	28.8	61.9	100%	5e-07
gi1527214 AAB807236.1	PSD-95 [Homo sapiens]	28.8	61.9	100%	5e-07
gi2813006 AAB84250.1	Tax interaction protein 15 [Homo sapiens] >q 119610660 qb EAW9C	28.8	61.9	100%	5e-07
gi159162899 LKEF_6	Chain A, Pd21 Of Sap90	28.8	38.8	100%	5e-07
gi218156338 NP_001136171.1	disks large homolog 2 isoform 1 [Homo sapiens]	28.8	59.0	100%	6e-06
gi151491250 CAH18860.1	hypothetical protein [Homo sapiens]	28.8	59.0	100%	6e-06
gi332164718 NP_001193998.1	disks large homolog 2 isoform 5 [Homo sapiens]	28.8	59.0	100%	6e-06
gi119595099 Bdmi75953.1	discs, large homolog 2, chapsyn-110 (Drosophila), isoform CRA_a [Hc	28.8	59.0	100%	6e-06
gi119595098 NP_001355.3	discs, large homolog 2 isoform 2 [Homo sapiens] >q 215274165 sp Q	28.8	59.0	100%	6e-06
gi1463026 AAB04949.1	channel associated protein of synapse [Homo sapiens]	28.8	59.0	100%	6e-06
gi221040586 Bdmi11970.1	unnamed protein product [Homo sapiens]	28.8	59.0	100%	6e-06
gi221040860 Bdmi12131.1	unnamed protein product [Homo sapiens]	28.8	59.0	100%	6e-06
gi221039974 Bdmi11750.1	unnamed protein product [Homo sapiens]	28.8	59.0	100%	6e-06

BMG744 1-30-12

Pragmatic selection of a peptide

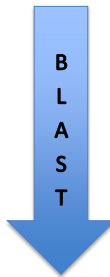
DLG4_HUMAN Mass: 80788 Score: 388 Queries matched: 18 emPAI: 0.68
 Disks large homolog 4 OS=Homo sapiens GN=DLG4 PE=1 SV=3

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
840	404.71	807.40	807.42	-0.02	1	19	0.61	1	K.AFDRATK.L
854	406.23	810.45	810.45	-0.00	1	1	24	3	K.RGFYIR.A
1135	436.21	870.40	870.41	-0.01	0	36	0.0049	1	R.ALFDYDK.T
1173	441.20	880.39	880.39	-0.00	0	40	0.0019	1	R.EYEIDGR.D
1730	519.25	1036.48	1036.49	-0.01	1	20	0.31	1	K.REYEIDGR.D
2022	557.79	1113.57	1113.57	-0.00	0	45	0.0013	1	K.NTYDVVYK.V
2042	562.24	1122.47	1122.47	-0.00	0	44	0.00038	1	K.DWGSSSGSQR.E
2048	563.30	1124.58	1124.59	-0.01	0	59	4.8e-005	1	K.IIPGAAAQDGR.L 2049
2125	578.79	1155.57	1155.58	-0.01	0	50	0.00032	1	K.DLLGEEDIPR.E
2349	418.22	1251.64	1251.66	-0.01	0	41	0.0026	1	R.NASHEQAALAK.N
2357	418.89	1253.65	1253.66	-0.02	0	38	0.0055	1	R.EVTHSAAVEALK.E
2484	438.91	1313.72	1313.73	-0.01	1	27	0.073	1	R.SLENVLEINKR.I
2558	452.23	1353.67	1353.68	-0.01	0	63	1.6e-005	1	K.HCILDVSAVAVR.R
2563	682.32	1362.62	1362.63	-0.01	0	95	6.9e-009	1	R.ANDLLSEFPDK.F
2601	462.90	1385.67	1385.69	-0.01	0	11	2.3	1	K.FGSCVPHTRPK.R
2715	505.28	1512.81	1512.83	-0.01	1	62	2e-005	1	R.KGQILSVNGVDLR.N
2737	513.59	1537.76	1537.77	-0.01	1	32	0.021	1	K.DLLGEEDIPRPR.R

MASCOT PROTEIN SUMMARY REPORT

BMG744 1-30-12

ANDLLSEFPDK



Accession	Description	Max score	Total score	Query coverage	E value
gi 4557229 NP_001356.1	disks large homolog 4 isoform 1 [Homo sapiens] >gi 3318653 qb AAC	41.4	41.4	100%	7e-08
gi 59148874 AAD56173.1	post-synaptic density 95 [Homo sapiens]	41.4	41.4	100%	7e-08
gi 119610659 EAW90253.1	discs, large homolog 4 (Drosophila), isoform CRA_c [Homo sapiens] >	41.4	41.4	100%	7e-08
gi 174588251 P78352.3	RecName: Full=Disks large homolog 4; AltName: Full=Postsynaptic de	41.4	41.4	100%	7e-08
gi 192447426 INP_001122299.1	disks large homolog 4 isoform 2 [Homo sapiens] >gi 119610661 qb E	41.4	41.4	100%	7e-08
gi 221041202 BAH12328.1	unnamed protein product [Homo sapiens]	41.4	41.4	100%	7e-08
gi 221041262 BAH12558.1	unnamed protein product [Homo sapiens]	41.4	41.4	100%	7e-08
gi 119610658 EAW90252.1	discs, large homolog 4 (Drosophila), isoform CRA_b [Homo sapiens] >	41.4	41.4	100%	7e-08
gi 73999148 AAH40523.1	DLG4 protein [Homo sapiens]	41.4	41.4	100%	7e-08
gi 218156338 INP_001136171.1	disks large homolog 2 isoform 1 [Homo sapiens]	35.4	35.4	91%	9e-06
gi 51491229 CAH18680.1	hypothetical protein [Homo sapiens]	35.4	35.4	91%	9e-06
gi 148939278 INP_004078.2	disks large homolog 1 isoform 2 [Homo sapiens] >gi 119573995 qb E	35.4	35.4	91%	9e-06
gi 5584361 AAAS0588.1	homolog of Drosophila discs large protein, isoform 2 [Homo sapiens]	35.4	35.4	91%	9e-06
gi 332164718 INP_001193688.1	disks large homolog 2 isoform 5 [Homo sapiens]	35.4	35.4	91%	9e-06
gi 148939628 INP_001091894.1	disks large homolog 1 isoform 1 [Homo sapiens] >gi 223590196 sp Q	35.4	35.4	91%	9e-06

Directed selection of a peptide

- In some experiments the choice of the peptide is not based on any previous MSMS data
 - The protein may have resulted from mRNA splicing or to nucleotide deletion within a gene, a premature stop codon, or to protease activity after a protein has been synthesized
 - In these cases, the peptide that might be formed is theoretical and its detection may not optimal – on the other hand, the scientific question being posed is critical

BMG744 1-30-12

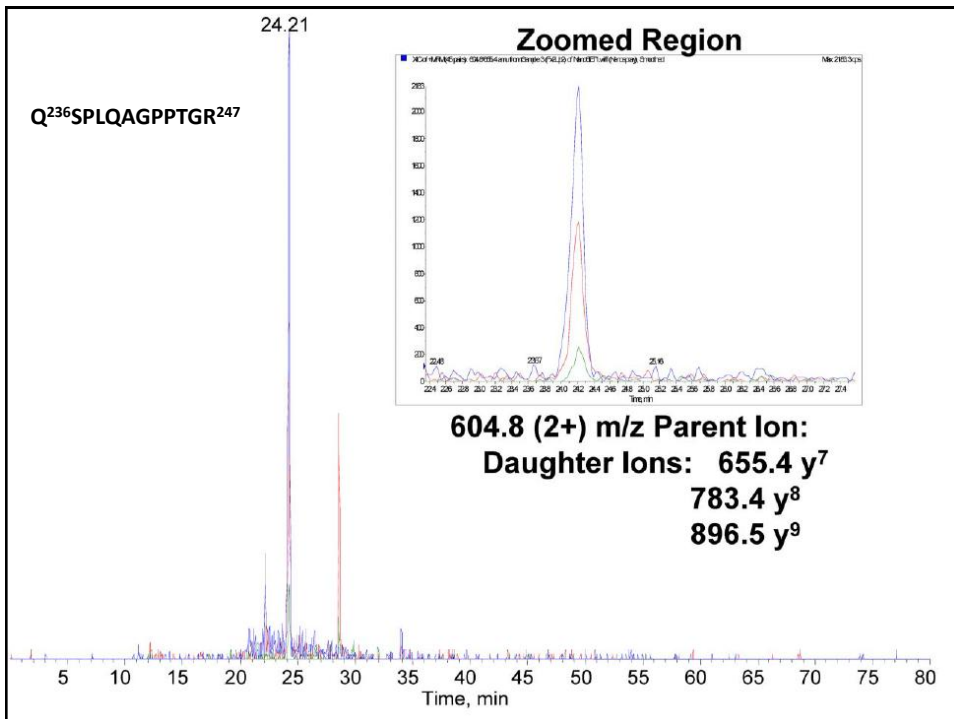
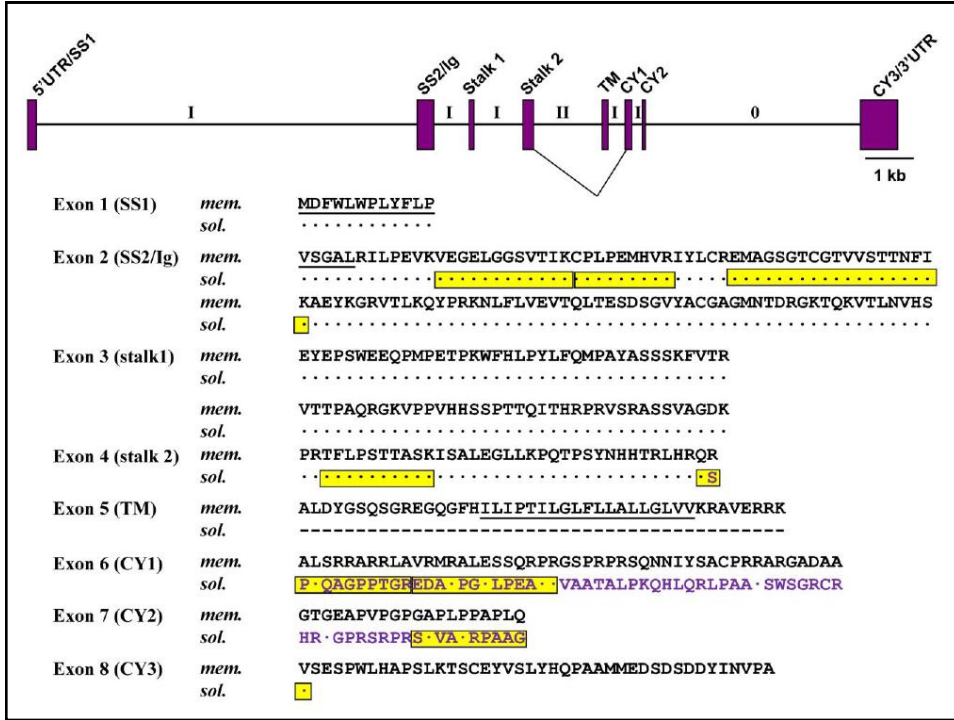
Soluble form of the FC μ R receptor

Enhanced levels of both the membrane-bound and soluble forms of IgM Fc receptor (Fc μ R) in patients with chronic lymphocytic leukemia

*Fu Jun Li,¹ *Yoshiki Kubagawa,² *Matthew K. McCollum,² Landon Wilson,³ Tomoko Motohashi,² Luigi F. Bertoli,⁴ James C. Barton,⁴ Stephen Barnes,³ Randall S. Davis,¹ and Hiromi Kubagawa²

In: Blood 2011;118:4902-4909

BMG744 1-30-12



Specificity of a peptide transition

- A BLAST search only tells us about sequence, not mass similarity
- Sherman et al. (2009) identified unique ion signature peptides
 - i.e., peptide molecular ions that give rise to fragment ions that cannot come from other peptides that pass through the mass filter (0.7 m/z wide) of a quadrupole analyzer set for the peptide of interest

BMG744 1-30-12

Testing peptide validity in mass space

ANDDLSEFPDK, $[M+2H]^{2+} = 682.3224$

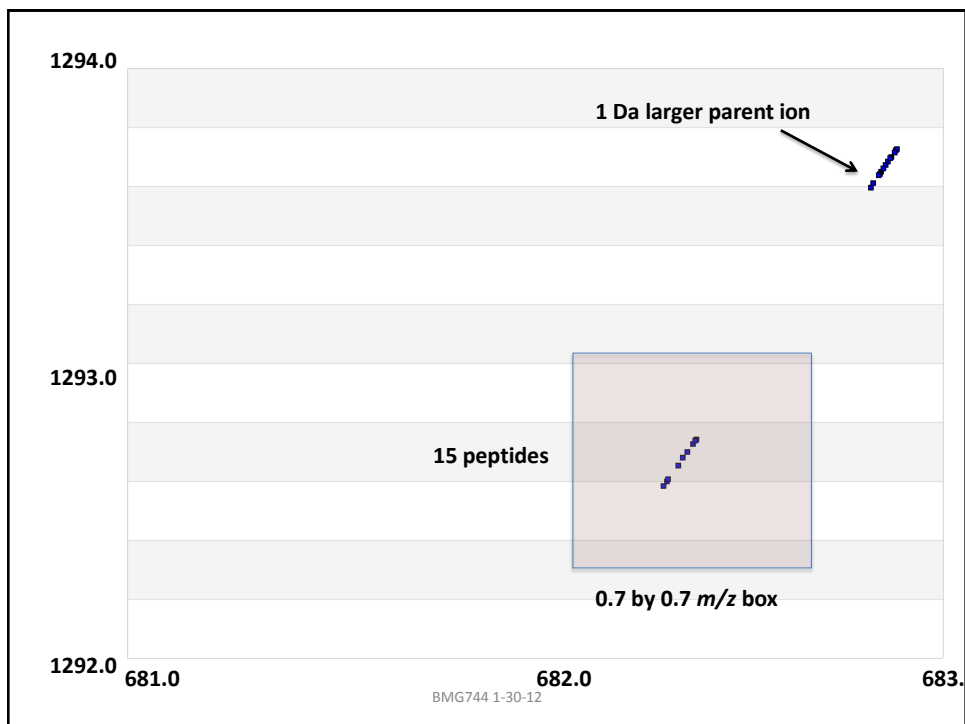
$Y_{11} = 1292.6005$

$Y_{10} = 1178.5576$

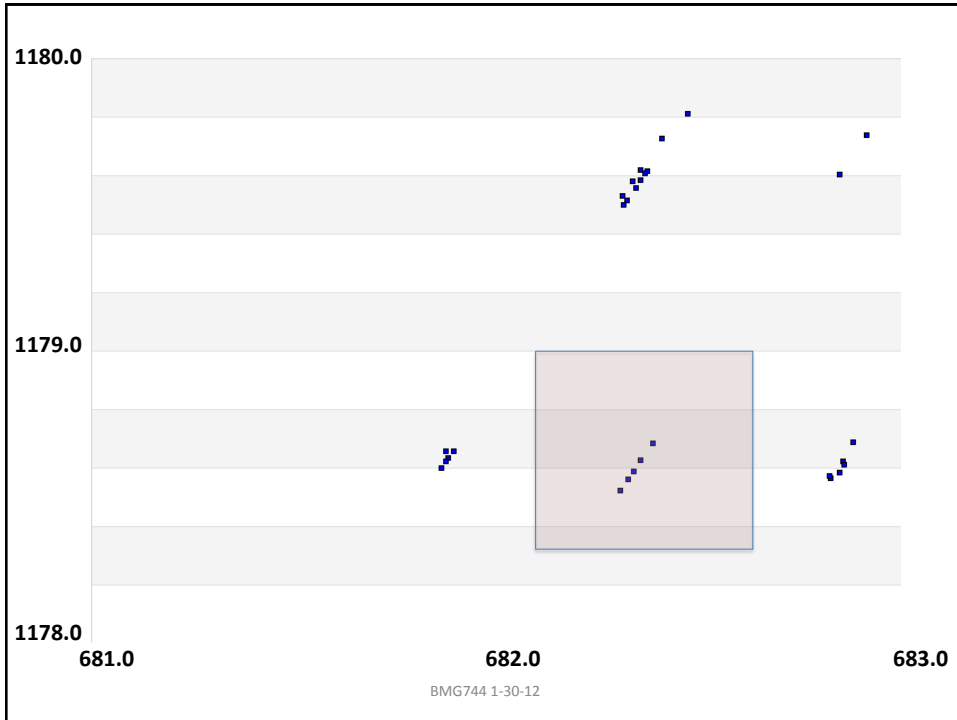
etc.

m/z mass min m/z mass max (x axis)
 ion mass min ion mass max (y axis)

BMG744 1-30-12



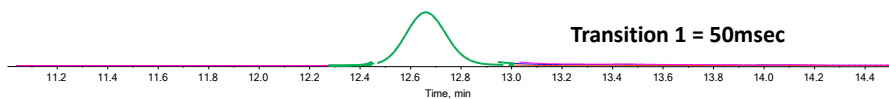
682.3148	1292.5847	ALDY YGLYDDR >sp Q8NDT2 RB15B_HUMAN Putative RNA-binding protein 15B OS=Homo sapiens GN=RBM15B PE=1 SV=3
682.3224	1292.5999	AIYYAWYEER >sp Q8TBY9 WDR66_HUMAN WD repeat-containing protein 66 OS=Homo sapiens GN=WDR66 PE=1 SV=2
682.3254	1292.6060	ANDDLLSEFPDK >sp P78352 DLG4_HUMAN Disks large homolog 4 OS=Homo sapiens GN=DLG4 PE=1 SV=3
682.3260	1292.6071	AFSTHAFSENPR >sp Q5TGY3 AHDC1_HUMAN AT-hook DNA-binding motif-containing protein 1 OS=Homo sapiens GN=AHDC1 PE=1 SV=1
682.3492	1292.6534	AADVAEALYSTPR >sp Q9BQW3 COE4_HUMAN Transcription factor COE4 OS=Homo sapiens GN=EBF4 PE=2 SV=2
682.3498	1292.6547	AQVPDTVFHHGR >sp Q9Y2G1 MRF_HUMAN Myelin gene regulatory factor OS=Homo sapiens GN=MRF PE=1 SV=3
682.3624	1292.6799	ADAALPVWPGGPR >sp Q3C1V9 YK041_HUMAN Putative uncharacterized protein ENSP00000334305 OS=Homo sapiens PE=5 SV=2
682.3730	1292.7010	APATPGAQLAPDVR >sp Q9NTN9 SEM4G_HUMAN Semaphorin-4G OS=Homo sapiens GN=SEMA4G PE=2 SV=1
682.3862	1292.7275	APVASVPPVHHPR >sp Q96EL1 C3orf54_HUMAN Uncharacterized protein C3orf54 OS=Homo sapiens GN=C3orf54 PE=2 SV=1
682.3855	1292.7261	ADPLHVALEVATK >sp Q9COH5 IRHG39_HUMAN Rho GTPase-activating protein 39 OS=Homo sapiens GN=ARHGAP39 PE=1 SV=2
682.3912	1292.7375	AGLGILHDIEGIR >sp Q9H4B0 OSGP2_HUMAN Probable O-sialoglycoprotein endopeptidase 2 OS=Homo sapiens GN=OSGEPL1 PE=2 SV=2
682.3912	1292.7375	AALVPTQAVPGSPR >sp P98095 FBLN2_HUMAN Fibulin-2 OS=Homo sapiens GN=FBLN2 PE=1 SV=2
682.3932	1292.7415	AQLP VV VFTFSR >sp Q15477 SKIV2_HUMAN Helicase SKI2W OS=Homo sapiens GN=SKIV2L PE=1 SV=3



682.3058	1178.5237	QQSSHYGQTDRl>sp Q6XPR3 RPTN_HUMAN Repetin OS=Homo sapiens GN=RPTN PE=1 SV=1
682.3254	1178.5629	ANDDLLSEFPDKl>sp P78352 DLG4_HUMAN Disks large homolog 4 OS=Homo sapiens GN=DLG4 PE=1 SV=3
682.3386	1178.5894	GQILGFWEERl>sp Q6NSX1 CCD70_HUMAN Coiled-coil domain-containing protein 70 OS=Homo sapiens GN=CCDC70 PE=2 SV=1
682.3568	1178.6257	NATALYHVEAFKl>sp Q9UNW1 MINP1_HUMAN Multiple inositol polyphosphate phosphatase 1 OS=Homo sapiens GN=MINPP1 PE=1 SV=1
682.3855	1178.6831	NALVSYSLVELRl>sp Q9UN72 PCDA7_HUMAN Protocadherin alpha-7 OS=Homo sapiens GN=PCDHA7 PE=1 SV=1

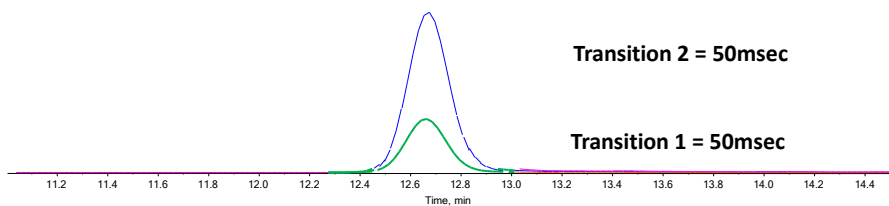
BMG744 1-30-12

Triple Quad vs TripleTOF



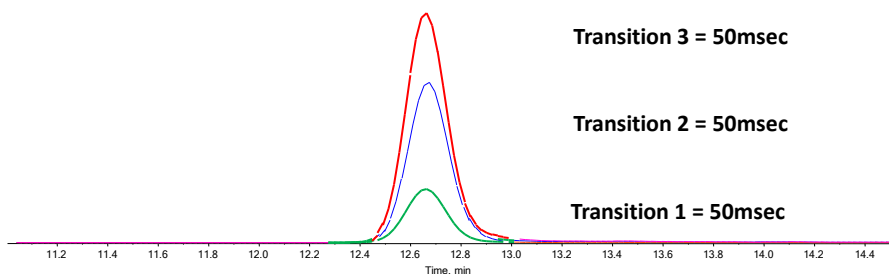
BMG744 1-30-12

Triple Quad vs TripleTOF



BMG744 1-30-12

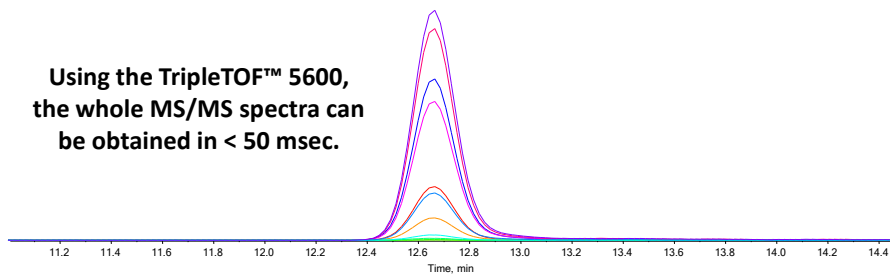
Triple Quad vs TripleTOF



BMG744 1-30-12

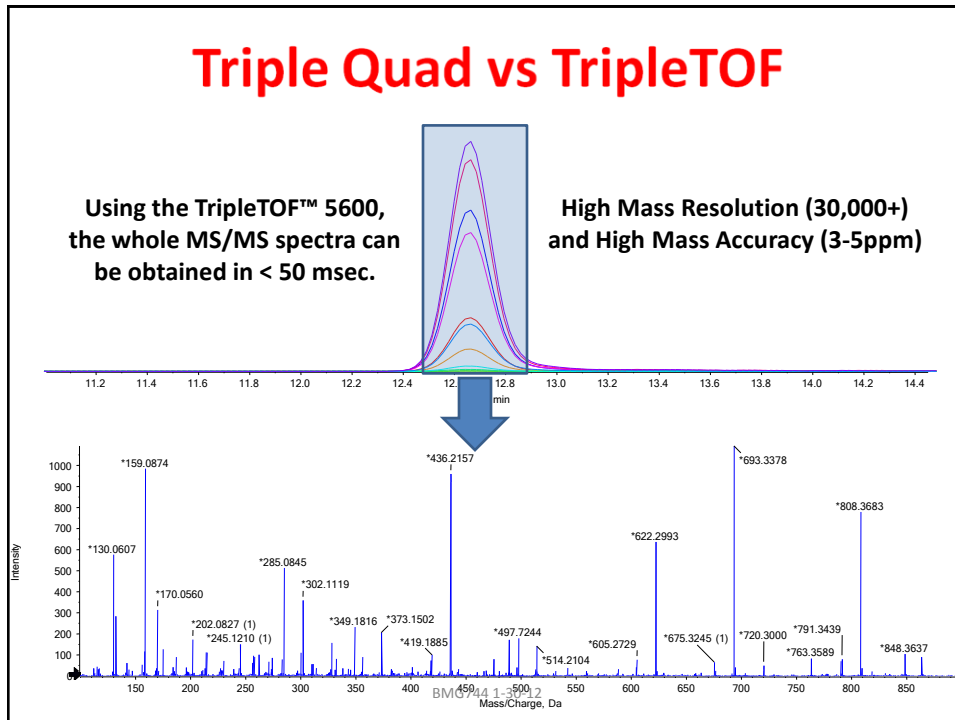
Triple Quad vs TripleTOF

Using the TripleTOF™ 5600,
the whole MS/MS spectra can
be obtained in < 50 msec.



BMG744 1-30-12

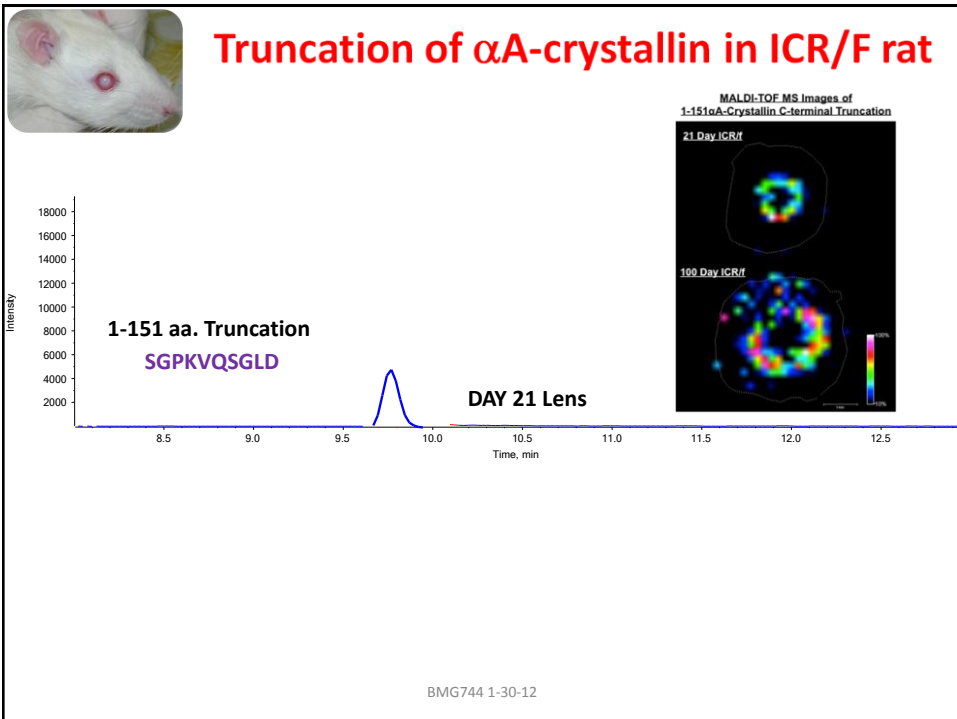
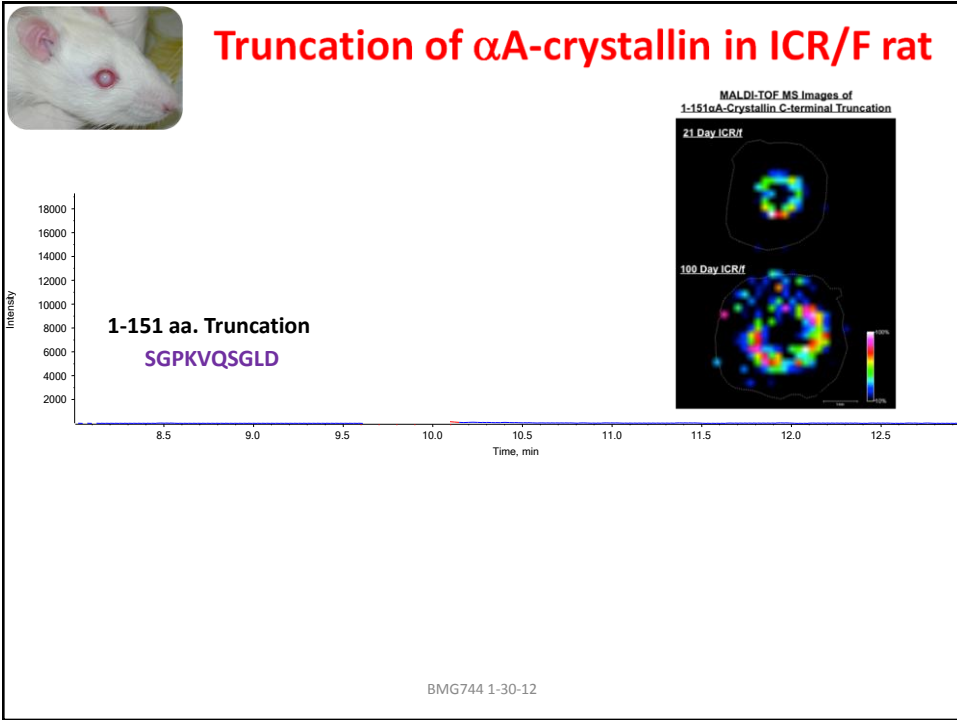
Triple Quad vs TripleTOF

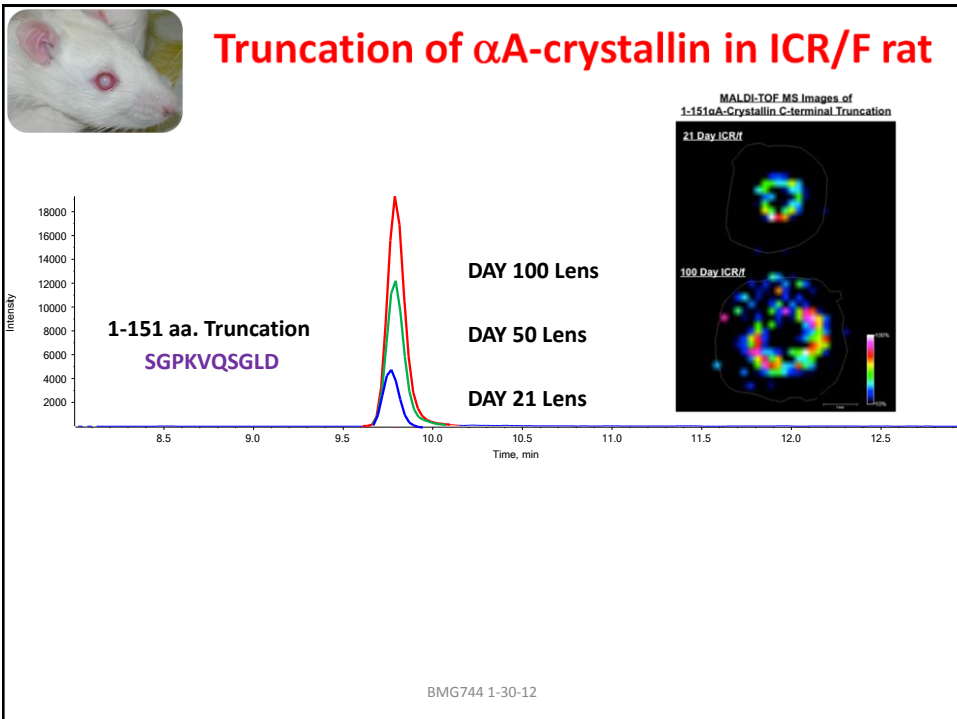
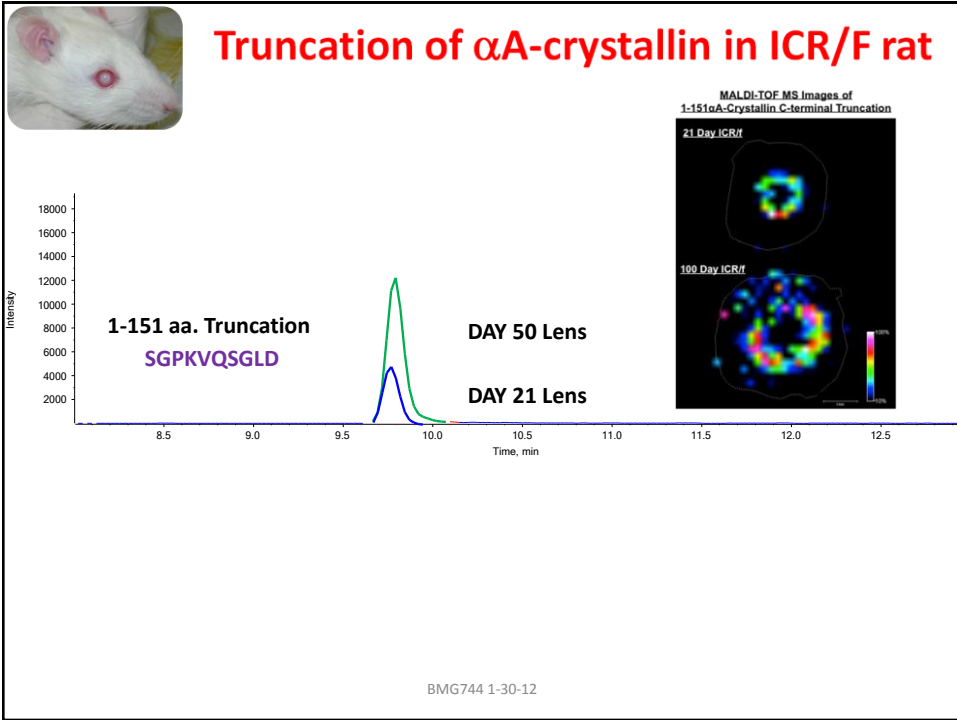


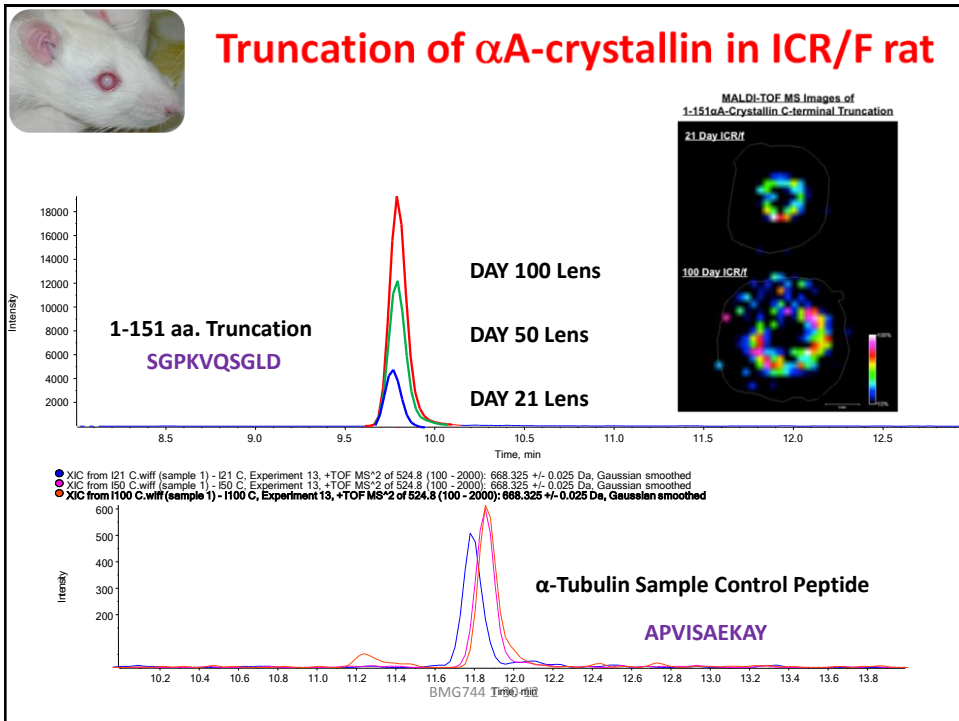
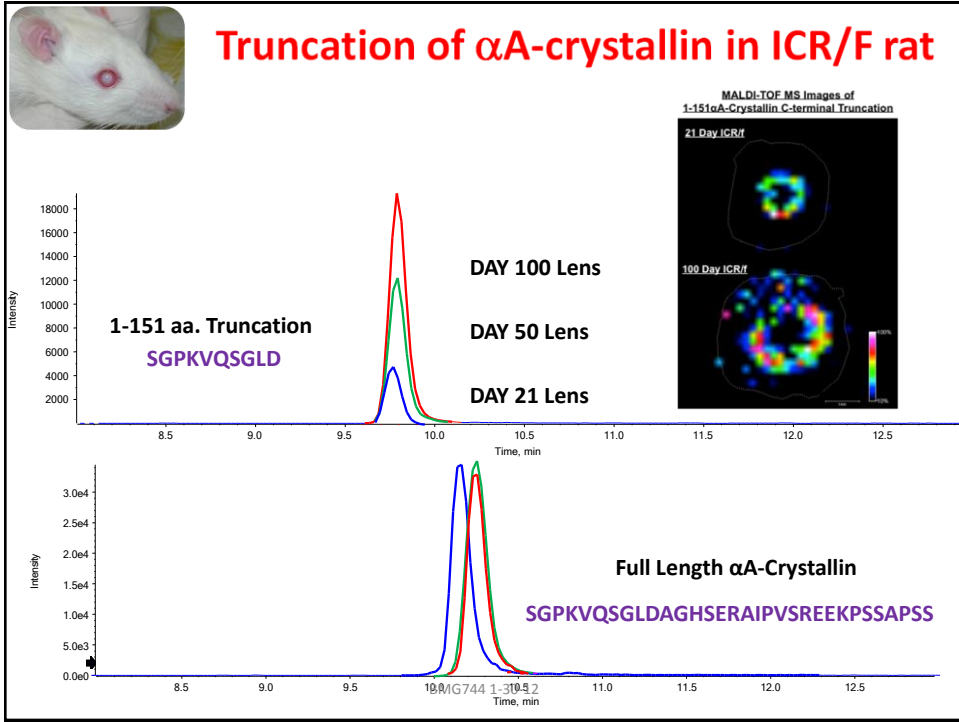
Verifying and quantifying C-terminal truncation

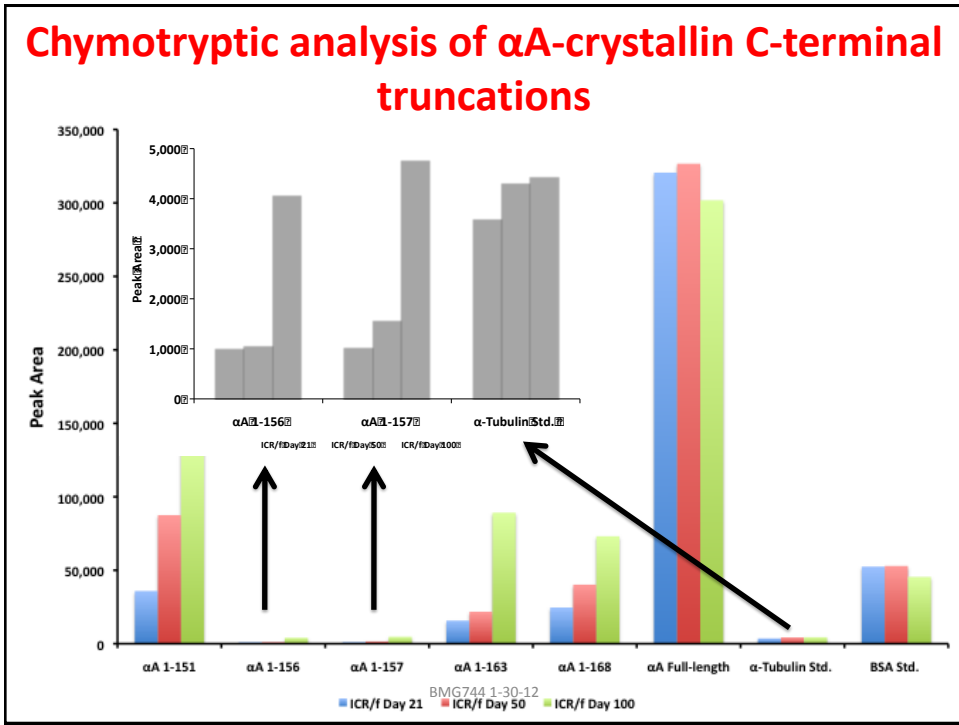
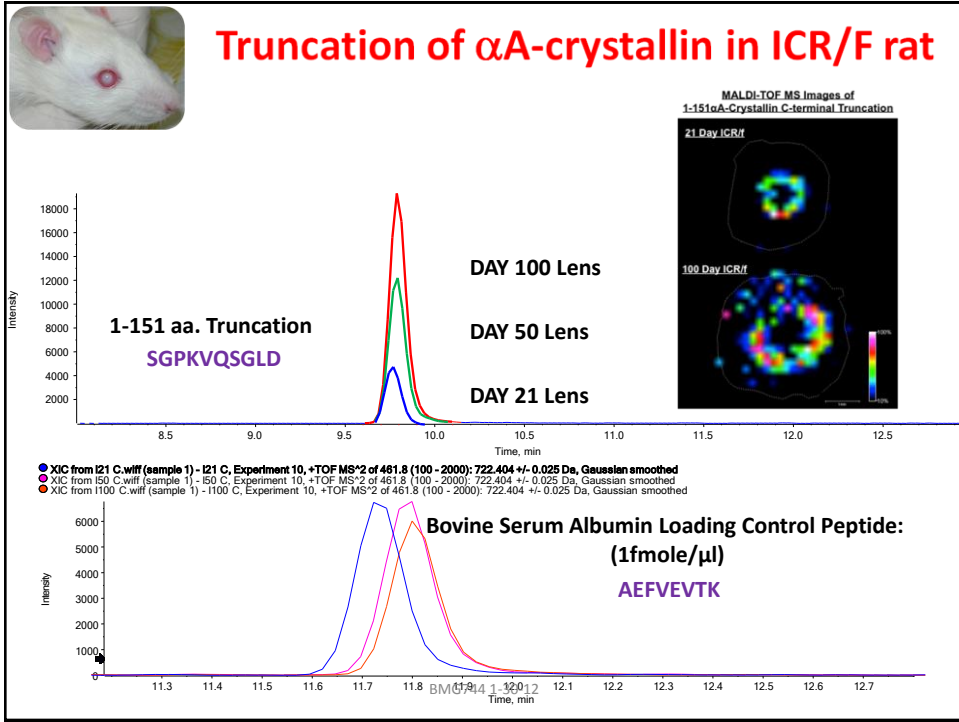
- In the rat full-length α A-crystallin is found endogenously at 173 amino acids. Previous MALDI-TOF Imaging and FT-ICR top-down MS experiments demonstrated the presence of multiple C-terminal truncations of the α A-crystallin.
- Full-length rat α A-crystallin has a chymotrypsin cleavage site at ^{141}Phe , which can be observed as an $[\text{M}+3\text{H}]^{3+}$ ion.
 - **FSGPKVQSGLDAGHSERAI PVSREEK PSSAPSS**
- Chymotryptic cleavages of C-terminal truncations:
 - **SGPKVQSGLD** (truncation at residue 151)
 - **SGPKVQSGLDAGHSE** (truncation at residue 156)
 - **SGPKVQSGLDAGHSER** (truncation at residue 157)
 - **SGPKVQSGLDAGHSERAI PVS R** (truncation at residue 163)
 - **SGPKVQSGLDAGHSERAI PVSREEK PS** (truncation at residue 168)

BMG744 1-30-12









NanoLC-MS of peptides and reproducibility of retention time

- Many people prepare their own capillary columns which sit between the nanoLC pump and the mass spectrometer
- The columns are subjected to the whims of of the packing procedure and of air conditioning in the mass spec laboratory

BMG744 1-30-12

Solutions to retention time variability

- Use machined columns on a Chip for reproducibility
- Controlled heating reduces solvent viscosity and hence back pressure
 - Leads to more rapid and reproducible retention times, and elution of hydrophobic peptides



Eksigent Nanoflex
and Chip column

BMG744 1-30-12

Reproducibility using the Eksigent-5600 system

