Targeted
Metabolomics &
Proteomics
Laboratory

# Taking a closer look at targeted MRM-MS

Stephen Barnes, PhD

MCLM 452; 4-7117

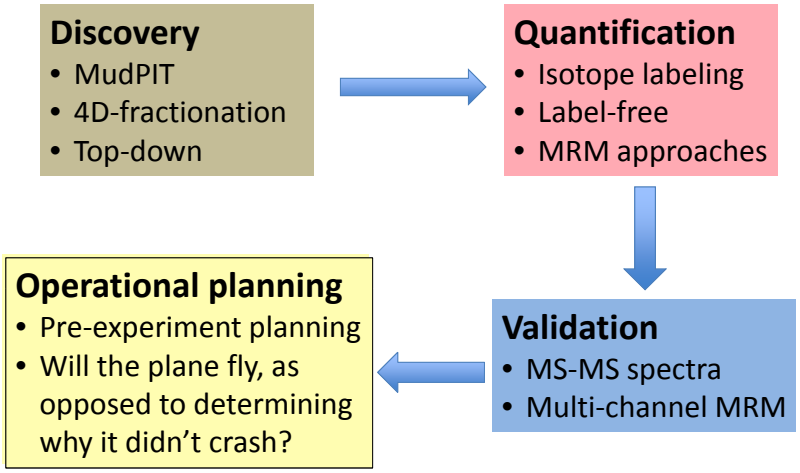
sbarnes@uab.edu



# 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 speed, high resolution, high mass accuracy analyzer

# Challenges in biological mass spectrometry






# 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**
  - **Others are a result of truncations, mRNA splicing**

# Selecting peptides for MRM

- Doing so empirically
  - Based on previously obtained MS/MS data on "your" mass spectrometer
- Using published/accumulated data
  - Skyline (optimizes for the most sensitive)
- Predicting suitable peptides
  - MRMPath/MRMutation (optimizes for biologically and informatically relevant)
- Carefully assessing mass space complexity
  - MRMSpace



# 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.NASHEQAAIALK.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.HCILDVSANAVR.R |
| 2563 | 682.32 | 1362.62 | 1362.63 | -0.01 | 0 | 95 | 6.9e-009 | 1 | R.ANDDLLSEFPDK.F |
| 2601 | 462.90 | 1385.67 | 1385.69 | -0.01 | 0 | 11 | 2.3 | 1 | K.FGSCVPHTTRPK.R |
| 2715 | 505.28 | 1512.81 | 1512.83 | -0.01 | 1 | 62 | 2e-005 | 1 | R.KGDQILSVNGVDLR.N |
| 2737 | 513.59 | 1537.76 | 1537.77 | -0.01 | 1 | 32 | 0.021 | 1 | K.DLLGEEDIPREPR.R |

MASCOT PROTEIN SUMMARY REPORT

**IIPGGAAAQDGR**

B L A S T

| Accession | Description | Max score | Total score | Query coverage | E value |
|---|---|---|---|---|---|
| gi\|4557529\|NP_001356.1 | disks large homolog 4 isoform 1 [Homo sapiens] >gi\|3318653\|gb\|AAC | 38.8 | 61.9 | 100% | 5e-07 |
| gi\|5918874\|AAD56173.1 | post-synaptic density 95 [Homo sapiens] | 38.8 | 61.9 | 100% | 5e-07 |
| gi\|119610659\|EAW90253.1 | discs, large homolog 4 (Drosophila), isoform CRA_c [Homo sapiens] > | 38.8 | 61.9 | 100% | 5e-07 |
| gi\|71658825\|P78352.3 | RecName: Full=Disks large homolog 4; AltName: Full=Postsynaptic de | 38.8 | 61.9 | 100% | 5e-07 |
| gi\|192447426\|NP_001122299.1 | disks large homolog 4 isoform 2 [Homo sapiens] >gi\|119610661\|gb\|E | 38.8 | 61.9 | 100% | 5e-07 |
| gi\|221041302\|BAH12328.1 | unnamed protein product [Homo sapiens] | 38.8 | 61.9 | 100% | 5e-07 |
| gi\|221041762\|BAH12558.1 | unnamed protein product [Homo sapiens] | 38.8 | 61.9 | 100% | 5e-07 |
| gi\|119610658\|EAW90252.1 | discs, large homolog 4 (Drosophila), isoform CRA_b [Homo sapiens] > | 38.8 | 61.9 | 100% | 5e-07 |
| gi\|119610657\|EAW90251.1 | discs, large homolog 4 (Drosophila), isoform CRA_a [Homo sapiens] | 38.8 | 61.9 | 100% | 5e-07 |
| gi\|1527215\|AAB07736.1 | PSD-95 [Homo sapiens] | 38.8 | 61.9 | 100% | 5e-07 |
| gi\|2613006\|AAB84250.1 | Tax interaction protein 15 [Homo sapiens] >gi\|119610660\|gb\|EAW90 | 38.8 | 61.9 | 100% | 5e-07 |
| gi\|159162599\|1KEF_A | Chain A, Pdz1 Of Sap90 | 38.8 | 38.8 | 100% | 5e-07 |
| gi\|218156338\|NP_001136171.1 | disks large homolog 2 isoform 1 [Homo sapiens] | 35.8 | 59.0 | 100% | 6e-06 |
| gi\|51491229\|CAH18680.1 | hypothetical protein [Homo sapiens] | 35.8 | 59.0 | 100% | 6e-06 |
| gi\|332164718\|NP_001193698.1 | disks large homolog 2 isoform 5 [Homo sapiens] | 35.8 | 59.0 | 100% | 6e-06 |
| gi\|119595499\|EAW75093.1 | discs, large homolog 2, chapsyn-110 (Drosophila), isoform CRA_a [Hc | 35.8 | 59.0 | 100% | 6e-06 |
| gi\|91199538\|NP_001355.2 | disks large homolog 2 isoform 2 [Homo sapiens] >gi\|215274165\|sp\|Q | 35.8 | 59.0 | 100% | 6e-06 |
| gi\|11463026\|AAB04949.1 | channel associated protein of synapse [Homo sapiens] | 35.8 | 59.0 | 100% | 6e-06 |
| gi\|221040586\|BAH11970.1 | unnamed protein product [Homo sapiens] | 35.8 | 59.0 | 100% | 6e-06 |
| gi\|221040860\|BAH12131.1 | unnamed protein product [Homo sapiens] | 35.8 | 59.0 | 100% | 6e-06 |
| gi\|221039974\|BAH11750.1 | unnamed protein product [Homo sapiens] | 35.8 | 59.0 | 100% | 6e-06 |

BMG744 1-31-14

# Pragmatic selection of a peptide

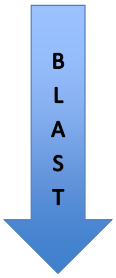
DLG4_HUMAN Mass: 80788 Score: 388 Queries matched: 18 emPAI: 0.68
Disks large homolog 4 OS=Homo sapiens GN=DLG4 PE=1 SV=3

| 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.NASHEQAAIALK.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.HCILDVSANAVR.R |
| 2563 | 682.32 | 1362.62 | 1362.63 | -0.01 | 0 | 95 | 6.9e-009 | 1 | R.ANDDLLSEFPDK.F |
| 2601 | 462.90 | 1385.67 | 1385.69 | -0.01 | 0 | 11 | 2.3 | 1 | K.FGSCVPHTTRPK.R |
| 2715 | 505.28 | 1512.81 | 1512.83 | -0.01 | 1 | 62 | 2e-005 | 1 | R.KGDQILSVNGVDLR.N |
| 2737 | 513.59 | 1537.76 | 1537.77 | -0.01 | 1 | 32 | 0.021 | 1 | K.DLLGEEDIPREPR.R |

**MASCOT PROTEIN SUMMARY REPORT**

BMG744 1-31-14

**ANDDLLSEFPDK**




| Accession | Description | Max score | Total score | Query coverage | E value |
|---|---|---|---|---|---|
| gi\|4557529\|NP_001356.1 | disks large homolog 4 isoform 1 [Homo sapiens] >gi\|3318653\|qb\|AAC | 41.4 | 41.4 | 100% | 7e-08 |
| gi\|5918874\|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\|71658825\|P78352.3 | RecName: Full=Disks large homolog 4; AltName: Full=Postsynaptic de | 41.4 | 41.4 | 100% | 7e-08 |
| gi\|192447426\|NP_001122299.1 | disks large homolog 4 isoform 2 [Homo sapiens] >gi\|119610661\|qb\|E | 41.4 | 41.4 | 100% | 7e-08 |
| gi\|221041302\|BAH12328.1 | unnamed protein product [Homo sapiens] | 41.4 | 41.4 | 100% | 7e-08 |
| gi\|221041762\|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\|73909118\|AAH40533.1 | DLG4 protein [Homo sapiens] | 41.4 | 41.4 | 100% | 7e-08 |
| gi\|218156338\|NP_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\|148539578\|NP_004078.2 | disks large homolog 1 isoform 2 [Homo sapiens] >gi\|119573995\|qb\|E | 35.4 | 35.4 | 91% | 9e-06 |
| gi\|558436\|AAA50598.1 | homolog of Drosophila discs large protein, isoform 2 [Homo sapiens] | 35.4 | 35.4 | 91% | 9e-06 |
| gi\|332164718\|NP_001193698.1 | disks large homolog 2 isoform 5 [Homo sapiens] | 35.4 | 35.4 | 91% | 9e-06 |
| gi\|148539628\|NP_001091894.1 | disks large homolog 1 isoform 1 [Homo sapiens] >gi\|223590196\|sp\|Q | 35.4 | 35.4 | 91% | 9e-06 |

# Selecting peptides

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

## 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 be optimal – on the other hand, the scientific question being posed is critical



## Searching pathways - MRMPath

- Proteins rarely operate all on their own, but rather in pathways or groups
- MRMPath 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

# Use of MRMPath

http://tmpl.uab.edu/MRMPath/

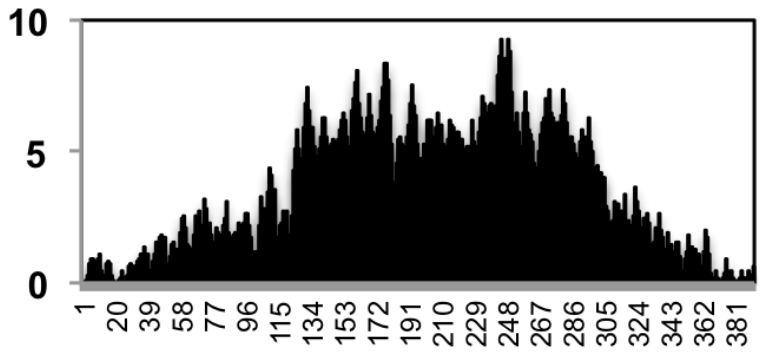
**Users select the protease they want to use, the pathway of interest and the species in which the research was carried out**

- **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**



# MRMutation

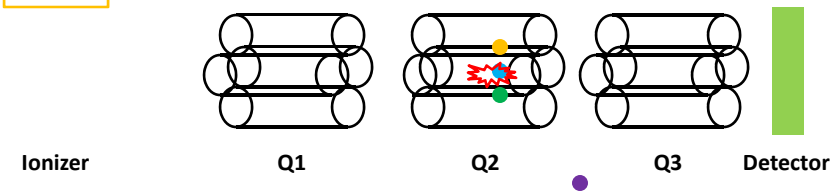
## Are there canonical protein sequences??




**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.**

# Multiple reaction ion monitoring

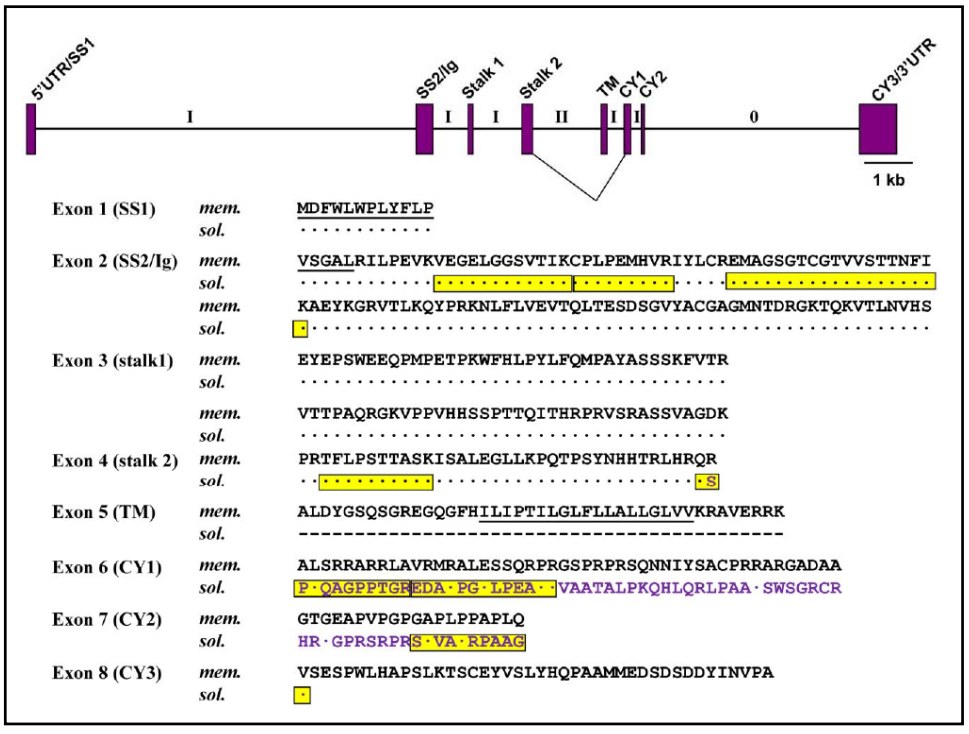

**Quantitative analysis of peptides in a complex mixture carried out using a triple quadrupole instrument**

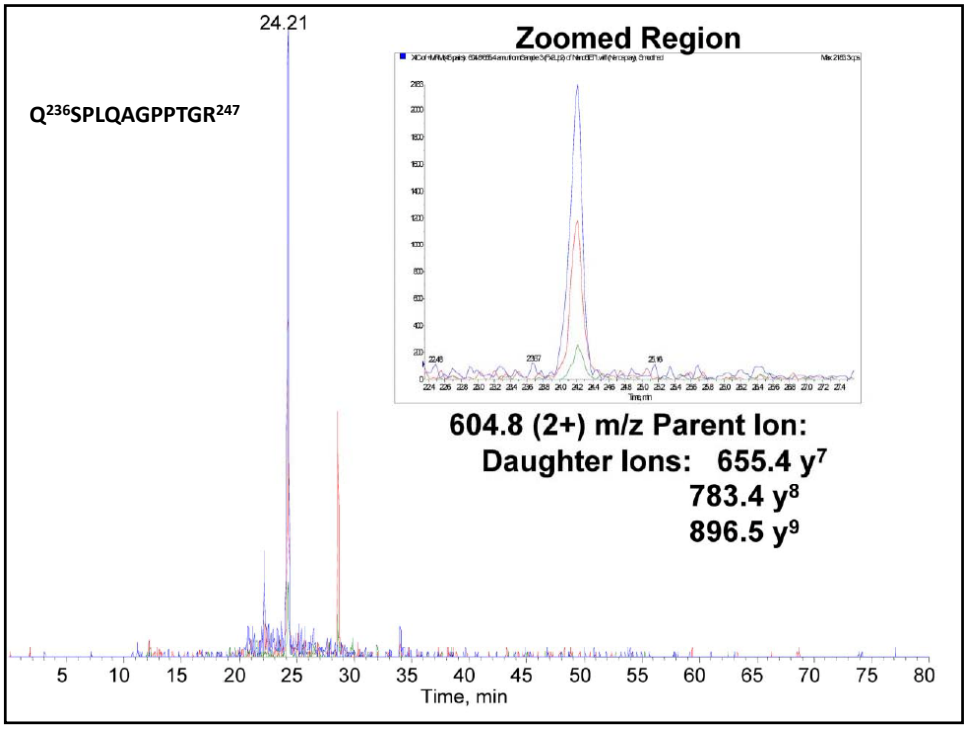



**Based on precursor ion/product ion pair(s)**

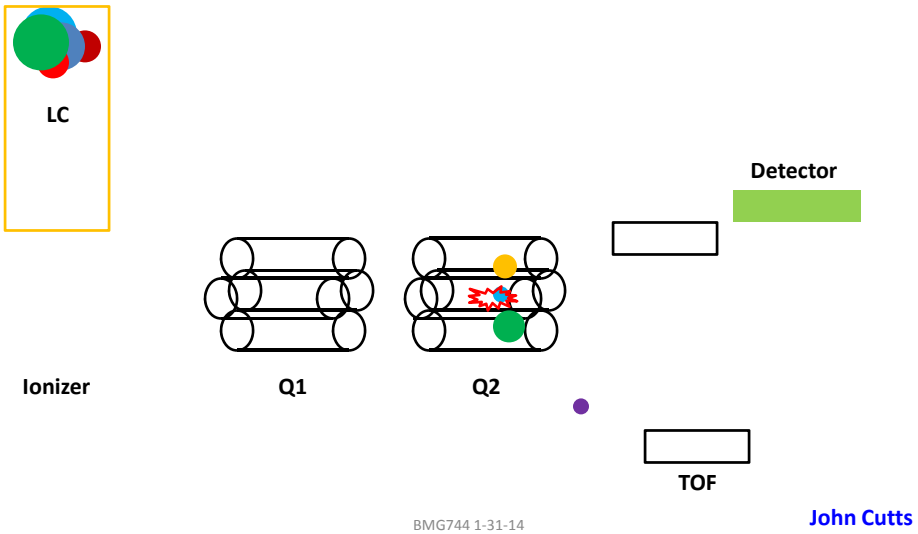


*Courtesy, John Cutts*





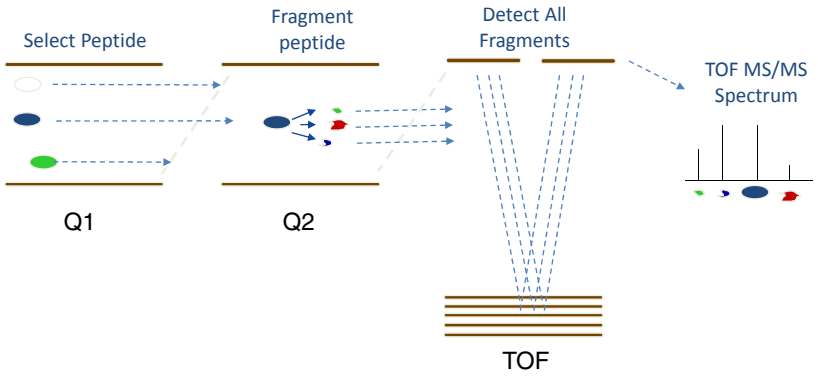

# Liquid Chromatography-Triple TOF Mass Spectrometry (LC-QQTOF MS)




BMG744 1-31-14

John Cutts

# Pseudo MRM Analysis



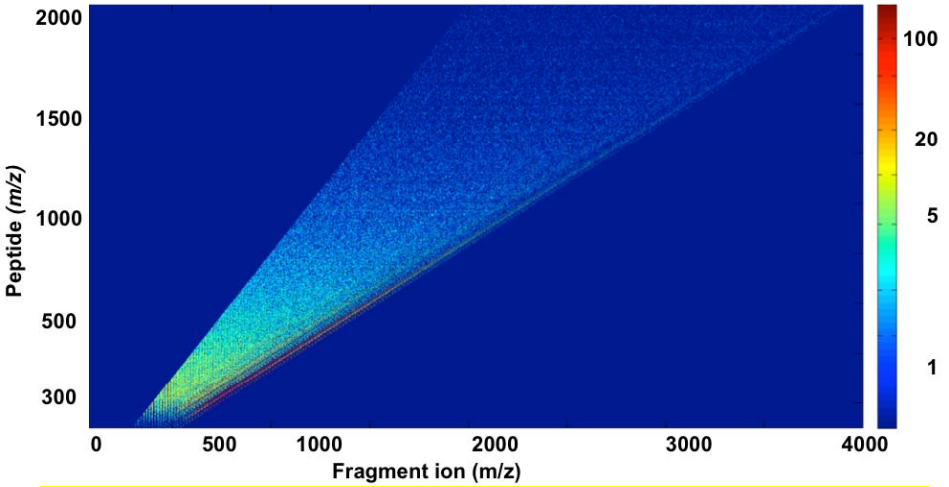

- **The key difference between the TripleTOF and the triple quad is that the entire MSMS spectrum is collected by the TripleTOF in a single 50 sec (or shorter) data acquisition – the selection of product ions is made post-data acquisition**
- **The mass accuracy of the product ions is 3-5 ppm**



# 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**

# How complex is the peptidome "soup"?




**"Theoretic" tryptidome of the human proteome for unmodified proteins has 3,883,436 b-ions and 4,230,819 y-ions**



# Goals

- To generate a web-based visual tool to evaluate the complexity of mass space for tryptic peptides derived from the human proteome
  - Building on the work of Jamie Sherman et al. (2009)
- In the region where the precursor ion and product ions for a "proteotypic" peptide have been chosen, to identify the names of other proteins that would satisfy the mass criteria
- To evaluate regions of mass uniqueness and the impact of reducing the size of the mass window

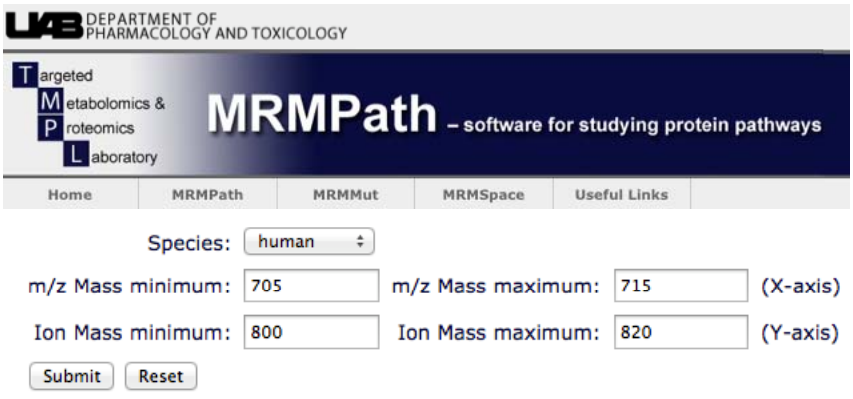
# General procedure

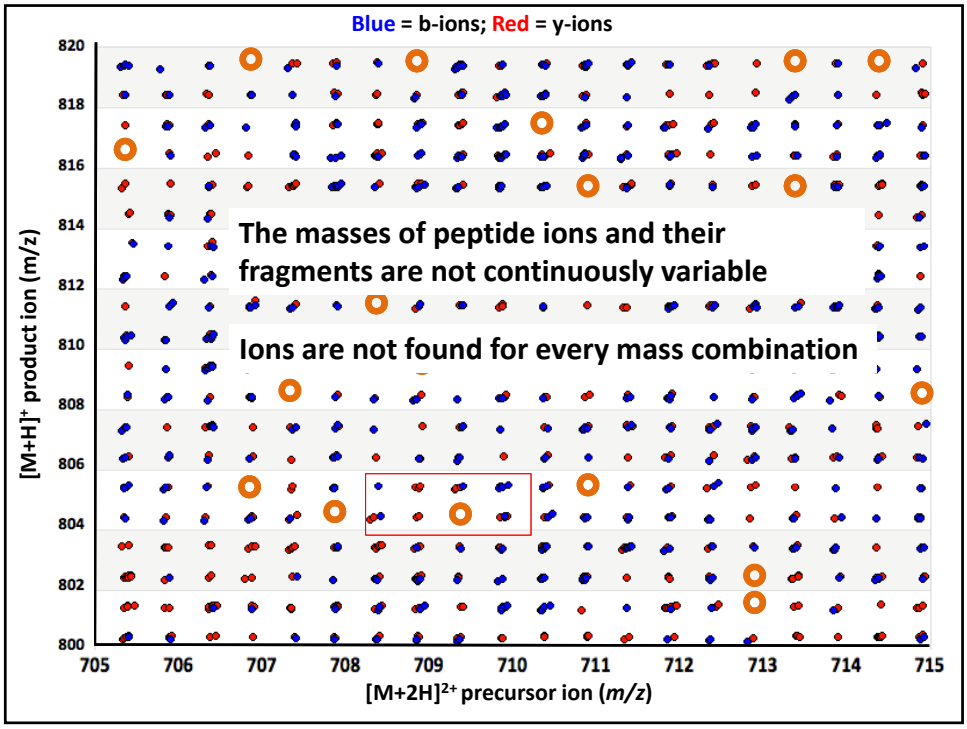
- We have data mined the Expasy.org/Uniprot web site to automatically
  - Download the amino acid sequences of the curated human proteome
  - Extract the known amino acid variants for each protein (e.g., human p53 has 1361 different amino acid mutations out of 393 aa residues)
  - Extract the curated posttranslational modifications and the amino acids on which they are located
- Develop web-based tools to inspect the data

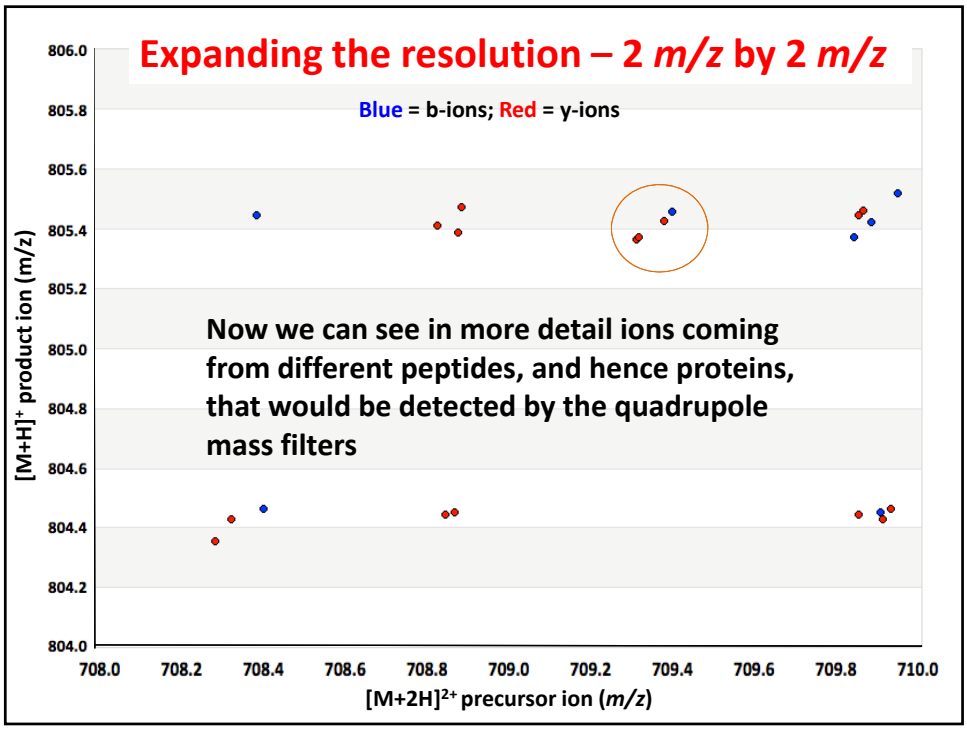


# Let's look at a small region of mass space

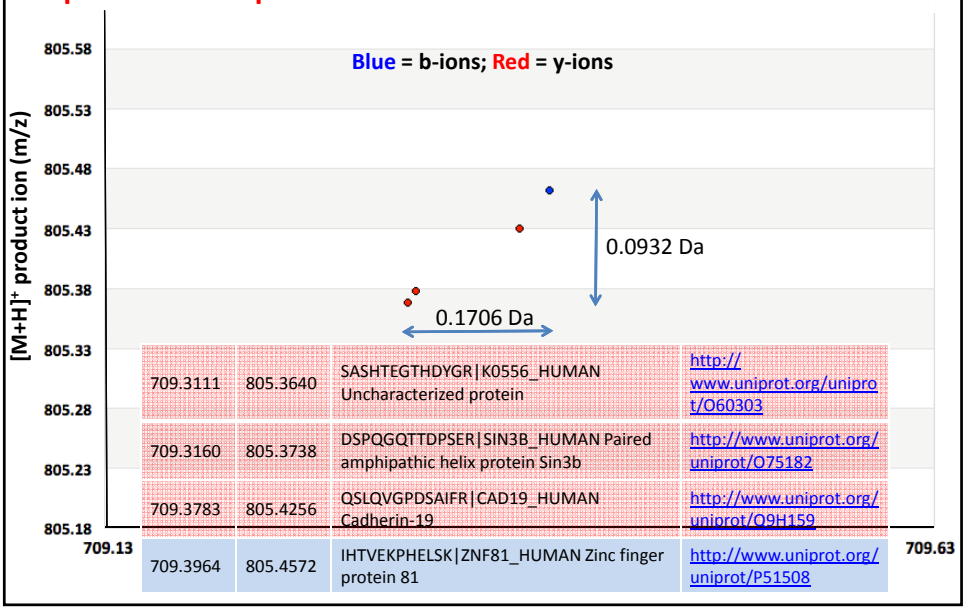
http://tmpl.uab.edu/MRMPath




Blue = b-ions; Red = y-ions
The masses of peptide ions and their fragments are not continuously variable
Ions are not found for every mass combination
[M+H]+ product ion (m/z)
[M+2H]2+ precursor ion (m/z)
820
818
816
814
812
810
808
806
804
802
800
705
706
707
708
709
710
711
712
713
714
715

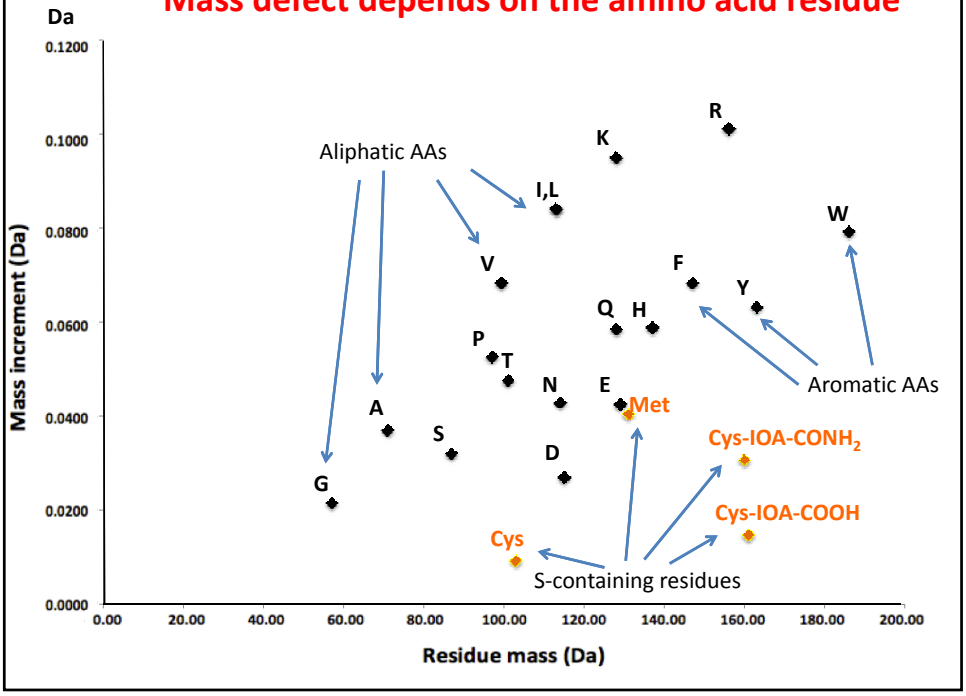

Expanding the resolution – 2 m/z by 2 m/z
Blue = b-ions; Red = y-ions
Now we can see in more detail ions coming from different peptides, and hence proteins, that would be detected by the quadrupole mass filters
[M+H]+ product ion (m/z)
[M+2H]2+ precursor ion (m/z)
806.0
805.8
805.6
805.4
805.2
805.0
804.8
804.6
804.4
804.2
804.0
708.0
708.2
708.4
708.6
708.8
709.0
709.2
709.4
709.6
709.8
710.0

## Maximum mass resolution (0.5 *m/z* by 0.5 *m/z*) reveals that the precursor and product ions are distributed over 100-200 mDa




| | | | |
|---|---|---|---|
| 709.3111 | 805.3640 | SASHTEGTHDYGR\|K0556_HUMAN Uncharacterized protein | http://www.uniprot.org/uniprot/O60303 |
| 709.3160 | 805.3738 | DSPQGQTTDPSER\|SIN3B_HUMAN Paired amphipathic helix protein Sin3b | http://www.uniprot.org/uniprot/O75182 |
| 709.3783 | 805.4256 | QSLQVGPDSAIFR\|CAD19_HUMAN Cadherin-19 | http://www.uniprot.org/uniprot/Q9H159 |
| 709.3964 | 805.4572 | IHTVEKPHELSK\|ZNF81_HUMAN Zinc finger protein 81 | http://www.uniprot.org/uniprot/P51508 |

## Mass defect depends on the amino acid residue



## Ions not to use in MRM experiments

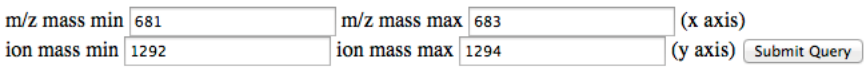
- $y_1$ ions are very redundant
  - For human tryptic peptides passing through the 710.8 *m/z* quadrupole filter, there are 90 that have Lys and 78 that have an Arg C-terminal residue
- The same applies to $b_{n-1}$ ions since they also result from the loss of Lys or Arg C-terminal residues
- $y_2$ and $b_{n-2}$ ions are also highly redundant, as are $y_{n-1}$, $y_{n-2}$ and $b_2$ ions

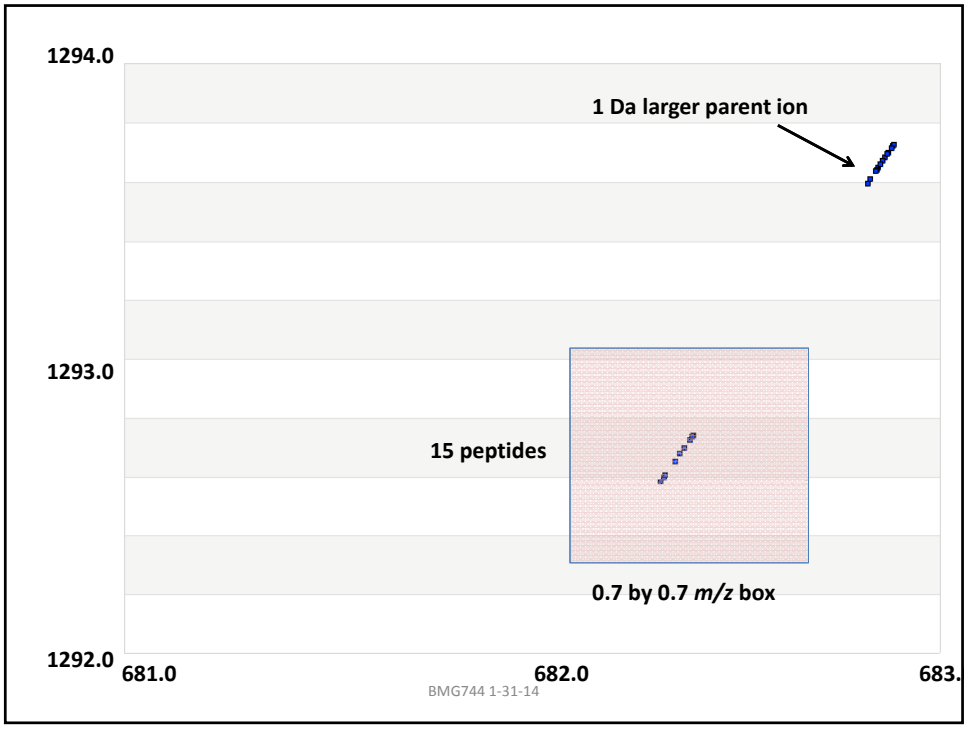


## Testing peptide validity in mass space

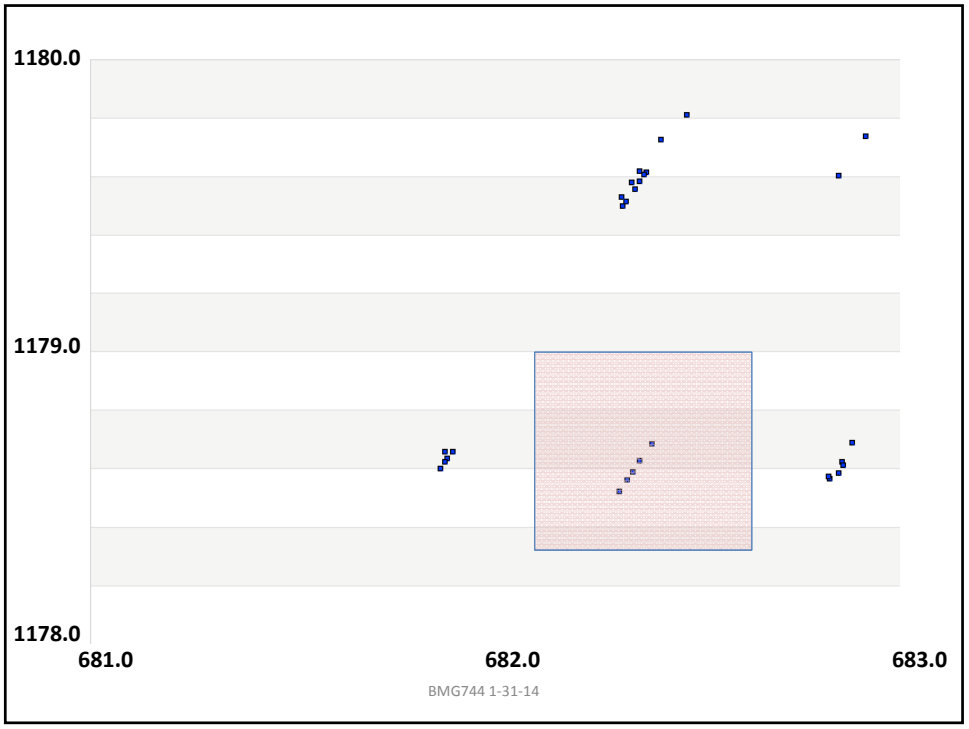
**ANDDLLSEFPDK, $[M+2H]^{2+}$ = 682.3224**

**$Y_{11}$ = 1292.6005**
**$Y_{10}$ = 1178.5576**
**etc.**





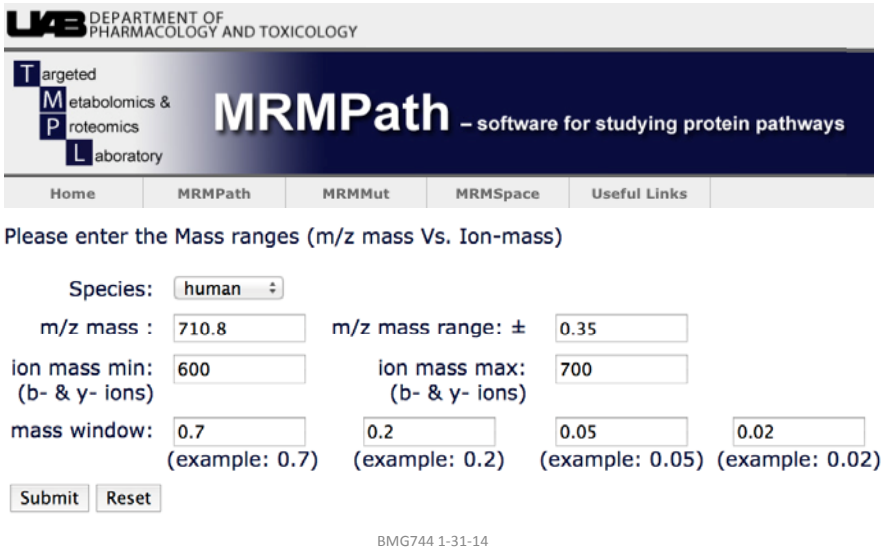

| | | |
|---|---|---|
| 682.3148 | 1292.5847 | ALDYYGLYDDR\|>sp\|Q8NDT2\|RB15B_HUMAN Putative RNA-binding protein 15B OS=Homo sapiens GN=RBM15B PE=1 SV=3 |
| 682.3224 | 1292.5999 | AIYYAWYEER\|>sp\|Q8TBY9\|WDR66_HUMAN WD repeat-containing protein 66 OS=Homo sapiens GN=WDR66 PE=1 SV=2 |
| 682.3254 | 1292.6060 | 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 | ADAALPVWPGGPGR\|>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\|CC054_HUMAN Uncharacterized protein C3orf54 OS=Homo sapiens GN=C3orf54 PE=2 SV=1 |
| 682.3855 | 1292.7261 | ADPLHVALEVATK\|>sp\|Q9C0H5\|RHG39_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 | AQLPVVVFTFSR\|>sp\|Q15477\|SKIV2_HUMAN Helicase SKI2W OS=Homo sapiens GN=SKIV2L PE=1 SV=3 |




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

BMG744 1-31-14

# Assessing uniqueness and redundancy using MRMSpace




BMG744 1-31-14

## Narrowing the mass window helps to distinguish different peptides

**human proteome - precursor (m/z) (710.8 ± 0.35 )**

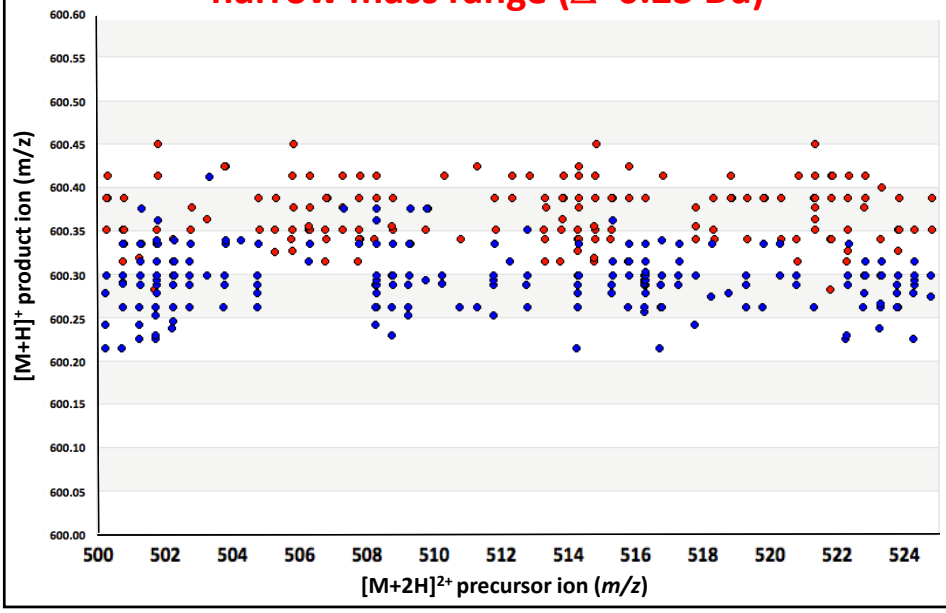
| product ion min | product ion max | Number of y-ions singles | Number of b-ions singles | Total number of y-ions | Total number of b-ions |
|---|---|---|---|---|---|
| **mass window : 0.7** | | | | | |
| 600 | 649.7 | 18 | 14 | 83 | 71 |
| 649.7 | 699.4 | 17 | 17 | 75 | 80 |
| **mass window : 0.2** | | | | | |
| 600 | 650 | 18 | 16 | 83 | 71 |
| 650 | 700 | 21 | 17 | 75 | 80 |
| **mass window : 0.05** | | | | | |
| 600 | 650 | 37 | 39 | 83 | 71 |
| 650 | 700 | 29 | 49 | 75 | 80 |
| **mass window : 0.02** | | | | | |
| 600 | 650 | 57 | 46 | 83 | 71 |
| 650 | 700 | 53 | 56 | 75 | 80 |

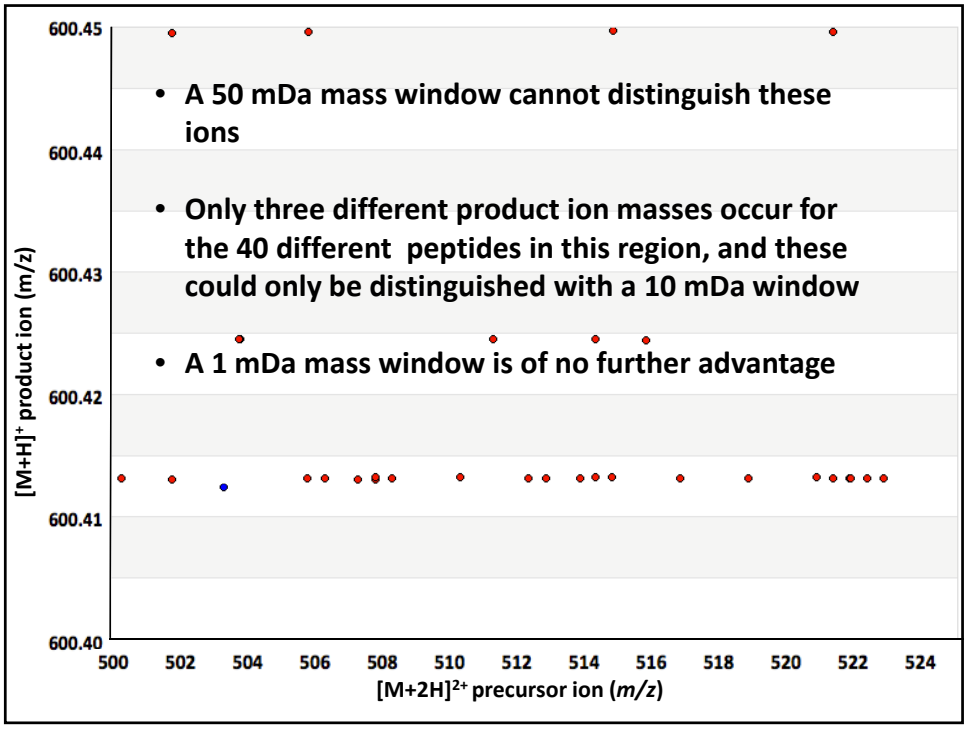
# SWATH – fast, deep MSMS

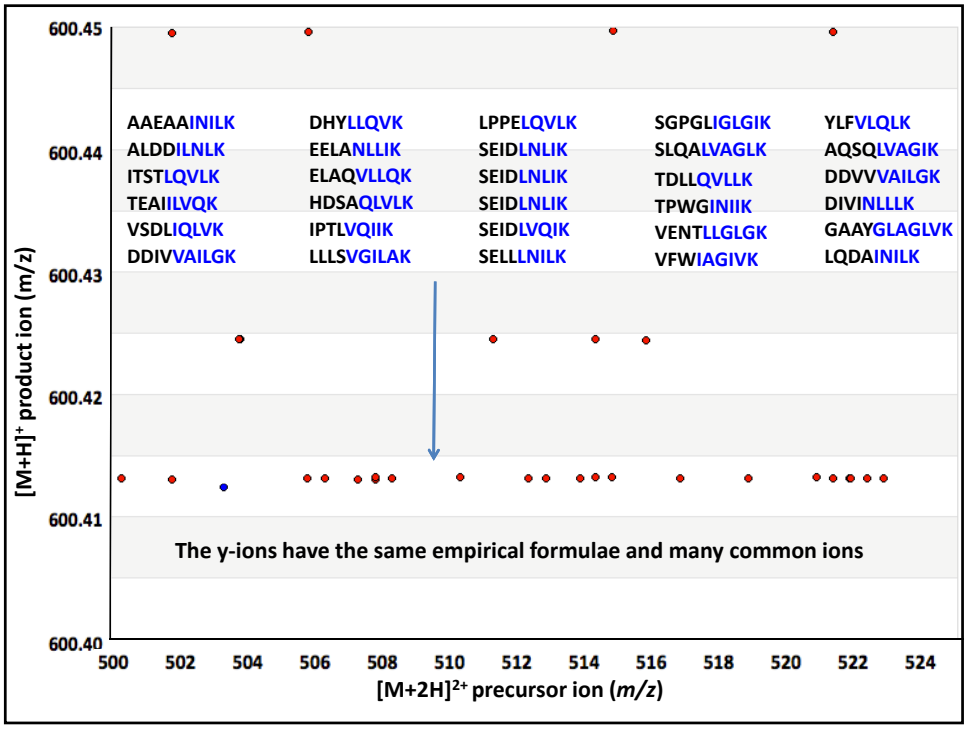
- This approach was described last year by Rudi Aebersold at a user meeting at ASMS where a 25 *m/z* mass window was used
  - Published in Mol Cell Proteomics (Gillet et al.)
  - All ions passing through the window were fragmented
  - If we select *m/z* 500-525 for the precursor ion window and *m/z* 600-600.7 for the product ion window, there are b- and y-ions from 470 proteins



## The ions associated with parent ions are in a narrow mass range (Δ=0.23 Da)




• A 50 mDa mass window cannot distinguish these ions
• Only three different product ion masses occur for the 40 different peptides in this region, and these could only be distinguished with a 10 mDa window
• A 1 mDa mass window is of no further advantage
600.45
600.44
600.43
600.42
600.41
600.40
[M+H]+ product ion (m/z)
500 502 504 506 508 510 512 514 516 518 520 522 524
[M+2H]2+ precursor ion (m/z)

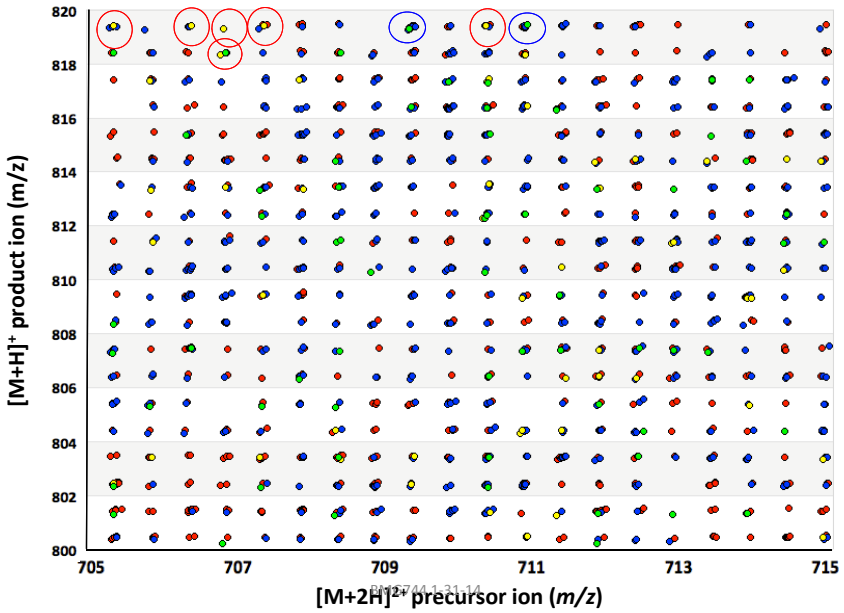

AAEAAINILK
ALDDILNLK
ITSTLQVLK
TEAIILVQK
VSDLIQLVK
DDIVVAILGK
DHYLLQVK
EELANLLIK
ELAQVLLQK
HDSAQLVLK
IPTLVQIIK
LLLSVGILAK
LPPELQVLK
SEIDLNLIK
SEIDLNLIK
SEIDLNLIK
SEIDLVQIK
SELLLNILK
SGPGLIGLGIK
SLQALVAGLK
TDLLQVLLK
TPWGINIIK
VENTLLGLGK
VFWIAGIVK
YLFVLQLK
AQSQLVAGIK
DDVVVAILGK
DIVINLLLK
GAAYGLAGLVK
LQDAINILK
The y-ions have the same empirical formulae and many common ions
600.45
600.44
600.43
600.42
600.41
600.40
[M+H]+ product ion (m/z)
500 502 504 506 508 510 512 514 516 518 520 522 524
[M+2H]2+ precursor ion (m/z)

# Effects of PTMs on mass space

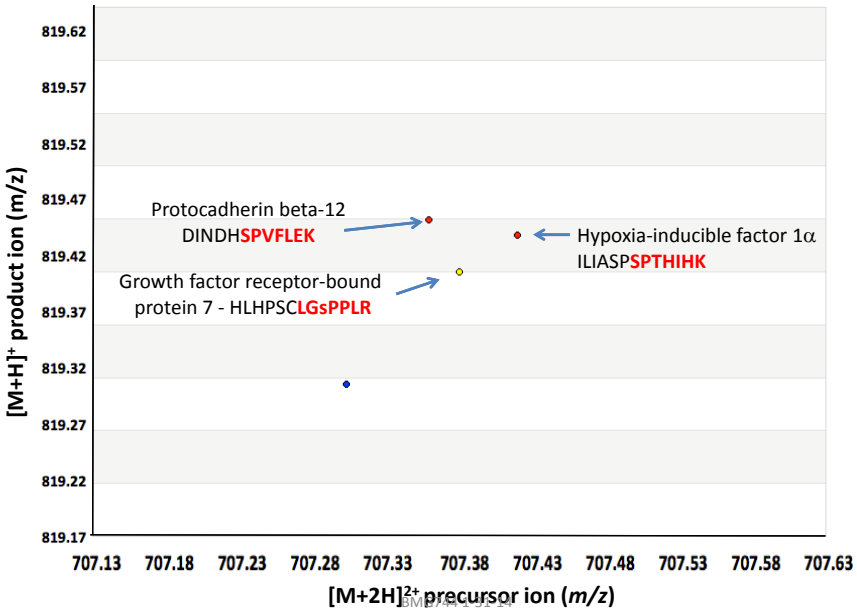
- ExPaSy.org lists 477 known modifications
  - We deleted 126 for our analysis since they were crosslinks and would be expected to substantially increase the mass of the peptide
  - Known sites of modification were added to the database



## Addition of PTMs into mass space – yellow and green circles



## Phosphorylation of GRFP-7 interferes with HIF-1α




BMG744 1-31-14

## Increasers of mass space complexity

- **Chemical**
  - Losses of $H_2O$, $NH_3$ and CO, $^{13}C$-isotope peaks, Multiply charged ions
- **Biochemical**
  - RP and KP sequences, RR, RK, KR and KK missed cleavages
  - A(cet)ylation of lysine groups, Deamidation of peptides
- **Biological**
  - ✧ Cleavages (these predominate in the lens) to produce subpeptides
  - ✧ Sequence variations in somatic tissues – NEXTGen deep sequencing and RNA-seq will overwhelm existing databases

BMG744 1-31-14

# Decreases in complexity

- **Absence of b- or y-ions**
  - Need to integrate the database with known fragmentation patterns for your instrument or instrument types
  - The abundance of the protein
- **Chromatography**
  - Better resolution will speed up the overall analysis and reduce need for stabilization



# Summary

- We have created web browser accessible software to inspect mass space of the human tryptidome
- Mass space for the human tryptidome in MRM-style experiments is very complex, but is not random since it has structure
- Redundancy is the highest for the smallest and largest product ions from a tryptic peptide – these should be avoided
- Narrowing the mass window for the analysis of product ions decreases redundancy
  - However, little is gained in going from 0.7 *m/z* to 0.2 *m/z*

## Conclusions

- MRMSpace reveals that low mass resolution triple quadrupole instruments have distinct limitations for analysis of complex samples from the human tryptidome
  - Presence/absence of b- and y-ions in the MSMS spectrum (would reduce complexity)
  - Differential mRNA splicing and the mutations expected in cancer and even normal (pre-disease) subjects (would increase complexity)



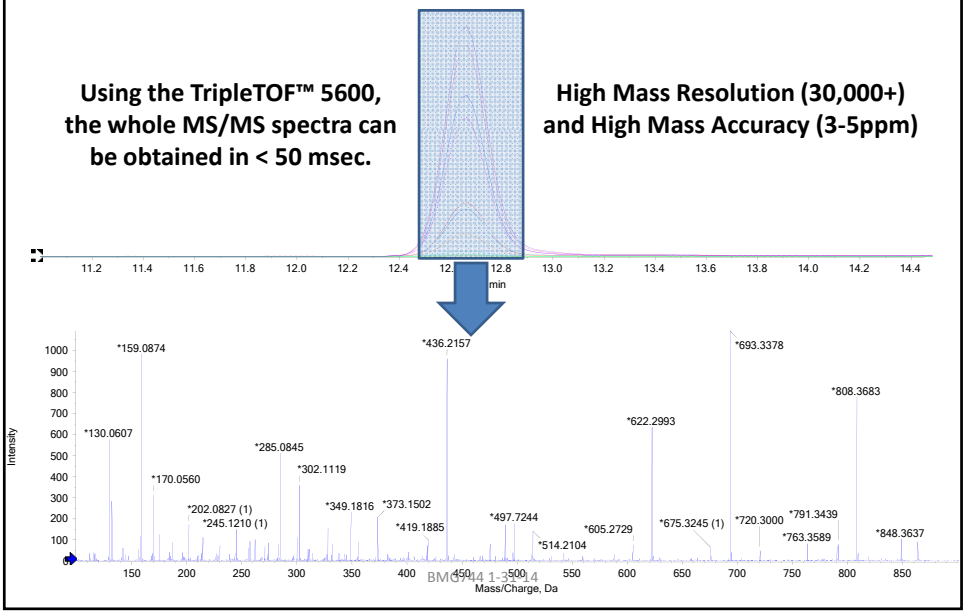
## What do we need?

- It's critical to collect the whole MSMS spectrum in one go, not just 3-4 channels of precursor ion/product ion pairs
- The MS/MS spectrum should be collected with the same sensitivity as ion pairs on a triple quadrupole instrument and in the same time-frame (let's say 10 msec – that would allow analysis of 100 peptides per sec)

# Triple Quad vs TripleTOF

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

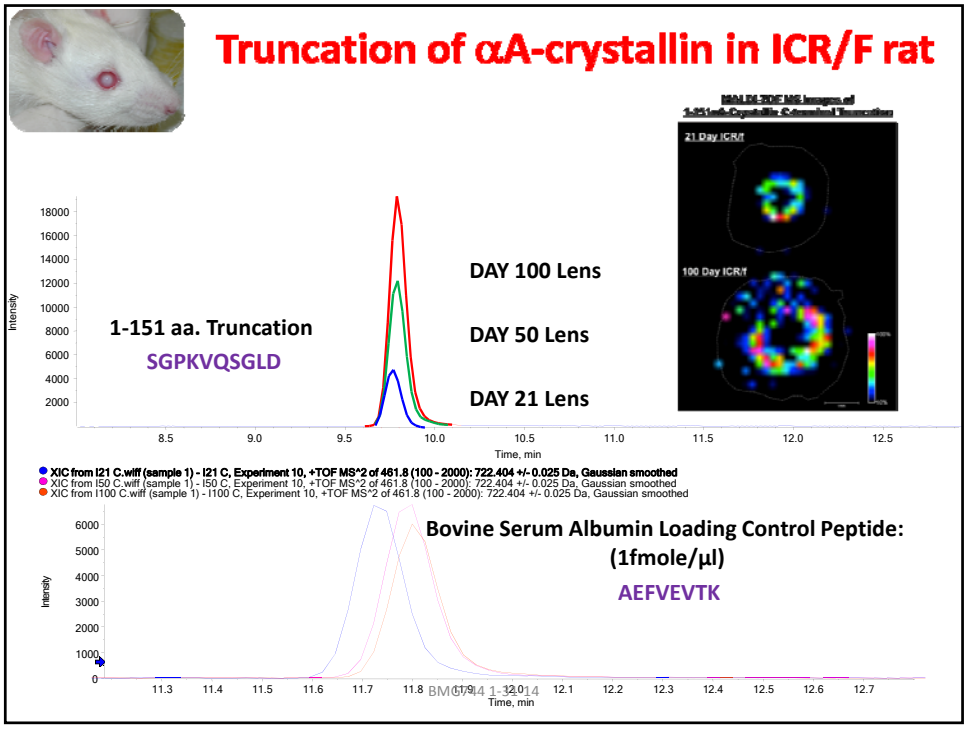
High Mass Resolution (30,000+) and High Mass Accuracy (3-5ppm)

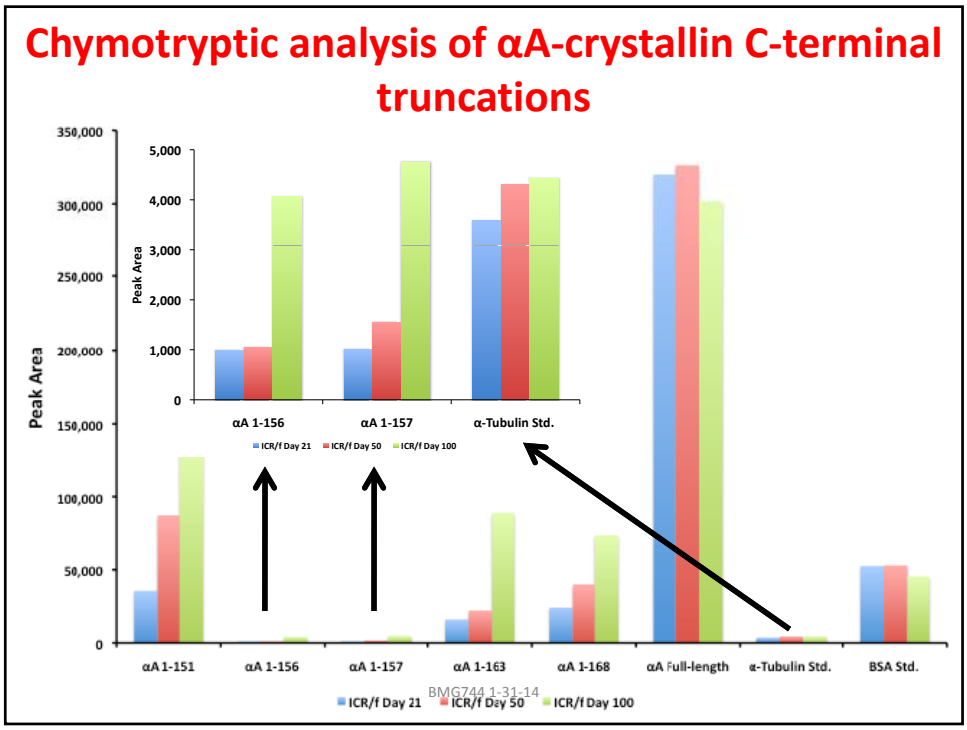





# Verifying and quantifying C-terminal truncation

- In the rat full-length αA-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 αA-crystallin.
- Full-length rat αA-crystallin has a chymotrypsin cleavage site at $^{141}$Phe, which can be observed as an $[M+3H]^{3+}$ ion.
  - **F**SGPKVQSGLDAGHSERAIPVSREEKPSSAPSS
- Chymotryptic cleavages of C-terminal truncations:
  - SGPKVQSGLD (truncation at residue 151)
  - SGPKVQSGLDAGHSE (truncation at residue 156)
  - SGPKVQSGLDAGHSER (truncation at residue 157)
  - SGPKVQSGLDAGHSERAIPVSR (truncation at residue 163)
  - SGPKVQSGLDAGHSERAIPVSREEKPS (truncation at residue 168)


Truncation of αA-crystallin in ICR/F rat
21 Day ICR/f
100 Day ICR/f
DAY 100 Lens
DAY 50 Lens
DAY 21 Lens
1-151 aa. Truncation
SGPKVQSGLD
Intensity
Time, min
XIC from I21 C.wiff (sample 1) - I21 C, Experiment 10, +TOF MS^2 of 461.8 (100 - 2000): 722.404 +/- 0.025 Da, Gaussian smoothed
XIC from I50 C.wiff (sample 1) - I50 C, Experiment 10, +TOF MS^2 of 461.8 (100 - 2000): 722.404 +/- 0.025 Da, Gaussian smoothed
XIC from I100 C.wiff (sample 1) - I100 C, Experiment 10, +TOF MS^2 of 461.8 (100 - 2000): 722.404 +/- 0.025 Da, Gaussian smoothed
Bovine Serum Albumin Loading Control Peptide:
(1fmole/µl)
AEFVEVTK
BMG744 1-31-14


Chymotryptic analysis of αA-crystallin C-terminal truncations
Peak Area
αA 1-156
αA 1-157
α-Tubulin Std.
ICR/f Day 21
ICR/f Day 50
ICR/f Day 100
αA 1-151
αA 1-156
αA 1-157
αA 1-163
αA 1-168
αA Full-length
α-Tubulin Std.
BSA Std.
BMG744 1-31-14

## 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



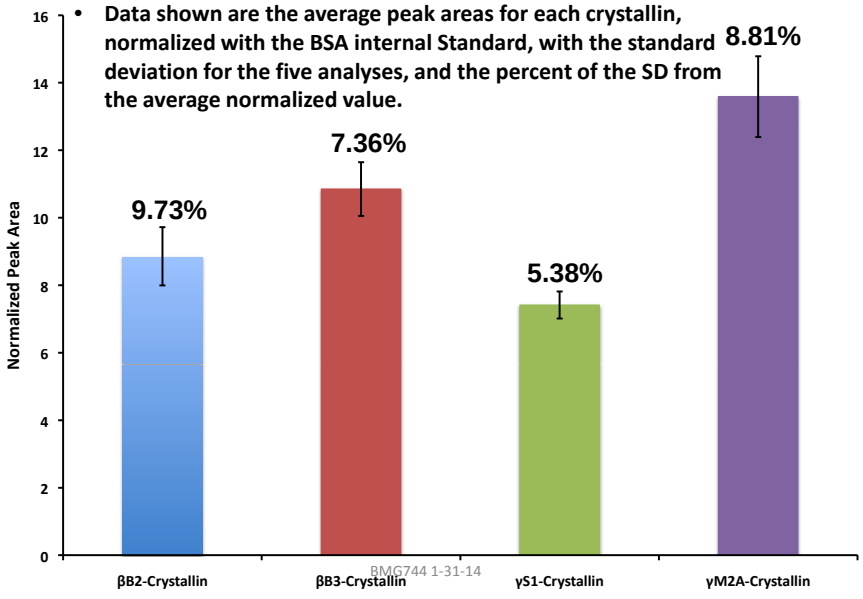
## Solutions to retention time variability

- 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

# Reproducibility using the Eksigent-5600 system

- Five injections of the same sample, from zebrafish diet study.
- Data shown are the average peak areas for each crystallin, normalized with the BSA internal Standard, with the standard deviation for the five analyses, and the percent of the SD from the average normalized value.




BMG744 1-31-14