

Meeting Increased Pressures to Detect and Quantify Host Cell Proteins in Biotherapeutics by LC/MS

Sean M McCarthy, PhD

- Residual host cell proteins (HCPs) from recombinant production are classified as process-related impurities.
 - HCPs can elicit an unpredictable immune response in patients.
 - Need to be monitored as part of regulatory guidelines
- European regulations in effect since 1997:
 - '6.2 Validation of the purification procedure - The ability of the purification process to remove other specific contaminants such as [host-cell proteins](#) ... should also be demonstrated'
- Current analytical methods (typically ELISA, gels, blots) are:
 - expensive,
 - subjective,
 - time-consuming to develop,
 - Require prior knowledge about the contaminant proteins
- Composition of HCPs is extremely heterogeneous and changes with production and purification procedures.



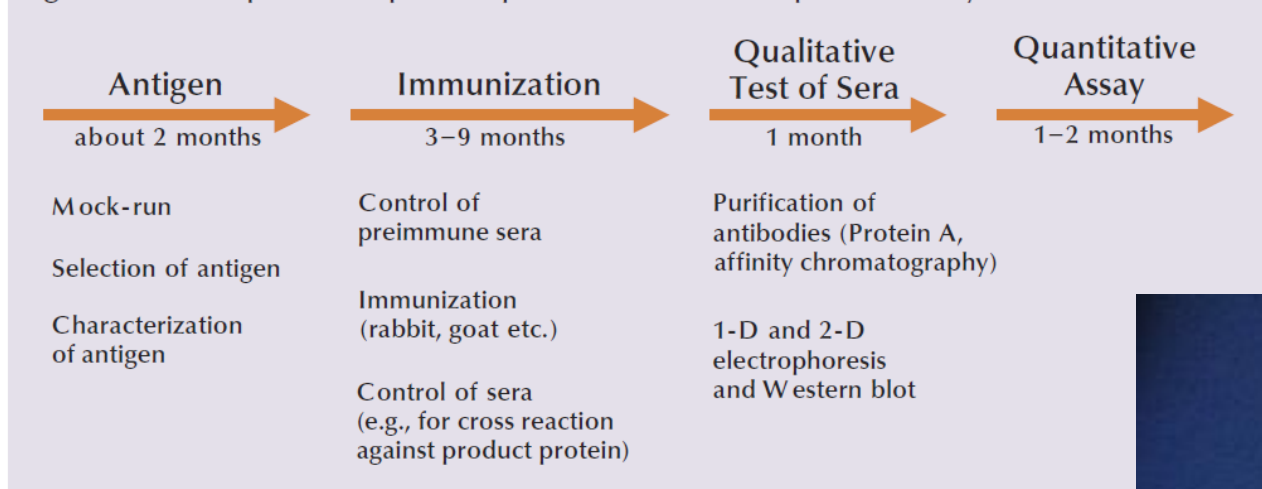
- During the development of Omnitrope, an immunogenicity issue emerged with an early version of the product.
- Up to 60% of patients enrolled in two clinical studies developed anti-hGH antibodies, and 100% developed anti-HCP antibodies.
- The cause was excess host cell protein levels, which was resolved by the manufacturer with purification process changes.

What is the cost of halting clinical trials or losing drug efficacy over time?



Immunoassays are sensitive but inflexible

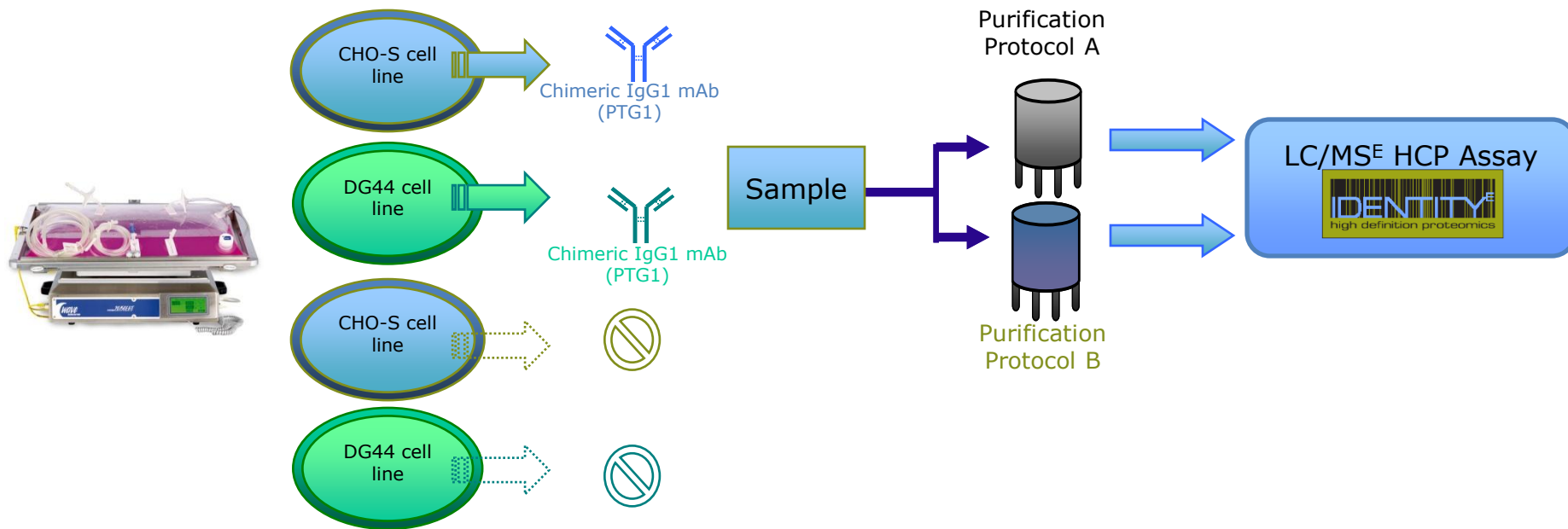
Figure 2. Development steps of a quantitative host cell protein assay



- **Process Changes may require new assay development.**
- **Hurdle to Biosimilar approval**
- **Regulators not all that happy with this current technology.**

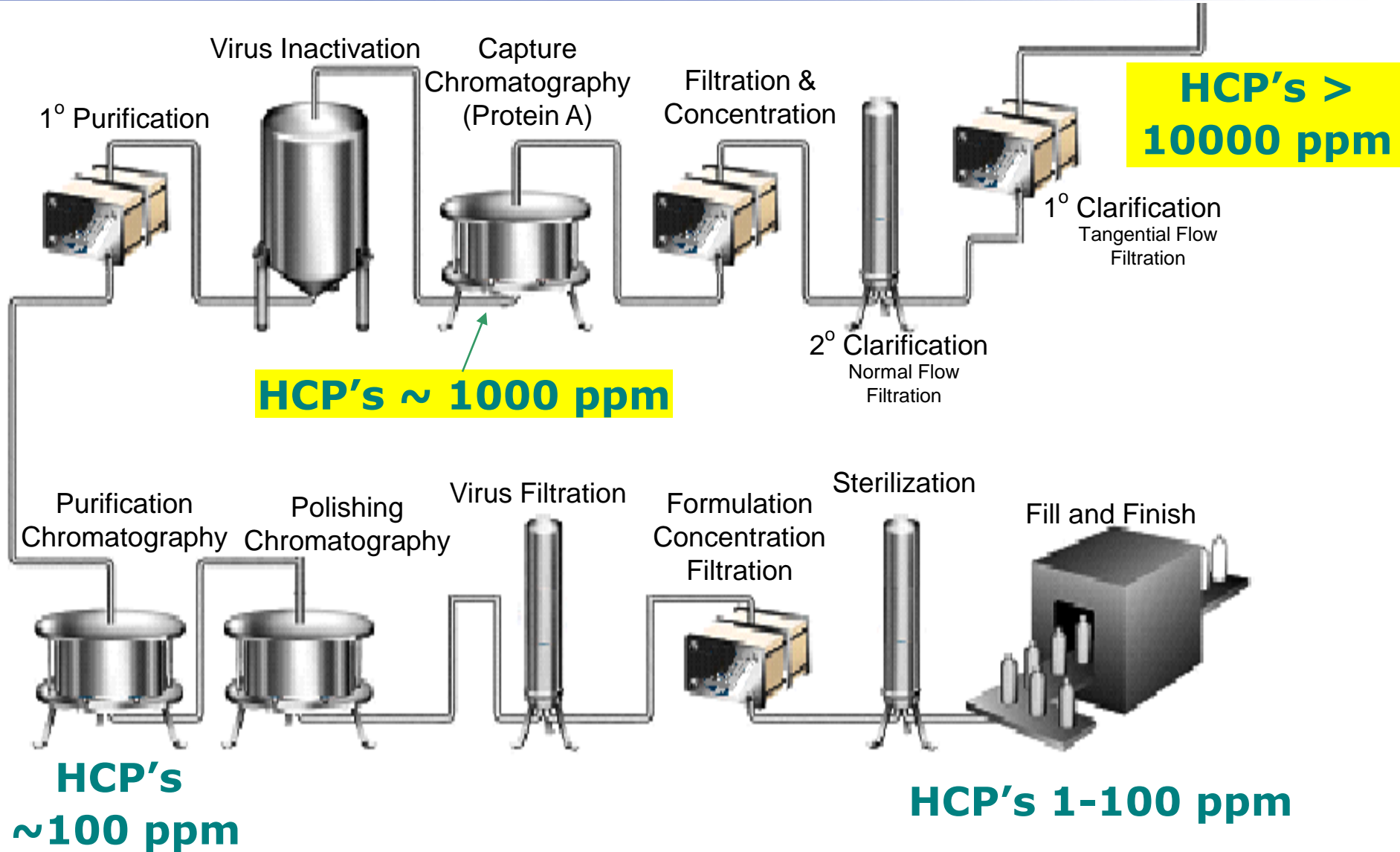


Biopharm International, Volume 13,
Number 6, pp. 38-45, May 2000



- Chimeric anti-phosphotyrosine IgG1 mAb (PTG1) expressed in 2 cell lines (CHO-S and DG44)
- Purified by Protein A chromatography (two different protocols).
- Two cell lines (DG44 and CHO-S) containing no PTG1 vector were also grown under the same conditions and purified on a Protein A column.
- Five proteins were spiked in PTG1
 - 4,000 fmoles LA (bovine beta-lactoglobulin), 800 fmoles PHO (rabbit glycogen phosphorylase b), 320 fmoles ADH (yeast alcohol dehydrogenase), 80 fmoles BSA (bovine serum albumin), and 16 fmoles ENL (yeast enolase).
- Samples analysed by LC/ MS^E

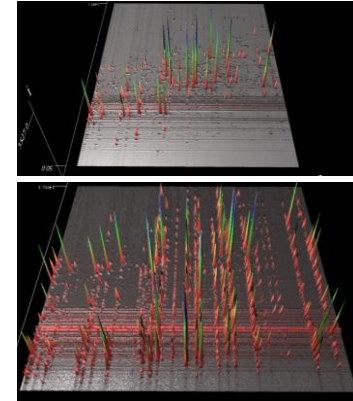
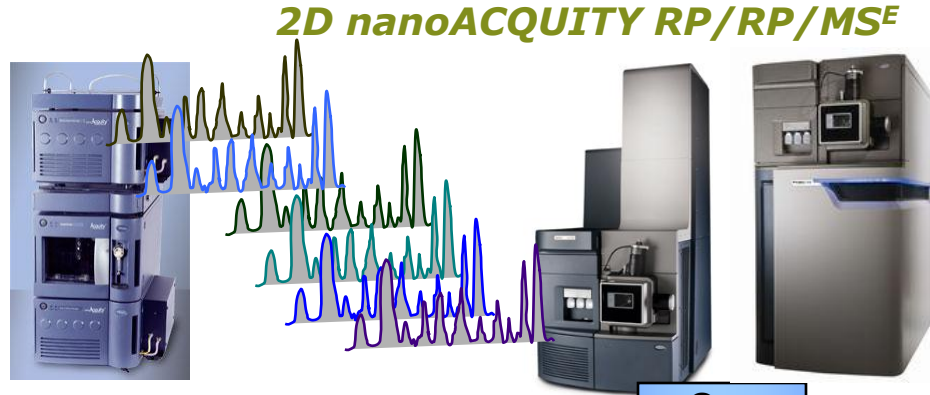
"Catalog" Host Cell Proteins (HCP's) early in downstream processing



Cataloging HCP's in a biotherapeutic protein sample



Tryptic Digestion



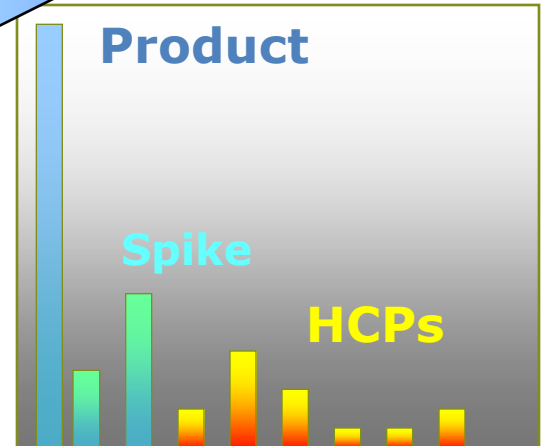
**Identify HCPs
Quantify by Hi3**



“Catalog” of
HCP Process-
impurity
proteins

**Routine
Sensitivity
~ 50 ppm**

ID no	Protein ID and species	Concentration (ppm or ng/ml) DG44 cells	CHO-S cell
1	78 kDa glucose regulated protein (hamster)	778	239
2	Actin isoforms (hamster)	3237	4275
3	Alcohol dehydrogenase (yeast) - 560 ppm (ADH)	90	188
4	Clusterin (mouse)	3306	5584
5	Cofilin 1 (mouse)	109	189
6	Elongation factor 1 and 2 (hamster)	1684	1247
7	Enolase (yeast) - 13 ppm (ENL)	23	16
8	Glyceraldehyde 3 phosphate dehydrogenase (hamster)	4594	6359
9	Glycogen phosphorylase (rabbit) - 530 ppm (PHO)	480	607
10	Heat shock 70 kDa protein (mouse)	9907	6532
11	Heat shock cognate 71 kDa protein (mouse)	628	322
12	Heat shock protein HSP 90 alpha(mouse)	266	874
13	alpha lactalbumin (bovine) - 1,200 ppm (LA)	654	725
14	Lipoprotein lipase (hamster)	1024	2358
15	Lysosomal alpha glucosidase (mouse)	2153	3385
16	Nucleolin (hamster)	1431	2891
17	Pyruvate kinase isozyme (mouse)	683	450
18	Serine protease HTRA1 (mouse)	741	952
19	Serum albumin precursor (bovine) - 100 ppm (BSA)	100	100
20	T complex protein subunits (mouse)	8120	10563
21	Tubulin alpha and beta chain (mouse)	2464	7158
	Percent composition for 21 common proteins	4.25	5.71

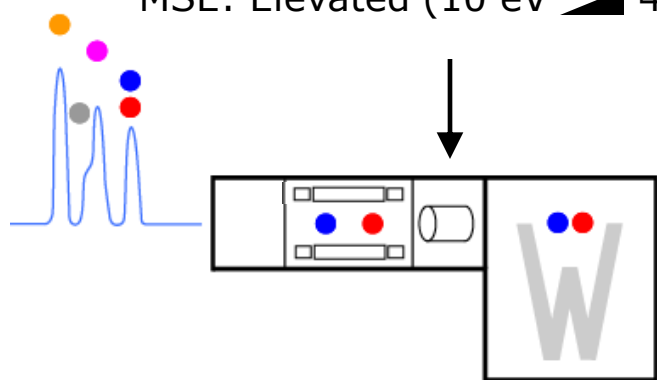


MS^E: An Alternate Scanning Methodology for Acquiring Peptide and Fragmentation Data

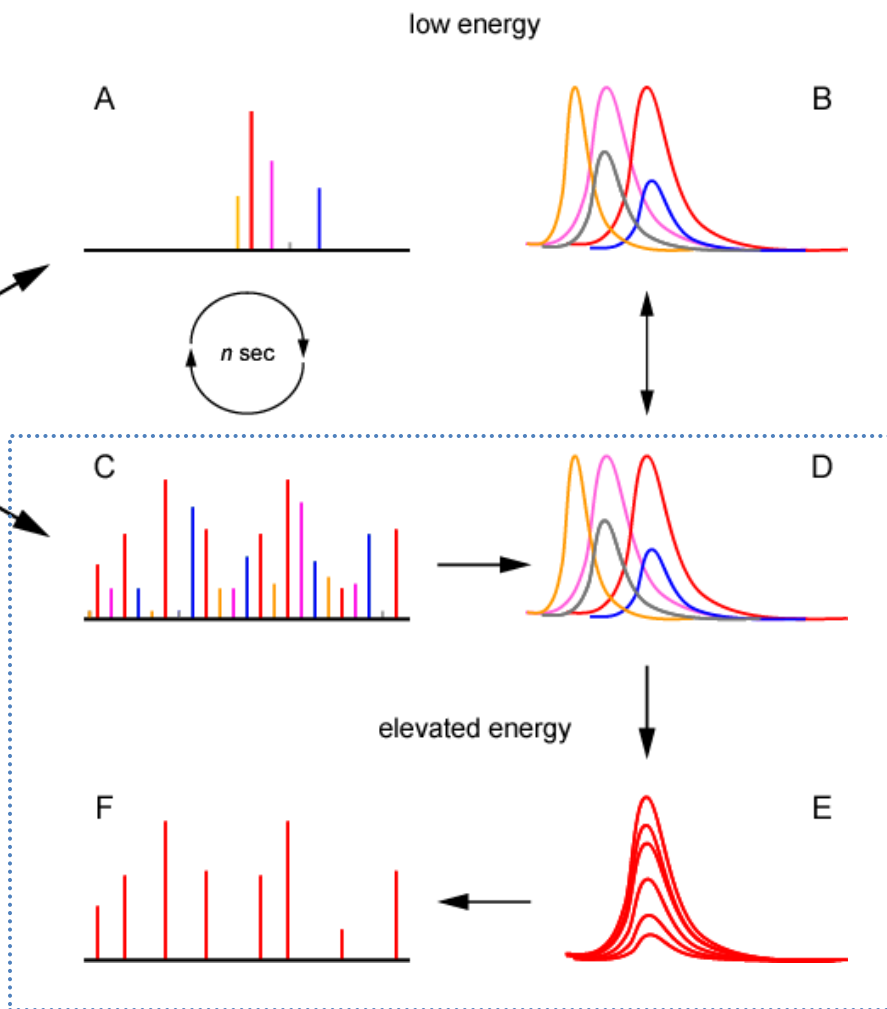
Collision Cell Energy Alternates:

MS: Low (5 eV)

MSE: Elevated (10 eV \blacktriangleleft 45 eV)



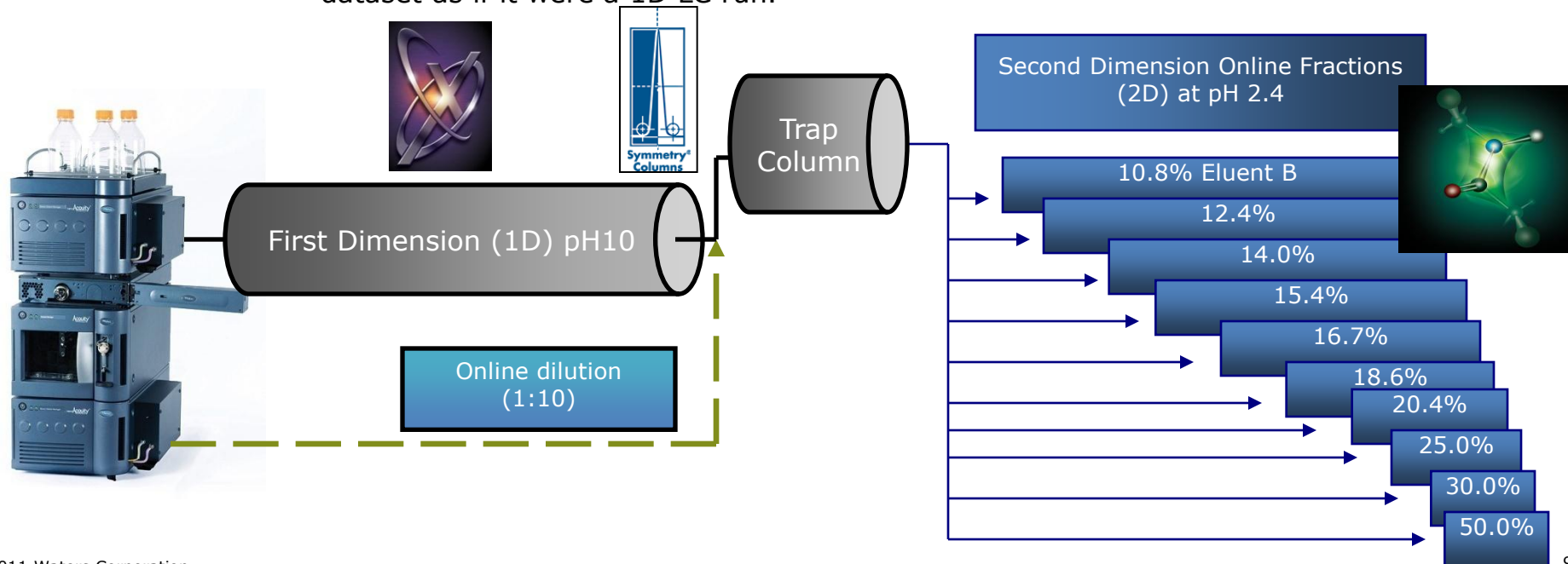
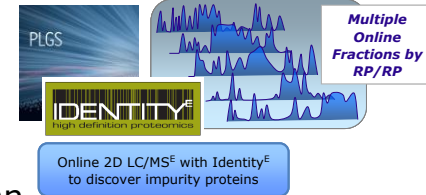
- Global Analysis (Accurate Mass)
- Minimize bias/selection of ions (\uparrow Reproducibility)
- Qualitative and Quantitative data from one analysis



LC Conditions for HCP discovery

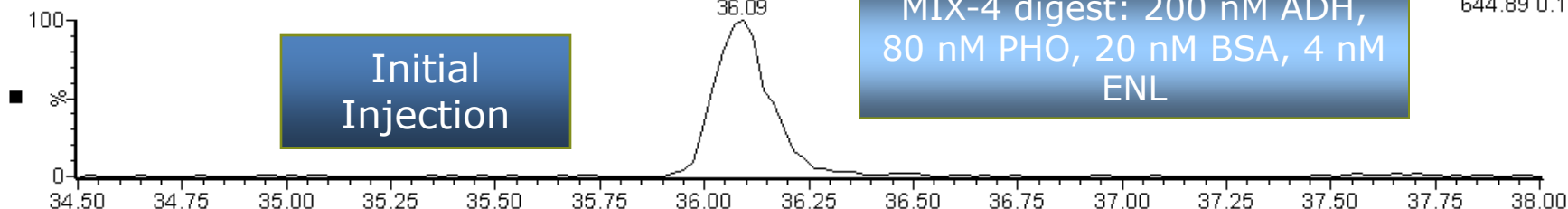
nanoACQUITY™ UPLC® system with 2D technology

- First Dimension (1D) pH=10: High Loading Capacity, high pH resistance (XBridge)
- Online dilution (1:10) of the eluent from 1D before analyte trapping onto the 2D column.
- Trap column: 5- μ m Symmetry for high carbon loading – facilitates peptide retention
- Second Dimension (2D) pH=2.4: analytical chromatography of each fraction in turn with longer gradient to maximize resolution for each fraction.
- Digital Fraction 'Merge': digitally group peptides that fall into different fractions by treating the dataset as if it were a 1D LC run.

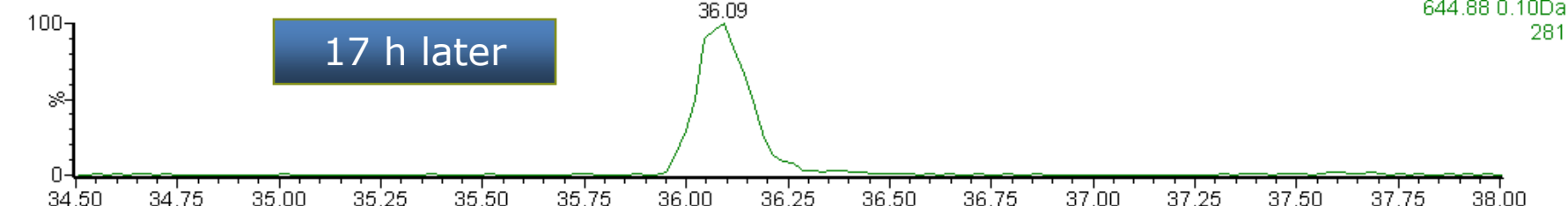


Reproducibility of 2D Chromatography: T43 ENL (VNQIGTLSESIK), 24 fmoles on column, Fraction 3/5 90 min gradients

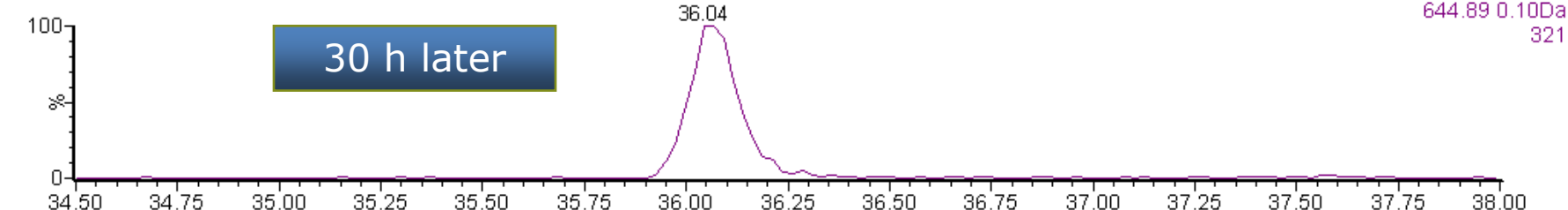
100218_UAA060_CD_07



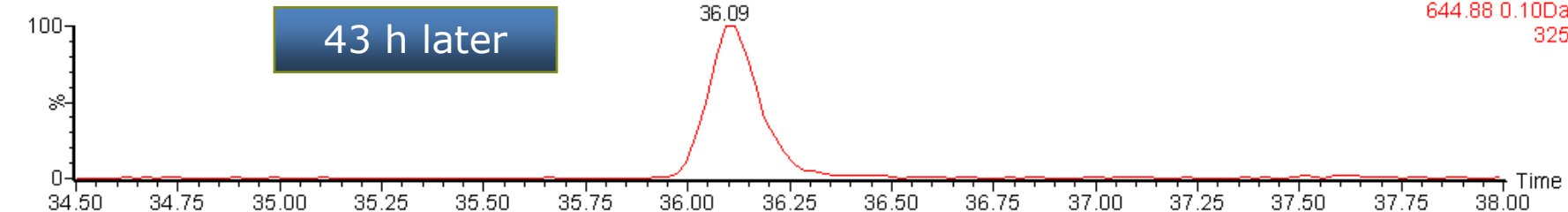
100219_UAA060_CD_04



100219_UAA060_CD_10



100220_UAA060_CD_04



Absolute Protein Quantitation (Hi3) by LC-MS^E

Basis: Intensity of Top 3 Peptides of each Protein

Absolute Quantification of Proteins by LCMS^E

A WRITE-UP PARALLEL MS ACQUISITION
Jeffrey C. Silva[‡], Marc V. Gorenstein[‡], Guo-Zhong Li[‡], Johannes P. C. Vissers[‡], and Scott J. Geromanos[‡]

Relative quantification methods have dominated the quantitative proteomics field. There is a need, however, to robustly absolute quantification studies to accurately model and understand the complex molecular biology that results in proteome variability among biological samples. A new method of absolute quantification of proteins is described. This method is based on the discovery of an unexpected relationship between MS signal response and protein concentration: the average MS signal response for the three most intense tryptic peptides per mole of protein is constant within a coefficient of variation of less than ±5%. Given an internal standard, this relationship is used to calculate a universal signal response factor. The universal signal response factor (universal) was shown to be the same for all proteins tested in this study. A controlled set of six exogenous proteins of varying concentrations was studied in the absence and presence of human serum. The absolute quantity of the standard protein was determined with a relative error of less than ±15%. The average MS signal responses of the three most intense peptides from each protein were plotted against their calculated protein concentrations, and this plot resulted in a linear relationship with an R² value of 0.9938. The analyses were applied to determine the absolute concentration of 11 common serum proteins, and these concentrations were then compared with known values available in the literature. Additionally written an unmodified *Escherichia coli* lysate, a subset of 20 identified proteins known to exist as functional complexes was studied. The calculated absolute quantities were used to accurately determine their stoichiometry. *Molecular & Cellular Proteomics* 5:144-156, 2006.

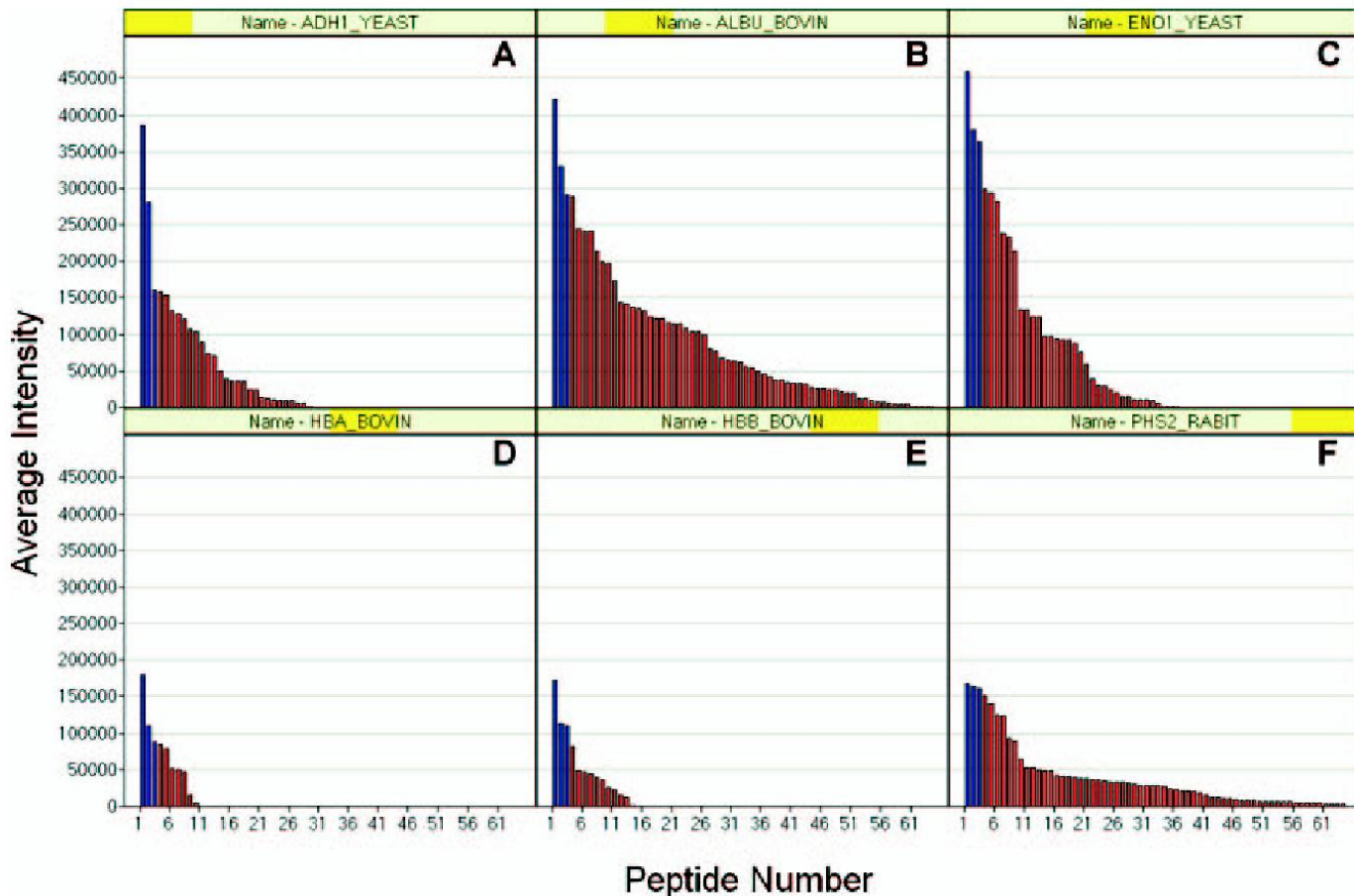
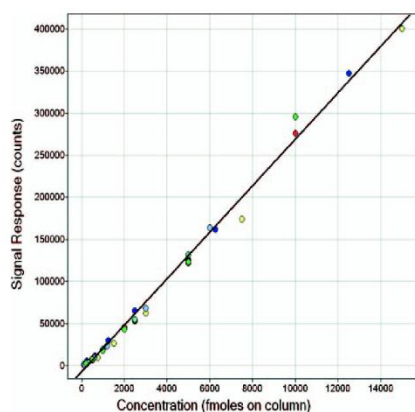
To date a majority of the quantitative proteomic analyses have been performed using stable isotope labeling strategies such as GCAT-01, TRAQ™ (GLAC) stable isotope labeling by amino acids in cell culture (SILAC), and ¹⁵N labeling. These methodologies require complex, time-consuming sample preparation and can be relatively expensive. Recently there have been numerous reports applying label-free methods to monitor the relative abundance of protein between different conditions (8-11). Relative quantification provides information regarding specific protein abundance changes between two conditions caused by an induced perturbation (environment-induced, drug-induced, and disease-induced). These studies require comparison of identical peptidic peptides in each of the two experiments to accurately determine relative ratios of the particular protein(s) of interest. Relative abundance values for each peptide to a given protein can then be obtained to quantitatively characterize the differential expression of proteins between different sample states. Many of these methods are based on determining the ratios of the peak area of identified peptides between different conditions. One critical factor limiting the quantitative reproducibility of these methods includes the ability to efficiently cluster the detected peptides. This in turn relies on the accuracy of the mass measurement and the chromatographic reproducibility. Although relative quantification monitors changes in protein abundance between two conditions, it does not determine the absolute quantity of these proteins. The ability to determine the absolute concentration of a protein (or proteins) present within a complex protein mixture is valuable for the understanding of the underlying molecular biology guiding the response to an applied perturbation. Cellular responses are often controlled through direct and indirect interactions of proteins present in the cell. These coordinated interactions allow the cell to communicate a response across many cellular compartments. The cell can thereby execute an efficient and expeditious requirement and production of critical proteins needed for adaptation. A method for determining the absolute quantity of proteins in a complex sample would enable determination of the stoichiometry of proteins within a sample and would facilitate understanding of the complicated biological network of cooperative protein interactions that guide cellular responses.

The study of proteins is crucial in understanding and combating disease through identification of proteins, discovering disease biomarkers, studying protein involvement in specific metabolic pathways, and identifying protein targets in drug discovery (1, 2). An important technique that is used in these studies to identify and quantify peptides and/or proteins present in simple and complex mixtures is ESI-LCMS.

From the Children's Hospital, Harvard Medical School, Boston, Massachusetts 02131-3009 and Waters Corporation, Milford, Massachusetts 01908-5000. Received: July 20, 2005, and in revised form: September 23, 2005. Published: MCP Papers in Press, October 11, 2005. DOI 10.1074/mcp.M50020-MCP2005

144 *Molecular & Cellular Proteomics* 5:1

© 2006 by The American Society for Biochemistry and Molecular Biology, Inc. This paper is available on-line at <http://www.mcponline.org>



***Molecular & Cellular Proteomics* 5:144–156, 2006.**

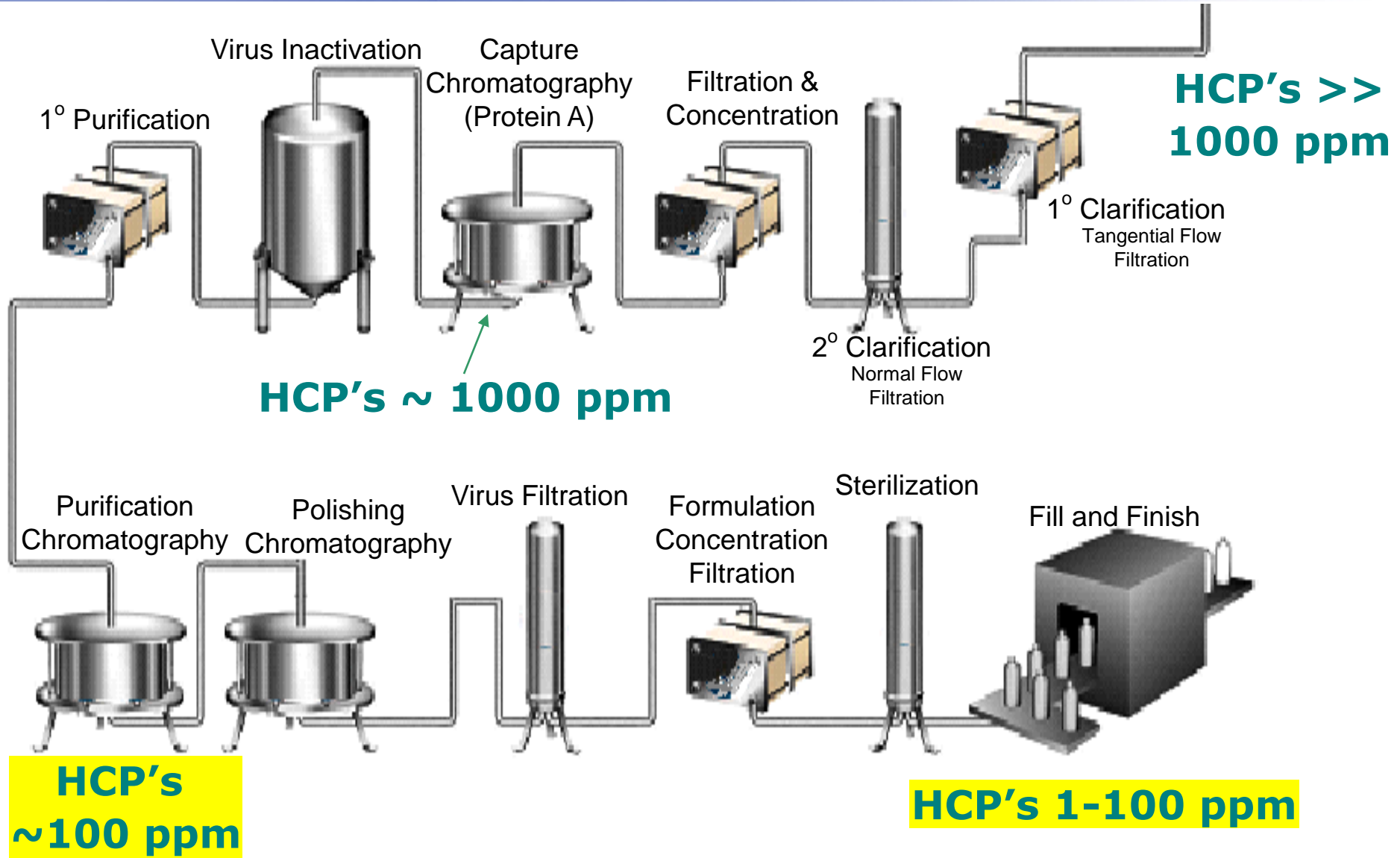
Example: Comparing HCPs in a therapeutic mAb isolated by two Protein A purification methods

		Protein ID and species	Protocol A (ppm or ng/mg)	Protocol B (ppm or ng/mg)
		40S ribosomal protein S3 (mouse)	663	561
		40S ribosomal protein S4 (mouse)	216	83
Post-Protein A Purification Step	Method A	Method B	580	1265
			2416	1844
No. proteins identified	49	73	560	690
			2467	1899
HCP Levels	1.0x	2.7x	82	87
			51	219
Total HCP by ELISA Assay (ppm)	1.0x	3.5x	1256	3204
			89	451
			3428	2386
			530	652
			7393	325
			468	448
			199	754
			1,200	1480
			764	731
		Lysosomal alpha glucosidase (mouse)	1607	541
		Nucleolin (hamster)	1068	14191
		Peroxiredoxin 1 (mouse)	122	229
		Procollagen C endopeptidase enhancer (mouse)	89	2995
		Pyruvate kinase isozyme (mouse)	510	1624
		S methyl 5 thioadenosine phosphorylase (mouse)	34	70
		Serine protease HTRA1 (mouse)	553	705
		Serum albumin precursor (bovine) - 80 fmoles IS	100	123
		T complex protein subunits (mouse)	6060	777
		Transcription factor HES 5 (mouse)	104	293
		Tubulin alpha and beta chain (mouse)	1839	10072

Determine:

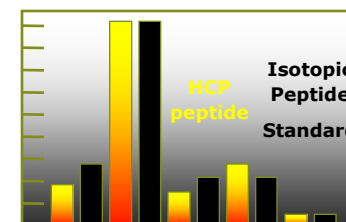
- ✓ **Total HCP Level**
- ✓ **HCP Complexity**
- ✓ **Which Proteins?**
- ✓ **At what Levels?**

Monitor Host Cell Proteins (HCP's) later in downstream processing using MRM



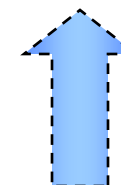
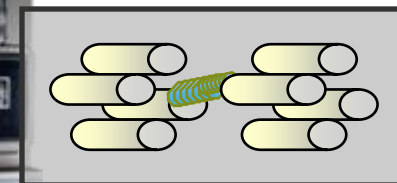
MRM Monitoring of HCP's

- Verify facilitates development of MRM methods for multiple HCP's via "signature" peptides
- Absolute quantification can be performed by using stable isotopically labeled peptide standards
- Assay extremely simple to update/ change**



Targeted MRM Quantification of HCP Signature Peptides

ID no	Protein ID and species	Concentration (ppm or ng/mg)	
		DG44 cells	CHO-S cells
1	78 kDa glucose regulated protein (hamster)	778	2354
2	Actin isoforms (hamster)	3237	4275
3	Alcohol dehydrogenase (yeast) - 560 ppm (ADH)	90	188
4	Clusterin (mouse)	3306	5584
5	Cofilin 1 (mouse)	109	189
6	Elongation factor 1 and 2 (hamster)	1684	1247
7	Enolase (yeast) - 13 ppm (ENL)	23	16
8	Glyceraldehyde 3 phosphate dehydrogenase (hamster)	4594	6359
9	Glycogen phosphorylase (rabbit) - 530 ppm (PHO)	480	607
10	Heat shock 70 kDa protein (mouse)	9907	6532
11	Heat shock cognate 71 kDa protein (mouse)	628	322
12	Heat shock protein HSP 90 alpha (mouse)	266	874
13	alpha lactalbumin (bovine) - 1,200 ppm (LA)	654	725
14	Lipoprotein lipase (hamster)	1024	2358
15	Lysosomal alpha glucosidase (mouse)	2153	3385
16	Nucleolin (hamster)	1431	2891
17	Pyruvate kinase isozyme (mouse)	683	450
18	Serine protease HTRA1 (mouse)	741	952
19	Serum albumin precursor (bovine) - 100 ppm (BSA)	100	100
20	T complex protein subunits (mouse)	8120	10563
21	Tubulin alpha and beta chain (mouse)	2464	7158
22		4.25	5.71



Discovery Phase (QTof)

Identify HCPs using LCMS^E



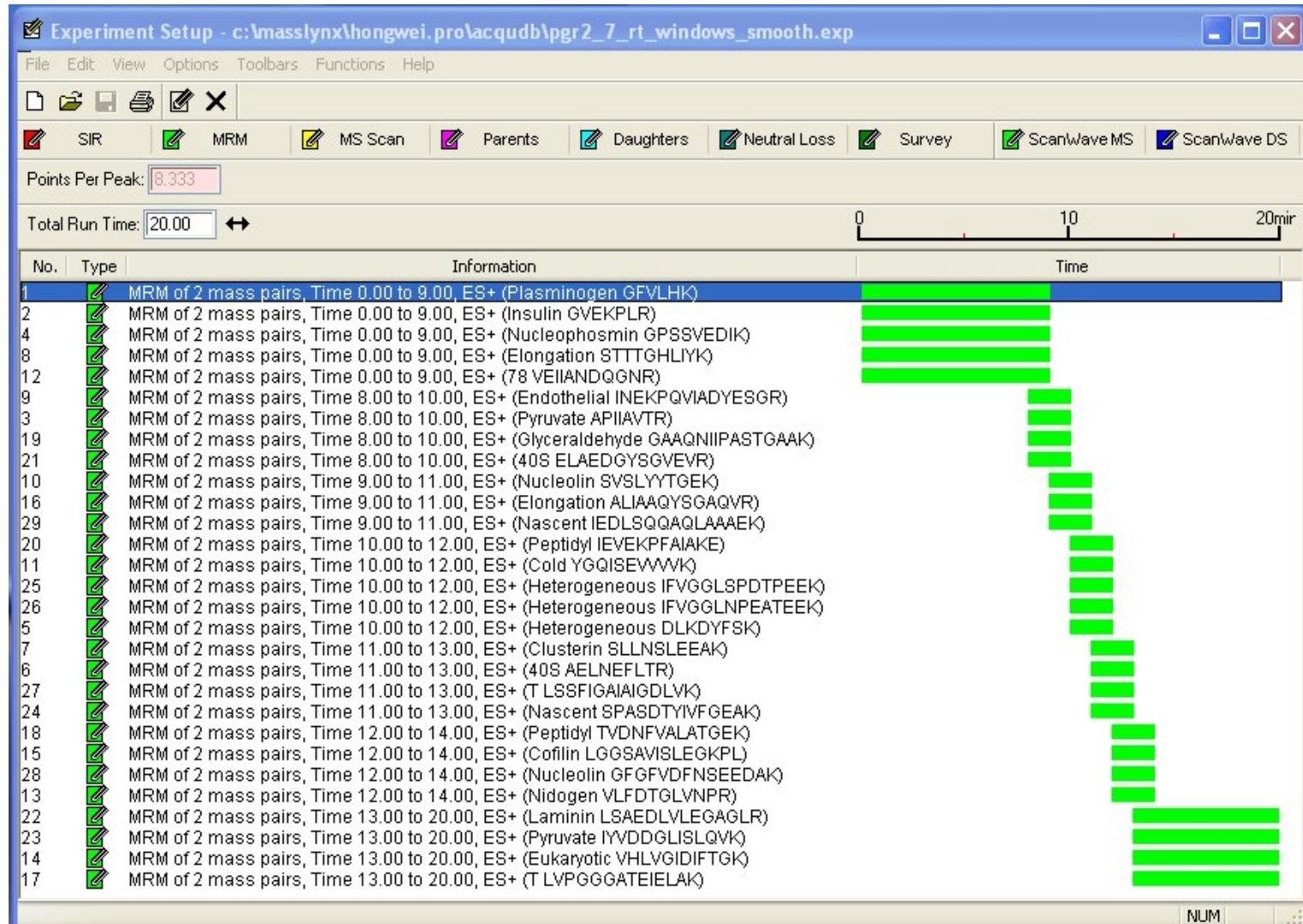
HCP Protein ID
Peptides (RT, Intensity)
MS^E Fragmentation

Monitoring Phase (QQQ)

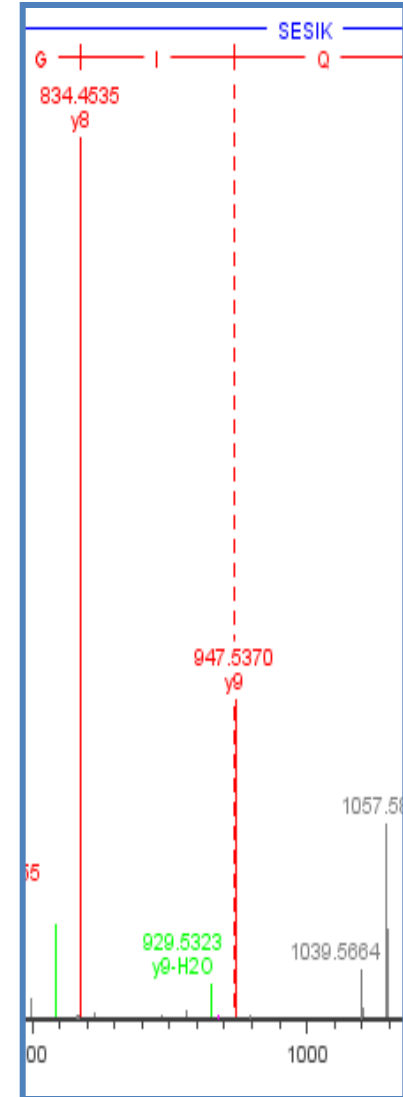
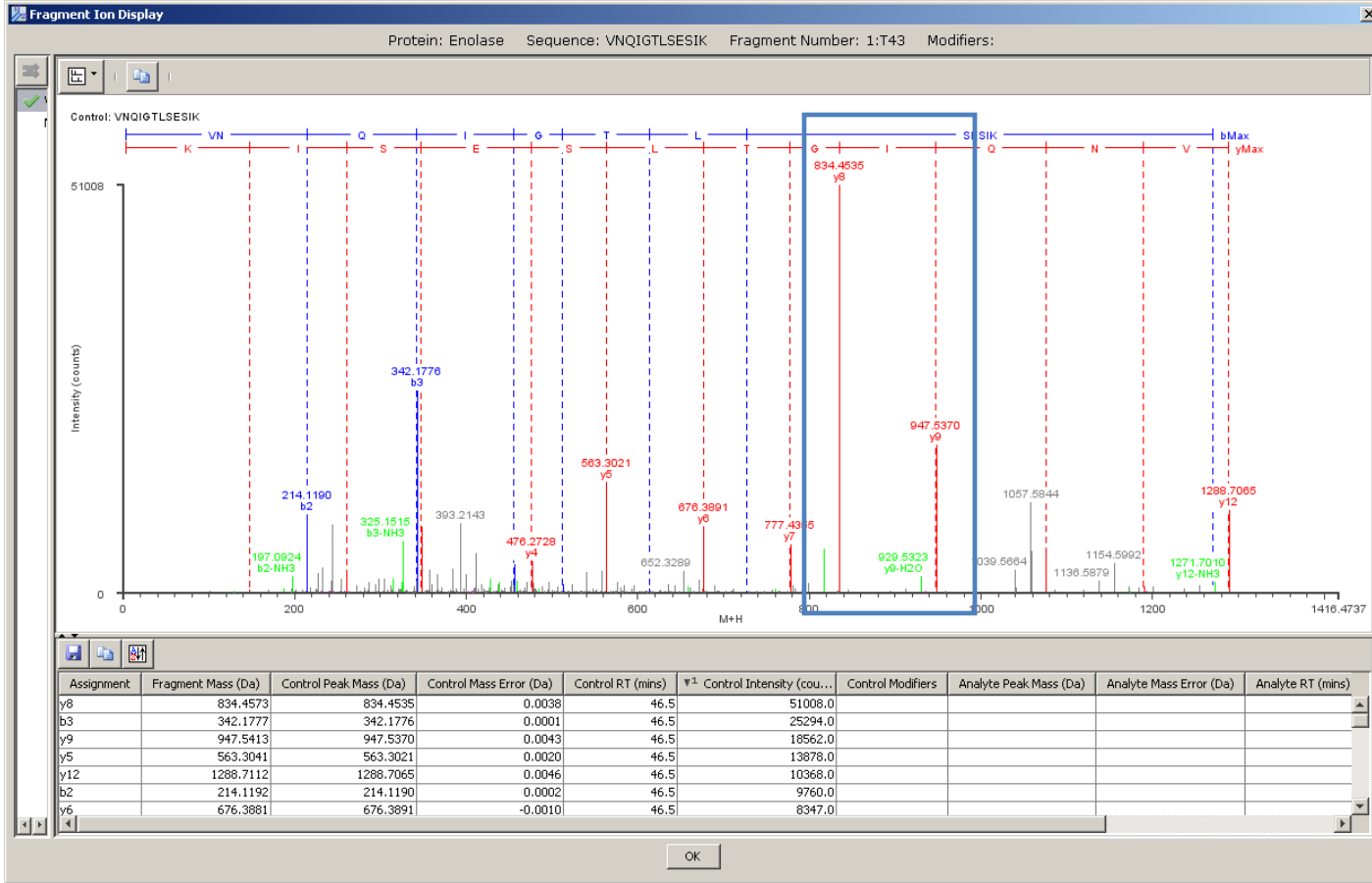
- Select 3-5 peptides per protein (best ionizing, good fragmentation)
 - Produce recombinant HCP?
- Produce synthetic peptides
- Identify best 2-3 MRM transitions per peptide (Infusion Experiment)
 - Optimize Cone Voltage
 - Optimize Collision Energy
- LC/MS method development
 - Spike peptides into biotherapeutic digest to identify product interferences
 - (Optional) Stable isotope labeled peptide for absolute quantitation
 - Investigate the linearity of the MRM assay



QQQ MRM method to monitor Top 20 HCP Proteins (29 Peptides, 58 Transitions) from method 1

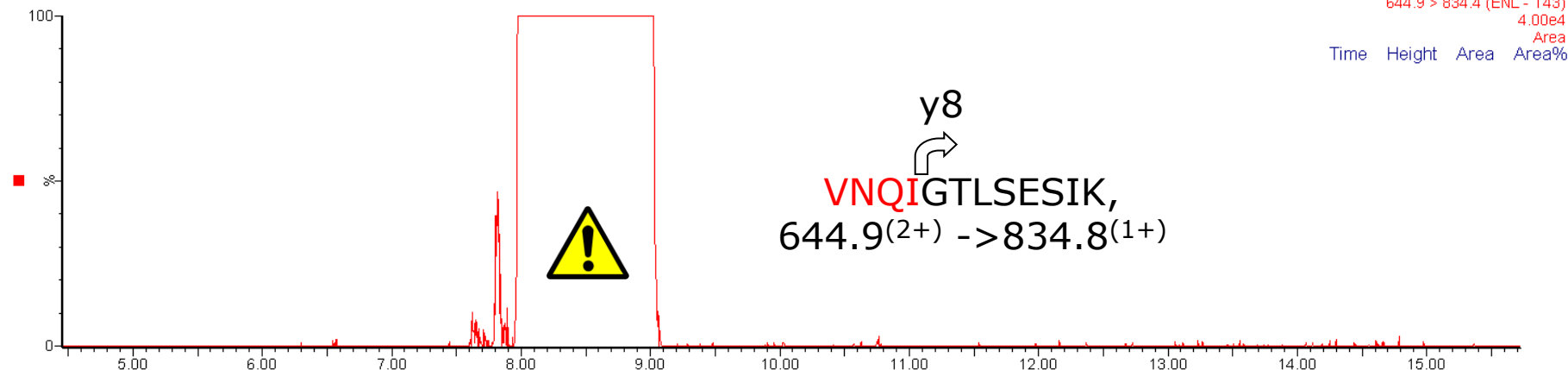


HCP T44 Peptide MS^E fragmentation

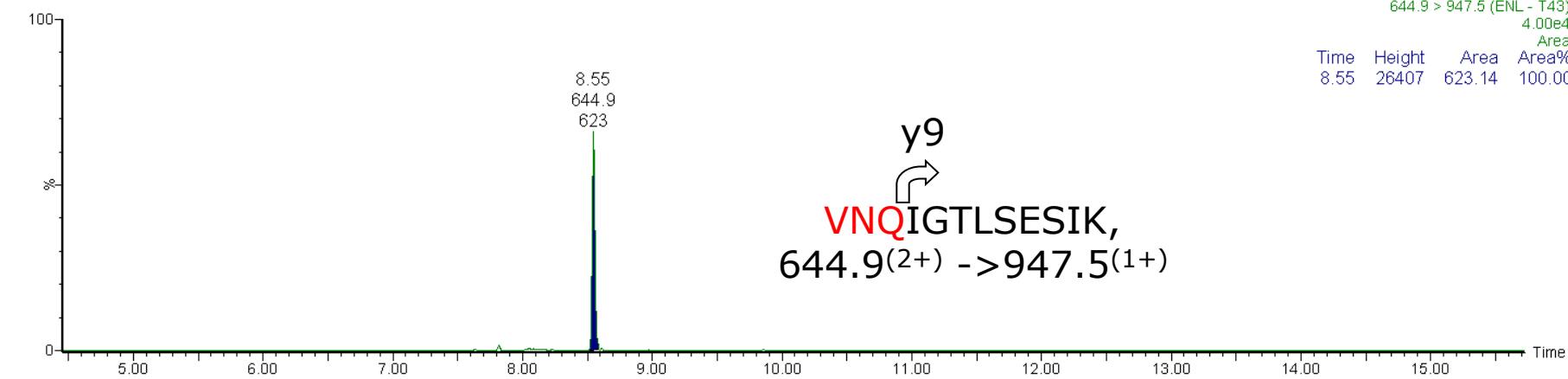


MRM Interference for a HCP Peptide MRM Transition

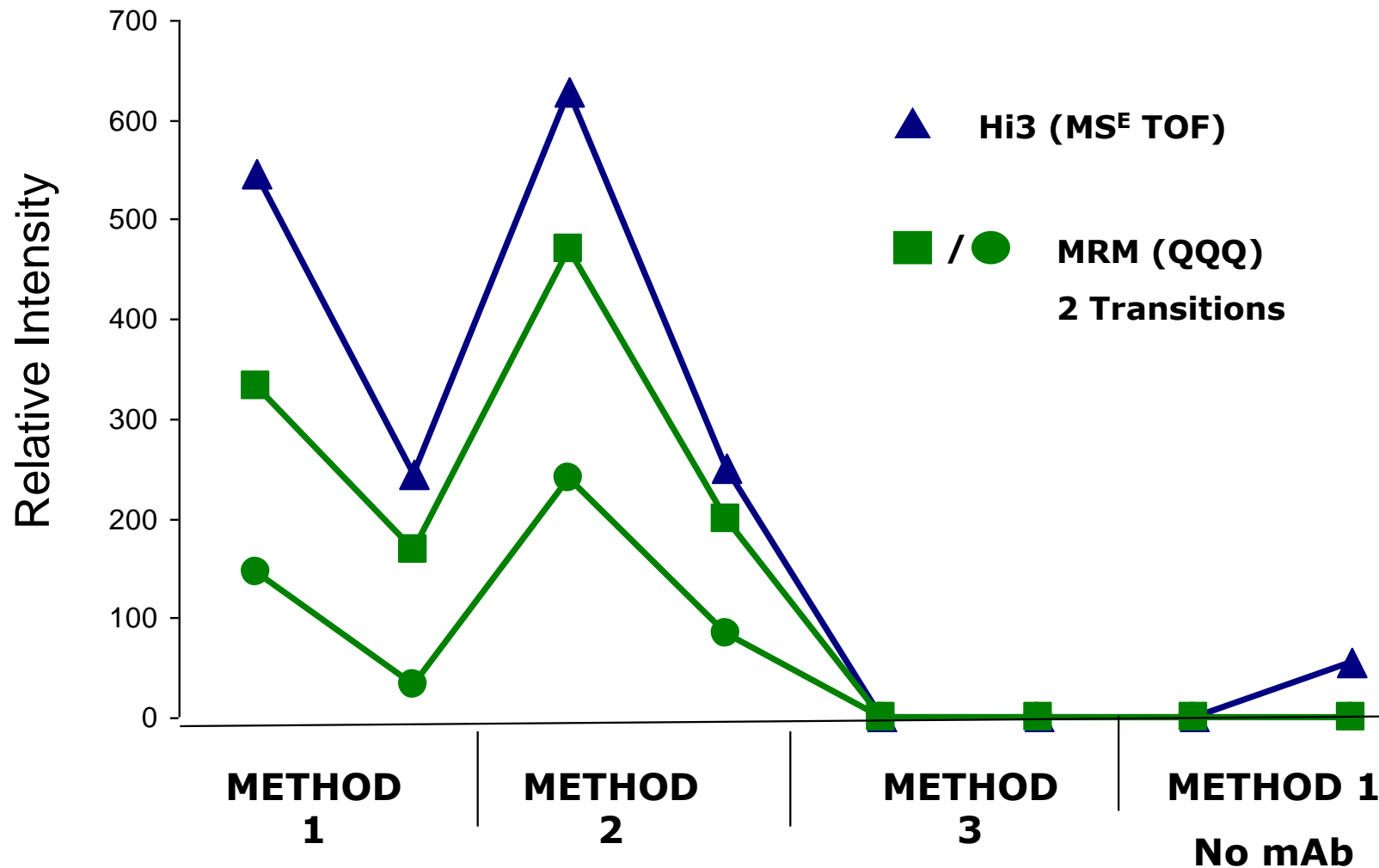
Herceptin + spiked proteins digest
090720_VBA053_CD_16



090720_VBA053_CD_18 Sm (Mn, 1x2)

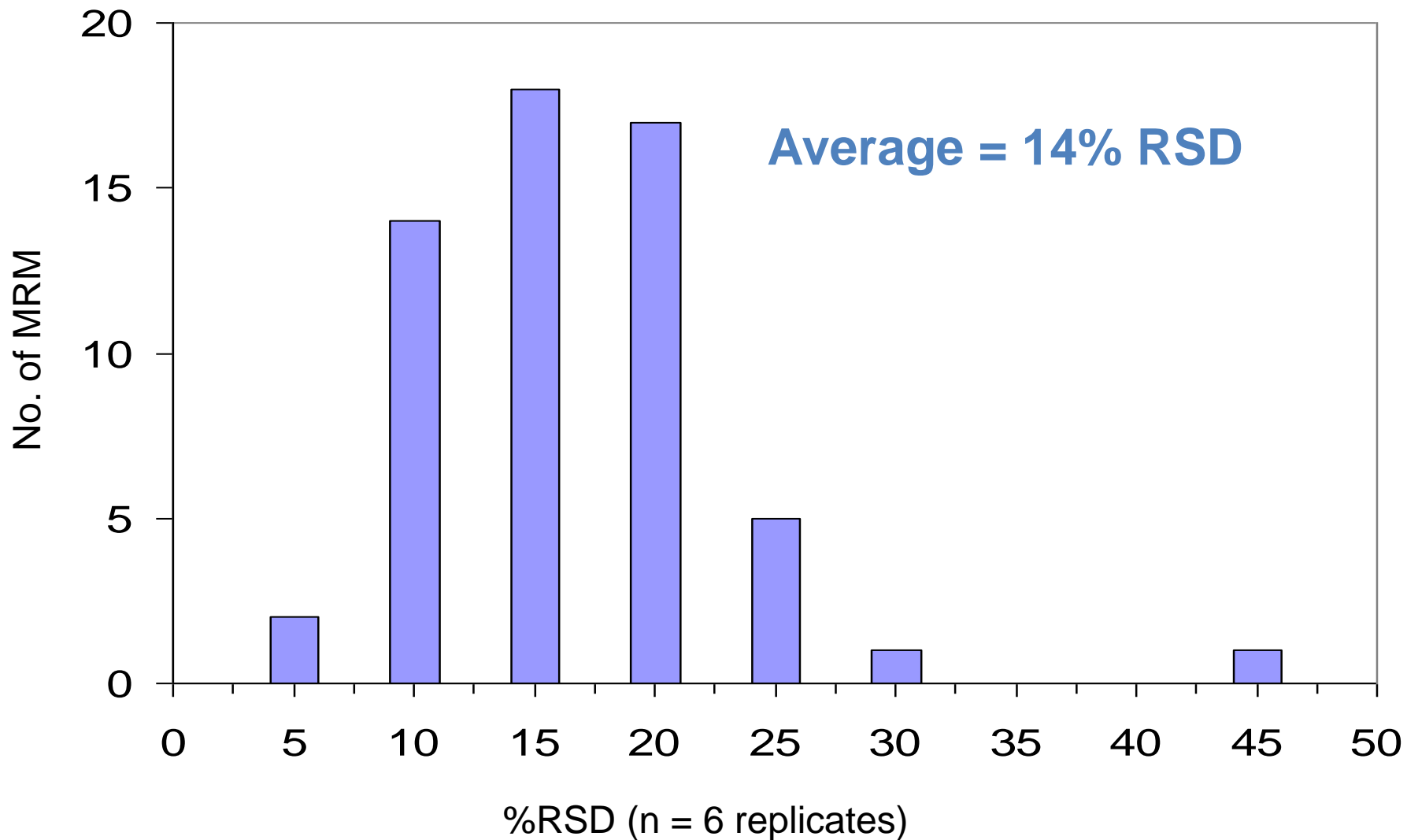


Hi3 vs. MRM quantitation data for a 58 kD HCP demonstrates common results



Analytical Reproducibility for 29 HCP Proteins (58 MRM Transitions)

Method 1 Sample



- Using the same technologies you use for protein characterization you can also identify HCP's in your product.
 - 2DLC/MS^E and the most accurate bioinformatics platform are all required to produce sensitive high quality results
 - Catalog (Identify and quantitate) and compare HCP's in early stages of processing.
- Easily developed MRM assays can monitor low ppm level HCP's in later processing steps and in the purified drug.
 - These MRM assays are readily modified to facilitate faster process development/improvement cycles.
 - Such MRM assays are commonly used today for regulated analyses, and could be readily validated for future QC needs.

Waters

THE SCIENCE OF WHAT'S POSSIBLE.™