

Assessing the Proportion of Post-Translationally Modified Peptides in the Immunopeptidome

Wilfred H. Tang¹, Christie L. Hunter¹,
Frank R. Rooney², Nicholas A. Williamson³,
Anthony W. Purcell³

¹Applied Biosystems, Foster City, CA

²Applied Biosystems, Melbourne, Australia

³University of Melbourne, Australia

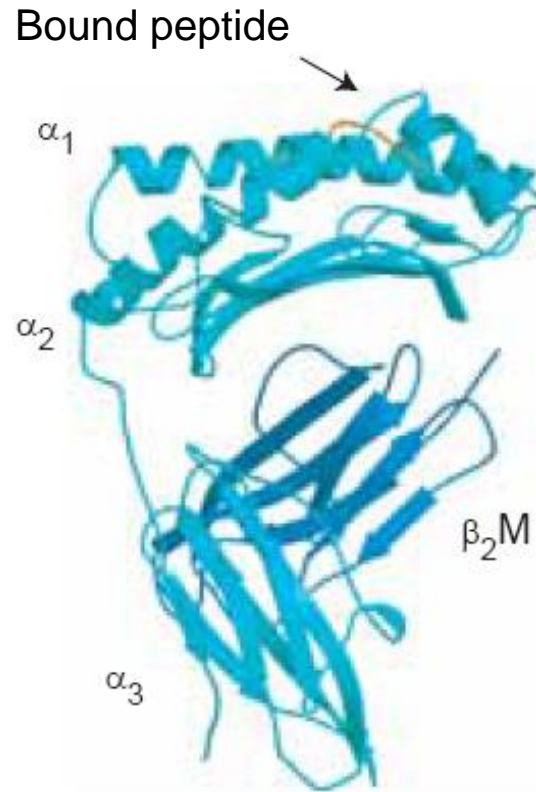
Outline

- What is the immunopeptidome?
- How do we study the immunopeptidome?
- Results

Immune System Provides Continual Surveillance

- Human leukocyte antigen (HLA) proteins display peptides on the cell surface
 - Cells break proteins into peptide fragments by diverse, non-specific proteolytic processes
 - HLA proteins bind to selected peptide fragments, and the HLA-peptide complexes are transported to the cell surface
- T cells continually scrutinize the HLA-peptide complexes displayed on cells throughout the body
 - Changes alert the immune system

Immunopeptidome



- Immunopeptidome is the collection of peptides bound to HLA
- Immunopeptidome plays a key role in immune defense
- Desirable to have a reliable method to map the immunopeptidome

Post-Translational Modifications Matter

Post-translational modifications associated with autoimmune responses

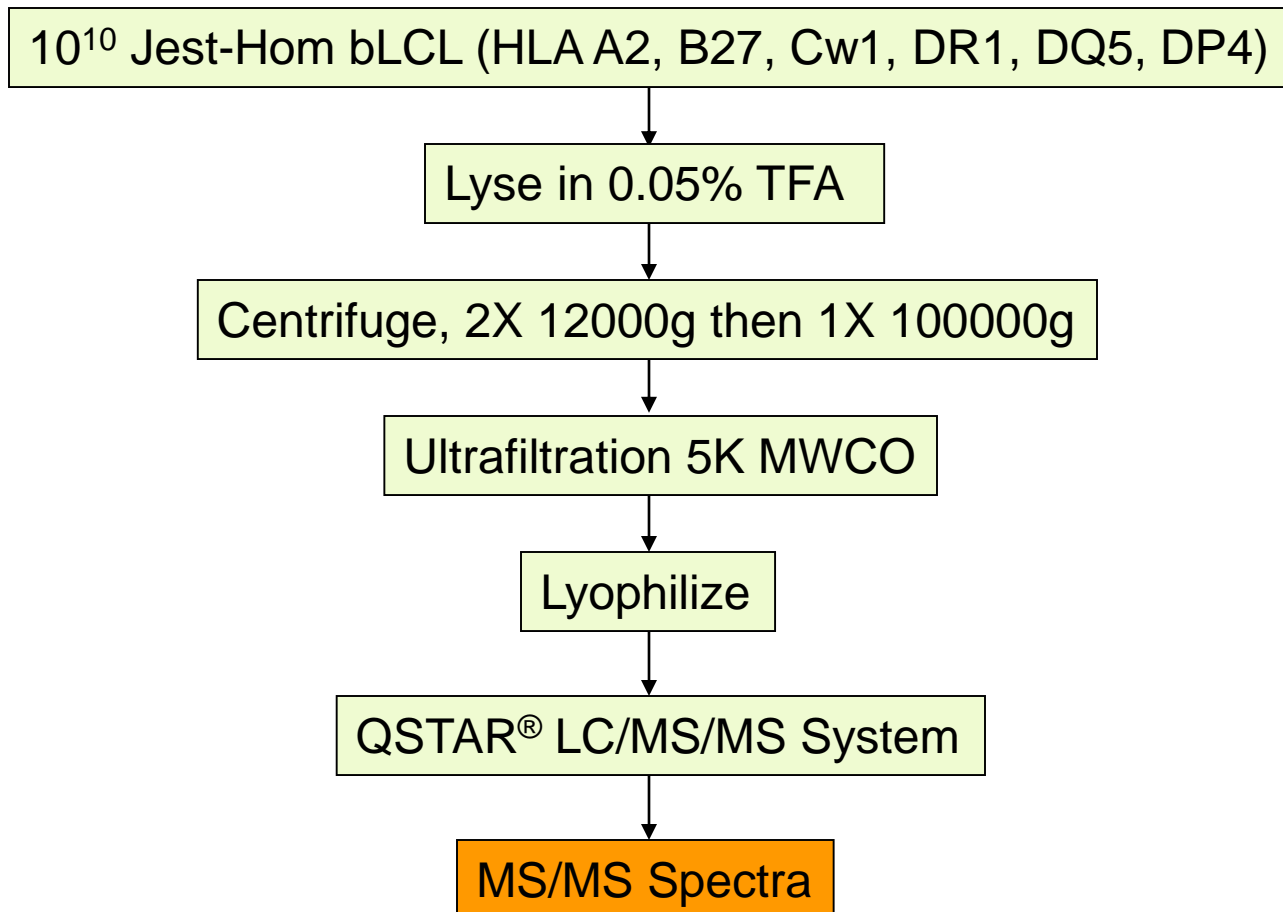
Disease	Modification	Antigen	Immune response		Model	Required for response	Refs
			T cell	B cell			
EAE and multiple sclerosis	Phosphorylation	α B-crystallin	Specific	ND	Mouse	No	13
	Deimination	MBP	Specific	Diverse	Rat	No	7
	Acetylation	MBP-Ac1-11	Specific	ND	Mouse	Yes	6
Collagen-induced arthritis	Glycosylation and hydroxylation	Type II collagen	Specific	ND	Mouse and rat	No	14 57
Rheumatoid arthritis	Deimination	Fibrin	ND	Specific	Human	ND	58
SLE	Phosphorylation	Multiple	ND	Diverse	Human	ND	40,12
	Isoaspartylation	snRNP D	Specific	Diverse	Mouse	Yes	8
	Mannose modification	Multiple	ND	Diverse	Mouse	Yes	59
	sDMA modification	Sm D1 and D3	ND	Specific	Human	ND	11
Celiac disease	Deamidation	Wheat gliadin	Specific	Specific to tTG	Human	Yes	60,61
Atherosclerosis	Lipid peroxidation	LDL and others	Specific	Diverse ^b	Mouse	Yes	62

^aAbbreviations: Ac, acetylated; EAE, experimental allergic encephalomyelitis; LDL, low density lipoprotein; MBP, myelin basic protein; ND, not determined; sDMA, symmetrical dimethyl arginine; SLE, systemic lupus erythematosus; snRNP, small nuclear ribonucleoprotein; tTG, tissue transglutaminase.

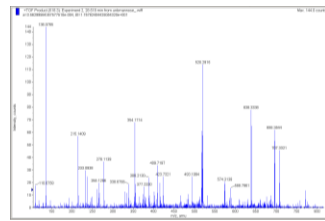
^bDepends on particular modification.

H. A. Doyle and M. M. Mamula, *Trends in Immunology*, **22**, 443 (2001).

Acid Extraction of HLA Peptides



Peptide Identification



MS/MS Spectrum

Database Search
Software

Peptide ID

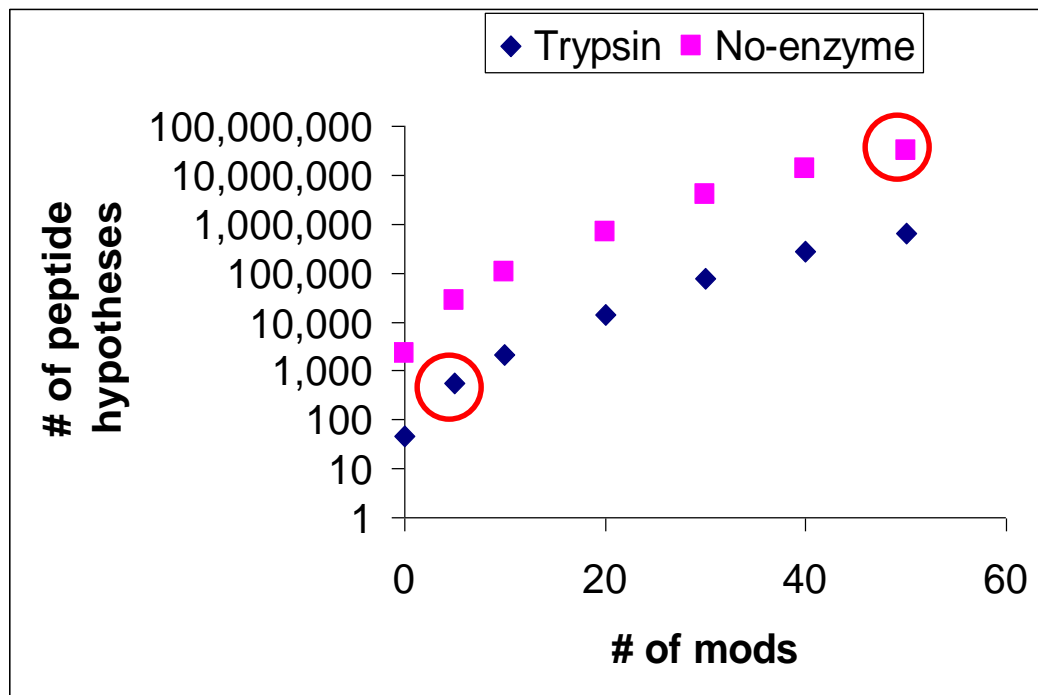
Conventional database search

- For each MS/MS spectrum:
 - Exhaustively enumerate **all** possible database peptides that match the experimental MS mass
 - Compare each candidate database peptide vs. experimental MS/MS spectrum
 - Report highest-scoring database peptide(s)
- Commonly used in shotgun proteomics experiments

HLA Peptide Identification

HLA peptides pose special challenges for database search

- No-enzyme search
- Post-translational modifications



Number of peptide hypotheses generated by exhaustive enumeration

Peptide mass: 1500.75 Da

MS Tolerance: 0.0015 Da (1 ppm)

*Database: UniProtKB Swiss-Prot
(Dec. 20, 2005)*

Searching Smarter – Paragon™ algorithm

- Don't score everything – search **selectively** rather than exhaustively
- For each MS/MS spectrum:
 - Use sequence tags to differentiate “hot spots” vs. “cold spots” in protein database
 - Search “hot spots” in database more thoroughly than “cold spots”
 - Conservatively, 99% of highly-confident peptides are found in hottest 10% of spots

No-Enzyme Search

Peragorion™ Algorithm Database Search

Small Protein:

SDPFEWILKWQASDHRGTYTESCRRCNNVEMSLWEIPLL

252 peptides total

(lengths 7-15)

HRGTYTES	SDHRGTY	NNVEMSLWE	DPFEWILKWQAS	LKWQASD	ILKWQAS
RCNNVEMSLWEI	SDHRGTYTES	TYTESCRRCNNV	DHRGTYTESCR	ILKWQASDHRGTYT	DPFEWILKWQA
VEMSLWEI	VEMSLWE	ILKWQASDHRGTY	FEWILKWQASDHR	ASDHRGTYTESC	QASDHRGTY
LKWQASDHRGTYTES	TYTESCR	RGTYTESCRRCNNV	PFEWILKWQ	RGTYTESCRRCNNVEMSL	ASDHRGTYTES
MSLWEIPLL	TESCRRCNNVEM	EWILKWQASDHRGTY	TYTESCRRCNNV	TYTESCRRCNNVEMSL	ESCRRCNNV
RCNNVEM	DHRGTYTESC	HRGTYTESCRRCNN	NNVEMSL	TYTESCRRCNNVEM	RCNNVEMSLWEI
SDPFEWI	ILKWQASDHRGT	ASDHRGTYTE	DHRGTYTESCR	TYTESCRRCNNVEM	TESCRRCNNVEMSL
SDPFEWILKWQAS	HRGTYTESCR	WILKWQASDH	WILKWQASDH	EWILKWQA	CRRCNNVEMSLW
FEWILKW	CNNVEMSLWEIPLL	RGTYTESCRRCNNV	SDHRGTYTESCR	EMSLWEIPLL	EMSLWEIPLL
RCNNVEMSLWEI	DPFEWIL	DPFEWILKWQASDH	DPFEWILKWQASDH	EMSLWEIPLL	WILKWQASDHRGTYT
SCRCNNVEMSLW	ASDHRGTYTESCR	QASDHRGTYTESC	QASDHRGTYTESCR	RCNNVEMSLW	LKWQASDHRGTYT
CNNVEMSLW	DPFEWILKWQASD	RGTYTESCR	RGTYTESC	SDPFEWILKWQASD	SCRCNNVEMSL
PFEWILKWQAS	TESCRRCNNVEM	ESCRRCNNVEM	SCRCNNVEM	EWILKWQASDHRGT	RGTYTESC
RGTYTESCR	WQASDHR	WQASDHR	PFEWILKWQA	KNQASDHRGTY	NNVEMSLWEIPLL
ESCRRCNNVEMSLWEI	PFEWILKWQASDHRG	NNVEMSLWEIPLL	EWILKWQASDH	KNQASDHRGTYTESC	ESCRRCNNVEMSLWEI
CNNVEMSLWEI	WQASDHRGTYTESC	NNVEMSLW	WQASDHRGTYTESC	NNVEMSLWEIPLL	EWILKWQASD
WQASDHRGTYTESCR	FEWILKWQ	ILKWQASDH	WILKWQASDHRG	WILKWQA	CNNVEMSLWEIPLL
DHRGTYTES	ASDHRGTYTESCR	EWILKWQASDHRG	YTESCR	DPFEWILKW	LKWQASDHR
SDPFEWILKWQA	SCRCNNVEM	WQASDHRGTYT	ILKWQASDHRG	GTYTESCR	QASDHRGTYTESCR
YTESCRRCNNVEM	YTESCRRCNNVEM	CNNVEMSLWEIPLL	QASDHRGTYTESC	SDHRGTYTESC	WILKWQAS
YTESCR	RCNNVEMSLWEIPLL	RGTYTESCR	SCRCNNVEMSL	GTYTESCRRCNNVEM	VEMSLWEIPLL
KNQASDH	EWILKWQAS	HRGTYTESCRRCNNVEM	DHRGTYTESCR	QASDHRG	CRRCNNVEMSLWEIPLL
KNQASDHRG	KNQASDHRGTYT	PFEWILKWQASDHR	YTESCRRCNNVEMSLW	DPFEWILK	ILKWQASD
DHRGTYT	ESCRRCNNVEM	KNQASDHRGTYTES	LKWQASDHRGTYTE	TYTESCR	CNNVEMSLWEI
RGTYTESCRRCNNV	HRGTYTESCR	RCNNVEMSLWEIPLL	WQASDHRGTYTES	SDPFEWILKWQ	RCNNVEMSLW
HRGTYTESCR	RCNNVEM	TESCRRCNNV	DPFEWILKWQASDHR	MSLWEIPLL	WQASDHRGTYTESC
NNVEMSLWEI	KNQASDHR	NNVEMSLW	NNVEMSLW	TESCRRCNNVEM	ASDHRGTYTESCR
DHRGTYTE	SLWEIPLL	DHRGTYTESCRRCNNV	DHRGTYTESCRRCNNV	EWILKWQ	DPFEWILKWQ
SCRCNNV	SCRCNNVEMSLWEI	TESCR	ILKWQASDHR	RGTYTESCR	ESCRRCNNV
TESCRRCNNVEMSLW	RCNNVEMSLWEI	PFEWILK	VEMSLWEIPLL	ASDHRGTYT	CRRCNNVEM
WILKWQASDHRGTY	ASDHRGT	VEMSLWEIPLL	ESCRRCNNVEMSLW	WQASDHRGTY	FEWILKWQASDHR
ASDHRGTY	KNQASDHRGTYTE	SDPFEWIL	SDPFEWIL	FEWILKWQASDHRGT	RGTYTESCRRCNNVEM
WILKWQASDHRGT	ESCRRCNNVEMSL	NNVEMSLWEI	NNVEMSLWEI	FEWILKWQASDHRGT	TESCRRCNNVEMSLW
ILKWQASDHRGTYTE	RGTYTESCR	SDHRGTYTESCR	SDHRGTYTESCR	RCNNVEMSL	YTESCRRCNNVEMSL
QASDHRGTYT	HRGTYTE	SDPFEWILK	SDPFEWILK	HRGTYTESCRRCNNV	LKWQASDHRGTY
LKWQASDHRGT	WILKWQASDHR	GTYTESCRRCNNVEM	GTYTESCRRCNNVEM	YTESCRRCNNVEMSL	SDPFEWILKW
WQASDHRGT	DHRGTYTESCRRCNN	HRGTYTESC	HRGTYTESC	ESCRRCNNVEMSLW	SDHRGTYTE
SDPFEWILKWQASDH	NNVEMSLWEI	PFEWILKWQASD	PFEWILKWQASD	KNQASDHRGT	WQASDHRGTYTESC
WILKWQASD	RGTYTESCRRCNN	SDHRGTYTESCR	SDHRGTYTESCR	TYTESCR	RCNNVEMSLWEIPLL
LKWQASDHRG	SDHRGTYTESCRRCNN	EMSLWEI	EMSLWEI	TYTESCRRCNNVEM	FEWILKWQA
PFEWILKWQASDH	QASDHRGT	FEWILKWQASDHRG	FEWILKWQASDHRG	SLWEIPLL	NNVEMSLWEI
PFEWILKW	TESCRRCNNVEM	CRRCNNVEM	LWEIPLL	CRRCNNVEMSLWEI	GTYTESCR

No-Enzyme Search Paragon™ Algorithm

Small Protein: **SDPFEWILKWQASDHRGTYTESCRRCNNVEMSLWEIPLL**

Hot spot

54 peptides near hot spot

(lengths 7-15)

HRGTYTES	SDHRGTY	NNVEMSLWE	DPFEWILKWQAS	LKWQASD	ILKWQAS
RCNNVEMSLWEI	SDHRGTYTES	TYTESCRRCNNV	DHRGTYTESCR	ILKWQASDHRGTYT	DPFEWILKWQA
VEMSLWEI	VEMSLWE	ILKWQASDHRGTY	FEWILKWQASDHR	ASDHRGTYTESC	QASDHRGTY
LKWQASDHRGTYTES	TYTESCR	RGTYTESCRRCNNV	FEWILKWQ	GYTESCRRCNNVEMS	ASDHRGTYTES
MSLWEIPLL	TESCRRCNNVEM	EWILKWQASDHRGT	YTESCRRCNNV	TYTESCRRCNNVEMSL	ESCRRCNNVE
RCNNVEM	DHRGTYTESC	HRGTYTESCRRCNN	HRGTYTESCR	TYTESCRRCNNVEM	NNVEMSLWEIP
SDPFEWI	ILKWQASDHRGT	ASDHRGTYTE	WILKWQASDH	DHRGTYTESCR	TESCRRCNNVEMSL
SDPFEWILKWQAS	HRGTYTESCR	RCNNVEMSLWEIPLL	WILKWQASDH	EMSLWEIPLL	RCNNVEMSLW
FEWILKW	CNNVEMSLWEIPLL	DPFEWIL	RGTYTESCRRCNNV	SDHRGTYTESCR	EMSLWEIPL
RCNNVEMSLWEIP	DPFEWIL	ASDHRGTYTESCR	DPFEWILKWQASDH	NNVEMSLWEIPLL	LKWQASDH
SCRCNNVEMSLW	ASDHRGTYTESCR	QASDHRGTYTESC	QASDHRGTYTESC	RCNNVEMSLW	RCNNVEMSLW
CNNVEMSLW	DPFEWILKWQASD	GYTESCR	GYTESC	SDPFEWILKWQASD	SCRCNNVEMSL
PFEWILKWQAS	TESCRRCNNVEM	ESCRRCNNVEM	SCRCNNVEM	EWILKWQASDHRGT	RGTYTESC
GYTESCR	WOASDHR	ESCRRCNNVEM	PFEWILKWQA	KNQASDHRGTY	KNQASDHRGTY
ESCRRCNNVEMSLWEI	PFEWILKWQASDHRG	NNVEMSLWEIPL	EWILKWQASDH	KNQASDHRGTYTESC	ESCRRCNNVEMSLWEI
CNNVEMSLWEI	WOASDHRGTYTESC	NNVEMSLW	YTESCR	NNVEMSLWEIPLL	NNVEMSLWEIPL
WOASDHRGTYTESCR	FEWILKWQ	ILKWQASDH	WILKWQASDHRG	WILKWQA	WILKWQASD
DHRGTYTES	ASDHRGTYTESCR	EWILKWQASDHRG	YTESCR	DPFEWILKW	CNNVEMSLWEIP
SDPFEWILKWQA	SCRCNNVE	QASDHRGTYT	QASDHRGTYTE	GYTESCR	LKWQASDHR
YTESCRRCNNVEM	YTESCRRCNNVE	CNNVEMSLWEIPL	SCRCNNVEM	SDHRGTYTESC	QASDHRGTYTESCR
TYTESCR	RCNNVEMSLWEIPLL	RGTYTESCR	SCRCNNVEM	GYTESCRRCNNVEM	WILKWQAS
KWQASDH	EWILKWQAS	HRGTYTESCRRCNNV	PFEWILKWQASDHR	QASDHRG	VEMSLWEIPL
KWQASDHRG	KWQASDHRGTYT	YTESCRRCNNVEMSLW	KWQASDHRGTYTES	DPFEWILK	RCNNVEMSLWEIPL
DHRGTYT	ESCRRCNNVEM	ESCRRCNNVEM	RCNNVEMSLWEIPL	TYTESCR	CNNVEMSLWE
GYTESCRRCNNV	HRGTYTESCR	RCNNVEMSLWEIPL	TESCRRCNNV	WOASDHRGTYTES	SDPFEWILKWQ
HRGTYTESCR	RCNNVEM	TESCRRCNNV	VEMSLW	MSLWEIP	MSLWEIP
NNVEMSLWEIP	KNQASDHR	NNVEMSLW	DHRGTYTESCRRCNNV	TESCR	TESCR
DHRGTYTE	SLWEIPLL	DHRGTYTESCRRCNNV	EWILKWQASDHR	EWILKWQ	EWILKWQ
SCRCNNV	SCRCNNVEMSLWE	TESCR	ILKWQASDHR	RGTYTESCR	ESCRRCNNV
TESCRRCNNVEMSLW	RCNNVEMSLWE	PFEWILK	VEMSLWEIP	ASDHRGTYT	RCNNVEM
WILKWQASDHRGTY	ASDHRGT	VEMSLWEIPLL	ESCR	WOASDHRGTY	FEWILKWQASDHRG
ASDHRGTY	KWQASDHRGTYTE	SDPFEWI	SDHRGTYT	RCNNVEMSL	RCNNVEMSL
WILKWQASDHRGT	ESCRRCNNVEMSL	NNVEMSLWEI	RCNNVEMSL	CNNVEMSL	CNNVEMSL
ILKWQASDHRGTYTE	RGTYTESCR	SDHRGTYTESCR	RGTYTES	RCNNVEMSL	RCNNVEMSL
QASDHRGTYT	HRGTYTE	SDPFEWILK	WOASDHRG	HRGTYTESCRRCNNV	LKWQASDHRGTY
LKWQASDHRGT	WILKWQASDHR	GYTESCRRCNNVEM	CNNVEM	YTESCRRCNNVEMSL	SDPFEWILKW
WOASDHRGT	DHRGTYTESCRRCNN	HRGTYTESC	RCNNVEMSLWEIP	ESCRRCNNVEMSLW	SDHRGTYTE
SDPFEWILKWQASDH	NNVEMSLWEI	PFEWILKWQASD	CRCCNNVEMSLWEI	KNQASDHRGT	KNQASDHRGT
WILKWQASD	RGTYTESCRRCNN	SDHRGTYTESCR	SCRCNNVEMSLWEI	ESCRRCNNVEMSLW	SDHRGTYTE
LKWQASDHRG	SDHRGTYTESCRRCNN	EMSLWEI	QASDHRGTYTES	TYTESCRRCNN	TYTESCRRCNN
PFEWILKWQASDH	QASDHRGT	FEWILKWQASDHRG	FEWILKWQASD	TYTESCRRCNNVEM	TYTESCRRCNNVEM
PFEWILKW	TESCRRCNNVE	CRCCNNVEM	LWEIPLL	SLWEIPL	SLWEIPL
				CRCCNNVEMSLWEI	CRCCNNVEMSLWEI
					GYTESCR

Post-Translational Modifications

Paragon™ Algorithm

Small Protein: **SDPFEWILKWQASDHRGTYTESCRCNNVEMSLWEIPLL**

Post-Translational Modifications

Paragon™ Algorithm

Small Protein:

SDPFEWILKWQASDHRGTYTESCR**CNN**VEMSLWEIPLL

Hot spot

Warm spot

Search for all modifications, likely and unlikely

Only search for likely modifications

Cold spot – *do not search*

Comparison of No-Enzyme Search Results

Immunopeptidome sample

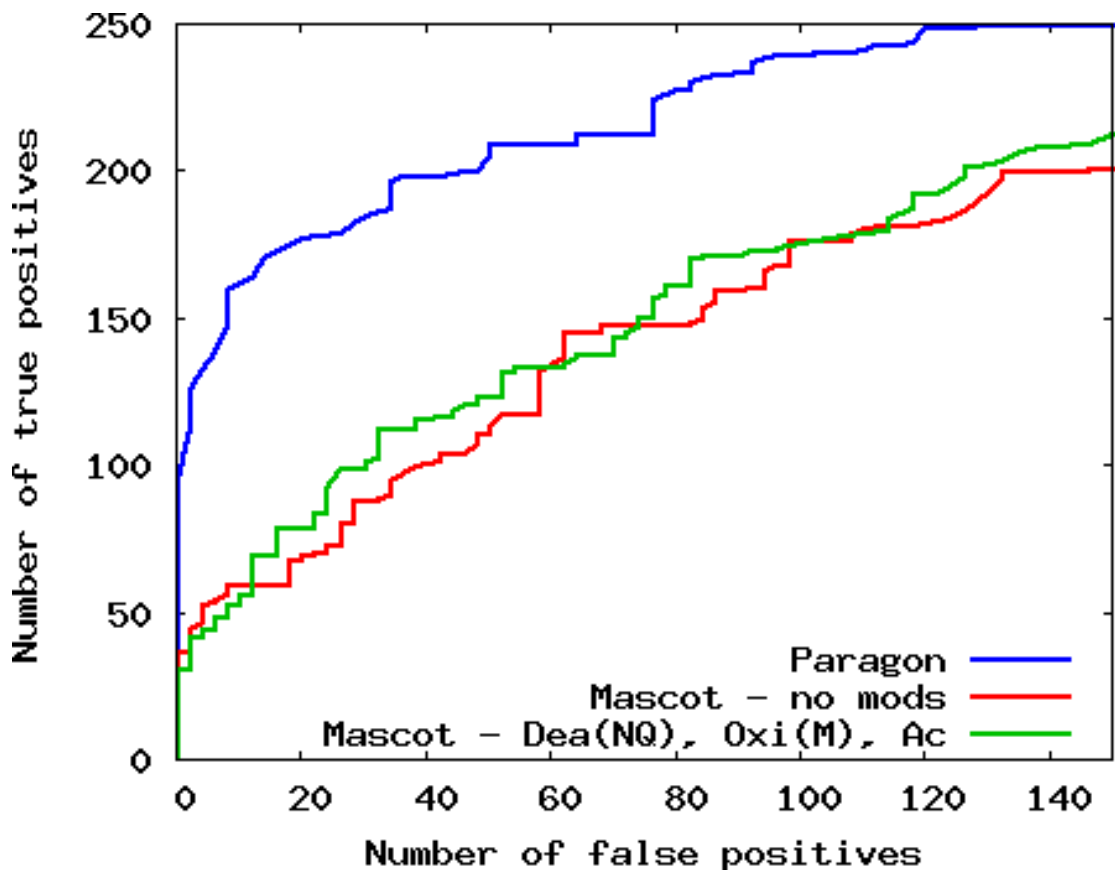
	Number of peptides identified [†]	Total search time [min]
Paragon™ Algorithm 549 features 173 modifications 376 amino acid substitutions (all possible)	160	109
Mascot No modifications	45	104
Mascot Deamidation (N, Q) Oxidation (M) Acetylation (protein N-term)	43	350

[†] Cumulative false positive rate of 5%, assessed by standard reverse database searching

Search database: UniProtKB Swiss-Prot (Dec. 20, 2005) + reversed protein sequences

Comparison of No-Enzyme Search Results

Immunopeptidome sample



Numeric ROC plot

Explaining the Difference in Database Search Results

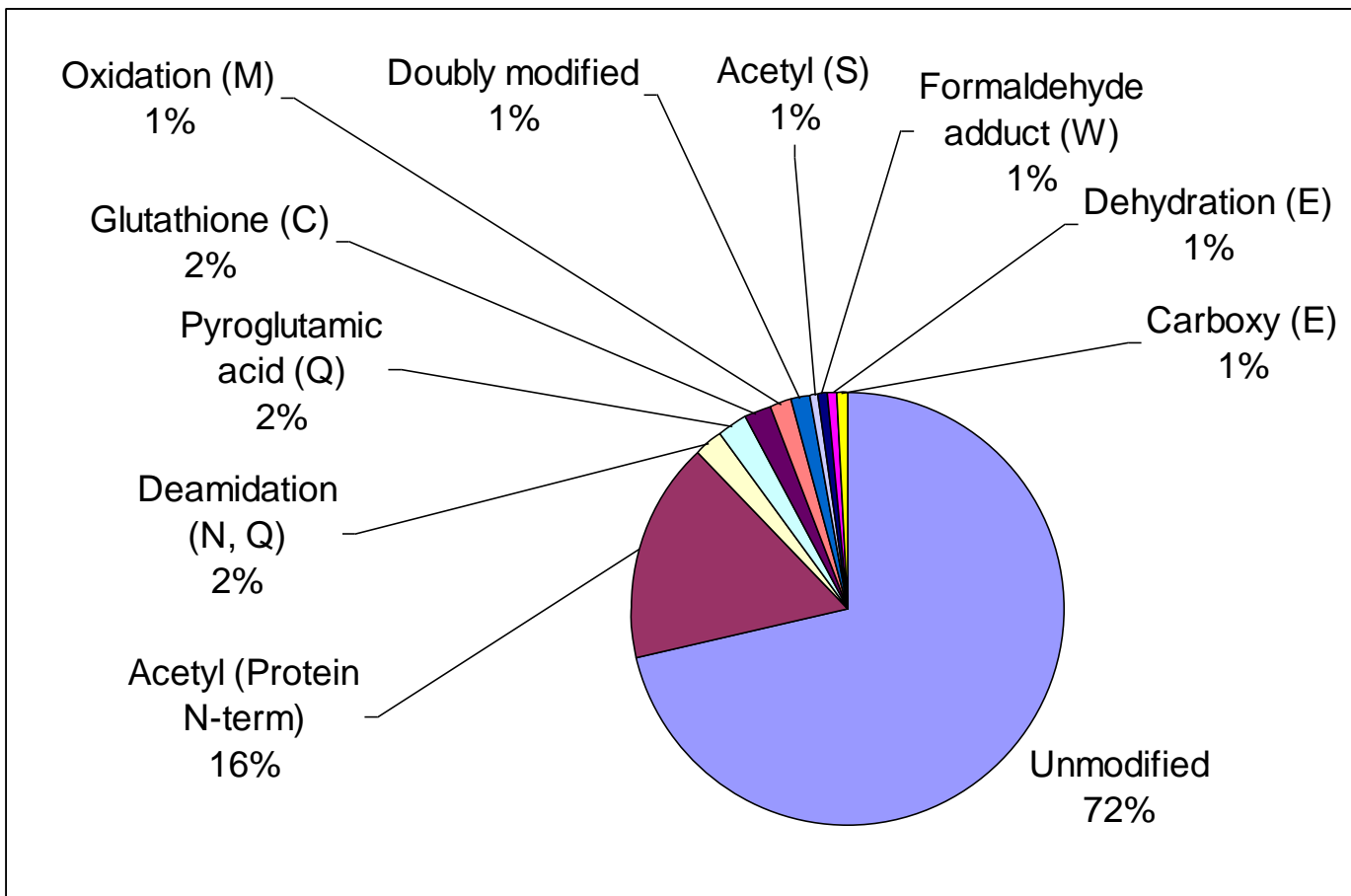
Paragon™ Algorithm

- Finds more PTMs
- **Selective searching**
 - **Scores fewer hypotheses**
 - **Less opportunities to be misled (less “noise”)**

Mascot

- Finds less PTMs
- **Exhaustive searching**
 - **Scores more hypotheses**
 - **More opportunities to be misled (more “noise”)**

Immunopeptidome Post-Translational Modifications Found

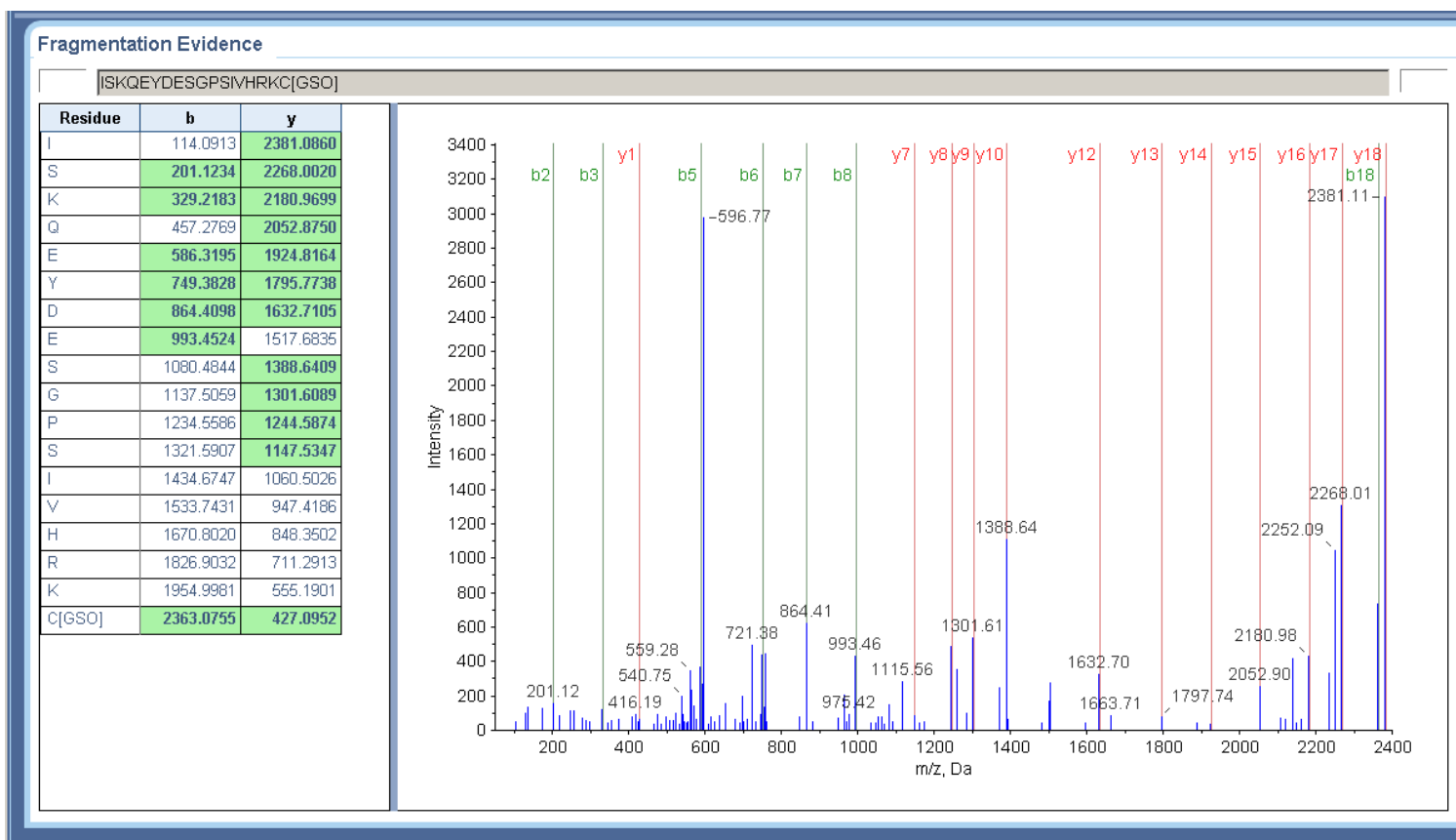


28% of peptides are modified

Peptides with confidence 95% or higher

Glutathionylation of Cysteine

ISKQEYDESGPSIVHRK**C[Glutathione]**



Glutathionylation of Cysteine

- Overlapping peptides

WISKQEYDESGPSIVHRK**C**

WISKQEYDESGPSIVHRK**CF**

ISKQEYDESGPSIVHRK**C**

ISKQEYDESGPSIVHRK**CF**

- We found glutathione on cysteine-374 of actin
- Cysteine-374 of actin is known to be reversibly glutathionylated in vivo

J. Wang et al., *J. Biol. Chem.*, **276**, 47763 (2001)

Conclusions

- We were able to look for a large variety of post-translational modifications in our immunopeptidome samples
- We discovered unanticipated modifications
- A surprisingly high percentage of the HLA peptides identified were modified
- Advantage of the Paragon™ algorithm lies not only in speed but also in reliability of results

Acknowledgments

Applied Biosystems

Sean Seymour
Ignat Shilov
Alpesh Patel
Alex Loboda
Sean Keating
Natalia Belyaeva
Bret Pehrson
Jim Bohannon
Robert Deutschman
Lauren Mansfield
Winnie Leung
Liliya Shilova
Vera Loboda
Lisa Schaechter
Dan Schaeffer
Lydia Nuwaysir

University of Melbourne

Kristine Ford
Anne Zhao

MDS Sciex

Stephen Tate

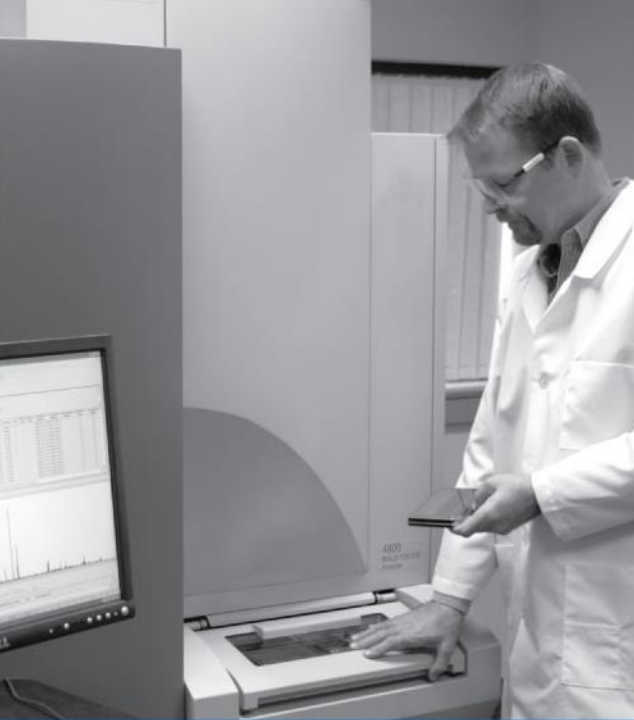
Trademarks and Licensing

Applera, Applied Biosystems and AB (design) are registered trademarks of Applera Corporation or its subsidiaries in the US and/or certain other countries.

Paragon is a trademark and QSTAR is a registered trademark of Applied Biosystems/MDS SCIEX, a joint venture between Applera Corporation and MDS Inc.

All other trademarks are the sole property of their respective owners.

© 2006 Applera Corporation and MDS Inc. All Rights Reserved.



Iterative search

Axiom 1 – For each identifiable protein in the original protein mixture, there will exist at least one detectable tryptic peptide with $n \leq n_{R1}$.

R. Craig and R. Beavis, “A method for reducing the time required to match protein sequences with tandem mass spectra,” *Rapid Commun. Mass Spectrom.*, **17**, 2310 (2003).

(n_{R1} = # of missed cleavages allowed)