

# Effects of Intelligent Data Acquisition and Fast Laser Speed on Analysis of Complex Protein Digests

## **AB SCIEX TOF/TOF™ 5800 System with DynamicExit™ Algorithm and ProteinPilot™ Software for Robust Protein Identification**

Aaron Booy<sup>1</sup>, Kathleen Lewis-Torpey<sup>2</sup>, Christie Hunter<sup>2</sup>  
<sup>1</sup>AB SCIEX, Canada, <sup>2</sup>AB SCIEX, USA

Although mass spectrometry based strategies have been highly successful in the identification of proteins in complex mixtures and in biomarker discovery experiments, sample complexity presents a difficult limitation for in-depth profiling. Many different peptides simultaneously eluting off the LC column with a very wide dynamic range means only a small fraction of the total number of peptides in a sample are actually sampled. This results in abridged proteome coverage and thus new acquisition strategies are needed in order to dig deeper in a shorter amount of time.

The AB SCIEX TOF/TOF™ 5800 System integrates a 1kHz variable rate laser, continuous stage motion, and the DynamicExit™ Algorithm software for faster, more reliable, in-depth data acquisition. The DynamicExit™ Algorithm increases the speed of acquisition by advancing to the next MS/MS experiment once sufficient fragmentation data has been collected for confident identification. This minimizes wasted acquisition time and sharply reduces the sample consumption rate allowing for a greater depth of coverage. Data acquisition rates are further decreased by implementing continuous motion



of the sample stage. By eliminating movement of the sample stage between sub spectrum acquisitions this increases sample throughput and increases the number of peptide identifications by allowing for a greater depth of coverage. These features coupled with a very fast acquisition rate enabled by a 1kHz laser enable rapid, in-depth protein identification experiments.

### **Key Features of AB SCIEX TOF/TOF™ 5800 System for Protein Identification in Complex Mixtures**

- The variable rate 1000Hz laser provides ~5-7 fold increase in MS/MS acquisition speed.
- DynamicExit™ Algorithm acquisition further increases MS/MS acquisition speed (~2x) while maintaining high data quality by intelligently adjusting the MS/MS acquisition time per spectrum based on spectral quality.
- Self-cleaning MALDI source enables regular source cleaning without breaking vacuum for increased instrument up-time.
- ProteinPilot™ software with the novel Paragon™ database searching algorithm identifies peptides from more MS/MS spectra with the ability to search for many peptide features simultaneously.

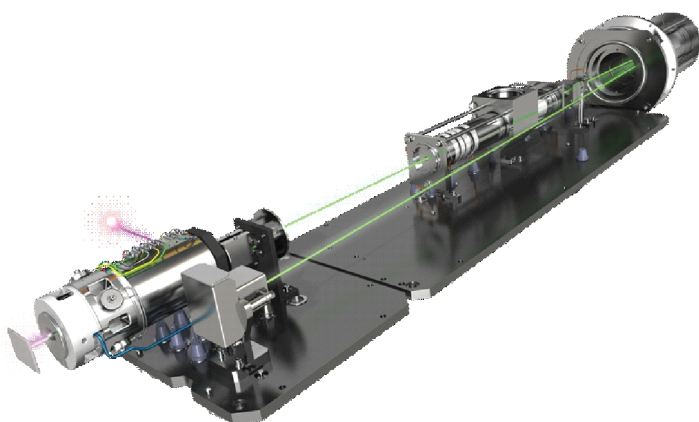


Figure 1. AB SCIEX TOF/TOF™ 5800 System Ion Path.

## Materials and Methods

**Sample Preparation:** A 500 ng tryptic digest of an E.coli lysate was subjected to LC MALDI reversed-phase fractionation using the Tempo™ LC MALDI Spotting System. Peptides were separated on a 100µm ID Chromolith CapRod column (Merck) with a 45-minute acetonitrile/0.1% TFA gradient at 2µL/minute. 720 spots were collected per LC run at a rate of 8s/spot. The matrix used was α-cyano-4-hydroxycinnamic acid (3.5mg/mL).

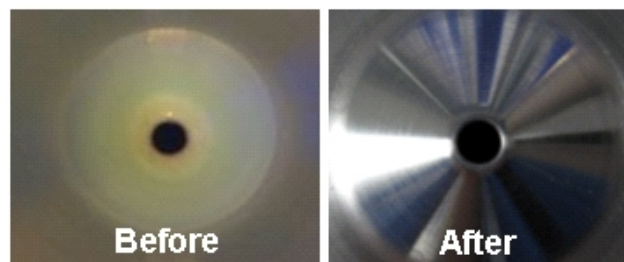
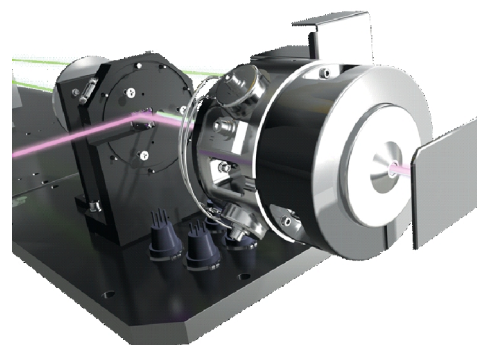
**4800 MALDI TOF/TOF™ System Acquisition parameters:** The MALDI stage motion was set to discrete motion in all acquisition modes (stop and move after each sub-spectrum). MS data was acquired at a laser repetition rate of 200Hz with 1000 laser shots/spectrum (50 laser shots/sub-spectrum). MS/MS data was acquired at 200Hz in 1kV MS/MS mode with 4000 laser shots/spectrum (25 laser shots/sub-spectrum) with the following TOF/TOF Series Explorer™ Stop Conditions:

- Max # shots per spectrum 4000
- Min # shots per spectrum 1000
- # MSMS fragments 8
- S/N of each fragment 75
- Interpretation Method: Top 25 precursors.

**TOF/TOF™ 5800 System Acquisition Parameters:** The MALDI stage was set to continuous motion mode. MS data was acquired at a laser repetition rate of 400Hz with 1000 laser shots/spectrum (250 laser shots/sub-spectrum). MS/MS data was acquired at 1000Hz in 1kV MS/MS mode with 4000 shots/spectrum (250 laser shots/sub-spectrum). Experiments were performed with or without the DynamicExit™ Algorithm using the following exit conditions:

- DynamicExit™ Algorithm: High spectral quality.
- Interpretation Method: Top 25 precursors.

**Database Searching:** Each dataset was processed with ProteinPilot™ Software 3.0 using a database of all E.coli proteins in UniProt as of Jan. 27, 2009 (4379) plus 77 common



**Figure 2. Self-Cleaning MALDI Source for Extended Operation.** After extensive LC MALDI acquisition, matrix can build up on the source. Regular source cleaning keeps the instrument clean and operating at the highest sensitivity.

contaminant proteins. The Paragon™ algorithm was used in Thorough mode with biological modifications and substitutions enabled<sup>1</sup>. False discovery rate (FDR) analysis was done by on-the-fly analysis of the reversed sequences using the embedded PSPEP tool<sup>2</sup>.

## The 5800 System Advantage

The 1kHz variable rate laser in combination with the continuous motion sample stage on the TOF/TOF™ 5800 System means acquisition rates ~7X faster than the 4800 instrument. Such speed increases may rapidly lead to contamination of both the source and laser mirror. In order to negate this effect, the 5800 instrument has a programmable self-cleaning source such that manual cleaning of the source is not necessary (Figure 2). The new laser allows variable repetition rates of 200 and 400Hz in MS mode, 5-400Hz in linear mode and 200, 400, and 1000Hz in MS/MS mode.

**Table 1. The 5800 advantage for Protein Identification.** The hardware improvements on the AB SCIEX TOF/TOF™ 5800 System increase the numbers of proteins and peptides identified in 1/7 of the acquisition time.

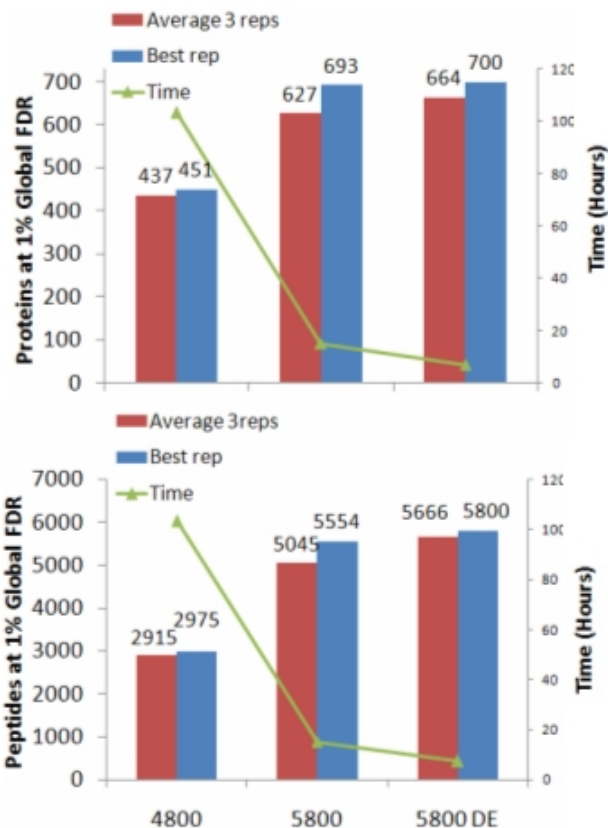
	4800				5800 + 1000 Hz Laser			
	Proteins Global FDR (1%)	Peptides Global FDR (1%)	Total Spectra	Time (hr)	Proteins Global FDR (1%)	Peptides Global FDR (1%)	Total Spectra	Time (hr)
Run 1	451	2945	10288	80.5	693	5554	11477	15.1
Run 2	418	2824	10680	111.0	555	4843	10259	15.6
Run 3	442	2975	9501	118.3	633	4739	11058	14.4
Avg of 3 runs	<b>437</b>	<b>2915</b>	<b>10156</b>	<b>103</b>	<b>627</b>	<b>5045</b>	<b>10931</b>	<b>15</b>

In addition to the faster laser, continuous stage motion and self-cleaning source, the 5800 is configured with a new ion mirror, a new detector and a new digitizer. The effect of all of these improvements on protein identification was assessed by comparing replicate LC MALDI acquisitions on an E.coli cell lysate digest (Table 1). Using the reverse database searching capability integrated in ProteinPilot™ Software, the protein/peptide FDRs were determined to further validate the identification results. The collective hardware improvements provide almost double the number of peptide identifications (1% global FDR) in a significantly shorter time frame (~7x faster).

### DynamicExit™ Algorithm Software

The DynamicExit™ Algorithm determines spectral quality based on an analysis of spectral noise and terminates the MS/MS acquisition once the spectrum reaches a user-defined quality level. MS/MS data on any one precursor is acquired until the spectrum reaches the desired quality level, then acquisition starts on the next precursor. This quality-driven dynamic acquisition both increases sample acquisition throughput and minimizes sample consumption per spot (reduced number of shots acquired per spectrum). Such an acquisition strategy facilitates more in-depth protein coverage per LC MALDI run in a reduced amount of time.

Three replicate LC MALDI runs on an E.coli cell lysate tryptic digest were analyzed with and without DynamicExit™ Algorithm (Table 2) on the TOF/TOF™ 5800 System. The use of the DynamicExit™ Algorithm provided slightly more peptide identifications in half the acquisition time. In addition to faster acquisition times, utility of the DynamicExit™ Algorithm results in less sample consumption and the possibility of deeper coverage through the analysis of more MS precursors (Figure 3).



**Figure 3. Summary of LC MALDI Acquisitions.** Almost double the amount of peptides were identified at the 1% Global FDR using the 5800 with the 1000Hz laser and DynamicExit™ Algorithm vs. the 4800 resulting in a 50% increase in the number of proteins identified.

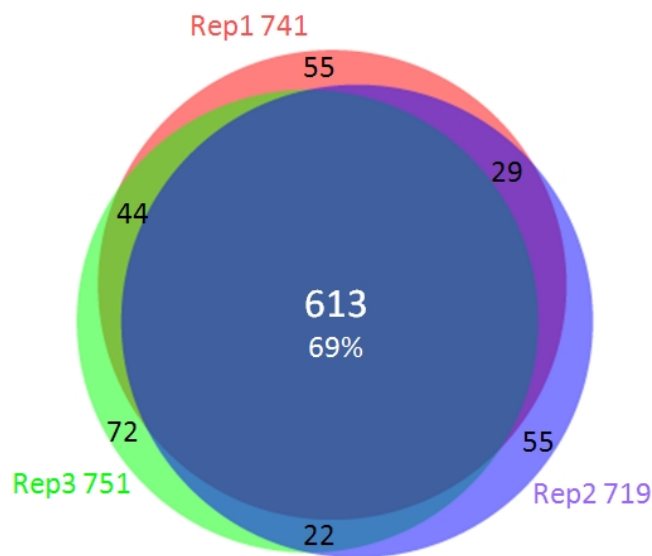
The combination of the fast acquisition rate of the 1kHz laser and the DynamicExit™ Algorithm provides a significant increase in the numbers of peptides and proteins identified in an LC MALDI experiment (Figure 3). Almost double the peptides were identified at 1% global FDR on average with the 5800 system over the 4800 system resulting in a 50% increase in the number of proteins identified.

**Table 2. Advantage of DynamicExit™ Algorithm.** This table illustrates the effectiveness of the DynamicExit™ Algorithm through the resulting decrease in acquisition time and sample consumption leading to more protein and peptide identifications at high confidence levels.

	5800 + 1000 Hz Laser				5800 + 1000 Hz + DynamicExit™ Algorithm			
	Proteins Global FDR (1%)	Peptides Global FDR (1%)	Total Spectra	Time (hr)	Proteins Global FDR (1%)	Peptides Global FDR (1%)	Total Spectra	Time (hr)
Run 1	693	5554	11477	15.1	700	5800	9572	7.0
Run 2	555	4843	10259	15.6	644	5504	9982	8.0
Run 3	633	4739	11058	14.4	649	5695	9333	6.6
Avg of 3 runs	627	5045	10931	15	664	5666	9629	7

## Reproducibility of LC MALDI Acquisition

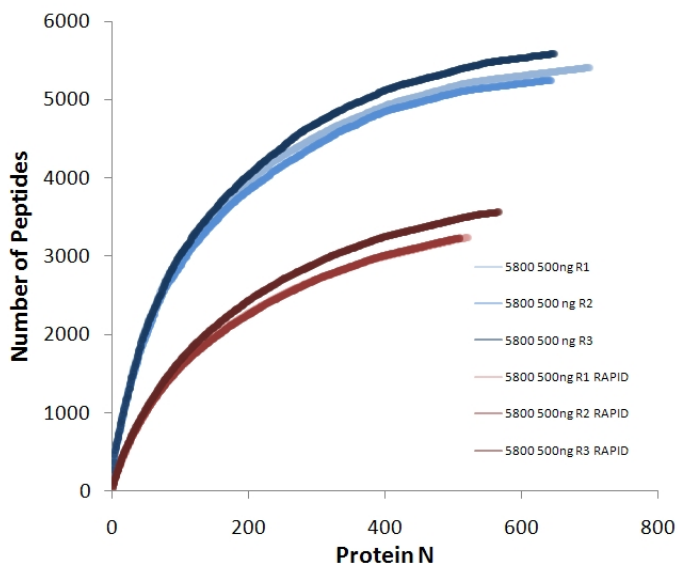
The advantage of the LC MALDI workflow is that the MS analysis is no longer dependent on the time frame of the LC. Using an LC MALDI approach, the rate of the collection of both the MS and MS/MS data is decoupled from the chromatographic separation, allowing as much or as little time as necessary for acquiring MS/MS spectra. Furthermore, more sophisticated selection of precursor ions ensures that all data is collected at the apex of the chromatographic peak, decreasing redundancy and increasing the number of peptides that can be identified from complex mixtures. Figure 4 illustrates the high degree of reproducibility between the 3 replicate runs. When comparing at the 5% global FDR, 69% of the proteins identified are common to all three and ~70-75% are common to any two runs.



**Figure 4. Venn Diagram** Illustrating the High Degree of Reproducibility of the 3 Replicate LC MALDI Acquisitions. Reproducibility within all 3 runs at the protein level is 70% and ~75% when comparing any two runs.

## Advantages of ProteinPilot™ Software

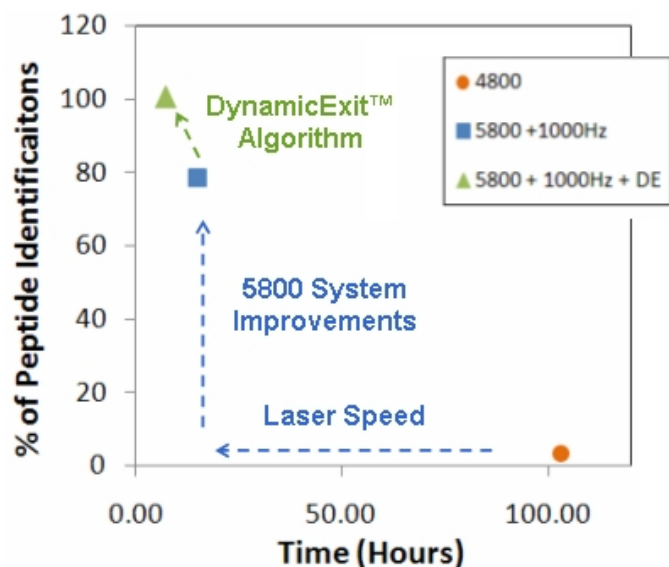
An increase in the number of MS/MS spectra requires more efficient and reliable database searching tools for protein identification. The Paragon™ Algorithm for database searching can efficiently and simultaneously consider about 150 biological modifications, sample preparation modifications and unexpected cleavages to identify more peptides from MS/MS data. The Pro Group™ Algorithm performs a statistical analysis on the peptides found to determine the minimal set of confident protein identifications. The software makes maximal use of MS/MS spectra, while minimizing the reporting of false protein identifications that can be common with other tools.



**Figure 5. Comparing Peptide Identifications using Rapid vs. Thorough Search Efforts.** Many more peptides are found per protein in a Thorough search due to the ability of Paragon™ Algorithm to find many modifications and non-conformant cleavages in a single search.

**Table 3. Biological Modifications and Non-Conformant Cleavages Found with Thorough Searching.** Thorough searching considers an additional 150 modifications and non-specific cleavages not normally considered in conventional search engines. This dramatically increases the number of peptides found, improves sequence coverage and more fully characterizes the identified proteins.

	5800 500 ng DyrExit rep1	5800 500 ng DyrExit rep2	5800 500 ng DyrExit rep3	Average	5800 500 ng DyrExit rep1 RAPID	5800 500 ng DyrExit rep2 RAPID	5800 500 ng DyrExit rep3 RAPID	Average
<b>Modifications Frequencies per Target</b>								
Carbamidomethyl(C)	35.0%	35.5%	35.2%	35.2%	36.0%	36.5%	36.0%	36.2%
Dethiomethyl(M)	37.8%	37.2%	34.4%	36.4%	0.0%	0.0%	0.0%	0.0%
Gln->pyro-Glu@N-term	34.6%	33.9%	35.0%	34.5%	39.4%	38.8%	40.1%	38.9%
Carbamidomethyl(H)	13.2%	13.9%	15.4%	14.2%	0.0%	0.0%	0.0%	0.0%
Glu->pyro-Glu@N-term	11.3%	11.6%	9.7%	10.9%	12.0%	11.7%	12.0%	12.9%
Carbamidomethyl@N-term	10.6%	9.4%	11.1%	10.4%	0.0%	0.0%	0.0%	0.0%
Dioxidation(W)	4.6%	4.9%	4.9%	4.8%	0.0%	0.0%	0.0%	0.0%
Deamidation(N)	3.9%	4.2%	4.6%	4.2%	4.3%	5.0%	5.5%	5.0%
Protein Terminal Acetyl@N-term	3.0%	5.5%	3.5%	4.0%	0.0%	0.0%	0.0%	0.0%
Oxidation(W)	4.7%	3.9%	2.9%	3.8%	0.0%	0.0%	0.0%	0.0%
Carbamidomethyl(K)	3.0%	3.0%	3.1%	3.0%	0.0%	0.0%	0.0%	0.0%
Oxidation(M)	1.2%	2.0%	1.5%	1.6%	2.4%	4.0%	3.0%	3.1%
Trp->Xynurenin(W)	2.0%	1.3%	1.4%	1.5%	0.0%	0.0%	0.0%	0.0%
Carbamidomethyl(E)	1.6%	1.1%	1.2%	1.3%	0.0%	0.0%	0.0%	0.0%
Carbamyl(M)	0.9%	1.4%	1.2%	1.2%	0.0%	0.0%	0.0%	0.0%
<b>Expected Digestion (over-cleavage)</b>								
% Expected termini	51.0%	50.9%	51.7%	51.2%	51.0%	51.0%	51.0%	51.0%
% Semi-specific (only one expected terminus)	5.1%	5.1%	5.2%	5.1%	0.0%	0.0%	0.0%	0.0%
% Non-specific (neither terminus expected)	0.1%	0.1%	0.1%	0.1%	0.0%	0.0%	0.0%	0.0%



**Figure 6. Advantages of Faster Laser Repetition Rate and DynamicExit™ Algorithm on the TOF/TOF 5800 System.** More peptides are found in a much faster acquisition time when a 1000Hz laser is combined with intelligent data acquisition on the very sensitive TOF/TOF™ 5800 System.

Two search modes are available in the ProteinPilot™ Software, Rapid and Thorough. In a Rapid search, the most common modifications and conformant enzymatic cleavages are considered, similar to what a typical search engine such as Mascot would produce (Table 3). A Thorough search considers an additional 150 different potential modifications and non-specific cleavages and therefore finds many more peptides per protein (Figure 5). The additional types of peptides that are found by the Thorough search mode are illustrated in Table 3. For example, there are a larger number of sample preparation modifications that are detected at a significant rate, such as non-typical sample oxidation products and a low percentage of peptides with semi-specific terminal residues (true tryptic cleavage at only one end).

## Conclusions

Herein we have demonstrated the value of the AB SCIEX TOF/TOF™ 5800 System for faster, more reliable in-depth protein profiling of complex mixtures.

- The faster laser, continuous motion sample stage and collective hardware improvements provide almost double the number of peptide identifications (2915 peptides with the 4800 system vs. 5045 peptides with the 5800 System at 1% global FDR) in a significantly shorter time frame (~7x faster).
- DynamicExit™ Algorithm provides an additional 2 fold increase in acquisition speed while ensuring the maximum number of high quality MS/MS spectra, less sample consumption, and therefore more peptide and protein identifications.
- The highly reproducible LC MALDI workflow is advantageous for biomarker discovery experiments.
- Programmable source cleaning provides more reliable and robust instrument uptime.
- The unique searching capabilities of the ProteinPilot™ software 3.0 provides more peptide and protein identifications by finding more modifications and more non-conformant cleavages with no loss in confidence.

The benefits of faster acquisition time and the ability to dig deeper into a sample will have a marked impact on the analysis of a wide range of samples, particularly those exhibiting a large dynamic range of analyte concentrations.

## References

1. Shilov IV et al. (2007) Mol. Cell Prot 6.9, 1638.
2. Tang WH et al. (2008) J. Prot. Res 7, 3661.
3. Improved Depth of Coverage for Single Spot Analysis AB SCIEX Technical Note 0921610-01.

For Research Use Only. Not for use in diagnostic procedures.

© 2010 AB SCIEX. The trademarks mentioned herein are the property of AB Sciex Pte. Ltd. or their respective owners. AB SCIEX™ is being used under license.

Publication number: 0921810-01