Analysis of Clinical Study Variables and Effect on Reproducibility Using the Lucid Proteomics System™

Enrique Dalmasso, Steve Roth, and Amanda Bulman, Bio-Rad Laboratories, Inc., Hercules, CA 94547

Introduction

Protein biomarker development programs center on the discovery, identification, and validation of differentially expressed proteins and peptides. These may serve as biological indicators of altered states resulting from disease, injury, or treatment. Examples include proteins released during cardiovascular injury, shed by or resulting from tumors, or modulated due to drug response. Scientific and clinical interest in novel protein biomarkers and improved panels for existing biomarkers continues to grow, with a goal of better decision making in disease diagnosis and prognosis and in drug discovery applications, such as the elucidation of disease pathways and drug discovery targets.

Proteomic studies are susceptible to bias from both preanalytical and analytical sources, such that observed differences in protein expression may be a result of bias rather than true biological differences related to the disease or treatment of interest. Care must be taken during study design and the development of study Standard Operating Procedures (SOPs) to minimize these biases for increased reproducibility and increased confidence in the results generated.

In this technical note we present a series of experiments addressing preanalytical variability, intra-assay and day-to-day reproducibility, consumable lot-to-lot reproducibility, and lab-to-lab reproducibility within a model clinical biomarker discovery study. The results clearly demonstrate the reproducibility of the proteomic profiles obtained using appropriate study design and experimental controls within the Lucid Proteomics System workflow, and in the manufacture of its components.

Materials and Methods

Sample Preparation

Reference serum samples were enriched for low-abundance proteins using the ProteoMiner™ protein enrichment kit (Bio-Rad Laboratories, Inc.) as described in the instruction manual (Bio-Rad Bulletin 10010636), but using a modified elution buffer (7 M urea, 2 M thiourea, 4% CHAPS, 25 mM Tris-HCl, pH >12.0). Spiked serum samples were prepared by the addition of cytochrome C (~12 kD) to reference serum, then aliquoted and frozen prior to shipment to testers.

Array Preparation

ProteinChip® cation exchange (CM10), anion exchange (Q10), and immobilized metal affinity (IMAC30) arrays (Bio-Rad) were prepared and samples were applied using 96-well ProteinChip bioprocessors (Bio-Rad), in order to contain volumes of 50–200 µl over each chromatographic spot. This containment allows for sample dilution in binding buffer and aids in compatibility with sample buffers containing components normally incompatible with mass spectrometry (high salt, urea, guanidine, etc.). The arrays were pre-equilibrated with their corresponding binding and washing buffers: 100 mM NaOAc, pH 4.0, for CM10 arrays, 50 mM Tris-HCl, pH 9.0, for Q10 arrays, and 100 mM sodium phosphate + 0.5 M NaCl, pH 7.0, for IMAC30 arrays.

Samples diluted 1:10 (ProteoMiner fractions) or 1:100 (unfractionated serum and plasma, or *Escherichia coli* lysate) in the appropriate binding buffer were added to individual bioprocessor wells and incubated for 30 min with shaking at room temperature. Arrays were washed three times with buffer followed by two times with deionized water. After drying, 1 µl of 50% saturated sinapinic acid (in 50% acetonitrile, 0.5% TFA) was applied two times to each spot and allowed to dry for at least 60 min before data collection.

System Qualification

The Lucid™ system qualification kit (Bio-Rad) was used to set up, optimize, and test the ultrafleXtreme or ultraflex MALDI TOF/TOF mass spectrometers (Bruker Daltonics) prior to data collection. The mass spectrometer methods were adjusted to maximize sensitivity and reproducibility of protein profiling in linear TOF mode, including setting of the required detector voltage and laser intensity. The kit includes the premade peptide standard and CM10 ProteinChip arrays, methods, and scripts required to enable evaluation, optimization, and monitoring of the instrument performance prior to initiation of a proteomic study. See the Lucid System Qualification Kit section for additional details.

Data Collection

Mass spectral data were collected with ultraflex series MALDI TOF/TOF mass spectrometers in linear mode using AutoXecute runs which were defined and exported from Lucid proteomics software (Bio-Rad) into flexControl (Bruker). Data

collection methods were mass calibrated with Bio-Rad's QC peptide array from the Lucid system qualification kit, which includes calibrant proteins spanning the 1–12 kD range. flexControl data acquisition methods were used to collect a minimum of 5,000 evenly distributed laser shots from each sample, representing a statistically significant number of shots to ensure the relative quantitation necessary for biomarker discovery studies.

Data Analysis

Raw data were imported into Lucid proteomics software for analysis. Spectra were first grouped into appropriate folders by profiling condition (array type, binding condition, matrix), then processed by baseline subtraction, filtering of electronic noise, setting values for noise calculation, and total ion current normalization. Total ion current normalization standardizes the intensities of a set of spectra in order to compensate for any spectrum-to-spectrum variations caused by minor differences in total protein concentration, array preparation, or data collection.

Automatic peak detection was performed using the Cluster Wizard feature within the Lucid proteomics software. Within specified mass ranges, individual peaks were labeled across all spectra within a folder and clustered based on their m/z values. Peaks meeting specified user-defined thresholds (in these studies, a minimum signal-to-noise ratio [S/N] of 5 and valley depth of 3) were automatically labeled. Peak clusters were created when a given peak exceeding the threshold was detected in 100% of spectra for reproducibility measurements, or in 20% of spectra for biomarker discovery.

Peak intensity coefficients of variation (CVs), automatically calculated by the Lucid proteomics software, were exported and used to calculate the reported median CV values. Hierarchical clustering and principal component analysis (PCA), two nonsupervised multivariate analysis tools within Lucid proteomics software, were used to detect potential bias linked to the day of array preparation. For the model clinical biomarker discovery study, biomarker candidates that exhibit significant differences between sample groups (for example, disease vs. control) were identified by univariate statistical analysis (*P* value, ROC/AUC, fold change). Additional data analysis capabilities of the Lucid proteomics software are described in Bio-Rad Bulletin 5951.

Results and Discussion

A series of experiments (details summarized in Table 1) addressed potential sources of preanalytical and analytical variability and their effects on reproducibility within biomarker studies. Examples of the effects of subtle differences in samples or sample handling highlight the importance of good study design and the need to control possible sources of preanalytical bias. Examples of monitoring reference serum samples in real studies emphasize the advantage of including reference samples to evaluate variability arising from the combination of analytical steps in a proteomics workflow. For the Lucid Proteomics System, these variables include sample prefractionation, array preparation, and data generation. In addition, with the consistent performance of the ProteinChip arrays and when system qualification is combined with appropriate sample handling and SOPs, differentially expressed proteins can be reproducibly detected by multiple independent laboratories.

Table 1. Summary of experiments performed to assess variability in biomarker discovery experiments.

Experiment	Sample	Array Chemistry	Bruker Instrument	Figure	Table
Preanalytical variability	*E. coli* lysate (fresh and frozen)	CM10	ultrafleXtreme	1A	–
	Human serum (2 different sources)	CM10	ultrafleXtreme	1B	–
Intra-assay reproducibility	ProteoMiner-enriched human serum	CM10, Q10, IMAC30	ultraflex II	2	2
Day-to-day reproducibility	ProteoMiner-enriched human serum	CM10, Q10, IMAC30	ultraflex II	3	3
Lot-to-lot reproducibility	Human plasma	CM10	ultrafleXtreme	4	–
Lab-to-lab reproducibility	Human serum (control and spiked)	CM10	7 different ultraflex IIs and IIIs	5	–

Preanalytical Variability

The sensitive and data-rich nature of proteomic profiling with the Lucid Proteomics System provides a clear advantage for revealing differences between samples or sample sets for biomarker discovery and can also help uncover, monitor, and troubleshoot undesired differences, for example, effects of sample handling or differences in the control or stock sample source. Observed differences in protein expression may be a result of preanalytical or analytical bias rather than true biological differences related to the disease or treatment of interest, therefore possible sources of bias must be controlled. If such variables are not well controlled, univariate and multivariate analyses within Lucid proteomics software can help define their possible effects on group classification and determination of biomarker candidates.

The effects of preanalytical variability are visible in the representative profiling spectra of *E. coli* lysates, where a number of low m/z peaks in the profile of a fresh sample are absent or decreased in the profile of a sample following one freeze/thaw cycle (Figure 1A). In a comparison of profiles for two human reference serum samples from two vendors, significant differences in relative peak intensity are apparent in several regions when referenced to the major peaks within the 6–7 kD range (Figure 1B).

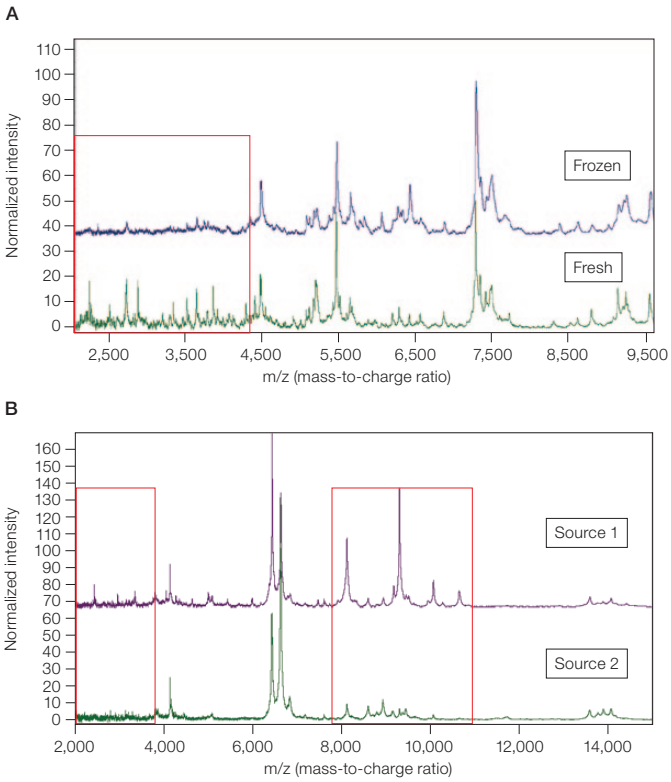


Fig 1. Preanalytical variability. Proteomic profiling can reveal undesired or unexpected differences between samples, for example, effects of sample handling or differences in the control or stock sample source. **A**, profiles of *E. coli* lysates on CM10 arrays reveal differences between a fresh sample and the same sample following one freeze/thaw cycle; **B**, profiles of reference serum samples on CM10 arrays reveal differences between two different suppliers.

Intra-Assay and Day-to-Day Reproducibility

With all analytical measurements, each step in the sample preparation to data collection to data analysis process contributes to overall performance as measured by reproducibility and accuracy. Therefore, each step must be independently optimized and controlled to generate reliable results and provide confidence in the validity of clinical biomarker studies. In these examples, sample prefractionation, the various array preparation steps, and linear TOF mass spectrometer data generation were all performed following well-defined SOPs. This controlled variability for such highly sensitive and data-rich protein profiling provides confidence in biomarker candidates discovered during complex biomarker discovery studies.

Within a clinical biomarker discovery study, 105 patient serum samples and a reference serum sample were processed in parallel by ProteoMiner enrichment. For each of three array chemistries used, the enriched reference serum sample was spotted on four different arrays for 5 nonconsecutive days. The similarity of the overall spectral profiles, or profile reproducibility, is clearly shown for both intra-assay (Figure 2) and day-to-day (Figure 3) comparisons. For each experimental day, median intra-assay peak intensity CVs ranged between 9.3 and 12.5% (Table 2). Peak intensity median CVs were also calculated to define day-to-day reproducibility for each tested array surface. Twenty replicates of the reference samples were randomized along with the patient samples and profiled over 5 separate but nonconsecutive days. Calculated CVs ranged between 15.3 and 17.6% (Table 3).

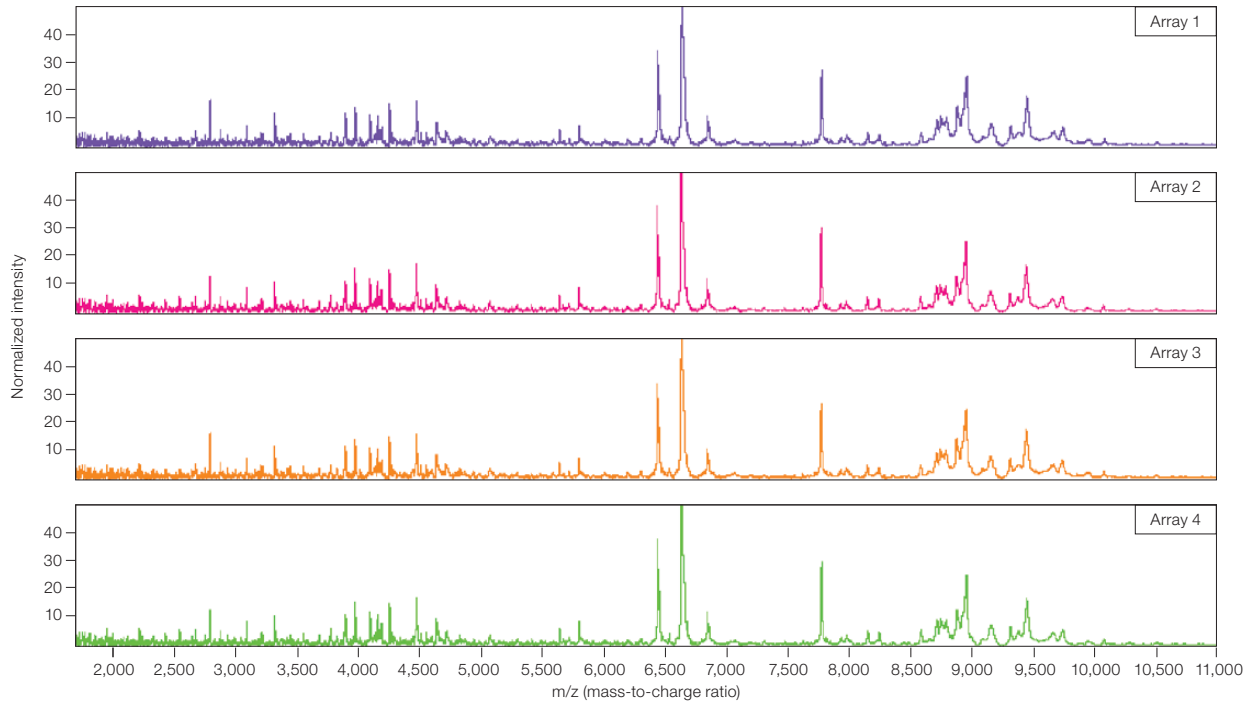
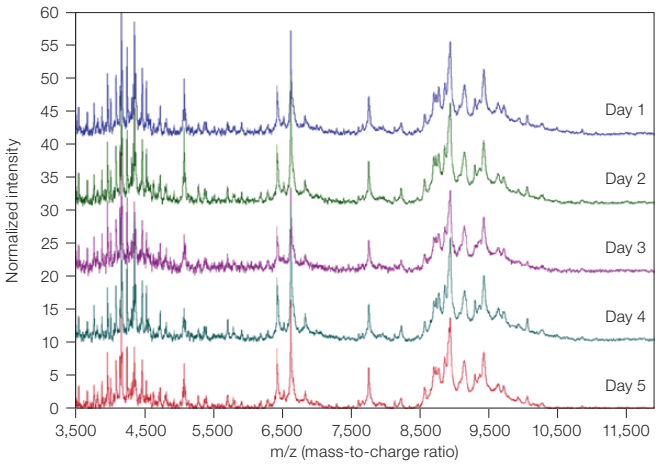


Fig 2. Intra-assay reproducibility. Reproducibility for ProteoMiner bead–enriched reference serum profiled on CM10 arrays across four distinct arrays from one of the 5 different experimental days. Data provided courtesy of Dr Stefan Lehr, Institute of Clinical Biochemistry and Pathobiochemistry, German Diabetes Center, Düsseldorf, Germany.

A



B

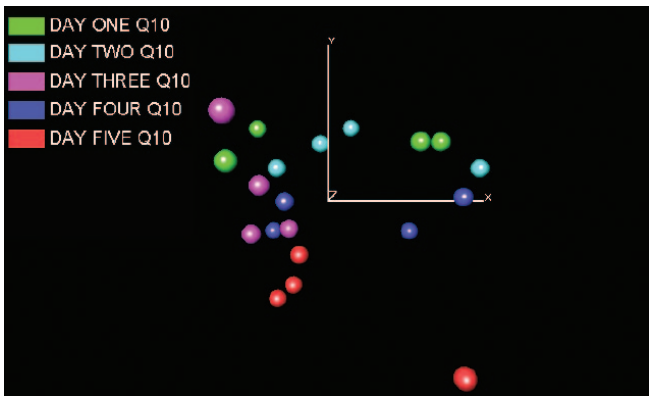


Fig. 3. Day-to-day reproducibility. Replicates of the ProteoMiner bead–enriched serum samples were randomized and profiled over 5 separate but nonconsecutive days. **A**, representative reference serum spectra obtained on Q10 arrays; **B**, nonsupervised multivariate analysis tools in Lucid proteomics software were used to detect potential bias linked to the day of array preparation but showed no significant data clustering. Data provided courtesy of Dr Stefan Lehr, Institute of Clinical Biochemistry and Pathobiochemistry, German Diabetes Center, Düsseldorf, Germany.

These examples illustrate that, when combining all analytical steps in the data generation workflow, including operator and time variables, median CVs of less than 20% can be attained without the need for internal spiking. Although very useful for assay of known proteins, internal spiking is undesirable for biomarker discovery, given the risk of possible profile alterations and masking of other proteins.

Lot-to-Lot Reproducibility

Manufacturing of the ProteinChip array surfaces follows good manufacturing processes (ISO 9001 certified), including production and QC of starting materials, automation during all manufacturing phases, and in-process and final functional QC. Figure 4 shows the reproducibility attainable across lots of arrays manufactured over a 2-year time period is comparable to the day-to-day reproducibility when repeating the same experiment using single lots of arrays. In this case, which adds different lots of arrays to the other experimental variables previously described, a median CV of 16% was calculated from the exported CVs for each of the individual 77 clusters across the different array lots. This enables and supports the performance of experiments over a period of time or between different laboratories utilizing the Lucid Proteomics System in biomarker research applications.

Table 2. Intra-assay reproducibility.

ProteinChip Array	Number of Peaks Detected (average/day)	Peak* Intensity per Day	
		Replicates	Median CV% (average/day)
Cation exchange (CM10)	87	4	9.3
Anion exchange (Q10)	74	4	12.5
Metal affinity (IMAC30)	103	4	11.6

* Peaks detected between 2.5 and 25 kD.

Table 3. Day-to-day reproducibility.

ProteinChip Array	Peak Intensity over 5 Days	
	Replicates	Median CV% (average/day)
Cation exchange (CM10)	20	17.6
Anion exchange (Q10)	20	16.4
Metal affinity (IMAC30)	20	15.3

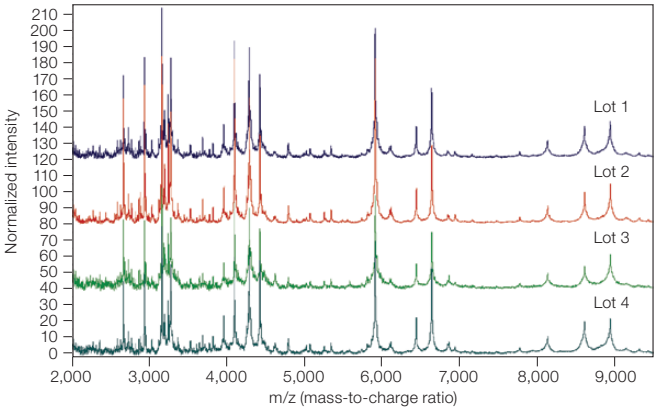


Fig. 4. Lot-to-lot reproducibility. Reproducibility of proteomic profiles across different lots of CM10 arrays of various manufacture dates and ages, tested on the same day using the same reference plasma sample.

Lab-to-Lab Reproducibility of Biomarker Discovery

The ultimate test of an analytical technique is whether the same results can be obtained by multiple operators at different laboratories. Investigators at seven laboratories were provided with all samples, supplies, SOPs, and a full performance of the instrument system qualification prior to analysis. Using their own MALDI TOF/TOF mass spectrometers to collect data and the Lucid proteomics software to detect peaks across the 3–70 kD mass range, they reported the m/z value for all peaks having a greater than 2-fold change comparing control and spiked samples. As a model test system for biomarker discovery, the seven independent laboratories reproducibly detected a single protein spiked into a reference serum sample (Figure 5), without any prior information regarding the spiked protein.

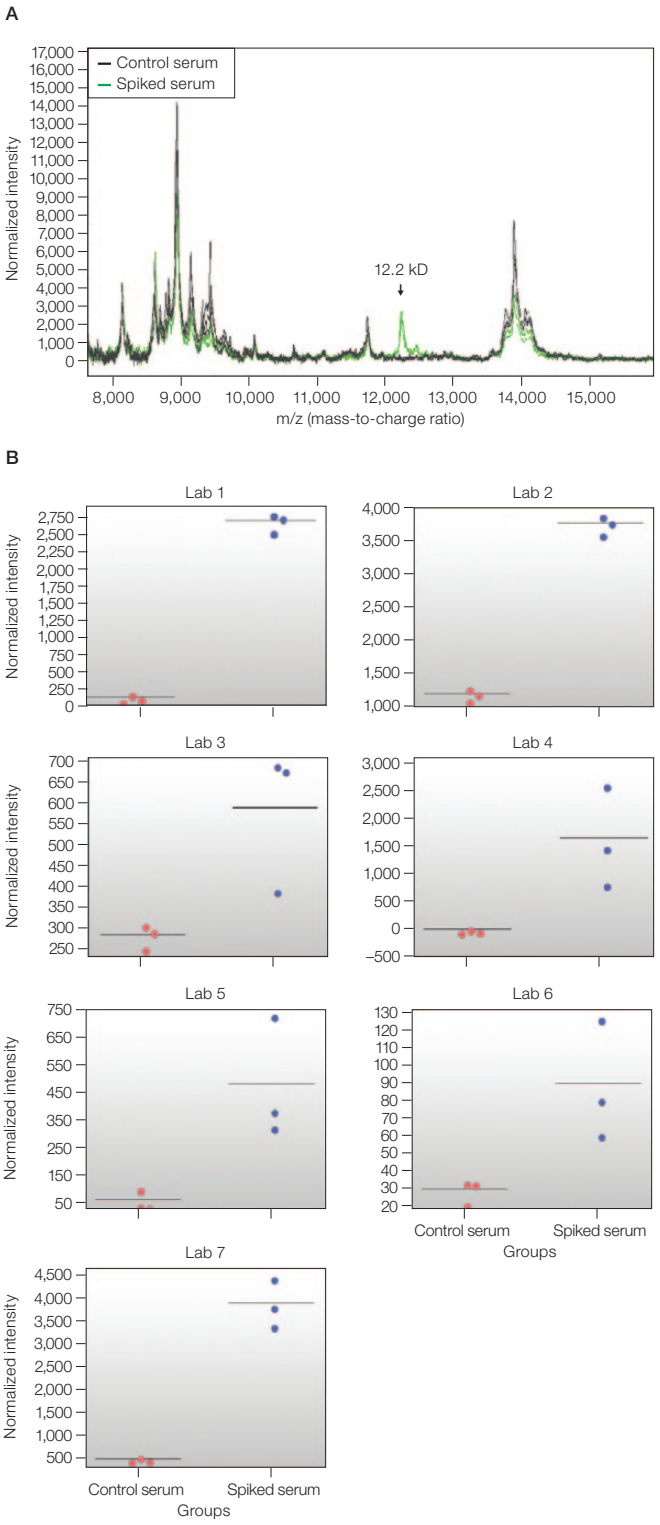


Fig. 5. Lab-to-lab reproducibility. Seven different independent labs worldwide reproducibly detected the same 12.2 kD protein previously spiked into a standard reference serum sample and profiled on CM10 arrays as shown in representative spectra for one laboratory (**A**) and scatter plots of the triplicates run by each laboratory (**B**).

Conclusions

Reproducibility of protein profiling is required for successful, scientifically sound clinical biomarker research. In these experiments, we have tested several potential sources of preanalytical and analytical bias that can impact the reproducibility, and therefore, the reliability, of proteomic profiling using the Lucid Proteomics System. Good study and experimental design, along with implementation of appropriate SOPs, increase the likelihood that observed differences in protein expression are true biological differences, rather than the result of preanalytical or analytical bias.

We have shown that uncontrolled preanalytical bias, such as sample source and sample handling, can contribute significant variability to proteomic profiles and that median peak intensity CVs less than 20% can usually be attained for reference samples processed alongside study samples. The parallel processing of a significant number of reference samples within a biomarker study is imperative, in order to monitor variability caused by the combination of individual analytical steps in the proteomic profiling workflow. Also, the consistent and reproducible proteomic profiling performance of different lots of arrays significantly increases the confidence in clinical studies performed at different periods of time and when comparing results across clinical studies performed by different laboratories.

Finally, using the Lucid system qualification kit to perform routine system optimization, system monitoring, and functionality tests minimizes analytical bias so users can achieve the above results by ensuring the validity and reproducibility of biomarker profiling.

References

Lucid™ proteomics software product information sheet. Bio-Rad Bulletin 5951.

ProteoMiner™ protein enrichment kits instruction manual. 2008. Bio-Rad Bulletin 10010636.

Information in this tech note was current as of the date of writing (2010) and not necessarily the date this version (rev A, 2010) was published.

Lucid System Qualification Kit

Reproducibility and detection sensitivity are ensured by using the Lucid system qualification kit. The kit provides standardization of Bruker MALDI TOF/TOF mass spectrometer performance over time, users, and laboratories by monitoring and adjusting key system performance parameters. The kit is used for both system setup and optimization during a one-time remote web session with a service engineer. The same kit is used for routine system standardization, qualification, and further optimization by the user. Premade ProteinChip arrays, methods, and scripts are provided to enable the setup and optimization steps, and Lucid proteomics software includes tools to rapidly analyze the data. Reports are provided for ease of monitoring the results and for convenient record-keeping.

During setup, system optimization and standardization are performed and specific methods are developed. With too low settings, all peaks are visible but the baseline drifts upward, making analysis and comparison of spectra very challenging. With too high settings, low intensity peaks are suppressed by an increase in noise level. When optimized, S/N ratios and peak intensities are high, with all peaks distinct from a stable baseline across the entire mass range (Figure S1).

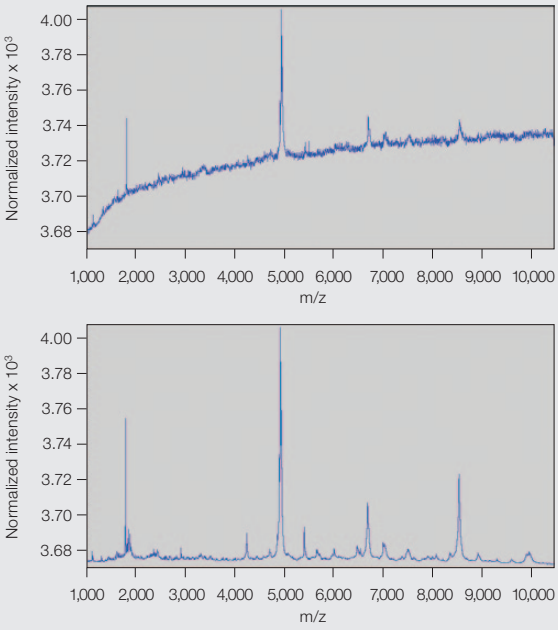


Fig S1. Example of the dramatic effect on profiling results by overall system optimization of several key instrument parameters.

After setup, routine system maintenance includes periodic optimization and component stability checks utilizing the calibration and CM10 reference arrays included with the kit. Performance of components may be optimized and monitored across their lifetime. Periodic overall system functionality tests are also required to ensure optimal performance and reproducibility. The Lucid system qualification kit enables qualification of both the mass accuracy and the sensitivity of the system, using the peptide standard array and the premade CM10 reference array, respectively. The software provides one-click analysis of the results and automatic creation of reports, documenting that the system provides the required accuracy for highly reproducible biomarker profiling results (Figure S2).

Standardized and optimized system performance are absolutely required in today's laboratories to ensure optimal study performance. For mass spectrometry, this is realized through optimization and monitoring of key performance parameters, as well as system functionality tests of mass accuracy and sensitivity. The Lucid system qualification kit provides the materials and scripts to minimize analytical bias and, thus, ensures the validity and reproducibility of biomarker profiling performed with Bruker MALDI TOF/TOF mass spectrometers.

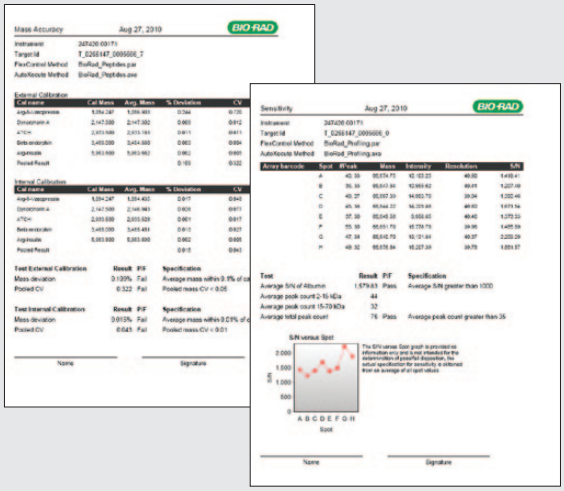


Fig S2. Automatically generated reports of system qualification results.

BIO-RAD

Bio-Rad Laboratories, Inc.

Life Science Group

Web site *www.bio-rad.com* ***USA*** *800 424 6723* ***Australia*** *61 2 9914 2800* ***Austria*** *01 877 89 01* ***Belgium*** *09 385 55 11* ***Brazil*** *55 31 3689 6600* ***Canada*** *905 364 3435* ***China*** *86 20 8732 2339* ***Czech Republic*** *420 241 430 532* ***Denmark*** *44 52 10 00* ***Finland*** *09 804 22 00* ***France*** *01 47 95 69 65* ***Germany*** *089 31 884 0* ***Greece*** *30 210 777 4396* ***Hong Kong*** *852 2789 3300* ***Hungary*** *36 1 459 6100* ***India*** *91 124 4029300* ***Israel*** *03 963 6050* ***Italy*** *39 02 216091* ***Japan*** *03 6361 7000* ***Korea*** *82 2 3473 4460* ***Mexico*** *52 555 488 7670* ***The Netherlands*** *0318 540666* ***New Zealand*** *0508 805 500* ***Norway*** *23 38 41 30* ***Poland*** *48 22 331 99 99* ***Portugal*** *351 21 472 7700* ***Russia*** *7 495 721 14 04* ***Singapore*** *65 6415 3188* ***South Africa*** *27 861 246 723* ***Spain*** *34 91 590 5200* ***Sweden*** *08 555 12700* ***Switzerland*** *061 717 95 55* ***Taiwan*** *886 2 2578 7189* ***United Kingdom*** *020 8328 2000*

Bulletin 6010 Rev A US/EG *10-1087 1010 Sig 1109*