Research Report

Multifactorial Screening Design and Analysis of SELDI-TOF ProteinChip® Array Optimization Experiments

BioTechniques 34:364-373 (February 2003)

H.C. Cordingley, S.L.L. Roberts, P. Tooke, J.R. Armitage, P.W. Lane, W. Wu, and S.E. Wildsmith
GlaxoSmithKline Pharmaceuticals, Hertfordshire, UK

ABSTRACT

Surface-enhanced laser desorption/ionization time-of-flight mass spectrometry is a powerful tool for rapidly generating protein expression data (peptide and protein profiles) from a large number of samples. However, as with any technology, it must be optimized and reproducible for one to have confidence in the results. Using a classical statistical method called the fractional factorial design of experiments, we assessed the effects of 11 different experimental factors. We also developed several metrics that reflect trace quality and reproducibility. These were used to measure the effect of each individual factor, and the interactions between factors, to determine optimal factor settings and thus ultimately produce the best possible traces. Significant improvements to output traces were seen by simultaneously altering several parameters, either in the sample preparation procedure or during the matrix preparation and application procedure. This has led to the implementation of an improved method that gives a better quality, reproducible, and robust output.

INTRODUCTION

The concept of examining and comparing expression data at the genomic and, more recently, proteomic levels for disease profiling and toxicity prediction screens has been gathering interest, particularly in the pharmaceutical industry (3,7,13,14). While the use of microarrays has been examined quite extensively for profiling the transcriptome (15,17), until recently, there has been little available in the proteomic arena that could provide the analysis of a large number of proteins simultaneously. 2D-PAGE has been, and generally still is, the favored tool for proteomic analysis, with high success rates for protein identification (6,10,11). However, it still remains a very resource-intensive process, both at the practical and at the data/image analysis levels, making it unsuitable for routine, medium- to high-throughput screening purposes. The surface-enhanced laser desorption/ionization time-of-flight mass spectrometry (SELDI-TOF MS) system potentially fills this niche, with the ability to generate panels of protein profiles from samples of interest in a relatively simple and rapid manner (4,8,9).

The technology is essentially matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS), based on the use of ProteinChip® arrays (Ciphergen Biosystems, Fremont, CA, USA) as the MALDI target. These protein biochips have been designed with modified areas that affinity capture subsets of proteins from complex biological samples (e.g., hydrophobic, weak cation exchange, strong anion exchange, and immobilized metal affinity chromatography surfaces). Once the proteins are bound, the protein biochips may be washed, removing contaminants such as buffer salts and detergents that prevent the effective ionization of proteins. This enables the profiling of proteins and peptides from complex biological samples. The mass spectra generated from various samples can then be analyzed and compared, with a new expression profile generated for each chip surface that is examined.

As with any proteomic analysis, only a subset of the proteins and peptides present in the original sample is examined. The protein subset present in the final analysis is dependent on a multitude of factors. These include the methodology used to extract proteins, the chip type used, the incubation and wash conditions, the type of matrix used and method of application, and the laser intensity settings.

While the technology is suitable (in principle) for screening based on protein and peptide profiles, the process needs to be optimized to distinguish true differences due to treatment. This entails generating the highest quality traces while maximizing reproducibility to ensure that differences in traces are true (not caused by artifacts), reproducible, and do not change over time. We have devised an approach to identify the best experimental method, based on the assessment of the quality and reproducibility of different combinations of factors. Eleven factors were chosen,

based on previous experience using the hydrophobic chip (Figure 1, A–K and Table 1).

Traditionally, we have used two buffers adapted from the 2D-PAGE methodology (9 M urea or 7.5 M urea with 2 M thiourea). We wished to compare them for their utility in SELDI-TOF. Despite the high urea concentration, proteins may still be degraded due to the presence of proteases in the protein lysate (J. Weekes, personal communication); therefore, we examined the presence or absence of protease inhibitors. Because high-molarity urea buffers are used, samples are maintained at room temperature throughout the sample preparation process to prevent precipitation at low temperature. However, this offers further opportunity for proteolysis, so we decided to examine whether the time that samples were held at room temperature during the homogenization procedure impacted significantly on trace quality or reproducibility. From experience with other biochemical methodologies (e.g., overlay assays, ELISA, Western blot analysis, and immunohistochemistry), it was anticipated that the manner (volume and concentration) in which a fixed amount of protein was presented to an affinity surface over a fixed time period might alter the output trace. To investigate this assumption, the volume/concentration was included as a factor for examination. If the manner of sample application were unimportant, then it would give more flexibility in the future, allowing for the use of more dilute samples if necessary (e.g., when using cell culture samples). The number and length of washes were also examined because these factors were thought likely to affect the output trace.

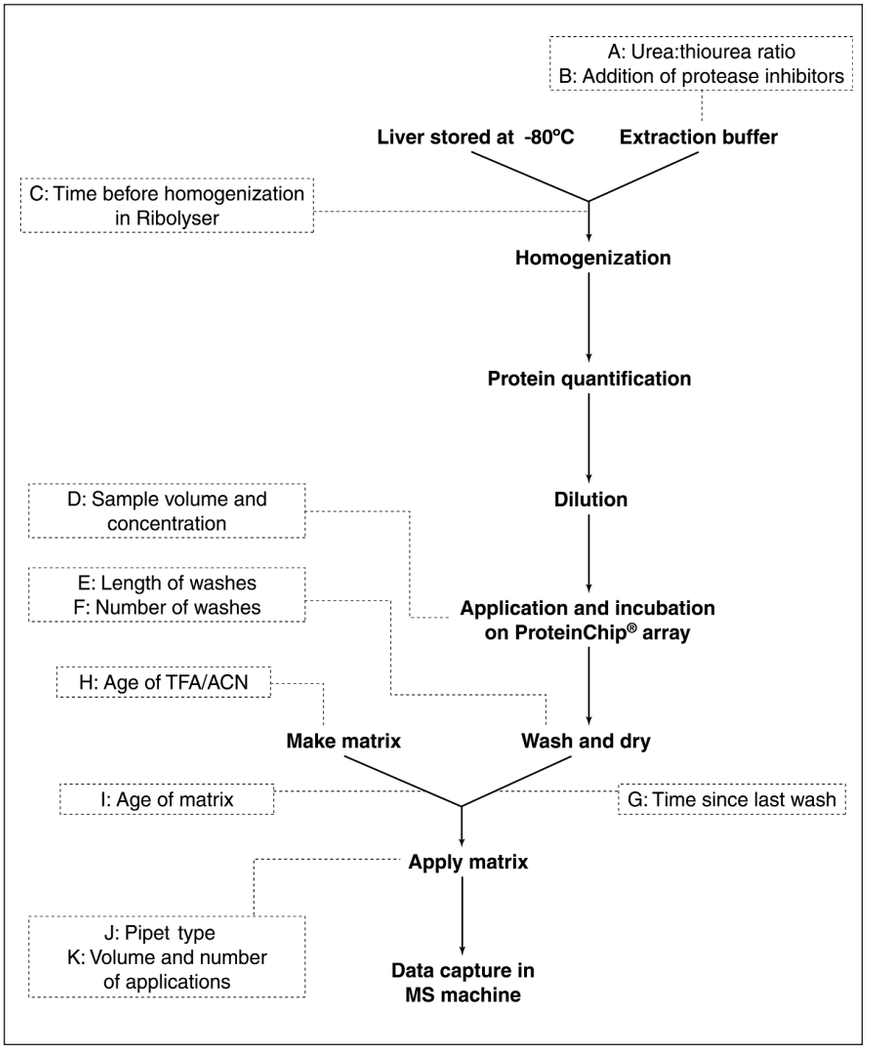

Figure 1. Outline of the SELDI-TOF preparation process.

When using SELDI-TOF, the matrix, or energy-absorbing molecule (EAM), is crucial to the output traces. When the EAM [sinapinic acid (SPA) in this study; Ciphergen Biosystems] is applied, the associated solvent solubilizes the proteins on the chip surface, which co-crystallize with the matrix as the solution dries. These crystals absorb the energy from the laser and generate the ionized proteins that are detected by the mass spectrometer (1). The quality of the matrix solution and reproducibility of deposition are integral to the quality and reproducibility of the traces observed. Because matrix degrades over time in the light, the effect of time on a range of matrix-associated factors (e.g., the age of the EAM solvents and the time elapsed between matrix preparation and application) was investigated to determine whether the time frames used had a detrimental impact on output traces. The time over which the chip was allowed to dry before the addition of EAM was also investigated, as proteins that are well dried onto the chip surface might be more difficult to resolubilize and thus adversely affect the traces generated. Because the matrix addition is so crucial in determining which proteins are ultimately detected with this technology, we also investigated the method of matrix application. In addition to determining whether the type of pipet used impacted the traces, we examined whether our traditional method of applying the matrix in two 0.5 µL aliquots was preferable to applying the matrix in a 1 µL application.

For each of these factors, up to three different settings were chosen (Table 1). A multifactorial experiment was designed, and the results were analyzed for trace quality and reproducibility.

MATERIALS AND METHODS

Practical Procedure

Figure 1 shows an outline of the SELDI process. The standard protocol

Table 1. Altered Factor Settings

ID	Factor	Type	Low level (-)	Midpoint	High Level (+)
A	Urea:thiourea ratio in extraction buffer	Continuous	9.5 M urea	8.5 M urea, 1 M thiourea	7.5 M urea, 2 M thiourea
B	Protease inhibitor[a]	Categorical	Absent	N/A	Present
C	Time between crushing and homogenization	Continuous	5 min	17.5 min	30 min
D	Concentration and volume of sample	Continuous	100 µL at 1 mg/mL	40 µL at 2.5 mg/mL	25 µL at 4 mg/mL
E	Length of washes	Continuous	0 min	5 min	10 min
F	Number of washes	Continuous	2	3	4
G	Time between last wash and addition of matrix	Continuous	5 min	32.5 min	60 min
H	Age of TFA/ACN mixture	Continuous	Immediate	3.5 days	1 week
I	Age of matrix	Continuous	5 min	32.5 min	60 min
J	Pipet type used for matrix deposition	Categorical	10 µL positive displacement	N/A	2 µL Standard Gilson
K	Number and volume of matrix depositions	Categorical	2 × 0.5 µL	N/A	1 × 1 µL

[a]Protease inhibitors (Calbiochem, Nottingham, UK) at final concentrations of 5 µg/mL aprotinin, 5 µg/mL pepstatin, 5 µg/mL leupeptin, 100 µg/mL 4-(2-aminoethyl) benzenesulfonylfluoride, HCL (AEBSF).

is given below, and variations of this are described in Table 1. Ultra-pure reagents, HPLC-grade water, and positive-displacement pipets were used throughout, unless otherwise stated.

Preparation of protein lysates. Liver tissue from untreated male Sprague-Dawley rats was stored at -80°C. Approximately 100 mg frozen tissue were weighed and gently crushed under liquid nitrogen. The resultant powder was placed into 1 mL room temperature extraction buffer [7.5 M urea, 2 M thiourea, 2% CHAPS, and 1% dithiothreitol (BDH Laboratory Supplies, Poole, UK)] and homogenized at room temperature using a Ribolyser (Hybaid, Teddington, UK) for 2 × 10 s on speed setting 4. Homogenization was minimized, and bursts were separated by 1 min intervals to prevent the overheating of the urea solution. Lysates were then removed from the ribolyser tubes and stored at -80°C. Protein concentration was determined using a modified Bradford Assay (12) and reagents from Bio-Rad Laboratories Ltd. (Hemel Hempstead, UK).

Incubation, spiking, and binding process. Eight-spot hydrophobic protein biochips (C_{16} surface H4 biochips; Ciphergen Biosystems) were used for this analysis. Protein lysates were defrosted, diluted to 2 mg/mL in extraction buffer, and centrifuged at 15 800× *g* for 5 min before use to remove any particulates (each aliquot was only used once to prevent freeze-thaw effects). Lysozyme (BDH Laboratory Supplies) was spiked into the sample to a final concentration of 75 µg/mL to act as a reference for normalizing to peak height. Lysozyme was selected because it appears at a molecular weight region where few other peaks are present. "Bioprocessors" (Ciphergen Biosystems) were used, which form separate wells above each spot on a protein biochip and enable loading of volumes up to 500 µL in each well. Sample (50 µL) was placed into each well of the bioprocessor, which was covered with Parafilm® to minimize evaporation and contamination, and then incubated on a circular shaker for 30 min (approximately 400 rpm) at room temperature. After incubation, the samples were removed by inverting the bioprocessor sharply. Three 5 min washes with 350 µL HPLC-grade water were performed, placing the chips on a rotary shaker during each wash. The bioprocessors were removed from the chips, which were allowed to dry at room temperature after carefully removing the excess water from the around the spots with a paper towel.

Matrix preparation and application process. The EAM used for these experiments was SPA. A saturated solution of SPA was prepared approximately 10 min before use by the addition of 400 µL of a freshly prepared 1:1 mixture of 100% acetonitrile (ACN) and 1% trifluoroacetic acid (TFA) (both from BDH Laboratory Supplies) to approximately 10 mg SPA in a light-resistant 1.5 mL microcentrifuge tube. This was mixed vigorously using a vortex and was then centrifuged for 5 min at 15 800× *g*. After air-drying, 2 × 0.5 µL EAM solution were pipetted onto each spot of a protein biochip, allowing time for the first application to dry before the addition of the second. Once the matrix was applied, protein biochips were kept in the dark until they were read using a consistent, automated data-collection protocol in the MS machine (PBS-II; Ciphergen Biosystems).

Data capture. Data were collected using the following settings: 2× warming shots at laser intensity 240 (not collected); collection of 5× laser intensity 230 every five positions between 20 and 80; high mass 50 000 Da; detector voltage 1800 V; and focus mass 12 000 Da.

Design of the Factorial Experiment

Due to the difficulty of examining 11 factors simultaneously in any detail, the factors were split into two groups: those related to sample preparation and application/binding (Figure 1, A–F, the "sample preparation" experiment) and those related to the formulation and deposition of the matrix (Figure 1, G–K, the "matrix experiment"). We thought it unlikely that there would be significant interactions between factors in the two groups. Two experimental designs were created for these factor groups; both were half-factorial designs (2^{6-1} for the sample preparation experiment and 2^{5-1} for the matrix experiment). For the sample preparation experiment, factors G–K were set at previously used defaults, as were factors A–F during the matrix experiment.

Center points were added to each design to confirm that a linear model was suitable for the continuous factors. These were positioned at the midpoint of each continuous factor and repeated at both levels of each categorical factor (Figure 2). All points were replicated four times (Table 1 and Reference 2). The sample preparation design therefore contained 136 measurements, and the matrix design contained 80 measurements. Ideally, the order in which the replicate measurements were taken would have been completely randomized to prevent time trends from influencing the results. However, this was not possible with this technology because of the collocation of eight samples on a single biochip, each of which must receive washes of identical duration and number. Therefore, the measurements were grouped by wash factor levels and randomized within these groups.

Data Preprocessing

Ciphergen ProteinChip software (version 2.0) was used for data preprocessing. Baseline subtraction was performed, followed by normalization of traces for intensity using either an internal or external standard peak. Peaks were then detected using the built-in peak detection algorithm, setting the minimum valley depth and minimum height above baseline to be twice the local noise, with a lower limit of 1000 Da. Data were exported to Microsoft® Excel® for the calculation of response variables.

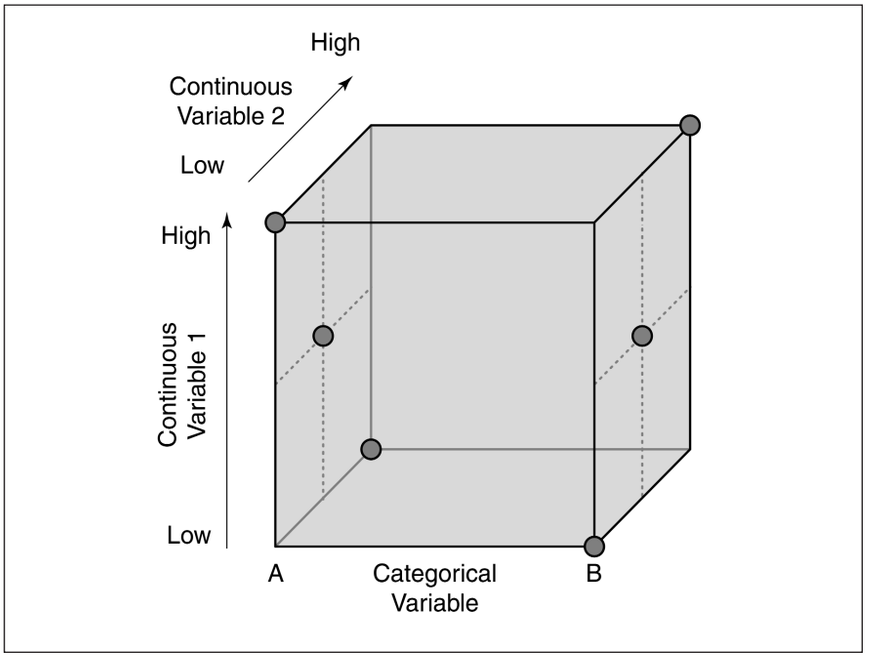

Figure 2. Illustration of midpoints in the design of factorial experiments. A simplified half-factorial experimental design shown here with one categorical factor and two continuous factors. Center points are included in the design, at the midpoints of each continuous factor, and repeated at each level of the categorical factor.

Response Variables

The response variable is a metric by which results are judged. For example, microarray technology conditions have been optimized based on the desired metric of signal-to-noise ratio (16). For the optimization of SELDI-TOF traces for screening purposes, a panel of criteria was identified that describes or contributes to an improved trace. These are: more peaks, "cleaner" peaks (i.e., sharp peaks that are well separated from each other, with smooth lines, with no secondary under lying peaks), peaks spread over a wider mass range (MR), more reproducible traces, traces better suited to an external spike, less noise, and fewer artifacts. In addition, the ease of the experimental procedure was considered. From this list, five metrics were selected that gave a quantitative representation of the aspects of the quality of a trace.

First, the number of peaks detected per trace (NP) was noted. This is important because the more datapoints one collects for analysis, the greater the likelihood of finding patterns of proteins that might reflect a result of interest. Second, the MR was calculated as the difference in the mass between the peaks detected at the highest and lowest molecular weight. This is also important because it is beneficial to examine as wide a range of proteins as possible. Two metrics were developed to assess "peak cleanness" (Figure 3). Peak cleanness 1 (PC1) gave a measure of the shape of the peaks: the average broadness of peaks relative to their mass. It was defined as

$$1/NP \sum M/W^{1/2},$$

(where NP is the number of peaks, M is the mass of a peak, $W^{1/2}$ is its width at half-height, and the summation is taken over all the peaks in the trace). Thus, the aim of this response variable was to

maximize the PC1 value. Peak cleanness 2 (PC2) gave a measure of the minimum (worst) peak resolution from a trace. The resolution between each consecutive pair of peaks was calculated according to the formula

$$\frac{2(M_2 - M_1)}{W_1^{1/2} + W_2^{1/2}},$$

where M is the mass of a peak and $W^{1/2}$ is its width at half-height. The minimum value of this quantity was determined for each trace. An increase in this response variable indicates an improvement in peak resolution. The Ciphergen software output of the preprocessed data occasionally gave no value for the width at half-height of some peaks; these peaks were excluded in the calculation of PC1 and PC2. Finally, for the sample preparation experiment, the suitability of each method for use with lysozyme as an external spike ["spike suitability" (SS)] was ascertained by measuring the minimum separation between the spike and adjacent peaks found on the trace (PC2 limited to a specific area of the trace).

On initial examination of the data, SS was noted to have a strongly positively skewed distribution. Therefore, analysis was performed on the $\log_{10}$ scale, and a check of the residuals confirmed that this transformation gave a response variable with a distribution that was reasonably unskewed and close to the normal.

Treatment of outliers. For two traces, a response variable was greater than the next largest replicate value by more than 10 times the range of the other replicates. These responses were excluded from further analysis.

Analysis of the Factorial Experiment

The experiments were analyzed with two objectives: (*i*) to determine the factors that were important for optimal trace quality and (*ii*) to determine those factors that were important to minimize variation. All analyses were carried out using Design Expert (versions 5.0.7 and 6.0.1; Stat-Ease, Minneapolis, MN, USA).

Optimization of response variables. Each response variable was analyzed using analysis of variance (ANOVA). The main effect and two- and three-way interactions for all factors were determined. Effects and interactions were reported if they were significant at the 0.01 level—a more stringent criterion than usual because of the large number of effects examined in each analysis (31 in the sample preparation experiment and 15 in the matrix experiment).

It should be noted that, because of the fractional nature of the experimental designs, some of the higher-level interaction results can be interpreted in two ways. For example, in the sample preparation experiment with six factors, when the three-way interaction of factors A, B, and C is examined, the three-way interaction of factors D, E, and F is examined at the same time. If the three-way interaction were found to be significant, then it would be impossible to distinguish whether this was due to the interaction of A, B, and C or D, E, and F. (Similarly, if ABF were examined, then this would be indistinguishable from CDE). In this case, the previously calculated main effect and two-way interaction results were examined, and the trio that included a factor with significant effect in these other results was reported as the likely source of the interaction.

For the matrix experiment with five factors, if G, H, and I were examined, then this would be confounded with an interaction between J and K. (Similarly, if GHK were examined, then this could not be distinguished from IJ). In this case, the two-way interactions were deemed more likely to have an effect than the three-way interactions.

Minimization of variation of response variables. To analyze the variation of the response variables, the coefficient of variation of each set of four replicates was calculated for each response variable. The advantages of using this measure are that (*i*) it is on the same scale as the response variables themselves (*contra* the variance) and (*ii*) that it is not correlated with the mean (*contra* the standard deviation), which would contradict the homoscedasticity assumptions behind ANOVA. The coefficients of variation were analyzed using ANOVA as described earlier. Again, two- and three-factor interactions were confounded in pairs, and these interactions are reported according to the preferences previously outlined.

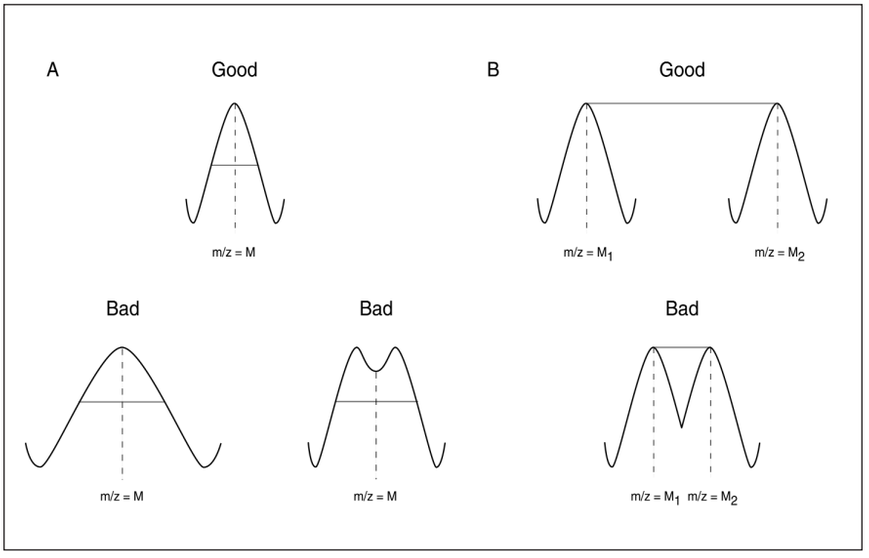

Figure 3. Explanation of PC1 and PC2 response variables. (A) PC1 measures the average broadness of peaks relative to their mass, and (B) PC2 measures the minimum resolution between two adjacent peaks on a trace.

RESULTS AND DISCUSSION

A comparison of a number of sample and matrix preparation protocols was performed, and the impact of these differences were noted, both in terms of

Table 2. Results of the Trace Quality (A) and Reproducibility (B) Optimization Experiments

	(A) Trace Quality				(B) Trace Reproducibility			
Response Variable	Significant factors and interactions	*P*-value	Best settings	Value obtained with best settings	Significant factors and interactions	*P*-value	Best settings	CV (%) obtained with best settings
NP	A	0.0047			K	0.0004	K+	26
	ABF	0.0006	A–, B+, F-	67.9				
	K	0.0031	K–	26.3				
MR	AB	0.0019			GK	0.0061	K–G+	3
	AC	0.0065	A–B+C–	20438	JH	0.0013	J+H+	3
PC1	—	—	—	—	—	—	—	—
PC2	CE	0.0053	E–C+	1.295	E	0.0083	E–	31
$\log_{10}$ (SS)	—	—	—	—	BE	0.0006	B–E+	28

Only factors that were found to be significant ($P < 0.01$) are shown.

trace quality and reproducibility for five different response variables.

Effects of Factors on Trace Quality Response Variables

Table 2 displays the significant results of the optimization for trace quality experiments ($P < 0.01$). Six of the 11 factors examined (A, B, C, E, F, and K) had a significant impact on one or more of the response variables. Only two of these, A (ratio of urea:thiourea) and K (number and volume of matrix depositions), had significant main effects, both on the number of peaks. Factors D, G, H, I, and J (concentration and volume of sample, time between last wash and matrix addition, age of TFA/ACN mixture, age of matrix, and pipet type used for matrix addition, respectively) had no significant effect on any response variable.

Sample preparation experiment. The 9 M urea buffer appears to give an improved number of peaks compared to the urea:thiourea mixture. The strong three-way interaction between A, B, and F ($P < 0.001$) indicates that the addition of protease inhibitors, combined with performing fewer washes, also increases the number of peaks detected. Protease inhibitor incorporation with the 9 M urea buffer (AB interaction) also leads to an improved MR for traces, as does minimizing the amount of time the sample is held at room temperature before homogenization (AC interaction). All of these findings support the hypothesis that the protein degradation of larger molecules can occur during the sample preparation process and that one should take care to minimize this for better quality traces.

The two-way interaction finding between C and E for PC2 (increasing time before homogenization while minimizing the length of wash, which leads to improved peak resolution) is more difficult to explain. Continuing the proteolysis hypothesis, increasing the time before homogenization would allow more time for protein degradation, reducing the number of peaks identified. This, in itself, is likely to lead to an increased PC2 value, as it is likely that the remaining peaks would be more separated on the resulting trace, but the impact of shorter washes is more difficult to rationalize. The washes used in this experiment were pure water washes, and one hypothesis that might help explain this phenomenon is that a short wash simply removes any unbound molecules, but a longer wash might allow an on/off binding equilibrium to be established between the chip surface and the wash solution. This equilibrium would not only allow for the rebinding of peptides that had become removed but, more importantly, would enable the binding of proteins under different molarity conditions.

The fact that factor D (concentration and volume of sample) had no significant effect on the results is intriguing because we anticipated that sample concentration might impact the subsets of proteins that bound to the chip surface, both specifically and nonspecifically. However, the results obtained suggest that the system is reasonably robust with respect to the amount of protein and time frame used here and, therefore, need not be critical for future work.

Matrix experiment. Only factor K (number and volume of matrix depositions) had any significant impact in this experiment; applying the EAM in two smaller volume applications significantly increased the number of peaks that were identified. Exactly how matrix transfers energy to proteins and peptides before flight remains a topic of debate (5). One possible explanation for this finding is that allowing two crystallization events somehow improves the matrix:protein interaction, which results in a better transfer of laser energy to the protein and in more molecules flying.

The fact that none of the other factors (G, H, I, or J) impacted trace quality was unexpected but demonstrates that the range of settings examined have no detrimental impact on traces.

Effects of Factors on Reproducibility of Trace Response Variables

Table 2 displays the significant results of the optimization for trace quality experiments ($P < 0.01$). Six of the 11 factors examined (B, E, G, H, J, and K) had a significant impact on one or more of the response variables. Only two of these, E (length of washes) and K

(number and volume of matrix depositions), had significant main effects, which were on the PC2 and NP response variables, respectively. Factors A, C, D, F, and I (ratio of urea:thiourea, time between crushing and homogenization, concentration and volume of sample, and age of matrix, respectively) had no significant effect on the reproducibility of any response variable.

Sample preparation experiment. Minimizing the length of wash (factor E) improves the reproducibility of PC2, yielding a CV of 31% (compared with 49% obtained with the worst settings). As for the trace quality experiment factor E finding, this is difficult to explain. However, it reinforces the trace quality experiment finding, indicating that short washes should be used.

Contrary to the finding above, a long wash time combined with the non-incorporation of protease inhibitors (the two-way BE interaction), appears to give the best reproducibility for spike suitability with a CV of 28% (compared with 55% obtained with the worst settings). An explanation for this might be that protease action removes peaks that would normally occur in the molecular weight range around 14 300 Da (the molecular weight of our lysozyme spike—supported by the strong ABF finding for the NP response variable from the trace quality experiment) and that a long wash removes these, giving a more open area for the spike. This finding shows that, although the response metrics were developed in an attempt to improve traces, one should take care to ensure that the results are interpreted in a coordinated way when deciding which combination of settings to use for future work.

It is surprising that none of the remaining sample preparation factors has an impact on reproducibility; however, this finding gives an indication of the robustness of the system and allows one to define settings that can be used without adversely affecting reproducibility.

Matrix experiment. Factor K (volume and number of matrix applications) was found to be significant for the NP response variable and the trace quality experiment. However, in contrast to the trace quality experiment finding, a single 1 µL deposition was shown to strongly improve reproducibility ($P < 0.001$), improving the CV to 26% (compared with 41% obtained using the worst settings). The two crystallization event method apparently increases the number of peaks identified, while one application of the matrix improves the reproducibility of the numbers of proteins and peptides that are detected. This latter finding could simply be a reflection of reduced pipetting error with the larger volume. More accurate, automated, reproducible matrix application methods are currently under investigation in our laboratory in an attempt to optimize for both the quality and reproducibility of the NP variable.

The application of matrix as two 0.5 µL aliquots was also found to have a significant impact on the reproducibility of the MR response variable when it was combined with extending the time since last wash (GK interaction). Although altering the settings for these factors had a statistically significant impact on reproducibility, even the worst setting for these two factors gave a CV of 8%, which meets our expectations (a CV of 3% was achieved with the best settings).

Similarly, the two-way interaction between using a standard 2 µL pipet and extending the solvent age (JH interaction) gives an improvement to the MR reproducibility, from a CV of 9% with the worst setting to 3% with the best setting, both of which are acceptable. The apparent improvement of the response variable with the age of TFA/ACN is unexpected because we anticipated that either this factor would have no effect or that minimizing the age of the mixture would give optimal results (because the solvents would not have deteriorated). This appears not to be the case, and this finding warrants further investigation.

Selection of Optimized Settings

An aim of this work was to determine and implement improved settings for future work. Of all the factors examined, only two were not significant for either the quality or reproducibility experiments for any response variable. These were D (sample volume and concentra-

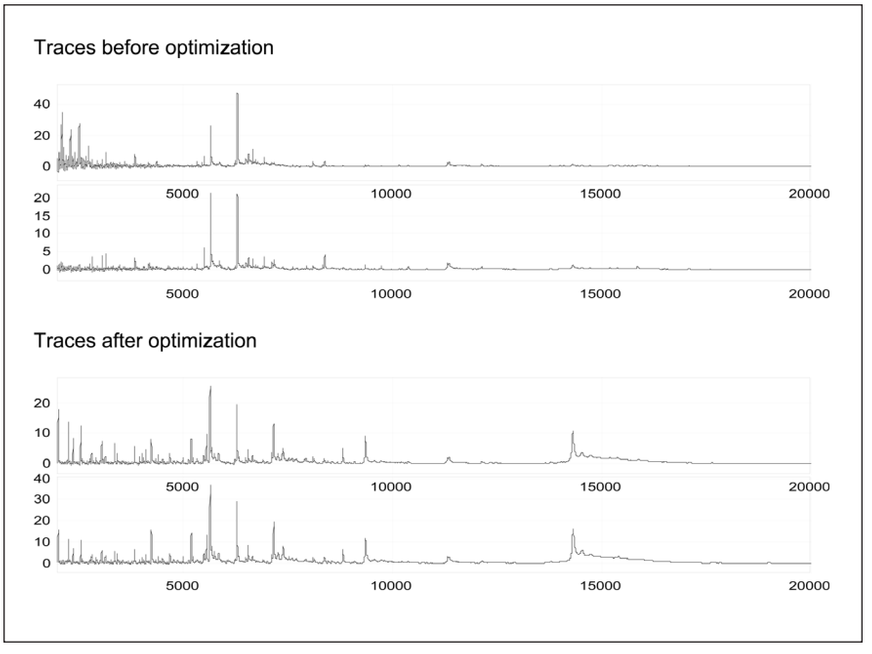

Figure 4. Comparison of SELDI traces using unoptimized and optimized conditions. (Top) Traces generated using unoptimized conditions as described in the Materials and Methods section. (Bottom) Traces generated using newly optimized conditions: 9.5 M urea (A–); protease inhibitors incorporated (B+); 5 min between crushing of sample and homogenzation (C–); two washes of 0 min (E–; F–); 60 min between last wash and addition of matrix (G+); Fresh TFA/ACN mixture (H–); Standard 2 µL Gilson pipet used to deposit matrix in two 0.5 µL aliquots (J+; K–). The sample was applied as 50 µL of a 2 mg/mL solution, and the matrix was prepared 10 min prior to use (D and I set as convenient).

tion) and I (age of matrix). This indicates that both of these factors can be changed as desired, within the range examined here, without impacting either the quality or reproducibility of the traces.

Of the remaining nine factors, the optimal settings for five of them can be readily determined from Table 2. These are to use 9 M urea extraction buffer (A–); minimize the number of washes to two (F–); maximize the time between the last wash and the matrix application (G+); use a 2 µL standard Gilson pipet (J+); and use older TFA/ACN mixture (H+). Because this latter finding was hard to rationalize scientifically and difficult to implement in the laboratory, an examination was made to determine the level of impact that altering this setting might have. Table 2 shows that the best setting, H+J+, gives a CV of 3% for the MR response variable; however, changing this to H–J+ (i.e., using freshly made TFA/ACN with a standard 2 µL Gilson pipet) results in a CV of only 7%. This was deemed acceptable, and the setting of H– (freshly made TFA/ACN) was adopted for future work.

The use of multiple response variables and the aim of optimizing the system for both quality and reproducibility mean that it is possible for certain factors to have different best settings for different response variables. This is the case for the last four factors (B, C, E, and K), and the results were examined in detail to determine which of the settings would give an effective compromise and be implemented in the laboratory in the future. When reviewing the settings in which there was conflict, a weighting system was used to determine which settings to take from this analysis for use. We felt that the NP and MR response variables were more important than the PC1 or PC2 variables, which again were more important than the SS variable. This resolved which settings to use for factors B (incorporation of protease inhibitors), C (minimizing time lapse before homogenization), and E (keeping the length of washes to a minimum). For the remaining factor that remained unresolved (K), it was decided that the quality experiment should override the reproducibility result, as the introduction of robotics might address the variability in subsequent experiments. Figure 4 shows example traces that were obtained using either the unoptimized conditions used prior to this work (Figure 4A) or the new optimized conditions obtained by this work (Figure 4B).

CONCLUSION

Using a statistical approach and devising multiple response variables that characterize SELDI-TOF traces, we have investigated the effect of multiple factors on the quality and reproducibility of SELDI-TOF data produced from rat liver lysates using H4 ProteinChips and derived the best settings for these factors from the range examined. These have been selected as: A–, B+, C–, E–, F–, G+, H–, J+, and K–, with D and I set as convenient (Table 1 and Figure 4).

While some unexpected results were obtained (especially results concerning the TFA/ACN mixture), most of the results are logical, if not anticipated. Sample preparation factors appear to affect trace quality more than reproducibility, while matrix deposition factors impact reproducibility more than trace quality. We anticipate that the introduction of robotics/laboratory automation will continue to improve the reproducibility of the system.

REFERENCES

1. **Beavis, R.C. and B.T. Chait.** 1996. Matrix-assisted laser desorption ionization mass-spectrometry of proteins. Methods Enzymol. *270*:519-551.
2. **Box, G.E.P., J.S. Hunter, and W.G. Hunter.** 1978. Statistics for Experimenters: an Introduction to Design, Data Analysis and Model Building. John Wiley & Sons, New York.
3. **Celis, J.E., M. Kruhøffer, I. Gromova, C. Frederiksen, M. Østergaard, T. Thykjaer, P. Gromov, J. Yu, et al.** 2000. Gene expression profiling: monitoring transcription and translation products using DNA microarrays and proteomics. FEBS Lett. *480*:2-16.
4. **Cordingley, H.C., G.E. Archer, S.E. Wildsmith, and P.J. Bugelski.** Proteomic analysis of hepatotoxicant treated rat liver using Ciphergen ProteinChip® SELDI-TOF technology. 2000. From Genome to Proteome: Knowledge acquisition and representation (4th Siena Meeting), 4–7 Sept 2000.
5. **Cotter, J.R.** 1997. Time-of-Flight Mass Spectrometry: Instrumentation and Applications in Biological Research. American Chemical Society, Washington DC.
6. **Fey, S.J. and P.M. Larsen.** 2001. 2D or not 2D. Two-dimensional gel electrophoresis. Curr. Opin. Chem. Biol. *5*:26-33.
7. **Fielden, M.R. and T.R. Zacharewski.** 2001. Challenges and limitations of gene expression profiling in mechanistic and predictive toxicology. Toxicol. Sci. *60*:6-10.
8. **Fung, E.T., V. Thulasiraman, S.R. Weinberger, and E.A. Dalmasso.** 2001. Protein biochips for differential profiling. Curr. Opin. Biotechnol. *12*:65-69.
9. **Jenkins, R.E. and S.R. Pennington.** 2001. Arrays for protein expression profiling: towards a viable alternative to two-dimensional gel electrophoresis? Proteomics *1*:13-29.
10. **Lopez, M.F.** 1999. Proteome analysis: I. Gene products are where the biological action is. J. Chromatog. B *722*:191-202.
11. **Rabilloud, T.** 2002. Two-dimensional gel electrophoresis in proteomics: old, old fashioned, but still it climbs up the mountain. Proteomics *2*:3-10.
12. **Ramagli, L.S. and L.V. Rodriguez.** 1985. Quantitation of microgram amounts of protein in two-dimensional polyacrylamide gel electrophoresis sample buffer. Electrophoresis *6*:559-563.
13. **Steiner, S. and N.L. Anderson.** 2000. Expression profiling in toxicology—potentials and limitations. Toxicol. Lett. *112-113*:467-471.
14. **Steiner, S. and F.A. Witzmann.** 2000. Proteomics: applications and opportunities in preclinical drug development. Electrophoresis *21*:2099-2104.
15. **Thomas, R.S., D.R. Rank, S.G. Penn, G.M. Zastro, K.R. Hayes, K. Pande, E. Glover, T. Silander, et al.** 2001. Identification of toxicologically predictive gene sets using cDNA microarrays. Mol. Pharmacol. *60*:1189-1194.
16. **Wildsmith, S.E., G.E. Archer, A.J. Winkley, P.W. Lane, and P.J. Bugelski.** 2001. Maximization of signal derived from cDNA microarrays. BioTechniques *30*:202-208.
17. **Wildsmith, S.E. and F.J. Elcock.** 2001. Microarrays under the microscope. Mol. Pathol. *54*:8-16.

Received 17 July 2002; accepted 14 November 2002.

Address correspondence to:

Dr. Hayley Cordingley
Safety Assessment, GlaxoSmithKline
The Frythe, Welwyn
Hertfordshire AL6 9AR, UK
e-mail: hayley_c_cordingley@gsk.com