Journal of Pharmaceutical and Biomedical Analysis 64–65 (2012) 8–15

Contents lists available at SciVerse ScienceDirect

Journal of Pharmaceutical and Biomedical Analysis

journal homepage: www.elsevier.com/locate/jpba

Development and validation of an alpha fetoprotein immunoassay using Gyros technology

Allison M. Given[a], Pamela M. Whalen[b], Peter J. O'Brien[a], Chad A. Ray[a,*]

[a] Pfizer Worldwide Research & Development, Pharmacodynamics, Metabolics and Dynamics, La Jolla, CA 92121, United States
[b] Pfizer Worldwide Research & Development, Oncology Research Unit, La Jolla, CA 92121, United States

ARTICLE INFO

Article history:
Received 15 December 2011
Received in revised form 1 February 2012
Accepted 3 February 2012
Available online 11 February 2012

Keywords:
Biomarker
Validation
Gyros
Alpha fetoprotein
Immunoassay

ABSTRACT

Circulating alpha fetoprotein (AFP) is a diagnostic and prognostic biomarker for hepatocellular carcinoma (HCC) with potential utility as a pharmacodynamic endpoint in rodent tumor models. This application is limited, however, by low sample volumes, highlighting the need for sensitive, sample-sparing biomarker assay methods. In order to improve the utility of AFP as an oncology biomarker, we developed a method for AFP using the Gyrolab™, an automated microimmunoassay platform. Commercially available antibodies were screened to identify optimal combinations that were then used in a multi-factorial design of experiments (DOE) to optimize reaction conditions. Analytical validation included assessments of accuracy and precision (A&P), and dilutional linearity/hook effect, as well as reagent and sample stability. The method is reliable, with total error, a measure of accuracy and precision, less than 30% for all concentrations tested. AFP concentrations were measurable in diseased mice and undetectable in normal mice. Therefore, this novel, low volume AFP immunoassay is suitable for pre-clinical drug development, where its miniaturized format facilitates serial sampling in rodent models of cancer.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

AFP is a 70 kDa glycoprotein produced during embryonic development by the yolk sac and embryonic hepatocytes [1]. The protein becomes repressed during adult life with the exception of patients suffering from liver pathologies including hepatitis C virus infections, cirrhosis of the liver, and some liver tumors, most notably hepatocellular carcinoma (HCC) [2]. This restricted expression makes AFP a valuable biomarker of disease, with a diagnostic sensitivity and specificity of approximately 60% and 90%, respectively at a cut-off of 20 ng/mL in the peripheral blood. The diagnostic sensitivity can improve to nearly 90% with the inclusion of additional markers such as squamous cell carcinoma antigen (SCCA) [3]. Circulating AFP concentrations can also be predictive of outcome. In hepatocellular carcinoma patients treated with concurrent radiation and chemotherapy, individuals showing a 50% decrease in circulating AFP following treatment had better overall and progression free survival compared to the AFP non-responders [4,5]. An early AFP response appears to be a useful predictor of anti-angiogenic response in patients with advanced HCC [6]. Tumor volume is correlated with serum AFP in a murine HCC xenograft model [7]. This relationship between AFP production and HCC tumor burden suggests that circulating AFP might also be a useful pharmacodynamic (PD) endpoint for oncology drug development. In order to meet the small volume demands of mouse tumor orthotopic models, we specifically chose to develop an assay on the Gyrolab immunoassay platform.

Pharmacodynamic (PD) studies benefit greatly from reliable analysis of multiple, serial samples from a single individual, which reduces biological variability and increases confidence in drug response profiles [9]. To this end, we developed an automated microimmunoassay for measurement of circulating AFP in samples from mouse orthotopic tumor models. To our knowledge, this report is the first example of the development of a PD biomarker assay on the Gyrolab platform.

2. Materials and methods

2.1. Materials

Monoclonal and polyclonal AFP-specific antibodies were purchased from R&D Systems (#MAB13691, AF1369), Genetex (#GTX23980, GTX77527, GTX19529) and Abcam (#ab91625). Biotin labeling, for antibodies used as capture reagents, was performed using Sulfo-NHS-LC-Biotin (Pierce, Rockford, IL) at a 12:1 molar ratio of Sulfo-NHS-LC-Biotin (#21327) to antibody per manufacture instructions. Alternately, antibodies were labeled with

* Corresponding author. Present address: 10628 Science Center Drive, San Diego, CA 92121, United States. Tel.: +1 858 526 4905; fax: +1 858 526 4325.
E-mail address: chad.a.ray@pfizer.com (C.A. Ray).

0731-7085/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.jpba.2012.02.001

Alexa Fluor 647 (Invitrogen, Carlsbad, CA) using a labeling kit (#A-20186) per manufacturer instructions. AFP reference standard, purified from human umbilical cord serum, was obtained from Fitzgerald Industries (#30R-AA031). Pooled SCID/beige/beige mouse serum, plasma and CD-1 mouse serum (normal mouse serum, NMS) were obtained from Bioreclamation, Inc. (Hicksville, NY) and stored at -80 °C. Rexxip A (#4820), Rexxip F (#4825) and Superblock buffers were purchased from Gyros AB (Upsala, Sweden) and Scytek Laboratories (West Logan, UT), respectively. All other reagents were considered noncritical.

2.2. Antibody labeling and screening

To determine the optimal capture and detection antibody pair to use for the AFP assay, seven commercial antibodies were tested (Table 1). Four of these antibodies were biotinylated and evaluated as capture reagents at a concentration of 100 μg/mL. Four antibodies were labeled with a fluorescent dye (Alexa Fluor 647; Invitrogen, Carlsbad, CA) and used as detection reagents at 20 μg/mL.

2.3. Gyrolab immunoassay

The Gyrolab platform uses compact discs (CD) fabricated with microfluidic channeling structures that enable parallel nanoliter-scale immunoassays. The Gyrolab instrument also incorporates automated liquid handling and a fluorescence detection system [8]. The Bioaffy 1000 CD contains twelve segments each comprised of eight microstructures. Each microstructure contains a sample delivery channel, a column packed with streptavidin-coated beads as well as a common reagent channel. Gyros technology utilizes capillary action as well as centrifugal force for delivery of reagents and samples. Within each microstructure are multiple hydrophobic barriers, allowing for containment of reagents and buffers and facilitating parallel processing of samples on the CD. These hydrophobic barriers are overcome by an initial short, high velocity spin, followed by a decrease and subsequent incremental increase in speed, which allows reagents and samples to pass through the column. The system can precisely dispense small sample volumes for up to 112 sample analyses. For example, the Gyrolab reproducibly dispenses 200 nL with a precision of 0.75% [8].

For the detection of AFP in mouse plasma, anti-human AFP antibodies were biotinylated or fluorescently labeled as previously described. Standard curves and quality controls (QCs) were prepared and diluted in 100% SCID/beige/beige mouse plasma prior to dilution. All standards, QCs and samples were placed on a Bioaffy 1000 CD and a modified Gyros 3-step method was applied. The 3-step or C–A–D method (Capture–Analyte–Detection) utilizes a two-wash program with a photomultiplier tube (PMT) setting of 1% (Supplemental Table 1).

Supplementary material related to this article found, in the online version, at doi:10.1016/j.jpba.2012.02.001.

2.4. Multi-factorial design of experiments for optimization

Design of experiments (DOE) has been used to optimize immunoassays previously [10,11]. A custom screening design in JMP 8.0 (SAS, Cary, NC) was created using four factors (capture and detection concentrations, buffer composition, and minimum required dilution). The buffer factor contained four options: Superblock, Superblock with 5% NMS, Superblock with 5% NMS and 500 mM sodium chloride (NaCl) and propriety Gyros buffers Rexxip A and F (F is used only for dilution of detection reagents). For each condition, a 7-point standard curve spanning from 4.55 to 4882 ng/mL was prepared in SCID/beige/beige mouse plasma and assayed in duplicate at 4-fold and 8-fold dilutions. In addition, quality control samples were prepared at two concentrations, 1000 ng/mL and 20 ng/mL, representing a high and low QC, respectively. Each QC was assayed 4 times per condition at both dilutions. The concentration values for each replicate were used to estimate the accuracy and precision which was represented by a single output parameter referred to as total error. The total error is the sum of the imprecision (coefficient of variation, %CV) and the absolute value of the inaccuracy (percent difference from theoretical, %PDT)—see Section 2.9 below. The objective was to minimize the total error and maximize the signal to noise ratio at the lowest standard concentration. All reagents were diluted manually and transferred into a single source plate. The Gyrolab Wizard was used to deliver the appropriate reagents to the correct sub-structures. Following the screening design, two factors were kept constant (buffer and dilution) and a response surface design was created using only the capture and detection concentration factors. Again, a 7 point standard curve in duplicate (2.52–1500 ng/mL) and 4 replicates of the high QC (1000 ng/mL) and low QC (3.5 ng/mL) were tested using three concentrations for capture (10, 55 and 100 μg/mL) and for detection (2, 11 and 20 μg/mL). The standard and QC concentrations were adjusted based on the results from the initial screening experiments. As in the screening design, the objective was to maximize the signal to noise ratio and minimize the total error.

2.5. DOE prediction confirmation experiment

A single accuracy and precision run was performed with a 7-point standard curve (2.52–1500 ng/mL) and ten replicates of both the high and low QCs (1000 and 3.5 ng/mL, respectively) to test the JMP prediction. Using Gyros regression software, the calculated concentrations were determined for each replicate.

2.6. Partial validation—accuracy and precision testing

The performance of the assay was determined over three days of accuracy and precision (A&P) testing. AFP was added into SCID/beige/beige mouse plasma to final concentrations of 1500–2.52 ng/mL by performing 2.9-fold serial dilutions. AFP was added into plasma to obtain QCs at 6 concentrations:—the upper limit of quantification (ULOQ), high quality control (HQC), mid quality control (MQC), low quality control (LQC), lower limit of quantification (LLOQ) and the dilutional quality control (DQC). The concentrations at these levels were as follows: 1250, 1000, 62.5, 10, 3.5 and 100,000 ng/mL, respectively. The ULOQ and LLOQ were lower and higher in concentration, respectively, to account for the inability of the Gyros regression software to extrapolate beyond the highest and lowest standard points. The DQC was prepared by adding AFP at 100,000 ng/mL; a 50-fold dilution from the original stock concentration. This control was included to represent a high concentration sample in 98% matrix, and then subsequently diluted an additional 1000-fold into matrix.

2.7. Partial validation—dilutional linearity/hook effect and carry-over

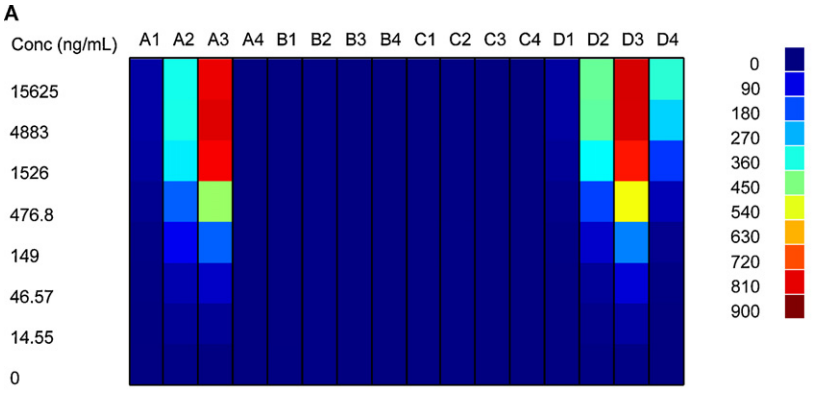
Dilutional linearity samples were prepared by spiking the AFP stock into SCID/beige/beige mouse plasma at a 50-fold dilution. After the initial dilution, the sample was serially diluted 4.5-fold into matrix from 100,000 to 0.59 ng/mL. Within this range, three of the samples (100,000–4938.3 ng/mL) were above the ULOQ, allowing for determination of the presence of a hook effect [12]. An acceptance criterion of %CV < 25% was set for all dilutional linearity/hook effect samples that were quantifiable within the range of the assay.

The potential for carry-over from samples containing high concentrations of AFP was addressed using a needle desorption

Table 1
AFP antibodies.

Antibody	Vendor	Catalog number	Capture	Detection
Anti-human AFP Mab	R&D Systems	MAB13691	X	X
Anti-human AFP Mab	Genetex	GTX23980	X	
Anti-human AFP Mab [6E6]	Abcam	ab91625	X	
Anti-human AFP Pab, Biotinylated[a]	R&D Systems	BAF1369	X	
Goat anti-human AFP Pab	Genetex	GTX77527		X
Rabbit anti-human AFP Pab	Genetex	GRX19529		X
Anti-human AFP Pab	R&D Systems	AF1369		X

[a] Labeled by manufacturer.



B

A1	A2	A3	A4	B1	B2	B3	B4	C1	C2	C3	C4	D1	D2	D3	D4
1.1	2.9	10.8	1.0	0.9	1.0	1.1	1.1	0.9	1.0	0.9	1.1	1.2	2.7	5.4	3.1

Fig. 1. Antibody pair screening heat map. Seven commercially available antibodies were paired in specific combinations and screened for reactivity. Four antibodies (A–D) were biotinylated and used for AFP capture. Four antibodies (1–4) were Alexa Fluor 647-labeled and used as detection reagents. The pairs were examined using full standard curves made by adding purified AFP into either buffer (data not shown) or SCID beige/beige mouse plasma. (A) The results from each experiment were entered into JMP software and plotted as a heat map, where change from low to high response is equivalent to change from the blue to the red end of the visible color spectrum. (B) The signal to noise ratio in the antibody pair screening experiment was determined by dividing the response of the lowest standard point by the response of the blank. The pair with the highest ratio was used throughout the rest of development, validation and sample analysis. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

protocol per manufacturer instructions. Briefly, the Gyrolab needles were washed in 0.5% sodium dodecyl sulfate (SDS) solution followed by a 50 mM glycine wash. Next, a plate containing 24 samples was loaded onto the instrument. The first eight samples contained blank SCID/beige/beige mouse plasma, the second eight samples contained SCID/beige/beige mouse plasma supplemented with 10,000 ng/mL of AFP followed by a third set of eight blank samples. All the samples were diluted 4-fold prior to plating.

Table 2
Summary of validation results.

Validation parameter	SCID beige/beige mouse plasma
Intra assay quality control A&P	3.2 to 23.6% TE −17.5 to 9.7% PDT 1.6–9.2% CV
Inter assay quality control A&P	5.4–20.3% TE −10.9 to 1.3% PDT 4.6–10.7% CV
Calibration range A&P (2.52–1500 ng/mL)	−0.9 to 1.4% PDT 1.5–11.8% CV
Stability	Benchtop: up to 4 h Gyrolab: up to 5 h 4 °C: up to 24 h Freeze Thaw: up to 4 cycles
Dilutional linearity Prozone	Up to 37,367-fold No hook effect up to 100,000 ng/mL

2.8. Partial validation—process temperature and freeze–thaw stability

All stability QCs were prepared by adding AFP to SCID/beige/beige mouse plasma. For process temperature stability, we compared the total error of fresh HQC and LQC samples to QCs incubated at ambient room temperature (ART) for 4 h, at 4 °C for 24 h and for 5 h at ambient instrument temperature (AIT). Samples that had gone through one to four freeze–thaw cycles were compared to freshly prepared HQC and LQC samples. One cycle consisted of freezing the sample at -80 °C for a minimum of 1 h followed by thawing at room temperature. An acceptance criterion of %CV <25% was set for all stability samples.

2.9. Calculations

%PDT and TE were calculated as follows:

$$\%\text{Difference from theoretical (\%PDT)} = \left[\frac{(\text{mean calculated concentration} - \text{theoretical concentration})}{\text{theoretical concentration}}\right] \times 100$$

$$\text{Total error (TE)} = (\%\text{CV} + |\%\text{PDT}|)$$

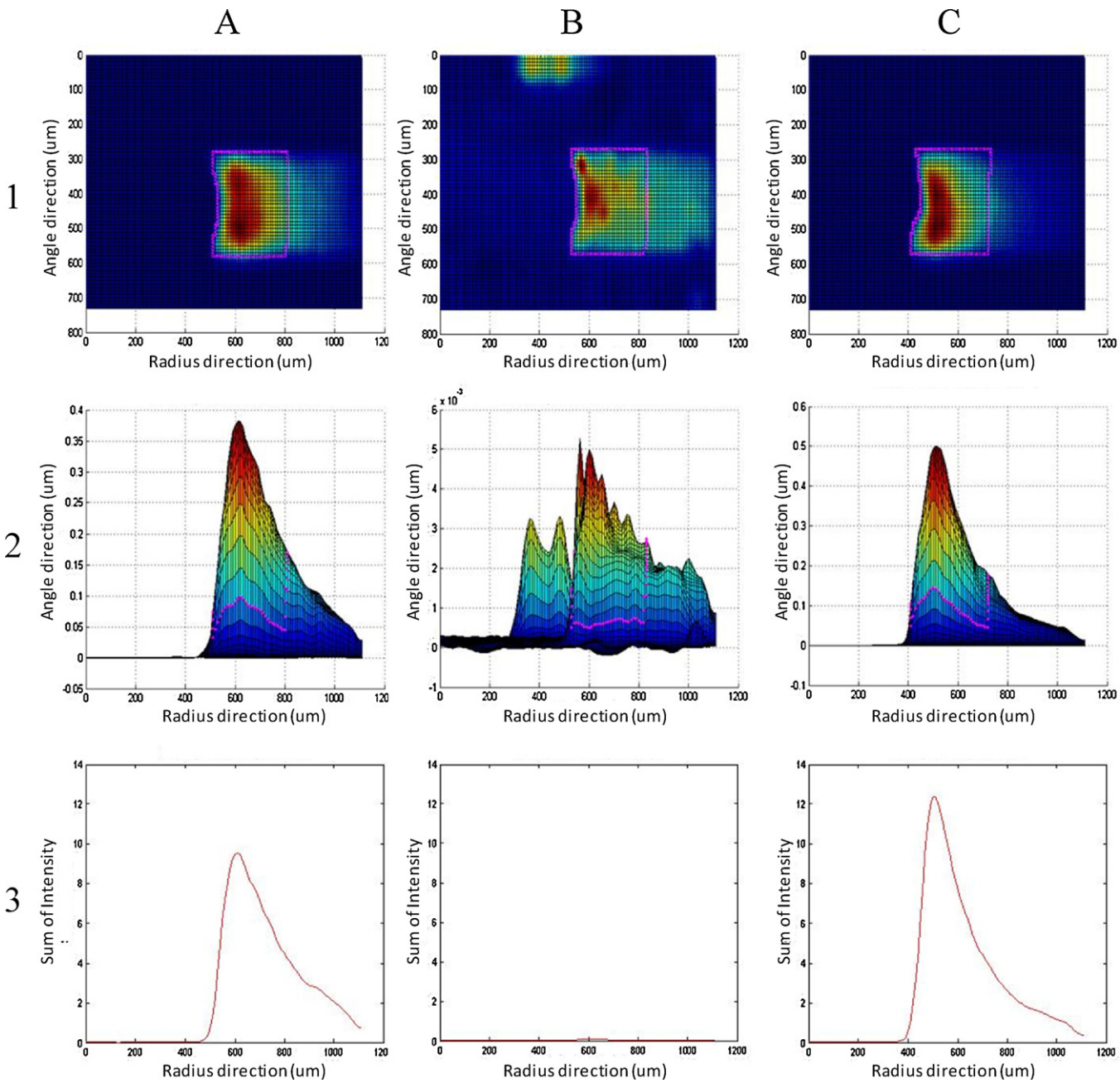


Fig. 2. Gyrolab Viewer images for antibody pair screening. The Gyrolab Viewer program provides various graphical and pictorial representations of column binding characteristics for each individual sample. This information can be incorporated into sample analysis to inform on microstructure and sample integrity as well as antibody reactivity and affinity. Columns A, B and C represent antibody pair combinations A3, C1 and D3, respectively. Rows 1, 2 and 3 show three image types from the software—a top down view of the peak using a heat map correlating to signal intensity, a side view of the same image and a graphical form, respectively. All images show intensity vs. radius direction. Image type 3 was auto scaled to the combination with the highest signal (C3). All images shown compare antibody pair combinations screened against the same sample.

3. Results

3.1. Selection of capture and detection antibody pairs

The labeled antibodies were tested in pairs and used to detect AFP in SCID/beige/beige mouse plasma. The responses for each pair were evaluated qualitatively using a heat map (Fig. 1A) as well as by examination of the signal to background ratio (signal to noise, S:N) at the lowest standard curve concentration (Fig. 1B). According to the heat map, two antibody pairs allow for robust detection of AFP in mouse plasma, A3 and D3. Although the D3 pairing produces the highest raw signal values, the signal to noise ratio is 2-fold higher for the A3 combination (5.4 vs. 10.8). The maximal and blank matrix raw responses for the A3 pairing were 804 and 1.9, respectively, while the responses for the D3 pairing were 823 and 5.46, respectively. Gyrolab Viewer software provides graphs and TIF images of the peak profiles for individual columns. Theoretically, the optimal peak shape is a sharp signal increase at the leading edge of the signal, followed by a relatively rapid decline. This pattern was reflected in the peak shape of the two antibody pairs that exhibited robust signals, A3 and D3 (Fig. 2). In contrast, the C1 antibody pairing (which had a low level response of approximately 2.0 and

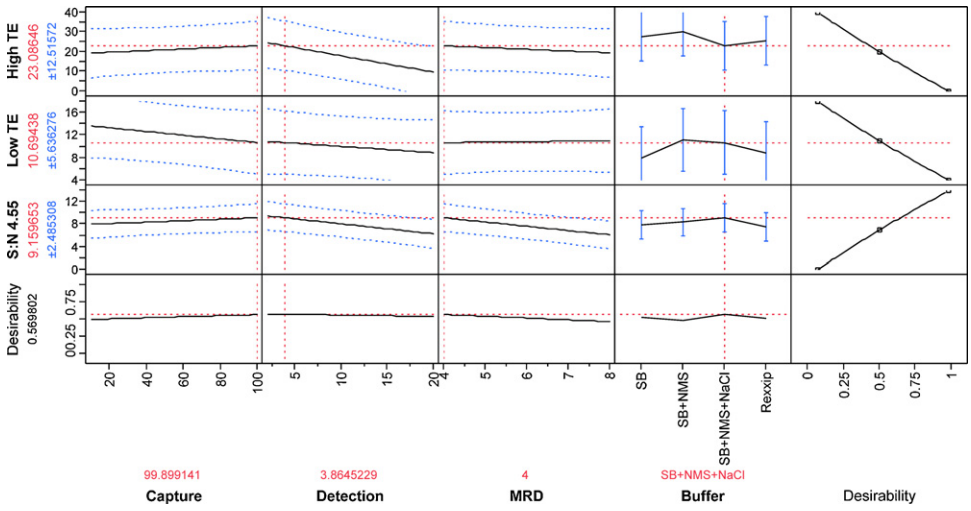


Fig. 3. DOE screening design results. Four factors were tested—capture antibody concentration (Capture), detection antibody concentration (Detection), minimum required dilution (MRD) and assay buffer (buffer). A high and low concentration was tested for each continuous factor. Four different assay buffers were tested—Superblock, Superblock with 5% NMS, Superblock with 5% NMS and 500 mM NaCl (SB + NMS + NaCl) and Gyros Rexxip buffers. Experiments were constructed in JMP and evaluated on the Gyrolab. The results were entered into JMP and the combination of factors that gave the greatest desirability—maximal signal to noise (S:N), minimal high QC total error (High TE) and minimal low QC total error (Low TE)—are represented by the prediction profiler. The optimized factors are displayed on the *x*-axis. The *y*-axis displays the prediction estimates and standard error when using the optimized conditions.

exhibited no significant change between blank and AFP-containing standards), shows a very uneven peak profile which is indicative of fluctuation close to background.

3.2. Evaluation of assay parameters using DOE

A custom multi-factorial screening design was used to test a subset of the 32 possible combinations of conditions in order to minimize the error and maximize the S:N ratio of the assay. The three continuous factors were capture and detection concentrations and MRD (minimum required dilution). A categorical factor was included to identify the best buffer composition. Each continuous factor was tested at a high and low level and four categories were tested for buffer composition. The goal of the model was to find the combination of the tested factors that would minimize the error and maximize the S:N ratio when a desirability function was applied [13]. In JMP software, this type of data is graphically represented by a prediction profiler, with the responses and factors applied across the *y*-axis and *x*-axis, respectively. Fig. 3 shows the prediction profiler after the desirability for each combination of factors was maximized to yield the lowest total error values and highest sensitivity. According to this prediction, the optimal format should be achieved by using antibody capture at 100 μg/mL, detection at 3.83 μg/mL, a 4-fold dilution and Superblock with 5% NMS and 500 mM NaCl.

In order to further characterize the best possible format, AFP assay development was continued using a response surface design [14]. For this analysis, the MRD and assay buffer conditions were kept constant from the previous experiment and additional concentrations for both the capture and detection steps were incorporated. The response surface model (RSM) provided a set of eight experiments that tested capture and detection antibody concentrations at three levels. As in the screening experiment, the High TE, Low TE and S:N data from these experiments were again evaluated as the output in JMP. The prediction profiler suggests that in order to maximize the desirability for each of the responses the assay should use a capture concentration of 100 μg/mL and a detection concentration of 11.1 μg/mL (Fig. 4). Also, nonlinearity was seen in the responses for the capture factor. Statistical predictions for the responses with these conditions were High TE of 11.98 ± 19.36, Low TE of 6.51 ± 33.48 and S:N of 4.21 ± 1.45. The TE ranges were relatively large due to two experiments where the conditions stretched the assay to limits that were not appropriate to obtain reliable data and subsequently led to increased error.

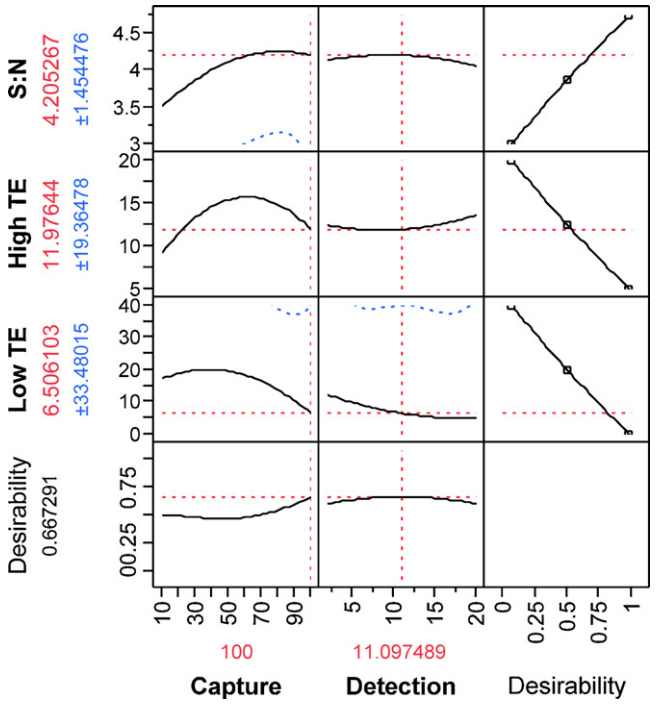


Fig. 4. Response surface design results. DOE in JMP was continued by performing a response surface design for capture and detection antibody concentrations. Three concentrations were tested for the capture (10, 55 and 100 μg/mL) and detection (2, 11 and 20 μg/mL) reagents. The combination of factors that gave the greatest desirability—maximal signal to noise (S:N), minimal high QC total error (High TE) and minimal low QC total error (Low TE)—are represented by the prediction profiler. The optimized factors are displayed on the *x*-axis. The *y*-axis displays the prediction estimates and standard error when using the optimized conditions.

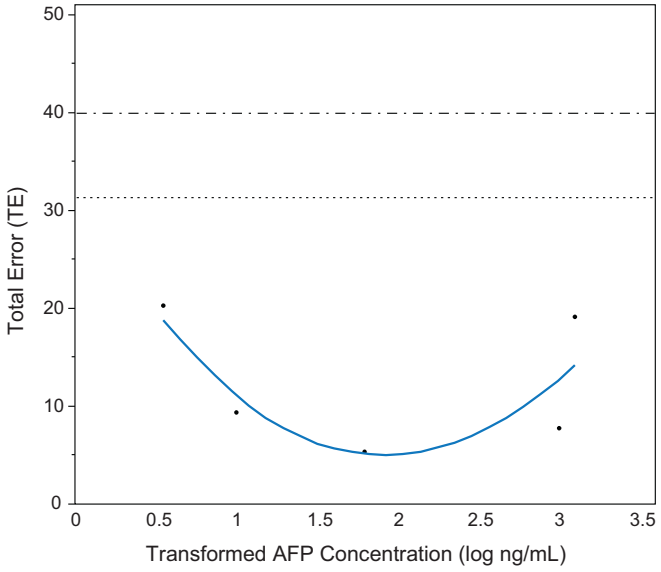


Fig. 5. Total error for quality controls during validation. Total error was calculated and compared against the transformed QC concentrations (1250 (ULOQ), 1000 (HQC), 62.5 (MQC), 10 (LQC) and 3.5 (LLOQ) ng/mL). Total error was highest at the ULOQ and LLOQ (19.2 and 20.3, respectively), reduced at the HQC and LQC (7.8 and 9.4, respectively) and lowest at the mid-point of the assay range or MQC (5.4). The total error estimates for High TE (dashed line) and Low TE (dotted line), calculated by JMP during DOE are graphed at approximately 31 and 40, respectively. All validation QC total errors were within these estimates.

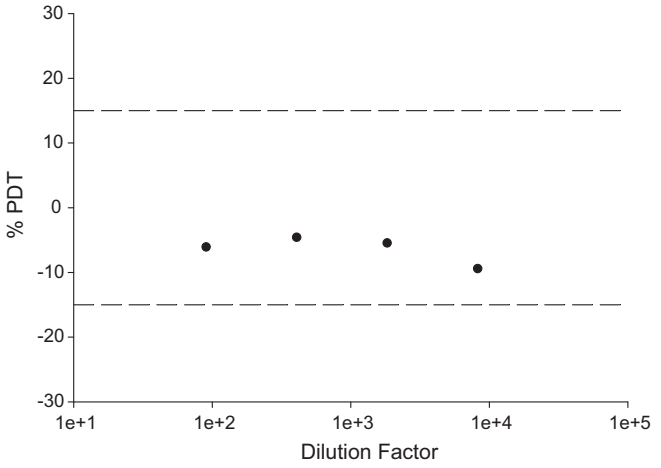


Fig. 6. Dilutional linearity testing. Samples were prepared with 100,000 ng/mL of exogenous AFP and diluted 4.5-fold into 100% SCID beige/beige mouse plasma. For all samples within assay range, the %PDT was within ±15% (dotted lines) demonstrating acceptable dilutional linearity.

3.3. DOE prediction confirmation of the finalized AFP biomarker assay format

To test the accuracy of the DOE prediction, AFP standards and five sets of QCs were run in duplicate and the observed results (TE and S:N) were compared against the predicted results. Total error prediction at the high level (1000 ng/mL) and low level (3.5 ng/mL) and the signal to noise ratio at 2.52 ng/mL were (as stated above), 11.98 ± 19.36, 6.51 ± 33.48 and 4.21 ± 1.45, respectively; compared to the observed results, which were 15.1 and 5.1 and 4.51. Each observed TE value was within the prediction limits as determined by JMP and were within ±27% of predicted.

3.4. Partial validation accuracy and precision

A summary of the parameters and the corresponding results for the partial validation of the Gyros AFP biomarker assay can be found in Table 2. The first step in the validation was accuracy and precision testing over three days by a single analyst. Fig. 4 shows a graphical representation of the A&P data, comparing the transformed QC concentration against the total error. For this biomarker assay, the predicted limits for the total error by JMP were ~40% and the a priori acceptance criteria for the total error were <30% at the HQC, MQC and LQC and <40% at the ULOQ and LLOQ. All QC levels fell within these acceptance criteria (DQC total error was 7.9%; not shown) and, for the HQC and LLOQ, the observed total errors of 7.8 and 20.3% respectively, were within the statistical parameters set by JMP DOE analysis (11.98 ± 19.36 and 6.51 ± 33.48). Also, the total error was lowest at the middle portion of the curve and increased as the upper and lower ends were approached (Fig. 5). The parabolic pattern indicated that an appropriate assay range was selected. This was particularly important, as sensitivity was required to detect early increases in AFP level and a higher ULOQ was also desirable to limit the number of sample dilutions.

3.5. Partial validation dilutional linearity/hook effect and carry-over

AFP concentrations in orthotopic mice were predicted to extend into the mg/mL range, necessitating the ability to dilute samples into the range of the assay and to determine if samples would need to be monitored for a prozone hook effect. During dilutional linearity testing, five of the samples containing exogenous AFP (1097.4–2.68 ng/mL) were within the range of the assay, while the final AFP sample (0.59 ng/mL) was below the quantitative limit (BQL). Hook effect sample concentrations were all above the quantitative limit (AQL) and a hook effect was not observed (data not shown). Dilutional linearity was confirmed to ~37,000-fold based on the %PDT—all samples within assay range had values between −10 and 8. As shown in Fig. 4, plotting 1/dilution against the AFP concentration illustrates the linear relationship of the data with an r^2 value of 0.999 and the average dilution factor between the calculated concentrations was 4.39 (Fig. 6).

The predicted AFP study sample concentrations necessitated the investigation of possible carry-over during sample delivery by the Gyrolab needles. Testing demonstrated that blank samples, delivered after samples containing high AFP concentrations, returned to near previously established baseline levels (Fig. 7). The Gyrolab uses ten needles–eight for sample delivery and two for reagent delivery. By grouping the samples in sets of eight, we ensured that the needles would be exposed to all three sample conditions in the appropriate order: blank, high analyte concentration, blank. Due to the small amount of AFP carry-over in combination with the high concentration of plasma in the samples (25%), CVs for sample replicates were monitored and the emergence of any upward trends in %CV would lead to the initiation of a needle cleaning protocol.

3.6. Partial validation stability testing

During analysis, samples may be left on the bench top or stored for short periods of time at 4 °C during run preparation. Multiple rounds of freezing can also occur if repeat analysis is called for due to run failure. For process temperature stability during plate based validations, the total error of HQC and LQC samples at ambient room temperature (ART) and at 4 °C is typically evaluated. However, the Gyrolab presents a unique microenvironment. The temperature within the Gyrolab instrument becomes significantly warmer than ART, leading to possible degradation of samples during extended

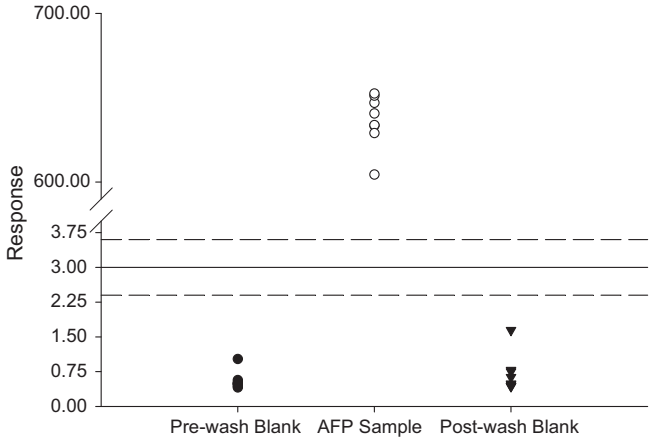


Fig. 7. Carry-over testing. Gyrolab needles were initially desorbed to remove any potential contaminants using 0.5% SDS followed by 50 mM Glycine solution. Blank samples (Pre-wash Blank) were analyzed to determine background followed by samples containing AFP at 10,000 ng/mL (AFP sample). A second set of blank samples (Post-wash Blank) were run to understand contamination from carry-over. The average response for the LLOQ of the assay is plotted (solid line) as well as 20% error (dashed line). All blank responses fell below this response, indicating that the increase in measured AFP in the Post-wash Blank was not significant and confirming that the wash system (Wash Buffer 1: 1× PBS with 0.1% Tween-20; Wash Buffer 2: 1.5 M NaCl in 20% ethanol; Pump Liquid: 1× PBS with 0.1% Tween-20) for the assay was acceptable.

runs. Therefore, process temperature stability was examined in AFP-containing samples under the following conditions: freshly prepared, 4 h at ART, 4 h at 4 °C, 24 h at 4 °C and 5 h inside the running Gyrolab. The %PDT ranges for the samples tested at these conditions were 3–17% at the HQC and −15 to 0% at the LQC. The results of the freeze–thaw stability testing are shown in Fig. 8. The %PDT ranges for the freeze–thaw samples were −8 to 17 at the HQC and −1 to 10 at the LQC, well within the acceptance criteria of ±30%.

3.7. In-study performance and results

The robustness of an assay is more thoroughly understood during sample analysis. This can be assessed by pass/fail rates and by monitoring the performance of the standards and QCs in-study.

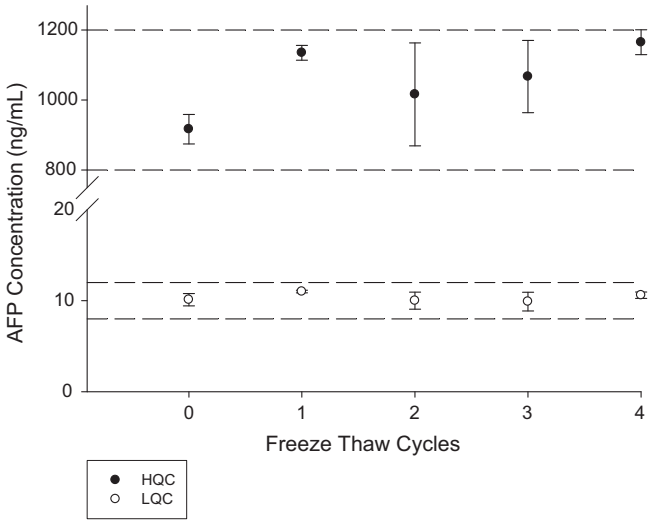


Fig. 8. Freeze–thaw stability. AFP was added into 100% Beige SCID mouse plasma at the HQC (1000 ng/mL) and LQC (10 ng/mL), aliquoted and stored at −80 °C for a minimum of 12 h. Samples were thawed between 1 and 4 times and subsequently analyzed on the Gyrolab in duplicate. All stability samples recovered within ±20% (dashed line) of the nominal value.

This data is also useful to compare back to the statistically determined DOE total error values, as it represents an overall increase in variability (time, reagent lot variability, etc.) and provides more information on the accuracy of the prediction. Analysis of >400 samples over 24 CD runs, resulted in total error values of 9.2, 8.0 and 10.2 for the HQC, MQC, and LQC, respectively. The assay had a pass rate of 100% with the deactivation of 3 individual QCs, one at each level, for the entire study. Prior to analysis of study samples, we measured AFP concentrations in ten normal SCID beige/beige mice. All individuals had AFP levels that were below the LLOQ of the assay (<2.52 ng/mL). In comparison, AFP levels in ten orthotopic mice ranged between 1300 and 230,000 ng/mL after ten days of vehicle treatment and 487,000 and 5,860,000 ng/mL after thirty days of vehicle treatment (data not shown).

4. Discussion

We have developed a microimmunoassay on the Gyrolab platform that is suitable for pre-clinical drug development. Specifically, we applied fit-for-purpose biomarker assay validation principles to ensure that the resulting assay was reliable enough to inform decisions regarding compound efficacy and accurately describe drug exposure-response relationships. In this process, antibodies were carefully selected and assay conditions were optimized using multi-factorial DOE. To test the suitability of the resulting method for the purposes at hand, a partial validation of key analytical factors was performed, and the assay was used to measure samples from a mouse orthotopic tumor model of HCC.

Antibody screening was relatively straightforward, as the analyte is well established, reagents are available through multiple vendors, and the natural protein exists in sufficient quantities to prepare reference standards. The screening process yielded two pairs that demonstrated acceptable dose-dependency and provided robust assay signals at the highest standard concentration. In order to select the final assay format, we evaluated assay signal at the low standard concentration compared to blank (S:N), matrix tolerance, and binding properties on the column. From these parameters we determined that A3 was the best pair with a S:N ratio 2-fold higher than D3. A3 behaved similarly across all matrices tested (plasma and serum), while D3 was unable to detect AFP in the presence of EDTA plasma (data not shown). The final characteristic, peak shape, was somewhat subjective; however, the instrument permits the user to visualize the binding reaction. For A3, the data demonstrated the prototypic peak—a sharp increase in signal at the leading edge followed by a relatively rapid decrease in binding. Understanding how to read and interpret these images can provide important information when choosing antibody pairs during assay development. However, emphasis on the peak shape should always be evaluated in the context of other performance characteristics including relative accuracy and precision of the assay standards and QCs.

Once the pair was identified, multi-factorial DOE was applied to define key assay factors. The optimization process incorporated a combination of screening design and response surface methodology using not only S:N ratio from the standard curve, but also total error calculated for quality controls. This output parameter is more realistic than using a standard curve alone and provides increased confidence in the estimates because it is more reflective of a true sample. The reliability of these estimates was confirmed in all three phases following optimization (prediction confirmation, pre-study validation, and in-study validation). One of the major advantages of using DOE is to identify conditions that improve the robustness of the assay. Pass rate is a reliable indicator of the relative robustness of the method and this assay has a 100% pass rate applying the industry standard 4-6-X approach to acceptance/rejection. The

other validation parameters (bench top stability, freeze thaw stability, and linearity) were acceptable for the intended purpose and stage of drug development. A unique aspect of the Gyrolab is the ability to process five CDs worth of samples in one run. However, since the instrument processes only one CD at a time, samples could potentially sit on the Gyrolab between ~20 min to 5 h before being delivered to the CD microstructures. The current hardware does not have a temperature controlled environment, so bench top stability must incorporate an evaluation of time on the Gyrolab instrument to adequately simulate the testing environment. AFP demonstrated no obvious trends or changes in immune-reactivity following a 5 CD run.

5. Conclusion

Taken together, the results show that the integration of DOE and Gyros technology facilitated rapid and successful method development (6 days from antibody screen to DOE prediction confirmation). The total run time of roughly 1 h enables faster development, as multiple experiments can be executed in a single day. The Gyrolab software wizard interface provides an easy-to-use mechanism for entering the different combinations of conditions defined during DOE. The on-board liquid handling reduces the likelihood of technical errors and increases pipetting speed. The combination of software and hardware has reduced many of the barriers of implementing DOE into immunoassay development, while the reduction in volume requirements allows for more information being obtained from limited quantity samples.

Acknowledgment

The authors would like to acknowledge Alison Joyce (Pfizer WRD) and extend their gratitude for her contributions to this manuscript.

References

[1] H. Yoshima, T. Mizuochi, M. Ishii, A. Kobata, Cancer Res. 40 (1980) 4276–4281.
[2] M. Soresi, C. Magliarisi, P. Campagna, G. Leto, G. Bonfissuto, A. Riili, A. Carroccio, R. Sesti, S. Tripi, G. Montalto, Anticancer Res. 23 (2003) 1747–1753.
[3] G. Giannelli, E. Fransvea, P. Trerotoli, M. Beaugrand, F. Marinosci, L. Lupo, G. Nkontchou, P. Dentico, S. Antonaci, Clin. Chim. Acta 383 (2007) 147–152.
[4] S.L. Chan, F.K. Mo, P.J. Johnson, E.P. Hui, B.B. Ma, W.M. Ho, K.C. Lam, A.T. Chan, T.S. Mok, W. Yeo, J. Clin. Oncol. 27 (2009) 446–452.
[5] B.K. Kim, S.H. Ahn, J.S. Seong, J.Y. Park, Y. Kim do, J.K. Kim, Y. Lee do, K.H. Lee, K.H. Han, Liver Int. 31 (2011) 369–376.
[6] Y.Y. Shao, Z.Z. Lin, C. Hsu, Y.C. Shen, C.H. Hsu, A.L. Cheng, Cancer 116 (2010) 4590–4596.
[7] C.M. Bagi, J. Christensen, D.P. Cohen, W.G. Roberts, D. Wilkie, T. Swanson, T. Tuthill, C.J. Andresen, Cancer Biol. Ther. 8 (2009) 856–865.
[8] P. Andersson, G. Jesson, G. Kylberg, G. Ekstrand, G. Thorsen, Anal. Chem. 79 (2007) 4022–4030.
[9] S.X. Peng, B.A. Rockafellow, T.M. Skedzielewski, N.D. Huebert, W. Hageman, J. Pharm. Sci. 98 (2009) 1877–1884.
[10] G.S. Sittampalam, W.C. Smith, T.W. Miyakawa, D.R. Smith, C. McMorris, J. Immunol. Methods 190 (1996) 151–161.
[11] C.A. Ray, V. Patel, J. Shih, C. Macaraeg, Y. Wu, T. Thway, M. Ma, J.W. Lee, B. Desilva, J. Pharm. Biomed. Anal. 49 (2009) 311–318.
[12] D. Rodbard, Y. Feldman, M.L. Jaffe, L.E. Miles, Immunochemistry 15 (1978) 77–82.
[13] G. Derringer, R. Suich, J. Qual. Technol. 12 (1980) 214–219.
[14] G.E.P. Box, D.W. Behnken, Technometrics 2 (1960) 455–475.