Journal of Pharmacological and Toxicological Methods 91 (2018) 43–49

Contents lists available at ScienceDirect

Journal of Pharmacological and Toxicological Methods

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

Research article

Analytical performance of a commercial multiplex Luminex-based cytokine panel in the rat

Tim Erkens*, Nick Goeminne, Ann Kegels, Martine Byloos, Petra Vinken

Mechanistic and Investigative Toxicology, Preclinical Development & Safety, Janssen Research & Development, A Division of Janssen Pharmaceutica NV, Turnhoutseweg 30, Beerse, Belgium

ARTICLE INFO

Keywords:
Biomarker
Cytokine
GLP
Multiplex
Rat
Validation

ABSTRACT

Introduction: Multiplex immunoassays are an important tool in biomarker research during preclinical drug development. However, information regarding analytical performance of commercial multiplex assays for animal species is often limited. To be able to correctly interpret study results, a fit-for-purpose validation approach is recommended. The objective of our study was to provide a realistic example of what level of validation can be expected from this type of assay, using a rat cytokine panel.

Methods: The analytical performance of a commercial Luminex-based multiplex assay comprising IFN-γ, IL-6, IL-10, IL-12p70, IP-10 and TNF-α was evaluated in Sprague-Dawley rat plasma and serum. Calibration curve, working range, precision, accuracy, selectivity, parallelism, dilutional linearity, prozone effect and sample stability were assessed.

Results: Analytical performance in plasma and serum was comparable. Precision and accuracy results for all analytes were acceptable with coefficient of variation (CV) and relative error (RE) often below 15%, except for serum IL-6. Selectivity results varied per analyte with several cytokines showing CV > 30% and no single minimum required dilution (MRD) could be identified. In addition, some striking differences between recombinant and endogenous protein results were observed. A pronounced prozone effect was detected for IP-10. Analytes in samples stored at − 70 °C were stable (RE < 30%) from 1 up to 6 months depending on the analyte.

Discussion: The results illustrate the challenges encountered during validation of commercial animal Luminex-based multiplex assays, revealing analytical limitations such as matrix and prozone effects. The Milliplex rat cytokine panel under investigation was deemed suitable for relative quantification of exploratory type biomarkers.

1. Introduction

Immunoassays are an important tool in biomarker research and are of great value for generating pharmacodynamic and safety data during drug development. Besides the classical ELISA assays, multiplex immunoassays have become widely used in biomarker research. Multiplexing multiple analytes into one panel has several advantages over singleplex assays, such as reduction in time, cost and sample volume. The latter is of particular interest in preclinical (safety) research where available sample volumes are often limited, especially in the commonly used rodent models. In addition, multiplexing enables identification of consistent patterns in a single sample, instead of relying on individual biomarker data.

Although there are a number of different platforms available for multiplexing (Tighe, Negm, Todd, & Fairclough, 2013), the current paper focuses on animal Luminex-based assays using color-coded beads. Over the last couple of years, the number of commercial Luminex-based assays fit for animal use is rapidly expanding. However, the information manufacturers have available on the analytical performance of the assays is relatively limited in our experience, despite the fact that this information is critical for correct study result interpretation. This observation is supported by several other publications (Belabani, Rajasekharan, Poupon, Johnson, & Bar-Or, 2013; Khan et al., 2015).

Recommendations for immunoassay validation based on a fit-for-purpose approach have been described extensively (Andreasson et al., 2015; Findlay & Dillard, 2007; Khan et al., 2015; Lee et al., 2006; Nowatzke, Cole, & Bowsher, 2010; Valentin, Ma, Zhao, Legay, & Avrameas, 2011). Currently, there are no fixed acceptance criteria for these kind of biomarker assays and one should be cautious not to simply apply the criteria for bioanalytical pharmacokinetic assays (Tighe, Ryder, Todd, & Fairclough, 2015; Timmerman, 2016). In addition, combining multiple analytes into one assay is likely to have an impact

* Corresponding author.
E-mail address: terkens1@its.jnj.com (T. Erkens).

https://doi.org/10.1016/j.vascn.2018.01.005
Received 15 May 2017; Received in revised form 20 November 2017; Accepted 17 January 2018

on the analytical performance and adds a layer of complexity to immunoassay validation compared to singleplex assays (Chowdhury, Williams, & Johnson, 2009; Ellington, Kullo, Bailey, & Klee, 2009; Jani et al., 2016). Recently, an excellent white paper discussed the many challenges, such as minimum required dilution (MRD), cross-reactivity and selectivity, encountered when using commercial multiplex ligand binding assays (Jani et al., 2016).

The fit-for-purpose validation approach distinguishes between exploratory and decision-making biomarkers, the latter requiring the most stringent validation process (Lee et al., 2006; Valentin et al., 2011). However, since no formal criteria exist, in practice a wide range of different levels of validation are being used for (safety) biomarkers, including in good laboratory practice (GLP) settings. The objective of the current study was to provide an in-depth evaluation of the analytical performance of a commercial multiplex immunoassay for the measurement of IFN-γ, IL-6, IL-10, IL-12p70, IP-10 (CXCL-10) and TNF-α on a Luminex platform in both rat plasma and serum. The rat is an important species in preclinical drug development and this multiplex cytokine assay is a valuable tool, both for monitoring safety and for mechanistic investigations of immune modulatory compounds. Cytokine biomarkers also have great translational potential towards clinical studies. Hence it is critical to identify potential strengths and weaknesses in order to assess the level of validation that can be achieved with this type of commercial multiplex assay. The following aspects were evaluated: calibration curve, working range, intra- and inter-batch precision/accuracy, selectivity, prozone (high dose hook effect), parallelism, dilutional linearity, and sample stability.

2. Materials and methods

The Milliplex MAP Rat Cytokine/Chemokine Magnetic Bead 6-plex Panel with IFN-γ, IL-6, IL-10, IL-12p70, IP-10, TNF-α (RECYTMAG-65K; Merck-Millipore) was used according to the manufacturer's instructions, using an automated magnetic plate washer (Bio-Tek 405 LS Microplate Washer). Briefly, the assay uses 25 μl of sample to capture an analyte on analyte-specific color-coded magnetic beads coated with capture antibodies. In a next step biotinylated detection antibodies are added, followed by an incubation with streptavidin-phycoerythrin. All measurements were performed on a Magpix Luminex instrument, using xPonent 4.2 (Luminex) and Bio-Plex Manager 6.1 (Bio-Rad) software. Experimental work was conducted by 2 different analysts, each performing separate runs. Throughout the study EDTA-plasma and serum collected from the carotid artery from male Sprague-Dawley rats (Charles River Laboratories; approximately 2–4 months of age) was used, which was obtained from the on-site AAALAC (Association for Assessment and Accreditation of Laboratory Animal Care) approved rodent facility according to the applicable animal welfare guidelines and legislation. Samples were stored at approximately − 70 °C within 0.5–1 h after collection and kept at this temperature when not in use. Sample analysis started within 0.5–1 h after taking samples out of the freezer, except for stability analysis for which details are described below. Unless specified otherwise, matrix samples were diluted 1:2 in assay buffer as specified in the kit insert. In addition to the standard provided in the kit, the following external recombinant rat proteins were used for spiking of sample matrix: IFN-γ (585-IF-100, R&D Systems), IL-6 (506-RL-010, R&D Systems), IL-10 (522-RL-005, R&D Systems), IL-12 (1760-RL-010, R&D Systems), IP-10 (E-65340, PromoKine), TNF-α (510-RT-010, R&D Systems). Acceptance criteria detailed below were considered acceptable for our intended use of the assay, taking into account anticipated biologic variability and available information in literature (Defawe et al., 2012; DeSilva et al., 2003; Jani et al., 2016; Valentin et al., 2011).

2.1. Calibration curve

Each run contained a calibration curve (in duplicate) of lyophilized recombinant protein diluted in assay buffer, prepared as defined in the kit insert. Over 20 calibration curves were evaluated, using an acceptance criterion of 20% (25% at highest and lowest standard point) for precision (CV%) of duplicate concentrations and for relative error (RE %) of the mean back-calculated concentration of each standard point versus its nominal concentration.

2.2. Validation samples

In total 8 different validation samples (VS) were used to cover the working range of the assay: 2 kit quality control samples containing recombinant proteins dissolved in water, and 3 EDTA-plasma and 3 serum samples with endogenous levels of the analytes of interest. For each of these VS, a nominal concentration was established by calculating the mean of triplicate measurements from 3 independent assay runs.

2.3. Intra- and inter-batch precision/accuracy

The intra-batch concentration precision (CV%) and accuracy (RE%) of the method was determined in an assay batch in which each VS was analyzed 9-fold. For inter-batch precision and accuracy each VS was analyzed in triplicate in 5 additional assay batches, and CV% and RE% were calculated using data from all 6 runs. A maximum of 30% for precision and accuracy was considered acceptable.

2.4. Selectivity

In order to detect any differential matrix effects (endogenous matrix components that could interfere with assay performance), selectivity was evaluated using 8–10 independent rat plasma and serum samples. Each sample was spiked using either recombinant kit standard or using a sample containing endogenous levels of the analytes of interest. The volume of the spiked material did not exceed 5% of the total sample volume. A CV of maximum 30% between concentrations was considered acceptable.

2.5. Prozone (high dose hook effect)

In the presence of prozone or high dose hook effect, falsely lower concentrations of analyte are measured in samples that actually contain high levels of analyte. To investigate a potential prozone effect, plasma and serum was either spiked using recombinant protein to obtain high concentrations above the upper limit of quantification (ULOQ) or samples with high endogenous levels of the analytes of interest were used. Samples were serially diluted with kit assay buffer.

2.6. Dilutional linearity

Dilutional linearity evaluates potential matrix effects and demonstrates if analyte concentrations above ULOQ can be diluted into the validated range of the assay. At least 3 independent plasma and serum samples were spiked with recombinant kit standard and serially diluted with kit assay buffer. An accuracy (RE) of maximum 30% compared to the primary (least diluted) sample was considered acceptable.

2.7. Parallelism

Parallelism evaluates if the standard concentration–response curve is parallel to the sample dilution–response curve using samples with the endogenous analyte in sample matrix. Parallelism was assessed in at least 3 samples by serial dilution of those analytes with appropriate endogenous concentrations, using assay buffer as diluent. An accuracy (RE) of maximum 30% compared to the primary (least diluted) sample was considered acceptable.

2.8. Sample stability

Sample stability was assessed in 3 plasma and 3 serum samples containing different levels of endogenous analyte. Each sample was analyzed in triplicate on each occasion. Short-term stability was evaluated at room temperature (4 h), and after 2–3 freeze-thaw cycles (4 h at room temperature followed by storage at −70 °C for at least 12 h). Long-term stability (−70 °C) was evaluated in samples stored for approximately 1, 3 and 6 months. A 30% accuracy (RE) compared to the nominal concentrations was considered acceptable.

3. Results

3.1. Calibration curve, working range, precision and accuracy

A 5-parameter logistic (PL) algorithm was used for all analytes to fit the 7 non-zero standard points set by the kit manufacturer (Supplementary Table S1). For IL-6, IL-12p70 and IFN-γ the lowest standard point was excluded in the majority of runs, because it did not meet the acceptance criteria or did not generate a detectable signal. The lower (LLOQ) and upper (ULOQ) limits of quantification were defined as the lowest and highest standard concentration, respectively, with acceptable accuracy and precision, which were supported by acceptable results for the validation samples. The LLOQ and ULOQ (corrected for a default 2-fold dilution factor) were set as indicated in Table 1. Serum IL-6 failed inter-assay accuracy for the low level validation sample (RE = −34.0%), and therefore no LLOQ could be established for serum. Intra- and inter assay precision and accuracy results in plasma and serum were acceptable for IL-10, IL-12p70, IFN-γ, IP-10 and TNF-α, with CV and RE often below 15% (Tables 2a and 2b). This was also the case for IL-6 in plasma. Inter-analyst CV for identical samples was similar to the inter-assay CV reported in Tables 2a and 2b. Bead counts for each analyte were above the minimum recommended 50 beads, and mostly ranged between 400 and 800 beads per analyte.

3.2. Selectivity

Overall, no acceptable selectivity could be demonstrated in plasma or serum for all analytes of the 6-plex panel. In 2-fold diluted serum spiked with recombinant kit standard, most results failed acceptance criteria (Table 3), while when spiked with a sample containing endogenous levels of protein, IFN-γ, IP-10 and TNF-α results were acceptable (CV < 30%). In the latter conditions, IL-10 failed selectivity, and endogenous IL-6 and IL-12p70 levels were too low for spiking. For 2-fold diluted plasma similar results were obtained, with IP-10 and TNF-α having a CV < 30% when endogenous protein was used for spiking. CV for plasma IL-6, IL-10, IL-12p70 and IFN-γ was > 30% when spiked with recombinant kit standard, and no samples with suitable endogenous levels were available for spiking. In 2-fold diluted samples, there was often a distinct under recovery compared to the theoretically spiked standard concentration for most analytes, and increasing the sample dilution to 4- or 8-fold did not improve selectivity

Table 1
LLOQ and ULOQ in plasma and serum.

Analyte	LLOQ[a] (pg/ml)	ULOQ[a] (pg/ml)	Kit manufacturer calibration curve range[a] (pg/ml)
IL-6	586[b]	600,000	146–600,000
IL-10	14.6	60,000	14.6–60,000
IL-12p70	97.7	100,000	24.4–100,000
IFN-γ	117	120,000	29.3–120,000
IP-10	19.5	20,000	4.9–20,000
TNF-α	4.9	20,000	4.9–20,000

[a] Corrected for a 2-fold sample dilution.
[b] For plasma only.

Table 2a
Summary of the intra- and inter-assay accuracy and precision results.

Analyte	VS[a]	Nominal concentration (pg/ml)	CV%[b]		RE%[b]	
			Intra	Inter	Intra	Inter
IL-6	QC1	6275	6.1	20.6	9.5	−8.7
	QC2	33,717	5.2	5.1	−3.6	−9.3
	P1	< 586	–	–	–	–
	P2	1052	25.4	17.2	−25.6	−10.6
	P3	62,577	2.0	5.4	−5.6	−8.4
	S1	1028	16.7	25.7	−0.3	−34.0
	S2	984	–	–	–	–
	S3	63,941	2.5	13.2	−16.7	−18.0
IL-10	QC1	667	6.4	20.2	8.8	−11.9
	QC2	3343	5.1	6.2	0.6	−8.6
	P1	220	7.6	8.7	11.6	1.1
	P2	312	8.1	8.9	5.2	0.0
	P3	1655	6.9	9.8	6.5	−2.5
	S1	240	9.4	9.3	19.0	4.3
	S2	679	1.9	8.9	2.9	−3.8
	S3	2964	4.7	10.9	−4.5	−10.5
IL-12p70	QC1	841	7.3	16.9	16.5	−0.8
	QC2	4176	5.5	6.7	7.9	−1.4
	P1	< 97,7	–	–	–	–
	P2	301	14.7	13.6	33.7[c]	20.2
	P3	173	10.3	17.9	12.4	17.1
	S1	595	7.4	16.4	39.7[c]	11.9
	S2	294	9.2	15.6	2.4	−14.0
	S3	306	4.1	14.2	22.3	11.7

[a] VS, validation sample; QC, kit quality control with recombinant protein in water; P, plasma sample with endogenous protein levels; S, serum sample with endogenous protein levels.
[b] Results indicated with "–" refer to concentrations below LLOQ or serum IL-6 for which LLOQ could not be defined.
[c] Intra-assay accuracy (RE) was > 30% due to one aberrant result in a run used to establish nominal validation sample concentrations. Inter-assay precision and accuracy did meet criteria for samples P2 and S1, and also the results of all other validation samples were within specifications. Therefore, IL-12p70 precision and accuracy results were considered acceptable.

Table 2b
Summary of the intra- and inter-assay accuracy and precision results.

Analyte	VS[a]	Nominal concentration (pg/ml)	CV%[b]		RE%[b]	
			Intra	Inter	Intra	Inter
IFN-γ	QC1	953	8.7	21.0	9.0	−11.7
	QC2	5075	6.0	6.5	4.5	−6.0
	P1	< 117	–	–	–	–
	P2	450	15.2	16.9	−21.3	12.2
	P3	11,301	7.5	11.6	3.0	0.6
	S1	361	14.0	26.7	−7.9	2.1
	S2	2030	2.7	11.0	12.4	12.1
	S3	15,141	2.9	7.9	6.9	4.7
IP-10	QC1	255	5.2	25.5	12.5	−12.8
	QC2	1299	5.1	7.1	3.2	−8.9
	P1	149	4.0	9.5	20.9	14.2
	P2	4190	2.1	8.4	2.3	−2.5
	P3	4500	2.0	10.9	2.4	−5.3
	S1	248	5.0	7.6	19.9	8.3
	S2	2348	1.4	12.3	6.3	−5.7
	S3	3783	2.5	13.2	1.2	−7.4
TNF-α	QC1	492	7.9	20.1	5.4	−6.0
	QC2	1903	7.0	6.0	−6.5	−3.1
	P1	154	14.0	10.7	10.4	16.8
	P2	32,2	13.6	13.8	−4.6	7.6
	P3	2316	6.8	12.6	−1.0	9.6
	S1	23,7	11.9	23.6	−2.5	−18.5
	S2	281	5.3	13.0	3.8	10.2
	S3	2630	5.6	11.7	−5.4	3.8

[a] VS, validation sample; QC, kit quality control with recombinant protein in water; P, plasma sample with endogenous protein levels; S, serum sample with endogenous protein levels.
[b] Results indicated with - refer to concentrations below LLOQ.

Table 3
Selectivity results in rat serum.

Analyte	Spiking source	Spiked concentration	Sample dilution	Mean measured concentration	SD	CV
		(pg/ml)		(pg/ml)	(pg/ml)	(%)
IL-6[a]	Recombinant	3750	2	< 586	–	–
	Recombinant	15,000	2	2362	726	31
IL-10	Recombinant	375	2	165	77	47
	Recombinant	1500	2	619	285	46
	Endogenous	–	2	88	103	117
	Endogenous	–	4	141	104	74
IL-12p70[a]	Recombinant	625	2	545	246	45
	Recombinant	2500	2	2028	637	31
IFN-γ	Recombinant	750	2	< 117[b]	–	–
	Recombinant	3000	2	744	376	50
	Endogenous	–	2	1028	235	23
	Endogenous	–	4	1042	204	20
IP-10	Recombinant	125	2	114	67	59
	Recombinant	500	2	287	80	28
	Endogenous	–	2	1284	150	12
	Endogenous	–	4	2157	317	15
TNF-α	Recombinant	125	2	87	29	34
	Recombinant	500	2	269	82	30
	Endogenous	–	2	137	36	26
	Endogenous	–	4	180	40	22

[a] No sample containing sufficient endogenous protein was available for spiking.
[b] In 6/10 samples concentration was < LLOQ, in the other 4 samples IFN-γ was only borderline detectable.

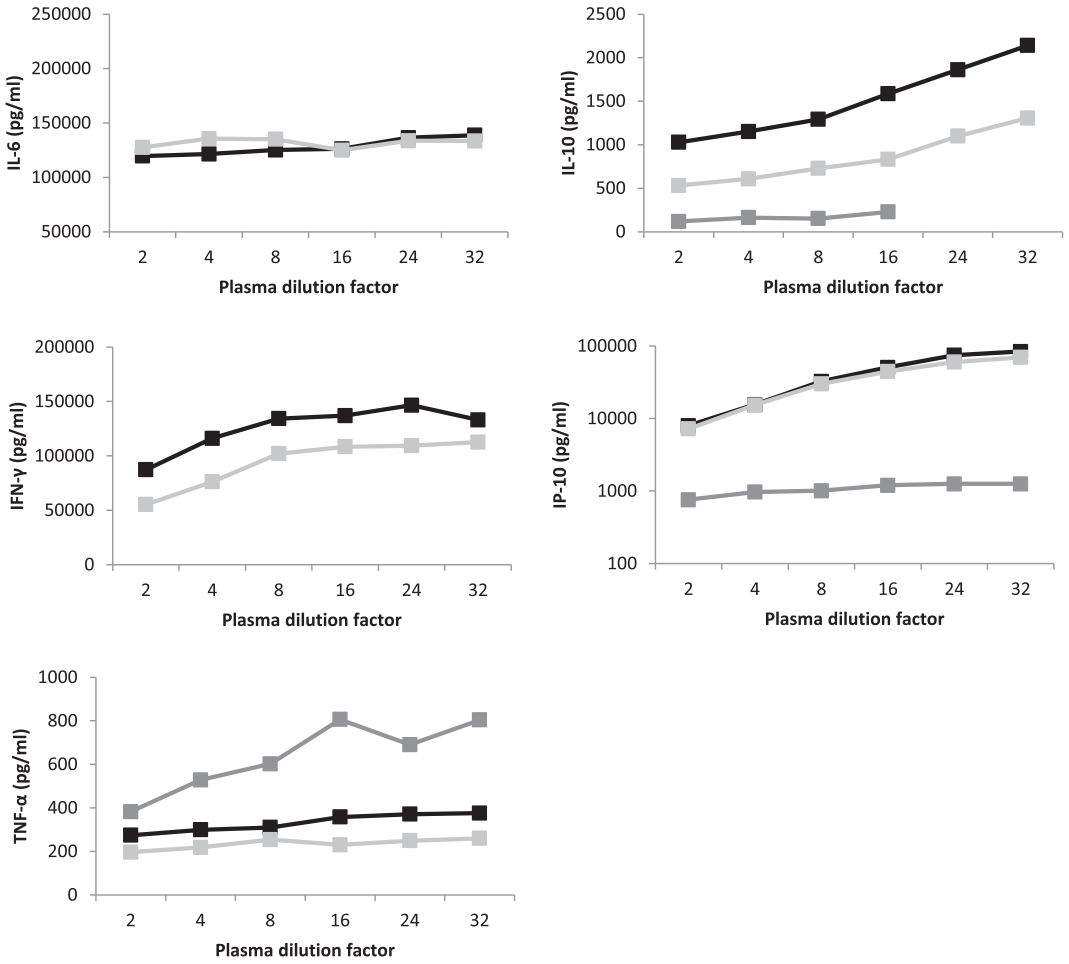

Fig. 1. Parallelism results in 3 different rat plasma samples with endogenous levels of the analytes of interest, using kit assay buffer as diluent. The different colors represent different samples. For IL-6 and IFN-γ only two samples with sufficiently high levels were available. IP-10 results (dark grey line) were confirmed in additional samples (data not shown), since the other two samples in this graph (black and light grey line) did not allow parallelism evaluation because of prozone interference. IL-12p70 levels were too low to evaluate.

results (Supplementary Table S2).

3.3. Parallelism and dilutional linearity

No single MRD suitable for all 6 analytes could be identified and sometimes data were not consistent between different samples (Fig. 1, Supplementary Fig. S1 and Supplementary Tables S3–S4). For IL-6 in plasma, parallelism was seen from a 2-fold dilution onwards. In both matrices, IFN-γ and IP-10 parallelism was observed from an 8-fold dilution onwards, while no consistent parallelism was observed for IL-10 or TNF-α from a 2-, 4- or 8-fold dilution onwards. Using kit Serum Matrix as a diluent for plasma gave similar results compared to kit assay buffer (data not shown).

Dilutional linearity also varied per analyte, often with similar results as seen for parallelism (Supplementary Fig. S2 and Supplementary Table S5). However, for IL-10 and IP-10 in serum, a linear range was observed starting at a 2-fold dilution, indicating a different behavior of the recombinant compared to the endogenous protein. No consistent IL-12p70 dilutional linearity was observed in plasma or serum, and no samples with sufficiently high endogenous levels were available for parallelism evaluation.

3.4. Prozone evaluation

Results for plasma and serum were similar. A pronounced prozone effect was detected for IP-10, both in sample matrix containing a high endogenous level and in sample matrix spiked with recombinant protein (Fig. 2 and Supplementary Table S6). In the endogenous sample, consistent results for IP-10 concentration were obtained from a 160-fold dilution onwards.

In endogenous samples no prozone effect for IL-6 and TNF-α was detected. When sample matrix was spiked with recombinant protein at even higher levels than shown in Fig. 2 and Supplementary Table S6, a prozone effect was present (data not shown), but it is uncertain if this finding is biologically relevant. No prozone effect was observed for IL-10 and IL-12p70, and concentrations for IFN-γ were too low to allow evaluation.

3.5. Sample stability

When all 6 analytes are multiplexed, the sample stability is determined by the least stable analyte. All analytes showed long-term stability for at least 1 month, in plasma and in serum (Fig. 3). Freeze-thaw stability varied per analyte (Supplementary Fig. S3). Four hour bench-top stability was not proven for IFN-γ in both matrices and for TNF-α in serum. Therefore, it is recommended to standardize and minimize the time samples are kept on the bench prior to storage or analysis.

4. Discussion

The analytical performance of the multiplex rat cytokine panel under investigation (IFN-γ, IL-6, IL-10, IL-12p70, IP-10 and TNF-α) applied on a Luminex platform was evaluated in rat serum and EDTA plasma, using a fit-for-purpose approach for assay validation (Khan et al., 2015; Lee et al., 2006; Valentin et al., 2011). Serum and EDTA plasma were selected because these are two typical matrices which are sampled in preclinical toxicity studies. Suitability of these matrices for cytokine analysis would allow optimal use of collected blood volume which is limited in rats.

For none of the 6 cytokines in this panel, a universal reference standard is available, as is the case for many commercial animal immunoassays. As a result, the assay only allows a relative quantification of the analytes. One has to keep in mind that using kits from other manufacturers will likely result in different analyte concentrations, as illustrated by Nechansky, Grunt, Roitt, & Kircheis (2008). Ideally, the working range of an assay including LLOQ and ULOQ is established based on matrix samples with respectively the lowest and highest analyte levels with acceptable results (Lee et al., 2006). However, in a multiplex setting it is difficult to obtain validation samples with desired concentrations for all analytes, especially in larger panels. This is not only a challenge when using samples with endogenous analyte levels, but also when spiking with recombinant kit standard. Selectivity results in Table 3 show that spiking sample matrix with kit standard resulted in an under recovery for some analytes, a finding which has also been reported for other immunoassays (Defawe et al., 2012; Staples, Ingram, Atherton, & Robinson, 2013). To overcome differences in concentration between analytes, additional validation samples for specific subsets of analytes might be used, but this could jeopardize practical feasibility since multiple levels of VS are required, and it would not resolve a lack of suitable samples to determine the ULOQ. On the other hand, the low level serum IL-6 validation sample failed inter-assay accuracy although the nearest calibration point was acceptable, which indicates that establishing LLOQ and ULOQ based on standard curve performance alone is not ideal either. Moreover, the lowest calibration point as defined in the kit insert did not meet our acceptance criteria for several analytes (Table 1). Therefore, we used a combination of the lowest and highest acceptable standard point supported by the VS to establish the limits of quantification. Based on our validation data, the working range for several cytokines differed substantially from what was indicated by the manufacturer. In general, the way the working range is established can differ between labs and consequently may influence potential application of the assay.

Overall, no acceptable selectivity could be demonstrated for all analytes of the 6-plex panel. Most selectivity results failed the acceptance criterion, with CV > 30% when recombinant kit standard was used for spiking 2-fold diluted matrix samples (which is the dilution recommended by the kit insert). Increasing the dilution did not improve selectivity results, but when instead a sample with sufficient endogenous protein levels was used for spiking, an improvement of the

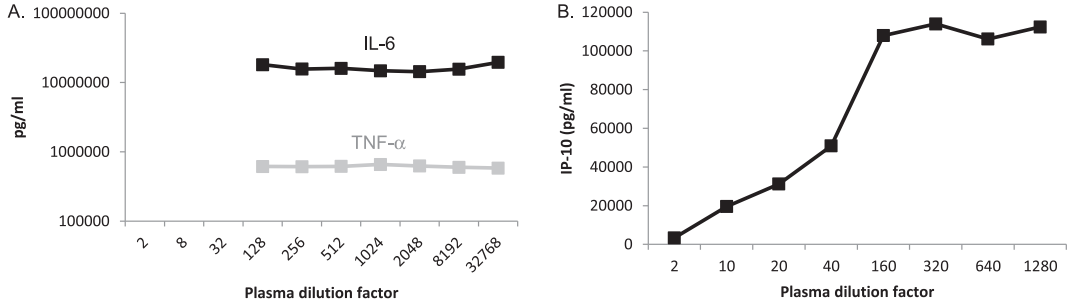

Fig. 2. Prozone evaluation results in rat plasma, using kit assay buffer as diluent. (A) Results for rat plasma spiked with recombinant protein. At the lower dilutions, IL-6 and TNF-α concentrations were above the ULOQ of the assay and could therefore not be plotted. (B) Results for rat plasma containing a high endogenous IP-10 level.

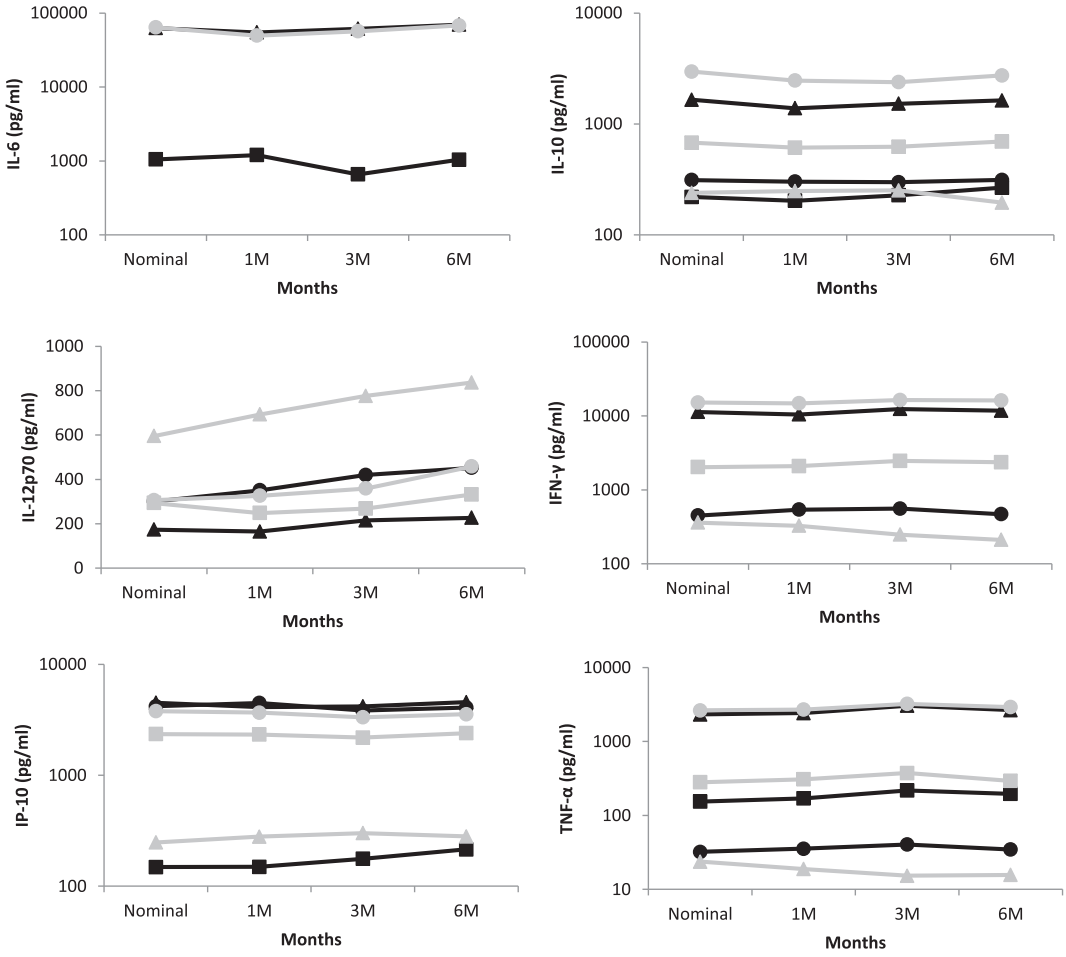

Fig. 3. Rat cytokine long-term stability in plasma and serum containing endogenous levels of the analytes of interest. Black lines represent plasma and grey lines serum, with different symbols used for different samples.

selectivity was seen for some cytokines. Although selectivity is often only evaluated by spiking a recombinant protein source, data show the importance of spiking with endogenous proteins, since for some analytes the outcome differed even though concentrations were similar (Table 3). Similar observations have previously been made for sample stability evaluation (Fraser, Fleener, Ogborne, & Soderstrom, 2015). Also, dilutional linearity and parallelism results differed between recombinant spiked and endogenous samples for some cytokines, as described in the results section. This potential discrepancy has a significant impact on the level of assay validation that can be achieved, and depending on the level required, the value of recombinant protein data to evaluate some analytical aspects can be questioned. As already mentioned, the challenge is to acquire samples with suitable endogenous levels for the analytes of interest, particularly when multiplexed.

Since parallelism data were available for all analytes except IL-12p70, the data obtained from samples with endogenous proteins were considered more relevant than dilutional linearity data generated using spiked recombinant proteins. However, no single minimum required dilution suitable for all 6 analytes within the panel could be identified in serum or plasma. As the MRD increases, assay sensitivity will decrease, which could compromise the usefulness of a particular multiplex assay. For this rat cytokine 6-plex assay, increasing the sample dilution above 2-fold did not improve selectivity results, and precision and accuracy data supported the 2-fold dilution. Insufficient parallelism implies that all study samples need to be analyzed using the same sample dilution. Depending on the purpose of the assay (exploratory versus critical decision making), these limitations might be acceptable since it has been demonstrated that the assay is capable of detecting cytokine increases. Especially if multiple analytes within a multiplex panel show consistent results, confidence in the assay will increase.

Since there are no formal standardized acceptance criteria for validation of multiplex assays (Tighe et al., 2015; Timmerman, 2016), the type of analytical aspects and how they are evaluated can differ significantly between laboratories in a preclinical GLP setting. As a result, the GLP validation of a commercial multiplex assay does not necessarily guarantee a level of analytical validation with high scientific rigor. It is however critical to be aware of the limitations of an assay, because also in a GLP setting the fit-for-purpose approach applies (Lee, 2009; Valentin et al., 2011). The question is what to do if some analytical aspects cannot be evaluated, because for instance suitable samples are not available or the spiked concentration of recombinant protein that can be achieved is not sufficiently high, as illustrated by the results from the current study. In those circumstances, it is important to assess the impact of any missing data and decide if the assay can still serve its intended purpose.

One example of an analytical aspect that is not always evaluated during validation, is the check for a prozone or hook effect. Prozone results in a falsely lower signal caused by high analyte concentrations (Selby, 1999). The results of the current study (Fig. 2 and Supplementary Table S6) demonstrate that IP-10 in this rat Milliplex assay is susceptible to prozone, not only in plasma or serum samples artificially spiked with high recombinant protein levels, but more importantly also in samples containing high endogenous levels. This effect can have a substantial impact on study result interpretation, as illustrated by the example in Fig. 4, in which the difference between two treatment

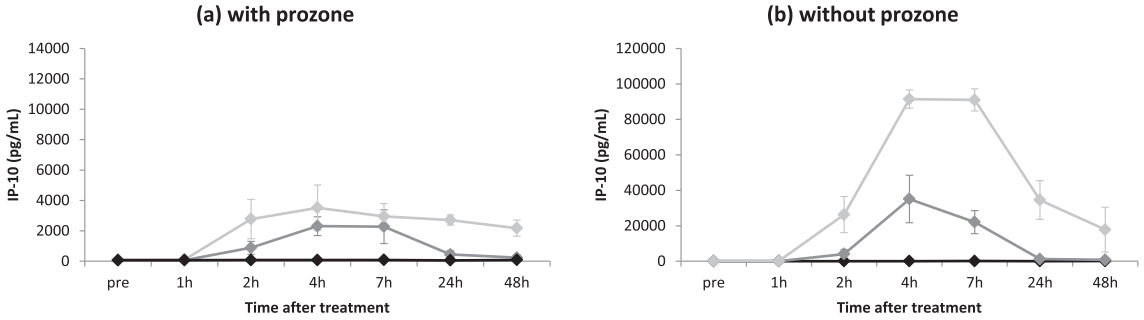

Fig. 4. Illustration of how IP-10 prozone interference can affect study result interpretation, with data from rats treated with a single oral dose of an immunostimulatory compound. The high dose group received a 10-fold higher dose compared to the low dose group. Plasma samples were collected from the tail vein. Lines represent mean data per group ($n = 3$) with standard deviations as error bars (control group: black; low dose group: dark grey; high dose group: light grey). (a) 2-fold diluted plasma samples with prozone interference; (b) 80-fold diluted plasma samples without prozone interference. Different scales were used on the y-axis for (a) and (b).

groups is obscured by prozone. In addition, IP-10 prozone interference further adds to the complexity of selecting a single MRD suitable for all 6 analytes in this rat multiplex panel. As shown in Figs. 2 and 4, a much higher sample dilution was needed to detect and avoid IP-10 prozone interference, but this strongly reduces the sensitivity for the other analytes and thereby the utility of the assay. Consequently, IP-10 may have to be removed from this multiplex panel, depending on the purpose of the experiment.

In our experience, characterization of the analytical performance of commercial animal multiplex immunoassays by the manufacturers is generally often limited and further enquiry at the vendor/manufacturer provides little additional information or insight. Although for the rat cytokine 6-plex assay used in this study the product insert mentions a few assay characteristics, insert data were mostly high level and insufficient details were provided which would enable verification of the results. In contrast to the serum spike recovery data from the kit insert, we found a distinct under-recovery of kit standard for most analytes in our panel (Table 3). The insert mentions that samples were spiked at various concentrations, but only a single mean recovery percentage is provided without any indication of variation within or between the different spiking levels. A thorough comparison with our own spike recovery data was not possible, because details on the spiking concentrations and spiking volume are lacking from the insert. The analytical performance of other multiplex Luminex kits, from different vendors, for other analytes, in different matrices and in various preclinical species (data not shown) was in our experience comparable to what was observed for the 6-plex assay used in the current study, including its weaknesses. Since these are commercial ready-to-use kits, further optimization of the assays in close collaboration with the kit manufacturers will be key in overcoming the issues identified.

In conclusion, the results from the current study illustrate the challenges and difficulties faced with during validation of commercial animal multiplex assays. The data provide a practical example of what level of validation can be achieved with this type of assay, and emphasize the importance of using validation samples with endogenous analyte levels rather than recombinant spiked material. The use of Luminex-based multiplex assays based on a fit-for-purpose approach can be of great added value in preclinical drug development, but it is important to characterize and be aware of analytical limitations such as matrix and prozone effects, and these have to be kept in mind when analyzing samples and interpreting study data. The Milliplex rat cytokine panel under investigation was deemed suitable for the relative quantification of more exploratory type biomarkers.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.vascn.2018.01.005.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial or not-for-profit sectors.

References

Andreasson, U., Perret-Liaudet, A., van Waalwijk van Doorn, L. J., Blennow, K., Chiasserini, D., Engelborghs, S., ... Teunissen, C. E. (2015). A practical guide to immunoassay method validation. *Frontiers in Neurology, 6*, 179.

Belabani, C., Rajasekharan, S., Poupon, V., Johnson, T., & Bar-Or, A. (2013). A condensed performance-validation strategy for multiplex detection kits used in studies of human clinical samples. *Journal of Immunological Methods, 387*, 1–10.

Chowdhury, F., Williams, A., & Johnson, P. (2009). Validation and comparison of two multiplex technologies, Luminex and Mesoscale Discovery, for human cytokine profiling. *Journal of Immunological Methods, 340*, 55–64.

Defawe, O. D., Fong, Y., Vasilyeva, E., Pickett, M., Carter, D. K., Gabriel, E., ... De Rosa, S. C. (2012). Optimization and qualification of a multiplex bead array to assess cytokine and chemokine production by vaccine-specific cells. *Journal of Immunological Methods, 382*, 117–128.

DeSilva, B., Smith, W., Weiner, R., Kelley, M., Smolec, J., Lee, B., ... Celniker, A. (2003). Recommendations for the bioanalytical method validation of ligand-binding assays to support pharmacokinetic assessments of macromolecules. *Pharmaceutical Research, 20*, 1885–1900.

Ellington, A. A., Kullo, I. J., Bailey, K. R., & Klee, G. G. (2009). Measurement and quality control issues in multiplex protein assays: A case study. *Clinical Chemistry, 55*, 1092–1099.

Findlay, J. W., & Dillard, R. F. (2007). Appropriate calibration curve fitting in ligand binding assays. *The American Association of Pharmaceutical Scientists Journal, 9*, E260–267.

Fraser, S., Fleener, C., Ogborne, K., & Soderstrom, C. (2015). When close is not close enough: A comparison of endogenous and recombinant biomarker stability samples. *Bioanalysis, 7*, 1355–1360.

Jani, D., Allinson, J., Berisha, F., Cowan, K. J., Devanarayan, V., Gleason, C., ... Stephen, L. (2016). Recommendations for use and fit-for-purpose validation of biomarker multiplex ligand binding assays in drug development. *The American Association of Pharmaceutical Scientists Journal, 18*, 1–14.

Khan, M. U., Bowsher, R. R., Cameron, M., Devanarayan, V., Keller, S., King, L., ... Lachno, D. R. (2015). Recommendations for adaptation and validation of commercial kits for biomarker quantification in drug development. *Bioanalysis, 7*, 229–242.

Lee, J. W. (2009). Method validation and application of protein biomarkers: Basic similarities and differences from biotherapeutics. *Bioanalysis, 1*, 1461–1474.

Lee, J. W., Devanarayan, V., Barrett, Y. C., Weiner, R., Allinson, J., Fountain, S., ... Wagner, J. A. (2006). Fit-for-purpose method development and validation for successful biomarker measurement. *Pharmaceutical Research, 23*, 312–328.

Nechansky, A., Grunt, S., Roitt, I. M., & Kircheis, R. (2008). Comparison of the calibration standards of three commercially available multiplex kits for human cytokine measurement to WHO standards reveals striking differences. *Biomarker Insights, 3*, 227–235.

Nowatzke, W., Cole, T. G., & Bowsher, R. R. (2010). Systematic analytical validation of commercial kits for the determination of novel biomarkers for clinical drug development. *Bioanalysis, 2*, 237–247.

Selby, C. (1999). Interference in immunoassay. *Annals of Clinical Biochemistry, 36*, 704–721.

Staples, E., Ingram, R. J., Atherton, J. C., & Robinson, K. (2013). Optimising the quantification of cytokines present at low concentrations in small human mucosal tissue samples using Luminex assays. *Journal of Immunological Methods, 394*, 1–9.

Tighe, P., Negm, O., Todd, I., & Fairclough, L. (2013). Utility, reliability and reproducibility of immunoassay multiplex kits. *Methods, 61*, 23–29.

Tighe, P. J., Ryder, R. R., Todd, I., & Fairclough, L. C. (2015). ELISA in the multiplex era: Potentials and pitfalls. *Proteomics Clinical Applications, 9*, 406–422.

Timmerman, P. (2016). Biomarker assay validation: Are we having the right discussions? *Bioanalysis, 8*, 2299–2303.

Valentin, M. A., Ma, S., Zhao, A., Legay, F., & Avrameas, A. (2011). Validation of immunoassay for protein biomarkers: Bioanalytical study plan implementation to support pre-clinical and clinical studies. *Journal of Pharmaceutical and Biomedical Analysis, 55*, 869–877.