Journal of Pharmaceutical and Biomedical Analysis 49 (2009) 1203–1212

Contents lists available at ScienceDirect

Journal of Pharmaceutical and Biomedical Analysis

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

Tartrate-resistant acid phosphatase (TRACP 5b): A biomarker of bone resorption rate in support of drug development: Modification, validation and application of the BoneTRAP® kit assay

Yuling Wu*, Jean W. Lee, Lennie Uy, Bassam Abosaleem, Han Gunn, Mark Ma, Binodh DeSilva

Pharmacokinetics and Drug Metabolism, Amgen Inc., Thousand Oaks, CA 91320, United States

ARTICLE INFO

Article history:
Received 10 December 2008
Received in revised form 23 February 2009
Accepted 3 March 2009
Available online 20 March 2009

Keywords:
TRACP 5b
Bone resorption biomarker
Immunoassay
Method validation
Drug development

ABSTRACT

A commercial kit assay of tartrate-resistant acid phosphatase (TRACP 5b) used for the diagnosis of bone resorption was modified with a 'Fit-For-Purpose' approach for drug development of anti-resorptive therapeutics. The modifications included changing the standard matrix from buffer to serum, using a consistent bulk reference material to prepare standards and quality controls (QC), and adding sample controls (SC) prepared from authentic sample pools.

Method validation experiments were conducted for: inter- and intra-assay accuracy and precision, establishment of SC, range finding of different population groups, selectivity tests, parallelism and stability. The analytical range was 1.00–10.0 U/L and the total errors of lower limit of quantification (LLOQ) and upper limit of quantification (ULOQ) validation samples were 8% and 21%, respectively. Data of range finding experiment showed that serum samples should be collected in tubes instead of bags. Selectivity results showed accurate spike recovery among the majority of test samples from target populations. Samples were demonstrated to be stable for up to four freeze/thaw cycles and for 24 months at $-70 \pm 10\,^{\circ}C$. Our results show that the modified TRACP 5b method is reliable for the quantification of TRACP 5b in human serum samples to support clinical trials of bone resorptive effect reflected by TRACP 5b activities. The method was robust with similar assay performance characteristics shown in three bioanalytical laboratories.

1. Introduction

Biomarkers of bone turnover have been used to facilitate the drug discovery and development process [1,2]. Improvements in the clinical outcome of bone fracture and physiological marker of bone mineral density (BMD) can take years to be demonstrated for drug effect. It is desirable if early predictive biomarkers can be used to aid drug development decisions and for patient treatment prognosis. For anti-resorptive treatments, assays that can reliably measure bone resorption can provide an early and informative index in clinical trials [3,4]. TRACP 5b is one of the bone resorption biomarkers that can be qualified for such use.

TRACP consists of two isoforms, 5a and 5b. Macrophages and dendritic cells appear to secrete only TRACP 5a [5]. TRACP 5b is highly expressed by bone-resorbing osteoclasts and secreted into the blood circulation as an active enzyme [6]. Higher TRACP 5b activity in blood indicates the increased bone resorption rate in patients with primary osteoporosis [7,8] and bone disease in cancer patients [9,10]. Moreover, TRACP 5b shows low diurnal variability, which allows flexible sample collection time [11]. TRACP 5b has proved to be a useful biomarker for monitoring the efficacy of anti-resorptive treatment, such as, alendronate [8], and for the diagnosis of cancer bone metastases [9,10].

Three methods were published for the measurement of TRACP 5b activities. One was a kinetic method, which can specifically measure TRACP 5b in the presence of heparin, a specific inhibitor for TRACP 5a [11]. The other two are immunoassay methods combining immunoreactive binding and enzyme action of TRACP 5b. One used α-naphthyl phosphate as a selective substrate for TRACP 5b [12], while the other used the pH optimum of TRACP 5b that distinguished from TRACP 5a [13]. The pH-selective immunoassay is commercially available and has been widely used. In addition to availability and its common use, another advantage of this kit assay is that the capture antibody specifically binds to the active TRACP 5b, and it does not bind to the inactive fragments or other molecules [13–15].

To demonstrate the predictive use of the bone resorption biomarker, the clinical studies may be carried out over several years

* Corresponding author. Current address: Translational Science, MedImmune Inc., One MedImmune Way, Gaithersburg, MD 20878, United States.
Tel.: +1 301 398 2593; fax: +1 301 398 7292.
E-mail address: wuyu@MedImmune.com (Y. Wu).

0731-7085/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.jpba.2009.03.002

to correlate the biomarker to the clinical outcome and or BMD. Large number of samples from many cohorts require sample analysis that may be performed at multiple labs. Therefore, a robust and reliable assay is needed to support the drug development. At the onset of the drug development program, the first generation TRACP 5b commercial kit was a research kit that was not adequate for reliable quantification. It had only three non-zero calibrators, which were insufficient to define the sigmoidal function for a ligand binding assay [16]; The regression method was linear; Kit calibrators and QCs were in buffer; and optical density (OD) was used as method acceptance criteria. In order to use the kit to produce reliable and robust quantitative results, the 'for research use only' kit method was modified and validated to meet the purpose of drug development.

This paper illustrates the method modification and validation conducted with the 'Fit-For-Purpose' approach as described by Lee et al. [17]. The method modifications included changing the assay matrix, using a consistent bulk reference material to prepare more standard levels to define the regression function, and to prepare validation samples (VS) to characterize assay performance for setting QC acceptance criteria. In addition, samples were pooled to form high and low SC to reflect authentic samples for stability and to monitor kit lot variability.

Although the intended application was to explore TRACP 5b as a biomarker of bone resorption, method validation was more rigorous than that for an exploratory biomarker due to the expectation of a large number of samples from clinical studies to be analyzed by multiple sites. The assay was applied to support the potential use of TRACP 5b as a pharmacodynamic biomarker to correlate the clinical effect of therapeutics in bone and cancer. The assay performance demonstrated method reliability for the quantification of TRACP 5b in human serum samples. The clinical data imply that TRACP 5b can be used as a biomarker of bone resorption in drug development.

2. Materials and methods

2.1. Materials

The BoneTRAP® assay kit was manufactured at IDS (Fountain Hills, AZ). The major components consist of kit calibrators of human recombinant TRACP, controls of human recombinant TRACP, microplate, releasing reagent, wash buffer, sample diluent, substrate buffer and substrate tablets. There were three standard levels in the first generation kit used at the early stage of the drug development program; the more recent kit had five standard levels. The bulk reference material, which was a human recombinant TRACP (~200 U/L), was purchased from the kit manufacturer. Human sera from different populations were purchased from Bioreclamation (Hicksville, NY).

2.2. Reagents preparation

2.2.1. Standards

To calibrate the bulk reference material, the reference material was diluted to the targeted QC concentrations of 7.50, 4.00 and 2.50 U/L. The enzyme unit was defined as μmoles of *p*-nitrophenyl phosphate hydrolyzed at 37 ± 3 °C for 1 h. The diluted samples were assayed against the kit calibrators. If the diluted sample results agreed with the nominal values, the bulk reference material concentration would be accepted. Once the calibration of the bulk reference material was confirmed, all subsequent validation experiments were performed using standards prepared from the bulk reference material.

An experiment was conducted to test matrix effect of human serum against standard and QCs in buffer. Standards were prepared in buffer. Two sets of VSs, each prepared in buffer or human serum and their concentrations determined against the standard curve in buffer. Because no difference in concentrations of VSs was observed between buffer and serum, it was concluded that human serum can be used for standards and VSs preparation in subsequent experiments.

For the accuracy and precision experiments, six standard sets of eight levels were independently prepared by spiking TRACP 5b reference material into human serum and analyzed in six assays. The standards were stored at −70 ± 10 °C. The nominal concentrations of the standard were: 1.00, 1.30, 2.00, 3.00, 5.00, 6.00, 8.00 and 10.0 U/L of TRACP 5b. A low anchor point, outside the range of quantification (0.50 U/L), was included to facilitate curve fitting.

2.2.2. Validation samples, quality controls and sample controls

The endogenous concentrations of TRACP 5b in multiple human serum lots were determined against the kit calibrator (initial screening). The blank control was pooled from lots with below quantifiable levels and saved for the preparation of standard and QC. A low and a high SC were formed by pooling serum lots with relatively low or high levels, respectively.

For accuracy and precision experiments, six sets of VS were independently prepared by spiking the bulk reference material into human serum and analyzed over six assays in replicates of six. Each VS preparation was stored at −70 ± 10 °C prior to analysis. The VS levels were designed to span the entire standard working range to evaluate assay performance. These included two possible levels of LLOQ and ULOQ, and three levels at low, mid and high of the standard curve. The nominal concentrations were: 1.00 (LLOQ1), 1.30 (LLOQ2), 2.50 (LQC), 4.00 (MQC), 7.50 (HQC), and 10.0 (ULOQ) U/L of TRACP 5b.

Because the reference material was a human recombinant TRACP 5b, samples of high and low endogenous levels were pooled to form the high and low SC to reflect the endogenous TRACP 5b. The average concentrations of the serum pools were determined during validation over six independent assays, each with five replicates.

2.3. Methods

2.3.1. Assay procedure

The microtiter plates supplied by the vendor were coated with mouse monoclonal anti-TRACP antibody. Sample diluent, standards, QC, VS, and SC were pipetted into the designated wells. All samples were assayed in duplicate. A releasing reagent was added to the wells and incubated at ambient room temperature (ART) for 60 ± 5 min with shaking on an orbital shaker set at 850–950 rpm (Titermix 100, Brinkmann, Westbury, NY). The antibody immobilized on the plate captured TRACP 5b in the samples. Following a wash, the *p*-nitrophenyl phosphate substrate solution was added to the wells and incubated at 37 ± 3 °C for 60 ± 5 min. TRACP 5b, a phosphatase, converted the colorless substrate into a colored compound. The color reaction was stopped with sodium hydroxide and the OD measured at 405 nm. The conversion of OD units to the concentration of TRACP 5b was achieved through a computer software mediated comparison to a standard curve on the same plate using a four-parameter logistic regression model with a weighting factor of $1/Y^2$ using Watson LIMS Version 7.0.0.01 (Thermo Scientific, Waltham, MA) data reduction package.

2.3.2. Selectivity test

Sixteen individual lots of normal human serum (eight male and eight female subjects) with endogenous values greater than 1.30 U/L were used to determine selectivity. TRACP 5b reference material was spiked into each lot at a nominal concentration of 2.50 U/L. The control (unspiked) and spiked samples were each analyzed in one replicate. The spike recovery was calculated by the

Table 1
Method modification summary.

Parameter	Kit assay (first generation)	Modified assay	Advantage of modification
Standards	Pre-made, supplied in the kit	Made from custom supplied bulk reference material	Use of consistent reference material throughout the drug development program and across multiple labs
Standard matrix	Buffer	Human serum	Same as the intended matrix of unknown samples
Number of standards	3 levels (1, 5, and 10 U/L)	7 levels (1.3, 2, 3, 5, 6, 8, and 10 U/L)	More points to define the nonlinear standard curve
Anchor point	None	1 anchor point (0.5 U/L)	Better curve fitting
Curve fit	Linear, no weighting	4-parameter, $1/Y^2$ weighting factor	Appropriate regression model for enzyme assays
Assay range	1–10 U/L	1–10 U/L	LLOQ and ULOQ verified from VS performance
Sensitivity	1 U/L	1 U/L	LLOQ defined with targeted total error
Assay VS/QC	3 controls, levels vary by kit lot	5VSs, 3QC levels made from bulk reference material	Multiple controls to track assay performance over the assay range
QC matrix	Buffer	Human serum	Same as the intended matrix of unknown samples
Sample controls (SC)	None	Serum pools of high and low levels prepared at Amgen	Endogenous controls to track sample stability and kit lot variability
Acceptance criteria	OD limits of blank and standards	4-6-25 (similar to the FDA guidance for bioanalytical assays supporting PK)	Statistically based

mean observed concentration subtracting the endogenous level of the control divided by the nominal spiked concentration.

The interference of the therapeutic candidate on the assay performance was evaluated by spiking the therapeutic into the high and low SC at concentrations of 10, 100, 1000, and 10,000 ng/mL. The measured TRACP 5b concentrations in high and low SC pools were evaluated.

2.3.3. Parallelism test

Serum lots were screened for their endogenous concentrations of TRACP 5b. Three lots with relatively high concentration of TRACP 5b were chosen for the test. Each lot was diluted with appropriate dilution factors so that the concentrations after dilution would be within the assay range. Lot 1 was diluted at 1.5-, 2- and 2.5-fold; lot 2 was diluted at 2- and 5-fold; and lot 3 was diluted at 5-, 6-, 8- and 10-fold.

2.3.4. Stability test

The high and low SC to reflect the endogenous TRACP 5b was used for freeze/thaw and $-70 \pm 10\,^{\circ}C$ stability testing. Three aliquots from each test condition were analyzed ($N = 3$). Their mean concentrations were compared with the reference range of ± 2 S.D. established during accuracy and precision experiments.

2.3.5. Clinical sample collection

Blood samples were collected from individuals using 10 mL collection tubes. Specimens were inverted gently for 5–10 times. Blood was allowed to clot for 30 min in upright position at ART. Once clotted, the specimen was centrifuged at $1800 \times g$ for 10 min. Serum was transferred into 2×3 mL cryovials and stored at $-70 \pm 10\,^{\circ}C$ until analysis.

3. Results

The TRACP 5b kit assay was modified in our laboratory. The method was validated in our laboratory and at two contract research laboratories (CRO) to support clinical studies over a time span that would require the use of multiple kit lots. The method modification and its advantages are summarized in Table 1. The rationale and details are discussed in the following sections.

3.1. Calibration of TRACP 5b reference standard

To minimize lot-to-lot variability of kit calibrators, we obtained a custom bulk reference material to provide a consistent source for multiple analytical sites to support long-term clinical studies. Instead of using the kit calibrators and controls, the standards and QCs were prepared from spiking the reference material into human serum. The reference material was first calibrated per manufac-

Table 2
Calibration of TRACP 5b reference standard.

QCs	Nominal concentration (U/L)	Calibrated concentration (U/L)	Mean calibrated concentration (U/L)	%AR
HQC	7.5	7.324 7.084	7.204	96
MQC	4	4.116 4.039	4.078	102
LQC	2.5	2.561 2.602	2.582	103

The bulk reference material was diluted to the targeted QC concentrations of 7.50, 4.00 and 2.50 U/L with sample diluent in the kit and assayed against the kit calibrators according to the kit method. % analytical recovery (%AR) was the mean calibrated concentration divided by the nominal concentration. Calculated concentration: 200 U/L (mean %AR × 200 U/L).

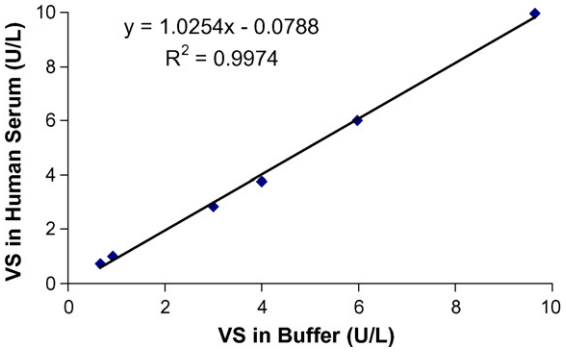

Fig. 1. Correlation plot of VSs in buffer and human serum. Validation samples were spiked in buffer and human serum at 0.70, 1.00, 3.00, 4.00, 6.00 and 10.0 U/L, respectively. The VS concentrations were determined with a buffer standard curve and plotted.

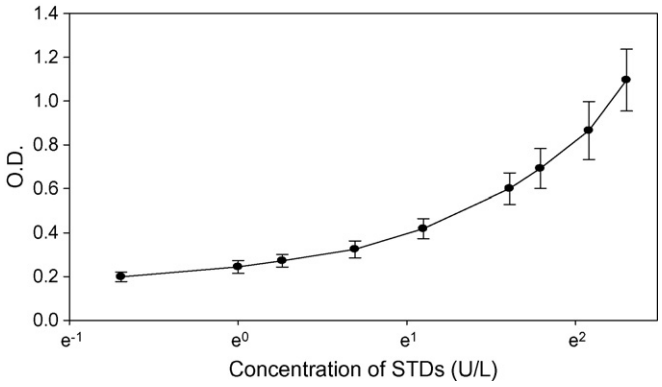

Fig. 2. Standard curve for TRACP 5b in human serum. Optical densities were determined from six assay runs and the mean values plotted vs. the concentration of standards (STDs) through a four-parameter regression (auto estimate) model with a weighting factor of $1/Y^2$ using Watson LIMS Version 7.0.0.01 data reduction package.

turer's recommendations against the kit calibrators. The average recovery (%AR) from the three levels of QCs was 100%, therefore, the concentration of the reference material provided by the vendor, 200 U/L was confirmed. This concentration was used for computing the nominal concentrations for standards and QCs (Table 2).

3.2. Comparison of validation samples in buffer and human serum

The standards and QCs provided in the kit were prepared in buffer, which was different from the intended matrix of the clinical samples. To support drug development, QCs should represent clinical samples and be prepared in the same matrix [17]. An experiment was conducted to compare VSs in buffer vs. those in serum against a buffer standard curve to assess matrix effect over the assay range. The correlation plot in Fig. 1 shows a slope of 1.025, intercept at 0.08 and a correlation coefficient of 0.997. Therefore, the results in serum correlated with those in buffer across all levels, indicating that the blank serum matrix can be used with no matrix effect or loss of assay sensitivity in comparison to the buffer matrix.

3.3. Method validation

3.3.1. Accuracy and precision experiments

Based on the above results, we proceeded to prepare standards and VSs in serum for method validation. The VS concentrations covered the entire standard curve for statistical evaluation. The *a priori* target acceptance criterion for VSs was set at 30% total error. The % total error (TE) of the VSs from the accuracy and precision experiments were used to determine the QC acceptance criteria for sample analysis.

3.3.2. Standard curve

The first generation kit had only three levels of calibrators, which is insufficient to define the sigmoidal function of a ligand-binding assay [16]. In addition, the kit calibrators may differ from one lot to another. Therefore, a bulk reference material was used to prepare consistent and sufficient standard points at eight levels plus one anchor point at the low end to better define the sigmoidal function. The kit method used linear regression without weighting (Table 1). With adequate standard points, the standard curve was best fit with a four-parameter regression model with a weighting factor of $1/Y^2$ (Fig. 2). Inter-assay % bias of the back-calculated standards ranging from −2% to 4% indicated excellent curve fitting.

3.3.3. Validation samples and quality controls

The six levels of VS covered the entire standard working range to evaluate assay performance statistics in the accuracy and precision experiment, including two possible levels of LLOQ at 1.0 and 1.30 U/L, ULOQ at 10.0 U/L and three QC levels. The accuracy and precision results shown in Table 3 defined an acceptable assay range of 1.00–10.0 U/L that is the same as the kit assay. The total error at LLOQ of 1.00 U/L and ULOQ of 10.0 U/L were 22% and 8%, respectively. The in-house data support the LLOQ at 1.00 U/L, which met the required *a priori* targeted total error of ≤30%; however, to be consistent with the contract laboratories' LLOQ of 1.30 U/L, the standard curve range was determined to be 1.30–10.0 U/L. The total error at 1.30 U/L was 21%.

The accuracy and precision experiments were conducted by two analysts using six sets of independently prepared standards and VSs. Fig. 3 shows the contributions of systemic error (% bias) and random error (imprecision) to total error (CV + |RE|). The inter- and intra-assay accuracy and precision data are shown in Table 3. There was an overall negative mean bias for the six runs, ranging from −3.8% to −9.6%. The imprecision tended to be higher toward the lower concentrations. Acceptance criteria for in-study were set based on the accuracy and precision performance: for standards, at least six standard levels must be within 20% of the nominal, and for QCs, four out of six QC levels should be within 20%.

Table 3
Validation sample accuracy and precision summary.

Characteristic	Statistic	Nominal concentration					
		VS1 (1.00 U/L)	VS2 (1.30 U/L)	VS3 (2.50 U/L)	VS4 (4.00 U/L)	VS5 (7.50 U/L)	VS6 (10.0 U/L)
Accuracy	Mean bias	−9.0	−8.2	−9.6	−7.0	−5.1	−3.8
Precision	Intrabatch (%CV)	6.4	6.2	3.7	3.6	3.1	3.0
	Interbatch (%CV)	12.6	12.6	7.5	6.1	4.7	4.2
Accuracy + precision	Mean + interbatch	21.6	20.8	17.1	13.1	9.8	8.0

Data from six accuracy and precision runs. Each VS were analyzed in six replicates. The intra- and inter-batch statistics were calculated according to DeSilva et al. [27]. Total error = accuracy + precision. $N = 36$.

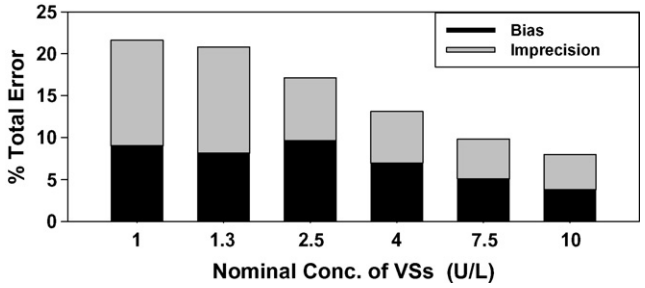

Fig. 3. Total error, bias and imprecision of validation sample. Data were from six accuracy and precision assays. Black bar represents % bias and gray bar represents imprecision, the sum approximates % total error.

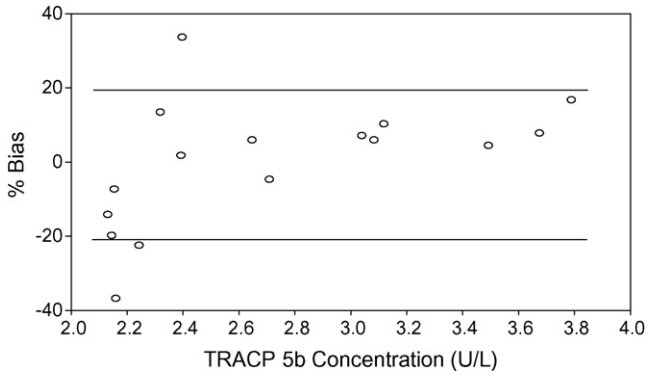

Fig. 5. Selectivity test by spiked recoveries in normal human serum. Serum from 16 (8 each from male and female subjects) individuals was tested by spiking 2.5 U/L TRACP 5b. The % recovery was calculated after subtraction from the endogenous concentration (*X*-axis). Each point represents the result from one individual serum. Three out of 16 results were beyond the ±20% bias boundaries.

3.4. High and low serum pools as sample controls

A human recombinant TRACP 5b reference material was used for the preparation of standards and VSs; the recombinant form may not truly represent the endogenous TRACP 5b. Therefore, samples of high and low endogenous levels were pooled as SC to reflect the endogenous TRACP 5b. The SC were used to test sample stability and track kit lot variability.

The SC concentrations were determined from nine independent runs with three replicates each ($N = 27$). The average concentration of TRACP 5b for the high pool was found to be 4.79 U/L. The reference range of ±2 S.D. was 2.14–7.45 U/L. The average concentration of TRACP 5b in the low pool was found to be 2.26 U/L. The reference range of ±2 S.D. range was 1.10–3.41 U/L. The same practice for setting up SC reference ranges was followed across laboratories.

3.5. TRACP 5b ranges in different human populations

The endogenous TRACP 5b levels were surveyed using individual sera from four different populations: (a) females aged 18–24 years ($N = 30$), (b) postmenopausal females ($N = 33$), (c) male aged 18–24 years ($N = 25$), and (d) male older than 50 years ($N = 32$). The samples were collected in bags from one supplier and ran at Lab 1 (at Amgen) and 2. There were no significant differences in the levels among the mean values of the four populations. However, about 84% samples tested (101 out of 120 samples) were below quantification limit (BQL).

Another range finding experiment was conducted at Lab 3 on sera from the similar populations. The samples were collected in serum separator tubes from the same supplier. All samples tested were quantifiable within the assay range (Fig. 4a). We suspected that the sample collection procedures may cause the differences in the observed levels. To further investigate this, Lab 2 repeated the range finding experiment with samples that were collected in the tubes from the same populations ($N = 20$ in each group) without changing the supplier. The data demonstrated that all samples tested were quantifiable (Fig. 4b). The ranges, 1.50–6.31 U/L (Fig. 4a) and 1.35–5.20 U/L (Fig. 4b) were comparable between the two labs for all the populations. The majority of samples were within the assay range. The result implies that the sample collection procedures may contribute to the discrepancies in TRACP 5b levels among the earlier results.

We also observed that results from the young males tend to have higher TRACP 5b levels than the young females. The postmenopausal samples also have a higher level than the young females. It was not clear about the age effect on the male population. However, the sampling size may be too small to draw a definite conclusion.

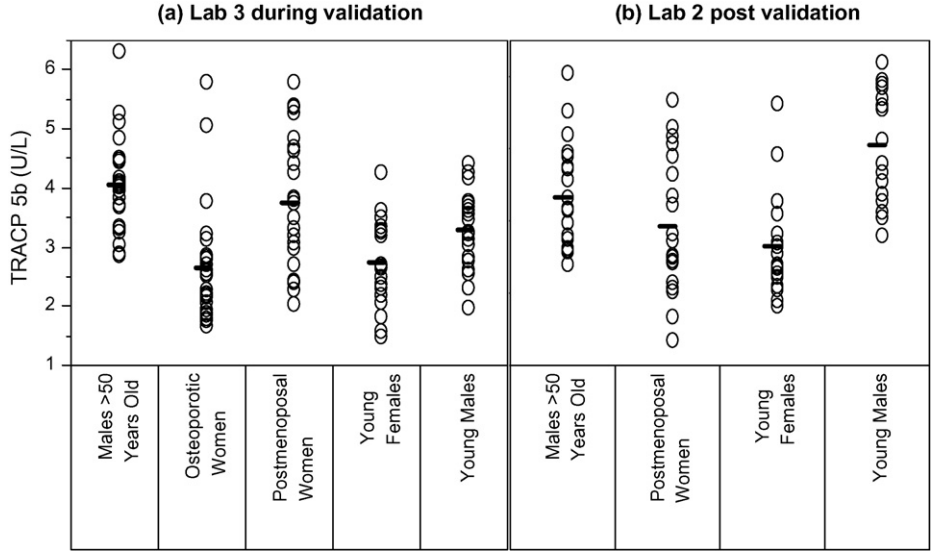

Fig. 4. TRACP 5b levels in human serum of different populations. Serum samples were collected into serum separator tubes from individuals of the target populations from the same supplier. (a) Range finding in lab 3 performed during validation. $N = 25$ for each group except for normal male and female groups $N = 20$. (b) Range finding in lab 2 performed post-validation. $N = 20$ for each group.

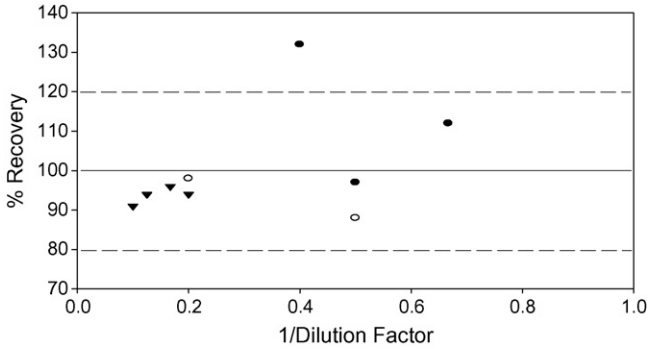

Fig. 6. Parallelism test with three endogenous serum lots. Solid circles: lot 1 of 4.285 U/L, diluted 1.5, 2.0 and 2.5-fold. Open circles: lot 2 AQL at ~10 U/L, diluted 2 and 5-fold. Triangles: lot 3 AQL at ~40 U/L, diluted 5, 6, 8 and 10-fold. Reference lines of recovery: solid 100%, dotted: 80 and 120% acceptance.

3.6. Selectivity (matrix effect) and specificity

To test possible matrix interference from individuals, spike recovery test was conducted on sera from 16 (8 each from male and female subjects) normal healthy individuals. TRACP 5b reference material was spiked into each lot at 2.50 U/L. The spike recovery was calculated after subtraction from the endogenous concentration and compared with the mean recovery of all lots. The mean recovery was 2.4 U/L (N = 16, CV% was 17%). The % bias against the endogenous concentrations (unspiked serum) is shown in Fig. 5. It ranged from −20% to 17% for 13 out of 16 lots (~81% of the test lots). The other three lots had % bias of −37%, −23% and 34%, respectively. The accurate spike recovery indicated that the majority of the test samples do not show matrix effect.

Since TRACP 5b biomarker is intended to be used to support a therapeutic under development, we evaluated if the therapeutic affected assay performance. Regardless of the drug concentrations (up to 10,000 ng/mL) added to the high or low SC samples, the TRACP 5b levels stayed within the 2 S.D. boundaries (data not shown). There was no effect of therapeutic on TRACP 5b determinations in either the high or low SC.

It is reported that the TRACP 5b commercial kit assay is highly specific for TRACP 5b without significant contribution by TRACP 5a [14,15].

Table 5
TRACP 5b validated method used in clinical studies.

Clinical studies	Study 1	Study 2	Study 3	Study 4
Data range (U/L)	<1.3–48	<1.3–12	<1.3–38	<1.3–24
Number of samples analyzed	863	2565	6736	8716
Number of sample data <1.3 U/L (<LLOQ)	19 (2%)	67 (3%)	41 (1%)	83 (1%)

Values in parentheses were the percentage of the number of samples at BQL out of total number of samples analyzed.

3.7. Parallelism

Because the standards were recombinant TRACP 5b prepared in a pooled serum lot, it was necessary to establish that the endogenous analyte behave similarly to the standard in this assay. Three serum lots with high endogenous TRACP 5b levels were tested for parallelism. The endogenous concentration of lot 1 was determined to be 4.29 U/L from six runs. It was diluted with the standard serum lot at 1.5, 2 and 2.5-fold for parallelism test (solid circles in Fig. 6). Two additional samples from clinical studies were identified to have concentrations at above quantification limit (AQL, >10.0 U/L) (lots 2 and 3). They were used for additional parallelism tests. The extrapolated concentrations of lots 2 and 3 were estimated to be ~10.0 and ~40.0 U/L, respectively. The appropriate dilution factors at 2 and 5 folds were applied for lot 2 (open circles in Fig. 6) and 5, 6, 8 and 10 folds for lot 3 (triangles in Fig. 6). All diluted sample results, except one (lot 1 at 2.5-fold), were within the acceptance criteria of ±20% (dotted line in Fig. 6). Therefore, the dilutional proportionality of the authentic sample results demonstrates parallelism.

Dilutional linearity was tested using samples spiked with the recombinant reference material into three individual serum lots at concentrations of 8.00, 10.0 and 12.0 U/L, and then serially diluted 1/2, 1/4, 1/5, 1/6, 1/8, 1/10, and 1/12. The expected concentrations after dilution were to fall within the calibration range. When the final diluted concentration was compared with the nominal, all serum lots showed recoveries within acceptance criteria of +20% bias at all dilutions (data not shown). The data demonstrated dilution linearity up to 12-fold dilution within the assay range.

3.8. Stability

To establish how many free/thaw cycle serum samples could be gone through and how long serum samples could be stored at

Table 4
Stability: freeze/thaw cycles and sample storage at −70 ± 10 °C.

Pools	Baseline[a] (U/L)			1 month		6 months		24 months	
	Mean	+2 S.D.	−2 S.D.	Result	Mean	Result	Mean	Result	Mean
Low pool	2.098	2.351	1.845	2.082		2.359		1.906	
				2.247	2.153	2.254	2.303	1.970	1.929
				2.130		2.296		1.910	
High pool	3.028	3.527	2.528	2.934		3.251		3.276	
				2.911	2.935	3.251	3.246	3.114	3.045
				2.959		3.236		2.745	

Pools	Freeze/thaw cycle 3		Freeze/thaw cycle 4		Freeze/thaw cycle 5		Freeze/thaw cycle 6	
	Result	Mean	Result	Mean	Result	Mean	Result	Mean
Low pool	1.948		1.855		1.679[b]		1.930	
	2.034	2.003	1.845	1.838[b]	1.677[b]	1.690[b]	1.760[b]	1.766[b]
	2.027		1.815[b]		1.713[b]		1.607[b]	
High pool	2.735		2.751		2.550		2.292[b]	
	2.854	2.800	2.815	2.838	2.530	2.506[b]	2.323[b]	2.370[b]
	2.812		2.949		2.438[b]		2.496[b]	

The unit for all the number in the table is U/L.
[a] Baseline endogenous TRACP 5b concentration determined in accuracy and precision.
[b] Outside ±2 S.D. range.

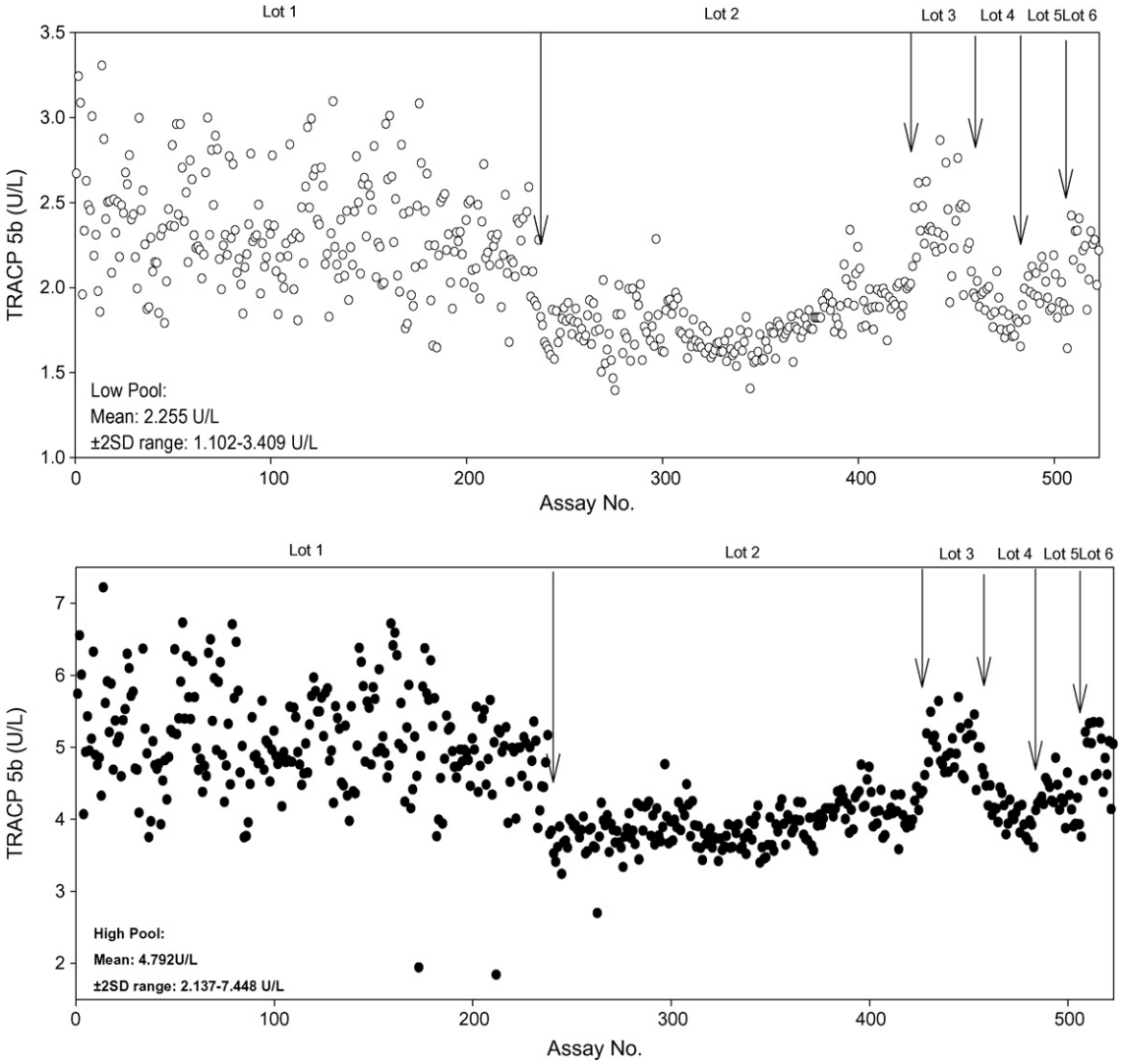

Fig. 7. Sample control pools used in multiple clinical studies during sample analysis. One single lot for each high and low serum pools were included on every single assay ($N = 2$, two wells per pool) during sample analysis over a time span of 36 months in multiple clinical studies. Mean value was calculated. The arrow referred the kit lot change. Upper panel: low pool; lower panel: high pool. $N = 238$, 190, 31, 22, 15 and 9, respectively for a total of 505 observations.

$-70 \pm 10\,^{\circ}C$ prior to analysis, high and low pools were used for stability testing. Three aliquots from each test condition and the corresponding control were analyzed ($N = 3$). The mean concentrations were compared with the reference range of ±2 S.D. established during accuracy and precision experiments.

In freeze/thaw stability experiments, high and low pools for cycle 3 and high pool in cycle 4 had measurable concentrations within the ±2 S.D. range (Table 4). For low pool at cycle 4, the measured mean concentration ($N = 3$) is slightly below the 2 S.D. limit (1.838 U/L vs. 1.845 U/L). This low value is caused by a single value out of the three aliquots being below the limit (1.815 U/L vs. 1.845 U/L). This could be due to the higher variability at the low end of the method. In view of above, stability is deemed to be acceptable for four freeze/thaw cycles. Measured concentration in both high and low pools for cycles 5 and 6 were outside the ±2 S.D. range (Table 4).

For sample storage at $-70 \pm 10\,^{\circ}C$ for 1, 6 and 24 months, the measured concentrations in both high and low pools were within the ±2 S.D. range established during accuracy and precision (Table 4).

The data indicated that the TRACP 5b in human serum is stable for up to four freeze/thaw cycles and for 24 months at $-70 \pm 10\,^{\circ}C$ (long-term stability beyond 24 months is ongoing).

3.9. Evaluation of kit lot variations by serum control during sample analysis

To support multiple clinical studies a large pool of the high and low SC were prepared, aliquotted and stored. A set of the SC was included in every assay during clinical sample analysis. Over a time span of about 36 months, six different kit lots were used among multiple studies. Fig. 7 shows the SC concentration chart during this time. Among the six lots used, lots 1 and 2 were used for about 200 runs each. There was no upward or downward trend of the concentrations, indicating stability of the endogenous analyte. There were similar shifts in both pools upon changes in kit lots. Thus, the SC provide valuable information for endogenous sample stability and for monitoring variability due to kit lot changes. For example, the change from lot 1 to 2 had a mean bias of 24% for both low and high pools, the inter-run variation of lot 2 results was less than that of lot 1. These variability should be taken into consideration during data assessment if the samples from patients were not analyzed using the same kit lot or within one run.

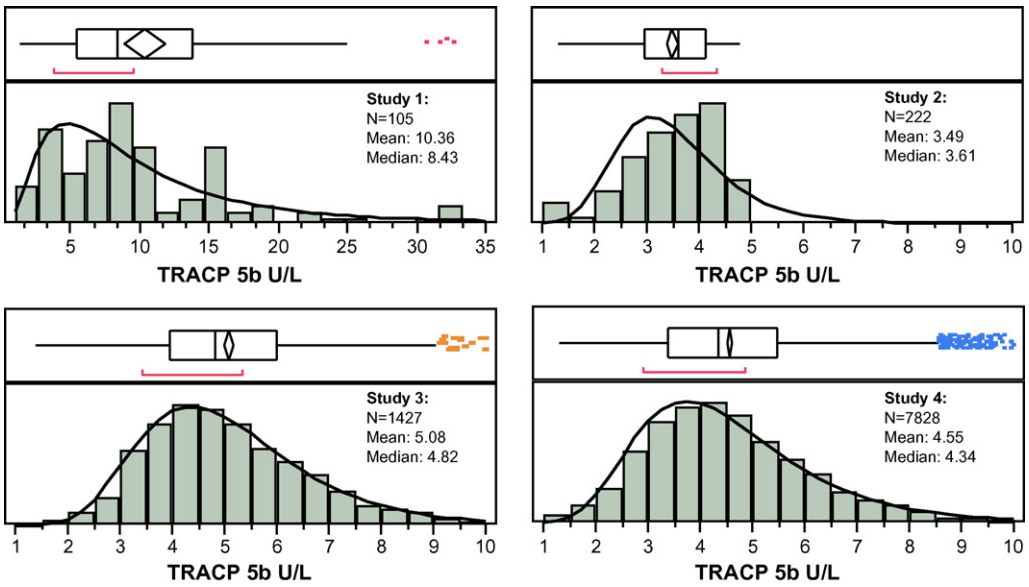

Fig. 8. Distributions of endogenous TRACP 5b levels of pre-dose samples in four clinical studies of different populations. The distributions were presented in mean and median. Solid line represents log normal distribution.

3.10. Application of TRACP 5b biomarker assay in clinical studies

TRACP 5b was used as an exploratory bone resorption biomarker in a clinical drug development for bone disease indication. The validated TRACP 5b method was applied to four clinical studies of different populations and patient numbers. Table 5 shows the endogenous TRACP 5b levels in human serum from pre- and post-dosed time points. The range was from <1.30 to 48.0 U/L. Most of the samples were quantifiable. Sample with BQL results comprised only 1% in all four studies (210 out of a total of 18,880 samples). Samples with AQL concentrations were diluted with the appropriate dilution factors demonstrated method's dilutional linearity and parallelism.

The distributions of endogenous TRACP 5b values of pre-dose samples from the four studies were presented in Fig. 8. The mean and median TRACP 5b levels were 10.4 and 8.43 U/L for study 1, 3.49 and 3.61 U/L for study 2, 5.08 and 4.82 U/L for study 3, and 4.55 and 4.34 U/L for study 4, respectively. The features of these sample distributions indicate that the validated method has suitable dynamic range and dilutional linearity to measure TRACP 5b levels in the clinical samples.

Patients were dosed with two anti-resorptive drugs through intravenous injections every 4 weeks for drug 1 and subcutaneous injections 12 weeks for drug 2. For each subject, the initial concentration of TRACP 5b (pre-dose) was set as baseline, and all subsequent values at each time point were expressed as a percentage change from the baseline. The median of percentage changes of all subjects over time by treatment is shown in Fig. 9. Treatment of drug 1 induces a significant 39% to 55% decrease in TRACP 5b over 57 weeks. A similar pattern, but even more significant decrease was observed in the group receiving drug 2 treatment over 33 weeks span, there is an apparent 42% to 70% drop in TRACP 5b activities. In the weeks of 45 and 57, TRACP 5b levels tend to come back to baseline, but still showed a 20% decrease. The % change in TRACP 5b activities reflected the effectiveness of therapeutics in the clinical studies. It demonstrated that TRACP 5b can be used as an anti-resorptive biomarker.

4. Discussion

Many commercial biomarker kits are used for biomarker measurements for the purpose of diagnosis [18,19]. As biomarkers are

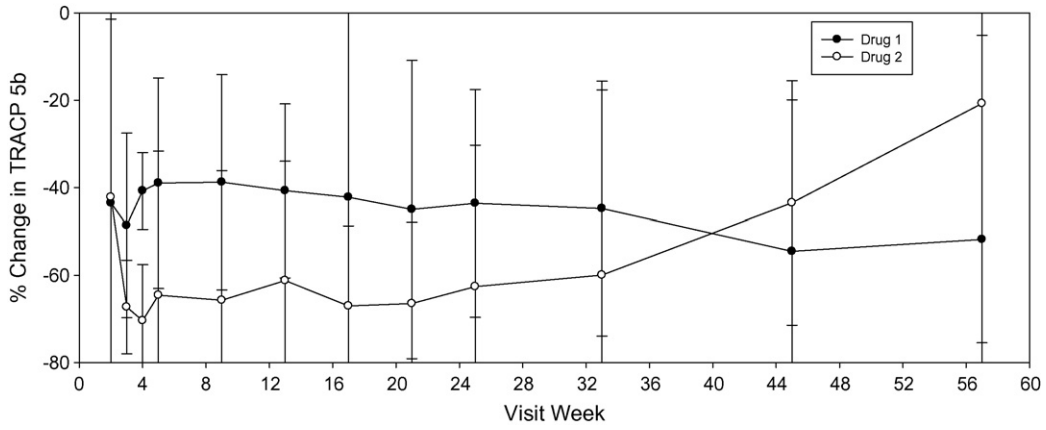

Fig. 9. Percentage changes of TRACP 5b during 57 months in the two treatment groups. Two groups of patients receiving anti-resorptive agents, drugs 1 and 2. The initial concentration of TRACP 5b (pre-dose) was set as baseline. The results are shown as the median ±2 S.D. The median is calculated based on the % change from baseline at same time point (visit week) in each group. The filled circle represents drug 1 treatment. The open circle represents drug 2.

widely used to support drug discovery and clinical development [20], the Fit-For-Purpose biomarker approach can provide reliable data meeting the demanding timeline and limited resources during drug development. Commercial biomarkers for diagnostic uses are often used for decision-making in drug development. Because the data for drug development are used for quantitative assessments of exposure/effect of the drug instead of a cutoff threshold of diagnosis, accurate and precise biomarker data are required. The analytical method of the commercial biomarker kit must be properly validated to meet the drug development application [17]. The BoneTRAP® assay kit was not designed specifically for drug development applications, thus gaps to meet this intended use had to be identified first. These gaps included: the need of a consistent source of reference material; sufficient standard points and regression model fit appropriate controls of spiked QCs as well as endogenous SC. The kit method was modified and then validated with the rigor to support clinical drug development sustainable by sample analysis from multiple analytical sites.

Two isoforms of TRACP 5a and 5b are circulating in the blood. TRACP 5a is derived from macrophages [5], while TRACP 5b is secreted from osteoclasts [6] as an active enzyme. TRACP 5a and TRACP 5b display different pH optimum, which are 5.2 and 5.8, respectively. The commercial kit chosen for this study has the advantage being immuno-selective for the specific molecule, as well as the enzyme reaction being selective for the 5b pH optimum, resulting in highly specific measurements for TRACP 5b [13–15].

One of the challenges for commercial kit utilization is kit lot variations, which may arise from kit calibrators or other components' change within the duration of a long clinical trial. For the TRACP 5b biomarker validation, we minimize variations by preparing standards and QC samples from one consistent source of high concentration reference material.

QCs were prepared by spiking a known amount of the recombinant reference material and used to accept or reject an analytical run. In addition, SC from sera pooled from individual samples to reflect the endogenous biomarkers in matrix. Their range (mean ± 2 S.D.) were established with multiple runs and used as a reference. We evaluated kit lot variability shift using SC during each run of sample analysis. The SC data were also used to monitor trends in stability of TRACP 5b in serum over time. We have established stability of TRACP 5b in serum for up to four freeze/thaw cycles and 24 months storage at $-70 \pm 10\,^{\circ}C$. The stability data provide useful information on handling samples prior to analysis to support drug development. Fig. 7 shows the SC assay fluctuations with apparent shifts at some kit lot changes. It is noted that the runs from kit lot 1 had higher variability than those from the other kit lots. The preparations of standards for the runs could cause the variability. The standards for the runs using kit lot 1 included the three kit calibrators plus adding more standards prepared from them, while the standards for runs from the other kit lots were prepared from the bulk reference material. The lower variability using the bulk reference material demonstrates the importance of implementation of consistent source of high concentration reference material. The use of the same SC set among studies and laboratories would add to conformance samples testing, enable long-term performance monitoring of kit lots, stability trend and as a basis to differentiate true study effect from analytical variability. We recommend that the same SC set should at least be used within one study.

Variations in the potency of the biological materials in different matrices and subtle methodological variations may interfere with the accuracy of measurements [21–23]. In the case of TRACP 5b, the original kit was developed in a buffer solution, while in clinical practice the matrix is serum. In our studies we showed that it was feasible to change the kit standard matrix from buffer to human serum while maintaining assay sensitivity and prepare QCs in serum with concentrations to over the entire range of the expected clinical samples.

Data from the initial range finding experiments (in labs 1 and 2) showed that about 84% samples were BQL, raising a concern that the assay was not sensitive enough. Since the TRACP 5b biomarker was intended to be used for exploratory purpose, we decided to move forward using the method for sample analysis. The later range finding experiment results from the lab 3 showed that no samples were BQL. Subsequent data from clinical studies showed that only 1% of four study samples were BQL (Table 5). The mean and median of TRACP 5b levels of pre-dose samples from four clinical studies (N = 105–7828) were also quantifiable (Fig. 8). All the pre-dose samples have a log normal distribution except for study 1 (population with advanced cancer). Study 1 samples had higher concentration (up to 35.0 U/L) that required sample dilution. All the other study sample levels were within the assay working range of 1–10 U/L. The dilutional linearity (spiked samples) and parallelism (authentic samples) data provided assurance on the accuracy of dilution for these clinical samples. The characteristics of the sample distributions determined by this assay suggest that the modified and validated method provided adequate sensitivity and functional range to measure TRACP 5b levels in our clinical samples.

Why were there discrepancies in un-dosed sample ranges during method validations among the three labs and in-study? One explanation is that the conditions of collection and storage of serum samples during validation may not be the same among the laboratories or during clinical trials. Samples obtained during validation experiments in labs 1 and 2 were from whole blood collected in bag for large volume, while samples used for validation in lab 3 or post-validation in lab 2 and clinical samples were collected in a standard "clot" tube (~10 mL). Fig. 4a and b shows that there were differences in TRACP 5b measurements dependent on sample collection procedures in either bags or tubes. Almost all the samples collected from tubes showed quantifiable results. It was reported that TRACP 5b lost enzymatic activity upon storage at higher temperature [24,25] and hemolysis [26]; both conditions may exist during the serum collection in bag. Our assay has established stability in serum for 4 h at ART (data not shown), which is in agreement with other literature [26], however, it is uncertain how long the whole blood sample collection can be kept at ART during clotting and serum separation. The common practice in the diagnostic sector to use samples from the Blood Bank should be evaluated in method development since the differences in samples collection and storage from clinical studies may cause artifacts. We recommend that sample collection for method validation experiments should be similar to that of clinical study, unless tested.

Matrix interference on assay performance was evaluated in normal healthy population, while parallelism was performed with patient samples. No matrix interference was detected from the normal or the targeted patient population samples. In addition, method robustness was evaluated during accuracy and precision experiments by two analysts performing the assays with multiple instruments over multiple days. The assay ruggedness was established with validation across multiple laboratories and sustained through successful runs in four clinical studies of 18,880 samples.

It has been reported that TRACP 5b is a promising new biomarker for the prediction of bone loss [26] and monitoring bone metastases in cancer patients [9,10]. We investigated if TRACP 5b can be used as a predictive bone resorptive biomarker on the effect of two anti-resorptive agents in a clinical study. TRACP 5b levels in patients receiving the therapeutics significantly decreased over 57 weeks. Our data showed that TRACP 5b can be a potentially predictive biomarker for monitoring the anti-resorptive responses.

In conclusion, we have modified a commercial TRACP 5b immunoassay kit through changing the assay matrix, adding more standards, and including both QCs and SC on each run. The assay

was analytically validated through a Fit-For-Purpose approach. The validated analytical range was 1.00–10.0 U/L. The ranges of endogenous TRACP 5b in different targeted populations were determined. SC pools were established and used in each run during clinical sample analysis to track the kit lot variability and sample stability. There was no matrix effect observed. No interference from the therapeutic drug was observed. Parallelism with authentic samples was demonstrated at dilutions up to 10. Samples were stable for up to four freeze/thaw cycles and 24 months at $-70 \pm 10\,^{\circ}C$. The validated method has shown to be appropriate in the measurement of TRACP 5b in clinical samples. TRACP 5b could be used as a biomarker of bone resorption.

Acknowledgements

We would like to acknowledge Cyndi Mayer, Frank Spriggs, Sariya Saiduddin, Rosagnya Rao for their participations in validation; Kimberly Klimko, Masood Khan (Covance), Sandra Beyl, Walter Bloome (Esoterix), Ed Lee, Winnie Sohn, Graham Jang and Mark Peterson for their contributions to this work.

References

[1] A.C. Looker, D.C. Bauer, C.H. Chesnut, C.M. Gundberg, M.C. Hochberg, G. Klee, M. Kleerekoper, N.B. Watts, N.H. Bell, Osteoporos. Int. 11 (2000) 467–480.
[2] E. Fink, C. Cormier, P. Steinmetz, C. Kindermans, Y. Le Bouc, J.C. Souberbielle, Osteoporos. Int. 11 (2000) 295–303.
[3] P.D. Delmas, Osteoporos. Int. 6 (2000) 66–76.
[4] K. Henriksen, L.B. Tanko, P. Qvist, P.D. Delmas, C. Christiansen, M.A. Karsdal, Osteoporos. Int. 18 (2007) 681–685.
[5] A.J. Janckila, D.H. Neustadt, Y.R. Nakasato, J.M. Halleen, T. Hentunen, L.T. Yam, Clin. Chim. Acta 320 (2002) 49–58.
[6] A.J. Janckila, K. Takahashi, S.Z. Sun, L.T. Yam, Clin. Chem. 47 (2001) 74–80.
[7] P. Gerdhem, K.K. Ivaska, S.L. Alatalo, J.M. Halleen, J. Hellman, A. Isaksson, K. Pettersson, H.K. Väänänen, K. Akesson, K.J. Obrant, J. Bone Miner. Res. 19 (2004) 386–393.
[8] A. Nenonen, S. Cheng, K.K. Ivaska, S.L. Alatalo, T. Lehtimäki, H. Schmidt-Gayk, K. Uusi-Rasi, A. Heinonen, P. Kannus, H. Sievänen, I. Vuori, H.K. Väänänen, J.M. Halleen, J. Bone Miner. Res. 10 (2005) 1804–1812.
[9] E. Terpos, J. de la Fuente, R. Szydlo, E. Hatjiharissi, N. Viniou, J. Meletis, X. Yataganas, J.M. Goldman, A. Rahemtulla, Int. J. Cancer 106 (2003) 455–457.
[10] J. Korpela, S.L. Tiitinen, H. Hiekkanen, J.M. Halleen, K.S. Selander, H.K. Väänänen, P. Suominen, H. Helenius, E. Salminen, Anticancer Res. 26 (2006) 3127–3132.
[11] R.A. Hannon, J.A. Clowes, A.C. Eagleton, A. Al Hadari, R. Eastell, A. Blumsohn, Bone 34 (2004) 187–194.
[12] A.J. Janckila, R.M. Simons, L.T. Yam, Clin. Chim. Acta 347 (2004) 157–167.
[13] J.M. Halleen, S.L. Alatalo, H. Suominen, S. Cheng, A.J. Janckila, H.K. Väänänen, J. Bone Miner. Res. 15 (2000) 1337–1345.
[14] J.M. Halleen. IDS Ltd documents.
[15] J.M. Halleen, H. Ylipahkala, S.L. Alatalo, A.J. Janckila, J.E. Heikkinen, H. Suominen, S. Cheng, H.K. Väänänen, Calcif. Tissue Int. 71 (2002) 20–25.
[16] M. Kelley, B. DeSilva, AAPS J. 18 (2007) 156–163.
[17] J.W. Lee, V. Devanarayan, Y.C. Barrett, R. Weiner, J. Allinson, S. Fountain, S. Keller, I. Weinryb, M. Green, L. Duan, J.A. Rogers, R. Millham, P.J. O'Brien, J. Sailstad, M. Khan, C. Ray, J.A. Wagner, Pharm. Res. 23 (2006) 312–328.
[18] M. Jung, S. Christgau, M. Lukoschek, D. Henriksen, W. Richter, Pathobiology 71 (2004) 70–76.
[19] P. Ravn, G. Neugebauer, C. Christiansen, Bone 30 (2002) 320–324.
[20] P. Garnero, P.D. Delmas, Curr. Opin. Rheumatol. 16 (2004) 428–434.
[21] B.N. Swanson, Dis. Markers 18 (2002) 47–56.
[22] C. Ricós, V. Alvarez, F. Cava, J.V. García-Lario, A. Hernández, C.V. Jiménez, J. Minchinela, C. Perich, M. Simón, J. Clin. Lab. Invest. 59 (1999) 491–500.
[23] C.G. Fraser, AACC Press, Washington, DC, 2001, p. 151.
[24] J.M. Halleen, T.A. Hentunen, M. Karp, S.M. Käkönen, K. Pettersson, H.K. Väänänen, J. Bone Miner. Res. 13 (1998) 683–687.
[25] Y.R. Nakasato, A.J. Janckila, J.M. Halleen, H.K. Vaananen, S.P. Walton, L.T. Yam, Clin. Chem. 45 (1999) 2150–2157.
[26] J. Lenora, K.K. Ivaska, K.J. Obrant, P. Gerdhem, Osteoporos. Int.
[27] B. DeSilva, W. Smith, R. Weiner, M. Kelley, J. Smolec, B. Lee, M. Khan, R. Tacey, H. Hill, A. Celniker, Pharm. Res. 20 (2003) 1885–1900.