Research Article

“Fit-for-Purpose” Method Validation and Application of a Biomarker (C-terminal Telopeptides of Type 1 Collagen) in Denosumab Clinical Studies

Jin Wang,^{1,2} Jean Lee,¹ Daniel Burns,¹ David Doherty,¹ Laura Brunner,¹ Mark Peterson,¹ and Binodh DeSilva¹

Received 8 January 2009; accepted 4 May 2009; published online 22 May 2009

Abstract. Biomarkers are used to study drug effects, exposure–response relationships, and facilitate early decision making during development. Denosumab, a fully human monoclonal antibody against receptor activator of nuclear factor- κ B ligand, profoundly inhibits bone resorption. C-terminal telopeptides of type I collagen (CTx), a bone resorption biomarker, provides early indications of denosumab effectiveness and informs protracted clinical outcomes (e.g., bone mineral density). Because of the dynamic relationship between denosumab and CTx, a precise and robust assay was desired. Thus, we adopted a fit-for-purpose approach to modify and validate a commercial CTx diagnostic kit to meet the intended applications of a quantitative pharmacodynamic biomarker for denosumab development. Seven standards were prepared to replace five calibrators provided in the kit. Three quality controls (QC) and two sample controls were used to characterize and monitor assay performance. Robotic workstations were used for standard and QC preparation and assay execution. Method validation experiments were conducted with rigor and procedures similar to those used for drug bioanalysis. The method demonstrated a linear range of 0.0490–2.34 ng/mL with four-parameter logistic regression. Inter-assay total error of validation samples in serum was $\leq 26.7\%$. Extensive tests were conducted on selectivity in sera from target populations, specificity, stability, parallelism, and dilutional linearity. Applications to samples from numerous clinical studies confirmed that the CTx method was reliable, robust, and fit for use as an early indicator of denosumab effectiveness. Refinement supported the confidence for use in pharmacokinetic/pharmacodynamic modeling, dose selections, correlation to clinical effects, and formulation bioequivalence work.

KEY WORDS: bone resorption marker; commercial immunoassay kit; method validation; pharmacodynamic (PD) biomarker; serum C-terminal telopeptides of type 1 (CTx).

INTRODUCTION

In bone-related diseases, bone resorption markers are extremely useful for disease diagnosis and for early indications of treatment responsiveness. Their use in drug development is advantageous, as changes in clinical endpoints, such as bone mineral density and bone fracture, require months to years, respectively, to manifest. C-terminal telopeptides of type I collagen (CTx) is a bone resorption biomarker that has been shown to be a highly effective biomarker tested for the detection of high bone turnover (1). It has been used as a biomarker for clinical development of drugs for osteoporosis and metastatic bone disease (2,3).

Denosumab is a fully human monoclonal antibody that binds to the receptor for nuclear factor κ B ligand (RANKL) and inhibits the binding of RANKL to its cognate receptor, RANK. This antibody causes the cells that resorb bone (i.e., osteoclasts) to undergo apoptosis, thereby exerting anti-resorptive effects with a mechanism different from that of

bisphosphonates (4). The bone is a dynamic tissue that is continuously formed and degraded. The degradation products of bone collagen are present in blood and urine, and measurement of the cross-linked diisomerized β CTx peptides in serum sample would logically reflect mature bone turnover (5). To support clinical development of denosumab, we utilized an immunoassay commercial kit (Serum Cross-Laps™) which uses highly specific monoclonal antibodies against the C-telopeptide fragments of type I collagen (6).

The objectives for using CTx as a pharmacodynamic (PD) marker in this development program were several, including pharmacokinetic (PK)–PD (PK–PD) modeling for dose selection, correlation of rapidly and dynamically responsive (to treatment) biomarkers to subsequent clinical events, and PD effect evaluation for formulation comparability assessments. For PK–PD modeling to expedite dose and regimen selection, reliable data of CTx with sufficient accuracy and precision are required to appropriately represent the drug effects. Therefore, more stringent criteria than a diagnostic biomarker are used for this program. Since the objectives were different, the direct adoption of the diagnostic commercial kit for drug development use was inappropriate. Instead, we followed a “fit-for-purpose” approach to modify the kit method, validate it, and implement it for clinical

¹ Department of PKDM, Amgen Inc., One Amgen Center Drive, Mail Stop 30E-3-C, Thousand Oaks, California 91320, USA.

² To whom correspondence should be addressed. (e-mail: jinw@amgen.com)

sample analysis to ensure it satisfies the objectives of the denosumab clinical development (7). The main modifications were as follows: (1) Standards were prepared from a high concentration reference material purchased from the same vendor to replace kit calibrators to assure standard consistency throughout the program. (2) More standard levels were added to better define the standard curve, and more quality control (QC) levels and authentic sample controls (SC) were incorporated to validate and monitor assay performance. (3) Immunoassay procedures were conducted on a robotic workstation.

The basic framework of method validation commonly used to support PK analyses in clinical studies was used as a basis for the method validation of CTx in human serum and to characterize assay accuracy, precision, sensitivity, selectivity, specificity, and stability (8). In addition, extensive validation experiments were conducted to support CTx as a biomarker in advanced application so that the data may be used for regulatory agency submission (9). As the study durations for the treatment of osteoporosis may be 3–5 years with large numbers of samples, specific considerations incorporated in the validation included evaluating patient population ranges, method ruggedness (within and between laboratories), consistencies in reference material and kit lots, and SC to monitor assay performance over the anticipated time span.

This paper illustrates the modification and method validation of a commercial diagnostic kit for the purpose of supporting the application of a PD biomarker to a drug development program, in this case CTx for denosumab. The developed analytical method was shown to be robust and able to sustain bioanalytical reliability across multiple analytical sites and clinical studies over the course of several years involving thousands of samples.

MATERIALS AND METHODS

Materials

The CTx reference standard (custom-ordered high standard) and other kit components were purchased from Nordic Bioscience Diagnostics (Herlev, Denmark). The kit components included: streptavidin-coated microtiter plate, biotinylated antibody, horseradish peroxidase (HRP)-conjugated antibody, incubation buffer, and 50× wash buffer. Tetramethylbenzidine (TMB) one component HRP Micro-well substrate was purchased from BioFX Laboratories (Owings Mills, MD, USA), 36 N H₂SO₄ from VWR International LLC (West Chester, PA, USA), sterile water from Hospira Inc (Lake Forest, IL, USA), and human serum from Bioreclamation Inc. (Hicksville, NY, USA). Deionized water was purified in-house.

Calibration Standards, Quality Control Samples, Validation Samples, and Serum Control Samples Preparation

Calibration Standards

The lyophilized CTx reference standard was reconstituted with sterile water and diluted with the supplied diluent for the standards to prepare nine levels of CTx standards at concentrations of 2.46 (or 2.79, as high anchor point), 2.34, 1.85, 0.820, 0.406, 0.205, 0.0800, 0.0490, and 0.0260 (low

anchor point) ng/mL. All standards were prepared on the day of assay using a Tecan Genesis Workstation 200 equipped with Gemini software v 3.5 (Tecan Group Ltd., Mannedorf, Switzerland).

Quality Controls and Validation Samples

All QCs and validation samples (VSs) were prepared in buffer in a similar way as the standards at concentrations of 2.34 (upper limit of quantification—ULOQ), 1.70 (high-level QC), 0.500 (mid-level QC), 0.150 (low-level QC), 0.0800 (low limit of quantification—LLOQ 1), and 0.0490 (LLOQ 2) ng/mL. The prepared QCs or VSs were assayed on the day of preparation. In addition to VSs in buffer, VSs in serum were prepared by spiking CTx reference standard into a normal human serum pool with the targeted endogenous CTx level at ~0.1 ng/mL. The VSs prepared in serum were stored at –70°C prior to analysis. The spiked concentrations of VS in serum were calculated by subtracting the basal endogenous CTx concentrations (baseline) from the measured concentrations.

Sample Controls

Low- and high-level SCs were prepared from pooling human sera that were prescreened to determine the endogenous CTx concentrations. Sera with low concentrations around 0.1 ng/mL were combined to form the low SC and those of >0.5 ng/mL for the high SC.

Clinical Sample Collection

Serum samples were collected from individuals using 10-mL serum separator glass tubes. Drawn blood was allowed to clot for at least 30 min at room temperature prior to centrifugation. Within 1 h of collection, samples were centrifuged at 2,000×g at 4°C for 15 min. The serum was transferred into an appropriately labeled cryovial and stored at –70°C.

Immunoassay Procedures

Each analytical run consisted of one set of standards and blank, two replicates of QCs at three levels, and SCs at two levels. Fifty microliters of standards, blank, QCs, SCs, or study samples and 150 µL of a mixture of a biotinylated anti-CTx antibody and a HRP-conjugated anti-CTx antibody were added sequentially to the streptavidin-coated plate by a Tecan Genesis Workstation 200. The plate was incubated for 120 min on a Brinkmann Titermix 100 (Brinkmann Instruments, Inc., Westbury, NY, USA) shaker set at ~300 rpm. Following a wash step, 100 µL TMB substrate was added to the plate. After incubation on the Brinkmann shaker at ~300 rpm for 15 min, 100 µL 2 N H₂SO₄ stop solution was added. The plate was read on a SpectraMax 340 PC plate reader (Molecular Devices, Sunnyvale, CA, USA) at 450 nM with reference at 650 nM.

Analytical Data Regression

The conversion of optical density (OD) units for the study samples, SCs, VSs, and QCs to concentrations was

performed using Watson LIMS (version 7.0.0.01, Thermo, PA, USA). The absorbance vs. concentration relationship was regressed according to a four-parameter logistic regression model with a weighting factor of $1/Y^2$. The back-calculated values of each standard point must meet the target acceptance criteria of within 20% of the nominal value, for at least 75% of the acceptable standard points. For sample analysis, two thirds of the QC should be acceptable within 20% of the nominal values.

Methods Improvement and Comparison

During the course of clinical development and assay use, an initial quantification method (method A) was used for early studies, and a slight improvement was made (method B) and used for later studies. The major difference between the two methods was data regression. Method A used a log-log transformation without anchor points, while method B used four-parameter logistic regression with anchor points and $1/Y^2$ weighting. Method A used a total “walk-away” automation, while method B used a modular automation program to allow flexibility and higher throughput. Method B was validated extensively and transferred to a contract research organization (CRO) to support subsequent clinical trials. All data reported here are those obtained via method B, except where noted in the method comparison section.

RESULTS

Method Validation

Standard Curve and Assay Range

Eight validation runs were performed over 3 days to determine assay range of the standard curve. Figure 1 shows the CTx dose-dependent response range from 0.0260 to 2.79 ng/mL. The standard curve showed acceptable inter-

Fig. 1. Composite CTx standard curve. Data from eight validation runs performed over 3 days. The standards were 2.46 (anchor point), 2.34, 1.85, 0.820, 0.406, 0.205, 0.0800, 0.0490, and 0.0260 (anchor point) ng/mL. Mean OD were plotted against concentrations with four-parameter logistic regression of $1/Y^2$ weighting. The error bars were standard deviation of the mean OD, which typically increase with increase responses in an immunoassay

assay precision of $\leq 7.5\%$ and accuracy in relative error (% RE) of $\leq 2.4\%$ in all back-calculated standard concentrations, excluding the anchor points. The validated assay range was from 0.0490 to 2.34 ng/mL, defined as from the lower limit of quantification to upper limit of quantification by the validation sample results (see next section). The low bias of the back-calculated standards confirmed the appropriateness of curve fitting by the four-parameter regression model with weighting factor of $1/Y^2$.

Accuracy and Precision

VSs with levels spanning the assay range were used to provide statistical performance on intra- and inter-run accuracy and precision from eight runs. Each run consisted of one set of standards and blank, two replicates of SCs at two levels, and three replicates each of six levels of VSs in buffer and in serum. To include a robustness component in this validation experiment, the runs were conducted over 3 days by three analysts at various tolerable incubation times and using two Tecan workstations.

Figure 2 shows the precision and accuracy data. The total error (TE) for the LLOQ2 VS in buffer and serum were 17.0% and 26.7%, respectively. Major error was derived from the inter-assay precision, which were 17.0% and 20.1% CV in buffer and serum, respectively. As the serum VS data are believed to represent the assay performance of clinical samples more realistically than those in buffer, LLOQ was defined to be 0.0490 ng/mL, which met the *a priori* acceptance criteria of $<30\%$ TE. The VS data indicated that the overall method performance was reliably accurate and precise to support clinical studies.

Survey of CTx Ranges in Targeted Populations

During CTx method validation, sera from normal and patient populations were surveyed. More than 50% of the serum from osteoporosis and breast cancer patients had CTx values below the LLOQ. However, these low concentrations did not reflect the pre-dose sample results from subsequent clinical studies of a larger population. Therefore, the ranges of target populations were updated by the data gathered from the clinical studies. The overall range was 0.0500–12.3 ng/mL. The median value of CTx in serum from target patient populations was 0.530 ng/mL for osteoporosis, 0.343 ng/mL for rheumatoid arthritis, 0.616 ng/mL for breast cancer, and 0.775 ng/mL for multiple myeloma (Table I).

Selectivity

To evaluate selectivity, 40 individual lots of normal human serum (20 male and 20 female) and 58 lots of male and female human sera with various disease states were tested for spike recovery of CTx. Each lot of serum was tested with and without the addition of CTx at 0.150 ng/mL. The spike recovery was calculated by subtracting the endogenous values of the unspiked samples from the spiked samples. No significant difference in percent bias on spike recovery was observed in normal serum vs. that of patients, although the latter tend to be more variable. Within the patient samples, the osteoporosis samples showed large

Fig. 2. Accuracy and precision of validation samples in buffer and human serum. The VSs were 2.34 ng/mL for ULOQ, 1.70 ng/mL for high QC, 0.500 ng/mL for mid QC, 0.150 ng/mL for low QC, 0.800 ng/mL for LLOQ1, and 0.0490 ng/mL for LLOQ2. The VSs were prepared in buffer (green- and blue-colored bars) or in human serum (red- and purple-colored bars). The total error bars were composed of the inter-assay accuracy (%RE, dotted areas) and inter-assay precision (%CV, downward lined areas)

variability with five of 18 samples exceeding 20% in bias, and rheumatoid arthritis samples showed a negative bias with five of ten samples exceeding 20% (Fig. 3).

Specificity

For advanced biomarker method validation, the ability to measure the biomarker in the presence of the therapeutic agent should also be evaluated. Denosumab as well as the structurally and functionally related compounds were tested for possible interference with CTx assay. The test compounds at various concentrations were spiked into the human serum samples. The observed results were compared to the controls, which were not spiked with the test compounds. No interference was observed by the addition of denosumab at up to 3,000 µg/mL. The target biomarker human RANKL (up to 1,120 µg/mL) and proximal biomarker osteoprotegerin (up to 260 µg/mL) had no effect on the assay. No effect was observed with the structurally similar compounds of C-

Fig. 3. Spike recovery of 0.145 ng/mL of CTx from normal and patient sera. CTx at 0.145 ng/mL was spiked into individual serum of various populations: breast cancer patients (BC, $N=20$), multiple myeloma patients (MM, $N=10$), normal female (F, $N=20$), normal male (M, $N=20$), osteoporosis patients (Osteo, $N=18$), and rheumatoid arthritis patients (RA, $N=10$). The solid line is the mean spike recovery of the normal female sera as a control comparison to other groups. The dotted lines represent the $\pm 20\%$ from the nominal spiked value. The box included data within 25 to 7th quantiles with the median line in the middle and extended values in the whiskers

terminal telopeptides type II up to 45.5 ng/mL and N-terminal telopeptides (NTx) up to 15 nM bone collagen equivalent (BCE). NTx at levels ≥ 15 nM BCE showed measurable CTx levels, probably due to the CTx contaminants in the NTx test materials that were purified from human bone.

Parallelism

Three parallelism test samples were prepared by pooling study samples with values from mid to near upper limit of quantification level of the standard curve. The test samples were diluted 1:1.5, 1:2, 1:5, and 1:10 with assay buffer and analyzed with the undiluted samples in a run. The high SC was also diluted and assayed as a control.

Table I. Serum CTx Level in Normal and Diseased Patient Population

Pre-study validation					
Disease state	Number	% Lot below LLOQ	Range (ng/mL)	Median (ng/mL)	
Osteoporosis	18	50.0	0.0540–0.544	0.249	
Rheumatoid arthritis	10	10.0	0.0860–0.886	0.219	
Breast cancer	10	60.0	0.0680–0.188	0.084	
Breast cancer metastasized	10	10.0	0.0520–0.563	0.178	
Multiple myeloma metastasized	10	10.0	0.0550–0.273	0.152	
Normal	40	2.5	0.0580–0.475	0.190	
In-study validation					
Disease state	No. of pre-dose sample	No. of pre-dose sample with conc. < LLOQ	Range (ng/mL)	Median (ng/mL)	No. of clinical study
Osteoporosis	9,410	21	0.0500–12.3	0.530	6
Rheumatoid arthritis	217	1	0.0570–1.09	0.343	1
Breast cancer	1676	2	0.0520–2.49	0.616	2
Breast cancer metastasized	271	1	0.0710–5.28	0.559	2
Multiple myeloma metastasized	119	1	0.0550–10.5	0.775	1

The parallelism results are shown in Fig. 4. The calculated concentrations (observed concentration $\times$ dilution factor) were plotted against the dilution factors. The flat profiles demonstrate that there was no effect due to dilution in buffer and that proportionality and immunoreactivity were not affected by the amount of endogenous sample or buffer. In addition to addressing the parallelism issue, the results also demonstrated dilutional linearity with all the test samples used. All diluted samples measured within 20% of the expected concentration calculated from the corresponding undiluted samples.

Stability

Extensive and rigorous stability tests on many conceivable conditions encompassing long studies were conducted for CTx. These conditions include subjecting CTx stability samples to room temperature, 4°C, and 37°C during sample processing and to multiple freeze/thaw cycles. Bench top stability was demonstrated at room temperature for 6 h, at 37°C for 1 h, and at 2–8°C for up to 22 h. The stability samples were thawed unassisted and then exposed to various bench-top conditions prior to assay. Freeze/thaw cycle stability was tested by comparing spiked or unspiked CTx samples subjected to four freeze/thaw cycles to a control set with one freeze/thaw cycle. One-year storage stability at –70°C and –20°C was performed using female postmenopausal serum samples at three levels near the high, mid, and low portion of the assay range. The data were compared to those of an initial set analyzed at the beginning of the long-term stability study. In addition, SC data generated from several multiyear clinical studies were used to establish long-term storage stability at –70°C.

All stability data from samples tested were within acceptance criteria. Figure 5 shows SC concentration data for long-term storage stability over a time span of 35 months. A trending analysis was performed using SAS version 9.3. The results showed a slightly upward trend with the concentration increase at a rate of 0.2 ng/mL every 10,000 days

Fig. 4. Parallelism test. Three authentic samples of relatively high concentrations (filled square, triangle, and circle symbols) and the HSC (open circles) were each diluted at 1.5-fold, twofold, fivefold, and tenfold with assay buffer. The observed concentrations times the dilution was plotted against 1/dilution factor

Fig. 5. In-study monitoring CTx sample control charts. **a**, HSC; **b**, LSC. Centerlines are the overall mean concentrations of LSC and HSC at 0.137 and 0.633 ng/mL, respectively. The upper and lower lines are the mean $\pm$ 2 SD values. Two lots of reference materials were used for standards and QC preparations during the time span of 3 years. Cross symbols are values from lot 1 (up to day 316) and square symbols are values from lot 2. Different shades (colors) represent different kit lots

for the low SC and 0.56 ng/mL every 10,000 days for the high SC. The trend lines indicate that the low SC would recover at 116% and high SC would recover at 110% at the 3-year time point. The small upward trend does not have a clinical meaningful impact on the study results considering the remarkable drug effect on CTx suppression (see “Clinical Applications”). We therefore concluded that CTx was stable without significant changes under the conceivable clinical and laboratory environments. The results are in agreement with the information provided by The National Committee for Clinical Laboratory Standards document which stated the 3-day stability at 4°C and greater than 3-year long-term stability at –20°C or –70°C for CTx samples (10).

Method Robustness and Ruggedness

Robustness tests were conducted on various incubation times, with multiple preparations of standard, VS, reagents, different matrix lots, batch sizes, assay days, instruments, and analysts. The tolerance boundaries of incubation times and batch size were also established. Statistical parameters on assay characteristics are defined in the acceptance criteria from these data so that out-of-specification assays can be identified.

Robustness of the assay was illustrated by the acceptable performance of eight batches of standards and VSs prepared and tested by three analysts over 3 days in two types of VS

matrices and using two Tecan Genesis 200 workstations. Various sample and substrate incubation times, batch size of one vs. two plates, and using different instruments (plate washer, plate reader, plate shaker, and incubator) were also tested during validation accuracy and precision evaluation. The results demonstrated characteristics of a robust assay with acceptable percent total error of $\leq 26.7\%$ for the LLOQ and $\leq 19.1\%$ for the other VSs in serum as shown in Fig. 2.

Method ruggedness was tested by between-lab performance of conformance samples. Thirty conformance samples plus two SCs were analyzed in-house and in a CRO lab. The results between methods and between laboratories are comparable. The geometric mean ratio of in-house/CRO was 1.16 and the lower and upper 90% confidence intervals were 1.13 and 1.19, respectively. There was an overall systemic positive bias of the in-house lab over the CRO. The magnitude of the bias was well within the method acceptance criteria.

The method ruggedness was also tested by a comparison of methods A and B with 30 conformance samples and two SCs assayed in-house using methods A and B. The geometric mean ratio of method B/method A was 1.10 and the lower and upper 90% confidence intervals were 1.06 and 1.14, respectively. The results demonstrate that assays by both methods are very similar on spiked and endogenous samples.

In-Study Assay Performance and Application

In-Study Variability Due to Reference or Kit Lot Changes

During in-study validation, two lots of CTx reference material and seven lots of CTx kits were used over the time span of 3 years to support more than 12 clinical studies. The individual studies were typically several years long in order to collect meaningful PD biomarker and physiological data such as bone mineral density and fracture. Samples from multiple sites were received and analyzed in batches at different time intervals; therefore, it was important to use the SC to monitor assay performance with the samples over time. The same set of high and low SC were assayed along with the study samples in each run for these 12 clinical studies. The SC charts in Fig. 5 show the high and low SC concentrations of more than 1,400 analytical runs by five analysts over 3 years. The mean value and lower and upper range limits shown in the figures were established during method validation. The reference ranges were determined by the mean ± 2 standard deviations from 27 validation runs. The SC data demonstrate the consistency of assay performance with respect to time, different operators, different instruments, and various reagent lots. Two reference material lots were used over the time span; this had no effect, since the mean percent difference of the one lot to the other was 6.3% and 3.2% for the low and high SC, respectively. There were also no significant changes due to kit lot changes; the biggest percent differences between the maximum mean over the minimum were 18.4% and 13.0% for low and high SC, respectively.

Clinical Applications

CTx measurements as a PD endpoint were generated from a clinical trial in osteopenic patients with a broad dose

range of denosumab and a single dose level of alendronate (a marketed bisphosphonate, standard of care) as the active control. Figure 6a shows the percent change from baseline in CTx following every six monthly subcutaneous doses of placebo, 14 or 60 mg denosumab, or 70 mg weekly administered oral alendronate. Figure 6b shows the denosumab serum concentrations for the doses examined. The suppression of CTx following denosumab administration was rapid (within days) and dramatic (approximately -90%). The reduction in suppression towards the end of the dosing interval and repeated suppression upon subsequent dosing suggests an apparent relationship between CTx and denosumab mechanism of inhibiting RANKL. Specifically, the 14-mg profile demonstrates a dynamic relationship between denosumab concentrations and the level of CTx.

CTx method was used to support PD comparability for bioequivalence studies supporting denosumab clinical development process improvements and scale-ups. Four comparability studies were conducted and the samples analyzed for

Fig. 6. **a** Pharmacodynamic profiles of patients dosed with denosumab in a phase 2 study. Doses were placebo (filled triangles), alendronate (open triangles), and denosumab Q6months, 14 and 60 mg (14 mg: filled circles, 60 mg: open circles). The Y-axis is mean CTx % change, with error bars of $\pm$ standard error of the mean; X-axis is time in months. **b** Pharmacokinetic profiles of patients dosed with denosumab in a phase 2 study. Doses were denosumab Q6months, 14 and 60 mg (14 mg: open triangles, 60 mg: filled triangles). The Y-axis is log mean denosumab serum concentration, with error bars of $\pm$ standard deviation; X-axis is time in months

PK and PD evaluation of their bioequivalence in denosumab exposure and comparability in PD response. The assay accuracy ranged from -0.6% to 9.0% for QC and from -1.0% to 7.9% for SC. The assay precision ranged from 2.7% to 9.4% for QC and from 4.9% to 10.6% for SC. The highest total error was 16.5% for low QC and 18.5% for low SC. The data show that the assay had sufficient accuracy and precision to support the stringent criteria of comparability studies when run under bioequivalency criteria and would enable the use of the CTx assay for PD comparisons.

DISCUSSION

Method Modification

The serum CrossLaps® enzyme-linked immunosorbent assay (ELISA) is a diagnostic kit intended for *in vitro* diagnostic use as an indication of human bone resorption (11). However, our utilization of CTx ELISA was to quantify a PD marker for supporting drug development program. The modifications to the CTx kit assay was needed to ensure the robust, accurate, and precise with minimal analytical and preanalytic variability. Table II summarizes the kit vs. modified method and the benefits of the modifications.

The CTx reference standard supplied by the kit was purified from concentrated urine samples, and the value (potency) could change depending on the purification batch. Because a “gold” standard was not available for CTx, an alternative was to use a bulk custom-made CTx reference material from the kit vendor to maintain consistency over the years of clinical trials.

For ligand binding assays, the standard curves are inherently nonlinear, which require more calibration points to define the curve function (12). It is recommended that a minimum of six non-zero concentrations be used when fitting a nonlinear (sigmoidal) concentration–response curve (8,13). We used the bulk stock reference material to prepare a nine-point standard curve, including two anchor points, sufficient to define the four-parameter logistic function.

For biomarker analysis, the controls reflecting authentic samples per the FDA guidance on bioanalytical method validation are lacking in commercial diagnostic kits (14). The use of the bulk stock reference material enabled us to prepare VS and QC to characterize assay performance and ensure assay control. The results of accuracy and precision experiments were used to define the sensitivity, assay range, and set acceptance criteria for QCs. For drug development support, the assay range is defined by the LLOQ to ULOQ with well-characterized parameters meeting *a priori* criteria and for which interpolated results are generated (13). For a diagnostic kit assay, the assay sensitivity is often defined using limit of detection (LOD). The LOD is the smallest amount that the method can reliably detect to determine presence or absence of an analyte. The LLOQ is the smallest amount the method can reliably measure quantitatively (15). The reliable quantification of CTx for clinical samples at the low concentration region is important because the drug is expected to decrease the CTx level. While the pre-dose sample results are of diagnostic interest, the post-dose results that are typically at low concentrations are of drug development interest. The drug effect is represented by percent change of post-dose sample results over that of the pre-dose. The CTx kit LOD was 0.020 ng/mL without defined accuracy

Table II. CTx Serum CrossLaps® ELISA vs. Amgen Modified Assay

Parameter	Kit CTx assay	Modified CTx assay	Advantage of modification
Standards	In liquid form, supplied in the kit	Lyophilized stock, custom supplied bulk reference material	Consistent reference standard throughout the drug development program
Number of non-zero standards	5 including a blank	9 including a blank, prepared in-house	More points to define the standard curve
Anchor point	None	2 anchor points	Better definition of the 4-parameters
Curve fit	Quadratic, no weighting	4-Parameter, $1/Y^2$ weighting	Better curve fit
Assay range	Not Defined (Standards 0.12–2.172 ng/mL, levels vary by kit lots)	0.0490–2.34 ng/mL	Consistent range with same standard levels
Sensitivity	LOD 0.020 ng/mL	LLOQ 0.0490 ng/mL	LLOQ defined by accuracy, precision and total error
Assay VS/QC	None	6VS, 3QC levels	Determine assay performance to set acceptance criteria.
Sample controls (SC) ^a	2 controls, levels vary by kit lot	High and low endogenous serum pools prepared, sufficient to monitor large no. of studies	Control in matrix to track stability and kit lot variability
Assay acceptance criteria	QC within established range	4-6-X ^b rule for each run	–
Automation application	None	Tecan Genesis 200 to prepare STD/QC and perform assay	Minimize human error

^a The endogenous pools of SC serve as additional control to monitor stability and assay performance shift due to kit lot changes. The assay QCs are used for assay acceptance

^b Out of six QCs, at least four must meet the acceptance criteria for the analytical run to be acceptable. The six QCs consist of two each at low, mid, and high concentration and at least one should be acceptable

and precision; the modified assay LLOQ of 0.0490 ng/mL was defined with accuracy of -6.6% RE, precision of 20.1% CV, and total error of 26.7% . We chose not to report data that were below the LLOQ for clinical samples.

Accuracy and Precision Evaluation

Method precision and accuracy are performance characteristics that describe the magnitude of random errors (variance) and systematic error (mean bias) associated with repeated measurements of the same homogenous (spiked) sample under specified conditions (15). Usually, for bioanalytical method validation, a well-defined reference standard material is spiked into the same sample matrix devoid of the analyte (blank), and the nominal value of the spiked concentration is used for the calculation of absolute accuracy and precision. For biomarker assays, endogenous analytes often exist in substantial amounts in matrix, such as the case of CTx in serum. It is challenging to find serum with blank levels of CTx for spiking; therefore, the calibration standards and QCs are prepared in buffer (standard diluent of the kit) instead of serum. The performance of the standard and QCs in buffer may not be the same as those in serum. It is necessary to assess the potential difference between the buffer matrix and serum matrix during the accuracy and precision experiments in method validation.

The accuracy and precision data were assessed using a validated accuracy and precision evaluation tool as described in the paper by DeSilva *et al.* (8). The matrix comparison was performed during the accuracy and precision evaluation for this method. The %RE, inter-assay %CV, and TE were assessed separately for the buffer matrix and serum matrix. The data in Fig. 2 show that the VS in serum had higher TE especially in the concentration levels at and below the low QC. The TE at the LLOQ 2 level was 17.0% vs. 26.7% for buffer and serum, respectively. For a biomarker assay, the LLOQ 2 at 0.0490 ng/mL was acceptable based on the 20.1% CV and -6.6% RE of VS in serum.

Use of Serum Sample Control

When it is difficult to find “blank” control matrix for standards or QC preparation for a biomarker assay, standards and QCs can be prepared in assay buffer if no matrix effect is demonstrated for the method (13). In addition, SCs prepared from pooling human sera were used to represent the endogenous analyte in serum samples. Figure 5 shows the SC data from in-study runs over a time span of 35 months. It is interesting to note that there were shifts in the mean values caused by changes in some kit lots. It is desirable to use the same reference lot and kit lot at least within one clinical study whenever possible; however, due to the long clinical trials of denosumab, this was not possible. Thus, the use of SC served as an important function to detect kit lot variation. In summary, the SC charts show that the assay is robust and the variability over time contributed by various analysts and runs, changes in kit lot, or reference material is well within the reference range.

Selectivity and Parallelism

Selectivity, commonly referred to as “matrix effect,” is analytical interference caused by nonspecific endogenous soluble binding factors presented in the sample matrix (16). Assay selectivity is the ability of the assay to discriminate the analyte unequivocally in the presence of components that may be expected to be present in the sample and alter assay results. Lack of selectivity can result in binding inhibition or enhancement. The results can appear as an over- or underestimation of the analyte concentration (17). Matrix components that cause interference may be dependent on the health status and populations as well as sample collection. Immunoassays are more prone to matrix effect since there is no extraction process. Assessing matrix effect is even more challenging for biomarker immunoassays because the variable presence of many binding proteins and their heterogeneity associated with the biomarkers are dependent on health status and sample collection conditions. More serum lots should be used to assess the matrix effect for advanced biomarker method validation than that of exploratory (18). For the CTx method, we used 40 lots of normal human serum for the spike-recovery experiment to evaluate the nonspecific interference. To assess the potential unknown binding factors in patient populations, we also tested at least ten lots each for the five targeted patient populations. No significant differences in spike recovery were observed in normal serum vs. that of patients. The results from patient samples tend to be more variable than that of the healthy. The majority recovered within 20% of the nominal value with the exception of rheumatoid arthritis patients tending to have a negative bias. However, the slight matrix effect indicated by results from the patient populations is much less than the dramatic change caused by denosumab to be a cause of concern.

The CTx standard supplied by the kit was a high-performance liquid chromatography purified material from urine; its components may not be the same as those in the individual serum samples. For example, there may be larger CTx peptide fragments in serum than those in urine. There may also be differences in the CTx fragments among individuals and within individual at difference times. In addition, because the standards were prepared in assay buffer, which was not the same matrix as the study samples, it is necessary to investigate whether there is similar immunoreactivity and proportionality between the standard in buffer and the heterogeneous CTx in serum by conducting parallelism test. Figure 4 shows that the calculated concentrations are similar for each test sample independent of the dilution factor. Therefore, the results indicated that the standards in buffer offer a valid proportional scale for the relative quantification of the endogenous CTx peptide fragments in serum.

The selectivity and parallelism experiments confirm the validity of standard prepared with the reference material in buffer for the relative quantification of CTx. Due to the heterogeneous nature of the reference material and the unknown CTx in the sample, the unit should be expressed as nanograms per milliliter of the reference material. The early publications using picomoles per liter is not appropriate because the molecular weight is unknown (5,19).

Stability

Protein and peptide biomarkers are susceptible to temperature-sensitive proteolysis. Sample storage stability should be performed under the similar conditions for the study sample measurement. The experiment should be conducted in unaltered representative matrix (20). Stability of protein biomarkers is also required to show that the assay is not compromised by pre-analytical factors, such as the variation in specimen collection and handling as well as the biological variation of the subjects to be tested. Stability information are gathered from method validation on short-term storage and process stability and extended through in-study validation for long-term storage stability (21).

As the matrix contains endogenous level, the biomarker stability samples can be prepared by pooling the matrices from the healthy and diseased populations. Special care should be given when preparing the stability sample pools. This includes using the same sample collection method, with the same anticoagulant, and storing the samples at the same storage condition, etc. We noticed that CTx levels were different between sera collected using bags vs. tubes. Therefore, the same procedure of tube collection as the study samples was requested to the serum supplier for sera screening and SC pooling. The large SC pools were aliquoted and stored at the same conditions as the study samples. An aliquot of high and low SC was analyzed in each assay with the study samples to generate the data for the trend analysis that allows the detection of out-of-trend shift due to other factors (such as a bad kit lot or other cause). We found the use of SC data from the in-study runs to be a very useful approach for CTx stability trend analysis.

Method Robustness and Ruggedness and Transfer to CRO

The analytical method intended to support years of clinical studies must be rugged and robust to ensure reproducibility. The use of automation helped to minimize human errors in standards and QC preparations and assay processing errors to improve assay repeatability within lab and reproducibility between labs. Tecan Genesis workstations 200 equipped with Gemini v3.5 software was used for standard, QC/VS preparation, and performing the assay. Transferring the automated method to the CRO was also easier than that of a manual method. The automated method was found to perform similarly at the CRO with the same robustness as in-house.

Clinical Applications

The pharmacodynamic profiles of CTx in Fig. 6a can be evaluated in comparison to the pharmacokinetic profiles in Fig. 6b. These figures demonstrate the dose-dependent effect of denosumab on bone resorption and the close relationship between drug exposure and response. Where the placebo dose group is reasonably flat and without drift over the 1-year period presented, the CTx levels in the denosumab dose groups suggest an apparent dynamic relationship with the serum levels of denosumab; CTx levels dropped precipitously upon administration and remained suppressed, while denosumab levels are maintained. Reversibility of effect of

denosumab was also demonstrated by the observation that CTx returned towards pre-dose levels as denosumab concentrations in the 14-mg dose group declined to insignificant levels. These observations are consistent with the known mechanism of action of denosumab, indicating that the drug effect is reversible, a potential advantage of denosumab.

The level of precision in the CTx assay was also important to the decision to use this biomarker assay in PD comparability evaluations to support comparability studies. These studies can have profound impact on a development program and, as such, require careful consideration when making formulation comparisons. While PK comparisons have well-defined criteria provided in regulatory guidance documents, PD comparability is still being defined for biologics (22). This makes PD comparison criteria somewhat more difficult to define in terms of statistics and study execution. Elimination of analytical noise that may compound the inherent intra- and inter-subject variability is therefore valuable in clarifying comparability assessments. In this instance, due to limitations imposed by the long duration of exposure following single-dose denosumab administration, instead of a crossover design, a parallel clinical design was used. While the biological variability was anticipated to be higher in the parallel design than in the crossover, the PD data were sufficient to establish comparability based on PD in four separate studies (serum CTx as an efficacy biomarker for development of anti-resorptive therapeutics for bone disorders; JW Lee *et al.*, unpublished data).

CONCLUSION

We modified and validated a commercial diagnostic kit of serum CTx with a fit-for-purpose approach to meet the need of advanced applications of a pharmacodynamic biomarker in drug development support. The modifications provided consistent performance in standard curve and controls for assay characterization and monitoring. Thorough tests were performed on selectivity of patient sera, sensitivity on buffer and serum, specificity against related peptides, and parallelism. The SC from authentic sample pools was used for extensive stability tests and detection of potential differences due to changes in reference or kit lot or analytical sites. The in-study results confirmed that the method was reliable and robust to enable the successful use of CTx as a PD biomarker to provide early drug responses in support of denosumab development.

ACKNOWLEDGMENT

We would like to thank Dr. Han Gunn, Lennie Uy, Joe Miller, Beth Johnson, Dr. Theingi Thway, Dr. Edward Lee, and Dr. Matthew Austin for their contributions to this work.

REFERENCES

1. Fink E, Cormier C, Steinmetz P, Kindermans C, Le Bouc Y, Souberbielle JC. Differences in the capacity of several biochemical bone markers to assess high bone turnover in early menopause and response to alendronate therapy. *Osteoporos Int.* 2000;11:295.

2. Delmas PD. Markers of bone turnover for monitoring treatment of osteoporosis with antiresorptive drugs. *Osteoporos Int*. 2000;11(Suppl 6):S66.
3. Cremers S, Garnero P. Biochemical markers of bone turnover in the clinical development of drugs for osteoporosis and metastatic bone disease: potential uses and pitfalls. *Drugs*. 2006;66:2031.
4. Bekker PJ, Holloway DL, Rasmussen AS, Murphy R, Martin SW, Leese PT, *et al*. A single-dose placebo-controlled study of AMG 162, a fully human monoclonal antibody to RANKL, in postmenopausal women. *J Bone Miner Res*. 2004;19:1059.
5. Christgau S, Bitsch-Jensen O, Hanover Bjarnason N, Gamwell Henriksen E, Qvist P, Alexandersen P, *et al*. Serum CrossLaps for monitoring the response in individuals undergoing antiresorptive therapy. *Bone*. 2000;26(5):505–11.
6. Rosenquist C, Fledelius C, Christgau S, Pedersen BJ, Bonde M, Qvist P, *et al*. Serum CrossLaps one step ELISA. First application of monoclonal antibodies for measurement in serum of bone-related degradation products from C-terminal telopeptides of type I collagen. *Clin Chem*. 1998;44:2281.
7. Lee JW, Devanarayan V, Barrett YC, Allinson J, Fountain S, Keller S, *et al*. Fit-for-purpose method development and validation for successful biomarker measurement. *Pharm Res*. 2006;23:312.
8. DeSilva B, Smith W, Weiner R, Kelley M, Smolec J, Lee B, *et al*. Recommendations for the bioanalytical method validation of ligand-binding assays to support pharmacokinetic assessments of macromolecules. *Pharm Res*. 2003;20:1885–900.
9. Lee JW, O'Brien P, Pan P, Xu R. Development and validation of ligand binding assays for biomarkers. In: Khan M, Findlay JWA, editors. *Ligand-binding assays: development, validation and implementation in the drug development arena*. NY, NY: Wiley; 2009 (in press).
10. NCCLS. Application of biochemical markers of bone turnover in the assessment and monitoring of bone disease; approved guideline. NCCLS document C48-A; 2004. ISBN 1-56238-539-9.
11. Kit insert. <http://www.idsplc.com/en-us/products/product.php?id=8521>. Accessed 11/9/08.
12. Shah VP. The history of bioanalytical method validation and regulations: evolution of a guidance document on bioanalytical method validation. *J AAPS*. 2007;9(1):E43–7.
13. Findlay JWA, Smith WC, Lee JW, Nordblom GD, Das I, DeSilva BS, *et al*. Validation of immunoassays for bioanalysis: a pharmaceutical industry perspective. *J Pharm Biomed Anal*. 2000;21:1249–73.
14. FDA. Guidance for industry on bioanalytical method validation: availability. *Fed Regist*. 2001;66:28526–7.
15. NCCLS. Protocols for determination of limits of detection and limits of quantitation; approved guideline. NCCLS document EP17-A; 2004. ISBN 1-56238-551-8.
16. Colburn W, Lee J. Biomarkers, validation and pharmacokinetic–pharmacodynamic modelling. *Clin Pharmacokinet*. 2003;42(12):997–1022.
17. Lee J, Ma H. Specificity and selectivity evaluations of ligand binding assay of protein therapeutics against concomitant drugs and related endogenous proteins. *J AAPS*. 2007;9(2):E164–70.
18. Lee JW, Hall M. Method validation of protein biomarkers in support of drug development or clinical diagnosis/prognosis. *J Chromatogr B*. 2009;877:1259–71 (special issue “Quantitative analysis of biomarkers by LC-MS/MS”).
19. Rosenquist C, Eledelius C, Christgau S, Pedersen BJ, Bonde M, Qvist P, *et al*. Serum CrossLaps one step ELISA. First application of monoclonal antibodies for measurement in serum of bone-related degradation products from C-terminal telopeptides of type I collagen. *Clin Chem*. 1998;44(11):2281–9.
20. Viswanathan CT, Bansal S, Booth B, DeStefano AJ, Rose MJ, Sailstad J, *et al*. Workshop/conference report—quantitative bioanalytical methods validation and implementation: best practices for chromatographic and ligand binding assays. *J AAPS*. 2007;9(1):E30–42.
21. Lee J, Wu Y, Wang J. Fit for purpose method validation and assays for biomarker characterization to support drug development. In: Bleravins M, Rhbari R, Jurima-Romet M, Carini C, editors. *Biomarkers in drug development: a handbook of practice, application and strategy*. New York: Wiley; 2009 (in press).
22. Guidance for Industry Statistical Approaches to Establishing Bioequivalence. U.S. Department of Health and Human Services Food and Drug Administration Center for Drug Evaluation and Research (CDER), January 2001.