Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Journal of Pharmaceutical and Biomedical Analysis 42 (2006) 261–271
JOURNAL OF PHARMACEUTICAL AND BIOMEDICAL ANALYSIS
www.elsevier.com/locate/jpba

Quantification of recombinant human parathyroid hormone (rhPTH(1-84)) in human plasma by immunoassay: Commercial kit evaluation and validation to support pharmacokinetic studies

Richard L. Sukovaty [a], Jean W. Lee [a,*], John Fox [b], Karen Toney [b], Damon I. Papac [b,1], Thomas A. Grover [b], David S. Wells [b]

[a] MDS Pharma Services, Lincoln, NE, United States
[b] NPS Pharmaceuticals, Salt Lake City, UT, United States

Received 15 February 2006; received in revised form 4 April 2006; accepted 6 April 2006
Available online 9 June 2006

Abstract

Immunoassays utilizing commercial kits designed for diagnostic use can be adapted and validated to meet Good Laboratory Practice (GLP) requirements to support pharmacokinetic (PK) studies. We illustrate in this paper a systematic approach for commercial kit evaluation and GLP-compliant method validation to establish selectivity, sensitivity, linearity, accuracy, precision and stability. Immunoassay kits for human parathyroid hormone (hPTH) quantification from three different vendors were assessed in a side-by-side comparison for their suitability for the PK analysis of recombinant humanPTH (rhPTH) in EDTA plasma. Two immunoradiometric (IRMA) assay kits and one immunoluminometric assay (ILMA) kit were evaluated. Since PTH is present as an endogenous component of human plasma, QC preparation in the biological matrix was handled differently than for a xenobiotic drug compound. The endogenous concentration of PTH was determined in plasma samples from 32 individual lots using the three kits. The lots with the lowest endogenous concentrations of PTH were selected, pooled to form the low QC and spiked with rhPTH to prepare the mid and high QCs. Four evaluation batches were run with each of the three commercial kits to evaluate reference standard linearity, and QC accuracy and precision. Selectivity against PTH peptide fragments PTH(7-84) and PTH(3-84) were assessed by cross-reactivity and accurate spike-recovery to the QC samples at two concentrations. One of the kits was chosen for full method validation because it had the lowest cross-reactivity against hPTH fragments (3-84) and (7-84), a wider dynamic range and the least total error. The accuracy and precision from six validation batches of the QCs were ≤ 8.1% C.V. and ≤ 7.9% R.E., respectively. Method robustness was shown by acceptable QC performance using a different batch of capture antibodies, through 12 batches of iodinated detection antibodies, and by the use of four analysts over 2 years of patient sample analysis from clinical trials.
© 2006 Elsevier B.V. All rights reserved.

Keywords: Human parathyroid hormone; PTH; Immunoassays; Commercial kit; GLP bioanalysis; Recombinant peptide drug

1. Introduction

Human parathyroid hormone (PTH) is a polypeptide with 84 amino acid residues that is the major regulator of systemic calcium homeostasis [1,2]. In the kidney, PTH acts to increase renal tubular Ca^{2+} reabsorption and synthesis of 1,25-dihydroxyvitamin D, which increases intestinal Ca^{2+} absorption. In bone, sustained elevation of PTH increases the number and activity of osteoblasts and osteoclasts, resulting in higher bone turnover and a decrease in bone mineral density [3,4]. In contrast, single daily injections of PTH produce a preferential increase of osteoblast activity, which results in a net increase in bone mineral density. PTH can be efficiently produced through both synthetic and biosynthetic routes [5]. Recombinant human PTH (rhPTH) is being developed for the treatment of osteoporosis. Most current osteoporosis therapies focus on the prevention of bone loss. The therapeutic approach with rhPTH is stimulation of new bone growth of normal composition and structure.

The first two N-terminal amino acids of PTH are required for PTH-1 receptor activation and the anabolic bioactivity of PTH

* Corresponding author at: 1, Amgen Center Drive, Thousand Oaks, CA 91320-1799, United States. Tel.: +1 805 447 9463; fax: +1 805 999027.
E-mail address: jwlee@amgen.com (J.W. Lee).
[1] Present address: Tandem Labs, 1121 East 3900 South, Salt Lake City, UT 84124, United States.

0731-7085/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.jpba.2006.04.023

in bone. Therefore, it is possible that N-terminally truncated fragments of PTH could be produced, which are inactive at the PTH-1 receptor, but will cross-react in some immunoassays for PTH. Such a fragment, putatively identified as PTH(7-84), has been reported to circulate in the plasma at significant concentrations [6], although structural verification is lacking. Therefore, to more accurately describe the pharmacokinetic (PK) profile of bioactive PTH, a selective immunoassay is required.

Sandwich immunoradiometric assay (IRMA) kits have been developed with immobilized antibodies against the C-terminus of PTH to capture the analyte from the sample, and radiolabeled antibodies against the N-terminus for detection. The specificity for these commercial kits was intended to be for intact PTH(1-84) [7,8]. An example of this type of kit is the Allegro Intact PTH kit from Nichols Institute, which has been widely used in clinical laboratories [9]. However, it was subsequently discovered that this generation of intact PTH kits did not adequately distinguish the full length PTH(1-84) from N-terminally truncated fragments of PTH, such as PTH(3-84) and PTH(7-84) [6]. For this reason, a newer generation of PTH assay kit with higher selectivity for PTH(1-84) was developed [10–13].

This paper describes a side-by-side comparison of two of the recent PTH kits versus the commonly used Nichols PTH kit and a systematic process of selection and method validation. The method was applied in PK studies to support the drug development of rhPTH. The use of a commercial immunoassay kit eliminates the time required to produce and characterize the antibody reagents in method development. However, method validation must be conducted to show that the kit developed for diagnostic use is suitable for the intended PK study application.

2. Materials

2.1. Chemicals and reagents

The rhPTH(1-84) reference standard and peptide fragment hPTH(3-84) were from NPS Pharmaceuticals (Mississauga, Ont., Canada). Human PTH(7-84) was from Bachem (King of Prussia, PA). Sequence identity was confirmed for hPTH(3-84) and hPTH(7-84) by molecular mass determination (MS) and amino acid analysis (data not shown). The lyophilized reference standard was 108.3 μg/vial, stored at −70 °C. Phosphate buffered saline (PBS) Stock (10×) was from Quality Biologicals (Gaithersburg, MD) or EM Science (Gibbstown, NJ). Bovine serum albumin (BSA), 50% NaOH, and 25% cetyltrimethylammonium chloride solution were from Sigma–Aldrich (St. Louis, MO). Water was purified in-house by a Barnstead NANOpure® system. Human EDTA plasma was from Biochemed (Winchester, VA). HNO_3 (16 N) and 30% H_2O_2 solutions were from Fisher (Fair Lawn, NJ).

2.2. Commercial kits

Kit A (Allegro intact PTH IRMA) was from Nichols (San Juan Capistrano, CA). Kit B (whole PTH(1-84) specific IRMA) was from Scantibodies (Santee, CA). These two kits consisted of PTH(1-84) calibrator standards, zero calibrator standard, PTH(1-84) quality controls, antibody coated beads, ^{125}I-anti-N-terminal PTH tracer, and wash concentrate. The wash concentrate was a 10× solution of surfactant in phosphate buffer saline (PBS) with 0.1% azide. Kit C (bioactive intact PTH immunoluminometric assay (ILMA)) was from Immutopics (San Clemente, CA). It consisted of the same components as the above two kits except that the detector antibody system consisted of acridinium labeled bioactive PTH(1-34) antibody, acridinium antibody diluting buffer, and 30× wash concentrate. All three vendors used capture antibodies immobilized on beads coated with goat polyclonal antibodies against the C-terminal 39-84 region of PTH. The basic difference among the kits is that the detector antibodies are specific against different N-terminus PTH regions as listed in Table 1. PTH in samples was captured by antibodies on the beads. Unbound material was washed from the beads with buffer solutions. Addition of the detector antibody formed a sandwich immunocomplex of anti-PTH(39-84)–PTH(1-84)–anti-N-terminal PTH. This sandwich-type recognition, utilizing capture antibody specific for the C-terminus of PTH and detection antibody specific for the N-terminus of PTH, provides very high assay specificity. In addition, the labeled detector antibodies provide the means for sensitive detection either by radioactive counting or by chemiluminescence. The bound ^{125}I was counted on a gamma counter for the IRMA methods. For the ILMA method, a chemiluminescence reaction was triggered by the oxidation of the bound acridinium ester by acidic hydrogen peroxide to an excited state. Upon the addition of NaOH, the ester returned to the ground state with light emission at 420–430 nm, which is detected by a luminometer.

Table 1
Commercial kits and the N-terminal specific detector antibodies

	Kit vendor	Detector antibodies
Kit A	Nichols Institute Diagnostics	^{125}I-anti-human PTH(1-34)
Kit B	Scantibodies Laboratories Inc.	^{125}I-anti-human PTH(1-6)
Kit C[a]	Immutopics	Acridinium anti-human PTH(1-12)

The capture antibodies were goat polyclonal antibodies against the C-terminal of PTH(39-84) immobilized on beads for all three kits. The detector antibodies were different for each kit.

[a] Kit C was an early stage assay specifically prepared by Immutopics for evaluation. This assay kit was different from the "Human Bioactive PTH 1-84 ELISA Kit" currently available from Immutopics.

3. Methods

3.1. Reagent preparation

BSA solutions of 2% and 0.5% were prepared in PBS. The concentrated wash buffers from the kits were diluted 10- or 30-fold with NANOpure® water and stored at room temperature until the kit expiration date. Kit C labeled detection antibody was reconstituted with the antibody-diluting buffer included in the kit. The trigger solution 1 for Kit C was 0.1 M of HNO_3 and 7.5% H_2O_2, the trigger solution 2 was 0.655% NaOH and 0.125% cetyltrimethylammonium chloride in water.

3.2. Preparation of calibration standards and quality control samples

3.2.1. Standards

Standards were prepared from rhPTH(1-84) and the kit calibrators included with the kit. The lyophilized calibrators were reconstituted with NANOpure® water and stored according to the kit brochure instructions. Standards of rhPTH(1-84) were prepared as follows: rhPTH(1-84) reference standard (108.3 μg/vial) was reconstituted with 1.0 ml of NANOpure® water to prepare the primary stock, which was diluted with 0.5% BSA to a secondary stock of 1000 ng/ml, and then further diluted to a tertiary stock of 100 ng/ml. The stock solutions were stored in polypropylene containers at −70 °C for up to 2 months.

3.2.2. QC preparations

All quality control samples (QCs) were prepared using fresh in-house blood draws from human volunteers. The blood samples were kept on ice and centrifuged promptly to separate the plasma, which was aliquoted and stored at −70 °C. The QCs included in the kits were not used. Thirty-two lots of human EDTA plasma were screened and evaluated using each of the three kits. The mean endogenous level of PTH in each lot was determined against the PTH standards of each kit. The plasma lots were selected and pooled to prepare a low QC with a calculated concentration approximately three to four times the lower limit of quantification (LLOQ). The middle and high QC levels were prepared by spiking a known amount of rhPTH(1-84) in each sample. A vial of rhPTH(1-84), separate from that used for the calibrator standards, was used to prepare the middle and high QCs. The QCs were distributed in 1.0 ml aliquots into polypropylene tubes and stored at −70 °C. The low QC values were the mean endogenous PTH concentrations for the pooled plasma determined in each of the three kits. Concentrations for the mid and high-level QCs were calculated as the sum of the endogenous PTH concentration and the concentration of spiked rhPTH(1-84).

3.3. Clinical sample collection

Blood samples were collected from individuals using 10 ml lavender-top collection tubes containing K_3EDTA (15% solution) from Beckon Dickinson (Franklin Lakes, NJ). Blood samples were immediately placed in an ice-water bath. Within 30 min of collection, samples were centrifuged at 2000 × *g* at approximately 5 °C for 15 min. The plasma samples were transferred into appropriately labeled cryovials and stored at −70 °C.

3.4. Immunoassay procedures

Each pre-study validation batch run consisted of at least a single set of calibration standards in duplicate and six replicates of QCs at three levels. For the in-study analytical run, each batch consisted of four QC replicates instead of six. All solutions, reagents, standards, QCs, and samples were brought to room temperature. To 200 μl of standard, sample (in duplicate) or QCs, 100 μl of labeled PTH antibody solution (^{125}I for Kit A and B, acridinium for Kit C) was added. After mixing, one antibody-coated bead was added to each test tube using a bead dispenser from Nichols. The tubes were covered and incubated at room temperature overnight (20–24 h for Kit A, 18–24 h for Kit B, or 16–24 h for Kit C) on an Eberbach (Ann Arbor, MI) shaker set at 170 rpm. Following the overnight incubation, the samples were transferred to a wash rack (Nichols) and 2.0 ml of wash solution was added to each tube and then decanted. The wash step was repeated two times for Kit A, and three times for Kit B and Kit C. The tubes were counted for 1 min on a gamma counter. For Kit C, the tubes were transferred to a luminometer after the injectors were primed with trigger solutions 1 and 2. Each tube was read using a program designed to add both trigger solutions and read for 2 s.

3.5. Cross-reactivity with rhPTH and PTH peptide fragments

Human recombinant PTH(1-84), and PTH peptide fragments (3-84) and (7-84) were tested for cross-reactivity and analyzed with each of the three commercial kits' calibrator PTH standard curve. The concentrations of the test compounds prepared in each kit's control matrix were: 10, 15, 30, 75, 150, 375, 750 and 1500 pg/ml for rhPTH and fragment (3-84); and 100, 200, 500, 1000, 2000, 5000, 10,000 and 20,000 pg/ml for fragment (7-84). Percent cross-reactivity was calculated by dividing the value obtained for PTH concentration by the nominal concentration of the test cross-reactant × 100. The mean of all observable data points was used to express an overall% cross-reactivity.

3.6. Interference from PTH peptide fragments

Possible interferences from PTH peptide fragment (3-84) or (7-84) were evaluated for each kit. Evaluation samples were prepared with rhPTH(1-84) at three concentrations followed by the addition of either PTH(3-84) or PTH(7-84) at two concentrations, or zero control standard (no peptide). The spiking concentrations of PTH(3-84) were 75 and 750 pg/ml and that of PTH(7-84) were 500 and 5000 pg/ml to test the possible concentration effect. The evaluation samples were analyzed against each PTH kit's calibration curve. The observed data of the peptide fragments were compared to the control to obtain the %control values.

3.7. Instrumentation

The samples analyzed by Kits A and B were counted on a Cobra II gamma scintillation counter from Perkin-Elmer (Boston, MA) optimized for detection of ^{125}I. The samples analyzed by Kit C were read on an AutoLumat Plus LB 953 Luminometer from EG&G Berthold, Berthold Technologies GmbH & Co. (Bad Wildbad, Germany), which was optimized for acridinium luminescence detection. The instrument settings were: volume injector P trigger solution 1 was set at 300 μl, volume injector M trigger solution 2 at 300 μl, delay injection P to M at 7 s, delay last injection to measure at 0 s, measuring time at 2 s, background measuring time at 2 s, maxi-

Table 2
Calibration curve linearity for Kits A, C and B

Kit A

	7.00 pg/ml	15.0 pg/ml	27.5 pg/ml	55.0 pg/ml	195 pg/ml	530 pg/ml	910 pg/ml	1820 pg/ml[a]
Evaluation experiment								
Mean	6.92	15.2	25.7	58.2	198	557	851	1820
C.V.%	10.4	7.5	7.0	4.7	3.4	2.2	4.7	4.5
R.E.%	−1.1	+1.3	−6.5	+5.8	+1.5	+5.1	−6.5	0.0
N	8	8	8	6	8	8	8	5[a]

Kit C

	8.00 pg/ml	15.0 pg/ml	27.0 pg/ml	46.5 pg/ml	100 pg/ml	310 pg/ml	1090 pg/ml
Evaluation experiment							
Mean	8.51	15.1	27.1	44.7	94.4	318	1090
C.V.%	7.8	5.9	7.7	9.1	9.4	5.1	1.6
R.E.%	+6.4	+0.7	+0.4	−3.9	−5.6	+2.6	0.0
N	7	7	7	8	7	8	8

Kit B

	6.10 pg/ml	15.4 pg/ml	47.0 pg/ml	140 pg/ml	670 pg/ml	2100 pg/ml
Evaluation experiment						
Mean	6.12	17.1	43.2	134	682	2100
C.V.%	14.6	4.8	3.4	4.9	4.0	1.8
R.E.%	+0.3	+11.0	−8.1	−4.3	+1.8	0.0
N	8	7	7	8	8	8
Validation experiment						
Mean	6.17	17.3	43.4	134	683	2100
C.V.%	13.8	4.7	2.9	4.1	3.2	1.5
R.E.%	+1.1	+12.3	−7.7	−4.3	+1.9	+0.0
N	9	9	11	12	12	12

Standard concentrations were those of the kit calibrators in pg/ml. Evaluation experiments are from 4 runs for the three kits; Kit B was chosen for validation experiments with 6 runs. Each batch run had one set of standards in duplicate. If a back-calculated value was ≥20% of the nominal value, the data point was not used in the regression.

[a] The 1820 pg/ml calibrator on Kit A occasionally had a replicate value with response greater than the quadratic equation maximal response. In those cases, a valid value could not be calculated by the program, and a value of 0.00 was reported. These data points were not included in the table.

mum background (RLU/s) at 0, and with automatic background subtraction on.

3.8. Analytical data treatment

The responses in counts per minute (CPM) or luminescence units were acquired using an OpenVMS® on AlphaServer® Systems Oracle® database. A weighted [(1/*x*), where *x* is the analyte concentration] quadratic method of data reduction was used to calculate the concentration of PTH. The quadratic equation was:

$$\text{Response} = A + (B \times \text{Conc}) + (C \times \text{Conc}^2).$$

4. Results and discussion

4.1. Evaluation of three commercial kits

The evaluation of the commercial kits included: (a) statistical assessment of four evaluation batches on the linearity of the kit standard calibrators, (b) accuracy and precision performance of the QC validation samples regressed against the kit standards, and (c) selectivity determined by immuno cross-reactivity with peptide fragments and percent interferences caused by these fragments at their expected in vivo concentrations.

4.1.1. Linearity and sensitivity

Method validation in support of therapeutic PK studies requires a priori acceptance criteria to be set for accuracy (relative error (%R.E.)) and precision (coefficient of variance (%C.V.)). The data presented in Table 2 shows acceptable linearity with C.V. of ≤20% and R.E. of ≤20% in all the kit calibrator levels. The acceptance criteria at the LLOQ for rhPTH was set at C.V. of ≤20% and R.E. of ≤±20% according to an MDS Pharma Services internal SOP. The standard range of Kit A was 7.0–1820 pg/ml, 6.1–2100 pg/ml for Kit B, and 8.0–1090 pg/ml for Kit C. The calibrator standards were evaluated for their linearity using a quadratic curve function for curve fitting. The quadratic fit was practical to adequately provide the “goodness of fit” in this case since the number of standard concentrations (only six) was insufficient to define a four or five parameter logistic mathematical function, especially for the asymptotes. Fig. 1 shows a typical Kit B calibration curve of response versus concentration.

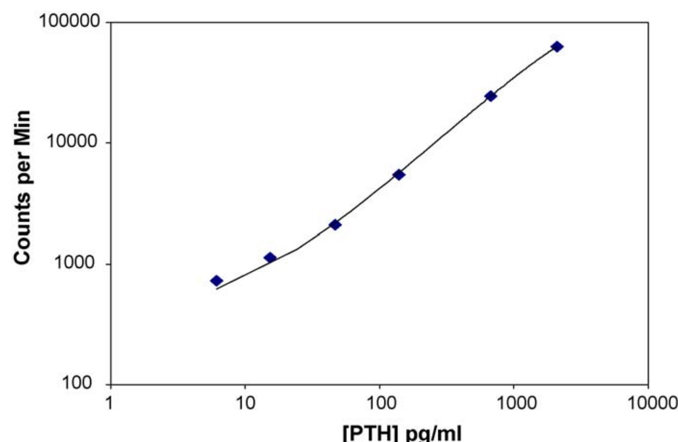

Fig. 1. A typical calibration curve plot from one validation batch run using Kit B.

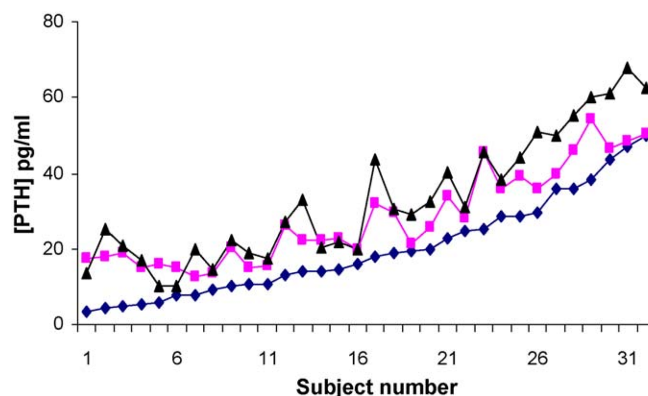

Fig. 2. Endogenous plasma levels of PTH in 32 human donor samples. Individual concentrations are arranged in ascending order based on values from Kit C. Triangles are values from Kit A, squares from Kit B and diamonds from Kit C.

4.1.2. Endogenous plasma PTH determination

The composition of the standard calibrators in the kits differed from the study samples in two ways: the matrix was stabilized human serum, instead of EDTA plasma in study samples; and the reference standard in the kits was synthetic human PTH, instead of a recombinant form. These differences were evaluated by spiking rhPTH in the zero calibrators from the kits. Similar results (data not shown) were obtained from native and recombinant forms of PTH, which indicated the kit standard calibrators could be used in the analysis of study samples. In addition, QCs were prepared in-house from freshly drawn EDTA blood samples mimicking the study samples from clinical trials. The stabilized human serum controls from the kits were discarded.

Blood samples were collected from 32 apparently healthy volunteers and individual PTH levels determined using the three kits. The assay results (mean of two replicates) of the endogenous levels, sorted in an ascending order based on the Kit C results, are shown in Fig. 2. The observed ranges in PTH concentration for the 32 donor samples were 3.21–49.8 pg/ml (Kit C), 12.5–54.3 pg/ml (Kit B), and 10.3–68.0 pg/ml (Kit A). The ranges were in agreement with the normal range of 10–65 pg/ml from 253 individuals cited in Kit A's brochure [14].

Table 3
Determination of QC accuracy and precision among four batches

Kit A			
Expected value	27.7 pg/ml (endogenous)	178 pg/ml	1507 pg/ml
Mean	24.1	214	1560
C.V.%	11.6	9.1	11.7
R.E.%	−13.0	+20.2	+3.3
Total error (%)	24.6	29.3	15.0
N	24	24	23
Kit C			
Expected value	20.1 pg/ml (endogenous)	170 pg/ml	1500 pg/ml[a]
Mean	16.3	191	1810
C.V.%	13.0	8.4	9.4
R.E.%	−18.9	+12.4	+20.7
Total error (%)	31.9	20.8	30.1
N	24	24	24
Kit B			
Expected value	23.4 pg/ml (endogenous)	173 pg/ml	1503 pg/ml
Mean	24.4	184	1460
C.V.%	9.1	4.8	3.9
R.E.%	+4.3	+6.4	−2.7
Total error (%)	13.4	11.2	6.6
N	24	24	24

In-house QC validation samples at three concentrations (in pg/ml) were run with the standard calibrators from Kits A, C and B. The expected value of the low QC was determined in an early batch run.

[a] QC1500 assayed at a DF = 2.

Individual lots with PTH concentrations approximately three to four times the LLOQ were pooled together to form the low QC. Mid and high QCs were prepared by spiking a part of the low QC pool with known amounts of rhPTH(1-84). The mean values of the low QC were determined for each kit from an initial evaluation run. They were 20.1, 23.4 and 27.7 pg/ml for Kits C, B and A, respectively. These observed basal values plus the spiked concentrations of rhPTH(1-84) provided the calculated values of the mid and high QCs. They were 170, 173 and 178 pg/ml for the mid QCs, and 1500, 1503 and 1507 pg/ml for the high QCs of Kits C, B and A, respectively. The observed mean values for the low, mid, and high QCs in each assay are listed in Table 3.

4.1.3. Precision, accuracy and total error

Because the analyte (PTH) is an endogenous molecule, it is not possible to use a nominal, theoretical value for the low QC. The determined concentration of the low QC was used for each kit as the target value to assess the accuracy performance of individual runs. However, this accuracy assessment on the low QC was not absolute. The values of the mid and high QCs were calculated from the basal level (low QC) plus the known amount of spiked rhPTH(1-84). Since the basal level was less than 16% of the spiked amount, the concentrations of the mid and high QCs were close to the nominal concentrations. The addition of rhPTH(1-84) in the preparation of the mid and high QCs also served to confirm the immuno-comparability (and subsequently stability) of the recombinant hPTH versus the endogenous form. Table 3 shows the QC data from four evaluation batches. Accuracy was marginally unacceptable for Kits A and C. Both kits had greater than 10% negative bias at the low QC and greater than 20% positive bias at either the mid or high QC. Kit B had the best accuracy at all three QC levels, with ≤+6.4% R.E. Imprecision was lowest with Kit B with 9.1, 4.8 and 3.9% C.V. for the low, mid and high QCs, respectively. The bioanalytical 4-6-20 rule (assay acceptance criterion) for each analytical batch requires that ≥4 out of the six QCs (two QCs at each of the low, mid and high levels) to be within 20% of the nominal values. Since QC assay performance is a culmination of variance and bias, total error would be a good overall assessment of QC acceptability [15]. When the total errors were examined for each kit method as shown in Table 3, the QC performance of Kits A and C with total error of 15.0–31.9% would not meet the acceptance criteria of ≤±20% for the target QC values. On the other hand, the Kit B results of 6.6–13.4% total error satisfied the 4-6-20 rule for bioanalytical run acceptance.

4.1.4. Selectivity

4.1.4.1. Cross-reactivity against peptide fragments. Human PTH peptide fragments (3-84) and (7-84) were tested for cross-reactivity and analyzed with each of the kits' calibrator PTH standard curve. Percent cross-reactivity was calculated from the mean of the assay values obtained at the eight different test peptide concentrations. If the observed concentration was below the quantifiable limit, the cross-reactivity was reported to be <0.1%. The immuno-reactivity of rhPTH(1-84) was similar to that of each kit's PTH(1-84) calibration standard. Kit A showed very high cross-reactivity with both peptide fragments, with 139% for hPTH(3-84) and 69% for hPTH(7-84); it would not be specific

Table 4
Interference of PTH(1-84) quantification by PTH(3-84) and PTH(7-84) peptide fragments

	Addition of peptide fragments				
	None (control)	(3-84) Fragment		(7-84) Fragment	
		75 pg/ml	750 pg/ml	500 pg/ml	5000 pg/ml
rhPTH concentration at 15 pg/ml					
Kit A (pg/ml)	12.4	423	1090	347	AQL
% Control	100	3411	8790	2798	–
Kit C (pg/ml)	11.7	42.7	471	17.7	48.7
% Control	100	365	4026	151	275
Kit B (pg/ml)	16.5	13.1	15.1	13.3	15.7
% Control	100	79	92	81	95
rhPTH concentration at 75 pg/ml					
Kit A (pg/ml)	96.9	195	1110	443	AQL
% Control	100	201	1146	457	–
Kit C (pg/ml)	62.5	104	558	73.4	102
% Control	100	166	893	117	163
Kit B (pg/ml)	75.9	70.4	71.1	73.3	76.2
% Control	100	93	94	97	100
rhPTH concentration at 750 pg/ml					
Kit A (pg/ml)	*799	*882	**1860	1110	AQL
% Control	100	110	233	139	–
Kit C (pg/ml)	*792	*916	1270	784	807
% Control	100	116	160	99	102
Kit B (pg/ml)	703	702	701	720	722
% Control	100	100	100	102	103

Each value was the mean of $N = 3$ determinations, except $N = 2$ for *, and $N = 1$ for the ** marked samples. AQL = Above Quantifiable Limits.

enough for PK bioanalytical work. These results are consistent with those reported in the literature [16–19]. Kit C had low a cross-reactivity of 1.3% with hPTH(7-84). However, the substantial cross-reactivity of 48% with hPTH(3-84) indicated a lack of absolute specificity for hPTH(1-84). Kit B had the lowest cross-reactivity from both peptide fragments, which indicated it was the most specific for hPTH(1-84) quantification.

An important element in these tests for cross-reactivity was to demonstrate equivalent immuno-reactivity of kit PTH(1-84) calibrator standards with the rhPTH(1-84) reference standard. Preparing QC samples with the rhPTH(1-84) reference standard would be an additional assurance. In-study QC data could be used to monitor and detect possible shifts in immuno-reactivity among different product batches of rhPTH(1-84) and kit calibrator lots during the course of drug development.

4.1.4.2. Interference by PTH fragments. C-terminal peptide fragments of PTH(1-84) can be present in plasma at higher concentrations than whole PTH(1-84), dependent on the health status of the subjects and serum calcium levels [20]. In order to assess the effects of C-terminal PTH peptide fragments on the quantification of whole PTH(1-84), an interference test was carried out on the three kits. Two different concentrations of PTH(3-84) and PTH(7-84) were added to samples of PTH(1-84) at three different concentrations to test for interference. The results are shown in Table 4. Kit A results indicated that the presence of either fragment affected the quantification of the PTH(1-84) due to cross-reactivity with the kit's detection antibody. The addition of PTH(3-84) resulted in greater interference than that of PTH(7-84). Interference was the cumulative effect of the immuno-cross-reactivity and the concentration of the interfering peptide in the samples. Kit C results indicated that the presence of PTH(3-84) affected quantification at all the concentrations of PTH(1-84) tested. The presence of PTH(7-84) affected the quantification of PTH(1-84) at lower concentrations (15 and 75 pg/ml), but not at a higher concentration (750 pg/ml). Kit B results indicated no observable interference by PTH(3-84) or PTH(7-84) at all the concentrations tested. These results supported the cross-reactivity results and confirmed that Kit B was the most specific for PTH(1-84) quantification.

4.1.5. Conclusion of kit evaluation

Among the three kits tested, Kit B produced results with the lowest cross-reactivity against the test C-terminal PTH fragments. QC samples assayed with Kit B were more accurate and reproducible than those assayed with the other two kits. Kit B also had the widest dynamic range of the tested kits. For these reasons, Kit B was selected for full method validation to support sample analysis from clinical trials of rhPTH.

4.2. Full validation of Kit B

The evaluation data from four batches had established sensitivity and specificity for Kit B. Two additional validation batches were evaluated to provide sufficient data points for statistical analysis on the assay performance parameters of precision, accuracy, and linearity. A total of six validation batches were evaluated over a period of more than 3 months. Considerable time was taken to negotiate and plan with the kit supplier to assure the same lot of capture and detector antibodies would be reserved in sufficient quantities and stored to insure stability over the duration of the drug development program. Additional testing to evaluate the matrix effects of individual plasma lots, the stability of PTH(1-84) in EDTA human plasma matrix, and method robustness were also performed on Kit B. The evaluation of method robustness included assays run by multiple analysts, with different lots of radioiodinated detector antibodies and two different batches of capture antibodies. It is important that a good communication and working relationship be established with the diagnostic supplier during a drug development program. It might take some time initially to clearly communicate the different needs in PK or pharmacodynamic studies from that of diagnostic purposes. Such clarification and resultant collaboration contributes to the subsequent success in sample analysis during long clinical trials.

4.2.1. Calibration standard linearity

The acceptance criteria for the calibrator standards were that the back-calculated values of the individual data points must be within 20% of the nominal value and at least 75% of the data points were acceptable. The between-batch variance of each calibrator standard's concentration should be $\leq$20%. The between-batch C.V. and R.E. of the standard calibrators are listed in Table 2, validation experiment of Kit B. All calibrator standard concentrations met the acceptance criteria. The correlation coefficients of the calibration curves were 0.9991–0.9998, with a mean value of 0.9996.

4.2.2. QC accuracy and precision

The within- and between-batch statistical data of mean, precision and accuracy were determined from six batches (Table 5). Each batch consisted of six replicates at each of the three QC concentrations. The precision and accuracy performance data for each batch met the acceptance criteria set for the method validation. They were well within 20%, the highest imprecision was 10.0% C.V. for the low QC in batch 2, while the highest bias was 12.4% R.E. for the mid QC in batch 3. The percent total error for each batch was calculated by adding the C.V. and the absolute R.E. The data indicated that five of the six batches of the low QC had total error of close to or higher than 10%, with batch 5 being close to 20%. For the mid and high QC, only batch 3 had a high value of 15.1%, all others were less than 9%. As can be seen from Table 5 the mid and high QCs easily met acceptance criteria, while the values obtained for the low QC met the acceptance criteria, but with higher assay variability and bias than the mid and high QC. The between-batch statistics had the same mean and R.E.% as the sum total column. The between-batch C.V. percentages were less than the sum total, since the variability of the sum total includes variance from the within- and between-batches. The between-batch variance was similar in magnitude to those of the within-batch variances. Batches 5 and 6 variances were similar to those of the early runs of batches 1–4. The data indicated that the overall method tolerance was

Table 5
Validation QC performance within- and between-batch statistics

	Batch no. (within-batch $N = 6$)						Inter-batch $N = 6$	Sum total $N = 36$
	1	2	3	4	5	6		
Low QC								
Mean (pg/ml)	24.6	25.2	24.8	23	20.8	22.7	23.5	23.5
C.V.%	5.39	10.0	8.12	3.21	8.97	6.13	7.00	10.90
R.E.%	5.06	7.62	5.77	−1.92	−10.97	−2.99	0.40	0.40
Total error (%)	10.5	17.7	13.9	5.1	19.9	9.1	12.7	11.3
Mid QC								
Mean (pg/ml)	186.8	177.3	194.5	177.3	174.2	176.7	181.1	181.0
C.V.%	0.67	2.18	2.7	2.89	1.57	2.31	2.10	4.70
R.E.%	8	2.5	12.43	2.5	0.67	2.12	4.70	4.60
Total error (%)	8.7	4.7	15.1	5.4	2.2	4.4	6.8	9.3
High QC								
Mean (pg/ml)	1463	1420	1517	1423	1433	1458	1450	1450
C.V.%	0.81	1.77	2.51	3.03	0.96	1.66	1.79	3.40
R.E.%	−2.64	−5.52	0.91	−5.3	−4.64	−2.97	−3.50	−3.50
Total error (%)	3.5	7.3	3.4	8.3	5.6	4.6	5.5	6.9

For each validation batch, mean, C.V.%, R.E.% and total error (%) (the sum of C.V.% and the absolute value of R.E.%) were calculated from six replicates of QCs. The inter-batch statistics were calculated from the within-batch data of the six validation runs ($N = 6$). The sum total statistics were calculated from all data points ($N = 36$).

good and the performance was reliable and adequate to support PK studies.

4.2.3. *Selectivity—lack of matrix effect*

To demonstrate a lack of interference by matrix, 10 individual lots of control human EDTA plasma were tested for spike recovery of rhPTH. An aliquot of each plasma lot was spiked with rhPTH at a concentration of 52.5 pg/ml. The spiked and unspiked samples were assayed. The spike recovery was calculated by subtracting the endogenous value of the unspiked samples from the spiked samples. The data in Table 6 indicated all 10 lots quantitated within 20% of the theoretical spiked value. While the basal values of the 10 individual lots varied from 14.9 to 54.3 pg/ml, the variance of the spike recovery among the 10 lots was only 6%. The mean value of the spike recovery was 59.3 pg/ml, showing a positive bias of 13% (mean spike recovery 113%). The lack of matrix effect was better reflected by the tight C.V. percentage and the range of recovery from the mean was from a negative 10% at the lowest lot to a positive 7% at the highest.

4.2.4. *Stability*

PTH is a peptide that can be hydrolyzed by proteolytic enzymes in the biological matrix. The stability of PTH in EDTA plasma was tested for tolerance to room temperature exposure and white lights during sample processing, and to multiple freeze/thaw cycles possible during sample reanalysis or shipping problems. Bench top stability was performed after QC samples were thawed and allowed to sit on the lab bench under white lights at room temperature prior to assay. Freeze/thaw cycle stability was tested by comparing QC samples subjected to four freeze/thaw cycles to a control set with one cycle. Storage sta-

Table 6
Validation test of lot-to-lot matrix effect-Spike recovery test of multiple plasma lots

Lot no.	A1 unspiked calc. conc. (pg/ml)	A2 spiked calc. conc. (pg/ml)	Spike recovered	
			Calc. conc. (A2-A1) (pg/ml)	% Recovered
1	22.2	84.3	62.1	118
2	14.9	76.9	62.0	118
3	22.3	80.3	58.0	110
4	16.2	77.9	61.7	118
5	35.8	98.0	62.2	118
6	54.3	112.7	58.4	111
7	29.8	88.4	58.5	111
8	45.5	98.5	53.0	101
9	26.1	80.3	54.3	103
10	18.0	81.2	63.2	120
Mean	28.5	87.8	59.3	113
C.V.%	–	–	6.0	6.0

Fresh EDTA plasma was drawn in-house from 10 individuals. To an aliquot of each lot, 52.5 pg/ml rhPTH was added. Duplicates of the unspiked and spiked samples were analyzed with Kit B according to the procedure. Percent spike recovery was calculated by $(A2 - A1)/52.5 \times 100$.

Table 7
Stability of PTH in human EDTA plasma

	Period or conditions	% of Control		
		QC 23.4 pg/ml	QC 173 pg/ml	QC 1503 pg/ml
Benchtop	44 h at RT under white light	101	91	90
Freeze/thaw	Four cycles	99	101	101
Long-term storage	90 weeks at −70 °C	97	102	104

Control samples are: at 0 h, one cycle of freeze/thaw, or the first validation batch for the benchtop, freeze/thaw, and long-term storage, respectively.

bility was tested for QC samples stored at −70 °C for a period that covered the time span from sample collection to analysis. The data was compared to those of an initial set analyzed at the start of validation.

The results of the stability tests are listed in Table 7. For bench top stability, the results were 90–101% of the controls after 44 h at room temperature. This is in contrast to the PTH stability in human serum given by the kit supplier, which listed a value of approximately 80% remaining after 8 h at room temperature. Human serum is the common matrix used in clinical laboratories because it is a "cleaner" matrix with less aggregation and performs well in automated procedures. Since peptides may be subjected to proteolysis, EDTA plasma might have the advantage of inhibiting the serine proteases. After four freeze/thaw cycles, the QCs quantitated at 99–101% of those of one freeze/thaw cycle. The long-term storage at −70 °C showed that PTH was stable up to 90 weeks. Fig. 3 plots long term storage over a time span of 26 months for low, mid and high QC samples. The mean values from the initial validation batch (month 0) were used as the control to calculate the "% of control" values. The acceptance criteria were ±20% of the control as marked by the lines at 0% and 120% in the plot. All the stability data were acceptable. No downward or upward trend was observed. The low QC showed more variability as predicted from the validation data; the values between 6 and 20 months suggested a downward trend in stability. However, this was not supported by the mid or high QC results, and the data from the 26-month sample confirmed that the phenomenon was caused by the variability of the low QC rather than a true degradation trend. For all other stability tests, it was more desirable to carry out the test within one analytical batch to eliminate the between-batch variability. The long-term storage stability is the only test in which comparisons were usually made between assay batches. In addition to between-batch variability, the preparative bias of the stored QC from the fresh set of standard calibrators could contribute to the discrepancies.

For commercial kits, it was important to have an up-front plan to assure the consistency of reagent supplies, including the calibrators as well as the antibodies. Different lots of protein or peptide drug reference could vary slightly in moisture content, aggregation, or other physical or chemical states. It was necessary to document the information on the different lots from the certificates of analysis. Instead of relying on the data from a single batch, it is important to examine trends in the long-term stability test results during long clinical trial programs. Therefore, QC charts should be maintained for such use. In Fig. 4 are QC charts on 90 batches of analytical runs over approximately 2 years. It is interesting to note that the shift around batch 45

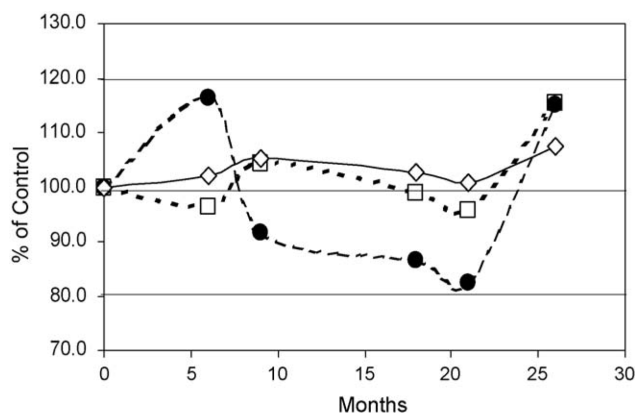

Fig. 3. Long-term storage stability data of QCs. Mean values of the QC determined at later test periods were compared to the original mean value, calculated as % of control. Symbols are circles for low QC, squares for mid QCs, and diamonds for high QCs. The upper and lower lines of 120% and 80% define the boundaries of acceptance for stability.

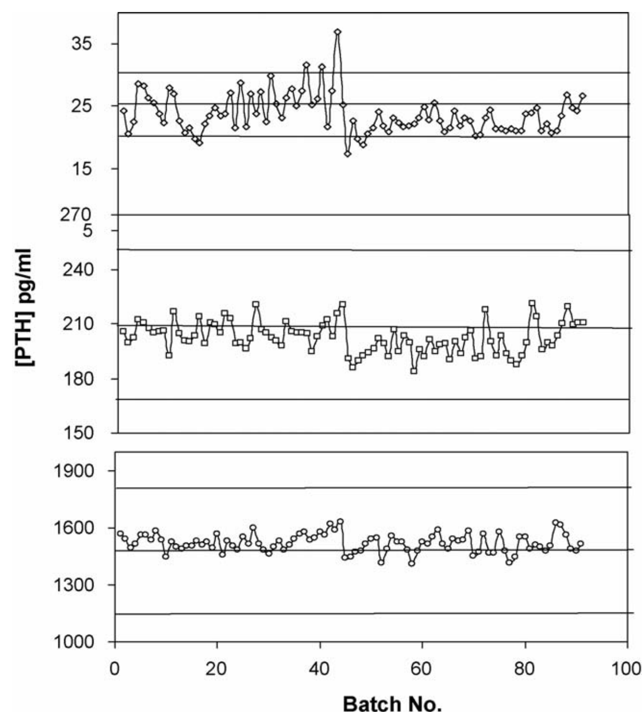

Fig. 4. In-study monitoring QC charts. Upper panel: low QC; middle panel: middle QC; lower panel: high QC. Centerlines are the validation QC target values of: 23.4, 196 and 1503 pg/ml. The upper and lower lines are the ±20% from the target QC values.

was caused by a systemic bias from a different standard kit lot. Therefore, the QC chart is important to track performance and changes due to systemic and random errors.

4.2.5. *Method robustness*

In using a commercial kit for immunoassay during the extended time span of a drug development program, the robustness of the method needs to be established. Therefore, tests were included in the pre-study validation on multiple lots of reagents, preparations, and analysts.

A second lot of capture antibody was tested to see if it would deliver the same assay performance as the first lot. Beads coated with either the validated lot of capture antibody or the second lot of capture antibody were obtained from the kit manufacturer. Two validation batches were run using the second lot of capture antibody, and the data were comparable to the validation data of the original lot of capture antibody. Results from 12 determinations (six in each batch) using the second lot of antibody had mean values of 18.9, 204 and 1520 pg/ml for the low, mid and high QCs, respectively. The corresponding C.V. values were 11.5%, 6.6%, and 6.9%. These data demonstrated that accuracy and precision using the second lot of capture antibody were acceptable and comparable to those of the first. In addition, selectivity against matrix lots was tested with eight lots of human EDTA plasma at endogenous levels and spiked at 210 pg/ml above the endogenous levels. The mean spike recovery was 108%, with a range from 99% to 120%; the C.V. was 6.1%, showing acceptable levels in eight out of eight lots tested. We concluded that the second lot of capture antibody was acceptable for use as a back up.

During in-study validation, 12 radio-iodination batches of the same lot of detector antibody were used; the performance parameters were similar among these batches. Since samples received at different times would be organized and grouped for analysis in batches, clear communication with the kit supplier on the exact timing and amounts needed for detector antibody labeling were maintained. Robustness of the assay was further illustrated by the acceptable performance of three analysts during pre-study validation and four analysts during the in-study validation for sample analysis over the time span of 121 weeks. The QC charts in Fig. 4 show the mean QCs of low, mid and high levels of 90 batches of analytical runs by the four analysts using the 12 batches of iodinated detector antibodies. The data demonstrate the consistencies of assay performance with respect to time, different operators, and various reagent lots.

5. Conclusion

The application of commercial diagnostic kits for PK studies eliminates the time required to produce and characterize antibodies and circumvents lengthy method development. However, adequate method evaluation and validation should be performed to show that the kits developed for diagnostic use would be suitable for pharmacokinetic, bioavailability and bioequivalence studies. Applications using commercial diagnostic kits under the Clinical Laboratory Improvement Amendments (CLIA) follow the Clinical and Laboratory Standards Institute (CLSI, formerly the National Committee for Clinical Laboratory Standards (NCCLS)) guidelines [21,22]. The validation of bioanalytical methods for PK studies follows guidance from the FDA under Good Laboratory Practices (GLP), which are not the same as CLIA [23–25]. Before diagnostic kits can be used for PK sample analysis, the kit's method should be adequately evaluated for suitability and validated in a GLP-compliant way. The basic components of immunoassay validation include selectivity (specificity), linearity, sensitivity, accuracy, precision, and stability [26,27]. In order to support a drug development program, which usually is conducted over a long time span and includes many sources of patient samples, uniformity and consistency of reagent supplies and method robustness should be planned up front and demonstrated in the supporting data of the clinical trial. The approach to method development and validation for macromolecular drugs using ligand-binding assays has been discussed in recent workshops and a position paper published by the AAPS Ligand Binding Assay Bioanalytical Focus Group [15]. Although our method validation was conducted several years before the publication of this paper, the process serves to illustrate a similar rational approach for kit evaluation and method validation, leading to subsequent robust and successful bioanalytical performance for a ligand-binding assay.

In order to utilize commercial immunoassay kits to support PK studies of an rhPTH drug compound, three PTH immunoassay kits were evaluated for linearity, sensitivity, selectivity, accuracy and precision in compliance with FDA guidance. Since PTH exists endogenously at various levels among different individuals, QC preparation in the biological matrix was handled differently than that of a xenobiotic drug compound. The endogenous concentration of PTH was determined in plasma samples from 32 individual lots using the three kits. The lower concentration lots were selected, pooled to form the low QC and spiked with rhPTH to prepare the mid and high QCs. Four evaluation batches were run with each of the three commercial kits to evaluate reference standard linearity, and QC accuracy and precision and selectivity against PTH peptide fragments PTH(7-84) and PTH(3-84). Kit B was chosen to support PK studies because of its best performance parameters. The method was validated in a GLP manner to establish selectivity against matrix effect, precision and accuracy, stability and method robustness. The Kit B method was successfully applied to the drug development program of rhPTH on thousands of clinical samples over more than 2 years.

Acknowledgements

We thank Jean Pedersen, Robert Klein, Randall G. Karl, and Jon Dobesh for their contribution to this work.

References

[1] E.M. Brown, in: J.P. Belezikian, Marcus, Levine (Eds.), The Parathyroids, Raven Press, New York, 1994, pp. 15–54.

[2] L.A. Fitzpatrick, J.P. Bilezikian, in: J.P. Belezikian, L.G. Raisz, G.A. Rodan (Eds.), Principles of Bone Biology, Academic Press, San Diego, 1996, pp. 339–346.

[3] J.F. Habener, J.T. Potts Jr., New Engl. J. Med. 299 (1978) 580–585, 635–644.
[4] J.S. Finkelstein, in: J.P. Belezikian, L.G. Raisz, G.A. Rodan (Eds.), Principles of Bone Biology, Academic Press, San Diego, 1996, pp. 993–1006.
[5] C.S. Tam, J.N. Heersche, T.M. Murray, J.A. Parsons, Endocrinology 110 (1982) 506–512.
[6] L.G. Raisz, C.H. Yajnik, R.S. Bockman, B.F. Bower, Ann. Intern. Med. 91 (1979) 739–740.
[7] D.L. Andress, D.B. Endres, S.M. Ott, D.J. Sherrard, Kidney Int. Suppl. 18 (1986) s87–s90.
[8] W.G. Wood, R. Butz, M. Casaretto, R. Hehrmann, H. Juppner, I. Marschner, C. Wachter, H. Zahn, R.D. Hesch, J. Clin. Chem. Biochem. 18 (1980) 789–795.
[9] P.C. Kao, N.S. Jiang, G.G. Klee, D.C. Purnell, Clin. Chem. 28 (1982) 69–74.
[10] G.V. Segre, J.T. Potts Jr., in: L.J. Degroot (Ed.), Endocrinology, W.B. Saunders, Philadelphia, 1995, pp. 1075–1093.
[11] J.T. Potts Jr., F.R. Bringhurst, T.J. Gardella, S.R. Nussbaum, G.V. Segre, H.M. Kronenberg, in: L.J. Degroot (Ed.), Endocrinology, W.B. Saunders, Philadelphia, 1996, pp. 920–965.
[12] S.R. Nussbaum, R.J. Zahradnik, J.R. Lavigne, G.L. Brennan, K. Nozawa-Ung, L.Y. Kim, H.T. Keutmann, W.A. Wang, J.T. Potts Jr., G.V. Segre, Clin. Chem. 33 (1987) 1364–1367.
[13] R.C. Brown, J.P. Aston, I. Weeks, J.S. Woodhead, J. Clin. Endocrinol. Metab. 65 (1987) 407–414.
[14] Intact PTH Directional Insert, Nichols Institute Diagnostics, kit insert, #36T-2170.
[15] B. DeSilva, W. Smith, R. Weiner, M. Kelley, J. Smolec, B. Lee, M. Khan, D. Tacey, H. Hill, A. Celniker, Pharm. Res. 20 (2003) 1885–1900.
[16] R. Lepage, L. Roy, J.H. Brossard, L. Rousseau, C. Dorais, C. Lazure, P. D'Amour, Clin. Chem. 44 (1998) 805–809.
[17] M.R. John, W.G. Goodman, P. Gao, T. Cantor, I. Salusky, H. Juppner, J. Clin. Endocrinal. Metab. 84 (1999) 4287–4290.
[18] P. Gao, S. Scheibel, P. D'Amour, M. John, S. Rai, H. Schmidt-Gayk, T. Cantor, J. Bone Miner. Res. 16 (2001) 605–614.
[19] J. Brossard, R. Lepage, H. Cardinal, L. Roy, L. Rousseau, C. Dorais, P. D'Amour, Clin. Chem. 46 (2000) 697–703.
[20] T.M. Murray, L.G. Rao, P. Divieti, F.R. Bringhurst, Endocr. Rev. 26 (2005) 78–113.
[21] Federal Register 57 FR 7002, (1992).
[22] Federal Register 42 CFR part 493, (2003).
[23] Guidance for industry on bioanalytical method validation: availability. Federal Reg., (2001).
[24] V.P. Shah, K.K. Midha, S. Dighe, I.J. McGilveray, J.P. Skelly, A. Yacobi, T. Layloff, C.T. Viswanathan, C.E. Cook, R.D. McDowall, K.A. Pittamn, S. Spector, J. Pharm. Sci. 81 (1992) 309–312.
[25] V.P. Shah, K.K. Midha, J.W.A. Findlay, H.M. Hill, J.D. Hulse, I.J. McGilveray, G. McKay, K.J. Miller, R.N. Patnaik, M.L. Powell, A. Tonelli, C.T. Viswanathan, A. Yacobi, Pharm. Res. 17 (2000) 1551–1557.
[26] J.W.A. Findlay, W.C. Smith, J.W. Lee, et al., J. Pharm. Biomed. Anal. 21 (2000) 1249–1273.
[27] K. Miller, R.R. Bowsher, A. Celniker, J. Gibbons, S. Gupta, J.W. Lee, S.J. Swanson, W.C. Smith, R.S. Weiner, Pharm. Res. 18 (2001) 1373–1383.