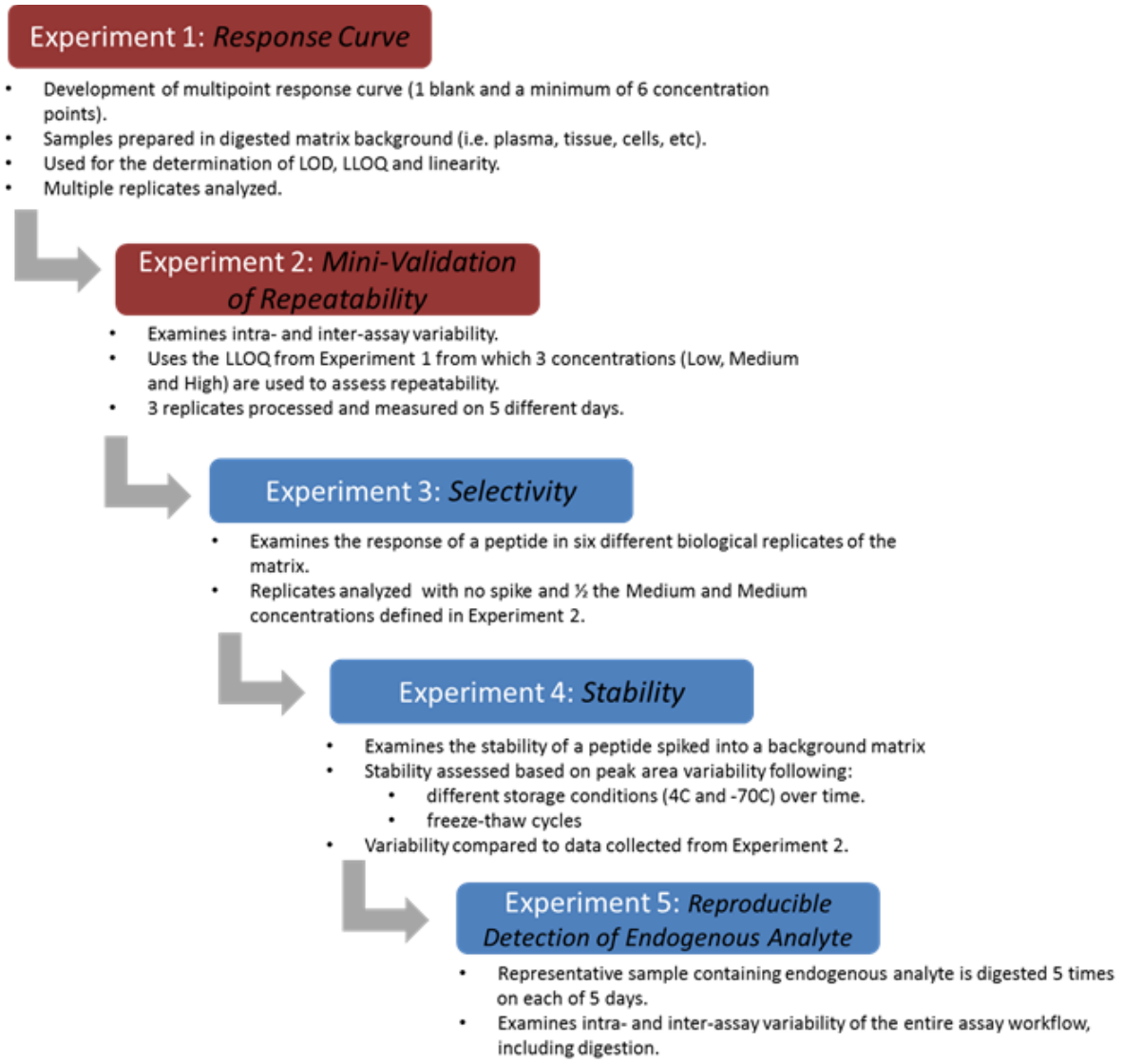
Overview of assay characterization for the CPTAC assay portal

CPTAC Assay Development Working Group

Summary

This guidance document provides a list of experiments that will help potential downstream users of assays developed in different laboratories feel more confident that investing time, money, and energy into adopting and deploying them will be beneficial. There are five experiments outlined. The first experiment involves the analysis of a response curve of peptides spiked into a representative background matrix. The purpose of this experiment is to evaluate linearity, the within-batch (intra-assay) precision of the quantitative LC-MS analysis of peptide in a complex mixture, the upper and lower limits of quantification, and to provide some data on selectivity. The second experiment provides data on reproducibility and validates the lower limit of quantification identified in the first experiment across many days, more akin to an actual experiment that is likely to span many days. The experiment



described here is designed with the hope that at least one concentration tested will have an imprecision (measured as coefficient of variation) of ≤20%. **These two experiments are the only experiments required for upload to the CPTAC assay portal.** The third experiment tests parallelism in multiple biological replicates of the matrix of interest. Parallelism is a measure of the influence of matrix components on the quantification of peptides in a complex mixture. Ion ratios are also used to establish selectivity in this and other experiments. The fourth experiment evaluates the stability of peptides after sample preparation and helps the downstream user understand if peptides can be left on the autosampler for a period of time or frozen prior to analysis. These first four experiments must be completed for an assay to be recognized with a higher level of qualification in the CPTAC program. The last experiment aims to demonstrate that endogenous proteins can be quantified in a relevant matrix. While this is not required to be uploaded to the portal or to receive a higher level of CPTAC qualification, **the ability of an assay to quantify endogenous protein in at least one matrix will be noted in the portal with a special endorsement** and is a signal to a downstream user that the standard operating procedure can be followed to obtain useful measurements in at least one matrix. A lack of endorsement should not be interpreted to mean that all potentially relevant matrices have been tested but instead, that the assay could be useful in a different matrix of interest.

Definitions

Assay

For the purposes of this document, an assay is defined as the quantification of a peptide in a complex mixture of peptides derived from the proteolytic (generally tryptic) digestion of a relevant sample matrix.

Workflow

Is defined as the sample processing entailed in performing the assay. Specific details for each assay workflow are to be provided in a standard operating procedure. Depending on the internal standard used in the assay, an assay workflow may or may not include proteolytic digestion. E.g., if a stable isotope-labeled internal standard peptide is used, then digestion is not included as part of the assay that is being validated in this guidance document; instead peptides are spiked into a background matrix of pre-digested sample. For assays that rely on “winged” peptides (i.e., peptides that have sequence specific amino acid overhangs on the amino- and carboxy-termini) or isotope-labeled protein, digestion is important. For any assay type, a detailed standard operating procedure for the digestion of the matrix of interest and all sample handling steps must be included on the portal.

Matrix

A “type” of matrix is defined as the range of sample types that will be encountered in a single typical experiment. As examples, a matrix type might include tissues (tumor and adjacent normal), serum samples (from diseased patients or healthy controls), or cell culture lysates (treated and untreated). For example, 6 “biological replicates” of a particular matrix type might include a serum sample from 6 different people (three with disease and three normals), or might include a cell line treated with ionizing

radiation (high dose), untreated cells, different batches of cells, and/or different passage of cells. For experiments #1 and #2, a single biological replicate of a single matrix type (or a pooled sample) would be acceptable. It is important to note that for some experiments, the background matrix may actually be a representative sample that has been proteolyzed according to a standard operating procedure, which is similar to the matrix that will be generated by proteolysis in an actual experiment. To this proteolyzed sample, peptides will be added to evaluate figures of merit for the assay. For experiment #5, the matrix will be undigested sample that would be used in an actual experiment to measure the endogenous concentration of protein.

Linear range

Linear response between the lower limit of quantification (LLOQ) and the upper limit of quantification (ULOQ, determined by the loss of linearity in experiment 1 or the highest concentration measured).

Reverse curve

In a matrix that has detectable amounts of endogenous protein that interferes with quantification of peptide at low concentrations, one alternative to estimate the LLOQ, ULOQ, and other parameters for the portal is to spike the matrix with isotope-labeled standard at variable concentrations.

Fragment ions / Transitions

It is generally recommended that the most intense ion is used to quantify the peptide and that the next two most abundant fragment ions are monitored to evaluate specificity of the assay. It is also possible to use the sum of multiple ions for the quantification of a peptide and three of those ions to calculate ion ratios to help evaluate specificity. In parallel reaction monitoring experiments, all fragment ions are detected and could be quantified simultaneously, with a subset of those fragment ions used to evaluate selectivity.

Standard operating procedure

A detailed description of the materials and equipment used and the steps taken to generate the data uploaded to the portal. This document should enable the reader to completely reproduce the data uploaded to the portal and should include vendor names of reagents and supplies, description of materials (*e.g.,* plastic vs. glass), purity of solvents, chromatographic columns used, *etc*.

Peak area ratios

The ratio of the chromatographic peak area of the endogenous peptide (or an unlabeled synthetic version of the peptide) to the chromatographic peak area of the internal standard peptide (generally a stable isotope-labeled synthetic version of the peptide). In the absence of endogenous or unlabeled synthetic peptide, the peak areas of labeled internal standard peptide can be reported for experiments 1-4.

Experiment 1. Response Curve

Overview: A single biological replicate or pool of matrix is generated via proteolytic cleavage and used for the development of a multipoint serial (or parallel) dilution of the peptide of interest. The goal of the response curve is to identify the lower limit of quantification (LLOQ), the limit of detection (LOD), and the linear range. The data will also be used to partially evaluate the intra-assay reproducibility and for the specificity of the assay.

Generation of the matrix: The matrix used for this experiment can also be used in Experiment 2 and Experiment 4 and can be aliquotted and frozen to be used for other experiments. The matrix should approximate as closely as possible the matrix that will be used during actual experiments. The matrix consists of a proteolytic digest (generally tryptic) of a complex mixture of proteins. For example, if the purpose of the final assay is to quantify the amount of a protein in human plasma, then the most acceptable matrix would be normal human plasma. Alternatively, one could use human serum or serum/plasma from another species. The latter might be most attractive if there is endogenous analyte in normal human plasma. As another example, if the purpose of the final assay is to quantify the amount of a protein in human breast tumors, then the most acceptable matrix would be human tumor tissue from a lumpectomy or from a xenograft. Alternatively, once could also use fibrofatty tissue from normal breast or a breast tumor cell line. The latter is less attractive in that it lacks the various cell types present in an actual tumor. Regardless of the final matrix used for this experiment, the proteolytic digestion should closely mimic the conditions used for the final assay (and should be detailed in a standard operating procedure that could be easily replicated by other laboratories as needed). Scaling up the digestion to make a bulk digest is desirably convenient, but it should be noted that the performance of the digestion at standard temperatures and incubation times could be impacted with larger digestion volumes. In some laboratories, internal standard peptide is added to the pooled matrix prior to aliquotting and addition of unlabeled peptide. This approach is acceptable for experiments 1, 3, and 4 but cannot be used in experiment 2 because this experiment attempts to simulate how the assay will actually perform in real use and the variability of the addition of internal standard will be a real source of variability in an actual experiment.

Formulation of the serial dilutions: Peptide should be diluted into an appropriate buffer (*e.g.*, phosphate buffered saline, ammonium bicarbonate, or, if peptides are not stable in buffer, buffer with carrier peptide or dilute matrix). The peptide concentrations of the serial dilutions (at least six points in addition to the blank should be made) should be designed so that when added to the matrix, the concentrations of peptide span from near the limit of detection of the peptide in buffer to 4-7 orders of magnitude higher than that concentration (the spiked volume, which is the same across all samples, should be at most 10% of the final volume of the spiked matrix). The blank should have the same volume of buffer added to it. Spiked matrix should have enough volume to inject 3-9 replicates (the blank will be injected ≥9 times and each other sample will be injected ≥3 times). It should be noted that stable isotope-labeled peptide can also be used to make the serial (or parallel) dilutions (*i.e.,* so-called reverse curve). This is particularly relevant for analytes for which endogenous concentrations of proteins are detectable in the matrix. Each sample is spiked with a constant amount of isotope-labeled peptide as

internal standard before analysis (or spiked with a constant amount of unlabeled peptide for the reverse curve).

Selection of peptides and internal standard peptides: There are four types of peptide and isotope-labeled internal standard peptides considered here for LC-MS assays. First, purified, well-characterized (*i.e.*, amino acid analysis, HPLC-UV analysis) peptides provide the most accurate measurement of the lower limit of quantification (LLOQ) and limit of detection (LOD). A separate guidance document for the characterization of these peptides is available. Second, crude peptides that have very little purification or characterization could be used as well. These peptides, often provided by the manufacturer in liquid form, can have significantly more or less peptide than reported by the manufacturer. For peptides in powder form that are weighed out in the laboratory, the empirical limits of detection and quantification are the upper limits of the actual (there might be less peptide in the spike, never more). As a result, peptides in powder form are preferred. Third, winged peptides (purified or not, well-characterized or not) that require digestion to liberate the peptide (or internal standard) of interest can be used and would be spiked prior to digestion. These winged peptides are not discussed in detail; however the experiments described could be adapted for their use. A fourth type of internal standard is an isotope-labeled protein; the experiments could be adapted as for the winged peptides. The type of standard is to be reported on the portal. It should be noted that the quality of internal standard peptides and proteins and impurities contained therein could affect the LOD and LLOQ of the assay.

LC-MS/MS analysis: For the purposes of the CPTAC program, an assay is defined as the quantification of a peptide in a complex tryptic digest. For Experiment 1, the lower limit of quantification will be calculated based on the variability of the analysis, which in this case will be the LC-MS/MS step. Samples will be injected multiple times on the same day or in the same experimental block. An effort could be made to inject from different vials, which would include vial-vial variability. In order to evaluate for carryover, samples will be injected in singlicate from low concentration to high concentration. A blank matrix sample will be injected ≥3 times prior to the first curve and injected ≥2 times after the high concentration for each of the triplicate curves. In an LC-MS experiment, at least three transitions should be monitored for each peptide.

Data analysis: CPTAC investigators have developed software tools that can be used with Skyline and Panorama (https://brendanx-uw1.gs.washington.edu/labkey/project/home/software/Skyline/begin.view) to calculate the figures of merit that will be required to upload an assay to the web portal. It is therefore required that data be analyzed in Skyline and uploaded to Panorama. The limit of detection will be determined from the blanks injected before the first curve, using the average plus 3 times the standard deviation of the blank signal. In some systems, no detectable signal will be observed in the blanks. In this case, the standard deviation of the signal detected in the lowest spiked sample should be used for this calculation instead. The lower limit of quantification will be determined in this experiment as the lowest concentration of peptide at which the imprecision of the assay [expressed as the coefficient of variation (CV), calculated as standard deviation/average] is <20%. The linearity of the assay (over any three points of the curve) will be assessed using the middle point(s). The observed concentration of the average of the three replicates of a middle concentration should be within 5% of

that predicted from best fit line passing through the other point(s). Alternatively, if there are five or more points that are being considered for linearity, one can fit a power function to the data ($y=Ax^n$, where y=peak area ratio or observed concentration, and x=the expected concentration). The data can be considered linear if $0.95<n<1.05$. In addition, when the response curve data are fit using linear regression on the linear scale, the standard error of the estimate of the slope of the regression should be less than 10% of the slope parameter. To evaluate for carryover, the peak area of the analyte peptide in the blank after the highest concentration is divided by the peak area in the high concentration sample and expressed as a percent. The three blanks, one after each high concentration sample, are averaged. In order to partially evaluate the specificity of the assay, the ratio of transitions can be determined in each sample at all concentrations (ratios of peak areas of different transitions of the same peptide). In all samples above the LLOQ, there should be no transition ratio that deviates more than 30% from the mean. The upper limit of quantification is determined as the highest concentration that lies on the linear part of the response curve.

Experiment 2. Mini-validation of repeatability

Overview: In this experiment, samples are run by LC-MS/MS at least 15 times to estimate the total variability of the LC-MS/MS experiment at a concentration near the LLOQ, which was determined in Experiment 1 to be the lowest concentration at which the intra-assay CV was <20%. The purpose of this experiment is to try to replicate the measurement of peptide in complex mixtures over multiple days, which more closely resembles an actual experiment. The total variability includes intra-assay and inter-assay variability and is traditionally validated using at least 20-25 days' worth of samples. This abridged experiment only approximates the variability of the assay in practice.

Formulation of the samples: A representative tryptic digestion matrix (could be the same as that made in Experiment 1) is freshly spiked on each day with three different concentrations of peptide (or isotope-labeled peptide if endogenous peptide is detectable in the blank matrix). These must not be the same spiked samples as those made in Experiment 1. The concentrations should approximate the working range of the assay. A general suggestion would be 1.5-3.0x LLOQ (Low), 50-100x LLOQ (Medium), and >100x LLOQ (High). The total volume of peptide in buffer that is spiked in should be ≤10% of the total volume of the spiked matrix. These samples will be injected ≥15 times, so enough sample volume should be made to facilitate this. Prior to analysis, each sample is spiked with isotope-labeled peptide as internal standard (or spiked with unlabeled peptide for the reverse curve-type experiment).

LC-MS/MS analysis: Each of these three samples is injected in at least triplicate in each of five days or five experimental blocks on different days. Different days are defined as two different calendar days with samples prepared ≥16 h apart. To avoid artificially minimizing variability, the order of the specimens should be randomized, meaning that the three concentrations should be interspersed randomly, as unknown samples would be randomized in an actual experiment. If carryover is a concern blanks should be inserted after the High and Medium samples before injecting a low.

Data analysis: The total variability will be estimated by the sum of squares. First, the intra-assay variability is calculated at each concentration as the CV of the three replicates on each of the five days.

The CVs determined for each of the five days is averaged (this is the average intra-assay CV). Second, the inter-assay variability is calculated at each concentration by determining the CV of the first injection of each concentration across the five days, then the second injection, and then the third (if run in triplicate each day, continue this process for the fourth replicate and so on if more replicates are injected each day). These three (or more) CVs are averaged (this is the average inter-assay CV). The total CV is calculated as the square root of the sum of (the average intra-assay CV)^2 and (the average inter-assay CV)^2. The "validated LLOQ" will be uploaded to the portal and is the lowest of the three concentrations at which the total estimated variability is <20%. The intra-assay variability, inter-assay variability, and the total variability will also be reported on the portal for each of the three samples. The transition ratios determined for each of the 45 injections should be within 30% of the mean, which is a second partial validation of the specificity of the assay.

Experiment 3. Selectivity

Overview: In this experiment, the response of peptide spiked into six different biological replicates of the matrix of interest is determined. If the response is similar in different biological replicates of the matrix (so called parallelism), then the assay can be considered specific for the analyte of interest.

Definition of a biological replicate: The purpose of this experiment is not to use different matrix types (*i.e.*, not urine and plasma or ovarian tumor and colon tumor). Instead, the goal is to use the same matrix derived from six different people or conditions. For example, six different biological replicates for serum would be serum samples from six different people. For cell lines, there might be six different breast cancer cell lines or the same cell line treated with six different perturbations (*e.g.*, PMA-treated, irradiated, insulin treated, cisplatin-treated, serum-starved, and hydrogen peroxide-treated). For tumors, there could be six different tumors or three different tumors and three paired adjacent normal tissue samples.

Formulation of the samples: Each biological replicate should be prepared in an identical fashion to real samples in the assay workflow. Each of the six biological replicates of matrix are prepared with no spike of analyte (*i.e.,* buffer only), spiked with the Medium concentration (50-100x LLOQ), and spiked with one-half of the Medium concentration (25-50x LLOQ) as defined in Experiment 2. Prior to analysis, each of these 18 samples is spiked with internal standard as dictated by the given assay workflow procedure. The 18 samples are prepared in at least duplicate.

LC-MS/MS analysis: The 18 samples are injected in a randomized fashion on the same day or in the same experimental block in at least duplicate. If carryover is a concern in the analysis of randomized samples (for an actual experiment, all unknowns would be randomized, this experiment should approximate a real experiment), then blanks can be inserted between samples.

Data analysis: The results from the duplicate injections for each sample are averaged and for each biological replicate are plotted on the linear scale. The slope of the line for each of the biological replicates is determined and each should be within 10% of the mean. In addition, the half-Medium concentration for each biological replicate should not be more than 10% different than that predicted

from the measurement of the Medium concentration-spiked sample. Lastly, the ion transition ratios for each of the injections of the samples in which peptide is detectable (presumably the 12 samples that have peptide spiked in) should be within 30% of the mean.

Experiment 4. Stability

Overview: The goal of this set of experiments is to determine if actual study samples can be processed and then left on the autosampler for a period of time or stored frozen before analysis by LC-MS/MS.

Formulation of samples: Representative matrix is spiked with the Low or Medium concentration as defined in Experiment 2. This sample should be divided into at least 12 aliquots.

LC-MS/MS analysis: (1) Put three aliquots of the sample (*i.e.,* in different injection vials) on the instrument and inject each vial in duplicate. At least six hours later, inject each of the three vials again. Repeat in at least 24 hours. The autosampler should be at the temperature that will be used during actual sample analyses (presumably 4-10°C). (2) Freeze the rest of the aliquots at -70°C. Remove three of them and allow them to thaw on the benchtop or on ice (this generally takes approximately 1 h, but thawing should be performed according to the standard operating procedure developed by each laboratory). Inject each of these three samples. This can be done in parallel with time zero in the refrigerated experiment above. (3) Remove three of the samples from the -70°C freezer and allow them to thaw on the benchtop at room temperature. Place them back in the freezer a second time and allow them to refreeze completely. Then, allow them to thaw again on the benchtop at room temperature or on ice (as determined by the given assay workflow). Inject each of these three samples. Again, this can be done in parallel with the refrigerated experiment above. (4) Wait four weeks and remove three aliquots from the freezer and allow to thaw on the benchtop or on ice. Inject each of these three samples.

Data analysis: When comparing the peak areas of each test to the time zero (*i.e.,* 6 h at 4°C, 24 h at 4°C, 4 weeks at -70°C, one freeze-thaw, two freeze-thaws), the variability of the six measurements should not exceed the variability determined for that concentration in Experiment 2 nor the variability measured for the six time zero analyses. In addition to peak areas, peak area ratios will also be reported on the web portal. It is possible for peak areas to be lower due to instrument performance, but peak area ratios should be stable across this experiment.

Experiment 5. Reproducible detection of endogenous analyte

Overview: In this experiment, the total variability of the entire assay workflow (including digestion and LC-MS) is estimated using a 5x5 experimental paradigm. This experiment provides the data needed for a special endorsement on the assay portal and demonstrates for at least one matrix that endogenous proteins are quantifiable. It should not be inferred that not being able to see endogenous peptide in one matrix means that the assay might not be useful in another matrix.

Formulation of samples: Five aliquots of the same sample (*e.g.,* plasma or homogenized tumor) is digested on each of five days according to the standard operating procedures associated with the given assay. Different days are defined as different calendar days, digestions separated by at least 16 h.

LC-MS/MS analysis: Each day, the five digests are injected. Replicate injections can be made if needed for improved precision. The digests should not be stored and analyzed on one day. The goal is to include the variability that will actually be observed when real samples are run over multiple days.

Data analysis: Peak area ratios for replicates are averaged. The data are analyzed as for Experiment 2.