Published in final edited form as:

Anal Chem. 2017 April 18; 89(8): 4550–4558. doi:10.1021/acs.analchem.6b05002.

Multiplex Substrate Profiling by Mass Spectrometry for Kinases as a Method for Revealing Quantitative Substrate Motifs

Nicole O. Meyer[†], Anthony J. O'Donoghue^{†,§}, Ursula Schulze-Gahmen[‡], Matthew Ravalin[†], Steven M. Moss[†], Michael B. Winter[†], Giselle M. Knudsen[†], and Charles S. Craik^{*,†}

[†]Department of Pharmaceutical Chemistry, University of California San Francisco, San Francisco, California 94158, United States

[‡]Department of Molecular and Cell Biology, University of California Berkeley, Berkeley, California 94720, United States

Abstract

The more than 500 protein kinases comprising the human kinome catalyze hundreds of thousands of phosphorylation events to regulate a diversity of cellular functions; however, the extended substrate specificity is still unknown for many of these kinases. We report here a method for quantitatively describing kinase substrate specificity using an unbiased peptide library-based approach with direct measurement of phosphorylation by tandem liquid chromatography–tandem mass spectrometry (LC–MS/MS) peptide sequencing (multiplex substrate profiling by mass spectrometry, MSP-MS). This method can be deployed with as low as 10 nM enzyme to determine activity against S/T/Y-containing peptides; additionally, label-free quantitation is used to ascertain catalytic efficiency values for individual peptide substrates in the multiplex assay. Using this approach we developed quantitative motifs for a selection of kinases from each branch of the kinome, with and without known substrates, highlighting the applicability of the method. The sensitivity of this approach is evidenced by its ability to detect phosphorylation events from nanogram quantities of immunoprecipitated material, which allows for wider applicability of this method. To increase the information content of the quantitative kinase motifs, a sublibrary approach was used to expand the testable sequence space within a peptide library of approximately 100 members for CDK1, CDK7, and CDK9. Kinetic analysis of the HIV-1 Tat (transactivator of transcription)-positive transcription elongation factor b (P-TEFb) interaction allowed for localization of the P-TEFb phosphorylation site as well as characterization of the stimulatory effect of Tat on P-TEFb catalytic efficiency.

*Corresponding Author: Phone: 415-476-8146. charles.craik@ucsf.edu.

ORCID

Charles S. Craik: 0000-0001-7704-9185

§Present Address

Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, San Diego, CA.

Notes

The authors declare no competing financial interest.

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.analchem.6b05002. Peptide library components, kinetic data graphs and tables, and detailed methods (PDF)

Table S1, MSP-MS library screen Peptide Reports (XLSX)

Table S3, Time course Peptide Reports (XLSX)

Table S5, PS-SCL Peak Integrations (XLSX)

Graphical Abstract

Protein kinases are a class of post-translational modifying (PTM) enzymes that are essential for carrying out a multitude of functions in the eukaryotic cell. Through addition of a phosphoryl group to a serine, threonine, or tyrosine residue on a substrate protein, kinases regulate processes such as signaling, cell cycle, and gene expression.^{1,2} Roughly one-third of the proteome is likely phosphorylated by one of the over 500 kinases that make up the human kinome.^{3,4} Despite their importance, many kinases and their endogenous substrates have yet to be described.^{3,5} Elucidating the biological functions of PTM enzymes such as kinases requires additional tools that allow for quantitative profiling of their activities in biological systems.

We wanted to test whether the substrate specificity of kinases could be probed using a small, defined synthetic peptide library to accurately report the endogenous specificity of the enzyme. We accomplished this by expanding upon a recently developed method known as multiplex substrate profiling by mass spectrometry (MSP-MS).⁶ This approach employs a 228-member, physicochemically diverse library of rationally designed substrates with maximum sequence diversity within a short tetradecameric peptide. Purified enzymes or complex biological mixtures can be added to the library, and the resulting modified peptide products (mass shifted by +80 Da from addition of a phosphate group) are directly monitored throughout the reaction by peptide sequencing via liquid chromatography–tandem mass spectrometry (LC–MS/MS). Using this highly sensitive method, quantitative substrate specificity motifs and catalytic efficiency values can be obtained for a kinase of interest. In this work, we demonstrate that the current library of 228 14-mer peptides contains modifiable sequences with suitable diversity to determine substrate specificity information for a representative set of evolutionarily distinct kinases.

Kinases generally phosphorylate flexible, unstructured regions of proteins with specificities strongly influenced by the primary amino acid sequence flanking the phosphoacceptor site.⁷ On the basis of these principles, we hypothesized that our linear peptide library would provide appropriate substrates for kinases. Co-crystal structures of kinases with a peptide substrate provide evidence that many kinases can accept a substrate in an extended linear conformation. On the basis of this evidence, linear peptides can be suitable kinase substrates,

and this has been exploited by many peptide-based approaches for profiling the substrate specificity of kinases.^{8–13}

Many kinase profiling technologies employ immobilized peptide arrays or pools of peptides in multiwell plates, allowing for the generation of specificity motifs.^{8,10–13,15,16} These methods have an advantage of being able to query a large number of unique substrates that would be impractical to employ in a single pooled assay. Immobilizing substrates may, however, prevent some of the normal enzyme–substrate interactions, and many of these methods can only detect a single phosphorylation event per substrate and/or not allow for localization of phosphorylation on the peptide substrates. Our MSP-MS assay, while initially limited in sequence diversity, is performed in solution, and localization of phosphorylation(s) on peptide substrates is easily achieved. Previous approaches employing in-solution libraries and phosphopeptide identification via LC–MS/MS have been able to establish optimal substrates by generating a peptide library based on the sequence of an existing peptide substrate.¹⁴ This strategy, while successful, is somewhat biased and often less diverse than a combinatorial library would be; furthermore, application of this strategy is less feasible if one is studying a kinase about which no specificity information is known. MSP-MS is distinct from other similar peptide library-based approaches like the KiC assay¹⁷ and those employing proteome-derived peptide libraries¹⁸ due to its diverse yet simple synthetic peptide library. Proteome-derived peptide library methods commonly involve phosphopeptide enrichment, a step that is unnecessary in MSP-MS. Peptide libraries derived from whole-cell lysates can be advantageous as they are large and diverse, but as a result necessitate a large number of library pools. In contrast, MSP-MS employs a defined and unbiased library with enough diversity to identify two to three key amino acid residues that can then direct development of a PS-SCL (positional scanning synthetic peptide combinatorial library) set to extend specificity motifs.

A variety of powerful proteomics-linked techniques, such as the chemical genetics^{19,20} and KESTREL (kinase substrate tracking and elucidation)²¹ approaches, allow for the direct labeling and identification of endogenous substrates.^{16,22–28} Approaches such as these have proven to be highly successful ways to characterize kinases and identify substrates, but are time- and labor-intensive and generally impractical for novel kinases. In contrast, MSP-MS, while unable to directly label an endogenous substrate, can be used to obtain initial substrate motifs rapidly. Several recent methods, such as the direct-KAYAK (kinase activity assay for kinome profiling)^{29,30} and HAKA-MS (heavy ATP kinase assay mass spectrometry)³¹ approaches, allow for quantitative detection of phosphorylation on substrates from cell lysates through the use of SILAC (stable isotope labeling with amino acids in cell culture) or radioactivity. MSP-MS is distinguished from the KAYAK by factors such as the use of label-free quantitation instead of SILAC, making the assay more straightforward and rapid, and the unbiased, physicochemical diversity of the MSP-MS library.

MSP-MS presents an alternative method for profiling PTM enzymes that is direct, rapid, unbiased, and allows for label-free quantitation, thereby avoiding the use of radioactivity. A straightforward initial MSP-MS experiment allows users to generate specificity motifs revealing the two to three key amino acid residues driving specificity, which then direct the generation of sublibraries for subsequent experiments. We envision that the development of

MSP-MS for use with other classes of PTM enzymes will impact the field of functional proteomics by affording sensitive and quantitative enzymatic signatures of PTM enzymes, either alone or in complex biological samples.

EXPERIMENTAL SECTION

Enzymes

Kinases were purchased or acquired as gifts: P-TEFb (U. Schulze-Gahmen, UC Berkeley),³² cSrc and LegK1 (Shokat lab, UCSF), protein kinase A (PKA; P6000S; New England Biolabs), CSNK1 γ 1 (PV3825; Thermo Fisher), CAM-KK2 (PV4206; Life Technologies), casein kinase II (CKII; P6010S; New England Biolabs), CDK1/cyclin B (P6020L; New England Biolabs), CDK2/cyclin A1 (PV6289; New England Biolabs), CDK7/cyclin H/MNAT1 (PV3868; Life Technologies), p42 MAP kinase (MAPK; P6080S; New England Biolabs), GCK (14–743; Millipore), BMPRII and ALK2 (Jura lab, UCSF), TbERK8 (Z. Mackey, Virginia Tech).

P-TEFb Immunoprecipitation

FLAG-CDK9 was immunoprecipitated from approximately 1×10^8 TTAC-8 cells (gift from Q. Zhou lab, UC Berkeley) grown in 10 T75 flasks as described previously.³³

Peptide Phosphorylation Assays

Initial MSP-MS assays were performed using 10–100 nM of each kinase and 0.5 μ M of each peptide in 50 mM Tris pH 7.4, 150 mM NaCl, 10 mM MgCl₂, and 1 mM ATP as described previously,⁶ except the library contained 104 additional tetradecapeptides designed using the same algorithm as published for a total of 228 synthetic peptides. The library was split into two pools of 114 peptides to optimize detection by LC-MS/MS. Ser/Thr-Pro sublibrary assays were performed under identical conditions except for the use of 3 μ M substrate, and PS-SCL assays used 10 μ M substrate. P-TEFb experiments $\pm$ Tat and/or AFF4 used conditions identical to previous kinase assays, using 100 nM P-TEFb, 100 nM HIV-1 Tat, and 830 nM AFF4.

Phosphopeptide Identification by LC-MS/MS Peptide Sequencing

Peptides were sequenced using an LTQ-Orbitrap Velos mass spectrometer (Thermo Scientific) as described previously.⁶ Detailed instrument settings and search parameters are provided in Supporting Information Methods S1.

Specificity Motif Generation

iceLogos to represent substrate specificity were generated as described previously,^{6,34} with one representation of each peptide phosphorylated by the kinase of interest input into the iceLogo program. “P” is the phosphorylation site, P – 1, P – 2, etc. is used to describe positions on the amino terminal side of the phosphorylation site, and P + 1, P + 2, etc. is used to describe positions on the carboxy terminal side. Heat maps for PS-SCL results were generated using catalytic k_{cat}/K_M values and the maximum percent conversion. Values were

clustered with the *x* algorithm in Cluster software and visualized using Java Treeview.^{35,36} Positive enrichment was visualized as blue and zero was set to yellow.

Design of Sublibraries for Ser/Thr-Pro-Specific Kinase Analyses

A sublibrary of 27 peptides containing a serine/threonine-proline motif and a positional scanning synthetic peptide combinatorial library were developed for in-depth quantitative assays for proline-directed kinases. Library design is described in Supporting Information Methods S2, and synthesis of the library peptides is described in Supporting Information Methods S3.

Kinetics Calculations

Reaction modeling to a kinetics equation was performed as described previously,⁶ with details specific to this experimental set described in Supporting Information Methods S4.

Safety Considerations

Peptide synthesis reagents are hazardous and must be disposed of separately from other waste. The 20% formic acid should be handled with care.

RESULTS AND DISCUSSION

In this study, we sought to test whether a defined, synthetic peptide library could be used to accurately probe kinase substrate specificity, with the goal of producing specificity information at negligible cost and with minimal sample manipulation. We previously developed a synthetic peptide library of 228 tetradecapeptides designed to maximize physicochemical diversity within a minimal sequence space. The current library uses 19 amino acid residues that excludes cysteine and methionine residues but includes norleucine (Nle). Every neighbor (XY) and near-neighbor (X*Y, X**Y) amino acid pair are represented in three or more peptides within the library. Initial use of this relatively small library followed by PS-SCL analysis allows for rapid identification of key substrate specificity determinants, avoiding the need for use of an inconveniently large combinatorial library.

Specificity profiling using MSP-MS is accomplished through addition of a kinase-containing sample (recombinant or immune-precipitated) to the peptide library in assay buffer with ATP. Aliquots are removed at selected time points, and phosphopeptide products are identified and quantified by LC-MS/MS (Figure 1a, Figure S1). Identified phosphopeptides are used to generate specificity profiles (i.e., iceLogos), and catalytic efficiency values can be calculated for individual peptide substrates in the library. Second-generation positional scanning synthetic peptide libraries are then generated based on the initial MSP-MS motif obtained, allowing for significant expansion of sequence diversity.

Establishment of Specificity Motifs for Diverse Kinases

To test the ability of MSP-MS to profile kinase specificity, we profiled at least one kinase from each branch of the kinome (Figure S2). Kinases profiled include PKA, MAPK, P-TEFb, cSrc, CSNK1 γ 1, CAMKK2, casein kinase II, CDK1/cyclin B, CDK2/cyclin A1, CDK7/cyclin H/MNAT1, and GCK (Figure 1b–h). Our method identified phosphorylation

by serine/threonine kinases and tyrosine kinases. These recombinant kinases phosphorylated 8–60 peptides (lists of identified phosphopeptides are in Table S1). Initial MSP-MS iceLogos presented in Figure 1 were generated using one representation of each peptide phosphorylated by the kinase of interest. MSP-MS profiling of these kinases showed that both serine/threonine and tyrosine kinases can produce a strong specificity motif highlighting the key two to three amino acids driving substrate specificity, exemplified by the MAPK substrate profile with a strong preference for proline in the P + 1 position relative to the phosphorylation site (Figure 1b). MSP-MS profiling of well-known kinases CKII and PKA revealed specificity motifs that aligned well with known substrate preferences.^{7–9} CKII favored acidic residues in the P + 1, P + 2, and P + 3 positions (Figure 1d), whereas PKA showed a preference for basic residues at the P – 3 and P – 2 positions and a small hydrophobic residue at the P + 1 position (Figure 1e). Comparison of the specificity data obtained for PKA and CKII to specificity information from a PS-SCL radiography-based approach proved that the data obtained from our experiment is comparable to other approaches, while being very rapid and easy to perform (Figure S3).⁸ Thus, initial MSP-MS analysis allows for rapid determination of two to three amino acids that dominate substrate specificity for a kinase of interest. These key amino acid specificity determinants were later used to guide the development of focused sublibraries to refine specificity motifs.

Substrate Profiling of Less-Characterized Kinases

To broaden the scope of MSP-MS analysis of the kinome, several less-characterized kinases were also studied, with the goal of aiding in endogenous substrate identification. Bone morphogenetic protein receptor type-2 (BMPR2) is a transmembrane protein integral to paracrine signaling.³⁷ BMPR2 is hypothesized to be autophosphorylated at Thr379 and Ser586, but that has yet to be confirmed. MSP-MS profiling of BMPR2 revealed a preference for similarly sized I/E/L residues at P + 1 and L at P + 2 (Figure 1h). This motif aligns with both of the potential autophosphorylation sites, Thr379 (SEVGT(Phospho)IRY-MAP) and Ser586 (KNRNS(Phospho)INYE). A related receptor kinase, ALK2, was also profiled using MSP-MS, and was found to phosphorylate 24 peptides (Figure 1i).

TbERK8 is an essential mitogen-activated protein kinase (MAPK) expressed in the bloodstream form of *Trypanosoma brucei*, the causative agent of sleeping sickness.³⁸ MSP-MS profiling of TbERK8 revealed phosphorylation of five library peptides (Table S1), with a consensus sequence S/T(Phospho)-X-R/K (Figure 1j). This motif aligns with two phosphorylation sites on a recently identified substrate, proliferating cell nuclear antigen (TbPCNA), which is phosphorylated by TbERK8 at Ser211 (ADACS(phospho)VRTH) and Ser216 (VRTHS-(Phospho)AKGKD)³⁹ These results show that MSP-MS screening for TbERK8 substrates allowed for initial identification of two key amino acid residues (S/T(Phospho)-X-R) as lead specificity determinants, consistent with its endogenous substrate sites.

Legionella pneumophila is a pathogenic bacterium that causes Legionnaires' disease in humans. *Legionella* kinase 1 (LegK1) is currently the only known *Legionella* protein with NF- κ -B-stimulating activity, phosphorylating NF- κ -B inhibitor α (NFK-BIA, IKBA) both in vitro and in vivo and potentially phosphorylating I- κ -B-p100 (NFKB2 p100).⁴⁰ LegK1 was

profiled using MSP-MS, identifying eight phosphopeptides (Table S1). An iceLogo generated from these phosphopeptides revealed a strong preference for acidic amino acids at P – 1, P + 1, and P + 2 (Figure 1k). The P – 1 specificity aligned with the phosphorylation sites within the known substrate IKBA, at Ser32 (RHDS(Phospho)GLD) and Ser36 (GLDS(Phospho)-MKDE).⁴⁰ The P – 1 and P + 1 specificity also aligned with the phosphorylation sites within a suggested substrate NFKB2 p100 at Ser713, Ser715, and Ser717 (PPTS₇₁₃DS₇₁₅DS₇₁₇-EGP). Peptides from the sequences of IKBA and NFKB2 p100 were synthesized and used in in vitro kinase assays with LegK1. LegK1 phosphorylated both peptides, with $k_{\text{cat}}/K_{\text{M}}$ values of 0.0037 ± 0.0018 nmol product/[min $\times$ nmol enzyme] for IKBA and 0.0076 ± 0.0026 nmol product/[min $\times$ nmol enzyme] for NFKB2 (Figure S4).

PS-SCL Analysis of Several Cyclin-Dependent Kinases Allows for Delineation of Their Specificity

A promising application of our method is delineation of substrate specificity preferences within a family of closely related kinases, allowing for development of a homologue-selective substrate. Initial MSP-MS analysis of a kinase of interest results in the identification of two to three amino acid residues that drive substrate specificity; however, assessing the roles of residues outside those key positions requires synthesis of a second-generation positional scanning library in which the specificity-determining positions are fixed. Preliminary MSP-MS analysis of several recombinant cyclin-dependent kinases (CDK1/cyclin B, CDK7/cyclin H/MNAT1, and CDK9/cyclin T1) confirmed that each of these kinases is proline-directed (Figure 2, parts b, d, and f)—iceLogo representations for these kinases are dominated by a serine/threonine-proline motif, and more detailed information is required to delineate their substrate specificities.

We therefore developed a PS-SCL derived from a common substrate for each CDK from the 228-member peptide library, RVYLTSPKAPES, to allow for delineation of the substrate specificities of these kinases (Figure 2a). Positions P – 5 through P – 1 and P + 2 through P + 6 were each varied in the parent substrate and substituted with a pool containing an equimolar mixture of 10 representative amino acids from distinct physicochemical groups (described in the Experimental Section), resulting in 10 pools of 10 peptides (Table S4). Catalytic efficiency values for each of the peptides in the PS-SCL assay were used to generate substrate specificity heat maps (Figure 2, parts c, e, and g). Comparison of the heat maps revealed distinct substrate preferences and allowed for identification of homologue-selective substrates. All three kinases had roughly the same activity toward the base peptide for the PS-SCL library, RVYLTSPKAPES, but substitution of P + 3 A to K resulted in a 20-fold better substrate for CDK1 ($k_{\text{cat}}/K_{\text{M}} = 3.5 \pm 0.63$ nmol product/[min $\times$ nmol enzyme]) than for CDK9 ($k_{\text{cat}}/K_{\text{M}} = 0.17 \pm 0.08$ nmol product/[min $\times$ nmol enzyme]) or for CDK7 (substrate is too poor to obtain $k_{\text{cat}}/K_{\text{M}}$ value) (Figure 2h). Compared to the parent peptide, RVYLTSPKAPES, changing the P + 3 residue from A to K leads to a 2.5-fold increase in the $k_{\text{cat}}/K_{\text{M}}$ value for CDK1 ($k_{\text{cat}}/K_{\text{M}}$ increases from 1.4 ± 0.24 to 3.5 ± 0.63 nmol product/[min $\times$ nmol enzyme]; Figure 2i). This kinetic analysis highlights the preference of CDK1 for a basic residue over a nonpolar residue at the P + 3 position and suggests how a coupled MSP-MS/PS-SCL approach can be used to distinguish between closely related kinases. In

principle, this same technique could be applied to any kinase family of interest in order to obtain extended substrate specificity motifs and identify preferred substrates.

Determination of Catalytic Efficiency Values for Peptide Substrates in Focused Sublibraries

We have presented an alternative profiling technique that avoids secondary detection methods (such as radiography or phosphospecific antibodies) through direct label-free quantitation and site localization of phosphorylations on a physicochemically diverse peptide library. CDK9/cyclin T1 (P-TEFb), CDK1/cyclin B, CDK2/cyclin A1, and CDK7/cyclin H/MNAT1 were confirmed to be proline-directed kinases in initial MSP-MS experiments, phosphorylating only peptides containing proline in the P + 1 position. To compare the catalytic efficiency of our MSP-MS peptide substrates with peptides derived from authentic substrates, a serine/threonine-proline sublibrary was developed, the sequences of which are listed in Table S4.

Each cyclin-dependent kinase (CDK) was assayed against the serine/threonine-proline sublibrary, and time-dependent formation of phosphorylated products was quantified using their precursor mass peak integrations by label-free quantitation (Figures S1 and S5). Data were fit to a first-order rate equation (Figure S6a), and catalytic efficiency ($k_{\text{cat}}/K_{\text{M}}$) values were extracted for each peptide (Table S6). P-TEFb phosphorylated natural substrate peptides with a 5–20-fold greater catalytic efficiency than any library peptides (Figure S6b, Table S6). For example, the natural substrate peptide YGSGSRTPLYGSQT had a $k_{\text{cat}}/K_{\text{M}}$ of 0.36 ± 0.02 nmol product/[min $\times$ nmol enzyme] compared to the best library peptide, HRRVYLTSPKAPES ($k_{\text{cat}}/K_{\text{M}} = 0.073 \pm 0.02$ nmol product/[min $\times$ nmol enzyme]). As evidenced by these results, MSP-MS allows for reproducible determination of catalytic efficiency values for each peptide substrate within the multiplex assay. With P-TEFb, we demonstrated that the MSP-MS library contains substrates that are poorer than authentic substrates, but the aggregated motif generated from multiple weak substrates was able to point to a first level of sequence specificity.

Positive Transcription Elongation Factor b (P-TEFb) Phosphorylates Ser-5

P-TEFb phosphorylates the C-terminal domain (CTD) of RNA polymerase II (RNAP II), which consists of 52 repeats of the heptad $Y_1S_2P_3T_4S_5P_6S_7$. Both Ser2 and Ser5 are potential P-TEFb substrates based on the presence of proline in the P + 1 position, but there is controversy as to which serine residue P-TEFb phosphorylates.⁴¹ An RNAP II CTD peptide consisting of two repeats of the YSPTSPS heptad was synthesized (YSPTSPSYSPTSPSR) and S2A and S5A alanine mutants of this peptide were also synthesized (YAP-TSPSYAPTSPSR, YSPTAPSYSPTAPSR) to confirm phosphorylation site localization. These CTD peptides were included in the serine/threonine-proline sublibrary for analysis. P-TEFb assays on this sublibrary revealed that recombinant P-TEFb phosphorylated Ser5 of the CTD peptide (Figure S7). P-TEFb also phosphorylated Ser5 of the CTD S2A peptide, and no phosphorylation was observed on the CTD S5A peptide. Our data showing Ser-5 as the P-TEFb phosphorylation site are in agreement with a recent study addressing P-TEFb specificity that also employed peptide substrates, although this

preference for Ser-5 may not necessarily translate to Ser-5 specificity within full-length RNAP II CTD in vivo.⁴²

P-TEFb mutants containing cyclin T1 variants with differently truncated C-terminal domains were assayed against the CTD WT and alanine mutant peptides to determine if the CTD of cyclin T1 controls P-TEFb substrate specificity. All truncated forms of cyclin T1 and full-length cyclin T1 still exclusively phosphorylated Ser5. The CTD of cyclin T1 was found to not alter P-TEFb selectivity, although longer cyclin T1 CTDs exhibited lower kinase activity (Figure S8).

The effect of cis/trans conformation of Pro within the RNAP II CTD was considered, based on the hypothesis that proline conformation and phosphorylation state may be related.^{43,44} Additional mutant CTD peptides were synthesized to determine if having P₃ or P₆ locked in the cis-conformation would alter the phosphorylation site. YSP(cis)TSPSYSP(cis)TSPS and YSPTSP(cis)SYSPTSP(cis)S peptides were synthesized using 5,5-dimethyl-pyrrolidine-2-carboxylic acid to simulate a cis-locked proline.⁴⁵ Proline conformation did not alter P-TEFb specificity, as Ser5 phosphorylation was still observed exclusively on each peptide. However, these “cis-locked” CTD peptides did exhibit diminished k_{cat}/K_M values as compared to the natural CTD peptide, with 17- and 13-fold lower values for *cis*-Pro3 CTD and *cis*-Pro6 CTD peptides, respectively (Figure S9, Table S7).

Immunoprecipitated P-TEFb

To test whether the phosphorylation site analysis with recombinant P-TEFb reproduced native P-TEFb specificity, P-TEFb complexes were immunoprecipitated from HEK293 cells stably expressing FLAG-tagged CDK9 (Figure 3a). Cyclin T1 binds tightly to CDK9, and this association was confirmed for FLAG-CDK9 by coimmunoprecipitation (Figure 3b). MSP-MS analysis of immunoprecipitated P-TEFb identified eight phosphopeptides. An iceLogo generated from this data (Figure 3c) revealed that recombinant and immunoprecipitated P-TEFb had matching substrate specificity (Figure 2f). Immunoprecipitated P-TEFb was also analyzed using the serine/threonine-proline sublibrary, and in agreement with recombinant P-TEFb results, immunoprecipitated P-TEFb only phosphorylated Ser5 of the CTD and CTD S2A peptides and did not phosphorylate CTD S5A.

In this experiment, the total yield of immunoprecipitated P-TEFb complex was estimated as 100 ng by densitometry of gel bands relative to standards, such that kinase assays could be performed with 20–30 nM enzyme. This is comparable to the minimum concentration of recombinant P-TEFb required to detect activity (10 nM; Figure S10). Recombinant P-TEFb MSP-MS assays were scaled down in a dilution series to determine the minimum amount of material that was practical for sample handling and analysis, and the limit of detection for P-TEFb was found to require 0.3 pmol, an amount easily attainable with a single immunoprecipitation experiment. The ability of our MSP-MS method to detect phosphorylation by an immunoprecipitated kinase speaks to the sensitivity and potential wide applicability of our method, allowing for profiling of kinases that cannot be produced recombinantly.

HIV-1 Tat Increases P-TEFb Catalytic Efficiency and Strengthens Substrate Specificity

HIV-1 Tat coordinates P-TEFb phosphorylation of the RNAP II CTD as well as two negative elongation factors, NELF and DSIF, to overcome the stalled transcription of the integrated viral genome (Figure 3d). Tat has been reported to increase P-TEFb catalytic efficiency as well as alter substrate specificity; however, the specificity claims were made using less sensitive assays employing phosphospecific antibodies.^{41,46–51} The scaffolding protein AFF4 has been suggested to have an inhibitory effect on P-TEFb activity.³² P-TEFb specificity and catalytic efficiency were therefore examined in the presence and absence of Tat and/or AFF4. Addition of Tat to the MSP-MS profiling assay of P-TEFb did not result in phosphorylation of any additional peptides from the 228-member library, and thus did not alter specificity (Table S1). However, kinetic analysis using the serine/threonine-proline sublibrary revealed that Tat increased the catalytic efficiency toward several substrate peptides that were already identified (Table S8). On the other hand, including the scaffolding protein AFF4 in this P-TEFb ± Tat experiment had no significant effect on either catalytic efficiency or substrate specificity (Table S9).

The effect of Tat on P-TEFb activity was nonuniform, as the catalytic efficiency of some peptides was enhanced more than others (Figure 3e). A weighted specificity motif was generated based on the sequences of peptides whose k_{cat}/K_M values were Tat-increased more than 2-fold (Figure 3f, Table S8). Alignment of this motif with RNAP II CTD Ser5 and Ser2 as potential phosphorylation sites indicates that Tat selectively increases P-TEFb catalytic efficiency toward peptides similar to the CTD Ser5 site; Ser5 as the modification site aligns with the weighted motif at six out of seven positions, whereas Ser2 as the modification site only aligns at three out of seven positions (Figure 3f). These results suggest that Tat strengthens P-TEFb specificity for Ser-5-like peptide substrates, but addition of the AFF4 scaffold protein does not alter intrinsic P-TEFb kinase activity.

CONCLUSIONS

In summary, we show that a chemically defined, diverse peptide library enables the rapid, sensitive, and quantitative profiling of a wide variety of kinases. While biased libraries are clearly useful for profiling specific kinases of interest, it is notable that our unbiased library was sufficiently diverse to allow for detection of phosphorylation events from numerous evolutionarily and substrate preference diverse kinases. As an initial screen, this rapid experiment provides enough key specificity information to direct the development of custom-designed libraries for an individual enzyme. The ability of our chemically diverse peptide library to capture the substrate specificity for proteases and now kinases suggests its potential use for other kinds of PTM enzymes.

Acknowledgments

We thank Jennifer Kung and Christopher Agnew (N. Jura lab, UCSF) for providing BMPR2 and ALK2 and Rebecca Levin (K. Shokat lab, UCSF) for providing cSrc. We would like to thank Huasong Lu (Q. Zhou lab, UC Berkeley) for the HEK293 cell line stably expressing FLAG-CDK9 with Dox-inducible HA-Tat, and for the P-TEFb mutants with various cyclin T1 truncations, and Peter Cimermancic (Sali lab, UCSF) for advice on computational aspects of the project. We also thank Kevan Shokat for helpful discussions and critical reading of the manuscript.

References

1. Chen Z, Gibson TB, Robinson F, Silvestro L, Pearson G, Xu B, Wright A, Vanderbilt C, Cobb MH. *Chem Rev.* 2001; 101:2449–2476. [PubMed: 11749383]
2. Johnson GL, Lapadat R. *Science.* 2002; 298:1911–1912. [PubMed: 12471242]
3. Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S. *Science.* 2002; 298:1912–1934. [PubMed: 12471243]
4. Cohen P. *Trends Biochem Sci.* 2000; 25:596–601. [PubMed: 11116185]
5. Walsh CT, Garneau-Tsodikova S, Gatto GJ. *Angew Chem, Int Ed.* 2005; 44:7342–7372.
6. O'Donoghue AJ, Eroy-Reveles AA, Knudsen GM, Ingram J, Zhou M, Statnekov JB, Greninger AL, Hostetter DR, Qu G, Maltby DA, et al. *Nat Methods.* 2012; 9:1095–1100. [PubMed: 23023596]
7. Songyang Z, Lu KP, Kwon YT, Tsai LH, Filhol O, Cochet C, Brickey DA, Soderling TR, Bartleson C, Graves DJ, et al. *Mol Cell Biol.* 1996; 16:6486–6493. [PubMed: 8887677]
8. Hutti JE, Jarrell ET, Chang JD, Abbott DW, Storz P, Toker A, Cantley LC, Turk BE. *Nat Methods.* 2004; 1:27–29. [PubMed: 15782149]
9. Pearson RB, Kemp BE. *Methods Enzymol.* 1991; 200:62–81. [PubMed: 1956339]
10. Wu JJ, Phan H, Lam KS. *Bioorg Med Chem Lett.* 1998; 8:2279–2284. [PubMed: 9873528]
11. Rychlewski L, Kschischo M, Dong L, Schutkowski M, Reimer U. *J Mol Biol.* 2004; 336:307–311. [PubMed: 14757045]
12. Rodriguez M, Li SSC, Harper JW, Songyang Z. *J Biol Chem.* 2004; 279:8802–8807. [PubMed: 14679191]
13. Luo K, Zhou P, Lodish HF. *Proc Natl Acad Sci U S A.* 1995; 92:11761–11765. [PubMed: 8524844]
14. Till JH, Annan RS, Carr SA, Miller WT. *J Biol Chem.* 1994; 269:7423–7428. [PubMed: 8125961]
15. Miller CJ, Turk BE. *Methods Mol Biol.* 2016; 1360:203–216. [PubMed: 26501912]
16. Amanchy R, Zhong J, Molina H, Chaerkady R, Iwahori A, Kalume DE, Grønborg M, Joore J, Cope L, Pandey A. *J Proteome Res.* 2008; 7:3900–3910. [PubMed: 18698806]
17. Huang Y, Thelen JJ. *Methods Mol Biol.* 2012; 893:359–370. [PubMed: 22665311]
18. Kettenbach AN, Wang T, Faherty BK, Madden DR, Knapp S, Bailey-Kellogg C, Gerber SA. *Chem Biol.* 2012; 19:608–618. [PubMed: 22633412]
19. Shah K, Liu Y, Deirmengian C, Shokat KM. *Proc Natl Acad Sci U S A.* 1997; 94:3565–3570. [PubMed: 9108016]
20. Ulrich SM, Kenski DM, Shokat KM. *Biochemistry.* 2003; 42:7915–7921. [PubMed: 12834343]
21. Knebel A. *EMBO J.* 2001; 20:4360–4369. [PubMed: 11500363]
22. Huang SY, Tsai ML, Chen GY, Wu CJ, Chen SH. *J Proteome Res.* 2007; 6:2674–2684. [PubMed: 17564427]
23. Xue L, Wang WH, Iliuk A, Hu L, Galan JA, Yu S, Hans M, Geahlen RL, Tao WA. *Proc Natl Acad Sci U S A.* 2012; 109:5615–5620. [PubMed: 22451900]
24. Holt LJ, Tuch BB, Villén J, Johnson AD, Gygi SP, Morgan DO. *Science.* 2009; 325:1682–1686. [PubMed: 19779198]
25. Coba MP, Pocklington AJ, Collins MO, Kopanitsa MV, Uren RT, Swamy S, Croning MDR, Choudhary JS, Grant SG. *Sci Signaling.* 2009; 2:ra19.
26. Yu Y, Anjum R, Kubota K, Rush J, Villen J, Gygi SP. *Proc Natl Acad Sci U S A.* 2009; 106:11606–11611. [PubMed: 19564600]
27. Troiani S, Uggeri M, Moll J, Isacchi A, Kalisz HM, Rusconi L, Valsasina B. *J Proteome Res.* 2005; 4:1296–1303. [PubMed: 16083279]
28. Morandell S, Grosstessner-Hain K, Roitinger E, Hudecz O, Lindhorst T, Teis D, Wrulich OA, Mazanek M, Taus T, Ueberall F, et al. *Proteomics.* 2010; 10:2015–2025. [PubMed: 20217869]
29. Kubota K, Anjum R, Yu Y, Kunz RC, Andersen JN, Kraus M, Keilhack H, Nagashima K, Krauss S, Paweletz C, et al. *Nat Biotechnol.* 2009; 27:933–940. [PubMed: 19801977]
30. Kunz RC, McAllister FE, Rush J, Gygi SP. *Anal Chem.* 2012; 84:6233–6239. [PubMed: 22724890]

31. Müller AC, Giambruno R, Weißer J, Májek P, Hofer A, Bigenzahn JW, Superti-Furga G, Jessen HJ, Bennett KL. *Sci Rep*. 2016; 6:28107. [PubMed: 27346722]
32. Schulze-Gahmen U, Upton H, Birnberg A, Bao K, Chou S, Krogan NJ, Zhou Q, Alber T. *eLife*. 2013; 2:e00327. [PubMed: 23471103]
33. He N, Liu M, Hsu J, Xue Y, Chou S, Burlingame A, Krogan NJ, Alber T, Zhou Q. *Mol Cell*. 2010; 38:428–438. [PubMed: 20471948]
34. Colaert N, Helsens K, Martens L, Vandekerckhove J, Gevaert K. *Nat Methods*. 2009; 6:786–787. [PubMed: 19876014]
35. de Hoon MJL, Imoto S, Nolan J, Miyano S. *Bioinformatics*. 2004; 20:1453–1454. [PubMed: 14871861]
36. Saldanha AJ. *Bioinformatics*. 2004; 20:3246–3248. [PubMed: 15180930]
37. Feng XH, Derynck R. *Annu Rev Cell Dev Biol*. 2005; 21:659–693. [PubMed: 16212511]
38. Valenciano AL, Ramsey AC, Mackey ZB. *Cell Cycle*. 2015; 14:674–688. [PubMed: 25701409]
39. Valenciano AL, Knudsen GM, Mackey ZB. *Cell Cycle*. 2016; 15:2827–2841. [PubMed: 27589575]
40. Ge J, Xu H, Li T, Zhou Y, Zhang Z, Li S, Liu L, Shao F. *Proc Natl Acad Sci U S A*. 2009; 106:13725–13730. [PubMed: 19666608]
41. Saunders A, Core LJ, Lis JT. *Nat Rev Mol Cell Biol*. 2006; 7:557–567. [PubMed: 16936696]
42. Czudnochowski N, Böskén CA, Geyer M. *Nat Commun*. 2012; 3:842. [PubMed: 22588304]
43. Xu YX, Manley JL. *Genes Dev*. 2007; 21:2950–2962. [PubMed: 18006688]
44. Xu YX, Hirose Y, Zhou XZ, Lu KP, Manley JL. *Genes Dev*. 2003; 17:2765–2776. [PubMed: 14600023]
45. Bielska AA, Zondlo NJ. *Biochemistry*. 2006; 45:5527–5537. [PubMed: 16634634]
46. Schüller R, Forné I, Straub T, Schrieck A, Texier Y, Shah N, Decker TM, Cramer P, Imhof A, Eick D. *Mol Cell*. 2016; 61:305–314. [PubMed: 26799765]
47. Buratowski S. *Mol Cell*. 2009; 36:541–546. [PubMed: 19941815]
48. Egloff S, Murphy S. *Trends Genet*. 2008; 24:280–288. [PubMed: 18457900]
49. Chapman RD, Heidemann M, Hintermair C, Eick D. *Trends Genet*. 2008; 24:289–296. [PubMed: 18472177]
50. Phatnani HP, Greenleaf AL. *Genes Dev*. 2006; 20:2922–2936. [PubMed: 17079683]
51. Meinhart A, Kamenski T, Hoepfner S, Baumli S, Cramer P. *Genes Dev*. 2005; 19:1401–1415. [PubMed: 15964991]
52. Mortz E, Krogh TN, Vorum H, Görg A. *Proteomics*. 2001; 1:1359–1363. [PubMed: 11922595]

Figure 1.

Depiction of the MSP-MS assay for kinases. (a) Recombinant or immunoprecipitated kinases are added to the peptide library. Aliquots are removed from the reaction at several time points, quenched, and then analyzed by LC-MS/MS. Phosphopeptides were identified and quantified at each time point, then used to develop specificity motifs and generate catalytic efficiency values for each peptide in the multiplex assay. (b–g) Validation of the use of MSP-MS for kinases using a variety of disparate, well-studied kinases and (h–k) characterization of less-studied kinases. All substrate signatures were developed using the phosphopeptides identified after 1200 min of kinase reaction with the peptide library. (a) GCK. (b) MAPK. (c) cSrc. (d) Casein kinase II. (e) Protein kinase A. (f) CAMKK2. (g) CSNK1γ1. (h) BMPR2. (i) ALK2. (j) TbERK8. (k) LegK1.

Figure 2.

Delineation of cyclin-dependent kinase substrate specificities. (a) Development of a positional scanning synthetic combinatorial library for further substrate specificity determination after initial MSP-MS characterization. A scaffold peptide is chosen from the library and two positions are fixed, based on the two key determinants of substrate specificity, as identified in MSP-MS. Every other position is then varied, one at a time, to a mixture of a diversity of amino acids representing all distinct physicochemical groups. Kinetic PS-SCL assays allow for generation of a more detailed substrate specificity heat map. (b) Substrate profile from MSP-MS analysis of CDK9/cyclin T1. (c) Heat map generated using maximum percent conversion values from CDK9 PS-SCL experiment. Positive enrichment was visualized as blue, and zero was set to yellow. Residue indicated in red is highlighted for comparison between kinases. (d) Substrate profile from CDK1/cyclin B. (e) Heat map generated using maximum percent conversion values from CDK1 PS-SCL experiment. (f) Substrate profile from CDK7/cyclin H/MNAT1. (g) Heat map generated using maximum percent conversion values from CDK7 PS-SCL experiment. (h) Differential phosphorylation of the “P + 3 K” peptide, RVYLTSPKKPES, by CDK1, CDK9, and CDK7, highlighting each enzyme’s preference for basic residues at the P + 3 position. P + 3 K is a better substrate for CDK1 than for CDK9 (20-fold difference in k_{cat}/K_M) or for CDK7 (substrate is too poor to obtain k_{cat}/K_M value). (i) CDK1 phosphorylation of PS-SCL scaffold peptide RVYLTSPKAPES and P + 3 A9 $\rightarrow$ K9 peptide RVYLTSPKKPES. Alteration of the P + 3 residue makes the peptide a 2.5-fold better substrate.

Figure 3.

MSP-MS analysis of recombinant and immunoprecipitated P-TEFb and the HIV-1 Tat-P-TEFb interaction. (a) Silver stain gel of FLAG-CDK9 immunoprecipitation from HEK293 cells stably expressing FLAG-CDK9. Lane contents: 1, ladder; 2, unbound to beads; 3, first wash; 4, FLAG eluent; 5, lysate. Indicated bands correspond to the correct molecular weight expected for FLAG-CDK9 and cyclin T1. In lane 4, 5 μ L out of the 110 μ L elution from the FLAG beads was loaded for SDS-PAGE and silver staining. This method has an approximately 50 fmol limit of detection,⁵² which for 124 kDa P-TEFb corresponds to approximately 5 ng as a lower limit of protein abundance. Total yield from this preparation

was estimated as approximately 100 ng. An amount of 0.4 ng of active P-TEFb is required per time point (Figure S10). (b) Anti-cyclin T1 Western blot of FLAG-CDK9 immunoprecipitation from HEK293 cells stably expressing FLAG-CDK9 reveals that cyclin T1 is the primary coimmunoprecipitant of FLAG-CDK9. Lane contents: 1, ladder; 2, unbound to beads; 3, FLAG eluent. (c) Substrate profile from immunoprecipitated P-TEFb, which aligns with recombinant P-TEFb substrate specificity. (d) Schematic representing the super elongation complex (SEC) in which Tat associates with P-TEFb and the scaffolding protein in order to coordinate P-TEFb phosphorylation of the RNA Pol II CTD and the two negative elongation factors NELF and DSIF, which allows the stalled RNA Pol II to resume transcription of the integrated viral genome. Residues phosphorylated by P-TEFb (RNA Pol II CTD Ser5; NELF Ser181, Thr272, Ser281, and Ser353; DSIF Thr775 and Thr784) and the proline residues driving specificity are represented in the figure in bold. (e) Kinetic experiments including HIV-1 Tat reveal that some peptides have an enhanced k_{cat}/K_M in the presence of Tat while others are relatively unchanged. (f) Weighted specificity motif generated using the sequences of peptides that Tat increased the k_{cat}/K_M values for by more than 2-fold. Alignment of this motif with RNAP II CTD Ser5 and RNAP II CTD Ser2 as the phosphorylation site indicates that Tat selectively increases P-TEFb catalytic efficiency toward peptides that are similar to CTD Ser5.