Title

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

Journal

Analytical Chemistry, 89(8)

ISSN

0003-2700

Authors

Meyer, Nicole O
O'Donoghue, Anthony J
Schulze-Gahmen, Ursula
et al.

Publication Date

2017-04-18

DOI

10.1021/acs.analchem.6b05002

Peer reviewed

Title: Multiplex Substrate Profiling by Mass Spectrometry for Kinases Reveals Quantitative Substrate Motifs.

Authors: Nicole O. Meyer[1], A.J. O'Donoghue [1,2], Ursula Schulze-Gahmen[3], Matthew Ravalin[1], Steven M. Moss[1], Michael B. Winter[1], Giselle M. Knudsen[1], and Charles S. Craik[1]*

[1]Dept. of Pharmaceutical Chemistry, University of California San Francisco, San Francisco, CA 94158. [2]Present address: Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, San Diego, CA. [3]Dept. of Molecular and Cell Biology, University of California Berkeley, Berkeley, CA.

*To whom correspondence should be addressed: Charles S. Craik, Univ. of California San Francisco, Genentech Hall, MC 2280, 600 16th Street, S-514, San Francisco, CA, 94158-2517, Tel.: 415-476-8146, email: charles.craik@ucsf.edu

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 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 microgram quantities of immunoprecipitated material, which will allow for wider applicability of this method. To increase the information content of the quantitative kinase motifs, a sub-library 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 – 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.

Introduction:

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, physico-chemically 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-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] Based on 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. Based on 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.[10,11,12,13,14,15]

There are a variety of tools available for profiling kinases. A number of peptide array and peptide library-based methods exist, allowing for generation of specificity motifs.[10,12,13,14,15,16,17] A variety of powerful proteomics-linked techniques, such as the chemical genetics [18,19] and KESTREL [20] approaches, allow for the direct labeling and identification of endogenous substrates.[21,22,23,24,25,26,27,28,29] Several recent methods,

such as the direct-KAYAK [30] and HAKA-MS [31] approaches, allow for quantitative detection of phosphorylation on substrates from cell lysates through the use of SILAC or radioactivity.

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. MSP-MS is distinct from other similar peptide-library approaches like the KiC assay [32] due to its diverse yet simple peptide library. 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 sub-libraries 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)[33], cSrc and LegK1 (Shokat lab, UCSF), Protein Kinase A (PKA; P6000S; New England Biolabs), CSNK1γ1 (PV3825; Thermo Fisher), CAMKK2 (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 TTAC-8 cells (gift from Q. Zhou lab, UC Berkeley) as described previously.[34]

Peptide phosphorylation assays.

Initial MSP-MS assays were performed using 10-100 nM of each kinase and 0.5 µM of each peptide (n=228) in 50 mM Tris pH 7.4, 150 mM NaCl, 10 mM $MgCl_2$, and 1 mM ATP as described previously[6]. Ser/Thr-Pro sub-library assays were performed under identical conditions except for the use of 3 µM substrate, and PS-SCL assays used 10 µM substrate. P-TEFb experiments +/- 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 Supplemental Methods 1.

Specificity motif generation.

iceLogos to represent substrate specificity were generated as described previously.[6,35] "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 carboxyl 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.[36,37] Positive enrichment was visualized as blue and zero was set to yellow.

Design of sub-libraries for Ser/Thr-Pro-specific kinase analyses.

A sub-library 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 Supplemental Methods 2 and synthesis of the library peptides is described in Supplemental Methods 3.

Kinetics calculations.

Reaction modeling to a kinetics equation was performed as described previously[6], with details specific to this experimental set described in Supplemental Methods 4.

Safety considerations.

Peptide synthesis reagents are hazardous and must be disposed of separately from other waste. 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 maximal specificity information at negligible cost and with minimal sample manipulation. We previously developed a synthetic peptide library of 228 tetradecapeptides designed to maximize physico-chemical 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 (Figure 1a).

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 liquid chromatography-tandem mass spectrometry (LC-MS/MS) (Figure 1b). 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 can then be generated for more accurate quantitation and expansion of sequence diversity based on the initial motif obtained.

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. Kinases profiled include PKA, MAPK, P-TEFb, cSrc, CSNK1γ1, CAMKK2, casein

kinase II, CDK1/Cyclin B, CDK2/Cyclin A1, CDK7/Cyclin H/MNAT1, and GCK (Supplemental Figures 1-6). Our method identified phosphorylation by serine/threonine kinases and tyrosine kinases. These recombinant kinases phosphorylated 8 to 60 peptides (lists of identified phosphopeptides are in Supplemental Table 1). MSP-MS profiling of these kinases showed that both serine/threonine and tyrosine kinases can produce a strong specificity motif, exemplified by the CDK1 substrate profile with a strong preference for proline in the P+1 position relative to the phosphorylation site (Supplemental Figure 1a). MSP-MS profiling of well-known kinases CKII and PKA revealed specificity motifs that aligned well with known substrate preferences.[10,7,11] CKII favored acidic residues in the P+1, P+2, and P+3 positions (Supplemental Figure 1c), 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 (Supplemental Figure 1d). GCK and CSNK1γ1 preferred to phosphorylate proteins with a small hydrophobic residue at P+1 (Supplemental Figure 1e, 1f). 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 next used to guide the development of focused sub-libraries to refine specificity motifs.

Substrate profiling of less characterized kinases.

To broaden the scope of MSP-MS analysis, 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.[41] 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 (Supplemental Figure 7a). This motif aligns with both of the potential autophosphorylation sites, Thr379 (SEVGT(Phospho)**I**RYMAP) and Ser586 (KNRNS(Phospho)**I**NYE). A related receptor kinase, ALK2, was also profiled using MSP-MS, and was found to phosphorylate 24 peptides (Supplemental Figure 7b).

TbERK8 is an essential mitogen-activated protein kinase (MAPK) expressed in the bloodstream form of *Trypanosoma brucei*, the causative agent of sleeping sickness.[42] MSP-MS profiling of TbERK8 revealed phosphorylation of five library peptides (Supplemental Table 1), with a consensus sequence

S/**T**(Phospho)-X-**R/K** (Supplemental Figure 7c). 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)V**R**TH) and Ser216 (VRTHS(Phospho)A**K**GKD) [43] 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-kappa-B-stimulating activity, phosphorylating NF-kappa-B inhibitor alpha (NFKBIA, IKBA) both *in vitro* and *in vivo* and potentially phosphorylating I-kappa-B-p100 (NFKB2 p100).[44] LegK1 was profiled using MSP-MS, identifying eight phosphopeptides (Supplemental Table 1). An iceLogo generated from these phosphopeptides revealed a strong preference for acidic amino acids at P-1, P+1, and P+2 (Supplemental Figure 7d). The P-1 specificity aligned with the phosphorylation sites within the known substrate IKBA, at Ser32 (RH**D**S(Phospho)GLD) and Ser36 (GL**D**S(Phospho)MKDE).[44] The P-1 and P+1 specificity also aligned with the phosphorylation sites within a suggested substrate NFKB2 p100 at Ser713, Ser715, and Ser717 (PPTS**D**S_{713}**D**S_{715}**D**S_{717}**E**GP). 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_{cat}/K_M values of 0.0037 ± 0.0018 nmol product / [min × nmol enzyme] for IKBA and 0.0076 ± 0.0026 nmol product / [min × nmol enzyme] for NFKB2 (Supplemental Figure 7e).

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. 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 3a, 3c, 3e) - 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 positional scanning-synthetic peptide combinatorial library (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.

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 ten representative amino acids from distinct physicochemical groups (described in methods), resulting in ten pools of ten peptides (Supplemental Table 4). Catalytic efficiency values for each of the peptides in the PS-SCL assay were used to generate substrate specificity heat maps (Figure 3b, 3d, 3f). Comparison of the heat maps revealed distinct substrate preferences and allowed for identification of isoform-selective substrates. For example, RVYLT**S**PK**K**PES, was a 20-fold better substrate for CDK1 (k_{cat}/K_M = 3.5 ± 0.63 nmol product / [min × nmol enzyme]) than for CDK9 (kcat/KM = 0.17 ± 0.08 nmol product / [min × nmol enzyme]) or for CDK7 (substrate is too poor to obtain k_{cat}/K_M value) (Figure 3g). Compared to the parent peptide, RVYLT**S**PK**A**PES, changing the P+3 residue from A to K leads to a 2.5-fold increase in the k_{cat}/K_M value for CDK1 (k_{cat}/K_M increases from 1.4 ± 0.24 to 3.5 ± 0.63 nmol product / [min × nmol enzyme]Figure 3h). This kinetic analysis highlights the preference of CDK1 for a basic residue over a non-polar 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

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 physico-chemically 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 sub-library was developed, the sequences of which are listed in Supplemental Table 4.

Each cyclin-dependent kinase (CDK) was assayed against the serine/threonine-proline sub-library and time-dependent formation of phosphorylated products was quantified using their precursor mass peak integrations by label-free quantitation (Supplemental Figures 8 and 9a). Data were fit to a first order rate equation (Supplemental Figure 9b), and catalytic efficiency (k_{cat}/K_M) values were extracted for each peptide. P-TEFb phosphorylated natural substrate peptides with a 5-20-fold greater catalytic efficiency than any library peptides (Supplemental Figure 9 and Supplemental Table 6). For example, the natural substrate peptide YGSGSR**T**PLYGSQT had a k_{cat}/K_M of 0.36 ± 0.02 nmol product / [min × nmol enzyme] compared to the best library peptide, HRRVYLT**S**PKAPES (k_{cat}/K_M = 0.073 ± 0.02 nmol product / [min × nmol enzyme]). k_{cat}/K_M values for CDK1/Cyclin B, CDK2/Cyclin A1, and CDK7/Cyclin H/MNAT1 for each peptide in the sub-library are reported in Supplemental Tables 7-9 and Supplemental Figures 10-12. As evidenced by these results, MSP-MS allows for reproducible determination of catalytic efficiency values for each peptide substrate within the multiplex assay. In this set of proline-directed kinases, 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.[46] 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 (YAPTSPSYAPTSPSR, YSPTAPSYSPTAPSR) to confirm phosphorylation site localization. These CTD peptides were included in the serine/threonine-proline sub-library for analysis. P-TEFb assays on this sub-library revealed that recombinant P-TEFb phosphorylated Ser5 of the CTD peptide (Figure

2a). 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*.[53]

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 (Supplemental Figure 13).

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.[54,55] Additional mutant CTD peptides were synthesized to determine if having P_3 or P_6 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.[56] 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 a 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 (Supplemental Figure 14 and Supplemental Table 10).

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. Cyclin T1 binds tightly to CDK9, and this association was confirmed for FLAG-CDK9 by co-immunoprecipitation (Figure 2c). MSP-MS analysis of immunoprecipitated P-TEFb identified eight phosphopeptides. An iceLogo generated from this data (Figure 2d) revealed that recombinant and

immunoprecipitated P-TEFb had matching substrate specificity (Supplemental Figure 6). Immunoprecipitated P-TEFb was also analyzed using the serine/threonine-proline sub-library 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 20 ug 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; Supplemental Figure 15). 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 pmoles, 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 2e). 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 phospho-specific antibodies. [47,48,49,50,51,52,46] The scaffolding protein AFF4 has been suggested to have an inhibitory effect on P-TEFb activity.[33] 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 (Supplemental Table 1). However, kinetic analysis using the serine/threonine-proline sub-library revealed that Tat increased the catalytic efficiency toward several substrate peptides that were already identified (Supplemental Table 11). 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 (Supplemental Table 12).

The effect of Tat on P-TEFb activity was non-uniform, as the catalytic efficiency of some peptides was enhanced more than others (Figure 2f). A weighted specificity motif was generated based on the sequences of peptides whose k_{cat}/K_M values were Tat-increased more than two-fold (Figure 2g, Supplemental Table 11). Alignment of this motif with RNAP II CTD Ser5 and Ser2 as potential phosphorylation sites indicates that Tat selectively increases P-TEFb catalytic efficiency towards 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 2g). 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.

Acknowledgements:

We thank Jennifer Kung and Christopher Agnew (N. Jura lab, UCSF) for providing BMPR2 and ALK2 and Rebecca Levin and Steven Moss (K. Shokat lab, UCSF) for providing cSrc and LegK1. 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, T. B., Robinson, F., Silvestro, L., Pearson, G., Xu, B., Wright, A., Vanderbilt, C., and Cobb, M. H. (2001) MAP kinases. *Chem. Rev. 101*, 2449–2476.
(2) Johnson, G. L., and Lapadat, R. (2002) Mitogen-activated protein kinase pathways mediated by ERK, JNK, and p38 protein kinases. *Science 298*, 1911–1912.
(3) Manning, G., Whyte, D. B., Martinez, R., Hunter, T., and Sudarsanam, S. (2002) The protein kinase complement of the human genome. *Science 298*, 1912–1934.
(4) Cohen, P. (2000) The regulation of protein function by multisite phosphorylation--a 25 year update. *Trends Biochem. Sci. 25*, 596–601.
(5) Walsh, C. T., Garneau-Tsodikova, S., and Gatto, G. J. (2005) Protein posttranslational modifications: the chemistry of proteome diversifications. *Angew. Chem. Int. Ed Engl. 44*, 7342–7372.
(6) O'Donoghue, A. J., Eroy-Reveles, A. A., Knudsen, G. M., Ingram, J., Zhou, M., Statnekov, J. B., Greninger, A. L., Hostetter, D. R., Qu, G., Maltby, D. A., Anderson, M. O., Derisi, J. L., McKerrow, J. H., Burlingame, A. L., and Craik, C. S. (2012) Global identification of peptidase specificity by multiplex substrate profiling. *Nat. Methods 9*, 1095–1100.
(7) Songyang, Z., Lu, K. P., Kwon, Y. T., Tsai, L. H., Filhol, O., Cochet, C., Brickey, D. A., Soderling, T. R., Bartleson, C., Graves, D. J., DeMaggio, A. J., Hoekstra, M. F., Blenis, J., Hunter, T., and Cantley, L. C. (1996) A structural basis for substrate specificities of protein Ser/Thr kinases: primary sequence preference of casein kinases I and II, NIMA, phosphorylase kinase, calmodulin-dependent kinase II, CDK5, and Erk1. *Mol. Cell. Biol. 16*, 6486–6493.
(8) Brinkworth, R. I., Breinl, R. A., and Kobe, B. (2003) Structural basis and prediction of substrate specificity in protein serine/threonine kinases. *Proc. Natl. Acad. Sci. U. S. A. 100*, 74–79.
(9) Johnson, L. N., Lowe, E. D., Noble, M. E., and Owen, D. J. (1998) The Eleventh Datta Lecture. The structural basis for substrate recognition and control by protein kinases. *FEBS Lett. 430*, 1–11.
(10) Hutti, J. E., Jarrell, E. T., Chang, J. D., Abbott, D. W., Storz, P., Toker, A., Cantley, L. C., and Turk, B. E. (2004) A rapid method for determining protein kinase phosphorylation specificity. *Nat. Methods 1*, 27–29.
(11) Pearson, R. B., and Kemp, B. E. (1991) Protein kinase phosphorylation site sequences and consensus specificity motifs: tabulations. *Methods Enzymol. 200*, 62–81.
(12) Wu, J. J., Phan, H., and Lam, K. S. (1998) Comparison of the intrinsic kinase activity and substrate specificity of c-Abl and Bcr-Abl. *Bioorg. Med. Chem. Lett. 8*, 2279–2284.
(13) Rychlewski, L., Kschischo, M., Dong, L., Schutkowski, M., and Reimer, U. (2004) Target Specificity Analysis of the Abl Kinase using Peptide Microarray Data. *J. Mol. Biol. 336*, 307–311.

(14) Rodriguez, M., Li, S. S.-C., Harper, J. W., and Songyang, Z. (2004) An Oriented Peptide Array Library (OPAL) Strategy to Study Protein-Protein Interactions. *J. Biol. Chem. 279*, 8802–8807.
(15) Luo, K., Zhou, P., and Lodish, H. F. (1995) The specificity of the transforming growth factor beta receptor kinases determined by a spatially addressable peptide library. *Proc. Natl. Acad. Sci. U. S. A. 92*, 11761–11765.
(16) Till, J. H., Annan, R. S., Carr, S. A., and Miller, W. T. (1994) Use of synthetic peptide libraries and phosphopeptide-selective mass spectrometry to probe protein kinase substrate specificity. *J. Biol. Chem. 269*, 7423–7428.
(17) Miller, C. J., and Turk, B. E. (2016) Rapid Identification of Protein Kinase Phosphorylation Site Motifs Using Combinatorial Peptide Libraries. *Methods Mol. Biol. Clifton NJ 1360*, 203–216.
(18) Shah, K., Liu, Y., Deirmengian, C., and Shokat, K. M. (1997) Engineering unnatural nucleotide specificity for Rous sarcoma virus tyrosine kinase to uniquely label its direct substrates. *Proc. Natl. Acad. Sci. U. S. A. 94*, 3565–3570.
(19) Ulrich, S. M., Kenski, D. M., and Shokat, K. M. (2003) Engineering a “Methionine Clamp” into Src Family Kinases Enhances Specificity toward Unnatural ATP Analogues. *Biochemistry (Mosc.) 42*, 7915–7921.
(20) Knebel, A. (2001) A novel method to identify protein kinase substrates: eEF2 kinase is phosphorylated and inhibited by SAPK4/p38delta. *EMBO J. 20*, 4360–4369.
(21) Amanchy, R., Zhong, J., Molina, H., Chaerkady, R., Iwahori, A., Kalume, D. E., Grønborg, M., Joore, J., Cope, L., and Pandey, A. (2008) Identification of c-Src tyrosine kinase substrates using mass spectrometry and peptide microarrays. *J. Proteome Res. 7*, 3900–3910.
(22) Huang, S.-Y., Tsai, M.-L., Chen, G.-Y., Wu, C.-J., and Chen, S.-H. (2007) A Systematic MS-Based Approach for Identifying in vitro Substrates of PKA and PKG in Rat Uteri. *J. Proteome Res. 6*, 2674–2684.
(23) Xue, L., Wang, W.-H., Iliuk, A., Hu, L., Galan, J. A., Yu, S., Hans, M., Geahlen, R. L., and Tao, W. A. (2012) Sensitive kinase assay linked with phosphoproteomics for identifying direct kinase substrates. *Proc. Natl. Acad. Sci. U. S. A. 109*, 5615–5620.
(24) Holt, L. J., Tuch, B. B., Villén, J., Johnson, A. D., Gygi, S. P., and Morgan, D. O. (2009) Global analysis of Cdk1 substrate phosphorylation sites provides insights into evolution. *Science 325*, 1682–1686.
(25) Coba, M. P., Pocklington, A. J., Collins, M. O., Kopanitsa, M. V., Uren, R. T., Swamy, S., Croning, M. D. R., Choudhary, J. S., and Grant, S. G. N. (2009) Neurotransmitters Drive Combinatorial Multistate Postsynaptic Density Networks. *Sci. Signal. 2*, ra19-ra19.
(26) Yu, Y., Anjum, R., Kubota, K., Rush, J., Villen, J., and Gygi, S. P. (2009) A site-specific, multiplexed kinase activity assay using stable-isotope dilution and high-resolution mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A. 106*, 11606–11611.
(27) Troiani, S., Uggeri, M., Moll, J., Isacchi, A., Kalisz, H. M., Rusconi, L., and Valsasina, B. (2005) Searching for Biomarkers of Aurora-A Kinase Activity: Identification of in Vitro Substrates through a Modified KESTREL Approach. *J. Proteome Res. 4*, 1296–1303.
(28) Morandell, S., Grosstessner-Hain, K., Roitinger, E., Hudecz, O., Lindhorst, T., Teis, D., Wrulich, O. A., Mazanek, M., Taus, T., Ueberall, F., Mechtler, K., and Huber, L. A. (2010) QIKS – Quantitative identification of kinase substrates. *PROTEOMICS 10*, 2015–2025.
(29) Kubota, K., Anjum, R., Yu, Y., Kunz, R. C., Andersen, J. N., Kraus, M., Keilhack, H., Nagashima, K., Krauss, S., Paweletz, C., Hendrickson, R. C., Feldman, A. S., Wu, C.-L., Rush,

J., Villén, J., and Gygi, S. P. (2009) Sensitive multiplexed analysis of kinase activities and activity-based kinase identification. *Nat. Biotechnol. 27*, 933–940.
(30) Kunz, R. C., McAllister, F. E., Rush, J., and Gygi, S. P. (2012) A high-throughput, multiplexed kinase assay using a benchtop orbitrap mass spectrometer to investigate the effect of kinase inhibitors on kinase signaling pathways. *Anal. Chem. 84*, 6233–6239.
(31) Müller, A. C., Giambruno, R., Weißer, J., Májek, P., Hofer, A., Bigenzahn, J. W., Superti-Furga, G., Jessen, H. J., and Bennett, K. L. (2016) Identifying Kinase Substrates via a Heavy ATP Kinase Assay and Quantitative Mass Spectrometry. *Sci. Rep. 6*.
(32) Huang, Y., and Thelen, J. J. (2012) KiC assay: a quantitative mass spectrometry-based approach for kinase client screening and activity analysis [corrected]. *Methods Mol. Biol. Clifton NJ 893*, 359–370.
(33) Schulze-Gahmen, U., Upton, H., Birnberg, A., Bao, K., Chou, S., Krogan, N. J., Zhou, Q., and Alber, T. (2013) The AFF4 scaffold binds human P-TEFb adjacent to HIV Tat. *eLife 2*, e00327.
(34) He, N., Liu, M., Hsu, J., Xue, Y., Chou, S., Burlingame, A., Krogan, N. J., Alber, T., and Zhou, Q. (2010) HIV-1 Tat and Host AFF4 Recruit Two Transcription Elongation Factors into a Bifunctional Complex for Coordinated Activation of HIV-1 Transcription. *Mol. Cell 38*, 428–438.
(35) Colaert, N., Helsens, K., Martens, L., Vandekerckhove, J., and Gevaert, K. (2009) Improved visualization of protein consensus sequences by iceLogo. *Nat. Methods 6*, 786–787.
(36) de Hoon, M. J. L., Imoto, S., Nolan, J., and Miyano, S. (2004) Open source clustering software. *Bioinforma. Oxf. Engl. 20*, 1453–1454.
(37) Saldanha, A. J. (2004) Java Treeview—extensible visualization of microarray data. *Bioinformatics 20*, 3246–3248.
(38) Ubersax, J. A., and Ferrell, J. E. (2007) Mechanisms of specificity in protein phosphorylation. *Nat. Rev. Mol. Cell Biol. 8*, 530–541.
(39) Chalkley, R. J., Baker, P. R., Medzihradszky, K. F., Lynn, A. J., and Burlingame, A. L. (2008) In-depth Analysis of Tandem Mass Spectrometry Data from Disparate Instrument Types. *Mol. Cell. Proteomics 7*, 2386–2398.
(40) Rush, J., Moritz, A., Lee, K. A., Guo, A., Goss, V. L., Spek, E. J., Zhang, H., Zha, X.-M., Polakiewicz, R. D., and Comb, M. J. (2005) Immunoaffinity profiling of tyrosine phosphorylation in cancer cells. *Nat. Biotechnol. 23*, 94–101.
(41) Feng, X.-H., and Derynck, R. (2005) SPECIFICITY AND VERSATILITY IN TGF-β SIGNALING THROUGH SMADS. *Annu. Rev. Cell Dev. Biol. 21*, 659–693.
(42) Valenciano, A. L., Ramsey, A. C., and Mackey, Z. B. (2015) Deviating the level of proliferating cell nuclear antigen in Trypanosoma brucei elicits distinct mechanisms for inhibiting proliferation and cell cycle progression. *Cell Cycle 14*, 674–688.
(43) Valenciano, A. L., Knudsen, G. M., and Mackey, Z. B. (2016) Extracellular-signal regulated kinase 8 of Trypanosoma brucei uniquely phosphorylates its proliferating cell nuclear antigen homolog and reveals exploitable properties. *Cell Cycle Georget. Tex 15*, 2827–2841.
(44) Ge, J., Xu, H., Li, T., Zhou, Y., Zhang, Z., Li, S., Liu, L., and Shao, F. (2009) A Legionella type IV effector activates the NF-kappaB pathway by phosphorylating the IkappaB family of inhibitors. *Proc. Natl. Acad. Sci. U. S. A. 106*, 13725–13730.
(45) Lesaicherre, M.-L., Uttamchandani, M., Chen, G. Y. J., and Yao, S. Q. (2002) Antibody-Based fluorescence detection of kinase activity on a peptide array. *Bioorg. Med. Chem. Lett. 12*, 2085–2088.

(46) Saunders, A., Core, L. J., and Lis, J. T. (2006) Breaking barriers to transcription elongation. *Nat. Rev. Mol. Cell Biol. 7*, 557–567.
(47) Schüller, R., Forné, I., Straub, T., Schreieck, A., Texier, Y., Shah, N., Decker, T.-M., Cramer, P., Imhof, A., and Eick, D. (2016) Heptad-Specific Phosphorylation of RNA Polymerase II CTD. *Mol. Cell 61*, 305–314.
(48) Buratowski, S. (2009) Progression through the RNA Polymerase II CTD Cycle. *Mol. Cell 36*, 541–546.
(49) Egloff, S., and Murphy, S. (2008) Cracking the RNA polymerase II CTD code. *Trends Genet. 24*, 280–288.
(50) Chapman, R. D., Heidemann, M., Hintermair, C., and Eick, D. (2008) Molecular evolution of the RNA polymerase II CTD. *Trends Genet. 24*, 289–296.
(51) Phatnani, H. P., and Greenleaf, A. L. (2006) Phosphorylation and functions of the RNA polymerase II CTD. *Genes Dev. 20*, 2922–2936.
(52) Meinhart, A., Kamenski, T., Hoeppner, S., Baumli, S., and Cramer, P. (2005) A structural perspective of CTD function. *Genes Dev. 19*, 1401–1415.
(53) Czudnochowski, N., Bösken, C. A., and Geyer, M. (2012) Serine-7 but not serine-5 phosphorylation primes RNA polymerase II CTD for P-TEFb recognition. *Nat. Commun. 3*, 842.
(54) Xu, Y.-X., and Manley, J. L. (2007) Pin1 modulates RNA polymerase II activity during the transcription cycle. *Genes Dev. 21*, 2950–2962.
(55) Xu, Y.-X., Hirose, Y., Zhou, X. Z., Lu, K. P., and Manley, J. L. (2003) Pin1 modulates the structure and function of human RNA polymerase II. *Genes Dev. 17*, 2765–2776.
(56) Bielska, A. A., and Zondlo, N. J. (2006) Hyperphosphorylation of Tau Induces Local Polyproline II Helix. *Biochemistry (Mosc.) 45*, 5527–5537.

Figures:

Figure 1.

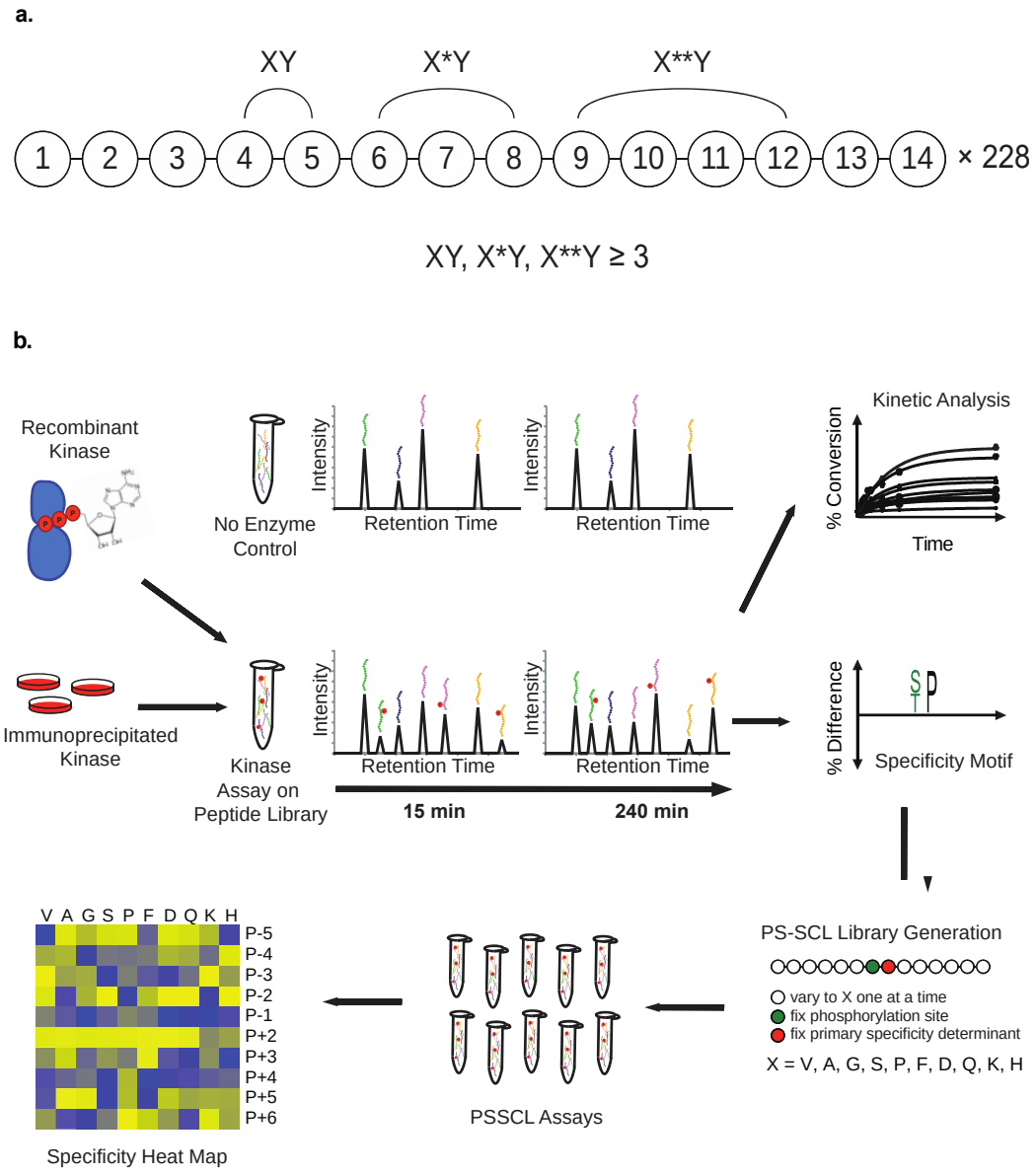


Figure 1. Design of a physico-chemically diverse library of synthetic peptides and execution of the Multiplex Substrate Profiling by Mass Spectrometry (MSP-MS) method for kinases. (a) Library design, incorporating three or more representations of every combination of every neighbor (XY) and near neighbor (X*Y, X**Y) pair of amino acids. All amino acids were included except for Cys, and nLeu was

substituted for Met.[6] (b) Depiction of the MSP-MS assay for kinases. 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.

Figure 2.

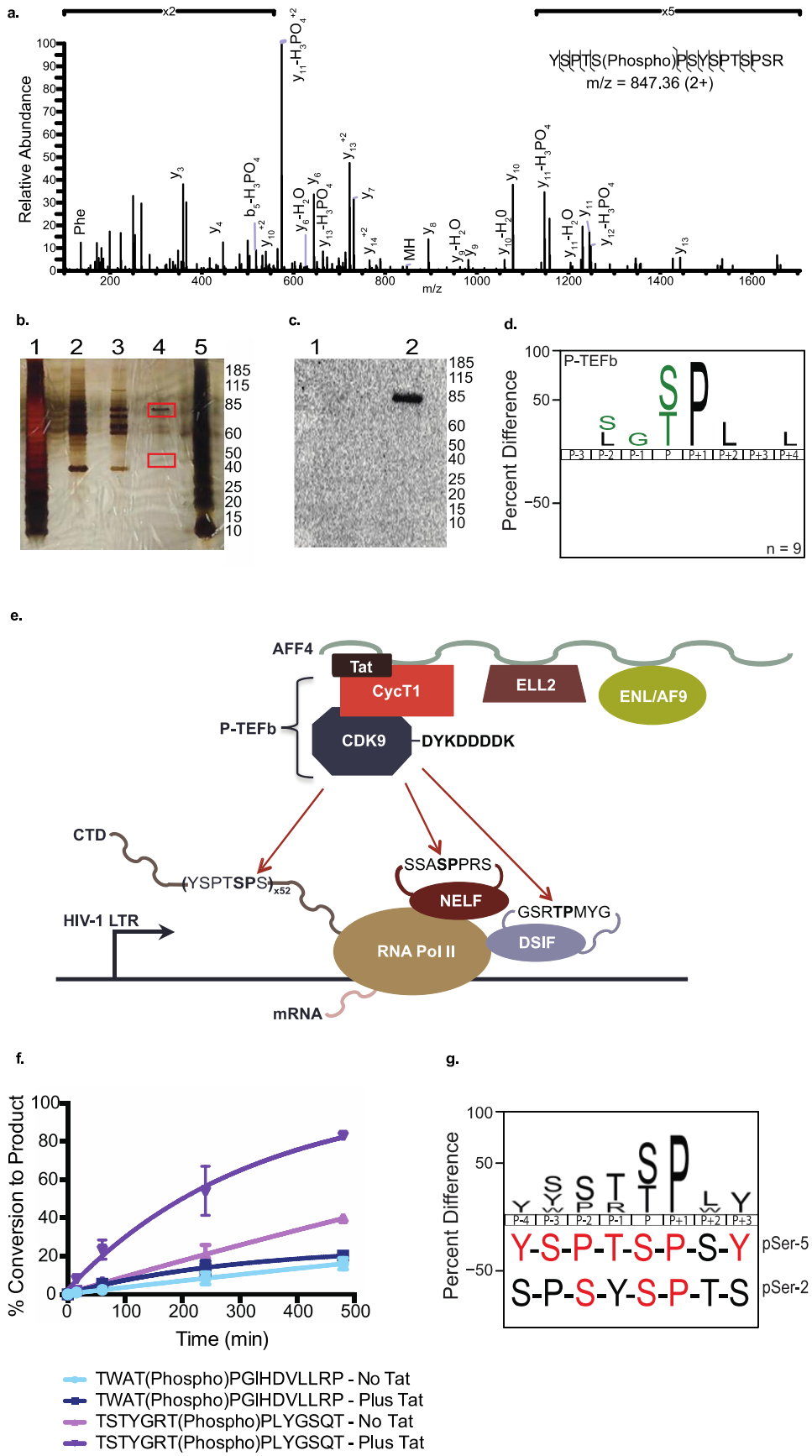
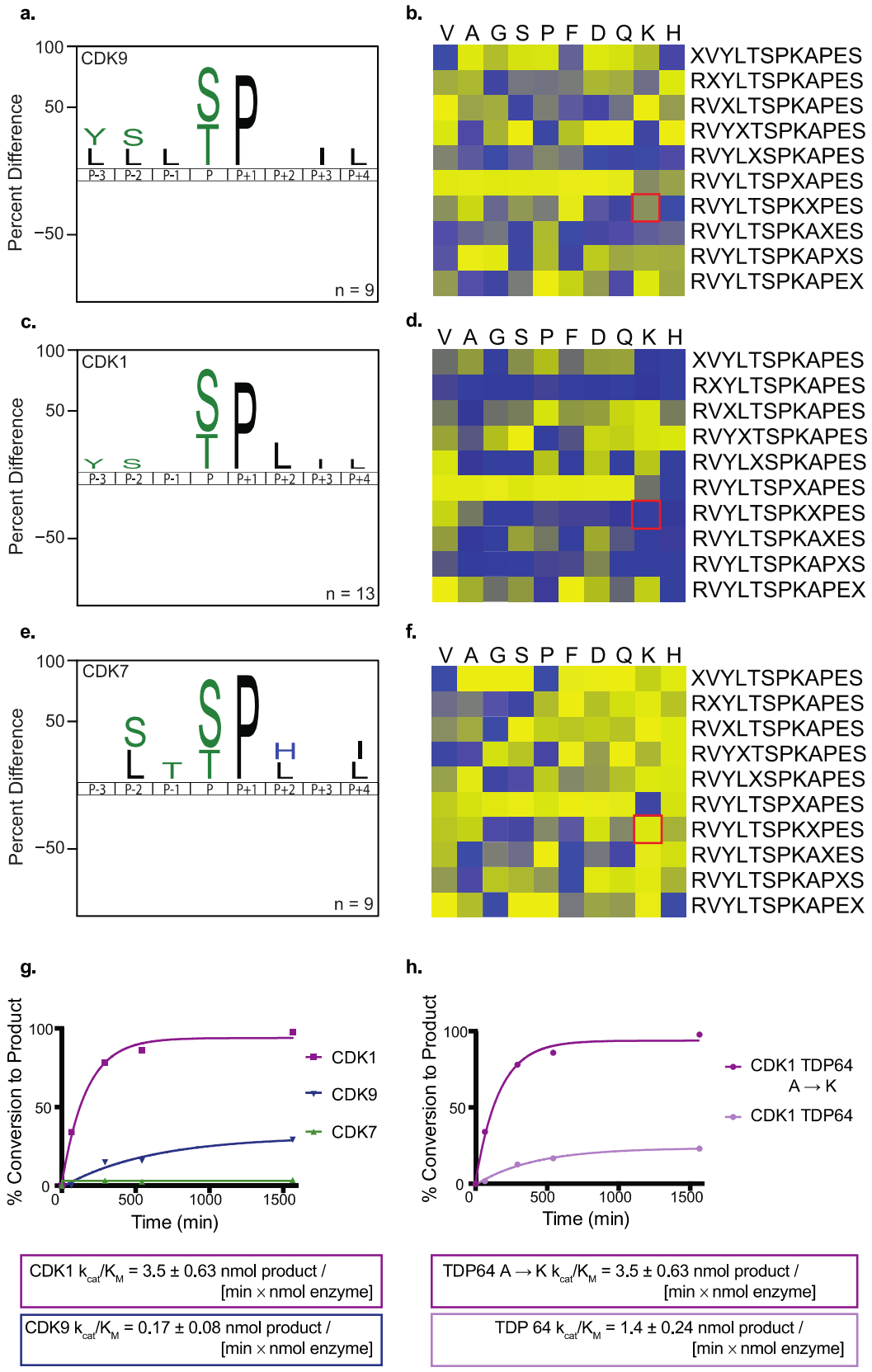

a.
x2
x5
YSPTS(Phospho)PSYSPTSPSR
m/z = 847.36 (2+)
Relative Abundance
m/z
b.
1 2 3 4 5
185
115
85
60
50
40
25
20
15
10
c.
1 2
185
115
85
60
50
40
25
20
15
10
d.
P-TEFb
Percent Difference
100
50
−50
P-3 P-2 P-1 P P+1 P+2 P+3 P+4
n = 9
e.
AFF4
Tat
CycT1
ELL2
ENL/AF9
P-TEFb
CDK9
DYKDDDDK
CTD
(YSPTSPS)x52
SSASPPRS
NELF
GSRTPMYG
DSIF
HIV-1 LTR
RNA Pol II
mRNA
f.
% Conversion to Product
Time (min)
TWAT(Phospho)PGIHDVLLRP - No Tat
TWAT(Phospho)PGIHDVLLRP - Plus Tat
TSTYGRT(Phospho)PLYGSQT - No Tat
TSTYGRT(Phospho)PLYGSQT - Plus Tat
g.
Percent Difference
P-4 P-3 P-2 P-1 P P+1 P+2 P+3
Y-S-P-T-S-P-S-Y
pSer-5
S-P-S-Y-S-P-T-S
pSer-2

Figure 2. MSP-MS analysis of recombinant and immunoprecipitated P-TEFb and the HIV-1 Tat-P-TEFb interaction. (a) Spectra of P-TEFb phosphorylation on the RNA Pol II CTD peptide, confirming Ser5 as the phosphorylation site. (b) 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. (c) Anti-Cyclin T1 western blot of FLAG-CDK9 immunoprecipitation from HEK293 cells stably expressing FLAG-CDK9 reveals that Cyclin T1 is the primary co-immunoprecipitant of FLAG-CDK9. Lane contents: 1 – ladder, 2 – unbound to beads, 3 – FLAG eluent. (d) Substrate profile from immunoprecipitated P-TEFb, which aligns with recombinant P-TEFb substrate specificity. (e) 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. (f) 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. (g) Weighted specificity motif generated using the sequences of peptides that Tat increased the k_{cat}/K_M values for by more than two-fold. Alignment of this motif with RNAP II CTD Ser5 and RNAP II CTD Serine2 as the phosphorylation site indicates that Tat selectively increases P-TEFb catalytic efficiency towards peptides that are similar to CTD Ser5.

Figure 3.



CDK1 k_{cat}/K_M = 3.5 ± 0.63 nmol product / [min × nmol enzyme]

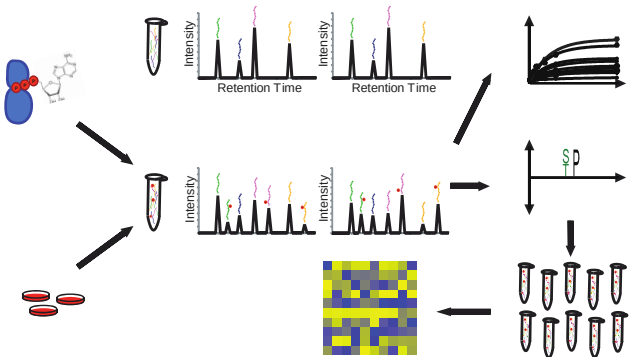
CDK9 k_{cat}/K_M = 0.17 ± 0.08 nmol product / [min × nmol enzyme]

TDP64 A → K k_{cat}/K_M = 3.5 ± 0.63 nmol product / [min × nmol enzyme]

TDP 64 k_{cat}/K_M = 1.4 ± 0.24 nmol product / [min × nmol enzyme]

Figure 3. Delineation of Cyclin-Dependent Kinase substrate specificities. (a) Substrate profile from MSP-MS analysis of CDK9/Cyclin T1. (b) 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. (c) Substrate profile from CDK1/Cyclin B. (d) Heat map generated using maximum percent conversion values from CDK1 PS-SCL experiment. (e) Substrate profile from CDK7/Cyclin H/MNAT1. (f) Heat map generated using maximum percent conversion values from CDK7 PS-SCL experiment. (g) Differential phosphorylation of the "P+3 K" peptide, RVYLT**S**PKKPES, 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).
(h) CDK1 phosphorylation of PS-SCL scaffold peptide RVYLT**S**PK**A**PES and P+3 A → K peptide RVYLT**S**PK**K**PES. Alteration of the P+3 residue makes the peptide a 2.5-fold better substrate.

for TOC only


Intensity
Retention Time
Intensity
Retention Time
Intensity
Intensity
SP