HIGH-THROUGHPUT IDENTIFICATION OF ONCOGENIC TYROSINE KINASE SUBSTRATE PREFERENCES TO IMPROVE METHODS OF DETECTION

by

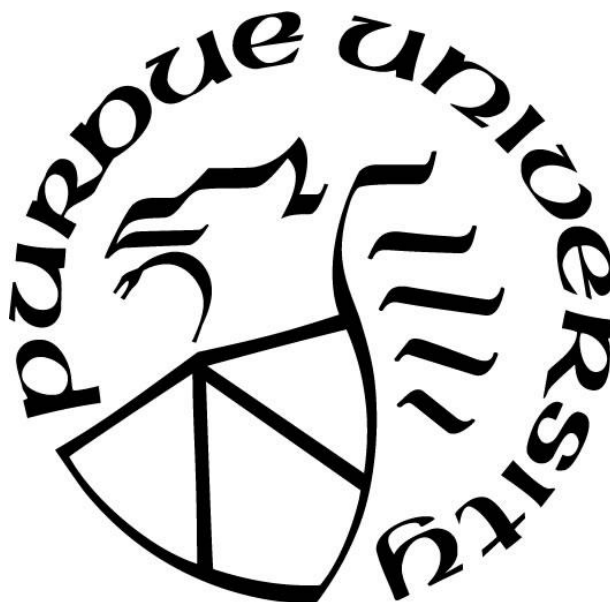**Minervo Perez**

A Dissertation

Submitted to the Faculty of Purdue University
In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

Department of Medicinal Chemistry and Molecular Pharmacology

West Lafayette, Indiana

December 2018

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Laurie L. Parker, Co-Chair

> Biochemistry, Molecular Biology and Biophysics

Dr. Weiguo Andy Tao, Co-Chair

> Biochemistry

Dr. Sophie Lelievre

> Basic Medical Sciences

Dr. Casey J. Krusemark

> Medicinal Chemistry and Molecular Pharmacology

Approved by:

> Dr. Jason R. Cannon
>> Head of the Graduate Program

I dedicate this dissertation to my friends, family and loved ones for never losing faith in me throughout my life.

ACKNOWLEDGMENTS

First and foremost, I would like to thank my advisor, Dr. Laurie L. Parker, for the wonderful seven years I spent in her guidance. Thank you for believing in me, for providing a nurturing environment and challenging me to become a better scientist. I could not have asked for a better mentor.

I would like to thank my thesis committee members for their continued support while at Purdue University and while I finished my studies at the University of Minnesota. I would like to thank each one of you for your time, feedback and contributions to my work. Dr. Andy W. Tao, I would like to thank you for accepting to become my thesis co-chair when I moved to Minneapolis.

My lab mates in the Parker lab, past and present, thank you for the wonderful experiences that we shared together. You guys provided a wonderful environment and support system for me to complete my work.

Special thanks to Dr. Andrew M. Lipchik, you continue to inspire me day in and day out. Thank you for taking me under you wing and providing me with the foundation to succeed in the completion of my graduate studies. Your words of encouragement and advice were invaluable during the last seven years.

Special thanks to John Blankenhorn, Erica Romero, Lindsay Breidenbach, Naomi Widstrom and Kevin Murray for your contributions to my dissertation work. I wish all of you the best of luck with your future endeavors.

I would like to thank my previous mentors and teachers at Benedictine University for believing in me: Dr. Allison Wilson, Dr. Monica Tischler and Dr. Ed Ferroni. Additionally, I would thank the Youth Leadership Academy in Elgin, IL for the leadership training and support they provided me during my time in middle school and high school.

I would like to thank the PULSE community at Purdue University for the constant support during my time at the West Lafayette campus. Special thanks to the University of Minnesota graduate program and community for making me feel welcomed and included during my time in Minneapolis.

To my family, my mother, Marcela, Rob, Robert and Allison—thank you for your support throughout my time away from family.

To Katelyn, I could not have done this without your constant support and words of encouragement. Thank you for everything you have done for me while we were apart from each other. I look forward to what the future will bring us.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Author: Perez, Minervo. PhD
Institution: Purdue University
Degree Received: December 2018
Title: High-Throughput Identification of Oncogenic Tyrosine Kinase Substrate
      Preferences to Improve Methods of Detection
Committee Chair: Laurie Parker; Weiguo Andy Tao, Co-Chairs

The use of computational approaches to understand kinase substrate preference has been a powerful tool in the search to develop artificial peptide probes to monitor kinase activity, however, most of these efforts focus on a small portion of the human kinome. The use of high throughput techniques to identify known kinase substrates plays an important role in development of sensitive protein kinase activity assays.

The KINATEST-ID pipeline is an example of a computational tool that uses known kinase substrate sequence information to identify kinase substrate preference. This approach was used to design three artificial substrates for ABL, JAK2 and SRC family kinases. These biosensors were used to design ELISA and lanthanide-based assays to monitor in vitro kinase activity. The KINATEST-ID pipeline relies on a high number of reported kinase substrates to predict artificial substrate sequences, however, not all kinases have the sufficient number of known substrates to make an accurate prediction.

The adaptation of kinase assay linked with phosphoproteomics technique was used to increase the number of known FLT3 kinase variant substrate sequences. Subsequently, a set of data formatting tools were developed to curate the mass spectrometry data to become compatible with a command line version of the KINATEST-ID pipeline modules. This approach was used to design seven pan-FLT3 artificial substrate (FAStides) sequences. The pair of FAStides that were deemed the most sensitive toward FLT3 kinase phosphorylation were assayed in increasing concentrations of clinically relevant tyrosine kinase inhibitors.

To improve the automation of the mass spectrometry data analysis and formatting for use with the KINATEST-ID pipeline, a streamlined process was developed within a bioinformatic platform, GalaxyP. The data formatting tools used to process the FLT3 mass spectrometry data were converted into compatible versions to execute within the GalaxyP

framework. This process was used to design four BTK artificial substrates (BAStide) to monitor kinase activity. Additionally, one of the BAStide sequences was designed in the lanthanide chelating motif to develop an antibody-free activity assay for BTK.

Lastly, a multicolored time resolved lanthanide assay was designed by labeling SYK artificial substrate and a SRC family artificial substrate to measure the activity of both kinases in the same kinase reaction. This highlighted the functionality of lanthanide-based time resolved assays for potential multiplexing assay development.

CHAPTER 1.   INTRODUCTION

1.1   Protein Kinases

Protein kinases regulate a vast number of cellular processes within the cell by playing a role in protein phosphorylation at serine, threonine, or tyrosine residues. To date, 538 genes that encode for 518 protein and 20 lipid kinases have been identified.[1,2] These kinases are primarily divided into 13 subfamilies characterized by the amino acid sequence and catalytic domain structure and consist of receptor or non-receptor kinases. Kinases contain additional functional domains to help regulate kinase interactions and activity. Extra- and/or intracellular machinery drives structural changes within the kinase molecule for protein kinase regulation. Of the 518 known protein kinases, 428 are serine/threonine (S/T) kinases while only 90 are tyrosine (Y) kinases. Protein kinases are categorized into membrane or non-membrane bound subtypes. Membrane-bound receptor kinases play an integral part in cell signaling. Upon ligand binding, the receptor kinases undergo a conformational change that exposes the kinase domain to the cytosol and initiates a signaling cascade.[3,4] Non-receptor kinases are found in the cytosol and act as second messengers to external stimuli.[5] The next section will overview the structure and substrate identification of protein kinases.

1.2   Kinase structure and substrate identification

Protein kinase families share structural features that dictate function. The N and C lobes and the kinase backbone[6] regions play integral roles in kinase activation and inactivation. The N-lobe's five β strands are important for the initial ATP interactions. The first three β-strand sequences are highly conserved and imperative for coordinating the nucleotide base and the phosphates of ATP for catalysis (p-loop).[6,7] The C-lobe's structure is predominantly α-helices and regulates substrate coordination while performing the phosphate transfer (activation loop). The β subdomain of the C-lobe contains the aspartic acid-phenylalanine-glycine (DFG) motif that is involved in coordinating the Mg^{2+}-ATP complex and is responsible for activating or inactivating kinases. Once the tyrosine residue within the activation loop is phosphorylated, the loop is stabilized, and the kinase switches

into its active state (Figure 1-1B).[6,7] The kinase backbone links the N and C lobes through a highly conserved motif. Two hydrophobic residues from both the N and C lobes create the hydrophobic link that assists in coordinating the ATP adenine ring for its interaction with the F-helix of the C-lobe. Though the catalytic domains of S/T and Y kinases are structurally similar, the depth of the catalytic domain drives tyrosine kinase specificity, which contain a deeper binding pocket to accommodate the bulky phosphor accepting tyrosine residue. Kinase specificity toward S/T or Y is also regulated by charge and hydrophobicity of the catalytic domain.[8]
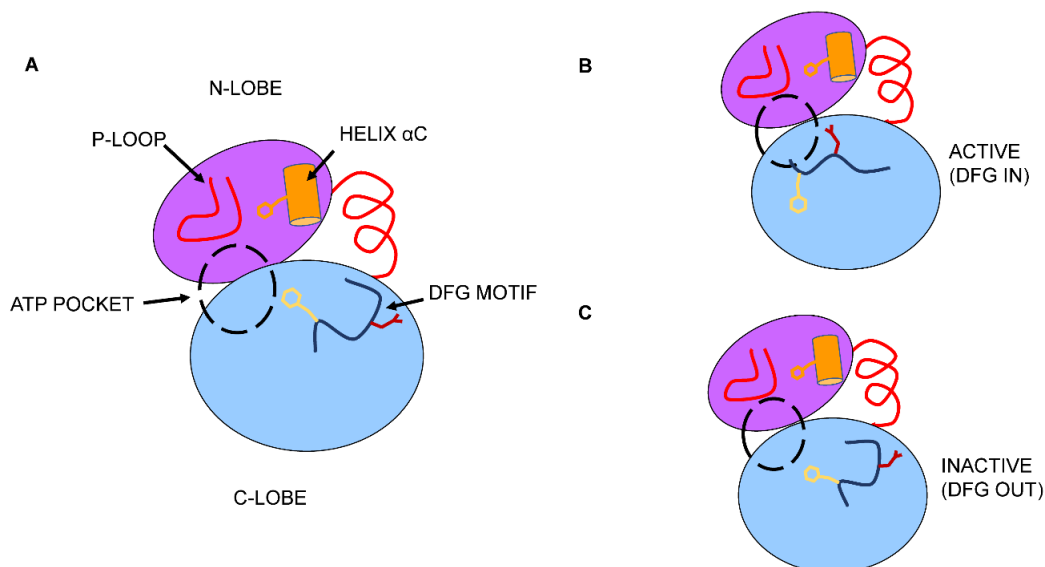
Figure 1.1. Overview of non-RTK kinase activation or deactivation.

1.2.1 Tyrosine kinases

Tyrosine kinases are classified into two categories: receptor tyrosine kinases and non-receptor tyrosine kinases. The activation of tyrosine kinases depends on their localization and their non-catalytic domains. Consequently, kinases within the same superfamily will contain structurally and functionally similar non-catalytic domains. The following subsections will briefly describe structure and activation of receptor tyrosine kinases (RTK) and non-receptor tyrosine kinases (non-RTK).

1.2.2   RTK

The twenty RTK subfamilies, which are activated by growth factors or ligands.[3] RTKs consists of extracellular, transmembrane and cytosolic domains. There are up to eighteen extracellular domains that vary between RTK subfamilies. The extracellular domain regulates RTK activation and can be activated via four primary modes: RTK-ligand crosslinking, RTK-ligand crosslinking with receptor rearrangement, multi-mode crosslinking and multi-mode adaptor crosslinking (Figure 1-2). The first mode of activation, RTK-ligand crosslinking, occurs when a dimeric ligand complex binds the receptor, driving two RTK molecules together to undergo autophosphorylation e.g. TRKA-Nerve growth factor complex (Figure1-2A). The KIT-Stem cell factor (SCF) ligand complex is an example of the second mode of RTK-ligand crosslinking with receptor rearrangement (Figure 1-2B). Under this mode, upon the binding of two SCF ligands, receptor homodimerization is initiated. Consequently, two Ig-like (D4 and D5) domains arrange the kinase molecule for autophosphorylation. Third, the FGFR-FGF-heparin protein complex is an example of multi-mode binding. To activate the FGF receptor (FGFR), the receptor-receptor, receptor-ligand, receptor-heparin and ligand-heparin must all be in direct contact (Figure 1-2C). The final RTK ligand mediated activation mode is the multi-mode adaptor crosslinking (Figure 1-2D). The Epidermal Growth Factor Receptor (EGFR) and insulin RTK subfamilies are examples. In this mode of activation, the extracellular domains of EGFR are folded upon themselves to maintain the receptor an autoinhibited state. One molecule of inhibited receptor dimerizes with another, driving a conformational shift to allow for dimeric ligand binding to the extracellular domain and subsequently initiate autophosphorylation.[3,9]
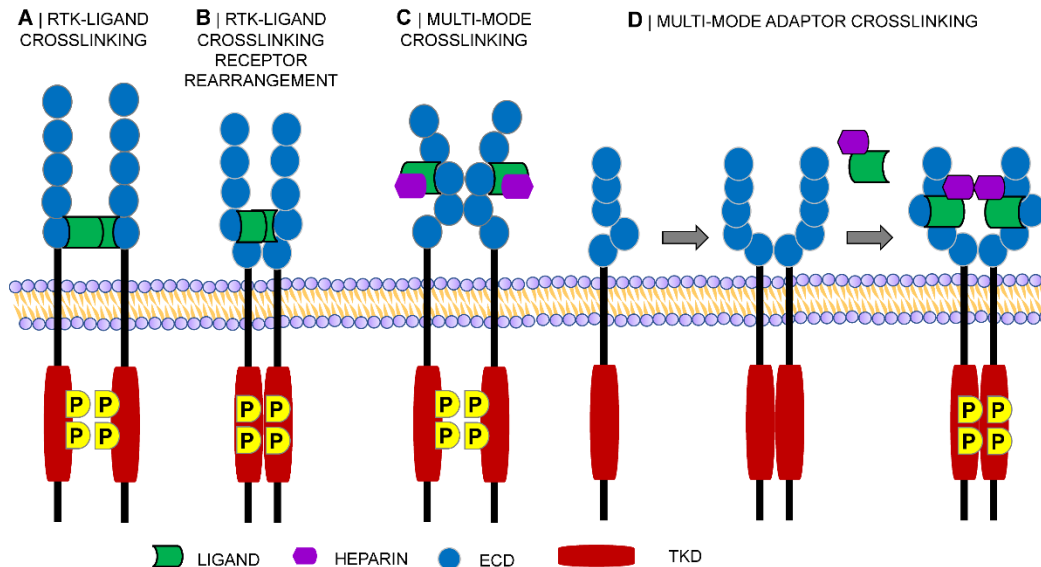
Figure 1.2 Tyrosine Kinase Activation

1.2.2.1 Non-RTKs

Unlike RTK, non-RTK contain non-catalytic domains that regulate the interaction between the kinase and its substrate. One example of non-catalytic domains controlling kinase activity is the SRC family of kinases (SFKs), which are heavily involved in cellular signaling. Aside from their kinase domain, the SFKs contain SRC homology 2 (SH2) and 3 (SH3) domains. SH2 domains bind phosphorylated tyrosine residues, while SH3 domains bind specific proline-rich sequences.[8] SFKs are often localized to the cell membrane through myristoylation and/or palmitoylation of the N-terminus.[10] However, not all non-RTK families contain SH2 or SH3 domains. They can contain other non-catalytic domains that regulate lipid binding, calcium signaling, binding partner interactions and/or cellular localization. The ABL kinase, which contains F-actin and DNA-binding domains to regulate its localization, is one example.[10]

Autophosphorylation or phosphorylation by another non-RTK regulate non-RTKs. For example, the SRC SH2 domain binds to the phosphorylated Tyr-527 (CSK) to hold the kinase in its inactive state.[8] Dephosphorylation of Tyr-527 allows the SH2 and SH3 domains to engage their intended target or specificity sequence (Figure 1-3).[8,10] Subsequently, autophosphorylation of Y416 in the activation loop leads to an active

kinase.[8,10] Phosphorylation at Y527 by CSK is a regulatory mechanism to control SRC catalytic activity.
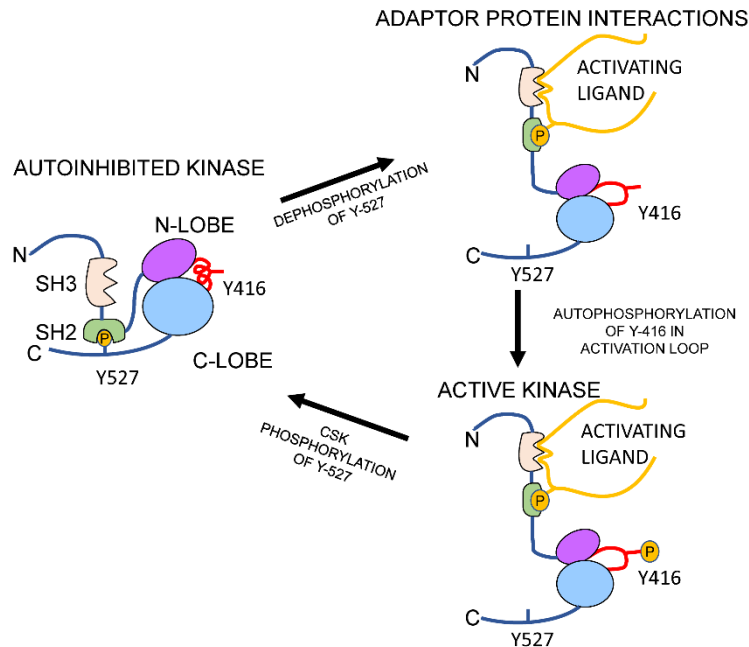
Figure 1.3 Non-receptor protein kinase regulation

1.2.3   Methods for determining kinase substrate specificity

Tyrosine kinases play a large role in many cellular processes, and identification of their biological substrates elucidates their involvement in specific signaling pathways. The influence of the amino acids flanking the phosphorylation site within a substrate was identified over four decades ago.[8] However, a biological substrate does not always contain the optimal sequence flanking the phosphorylation site[11] because other processes regulate a kinase's localization and function. Therefore, identifying the determinants of substrate specificity for particular kinases has been the subject of intensive research.[8] These discoveries led to the acceptance of *in vitro* kinase assays as a tool used to identify and validate a kinase's sequence preference. *In vitro* high-throughput screens using the genomes of simple organisms, such as yeast and mammals, have attempted to identify kinase substrates at the protein and peptide level.[12] These high-throughput methods for rapid identification of kinase substrates include genetically encoded libraries (phage display, mRNA-fusion and cDNA display peptide libraries), synthetic peptide libraries

(positional scanning, oriented and one bead-one-peptide libraries), computational screens, and mass spectrometry assays.[13] While these methods have viable applications, they also contain caveats that limit their practical application. This section will briefly overview the strengths and weaknesses of these techniques.

1.2.3.1   Genetically-encoded peptide libraries

A widely used approach for identifying a protein's binding affinity or substrate preference is genetically encoded libraries. This process involves genetically altering a simple model system such as worms, flies or yeast to either knockdown or overexpress the kinase of interest to identify interacting partners that are then validated *in vitro*.[12,14] Further technological advancement led to the use of genetically encoded peptide libraries (Phage display, mRNA fusion peptide libraries, cDNA display libraries and DNA/PNA programmed synthetic libraries) in an *in vitro* setting.[12,14–16] This approach has been an invaluable tool for identifying biologically relevant kinase substrates and kinase specificity.[14] However, these approaches are time and labor intensive as they require further steps such as polymerase chain reaction amplification and DNA/RNA sequencing to identify the phosphopeptide sequence. Once the sequence is identified, it undergoes a final validation step to determine if the peptide is a kinase substrate. These approaches are prone to contain false positive sequence identification that can be induced by support (phage) or identifier (DNA/RNA) interaction with the kinase of interest.

1.2.3.1.1   Phage display libraries

Phage display libraries have been successful in identifying protein kinase substrates and their specificity.[12,14,17] This process involves cloning cDNA into phage-expressing vectors to create a bacteriophage expressing a random or specific peptide sequence.[17] *E. coli* are then infected with the bacteriophage.[17] After undergoing multiple rounds of selection, the phage library is amplified, and *E. coli* are infected with phages to create a phage library stock. [12,14,17] Subsequently, the phage library is plated on nitrocellulose membranes or combined to form a combinatorial mixture of multiple phages.[12,14] The phages are then assayed *in vitro* with the kinase of interest and radioactive or non-

radioactive ATP.[12,14] Phosphorylation is detected through radiography or antibody-based phosphorylation detection.[12,14]
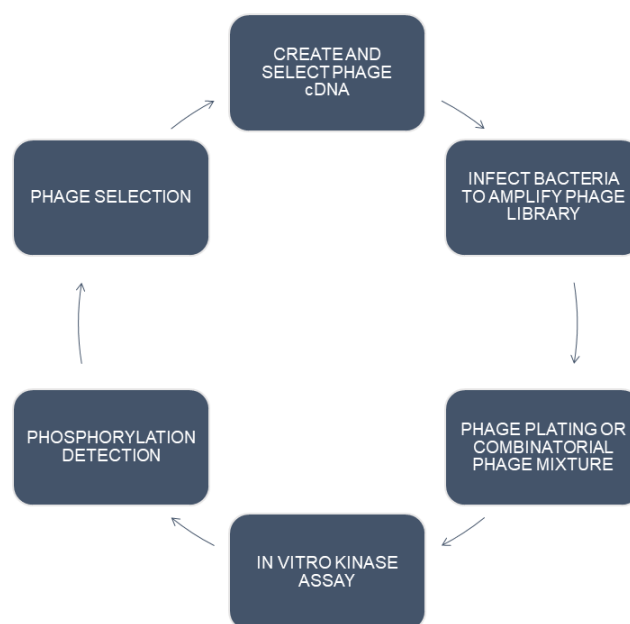
Figure 1.4. Phage display peptide library concept overview

The positive hits from the phage library undergo multiple rounds of selection and kinase treatment (Figure 1-4).[12,14] Finally, the lead sequences are synthesized and assayed *in vitro* in solution to validate them as kinase substrates.[12,14] While there has been success using phage display peptide libraries, biological and technological limitations constrain them. Low copy number of the target peptide displayed on the phage surface limits the process to a binary outcome preventing phosphopeptide quantification.[13,17] The multiple rounds of phage validation needed are time and labor intensive. Finally, the large quantity of recombinant kinase required can be a cumbersome process itself.

1.2.3.1.2 mRNA/cDNA fusion peptide libraries

mRNA/cDNA fusion peptide libraries have been used *in vitro* to discover kinase substrates. One example is ABL kinase.[18–20] Conceptually, the process is similar to phage display. The generation of the mRNA-peptide fusions occurs *in vitro* allowing for a 10,000-fold higher throughput process than phage display.[19] This technique requires an mRNA/cDNA-peptide or protein fusion through a puromycin linkage (mRNA/cDNA

sequence of target peptide-puromycin linker-solid support). mRNA-peptide fusion libraries have been predominantly used to identify DNA binding or protein-protein interacting peptides with length ranges between 10 and 110 amino acids.[19] dsDNA is transcribed into mRNA and ligated enzymatically to a DNA-puromycin oligonucleotide to form a complex encoding for the target peptide sequence.[19] Following purification, the mRNA/cDNA-peptide fusions are incubated with an anti-phosphotyrosine antibody to remove molecules that non-specifically bind to the antibody before incubation with the kinase of interest.[19] Following kinase incubation, the mRNA/cDNA-peptide fusions are incubated with an anti-phosphotyrosine antibody to enrich for phosphorylated sequences.[19] Lastly, the hit mRNA/cDNA-peptide fusion is amplified, and the dsDNA is sequenced to identify the phosphopeptide sequence.[19] mRNA-peptide fusion libraries are a valuable tool for *in vitro* identification of kinase substrates, but the use of puromycin linker/selection marker limits the experimental conditions.[21–23] Puromycin mimics the ribosome substrate and prevents translation. The mRNA-peptide fusion technique is primarily limited by the translational efficiency of the translation process and degradation by ribonucleases.[22,23]
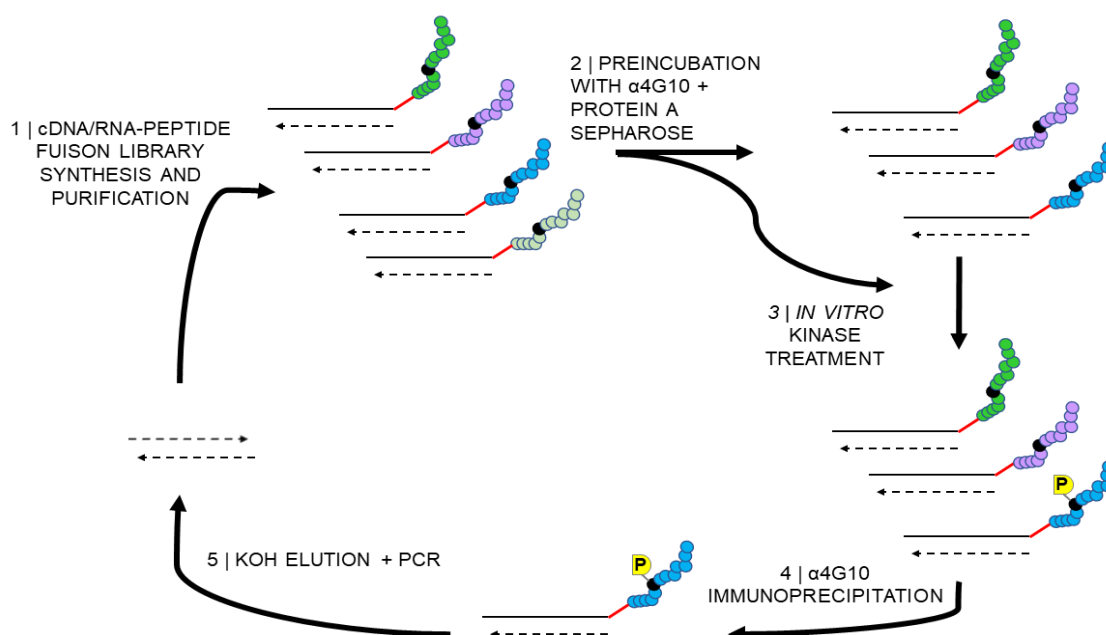
Figure 1.5. DNA/RNA concept overview

cDNA-fusion libraries are conceptually similar when compared to the mRNA-fusion technique. Unlike the mRNA approach, the cDNA-fusion approach has been shown

to be an experimentally robust and effective way to identify kinase substrates.[22,23] The results obtained through this technique, however, do not translate effectively into an *in vivo* setting because the sequences identified through this process do not account for additional post-translational modifications that might play a role in kinase substrate identification.

1.2.3.2 Synthetic peptide libraries
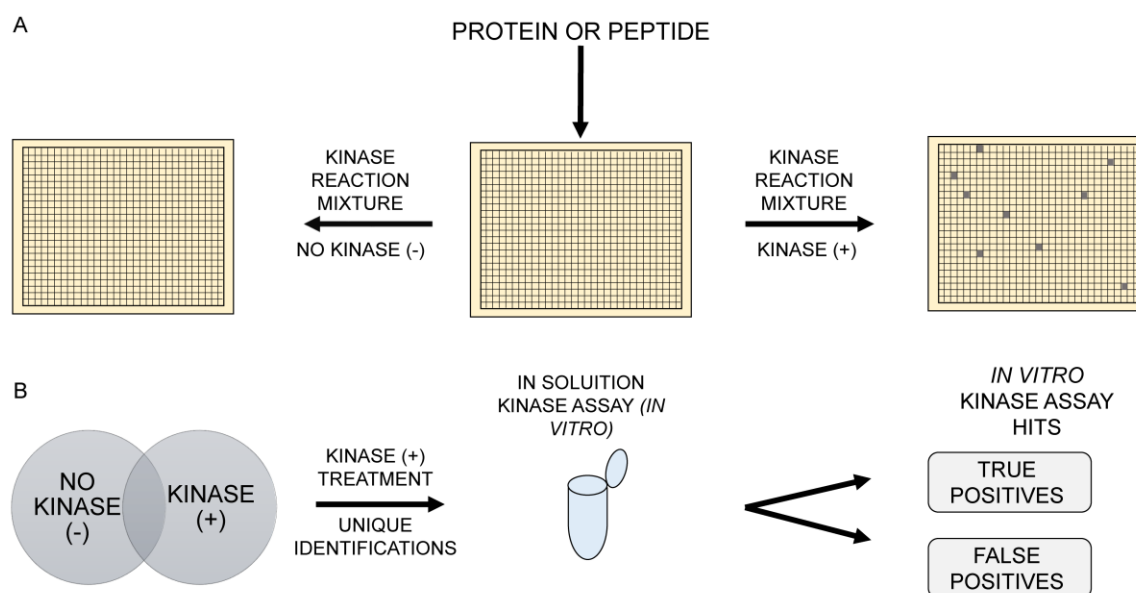
Figure 1.6. SPCL overview

Synthetic peptide libraries are a powerful tool that have been used to identify kinase substrate preference in yeast and human proteomes.[12,14,24–26] Unlike genetically encoded peptide libraries, which are limited to natural amino acids, synthetic peptide libraries can incorporate un-natural amino acids and derivatives. Generally, synthetic libraries are divided into positional scanning synthetic peptide combinatorial libraries (PS-SPCL) or one bead one compound (OBOC) categories. Original iterations of PS-SPCL were bound to membrane solid supports allowing for rapid peptide library synthesis and discovery of a kinase's preferred substrate preference.[24,27] Although synthesizing PS-SPCL on membrane solid supports enables fast library synthesis, this approach introduced false positive identifications by limiting the kinase's mode of biding to the peptide when compared to in solution *in vitro* assays. Thus, PS-SPCL phosphorylated peptides require a subsequent in-solution validation step. Technological advancements, however, have developed non-

membrane bound PS-SPCL that allow for the libraries to be aliquoted for multiple uses decreasing the rate of false positive sequence detection. PS-SPCL are limited by their lack of commercial availability and hit sequences require multiple rounds of validation. Similar to PS-SPCL, OBOC peptide libraries are not limited to natural amino acids; peptoid libraries have been applied to identify protein-protein interactions, and PTMs have been incorporated in library design.[24] OBOC, although useful, comes with its own set of limitations. Peptides are synthesized on beads using chemical linkers that can produce steric hindrance and prevent kinase activity.[24] Additionally, a single bead can contain up to 10^{13} copies of a single peptide that can increase the local concentration and lead to detection of poor peptide sequence.[24] Furthermore, the peptide sequence found on each bead is not readily identifiable and requires Edman sequencing or mass spectrometry sequence   identification.
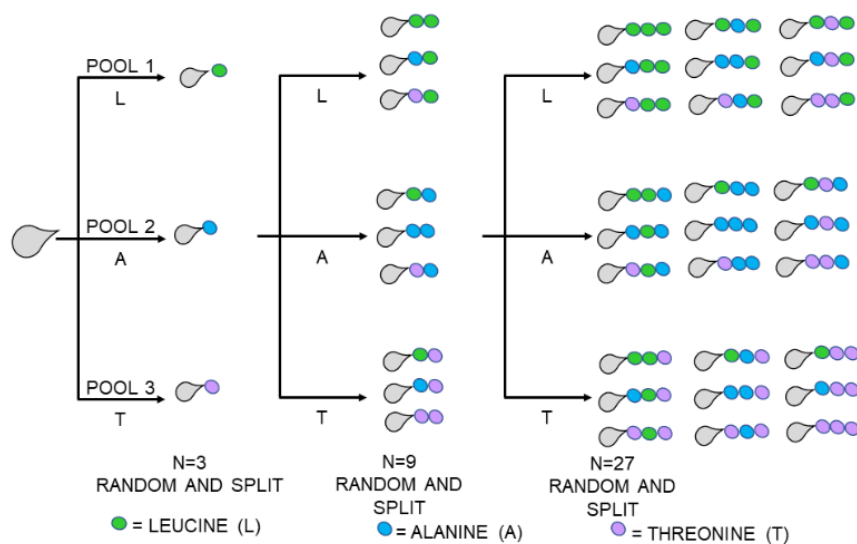
Figure 1.7. One bead one peptide library generation overview

1.2.3.3   DNA/PNA programmed synthetic libraries

The peptide nucleic acid (PNA) method combines mRNA/cDNA microarrays, fluorescently labeled antibodies and mix encoding.[15] Each peptide in the microarray is engineered in the following scaffold: Fluorescein-PEG spacer-F-Q-X-X-Y-X-X-I-K-PEG linker-PNA, where X denotes any amino acid except tyrosine to prevent multiple

phosphorylation sites within the same peptide.[15] The PNA microarray is then treated *in vitro* with the kinase of interest.[15] Phosphorylation is detected with an anti-phosphotyrosine (4G10) antibody followed by an anti-mouse secondary antibody conjugated with Cy3.[15] The ratio of Cy3/Fluorescein is used to normalized varying peptide concentrations.[15] Additionally, the Cy3/Fluorescein ratio can be used for relative phosphopeptide quantification and subsequent extrapolation of kinase sequence preference.[15] The PNA encoded peptide library has successfully recapitulated the preferred substrate motif for ABL kinase.[15] Additionally, this process has been successfully implemented to elucidate the sequence preference for the HER2 and VEGFR2 RTKs.[15] This process for interrogating a kinase's preferred substrate sequence does not limit the creation of the peptide library to natural amino acids, as a previous study showed that D-amino acids can be incorporated into this method.[15] Although PNA/DNA-encoded peptide libraries have been used to identify kinase substrate preference, they also have limitations. The PNA/DNA-encoded library technique does not address other PTMs that might regulate kinase substrate interactions. Additionally, as implemented, this process contains fixed amino acids at position -4 (F), position (-3), position 3 (I) and at position 4 (K), with respect to the phosphotyrosine. This can bias the preferred sequence motif of a kinase of interest. Subsequent *in vitro* validation of the hits is necessary to verify them as actual substrates of the target kinase.
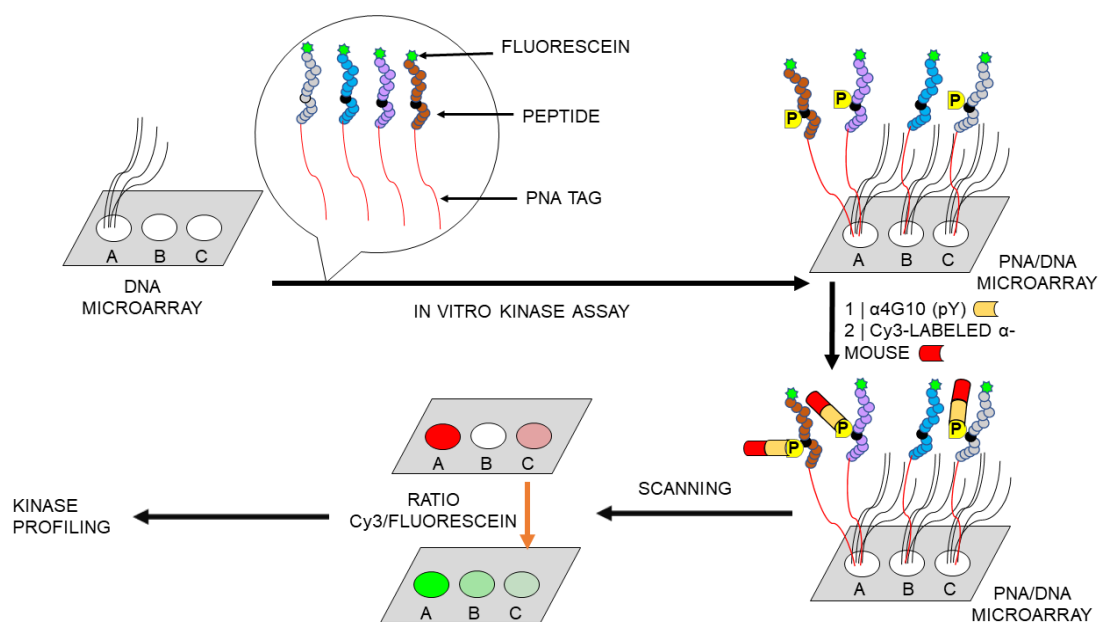
Figure 1.8. PNA peptide library concept overview

1.2.3.4   Computational methods for predication of kinase specificity

The advancement of kinase substrate identification technologies through either biological (phage display or DNA encoded) or synthetic (PS-SPCL or OBOC) libraries has paved the way to a new approach: deciphering the optimal kinase sequence through computational biology. To centralize the information produced using biological or synthetic libraries, large proteomic databases have been developed to store sequence information for enzyme substrates and protein-protein interactions.[28–31] These databases allow researchers to filter and retrieve data that are relevant to specific searches such as kinase substrates. Moreover, these databases use consensus motif analysis tools to predict kinase substrates. [28–31] This approach uses the amino acid sequences found in the database and identifies the reoccurring amino acids and their respective positions relative to the phosphorylation site. Additionally, machine learning algorithms can be developed using this functionality to identify the kinase's preference for specific amino acids within the substrate motif. Algorithm development utilizes sequence similarity-based clustering and higher-order machine learning tools such as statistical and mathematical modeling. The

sections below provide a brief overview of the strengths and weaknesses of these approaches.

1.2.3.4.1 Position specific scoring matrices (PSSM)

A computational motif analysis approach centers on pattern-based detection and pioneering. Examples include Scansite and Motif-X.[32,33] PSSMs are useful tools for identifying consensus sequences from the high-throughput screens described previously.[33–35] The "hit" sequences are used as references to identify similar sequences within the proteome, which can then be merged to define consensus sequences. This has identified substrate sequences for many protein kinases.[33–35] These algorithms assign a numeric value to each amino acid flanking the phosphosite (S, T or Y) to create a percent scoring threshold. The percent scoring threshold is used to make a binary decision of sequence similarity and matches within the proteome. The PSSM approach, however, is highly dependent on the training dataset from peptide libraries to identify a clear kinase preference.

1.2.3.4.2 Sequence similarity-based clustering

Sequence similarity-based clustering is a powerful tool to identify sequence homology between protein or peptide sequences.[36] Sequence clustering uses blocks substitution matrices (BLOSUM) derived from protein sequence information. Common similarity matrices are BLOSUM62, -80 and -100.[36] The matrix name is defined by the threshold used to generate the underlying similarity table, or in other words, the proteins used to generate the matrix share 62, 80 or 100 percent sequence similarity.[36,37] Examples of BLOSUM-based clustering are phosphorylation set enrichment analysis (PSEA) and PostMOD used to identify consensus kinase substrate motifs. PSEA is a derivative of gene set enrichment analysis (GSEA) that was used to identify substrate sequences for individual kinases, kinase families or kinase groups.[38] The PostMOD creators used the Phospho.ELM database as input for their matrix and compared it to competing algorithms. Unlike PSEA, the creator of PostMOD developed a new matrix scoring system to reduce background noise, improve performance over the BLOSUM62-based matrices and create a simpler alternative to high order machine learning methods.[39] Algorithm performance metrics were

calculated for PostMOD and competing algorithms to validate PostMOD as a viable alternative to complex modeling techniques such as machine learning algorithms,[39] which are discussed in the next section. Although these approaches have proven useful in identifying kinase substrates, they depend on known substrate motifs to match an unknown peptide sequence as a substrate. In addition, sequence similarity-based approaches are limited to binary predictions that determines if a peptide sequence is or is not a peptide substrate and does not consider substrate turnover efficiency.

1.2.3.4.3 High order machine learning (HOML) models

Machine learning algorithms have been employed to identify enzyme substrates *in vitro* and *in vivo*. These tools use multiple layers of information between kinase and protein interaction, kinase localization and protein function to learn underlying relationships that govern kinase activity.[14,40] A machine learning algorithm can be developed using artificial neural networks (ANN) (NETPHOSK),[40,41] hidden markov models (HMM) (KINASEPHOS), probabilistic models (Bayesian decision theory, BDT), support vector machine (SVM)[42] (PredPhospho),[43] random forest (Phosphopredict)[40,44] and conditional random fields (CRF) (CRPhospho).[45]

Using ANNs requires the input of known sequence information from phosphorylation sites (true positive) and non-phosphorylation sites (true negative).[40,41] The network is then trained to weigh the positive input over the negative input to produce an efficient working model.[40,41] However, the multiple layers or dimensions applied to ANNs have revealed their struggles with large datasets.

HMMs work slightly different in where the input database contains both positive and negative substrate sequences of a kinase, which the algorithm separates into two different lists. HMMs create multiple models based on the data and selects the most accurate one using the k-fold cross validation and leave-one-out model.[46] HMMs do not perform homology reduction and do not account for redundant sequences.[46] Additionally, HHMs suffer from overfitting the model to the data. Furthermore, HMM models do not include structural information when identifying the best performing model.[46]

Probabilistic models use high order statistical probability (PhosD)[47] or Bayesian statistics (BDT)[40] to develop algorithms that calculate the probability of the four flanking

amino acids with respect to the phosphosite (S, T or Y) for each sequence within the input dataset's (positive and negative) substrates. Additionally, PhosD incorporates kinase domain and substrate protein interactions to determine a protein as a biological substrate.[47] However, when incorporating protein-protein interactions between kinase and protein substrate from proteomic databases (Phosphosite, Phospho.ELM or UniProt) the model's ability to correctly identify a biological substrate (recall) decreases. [47] Probabilistic models are highly dependent on the number of positive substrate datasets to make accurate kinase predictions.

SVM-based models identify kinase substrates at the kinase family level through statistical learning theory (PredPhospho).[43,44] SVM models treat the input dataset as individual points in space to identify a mathematical model that separates the data into two or more classifiers. Following construction of the mathematical function, SVM models use the data points (support vectors) closest to the mathematical function as reference points to identify the appropriate mathematical function. The mathematical function is determined when the distance between support vector and the mathematical function are symmetric. Once the mathematical function is identified, it can then be used to classify unknown data points into a predicted class (Figure 1-9B).

Conditional random fields-based models focus their algorithms on the positive substrate datasets to understand and recognize the properties of the phosphosite flanking amino acids to identify a distinct pattern for a positive substrate.[45] The algorithm then uses the negative dataset to establish a threshold that determines if a sequence is a positive kinase substrate.[45]

The random forest machine learning algorithm was used to develop the Phosphopredict algorithm. This algorithm identifies weak relationships ("trees") or associations to create an overarching strong relationship ("forest").[40] The Phosphopredict algorithm uses sequence information, secondary structure, predicted solvent availability and biological functions to generate a strong algorithm.[40] The developers of Phosphopredict used the Phospho.ELM, PhosphoSitePlus and UniProt databases to predict new and uncharacterized phosphorylation sites for twelve kinase families.[40]

HOML models can account for influential relationships between reoccurring amino acids within a substrate sequence, protein-protein interactions and cellular localization to

identify known and novel kinase substrates.[40,42,43,47] HOML models have been useful tools to identify kinase substrate sequences and biological substrates, but they still have their limitations. Machine learning algorithms are predominantly used to identify phosphorylation sites within biologically relevant kinase substrates. This can be a limiting factor when predicting substrates for poorly studied protein kinases. Furthermore, as the complexity of the machine learning algorithm increases so does the ability to understand and decipher the decision making of the model.[41]
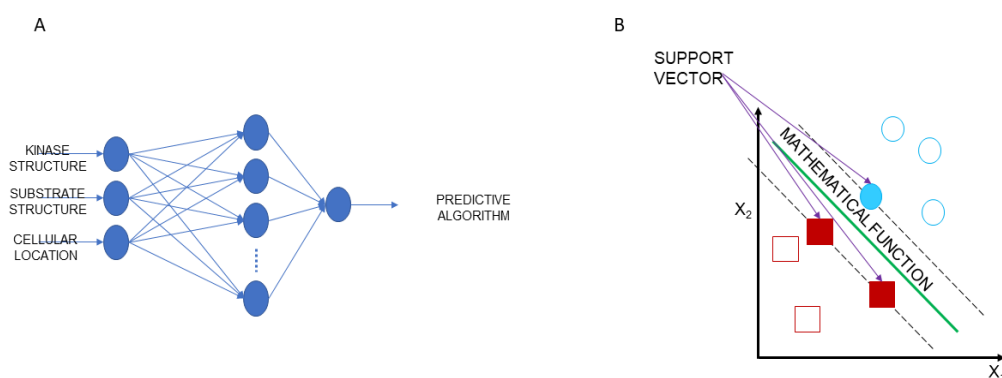
Figure 1.9. High order machine learning concept for artificial neural networks (ANN; left) and support vector machines (SVM; right)

1.3 The role of protein kinases in diseases

Protein kinases regulate most of the cellular processes that are essential for cell survival, proliferation and programmed cell death. Disturbances in these signaling pathways lead to signaling abnormalities that upregulate or suppress vital cellular processes leading to (but not limited to) skeletal, immunological, neurological, autoimmune diseases and cancer.[48,49] The original hallmarks of cancer were termed: sustaining proliferative signaling, evading growth suppressors, activating invasion and metastasis, limitless replicative potential, sustained angiogenesis and evading aptoptosis.[50] The hallmarks of cancer have been updated to included emerging (deregulating cellular energetics and avoiding immune destruction) and enabling characteristics (genome instability/mutations and tumor-

promoting inflammation).[51] Protein kinases contribute to the transformation of a cell population and the gain of the cancer hallmarks (ie., evading growth suppressors, limitless replicative potential, evading apoptosis, etc.),[50,51] allowing cells to circumvent normal inhibitory processes and promote disease.

1.3.1 The involvement of the human kinome in cancer

1.3.1.1 Specific examples of kinases implicated in cancer

Protein kinases are critical for sustaining proliferative singling, evading growth suppressors, resisting cell death, inducing angiogenesis and deregulating cellular metabolism in cancerous cells.[48–51] For example, the platelet derived growth factor receptor (PDGFR) is overexpressed in cancers to promote sustained proliferative signaling.[48,51,52] The PDGFR signaling pathway is an example of how cancer cells become over reliant on unmutated cellular processes in tumorigenesis. Another example of a kinase implicated in cancer is epidermal growth factor receptor (EGFR). EGFR plays a role in evading growth suppression by regulating the phosphoinositol-3-kinase (PI3K), AKT, and mammalian target of rapamycin (mTOR) pathway in cancerous cells.[3,51] Cancer cells hijack cellular pathways to supplement their environment with nutrients and dispose of waste. The most common pathway used to drive angiogenesis is the vasculature endothelial growth factor A (VEGF-A) and thrombospondin-1.[51,53,54] The VEGF-A ligand is often overexpressed in cancer cells. VEGF-A induces the activity of three members of the VEGF receptor tyrosine kinase family to initiate and regulate pro-angiogenesis pathways like ERK, AKT and MAPK.[53] Further, protein kinases are heavily involved in the regulation of another cancer hallmark, alteration of cellular metabolism. Such kinases include but are not limited to AMP-activated protein kinase (AMPK), liver kinase B1 (LKB1) and PI3K.[55,56] Upregulation of signaling pathways, however, is not the only driver of tumorigenesis. Genetic abnormalities can produce oncogenic mutations in protein kinases resulting in hyperactive kinases.

1.3.1.2 Protein kinase abnormalities in cancer

The implication of kinases in cancer has led to a sequencing effort to identify reoccurring kinase mutations across an array of cancers.[57–60] These studies have uncovered

many known and unknown mutations.[57–60] The most common protein kinase abnormalities are chromosomal perturbations and function-altering mutations. Chromosomal perturbations occur because of a chromosomal instability leading to translocation of two chromosomes. The most well-known example is the translocation of chromosomes 9 and 22 forming the "Philadelphia chromosome." In this translocation, the breakpoint cluster region (BCR) and the non-RTK Abelson tyrosine kinase (ABL) genes are fused.[57–60] This encodes for the fusion protein BCR-ABL which has a constitutively active catalytic kinase domain that drives chronic myelogenous leukemia progression.[18,61,62] Function-altering mutations are subdivided into substitution or deletion categories. These alter protein kinase activity by introducing an amino acid substitution (gain-of-function) or deletion of segments within protein sequence of the protein kinase.[60] For example, amino acid substitution of the amino acid residue located four C-terminal residues from the glycine of the DFG motif (N terminus-XX**DFG**XXX**Y**XX**-**C terminus) within the activation loop of the kinase domain leads to constitutively active kinase.[59,60] Examples of deletion gain-of-function mutations are observed in BRAF, EGFR and HER2 where a four to nine amino acid segment, near the activation loop, is lost and results in a constitutively active kinase.[59,63]

1.3.1.3   Understudied protein kinases and their role in disease progression

Oncogenic cells become dependent on signaling pathways regulated by protein kinases to drive tumorigenesis and drug resistance. Examining the PhosphoSitePlus kinase substrate database, the majority of research efforts have been focused on a small subset of the human kinome (Figure 1-10).[64] The dataset contained 385 kinases with over 13,701 combined kinase substrate identifications from over ~340,000 combined "records" or studies, which have been identified through mass spectrometry experiments by researchers or Cell Signaling Technology.[64] Currently, the PhosphoSitePlus database does not have any reported substrates for 133 of the 518 protein kinases.[64] High-throughput sequencing studies are unveiling new insight into the role protein kinases play in cancer and diseases, which will clarify if the 133 kinases with no reported substrates are involved in tumorigenesis. Identifying the protein substrates for these understudied kinases will

provide insight into their biological roles, and enable development of assays that will help determine if they are viable drug targets.[65]
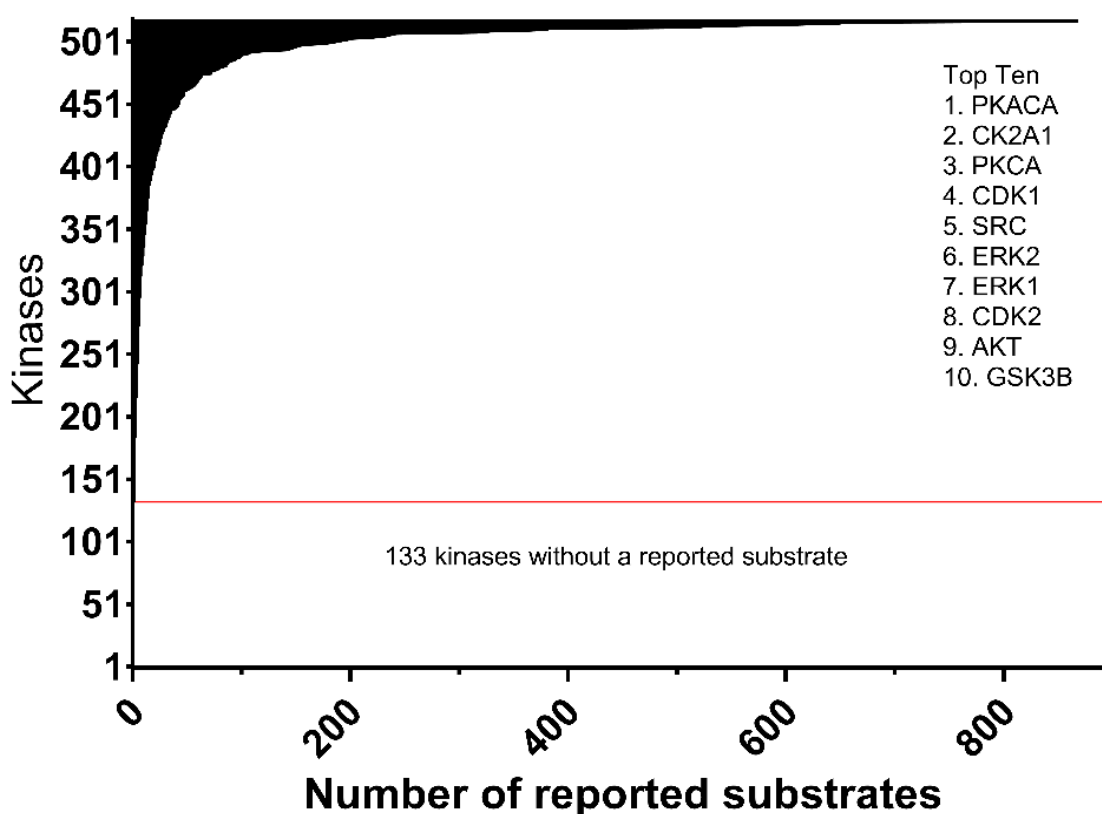
Figure 1.10. Number of reported kinase substrates per kinase

1.3.2 Monitoring kinase activity and drug discovery

The success of imatinib as a treatment for BCR-ABL positive chronic myeloid leukemia has led to an increase in kinase inhibitor development. Since the approval of imatinib in 2001, 37 kinase inhibitors have received Food and Drug Administration (FDA) approval.[66] However, most of the approved kinase inhibitors are not kinase specific and contain non-kinase targets, which have contributed to adverse side-effects and/or drug resistance.[67] Moreover, a large proportion of kinases and their biological roles remain understudied, which highlights the potential to discover new drug targets.

Small molecule inhibitors have been developed to target RTK, however, these inhibitors contain off-target activity that ellicit adverse side effects. Additionally,

resistance can be acquired independent of the RTK target by increased expression of the receptor or the receptor ligand. The initial success and limitations of targeting RTKs has shifted the focus to identifying new pathways that are indirectly involved in tumorigenesis but are cellular processes that oncogenic cells are depend upon such as transcription and immune system regulation.[66]

Examples of current kinase targets that regulate transcription include the cyclin-dependent kinases (CDK7-9, 12 and 13).[66] Immune system regulators include CDK8, colony stimulating factor-1 receptor (CSF1R), PI3K (γ and δ) and the TAM kinases family (AXL, TYRO3 and MERTK).[66] Currently, only CDK 7-9 have a reported substrate peptide available to monitor kinase activity while the rest do not. This is a short list of kinases that are involved in emerging signaling processes involved in cancer drug resistance.[66] Targeting cancers that are addicted to these kinase driven pathways requires effective ways to monitor kinase activity.

1.3.3   The impact of developing efficient kinase substrates to monitor kinase activity

Protein kinases are known to give cancerous cells advantages in cellular process that aid tumorigenesis and evade chemotherapy-induced death.[50,51] This process can be mediated through increased copies of the kinase molecule or function-altering mutations. As described in prior sections, function-altering mutations can be single amino acid substitution, abnormal sequence insertion or deletion of inhibitory protein domains.[57–60] The dependency of cancerous cell on overactive kinase signaling made kinases viable drug targets.

Small molecule inhibitors have been developed to target oncogenic kinases and improve chemotherapy outcomes in cancer patients. These inhibitors contain multiple modes of action to target kinases.[68–70] Small molecules that are efficacious towards wild type kinases may not inhibit mutant kinases.[66,71] Inhibitor resistance can be mediated by a secondary mutational event that prevents the small molecule from binding to the kinase molecule. To combat drug resistance and reduce off-target activity, we must understand how function-altering mutations affect kinase mechanisms that requires an understanding of normal modes of action. [66,71] In addition, kinase inhibitors can be used to target cellular

pathways that promote alternative resistance mechanism such as increased cellular metabolism or immune response.[66]

New kinase inhibitors are needed to target these kinases of interest. Advancements in HTS technology have identified a variety of new kinase inhibitors, however, there is a need to identify specific and selective kinase inhibitors.[72] Determining the preferred peptide sequence that a kinase will phosphorylate is beneficial on many fronts. In the absence of an ideal substrate sequence, polymer substrate sequences have been used to probe kinase activity.[72–75] Polymerized substrates are comprised of a phosphor accepting amino acid residue surrounded by a series of either basic or acidic amino acids.[73] Good substrates can help lower the cost of HTS by requiring less recombinant kinase.[76,77] To identify an efficient kinase substrate, the affinity of a kinase for a substrate (K_m) can be used.[78,79] K_m is defined as half the substrate concentration required to saturate the enzyme that is acting upon it. A substrate containing a low K_m value indicates that a lot less substrate is required to saturate the kinase of interest (V_{max}), while a substrate containing a high K_m value requires a larger concentration of substrate to saturate the kinase.

Polymer substrates have been used successfully in HTS to monitor kinase activity,[73] however, as discussed in earlier sections, some kinases have preferences for uncharged amino acids at specific positions with respect to the phosphorylated residue.[76,80–82] Poor kinase and substrate compatibility leads to low kinase activity on the substrate that affects the dynamic range of assay readout and are discussed in section 1.5. The promiscuity of polymer substrates prevents their use in cellular assays.[74] Identifying the preferred substrate for a kinase of interest can be used to develop efficient kinase substrates for sensitive activity assays. Additionally, understanding kinase substrate preference information can then be leveraged to develop kinase specific peptide substrates to monitor cellular kinase activity.[77,83] In turn, specific kinase substrates can be coupled with fluorescence applications to monitor kinase activity of function-altering mutations by developing kinase specific biosensors, which are discussed in section 1.5.4. Finally, the optimal sequence that a kinase phosphorylates can be used to identify potential sequences in a biological setting that a kinase might phosphorylate,[77] which can be identified using machine learning algorithms. Identifying the preferred sequence that a given kinase phosphorylates can benefit drug discovery by lowering cost of HTS, improve assay

readouts and help elucidate kinase biology through biosensor assays, which can be applied to understand abnormal protein kinase signaling in human diseases.

1.4   Mass spectrometry

Mass spectrometry is the study of analytes based on their mass to charge (m/z) ratio. Mass spectrometry (MS) is an analytical tool that has been used to analyze complex cell lysate mixtures from flies, yeast and mammals (shotgun mass spectrometry).[84] Additionally, MS has been an integral technology to study post-translational modifications (PTM) of proteins (tandem mass spectrometry).[84] In this technique, a liquid chromatography (LC) system is coupled to a mass spectrometer, which contains an ion source, mass detector and mass analyzer. The LC system retains and fractionates a sample mixture for MS analysis. Following sample elution from the LC column the mass spectrometer ionizes the sample using electrospray ionization (ESI) that is primarily used for complex mixtures. Upon ionization, the sample enters the mass analyzer (Q1) where multiple scans a second are performed to determine the most abundant ions in the mixture. The rapid scan rates allow for multiple snapshots of the eluting peptide mixture, and tandem mass spectrometry is used to obtain peptide sequence information.

1.4.1   Tandem mass spectrometry

In a general mass spectrometry experiment, peptides a eluted form the liquid chromatography column, converted into a gas phase and injected into the mass spectrometer were their mass to charge ratio is detected through a mass analyzer.[84,85] Tandem compatible mass spectrometers (ion trap) couple multiple mass analyzers in line to filter, fragment and scan for product ions (Figure 1-11A).[86] In a tandem mass spectrometry experiment, a survey scan (MS1) is performed to identify the ions (parent ions) present in mixture and the information is used to trigger a second scan (MS2) on the detected ions. The MS2 scan can be triggered to fragment all the detected ions in the mixture (top speed) or only trigger an MS2 scan on the most abundant ions (top-N).[86,87] Following the MS1 scan, the selected MS1 parent ions (top-N or top speed) are captured in the ion-trap mass analyzer (Q2) for peptide fragmentation (product ions). The product ions are then scanned in the third mass analyzer (Q3).[86,87] Early iterations of ion trap mass

analyzers were essential for identification of PTM in complex mixtures but their low mass accuracy, scan speed and resolving power limited identifications.[86] The introduction of the orbitrap mass analyzer increased low mass accuracy, scan speed and resolving power that hindered ion trap instruments.[86,88] The mass analyzer in an orbitrap mass spectrometer consists of four components: a quadrupole (Q) mass analyzer, ion-trap, collision cell (coll. cell) and an orbitrap mass analyzer (Figure 1-11B). In a tandem MS experiment performed on orbitrap mass analyzers, following the MS1 scan (Q1), selected ions are moved to the ion-trap for ion accumulation and are then transferred to the collision cell for peptide fragmentation. Subsequently, the fragment ions are returned to the ion trap to be transferred to the orbitrap mass analyzer.[88] The orbitrap mass analyzer contains an inner electrode that provide current that causes the fragment ions to oscillate in the trap, while the outer electrodes create the ion trapping field and detect the oscillating frequency of the fragment ions (Figure 1-11B).[88] The oscillating frequencies are used to identify a fragment ion's m/z values.[88]

Development of the orbitrap mass analyzer technology has transformed how PTM are identified with higher mass accuracy and resolving power.[86,88,89] Although protein phosphorylation is the most common PTM in the cell, phosphorylated peptides (phosphopeptides) are still in low abundance and the added phosphate residue in the phosphopeptides reduces their ability to ionize effectively.[90] Affinity enrichment techniques have been developed to improve identification of phosphopeptide by mass spectrometry and are discussed in the next section.
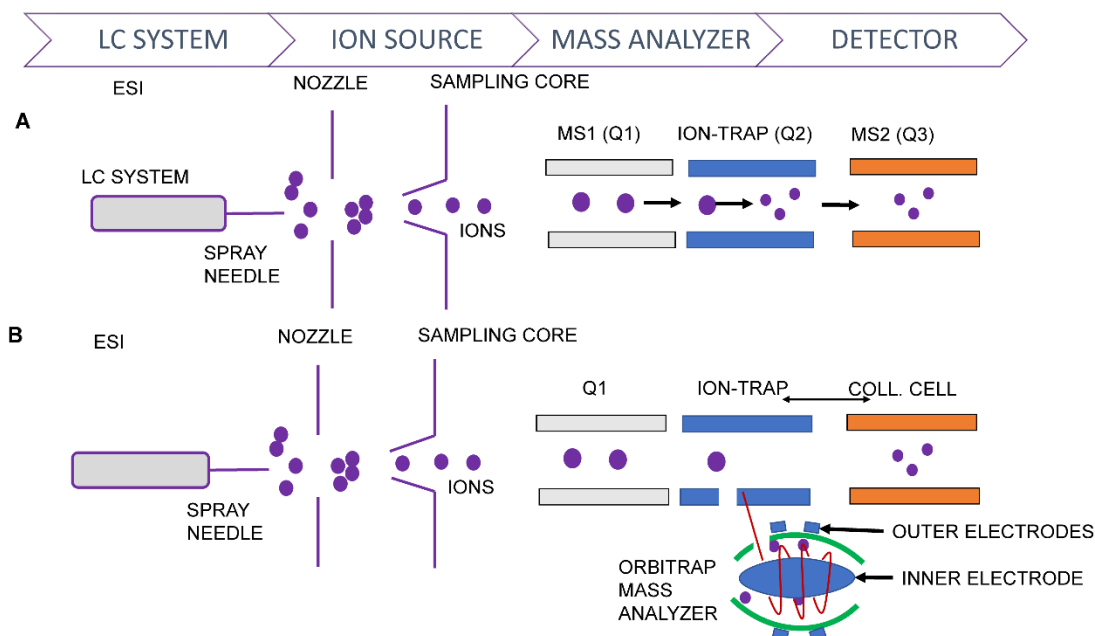
Figure 1.11. Tandem mass spectrometry overview.

Traditional LC-MS/MS concept overview (A) and Orbitrap tandem mass spectrometry overview (B)

1.4.1.1 Affinity enrichments

To overcome the low abundance of phosphopeptides in complex cell mixtures, affinity enrichment techniques such as antibody-based or immobilized metal affinity chromatography (IMAC) and titanium oxide (TiO$_2$) nanopolymer complexes have been employed.[42,90–92] Fe(III) and Ga(III)-based IMAC technology, which were developed to enrich histidine and cysteine-rich peptides, also effectively enrich for phosphopeptides.[42] These techniques are widely used prior to mass spectrometry analysis for phosphopeptide enrichment.[42,90,92,93] Together, the advancements in mass spectrometry and sample preparation have improved identification of phosphorylated peptides (phosphoproteomics) and has become a viable tool detecting new kinase substrates and their phosphorylation sites.[12,90,93–95] The multiple sample handling techniques that are required, however, can introduce variation that is mass spectrometer independent. The most recent introduction of polymer-based metal ion affinity (PolyMAC) capture further advanced the enrichment process for phosphopeptides.[90] The PolyMAC enrichment has been shown to double the amount of phosphopeptides captured and identified from cellular lysate when compared to

TiO$_2$ beads.[90] Additionally, the PolyMAC enrichment has demonstrated greater reproducibility.[90]

1.4.2   Kinase substrate motif identification through mass spectrometry

Phosphoproteomics is a widely employed technique to identify phosphorylated peptides and has been used to identify kinase substrates in a high-throughput manner. Several approaches employing phosphorylated peptide libraries analyzed by mass spectrometry have been developed to evaluate kinase substrate preference.[96–100] These approaches successfully identified new biologically relevant kinase substrates[97] or specific substrate motif for an array of kinases.[96–100] These techniques demonstrated that treating cell lysates or synthetic peptide libraries with recombinant kinases and enriching for phosphopeptides that it was possible to increase the list of known substrates for a kinase of interest.

The KAYAK approach used a focused synthetic peptide library that consisted of known artificial kinase substrate sequences or sequences that are known to be phosphorylated endogenously by certain kinases (S/T or Y) from the Swiss-Prot (UniProt) database. The unphosphorylated peptides were synthesized in a 96-well plate format using unlabeled amino acids. The corresponding phosphorylated control peptide library was synthesized using an isotope labeled proline residue that was coded into all the sequences synthesized.[98] To monitor and quantify the kinase activity of the PI3K and MAPK signaling pathway, cellular stress (starvation, EGF stimulation/inhibition) was induced *in vitro* and in cancer cell models.[98] Following lysate incubation of the unlabeled peptide library, the phosphopeptide standards (isotope labeled proline) were spiked into the quenched lysate reaction mixture and 45 were pooled for phosphopeptide enrichment and mass spectrometry analysis.[98] This approach allowed the authors to observe activation of the PI3K and MAPK pathways. This approach, however, has its limitations. The focused library must contain good substrates for the kinases within the signaling pathway of interest, and it requires prior kinase substrate preference knowledge to make an effective library.[98] Additionally, synthesizing 90 isotope labeled-phosphopeptide control library is cost and time extensive.

Rapid determination of multiple linear kinase substrate motifs by mass spectrometry is another phosphoproteomic approach. In this study, the peptide library was derived from cell lysate. The lysate underwent proteolytic cleavage, phosphatase treatment, and strong cation exchange purification prior to *in vitro* kinase treatment. Subsequently, phosphopeptides are enriched using techniques described above and subjected to mass spectrometry and motif analysis through GrMFPh algorithm.[100] This process has been used to identify the kinase substrate preference for a small panel of protein kinases that included PIM1 and 2, CLK3, DURK1α, HASPIN, CAMKK2β, BMPR2 and the PLK kinase family.[100] Additionally, Gerber and colleagues interrogated the optimal kinase sequence for HASPIN and BMPR2 by mutating the phosphor-accepting residue to either S, T or Y to determine chemical affinity of the kinases towards the phosphor-accepting residue.[100] The experimental results demonstrated that HASPIN prefers to phosphorylate threonine over serine while not phosphorylating tyrosine.[100] BMPR2 phosphorylated all three of the BMPR2-tide derivative sequences but demonstrated preference towards S/T.[100] Although this work successfully elucidated the panel of kinases' preferred substrate motifs, the phosphorylated sequences were not used to identify new biologically relevant kinase substrates.

The serine-oriented human library of peptides (SERIOHL) were generated by using known phosphorylated serine (phosphoserine) centered sequences that were encoded into bacterial plasmids. The peptides are expressed in *Escherichia coli* as a cost efficient way to generate a large quantity of peptide libraries.[99] The SERIOHL-KILR process successfully identified the preferred substrate motif for PKA, PKCβ wild type and mutant kinases. However, this process does not account for endogenous PTM that are found in endogenous cell lysates that could play a role in kinase substrate recognition by non-catalytic kinase domains. Furthermore, the SERIOHL-KILR method is limited to serine-centered sequences.

The KALIP process was designed to identify non-native and native kinase substrates for the non-RTK SYK in DG75 and MDA-MB-231 cell models.[97] The KALIP sample work-up is a two-stage process of identifying non-native and native SYK substrates. To identify biologically relevant targets, cells with or without SYK were lysed, prepared for MS, and subjected to PolyMAC enrichment.[97]

To identify non-native substrates, cells with or without SYK were lysed, trypsin digested, and the digests were subjected to phosphopeptide enrichment (PolyMAC). Following alkaline phosphatase treatment, lysate was treated with recombinant EGFP-SYK and subjected to a second phosphopeptide enrichment prior to tandem mass spectrometry analysis.[97] The KALIP technique discovered five new endogenous SYK substrates and helped define SYK's role in multiple non-B cell receptor signaling pathways.[97] The KALIP process, however, does not account for other endogenous kinases that might phosphorylate the identified substrates found using this technique. This would require further validation through additional biochemical techniques. Additionally, the tryptic digestion creates peptide sequences that do not contain lysine or arginine residues close to the phosphorylation site and if they did, they would be too small for mass spectrometer detection. The results identified would artifactually show that the kinase of interest does not prefer positively charged residues close to the phosphorylation site. The tryptic peptide substrate library would benefit kinases that prefer acidic residues close to the phosphorylation site such as SYK.

1.5   Methods for measuring kinase activity

Monitoring kinase activity in a controlled environment allows for direct interpretation of kinase interacting compounds or inhibitors that alter kinase activity. Molecular biology and biochemical techniques have previously been used to monitor kinase activity using three major approaches: measuring the concentration of the co-substrate ATP, the co-product ADP, and phosphorylation state of the peptide substrate.[101,102] These approaches have their strengths and weaknesses that will be highlighted in the next sections below.
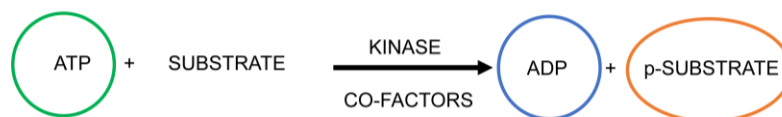
Figure 1.12. Three approaches to infer kinase activity.

Measuring the decrease in ATP concentration following reaction completion (green), ADP production (blue) or the phosphorylation state of the substrate (orange).

1.5.1 ATP-dependent kinase activity measurement

Measuring the concentration of ATP allows for antibody free detection of kinase activity. Commercially available ATP measuring kits use a luciferase reporter assay to measure ATP concentration, which is reduced following kinase reaction that increases the concentration of ADP.[101,103] Briefly, following kinase incubation, the unconverted ATP concentration is reacted with D-luciferin and treated with luciferase to generate oxyluciferin, which can be measured by luminescent measurements.[103] ATP-luciferase assays are highly compatible with HTS and can be used with protein and lipid kinases. Two caveats of these assays are interference by various compounds used with *in vitro* assays such as sodium chloride. Additionally, false negatives induced by luminescent kinase inhibitors.[101] Additionally, these assays require a substantial amount of substrate turnover for adequate decrease in luminescent signal. Ultimately, these approaches are also limited by the rate of substrate turnover and require optimal kinase-substrate compatibility, which is a limiting factor for understudied kinases.

1.5.2 ADP-dependent kinase activity measurement

Measuring the production of ADP has been used in commercially available detection kits to monitor kinase activity (Figure 1-12).[102] These kits measure ADP presence through enzymatic or antibody-based techniques such as luciferase (ADP-Glo) or fluorescence.[102] The ADP-Glo assay is a two-step process. The first step introduces an enzyme (reagent 1) that quenches the kinase reaction and removes residual ATP.[102] The

second step converts the ADP generated during the kinase reaction to ATP by using a second enzyme (reagent 2).[102] Subsequently, ATP levels are measured via a luciferase/luciferin reaction. The luminescence intensity therefore correlates to ADP levels.[101,102] The ADP-Glo measurements can be affected by temperature fluctuations, high concentrations of kinase co-factors (NaCl, $CaCl_2$, Calmodulin, etc.), chemicals (DTT, $NaVO_3$) and detergents (Tergitol-NP-9, Tween-20 and Triton X-100).[102]

The limitations of the luciferase-based assays led to developments of fluorescence-based assays through conjugated antibodies (labeled with lanthanide or horseradish peroxidase) to detect the presence of ADP following the kinase reaction.[101] The assay requires an initial fluorescence measurement prior to initiating the kinase reaction to obtain the positive control (ATP) values that are then compared through endpoint or kinetic measurements.[101] Quantification is achieved with a calibration curve.[101] The ADP generated during the kinase reaction will displace the fluorescently-labeled ADP from the anti-ADP antibody, reducing the signal. Kinase activity is interpreted as inversely proportional to the loss of fluorescent signal. However, the antibody used to detect ADP has non-specific activity towards ATP at millimolar concentrations that limits the range of ATP concentrations used in the kinase reaction.

Extrapolating kinase activity from the concentration of ATP and production of ADP are promising approaches because they do not completely rely on kinase-substrate compatibility. However, these assays contain limitations such as decreased dynamic range if not enough substrate is turned over when measuring ATP concentration or interference from kinase assay cofactors, reducing agents and detergents when monitoring the production of ADP. Thus, the conventional approach of measuring the phosphorylation state of the kinase substrate has been the predominant approach to monitor kinase activity. The following section is an overview of this approach.

1.5.3   Phosphopeptide-dependent kinase activity measurement

The classical methods for detecting kinase activity measure the amount of phosphorylated peptide substrate through radiometric, Enzyme-Linked ImmunoSorbent Assay (ELISA) and lanthanide chelating assays.[80,83,101,104,105] Radiometric assays have the

advantage of being antibody-free, which is a disadvantage for ELISA and lanthanide-based assays in addition to being sequence dependent and are discussed below.

1.5.3.1 Radiometric assays

Radiometric assays were the earliest technologies developed for monitoring kinase activity.[101] These assays utilize radiolabeled ATP at the γ-phosphate to determine enzymatic parameters.[101,104] Radiometric-based assays measure direct kinase activity and do not require sequential steps (ATP/ADP reagent addition or antibody incubation). This process requires the phosphopeptide to contain a net positive charge of 2 or 3 at pH below 2 and is achieved by incorporating an N-terminal lysine residue during peptide substrate design.[104] Following the kinase assay, the peptide substrates are bound to phospho-cellulose paper and washed thoroughly to remove excess ATP to ensure the measured signal comes from the phosphopeptide.

Although this is the "gold standard," this technique has logistical and biological limitations as radiometric assays require special training, separate regulated work areas and costly reagent disposal.[104] Moreover, this technique assumes that the positively charged lysine residues do not interfere with kinase substrate interactions, however, they might alter acidophilic kinase activity. Consequently, these limitations led to the discovery of additional methods to measure phosphopeptide quantities.

1.5.3.2 Antibody-based phosphopeptide detection

Antibody-based detection method is widely used to measure protein phosphorylation in immunoblotting and ELISA-based assays. Following the kinase assay, the quenched samples are added to a 96 or 384-well plate allowing the phosphopeptide to bind to the bottom of the plate via an enrichment tag within the peptide sequence. One common enrichment strategy exploits the high binding affinity between biotin-avidin. Once the peptides are bound, they are incubated with an anti-phosphate primary antibody for S/T/Y residues and subsequently incubated with an anti-secondary IgG that is conjugated with horseradish peroxidase.[83] The phosphopeptide-primary-secondary-HRP complex is then incubated in the presence of HRP substrate and subjected to fluorescence

measurements. The increase in fluorescence signal is directly proportional to the amount of phosphopeptide bound through the biotin-avidin complex.

However, the ELISA-based method contains technical limitations. This technique is most useful for detecting phosphotyrosine residues because of the specificity, sensitivity and commercial availability of anti-phosphotyrosine antibodies. Although anti-S/T antibodies exist, they lack the ability to bind a broad range of phosphorylated serine/threonine containing sequences. Phosphoserine/threonine antibodies are more specific to the amino acid sequence surrounding the phosphorylation site when compared to phosphotyrosine residues.[106]

1.5.3.3 Lanthanide chelating phosphopeptide detection

Proteins in the human proteome are known to coordinate and bind metal ions to perform their proteome function.[105,107,108] One example is the alpha (α)-synuclein protein that binds calcium ions (Ca^{2+}), among others, to aggregate and is involved in the progression of Parkinson's disease.[109] Binding of metal ions is mediated by phosphorylation of Y125, S129, Y133 and Y136 residues within the c-terminal region of α-synuclein. Specifically, the Y125 centered peptide has been shown to bind metal ions efficiently.[109] Calcium contains similar ionic radii to lanthanide metals, which have been shown to displace Ca^{2+} in tissue to alter bone integrity and block Ca^{2+} pumps.[110]

Lanthanides are rare earth metals that are characterized as strong Lewis acids with unique chemical properties that have been exploited to become photoluminescent, bioimaging and biosensing probes.[110–112] One example is the development of terbium-based time-resolved luminescence kinase activity assays (Figure 1-13).[80,81,105,108] To this effect, a SYK artificial substrate (SAStide) was centrally aligned to the α-synuclein calcium binding peptide tyrosine residue (Y125).[105] SYK activity was measured in the presence of piceatannol, a SYK TKI, using the SAStide peptide.[105] The lanthanide binding process was then incorporated into the KINATEST-ID pipeline (chapter 2) to design artificial peptide substrates for the SRC-kinase family (SFAStide), ABL (ABAStide) and JAK2 (JAStide-E) kinase.[80]
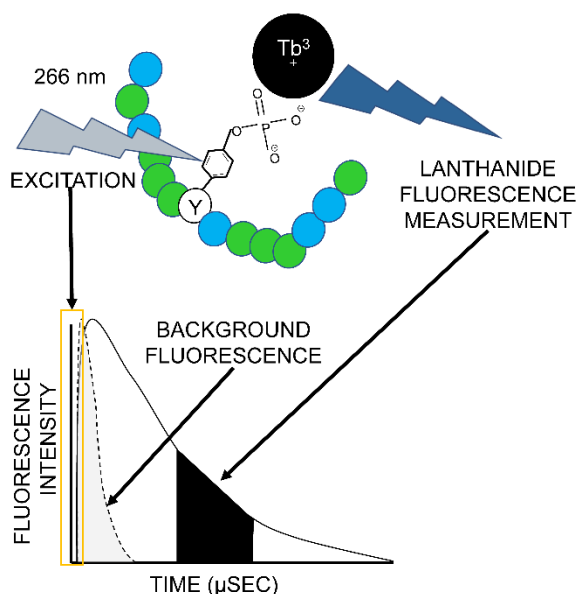
Figure 1.13. Time-resolved luminescence detection of kinase activity through terbium chelation.

Adapted from reference 89.

Lanthanide-based detection assays require certain considerations during peptide substrate design such as requiring acidic residues a position -6, -4, 1, 5 and 6 flanking the tyrosine residue, which are required to form the terbium chelating motif. The fixed acidic residues at their respective positions might reduce the substrate turnover when assayed against the kinase of interest. Lanthanide chelating assays are antibody-free, low cost and high throughput screening compatible that makes them an attractive alternative to monitor tyrosine kinase activity.

1.5.4 Kinase substrates as biosensors to monitor cellular kinase activity

Protein kinase substrates are used on multiple fronts to help elucidate a kinase of interest's biological role. Coupling peptide substrates to fluorescent probes to monitor the in cellular kinase activity has been reported for a variety of kinases.[113–115] This approach can be achieved through two major approaches. Substrates can be genetically encoded or exogenously introduced using a cellular penetrating sequence. Artificial peptide sequences can be coupled to fluorescent probes such as fluorescent proteins, organic dyes or nanoparticles.[111,113–115]

Conjugating kinase substrates to fluorescent probes broadens their use to study biologically relevant kinase activity. Fluorescent proteins (FP) are used to study protein in a biological setting. Self-reporting biosensors have been employed to study kinase activity. This is achieved by genetically encoding a kinase substrate into a reporter scaffold consisting of a photon donor, a phosphorylation sensing probe, substrate and photon acceptor. Examples of photon donors and acceptor are genetically encoded green fluorescent protein and derivatives.[115] Upon substrate phosphorylation, the phosphorylation sensing probe, such as an SH2 domain, folds upon itself and brings the photon donor and acceptor into proximity. Through fluorescence resonance energy transfer (FRET), the photon donor is then excited at a specific wavelength and transfers energy to the photon acceptor. The photon accept emits at specific wavelength and emission energy is measured through fluorescence microscopy.

One example of FRET-based technologies used for in cell assays to monitor kinase activity is A-kinase anchor proteins (AKAR).[113] To develop the AKAR assay, a DNA plasmid encodes for the PKA kinase substrate within a biosensor scaffold and was used to monitor protein kinase A activity. Cyan fluorescent protein was used as the photon donor and yellow fluorescent protein as the photon acceptor. Through this approach, PKA activity was monitored in real time upon stimulation of the PKA signaling pathway.[113] A second example of technology using genetically encoded biosensors is the fluorescence fluctuation increase by contracts (FLINC).[114] Unlike FRET-based approaches, FLINC measures the fluorescence emissions per pixel or fluctuations of the FP from cell images.[114] FLINC biosensors are designed in the AKAR scaffold using FLINC specific fluorescent proteins. Upon substrate phosphorylation, the phosphorylated peptide sensor binds to the phosphorylated residue and the biosensor folds upon itself bringing the two FPs into proximity. This technique was used to monitor activity and identify the localization of PKA and ERK kinases.[114]

A second approach for monitoring cellular kinase activity is to use artificial peptide substrates labeled with organic dyes.[115] Following conjugation, the substrate-probe complexes are delivered inside of the cell through membrane penetrating sequences. The conjugated substrate-probe complexes can be toxic to the cells. In addition, self-reporting

biosensors can be localized within the cell using localization tags such as nuclear localization or exclusion sequences.[115]

Conjugating kinase substrates with fluorescent probes to monitor kinase activity contains limitations. One limitation is the lack of substrate specificity within a kinase family, which are categorized by their protein structure and sequence similarity. The high degree of similarity between catalytic domains within the same family makes generating specific substrates a difficult task. Substrate specificity is important when using genetically encoded or probe conjugated kinase substrates to monitor cellular kinase activity. A second limitation is the lack of reported and efficient kinase substrates for the majority of the human kinome.[64] The approaches discussed above highlight the importance of identifying ideal substrates for understudied kinases.

1.6   Dissertation objectives

As described above, experimental and *in silico* approaches have been used to identify kinase substrate preferences, predict and to design peptide substrates to monitor protein kinase activity. However, the high throughput identification of kinase substrates required laborious and expensive approaches that required multiple phases of validation. Herein, strategies for an automated, high throughput process for tyrosine kinase substrate identification, artificial substrate design, and kinase activity monitoring will be discussed.

Chapter two of this thesis contains a published manuscript on which I was second author, describing the KINATEST-ID pipeline and how my work contributed to the creation of this method. The KINATEST-ID pipeline is a set of modular steps used to identify, generate *in silico,* and validate a kinase family's preferred kinase substrate through a terbium chelation time resolved assay.

The work described in chapter three continues the application of the KINATEST-ID pipeline. In this chapter, we adapted a streamlined phosphoproteomic process to increase the number of known FLT3 kinase variant substrates and generated a set of data formatting tools to facilitate the use of phosphoproteomic data into an updated version of the KINATEST-ID pipeline. FLT3 is a RTK involved in acute myeloid leukemia and it is a major role player in drug resistance and disease relapse. Currently, there are many efforts to elucidate the FLT3 RTK signaling pathway in AML but with no reported efficient

artificial substrate. Through this work, we employ the kinase assay linked with phosphoproteomics technique for rapid kinase substrate identification and my efforts centered on establishing the KALIP technique as a viable process to identify FLT3 kinase variant substrates for incorporation into the KINATEST-ID pipeline. Through this process, we identified the FLT3 substrate preference to design a set of FLT3 artificial substrates (FAStide), of which two substrates were used in TKI dose response assays to monitor FLT3 kinase variant activity (Chapter 3).

The work described in chapter four was a collaboration with a Masters student in our group, John Blankenhorn, and focuses on incorporating the data formatting and KINATEST-ID tools, developed in chapter three, into the GalaxyP framework that allowed for autonomous processing of the phosphoproteomic data to develop a lanthanide-based assay to monitor Bruton's tyrosine kinase (BTK) activity. BTK is a non-receptor tyrosine kinase involved in leukemia progression and disease relapse. BTK does not have a reported artificial substrate that can be used to monitor kinase activity, limiting drug discovery efforts. This work outlines the use the KALIP technique to increase number of BTK substrates that are then incorporated into GalaxyP-based KINATEST-ID pipeline. The updated KINATEST-ID workflow in GalaxyP also uses the BLOSUM-100 matrix to align a BTK artificial substrate (BAStide) with a lanthanide binding motif for antibody free time resolved kinase activity detection (chapter four).

The final chapter of this work is a description of my contributions in the development of a multiplexed and antibody free detection method for multiple tyrosine kinase activities published in Analytical Chemistry, on which I was second author.

CHPATER TWO:
- DEVELOPMENT OF THE KINATEST-ID PIPELINE FOR SUBSTRATE PREFERENCE IDENTIFICATION
- 2nd AUTHOR CONTRIBUTIONS

CHAPTER THREE:
- A PHOSPHOPROTEOMICS APPROACH TO IDENTIFY KINASE SUBSTRATES FOR KINATEST-ID IDENTIFICATION OF SUBSTRATE MOTIF
- 1st AUTHOUR CONTRIBUTIONS

CHAPTER FOUR:
- STREAMLINED WORKFLOW FOR MS DATA AND KINATEST-ID ANALYSIS WITHIN THE GALAXYP FRAMEWORK
- 1st AUTHOR CONTRIBUTIONS

CHAPTER FIVE:
- DEVELOPMENT OF AN ANTIBODY-FREE MULTICOLORED TB ASSAY FOR KIASE ACITIVITY DETECTION
- 2nd AUTHOR CONTRIBUTIONS

Figure 1.14. Thesis chapter concept map

CHAPTER 2.   KINATEST-ID™: A PIPELINE TO DEVELOP PHOSPHORYLATION-DEPENDENT TERBIUM SENSITIZING KINASE ASSAYS

2.1   Contributions to this work

Chapter two is the KINATEST-ID manuscript in its published form. I have received copyright approval to incorporate the manuscript into my dissertation. The manuscript describes a series of steps that uses reported kinase substrate information to create a set of modules that are used to identify, predict, filter, align and validate artificial peptide substrates for a kinase of interest to create an antibody-free activity assay.

During the first three years of my graduate school career, I worked under the guidance of Dr. Laurie L. Parker and a lab alumnus, Dr. Andrew M. Lipchik. They afforded me the opportunity to take the lead in the development, synthesis and validation of the SRC family kinase artificial substrate (SFAStide-A) reported in the publication of the KINATEST-ID pipeline.

The processed included searching proteomic databases to identify Lyn and other SCR family kinases substrates that were used to generate a small library of peptide substrates that were used as input for the KINATEST-ID workbooks, which predicted a series of candidate peptide sequences. With the support and encouragement from Dr. Parker and Dr. Lipchik, I identified and synthesized what eventually became the SFAStide-A sequence. Subsequently, I carried out the reminder of the KINATEST-ID pipeline steps to align the SFAStide into the terbium binding motif to create an in vitro terbium-based assay for detection of SRC family kinase activity.

Reprinted (adapted) with permission from (Lipchik AM, Perez M, Bolton S, et al. KINATEST-ID: A pipeline to develop phosphorylation-dependent terbium sensitizing kinase assays. J Am Chem Soc. 2015;137(7):2484-2494. doi:10.1021/ja507164a.). Copyright (2015) American Chemical Society.

Andrew M. Lipchik[1], Minervo Perez, Scott Bolton, Vasin Dumrongprechachan, Steven B. Ouellette[2], Wei Cui, Laurie L. Parker[3]*

Department of Medicinal Chemistry and Molecular Pharmacology, College of Pharmacy, Center for Cancer Research, Purdue University, West Lafayette, IN 47907

Present addresses: [1]Department of Genetics, Stanford University School of Medicine, Stanford, CA, 94305; [2]KinaSense LLC, West Lafayette, IN 47906; [3]Department of Biochemistry, Molecular Biology and Biophysics, College of Biological Sciences, University of Minnesota Twin Cities, Minneapolis, MN 55455
*Corresponding author: llparker@umn.edu

2.2   Abstract

Non-receptor protein tyrosine kinases (NRTKs) are essential for cellular homeostasis, and thus are a major focus of current drug discovery efforts. Peptide substrates that can enhance lanthanide ion luminescence upon tyrosine phosphorylation enable rapid, sensitive screening of kinase activity, however design of suitable substrates that can distinguish between tyrosine kinase families is a huge challenge. Despite their different substrate preferences, many NRTKs are structurally similar even between families. Furthermore, the development of lanthanide-based kinase assays is hampered by incomplete understanding of how to integrate sequence selectivity with metal ion binding, necessitating laborious iterative substrate optimization. We used curated proteomic data from endogenous kinase substrates and known Tb^{3+}-binding sequences to build a generalizable *in silico* pipeline with tools to generate, screen, align and select potential phosphorylation-dependent Tb^{3+}-sensitizing substrates that are most likely to be kinase specific. We demonstrated the approach by developing several substrates that are selective within kinase families and amenable to HTS applications. Overall, this strategy represents a pipeline for developing efficient and specific assays for virtually any tyrosine kinase that use high throughput screening-compatible lanthanide-based detection. The tools provided

in the pipeline also have the potential to be adapted to identify peptides for other purposes, including other enzyme assays or protein binding ligands.

2.3   Introduction

Protein kinases catalyze the reversible phosphorylation of proteins and play a ubiquitous role in the regulation of signal transduction pathways directing cellular processes including proliferation, survival and adhesion. Phosphorylation of a protein can result in changes in activity, conformation, and stability as well as facilitate protein-protein interactions through phospho-recognition domains. The human genome encodes more than 500 protein kinases, 32 of which are non-receptor tyrosine kinases (NRTKs).[4] This group of kinases has diverse roles in integrating signaling events initiated at the plasma membrane, including regulation of cell shape, motility, proliferation, and survival. NRTK deregulation occurs frequently in cancer through a variety of mechanisms including overexpression, gain-of-function mutation, or loss of negative regulators.[116–118] The association of many NRTKs with cancer and inflammatory disease has led to large drug discovery efforts, resulting in the development of 24 FDA-approved small molecule NRTK inhibitors since 2001.[119] However, despite their established clinical importance, approved inhibitors target only a small subset of NRTKs (5 out of 32). A major factor impeding development of kinase inhibitors is the difficulty in producing compounds that are highly specific, and several promising kinase inhibitors have failed clinical trials due to unanticipated off-target effects. Therefore, the development of broad-based tools that allow for sensitive detection of kinase activity has important applications in profiling kinase inhibitor specificity.

Typical strategies for monitoring kinase activity use radioactive ATP, antibodies, or proteomics to detect phosphorylation of native substrates.[120–122] While these methods have successfully generated a wealth of information about kinase activity, each suffers from several disadvantages. For example, redundancy among even otherwise disparate kinases can also confound the assignment of endogenous phosphorylation sites to a specific enzyme. Artificial peptide substrates offer an attractive strategy for examining kinase activity either *in vitro* or in intact cells, due to their diverse chemistries, compatibility with a wide variety of detection platforms, and their ability to directly report the function of a particular enzyme. A variety of detection methods have been utilized for assaying artificial

substrates, including capillary electrophoresis, voltammetry, mass spectrometry, antibody-based detection (e.g. ELISA), light scattering based methods using SERS and RLS, and fluorescence-based methods such as chelation enhanced fluorescence (CHEF), FRET and fluorescence quenching.[123,124] (still need one source)In particular, CHEF methods that sensitize lanthanide ions such as terbium (Tb^{3+}) in a phosphorylation-dependent manner can enable high sensitivity and analytical reproducibility. Previously, we described the application of a kinase specific peptide substrate (SAStide) for the sensitive detection of spleen tyrosine kinase (Syk) activity *in vitro* through phosphorylation-dependent enhanced sensitization of Tb^{3+} luminescence.[105] The luminescence signal is generated when phosphorylation of the tyrosine residue results in exclusion of water and completion of the Tb^{3+} coordination sphere. Phosphorylation also alters the excitation wavelength of the aromatic side chain, increases the binding affinity for the peptide, and increases the luminescence lifetime, resulting in a large increase in signal to noise (16-fold in the case of SAStide). However, other than this example of a serendipitous case, most CHEF substrates are designed primarily to achieve optimal metal binding, which often comes at the expense of kinase selectivity and enzyme kinetics. Currently there is no general, streamlined method to identify and develop novel substrates that are simultaneously specific for an individual kinase and strong metal chelators. To develop such an approach, both elements (specificity and binding) must be taken into account.

In this report, we present a pipeline to develop peptide substrates for tyrosine kinases (using the NRTKs as a model system) that are compatible with phosphorylation dependent sensitization of Tb^{3+} (Fig. 1). We employed curated collections of known endogenous substrate sequences and data from positional scanning peptide library microarrays to develop an *in silico* positional scoring matrix model that enabled the rapid identification of selectivity determinants and assessed the relative importance of maintaining certain residues at each position. We used this information and Tb^{3+}-binding motif alignment as sequence-space-filtering criteria to narrow down the potential substrate library generated from the motif for a given kinase. This yielded a manageable handful of sequences that could be empirically tested and thoroughly characterized. We applied this pipeline to generate biosensors for Abl, Jak2, and Src-Family tyrosine kinases and

demonstrated HTS assays using the Abl substrate against a small molecule library to identify novel Abl inhibitors.
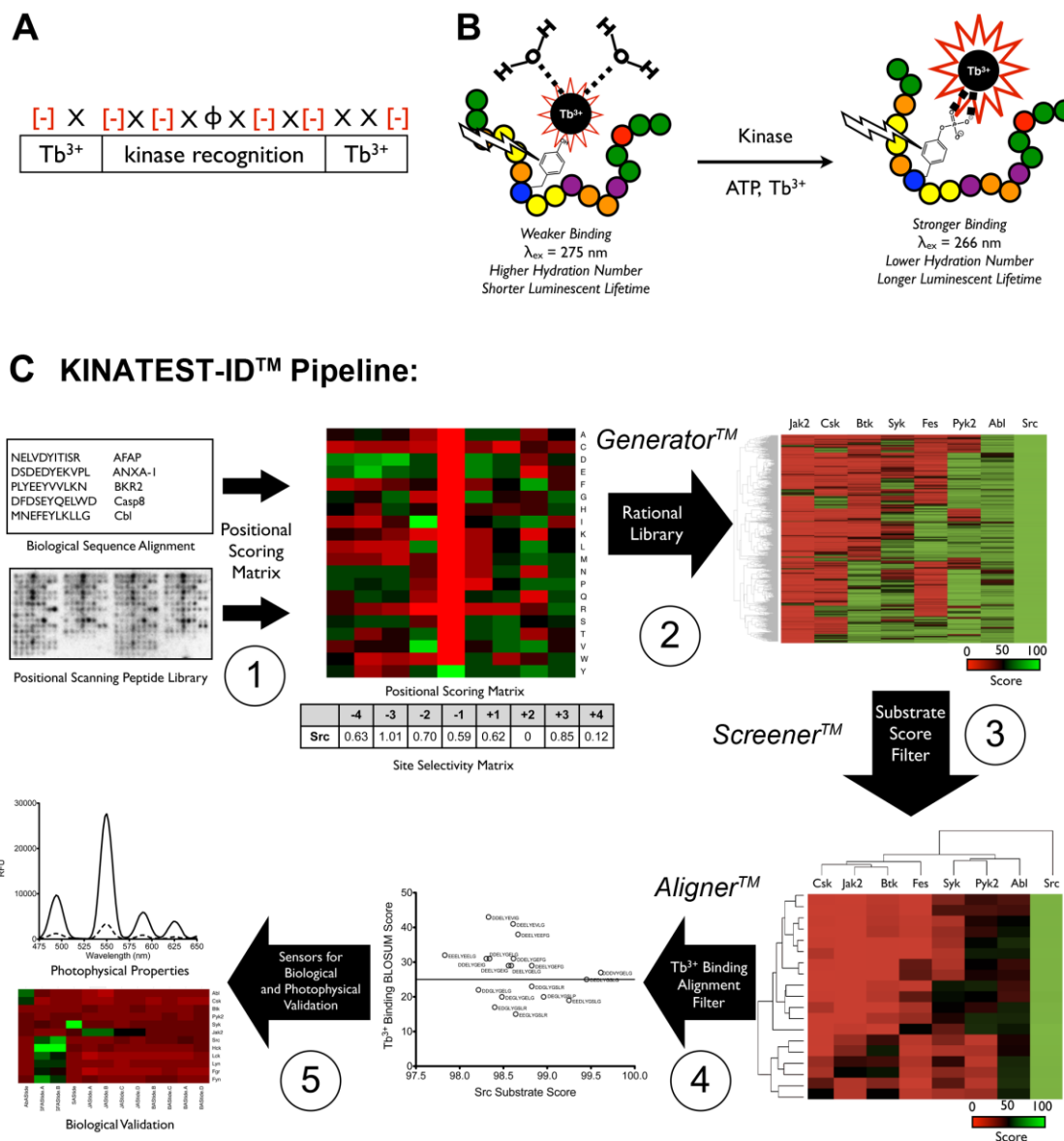
Figure 2.1 Design and development of phosphorylation-dependent enhanced Tb^{3+} luminescence tyrosine kinase peptide biosensors.

Figure 2.1 continued

Design and development of phosphorylation-dependent enhanced Tb^{3+} luminescence tyrosine kinase peptide biosensors.

A) General biosensor design strategy for kinase biosensors capable of phosphorylation induced enhanced Tb^{3+} luminescence, where X is any amino acid, Φ is a hydrophobic antenna containing residue and [-] is an acidic amino acid. B) The detection strategy using the phosphorylation-dependent physical changes in the biosensors that result in enhanced Tb^{3+} luminescence. C) To develop a kinase specific peptide based biosensor, we first obtain all known phosphorylated substrates for a given kinase as the foreground as well as all unphosphorylated tyrosine centered sequences for the substrates and validated proteins that interact with the kinase as the background. Data from positional scanning peptide library screens from the Turk laboratory were also included.[76] (1) A positional scoring matrix, where values represent the preference for each amino acid at every position, and a site-selectivity matrix (SSM), representing the degree to which a given position "requires" a given amino acid, are generated from these data. SSM values are centered at one; values greater than one reflect a strong preference for a particular amino acid at that position and values less than one reflect a lack of preference. (2) A library of sequences were generated *in silico* based on substrate preferences at each positions using the site selectivity score (using the *"Generator^TM"* tool). (3) The library is scored against the kinase of interest as well as all other tyrosine kinases and clustered using bidirectional Euclidian distance and filtered to remove any nonspecific or nonsubstrate sequences for the kinases based on the PSM scores (using the *"Screener^TM"* tool). Scores are on a scale from 0 to 100, where binary classification (of "Substrate" or "Non-substrate") was determined based on threshold values through cross-validation. (4) The remaining sequences are scored using a BLOSUM matrix to assess the similarity to the phosphorylation-dependent Tb^{3+}-binding α-syn Y125 peptide[124, 125] (using the *"Aligner^TM"* tool), which enables filtering out of sequences that are predicted to be selective substrates but not to match the Tb^{3+} motif inherent in the target sequence (which in this case was the best-characterized model, the α-syn Y125 peptide, but could be another Tb^{3+}-binding sequence of interest). (5) The remaining sequences are validated empirically for kinase specificity and photophysical properties associated with Tb^{3+} luminescence. For each relevant step, the score similarity for each kinase (columns) and sequence similarity to one another across kinases (rows) were clustered using bidirectional Euclidian distance.

2.4   Experimental Section

2.4.1   Positional Scoring Matrix (PSM) and Site Selectivity Matrix (SSM) generation.

A blank "substrate informatics sheet" that can be used to perform the functions that yield the PSM and SSM, and workbooks for the Generator™, Screener™, and Aligner™ tools are provided as supplementary files. The calculations in the workbook were implemented as follows:

2.4.1.1   Positional Scanning Peptide Library

To combine the PSPL data from the Turk laboratory with the endogenous substrate information in the filtering algorithm, peptide phosphorylation signals for each array were quantified based on the median intensity for each spot. The median intensity values were then background corrected and signal intensity were then normalized by the following equation:

$$Z_{i,j} = m \leftrightarrow \frac{S_{i,j}}{\Sigma S_{j,i}}$$

where Z_{ij} stands for the normalized score of amino acid j at position i having a signal score S_{ij} and m stand for the total number of amino acids. S_{ci} is the signal score of amino acid j at position i where i is defined in the summation of all the m amino acids.

2.4.1.2   Positional Probability (from endogenous substrates)

We computed the probability matrix, PM, as follows. It is experimentally known that kinase k phosphorylates n substrates (n_1, n_2, …, n_n) consisting of nine amino acids, four on each side of the phosphorylation site. The frequency of each amino acid at each position in the collection of substrates was computed, $f_{j,i}$, where j is amino acid (A, C, …, W, Y) at position i (-4, -3, …, 1). Due to the limitation of identified substrates for some kinases, when $j = 0$ for those amino acids the value of $j = 1/n$, where n is the number of substrate sequences for kinase k. The matrix values were computed by comparing the observed frequency, $f_{i,j}$, within the substrates to the expected frequency (background frequency), $b_{i,j}$, derived from the frequency of each amino acid in each protein containing a substrate sequence as well as non-phosphorylated interacting proteins (obtained from the

Protein Information Resource (http://pir.georgetown.edu).[125] This allowed for the background of amino acids to reflect the proteins with which the kinase naturally interacts. We constructed the probability matrix 20 x 9 for each amino acid and position defined as $s_{i,j} = f_{i,j} / b_{i,j}$.

2.4.1.3 Positional Scoring Matrix

The two individual matrices, PM and PSPLM, were then multiplied together to form the positional scoring matrix, PSM. The value for each amino acid can then be used to identify favorable and unfavorable residues at each position. Values greater than 0.9 were considered favorable or permissive for the kinase, while values less than 0.9 were consider unfavorable or impermissive.

For an nonapeptide of a given amino acid sequence the product of all $s_{i,j}$ values yields the raw probability score, S_R.

$$S_R = \prod_{i=1}^{8} S_{i,j}$$

The raw score was normalized by probability of any nonapeptide being a substrate for kinase k, P_s. P_s was determined by the number kinase substrates collected n plus the number of significantly favorable amino acid from the PSPL compared to the total number tyrosine centered nonapeptides seen in substrate and interacting proteins and the 200 peptides from the PSPL for kinase k.

$$P_s = \frac{n + x}{n_T + 200}$$

$$S = \frac{S_R}{S_R + \dfrac{1}{P_S}}$$

The site selectivity matrix was determined by the ratio of the number of significantly abundant residues found at the subsite, $n^{sig}_{i,j}$, to that the expected abundance from a random distribution, $n^{sigaa}_{i,j}$, multiplied by the ratio of the number of the number of significantly abundant at the subsite to the total number of residues, $n^{aa}_{i,j}$.

$$S_i = \frac{\sum_j n_{i,j}^{sig}}{\sum_j n_{i,j}^{sigaa}} \times \frac{\sum_j n_{i,j}^{sig}}{\sum_j n_{i,j}^{aa}}$$

An amino acid was defined as being significantly abundant if its frequency was found to be greater than two standard deviations above the mean.

2.4.2   Generation of kinase-focused virtual peptide libraries.

Kinase focused virtual (i.e. *in silico*) peptide libraries were generated using the Generator™ tool based on the values of the PSM. All $s_{i,j} > 0.9$ were chosen as potential residues at each position. Combinatorial peptide sequences were generated from these residues and scored against each kinase using the Screener™ tool. Those peptides that scored positive for the kinase (or kinase family) of interest and negative for all other kinases (or kinase families) were then selected and added to virtual "focused libraries" in the Aligner™ Excel spreadsheet for further screening.

2.4.3   Terbium Binding *in silico* Screening.

Following the generation of focused putative kinase substrate libraries, sequences were filtered for the potential to bind terbium in a phosphorylation-dependent manner using the Aligner tool™. A BLOSUM62 matrix was used to generate a sequence similarity score between the focused library of potential kinase substrates and the known terbium sensitizing sequence α-syn Y125 (DPDNEAYEMPSEEG).[126,127] The top several sequences (as desired) were chosen for further empirical evaluation.

2.4.4   Peptide synthesis and purification.

Peptides were synthesized using a Protein Technologies Prelude parallel peptide synthesizer on Rink-amide resin (Peptides International, Louisville, KY). Coupling of standard Fmoc-protected amino acids (Peptides International, Louisville, KY) was achieved with HCTU (Peptides International, Louisville, KY)(100 mM) in the presence of NMM (Sigma-Aldrich, St. Louis, MO) (400 mM) in DMF (EMD Millipore, Billerica, MA) for two 10 min couplings. Fmoc deprotection was performed in 20% piperidine (Sigma-Aldrich, St. Louis, MO) in DMF for two 2.5 min cycles. Peptides were purified to >90% purity by preparative C18 reverse-phase HPLC (Agilent 1200 series) using a linear gradient

5%-38% acetonitrile/0.1%TFA and water/0.1%TFA and characterized using HPLC-MS (ThermoFinnegan Accela-LTQ).

2.4.5   *In vitro* kinase assays (Tb^{3+} luminescence).

Recombinant kinases; Abl, Src, Lyn, Csk, Jak2 and Hck (Millipore) and Syk, Btk, Fyn, Pyk2, and Fgr were expressed as described elsewhere.[76] Recombinant kinases were incubated with the kinase reaction buffer (100 µM ATP, 10mM $MgCl_2$, 125 ng/µL BSA and 25 mM HEPES pH 7.5, total volume 180 µL) containing 12.5 µM biosensor at 30°C. Aliquots (20 µL) were taken at designated time points (0.5, 5, 10, 15, 30, 45 and 60 min) and quenched in 6 M urea (20 µL). The quenched samples were then treated with the luminescence buffer (500 µM Tb^{3+} and 500 mM NaCl, 10 µL) for a total volume of 50 µL (final concentrations of sample components: 2.4 M urea, 40 µM ATP, 4 mM $MgCl_2$, 50 ng/µL BSA and 10 mM HEPES pH 7.5). Time-resolved luminescence emission spectra were collected on a Biotek Synergy4 plate reader equipped with a monochromator at room temperature in black 384-well plates (Greiner Fluorotrac 200). Spectra were collected between 450 and 650 nm in 1 nm increments with 1 ms collection time and 10 reading per data point at a sensitivity of 180 after excitation at 266 nm with a Xenon flash lamp followed by a delay of 50 µs. area under each spectrum was integrated using GraphPad Prism. An additional aliquot (2 µL) of the kinase reaction mixture was taken at each time point for validation of phosphorylation using an ELISA-based chemifluorescent assay as previously described.[83]

2.4.6   Chemifluorescent detection of phosphorylation.

Each aliquot was quenched with 0.5 M EDTA and incubated in a 96-well Neutravidin coated plate (15 pmol biotin binding capacity per well, Thermo Scientific) in Tris-buffered saline (TBS, 25mM Tris-HCl and 150mM NaCl) containing 0.1% BSA and 0.05% Tween 20 for 1h. Following incubation, each well was washed with the TBS buffer and then incubated with mouse anti-phosphotyrosine monoclonal antibody 4G10 (Millipore, 1:10,000 dilution in TBS buffer) for 1h. Following incubation, each well was washed with TBS buffer and incubated with horseradish peroxidase-conjugated goat anti-mouse immunoglobulin G (IgG) secondary antibody (Abcam) (1:1000 dilution) for 1h.

Wells were then washed and treated with Amplex Red reaction buffer (Amplex Red reagent, Invitrogen, 20 mM H_2O_2 and sodium phosphate buffer) for 30 min. Fluorescence was measured using a Synergy4 multiwell plate reader (Biotek) with an excitation wavelength of 532 nm and emission wavelength of 590 nm.

2.4.7 Dose-response inhibition assay.

Kinase (15 nM) was incubated with the kinase reaction buffer described above in the presence of DMSO (vehicle) or varying concentrations of kinase inhibitors (nilotinib, bosutinib, ruxolitinib) at 30°C for 10 min prior to the start of the reaction by adding the peptide substrate. The reaction was started with the addition of biosensor (37.5 µM, total reaction volume 20 µL). Each reaction was quenched after 30 min in 6 M urea (20 µL). The samples were then treated with the luminescence buffer (500 µM Tb^{3+} and 500 mM NaCl, 10 µL) for a total volume of 50 µL. Time-resolved luminescence spectra were collected as described above and the area under the emission curve determined. The IC_{50} value for each inhibitor was determined by fitting data to equation below where inhibition$_{max}$ is the bottom plateau of the curve, inhibition$_{min}$ is the top plateau of the curve, the Hill slope is the steepness of the curve and X is the concentration of the inhibitor.

$$y = \frac{Inhibition_{max} + (Inhibition_{min} - Inhibition_{max})}{1 + 10^{((logIC_{50} - X) * Hill\ Slope))}}$$

2.4.8 High-throughput screening assay.

Abl kinase (3 nM) was incubated with the kinase reaction buffer described above in the presence of DMSO (vehicle), imatinib (positive control) or a single compound from the GSK PKIS library (10 µM), at 30°C for 30 min prior to the start of the reaction by adding the peptide substrate. The reaction was started with the addition of the biosensor AbAStide (12.5 µM, total reaction volume 20 µL). Each reaction was quenched after 1 h in 6 M urea (20 µL). The samples were then treated with the luminescence buffer (500 µM Tb3+ and 500 mM NaCl, 10 µL) for a total volume of 50 µL. Time-resolved luminescence emission intensities were collected at the maxima of the four emission peaks and summed together to give total signaling for each well using the instrument settings described above.

Percent inhibition was determined using the positive inhibition control, imatinib, and the negative inhibition control, DMSO.

2.4.9 Growth inhibition curves.

K562 cells were seeded into 96-well plates at 10,000 cells per well in Iscove's Modified Dulbecco's Medium supplemented with 10% fetal bovine serum and pen/strep. The cells were dosed with the indicated inhibitor at the indicated concentrations (n = 4), and allowed to incubate for 3 days at 37° C. Following incubation, XTT reagent (ATCC) was added according to manufacturers protocol, and allowed to incubate at 37° C for 3 hours. Absorbance at 475 nm was measured on a Biotek Synergy4 plate reader. Values were calculated as percent of vehicle (0.1% DMSO), plotted in Graphpad Prism 6, and IC50 values generated by fitting a variable slope (four parameter) curve.

2.4.10 High-Throughput Screening Calculations.

The Z' factor was calculated according to Eq. 2.

$$Z' = \frac{\left(\mu_{pos} - \frac{3\sigma_{pos}}{\sqrt{n}}\right) - \left(\mu_{neg} - \frac{3\sigma_{neg}}{\sqrt{n}}\right)}{\mu_{pos} - \mu_{neg}} \quad [2]$$

The signal window was calculated according to Eq. 3

$$SW = \frac{\left(\mu_{pos} - \frac{3\sigma_{pos}}{\sqrt{n}}\right) - \left(\mu_{neg} - \frac{3\sigma_{neg}}{\sqrt{n}}\right)}{\frac{\sigma_{pos}}{\sqrt{n}}} \quad [3]$$

where n is the number of replicates, μ_{pos} and μ_{neg} are the average luminescence of the positive (phosphorylated peptide or uninhibited) and negative (unphosphorylated peptide or control inhibitor-treated) controls respectively; σ_{pos} and σ_{neg} are the standard deviation of the positive and negative controls.

2.5   Results

2.5.1   KINATEST-ID™: a substrate peptide sequence space filtering pipeline

Inspired by the general design rules of previous CHEF-based sensors for detection of kinase activity (i.e. Sox-Mg^{2+} and EF-hand-Tb^{3+} sensors), we aimed at developing a general approach to design biosensors for tyrosine kinase activity analysis using phosphorylation-dependent enhanced Tb^{3+} luminescence. Sensors were designed to combine nonreceptor tyrosine kinase substrate specificity with the excitation and chelation elements governing Tb^{3+} luminescence. To achieve this, each kinase biosensor was developed to contain an optimized substrate sequence with an embedded Tb^{3+} coordination motif, similar to that previously identified from the 14-residue fragment of α-synuclein surrounding Y125. (Figure 1A) Based on our previous work, we hypothesized that phosphorylation-dependent physical changes in the biosensor would enable enhanced Tb^{3+} luminescence of the phosphorylated biosensor compared to the unphosphorylated form (Figure 1B).[105]

For each kinase, a focused virtual library of peptide biosensors was designed, optimized and selected *in silico* through a bioinformatic pipeline, KINATEST-ID™ (Kinase Terbium Emission Sensor Identification) comprised of three data processing tools: Generator™, Screener™, and Aligner™, implemented in Microsoft Excel workbooks that are available as supplementary files (Fig 1C). This method starts with the generation of a positional scoring matrix (PSM) (as described below) for a given kinase, which uses highly curated, biologically validated phosphorylation sites for individual kinases as well as empirically observed effects of amino acids from positional scanning peptide microarray data (unpublished, shared by Turk and co-workers) (Figure 1C Step 1). This matrix represents the relative preference the kinase has for each amino acid at each position within the sequence, yielding comparable preference motifs to those generated by state-of-the-art phosphosite prediction algorithms (e.g. NetPhorest and M3)[128,129] (Tables S1-S7). A site selectivity matrix (SSM) (evaluating the importance of a particular site in the sequence to the preference of the kinase for that substrate) was also generated using the data (Table S8). These matrices were used to guide the generation of a focused *in silico* library of possible kinase-specific peptide substrates using the "Generator™" tool, where the motifs derived from the set of amino acids that were represented at >2 standard deviations from the mean

were used to generate a list of all possible permutations of that set of amino acids at their respective positions (Figure 1C Step 2).

Each sequence in the focused library was given a score based on the PSM (which takes into account both the endogenous and positional scanning peptide library data) for the given kinase, as well as a score for all other kinases included in the analysis using their respective PSMs, using the "Screener™" tool, which effectively cross-references each sequence for its predicted selectivity amongst the kinases included in the analysis (Figure 1C Step 2). The focused library was then filtered using Screener™ based on classifying the sequences as predicted "substrates" or "nonsubstrates" for each kinase as well as "specific" or "nonspecific" for the given kinase. All nonsubstrate and nonspecific sequences were then filtered from the library. (Figure 1C Step 3) Cutoff scores for classifying the sequences as substrates or nonsubstrates for each kinase were selected based on the algorithm training parameters to give the lowest false discovery rate for the kinase of interest and the highest sensitivity for all off target kinases (Table S9). While not necessarily providing hard cutoffs, this at least ensured that all remaining sequences in the library would have a maximal likelihood to be substrates for the desired kinase and not for the other kinases. The remaining sequences were compared to the atypical Tb^{3+} sensitizing peptide derived from the α-synuclein Y125 center peptide using BLOSUM sequence alignment scoring using the "Aligner™" tool. (Figure 1C Step 4) Sequences with a BLOSUM score below the threshold of 25% similarity were considered "non-optimal" binders, however some were synthesized for testing to evaluate the predictive capabilities of the alignment score. Sequences could also be optimized for Tb^{3+} binding by changing amino acids at positions that are less important for substrate recognition (based on the site selectivity scores). This ultimately yielded a compressed library of potential kinase-specific peptide substrate sequences that were also likely to sensitize Tb^{3+} luminescence, from which a handful of the top ranked sequences were chosen for studies to demonstrate kinase specificity and Tb^{3+} sensitization. The site selectivity matrix was used when deciding the priority for sequences to empirically test, since it enabled more optimal balancing of both Tb^{3+} binding residues and residues the kinase preferred at specific sites (Figure 1C Step 5). Accordingly, particular sites that lack selectivity (thus having more flexibility for a given amino acid at

that position) but are required for Tb^{3+} binding could be substituted with the appropriate Tb^{3+} binding residue, as opposed to a residue suggested by the catalytic preference motif.

2.5.2 Design of Abl, Jak2 and Src-family kinase substrate biosensors.

To demonstrate its utility, KINATEST-ID™ was applied to generate Tb^{3+}-sensitizing biosensors predicted to be specific for Abl, Jak2, and Src-family kinases. Initial potential substrate sequence libraries were generated by determining each kinase's preference motif using the *in silico* model and listing all possible permutations of that motif in a virtual library using Generator™. These virtual libraries started with ~43,000, 92,000 and 5,500 sequences for Abl, Jak2 and Src-family kinases, respectively. These libraries were then filtered with Screener™ by PSM scores for each kinase in the analysis to remove sequences with favorable predictions for other kinases (i.e. nonspecific) and unfavorable predictions for the target kinase (i.e. nonsubstrate), which drastically reduced the library size by ~99% for each kinase. The Tb^{3+} binding alignment score filter was then applied using Aligner™, which reduced the size of the libraries by a further ~50%, leaving libraries ranging in size from 11-250 sequences. The remaining sequences for each of the kinases contained the identified kinase substrate motifs as well as the α-syn Y125 Tb^{3+} binding motif or slight shifts in that motif.

From these libraries, several sequences were selected to evaluate empirically for each kinase as kinase artificial substrate peptides (KAStides) for Abl (AbAStide), Jak2 (JAStide), and Src-family (SFAStide) kinases (Fig. 2A). Sequences from the pipeline were selected on the basis of highest predicted selectivity for the given kinases, and higher Tb^{3+} binding alignment scores (although a selection with a range of lower binding scores was also included, in order to test the relationship between alignment score and binding affinity). The specificity of these sensors was assessed by screening the peptides against a panel of kinases representing at least one member of each family of nonreceptor tyrosine kinases. The ability of the kinases to phosphorylate a given peptide was determined using an endpoint *in vitro* kinase assay. Phosphorylation of each peptide was determined quantitatively using chemifluorescent ELISA.[83] Relative fluorescence units (representing the amount of phosphorylated peptide present) were measured and percent phosphorylation

was interpolated from a calibration curve generated from synthetically phosphorylated peptide. (Figure 2B, Supporting information Figure S1.)

A

Kinase	Substrate	Sequence	Abl	Btk	Csk	Fes	Jak2	Pyk2	SFK	Syk	Tb³⁺
Abl	AbAStide	GGDEDEVAYQPFEDGGK(biotin)GG	**92.9**	0.00	9.49	5.59	0.00	5.53	27.5	19.4	25
Jak2	JAStide-A	GGDNDPDRYIRTEDEGGK(biotin)GG	18.7	0.00	0.00	5.26	**90.6**	0.00	0.410	2.03	25
	JAStide-B	GGEDEGDRYLKLEEDGGK(biotin)GG	0.879	1.34	0.00	3.59	**98.9**	0.100	2.35	10.5	13
	JAStide-C	GGEDEGDRYVQLDEGGK(biotin)GG	2.97	0.120	0.0682	0.544	**99.5**	5.04	7.99	0.330	12
	JAStide-D	GGEDPKFYVQLDEGGK(biotin)GG	26.1	0.0351	0.119	1.73	**98.8**	0.00	12.2	0.276	12
	JAStide-E	GGDNDPDEYITLDEDGGK(biotin)GG	34.4	1.14	0.0330	0.770	**99.7**	0.00	79.0	79.7	25
SFK	SFAStide-A	GGEEDEDIYEELDEPGGK(biotin)GG	34.7	19.7	65.1	51.8	0.00	86.9	**99.6**	98.8	26
	SFAStide-B	GGDNEGDVYDFVEDEGGK(biotin)GG	37.0	26.3	1.47	66.9	0.351	**99.3**	91.1	1.45	27
Syk	SAStide	GGDEEDYEEPDEPGGK(biotin)GG	83.2	0.940	51.0	17.3	0.00	13.0	90.5	**99.9**	38

Figure 2.2 Identification, validation, and characterization of kinase specific biosensors using KINATEST-IDTM.

Figure 2.2 continued

Identification, validation, and characterization of kinase specific biosensors using KINATEST-ID^{TM}.

A) The kinase substrate sequences selected for further evaluation and their prediction scores for the panel of kinases used in the assay, and Tb^{3+} alignment score. For kinase substrate prediction, scores >90 generally reflect "positive" substrates, whereas scores lower than 90 reflect some similarity with that kinase's preferred motif but were below the thresholds defined by the Screener™ tool (i.e. lowest false discovery rate and highest sensitivity for the off-target kinases). Tb^{3+} alignment scores >25 were considered "positive" for Tb^{3+} binding, and generally, higher alignment scores correlated with longer luminescence lifetime of the phosphopeptide-Tb^{3+} complex (and thus higher occupancy of the chelated vs. hydrated form of Tb^{3+} in the equilibrium) (Fig. S4). B) Screening of kinase substrates against a panel of purified recombinant kinases (3 nM each, except 250 nM Csk which was the amount of recombinant Csk required to phosphorylate the positive control Src Y530-centered peptide in characterization experiments, data not shown) using ELISA–based chemifluorescence detection. Color-coded values represent the mean of experiments performed in triplicate (with individual graphs shown in Fig. S1). Substrate phosphorylation specificity per kinase (rows) and kinase specificity per substrate (columns) were clustered using bidirectional Euclidian distance. C-E) Recombinant, active Abl, Jak2 and Lyn (3 nM) were used to carry out the kinase reactions with 100 μM ATP and increasing concentrations of AbAStide, JAStide-E and SFAStide-A. Reaction progress was monitored using ELISA–based chemifluorescence detection. Initial rates of phosphorylation of the kinase specific biosensors (picomoles of phosphorylated product per minute) for AbAStide (C) JAStide-E (D) and SFAStide-A using Lyn (E) were calculated and fitted to the Michaelis-Menten equation. Values represent the mean ±SEM of experiments performed in triplicate.

All sequences were specific substrates of the intended kinases/kinase families (Fig. S1). AbAStide did display some nonspecific phosphorylation by Csk; however this was only observed at a very high concentration of Csk enzyme (250 nM, 83-fold greater than that used for the Abl assay). This suggested that, while not explicitly measured, the k_{cat} and catalytic efficiency of AbAStide for Csk are most likely significantly lower than those for Abl. Analysis of Jak2 preference amongst the pool of substrates and kinases tested demonstrated that JAStide-E was the most efficient, with significantly more phosphorylation by JAK2 compared to the other potential JAK substrate sequences (P<0.0001 for JAStide-A and D and P<0.001 for JAStide-B and C). This was consistent with predicted preferences in Jak2 substrates, for which the -1 position demonstrated the greatest preference for acidic residues (JAStide-E) and reduced favorability for arginine (JAStide-A, B, C) and phenylalanine (JAStide-D). The SFAStides displayed comparable levels of phosphorylation across all Src family kinases, while maintaining selectivity against all other families. The variation in residue chemical properties between the sequences at the -3 and +2 positions demonstrated that SFKs tolerate substitutions at these positions with little effect on phosphorylation (which is in accordance with the positional selectivity matrix results). To our knowledge, these sequences are the first reported Jak2 specific substrates (JAStide-A-E), and the first demonstration of family-spanning specific substrates for Src-family kinases (SFAStide A and B). Notably, the core kinase recognition sequence (DEDIYEELD) in the substrate we term SFAStide-A has been previously identified as an optimal Lyn kinase motif,[130] however it has not previously been analyzed in the context of the entire Src family. This gave us further confidence in the ability of our upstream informatic approach to identify appropriate substrate sequences, and also supported the importance of validating peptide substrates across a panel of kinases.

AbAStide, JAStide-E and SFAStide-A were selected for further characterization based on their specificities and efficiency of phosphorylation by their designated kinases. These biosensors were characterized using steady-state kinetics to model the kinase-substrate interaction and subsequent phosphorylation of the substrates. The initial *in vitro* reaction velocities for each kinase-substrate pair were measured and fitted to the Michaelis-Menten equation to derive K_m, V_{max}, turnover number (k_{cat}) and catalytic efficiency (k_{cat}/K_m). (Figure 2C-E, Table S13, Supporting Information Figure S2) Overall, the kinetic

parameters fell between those found in previously reported "optimal" substrate and those for endogenous substrates (which are often relatively low efficiency as short peptides when isolated from their protein context). The K_m values for all the peptides were within ~50-200 µM, lower than is typically observed for endogenous substrates,[131] but about ~2-5-fold greater than for relevant "optimal" substrates. Several reportedly Src specific peptide substrates have been developed using one-bead-one-peptide and oriented peptide libraries with K_m values between 20-55 µM, similar to SFAStide-A (K_m = 62 µM). AbAStide (K_m = 99 µM) exhibited a substantially increased K_m compared to the optimal substrate Abltide (K_m = 4 µM) but comparable to the endogenous substrate CrkL Y207 (K_m = 134 µM). Since JAStide-E (K_m = 186 µM) represents the first report of an unnatural specific substrate for a JAK kinase, we could only compare it to the commonly used endogenous phosphorylation site STAT5 Y694 (K_m = 306 µM), relative to which JAStide-E's K_m was 2-fold lower. The catalytic efficiencies for AbAStide, JAStide-E, and SFAStide-A sequences were excellent, comparable to those reported for the "optimal" kinase substrates. These results demonstrated that KINATEST-ID™ is capable of identifying sequences with a high likelihood of being selective substrates that have comparable kinetic parameters to the optimal substrates previously identified using traditional, fully empirical methods.

2.5.3 Tb^{3+} luminescence characterization of KINATEST-ID™ identified biosensors

The biosensors that displayed appropriate specificity in the screening panel were further evaluated for phosphorylation-dependent enhanced Tb^{3+} luminescence. Phosphorylated and unphosphorylated forms of the peptides were synthesized and Tb^{3+} luminescence emission was analyzed. Steady-state measurements of the biosensors revealed a modest range of enhancement (~1-2 fold) in Tb^{3+} luminescence upon phosphorylation. (Table S14) However, as we have previously observed for a Syk-specific peptide substrate (SAStide),[105] time-resolved measurements significantly improved the enhancement of Tb^{3+} luminescence to the range of ~5-11 fold (approximately 3-5 fold improvement over steady-state measurements). As in that previous work, the enhancement of Tb^{3+} luminescence could be attributed to the differences in properties of the sensors including excitation wavelength (266 nm for the phosphorylated vs. 275 nm for the unphosphorylated), binding affinity, luminescence lifetime and hydration number (Table

S14), which validated the phosphorylation-dependent design of the sensors. The unphosphorylated sequences exhibited binding constants (K_d) in the range of 9-80 μM, which were substantially weaker than the range of 1-12 μM observed for the phosphorylated forms (Table S14). The luminescence lifetimes of the all the biosensors were increased by an amount in the range of 100-200 μsec upon phosphorylation, enabling high signal to noise through time-resolved detection. Interestingly, these lifetimes appeared to be correlated with the Tb^{3+} binding sequence alignment score (Fig. S4), suggesting that the alignment parameter may be useful as a predictive measure for choosing sequences for further characterization as phosphorylation sensitive biosensors since longer lifetimes tended to result in better signal to noise. Overall, these results showed that this general design strategy can be applied to diverse tyrosine kinase substrates, and that these predicted substrates exhibit robust Tb^{3+} luminescence sensitization with photophysical properties consistent with the anticipated detection mechanism.

2.5.4 *In vitro* time-resolved Tb^{3+} luminescence-based detection of tyrosine kinase activity

AbAStide, JAStide-E, and SFAStide-A were further characterized for *in vitro* time-resolved Tb^{3+}-luminescence-based detection of kinase activity. Conditions for optimal detection and calibration curves (using various ratios of phosphorylated and unphosphorylated forms of the sensors) were established in the kinase reaction buffer to account for potential interference from assay buffer components. All sensors displayed linear increases in Tb^{3+} luminescence with increasing percent phosphorylation allowing for quantitative determination of phosphorylation (Supporting Information Figure S3). High-throughput screening parameters were also derived from the calibration curves, including the Z' factor and signal widow (SW), reflecting assay robustness. All sensors displayed appropriate parameters (Z' factor < 0.5 and SW < 2) for application in HTS screening assays (Table S15)

Quantitative *in vitro* kinase activity assays were performed using AbAStide, JAStide-E, and SFAStide-A and recombinant kinases over a 60-minute time course. Percent phosphorylation was interpolated from calibration curves, and followed the trends for those obtained using the quantitative ELISA-based read out (Fig. S5). (Figure 3 A-C) Dose-response inhibition of Abl, Jak2, and Hck kinase activity by the inhibitors imatinib,

ruxolitinib, and dasatinib, respectively, was then assayed in an inhibitor dilution series from 10 pM to 500 µM. Luminescence emission spectra were collected and normalized to the vehicle (DMSO) control and reported as percent control. The observed IC_{50} values were 3.9 ± 1.3 nM, 2.9 ± 1.4 nM, and 2.3 ± 1.6 nM for imatinib/c-Abl, ruxolitinib/JAK2, and dasatinib/Hck, respectively. These values are in agreement with those reported in the literature for each drug/kinase combination.[132–134] The Z' factor and SW for these assays were sufficient for HTS at some concentrations of inhibitor, indicating that characterization of the behavior at a given degree of inhibition will be necessary for optimizing screening assays (Table S16). The AbAStide biosensor was selected for further validation in an *in vitro* HTS for inhibitors of c-Abl.
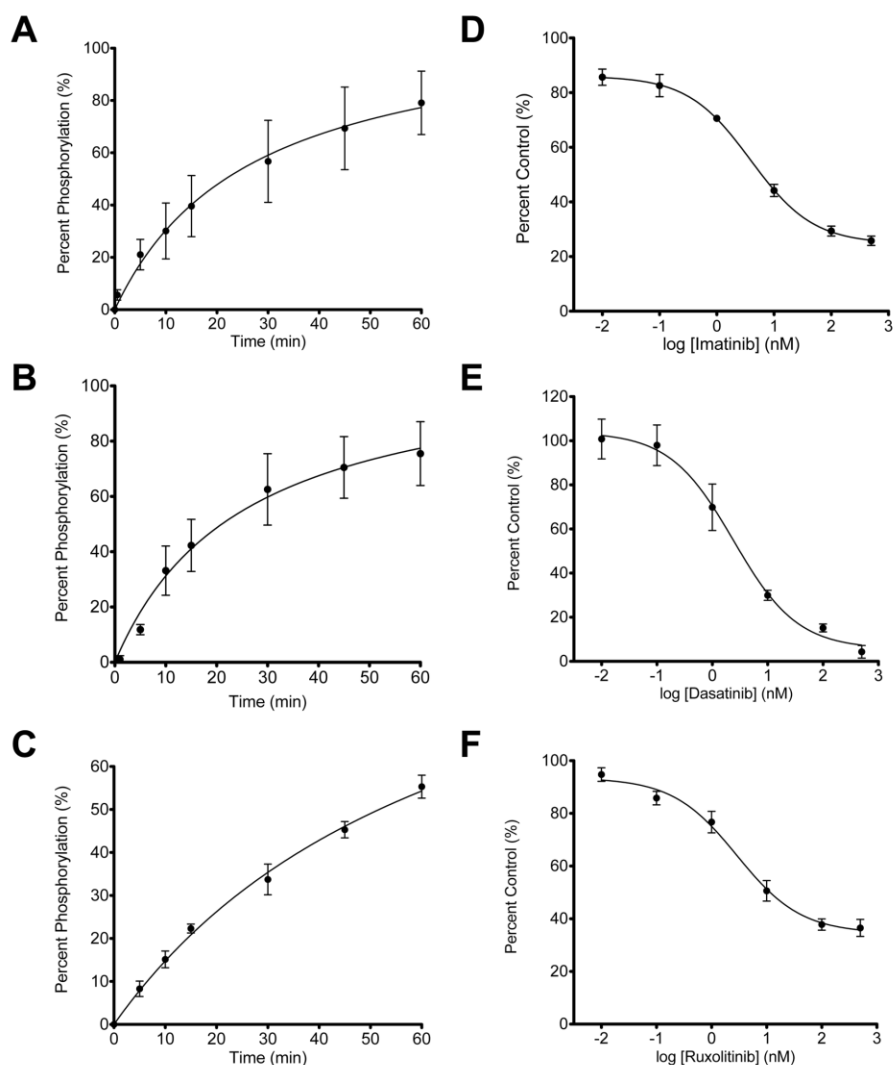
Figure 2.3 Quantitative time-resolved phosphorylation-enhanced Tb3+ luminescence detection of nonreceptor tyrosine kinase activity and inhibition.

Kinase reaction progress curves for Abl (A) Hck (B) and Jak2 (C). Dose-response inhibition of Abl with imatinib, Hck with dasatinib, and Jak2 with ruxolitinib (D-F). Kinase reactions were performed with kinase reaction buffer containing ATP, $MgCl_2$, HEPES, pH 7.5 and 15 nM recombinant kinase. IC_{50} values were determined values were generated by fitting the data to a variable slope (four parameter) curve. Data represent the average $\pm$ SEM of experiments performed in triplicate.

2.5.5   Application of AbAStide to High Throughput Screening for Small Molecule Inhibitors

Replicate in vitro kinase assays were performed in a 384-well plate format in the presence (N=96) or absence (N=96) of imatinib to evaluate reproducible detection of Abl activity using AbAStide.   To evaluate assay quality, positive and negative controls (containing the phosphorylated form and unphosphorylated form of the biosensor, respectively, N=96 for each) were also analyzed. Detection of AbAStide phosphorylation was robust and reproducible. (Fig. 4A) The Z' factor and SW for the kinase reaction replicates were 0.56 and 84 respectively, comparable to those for control well readings, demonstrating sufficient performance for use in high-throughput screening. We leveraged this in a high-throughput screen using the GSK PKIS library, which consists of 364 compounds arrayed in 96-well plates as single compounds at 10 mM in DMSO (available to the research community upon request, see cited reference).[135] The library was screened at a constant 1:1000 dilution, with 10 μM final concentration of compound in each well (1% DMSO). Compounds were incubated with the kinase for 30 minutes prior to start of the kinase reaction, which was initiated by the introduction of the biosensor substrate. The kinase reaction was allowed to proceed for one hour before being quenched with the Tb3+ luminescence buffer (containing urea and Tb3+). The time-resolved Tb3+ emission intensity was measured and the "percent inhibition" was determined compared to the biological positive and negative controls (known inhibitor imatinib and no inhibitor, respectively) (Fig. 4B). Primary hits were identified as compounds reducing Tb3+ luminescence by greater than 3-fold (the top 5% most potent inhibitors, which were the top 18 compounds). These top 5% primary hits were tested in a secondary screen using the same kinase reaction conditions, but employing a chemifluorescent ELISA-based detection instead of Tb3+-based detection. (Fig. S6) The secondary screen confirmed that all of the hits inhibited Abl kinase activity by at least 50% compared to vehicle (Table S17).
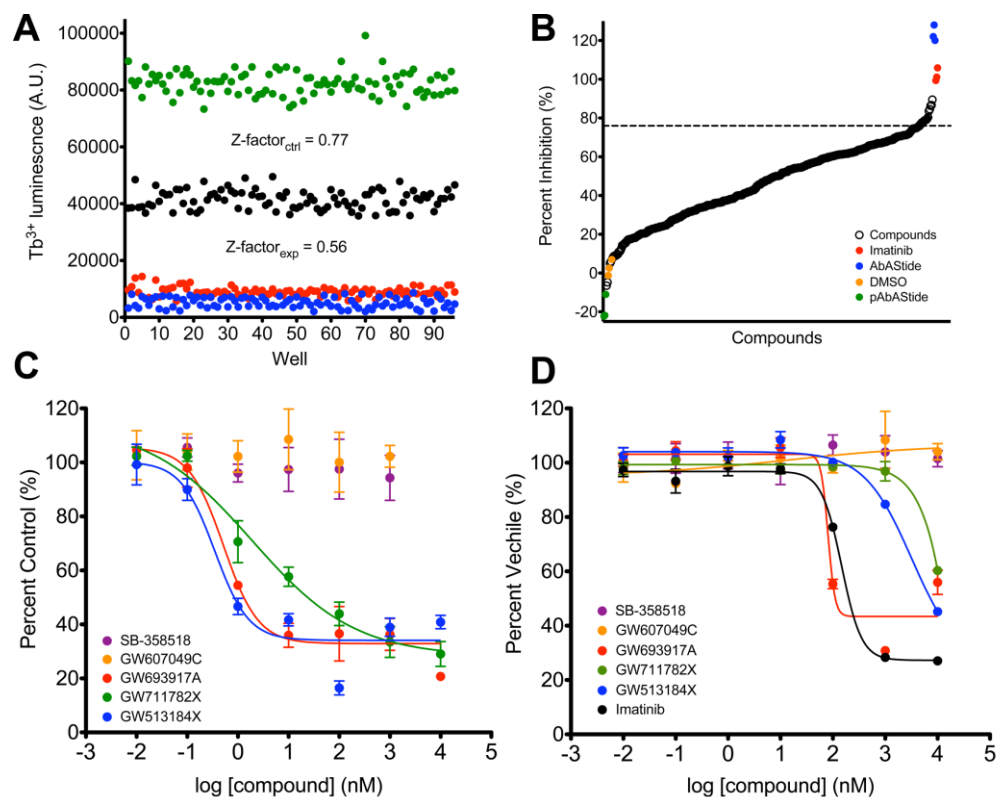
Figure 2.4. A high-throughput chemical screen using AbAStide biosensor identifies inhibitors of Abl tyrosine kinase.

A) The AbAStide in vitro kinase assay shows highly reproducible signal upon imatinib treatment. Green: synthetically phosphorylated peptide (positive control); Black: kinase reaction; Red: kinase reaction + imatinib; Blue: unphosphorylated peptide (negative control). B) Distribution of Abl inhibition identified in a high-throughput screen performed with the GSK PKIS library using AbAStide-sensitized Tb3+ luminescence. Green: synthetically phosphorylated peptide (positive control); Black: GSK PKIS compounds; Red: kinase reaction + imatinib (positive control); Blue: unphosphorylated peptide (negative control); Orange: DMSO (vehicle control). C) Dose-response inhibition of Abl kinase activity by selected compounds from the GSK PKIS, including the top three hits and two non-hits as negative controls. The extent of biosensor phosphorylation was interpolated from an externally generated calibration curve (not shown) and normalized to vehicle (DMSO) control. Data represent the average ± SEM of experiments performed in triplicate. D) XTT cytotoxicity assay for selected compounds (as in 4C), showing potencies in K562 cells. IC50 values were generated by fitting the data to a variable slope (four parameter) curve. Data represent the average ± SEM of experiments performed in triplicate.

The three most potent inhibitors from the HTS and validation screens were GW693917A, GW711782X and GW513184X, developed to target TIE2/VEGFR2, ALK5 and GSK3β respectively. These and two negative compounds (SB-358518 and GW607049C) were selected for further evaluation *in vitro* using a dose-response kinase assay with AbAStide to demonstrate the selectivity of the assay for identifying compounds correctly as inhibitors or noninhibitors. All three hits potently inhibited Abl kinase activity, with IC_{50} values of 0.52 nM, 1.91 nM, and 0.35 nM for GW693917A, GW711782X and GW513184X respectively, while the negative compounds gave no inhibition of Abl. (Fig. 4C) To determine whether the results of the *in vitro* inhibition studies translate to a CML model, the compounds were tested in cellular viability assays against the human CML cell line K562. Cellular IC_{50} for GW693917A was comparable to imatinib, at 81 nM compared to 147 nM. GW711782X and GW513184X were less potent in the cell viability assay, at 20 µM and 3.24 µM respectively. (Fig. 4D) Together these results demonstrate proof-of-concept that this strategy can produce an effective HTS assay for drug discovery applications.

2.6   Discussion

Synthetic peptide libraries are commonly used to identify determinants of kinase substrate specificity, these methods can be laborious to perform and require substantial quantities of purified kinase, which can limit widespread application. Here, we addressed these challenges by developing a straightforward computational strategy (KINATEST-ID™) which combines the identification of kinase specificity determinants with the prediction of kinase-substrate phosphorylation and peptide:Tb^{3+} complex formation, and used it to generate NRTK-specific biosensors for phosphorylation-dependent time-resolved Tb^{3+} luminescence detection. Traditionally, fluorescence-based kinase sensors have been generated through empirical design and iterative optimization, which slows down the pipeline for assay development. The design rules applied in KINATEST-ID™ facilitate substrate discovery by providing a set of *in silico* filters for sequence selection. The final sequences for AbAStide, SFAStide-A, and JAStide-E reported here demonstrate the utility of the design rules, yielding strong family-based selectivity and Tb^{3+} luminescence enhancement. Design of Tb^{3+} luminescence-based reporters of kinase

activity has previously been difficult to streamline since Tb^{3+} binding motifs are not trivially compatible with all kinase preference motifs. Moreover, overall similarity in consensus sequences among NRTKs necessitates a tradeoff between optimal activity and specificity. By taking into account the importance of each given site in a substrate sequence to the recognition and selectivity of the cognate kinase, we successfully achieved a balance between the confounding factors involved. These substrates exhibited robust dynamic ranges and signal to noise, and their potential for high-throughput assay compatibility was demonstrated in an inhibitor screen. As efforts to expand the characterization of kinase-specific phosphoproteomes increase through the application of recently developed methods,[97,100] the information available for generating the motifs, PSMs, and Screener™ selection for additional kinases will also expand. In next-generation applications of these sequences, incorporating docking motifs that target protein interaction domains could further increase the efficiency and potentially the selectivity of phosphorylation. Such modular substrates have previously been designed incorporating the D-domain and DEF-sites of Erk as well as the SH2 and SH3 domains of Abl and Hck.[136–138] We are currently pursing the application of these substrates in more complex mixtures of proteins, based on our previous work developing cell-deliverable substrates for Abl and Syk kinases,[83,139] in order to exploit their selectivity to measure the activity and inhibition of specific kinases in a heterogeneous environment.

Overall, while we validated the method with well-known kinase drug targets (Abl, Src-family, and JAK2) as a model system, the generality of the approach suggests that the KINATEST-ID™ strategy should be able to be applied to develop new assays for other kinases that are currently underexplored in drug development. Even though we focused here on tyrosine kinases, the *in silico* focused library generation tools could be used to develop new artificial peptide substrates for serine/threonine (S/T) kinases as well. Such substrates could be employed in any type of phosphorylation read-out, however the BLOSUM alignment component could be used with any detection-related motif desired. For example, Tb^{3+}-based S/T kinase detection requires a sensitizing chromophore such as tryptophan (W); accordingly, previously reported Tb-sensitizing, W- or unnatural amino acid-containing sequences (such as those reported by the Zondlo and Imperiali groups)[140–144] could be used for the BLOSUM matrix to focus and filter the virtual library for empirical

evaluation. This generality should make KINATEST-ID™ a useful approach to streamline the development of peptide-based kinase assays, as well as for broader applications towards other enzyme substrates or binding ligands for which sufficient training data are available.

2.7   Acknowledgements

2.8   Associated Content

2.8.1   Supporting information.

Additional characterization data for peptides, detailed motif characterization tables for the PSM, data from Michaelis-Menten kinetics characterization, and further information about HTS experiments is provided as supporting information. This material is available free of charge via the internet at http://pubs.acs.org.

CHAPTER 3.   HIGH-THROUGHPUT IDENTIFICATION OF FLT3 WILD-TYPE AND MUTANT KINASE SUBSTRATE PREFERENCES AND APPLICATION TO DESIGN OF SENSITIVE IN VITRO KINASE ASSAY SUBSTRATES

3.1   Abstract

Acute myeloid leukemia (AML) is an aggressive disease that is characterized by abnormal increase of immature myeloblasts in blood and bone marrow. The FLT3 receptor tyrosine kinase plays an integral role in haematopoiesis. One third of AML diagnoses exhibit gain-of-function mutations in FLT3, with the juxtamembrane domain internal tandem duplication (ITD) and the kinase domain D835Y variants observed most frequently. Few FLT3 substrates or phosphorylation sites are known, which limits insight into FLT3's substrate preferences and makes assay design particularly challenging. We applied *in vitro* phosphorylation of a cell lysate digest (adaptation of the Kinase Assay Linked with Phosphoproteomics (KALIP) technique and similar methods) for high-throughput identification of substrates for three FLT3 variants (wild-type, ITD mutant, and D835Y mutant). Incorporation of identified substrate sequences as input into the KINATEST-ID substrate preference analysis and assay development pipeline facilitated the design of several peptide substrates that are phosphorylated efficiently by all three FLT3 kinase variants. These substrates could be used in assays to identify new FLT3 inhibitors that overcome resistant mutations to improve FLT3-positive AML treatment.

3.2   Introduction

Acute myeloid leukemia (AML) is an aggressive cancer with a diverse genetic landscape. The FLT3 gene encodes for a receptor tyrosine kinase (FLT3) that regulates hematopoiesis and perturbations to its signaling pathways appear to promote AML disease progression. In fact, FLT3 is implicated as a major factor in AML relapse.[145] Thirty percent of AML cases have mutations to FLT3 that lead the kinase to be constitutively active,[146,147] most commonly to the juxtamembrane domain and the kinase domain.[146,148,149] Internal tandem duplication (FLT3-ITD) in the juxtamembrane or the first tyrosine kinase domain (TKD) occurs when a segment is duplicated (head to tail) leading to the loss of repressive

regions in the protein.[150] A second common mutation is a substitution of aspartic acid 835 to a tyrosine residue (D835Y) in the TKD. Both ITD and TKD mutants can activate and dimerize with the wild type FLT3.[151] The effects of these mutations on FLT3 signaling are still unclear, but one possibility is that mutant FLT3-TKD and FLT3-ITD activate alternative signaling pathways, or activate standard FLT3 pathways aberrantly, compared to the WT. Mutations to FLT3 are correlated with poor long-term prognosis[152,153] and while patients with FLT3 mutations achieve similar initial disease remission to those with wild-type FLT3, they have an increased risk for relapse.[146,152,154] *In vitro* studies show that FLT3-ITD mutant-expressing cell lines are resistant to cytosine arabinoside (the primary AML therapeutic).[152] These findings prompted the use of a combinatorial approach to AML therapies to include FLT3 tyrosine kinase inhibitors (TKIs), which are frequently initially successful but often lead to FLT3 inhibitor resistance and subsequent disease relapse.

The current FDA approved TKIs used to inhibit FLT3 were not developed specifically to target FLT3.[155–157] Sorafenib is a type II pan-TKI which is FDA approved for use in combinatorial approaches with AML chemotherapy, but elicits no response in FLT3 variants with tyrosine kinase domain mutations.[152,158–161] Efforts to develop FLT3 mutant-specific TKIs lead to the discovery of the type II TKI quizartinib, which can inhibit the FLT3-ITD mutant and is currently undergoing phase III clinical trials for AML.[162] Despite quizartinib's efficacy towards FLT3-ITD, it has no activity against FLT3-TKD point mutations and thus these mutations are the primary mode of quizartinib monotherapy resistance.[162–165] Quizartinib also has potent activity towards Platelet Derived Growth Factor receptor (PDGFR) and c-KIT kinases, and produces side effects that may be related to their inhibition in patients undergoing a FLT3 TKI regimen.[166,167] Crenolanib, a TKI designed to target the α and β isoforms of PDGFR, has demonstrated activity against a broad range of FLT3 mutations.[145,168] Unlike quizartinib, crenolanib does not inhibit c-KIT (the main kinase implicated in undesirable side effects of quizartinib) at safe plasma concentrations, and is undergoing phase II clinical trials in relapsed AML patients with a driver FLT3 mutation (clinical trial identifier NCT01657682).[166,169] Recent reports have shown that secondary point mutations within the kinase domain of FLT3 can reduce

crenolanib's clinical efficacy suggesting it is only a matter of time until crenolanib resistant mutations are found in a clinical setting.[166,169]

The complex abnormality landscape of AML reduces the possibility that a single FLT3 TKI would be a viable monotherapy for AML. Although crenolanib is a promising TKI, efficient development of new inhibitors will require better assays than those currently available, and adaptable strategies that effectively screen inhibitors to target mutant forms of FLT3 are especially needed.[162] Since very little is known about FLT3 substrate preferences, there are few options available when designing FLT3 activity assays. The current activity tests are limited by inefficient phosphorylation activity, and/or their phosphorylation by the mutant variants has not been characterized. In this manuscript, we describe the development of several novel and efficient peptide substrates for FLT3 and two clinically-significant mutant variants (the ITD and D835Y mutants). We adapted the "Kinase Assay Linked with Phosphoproteomics" (KALIP)[90,97] strategy (from the Tao lab) to perform high-throughput determination of FLT3's preferred peptide substrate motif in a manner similar to other previously reported methods (e.g. Kettenbach *et al* from the Gerber group).[100] In these approaches, a cell lysate digest is stripped of endogenous phosphorylation and used in a kinase reaction as a pseudo-"library" of peptides to determine kinase substrate preferences by identifying phosphorylated sequences via enrichment and mass spectrometry (ideal for high-throughput analysis of many substrates simultaneously without requiring radioactivity or other labeling).[170,171] We then used the identified substrate preferences to rationally design a panel of candidate peptides incorporating key sequence features predicted to make them favorable for phosphorylation by the FLT3 kinase variants, following our previously reported substrate development pipeline KINATEST-ID.[80] We demonstrated that these substrates enable efficient inhibitor screening for all three forms of FLT3. These peptides could be used in many different types of drug discovery settings to more rapidly and efficiently screen for and validate FLT3 inhibition.

3.3   Materials and Methods

3.3.1   Cell Culture and Endogenous Peptide Sample Preparation

KG-1 cells (ATCC) were maintained in IMDM media (Gibco) supplemented with 10% heat inactivated fetal bovine serum (FBS), 1% penicillin/streptomycin in 5% CO_2 at 37 °C. KG-1 cells were washed with 30 mLs of phosphate buffered saline (PBS) 5 times. The cells were then pelleted at 1,500 RPM for 5 minutes and lysed with buffer containing 8 M urea, 0.1 M ammonium bicarbonate pH 8.5, 20% acetonitrile (ACN), 20 mM dithiothreitol (DTT), and 1X Pierce Phosphatase Inhibitor tablet (Roche) pH 8.0. Lysed cells were incubated on ice for 15 minutes and then were subjected to probe sonication to shear the DNA. Lysates were treated with 40 mM iodoacetamide and incubated at room temperature (protected from light) for 60 minutes. Samples were then centrifuged at 15,000 RPM for 30 minutes to remove cellular debris. Urea concentration was diluted to 1.5 M using 50 mM ammonium bicarbonate buffer (pH 8.0) and the samples were set up for trypsin digestion at a 1:50 trypsin (ThermoScientific) ratio and incubated at 37 ˚C overnight. Trypsin digestion was quenched by adding 10% trifluoroacetic acid (TFA) in water to lower the pH below 3. Subsequently, the tryptic digest was desalted using hydrophilic-lipophilic balanced copolymer (HLB) reverse phase cartridges (Waters) and vacuum dried.

3.3.2   Alkaline Phosphatase Treatment

Samples were reconstituted in alkaline phosphatase dephosphorylation buffer containing 50 mM tris(hydroxymethyl)aminomethane hydrochloride (Tris-HCL), 0.1 mM Ethylenediaminetetraacetic acid (EDTA) at pH 8.5. Alkaline phosphatase (6 U, Roche) were added to each sample followed by incubation for 90 minutes at 37 ˚C. The reaction was quenched by incubating the samples in 75 ˚C for 15 minutes (Figure 3-2).

3.3.3   KALIP Recombinant FLT3 Kinase Assays

Recombinant kinases were purchased from EMD Millipore (WT, PN: PV3182; FLT3-D835Y, PN: PV3967; FLT3-ITD, PN: PV6190). The samples were briefly vortexed and aliquoted into two equal parts. Peptide samples were reconstituted in kinase reaction buffer containing 50 mM Tris HCL, pH 7.5, 10 mM $MgCl_2$, 1 mM DTT, 1 mM Na_3VO_4

and 2 mM adenosine 5'-triphosphate (ATP). The kinase (or water for control) was added last to each sample and incubated for 16 hours (16 H) at 37 ˚C. The FLT3-WT treatment contained an additional two-hour (2H) time point. The reaction was quenched by bringing up the concentration of TFA to 0.5% and desalted using Oasis HLB 1cc cartridge columns with 30-micron particle size and 30 milligram sorbent.

3.3.4   PolyMAC Enrichment

The phosphopeptide enrichments were carried out according to manufacturer's instructions (Tymora Analytical, West Lafayette, IN).[90,97] The enrichment kit is made up of four components: 1) loading buffer, 2) PolyMAC magnetic beads, 3) wash buffer 1 and 2, and 4) elution buffer. In brief, the dried peptides were resuspended in Loading Buffer and 100 µL of PolyMAC capture beads were added to the mixture. The phosphopeptide-PolyMAC mixture was mixed at 700 RPM for 30 minutes. Subsequently, the mixture was centrifuged briefly and placed on a magnetic rack to remove the un-phosphorylated peptide solution. The beads were washed twice with wash buffer 1 and rocked for 5 minutes at 700 RPM. The phosphopeptide-PolyMAC complex was placed on the magnetic stand until beads were immobilized by the magnet, and the supernatant was discarded. The process was repeated using wash buffer 2. The phosphopeptides were eluted from the capture beads using 300 µL of elution buffer and then vacuum dried.

3.3.5   LC-MS/MS Data Acquisition

Samples were reconstituted in 25 µL of mass spectrometry loading buffer (98/2/0.5%; H_2O/ACN/formic acid (FA)) and centrifuged for 30 minutes at 15,000 RPM. A 20 µL aliquot was transferred to a low binding safe-lock microtube (Eppendorf). A 2.5 µL aliquot was loaded on a ThermoScientific Easy NanoLC LC 1000 system. The reverse-phased HPLC peptide separation was performed using a 100 µm inner diameter Picotip emitter column packed in-house with 1.9 µm C18 ReproSil-Pur sorbent. The mobile phase consisted of 0.1% formic acid in ultra-pure water (Solvent A) and 0.1% formic acid in acetonitrile (Solvent B). Samples were run over a linear gradient (2-30% solvent B; 60 minutes) with a flow rate of 200 nL/min into a high resolution Orbitrap Fusion Tribrid Mass Spectrometer, operated using data dependent mode at a resolution of 60,000 with a

scan range of 300-1500 m/z. After each round of precursor detection, an MS/MS experiment was triggered on the top 12 most abundant ions using High Collision Dissociation (HCD). The mass analyzer parameters were set between two and seven charge states with a dynamic exclusion time of 15 seconds.

3.3.6   Data Analysis

3.3.6.1   Phosphopeptide identification

The Orbitrap Fusion mass spectra files were searched against a merged version of the reviewed human Uniprot database downloaded from uniprot.org (2/27/2017; 20,202 entries) and the cRAP database (common lab contaminants; downloaded from (thegpm.org/crap/) on 2/27/2017) using the Paragon algorithm in the ProteinPilot 5.0 proteomic search engine within the Galaxy-P pipeline to create a Distinct Peptide Report (the output report from ProteinPilot 5.0).[172–174] Peptide precursor mass tolerance was set at 0.02 Da and MS/MS tolerance was set to 0.1 Da. Proteomic database search parameters included trypsin digestion, urea denaturation, phosphorylation emphasis, iodoacetamide fixed modification to cysteine residues, and variable biological modifications. False discovery rate (FDR) analysis was activated for each individual search. ProteinPilot 5.0 used a reverse database as the decoy to calculate the false discovery rate (FDR) for each independent search.[175] We set the global 1% FDR score as our cutoff threshold.

3.3.6.2   Data processing and KINATEST-ID substrate candidate prediction.

3.3.6.2.1   Streamlined data processing of LC-MS data as input for KINATEST-ID algorithm substrate design

A series of novel scripts were developed to prepare and analyze the results from KALIP to design potential substrates in the KINATEST-ID platform. To extract and reformat the phosphopeptide sequences from the ProteinPilot distinct peptide report, we created the KinaMine program and GUI that extracts all sequences from a ProteinPilot 5.0 (SCIEX) Distinct Peptides Report output file that have phosphorylated tyrosine residues identified at a 99% confidence (1% FDR), and creates "Substrate" and "Substrate Background Frequency (SBF)" files, which contain the observed substrate sequences and the UniProt (uniprot.org) accession numbers and calculated representation of all amino

acids for the proteins from which substrate sequences were identified, respectively. We created the "commonality and difference finder.r" script to identify the phosphopeptides from the "substrates" and SBF files that are shared by all of the FLT3 kinase variants, and generated the "SHARED-16H" substrate and SBF files. We extracted the UniProt accession numbers from the SBF lists and used them to download a customized FASTA file from the UniProt website that contained entries only for those protein sequences, and converted that into .csv format using the "FASTAtoCSV" script. We created the "Negative Motif Finder.r" script to extract (and *in silico* trypsin digest) all tyrosine centered sequences present in any of the "background" proteins, and compared them to the substrate list (KinaMINE output) to return the sequences that were not observed in the phosphoproteomics data as a best estimate of "non-substrates."

3.3.6.2.2 Extraction and reformatting of phosphopeptide sequences from peptide ID results

The KinaMINE data formatter (Kinamine.jar) uses the Distinct Peptide Report and the FASTA file that was used in the proteomics search engine as input, filters the peptides from the report with a threshold of 1% FDR, to consolidate the sequences of all peptides that were phosphorylated in the experiment. It then outputs a .csv table (the "Positive Substrates" file, which is named by the user at the time of running the script) of those tyrosine-phosphorylated sequences, with each amino acid separated into an individual column and the phosphotyrosine aligned. This table also contains the accession number of the protein each peptide was from, which is used to extract the sequences of those proteins from the inputted FASTA file and calculate the "Substrate Background Frequency" (frequency of the 20 canonical amino acids found in each of the proteins individually; SBF), also output as a .csv. This .csv file also reports the total number of tyrosine residues within those protein sequences and the number of those tyrosine residues that were observed as phosphorylated in the experiment for subsequent use in determining FLT3's "normalization score" in the Screener module of KINATEST-ID (described below).

3.3.6.2.3 Phosphopeptide list comparison filtering

To select the sequences that were phosphorylated in common between the WT and the two mutant forms of FLT3, we developed a filtering script in R ("Similarity and

Difference Finder.R") to extract sequence lists and generate corresponding Substrate Background Frequency tables for the proteins corresponding to the selected peptides. This script provides either the intersection or symmetric difference between those sets as two new output tables containing only the information relevant to the sequences desired.

3.3.6.2.4 Approximating most likely "true negative" sequence list from substrate dataset

The accession numbers for proteins that remain in the Substrate Background Frequency list after the previous filter are submitted to the reviewed human Uniprot/SwissProt database (http://uniprot.org/uploadlist/) to generate a FASTA file containing the sequences of those proteins. The FASTA file is converted separately to .csv format using a script obtained from (https://www.researchgate.net/post/Converting_a_fasta_file_to_a_tab-delimited_file10). This file and the filtered Positive Substrates list file (generated as described in the previous section) are used as input for the "NegativeMotifFinder.R" to extract additional tyrosine-containing sequences from those proteins that could in principle have been phosphorylated but were not detected (outputting a "Negative Motifs" .csv file that is named by the user upon running the script). "Negative Motifs" files and corresponding "Positive Substrates" files are later used by the Kinatestpart1.R script to calculate Matthews Correlation Coefficient (MCC) values that give a general threshold for which peptides will or will not be phosphorylated by the kinase of interest.

3.3.6.2.5 KINATEST-ID streamlined processing in R

Using Substrates, Substrate Background Frequency, Non-substrate Motifs, and Screener.csv, the scripts "Kinatestpart1.R" and "Kinatestpart2.R" were written to replicate the functionality of the KINATEST-ID workbooks previously described,[80] including determining over- and under-representation of particular amino acids (the *standard deviation table/positional scoring matrix*) and/or side chain chemical properties (the *site selectivity matrix*) at particular positions relative to the tyrosine to define a preferred substrate motif, permutation of that motif into a list of all possible combinations of the preferred amino acids at their given positions (*Generator*), and scoring of those sequences against the targeted kinase's positional scoring matrix model as well as those for a panel

of other off-target kinases (*Screener*). These scripts ultimately create three .csv output files named by the user and containing the following information, respectively (all consistent with steps and components of the original published KINATEST-ID implementation)[80]: 1) the Standard Deviation and the AA Percent Tables; 2) the Site Selectivity Matrix, the Endogenous Probability Matrix (EPM) (which gives the scoring function used to calculate scores for a given sequence via the positional scoring matrix model defined by a given input dataset), the Normalization Score and the MCC Characterization Table; and 3) the list of predicted substrates ranked according to lowest "off-target" kinase scores using the Screener comparisons. For more information about these functions see the previously published description of KINATEST-ID.[80]

3.3.6.3 Peptide Synthesis and Purification

Peptides were synthesized using a Protein Technologies SymphonyX synthesizer using 4-methyl benzhydrylamine resin (Iris Biotech GMBH). Standard Fmoc-protected amino acid (AA) coupling occurred in the presence of 95 mM HCTU (Iris Biotech GMBH) and 200 mM N-methylmorpholine (Gyros Protein Technologies; S-1L-NMM) over two 20-minute coupling cycles. Fmoc deprotection occurred in the presence 20% piperidine in dimethylformamide (DMF, Iris Biotech GMBH;) over two 5-minute cycles. The peptides were purified to >95% purity by preparative C18 reverse phase HPLC (Agilent 1200 series) over a 5-25% acetonitrile/0.1% TFA and water/0.1% TFA gradient and characterized using HPLC-MS (Agilent 6300 MSD). Peptide substrates were dissolved in a PBS solution containing 5% dimethyl sulfoxide (DMSO). Absorbance measurements at 280 nm wavelength were used to determine the peptide concentration using the Beer-Lambert law (for which peptide extinction coefficients were calculated using Innovagen's peptide property calculator (https://pepcalc.com/).

3.3.6.4 In Vitro Kinase Assays

Recombinant kinases were purchased from SignalChem (FLT3, FLT3-D835Y, FLT3-ITD, Mast/stem cell growth factor receptor KIT, platelet derived growth factor receptor beta (PDGFRβ), anaplastic leukemia kinase (ALK), proto-oncogene tyrosine-protein kinase SRC, tyrosine protein kinase LYN and Bruton's tyrosine kinase (BTK).

Kinases were diluted to approximately 120 nM in kinase dilution buffer (20 mM 4-morpholinepropanesulfonic acid (MOPS) pH 7.5, 1 mM EDTA, 0.01% Brij-35, 5% Glycerol, 0.1% beta-mercaptoethanol and 1 mg/mL bovine serum albumin (BSA). Recombinant kinases (20 nM) were pre-incubated for 15 minutes in reaction buffer containing 25 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) pH 7.5, 10 mM $MgCl_2$, 100 µM ATP, 3 mM DTT, 3 µM $NaVO_3$ at 37 °C and 5% DMSO. The assay reaction was started by adding the peptide substrate to a 37.5 µM reaction concentration in a 30 µL volume. Sample aliquots (10 µL) were quenched by combining 1:1 with 30 mM EDTA at 2 and 60-minute timepoints (or 30-minute timepoint for TKI dose response assays).

3.3.6.5 Chemifluorescence Detection of Phosphorylation

Quenched aliquots were incubated in a 96-well streptavidin coated plate (125 pmol binding capacity, ThermoScientific) in Tris buffered saline (25 mM Tris-HCL and 150 mM NaCl) with 0.05% Tween 20 (TBS-T) containing 5% w/v skim milk for 1 h.[83] Subsequently, each well was washed with TBS-T and incubated with antiphosphotyrosine mouse monoclonal antibody 4G10 (MilliporeSigma, 1:10,000 dilution in TBS-T). Following incubation, wells were washed with TBS-T and incubated with horseradish peroxidase-conjugated rabbit anti-mouse immunoglobulin secondary antibody (Abcam, 1:15,000 dilution in TBS-T) for 1h. The wells were washed (TBS-T and 50 mM sodium phosphate buffer) and then treated with Amplex Red (Invitrogen, Carlsbad, CA) reaction buffer (0.5 mM AR, 20 mM H_2O_2 and sodium phosphate buffer) for 30 minutes. Fluorescence measurements were taken on a Neo2 microplate reader (Biotek, Winooski, VT) with an excitation wavelength of 530 nm and emission wavelength of 590 nm. The IC_{50} values were calculated by fitting the data to the equation below, were the $inhibition_{max}$ represents the lower plateau of the curve while the $inhibition_{min}$ pertains to the upper plateau, X represents inhibitor concentration and the steepness of the curve was set to a standard hill slope of negative one.

$$y = \frac{inhibition_{max} + (inhibition_{min} - inhibition_{max})}{1 + (X/IC_{50})}$$

3.3.6.6   Experimental design and statistical rationale
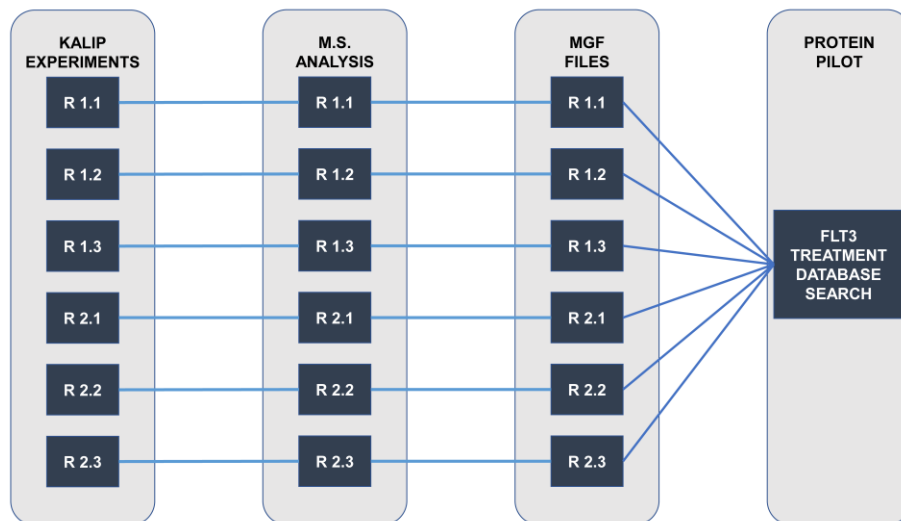
Figure 3.1. Schematic representation of raw mass spectrometer file combination for ProteinPilot database searches.

Each KALIP kinase treatment (WT, D835Y and/or ITD) was performed with three biological replicates (R1). The KALIP process was then repeated later to generate a second independent KALIP technical experiment (R2). Replicates were individually analyzed on the mass spectrometer and then converted to MGF files, ProteinPilot 5.0 database search consisted of six mass spectrometer files for each kinase treatment (no kinase or kinase treatment).

The trypsin digestion "library" preparation was performed for two independent replicates, each of which was subjected to an FLT3 kinase variant enzyme in triplicate (Figure 3-1). Upon file conversion of raw mass spectrometer files into a MGF format, the resulting six kinase treated files per FLT3 variant were combined into one ProteinPilot 5.0 protein identification search. These proteomic database search results were processed as described above (KinaMine and R data formatting scripts) to generate the KINATEST-ID input lists (Figure 3-2 step 5 and Figure 3-4).

A Dixon's Q test with an α value of 0.05 and a ROUT test with a q value set at 5 percent were performed to identify experimental outliers in the inhibitor dose response assay. Fluorescence measurements per well were normalized to the vehicle-only (DMSO) control signal, plotted using Prism software (GraphPad, La Jolla, CA) and fitted to a non-linear equation described above. An exact sum-of-squares F test with a p value set at 0.05

was performed to identify differences in the reported IC_{50} curves for each TKI against the FLT3 kinase variants.
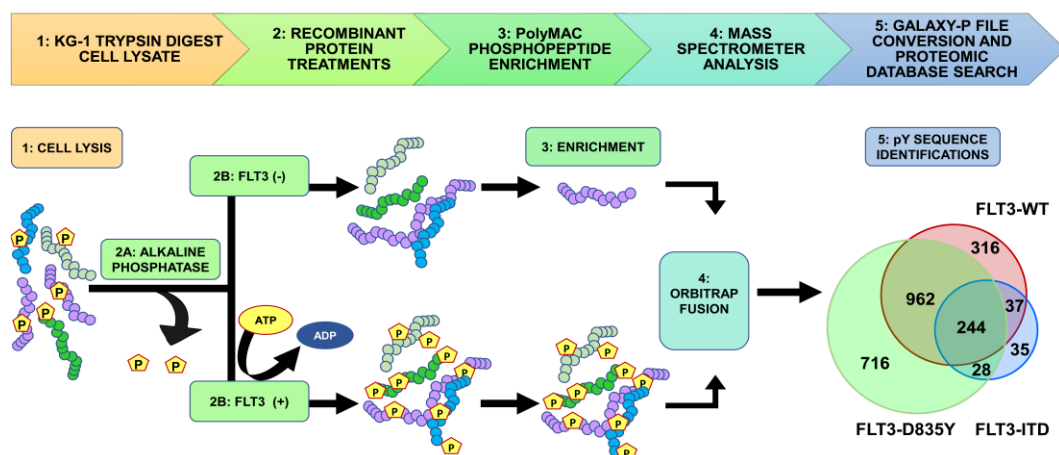
Figure 3.2. In vitro phosphorylation, enrichment and identification of substrate peptides from cell lysate.

To identify a larger number of FLT3 kinase substrates than was previously available, we subjected KG-1 cell lysate to trypsin digestion (Step 1). The tryptic digest was then treated with alkaline phosphatase (Step 2A). Subsequently, the sample was split into two equal parts (Step 2B) prior to in vitro recombinant FLT3 treatment (FLT3-WT, FLT3-D835Y and FLT3-ITD). The samples were enriched with PolyMAC magnetic beads (Step 3) and an aliquot was analyzed on an Orbitrap-Fusion mass spectrometer (Step 4). The mass spectrometer files were uploaded to the Galaxy-P proteomics pipeline for file conversion and ProteinPilot database search (Step 5).

3.4 Results

3.4.1 *In vitro* kinase reaction to identify substrates for input/analysis with the KINATEST-ID pipeline

Similar to the KALIP method[97] and others,[176] we used trypsin-digested cell lysate as a non-randomized peptide "library" to determine FLT3 kinase substrate preferences. Briefly, AML KG-1 cells were grown to log phase, lysed with a urea lysis buffer and digested with trypsin as described above. Following trypsin digest, peptides were treated with alkaline phosphatase to remove endogenous phosphorylation from tyrosines. The phosphatase-treated digest was then divided into aliquots that were processed in parallel: one treated with kinase reaction mixture but no kinase ("FLT3-"), alongside the kinase reaction for each version of the kinase (recombinant FLT3-WT, D835Y or ITD kinase,

respectively, "FLT3+") (Figure 3-2). The kinase reactions were performed for 16 hours (WT, D835Y and ITD) as described in the original KALIP protocol.[97] Following kinase treatment, phosphopeptides were enriched using the soluble polyMAC dendrimer[90,95,97] and analyzed on an Orbitrap Fusion mass spectrometer. Sequence and phosphorylation site identification was performed using the ProteinPilot 5.0 software on the GalaxyP platform (https://galaxyp.msi.umn.edu).

Overall, we identified more than 10-fold more substrates for FLT3 and the two mutant variants than we had curated for most of the kinases we had evaluated in our previous publication using KINATEST-ID, in which substrate numbers ranged from ~15 to ~170.[80] Using a relatively strict identification quality cut-off of 1% false discovery rate (FDR), we observed 1559 phosphorylated peptides from the kinase reaction with the WT FLT3 treatment, 2010 from the FLT3-D835Y mutant, and 344 from the FLT3-ITD mutant. Of these, 244 overlapped in common from each reaction (Figure 3-2). The FLT3-D835Y mutation stabilizes the activation loop within the kinase domain, leading to a stable, constitutively active kinase,[149,154] which likely explains the higher number of phosphopeptides observed in that reaction. The FLT3-ITD mutant is an in-frame gene sequence duplication that encodes for the amino acid segment connecting the juxtamembrane and tyrosine kinase domains. While the WT and D835Y versions of the kinase used in these KALIP experiments contained only a His tag (as described by the vendor), the ITD mutant was also tagged with a GST tag. Due to its larger molecular weight relative to the WT and D835Y variants, less of this kinase was used in the KALIP reactions, which likely accounts for the lower number of substrates observed for that variant. Nevertheless, compared to the manual curation of the literature-reported substrates and phosphoproteomic databases that was implemented previously,[80] even for the ITD mutant our approach identified many more phosphopeptide sequences than before, which could be used as input for the KINATEST-ID pipeline.

We used either the full list of substrate sequences from each kinase reaction ("WT-16H", "D835Y-16H" or the "ITD-16H" datasets), or the 244 common substrate sequences that had been identified from all three of the FLT3 variants' reactions (the "SHARED-16H" dataset), as separate inputs for initial positional scoring matrix (PSM) analyses in KINATEST-ID. The shared substrate sequences were extracted from the

phosphoproteomics data outputs using data processing and analysis tools, the KinaMine and Commonality and Difference Finder and the Kinatest part 1.r script (Figure 3-3), to generate PSMs of amino acid preference motifs from each dataset (Figure 3-4). We observed subtle, but likely functionally insignificant, differences in amino acid over- and under-representation at the positions -4 to +4 relative to the phosphorylated tyrosine for each of the different FLT3 variants. Generally for all variants, acidic amino acids were overrepresented and basic amino acids were underrepresented N-terminal to the phosphotyrosine, while hydrophobic amino acids and glutamine/asparagine were slightly overrepresented C-terminal to the phosphotyrosine. Given the lack of substantial differences between the PSMs for the WT and mutant variants, we focused on the substrates that were observed in common for all three kinase variants to move forward with design of novel peptides that could be used as substrates in FLT3 activity assays.
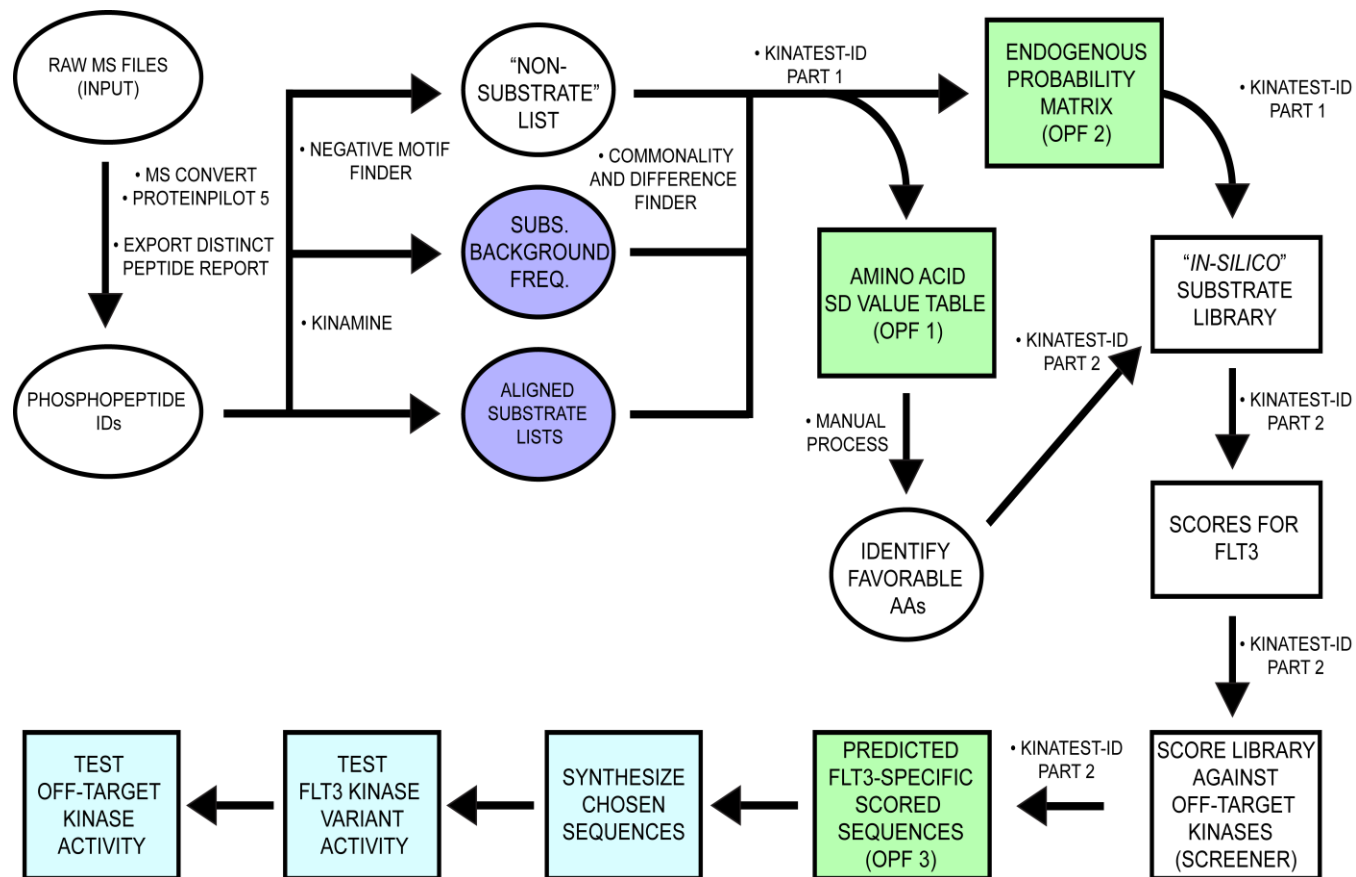
Figure 3.3   Conceptual overview of our KALIP data processing and formatting for KINATEST-ID incorporation to develop FLT3 artificial substrates.

Figure 3.3. Continued

Conceptual overview of our KALIP data processing and formatting for
KINATEST-ID incorporation to develop FLT3 artificial substrates.

After conversion and peptide/protein ID in the Galaxy-P proteomic pipeline, the
KinaMine data formatter tool extracted the confidently-identified (1% FDR)
sequences that contained a phosphorylated tyrosine residue (pY), centrally aligned
the sequences to the tyrosine of interest ("Aligned substrate lists") and extracted the
UniProt accession number with the accompanying proteins' amino acid
composition file ("Subs. Background Freq."). The Negative motif finder script uses
the Uniprot accession numbers to generate a list of tyrosine-containing potential
tryptic peptides that were not observed in the phosphopeptide dataset. The scripts
KinatestID part1.r and -part2.r processed those input files to identify substrate
preferences and generate a ranked list of candidate sequences as potential FLT3
substrates. Kinatestid-part2.r then scores the input substrate and non-substrate lists
and outputs as two additional files. After performing this workflow on
phosphopeptide data from FLT3 kinase reactions, chosen candidate sequences were
synthesized and validated in vitro against the FLT3 kinase variants. The candidate
sequences were then assayed in vitro against a panel of kinases to determine off-
target kinase activity. In silico/predictive steps are illustrated in white/green.
Empirical steps of synthesizing peptides and characterizing FLT3 activity and
specificity are depicted in light blue.

3.4.2 KINATEST-ID-based design of novel FLT3 Artificial Substrate peptides (FAStides)

We then used the "SHARED-16H" dataset and employed the next steps of the KINATEST-ID approach to design a set of candidate sequences for synthesis and biochemical testing. This process used the KINATEST-ID "Generator" tool (via the Kinatest part 2.r script) to create a list of sequences comprising all the permutations of the amino acids overrepresented at each position by at least two standard deviations from the mean (Figure 3-4A-B). One caveat was that while tyrosine was observed as overrepresented at -1 and +1 to the phosphotyrosine, we chose to exclude it from the preference motif for this iteration of substrate design, due to the potential ambiguity in assay signal that could ultimately be introduced by having more than one phosphorylatable residue in the designed substrates. Additionally, F was included as an option at position +1 (despite having relatively low representation) as a hydrophobic alternative to the more highly represented I and V in an attempt to provide better specificity, due to the high frequency of those two amino acids at that position in the motifs for other tyrosine kinases. Permutation of the motif was then followed by scoring of the resulting 19,201 sequences against the PSMs for WT FLT3, the two mutant variants, and a panel of other kinases,[80] using the KINATEST-ID "Screener" tool (Figure 3-3). The sequences and their scores against the PSMs are summarized in Figure 3-4C while their "off-target" kinase PSM scores are summarized in Figure 3-7G. We chose one set of sequences that scored well for FLT3 but scored poorly for other kinases (sequences A, D, G and H). Since the highest scoring sequences for FLT3 also scored well for several other off-target kinases in the Screener panel, we selected another set of those sequences (sequences B, C, E and F) to have a higher likelihood of obtaining an efficient (though potentially not FLT3-specific) substrate. Additionally, we synthesized two control sequences that were previously reported to be phosphorylated by FLT3 (ABLtide and FLT3tide).[177,178]
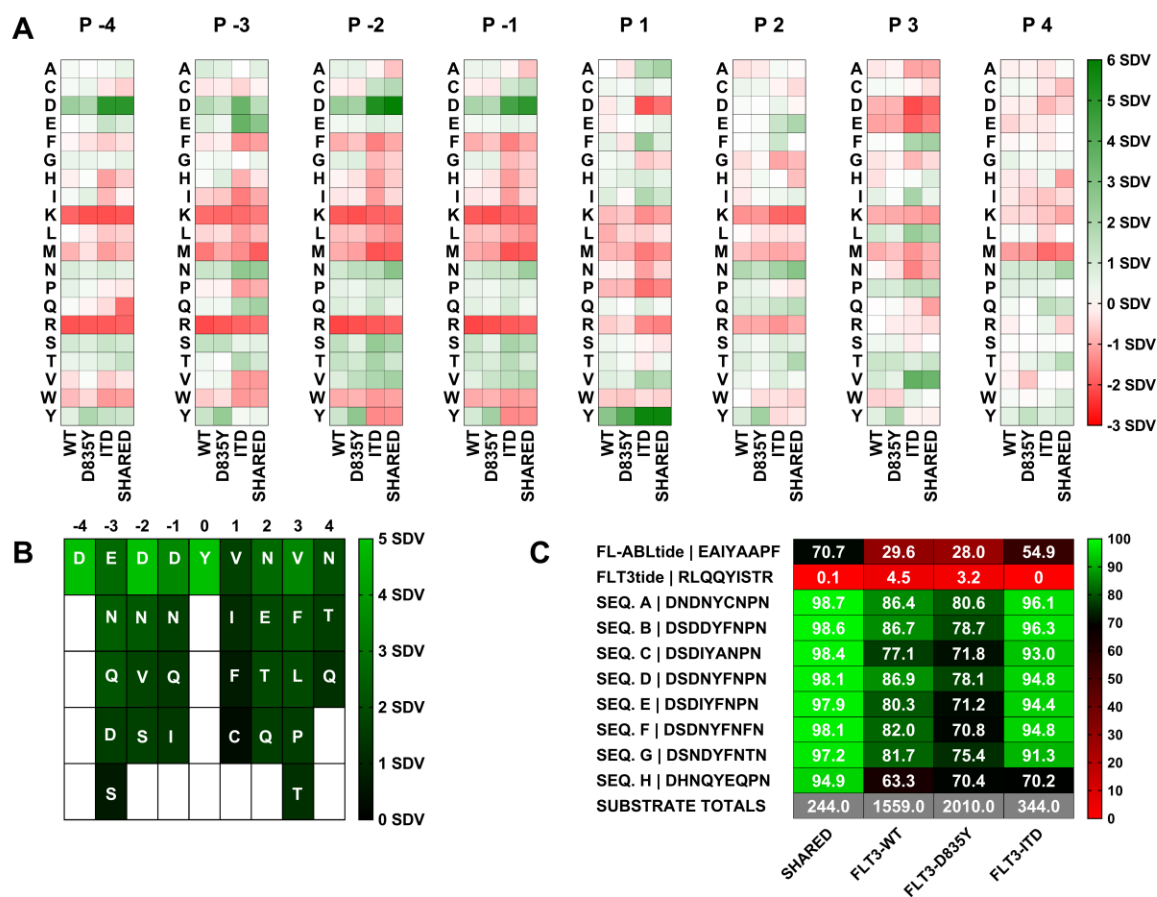
Figure 3.4 Positional preferences, motif, and substrate candidate mini-library for FLT3-WT, FLT3-D835Y and FLT3-ITD

(A) Observed representation of each amino acid at each position (-4 to +4 relative to phosphotyrosine) in the individual phosphoproteomics datasets for the three FLT3 variants (WT, D835Y, and ITD) and for the sequences shared in all three datasets (Shared). Green = over-represented, white = neutral, red = under-represented. (B) Table summarizing the positional preferences (>1 standard deviation from the mean) used in the "Generator" tool portion of Kinatest part 2.r for permutations to rationally design candidate substrate sequences. Representation of the given amino acids at respective positions is shown via color scale for reference. Sequences of selected, synthesized candidate substrates and controls and scores (illustrated with color scale) against the respective scoring models for each dataset (shared and individual variants).

3.4.3   *In vitro* validation of FAStide sequences as FLT3 kinase variant substrates

We tested the candidate sequences via *in vitro* kinase assays using anti-phosphotyrosine antibody 4G10 in a chemifluorescent Enzyme-Linked Immunosorbent assay (ELISA) as the readout to detect peptide phosphorylation, with sample aliquots

quenched at 2 and 60 minutes. Figure 3-5A shows that the sequences that were most efficiently phosphorylated by FLT3-WT, D835Y and ITD contained the DXDXYXNXN motif. Figure 3-5A shows that S was well-tolerated at position -3 (while N and H were less tolerated), both N and D were tolerated at position -1, and F was the preferred amino acid at position +1 while A was not well tolerated. Sequences that contained F, P or T residues at position +3 were phosphorylated by all FLT3 kinases. The control peptides (FL-ABLtide and FLT3tide) were also scored against the PSMs for all of the datasets (Figure 3-4C) and assayed in parallel with the FAStide sequences against the FLT3 kinase variants. The previously reported substrate FLT3tide[177] scored poorly against our models and was a poor FLT3 substrate in our assays (Figure 3-5A). ABLtide, a previously reported FLT3 substrate,[178] scored moderately against the SHARED-16H dataset model, and performed moderately as a substrate.

3.4.4 Evaluating the relationship between substrate input datasets and resulting PSM model scores vs. biochemical assays

FAStide sequences A, B, D, E, F and G generally had higher PSM scores than C and H, and were all phosphorylated more efficiently than C and H in the assays. The FLT3tide reference peptide scored poorly in all the matrices and was phosphorylated very poorly in the assays, and FL-ABLtide scored moderately and was phosphorylated moderately relative to the best of the FAStide sequences in the assays. For the longer KALIP kinase reactions (16 hours), the strongest correlations were found between each variant's biochemical assay results at 60 min and the WT-16H dataset-derived PSM scores (Figure 3-4C). The FAStide sequences scored the lowest with the D835Y-16H PSM model, which contained the largest substrate list (2010 substrates), and their scores had lower Spearman correlations with assay results from the WT, D835Y, and ITD variants, respectively. This is most likely attributable to a combination of the larger size of the input dataset for that PSM and that the FAStides were designed based on the selected subset of that data shared with the other two variants, rather than the entire dataset used to derive this PSM model. The FAStide sequences received higher scores using PSM models with the smallest substrate lists (SHARED-16H and ITD-16H), which was also primarily an artifact of all or nearly all of the sequences in those smaller datasets being used to design the FAStides in the first place. Sequence scores derived from PSM models with the lowest

number of input substrates were less correlated with the biochemical assay results, which might arise from smaller dataset artifacts or some other, as yet unidentified factor affecting the accuracy of the predictive models from those datasets.

3.4.5 Evaluating the length of time of *in vitro* kinase treatment on substrate motif prediction

We also examined the effect of reaction time in the kinase treatment step on the ability of the KALIP-KINATEST-ID process to identify efficient substrates that can be used in enzyme assays. We performed a two-hour kinase reaction using FLT3-WT kinase and processed the data (referred to hereafter as WT-2H) as described above to determine if the KALIP kinase treatment time affected 1) the characteristics of the preference motif arising from a given dataset, and 2) its utility for subsequent substrate design. We identified 888 phosphopeptides from the 2H KALIP FLT3-WT kinase treatment (relative to 1559 for the 16-H treatment as described above). We compared the "WT-2H" to the "WT-16H" substrate list and found 559 sequences shared by both. These sequences, referred to as the "WT-OVERLAP" substrate and background frequency lists, represented sequences that were likely to have been phosphorylated rapidly and robustly (Figure 3-5B). The corresponding SDV values were compared to those for each substrate list from the WT-2H and WT-16H experiments, as shown in Figure 3-6.

Overall, the preferences at each position as represented by the SDV tables were similar (Figure 3-6), however subtle differences in the WT-2H dataset from the two hour incubation resulted in a scoring model that appeared to more accurately reflect the substrate phosphorylation efficiency for the WT kinase as observed in the biochemical experiments relative to the models derived from the WT-16H dataset. PSM scores derived from the WT-2H, WT-16H and WT OVERLAP datasets were compared with the assay results for the eight FAStides and two control peptides (Figure 3-5A). Spearman correlations are shown in Figure 3-5C. Similar to the WT-16H, the assay signal at 60 min and the PSM scores generated via the WT-2H and WT-OVERLAP datasets were very highly correlated, appearing slightly stronger for WT-2H than for WT-OVERLAP and WT 16H. This may indicate that shorter KALIP kinase treatment time is better for determining more efficiently (i.e. rapidly) phosphorylated substrates, and that longer treatment may "dilute" the substrate preference motif with less efficient sequences.
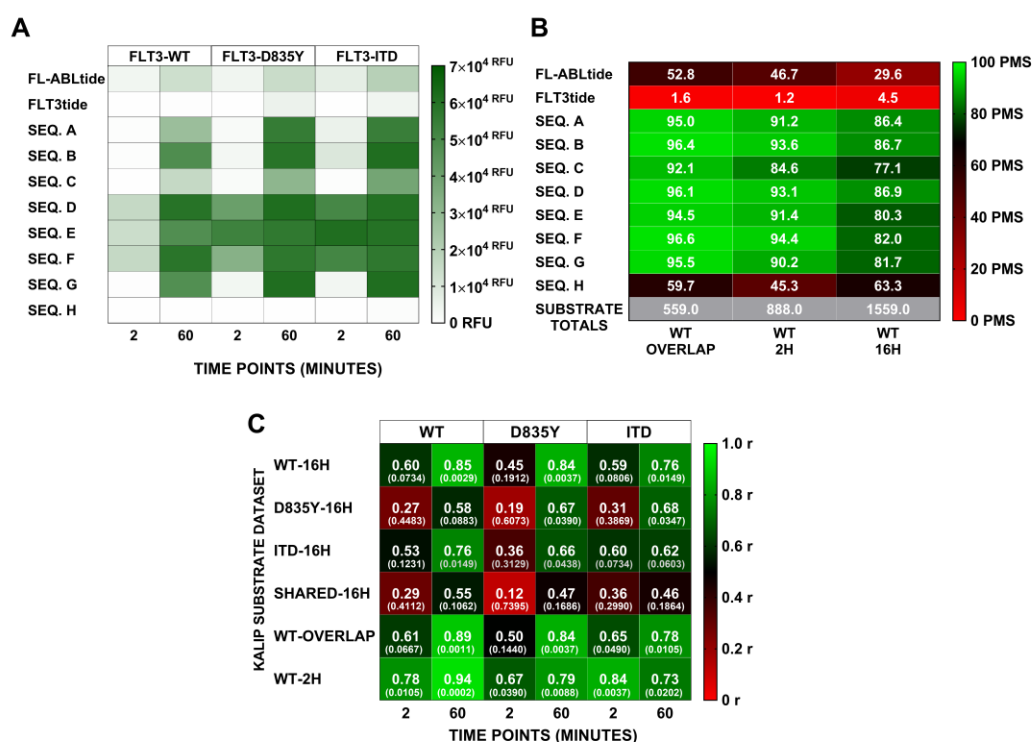
Figure 3.5 FLT3 kinase variant activity assays and scoring model correlation.

(A) FLT3 (WT, D835Y and ITD) in vitro kinase assay results for the candidate sequences (A, B, C, D, E, F, G and H). Each FLT3 variant (columns WT, D835Y, ITD) was reacted with each peptide (rows), with aliquots taken at 2 and 60 min (sub-columns). Phosphorylation levels indicated by RFU from ELISA detection, illustrated via color scale (green = high phosphorylation, white = low/no phosphorylation). (B) Effect of in vitro KALIP kinase assay incubation time on KINATEST-ID scores for candidate sequences. Positional scoring model scores for each sequence (rows) against the models (columns) derived from all phosphopeptides observed in the WT 2H or 16H KALIP kinase reaction, respectively, or just those observed in both (WT OVERLAP). Color scale from red (low PMS score) to green (high PMS score). (C) Spearman's non-parametric correlation test to measure correlation between the KINATEST-ID's scoring system and the in vitro activity assay results. r values shown in bold, above p-values denoted in parenthesis. Color scale indicates low (red) to high (green) Spearman r values.
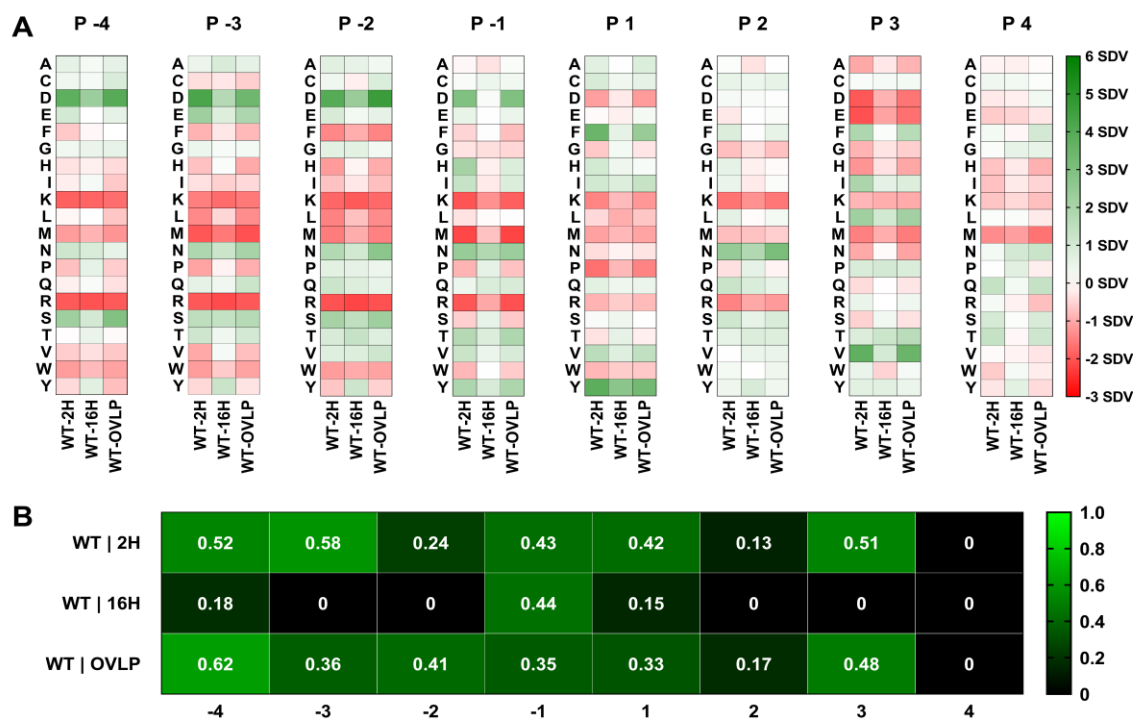
Figure 3.6 Heat map representation of FLT3-WT time course KALIP experiment, Site Selectivity Matrix and artificial substrate library sequence scoring comparison.

(A) Observed representation of each amino acid at each position (-4 to +4 relative to phosphotyrosine) in the individual phosphoproteomics datasets for the kinase treatments at two hours (WT-2H) or sixteen hours (WT-16H), or for the sequences shared in the two datasets (WT-OVLP). Green = over-represented, white = neutral, red = under-represented. To summarize, differences were modest between the two treatment times. (B) We compared the three substrate lists' SSM values to identify positions with a value greater than 1, which is the previously reported threshold used to consider a position as "significant."[80] None of the KALIP dataset SSMs contained a position with a value greater than one, suggesting that all positions exhibited some flexibility for which particular amino acid was present.

3.4.6 *In vitro* characterization of FAStide FLT3 specificity

While *in vitro* recombinant kinase assays using the novel peptides identified here would not require exquisite specificity since they would typically employ purified kinase, we wanted to perform a limited assessment of the "off-target" phosphorylation of these peptides using a small panel of other kinases: KIT, PDGFRβ, ALK, SRC, LYN and BTK. KIT and PDGFRβ are kinases for which the substrate preference motif has not yet been identified, however, they are within the same kinase family as FLT3 and based on observations for other kinases,[80] were likely to phosphorylate the same FAStides as FLT3. Based on the previously published PSM scoring models[80] (derived from manual substrate list curation and not the KALIP process), SRC was predicted to phosphorylate sequences B, C, E and F while LYN is predicted to phosphorylate sequences A and C. BTK was not expected to phosphorylate any of the FAStide sequences. Additionally, our panel included the receptor tyrosine kinase, ALK, which we previously observed to phosphorylate similar sequences to those phosphorylated by SRC.[81] To ensure each of those recombinant kinases was active, we performed an *in vitro* kinase assay with several reference peptides that have been previously characterized in our laboratory for the kinases in the panel. The "universal" tyrosine kinase substrate that we previously reported U5 (DEAIYATVA)[34] was the reference peptide chosen for KIT, PDGFRβ and BTK, and an ALK substrate we previously reported (ALAStide)[81,179] was used for ALK (Figure 3-7A-F). SFAStide-A (DEDIYEELD)[80] was used as the reference peptide for SRC and LYN kinases. Briefly, the kinases were preincubated with kinase reaction mixture and the *in vitro* reaction was initiated by the addition of substrate peptide. The samples were quenched at 2 and 60-minute time points as described above. Phosphorylation was measured using the previously described ELISA-based assay,[80,83] and results are shown in Figure 3-7.

Overall, the sequences with the least off-target phosphorylation in this panel were FAStide-A, which was phosphorylated moderately by c-KIT and PDGFRβ (both FLT3 family members) and LYN after 60 minutes but not the others in the panel, and FAStide-G, which was phosphorylated to a relatively low degree after 60 minutes only by c-KIT and SRC (Figure 3-7A and 3-7E). FAStide-B, -D, -E and -F were phosphorylated by more of the off-target kinases by 60 min and FAStide-C was a robust substrate for both SRC and LYN. PDGFRβ and BTK, on the other hand, did not phosphorylate any of the artificial

sequences over a 60-minute incubation. The off-target *in vitro* kinase assay results were mostly but not entirely consistent with the Screener predictions, which is not surprising given that Screener is limited by the PSM models built into its cross-referencing algorithm—the main caveat is that the PSM models in Screener all come from the previously developed KINATEST-ID package[80] that did not have KALIP phosphoproteomics data as input. A future goal is to update current kinase PSMs with newly generated KALIP data, as well as adding data from more kinases to improve the cross-referencing depth.
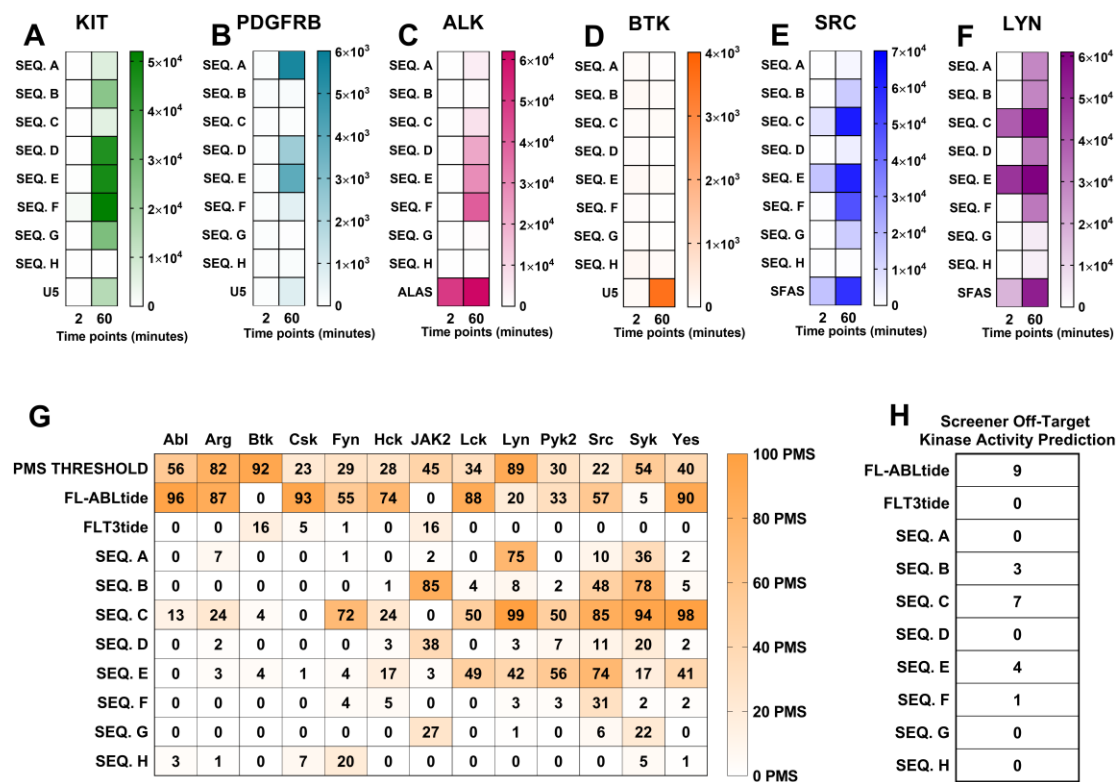
Figure 3.7 The focused peptide library of kinases predicted to tolerate a given sequence as a substrate.

(A-F) Phosphorylation level for each peptide (rows) at two time points (2 and 60 min, columns) is shown separately for each kinase, with color scales normalized to minimum/maximum RFU values from ELISA detection for that kinase. FAStide candidate data shown in rows 1-8, reference peptide data show in row 9 as a positive control for kinase activity. (G) PMS scores for each FAStide candidate and two reference peptides (rows) for 13 kinases (columns) previously built into the KINATEST-ID Screener tool from literature curated substrate input lists. Scores are indicated by color scale from low (white) to high (orange). (H) "Off-target" activity prediction from Screener tool, indicating the number of kinases predicted to tolerate a given sequence as a substrate.

3.4.7 Detection of FLT3 kinase variant inhibition through FAStide *in vitro* phosphorylation

To demonstrate how the FLT3 artificial substrates can be used to monitor TKI efficacy, we performed dose-response (DR) assays for FLT3-WT, FLT3-D835Y and FLT3-ITD with sorafenib, quizartinib and crenolanib, three TKIs that have been characterized against the three FLT3 variants.[164,166,169,180] FAStide-E and FAStide–F were chosen for the DR assays due to their efficient phosphorylation by all three FLT3 kinase variants, and employed in parallel experiments. Each FLT3 kinase variant was pre-incubated in the kinase reaction mixture (containing ATP) with the respective TKI (0.00001 to 100 nM) for 15 minutes at 37˚C without substrate, and the kinase reaction was initiated via the addition of the substrate (37.5 µM). Reactions were quenched after 30 min and wells analyzed using ELISA, as described above. Fluorescence values (relative fluorescence units, RFU) were collected and normalized to values for wells containing vehicle control (DMSO). In general, both substrates exhibited dose-response curves and IC_{50} values that were consistent with what was expected for the given inhibitor against each FLT3 variant, with one notable exception (further described below) (Figure 3-8, Table 1). All three inhibitors potently inhibited WT FLT3 (with IC_{50} values in the ~1-30 pM range). Potency towards the ITD mutant was lower for all three inhibitors (IC_{50} values between ~40-800 pM), with crenolanib more potent than the other two. The D835Y mutant's dose response curves were also as expected for sorafenib and crenolanib, with sorafenib being significantly less potent than it was against the WT (~200-250-fold higher IC_{50}), while crenolanib maintained pM-range IC_{50}. For quizartinib, on the other hand, dose response curves were different for the assays performed using FAStide-E compared to FAStide-F. Quizartinib is a type II inhibitor, known to bind to the inactive "DFG-out" conformation of the kinase. Bulky, hydrophobic mutations at position 835 in FLT3 are thought to confer resistance to quizartinib, due to the effects of the side chains on the structure and dynamics of the DFG loop in the kinase domain, with the extra steric bulk disrupting the stability of the inactive conformation.[181,182] The FAStide-E quizartinib dose response results for the D835Y mutant were consistent with this model, with essentially no significant inhibition even at concentrations as high as 100 nM (>10,000-fold higher than the IC_{50} for quizartinib against the WT FLT3). However using FAStide-F as the substrate, inhibition was observed

in the same IC_{50} range as for sorafenib. This suggests that substrate interactions may affect inhibitor binding stability, perhaps by playing a role in DFG-in/-out dynamics.

FAStide E
DSDIYFNPN

FAStide F
DSDNYFNFN
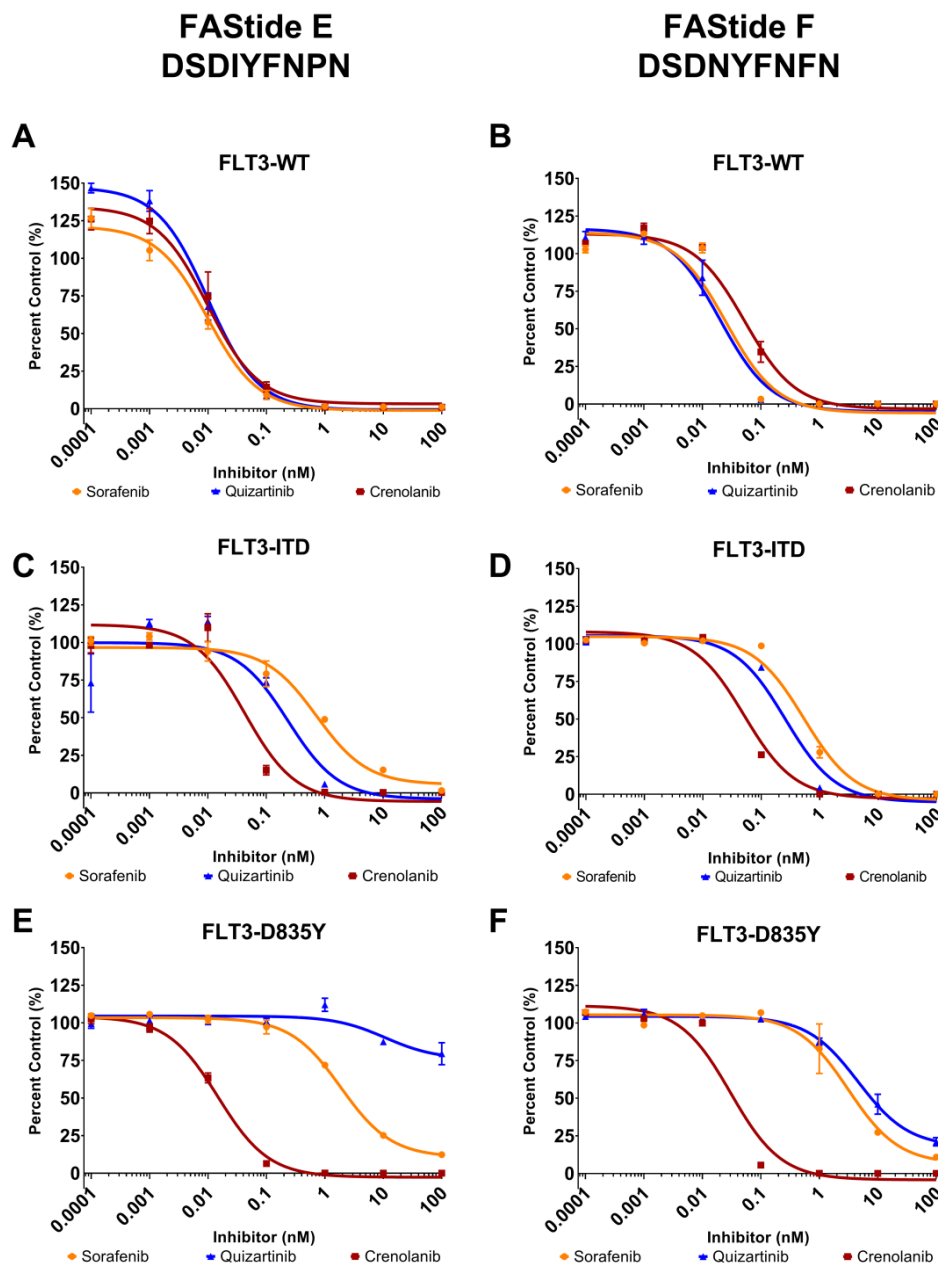
Figure 3.8 Monitoring FLT3 kinase activity and inhibition by clinically relevant tyrosine kinase inhibitors (TKI).

Dose-response assays were performed using two substrates, FAStide-E (A, C, E) and FAStide-F (B, D, F) in the presence of increasing TKI concentrations (sorafenib, quizartinib and crenolanib). Data show RFU values for each inhibitor normalized to vehicle control (DMSO; representative of 6 independent reactions) and plotted as percent control. IC50 values were calculated from fitting the data to a fixed slope (three parameter) curve in GraphPad Prism. Data points/error bars represent mean ±SEM.

Table 3-1. IC_{50} values measured by monitoring the phosphorylation of FAStide-E or -F in ELISA-based assays.

Table 1: TKI IC_{50} VALUES (pM) +/- S E						
TKI	FLT3-WT		FLT3-ITD		FLT3-D835Y	
	Seq. E	Seq. F	Seq. E	Seq. F	Seq. E	Seq. F
SORAFENIB	8.7±3.5	13±11	780±300	550±81	1,800±300	3,200±940
QUIZARTINIB	8.6±1.3	32±20	190 ±110	266 ±49	NA	4,700±870
CRENOLANIB	13±3.8	1.3±0.92	45 ±14	52 ±8.9	15±1.4	32±7.9

3.5   Discussion

Drug resistance in AML has been a major factor in the poor 5-year survival and clinical remission rates. Treatments targeting patients with FLT3-positive AML have seen promising results, but inhibitor resistance has been detrimental to clinical efficacy. FLT3 remains a viable drug target in AML,[72,178] however, none of the current TKIs used in therapy are FLT3 specific and even once those are developed, rapid emergence of mutations that abrogate drug binding will be a continuing challenge.[155,156,183,184] This highlights the importance of having efficient assays that can be used as tools to identify specific and selective TKIs that target FLT3 and mutant variants. In this work, we coupled *in vitro* kinase reactions on cell lysate digests with the KINATEST-ID pipeline to design, synthesize and validate a panel of sequences to detect the activity of a kinase that has few known substrates. Our process created a novel panel of peptides that can be used in kinase assays and provide higher phosphorylation efficiency than previously reported substrates. These findings demonstrate how the streamlined combination of KALIP and the KINATEST-ID pipeline can be used to identify novel artificial kinase substrates.

The original KINATEST-ID pipeline[80] relied on literature-validated sequences, including some from proteomic databases and positional scanning peptide library (PSPL) assays, as input for the matrices. Each sequence was manually curated by further literature examination (looking for corroboration of upstream kinase evidence via e.g. testing of non-phosphorylatable A/F mutations). This severely limited the number of substrates that could be included in the "true positive" input list, which likely resulted in less accurate

predictions of optimal substrate sequences. While this was sufficient to develop effective substrates for several kinases that showed reasonable degrees of selectivity for their targets,[80] it was not optimal—and further, if a given kinase did not have sufficient known substrates or PSPL assay data available then it was not possible to make any prediction at all. Using cell lysate digest as a peptide library for high-throughput identification of FLT3 kinase variant substrates enabled a large increase in the numbers of *bona fide* substrate sequences that could be used to build positional scoring models.

The improved positional scoring matrix models developed from these large, empirically detected substrate sequence datasets enabled a prediction for FLT3's preferred amino acid motif, which was then used to design several potential novel substrate peptides. Scores from the positional scoring matrix models correlated well with the relative biochemical behavior of the novel substrates, especially when the input dataset comprised sequences that were observed as phosphorylated after a short kinase incubation time (2 h, which is closer to the reaction time scale used for biochemical assays in practice). This suggests that even though endogenously-derived tryptic peptide libraries are somewhat biased relative to randomized/unbiased synthetic libraries[170] (given the sequence constraints imposed by their genomic origin), they are still able to provide sufficient sequence diversity to enable discovery of hundreds to thousands of substrates and accurately reveal substrate preferences for a given kinase. It also suggests that although performing the reaction at the protein level[171] may be better for developing prediction models for identifying endogenous protein substrates, performing substrate preference analyses at the peptide level *in vitro* is sufficient for designing peptide probes.

Intriguingly, we observed substrate-dependent inhibition for the well-characterized TKI quizartinib against the FLT3 D835Y mutant. This suggests that the particular substrate used in a screening assay might bias the interpretation of whether an inhibitor is or is not potent against a given enzyme. This highlights the importance of the substrate in inhibitor assays, and suggests that expanding the range of efficient substrates available as drug discovery tools would be beneficial. Work is ongoing to determine whether this is specific to the FLT3 D835Y mutant or is a more general issue for kinases. Other next steps will be to expand the application of this approach to the kinases previously built into the KINATEST-ID "Screener" panel,[80] in order to improve the accuracy of the positional

scoring matrix models for the "off-target" kinases and achieve better predictions of selectivity during the substrate design process. While the original, previously published Screener panel was accurate enough to offer the practical ability to pre-filter a large list of potential sequences down to a more manageable number, clearly the selectivity prediction was limited by the same factors (comprehensiveness of the input dataset) as the preference prediction. Ongoing efforts to apply the KALIP adaptation approach reported here to more kinases should facilitate improvement of this aspect, as well.

In summary, in this work we demonstrate the utility of generating a large dataset of *bona fide* substrate information, using a relatively cheap and easily produced peptide pseudo-"library" derived from cellular proteins via proteolytic digest, for defining substrate preference motifs and scoring models that enable design of efficient peptide substrates for kinase enzymes. This strategy enabled discovery of multiple substrates, some of which may influence inhibitor interactions with the enzyme and affect conclusions about inhibitor efficacy. We also anticipate that this process can be applied to orphan kinases for which little to nothing is known, to first identify substrate sequences through the *in vitro* proteolytic peptide library kinase reaction followed by prediction of the preference motifs from those data. Those motifs could be used to design, synthesize and validate artificial substrates, which can assist in chemical biology and drug discovery efforts to identify novel and potent inhibitors to study their biology and/or become therapeutic leads. Furthermore, this workflow could potentially be generalized to any enzyme-driven disease for which substrate preference data can be determined from proteolytically or synthetically prepared peptide libraries[185] and used to design novel substrates for use in assays. This will greatly enhance the generalization of the novel substrate probe design process we initially implemented in our first report of KINATEST-ID,[80] broadening the scope for drug discovery assay development.

CHAPTER 4.   DEVELOPMENT OF A GALAXYP WORKFLOW FOR HIGH-THROUGHPUT IDENTIFICATION OF BTK KINASE SUBSTRATE PREFERENCE AND DESIGN OF ANTIBODY AND ANTIBODY-FREE ACTIVITY ASSAYS

4.1   Abstract

Protein kinases play an integral role in the addiction of cancer cells to pro-survival regulatory mechanisms to drive tumorigenesis. BTK is involved in pro-survival signaling pathways in B-cell signaling and malignancies. BTK has been observed to be an important kinase that non-B cell malignancies and solid tumors also use to promote survival, which has made it viable drug target. The limited information known about BTK's preferred phosphorylation motif sequence limits the ability to develop sensitive activity assays that can be used in drug discovery to develop new kinase inhibitors. We applied the phosphoproteomic approach described in chapter three to increase the known number of BTK substrates. We incorporated the data formatting tools developed in chapter three to process and perform the KALIP-KINATEST-ID analysis within the GalaxyP framework. This streamlined processed was used to generate sensitive antibody-free assays to monitor BTK kinase activity.

4.2   Introduction

The *BTK* gene was identified in the 1980s as a gene encoding for a non-receptor tyrosine kinase that is now part of the Tec protein kinase family.[186] Under normal circumstances, BTK is localized to the cytosol and autoinhibited by intra-domain binding. BTK contains a plecksrtin homology and Tec homology (PHTH), Src homology 2 (SH2), SH3, and kinase domains.[187–189] BTK is activated and recruited to the cell membrane when the PH segment, of the PHTH domain, binds to phosphatidylinositol-3,-4,-5-triphosphate (PIP3).[187–189] Deletion of the PHTH domain does not alter kinase activity, but it is required for BTK substrate recognition.[187] PHTH binding of PIP3 allows for the SH2 and SH3 domains to become available to the cytosolic proteins.[190] The SH2 domains binds to phosphotyrosines while the SH3 binds to proline rich regions. SRC family and spleen

tyrosine kinases (SYK) phosphorylate tyrosine 551 in the kinase domain leading to autophosphorylation of Y223 within the SH3 domain.[189]

BTK is predominantly expressed in B cells and plays a role in B cell receptor (BCR), chemokine, and Toll-like receptor (TLR) signaling. Upon ligand binding to the BCR, SRC family kinases phosphorylate immunoreceptor tyrosine-based activation motifs (ITAMs) leading to SYK cell membrane localization. This signaling cascade recruits PI3K increasing PIP3 concentration and phosphorylation enabling the recruitment BTK to the cell membrane. Upon recruitment, BTK is phosphorylated by SRC and SYK family kinases.[186,189] Once activated, BTK phosphorylates PLCγ2 to initiate and regulate production of inositol triphosphate and diacylglycerol to regulate intracellular Ca^{2+} levels through RAS signaling.[189] Additionally, BTK regulates BCR intracellular localization by phosphorylating the actin regulator Wiskott-Aldrich syndrome protein (WASP) and the small guanine nucleotide exchange factor, VAV.[189] Cellular membrane localization of BTK leads to interaction with chemokine receptors CXCR4 and CXCR5. It is hypothesized that direct interaction of the chemokine receptors and BTK is mediated by the G_α and G_β subunits by binding to the PHTH domain of BTK. Following chemokine receptor binding, Y551 is exposed and susceptible to SYK and SRC family kinase phosphorylation that leads to BTK autophosphorylation and initiation of the PLCγ2 and RAS signaling pathways.[189] BTK mediates TLR signaling by directly interacting with the Toll/IL-1 domain of the receptor or by interacting with myeloid differentiation primary response 88 (MYD88) that is a downstream adaptor protein of most TLRs.[189,191] Additionally, BTK has been shown to phosphorylate TLR3 at Y759 during antiviral cellular response.[192] Since BTK is involved in regulating a variety of signaling pathways, it has been shown to play a central role in a wide variety of diseases from cancers to autoimmune diseases.[189,193,194]

During its discovery in 1993, BTK deficiency was attributed to be the driving force of XLA disease progression.[186,189] Since then, BTK has been implicated in initiation and maintenance of B-cell malignancies, leukemias, solid tumors, and autoimmune disorders. Unlike protein kinases in other diseases, there are no gain-of-function mutations that create an over-active BTK kinase in BTK-driven diseases. This is attributed to the multiple layers of autoinhibition by the PHTH domain binding along with the SH3 domain binding to the proline-rich linker region that joins the PHTH and SH3 domains.[187,190] Overreliance of

cancer cells in B-cell malignancies on the BTK pathway is a classical hallmark of cancer. Chronic lymphocytic leukemia cells overexpress BTK, while Mantle cell lymphomas overexpress LYN and SYK kinases, a part of the BTK activation mechanism (hyperphosphorylation of BTK Y223).[186,189] Activated B-cell like diffuse large B cell lymphomas (ABC-DLBCL) depend on NF-κB signaling that is regulated by BTK mediation of BCR and TLR signaling.[189] Additionally, B-cell activation is a classical maker in rheumatoid arthritis that has led to the use of TKIs for the treatment of rheumatoid arthritis.[193]

Although expression of BTK is predominantly observed in B-cells, a truncated BTK isoform, where a portion of the PH segment of the PHTH domain is deleted, was discovered in colon cancer.[194] Deletion of the PHTH regulatory domain has been shown to increase BTK kinase activity[187] and coincides with the observations seen in colon cancer tumors and cells.[194] Furthermore, the truncated BTK isoform was shown to play a prominent role in survival and proliferation of colon cancer cells.[194] Thus, the involvement of BTK in a variety of diseases has made it an attractive drug target.

Kinase inhibitors can be classified by their binding mode.[68–70,195] The first mode of binding occurs in the kinase domain ATP binding pocket and can be further characterized by the DFG motif orientation. Type I inhibitors bind to the DFG-in orientation of active kinases, while type II inhibitors bind to the DFG-out orientation of inactive kinases. Additionally, type II inhibitors bind to an allosteric hydrophobic binding pocket within the kinase domain adding a layer of selectivity. The second binding mode consists of inhibitors binding non-competitively to a small allosteric pocket within the ATP binding site and are classified as type III inhibitors. In the third binding mode, type IV inhibitors bind allosterically to stabilize the inactive kinase conformation. Finally, the fourth and most efficacious binding mode of BTK inhibitors consists of covalent inhibitors that render the kinase catalytically inactive.[189,193,196,197]

Two FDA-approved inhibitors, ibrutinib and acalabrutinib, used to treat BTK-driven diseases irreversibly bind to cysteine 481 within the catalytic domain to render the kinase inactive. During clinical trials, the overall response rate for ibrutinib treatment was 60% thus leading to FDA approval for treatment of chronic lymphocytic leukemia (CLL) and mantle cell lymphoma (MCL).[189,196,197] While, ibrutinib is efficacious towards BTK, it

contains off-target activity against LYN, epidermal growth factor receptor (EGFR), janus kinase 3 (JAK3) and other TEC family kinases (ITK and TEC)[189,196,197] leading to severe side effects such as bleeding, rash, diarrhea and atrial fibrillation.[196] Additionally, the point mutation C481S within BTK's kinase domain prevents ibrutinib binding resulting in drug resistance. Due to ibrutinib's adverse side effects, efforts to develop BTK specific inhibitors led to the discovery of acalabrutinib.[198] Acalabrutinib is structurally similar to ibrutinib however it displays higher specificity for BTK.[197] Clinical trials with ibrutinib resistant CLL and MCL patients are underway to determine acalabrutinib efficacy. Additionally, there are current clinical studies underway to determine the efficacy of reversible and specific BTK inhibitors to resistant mutants. [196,197]

Unlike B-cell malignances, the role of BTK has not been elucidated in myeloid cell signaling. BTK is expressed in over 80% of AML cases and has been identified as a downstream substrate of FLT3-ITD.[199] High levels of BTK phosphorylation were observed in FLT3-ITD positive AML cell lines, and patient samples were disrupted by treating cells with quizartinib (FLT3-ITD TKI). Furthermore, FLT3-ITD negative cells stimulated with FLT3 ligand did not alter BTK phosphorylation levels, suggesting BTK is not a substrate for WT FLT3. Interestingly, when FLT3-ITD positive cells were treated with quizartinib and ibrutinib cell death was amplified. To determine the involvement of BTK in chemotherapy resistant AML cell lines and patient primary cell cultures were treated with ibrutinib and were susceptible to BTK inhibition, implying that chemotherapy-resistant FLT3-ITD negative AML cells rely on BTK to evade cell death. In chemotherapy-resistant FLT3-ITD negative AML cells, TLR9 signaling was identified as the primary diver of increased BTK activity. Although the role of BTK in B-cell malignancies has been identified, the role it plays in other diseases is still being elucidated and highlights the importance of potent BTK inhibitors.

Biochemical tools, such as sensitive assays to measure kinase activity, that can aid in the identification of new specific BTK inhibitors would benefit drug discovery efforts. Development of reliable high throughput assays to monitor kinase activity requires efficient kinase substrates. Unfortunately, none have been reported for BTK. Enzyme substrate specificity has been widely studied, but with recent technological advances such as high-resolution mass spectrometry, researchers are now able to process large scale

datasets to identify and predict enzyme substrate sequences. In this chapter, we present an automated process for phosphoproteomic and KINATEST-ID data analysis within the GalaxyP environment which resulted in the discovery of four sensitive BTK kinase substrates that were used to develop an antibody free activity assay for BTK.

4.3 Materials and methods

4.3.1 Cell culture and endogenous peptide sample preparation

Cell culture and endogenous peptide sample preparation was carried out as described in chapter three with slight modification. KG-1 cells (ATCC) were maintained in IMDM media (Gibco) supplemented with 20% heat inactivated fetal bovine serum (FBS), 1% penicillin/streptomycin in 5% CO_2 at 37 °C.

4.3.2 Alkaline phosphatase

The alkaline phosphatase treatment was carried out as described in chapter three.

4.3.3 KALIP recombinant BTK kinase assay

Recombinant BTK KALIP kinase assay was carried out as described in chapter three using a two-hour incubation.

4.3.4 LC-MS/MS data acquisition

LC-MS/MS data acquisition was carried out as described in chapter three with a modification to the LC elution time interval. The mobile phase consisted of 0.1% formic acid in ultra-pure water (Solvent A) and 0.1% formic acid in acetonitrile (Solvent B). Samples were run over a linear gradient (5-30% solvent B; 80 minutes). The subsequent steps are as described in chapter three.

4.3.5 Data analysis

The KinaMINE and scripts in R studio developed in chapter 3 to identify, extract and format phosphopeptide sequences to incorporate them into the KINATEST-ID pipeline were converted into GalaxyP compatible files described below.

4.3.5.1   Script conversion into XML for compatible upload to the GalaxyP environment

Tools can be incorporated into the GalaxyP environment but require their conversion into a command line execution process within the Linux operating system interface.[172,200] This required the creation of a set of instructions in the form of an XML file, which contains pertinent input file information such as number of inputs and type of inputs (text or FASTA, etc.). Additionally, the XML file will contain the tool directory and tool initiation instructions (https://galaxyproject.org/admin/tools/add-tool-tutorial/). The tools and corresponding XML instructions are "wrapped" together into a GZ file format that is used by GalaxyP, which handles tool dependencies dictated by the language the tool is written in such as in R studio or in Java.

4.3.5.2   File conversion workflow

Msconvert and MGF file formatter tools were used to convert raw mass spectrometer files into mzml and MGF files for input into ProteinPilot version 5. Raw mass spectrometer files were combined into a "dataset collection"[200] and processed through the workflow as described in chapter three.

4.3.5.3   Proteomic database Workflow 2

The MGF files were searched against a merged FASTA database (reviewed human Uniprot and the cRAP databases) using ProteinPilot version 5 as described in chapter three. For each database search, the Group file extractor tool extracted the distinct peptide report (DPR) and corresponding search summary excel file. Then, the DPR and the merged FASTA database were used as input for the KinaMine -7 to 7 tool within Galaxy-P to generate the substrate and substrate background frequency lists. The FDR threshold was set at <1% global FDR.

4.3.5.4   KINATEST-ID Workflow

The KINATEST-ID workflow was created to recapitulate the processes described in chapter 3. In brief, to approximate the list of tyrosine centered peptide sequences that could have been potentially phosphorylated but were not, the Negative Motif Finder -7 to 7 function within Galaxy-P was used. This list was generated by uploading the Positive

Substrates file, Substrate Background Frequency File, and a human proteome reference. The human proteome reference was created in Galaxy-P and a file conversion was performed using the FASTA to Tabular function in Galaxy-P. Then once the database is in tabular format, the updated version of the Negative Motif finder -7 to7 identifies the UniProt accession numbers of the proteins that contain the identified BTK substrates and then extracts them from the tabular human UniProt database, and then performs an *in-silico* trypsin digest to generate a list of all tyrosine centered sequences possible. The tool then compares this list with the sequences found in the positive substrate list and returns the sequences that were not in the positive substrate lists.

4.3.6   Peptide synthesis

Peptides were synthesized as described in chapter three.

4.3.7   BAStide in vitro kinase assay

Assays with recombinant BTK kinase were carried out as described in chapter three. In brief, the assay reaction was started by adding the peptide substrate to a 37.5 µM final reaction concentration in a 50 µL volume. Sample aliquots (10 µL) were quenched by combining 1:1 with 30 mM EDTA at 4, 15, 30, 45 and 60-minute timepoints.

4.3.8   ELISA-based in vitro kinase assay

ELISA-based assays were carried out as described in chapter three with slight modifications. Peptide phosphorylation was detected by incubation with horseradish peroxidase-conjugated antiphosphotyrosine mouse monoclonal antibody 4G10 (MilliporeSigma, 1: 5,000 dilution in TBS-T). The subsequent steps were carried out as described in chapter three.

4.3.9   Terbium-based in vitro kinase assay

Recombinant BTK kinase was purchased and diluted as described in chapter three. In brief, the kinase was pre-incubated with kinase reaction mixture as described in chapter three at room temperature. The kinase reaction was started by adding the BAStide-D substrate to a final concentration of 20 µM in a 20 µL volume at room temperature. Sample

aliquots were quenched by combining 1:1 with 6M urea at 0.5, 5, 10, 15, 30, 45, 60 and 90-minute timepoints.

4.3.10 Terbium-based phosphorylation detection assay

Terbium luminescence assay data collection was carried out as described in chapter two. Time resolved emission spectra were collected on a Biotek Neo2 plate reader with a monochromator at room temperature. Spectra was collected as described in chapter two with a sensitivity (gain) setting of 230.

4.4 Results

4.4.1 In vitro kinase reaction to identify substrates for input/analysis with the KINATEST-ID pipeline

We applied our adapted KALIP strategy to increase the known number of BTK kinase substrates. The KALIP workflow was carried out as described in chapter 3 (FLT3) with some slight modifications. KG-1 cell lysates were trypsin digested and desalted using HLB cartridges. After a speed vacuum drying step, the samples were reconstituted in alkaline phosphatase reaction buffer and treated with alkaline phosphatase for 90 minutes. The phosphatase reaction was quenched by incubating the reaction mixture at 75° Celsius (C) over 30 minutes. The samples were divided into two aliquots, one that was treated with reaction mixture with no kinase while the other was treated with recombinant BTK. The samples were incubated for two hours at 37 degrees C and quenched by acidifying the lysate to a pH below 1. The samples were again desalted with HLB cartridges. Then, the lysates were subjected to phosphopeptide enrichment (PolyMAC) and subsequently analyzed on the mass spectrometer. The mass spectrometer files were then uploaded to GalaxyP platform (http://galaxyp.msi.umn.edu/)

Our aim was to generate an automated data analysis process and to accomplish this we created three workflows within Galaxy P. They are the file conversion, proteomic database and the KINATEST-ID workflows (Figure 1). Workflows within the Galaxy P environment are automated processes that are developed to perform a series of steps in chronological order. Workflow outputs are then used as input for downstream workflows that are initiated once the input file is created. All three of our workflows were initiated in

parallel. The file conversion workflow was developed to automatically convert raw mass spectrometer files into the mascot generic format (MGF) intended for use with ProteinPilot version 5. The proteomic database workflow was initiated once the corresponding MGF files were created by the file conversion workflow. The proteomic database preforms two steps. The first step carries out the proteomic database search while the second step extracts (group file extractor) the distinct peptide report (DPR) from the search file. The DPR is the input required for the KinaMINE data formatter tool that identifies the phosphotyrosines at a corresponding confidence score. The KINATEST-ID workflow recapitulates the data formatting steps of the scripts developed in chapter 3 with the added functionality of working with 7 amino acids N- and C-terminal to the tyrosine of interest. The inputs required are the distinct peptide report and the merged human UniProt-cRAP database in a tabular format. Upon completion of the KINATEST-ID workflow, three output files (OPF) were created. The first OPF contains the amino acid standard deviation values (SDV), the second contained the endogenous probability matrix (EPM), and third file contained the predicted BTK-specific scored sequences.

The original KINATEST-ID pipeline relied on literature and proteomic databases for substrate identification, of which only 21 were reported (Figure 2A). The literature-based KINATEST-ID algorithm failed to produce an artificial peptide substrate sequence that implied that we did not have sufficient information to identify BTK's preferred substrate sequence. Using the three workflows described above, we identified 101 sequences phosphorylated by BTK. Additionally, we processed the BTK proteomic search results using the command line process (chapter 3), and we identified 95 substrates. The SDV and the site selectivity matrices (SSM) are illustrated in figure 2. Overall, there were slight differences in the command line and the GalaxyP based KINATEST-ID SDV and SSM matrices. The larger differences were observed between the literature-based matrices when compared to the Galaxy P or command line process of the KALIP data. Amino acids were deemed statistically significant if they were observed two standard deviations above or below the mean. Each of the positions analyzed, within the SDV table, we identified instances where amino acids observed to be under-represented in the literature algorithm was observed to be a favorable or statistically significant amino acid. These observations in conjunction with the lack of artificial substrates identified using the literature-based

KINATEST-ID algorithms implied that the KALIP derived algorithms identified BTK substrate characteristics that could be used to develop artificial substrates.
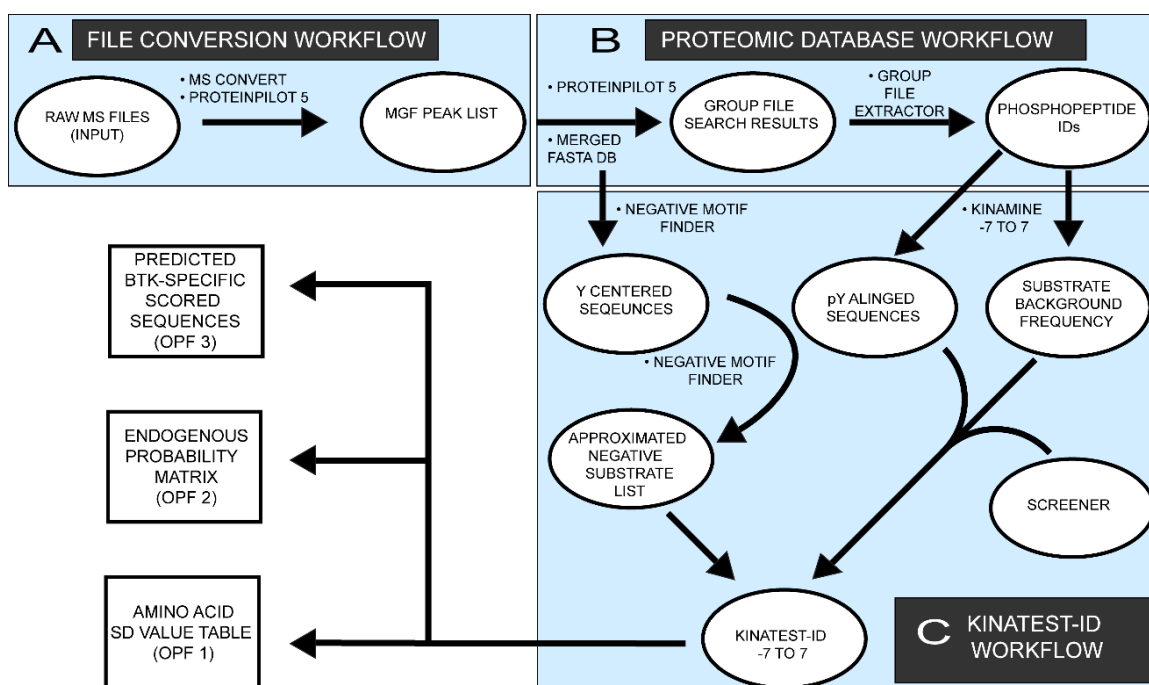
Figure 4.1. Schematic representation of the GalaxyP KINATEST-ID workflows

4.4.2 KINATEST-ID based design of novel BTK Artificial Substrate peptides (BAStides)

Using the GalaxyP derived KINATEST-ID output files, we identified statistically significant amino acids and their respective positions (Figure 2B and 2C). To generate the sequence permutations, the KINATEST-ID workflow identified all the amino acids that contains a SDV above 2 at each position and selected the AA with the highest value for positions that did not contain a significant AA. The KINATEST-ID workflow generated over 1.4 million sequence permutations (Output file 3) using the information from the SDV table (OPF1). To narrow the number of sequence permutations, we focused on the four amino acids N- and C-terminal to the phosphotyrosine. In addition to the SDV, we considered the SSM that identified positions -4, -3, -2, -1 and 1 as potential significant (SSM value greater than 1) positions with SSM values of 0.8, 1.5, 1.2, 2.2 and 0.8 respectively. Position -4 contained aspartic acid (D) and asparagine (N), while position -3 contained D, glutamic acid (E), phenylalanine (F) and glutamine (Q). Position -2 contained

D, N, Q and serine (S), while position -1 contained phenylalanine (F), leucine (L) and isoleucine (I). Position 1 contained A and E while threonine (T), valine (V) and lysine (K) were fixed at positions 2, 3 and 4 respectively. We used the command line based KINATEST-ID (chapter 3) to generate the possible 96 sequence permutations, of which 8 were predicted to be specific BTK substrates. The top two scoring sequences were selected to be BAStide-A (DFQLYATVK) and BAStide-B (NFQLYATVK). Additionally, the SSM indicated that BTK required a statistically significant amino acid at position -1 to phosphorylate a candidate sequence and to test this hypothesis we made a L to tryptophan (W) mutation to BAStide-B to create BAStide-C (NFQWYATVK). W was chosen because it contained a -0.95 SDV at position -1. Lastly, we mutated BAStide-A at position 1 (A to E) to create BAStide-E (EDDFQLYETVKEE), which was incorporated in the terbium-binding motif discussed later.

The BAStide sequences were scored against the literature, command line and the GalaxyP based KINATEST-ID models (Figure 2D). Although the literature-based KINATEST-ID model analyzed SDV for seven amino acids N- and C-terminal to the tyrosine of interest, the scoring function (Endogenous probability matrix and sequence scoring in chapter 2) only considered four amino acids. Consequently, we took two approaches for scoring the BAStide sequences against the GalaxyP model. We scored the sequences by considering four (GalaxyP -4 to 4) or seven (GalaxyP -7 to7) N and C terminal amino acids with respect to the tyrosine. BAStide A, B and D scored well in the GalaxyP -4 to 4 scoring approach while BAStide-C scored poorly. All the BAStide sequences scored poorly in the GalaxyP -7 to 7 approach that can be attributed to the N-terminal glycine residues at positions -6 and -5, which were observed as underrepresented and impacted the sequence score negatively. BAStides A-C did not contain an amino acid at position -7 and the blank space was assigned a value of 1, which is considered neutral value within the EPM and scoring functions. All the BAStide sequences contained a sequence score of zero in the literature-based model, which indicated that we did not have sufficient information to make an accurate sequence prediction. The BAStide sequences were then synthesized and prepared for biochemical assays discussed in the section below.
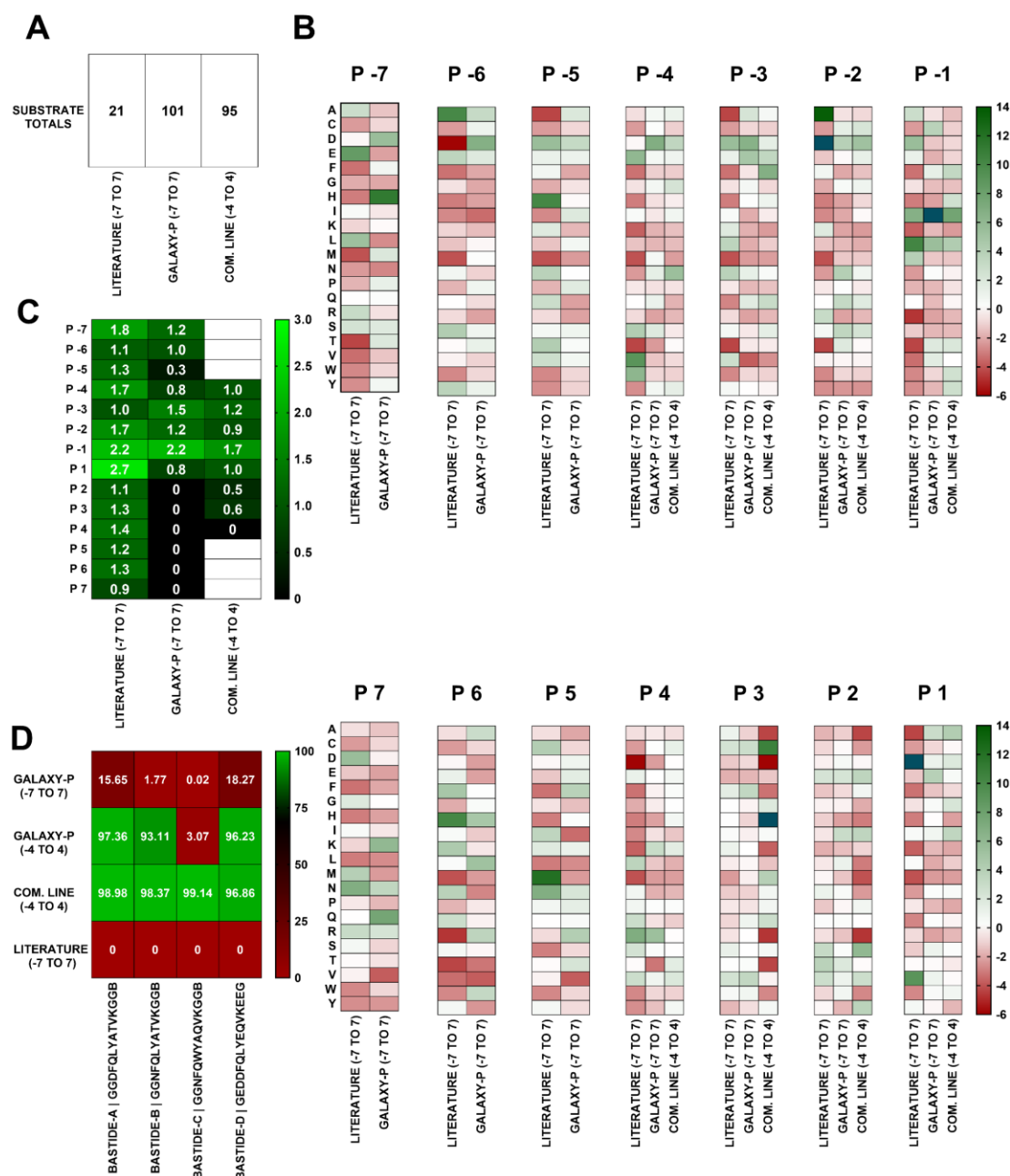
Figure 4.2. BTK phosphopeptide identifications, positional preferences and KINATEST-ID model sequence scoring.

4.4.3 *In vitro* validation of BAStide sequences as BTK kinase substrates

The BAStide candidate sequences were assayed *in vitro* against recombinant BTK over a 60-minute incubation with sample aliquots quenched at 4, 15, 30, 45 and 60 minutes. A previously reported BTK substrate, universal peptide-5[179] was used as our positive

activity control. Phosphorylation was detected using an anti-phosphotyrosine 4G-10 antibody conjugated to horseradish peroxidase in a ELISA-based assay. Overall, BTK phosphorylated BAStides-A, -B and -D more efficiently than peptide-5, while BAStide-C was observed to be a less efficient substrate (Figure 4-3). In our GalaxyP -4 to 4 scoring model of KINATEST-ID, BAStide-A was the highest scoring sequence followed by BAStide-D and -C respectively, while BAStide-C was predicted to be a poor substrate for BTK.

Our results indicate that BTK prefers aspartic acid (SDV 6.85) over asparagine (SDV 1.32) at position -4, which was not considered a significant position (SSM value of 0.8). We observed that BTK activity was reduced over a 60-minute in vitro incubation by mutating a statistically significant amino acid for a non-significant one at position -4. A similar effect was observed between BAStide-B and -C, which was designed to determine the effects of the site selectivity matrix on BTK kinase activity upon its preferred substrate motif. The SSM identified position -1 as the most important site that required a statistically significant amino acid in a substrate to be phosphorylated by BTK, which was supported by our L (5.47 SDV) to W (-0.96 SDV) mutation assay results. A similar observation was seen between BAStide-A and -D because the sequences contained multiple differences to draw a robust conclusion between the alanine to glutamic acid at position 1.
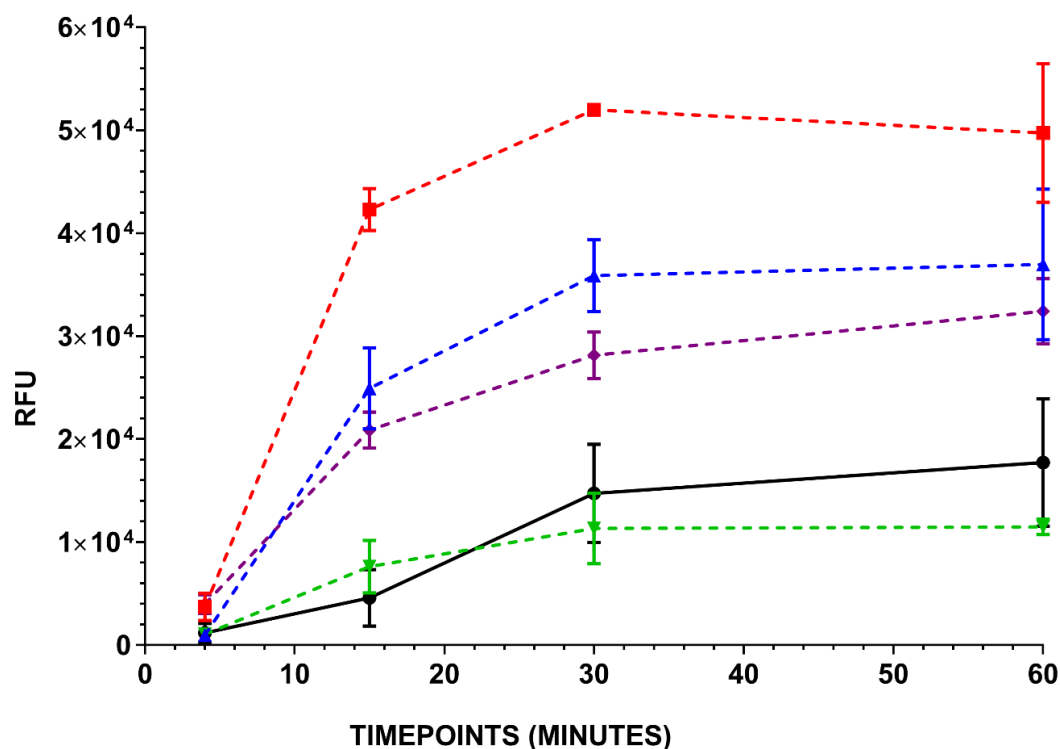
Figure 4.3. BTK phosphorylation of BAStide candidate sequences in vitro.
Data represents averages $\pm$ SEM of experiments performed in triplicate.

4.4.4 Evaluating the correlation between activity and scoring model using the universal peptide substrates

To determine statistical correlation between KINATEST-ID scoring models and BTK kinase activity against BAStide and universal peptide substrates, we performed a Spearman's non-parametric test as described in chapter three. In our ELISA-based activity assays, we used peptide 5 as our reference substrate to determine baseline BTK kinase activity (Figure 4-3). BAStides-A, B and D were observed to be more efficient substrates than peptide 5 and BAStide-C over a 60-minute incubation.

To correlate our KINATEST-ID scoring models to BTK biochemical activity, we scored the panel of universal substrates reported by Marholz and colleagues[179] using our KINATEST-ID models (Figure 4-4A). To compare the BTK activity between the universal peptide data and our ELISA activity assay, we reprocessed and normalized the 30-minute time point data, reported by Marholz and colleagues, to create a 'peptide 5' percent activity

signal (Figure 4-4B). Subsequently, we normalized the BAStides 30-minute raw data point to the signal of peptide 5 (Figure 4-4C). The 'peptide 5' percent activity signal was then correlated to the KINATEST-ID scoring models using spearman's non-parametric test as described in chapter three. We observed a modest positive correlation between the GalaxyP and command line scoring models, which used the four N- and C-terminal amino acids, with respect to the tyrosine residue to generate the positional matrix score. These results demonstrate that the GalaxyP-based KINATEST-ID algorithm's positional matrix score is moderately correlated to efficient artificial BTK substrates.
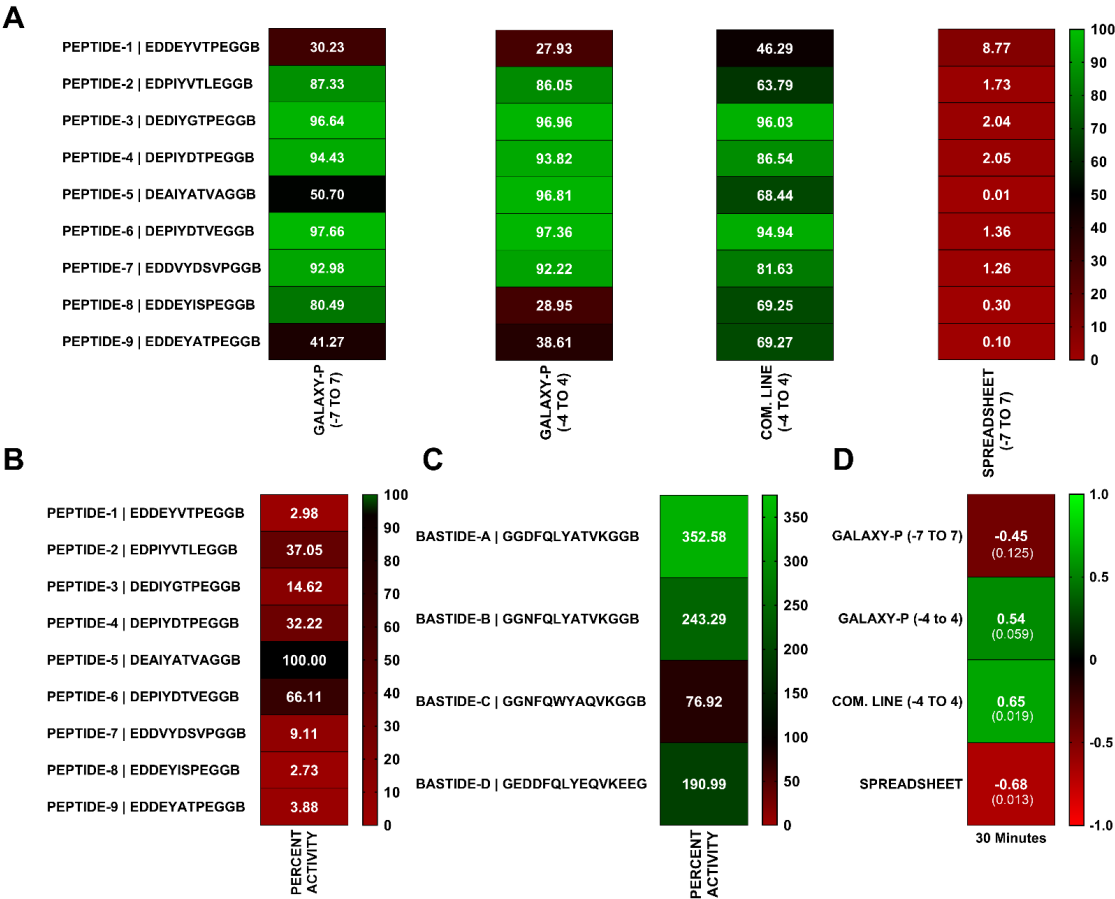
Figure 4.4. Universal peptide KINATEST-ID model scoring (4-4A).

Universal peptide (4-4B) and BAStide (4-4C) substrate data were normalized to peptide 5 signal at a 30-minute timepoint. Universal peptide data reprocessed from Marholz et al.

4.4.5   Monitoring BTK kinase activity through a time resolve terbium chelating assay

To demonstrate how the GalaxyP workflows can be automated for KALIP and KINATEST-ID data analysis to develop antibody-free kinase assays, we aligned the preferred BTK substrate sequence into the terbium binding motif. During BAStide design, we performed sequence alignment using a known terbium binding peptide (α-synuclein Y125) to identify sequences that were predicted to be BTK substrates (Figure 4-5A). A terbium chelating peptide contains acidic residues at position -6, -4, 1, 5 and 6 with respect to the tyrosine residue, which completes the terbium binding motif upon phosphorylation. To develop a BTK substrate that can chelate the Tb^{3+} ion, we incorporated acidic residues at the corresponding positions within the sequence of BAStide-D (Figure 4-5A) that was validated as a BTK substrate through an ELISA-based assay (Figure 4-3). The phosphorylated version of BAStide-D (pBAStide-D) was synthesized to determine if it can chelate terbium in a phosphorylation dependent manner. A calibration curve with increasing percent (0 to 100%) of pBAStide in the presence of kinase assay reagents (40 µM ATP, 0.4 µM $NaVO_4$, 10 mM HEPES and 4 mM $MgCl_2$) demonstrated the increase in terbium signal with increasing amount of phosphopeptide present (Figure 4-5B). The spectrum measurements were plotted on GraphPad Prism software to calculate the area under the curve (AUC) values that were fitted to a liner regression line (Figure 4-5C). BAStide-D was assayed against BTK and samples were quenched at 0.5, 5, 10, 15, 30, 45, 60 and 90-minute timepoints with 6 M urea. The quenched samples were then treated with luminescence buffer (10 mM HEPES, 100 mM NaCl and 100 µM Tb^{3+}) and measurements were taken with an excitation wavelength of 267 nm, 1000 µsec, 50 µsec delay time with 10 measurements per data point. Using the calibration curve linear regression line, we extrapolated the corresponding percent phosphorylation of BAStide-D (Figure 4-5D and 4-5E). Our results demonstrate an antibody-free assay to monitor BTK kinase activity. The terbium-based readout can now be used to monitor BTK activity in the presence of kinase-inhibitors.
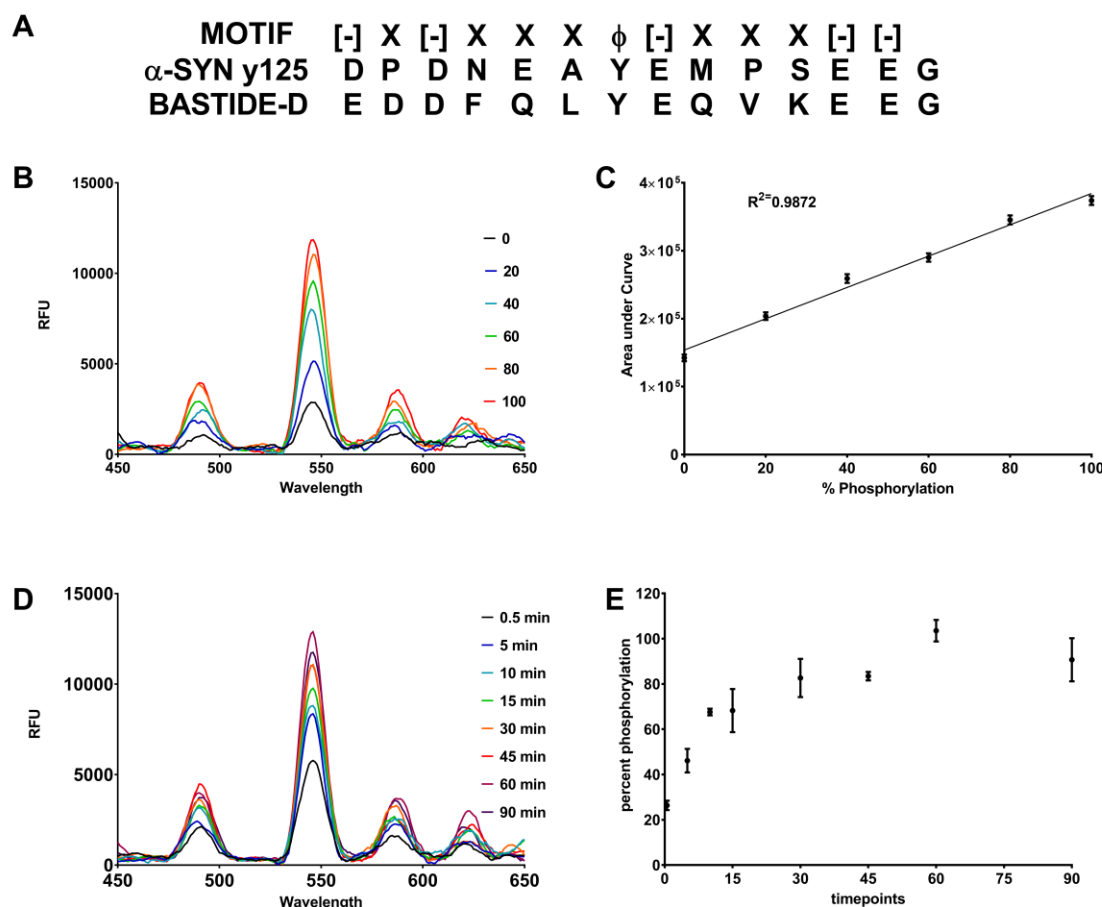
Figure 4.5. BAStide-D terbium assay design, validation and signal detection.

Comparison of BAStide-D with the alpha synuclein Y125 and required Tb residues (A). (B) Smooth of calibration curve terbium binding of phosphorylated BAStide-D and area under the curve (AUC) from the raw curves. (D) Smooth of in vitro kinase assay using BAStide-D with BTK over a 90-minute incubation (B) and extrapolate percent phosphorylated BAStide-D (E). Data represents averages $\pm$ SEM of experiments performed in triplicate (C).

4.5 Discussion

BTK was identified as a viable drug target due to the role it plays in disease progression of immunological disorders, B-cell malignancies and solid tumors. BTK inhibitors are predominantly efficacious, but they contain severe side effects driven by off target activity (ibrutinib).[189,196,197] Additionally, resistance to ibrutinib and acalabrutinib mediated by the C481S mutation that prevents covalent inhibitor binding highlights the need to identify next generation TKIs to combat drug resistance. However, efficient assays

to monitor BTK kinase activity in a dose dependent manner are needed. In this chapter, we created an automated process within the GalaxyP framework to analyze experimental (KALIP)[97] and computational (KINATEST-ID)[80] results to identify, extract and format kinase peptide substrates. The KALIP technique is a high throughput tool that is used to identify large numbers of kinase substrates with experimental uniformity.[95,97] We converted the KINATEST-ID[80] and relevant data formatting scripts in R studio (chapter 3) to be compatible with the GalaxyP environment. By using the tools available in the GalaxyP environment we created the file conversion, proteomic database and KINATEST-ID workflows. This process produces three output files that identify the preferred amino acids (OPF1) at their respective positions, generates a scoring matrix (OPF2), and returns a focused peptide library of potential artificial substrates (OPF3). We applied our Galaxy-P based KINATEST-ID pipeline to identify BTK's preferred substrate motif to validate four artificial substrates (BAStides).

To validate our GalaxyP workflow, we manually processed our data using the scripts developed in chapter 3 and compared both approaches to the previously reported literature-based KINATEST-ID[80] results. Additionally, we chose to compare the models to identify differences between the phosphoproteomic and literature models that prevented the literature model from predicting BTK artificial substrate sequences. In addition, we used our models to score the reported universal peptide substrate panel,[179] of which five sequences were found to be phosphorylated by BTK. We used universal peptide-5 as our BTK activity control in our biochemical assays. The phosphorylation signal for our reported BAStides and the signal from the universal peptide report[179] were normalized using the signal from peptide-5 to generate a comparable percent activity signal (Figure 4-4). We used a non-parametric Spearman correlation test to identify a relationship between the substrates' positional matrices score and normalized data signals. We identified a positive correlation between the GalaxyP and the command line data analysis models when considering the four amino acids N- and C-terminal to the tyrosine of interest. Our phosphoproteomic models successfully identified the BAStide and universal peptide substrates that were phosphorylated by BTK.

Our goal was to demonstrate that our GalaxyP workflow can analyze phosphoproteomic data to produce sensitive substrates that can be used to develop high

throughput assays to monitor BTK kinase activity. The original KINATEST-ID pipeline contained an aligner module that scored artificial sequences on their ability to chelate terbium (Tb) ions, which allowed for antibody free detection of kinase acitivity.[80] We designed BAStide-D in the corresponding terbium binding motif that allowed for phosphorylation-dependent terbium signal detection. Subsequently, we assayed BAStide-D in the presences of BTK over a ninety-minute incubation and successfully measured phosphorylation dependent Tb signal. Currently, the aligner module is not part of the GalaxyP KINATEST-ID workflow, but work is being done to convert the aligner module into a GalaxyP compatible script.

In summary, we adapted the phosphoproteomics data formatting and KINATEST-ID scripts into the GalaxyP framework to create workflows that automate the analysis of KALIP derived data to produce the three output files from the KINATEST-ID scripts. The creation of workflows within the GalaxyP framework decreases user induced errors and the capability for automated analysis of multiple kinase treatment datasets. Although the current iteration of the GalaxyP KINATEST-ID workflow is functional, the process requires increased functionality updates. Our workflow is only compatible with ProteinPilot 5 and needs support for other proteomic search engines available in the GalaxyP environment. Furthermore, implementation of the commonality and difference finder script would allow users to identify symmetric and non-symmetric sequence identifications within multiple datasets and to compare technical replicates. Additionally, a comparison feature would allow us to compare the proteomic search results for the same dataset to identify sequences seen in more than one proteomic search engine. Development of an automated tool that identifies and extracts the confidence level score for each search without user input is required for further automation within the GalaxyP environment. The processing and functionality of the KINATEST-ID workflow would benefit from user input to pre-determine the amino acid selection used to generate the sequence permutations of potential kinase substrates. This feature would speed up the processing time to create an in-silico peptide library with the desired sequence constraints such as the Tb binding motif.

Therefore, the GalaxyP KINATEST-ID workflow can be used to identify artificial substrates for poorly studied tyrosine kinases with the capability of high throughput Tb based assay design, which can be used to identify and develop selective and specific kinase

inhibitors. The current format of our GalaxyP workflow can be easily manipulated and applied to S/T kinases. The reported process does not depend on defined kinase biology or reported substrates, which makes it a valuable tool in drug discovery efforts. Thus, we anticipate this process can be altered to identify substrates for orphan kinases or enzymes to aid drug discovery efforts.

CHAPTER 5.   MULTI-COLORED, TB3+-BASED ANTIBODY-FREE DETECTION OF MULTIPLE TYROSINE KINASE ACTIVITIES

5.1   Contributions to this work

Chapter five is the published form of a multicolored terbium-based antibody-free detection method for tyrosine kinase activities manuscript in the Analytical Chemistry journal. I have obtained copyright permission from the journal to include the published form of the manuscript into my dissertation document.

In continued guidance of Dr. Laurie L. Parker and Dr. Andrew M. Lipchik, I resumed my contributions in demonstrating how the SCR kinase family substrate can be used in an antibody-free detection assay. My contributions centered on the validation of the fluorophore labeled SFAStide-A substrate through terbium and ELISA based detection methods.

Reprinted (adapted) with permission from (Lipchik AM, Perez M, Cui W, Parker LL. Multicolored, Tb(3)(+)-Based Antibody-Free Detection of Multiple Tyrosine Kinase Activities. *Anal Chem*. 2015;87(15):7555-7558. doi:10.1021/acs.analchem.5b02233.). Copyright (2015) American Chemical Society."

Andrew M. Lipchik[‡], Minervo Perez[§], Wei Cui[§] and Laurie L. Parker[§]*

Department of Medicinal Chemistry and Molecular Pharmacology, College of Pharmacy and Purdue Center for Cancer Research, Purdue University, 201 S. University Street, West Lafayette, IN 45707 (USA)

*Corresponding Author

Fax: (612) 625-2163

E-mail: llparker@umn.edu

5.2   Abstract

Kinase signaling is a major mechanism driving many cancers. While many inhibitors have been developed and are employed in the clinic, resistance due to crosstalk

and pathway reprogramming is an emerging problem. High-throughput assays to detect multiple pathway kinases simultaneously could better model these complex relationships and enable drug development to combat this type of resistance. We developed a strategy to take advantage of time-resolved luminescence of Tb^{3+}-chelated phosphotyrosine-containing peptides, which facilitated efficient energy transfer to small molecule fluorophores conjugated to the peptides to produce orthogonally-colored biosensors for two different kinases. This enabled multiplexed detection with high signal to noise in a high-throughput-compatible format. This proof-of-concept study provides a platform that could be applied to other lanthanide metal and fluorophore combinations to achieve even greater multiplexing without the need for phosphospecific antibodies.

5.3   Introduction

Numerous leukemias and lymphomas have been characterized by the clonal expansion of B-lymphocytes due to the deregulation of the B-cell receptor signaling pathway.[201,202] Tyrosine kinases Lyn, Syk and Btk are the main signal transducers in this pathway, making them popular therapeutic targets for small molecule inhibitors.[203] Despite the identification of this pathway as the cause of disease, effective therapeutic options targeting the B-cell receptor pathway and/or these kinases are still relatively limited. Often these kinase activities are dependent on each other, which can affect the efficacy of inhibitor drugs targeting individual enzymes. There is a need for new detection strategies that offer sensitive and specific detection of multiple kinase activities that can enhance the depth of information obtained in a screening assay, monitoring more than one signal simultaneously and mimicking reconstitution of the relevant pathways.

Förster resonance energy transfer (FRET) based assays have been developed to monitor multiple dynamic cellular processes simultaneously in a single assay.[204–208] However, while useful in some applications, FRET based methods that use organic fluorophores or fluorescent proteins as both the donor and acceptor suffer from limitations including small dynamic ranges, small Stokes shifts/wide emission peaks resulting in spectral bleed through, and the requirement for genetic engineering and expression of protein fluorophores. Lanthanides (Ln^{3+}) have been explored as probes in biological assays

for the detection of ligand binding, enzyme activity, and protein-protein interactions due to their unique optical properties.[209–217]

Compared to organic fluorophores and fluorescent proteins, Ln^{3+} have narrow emission bands, large Stokes shifts, and long photoluminescence lifetimes, enabling time-resolved analysis, high sensitivity and specificity of detection due to reduced interference from short-lived background fluorescence. These also allow multiplexed detection via the multiple distinct, well-resolved emission bands that can be exploited for luminescence resonance energy transfer (LRET) to more than one acceptor fluorophore, chosen such that the emission profiles do not overlap (e.g. Fig. 1A). Existing examples of this strategy rely on antibodies for detection, with either the substrate or a substrate-specific antibody tagged with a small molecule fluorophore for emission, and a phosphospecific antibody labeled with a chelated lanthanide for detecting phosphorylation via donation to the small molecule fluorophore.[218–221] These strategies are therefore limited to the antibodies available for a given substrate modification, and subject to the costs and handling issues presented by such immunodetection workflows.
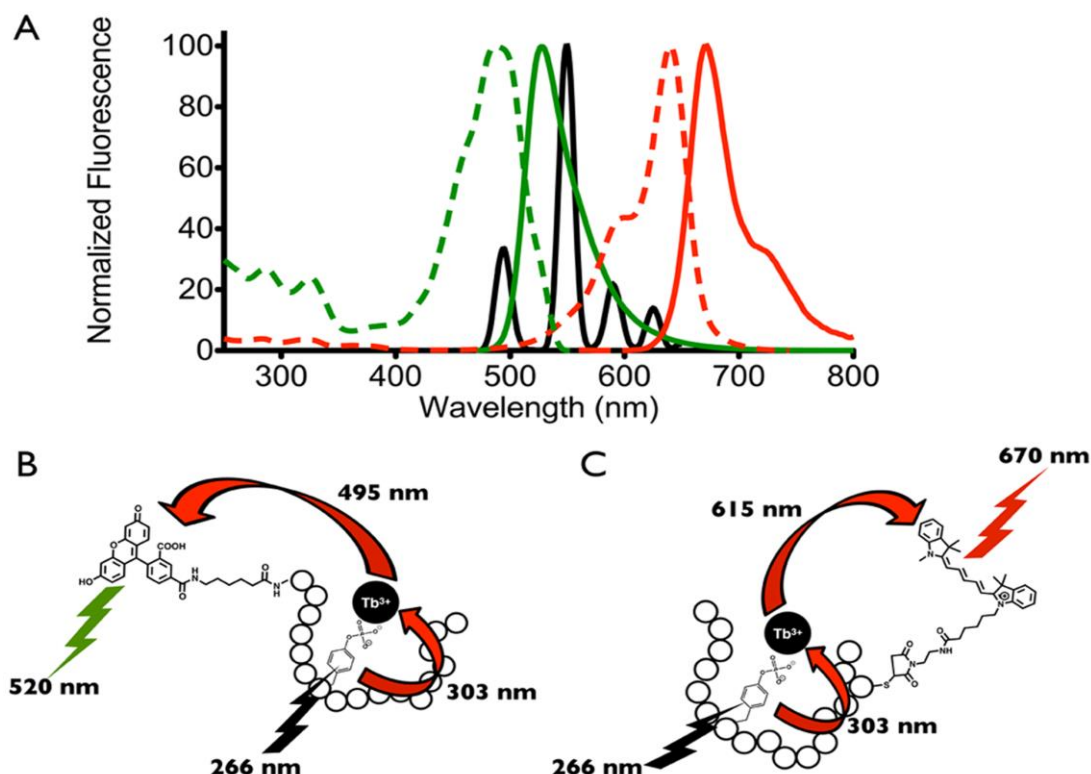
Figure 5.1. Multiplexed detection using time-resolved lanthanide-based resonance energy transfer (TR-LRET) and fluorophore conjugated peptide biosensors.

(A) Emission spectrum of phosphopeptide-Tb3+ complex (black), excitation (dashed lines) and emission (solid lines) spectra of the two acceptor fluorophores 5-FAM (green) and Cy5 (red). Schematic illustrating TR-LRET detection of Lyn (B) and Syk (C) tyrosine kinase activities using the 5-FAM-SFAStide-A (5-FAM-Ahx-GGEEDEDIYEELDEPGGKbiotinGG) and SAStide-Cy5 (GGDEEDYEEPDEPGGCCy5GG) biosensors respectively.

Previously, we demonstrated development of peptide biosensors capable of detecting tyrosine kinase activity through phosphorylation-enhanced terbium (Tb^{3+}) luminescence.[80,81,105] Here we show extension to a multiplexed detection platform for simultaneous monitoring of multiple tyrosine kinase activities (Lyn and Syk) via SFAStide-A and SAStide substrates (sequences given in Table 1).[80,105] Multi-colored detection was achieved through time-resolved luminescence energy transfer (TR-LRET) by employing the phosphopeptide-Tb^{3+} complexes as the energy donors and the conjugated fluorophores cyanine 5 (Cy5) and 5-carboxyfluorescein (5-FAM) respectively, as the energy acceptors (Figure 1A).

Table 5-1. Peptide biosensor sequences[a][b]

[a] 5-FAM=5-carboxyfluorescein; Ahx=6-aminohexanoic acid; K_b=biotinyl-L-lysine; C_{Cy5}=cysteine thiol conjugated with Cy5. [b] Sequence segments represented in bold are the core kinase recognition/Tb^{3+}-chelation residues of the biosensor.

Name	Kinase	Sequence
5-FAM-SFAStide-A	Src-family	5-FAM-Ahx-**GGEEDEDIYEELDEP**GGK$_b$GG
SAStide-Cy5	Syk	**GGDEEDYEEPDEP**GGC$_{Cy5}$GG

5-FAM was selected as the acceptor to couple with the pSFAStide-A-Tb^{3+} complex because its broad excitation peak at 495 nm matches well with the $^5D_4 \rightarrow {}^7F_6$ emission band of Tb^{3+} centered at 495 nm. Sensitized excitation of the phosphorylated 5-FAM-SFAStide-A-Tb^{3+} complex through phosphotyrosine triggers energy transfer to 5-FAM, giving emission from 5-FAM at its characteristic wavelength (~520nm), which falls in a relatively "empty" region of the Tb^{3+} emission spectrum (Figure 1B). Similarly, detection of pSAStide-Cy5-Tb^{3+} complex is achieved based on the overlap of the Cy5 excitation band with the $^5D_4 \rightarrow {}^7F_4$ and $^5D_4 \rightarrow {}^7F_3$ emission bands of Tb^{3+} centered at 595 nm and 620 nm, giving Cy5 emission at its characteristic wavelength (~670 nm) which is also free of interference from Tb^{3+} emission (Figure 1C).

Phosphorylated and unphosphorylated forms of SAStide-Cy5 and 5-FAM-SFAStide-A were synthesized as controls. As we characterized in our previous work, phosphorylation of the peptide substrates resulted in physiochemical and photophysical changes in the peptide-Tb^{3+} complex that enable detection of kinase activity. These changes include enhancing the Tb^{3+} binding affinity, reducing the Tb^{3+} chelate hydration number, increasing the Tb^{3+} luminescence lifetime, and shifting the excitation wavelength of tyrosine.[80,81,105]
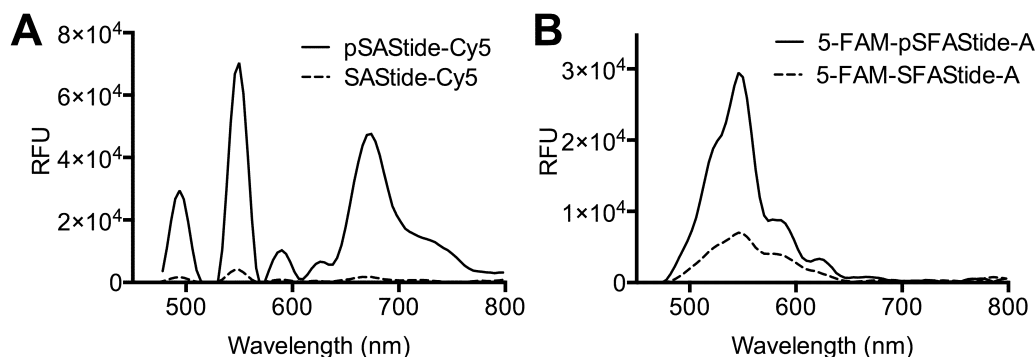
Figure 5.2. Time-Resolved Lanthanide-based Resonance Energy Transfer (TR-LRET) detection of phosphorylation-dependent signals and fluorescence cross-interference.

(A) Time-resolved luminescence emission spectra for SAStide-Cy5 (dashed line) and pSAStide-Cy5 (solid line). (B) 5-FAM-SFAStide-A (dashed line) and 5-FAM-pSFAStide-A (solid line). Spectra were collected from 15 µM peptide in the presence of 100 µM Tb3+ in 10 mM HEPES, 100 mM NaCl, pH 7.5, λex = 266 nm, in 50 µL total volume, 1 ms collection time, 50 µs delay time, and sensitivity 180. Data represent the average of experiments performed in triplicate.

Time-resolved analysis of each peptide biosensor in the presence of Tb^{3+} gave the four characteristic luminescence emission peaks from Tb^{3+} as well as the fluorescence emission peak from the conjugated fluorophore label (Figure 2A, B). Quantitative comparison of the emission spectra between the phosphorylated and unphosphorylated biosensors showed a 25-fold increase in intensity at the Cy5 emission maximum (λ_{670}) for pSAStide-Cy5 (Figure 2A), and a 3.9-fold increase in intensity at the 5-FAM emission maximum (λ_{520}) for 5-FAM-pSFAStide-A (Figure 2B). Control experiments in the presence and absence of Tb^{3+} showed that excitation of pSAStide-Cy5 at 266 nm was Tb^{3+}- and therefore LRET-dependent rather than arising from direct excitation of the fluorophore. Excitation of 5-FAM-pSFAStide-A at 266 nm was also Tb^{3+}/LRET-dependent, but also showed some low-level background excitation at ~330 nm even in the absence of Tb^{3+} (Supporting Information S5), which we speculate can be attributed to delayed fluorescence or phosphorescence of the fluorescein, possibly from peptide adsorbed onto the surface of the well since such longer lived emission (up to the ms range) has been previously observed for fluorescein particularly when adsorbed onto surfaces or in solid state environments.[222–224] However this background excitation of 5-FAM would not substantially affect the LRET readout for the assay since excitation is performed at 266 nm, which did not show any

signal for excitation in the absence of Tb^{3+}. Accordingly, the relevant changes in the intensity of the fluorophore signals upon phosphorylation of their respective peptides provide sensor-specific spectral features that can be monitored to determine phosphorylation of the sensors and consequently kinase activity.

In order to achieve multiplex detection in the same sample, the reaction and detection conditions needed to be optimized to have limited cross-interference between sensors. Cross-interference was evaluated by analyzing the fluorophore signal from an unphosphorylated sensor in the presence of the other phosphorylated biosensor. To accomplish this, the concentrations of the biosensors and Tb^{3+}, as well as the delay time, were varied and TR-LRET spectra collected. Quantification was accomplished by Gaussian fitting of the fluorophore emission peaks and integrating the resulting curves for each peak (Supporting Information S6). Under the optimized conditions, the TR-LRET spectra for each phosphorylated biosensor displayed minimal signal from cross-interfering fluorophore, while giving significantly stronger signal for the desired fluorophore (Supporting Information S7). TR-LRET distance parameters were also characterized (Supporting Information S8 and Table 1).

Next, a calibration curve was plotted to show the quantitative relationship between sensor phosphorylation and its corresponding TR-LRET signal for each sensor (Supporting Information S9). Experiments were performed in the presence of the unphosphorylated form of the other biosensor and the kinase reaction buffer (to best mimic the conditions of a multiplexed kinase reaction). Proportion of phosphorylated peptide was quantitatively determined by integrating the signal centered at 520 nm for 5-FAM and 670 nm for Cy5. The high signal to noise ratio observed in the initial control experiments was maintained in the presence of the reaction buffer with 7.6:1 for SAStide-Cy5 and 5.8:1 for 5-FAM-SFAStide-A. Z'-factor and signal window (SW) values were calculated and shown to be appropriate for HTS with Z'-factor values of 0.72 and 0.78, and SW of 13.27 and 12.65, for SAStide-Cy5 and 5-FAM-SFAStide-A, respectively. Details of these calculations are provided in the supporting information.

After establishing the relationship between sensor phosphorylation and TR-LRET signal, we employed the two biosensors in a kinase assay. Analysis of Syk and Lyn activities *in vitro* was accomplished using the purified kinases with the kinase reaction

buffer and detection conditions described in the supporting information. Briefly, after pre-incubation of the kinases with the reaction buffer for 10 minutes, the reaction was initiated by the addition of the biosensor(s). Aliquots were removed from the reaction, quenched with urea, treated with Tb^{3+}, and brought to a volume of 100 µL. In the presence of only one or the other of the kinases, TR-LRET emission spectra for each respective biosensor displayed an increase in the conjugated dye's fluorescence signal (with minimal bleed through or background interference from the fluorophore attached to the other biosensor) over the time course of the reaction (Figure 3A-D). These results confirmed the relative specificity of each biosensor for its individual kinase, in agreement with previously reported results from our laboratory for SAStide and a separate assay using ELISA-based chemifluorescence detection for SFAStide-A (Supporting Information S10).[83] Finally, to demonstrate multiplex detection, both biosensors were incubated with both kinases in a single reaction. A simultaneous increase in intensity for both fluorophores was seen over the time course, indicating an increase in phosphorylation of both peptides (Figure 3E).
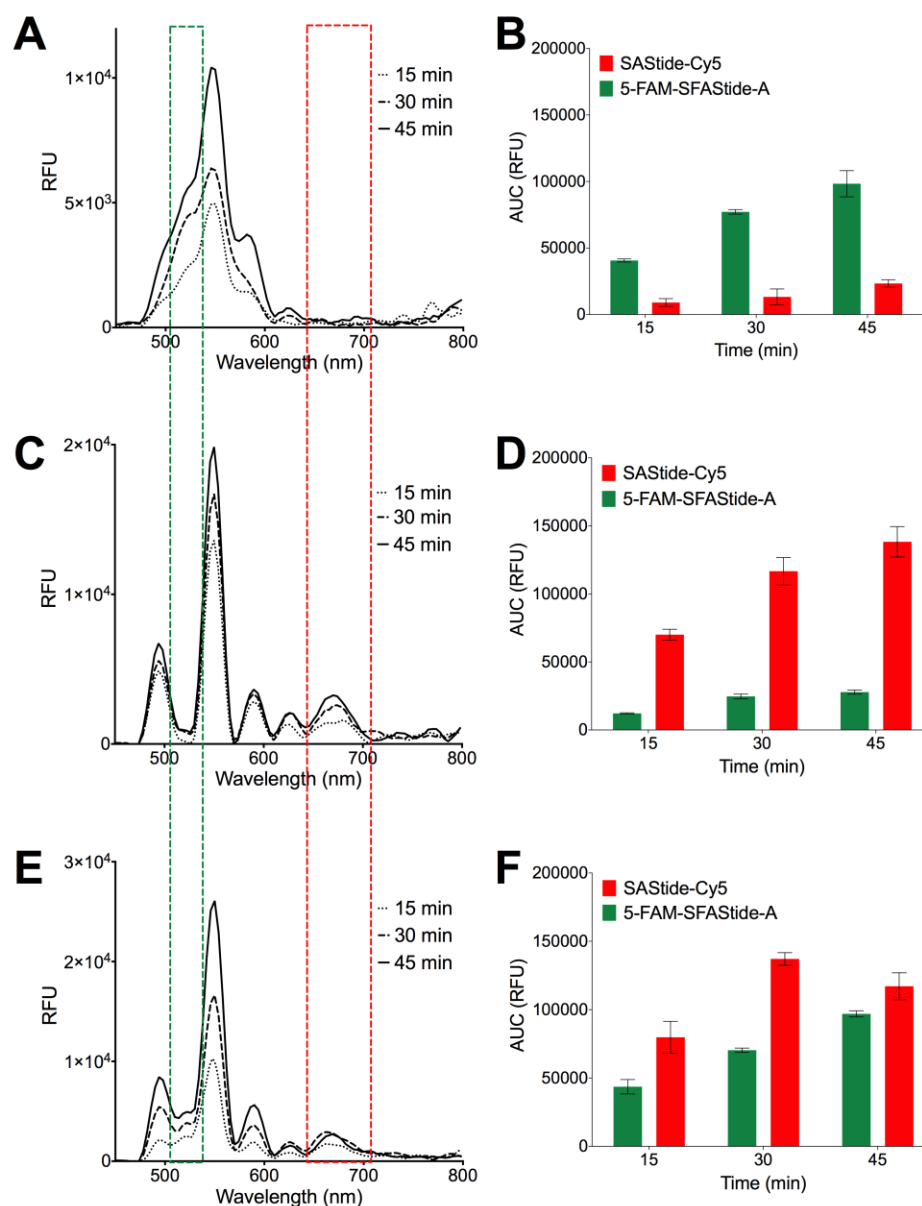
Figure 5.3 Simultaneous multiplexed in vitro detection of Syk and Lyn kinase activities.

(A) In vitro Lyn assay luminescence emission spectra in the presence of both 5-FAM-SFAStide-A and SAStide-Cy5. (C) In vitro Syk assay luminescence emission spectra in the presence of both 5-FAM-SFAStide-A and SAStide-Cy5. (E) In vitro Lyn and Syk assay luminescence emission spectra the presence of both 5-FAM-SFAStide-A and SAStide-Cy5. (B, D, F) Quantification of 5-FAM-SFAStide-A signal and SAStide-Cy5 signal for each assay. The green and red boxes represent the approximate spectral regions represented in the Gaussian fitted curves used to integrate the signal (see supporting information S6 for more detail). Assays were performed in the presence of 2.5 µM 5-FAM-SFAStide-A, 12.5 µM SAStide-Cy5, Lyn, Syk or both kinases (15 nM), 100 µM ATP, 10 mM MgCl2 and 0.2 ng/µL BSA.

In summary, we have presented the development of a platform for detection of kinase activity that leverages the overlap of the multiple distinct emission bands of Tb^{3+} with orthogonal fluorescently labeled peptide substrates that are capable of phosphorylation-enhanced Tb^{3+} luminescence. Multiplexed kinase activity detection has remained a challenge in the field, with only a few examples of successful implementation. The Lawrence group accomplished dual kinase detection using the environmentally sensitive fluorophores oxazine and cascade yellow conjugated to peptide substrates for the Lyn and Abl kinases, respectively.[225] Unfortunately, most environmentally-sensitive fluorophores are limited in their application in more complex or higher throughput systems by small dynamic ranges and problems with background fluorescence.

A key point is that the approach presented here circumvents some of the limitations of antibody-based TR-FRET/LRET approaches and complements the previous strategies, enabling direct sensing of phosphate incorporation to the biosensors—avoiding the need for antibody labels and giving high signal-to-noise, streamlining the path from enzyme reaction to assay read-out. This strategy should be compatible with other kinases and fluorophores to increase the number of activities monitored in a single reaction, setting the stage for pathway-based drug screening to target signaling pathway reprogramming in inhibitor resistance. Future application to real-time activity monitoring would further extend the utility of this strategy.

ASSOCIATED CONTENT

Supporting Information. Detailed experimental conditions, peptide characterization, TR-LRET parameters (e.g. lifetimes, etc.) and additional supporting data as referenced in the text are provided. This material is available free of charge via the Internet at http://pubs.acs.org.

Conflict of Interest Disclosure

Dr. Parker and Dr. Lipchik own equity in and serve on the Scientific Advisory Board of KinaSense. Dr. Parker's relationship with KinaSense has been reviewed and managed by the University of Minnesota in accordance with its conflict of interest polices.

AUTHOR INFORMATION

Present Addresses

‡Department of Genetics, Stanford University School of Medicine, Stanford, CA, 94305

§Department of Biochemistry, Molecular Biology and Biophysics, College of Biological Sciences, University of Minnesota Twin Cities, Minneapolis, MN 55455

CHAPTER 6.   CONCLUSION

6.1   Concluding remarks

Protein kinases regulate a variety of cellular functions that help maintain cellular homeostasis. Understanding a kinase's biological role in cellular signaling can aid in identifying uncontrolled kinases. Abnormal kinase signaling can lead to a variety of human diseases, such as cancer, which become addicted to over active kinases to promote disease progression and tumorigenesis.

Currently, a majority of the human kinome is understudied and the lack of knowledge limits drug discovery efforts to identify potent and specific inhibitors. Kinase inhibitors are effective tools used to combat kinase driven diseases, however, off-target activity can induce undesired side-effects that limit the use of potent kinase inhibitors. Understanding how understudied kinases are involved in cellular signaling will help focus drug discovery efforts towards kinases that are involved in human diseases.

Conversely, non-specific kinase inhibitors can be viable tools to help elucidate unknown kinase biology. Non-specific kinase inhibitors that have activity towards understudied kinases can be used as a starting point to develop specific and selective inhibitors to clarify a kinase's biological role. Tools that can be used to monitor kinase activity are needed to determine the efficacy of new kinase inhibitors. Developing sensitive kinase activity assays depends on the efficiency of a kinase to phosphorylate a peptide substrate. Many approaches are available to identify a kinase's substrate preference and have been used to determine a kinase's preferred sequences for phosphorylation. The advantages and limitations of those approaches are highlighted in chapter one. In brief, these experimental approaches are either time intensive or cost prohibitive to perform on many kinases. Incorporating computational approaches to predict biological or artificial substrates using existing biological data have been attempted, with varying success.

There is a need for an efficient pipeline that can be used to identify a kinase's preferred substrate sequence. In chapter two, my contributions helped establish and validate the KINATEST-ID pipeline as a computational tool to use known kinase substrates from proteomic databases to predict artificial peptide sequence for antibody free

assay design. One limitation to the original KINATEST-ID pipeline is that it depended on known kinase substrates to make an accurate substrate prediction.

To this end, in chapter three, we adapted an in vitro linear kinase assay approach coupled with mass spectrometry for high throughput identification of kinase substrates. We used this approach to increase the list of known substrates for FLT3 and two clinically relevant kinases. We developed a series of data formatting tools to process the mass spectrometry data to make it compatible with the KINATEST-ID algorithms. This approach was used to validate seven artificial sequences as pan-FLT3 kinase substrates.

The focus of chapter four centered on establishing an automated process, within the GalaxyP framework, to analyze the mass spectrometry data and identify candidate substrate sequences with minimal user feedback. We used this approach to design antibody and lanthanide-based activity assays for BTK kinase.

In chapter five, my contributions led to validating a multicolored lanthanide-based multiplexed assay that can be used to monitor kinase specific activity. The reported antibody-free assay can benefit drug discovery efforts by having an efficient assay to monitor the specificity of a kinase inhibitor in an in vitro system. The results in chapter five were a validation of how lanthanide sensitizing peptides can be used in this setting.

Similar approaches coupling in vitro kinase reactions with mass spectrometry as a readout have been used to identify kinase substrates, however, our application couples mass spectrometry with the KINATEST-ID algorithms. Coupling phosphoproteomics with the KINATEST-ID pipeline creates a stream lined workflow to develop detection assays for understudied kinases. In turn, these sensitive activity assays can be used to identify new and specific kinase inhibitors to combat drug resistance in human diseases.

REFERENCES

1.   Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409,** 860–921 (2001).

2.   Hanks, S. K. Genomic analysis of the eukaryotic protein kinase superfamily: A perspective. *Genome Biol.* **4,** (2003).

3.   Lemmon, M. A. & Schlessinger, J. Cell signaling by receptor tyrosine kinases. *Cell* **141,** 1117–34 (2010).

4.   C. Manning, D.B. Whyte, R. Martinez, T. hunter, S. S. The Protein Kinase Complement of the Human Genome: EBSCOhost. *Sciencemag.Org* **298,** 1912–1934 (2002).

5.   Nolen, B., Taylor, S. & Ghosh, G. Regulation of Protein Kinases. *Mol. Cell* **15,** 661–675 (2004).

6.   Taylor, S. S. & Kornev, A. P. Protein kinases: evolution of dynamic regulatory proteins. *Trends Biochem. Sci.* **36,** 65–77 (2011).

7.   Huse, M. & Kuriyan, J. The conformational plasticity of protein kinases. *Cell* **109,** 275–282 (2002).

8.   Ubersax, J. a & Ferrell, J. E. Mechanisms of specificity in protein phosphorylation. *Nat. Rev. Mol. Cell Biol.* **8,** 530–41 (2007).

9.   Schlessinger, J. Receptor Tyrosine Kinases: Legacy of the First Two Decades. *Cold Spring Harb. Perspect. Biol.* **6,** a008912–a008912 (2014).

10.   Hubbard, S. R. & Till, J. H. Protein tyrosine kinase structure and function. *Annu. Rev. Biochem.* **69,** 373–98 (2000).

11.   Liu, B. A. *et al.* SH2 Domains Recognize Contextual Peptide Sequence Information to Determine Selectivity. *Mol. Cell. Proteomics* **9,** 2391–2404 (2010).

12.   Xue, L. & Tao, W. A. Current technologies to identify protein kinase substrates in high throughput. *Front. Biol. (Beijing).* **8,** 216–227 (2013).

13.   Lipchik, A. M. & Lafayette, W. Development of Tyrosine Kinase PeptideBiosensors and Methods for Detection. (Purdue University, West Lafayette, IN 47906, 2013).

14.   Johnson, S. A. & Hunter, T. Kinomics: methods for deciphering the kinome. *Nat. Methods* **2,** 17–25 (2005).

15. Pouchain, D., Díaz-Mochón, J. J., Bialy, L. & Bradley, M. A 10,000 member PNA-encoded peptide library for profiling tyrosine kinases. *ACS Chem. Biol.* **2,** 810–8 (2007).

16. Franzini, R. M., Neri, D. & Scheuermann, J. DNA-encoded chemical libraries: Advancing beyond conventional small-molecule libraries. *Acc. Chem. Res.* **47,** 1247–1255 (2014).

17. Schmitz, R., Baumann, G. & Gram, H. Catalytic specificity of phosphotyrosine kinases Blk, Lyn, c-Src and Syk as assessed by phage display. *J. Mol. Biol.* **260,** 664–677 (1996).

18. Cujec, T. P., Medeiros, P. F., Hammond, P., Rise, C. & Kreider, B. L. Selection of v-abl tyrosine kinase substrate sequences from randomized peptide and cellular proteomic libraries using mRNA display. *Chem. Biol.* **9,** 253–264 (2002).

19. Takahashi, T. T., Austin, R. J. & Roberts, R. W. mRNA display: Ligand discovery, interaction analysis and beyond. *Trends Biochem. Sci.* **28,** 159–165 (2003).

20. Kozlov, I. A. *et al.* A highly scalable peptide-based assay system for proteomics. *PLoS One* **7,** (2012).

21. Mochizuki, Y., Kumachi, S., Nishigaki, K. & Nemoto, N. Increasing the library size in cDNA display by optimizing purification procedures. *Biol. Proced. Online* **15,** 1–5 (2013).

22. Galán, A. *et al.* Library-based display technologies: Where do we stand? *Mol. Biosyst.* **12,** 2342–2358 (2016).

23. Kurz, M., Gu, K., Al-Gawari, A. & Lohse, P. A. cDNA - protein fusions: covalent protein - gene conjugates for the in vitro selection of peptides and proteins. *Chembiochem* **2,** 666–72 (2001).

24. Gray, B. P. & Brown, K. C. Combinatorial peptide libraries: Mining for cell-binding peptides. *Chem. Rev.* **114,** 1020–1081 (2014).

25. Ptacek, J. *et al.* Global analysis of protein phosphorylation in yeast. *Nature* **438,** 679–684 (2005).

26. Chen, C., Nimlamool, W., Miller, C. J., Lou, H. J. & Turk, B. E. Rational Redesign of a Functional Protein Kinase-Substrate Interaction. *ACS Chem. Biol.* acschembio.7b00089 (2017). doi:10.1021/acschembio.7b00089

27. Chen, C. & Turk, B. E. Analysis of serine-threonine kinase specificity using arrayed positional scanning peptide libraries. *Curr. Protoc. Mol. Biol.* **Chapter 18,** 1–15 (2010).

28. Dinkel, H. *et al.* Phospho.ELM: A database of phosphorylation sites-update 2011. *Nucleic Acids Res.* **39,** 261–267 (2011).

29. Keshava Prasad, T. S. *et al.* Human Protein Reference Database--2009 update. *Nucleic Acids Res.* **37,** D767–D772 (2009).

30. Sigrist, C. J. A. *et al.* PROSITE, a protein domain database for functional characterization and annotation. *Nucleic Acids Res.* **38,** 161–166 (2009).

31. Hornbeck, P. V., Chabra, I., Kornhauser, J. M., Skrzypek, E. & Zhang, B. PhosphoSite: A bioinformatics resource dedicated to physiological protein phosphorylation. *Proteomics* **4,** 1551–1561 (2004).

32. Schwartz, D. & Gygi, S. P. An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets. *Nat. Biotechnol.* **23,** 1391–1398 (2005).

33. Obenauer, J. C., Cantley, L. C. & Yaffe, M. B. Scansite 2.0: Proteome-wide prediction of cell signalling interactions using short sequence motifs. *Nucleic Acids Res.* **31,** 3635–3641 (2003).

34. Fujii, K. *et al.* Kinase peptide specificity: Improved determination and relevance to protein phosphorylation. *Proc. Natl. Acad. Sci.* **101,** 13744–13749 (2004).

35. Yaffe, M. B. *et al.* A motif-based profile scanning approach for genome-wide prediction of signaling pathways. *Nat. Biotechnol.* **19,** 348–353 (2001).

36. Henikoff, S. & Henikoff, J. G. Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci.* **89,** 10915–10919 (1992).

37. Pearson, W. R. Selecting the right similarity-scoring matrix. *Curr. Protoc. Bioinforma.* 1–9 (2013). doi:10.1002/0471250953.bi0305s43

38. Suo, S. B., Qiu, J. D., Shi, S. P., Chen, X. & Liang, R. P. PSEA: Kinase-specific prediction and analysis of human phosphorylation substrates. *Sci. Rep.* **4,** 1–10 (2014).

39. Jung, I., Matsuyama, A., Yoshida, M. & Kim, D. PostMod: Sequence based prediction of kinase-specific phosphorylation sites with indirect relationship. *BMC Bioinformatics* **11,** 1–10 (2010).

40. Song, J. *et al.* PhosphoPredict: A bioinformatics tool for prediction of human kinase-specific phosphorylation substrates and sites by integrating heterogeneous feature selection. *Sci. Rep.* **7,** 1–19 (2017).

41. Blom, N., Sicheritz-Pontén, T., Gupta, R., Gammeltoft, S. & Brunak, S. Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics* **4,** 1633–1649 (2004).

42. Walker, J. M. *Phospho-Proteomics*. **527,** (Humana Press, 2009).

43. Kim, J. H., Lee, J., Oh, B., Kimm, K. & Koh, I. S. Prediction of phosphorylation sites using SVMs. *Bioinformatics* **20,** 3179–3184 (2004).

44. Awad, M. & Khanna, R. *Efficient learning machines: Theories, concepts, and applications for engineers and system designers. Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers* (2015). doi:10.1007/978-1-4302-5990-9

45. Dang, T. H., Van Leemput, K., Verschoren, A. & Laukens, K. Prediction of kinase-specific phosphorylation sites using conditional random fields. *Bioinformatics* **24,** 2857–2864 (2008).

46. Huang, H. D., Lee, T. Y., Tzeng, S. W. & Horng, J. T. KinasePhos: A web tool for identifying protein kinase-specific phosphorylation sites. *Nucleic Acids Res.* **33,** 226–229 (2005).

47. Qin, G. M., Li, R. Y. & Zhao, X. M. PhosD: Inferring kinase-substrate interactions based on protein domains. *Bioinformatics* **33,** 1197–1204 (2017).

48. Lahiry, P., Torkamani, A., Schork, N. J. & Hegele, R. A. Kinase mutations in human disease: Interpreting genotype-phenotype relationships. *Nature Reviews Genetics* **11,** 60–74 (2010).

49. Fleuren, E. D. G., Zhang, L., Wu, J. & Daly, R. J. The kinome 'at large' in cancer. *Nat. Rev. Cancer* **16,** 83–98 (2016).

50. Hanahan, D. & Weinberg, R. A. The hallmarks of cancer. *Cell* **100,** 57–70 (2000).

51. Hanahan, D. & Weinberg, R. A. Hallmarks of Cancer: The Next Generation. *Cell* **144,** 646–674 (2011).

52. Verstraete, K. & Savvides, S. N. Extracellular assembly and activation principles of oncogenic class III receptor tyrosine kinases. *Nat. Rev. Cancer* **12,** 753–766 (2012).

53. Simons, M., Gordon, E. & Claesson-Welsh, L. Mechanisms and regulation of endothelial VEGF receptor signalling. *Nat. Rev. Mol. Cell Biol.* **17,** 611–625 (2016).

54. Shibuya, M. Vascular Endothelial Growth Factor (VEGF) and Its Receptor (VEGFR) Signaling in Angiogenesis: A Crucial Target for Anti- and Pro-Angiogenic Therapies. *Genes Cancer* **2,** 1097–1105 (2011).

55. Cairns, R. a, Harris, I. S. & Mak, T. W. Regulation of cancer cell metabolism. *Nat. Rev. Cancer* **11,** 85–95 (2011).

56. Shackelford, D. B. & Shaw, R. J. The LKB1-AMPK pathway: Metabolism and growth control in tumour suppression. *Nat. Rev. Cancer* **9,** 563–575 (2009).

57. Klijn, C. *et al.* A comprehensive transcriptional portrait of human cancer cell lines. *Nat. Biotechnol.* **33,** 306–312 (2015).

58. Stransky, N., Cerami, E., Schalm, S., Kim, J. L. & Lengauer, C. The landscape of kinase fusions in cancer. *Nat. Commun.* **5,** 1–10 (2014).

59. Kumar, R. D. & Bose, R. Analysis of somatic mutations across the kinome reveals loss-of-function mutations in multiple cancer types. *Sci. Rep.* **7,** 1–12 (2017).

60. Torkamani, A., Verkhivker, G. & Schork, N. J. Cancer driver mutations in protein kinase genes. *Cancer Lett.* **281,** 117–127 (2009).

61. Azam, M., Seeliger, M. A., Gray, N. S., Kuriyan, J. & Daley, G. Q. Activation of tyrosine kinases by mutation of the gatekeeper threonine. *Nat. Struct. Mol. Biol.* **15,** 1109–1118 (2008).

62. Khatri, A., Wang, J. & Pendergast, A. M. Multifunctional Abl kinases in health and disease. *J. Cell Sci.* **129,** 9–16 (2016).

63. Foster, S. A. *et al.* Activation Mechanism of Oncogenic Deletion Mutations in BRAF, EGFR, and HER2. *Cancer Cell* **29,** 477–493 (2016).

64. Hornbeck, P. V. *et al.* PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Res.* **43,** D512–D520 (2015).

65. Drewry, D. H. *et al.* Progress towards a public chemogenomic set for protein kinases and a call for contributions. *PLoS One* **12,** e0181585 (2017).

66. Ferguson, F. M. & Gray, N. S. Kinase inhibitors: the road ahead. *Nat. Rev. Drug Discov.* **17,** 353–377 (2018).

67. Munoz, L. Non-kinase targets of protein kinase inhibitors. *Nat. Rev. Drug Discov.* **16,** 424–440 (2017).

68. Treiber, D. K. & Shah, N. P. Ins and outs of kinase DFG motifs. *Chem. Biol.* **20,** 745–746 (2013).

69. Garuti, L., Roberti, M. & Bottegoni, G. Non-ATP competitive protein kinase inhibitors. *Curr. Med. Chem.* **17,** 2804–2821 (2010).

70. Wu, P., Nielsen, T. E. & Clausen, M. H. FDA-approved small-molecule kinase inhibitors. *Trends Pharmacol. Sci.* **36,** 422–439 (2015).

71. Zhang, J., Yang, P. L. & Gray, N. S. Targeting cancer with small molecule kinase inhibitors. *Nat. Rev. Cancer* **9,** 28–39 (2009).

72. Anastassiadis, T., Deacon, S. W., Devarajan, K., Ma, H. & Peterson, J. R. Comprehensive assay of kinase catalytic activity reveals features of kinase inhibitor selectivity. *Nat. Biotechnol.* **29,** 1039–1045 (2011).

73. Racker, E. Use of synthetic amino acid polymers for assay of protein-tyrosine and protein-serine kinases. *Methods Enzymol.* **200,** 107–11 (1991).

74. Blouin, J., Roby, P., Arcand, M., Beaudet, L. & Lipari, F. Catalytic Specificity of Human Protein Tyrosine Kinases Revealed by Peptide Substrate Profiling. *Curr. Chem. Genomics* **5,** 115–121 (2011).

75. Tellez, R., Gatica, M., Allende, C. C. & Allende, J. E. Copolymers of glutamic acid and tyrosine are potent inhibitors of oocyte casein kinase II. *FEBS Lett.* **265,** 113–6 (1990).

76. Deng, Y. *et al.* Global analysis of human nonreceptor tyrosine kinase specificity using high-density peptide microarrays. *J Proteome Res* **13,** 4339–4346 (2014).

77. Miller, C. J. & Turk, B. E. Kinase Screening and Profiling. **1360,** 203–216 (2016).

78. Glickman, J. F. *Assay Development for Protein Kinase Enzymes*. *Assay Guidance Manual* (2004). at <http://www.ncbi.nlm.nih.gov/books/NBK91991/>

79. Brooks, H. B. *et al.* Basics of Enzymatic Assays for HTS. *Assay Guid. Man.* 3–4 (2012). doi:NBK92007

80. Lipchik, A. M. *et al.* KINATEST-ID: A pipeline to develop phosphorylation-dependent terbium sensitizing kinase assays. *J. Am. Chem. Soc.* **137,** 2484–2494 (2015).

81. Cui, W. & Parker, L. L. A time-resolved luminescence biosensor assay for anaplastic lymphoma kinase (ALK) activity. *Chem. Commun.* **51,** 362–365 (2015).

82. Miller, C. J. & Turk, B. E. Homing in: Mechanisms of Substrate Targeting by Protein Kinases. *Trends Biochem. Sci.* **43,** 380–394 (2018).

83. Lipchik, A. M., Killins, R. L., Geahlen, R. L. & Parker, L. L. A peptide-based biosensor assay to detect intracellular Syk kinase activation and inhibition. *Biochemistry* **51,** 7515–7524 (2012).

84. Aebersold, R. & Mann, M. Mass spectrometry-based proteomics. *Nature* **422,** 198–207 (2003).

85. Glish, G. L. & Vachet, R. W. The basics of mass spectrometry in the twenty-first century. *Nat. Rev. Drug Discov.* **2,** 140–150 (2003).

86. Domon, B. & Aebersold, R. Mass spectrometry and protein analysis. *Science* **312,** 212–7 (2006).

87. Tong, M. & Seeliger, M. A. Targeting conformational plasticity of protein kinases. *ACS Chem. Biol.* **10,** 190–200 (2015).

88. Eliuk, S. & Makarov, A. Evolution of Orbitrap Mass Spectrometry Instrumentation. *Annu. Rev. Anal. Chem.* **8,** 61–80 (2015).

89. Lesur, A. & Domon, B. Advances in high-resolution accurate mass spectrometry application to targeted proteomics. *Proteomics* **15,** 880–890 (2015).

90. Iliuk, A. B., Martin, V. a, Alicie, B. M., Geahlen, R. L. & Tao, W. A. In-depth analyses of kinase-dependent tyrosine phosphoproteomes based on metal ion-functionalized soluble nanopolymers. *Mol. Cell. Proteomics* **9,** 2162–2172 (2010).

91. Dengjel, J., Kratchmarova, I. & Blagoev, B. Receptor tyrosine kinase signaling: A view from quantitative proteomics. *Mol. Biosyst.* **5,** 1112–1121 (2009).

92. Larsen, M. R., Thingholm, T. E., Jensen, O. N., Roepstorff, P. & Jørgensen, T. J. D. Highly Selective Enrichment of Phosphorylated Peptides from Peptide Mixtures Using Titanium Dioxide Microcolumns. *Mol. Cell. Proteomics* **4,** 873–886 (2005).

93. Larance, M. & Lamond, A. I. Multidimensional proteomics for cell biology. *Nat. Publ. Gr.* **16,** 269–280 (2015).

94. Hebert, A. S. *et al.* The One Hour Yeast Proteome. *Mol. Cell. Proteomics* **13,** 339–347 (2014).

95. Xue, L., Wang, P., Cao, P., Zhu, J.-K. & Tao, W. A. Identification of ERK1 Direct Substrates using Stable Isotope Labeled Kinase Assay-Linked Phosphoproteomics. *Mol. Cell. Proteomics* **1,** 3199–3210 (2014).

96. Kubota, K. *et al.* Sensitive multiplexed analysis of kinase activities and activity-based kinase identification. *Nat. Biotechnol.* **27,** 933–40 (2009).

97. Xue, L. *et al.* Sensitive kinase assay linked with phosphoproteomics for identifying direct kinase substrates. *Proc. Natl. Acad. Sci.* **109,** 5615–5620 (2012).

98. Yu, Y. *et al.* A site-specific, multiplexed kinase activity assay using stable-isotope dilution and high-resolution mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 11606–11 (2009).

99. Barber, K. W. *et al.* Kinase Substrate Profiling Using a Proteome-wide Serine-Oriented Human Peptide Library. *Biochemistry* **57,** 4717–4725 (2018).

100. Kettenbach, A. N. *et al.* Rapid Determination of Multiple Linear Kinase Substrate Motifs by Mass Spectrometry. *Chem. Biol.* **19,** 608–618 (2012).

101. Jia, Y., Quinn, C., Kwak, S. & Talanian, R. Current In Vitro Kinase Assay Technologies: The Quest for a Universal Format. *Curr. Drug Discov. Technol.* **5,** 59–69 (2008).

102. Li, H. *et al.* Evaluation of an Antibody-Free ADP Detection Assay: ADP-Glo. *Assay Drug Dev. Technol.* **7,** 598–605 (2009).

103. Kashem, M. A. *et al.* Three Mechanistically Distinct Kinase Assays Compared: Measurement of Intrinsic ATPase Activity Identified the Most Comprehensive Set of ITK Inhibitors. *J. Biomol. Screen.* **12,** 70–83 (2007).

104. Hastie, C. J., McLauchlan, H. J. & Cohen, P. Assay of protein kinases using radiolabeled ATP: A protocol. *Nat. Protoc.* **1,** 968–971 (2006).

105. Lipchik, A. M. & Parker, L. L. Time-resolved luminescence detection of spleen tyrosine kinase activity through terbium sensitization. *Anal. Chem.* **85,** 2582–2588 (2013).

106. Fílla, J. & Honys, D. Enrichment techniques employed in phosphoproteomics. *Amino Acids* **43,** 1025–1047 (2012).

107. Mueller, M.-T. *et al.* Combined targeted treatment to eliminate tumorigenic cancer stem cells in human pancreatic cancer. *Gastroenterology* **137,** 1102–13 (2009).

108. Lipchik, A. M., Perez, M., Cui, W. & Parker, L. L. Multicolored, Tb(3)(+)-Based Antibody-Free Detection of Multiple Tyrosine Kinase Activities. *Anal. Chem.* **87,** 7555–7558 (2015).

109. Lu, Y., Prudent, M., Fauvet, B., Lashuel, H. A. & Girault, H. H. Phosphorylation of α-synuclein at Y125 and S129 alters its metal binding properties: Implications for understanding the role of α-synuclein in the pathogenesis of Parkinson's disease and related disorders. *ACS Chem. Neurosci.* **2,** 667–675 (2011).

110. Martinez-Gomez, N. C., Vu, H. N. & Skovran, E. Lanthanide Chemistry: From Coordination in Chemical Complexes Shaping Our Technology to Coordination in Enzymes Shaping Bacterial Metabolism. *Inorg. Chem.* **55,** 10083–10089 (2016).

111. Zhang, K. Y. *et al.* Long-Lived Emissive Probes for Time-Resolved Photoluminescence Bioimaging and Biosensing. *Chem. Rev.* **118,** 1770–1839 (2018).

112. Sumaoka, J., Akiba, H. & Komiyama, M. Selective Sensing of Tyrosine Phosphorylation in Peptides Using Terbium (III) Complexes. *Int. J. Anal. Chem.* **2016,** (2016).

113. Zhang, J., Ma, Y., Taylor, S. S. & Tsien, R. Y. Genetically encoded reporters of protein kinase A activity reveal impact of substrate tethering. *Proc. Natl. Acad. Sci.* **98,** 14997–15002 (2001).

114. Mo, G. C. H. *et al.* Genetically encoded biosensors for visualizing live-cell biochemical activity at super-resolution. *Nat. Methods* **14,** 427–434 (2017).

115. Lavogina, D., Kopanchuk, S. & Viht, K. Dissection of Protein Kinase Pathways in Live Cells Using Photoluminescent Probes: Surveillance or Interrogation? *Chemosensors* **6,** 19 (2018).

116. Conde-Perez, A. & Larue, L. PTEN and melanomagenesis. *Futur. Oncol.* **8,** 1109–1120 (2012).

117. Ren, R. Mechanisms of BCR-ABL in the pathogenesis of chronic myelogenous leukaemia. *Nat. Rev. Cancer* **5,** 172–183 (2005).

118. Tefferi, A., Skoda, R. & Vardiman, J. W. Myeloproliferative neoplasms: contemporary diagnosis using histology and genetics. *Nat. Rev. Clin. Oncol.* **6,** 627–37 (2009).

119. Uitdehaag, J. C. M. *et al.* A guide to picking the most selective kinase inhibitor tool compounds for pharmacological validation of drug targets. *Br. J. Pharmacol.* **166,** 858–876 (2012).

120. Irish, J. M., Czerwinski, D. K., Nolan, G. P. & Levy, R. Kinetics of B cell receptor signaling in human B cell subsets mapped by phosphospecific flow cytometry. *J. Immunol.* **177,** 1581–1589 (2006).

121. Bendall, S. C. *et al.* Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. *Science (80-.).* **332,** 687–696 (2011).

122. Monetti, M., Nagaraj, N., Sharma, K. & Mann, M. Large-scale phosphosite quantification in tissues by a spike-in SILAC method. *Nat. Methods* **8,** 655–8 (2011).

123. Fernandes, N. & Allbritton, N. L. Effect of the DEF motif on phosphorylation of peptide substrates by ERK. *Biochem. Biophys. Res. Commun.* **387,** 414–418 (2009).

124. Zhang, J., Ma, Y., Taylor, S. S. & Tsien, R. Y. Genetically encoded reporters of protein kinase A activity reveal impact of substrate tethering. *Proc. Natl. Acad. Sci.* **98,** 14997–15002 (2001).

125. Wu, C. H. The Protein Information Resource. *Nucleic Acids Res.* **31,** 345–347 (2003).

126. Liu, L. L. & Franz, K. J. Phosphorylation of an α-Synuclein Peptide Fragment Enhances Metal Binding. *J. Am. Chem. Soc.* **127,** 9662–9663 (2005).

127. Liu, L. L. & Franz, K. J. Phosphorylation-dependent metal binding by α-synuclein peptide fragments. *J. Biol. Inorg. Chem.* **12,** 234–247 (2007).

128. Miller, M. L. *et al.* Linear motif atlas for phosphorylation-dependent signaling. *Sci. Signal.* **1,** 1–11 (2008).

129. Newman, R. H. *et al.* Construction of human activity-based phosphorylation networks. *Mol. Syst. Biol.* **9,** 655–655 (2014).

130. Ruzzene, M. *et al.* Sequence specificity of C-terminal Src kinase (CSK)--a comparison with Src-related kinases c-Fgr and Lyn. *Eur J Biochem* **246,** 433–439 (1997).

131. Hantschel, O. *et al.* BCR-ABL uncouples canonical JAK2-STAT5 signaling in chronic myeloid leukemia. *Nat. Chem. Biol.* **8,** 285–293 (2012).

132. Quintas-Cardama, A. *et al.* Preclinical characterization of the selective JAK1/2 inhibitor INCB018424: therapeutic implications for the treatment of myeloproliferative neoplasms. *Blood* **115,** 3109–3117 (2010).

133. O'Hare, T. *et al.* In vitro activity of Bcr-Abl inhibitors AMN107 and BMS-354825 against clinically relevant imatinib-resistant Abl kinase domain mutants. *Cancer Res* **65,** 4500–4505 (2005).

134. Manley, P. W. No Title.

135. Drewry, D. H., Willson, T. M. & Zuercher, W. J. Seeding collaborations to advance kinase science with the GSK Published Kinase Inhibitor Set (PKIS). *Curr Top Med Chem* **14,** 340–342 (2014).

136. Sheridan, D. L., Kong, Y., Parker, S. A., Dalby, K. N. & Turk, B. E. Substrate discrimination among mitogen-activated protein kinases through distinct docking sequence motifs. *J Biol Chem* **283,** 19511–19520 (2008).

137. Tae Ryong Lee, Till, J. H., Lawrence, D. S. & Miller, W. T. Precision substrate targeting of protein kinases v-Abl and c-Src. *J. Biol. Chem.* **270,** 27022–27026 (1995).

138. Miller, W. T. Determinants of Substrate Recognition in Nonreceptor Tyrosine Kinases. *Acc. Chem. Res.* **36,** 393–400 (2003).

139. Placzek, E. A., Plebanek, M. P., Lipchik, A. M., Kidd, S. R. & Parker, L. L. A peptide biosensor for detecting intracellular Abl kinase activity using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Anal Biochem* **397,** 73–78 (2010).

140. Balakrishnan, S. & Zondlo, N. J. Design of a Protein Kinase-Inducible Domain. *J. AM. CHEM. SOC* **128,** 5590–5591 (2006).

141. am Ende, C. W., Meng, H. Y., Ye, M., Pandey, A. K. & Zondlo, N. J. Design of lanthanide fingers: compact lanthanide-binding metalloproteins. *ChemBioChem* **11,** 1738–47 (2010).

142. Nitz, M., Franz, K. J., Maglathlin, R. L. & Imperiali, B. A Powerful Combinatorial Screen to Identify High-Affinity Terbium(III)-Binding Peptides. *ChemBioChem* **4,** 272–276 (2003).

143. Nitz, M. *et al.* Structural origin of the high affinity of a chemically evolved lanthanide-binding peptide. *Angew Chem Int Ed Engl* **43,** 3682–3685 (2004).

144. Reynolds, A. M., Sculimbrene, B. R. & Imperiali, B. Lanthanide-binding tags with unnatural amino acids: sensitizing Tb3+ and Eu3+ luminescence at longer wavelengths. *Bioconjug Chem* **19,** 588–591 (2008).

145. Leick, M. B. & Levis, M. J. The Future of Targeting FLT3 Activation in AML. *Curr. Hematol. Malig. Rep.* **12,** 153–167 (2017).

146. Stirewalt, D. L. & Radich, J. P. The role of FLT3 in haematopoietic malignancies. *Nat. Rev. Cancer* **3,** 650–65 (2003).

147. Pozarowski, P. & Darzynkiewicz, Z. Analysis of cell cycle by flow cytometry. *Methods Mol. Biol.* **281,** 301–11 (2004).

148. Leung, A., Man, C.-H. & Kwong, Y.-L. FLT3 inhibition: a moving and evolving target in acute myeloid leukemia. *Nat. Rev. Leuk.* **27,** 260–268 (2013).

149. Yamamoto, Y. Activating mutation of D835 within the activation loop of FLT3 in human hematologic malignancies. *Blood* **97,** 2434–2439 (2001).

150. Yoshimoto, G. *et al.* FLT3-ITD up-regulates MCL-1 to promote survival of stem cells in acute myeloid leukemia via FLT3-ITD – specific STAT5 activation. *Blood* **114,** 5034–5044 (2009).

151. Kiyoi, H., Ohno, R., Ueda, R., Saito, H. & Naoe, T. Mechanism of constitutive activation of FLT3 with internal tandem duplication in the juxtamembrane domain. *Oncogene* **21,** 2555–2563 (2002).

152. Swords, R., Freeman, C. & Giles, F. Targeting the FMS-like tyrosine kinase 3 in acute myeloid leukemia. *Leukemia* **26,** 2176–2185 (2012).

153. Kim, Y. *et al.* Quantitative fragment analysis of FLT3 -ITD ef fi ciently identifying poor prognostic group with high mutant allele burden or long ITD length. *Nature* **5,** e336-7 (2015).

154. Smith, C. C., Lin, K., Stecula, A., Sali, A. & Shah, N. P. FLT3 D835 mutations confer differential resistance to type II FLT3 inhibitors. *Leukemia* 1–3 (2015). doi:10.1038/leu.2015.165

155. Moore, a S. *et al.* Selective FLT3 inhibition of FLT3-ITD+ acute myeloid leukaemia resulting in secondary D835Y mutation: a model for emerging clinical resistance patterns. *Leukemia* **26,** 1462–1470 (2012).

156. Grunwald, M. R. & Levis, M. J. FLT3 inhibitors for acute myeloid leukemia: a review of their efficacy and mechanisms of resistance. *Int. J. Hematol.* **97,** 683–94 (2013).

157. Daver, N. *et al.* Review Article Secondary mutations as mediators of resistance to targeted therapy in leukemia. *Blood* **125,** 10–20 (2015).

158. Metzelder, S. K. *et al.* High activity of sorafenib in FLT3-ITD-positive acute myeloid leukemia synergizes with allo-immune effects to induce sustained responses. *Leukemia* 2353–2359 (2012). doi:10.1038/leu.2012.105

159. Lindblad, O. *et al.* Aberrant activation of the PI3K/mTOR pathway promotes resistance to sorafenib in AML. *Oncogene* **35,** 5119–5131 (2016).

160. Weisberg, E., Sattler, M., Ray, A. & Griffin, J. D. Drug resistance in mutant FLT3-positive AML. *Oncogene* **29,** 5120–5134 (2010).

161. Renneville, A. *et al.* Cooperating gene mutations in acute myeloid leukemia: a review of the literature. *Leukemia* **22,** 915–31 (2008).

162. Larrosa-Garcia, M. & Baer, M. R. FLT3 Inhibitors in Acute Myeloid Leukemia: Current Status and Future Directions. *Mol. Cancer Ther.* **16,** 991–1001 (2017).

163. Wander, S. a, Levis, M. J. & Fathi, A. T. The evolving role of FLT3 inhibitors in acute myeloid leukemia: quizartinib and beyond. *Ther. Adv. Hematol.* **5,** 65–77 (2014).

164. Williams, A. B. *et al.* Mutations of FLT3/ITD confer resistance to multiple tyrosine kinase inhibitors. *Leukemia* **27,** 48–55 (2013).

165. Zarrinkar, P. P. *et al.* AC220 is a uniquely potent and selective inhibitor of FLT3 for the treatment of acute myeloid leukemia (AML). *October* **114,** 2984–2992 (2009).

166. Galanis, A. *et al.* Crenolanib is a potent inhibitor of flt3 with activity against resistance-Conferring point mutants. *Blood* **123,** 94–100 (2014).

167. Cortes, J. E. *et al.* Phase I Study of Quizartinib Administered Daily to Patients With Relapsed or Refractory Acute Myeloid Leukemia Irrespective of FMS-Like Tyrosine Kinase 3–Internal Tandem Duplication Status. *J. Clin. Oncol.* **31,** 3681–3687 (2013).

168. Zimmerman, E. I. *et al.* Crenolanib is active against models of drug-resistant FLT3-ITD 2 positive acute myeloid leukemia. *Myeloid Neoplasia* **122,** 3607–3615 (2013).

169. Smith, C. C. *et al.* Crenolanib is a selective type I pan-FLT3 inhibitor. *Proc. Natl. Acad. Sci.* **111,** 5319–5324 (2014).

170. Meyer, N. O. *et al.* Multiplex Substrate Profiling by Mass Spectrometry for Kinases as a Method for Revealing Quantitative Substrate Motifs. *Anal. Chem.* **89,** 4550–4558 (2017).

171. Lubner, J. M., Balsbaugh, J. L., Church, G. M., Chou, M. F. & Schwartz, D. Characterizing Protein Kinase Substrate Specificity Using the Proteomic Peptide Library (ProPeL) Approach. *Curr. Protoc. Chem. Biol.* **10,** e38 (2018).

172. Afgan, E. *et al.* The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res.* **46,** W3–W10 (2018).

173. Boekel, J. *et al.* Multi-omic data analysis using Galaxy. *Nat. Biotechnol.* **33,** 137–139 (2015).

174. Shilov, I. V *et al.* The Paragon Algorithm , a Next Generation Search Engine That Uses Sequence Temperature Values and Feature Probabilities to Identify Peptides from Tandem Mass Spectra * □. *Mol. Cell. Proteomics* **6,** 1638–1655 (2007).

175. Tang, W. H., Shilov, I. V. & Seymour, S. L. Nonlinear Fitting Method for Determining Local False Discovery Rates from Decoy Database Searches. *J. Proteome Res.* **7,** 3661–3667 (2008).

176. Hutti, J. E. *et al.* A rapid method for determining protein kinase phosphorylation specificity. *Nat. Methods* **1,** 27–29 (2004).

177. Böhmer, F.-D. & Uecker, A. A substrate peptide for the FLT3 receptor tyrosine kinase. *Br. J. Haematol.* **144,** 127–30 (2009).

178. Warkentin, A. A. *et al.* Overcoming myelosuppression due to synthetic lethal toxicity for FLT3-targeted acute myeloid leukemia therapy. *Elife* **3,** 1–17 (2014).

179. Marholz, L. J., Zeringo, N. A., Lou, H. J., Turk, B. E. & Parker, L. L. In Silico Design and in Vitro Characterization of Universal Tyrosine Kinase Peptide Substrates. *Biochemistry* **57,** 1847–1851 (2018).

180. Liegel, J., Courville, E., Sachs, Z. & Ustun, C. Use of sorafenib for post-transplant relapse in FLT3/ITD-positive acute myelogenous leukemia: Maturation induction and cytotoxic effect. *Haematologica* **99,** e222–e224 (2014).

181. Smith, C. C. *et al.* Heterogeneous resistance to quizartinib in acute myeloid leukemia revealed by single-cell analysis. *Blood* **130,** 48–58 (2017).

182. Zorn, J. A., Wang, Q., Fujimura, E., Barros, T. & Kuriyan, J. Crystal Structure of the FLT3 Kinase Domain Bound to the Inhibitor Quizartinib (AC220). *PLoS One* **10,** e0121177 (2015).

183. Gozgit, J. M. *et al.* Potent activity of ponatinib (AP24534) in models of FLT3-driven acute myeloid leukemia and other hematologic malignancies. *Mol. Cancer Ther.* **10,** 1028–1035 (2011).

184. Leung, a Y. H., Man, C.-H. & Kwong, Y.-L. FLT3 inhibition: a moving and evolving target in acute myeloid leukaemia. *Leukemia* **27,** 260–268 (2012).

185. Ivry, S. L. *et al.* Global substrate specificity profiling of post-translational modifying enzymes. *Protein Sci.* **27,** 584–594 (2018).

186. Smith, C. I. E. From identification of the BTK kinase to effective management of leukemia. *Oncogene* **36,** 2045–2053 (2017).

187. Lowry, W. E., Huang, J., Lei, M., Rawlings, D. & Huang, X.-Y. Role of the PHTH Module in Protein Substrate Recognition by Bruton's Agammaglobulinemia Tyrosine Kinase. *J. Biol. Chem.* **276,** 45276–45281 (2001).

188. Brown, J. R. *et al.* The Bruton tyrosine kinase inhibitor ibrutinib with chemoimmunotherapy in patients with chronic lymphocytic leukemia. *Blood* **125,** 2915–2923 (2016).

189. Hendriks, R. W., Yuvaraj, S. & Kil, L. P. Targeting Bruton's tyrosine kinase in B cell malignancies. *Nat. Rev. Cancer* **14,** 219–232 (2014).

190. Joseph, R. E., Wales, T. E., Fulton, D. B., Engen, J. R. & Andreotti, A. H. Achieving a Graded Immune Response: BTK Adopts a Range of Active/Inactive Conformations Dictated by Multiple Interdomain Contacts. *Structure* **25,** 1481–1494.e4 (2017).

191. Page, T. H. *et al.* Bruton's tyrosine kinase regulates TLR7/8-induced TNF transcription via nuclear factor-κB recruitment. *Biochem. Biophys. Res. Commun.* **499,** 260–266 (2018).

192. Lee, K.-G. *et al.* Bruton's tyrosine kinase phosphorylates Toll-like receptor 3 to initiate antiviral response. *Proc. Natl. Acad. Sci.* **109,** 5791–5796 (2012).

193. Whang, J. A. & Chang, B. Y. Bruton's tyrosine kinase inhibitors for the treatment of rheumatoid arthritis. *Drug Discov. Today* **19,** 1200–1204 (2014).

194. Grassilli, E. *et al.* A novel oncogenic BTK isoform is overexpressed in colon cancers and required for RAS-mediated transformation. *Oncogene* **35,** 4368–4378 (2016).

195. Wu, P., Nielsen, T. E. & Clausen, M. H. Small-molecule kinase inhibitors: an analysis of FDA-approved drugs. *Drug Discov. Today* **21,** 5–10 (2016).

196. Wu, J., Liu, C., Tsui, S. T. & Liu, D. Second-generation inhibitors of Bruton tyrosine kinase. *J. Hematol. Oncol.* **9,** 80 (2016).

197. Liang, C. *et al.* The development of Bruton's tyrosine kinase (BTK) inhibitors from 2012 to 2017: A mini-review. *Eur. J. Med. Chem.* **151,** 315–326 (2018).

198. Wu, J., Zhang, M. & Liu, D. Acalabrutinib (ACP-196): a selective second-generation BTK inhibitor. *J. Hematol. Oncol.* **9,** 21 (2016).

199. Oellerich, T. *et al.* FLT3-ITD and TLR9 use Bruton tyrosine kinase to activate distinct transcriptional programs mediating AML cell survival and proliferation. *Blood* **125,** 1936–1947 (2015).

200. Jagtap, P. D. *et al.* Flexible and accessible workflows for improved proteogenomic analysis using the galaxy framework. *J. Proteome Res.* **13,** 5898–5908 (2014).

201. Küppers, R. Mechanisms of B-cell lymphoma pathogenesis. *Nat. Rev. Cancer* **5,** 251–262 (2005).

202. Nogai, H., Dörken, B. & Lenz, G. Pathogenesis of non-Hodgkin's lymphoma. *J. Clin. Oncol.* **29,** 1803–1811 (2011).

203. Mahadevan, D. & Fisher, R. I. Novel Therapeutics for Aggressive Non-Hodgkin's Lymphoma. *J. Clin. Oncol.* **29,** 1876–1884 (2011).

204. Peyker, A., Rocks, O. & Bastiaens, P. I. H. Imaging activation of two Ras isoforms simultaneously in a single cell. *ChemBioChem* **6,** 78–85 (2005).

205. Galperin, E., Verkhusha, V. V & Sorkin, A. Three-chromophore FRET microscopy to analyze multiprotein interactions in living cells. *Nat. Methods* **1,** 209–217 (2004).

206. Kienzler, A. *et al.* Novel three-color FRET tool box for advanced protein and DNA analysis. *Bioconjug. Chem.* **22,** 1852–1863 (2011).

207. Piljic, A. & Schultz, C. Simultaneous Recording of Multiple Cellular Events by FRET. *ACS Chem. Biol.* **3,** 156–160 (2008).

208. Ding, Y., Ai, H., Hoi, H. & Campbell, R. E. Förster Resonance Energy Transfer-Based Biosensors for Multiparameter Ratiometric Imaging of Ca 2+ Dynamics and Caspase-3 Activity in Single Cells. *Anal. Chem.* **83,** 9687–9693 (2011).

209. Hermanson, S. B. *et al.* Screening for Novel LRRK2 Inhibitors Using a High-Throughput TR-FRET Cellular Assay for LRRK2 Ser935 Phosphorylation. *PLoS One* **7,** e43580 (2012).

210. Jeyakumar, M. & Katzenellenbogen, J. A. A dual-acceptor time-resolved Föster resonance energy transfer assay for simultaneous determination of thyroid hormone regulation of corepressor and coactivator binding to the thyroid hormone receptor: Mimicking the cellular context of thyroid hormone act. *Anal. Biochem.* **386,** 73–78 (2009).

211. Jeyakumar, M., Webb, P., Baxter, J. D., Scanlan, T. S. & Katzenellenbogen, J. A. Quantification of Ligand-Regulated Nuclear Receptor Corepressor and Coactivator Binding, Key Interactions Determining Ligand Potency and Efficacy for the Thyroid Hormone Receptor †. *Biochemistry* **47,** 7465–7476 (2008).

212. Sculimbrene, B. R. & Imperiali, B. Lanthanide-Binding Tags as Luminescent Probes for Studying Protein Interactions. *J. Am. Chem. Soc.* **128,** 7346–7352 (2006).

213. Vuojola, J., Syrjänpää, M., Lamminmäki, U. & Soukka, T. Genetically Encoded Protease Substrate Based on Lanthanide-Binding Peptide for Time-Gated Fluorescence Detection. *Anal. Chem.* **85,** 1367–1373 (2013).

214. Weitz, E. A., Chang, J. Y., Rosenfield, A. H. & Pierre, V. C. A Selective Luminescent Probe for the Direct Time-Gated Detection of Adenosine Triphosphate. *J. Am. Chem. Soc.* **134,** 16099–16102 (2012).

215. Rajapakse, H. E. *et al.* Time-resolved luminescence resonance energy transfer imaging of protein-protein interactions in living cells. *Proc Natl Acad Sci U S A* **107,** 13582–13587 (2010).

216. Hussong, R. & Hildebrandt, A. *Signal processing in proteomics. Proteome Bioinformatics, Methods in Molecular Biology* **604,** (2010).

217. Yapici, E., Reddy, D. R. & Miller, L. W. An Adaptable Luminescence Resonance Energy Transfer Assay for Measuring and Screening Protein-Protein Interactions and their Inhibition. *ChemBioChem* **13,** 553–558 (2012).

218. Hildebrandt, N., Wegner, K. D. & Algar, W. R. Luminescent terbium complexes: Superior Förster resonance energy transfer donors for flexible and sensitive multiplexed biosensing. *Coord. Chem. Rev.* **273–274,** 125–138 (2014).

219. Kim, S. H., Gunther, J. R. & Katzenellenbogen, J. A. Monitoring a coordinated exchange process in a four-component biological interaction system: Development of a time-resolved terbium-based one-donor/three-acceptor multicolor FRET system. *J. Am. Chem. Soc.* **132,** 4685–4692 (2010).

220. Horton, R. A. & Vogel, K. W. Multiplexing terbium- and europium-based TR-FRET readouts to increase kinase assay capacity. *J Biomol Screen* **15,** 1008–1015 (2010).

221. Kupcho, K. R. *et al.* Simultaneous monitoring of discrete binding events using dual-acceptor terbium-based LRET. *J Am Chem Soc* **129,** 13372–13373 (2007).

222. Carmichael, I., Helman, W. P. & Hug, G. L. Extinction Coefficients of Triplet–Triplet Absorption Spectra of Organic Molecules in Condensed Phases: A Least-Squares Analysis. *J. Phys. Chem. Ref. Data* **16,** 239–260 (1987).

223. Roalstad, S., Rue, C., LeMaster, C. B. & Lasko, C. A Room Temperature Emission Lifetime Experiment for the Physical Chemistry Laboratory. *J. Chem. Educ.* **74,** 853 (1997).

224. Lam, S. K. & Lo, D. Time-resolved spectroscopic study of phosphorescence and delayed fluorescence of dyes in silica-gel glasses. *Chem. Phys. Lett.* **281,** 35–43 (1997).

225. Wang, Q. *et al.* Multicolor monitoring of dysregulated protein kinases in chronic myelogenous leukemia. *ACS Chem Biol* **5,** 887–895 (2010).

VITA

MINERVO PEREZ

EDUCATION

Purdue University, West Lafayette, Indiana, USA   **Expected Fall 2018**
Graduate Student, PULSe

Benedictine University, Lisle, Illinois, USA
B.S., Health Science   **2011**
Academic Advisor: Monica Tischler, Ph.D.
Elgin Community College, Elgin, Illinois, USA
General Education   **2011**

RESEARCH EXPERIENCE

Graduate Research Specialist, University of Minnesota   **2015-Present**
Department of Biochemistry, Molecular Biology and
Biophysics
Kinase assay linked with phosphoproteomic adaptation
and implementation with the KINATEST-ID Platform.

Graduate Research Assistant, Purdue University   **2012-Present**
Purdue University Interdisciplinary Life Sciences Program
Involved in the development of the KINATEST-ID Platform.

Graduate Assistant, Purdue University   **2012-2014**
The Office of Interdisciplinary Graduate Programs
Purdue University Interdisciplinary Life Sciences Program
Social media coordinator.

Research Assistant/Laboratory Technician, Purdue University   **2011-2012**
Parker Lab, Department of Medicinal Chemistry and
Molecular Pharmacology.
Exploring substrate selectivity through a bioinformatics
approach for peptide-based biosensor development.

Undergraduate Research Assistant, Purdue University   **Summer 2011**
Summer Research Opportunity Program (SROP).

Undergraduate Research Assistant, Benedictine University   **Spring 2011**
Biochemistry and Molecular Biology Cell Biology
Laboratory.

TEACHING EXPERIENCE

Organic Chemistry Laboratory Graduate Teaching Assistant
Purdue University, College of Pharmacy   **Spring 2014**
Department of Medicinal Chemistry and Molecular
Pharmacology

MEMBERSHIPS

American Society for Mass Spectrometry (ASMS)	**2018-Present**
American Association For The Advancement Of Science (AAAS)	**2017-Present**
US Human Proteomics Organization (US HUPO)	**2012-Present**
American Chemical Society (ACS)	**2012-2013**

AWARDS

US-HUPO Poster Presentation Award	**2017**
Travel Award, NCI CRCHD Professional Development Workshop	**2016**
US HUPO Travel Stipend, US HUPO 12TH Annual Conference	
Travel Award, NCI CRCHD Professional Development Workshop	**2016**
Special Recognition for First-Year Poster Award	
Purdue University	**2016**

WORKSHOPS & CAREER DEVELOPMENT

NCI Graduate Student Recruitment Program	**Spring, 2018**
NCI CRCHD Professional Development Workshop	**Spring, 2016**
Emerge Bioscience Career Development	**Spring, 2016**
CBC Summer workshop in Proteomics and Informatics	**Summer, 2012**

RESEARCH INTEREST

Keywords: Proteogenomics, Cancer Detection and Diagnostics, Prostate Cancer, Pancreatic Cancer, Leukemia and Enzymes.

PUBLICATIONS (PUBLISHED)

Lipchik AM, **Perez M**, Cui W, Parker LL. Multicolored, Tb^{3+}-Based Antibody-Free Detection of Multiple Tyrosine Kinase Activities. Anal Chem. 2015 Aug 4;87(15):7555-8.

Lipchik AM, **Perez M**, Bolton S, Dumrongprechachan V, Ouellette SB, Cui W, Parker LL. KINATEST-ID: a pipeline to develop phosphorylation-dependent terbium sensitizing kinase assays. J Am Chem Soc. 2015 Feb 25;137(7):2484-94.

Gui J, Liu B, Cao G, Lipchik AM, **Perez M**, Dekan Z, Mobli M, Daly NL, Alewood PF, Parker LL, King GF, Zhou Y, Jordt SE, Nitabach MN. A tarantula-venom peptide antagonizes the TRPA1 nociceptor ion channel by binding to the S1-S4 gating domain. Curr Biol. 2014 Mar 3;24(5):473-83.

PRESENTATIONS

Perez, M., Blankenhorn, J. Murray, K., Tao, W.A., Parker, LL. US HUPO Annual Conference, Minneapolis, "Adaptation of KALIP for the Development and Prediction of Artificial Peptide Substrates to Monitor FMS-Like Tyrosine kinase 3 (FLT3) Activity." March 12, 2018. Lighting Talk.

Perez, M., Blankenhorn, J. Murray, K., Tao, W.A., Parker, LL. US HUPO Annual Conference, Minneapolis, "Adaptation of KALIP for the Development and Prediction of Artificial Peptide Substrates to Monitor FMS-Like Tyrosine kinase 3 (FLT3) Activity" March 12, 2018. Poster.

Perez, M., Tao, W.A., Parker, LL. Biochemistry, Molecular Biology and Biophysics Department Retreat, Lake Itasca, MN, "The Development of FLT3 Artificial Peptide Substrates (FAStide) Through Kinase Assay Linked with Phosphoproteomics (KALIP)." September 30, 2017. Talk.

Perez, M., Tao, W.A., Parker, LL. Celebrating 21 Years of the CURE Program, National Cancer Institute Center to reduce Cancer Heath Disparities (CRCHD), Bethesda, MD, "Identification of mutant FMS-like tyrosine kinase 3 substrates using KALIP." June 27, 2017. Poster.

Perez, M., Tao, W.A., Parker, LL. US HUPO 14th Annual Conference, San Diego, "Identification of mutant FMS-like tyrosine kinase 3 substrates using KALIP." March 21, 2017. Lighting Talk.

Perez, M., Hsu, C., Tao, W.A., Parker, LL. US HUPO 14th Annual Conference, San Diego, "Identification of mutant FMS-like tyrosine kinase 3 substrates using KALIP." March 21, 2017. March 21, 2017. Poster.

Perez, M., Hsu, C., Tao, W.A., Parker, LL. National Cancer Institute Center to reduce Cancer Heath Disparities (CRCHD) 2016 Professional Development and Mock Review Workshop, Bethesda, "Identification of mutant FMS-like tyrosine kinase 3 substrates using KALIP." May 24, 2016. Poster.

Perez, M., Hsu, C., Tao, W.A., Parker, LL. US HUPO Conference, Boston, "Identification of mutant FMS-like tyrosine kinase 3 substrates using KALIP." March 14, 2016. Poster.

Perez, M., Hsu, C., Tao, W.A., Parker, LL. US HUPO Conference, Tempe, "Identification of FMS-like tyrosine kinase 3 substrates using KALIP." March 15, 2015. Poster.

Perez, M., Lipchik, AM, Parker LL. "Pursuit of Substrate Specificity." Purdue University Life Science's (PULSe) Spring Reception, 2013 Apr 1, West Lafayette, IN. Poster.