2012

Approaches to Study Small Molecule Inhibitors and Their Targets

Sarah Anne Wacker

APPROACHES TO STUDY SMALL MOLECULE INHIBITORS AND THEIR TARGETS

A Thesis Presented to the Faculty of

The Rockefeller University

in Partial Fulfillment of the Requirements for

the degree of Doctor of Philosophy

by

Sarah Anne Wacker

June 2012

APPROACHES TO STUDY SMALL MOLECULE INHIBITORS AND THEIR TARGETS

Sarah Anne Wacker
The Rockefeller University 2012

Bioactive small molecules are valuable tools to understand and manipulate biological pathways. In order to be effective as either probes for understanding cell biology or as clinical drugs, the small molecule's mechanism of action must be characterized. However, identifying a small molecule's target and characterizing its interaction with that target remain major challenges in the chemical biology field. In this thesis, I describe methods to improve the process of drug target identification and binding site characterization.

In order to identify the target of a small molecule, I have developed an approach in which multiple drug resistant clones are isolated and transcriptome sequencing is used to find mutations in each clone. Further analysis of mutations common to more than one drug-resistant clone can identify a drug's physiological target and indirect resistance mechanisms. In proof-of-concept studies I analyze clones resistant to two cytotoxic anti-cancer drugs, BI 2536 and bortezomib. For both compounds I detect mutations in the known target that confer resistance to the drug. Unlike other target identification methods, this approach can establish a genetic proof of the target in human cells.

I have also developed a method to characterize a small molecule's binding site after its target is known. In this method, called Stable Isotope Labeled Inhibitors for Crosslinking (SILIC), structure-activity relationship data is used to design inhibitor analogs that incorporate a photo-crosslinking group along with either natural or heavy stable isotopes. An equimolar mixture of these inhibitor analogs is crosslinked to the target protein to yield a robust signature for identifying inhibitor-modified peptide fragments in complex mass spectrometry data. I applied this approach to an ATP-competitive inhibitor of kinesin-5, a widely conserved motor protein required for cell division. This analysis, along with mutagenesis studies, suggests that the inhibitor binds at an allosteric site in the motor protein.

Acknowledgements

The Kapoor laboratory was the ideal setting for me to conduct my graduate research. The breadth of research conducted in the lab allowed me to explore a variety of scientific topics and to learn how to use an assortment of experimental tools. I would like to thank my advisor, Tarun Kapoor, for running a laboratory that can be both nurturing for graduate students and intellectually stimulating. Tarun is a fantastic scientist and mentor. Throughout my graduate studies he pushed me to think more deeply about my specific research projects and yet also encouraged me to learn about a broad range of scientific disciplines. I thank Tarun for motivating me to remain interested in science. Additionally, I would like to thank all members of the Kapoor laboratory, both past and present, for creating a congenial work environment and providing advice on my research projects. I am especially grateful to Emily Foley, Radhika Subramanian, Xiang (David) Li, and Lei Tan for thoughtful discussions and scientific advice.

All of the research described in this thesis has been a collaborative effort. I would like to thank Olivier Elemento at Weill Cornell Medical College for conducting all of the bioinformatics analyses for the target identification approach. Ben Houghtaling taught me how to do cell culture and did the first resistance selections with BI 2536. Xiang Li came up with the concept of the SILIC approach and conducted the mass spectrometry experiments after crosslinking. Rachel Fleisher and Sudhir Kashyap synthesized PCTP; Sudhir also synthesized compounds **1** and **2**. Benjamin Kwok and Joshua Weinger both helped me understand “kinesin-ology” and learn the techniques used for studying the kinesin-5 motor.

All of the mass spectrometry experiments were completed in Brian Chait’s laboratory. Zachary Quinkert conducted native mass spectrometry experiments of kinesin-5. Kelly Molloy showed me how to prepare samples for mass spectrometry and allowed me to use many of her reagents. Yinyin Li helped with peak identification and analysis of the mass spectrometry data from the SILIC project.

I relied on the Genomics Resources Core Facility of Weill Cornell Medical College for the Illumina sequencing. For both projects in my thesis I utilized instruments in the High-Throughput Screening Center at Rockefeller University. I would like to especially thank Deena Oren and the Structural Biology Resource

Center at Rockefeller University. Deena was extremely patient and helpful in teaching me about crystallography and spent many hours with me collecting and processing diffraction data.

I would like to thank the members of my faculty advisory committee, Tom Muir and Shai Shaham, for providing me with interesting discussions and helpful suggestions during our annual meetings. I also am thankful to Paramjit Arora for attending my thesis defense and helping to create an intellectually stimulating environment.

I am also grateful for the Dean's office and its staff for making all of the logistical aspects of graduate life so straightforward. Their assistance has allowed me to concentrate my efforts on the science.

Finally I would like to thank Andrés who has encouraged, supported, and loved me throughout my graduate experience. He is my model of what a scientist should be.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 3

Chapter 4

Appendix 2

LIST OF TABLES

Introduction: Techniques for target validation

Many bioactive small molecules function by binding to proteins and affecting those proteins' cellular function(s). These small molecules are valuable tools to study the proteins on which they act (Peterson and Mitchison, 2002), and small molecule "bioprobes" have been successfully used to characterize the roles of proteins involved in a multitude of cellular processes. They are especially useful to probe protein functions since they act on a fast timescale and their effects are frequently reversible (Lampson and Kapoor, 2006). In addition to functioning as tools to study cellular biology, small molecules can be developed into clinical therapeutics.

A clear understanding of the mechanism of action of a small molecule is essential if that compound is to function as a research tool or be developed into a therapeutic agent. This is especially important for a small molecule bioprobe: if the proteins it interacts with are not known, then the compound cannot be used to further understand those proteins. It is also advantageous to identify the protein targets of a drug as this knowledge informs on its potential clinical efficacy. Information regarding a drug's mechanism of action allows for optimization of the drug's specificity during development and, while the drug is in the clinic, provides guidance for potential combination therapies, likely side effects due to additional targets, and insight into drug resistance.

The process of connecting small molecules with their cellular targets is a major challenge in the fields of chemical biology and drug discovery. This chapter will discuss the major methods available for identifying small molecule targets, with a focus on the advantages and disadvantages of each technique.

Methods to identify the targets of small molecules

There are numerous methods that can be used to partially elucidate a small molecule's mechanism of action. Identifying the optimal method to determine a drug's relevant target(s) depends on a number of factors, particularly on the organisms in which a drug is active and on the ease with which a drug can be chemically manipulated. Current methods for target identification can be broadly divided into two categories: biochemical strategies and chemical genetics approaches.

Biochemical strategies

A common approach to target identification relies on *in vitro* readouts of protein activity or small molecule binding (Sleno and Emili, 2008). This is how cyclophilin was identified as the target of cyclosporine A (Handschumacher et al., 1984), how the target of FK506 was determined to be FKBP (Harding et al., 1989), and how histone deacetylases were identified through binding to trapoxin (Taunton et al., 1996). Biochemical target identification strategies have been in use for decades (Cuatrecasas et al., 1968; Reynolds et al., 1978; Rix and

Superti-Furga, 2009) and the basic strategy has benefitted from many technical advances.

Affinity chromatography to isolate binding partners

The proteins that bind directly to a small molecule can be identified through affinity chromatography (**Figure 1.1**). In this method, the small molecule of interest is attached to a solid support and then used to "pull-out" binding partners from cellular extracts. Typically, the small molecule is immobilized to the solid support through a suitable functional group (carboxylic acid, primary amine, or hydroxyl group) or linking molecule. This "affinity matrix" is then combined with a cell or tissue lysate in order to bind proteins. The affinity matrix is stringently washed to remove non-specific binding proteins. Finally the bound proteins are eluted from the solid support by denaturation. The eluted proteins are then identified, usually by mass spectrometry.

While biochemical purification of proteins that bind small molecules has been useful in the identification of many target proteins, the approach has a number of limitations. It can be challenging to generate analogs of a small molecule capable of attaching to the support resin, since this generally requires addition of a separate functional group to the small molecule. The small molecule must be modified in a way that its mechanism of action is unchanged. The small molecule must also be fully accessible to all potential target proteins, and this accessibility can be affected by the chemical modifications required for resin

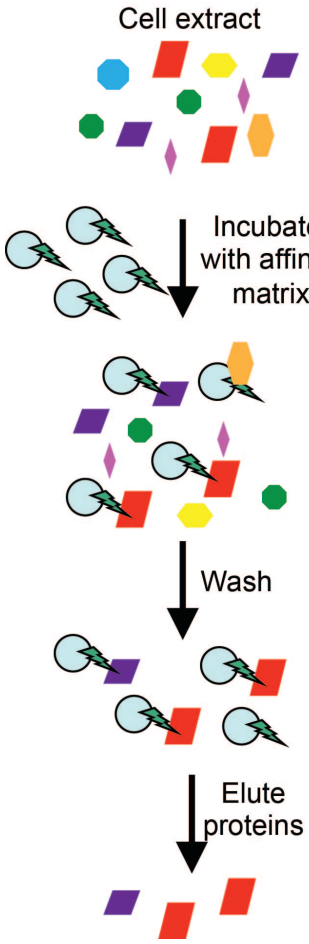


Figure 1.1 Schematic highlighting the key steps of affinity chromatography. A cellular extract is incubated with an affinity matrix consisting of the small molecule of interest (dark green thunderbolt) attached to a support resin (light blue circle). The small molecule interacts with proteins. The affinity matrix is washed to remove unbound and non-specific proteins. The proteins are eluted from the affinity resin and must be identified.

attachment. Ensuring these two basic requirements is difficult when the mechanism of action is unclear and the molecular binding requirements are unknown.

Additional challenges arise during the identification of binding proteins from an affinity chromatography experiment. Even with vigorous washing, it is difficult to eliminate all non-specific binding proteins. Furthermore, the small molecule—target interaction must be of high affinity in order to persist throughout the washing steps. Finally, not all targets are soluble in the media used for binding and, thus, the relevant binding partners for a small molecule may not be amenable to bioinformatics identification.

To discriminate between real binding partners and non-specific binding proteins, the eluted proteins can be compared with those found in a negative control sample. Ideally, a non-active small molecule analog is examined in parallel with the active small molecule (Oda et al., 2003). Since inactive compounds are frequently not available, it is also possible to use the resin alone or the solid support with a similar linking group as control conditions. Alternatively, chemically distinct molecules may be used to identify and deprioritize proteins that frequently appear in affinity chromatography experiments. It is also possible to compare proteins eluted after serial binding of lysate to the affinity matrix, which assumes that specific binding proteins will be the first to be depleted from the lysate (Yamamoto et al., 2006).

Even with proper negative controls, it can be difficult to determine whether, and by how much, proteins are enriched in the sample containing the active compound over the control sample. To more easily discriminate between protein levels in the two samples, stable isotope labeling of proteins has been employed (Ong et al., 2009). This requires two cell lysate samples: one natural sample and the other containing proteins labeled with heavy isotopes. The two protein extracts are then incubated with either the drug-bound affinity matrix or the negative control. After elution, the samples are combined and processed together by mass spectrometry. As proteins from the heavy isotope sample are heavier than proteins from the unlabeled sample by a known mass, this change in mass allows proteins associated with the affinity matrix to be easily distinguished from proteins that are present in similar levels in the negative control. The addition of stable isotopes allows for better detection of low abundance proteins and prioritization of more promising candidates because of the quantitative detection (Chan et al., 2009).

Another strategy for identifying target proteins with low abundance or which have lower affinity for the drug is by creating covalent attachments between the small molecule and its target proteins. This makes it easier to wash away non-specific binding proteins without losing real target proteins that are of low abundance or have weak association kinetics with the bioactive molecule. This approach was useful in identifying the target of the anti-angiogenic compound fumagillin (Sin et al., 1997).

Activity-based protein profiling

In addition to affinity-matrix purification of drug binding partners, monitoring the activity of potential binding proteins can be used to determine the selectivity of a small molecule. Activity-based protein profiling (ABPP), a technique originally developed to identify which members of a protein family are active (Liu et al., 1999), has been adapted for target identification (Cravatt et al., 2008). ABPP relies on a probe, separate from the small molecule inhibitor, that targets the active site of all proteins in a defined enzyme family. This probe can be used to isolate or visualize active members of this family. The technique can be used for target identification if the small molecule inhibits proteins so that they no longer bind to the active-site-directed probe. By comparing the active proteins in a control protein sample with those in a sample pretreated with the small molecule inhibitor, one can infer which proteins' activities have been impacted by the drug.

While ABPP has the benefit of providing information on protein activity, it has a number of disadvantages compared with affinity chromatography. ABPP is only useful if the general target class is already known. Although ABPP identifies how selective a compound is for a given class of enzymes, it does not provide any data on unrelated proteins. For enzymes whose activities rely on one another, such as kinases in a signaling pathway, it is not possible to know whether a protein is inactive because it is directly inhibited by the small molecule

or the small molecule inhibited an upstream protein. For these reasons, there are limited examples of ABPP being used for target identification (Cravatt et al., 2008; Leung et al., 2003).

When a biochemical target identification strategy is successful, a small number of specific binding proteins are identified; these proteins must be further tested in cells to determine if they are associated with the drug's biological activity. This represents a substantial limitation of *in vitro* approaches: although they can identify direct binding partners, more research is needed before these binding partners are associated with the drug's activity in cells. This limitation is magnified by potential discrepancies between the effect of a drug on a protein and the phenotype after knockdown of a protein's levels by RNA interference (RNAi) (Weiss et al., 2007). As RNAi is a common method for validating drug targets in cells, any discrepancies between the two techniques can lead to uncertainty as to the drug's true target. Due to its limitations, more scientists are using biochemical studies in combination with cell-based approaches to reach a greater understanding of drug effectiveness (Rix and Superti-Furga, 2009). Some cell-based approaches for target identification are described in the next section.

Chemical Genetics Approaches

Genetics-based approaches to drug target identification are useful alternatives to methods that rely on biochemical techniques which later require validation in cells. The concept behind chemical genetics approaches is simple:

using a small molecule in combination with genetic perturbations can identify the genetic perturbations, and therefore the genes, which suppress or enhance the compound's effects. These genes are likely to be a part of the pathway on which the drug acts and may be the drug's direct target. Since the genes are identified in cells based upon a change in responsiveness to the drug, the predicted targets are more likely to be physiologically relevant, which is a major advantage compared with a biochemical strategy.

Studies in Model Organisms

Genetic screening in model organisms is a well-established approach for drug target identification. The most frequently utilized model organism for target identification is the yeast *Saccharomyces cerevisiae*, for which advanced genetic tools are available (Ho et al., 2011).

One tool for genetic studies in *S. cerevisiae* is a gene deletion collection of ~6000 heterozygous diploid and ~5000 viable haploid gene deletion strains (Winzeler et al., 1999). Each strain in this library has been tagged with a unique twenty base pair sequence that acts as a "molecular barcode." Thus, strains that survive competitive growth assays can be easily identified.

The yeast gene deletion collection is utilized in haploinsufficiency profiling studies (Giaever et al., 1999). These studies are based on the idea that a drug will more potently affect organisms that have fewer copies of the target gene. This can be the case when the inhibitor prevents a protein from completing an

essential role in the cell; thus, if there are fewer copies of the protein that can take on this role, less inhibitor is needed to compromise its function. In this technique a pool of bar-coded heterozygous yeast strains are grown competitively in the presence of a drug. After growth selections, DNA microarrays are used to identify the strains with compromised growth in the presence of the drug (Giaever et al., 1999; Ho et al., 2011). These strains are likely to be haploinsufficient for genes that encode proteins targeted by the small molecule of interest.

Like many genetics-based approaches, haploinsufficiency profiling can identify both direct drug targets and other components of the relevant signaling pathways. Furthermore, the method is unbiased if the drug inhibits growth. However, haploinsufficiency profiling is unlikely to be effective in identifying drug-binding partners whose function is not impaired by the compound of interest. Additionally, not all target genes are more sensitive to a compound when only one of two gene copies is deleted. In these cases a more severe reduction in target gene expression may be necessary.

While haploinsufficiency profiling identifies gene deletions that further impair cells after drug treatment, other strategies identify genes that, when overexpressed, decrease the potency of a particular compound (**Figure 1.2**) (Luesch et al., 2005). An example of this strategy examines the growth after drug treatment of pools of yeast overexpressing different proteins. This method can identify both direct targets and other gene products in the relevant signaling

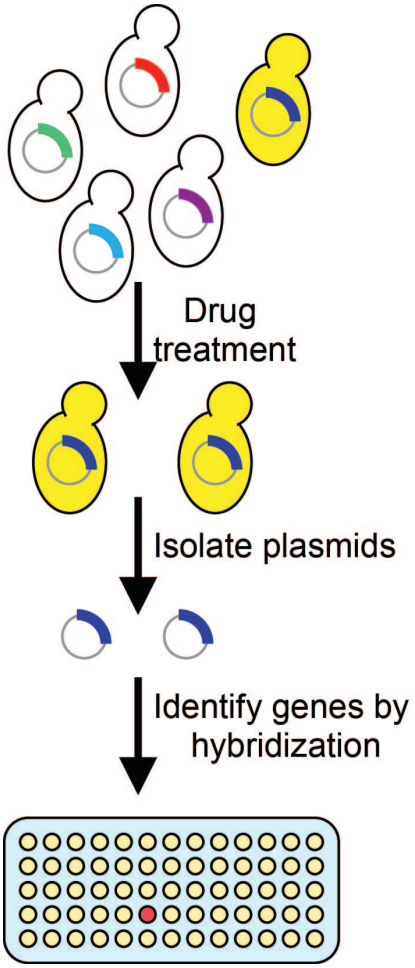


Figure 1.2 Schematic highlighting the key steps of yeast overexpression approaches. A collection of yeast strains, each containing an extra copy of a different gene, are treated with the small molecule of interest at a concentration that impairs the growth of normal yeast. Yeast containing an additional copy of the small molecule's target may grow better than wild type yeast in the drug. Plasmids are isolated from all yeast strains and genes that conferred a growth advantage are more abundant. These genes are identified through hybridization to microarrays.

pathway, as was the case for the drug rapamycin (Butcher et al., 2006). As with gene deletion studies, this basic approach is limited when changes in gene dosage do not impact the efficacy of a small molecule. Furthermore, a large number of candidate genes are often identified.

Forward genetic studies, in which unknown mutations that alter responsiveness to a drug are selected for and characterized, have also been highly valuable for drug target identification. During a forward genetics screen, random mutations are commonly introduced into the genome and mutants that are resistant to the drug are selected. This approach was used to identify the target of rapamycin (TOR) proteins TOR1 and TOR2 and to demonstrate that resistance to rapamycin can occur through mutations in TOR1 and TOR2, along with FPR1 (Heitman et al., 1991). The method is unbiased and can identify a variety of types of mutation (for example, gene deletion or single nucleotide variations). This strategy is especially advantageous because specific mutations are often identified that, if shown to cause resistance to the target *in vitro*, provide strong evidence for a direct target (the gold standard in the field).

The yeast three-hybrid (Y3H) approach takes advantage of model organisms but is able to identify proteins that directly bind a small molecule of interest (Licitra and Liu, 1996). The Y3H system is a variation on the technique of yeast two-hybrid, adapted to identify drug-protein interactions. As its name suggests, Y3H requires a third component: the drug is linked to a ligand that brings together the two protein molecules. When a protein binds to the small

molecule of interest, the proximity of a transcriptional activator, which is linked to the drug-binding protein, results in transcription of a reporter gene.

The Y3H approach has several advantages. Diverse cDNA libraries from different organisms (or tissues) can be examined with the approach. It is possible to directly identify the binding partners of the small molecule. Moreover, unlike biochemical affinity-based methods, proteins of low abundance or that have weak affinity for the small molecule can be identified.

There are also challenges associated with the Y3H approach. As the technique relies on a derivative of the drug that is linked to a ligand, the design and synthesis of this molecule may require a substantial research investment. As yeast frequently have high levels of multi-drug resistance, the intracellular concentration of many small molecules is reduced, decreasing the sensitivity of the method. Many false positives are identified in Y3H; thus, it can be challenging to identify which of the "hits" represent real interactions.

Recent methodological advances have overcome some disadvantages of the Y3H approach. In order to increase the effective intracellular concentration of the drug, *S. cerevisiae* strains have been created with decreased levels of multi-drug pumps (Kolaczkowski et al., 1998). The number of false positives has been reduced by identifying cells with contaminating plasmids and by checking for spontaneous conversion of bait proteins into transcriptional activators (Vidalain et al., 2004). Using a SNAP-tag system allows for the Y3H drug derivative to also be tested by affinity chromatography, providing validation of binding partners

(Chidley et al., 2011). Using these methods, the target of the anti-inflammatory drug sulfasalazine was identified (Chidley et al., 2011).

Some studies combine multiple genetics methodologies for more robust target identification. For example, combining deletion and overexpression libraries allowed Hoon and coworkers to better understand the mechanism of action of two protein phosphatase inhibitors, as each technique further narrowed down the potential targets (Hoon et al., 2008). Likewise, Ho and coworkers made use of both gene overexpression and forward genetics methods, by creating a library of open reading frames to complement and identify mutated genes that result in drug resistance (Ho et al., 2009). This technique allowed for the identification of drug-resistant mutations in the target of theopalauamide, a natural product with potent antifungal activities (Ho et al., 2009). In addition to *S. cerevisiae*, other model organisms are becoming useful for target identification studies, including *S. pombe* (Nishimura et al., 2010), bacteria (Pathania et al., 2009), and *C. elegans* (Burns et al., 2006; Kaletta and Hengartner, 2006).

Chemical genetic approaches in model organisms have been useful for a number of case studies and are ideal ways to study drugs that target these organisms. A major limitation of studies in model organisms, however, is the potential for divergence between their target proteins and those which exist in humans. This divergence exists both in the types of targets (many human proteins do not have homologs in *S. cerevisiae* and other organisms) and in the sequence of the targets (because of variation in amino acid sequence among

clear homologs, some drugs are not effective against both yeast and human counterparts). Another drawback is that for many drugs, particularly anti-cancer agents, the relevant setting is a diseased state. This suggests that important targets may not exist in the model organism. Finally, as yeast have high expression of drug pumps, many small molecules do not effectively get into these organisms.

It is possible to overcome some of these challenges. For example, targets from the relevant species can be expressed in model organisms (Arnoldo et al., 2008). Alternatively, some human diseases can be mimicked in other organisms before screening for drug targets (Hartwell et al., 1997). However, there has been a recent push to conduct more target identification studies in human cells.

Techniques using human cells

Due to the biology of human cells and the human reproductive system, along with the complexity of the human genome, there are not as many genetic resources available for studying human cells as there are for model organisms. Thus, target identification studies using human cells are relatively few. Nevertheless, several new technologies have been developed and it is likely that these approaches will rapidly alter the chemical biology field.

Whole-genome transcript profiling is one of the more promising target identification techniques available for studying compounds in human cells. The method is based on the idea that if drugs have common mechanisms of action,

then they are likely to produce similar effects on cells. Likewise, if a drug inhibits a protein, genetic deletion of that protein may have a similar effect (Lehar et al., 2008). By matching the transcriptional pattern of cells after treatment with many small molecules, it should be possible to group compounds that elicit similar effects. These groups will then suggest families of compounds that act on the same pathways. An early example of this strategy was conducted in yeast (Hughes et al., 2000). Databases of the effects of chemical treatment were used to identify cellular pathways affected by drug treatment. These pathways could be determined by pattern matching, even in some cases when the effect was very subtle (Gunther et al., 2003).

The most comprehensive pattern-matching project to date is the Connectivity Map database at the Broad Institute (Lamb et al., 2006). This method has been used in several studies, including identifying HSP90 as the target of gedunin and celastrol (Hieronymus et al., 2006). This project conducted high-throughput screening of various human cells treated with small molecules or genetic perturbations (e.g. RNAi or genetic ablation). The transcriptional profiles of the effects of these perturbations were recorded and have been stored in a public reference catalog. Importantly, data collected by the Connectivity Map was obtained in a manner meant to reduce experimental variability, which is a common limitation of transcriptional profiling methods (Lamb, 2007; Lamb et al., 2006). This database can now be queried to identify agents (either genetic or chemical) that result in a similar pattern of gene expression as the compound of

interest. This leads to hypotheses of how the small molecule may act, based upon the known function of the genes and small molecules profiled in the Connectivity Map database.

Compared to other approaches, such as affinity chromatography, the Connectivity Map has both advantages and disadvantages. The proteins and pathways resulting from transcriptional profiling are physiologically relevant in human cells, which is not possible for biochemical approaches. However, the transcriptional data in the Connectivity Map typically represent networks of many proteins and translating the hits to a specific protein target is challenging. Another difficulty is that not all compounds elicit transcriptional changes. For those small molecules that do result in a unique transcriptional profile, there may not be a compound or genetic perturbation in the database that creates a similar effect. Until more agents are analyzed and cataloged, this will likely be a major challenge.

In order to expand the applicability of pattern-matching there has been a call to couple transcriptional profiling with other cellular responses (Feng et al., 2009). Methods, such as high-throughput microscopy, have been successfully used to characterize the dose-dependent effects of diverse small molecules on multiple cancer cells (Perlman et al., 2004). The resulting data can also be grouped based on similarities between compounds, providing a starting point for assigning targets to previously uncharacterized molecules. By combining multiple

phenotypic readouts it is more likely that the predicted target pathways will be reliable and the number of “hits” should be reduced.

As various technologies improve, it is becoming more reasonable to apply the concepts behind chemical genetic screens in model organisms to human cells. One way this has been accomplished is by using RNAi screens to aid in target identification (Brummelkamp et al., 2006; Luo et al., 2008). As with gene deletion in yeast, depleted proteins important for the drug’s mechanism of action should impact the phenotype of this drug, frequently affecting cell survival. There are additional limitations, however, of using RNAi instead of a gene deletion. These include lack of specificity of the RNAi for its protein and varying knockdown efficiencies of RNAi. While redundancy within the RNAi library helps discern the effects of specific proteins, varying reductions of protein levels after RNAi remains a challenge.

Each of the afore-described techniques has been valuable for the identification of some targets, but, as mentioned, they all have limitations. As technology advances, some limitations are being overcome. Depending upon the drug, the organisms it is active on, and, of course, its elusive target, the ideal strategy varies.

For many diseases, highly effective drugs require a differential effect in disease and normal cells. For example, cancer cells are genetically distinct from normal cells and taking advantage of this difference may allow a particular drug

to be effective in only the diseased cell with few side effects. Thus it is important that as methods for target identification advance, studies of the mechanism of action of drugs take place in models that are closest to the treatment context. This will likely involve approaches that are sufficiently flexible to be used not only in different organisms but also in different cell and tissue types.

Methods to identify drug-binding sites

The value of the small molecule both as a bioprobe and as a drug can be limited by a lack of understanding of its mechanism of inhibition and mode of target protein binding. Without these data, it can be difficult to improve potency, evaluate specificity, and fully explain cellular phenotypes resulting from drug treatment. Furthermore identifying binding sites can lead to information about the enzyme's mechanism that can only be guessed at by kinetic experiments. Methods to identify the binding site of a small molecule inhibitor include structural studies, computational docking, and photo-crosslinking.

Structural studies that characterize the interaction between small molecules and their protein targets rely on two major techniques: x-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy. Both of these techniques provide data that can be used to build a representation of the drug—target complex at atomic resolution.

X-ray crystallography studies are the most common means of obtaining structural information on proteins in complex with small molecules (Blundell and

Patel, 2004). This technique relies on forming crystals of the protein bound to the small molecule (or soaking crystals of the apo protein in the small molecule inhibitor) and then diffracting x-rays off of the crystal to obtain a map of the protein's structure (Hassell et al., 2007). The resulting data provide a robust and highly accurate model of the specific interactions between the small molecule and its target protein. However, even for established systems, the rate of failure for generating structures of protein—inhibitor complexes can be substantial and is highly compound-dependent (Chung, 2007). NMR spectroscopy can be used to determine structures of target—ligand complexes for proteins that may not be amenable to crystallographic studies (Pellecchia et al., 2008). However, macromolecular NMR is generally limited to relatively small proteins.

Structure-based computational modeling is an increasingly popular approach to predict small molecule binding sites on proteins. A variety of computational techniques exist, of which the most basic is docking (Mobley and Dill, 2009). Docking uses known protein and small molecule structures to explore many potential ligand-binding sites. These different possibilities are compared based upon likely changes in free energy upon binding. In docking approaches, the protein is typically considered to be rigid, which is a limitation of the method as binding pockets can be hidden and may only reveal themselves when the appropriate ligand is bound (Laskowski et al., 2009). Furthermore, the accuracy of docking methods for identifying protein interaction surfaces are frequently low (Wass et al., 2011). Other computational modeling tools have been developed to

improve upon the accuracy of docking and allow for more unbiased models to be created (Mobley and Dill, 2009). However, these methods frequently rely on molecular dynamics simulations and are computationally intensive.

Photo-crosslinking of the small molecule to its protein target is a valuable technique for studying drug—protein interactions. This technique requires the small molecule to be engineered so it contains a light-activated crosslinking group (for example, a benzophenone group, diazirene, or azido group). Treatment of these groups with light results in highly reactive intermediates that will form covalent attachments with nearby amino acids. To identify how the small molecule binds to the protein, the complex is digested and each peptide is examined by mass spectrometry to determine whether it contains a small molecule adjunct (Robinette et al., 2006). This method, like affinity chromatography, is limited by the design and synthesis of appropriate small molecule analogs. It can also be challenging to identify peptides attached to the small molecule in complicated mass spectra.

In addition to these methods, other biochemical and biophysical strategies can be used to study the interactions between a drug and its target. Mutagenesis studies are frequently utilized to validate the binding sites suggested by the methods described above. Analysis of the location of resistance-conferring mutations in a protein's structure can also predict potential interaction surfaces (Luo et al., 2007). Binding sites can also be computationally predicted from

protein sequence, conservation, and structural features of related proteins (Wass et al., 2011).

Thesis overview

The development of new methodological tools to better understand small molecule—protein interactions is an essential part of improving the use of bioactive compounds both as drugs and as bioprobes. Identifying small molecule targets and characterizing the binding site of a small molecule can provide unexpected opportunities to improve a compound's selectivity and potency. While there are a number of techniques available for both of these purposes, all have limitations. The goal of my thesis research was to develop new ways of studying small molecules to expand the chemical biology toolbox.

The topic of chapter 2 is a method that allows for the identification of small molecule targets in human cells by examining drug resistance. I will describe my proof-of-concept studies that suggest combining transcriptome sequencing and bioinformatics allows for a robust and general way of identifying drug targets.

In chapter 3 I describe another method, which builds upon photo-crosslinking studies, to help determine how a small molecule and protein interact. I use this technique to investigate the binding of a small molecule inhibitor of an essential kinesin protein.

A general analysis of my thesis and ideas for future experiments are discussed in chapter 4.

Chapter 2: Using transcriptome sequencing to identify mechanisms of drug action and resistance

Summary

As described in chapter 1, identifying the target(s) of a small molecule remains a major challenge in drug development and chemical biology. Here, I discuss an approach that combines next-generation sequencing and bioinformatics analyses to identify mutations associated with drug resistance, including mutations in the drug's target.

In this method, multiple drug resistant clones are isolated from a heterogeneous population of human cells. These resistant clones are analyzed by transcriptome sequencing along with the parental population to identify mutations unique to the drug-resistant clones. I show that further analysis of mutations common to more than one independent clone can lead to a drug's physiological target and indirect resistance mechanisms.

In proof-of-concept studies, I analyze clones resistant to two cytotoxic anti-cancer drugs, BI 2536 and bortezomib. In the case of BI 2536 I identify two mutations within the drug's known target, Plk1, that confer resistance to BI 2536 both in cells and *in vitro*. For bortezomib, I obtain a list of five genes that are mutated in more than one clone, including the drug's known target, PSMB5. I show that the mutations in PSMB5 are sufficient for conferring resistance to

bortezomib in cells. This approach has a number of advantages over other target identification methods, particularly that it can establish a genetic proof of the target in human cells.

Introduction

One of the major problems in developing drugs or chemical probes is the difficulty in finding their physiological targets (Burdine and Kodadek, 2004). Currently, approaches to analyze how a drug acts fall into two classes. The first category includes strategies that rely on model organisms that are amenable to genetic manipulations (Chan et al., 2009; Giaever et al., 1999; Ho et al., 2009). However, many drugs active in humans are inactive in these model organisms, possibly due to multi-drug resistance mechanisms and target divergence (Walker, 1982). The second category includes affinity-based methods that are used to identify proteins that bind the drug (Ong et al., 2009; Rix and Superti-Furga, 2009). These approaches are generally effective when the drug is potent and the targets are reasonably abundant *in vivo*. However, establishing that a drug-binding protein is the relevant physiological target depends on correlations between chemical inhibition *in vitro* and protein knockdown phenotypes. These correlations often prove to be misleading for several reasons, including differences between activity inhibition, which can be acute, and the phenotypes associated with loss of a protein, which can be indirect or result from cumulative effects (Weiss et al., 2007). Currently, we lack an approach that addresses these

limitations and can be directly used to discover mechanisms of drug action in human cells.

The gold standard in identifying a drug's target is achieved when two criteria are met. First, it must be shown that resistance to a drug in a physiological context occurs through mutations in the candidate target protein. Second, in the case of direct targets, these mutations should suppress inhibition of the target's activity by the drug. This standard is met for a few drugs, such as Gleevec, for which focused analyses of the expected target led to the identification of these mutations (Azam et al., 2003; Gorre et al., 2001). However, the large size and complexity of the human genome has limited the unbiased analyses of genetic mechanisms conferring drug resistance. Thus, general approaches to find drug targets through examining drug resistance in human cells are not currently available.

My colleagues and I sought to develop an approach that could achieve gold standard validation of a drug's target and could be applied to relevant human cell types (e.g. cancer cells). To accomplish this, next-generation sequencing and bioinformatics analyses were used to identify mutations in drug resistant clones isolated from cultured human cell lines. While several differences in the sequences of individual drug-resistant clones are likely, if mutations occurred at high frequency in the drug's direct target, then analyzing mutations common to multiple clones might focus the analysis on a handful of potential drug targets. When combined with additional biochemical analyses, these data could

lead to the gold standard validation of the drug's physiological target. As a proof of concept I analyzed the mechanisms that confer resistance to two drugs, BI 2536 and bortezomib (**Figure 2.1**), whose targets are known.

Proof-of-concept studies with BI 2536

BI 2536 is a dihydropteridinone that is currently in clinical trials and inhibits Polo-like kinase 1 (Plk1), a major cell cycle regulator (Lenart et al., 2007). To isolate drug-resistant clones I used a human colon cancer cell line, HCT-116, which is deficient in mismatch repair and has low expression of multi-drug resistant (MDR) pumps (Teraishi et al., 2005). Therefore, HCT-116 behaves like a mutagenized cell line, facilitating the rapid identification of mutations that confer drug resistance (Girdler et al., 2008; Glaab and Tindall, 1997). From fifteen BI 2536-resistant clones (clones were isolated from selections with 10 nM BI 2536, LD_{50}: 3.9 ± 2.8 nM), I randomly selected six clones for transcriptome sequencing (also called RNA-seq). The LD_{50}s for BI 2536 were 3-9 fold higher in these clones than in the parental cell line (**Figure 2.2**). In parallel, the parental cell line was also processed for transcriptome sequencing.

The sequencing data were analyzed by identifying single nucleotide variations and short insertions/deletions (indels) in each clone and only those present in coding sequences were further considered. 6-14 single nucleotide variations were significantly increased (with a 0.5% false discovery rate) in the BI 2536-resistant clones, compared to the parental cell population (**Appendix 1,**

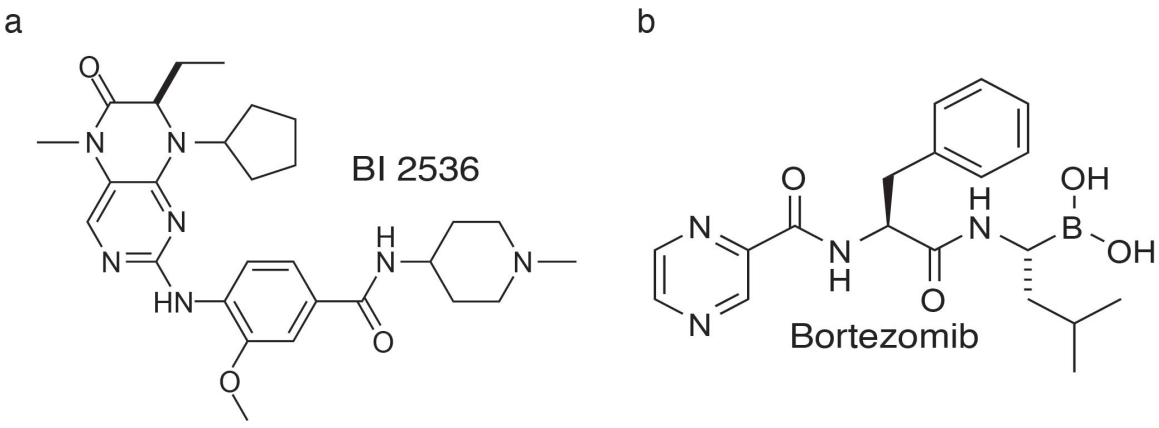
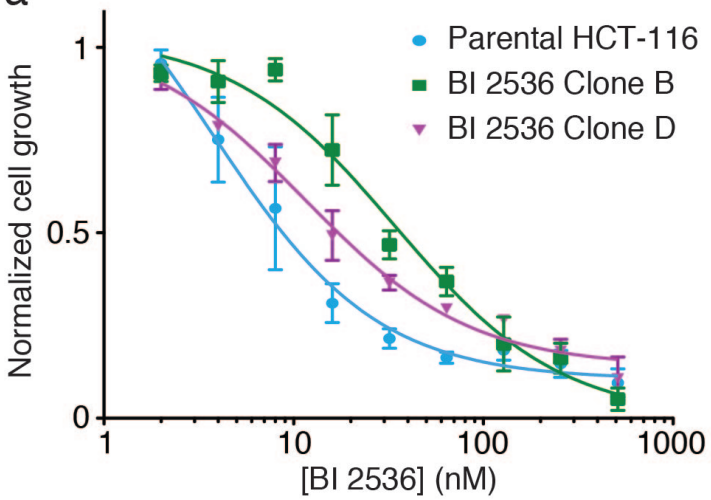


Figure 2.1 Structures of cytotoxic inhibitors BI 2536 (a) and bortezomib (b).

a



b

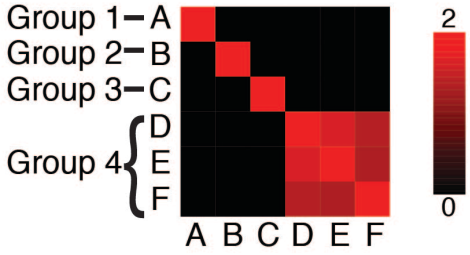
Clone	LD_{50} (nM)
Parental	3.9 ± 2.8
A	34.9 ± 13.0
B	33.5 ± 2.6
C	27.2 ± 8.6
D	14.2 ± 7.8
E	21.1 ± 5.7
F	27.5 ± 13.3

Figure 2.2 BI 2536 resistant clones were obtained through selection with drug. (a) Median lethal dose values (LD_{50}s) were obtained by measuring the number of metabolically active cells after treatment with a range of concentrations of BI 2536. **(b)** BI 2536 LD_{50}s calculated for the parental HCT-116 cells and the six BI 2536-resistant clones processed for transcriptome sequencing (mean ± sem, n = 3).

Tables A1-A6; no indels were found). Similar clones were then identified and merged into groups using a clustering approach. This clustering analysis determined how similar two clones were based on the number of single nucleotide variations they had in common. Among the six clones, A, B, and C were found to be independent (referred to as groups 1, 2, and 3, respectively), while D, E, and F formed a single group (group 4, **Figure 2.3a**). I further focused my analysis on genes that were mutated in more than one independent BI 2536-resistant group. While there were no genes common to all groups, PLK1 was the only gene mutated in more than one group (**Figure 2.3b**).

Two distinct mutations in the Plk1 protein were identified, G63S and R136G, both of which map to the binding site of BI 2536 in the crystal structure of Plk1 (**Figure 2.3b, c**) (Kothe et al., 2007). To examine whether these mutations are sufficient for conferring resistance to BI 2536, I generated stable cell lines expressing GFP-tagged full-length Plk1 wild type, G63S, or R136G. Independent cell lines were used as secondary mutations were less likely to be common between these lines and the parental HCT-116 cells. The R136G mutation suppressed BI 2536 toxicity in hTERT-RPE1 cells (**Figure 2.4a**), and the G63S mutation suppressed toxicity both in hTERT-RPE1 (**Figure 2.4a**) and in HeLa cells (**Figure 2.4b**). These data suggest that Plk1 is the major physiological target of BI 2536. Published data show that the R136G mutation has reduced sensitivity to BI 2536 *in vitro* (Scutt et al., 2009). To determine whether the G63S mutation also reduces sensitivity to BI 2536 *in vitro*, I expressed full-length Plk1

a



b

Gene	Amino Acid Mutation	BI 2536 Group 1	2	3	4
PLK1	R136G	X	X		
PLK1	G63S			X	

c

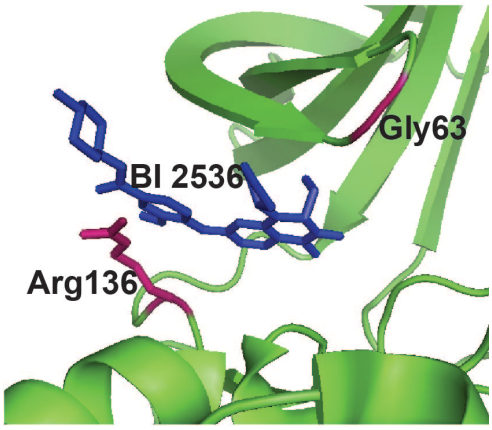


Figure 2.3 Bioinformatics analysis of BI 2536-resistant clones. (a) Graph-based analysis of similarities (0 = low similarity, 2 = high similarity) between BI 2536-resistant clones allows clones D, E, and F to be grouped together (group 4). **(b)** Genes and the coding mutations identified in more than one independent BI 2536-resistant group. **(c)** Amino acids Arg136 and Gly63 in Plk1, both mutated in BI 2536-resistant clones, are proximal to the BI 2536 binding site (PDB: 2RKU) (Kothe et al., 2007).

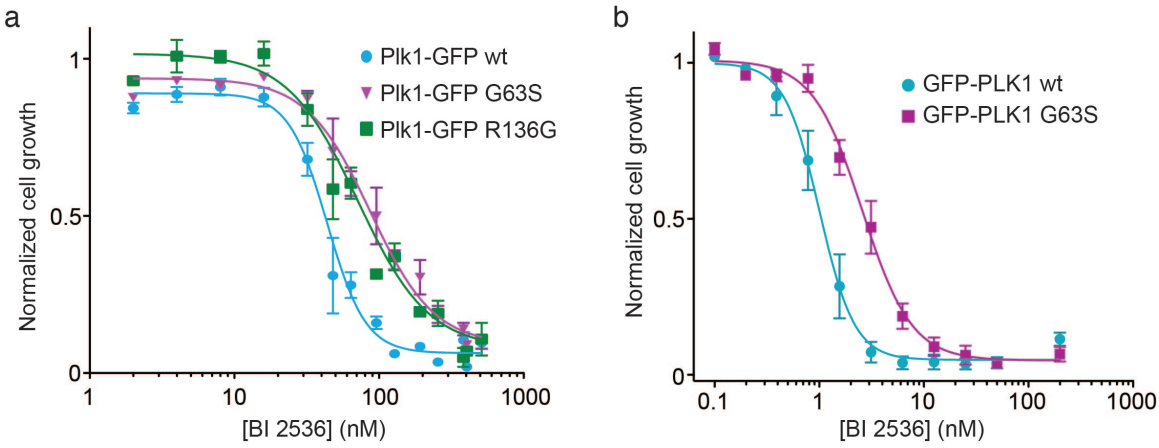


Figure 2.4 The effects of BI 2536 exposure on human cells expressing Plk1 wild type and mutant proteins. (a) Proliferation assay showing the effects of BI 2536 exposure on hTERT-RPE1 cells stably expressing Plk1-GFP wild type (wt), Plk1-GFP G63S, or Plk1-GFP R136G. LD_{50}s: 44 ± 5 nM (GFP-Plk1 wt), 83 ± 9 nM (GFP-Plk1 G63S), 76 ± 8 nM (GFP-Plk1 R136G); n = 6, mean ± sem, $p < 0.01$ for both, two-tailed paired t-test. **(b)** Proliferation assay showing the effects of BI 2536 exposure on HeLa cells stably expressing Plk1-GFP wt and Plk1-GFP G63S. LD_{50}s: 1.0 ± 0.2 nM (GFP-Plk1 wt) and 2.7 ± 0.3 nM (GFP-Plk1 G63S); n = 4, mean ± sem. $p < 0.05$, two-tailed paired t-test.

wild type and G63S proteins in insect cells and used these proteins in kinase assays (**Figure 2.5**). While the G63S protein was less active than wild type Plk1 (~5-fold), it also was substantially resistant to BI 2536 (**Figure 2.5b**). Therefore, these data provide the gold standard proof that Plk1 is the target of BI 2536.

BI 2536-resistant group 4 (representing clones D, E, and F) did not have a mutation in the PLK1 gene, suggesting a different resistance mechanism. Group 4 clones suppress not only cell growth, but also the characteristic cell division phenotype associated with the loss of Plk1 activity. In particular, the reduction in bipolar spindles, associated with BI 2536 treatment (**Figure 2.6a**), is suppressed in these clones (**Figure 2.6b**). This raised the possibility that increased drug efflux is a potential resistance mechanism in these clones. An unbiased survey of all transcript levels (see **Appendix 1**), which can be determined from the RNAseq data, revealed that ABCB1 (P-glycoprotein, a drug efflux transporter) was among the most highly over-expressed mRNAs in group 4 clones (**Figure 2.6c**). Consistent with the hypothesis that the intracellular concentration of BI 2536 was lowered by drug efflux, group 4 clones had reduced sensitivity to the chemically-unrelated drug taxol, a compound known to be transported by the ABCB1 pump (Yusuf et al., 2003) (**Figure 2.6d**). While further experiments are needed to establish if BI 2536 resistance in group 4 clones is via ABCB1 drug efflux, my data suggest that this approach can lead to testable hypotheses for indirect mechanisms of drug resistance.

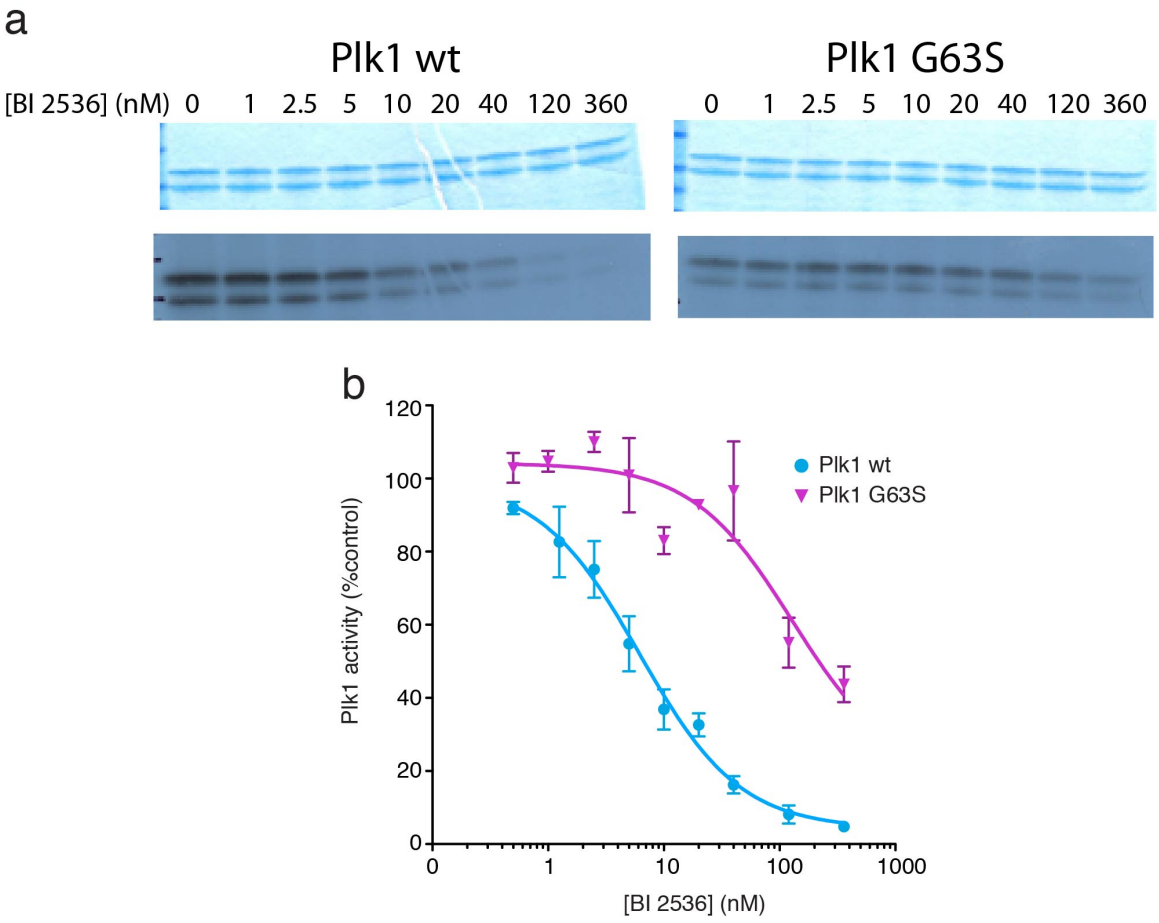


Figure 2.5 Kinase assay showing response of Plk1 wild type and Plk1 G63S to BI 2536 *in vitro*. (a) Recombinant full-length Plk1 wild type (wt) and Plk1 G63S can phosphorylate casein (which runs as two bands on a 4-20% gel), as analyzed be transfer of γ-^{32}P. BI 2536 inhibits this activity in a dose dependent manner for both proteins. **(b)** Quantification of phosphorylated casein bands were quantified, with activity reported as a percentage of the control (5% DMSO). IC_{50}s: 6.2 ± 1.3 nM (Plk1 wt), 128 ± 3 nM (Plk1 G63S); n = 3, mean ± sem.

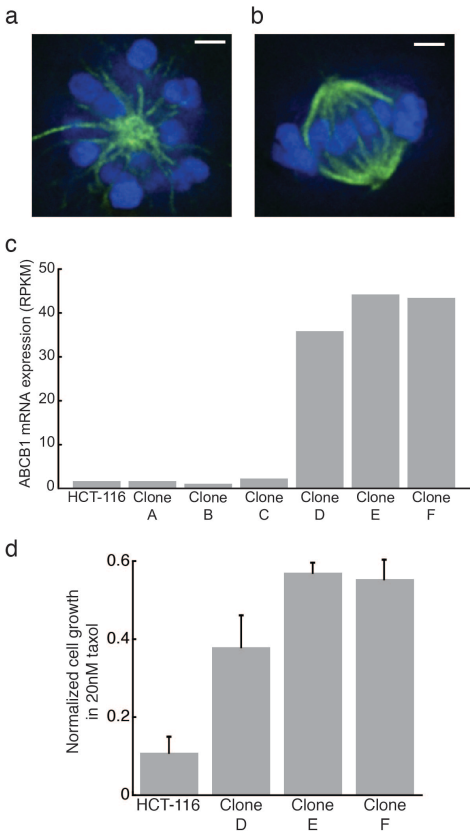


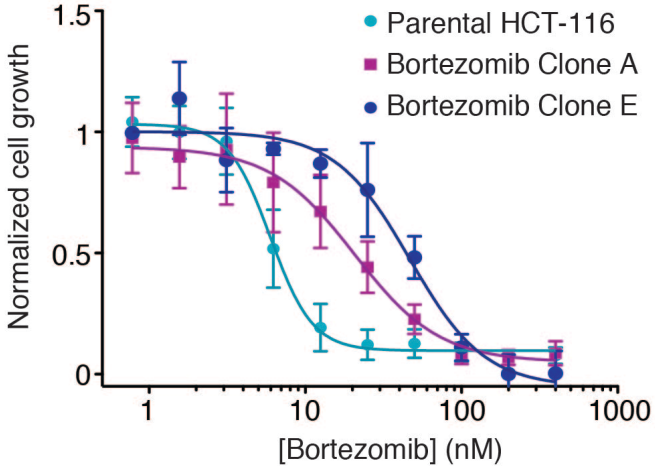
Figure 2.6 Analysis of the mechanism of resistance of group 4 clones. (a-b) Analysis of microtubule (green) and chromosome (blue) organization in dividing cells treated with BI 2536 (50 nM). Normal bipolar spindles were observed in only 11 ± 3% of HCT-116 cells. A typical monopolar spindle observed in these cells is shown **(a)**. In comparison, 46 ± 4% of the spindles in the drug-resistant clone E were bipolar, as shown **(b)**. Scale bar, 2 μm. **(c)** Expression levels of the ABCB1 mRNA in the parental HCT-116 cells and the six BI 2536-resistant clones. Levels are measured as the number of reads per kilobase, per million reads (RPKM). **(d)** Proliferation assay showing the effect of 20 nM taxol on HCT-116 parental cells and clones D, E, and F, normalized to untreated cells (n = 3, mean ± s.d.).

Proof-of-concept studies with bortezomib

I next examined if this approach could be applied to another drug. Bortezomib (**Figure 2.1b**) is a boronic acid dipeptide that inhibits the proteasome through its subunit PSMB5 and is used clinically to treat multiple myeloma and mantle cell lymphoma (Chen et al., 2011). Nineteen clones were isolated from HCT-116 cells grown in the presence of bortezomib (8 – 12 nM, LD_{50}: 6.3 ± 0.9 nM). Five clones, with reduced bortezomib sensitivity (LD_{50}s 2.4 – 6.5 fold higher, **Figure 2.7**), were processed for transcriptome sequencing. 15 – 28 single nucleotide variations were identified from each clone (**Appendix 1, Tables A7-A11**). Using clustering analysis (**Figure 2.8a**) clones A, B, C, and D were grouped together (group 1) and clone E was found to be independent (group 2). Five genes were mutated in both bortezomib-resistant groups (**Figure 2.8b**). The only gene with two distinct mutations (M104V and A108T) was PSMB5, the known target of bortezomib. The existence of two distinct resistance mutations would make PSMB5 the highest priority gene for further analysis if the target of bortezomib was not already known. However, the remaining four genes would also need to be examined as potential targets.

I found that expression of GFP-tagged PSMB5, carrying either the M104V or A108T mutations, suppressed bortezomib sensitivity in an independent cell line (hTERT-RPE1, **Figure 2.9a**). This is consistent with previous reports that A108T confers bortezomib-resistance (Lu et al., 2008; Oerlemans et al., 2008). As both of the mutations identified in PSMB5 map to the drug's binding site, I

a

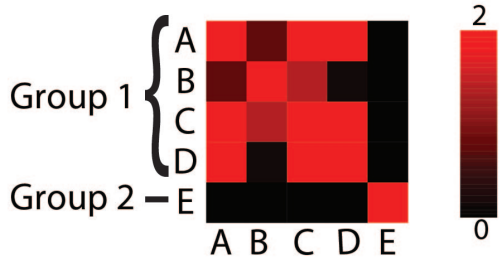


b

Clone	LD_{50} (nM)
Parental	6.3 ± 0.9
A	19.4 ± 6.7
B	15.3 ± 6.2
C	17.8 ± 6.0
D	17.2 ± 5.4
E	41.5 ± 9.5

Figure 2.7 Bortezomib-resistant clones were obtained through selection with drug. (a) Median lethal dose (LD_{50}s) were obtained by measuring the number of metabolically active cells after treatment with a range of concentrations of bortezomib. **(b)** Bortezomib LD_{50}s calculated for the parental HCT-116 cells and the five bortezomib-resistant clones processed for transcriptome sequencing (mean ± sem, n = 3).

a



b

Gene	Amino Acid Mutation	Bortezomib Group 1	Bortezomib Group 2
PSMB5	A108T	X	
PSMB5	M104V		X
PPP1R10	R58H	X	X
CDK5RAP3	V21G	X	X
ETHE1	S19A	X	X
ID1	S8R	X	X

Figure 2.8 Bioinformatics analysis of bortezomib-resistant clones. (a) Graph-based analysis of similarities (0 = low similarity, 2 = high similarity) between bortezomib-resistant clones allows clones A, B, C, and D to be grouped together (group 1). **(b)** Genes and the coding mutations identified in more than one independent bortezomib-resistant group.

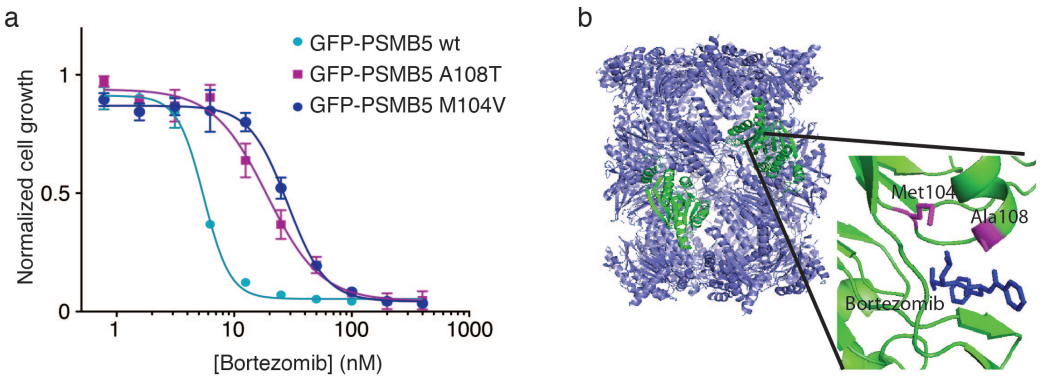


Figure 2.9 Characterization of PSMB5 mutations found in bortezomib-resistant clones. (a) Proliferation assay showing the effect of treatment with bortezomib on hTERT-RPE1 cells that are stably expressing GFP-PSMB5 wild type (wt), GFP-PSMB5 M104V, or GFP-PSMB5 A108T, in addition to the endogenous wild type gene. LD_{50}s: 5.55 ± 0.09 nM (GFP-PSMB5 wt), 28.7 ± 2.8 nM (GFP-PSMB5 M104V), 19.2 ± 2.8 nM (GFP-PSMB5 A108T); n = 3, mean ± sem, both $p < 0.05$, two-tailed paired t-test. **(b)** Amino acids Met104 and Ala108 in Pre2 (yeast homolog of PSMB5), a subunit of the 20S proteasome, are proximal to the bortezomib binding site (PDB: 2F16) (Groll et al., 2006).

hypothesize that they directly suppress drug interactions (**Figure 2.9b**) (Groll et al., 2006). While additional biochemical tests are needed to examine this further, my data indicate that the method described here can efficiently lead to resistance mechanisms that include a drug's direct target.

Directed analysis of frequency of mutations in target

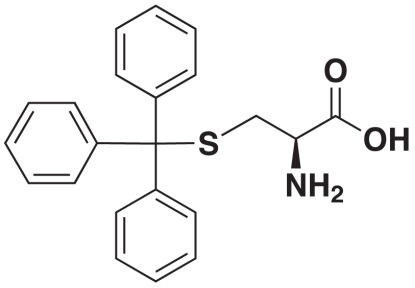
My analysis thus far indicates that for this approach to be applicable, resistance via mutations in a drug's direct target must occur at high frequency in drug-resistant clones. To examine how frequently this occurs, I examined the sequence of the PLK1 gene in each of the nine BI 2536-resistant clones that had not already been processed by RNA-seq. The frequency at which the PLK1 gene was mutated was ~45% (4 of 9 clones, **Table 2.1**).

I next analyzed two kinesin-5 inhibitors, S-Trityl-L-cysteine (STLC), which is known to be selective (Skoufias et al., 2006), and 4-(2-(1-phenylcyclopropyl) thiazol-4-yl)pyridine (PCTP), which has been shown to inhibit other related motor proteins *in vitro* (Rickert et al., 2008). I obtained resistant clones for these inhibitors (1 μM STLC in selections, LD_{50}: 0.63 ± .06 μM; 8 – 12 μM PCTP in selections, LD_{50}: 2.2 ± 0.1 μM). Kinesin-5 mutations were found in ~30% of the STLC-resistant clones (4 of 14 clones, **Figure 2.10**) and in ~15% of PCTP resistant clones (3 of 22 clones, **Figure 2.11**). These data indicate that resistance in a drug's direct target occurs at high frequency when a drug has one

Table 2.1: RT-PCR and Sanger sequencing of the PLK1 gene in nine BI 2536-resistant clones.

Clone	PLK1 Mutation
1	None
2	None
3	R136G
4	R136G
5	R136G
6	None
7	None
8	None
9	R136G

a

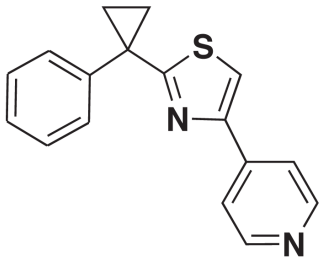


b

Clone	Kinesin-5 Mutation
1	E123K
2	None
3	None
4	None
5	None
6	None
7	A334V
8	None
9	H354R
10	None
11	None
12	None
13	A103V
14	None

Figure 2.10 Characterization of STLC-resistant clones. (**a**) Chemical structure of the kinesin-5 inhibitor S-trityl-(L)-cysteine (STLC). (**b**) Fourteen clones were isolated from selections with 1 μM STLC (LD_{50}: 0.63 ± 0.06 μM). RT-PCR and Sanger sequencing identified mutations in the kinesin-5 RNA of 4/14 clones.

a



b

Clone	Kinesin-5 Mutation
1	None
2	None
3	None
4	None
5	None
6	None
7	None
8	None
9	None
10	S269N
11	S269N
12	Y104H
13	None
14	None
15	None
16	None
17	None
18	None
19	None
20	None
21	None
22	None

Figure 2.11 Characterization of PCTP-resistant clones. (**a**) Chemical structure of the kinesin-5 inhibitor 4-(2-(1-phenylcyclopropyl)thiazol-4-yl)pyridine (PCTP). (**b**) Twenty-two clones were isolated from selections with 8 – 12 μM PCTP (LD_{50}: 2.2 ± 0.1 μM). RT-PCR and Sanger sequencing identified mutations in the kinesin-5 RNA of 3/22 clones.

major physiological target (as is the case for STLC, BI 2536 and bortezomib). When a drug has multiple targets (for example, PCTP) it is likely that resistance in a single target will be less frequent and thus this approach may be limited as a greater number of clones have to be sequenced in order to identify genes with mutations in more than one clone.

Selections with additional compounds

While selecting for resistant clones to BI 2536, bortezomib, STLC, and PCTP I also attempted to select for resistance to additional compounds. These compounds include two molecules with known targets, the microtubule inhibitor taxol (Horwitz, 1994) and the Aurora kinase inhibitor hesperadin (Hauf et al., 2003) (**Figure 2.12a** and **b**). I examined these compounds at the same time as bortezomib with the aim of characterizing the generality of my approach. Additionally, I selected for resistant clones with compounds that I hoped would allow me to use my method to identify novel targets. These compounds include ARQ-197 (**Figure 2.12c**), which is published as a c-Met inhibitor (Munshi et al., 2010), but may have additional targets and two cytotoxic compounds without clearly defined targets, BE-54017 and Cladoniamide A (**Figure 2.12d, e**). Each of these drugs failed at some point during the resistance selection process and likely did so for a variety of reasons. My results with these inhibitors suggest some limitations of my method.

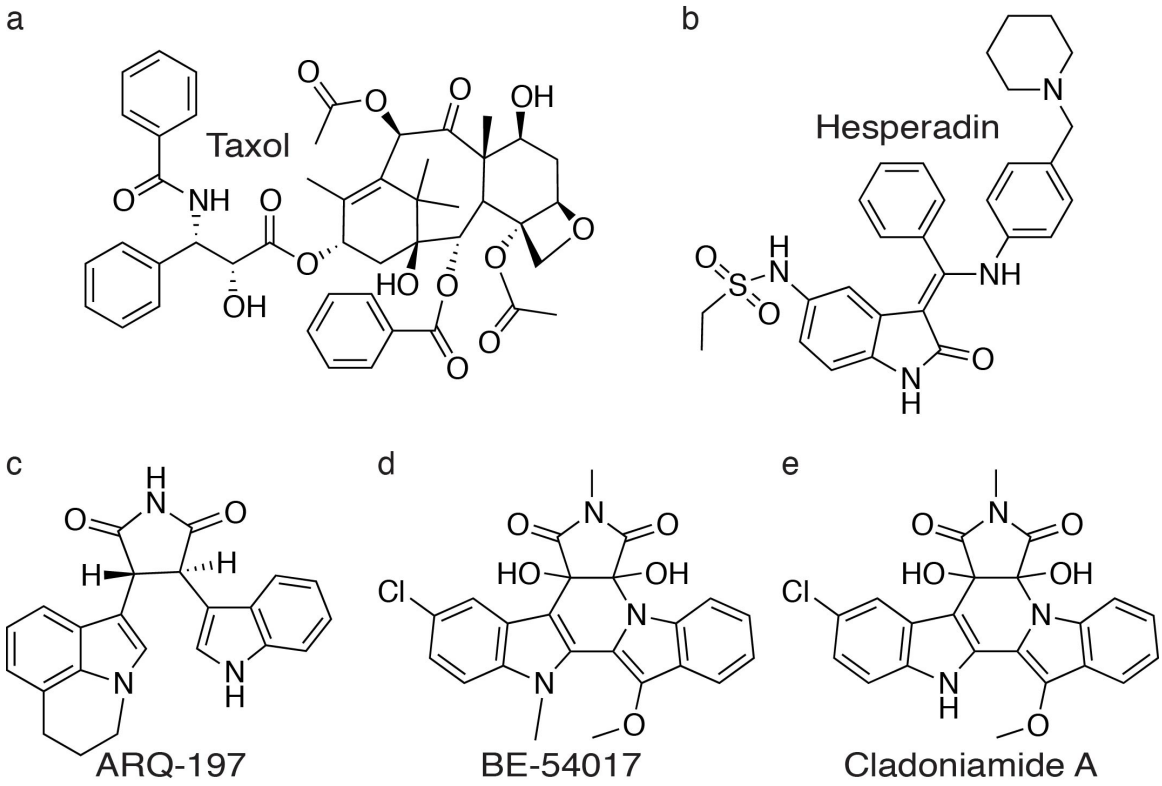


Figure 2.12 Structures of small molecules taxol (a), hesperadin (b), ARQ-197 (c), BE-54017 (d), and Cladoniamide A (e).

Selections with the Aurora inhibitor hesperadin were conducted at compound concentrations of 0.1 – 3 μM. At 3 μM all cells died and the plates were completely clear; this is a common feature of especially high drug concentrations and was also observed during selections with bortezomib and STLC. At concentrations between 400 nM and 2 μM, a few cells remained on each selection plate, with more cells remaining on plates with lower concentrations of hesperadin. These cells did not divide and no colonies formed. At very low concentrations of hesperadin (100 – 300 nM) much of the plate was covered with cells and it was impossible to pick individual colonies. The observations with the high (3 μM) and low (100 – 300 nM) concentrations of hesperadin are common and not surprising. However, I have not previously observed plates that contain only non-dividing single cells, as were present in the mid-concentration hesperadin selections. The presence of these individual cells that precluded growth of resistance colonies is surprising as similar selections have been successful with another Aurora kinase inhibitor, ZM447439 (Girdler et al., 2008). This difference between the two Aurora kinase inhibitors may be due to variations in the mechanism of action of the two drugs. These findings suggest that for drugs that elicit certain cell states (for example, senescence, not cell death) my approach may be unsuccessful. Fortunately, this should be able to be determined at an early stage of the analysis (during resistance selection).

Selections with taxol (2.5 – 7.5 nM) resulted in the isolation of twenty-eight taxol-resistant clones. Before further analysis, I randomly selected clones for

testing and found they were 2 – 20-fold less sensitive to taxol than the parental cells (**Figure 2.13a**). Interestingly, all of the clones tested were also, to a similar extent, less sensitive to BI 2536 (**Figure 2.13b**). This suggests that these clones have multi-drug resistance. This finding is not surprising as the most common clinical resistance mechanism for taxol is also multi-drug resistance, specifically up-regulation of the drug pump ABCB1 (Horwitz et al., 1993). As all clones tested displayed this same multi-drug resistance phenotype, I hypothesize that some genes, such as those that encode highly conserved and essential proteins like tubulin, are not amenable to spontaneous mutation, as required for my approach. For these target proteins, my method will not likely be effective.

Selections with ARQ-197 (750 – 850 nM) resulted in thirty-six clones. Upon further testing, all of the tested clones had LD_{50} values that were 1.5-fold or less the value of the parental LD_{50} (**Figure 2.14**). Furthermore, as there was such a small change in drug sensitivity over the parental cells, it was unclear whether these clones had altered sensitivity to taxol (a 1.5-fold change in taxol's LD_{50} could not be robustly identified). ARQ-197, which is published as a c-Met inhibitor (Munshi et al., 2010), elicits cellular signaling phenotypes that are not fully consistent with c-Met inhibition (personal communication, Pasi A. Jänne). Furthermore, the lack of strongly resistant ARQ-197 clones suggests that the compound may have multiple targets. To further characterize the potential target(s) of ARQ-197 I examined its phenotype in HCT-116 cells by microscopy. I found that these cells displayed significant reductions in microtubule levels, a

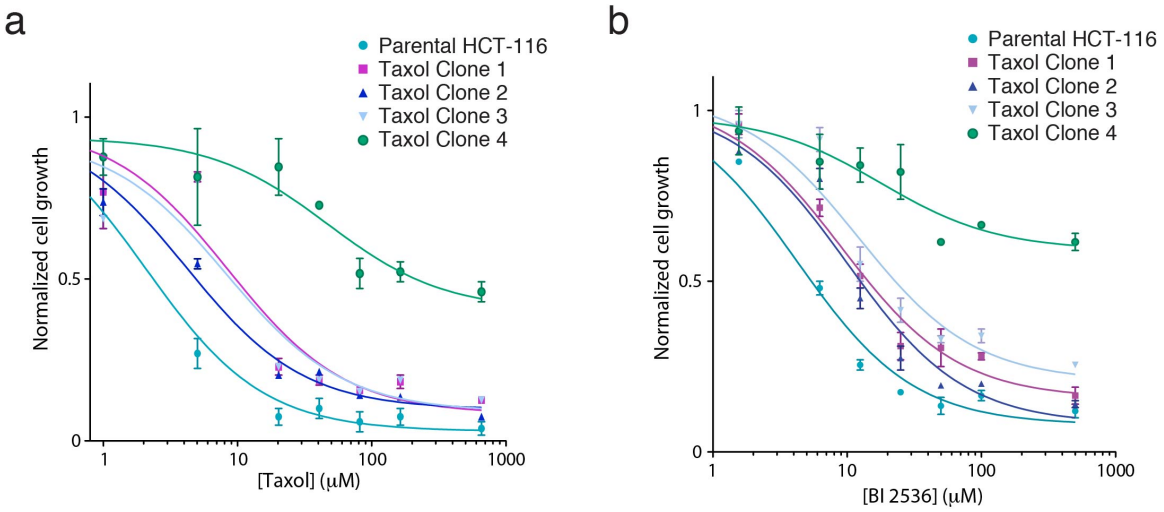


Figure 2.13 Taxol-resistant clones were obtained through selection with drug. Median lethal dose values (LD_{50}s) were obtained by measuring the number of metabolically active cells after treatment with a range of concentrations of taxol **(a)** or BI 2536 **(b)** (mean ± sem, n = 2).

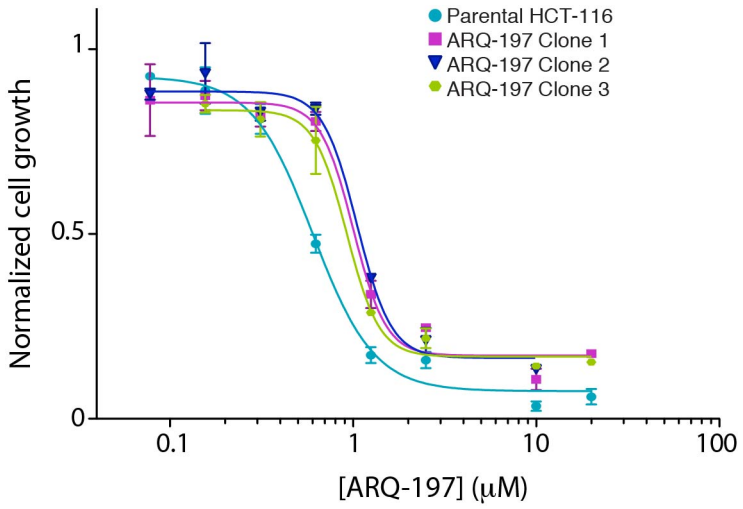


Figure 2.14 ARQ-197-resistant clones were obtained through selection with drug. Median lethal dose values (LD_{50}s) were obtained by measuring the number of metabolically active cells after treatment with a range of concentrations of ARQ-197 (mean ± sem, $n \geq 2$).

phenotype that closely mimics cell treatment with the microtubule poison nocodazole (**Figure 2.15**). While I have not been able to show an effect of ARQ-197 on microtubules *in vitro*, my cellular data support the idea that ARQ-197 has a target other than c-Met.

The compounds BE-54017 and Cladoniamide A were brought to my attention by Fang-Yuan Chang in Sean Brady's laboratory at Rockefeller University, who is interested in their biosynthesis. As the mechanisms of action of these cytotoxic compounds are unknown, I hoped that my method might help to elucidate their targets. Selections with BE-54017 (200 – 400 ng/ml) and Cladoniamide A (150 – 200 ng/ml) allowed for several colonies to form. However, the obtained clones were not able to survive multiple passages in the appropriate concentration of drug. As no clones remained, these compounds were not further pursued. Together my data suggest that this approach should be successful for drugs that have single easily-mutated targets, but that my method will be less successful when a drug has multiple targets. When the method will not be successful, it is likely that there will be an indication of this during an early stage (i.e. before transcriptome sequencing).

Discussion

Along with my collaborators, I have developed a method to identify the targets of drugs in human cells by examining resistance mechanisms. This strategy has the following key steps (**Figure 2.16**). First, genetically

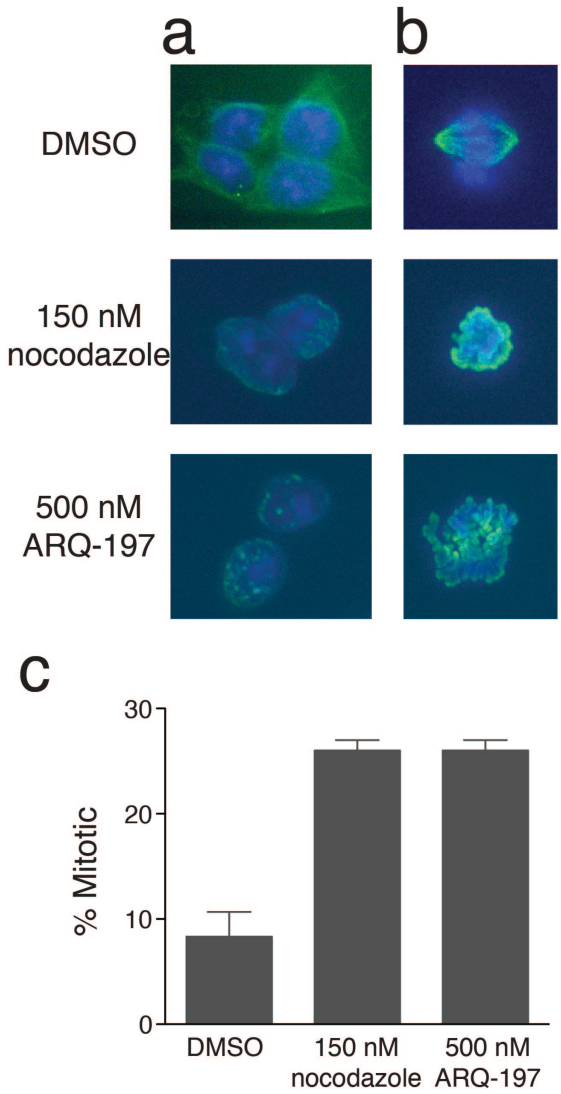


Figure 2.15 Effect of ARQ-197 on HCT-116 cells. Analysis of microtubule (green) and chromosome (blue) organization in interphase **(a)** and dividing cells **(b)** treated with DMSO, nocodazole (150nM), or ARQ-197 (500 nM) for four hours. **(c)** The percentage of cells that are mitotic after each treatment is plotted (mean ± sem, $n \geq 2$).

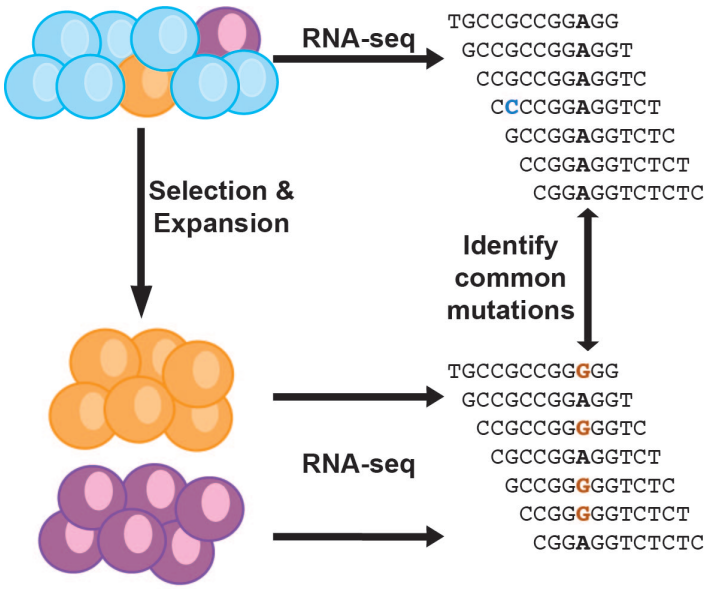


Figure 2.16 Schematic highlighting the key steps of the approach. Selecting and expanding drug-resistant clones from a heterogeneous parental population. Massively parallel sequencing of mRNA from multiple drug-resistant clones and parental (untreated) cells. Bioinformatics analyses to identify genes that are mutated with high frequency. A subset of sequencing reads for a BI 2536-resistant clone (clone A) and the HCT-116 parental cells are shown. The ~50% mutation frequency indicates heterozygosity.

heterogeneous human cells are treated with a drug at concentrations that kill most cells. Then, multiple drug resistant clones are selected and expanded. Based on my analysis with BI 2536, clones with multi-drug resistance can be excluded by testing for reduced sensitivity to unrelated compounds. The remaining clones are processed for transcriptome sequencing, along with the parental (untreated) cell population. Mutations in coding sequences of the resistant clones, not observed in the parental cell population, are identified. Mutations that arise in more than one independent clone suggest resistance mechanisms. Biochemical and cell biological assays can then be used to examine which of these resistance mechanisms represent the drug's direct target and which are indirect resistance mechanisms.

My strategy has some advantages over other approaches to identify a drug's target. First, unlike many other target identification approaches, this method does not rely on chemical modifications of the drug of interest. This can be important when small changes in a drug's chemical composition can alter its mechanism of action (Wood et al., 2010). Second, my approach can be applied to any cells that can be grown in culture, enabling cell-type specific analyses, which would be particularly useful if a drug is toxic in specific tissues. While this approach is currently limited to analyzing cytotoxic drugs, in principle it should be possible to extend its use to non-toxic drugs. Furthermore, while my approach focuses only on discovering single nucleotide variations and insertions/deletions and thus does not report on all potential mechanisms of resistance, these

limitations can be addressed by combining the approach with other genomic methods.

In summary, this combined experimental and computational approach has the potential to reveal the physiological on-targets of a drug in disease cells, unintended off-targets in healthy cells, and can reveal cellular mechanisms of drug resistance. These findings can impact chemical modifications of drugs to improve efficacy and limit toxicity. When unanticipated drug targets are found new uses of the drugs may also be suggested.

Chapter 3: Examining the mechanism of action of a kinesin inhibitor using Stable Isotope Labeled Inhibitors for Crosslinking (SILIC)

Summary

Once a small molecule's target is identified, a remaining challenge is characterizing the chemical inhibitor's binding site within its protein target. This is especially difficult if the target—inhibitor interaction requires multi-protein mixtures and high-resolution structural studies are not straightforward. Building upon previous research involving photo-crosslinking and incorporation of mixtures of stable isotopes, I have helped to develop a method, **S**table **I**sotope **L**abeled **I**nhibitors for **C**rosslinking (SILIC), for mapping a small molecule inhibitor's binding site in its target protein. In SILIC, structure-activity relationship data is used to design inhibitor analogs that incorporate a photo-crosslinking group along with either natural or heavy isotopes. An equimolar mixture of these inhibitor analogs is crosslinked to the target protein to yield a robust signature for identifying inhibitor-modified peptide fragments in complex mass spectrometry data. As a proof of concept, I applied this approach to an ATP-competitive inhibitor (called 4-(2-(1-phenylcyclopropyl)thiazol-4-yl)pyridine (PCTP)) of kinesin-5, a widely conserved motor protein required for cell division and an anti-

cancer drug target. This crosslinking analysis, along with mutagenesis studies, suggests that the inhibitor binds at an allosteric site in the motor protein.

Introduction

Photo-crosslinking of small molecules to proteins has been used to trap drug—target protein interactions in complex protein mixtures (Chen et al., 2002; Seiffert et al., 2000; Shorr et al., 1982). Identifying drug targets and mapping drug binding sites after photo-crosslinking typically relies on systematic mass spectrometry based analyses of digested protein fragments to identify those with a small molecule adduct (Al-Mawsawi et al., 2006; Wood et al., 2010). While there are examples of the successful use of this approach, the general applicability of the method has been limited as crosslinking is often sub-stoichiometric (Hermanson, 2008), and the different possible inhibitor—peptide adducts can be difficult to detect in complex mass spectra. One strategy to address this involves generating inhibitor analogs with an affinity tag for capturing the inhibitor—peptide adducts (Salisbury and Cravatt, 2007). In many cases, however, the inhibitor's dual modifications, for photo-crosslinking and affinity-capture, can alter the compound's mechanism of action.

As an alternative approach to identify inhibitor—protein adducts within complex mass spectra, inhibitor analogs can be generated such that they carry a unique isotope pattern (Kelleher et al., 1997; Weerapana et al., 2010). The incorporation of natural and heavy stable isotopes into a benzophenone photo-

crosslinker moiety appended to the inhibitor of interest has been shown to aid the identification of its target in a proof-of-concept study (Lamos et al., 2006). However, the method is not likely to be useful for mapping an inhibitor's binding site. This is, in large part, due to the crosslinking group being incorporated via a linker, so that it is a significant distance from the functional groups that are likely to make key contacts with the target's binding site.

Building on these studies, my colleagues and I have developed a method, which we call **S**table **I**sotope **L**abeled **I**nhibitors for **C**rosslinking (SILIC), for mapping small molecule—protein binding sites. In the first step of this approach a photo-crosslinking group is incorporated into the inhibitor of interest (**Figure 3.1**), guided by available structure-activity relationship (SAR) data. Importantly, the photo-crosslinking group is appended at a site that does not change the inhibitor's mechanism of action, but is in the closest proximity possible to the inhibitor's activity-conferring functionality, so as to increase the probability that crosslinks are at, or near, the protein's inhibitor-binding pocket. Natural and heavy isotope containing inhibitor analogs, which have a mass difference of a few daltons but otherwise identical physical properties, are then generated. The multi-protein complex to be analyzed is then incubated with a 1:1 mixture of natural and heavy inhibitor. After photo-crosslinking and protein digestion, the resulting mixture of peptide fragments is separated by HPLC and analyzed using high-resolution mass spectrometry. The resulting mass spectra can be comprised of thousands of peaks and the peptide-inhibitor adduct is likely to be

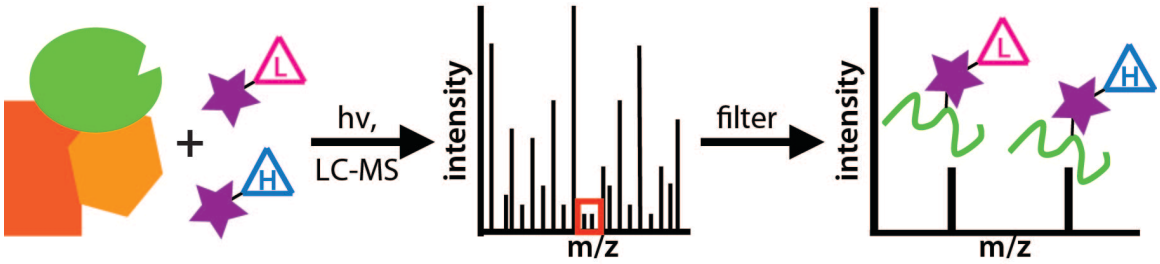


Figure 3.1 Schematic for SILIC. Inhibitors, with a photo-crosslinking substituent and natural (N) or heavy (H) isotopes, are mixed with a complex of proteins, including the target (green). After UV-crosslinking, protein digestion, and LC-MS one obtains complex mass spectra that can be filtered based upon a 'signature' of peaks with the expected mass difference and equal signal intensities.

of low abundance due to sub-stoichiometric labeling. The peptide-inhibitor adduct is identified when a pair of peptides that co-elute in the LC have the expected mass difference and essentially equal signal intensity. Finally, guided by these data, site-directed mutagenesis experiments can be designed to further examine the inhibitor-binding sites identified by SILIC.

Mechanism of action of PCTP

As a proof of concept, I examined the small molecule (4-(2-(1-phenylcyclopropyl)thiazol-4-yl)pyridine (PCTP, **Figure 2.11a)**, an inhibitor of kinesin-5 (Rickert et al., 2008). Kinesins, which comprise a family of over forty proteins, are motor proteins that move cargo along microtubules, polymers of the cytoskeletal protein tubulin (Miki et al., 2005; Vale and Milligan, 2000). The kinesin-5 family is required for the assembly of the microtubule-based apparatus necessary for cell division, which is called the spindle (Sharp et al., 2000). Inhibitors of kinesin-5 have provided valuable insight into mechanisms of cell division and have entered clinical trials as anti-cancer drugs (Duhl and Renhowe, 2005; Mayer et al., 1999). Kinesin-5 inhibitors that are in clinical trials, and have been used for cytological experiments, bind an allosteric site not conserved in other kinesins (Maliga et al., 2002; Yan et al., 2004). These inhibitors are not competitive with respect to ATP (Cochran and Gilbert, 2005). Recently, ATP-competitive inhibitors of kinesin-5 have been reported, including PCTP (Luo et al., 2007; Rickert et al., 2008). Of these, PCTP stood out as it has been shown to

inhibit other kinesin proteins (Rickert et al., 2008). Since the ATP-binding site is the most conserved feature in kinesins, I was interested in using PCTP as a possible starting point for developing new inhibitors of other kinesins. However, the binding site of PCTP in kinesin-5 or any of the kinesins it inhibits is not known. Both other research groups and I have unsuccessfully attempted to obtain a crystal structure of the PCTP-kinesin-5 complex (Rickert et al., 2008) (see **Appendix 2**).

To map the binding site of PCTP, I first analyzed its mechanism of action. Interestingly, PCTP inhibits steady state ATP hydrolysis by kinesin-5's motor domain (residues 1-368, expressed in bacteria) 25-times more potently when microtubules, the motor protein's tracks, are present in the reaction (IC_{50}: PCTP + microtubules + kinesin-5 = 1.2 ± 0.3 μM; PCTP + kinesin-5 = 30 ± 7 μM; **Figure 3.2a**). I next examined inhibition of kinesin-5 driven microtubule gliding. These assays require protein constructs that are larger than those consisting of the monomeric ATPase domain. I found that full-length homotetrameric kinesin-5 (from *Xenopus laevis*, expressed in insect cells) drives microtubule gliding at 19.6 ± 4.5 nm/s (1 mM MgATP). Remarkably, even at 50 nM of PCTP, the motor activity of kinesin-5 was completely inhibited (**Figure 3.2b**). As this concentration of PCTP is much lower than its IC_{50}, this finding suggests that tightly bound motor protein-microtubule complexes are formed in the presence of PCTP and these complexes act as brakes against other active motor protein molecules to stop microtubule motion. This state of tight kinesin—microtubule binding, referred to

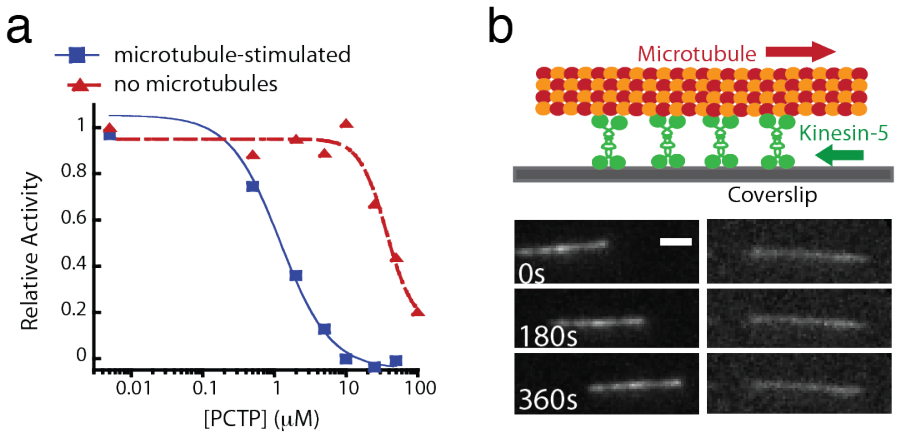


Figure 3.2 PCTP inhibition of kinesin-5 is affected by microtubules. (a) Representative curve from a steady-state ATP hydrolysis assay showing the response of kinesin-5 to PCTP, in the presence and absence of microtubules. IC_{50}s: 1.2 μM ± 0.3 (kinesin-5 + microtubules), 30 μM ± 7 (kinesin-5, nomicrotubules)**.** IC_{50} values are averages ± s.d. (n = 3). **(b)** In the presence of MgATP (1 mM) and DMSO, homotetrameric kinesin-5 drives microtubule gliding at an average rate of 19.6 ± 4.5 nm/s (left panel, n > 25). PCTP (50 nM) inhibits this activity (right panel, n > 30). Scale bar is 2 μm.

as rigor, is seen when kinesin proteins are inhibited by a non-hydrolyzable ATP analog (Vale et al., 1985). Experiments by Mitchison and colleagues using microtubule-pelleting assays confirm my findings that PCTP induces rigor (Groen et al., 2008). Together, these ATPase and microtubule gliding assay data suggest that the binding mode of PCTP to kinesin-5 should be examined in the presence of microtubules and may help explain why I had difficulty obtaining a crystal structure of the kinesin motor domain with PCTP in the absence of microtubules. Thus, analyzing the PCTP-kinesin-5 interaction in the presence of microtubules provides an interesting case for developing and applying the SILIC approach.

Examination of PCTP binding with SILIC

Guided by available SAR data for PCTP (Rickert et al., 2008), a colleague, Sudhir Kashyap, designed and synthesized analogs of PCTP, which contain an azide substituent on the phenyl ring as a photoreactive group and have natural (hydrogen, H; compound **1**) and heavy (deuterium, D; compound **2**) isotopes (**Figure 3.3a**). Synthesis of compound **1** was based on the published procedure for PCTP, but started with 4-bromophenyl-acetonitrile to position the aryl azide group (Rickert et al., 2008). To incorporate deuterium atoms that generate a 4 Da mass difference compared with **1**, toluene-d_8 was used as a starting material to synthesize the substituted phenyl ring moiety in **2**. Notably, the introduction of the crosslinking group in PCTP did not affect the potency of the compounds and, as

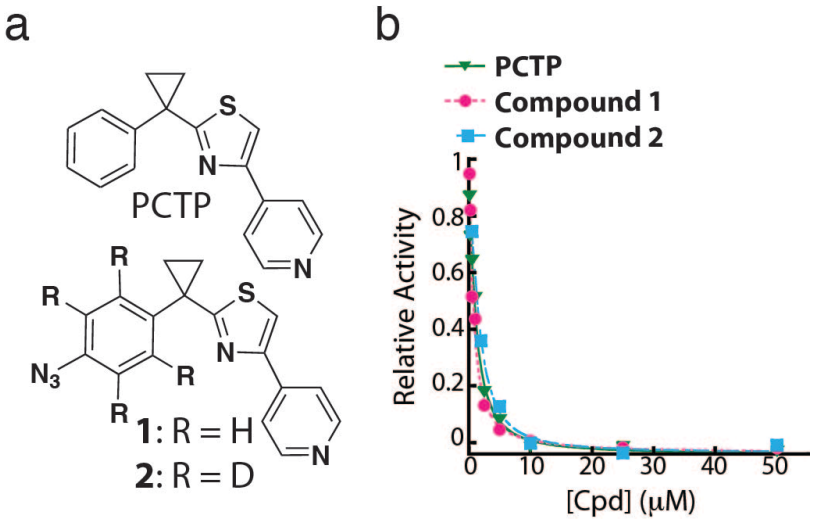


Figure 3.3 SILIC analogs of PCTP. (a) Chemical structure of PCTP and the natural and heavy crosslinkable analogs generated for SILIC. **(b)** Potency of compounds **1** and **2** compared with PCTP in inhibiting kinesin-5 activity, as examined using a steady-state microtubule-stimulated ATP hydrolysis assay. IC_{50}s: 1.2 μM ± 0.3 (PCTP), 1.2 μM ± 0.4 (compound **1**), 1.0 μM ± 0.4 (compound **2**). IC_{50} values are averages ± s.d. (n = 3).

expected, the natural and heavy analogs have similar activities against kinesin-5 (**Figure 3.3b**, IC_{50}: **1** = 1.2 ± .4 μM; **2** = 1.0 ± .4 μM).

To conduct crosslinking, a 1:1 mixture of **1** and **2** (3 mM each) was incubated with kinesin-5 motor domain and microtubules. After UV irradiation at 254 nm for 30 min, the reaction mixture was resolved by SDS-PAGE. Following in-gel digestion, the peptide mixture was separated and analyzed by LC-MS/MS by my colleague Xiang Li. Computer-based analysis with the software MaxQuant (Cox and Mann, 2008) was used to efficiently detect peaks with a mass difference of 4 Da that eluted at the same time from the LC column. Manual analysis of potential peaks was also conducted. In independent experiments (n = 3), a single peptide was identified with equal intensity peaks corresponding to the expected mass of the peptide plus the mass of **1** and **2** (**Figure 3.4a**). This peptide is likely to be the major inhibitor—peptide adduct. Further analysis by MS/MS identified this peptide as a fragment corresponding to Ser^{120} – Arg^{138} of kinesin-5 and the crosslinking site of the inhibitors is very likely at one of three amino acid residues (Tyr125, Thr126, or Trp127) (**Figure 3.4b**).

I next examined whether excess PCTP can suppress the crosslinking of **1** (or **2)** to kinesin-5. For this experiment, I crosslinked kinesin-5 with **1** (3 μM) in the presence of excess PCTP (45 μM) and, in parallel, crosslinked kinesin-5 with **2** (3 μM) alone. These samples were then mixed and processed for mass spectrometry analysis. This protocol provided a quantitative readout of the competition, as only one set of peaks in the mass spectrum should be

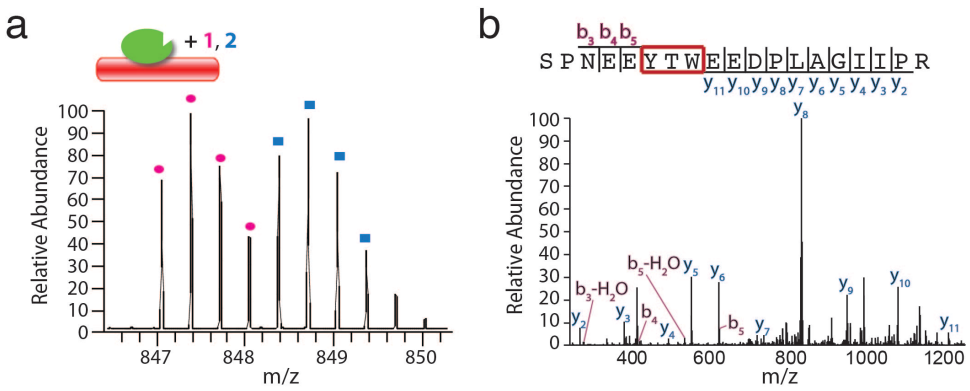


Figure 3.4 Crosslinking of PCTP analogs to kinesin-5. (a) A representative mass spectrum of the 'signature' peaks that denote peptide fragments of kinesin-5 crosslinked by equimolar **1** (pink circles) and **2** (blue squares) in the presence of microtubules. Cartoon shows kinesin-5 (green) and microtubules (red) with crosslinking compounds. **(b)** The MS/MS spectrum and summary of fragmented ions of the crosslinked peptide.

suppressed. As shown in **Figure 3.5**, PCTP competes with **1**, suggesting that PCTP also binds at a site proximal to residues Tyr125, Thr126, and Trp127.

In the three-dimensional structure of the kinesin-5 motor domain, residues Y125, T126, and W127 map to a portion of loop-5 (**Figure 3.6**, PDB: 2WOG) (Kaan et al., 2010). This loop is part of an allosteric binding site for several other kinesin-5 inhibitors (Kaan et al., 2010; Yan et al., 2004), including S-Trityl-L-cyteine (STLC, IC_{50} = 1 μM, steady-state ATPase assay without microtubules (DeBonis et al., 2004)). As these inhibitors bind this pocket in the absence of microtubules, I examined if PCTP did the same. To test this, I repeated the crosslinking and competition experiments in the absence of microtubules and found the same single crosslinking site (**Figure 3.7**). Furthermore, I found that addition of excess STLC (50 mM) prevented crosslinking of **1** (5 mM) to kinesin-5, relative to the reference sample (kinesin-5 crosslinked to **2**, **Figure 3.8**). Together, these data suggest that PCTP, like STLC, binds in the allosteric loop-5 site.

Site-directed mutagenesis to characterize PCTP—kinesin-5 interaction

I next used site-directed mutagenesis to analyze whether inhibition of kinesin-5 by PCTP was sensitive to changes in the allosteric pocket. A leucine-214 to alanine (L214A) mutation, in this allosteric binding site of kinesin-5 (**Figure 3.6**), is known to suppress STLC inhibition (Brier et al., 2006). Using the steady-state ATP hydrolysis assay, both with and without microtubules, I found

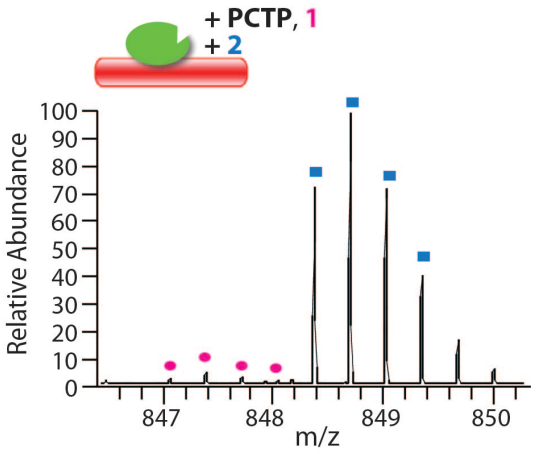


Figure 3.5 PCTP competes with 1 for kinesin-5 binding. When **1** (3 μM) is crosslinked in the presence of PCTP (45 μM), there is near complete loss of crosslinked peptide, relative to that observed in the presence of **2** (3 μM) alone.

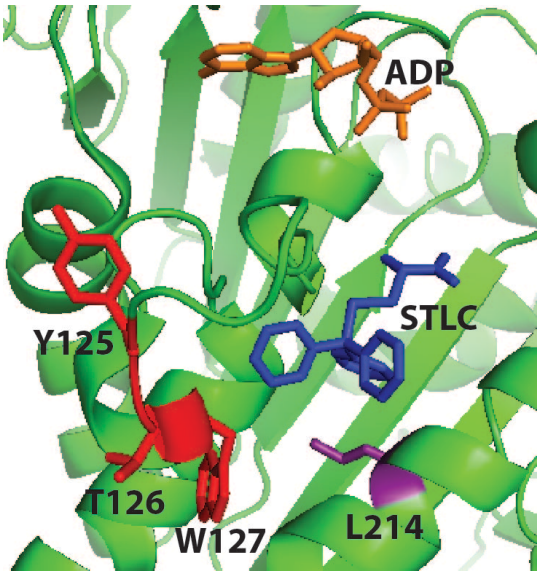


Figure 3.6 The loop-5 region of the kinesin-5 motor domain. The kinesin-5 motor domain (green, PDB: 2WOG) (Kaan et al., 2010) in complex with ADP (orange) and STLC (blue). The crosslinked residues Y125, T126, and W127 are shown in red, while residue L214 is in purple.

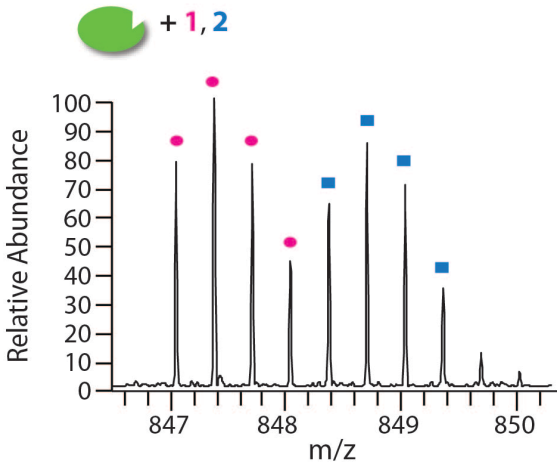


Figure 3.7 Mass spectrum of peptide—inhibitor adduct after crosslinking of kinesin-5 with 1 and 2 in the absence of microtubules. The 'signature' peaks, with the expected isotopic distribution and mass differences indicate crosslinking occurs without microtubules.

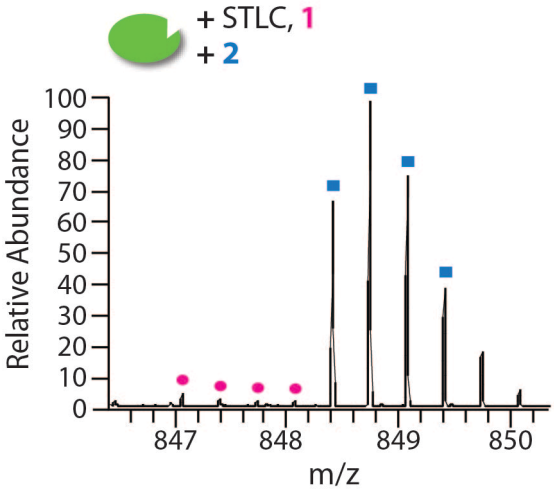


Figure 3.8 STLC competes with 1 for kinesin-5 binding. When **1** (5 μM) is crosslinked in the presence of STLC (50 μM), there is near complete loss of crosslinked peptide, relative to that observed in the presence of **2** (5 μM) alone.

that this mutation led to a ten-fold increase in the potency of PCTP (**Figure 3.9**). Importantly, crosslinking of **1** and **2** to kinesin-5, which had the L214A mutation, resulted in identification of the same binding site (**Figure 3.10**). In addition, I find PCTP is an ATP-competitive inhibitor of the kinesin-5 L214A mutant (**Figure 3.11**), indicating that the mode of inhibition is not altered by this mutation. These data provide further evidence supporting the model that PCTP binds to the loop-5 site.

In chapter 2, I described selections for clones resistant to PCTP. These clones contained two distinct mutations in kinesin-5 that are correlated with resistance to PCTP in cells (**Figure 2.11**, **Figure 3.12**). These mutations, Y104H and S269N, do not map to the loop-5 region of kinesin-5 but do map together in the three-dimensional structure of kinesin-5 (**Figure 3.13,** PDB: 2WOG) (Kaan et al., 2010). I decided to examine whether these mutations confer PCTP resistance to kinesin-5 *in vitro*. I found that both mutations confer resistance to PCTP in a steady-state microtubule-stimulated ATPase assay with the motor domain of kinesin-5 (**Figure 3.14a).** Interestingly, in the absence of microtubules, these mutations do not confer resistance to PCTP (**Figure 3.14b**). Kinesin-5 S269N is actually more sensitive to PCTP than the wild type protein in the absence of microtubules. To determine whether kinesin-5 S269N may significantly alter the structure of kinesin-5 I conducted crystallography studies (see **Appendix 2**).

During the mass spectrometry peak finding analysis, I carefully examined the peptides expected from the region of the protein containing Y104 and S269;

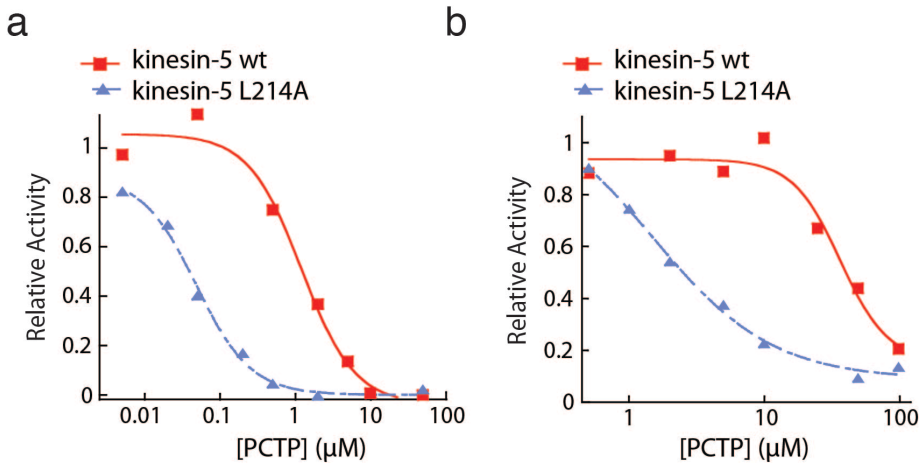


Figure 3.9 Effect of PCTP on kinesin-5 L214A. (a) Representative curve from a microtubule-stimulated steady-state ATP hydrolysis assay showing the response of kinesin-5 wild type (wt) and L214A to PCTP. IC_{50}s: 1.2 μM ± 0.3 (kinesin-5 wt), 110 nM ± 60 (kinesin-5 L214A)**. (b)** Representative curve from a steady-state ATP hydrolysis assay in the absence of microtubules showing the response of kinesin-5 wild type (wt) and L214A to PCTP. IC_{50}s: 30 μM ± 7 (kinesin-5 wt), 2.4 μM ± 0.4 (kinesin-5 L214A)**.** IC_{50} values are averages ± s.d. (n = 3).

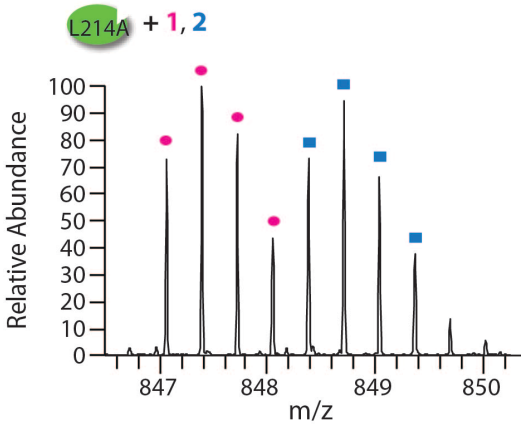


Figure 3.10 Crosslinking of PCTP analogs to kinesin-5 L214A. Mass spectrum of peptide-inhibitor adduct after crosslinking of kinesin-5 L214A with **1** and **2** (no microtubules). The 'signature' peaks, with the expected isotopic distribution and mass differences indicate crosslinking occurs to the mutant protein.

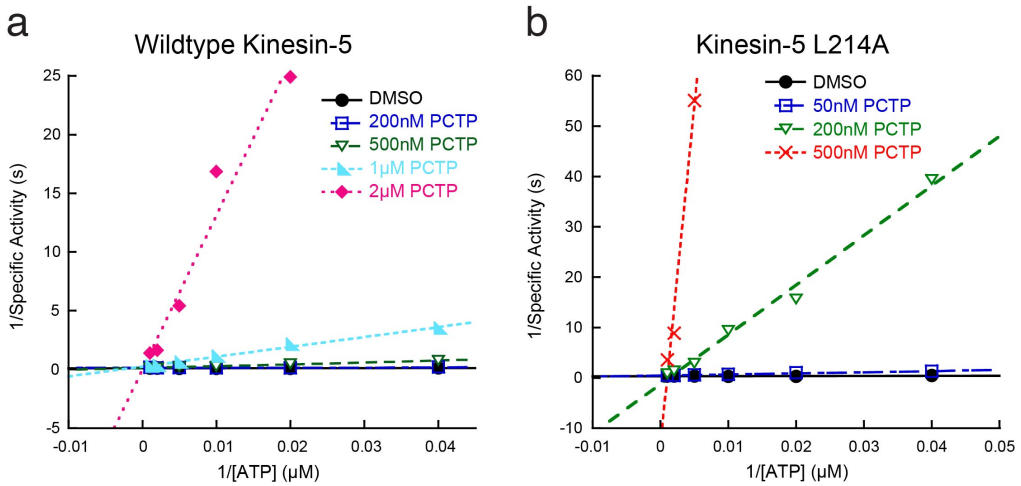


Figure 3.11 Lineweaver-Burk plots of kinesin-5 wild type (a) and L214A mutant (b). Plots were made using microtubule-stimulated ATPase activity at 25 μM, 50 μM, 100 μM, 200 μM, 500 μM, and 1 mM ATP. Individual linear regressions were calculated for each inhibitor concentration. Data values are average of 3.

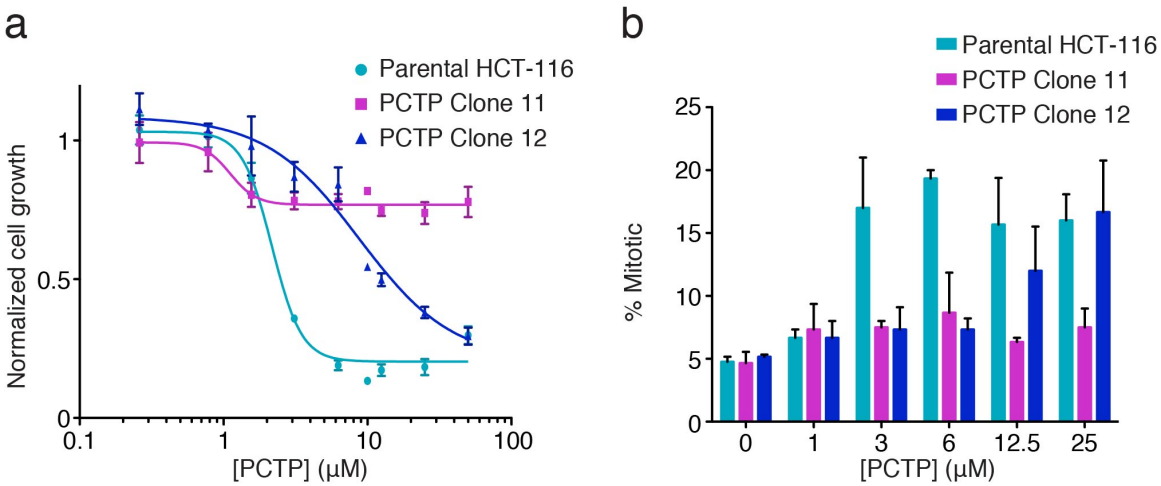


Figure 3.12 PCTP-resistant clones were obtained through selection with drug. (a) Median lethal dose values (LD_{50}s) were obtained by measuring the number of metabolically active cells after treatment with a range of concentrations of PCTP. (b) HCT-116 parental cells and clones 11 and 12 were treated with PCTP for five hours before methanol fixation. Fixed cells were stained for tubulin and DNA before they were counted under the microscope. The percentage of cells that were mitotic is plotted (mean ± sem, n = 3).

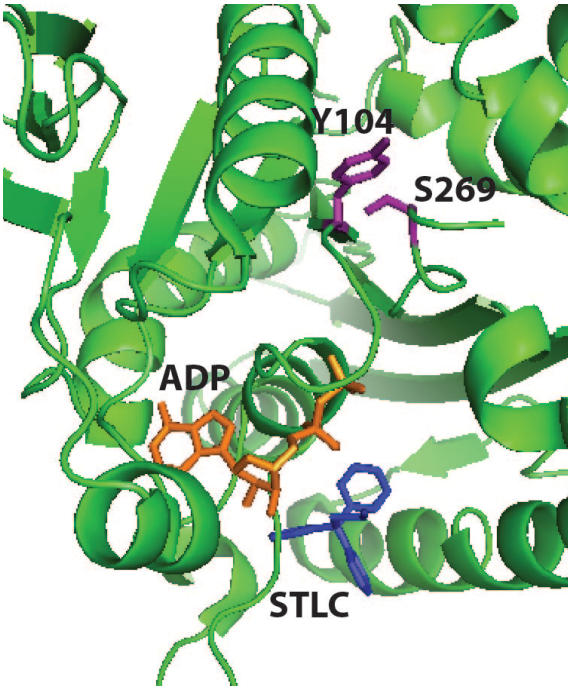


Figure 3.13 Residues Y104 and S269 in the motor domain of kinesin-5. The kinesin-5 ATPase domain (green, PDB: 2WOG) (Kaan et al., 2010) in complex with ADP (orange) and STLC (blue). Residues Y104 and S269 are shown in purple.

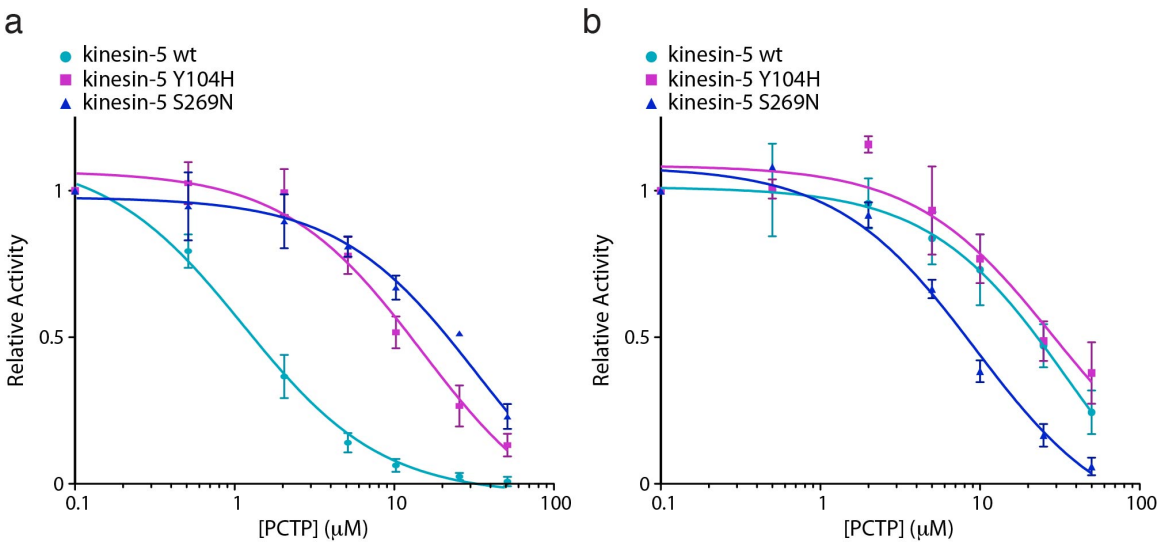


Figure 3.14 Effect of PCTP on kinesin-5 Y104H and S269N. (a) Microtubule-stimulated steady-state ATP hydrolysis assay (plotted values are mean ± sem) showing the response of kinesin-5 wild type (wt), Y104H, and S269N to PCTP. IC_{50}s: 1.2 μM ± 0.3 (kinesin-5 wt), 14 μM ± 2 (kinesin-5 Y104H), 32 μM ± 3 (kinesin-5 S269N)**. (b)** Steady-state ATP hydrolysis assay in the absence of microtubules (plotted values are mean ± sem) showing the response of kinesin-5 wild type (wt), Y104H, and S269N to PCTP. IC_{50}s: 30 μM ± 7 (kinesin-5 wt), 28 μM ± 2 (kinesin-5 Y104H), and 9.3 μM ± 3 (kinesin-5 S269N)**.** Values are averages ± s.d. (n = 3-5).

however, I never saw any indication of crosslinking. Thus, my interpretation is that the kinesin-5 site represented by Y104 and S269 is an allosteric regulatory site that moves upon microtubule binding, while the loop-5 region is the binding site of PCTP. Thus my model is that PCTP binds at or near the site bound by other kinesin-5 inhibitors that are not ATP-competitive.

Discussion

In summary, I've demonstrated that SILIC can be used to map the binding site of an inhibitor to its target. This approach should be effective in analyzing inhibitor-target interactions, particularly when structural studies are intractable. A potential limitation of the approach is that the incorporation of a photo-crosslinking moiety, such as an aryl azide, may not always be feasible. However, aryl groups appear at high frequency in bioactive small molecules. In these cases, SAR studies could allow the inclusion of an azide moiety such that it does not alter the inhibitor's mechanism of action and yet is proximal to the target's binding site residues. In the future this approach may be extended to more complex cellular contexts to examine whether cellular physiology alters the mode of drug—target interaction and mechanism of drug action.

Chapter 4: Conclusions and Discussion

In the preceding chapters I have presented two methods that I hope will benefit the study of small molecule inhibitors and the proteins to which they bind. In this section of my thesis I will describe in greater detail some of my findings and their implications. For chapter 2 I will present ideas for further improving and expanding the scope of my method for target identification. I will also analyze my results regarding kinesin-5 and PCTP that were reported in chapter 3.

My approach to target identification

In chapter 2, I described an approach that combines next-generation sequencing technology with bioinformatics analyses in order to examine pathways and proteins involved in drug resistance and action. My studies with the drugs BI 2536 and bortezomib indicate that, by analyzing drug-resistant clones through transcriptome sequencing and bioinformatics, a short list of potential drug targets can be obtained. Biochemical and cytological studies are then used to validate which of these proteins are direct targets and which represent indirect mechanisms of action.

While others have previously isolated resistant clones from human cells and analyzed resistance mutations in a targeted fashion (Girdler et al., 2008), before I began this study it was not clear whether resistance mutations could be

identified in an unbiased manner in human cells. My approach takes advantage of next-generation DNA sequencing methods and robust bioinformatic analyses to overcome the large size and complexity of the human genome and the genetic heterogeneity associated with many human cell lines, particularly those which can be grown in culture. By isolating multiple drug resistant clones and grouping those that are closely related, it is possible to obtain a short list of genes which represent potential drug targets.

Importantly, the use of RNA-seq provides information on single nucleotide variations, insertions/deletions, and mRNA transcript levels (Wang et al., 2009). I have shown that both the mutational analysis (which includes single nucleotide variations and insertions/deletions) and the transcript level analysis can be crucial to the discovery of drug resistance mechanisms. A limitation of the method, though, is that only genes with sufficient sequencing coverage will be identified with my approach. If a gene is carrying a mutation but is expressed at levels below the detection threshold in either the drug-resistant clone or the parental cell line, it will not be detected by the bioinformatics analyses. In order to detect rare transcripts the sequencing depth would need to be increased.

An essential feature of my approach is conducting transcriptome sequencing of the parental cell line and using these data to identify mutations present only in resistant clones. The sequenced parental population should be as close as possible to the population of cells used for resistant selections. In my experiments with bortezomib I used a population of cells for selections that were many passages past the cells processed for RNA-seq (only one parental HCT-

116 population – which was closest to the population selected with BI 2536 – was analyzed by transcriptome sequencing). I conjecture that this oversight is responsible for multiple genes being mutated in all bortezomib clones (**Figure 2.8b**) and that the four non-PSMB5 genes were likely mutated with high frequency in the actual parental population. This prediction needs to be verified by analyzing the exact parental population used for bortezomib selection by RNA-seq. In future applications of this approach, heed should be taken to sequence the exact population of cells used for drug selection.

Drugs with multiple targets

A potential limitation of this approach to target identification is that it may be unsuccessful for compounds with more than one target; however, this depends upon the particular targets. If the targets are redundant, in that mutation of any target rescues cell growth, then sequencing a large number of clones would likely identify the targets. While this would require a significant research investment and the number of clones needed is unclear, our method should, in principle, be successful.

In the case of a drug that causes toxicity through many distinct pathways, each target would need to be mutated to rescue growth. It is unlikely that mutations in all targets would occur simultaneously and, thus, it may be difficult to select for and obtain clones that had drug-specific resistance. I predict this is the case for ARQ-197, BE-54017, and Cladoniamide A, but further research is needed to test this hypothesis.

When a compound has multiple targets but only one of these results in cell death then only a mutation in this target would be required for resistance. In this scenario, only the cytotoxicity-conferring target would be detected.

If a small molecule has multiple targets, but affects each target with a different potency, my method would likely only identify the most sensitive target. This is due to conducting selections with concentrations of compounds just high enough to kill most cells. However, if higher drug concentrations were used, it may be possible to identify more targets. Interestingly, my proof-of-concept studies with bortezomib provide some insight into this issue. In addition to inhibiting PSMB5, bortezomib has been shown to inhibit PSMB1, another subunit of the proteasome, in cells (Berkers et al., 2005). A brief examination of the other mutations identified in each clone (those which were not common to multiple independent clones), showed that PSMB1 is also mutated in two of the clones (A and D, **Appendix 1, Tables A7 and A10**). While I have not examined whether the PSMB1 mutation (R128H) confers resistance to bortezomib, or added resistance in a PSMB5 mutant background, its presence intimates that deeper analysis of the sequencing data obtained through our method might suggest secondary targets of a drug. In my study, bortezomib-resistant clones A and D, which contain the mutation in PSMB1, do not have altered LD_{50} values when compared with clones B and C (which contain only the A108T mutation in PSMB5, **Figure 2.7b**). Exploring whether or not the PSMB1 mutation affects the sensitivity of the proteasome to bortezomib is an interesting area for further study.

Expanding the scope of my approach

In my proof-of-concept studies, I only conducted resistance selections based on cytotoxicity. Nevertheless, I do not believe my approach is limited to cytotoxic compounds. By using a mismatch repair deficient cell line, the parental population of cells has inherent genetic heterogeneity. Thus, phenotypic or reporter-based readouts could, in principle, be used to select clones with reduced drug response, eliminating the need to select for cell growth. High-throughput screens have been developed to identify small molecules that elicit a desired phenotype, such as the transcriptional activation of a luciferase reporter (Huang et al., 2009). These same systems should be able to select for drug-resistant clones when coupled with a cell sorting system (for example, fluorescent activated cell sorting). Unfortunately, using a reporter-based readout would introduce bias into the approach, as only targets in the pathway resulting in activation of the reporter would be identified. Further experimentation is needed to confirm whether my target identification method can be used for non-cytotoxic drugs.

My approach for target identification relies on transcriptome sequencing, which provides information on mutations in mRNA and changes in gene expression. In order to identify other potential mechanisms of resistance, such as mutations in promoter sequences and epigenetic changes, my strategy can be combined with other genomic methods, such as full genome sequencing, bisulfite conversion of unmethylated cytosines followed by sequencing to detect DNA

methylation, or chromatin immunoprecipitation followed by sequencing to evaluate transcription factor binding sites (Johnson et al., 2007; Laird, 2003). Genomics methodologies are rapidly advancing and becoming more accessible to general users (Wong et al., 2011). Thus, as research advances it will be beneficial to compare the various sequencing strategies and identify the most effective technologies for identifying resistance mechanisms.

In addition to coupling this method with other genomics technologies, I believe it would be beneficial to explore how to further utilize the transcriptome data. Currently, my approach systematically investigates coding mutations (single nucleotide variations and insertions/deletions) to identify those found in more than one independent clone. The gene expression data, however, were only examined after the mutation analysis did not identify a mechanism of resistance for all BI 2536-resistant clones. While the current bioinformatic strategy was able to successfully identify the known target of both BI 2536 and bortezomib, I feel that the overall approach may benefit from a systematic analysis of the gene expression data. I envision analyzing the transcript abundance data in the same way as the single nucleotide variation data, by identifying genes up-regulated or down-regulated in more than one independent clone. However, it is not clear how great of a change in gene expression is necessary to be physiologically relevant. The current analysis reports on all genes with three-fold changes in expression compared to the parental population (**Appendix 1, Tables A.12 – A.33**). As each clone had 3 – 15-times more genes with altered transcript abundance than with single nucleotide variations, more research is needed to identify a gene-

expression cut-off that is biologically meaningful and also tractable to bioinformatics analyses. Further research is also needed to identify a process for systematically analyzing the gene expression data.

Advantages of my approach

Even without additional technological or bioinformatics developments, the approach I have helped develop has several key advantages over other target identification methodologies. My approach allows for simultaneous identification of both drug targets and unrelated mechanisms of drug resistance in human cells. Unlike affinity-based approaches, this method can be used for drugs that are not very potent and the strategy does not rely on chemical modifications of the bioactive compound, which can alter its mechanism of action. The approach should be able to analyze mechanisms of resistance in any human cell that can be grown in culture, enabling cell-type specific analyses. The strategy identifies targets without bias to any target class or pathway. Finally, the data provided by this method, when combined with biochemical analyses, identify targets with a level of proof that meets the gold standard of the field.

Kinesin-5—small molecule interactions

In chapter 3 I describe a method that incorporates heavy and light isotopes into small molecules to improve their detection after crosslinking to a protein. I use this strategy to investigate how the small molecule PCTP interacts with its target, kinesin-5.

Kinesins are motor proteins that can use ATP hydrolysis to drive movement along microtubules. A decade ago, the small molecule monastrol was found to inhibit the kinesin-5 protein (Mayer et al., 1999); this was the first chemical inhibitor that targeted a protein, other than tubulin, needed for mitotic spindle assembly. This discovery helped catalyze the development of drugs against kinesin-5, some of which are currently in clinical trials (Huszar et al., 2009). Most kinesin-5 inhibitors, including monastrol and STLC, do not compete with ATP and bind to an allosteric region, called loop-5. These kinesin-5 inhibitors are also highly selective for kinesin-5. In contrast, PCTP and its close analogs have been shown to inhibit the kinesin proteins MCAK and Kif3A in addition to kinesin-5 (Rickert et al., 2008). My original goal was to improve the specificity of PCTP for kinesins other than kinesin-5. To accomplish this, I wanted structural information regarding the binding site of PCTP, which led to the development of the SILIC technique.

The finding that PCTP crosslinks to the loop-5 binding site and is competitive with the loop-5 inhibitor STLC was surprising for two main reasons. First, most nucleotide-competitive inhibitors bind to the nucleotide-binding site, not a separate allosteric site (Zhang et al., 2009). Thus, I had expected to see PCTP crosslink to the ATP-binding pocket of kinesin-5. The second reason it was unexpected that PCTP crosslinked to the loop-5 pocket is that this loop is not conserved in other kinesins. The sequence of kinesin-5 contains an additional 5 – 12 amino acids in loop-5 compared with other kinesins, indicating that the loop is

not as elongated in other kinesins, including MCAK and Kif3A, which are also sensitive to PCTP. This suggests that more studies are needed to determine how PCTP interacts with other kinesin proteins, particularly MCAK. These studies could also utilize the SILIC method.

Interestingly, another ATP-competitive inhibitor of kinesin-5 has also been shown to bind outside of the nucleotide-binding pocket (Luo et al., 2007). GSK-1 has been proposed to bind to an allosteric site that is distinct from the binding site of PCTP and STLC, based on a combination of resistant cell lines, site-directed mutagenesis, and inhibitor crosslinking experiments. Further studies with both PCTP and GSK-1 are needed to reveal how binding at allosteric sites results in ATP-competitive inhibition that leads to a tightly bound microtubule-motor protein complex.

One approach to investigate how PCTP drives tight binding of kinesin-5 to microtubules is to examine mutations that affect the PCTP—kinesin-5 interaction. I identified two mutations in kinesin-5, Y104H and S269N, which were correlated with resistance to PCTP in human cells. Both mutations individually confer resistance to PCTP in the presence of microtubules *in vitro*. However, in the absence of microtubules these mutant proteins were not resistant to PCTP and one of the mutations, S269N, was more sensitive to the compound. The residues Y104 and S269 map together in the structure of kinesin-5 and are mostly conserved among kinesin proteins (**Figure 3.12, Figure 4.1**). I believe an

interesting area of further study would be examining how this site is important in PCTP inhibition of kinesin-5 and microtubule binding.

I identified four mutations in kinesin-5 associated with resistance to STLC, but have not yet verified whether any of these mutations are sufficient to confer resistance to STLC. An initial examination of these mutations indicates that they are scattered throughout the kinesin-5 motor domain (**Figure 4.2**). Notably, one of the mutations associated with STLC resistance, A103V, is in the same region of the protein as the PCTP-resistant mutations. I believe all of these mutations warrant further investigation as they may lead to insights into the mechanism and regulation of kinesin-5 inhibitors.

Perspective

While several approaches are available to identify the proteins a small molecule binds to and to characterize their interaction, they all have limitations. In my thesis work I developed two methods that overcome some of these limitations: a chemical genetics strategy facilitates the identification of small molecule targets in human cells; and a method for characterizing drug-binding sites simplifies existing techniques that use small molecule crosslinking. I expect these two approaches to become an important part of the chemical biology toolbox and to help the process of drug discovery.

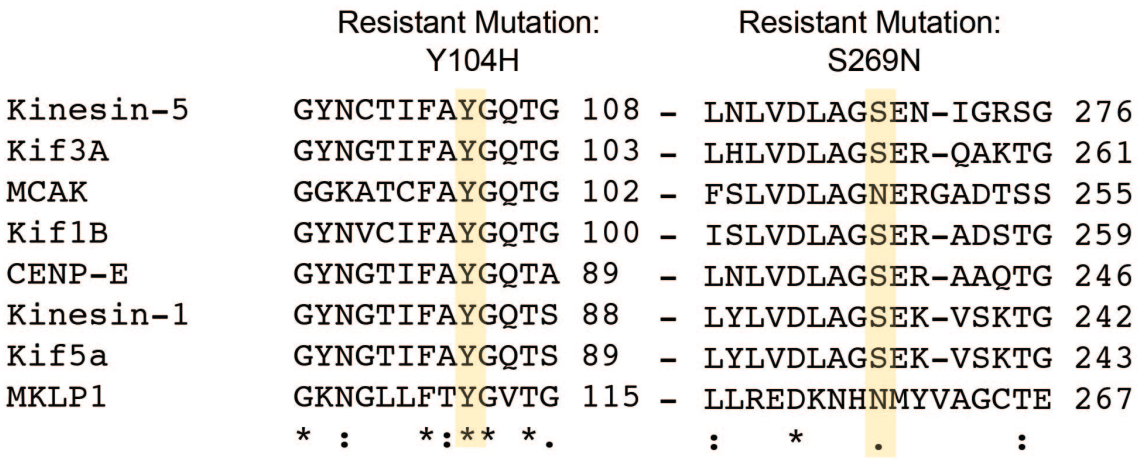


Figure 4.1 Clustal W alignment of kinesin sequences near kinesin-5 residues Y104 and S269. The alignment uses sequences of selected kinesin family members. Kinesin-5, Kif3A, and MCAK are susceptible to inhibition by PCTP, while Kif1B, CENP-E, Kinesin-1, Kif5a, and MKLP1 are not targets of PCTP (Rickert et al., 2008).

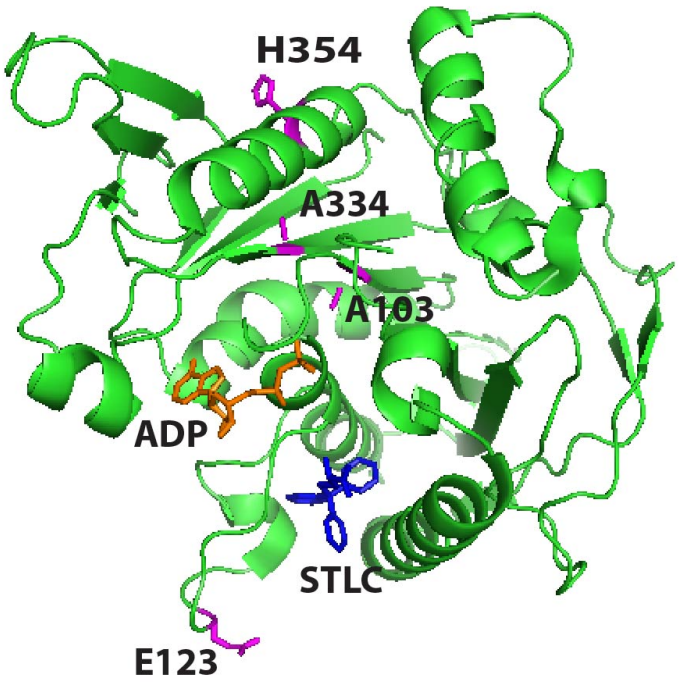


Figure 4.2 Location of kinesin-5 mutations associated with STLC resistance. The kinesin-5 ATPase domain (green, PDB: 2WOG) (Kaan et al., 2010) in complex with ADP (orange) and STLC (blue). Residues A103, E123, A334, and H354 are shown in purple.

Appendix 1

Table A.1.1: Significant mutations identified in BI 2536-resistant clone A

Gene	Amino Acid Change	# Reads Parental	# Reads Clone	Adjusted P
GNB2L1	C182Y	0/76	39/74	9.669E-12
RPLP2	S86P	0/77	33/74	3.942E-09
ACLY	L998P L1008P	0/45	25/41	1.967E-07
NAMPT	P417Q	0/41	26/44	7.503E-07
EIF3L	T354M	0/58	26/56	7.503E-07
CAPN1	M594I	0/55	24/52	0.000002499
VKORC1	F131S	0/46	21/41	0.000008499
LTA4H	P2S	0/42	16/27	0.00001429
NDUFV2	V22I	0/40	16/31	0.0001378
DPP7	D375V	0/41	16/32	0.0001594
CTNNB1	T297M	0/44	15/32	0.0002658
CCDC47	L163R	0/31	22/48	0.001517
PLK1	R136G	0/43	15/39	0.002645
PPP1R2	T92I	0/25	14/25	0.004361

Table A.1.2: Significant mutations identified in BI 2536-resistant clone B

Gene	Amino Acid Change	# Reads Parental	# Reads Clone	Adjusted P
H1F0	P142L	0/63	33/66	0.000000018
MAPK1IP1L	S211L	0/34	20/33	0.00005697
CALM2	G62S	0/54	22/55	0.00008785
UBE2R2	D226N	0/41	20/40	0.0001057
PLK1	R136G	0/43	18/38	0.0001766
EPRS	V530I	1/38	20/36	0.0003368
KITLG	K152N	0/50	19/49	0.0003368
PPIH	G175V	0/41	16/36	0.0008105
EFTUD2	A511V A546V	0/29	15/27	0.001467

Table A.1.3: Significant mutations identified in BI 2536-resistant clone C

Gene	Amino Acid Change	# Reads Parental	# Reads Clone	Adjusted P
GAL	A39T	0/46	18/31	0.00001633
UQCRC2	E274G	0/48	21/44	0.00004344
SF1	V299A V450A V325A	1/50	22/45	0.0001135
MCM5	R61W	0/36	17/36	0.001492
PLK1	G63S	0/34	12/22	0.002141
SCO2	R245I	0/27	11/17	0.002611
GEMIN4	L836S	0/36	10/19	0.004542

Table A.1.4: Significant mutations identified in BI 2536-resistant clone D

Gene	Amino Acid Change	# Reads Parental	# Reads Clone	Adjusted P
PPP1CA	D168G D212G D223G	0/51	20/42	0.00003813
SYNCRIP	D395Y D360Y D297Y	0/45	19/41	0.0001619
PYCR1	A8D	0/46	17/39	0.0003439
CHD4	G223R	0/48	13/28	0.0004466
ARF3	A27T	0/39	16/34	0.0005887
PTPRF	G70R	0/30	18/35	0.001079

Table A.1.5: Significant mutations identified in BI 2536-resistant clone E

Gene	Amino Acid Change	# Reads Parental	# Reads Clone	Adjusted P
PPP1CA	D168G D212G D223G	0/51	23/51	0.00005968
PYCR1	A8D	0/46	16/37	0.0008969
ARF3	A27T	0/39	14/28	0.0009591
TCF3	G302D	0/29	18/32	0.0009591
TAPBP	P149H P1141H	0/26	13/20	0.001343
SYNCRIP	D395Y D360Y D297Y	0/45	16/40	0.001343
C14orf4	R21H	0/37	12/23	0.001523
BAT3	S602F S1164F S1170F	1/31	14/22	0.002034
TCF7L2	A521S A504S A527S	1/41	14/28	0.003563
LTBP3	S349Y S466Y	0/37	13/29	0.003803
MACROD1	P142L	0/30	13/25	0.003928

Table A.1.6: Significant mutations identified in BI 2536-resistant clone F

Gene	Amino Acid Change	# Reads Parental	# Reads Clone	Adjusted P
PPP1CA	D168G D212G D223G	0/51	26/47	4.303 E-07
ARF3	A27T	0/39	21/34	0.000004478
KIAA2013	S379G	0/31	14/22	0.0003868
CD151	Y62C	1/57	14/31	0.00106
YARS	I118V	0/35	16/34	0.002154
SYNCRIP	D395Y D360Y D297Y	0/45	19/54	0.002802
BAT3	S602F S1164F S1170F	1/31	16/27	0.003037

Table A.1.7: Significant mutations identified in bortezomib-resistant clone A

Gene	Amino Acid Change	# Reads Parental	# Reads Clone	Adjusted P
PSMB5	A5T A108T	0/58	45/89	1.575E-08
PSMB1	R128H	0/56	42/83	3.583E-08
CAPN1	R87C	0/37	45/77	6.763E-07
PDCD6	R143W	0/38	35/68	0.00001485
IER5L	T266A	0/50	29/69	0.00003484
SNAPC2	L126F	0/28	25/40	0.00008129
C19orf33	A39P	0/35	36/74	0.00008494
SF1	V299A V450A V325A	1/50	34/80	0.000136
NUMA1	P1932L	0/25	32/54	0.000208
RRM1	P603H	0/40	30/70	0.0002232
MKI67	F1438V F1798V	0/41	25/57	0.0002376
KIAA0664	L167P	0/37	26/56	0.0002484
PPA2	M32T	0/32	18/32	0.0003507
PTP4A1	R93C	0/38	33/80	0.0004211
C11orf59	Y140N	1/37	34/70	0.0004873
NCAPG	A17G	1/37	13/20	0.0007323
EIF3B	G39E	0/32	17/32	0.0007784
MYO1C	V837M V821M V856M	0/31	33/77	0.001635
DHX38	P719T	0/22	27/49	0.002371
PMPCB	R238T	0/26	29/61	0.00284
ARHGAP29	I77V	0/25	33/70	0.00327
UBE2B	W96R	0/23	28/55	0.003655

Table A.1.8: Significant mutations identified in bortezomib-resistant clone B

Gene	Amino Acid Change	# Reads Parental	# Reads Clone	Adjusted P
PSMB5	A5T A108T	0/58	47/93	9.671E-09
IER5L	T266A	0/50	39/78	3.606E-07
MYH9	L906P	0/48	37/73	5.599E-07
CAPN1	R87C	0/37	42/74	0.000001111
VARS	A1291T	0/32	50/88	0.000004129
SF1	V299A V450A V325A	1/50	37/79	0.000017
MYO1C	V837M V821M V856M	0/31	45/85	0.00003752
RRM1	P603H	0/40	28/59	0.0000583
SFRS13A	R590G R952G	0/31	35/67	0.00007582
KIAA0664	L167P	0/37	26/53	0.00008169
PDIA4	A125V	0/52	29/85	0.0002119
PPA2	M32T	0/32	18/32	0.0002535
MKI67	F1438V F1798V	0/41	22/51	0.0003204
ETHE1	S19A	0/26	18/29	0.0004372
AP1M2	F107L	0/23	40/72	0.0005007
PDCD6	R143W	0/38	29/72	0.000577
AP1M2	I109T	0/23	35/67	0.001409
KIF1C	R1030H	0/25	29/59	0.002085
SPAG5	S353P	0/33	16/34	0.002085
POLG	A1105V	0/23	26/49	0.00212
ARHGAP29	I77V	0/25	28/59	0.003322
PPP1R10	R58H	0/22	32/63	0.003322
MAST2	T652M	0/21	21/37	0.003405
PMPCB	R238T	0/26	28/62	0.003994
APTX	L151P L63P L165P	0/18	29/50	0.004052
ABI2	R28W	0/19	23/40	0.004788

Table A.1.9: Significant mutations identified in bortezomib-resistant clone C

Gene	Amino Acid Change	# Reads Parental	# Reads Clone	Adjusted P
PSMB5	A5T A108T	0/58	34/63	2.582E-08
PDCD6	R143W	0/38	27/45	0.000002127
C19orf33	A39P	0/35	29/48	0.0000033
CAPN1	R87C	0/37	32/58	0.000005496
MYO1C	V837M V821M V856M	0/31	31/56	0.00004884
SF1	V299A V450A V325A	1/50	26/57	0.0001033
ID1	S8R	0/52	15/32	0.0001033
IER5L	T266A	0/50	16/35	0.0001141
MKI67	F1438V F1798V	0/41	19/42	0.0002948
KIAA0664	L167P	0/37	15/28	0.0002948
C11orf59	Y140N	1/37	26/51	0.0003872
BAZ1B	Y1358H	1/42	18/35	0.0004709
SNAPC2	L126F	0/28	13/21	0.0007291
ARHGAP29	I77V	0/25	28/55	0.001162
PPP1R10	R58H	0/22	19/32	0.001423
PPA2	M32T	0/32	11/20	0.001863
PTTG1	T3N	1/34	16/30	0.002387
PMPCB	R238T	0/26	23/51	0.003889
MRPL40	A161V	0/29	17/38	0.003892
RRM1	P603H	0/40	16/46	0.004319
PRNP	C115G V115G	0/27	15/31	0.004319

Table A.1.10: Significant mutations identified in bortezomib-resistant clone D

Gene	Amino Acid Change	# Reads Parental	# Reads Clone	Adjusted P
PSMB5	A5T A108T	0/58	34/66	4.927E-08
PSMB1	R128H	0/56	35/74	3.127E-07
RRM1	P603H	0/40	25/41	0.000001578
PTP4A1	R93C	0/38	37/67	0.000002102
C19orf33	A39P	0/35	30/49	0.000002102
PDCD6	R143W	0/38	27/49	0.000007936
C11orf59	Y140N	1/37	29/50	0.00003677
BAZ1B	Y1358H	1/42	20/35	0.00007982
SF1	V299A V450A V325A	1/50	26/57	0.00008717
MYO1C	V837M V821M V856M	0/31	28/53	0.000119
CAPN1	R87C	0/37	27/59	0.0001627
KIAA0664	L167P	0/37	17/34	0.00034
UBE2B	W96R	0/23	27/46	0.000427
PTTG1	T3N	1/34	14/21	0.000427
NUMA1	P1932L	0/25	18/29	0.000427
CDK5RAP3	V21G	0/33	13/22	0.000427
SNAPC2	L126F	0/28	14/22	0.000427
ARHGAP29	I77V	0/25	31/59	0.0005902
MRPS35	C199Y	0/30	23/48	0.0008332
PMPCB	R238T	0/26	22/42	0.001105
MKI67	F1438V F1798V	0/41	15/36	0.001229
PPA2	M32T	0/32	13/25	0.001465
NOP14	T658M	0/24	15/26	0.002046
DHX38	P719T	0/22	16/28	0.003174
MARCH6	W783R	0/20	26/48	0.00318
ABI2	R28W	0/19	14/22	0.00402
PLEC	A602G A761G A624G A592G A628G A610G A651G	0/29	12/24	0.004292
TUFM	A6G	0/43	7/14	0.004936

Table A.1.11: Significant mutations identified in bortezomib-resistant clone E

Gene	Amino Acid Change	# Reads Parental	# Reads Clone	Adjusted P
STT3A	R662S	0/43	31/59	0.000001715
CHD4	R1489H	1/43	33/58	0.000003451
PSMB5	M1V M104V	0/55	28/70	0.0000142
PYCR1	S232N	0/57	25/73	0.0001192
EGR1	P467T	0/38	36/86	0.0001875
ITM2B	A15T	0/29	26/48	0.000218
NCAPH2	S63Y	0/27	17/30	0.001011
PPP1R10	R58H	0/22	23/39	0.001151
UBE2Z	R321H	0/39	23/60	0.001302
CDK5RAP3	V21G	0/33	15/30	0.001328
SIAH2	A191T	0/28	15/27	0.00142
PUF60	A311P A285P A328P	0/47	12/29	0.001505
ID1	S8R	0/52	14/41	0.001946
ETHE1	S19A	0/26	9/13	0.002428
STT3A	V633A	0/43	20/61	0.00326

Table A.1.12: Genes with increased expression in BI 2536-resistant clone A

Gene	Increased Abundance (log 2)
NRP1	2.53
WFDC2	2.09
DNER	1.63
C3orf57	1.6
CPA4	1.59
KRT23	1.59
LHB	1.58
TP53I3	1.55
MDGA1	1.54
NR4A2	1.53
PARP10	1.5

Table A.1.13: Genes with decreased expression in BI 2536-resistant clone A

Gene	Decreased Abundance (log 2)
RPS4Y1	-5.49
GPR110	-2.61
EIF1AY	-2.28
RGS2	-2.26
FAM107B	-2.25
ATP2C2	-2.08
GALNT3	-2
CYorf15A	-1.93
BAMBI	-1.92
SEC31A	-1.9
EHF	-1.88
SOX4	-1.88
KRT7	-1.86
RBPMS	-1.84
TACSTD2	-1.82
HNF1B	-1.81
PODXL2	-1.76
GRHL2	-1.75
DDX3Y	-1.68
ZNF608	-1.62
MNS1	-1.58
HOXA4	-1.58

Table A.1.14: Genes with increased expression in BI 2536-resistant clone B

Gene	Increased Abundance (log 2)
LCN2	2.36
FGFBP1	1.55

Table A.1.15: Genes with decreased expression in BI 2536-resistant clone B

Gene	Decreased Abundance (log 2)
RPS4Y1	-5.75
EIF1AY	-2.39
MCAM	-2.36
TACSTD2	-2.32
ANXA10	-1.95
CYorf15A	-1.93
HIST1H4A	-1.9
APPL1	-1.89
ALDH16A1	-1.77
TMEM59L	-1.77
NOVA2	-1.75
NGFR	-1.73
KIAA1543	-1.7
KISS1R	-1.69
C1QL1	-1.66
PALM3	-1.63
CCDC78	-1.62
NOTUM	-1.56
MRC2	-1.55
FABP6	-1.54
YJEFN3	-1.53
DDX3Y	-1.5

Table A.1.16: Genes with increased expression in BI 2536-resistant clone C

Gene	Increased Abundance (log 2)
PRF1	2.96
C3orf57	2.13
LEMD1	2
EGR3	1.94
TSPAN1	1.93
CSTA	1.88
FOSB	1.85
LCN2	1.74
ARRDC4	1.7
JUN	1.66
LHB	1.64
SYNPO	1.6
SERPINE1	1.58
TNNC1	1.57
MMP7	1.56
LOC730755	1.56
ABCA3	1.55
TACSTD2	1.52
KLRC4	1.51
PADI3	1.5
C6orf223	1.5
CHAC1	1.5

Table A.1.17: Genes with decreased expression in BI 2536-resistant clone C

Gene	Decreased Abundance (log 2)
RPS4Y1	-5.75
CALD1	-2.9
EIF1AY	-2.39
CYorf15A	-2.07
TM9SF3	-1.96
HECTD3	-1.95
DIAPH2	-1.77
DDX3Y	-1.68
FGF22	-1.68
CCDC69	-1.58
SLC10A7	-1.54

Table A.1.18: Genes with increased expression in BI 2536-resistant clone D (pages 107-108)

Gene	Increased Abundance (log 2)
KRTAP3-1	4.34
MYBPH	3.9
KLK5	3.87
ABCB1	3.83
LCN2	3.6
HIST1H2AC	3.01
LOC730755	2.97
MYL2	2.88
HIST1H2BD	2.79
IFIT2	2.74
COL17A1	2.66
KRTAP2-1	2.62
KLK6	2.61
CPA4	2.5
SUSD2	2.42
KLK7	2.4
IFIT1	2.36
KRT81	2.36
GAST	2.35
OASL	2.31
TP53I3	2.31
HIST1H2BC	2.25
TSPAN1	2.24
TNFRSF10C	2.15
NRP1	2.14
KRT37	2.14
KRT13	2.12
HIST2H2AA3	2.1
WFDC2	2.09
LOC402778	2.05
KLK10	2.04
IFI44	2.03
SERPINE1	2
KRTAP2-4	2
SRPX2	1.99
PRF1	1.98
PVRL4	1.98
SULF2	1.96
HIST1H2BJ	1.95
IFI27	1.94
WNT7A	1.92
HIST1H1C	1.92
C15orf52	1.9
RRAD	1.89
BTBD19	1.89
LOXL4	1.88
JUN	1.88

CLIC3	1.87
TRIM22	1.86
DKK4	1.85
C10orf114	1.84
HIST2H2BE	1.8
C1orf116	1.79
GRIN2C	1.79
HIST1H2BE	1.78
SHC2	1.75
HPD	1.73
IFI6	1.73
EBI3	1.72
WFDC3	1.72
S100A5	1.72
GALNT5	1.72
MAP6	1.69
HIST2H2AA4	1.68
HRNR	1.68
CFB	1.66
KRT6A	1.66
SPARC	1.63
TNNC1	1.62
EFCAB4A	1.62
IDUA	1.6
KRT14	1.58
SEMA3B	1.58
KRT17	1.58
ARRDC4	1.56
S100A3	1.55
SAMD4A	1.54
LAMA3	1.54
HIST1H2BK	1.54
LAMB3	1.53
PLEKHB1	1.53
HIST1H3H	1.52
CAPS	1.52
HLA-DRB1	1.51
AQP3	1.51
IFIT3	1.51

Table A.1.19: Genes with decreased expression in BI 2536-resistant clone D

Gene	Decreased Abundance (log 2)
RPS4Y1	-5.75
LMAN1	-2.94
EIF1AY	-2.28
ANXA10	-2.05
CYorf15A	-1.93
FOS	-1.79
GRHL2	-1.79
PTH2	-1.77
C1orf61	-1.7
INSIG1	-1.69
DDX3Y	-1.68
NRTN	-1.56
CENPM	-1.55
GALNT3	-1.53
TNFRSF13C	-1.53
GPR110	-1.53
TMEM149	-1.5

Table A.1.20: Genes with increased expression in BI 2536-resistant clone E

Gene	Increased Abundance (log 2)
ABCB1	4.12
TACSTD2	3.46
IFI27	2.64
ODZ3	2.58
TNNC1	2.38
GALNT5	2.21
CLIC3	2.14
MAP6	2.08
KRT81	2.07
MYBPH	2.04
C1orf116	1.98
LOC730755	1.96
LEMD1	1.95
HLA-DRB1	1.94
CPA4	1.92
IL8	1.88
HIST1H2BD	1.85
SPRR2D	1.85
HLA-DPA1	1.84
CTGF	1.72
IFI6	1.7
SELENBP1	1.68
NRG1	1.65
PRKAA2	1.63
GPR87	1.62
SHC2	1.61
SERPINE1	1.6
TSPAN1	1.58
HIST1H2BC	1.58
CXCL1	1.58
LOXL4	1.55
HIST1H2AC	1.54
TP53I3	1.5

Table A.1.21: Genes with decreased expression in BI 2536-resistant clone E

Gene	Decreased Abundance (log 2)
RPS4Y1	-5.75
LMAN1	-2.31
EIF1AY	-2.28
GPR110	-2.12
CYorf15A	-2.07
IAH1	-1.89
ZBED2	-1.86
DDX3Y	-1.78
DHRS3	-1.56
ANXA10	-1.56
HAS3	-1.52

Table A.1.22: Genes with increased expression in BI 2536-resistant clone F

Gene	Increased Abundance (log 2)
ABCB1	4.1
MYBPH	3.43
DCLK1	2.45
MAP6	2.25
HIST1H2AC	2.2
LOC730755	2.17
HIST1H2BD	2.17
HIST1H1C	2.16
TSPAN1	1.95
IFI27	1.85
IFIT1	1.8
IL8	1.79
LEMD1	1.78
IFIT2	1.77
PRSS33	1.75
TP53INP1	1.72
VIM	1.71
EBI3	1.58
HIST1H2BC	1.58
TP53I3	1.57
CLIC3	1.54
AQP3	1.53
CTGF	1.52
C3orf57	1.51

Table A.1.23: Genes with decreased expression in BI 2536-resistant clone F

Gene	Decreased Abundance (log 2)
RPS4Y1	-5.75
EIF1AY	-2.39
LMAN1	-2.31
FLNC	-2.23
CYorf15A	-2.07
GRHL2	-2.01
GALNT3	-2
ANXA10	-1.95
MAL2	-1.79
C2orf82	-1.71
ZBED2	-1.68
ATP2C2	-1.57
AIM1	-1.56
DDX3Y	-1.54
NRTN	-1.54
FABP6	-1.54

Table A.1.24: Genes with increased expression in bortezomib-resistant clone A (pages 114-115)

Gene	Increased Abundance (log 2)
LCN2	3.37
C12orf39	3.32
IL20RB	2.99
CLGN	2.54
FAM129A	2.54
NRP1	2.41
C9orf150	2.34
CRYBA1	2.34
LEMD1	2.26
THSD4	2.25
FOSB	2.24
JUN	2.11
LOC375190	2.1
MYH15	2.09
KLHDC7B	2.07
SHC2	2
SERPINB8	1.99
LAMP3	1.97
SERPINA5	1.96
TCP11L2	1.94
SEPP1	1.92
SLFN5	1.91
BEX4	1.91
CCDC68	1.9
ABCA3	1.83
KLRK1	1.82
C21orf56	1.81
CTGF	1.81
PADI3	1.8
ARRDC4	1.8
HLA-DRA	1.77
ABCC3	1.77
FHIT	1.76
TSPAN1	1.75
KLF9	1.75
CHAC1	1.75
FHL1	1.73
SYNPO	1.69
PRF1	1.68
JHDM1D	1.67
PKIB	1.66
C10orf11	1.66
LMBRD1	1.65
CCDC146	1.65
GCNT3	1.65

KLRC2	1.64
LAMC2	1.64
CLIP4	1.62
MAT1A	1.61
MAML3	1.61
SAMD3	1.6
EHBP1	1.6
DPP4	1.58
KLRC4	1.58
PGCP	1.56
PTPDC1	1.56
ADSSL1	1.56
EGR3	1.56
FUT8	1.56
LGALS2	1.54
SH3D19	1.54
GLDN	1.54
TMEM106B	1.54
YPEL5	1.53
PAQR8	1.52
TTC13	1.52
C6orf223	1.51
PCMTD1	1.51
NR4A2	1.5
EXD3	1.5
MXD1	1.5

Table A.1.25: Genes with decreased expression in bortezomib-resistant clone A

Gene	Decreased Abundance (log 2)
RPS4Y1	-5.75
EIF1AY	-2.66
HIST1H4A	-2.22
DDX3Y	-2.17
CYorf15A	-2.07
SLC43A3	-1.94
PXDN	-1.92
ACTA1	-1.73
ZNF704	-1.68
PODXL2	-1.65
PALM3	-1.63
ROMO1	-1.61
RPS29	-1.6
SDF2L1	-1.6
HSPH1	-1.58
ANP32C	-1.56
MT1H	-1.54
FAM166A	-1.51
C4orf48	-1.5

Table A.1.26: Genes with increased expression in bortezomib-resistant clone B (pages 117-119)

Gene	Increased Abundance (log 2)
IL20RB	2.92
NRP1	2.81
C12orf39	2.67
SERPINA5	2.49
LEMD1	2.43
KNG1	2.39
FAM129A	2.36
CLGN	2.26
PADI3	2.25
C9orf150	2.24
BEX4	2.15
APOD	2.12
KLHDC7B	2.1
SERPINB8	2.09
C3orf57	2.04
HLA-DPA1	2.03
TSPAN1	2.02
MYH15	2
FHL1	1.99
THSD4	1.96
SERPING1	1.95
CTHRC1	1.95
SEPP1	1.94
HLA-DRA	1.94
LOC375190	1.94
JHDM1D	1.94
KLRC2	1.92
CCDC68	1.92
SHC2	1.91
TCP11L2	1.89
FUT8	1.89
SATB1	1.89
BST2	1.87
SERPINE1	1.86
PRF1	1.86
ETV1	1.85
C10orf11	1.83
NEBL	1.82
KLK7	1.81
LMBRD1	1.79
F10	1.78
MAML3	1.77
ADSSL1	1.77
KLRK1	1.75
CD22	1.75

SPARC	1.75
PLAC1	1.74
SLFN5	1.74
KLF9	1.74
KIAA0040	1.74
LAMP3	1.71
VSNL1	1.71
KLRC3	1.7
LOC402778	1.7
SYNPO	1.7
ABCA3	1.69
PGCP	1.69
C21orf56	1.68
PKIB	1.68
EGR3	1.68
LRRC6	1.67
PLEKHB1	1.67
MAT1A	1.67
FHIT	1.67
ACTA2	1.67
YPEL5	1.66
DES	1.66
GJB4	1.66
TP53INP1	1.64
RGS6	1.64
DMGDH	1.64
LGALS2	1.63
SAMD3	1.63
HLA-DRB1	1.63
SGPP2	1.62
NR4A2	1.6
PCMTD1	1.6
CHAC1	1.59
HIST1H2AC	1.58
ABCC3	1.58
TMEM132B	1.58
LCN2	1.58
KIAA1370	1.57
TSC22D3	1.57
CLIP4	1.57
GNG13	1.57
SERPINF1	1.56
PTGDS	1.56
CCPG1	1.56
KLRC4	1.56
SPHAR	1.56
DEPDC6	1.55
GPM6A	1.54
TBC1D19	1.53
KRCC1	1.53
PCBP3	1.53

ZNF804A	1.53
CTSO	1.52
WWOX	1.51
C5orf53	1.51
LYSMD2	1.5

Table A.1.27: Genes with decreased expression in bortezomib-resistant clone B

Gene	Decreased Abundance (log 2)
RPS4Y1	-5.75
EIF1AY	-2.66
HIST1H4A	-2.22
DDX3Y	-2.11
CYorf15A	-2.07
SLC43A3	-2.03
GPR110	-1.89
PODXL2	-1.87
GNB3	-1.85
HNF1B	-1.81
HIST1H4E	-1.78
APOC4	-1.63
RPS29	-1.62
HIST1H2AB	-1.58
ACTA1	-1.56
SAP25	-1.51

Table A.1.28: Genes with increased expression in bortezomib-resistant clone C

Gene	Increased Abundance (log 2)
C12orf39	3.55
IL20RB	3.14
FAM129A	2.59
LCN2	2.47
CLGN	2.3
NRP1	2.29
SLFN5	2.24
MYH15	2.13
LAMP3	2.05
THSD4	2.02
SERPINB8	2
C9orf150	1.98
LEMD1	1.96
FOSB	1.94
TSPAN1	1.87
JUN	1.87
PDE2A	1.84
SERPINA5	1.8
TCP11L2	1.79
BEX4	1.78
PLAU	1.77
KLHDC7B	1.77
PRF1	1.77
SERPINE1	1.74
NT5E	1.69
CTGF	1.69
CLIP4	1.67
LOC375190	1.61
CD22	1.61
ARRDC4	1.6
CCDC68	1.6
C21orf56	1.58
JHDM1D	1.58
TMEM106B	1.58
VSNL1	1.56
TP53INP1	1.55
LAMC2	1.55
CHAC1	1.55
ABCA3	1.54
EHBP1	1.54
C4orf34	1.51
DMGDH	1.5

Table A.1.29: Genes with decreased expression in bortezomib-resistant clone C

Gene	Decreased Abundance (log 2)
RPS4Y1	-5.75
PXDN	-2.73
SLC43A3	-2.69
EIF1AY	-2.66
HIST1H4A	-2.22
DDX3Y	-2.17
C4orf48	-2.16
CYorf15A	-2.07
PODXL2	-2
PRR22	-1.99
HES4	-1.98
PALM3	-1.83
LRRC26	-1.83
CORO1A	-1.81
EFNA2	-1.78
FZD1	-1.78
APOC4	-1.77
HSPH1	-1.77
GNB3	-1.76
NOVA2	-1.75
HAUS7	-1.74
IGSF9	-1.72
NPB	-1.72
ICAM1	-1.72
AIM1L	-1.72
HIST1H4E	-1.69
FXYD1	-1.69
GLI2	-1.67
ACTA1	-1.64
SDF2L1	-1.63
ANP32C	-1.63
CTRL	-1.58
ZNF704	-1.58
TMEM145	-1.58
NUDT1	-1.58
WDR90	-1.55
EPN3	-1.53
ASB16	-1.53
CCDC28B	-1.52
HSPA2	-1.52
METRN	-1.52
SAP25	-1.51
WDR4	-1.51

Table A.1.30: Genes with increased expression in bortezomib-resistant clone D

Gene	Increased Abundance (log 2)
IL20RB	2.86
C12orf39	2.61
TSPAN1	2.22
SLFN5	2.14
C2orf78	2.09
KLK5	2.08
MYH15	2.04
C9orf150	2.04
FAM129A	2.01
KLK7	1.95
NRP1	1.93
SERPINA5	1.9
HLA-DRA	1.89
KLK6	1.88
PADI3	1.82
THSD4	1.81
C3orf57	1.74
PBX1	1.73
SLC7A11	1.72
ASNS	1.68
PKIB	1.67
PLAU	1.65
SERPINB8	1.65
SEPP1	1.62
C4orf34	1.62
C4orf32	1.6
BEX4	1.6
ARRDC4	1.58
KLK10	1.57
CLGN	1.54
CLIP4	1.54
HIST1H2AC	1.53
PRF1	1.52
HYAL1	1.52
ACTA2	1.5

Table A.1.31: Genes with decreased expression in bortezomib-resistant clone D

Gene	Decreased Abundance (log 2)
RPS4Y1	-5.75
EIF1AY	-2.66
SLC43A3	-2.34
HIST1H4A	-2.22
PALM3	-2.15
DDX3Y	-2.11
CYorf15A	-2.07
PODXL2	-2
LRRC26	-2
ACTA1	-1.92
C2orf82	-1.91
WNK2	-1.83
C4orf48	-1.83
CCDC28B	-1.82
HIST1H4E	-1.78
APOC4	-1.77
MUSTN1	-1.76
SH2B2	-1.75
GPR110	-1.75
SAP25	-1.74
RBPMS	-1.74
HNF1B	-1.73
BGLAP	-1.73
PPP2R3B	-1.71
GNB3	-1.68
GLI2	-1.67
TMEM121	-1.67
GPR157	-1.58
NFKBIL2	-1.58
C19orf76	-1.57
KLHL17	-1.56
YJEFN3	-1.56
ANP32C	-1.56
MT1H	-1.54
CTRL	-1.53
ZNF488	-1.53
AMH	-1.53
CBX2	-1.53
FZD1	-1.52
KRT7	-1.51

Table A.1.32: Genes with increased expression in bortezomib-resistant clone E **(pages 125-126)**

Gene	Increased Abundance (log 2)
WFDC2	3.2
LEMD1	2.67
FOSB	2.65
SERPINA5	2.62
KLRK1	2.6
JUN	2.6
FGF9	2.24
EGR3	2.23
KLRC4	2.19
IL8	2.18
SEPP1	2.11
GPR15	2.05
VSNL1	2.03
C10orf10	2.03
KIAA1199	2.02
KLRC2	2.02
RCAN2	1.97
MAT1A	1.95
NR4A2	1.94
ABCC3	1.93
SYNPO	1.91
KIAA0040	1.86
KLRC3	1.85
PADI3	1.82
EGR2	1.81
PLEKHB1	1.8
SCG2	1.76
TSPAN1	1.76
KLF9	1.74
CTHRC1	1.73
SPOCK3	1.7
DDIT4L	1.67
PKIB	1.66
LMBRD1	1.65
MAML3	1.61
SATB1	1.61
C10orf11	1.61
SHC2	1.58
PGCP	1.58
TSC22D3	1.57
CCDC146	1.57
JHDM1D	1.56
DUSP1	1.56
TCP11L2	1.56
ENPP5	1.54

GLDN	1.54
LRRC6	1.53
LOC375190	1.53
ST6GALNAC3	1.53
CTSO	1.52
BTC	1.51
CYR61	1.51
CTBS	1.5

Table A.1.33: Genes with decreased expression in bortezomib-resistant clone E

Gene	Decreased Abundance (log 2)
RPS4Y1	-5.75
KRT7	-3.81
EIF1AY	-2.66
HIST1H4A	-2.49
ZBED2	-2.32
TACSTD2	-2.32
DDX3Y	-2.11
CYorf15A	-2.07
C4orf48	-1.95
PALM3	-1.86
ACTA1	-1.73
SDF2L1	-1.68
C2orf82	-1.66
PIF1	-1.64
CBX2	-1.61
HAUS7	-1.6
LIPE	-1.56
IGFBP6	-1.54
GAL	-1.5

Appendix 2

Studies toward the structural analysis of the kinesin-5–PCTP interaction

My initial goal in studying the interaction of PCTP and kinesin-5 was to obtain a crystal structure of the protein—inhibitor complex. The kinesin-5 motor domain has been crystallized by several other groups both in the presence and absence of inhibitors (Kaan et al., 2010; Parke et al., 2010; Roecker et al., 2007; Turner et al., 2001; Yan et al., 2004). However, it had been reported that a group at Merck Research Laboratories was not able to solve the structure of the kinesin-5—PCTP complex (Rickert et al., 2008). As it was unlikely that I would be successful conducting crystallography studies where a large pharmaceutical company had failed, I decided to try another strategy. All published structures of kinesin-5 have been in the presence of nucleotide, either ADP or AMP-PNP (a non-hydrolyzable analog of ATP). Thus, I attempted to crystallize the PCTP—kinesin-5 complex using two separate kinesin-5 samples: kinesin-5 bound to ADP and kinesin-5 in the absence of ADP (the apo-kinesin-5 protein). As PCTP is ATP competitive, I hoped that this dual strategy would be more successful in generating structures of the protein—inhibitor complex.

To obtain apo-kinesin-5 protein, the kinesin-5 protein was purified in the absence of ATP and $MgCl_2$, which are typically included in kinesin purification buffers. The protein was also incubated with relatively high concentrations of

EGTA and EDTA, with the goal of chelating residual Mg ions and destabilizing ADP that was bound to kinesin-5. This procedure, which was based on published results (Cochran and Gilbert, 2005), produced kinesin-5 that, when examined by native mass spectrometry, was consistent with it being the apo form of the protein (**Figure A.2.1**). I have not, however, fully determined whether this apo-kinesin-5 protein is homogeneously lacking nucleotide.

Large quantities of the kinesin-5 motor domain were purified both bound to ADP and with an empty nucleotide-binding pocket. Crystallization trials were conducted with both constructs using published kinesin crystallization conditions and standard sparse-matrix crystallization screens. From previously published kinesin-5 conditions, I obtained imperfect crystals of kinesin-5 with ADP, but not with apo-kinesin-5 (**Figure A.2.2a**). Microseeding from the original crystals was conducted in order to generate single crystals of kinesin-5 with ADP (**Figure A.2.2b**). These crystals diffracted below 2.8 Å on the Rockefeller Structural Biology Resource Center beamline; however, these crystals were never used to solve a structure as the kinesin-5—ADP structure has already been published (Turner et al., 2001). These crystals could potentially be used for soaking with PCTP to determine if it is possible to obtain a structure of the kinesin-5—PCTP complex. No crystals were obtained under any condition with the apo-kinesin-5 protein. Furthermore, when PCTP was added to the crystallization wells with kinesin-5—ADP, crystals did not form. These trials suggest that it will be difficult to obtain a crystal structure of apo-kinesin-5 or kinesin-5 in complex with PCTP.

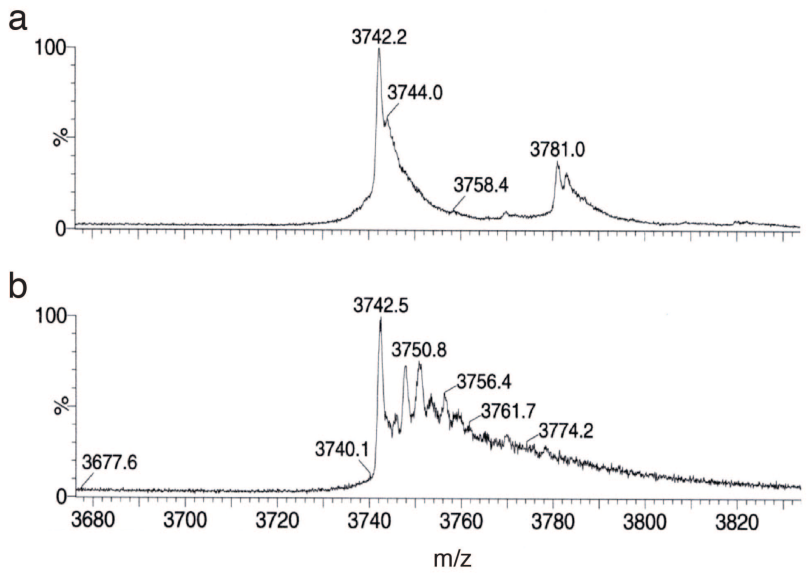


Figure A.2.1 Native mass spectrometry of kinesin-5. (a) Native mass spectrum of kinesin-5 with ADP (z = 11). The 3742.2 peak represents the mass of kinesin-5 alone (41,164 Da), while the 3781.0 peak represents kinesin-5 bound to ADP (41591 Da). **(b)** Native mass spectrum of apo-kinesin-5 (z = 11). The 3742.5 peak represents the mass of kinesin-5 alone (41,167 Da). Native mass spectrometry experiments were conducted by Zachary Quinkert.

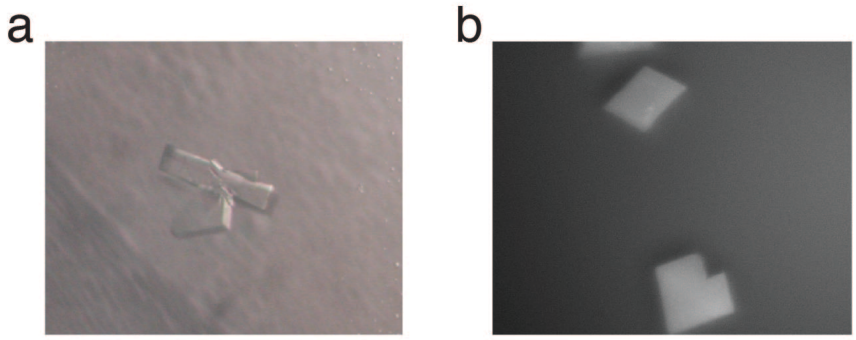


Figure A.2.2 Crystallization of kinesin-5 with ADP. (a) Crystals of kinesin-5 with ADP grew in hanging drops using 1:1 parts well buffer (100 mM MES, pH 5.5 – 6.0, 21 – 24% PEG 3350, and 200 mM $NaNO_3$) and protein (7 – 14 mg/ml) at 4°C. **(b)** Single crystals of kinesin-5 with ADP grew from microseeds of previous crystals in hanging drops using 1:1 parts well buffer (100 mM MES, pH 5.6, 14% PEG 3500, and 200 mM $NaNO_3$) and 3.5 mg/ml protein at 4°C.

Studies toward the structural analysis of the kinesin-5 S269N mutant

In order to better characterize the effects of mutation in the region of kinesin-5 that contains residues Y104 and S269, crystallization of the kinesin-5 S269N motor domain was attempted. The kinesin-5 S269N protein (bound to ATP) formed crystals in similar conditions as the wild type kinesin-5, after seeding with kinesin-5 microseeds (**Figure A.2.3**). Initially kinesin-5 wild type microseeds were used; the wild type protein was then diluted with a second round of microseeding using the kinesin-5 S269N crystals. Kinesin-5 S269N crystals consistently formed clusters that were difficult to break apart.

A full data set from kinesin-5 S269N that diffracted below 2 Å was collected at the National Synchrotron Light Source at the Brookhaven National Laboratories. With the help of Deena Oren at Rockefeller's Structural Biology Resource Center, I used these data to solve the structure of kinesin-5 S269N using molecular replacement of the wild type kinesin-5 structure (Turner et al., 2001). The structure is partially refined and currently has an R_{free} factor of 25.13% (**Table A.2.1**). The electron density map of kinesin-5 S269N is consistent with the protein containing the mutated residue (**Figure A.2.4a**). With minimal further refinement, this structure is suitable to be deposited in the Protein Data Bank. In comparing my structure of kinesin-5 S269N with the published structure of kinesin-5 wild type the main variations are in loops with high B factors that are presumed to have some inherent flexibility (**Figure A.2.4b**). I could not detect any changes that I believed were likely to impact the binding of PCTP.

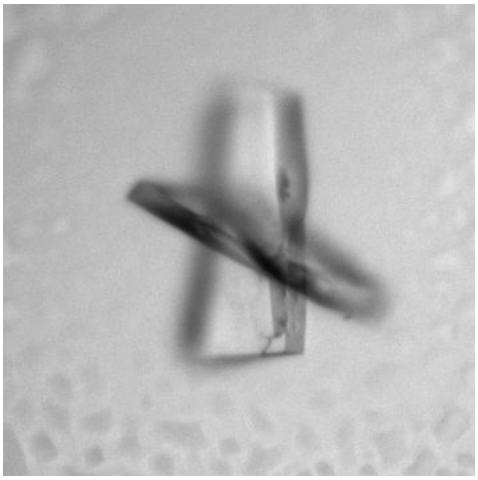


Figure A.2.3 Crystallization of kinesin-5 S269N with ADP. Crystals of kinesin-5 S269N with ADP grew from kinesin-5 microseeds in hanging drops using 1:1 parts well buffer (100 mM MES, pH 5.5, 16 – 20% PEG 3350, and 100 – 300 mM $NaNO_3$) and protein (3.5 – 7.5 mg/ml) at 4°C.

Table A.2.1 Data collection and refinement statistics for kinesin-5 S269N structure. This structure may need further refinement.

Collection	
Synchrotron	NSLS
Beamline	X29A
Space Group	$P2_1$
a, b, c, ß (°)	52.38, 78.22, 92.58, 93.06
Resolution (Å)	50.0 – 1.95 (2.01 – 1.95)
R_{sym} (%)	9.6% (53.6%)
<I/s(I)>	17.5 (3.6)
Observed reflections	125,417
Unique reflections	54,417
Completeness	99.8% (100%)
Redundancy	6.2
Refinement	
r.m.s.d. bond lengths (Å)	.0138
r.m.s.d. bond angles (°)	1.630
R_{factor} (%)	19.70%
R_{free} (%)	25.13%
Water atoms	264

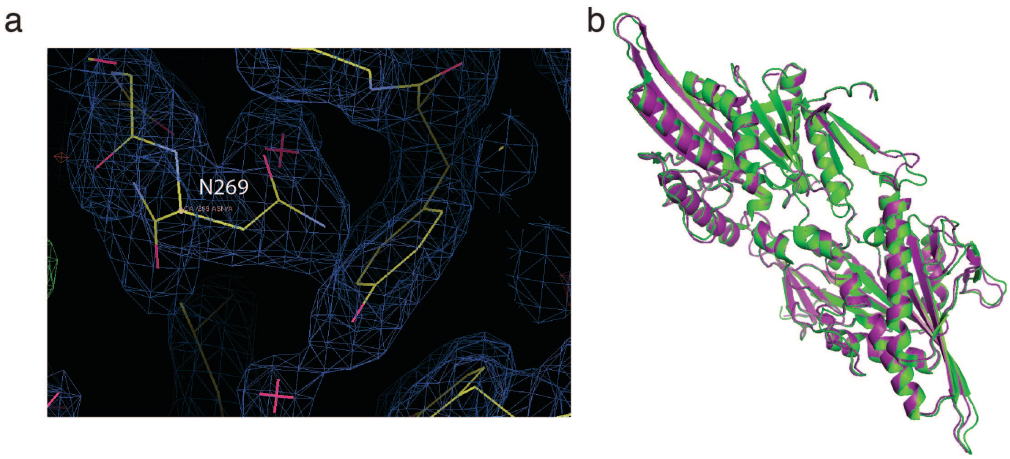


Figure A.2.4 Structure of kinesin-5 S269N. (a) The electron density map of kinesin-5 S269N is consistent with residue 269 being an asparagine. **(b)** An alignment of kinesin-5 wild type (PDB: 1II6, green) (Turner et al., 2001) and kinesin-5 S269N (pink).

Materials and Methods

Cell culture conditions

HCT-116 cells and clonal lines were cultured in McCoy's 5A medium (Invitrogen). hTERT-RPE1 cells were cultured in Dulbecco's Modified Eagle's Medium/F12 1:1 nutrient mix (Invitrogen), while HeLa and 293-Ampho cells were cultured in Dulbecco's Modified Eagle's Medium (Invitrogen). All cultures were supplemented with 10% FBS (Atlanta Biologicals) and penicillin-streptomycin (100 U/ml and 100 ug/ml, respectively, Invitrogen) and grown at 37°C in a humidified chamber with 5% CO_2. hTERT-RPE1, HeLa, and 293-Ampho cells were also supplemented with 1% MEM non-essential amino acids (Invitrogen) and 2 mM l-glutamine (Invitrogen).

Molecular biology and cloning

Human PLK1 (image clone ID# 2822226, Open Biosystems) or human PSMB5 (image clone ID# 4795732, Open Biosystems) was cloned into a pMSCV_puro vector (Clontech) with an N-terminal GFP – PreScission protease site compatible with the Gateway cloning system (Invitrogen). Site-directed mutagenesis to generate the PLK1 R136G and G63S mutations or the PSMB5 M104V and A108T mutations was performed using QuickChange (Stratagene) according to the manufacturer's instructions. DNA encoding the wild type and

mutant proteins was used to generate stable cell lines through retroviral infection. Retroviruses were packaged in 293-Ampho cells. hTERT-RPE1 and HeLa cells were infected by retrovirus with 4 μg/ml polybrene (Sigma) and selected by puromycin (Sigma).

The human kinesin-5 ATPase domain (residues 1-368, Pubmed accession: NP_004514 from Open Biosystems) was cloned into a pET28a vector containing a PreScission protease cleavable N-terminal His-tag. This adds a 'GP' sequence at the N-terminus of the native protein sequence after removal of the His tag by PreScission proteolysis. The L214A, Y104H, and S269N mutant was generated from this plasmid using QuickChange site-directed mutagenesis kit (Stratagene). For *in vitro* experiments, the full-length human Plk1 wild type and G63S mutant proteins were also cloned into pFastbac Htb baculovirus expression vectors (Gibco) with a TEV protease cleavable N-terminal His-tag.

Preparation of fixed cells and imaging

For microscopy, cells were plated on glass coverslips coated with poly-D-lysine (Sigma). The following day, cells were incubated in media containing DMSO or 50 nM BI 2536 for 5 hours in standard culture conditions. Cells on coverslips were fixed in methanol at -20°C for ten minutes and washed in TBS-Tx (TBS with 0.1% Triton-X-100) before they were blocked in AbDil (TBS-Tx with 1% BSA) for one hour at room temperature. Cells were stained with FITC-conjugated anti-tubulin DM1a antibody (Sigma) at a dilution of 1:1500 in AbDil and Hoescht.

Fixed cells were imaged with a Nikon TE2000 microscope (Morrell Instruments) or a DeltaVision Image Restoration Microscope (Applied Precision Instruments) and deconvoluted in SoftWoRx.

Chemical compounds

BI 2536 (> 99% pure) was purchased from Selleck chemicals. Bortezomib (99% pure) was purchased from LC Laboratories. S-trityl-L-cysteine (STLC, 97% pure) and taxol (> 95% pure) were purchased from Sigma. Hesperadin (> 95% pure), 4-(2-(1-phenylcyclopropyl)thiazol-4-yl)pyridine (PCTP, > 95% pure) and PCTP analogs 1 and 2 (both > 95% pure) were synthesized by a lab member, Sudhir Kashyap (Wacker et al., 2011). ARQ-197 (> 98% pure) was purchased from Active Biochemicals. BE-54017 and Cladoniamide A were obtained from Fang-Yuan Chang.

Selection of resistant clones

Resistant clones were generated by plating 0.5 – 1.0 x 10^6 HCT-116 cells into 10 cm culture dishes with media containing 10 nM BI 2536, 1 μM STLC, 8 – 12 μM PCTP, 8 – 12 nM bortezomib, 0.1 – 3 μM hesperidin, 2.5 – 7.5 nM taxol, 750 – 850 nM ARQ-197, 200 – 400 ng/ml BE-54017, or 150 – 200 ng/ml Cladoniamide A. Drug concentrations were determined by either using the LD_{90} or experimentally determining a concentration of drug that killed all cells except a few which would grow into colonies. After cells had adhered, media with

compound was exchanged every three days for three - five weeks. Most cells did not survive, but a few per plate grew into colonies (less than 20 colonies were found on each plate). Colonies were picked by placing a cloning cylinder around each clone and then treating with 50 µl 0.05% trypsin-EDTA for five minutes. Cells were then transferred to a single well of a 24-well plate containing media with compound. As cells became confluent they were transferred to larger culture dishes, always with media containing drug at the same concentration as the selections.

Cell proliferation assays and calculation of LD_{50} values

In order to quantify cell growth in the presence of drug, cells (1000 in 100 µl of media per well) were plated in a flat-bottomed 96-well plate and treated the next day with various concentrations of the appropriate compound, in duplicate. After three days, cell proliferation was determined using a WST1 assay (Millipore) according to the manufacturer's instructions. Normalized cell proliferation was calculated as the change in the number of cells at each concentration compared to control wells. The assay was repeated two - six independent times. Normalized cell proliferation values were plotted and LD_{50} values were obtained by curve fitting with Prism. A two-tailed paired t-test was used to determine statistical significance.

RNA purification and RT-PCR

Total RNA was isolated from cells using the RNeasy Mini Kit (Qiagen) according to the manufacturer's instructions. Full-length Plk1 or kinesin-5 cDNA was synthesized and amplified from total RNA using Plk1 or kinesin-5 specific primers and the SuperScript III One-Step RT-PCR System (Invitrogen).

RNA-seq library construction

Following isolation, total RNA integrity was checked using an Agilent Technologies 2100 Bioanalyzer and an RNA Integrity Number greater than 8 was required for further processing. Library construction was performed according to standard Illumina protocols with Illumina reagents. Briefly, mRNA was purified from total RNA using magnetic beads: 5-100 ng of total RNA were heated to disrupt the secondary structures and then added to pre-prepared Sera-mag Magnetic Oligo(dT) Beads. After washing, 10mM Tris-HCI was added to the beads, the samples were heated, and mRNA was eluted. The mRNA was then fragmented using divalent cations under elevated temperatures. The cleaved RNA fragments were copied into first strand cDNA using reverse transcriptase and random primers. mRNA was removed by RNaseH and a replacement strand was synthesized to generate double-strand cDNA. The overhangs resulting from fragmentation were converted into blunt ends by T4 DNA polymerase and DNA polymerase I Klenow fragment. An ‘A’ base is added to the 3' end of the blunt phosphorylated DNA fragments to prepare them for ligation to the adapters,

which have a single ‘T’ base overhang at their 3' end. Illumina adapters were ligated to the ends of the DNA fragments, preparing them to be hybridized to a flow cell. Ligation reaction products were purified on an agarose gel. A 200 ± 25 bp size-range of templates was selected for downstream enrichment. The cDNA fragments with adapters on both ends were amplified by PCR with primers complementary to the adapters. Size, purity and concentration of the library were checked on an Invitrogen Qubit Fluorometer using the Quant-IT dsDNA HS Assay Kit and on an Agilent Technologies 2100 Bioanalyzer using their High Sensitivity DNA Kit.

High-throughput sequencing

Sequencing was performed using both Illumina GAIIx (for BI 2536 clones) and HiSeq2000 (for bortezomib clones) machines. For GAIIx, the protocols for the Illumina Single-Read Cluster Generation Kit were used for cluster generation on the Cluster Station. The targeted samples were diluted to ten nanomoles and denatured with sodium hydroxide. Ten picomoles of each target-enriched sample and control was loaded into separate lanes of the same flow cell, hybridized onto the flow cell, and isothermally amplified. After linearization, blocking, and primer hybridization, sequencing was performed for 40 cycles on the Illumina 36 Cycle Sequencing Kit v4 with version 7.0 sequencing protocols. Raw image data was converted into base calls using the Illumina pipeline v1.6 with default parameters. Rigorous quality control was performed using data from

reports generated by the Illumina pipeline. For HiSeq2000, a similar protocol was used, and some of the clones were sequenced on the same lane (3-plex). After quantifying and checking the size and purity of the product, multiplexed DNA libraries were normalized to 10 nM and then sample libraries were pooled together in equal volumes. 7 pM of each pooled DNA library template was amplified on an Illumina cBot instrument involving immobilization and 3' extension, bridge amplification, linearization and hybridization, then sequenced on the Illumina HiSeq2000 sequencer using 51 cycles.

Alignment of RNA-seq reads

RNA-seq reads were aligned to RefSeq transcript sequences downloaded from the UCSC Genome Browser in June 2010, using the BWA program with default parameters (Li and Durbin, 2009). Out of 25-100 million reads obtained in each run, 74-83% could be mapped to RefSeq transcripts. Clonal reads, i.e. multiple reads mapping at the same position and same orientation in a transcript, were collapsed into a single read (Chepelev et al., 2009). Following mapping to RefSeq transcripts to identify reads mapping to exons and across known exon junctions, all mapped reads were remapped to the reference human genome using custom programs and based on the June 2010 RefSeq gene annotation.

Overall bioinformatics strategy

The following strategy was used to determine which genetic variants increased their relative abundance in the expanded clones compared to the original cell population. First, single nucleotide variations (SNVs) and insertions/deletions (indels) in the sequenced mRNAs of each clone were identified. Next, the relative abundance of these variants was compared to the relative abundance in the same variants (at the same location) in the original cell population. Only variants whose relative abundance had increased significantly (after correction for multiple hypothesis testing) in the expanded clone were retained. Finally, variants with unlikely functional impact, e.g. synonymous variants and variants in 5'UTRs and 3'UTRs were filtered out. All bioinformatics analyses were conducted by my collaborator, Olivier Elemento. All computer programs and scripts used for the analysis, together with a comprehensive description of how to use them are available at his website: http://icb.med.cornell.edu/wiki/index.php/Elementolab/TargetID

Expression and purification of full-length Plk1

Full-length Plk1 was produced in Sf9 insect cells through a baculovirus expression system. Cells were resuspended in buffer containing 50 mM phosphate (pH 8), 300 mM KCl, 10 mM imidazole, 0.1% Tween-20, 10% glycerol, 5 mM $MgCl_2$, 1 mM ATP, and HALT protease inhibitor cocktail (Pierce), then lysed by sonication. The lysate was clarified by centrifugation and applied to Ni-

NTA resin (Qiagen). Resin was further washed with a buffer containing 50 mM phosphate (pH 8), 250 mM KCl, 20 mM imidazole, 10% glycerol, 1 mM $MgCl_2$, 50 μM ATP, and HALT. Protein was eluted from the Ni-NTA resin using the same buffer as the wash steps but that contained 250 mM imidazole and 150 mM KCl. Samples containing protein were identified with a Bio-Rad protein assay and then combined. The protein sample was combined with AcTEV protease (20 units, Invitrogen) and then dialyzed into Plk1 storage buffer containing Protein was concentrated and dialyzed into buffer containing 10 mM HEPES (pH 7.5), 150 mM KCl, 10 mM $MgCl_2$, 10% sucrose and 500 μM dithiothreitol.

Kinase assay

Plk1 activity was assayed in kinase buffer (20 mM HEPES (pH 7.5), 10 mM $MgCl_2$, 5 mM dithiothreitol, 1 mM EGTA, 10 mM NaF, and 20 mM β-glycerophosphate). Casein (3 μg), recombinant full-length human Plk1, and the appropriate amount of compound in DMSO (5% (v/v) final) were added to half the reaction volume (10 μl) and allowed to incubate for 15 min on ice. Then 10 μl of kinase buffer containing 150 μM [γ-^{32}P]ATP was added. Reactions were mixed, incubated for 20 min at 30°C and stopped by the addition of Laemmli sample buffer and heating at 80°C for ten minutes. Proteins were resolved by SDS-PAGE (4-20% tris-glycine gel). Gels were dried and visualized by autoradiography. Gel bands were quantified with ImageJ software (National Institute of Health).

Expression and purification of kinesin-5

Expression of kinesin-5 motor domain in *Escherichia coli* BL21(DE3) Rosetta cells (Novagen) was induced with 0.5 mM isopropyl-1-thio-β-d-galactopyranoside. After growth at 18°C for 16 – 22 hours, cells were harvested. Cells were resuspended in buffer containing 50 mM phosphate (pH 8), 250 mM KCl, 7.5 mM imidazole, 0.1% Tween-20, 0.5 mM $MgCl_2$, 0.5 mM ATP, 10 mM β-mercaptoethanol (BME), 2 mM benzamidine-HCl, 1 mM phenylmethanesulfonyl fluoride (PMSF), and HALT protease inhibitor cocktail (Pierce), then lysed by sonication. The lysate was clarified by centrifugation and applied to Talon resin (Clontech). Resin was further washed with the lysis buffer before protein was removed from the column by overnight cleavage at 4°C with GST-PreScission protease. After cleavage, the sample was incubated with glutathione resin to remove PreScission protease. Kinesin-5 protein was concentrated and dialyzed into crystallization buffer (50 mM Pipes (pH 6.8), 50 mM KCl, 0.2 mM $MgCl_2$, 0.1 mM ATP, 1 mM EGTA, and 0.1 mM Tris-(2-carboxyethyl)phosphine (TCEP) HCl). Wild type kinesin-5 and kinesin-5 S269N for crystallography studies were further purified by size exclusion chromatography (Superdex 200; Amersham Pharmacia Biotech) before the dialysis step. Samples for ATP hydrolysis assays and crystallography were stored at -80°C until needed.

Apo-kinesin-5 was expressed as above with a few modifications. Cells were resuspended in buffer containing 50 mM phosphate (pH 8), 250 mM KCl, 7.5 mM imidazole, 0.1% Tween-20, 10 mM BME, 2 mM benzamidine-HCl, 1 mM

PMSF, and HALT protease inhibitor cocktail (Pierce). After binding to the TALON resin, subsequent cleavage with PreScission protease, and removal of PreScission protease, the kinesin-5 protein sample was incubated with a final concentration of 5 mM EDTA and 5 mM EGTA for 30 minutes. Afterwards EDTA and EGTA were removed by dialysis into “apo crystallization buffer” (50 mM Pipes (pH 6.8), 50 mM KCl, 1 mM EGTA, and 0.1 mM TCEP-HCl).

Full-length *Xenopus laevis* kinesin-5 used in the microtubule gliding assays was prepared as described previously (Weinger et al., 2011). Bovine tubulin was purified and labeled with X-Rhodamine according to published protocols (Hyman, 1991).

Steady-state ATP hydrolysis and microtubule motility assays

For steady-state ATP hydrolysis assays and crosslinking experiments, microtubules were polymerized from recycled tubulin in BRB80 buffer (80mM Pipes (pH 6.8), 1 mM $MgCl_2$, 1mM EGTA) containing 10% DMSO, 1 mM dithiothreitol, and 1 mM GTP at 37°C for 25 min. Microtubules were pelleted over a 40% glycerol cushion and resuspended in BRB80 with 20 μM taxol.

ATP hydrolysis experiments used an assay that coupled ATP hydrolysis to oxidation of NADH (De La Cruz et al., 2000). Reactions were performed in BRB80 (reactions with microtubules) or 200 mM potassium phosphate (pH 8, reactions without microtubules) along with 1 mM dithiothreitol, 0.1 mg/ml bovine serum albumin (Fisher Scientific), 2.5% DMSO, 1 mM MgATP, 2.5 mM reduced

β-Nicotinamide adenine dinucleotide (disodium salt, Sigma), 200 U/ml L-lactic dehydrogenase (from rabbit muscle, Sigma), 1000 U/ml pyruvate kinase (from rabbit muscle, Sigma), 5 mM Phospho(enol)pyruvic acid (monopotassium salt, Sigma), and 400 nM microtubules. Kinesin-5 motor domain was added to begin reactions, which were conducted in 50 µl volumes in 384-well plates. Time courses of fluorescence decrease upon NADH oxidation were obtained using a FlexStation microplate reader (λ_{ex} = 340 nm, 440 nm emission filter). Normalized specific activity values were plotted and IC_{50} values were obtained by curve fitting with either Kaleidagraph or Prism.

For microtubule gliding assays, X-Rhodamine labeled microtubules were polymerized from GMPPCP seeds in BRB80 and stabilized with 10 µM taxol. Microtubule gliding assays were performed using TIRF microscopy on an inverted microscope (Axiovert; Carl Zeiss) (Kapitein et al., 2008) and conducted as described previously (Kwok et al., 2004).

Photo-crosslinking reactions

PCTP analogs **1** and **2** (40x final, in DMSO) were combined with kinesin-5 ATPase domain (1.5 µM) in 50 mM HEPES (pH 7, no microtubules) or in BRB80 with 3.3 µM microtubules (polymerized as in the ATP hydrolysis assays). Samples were irradiated at 254 nm using a Spectroline ENF 260C UV lamp for 30 min at 4°C (no microtubules) or room temperature (with microtubules). Reactions were stopped by the addition of dithiothreitol (12.5 mM final) and gel

loading buffer. For competition experiments, **1** was pre-mixed with PCTP or STLC in DMSO to obtain a 40x solution of each compound.

Sample preparation for mass spectrometry

Crosslinked samples, combined with gel loading buffer and dithiothreitol, were briefly heated (10 min, 70°C), and then treated with iodoacetamide (125 mM) for 30 min to alkylate all reduced cysteines. Proteins were separated and immobilized on Bis-Tris gels (Invitrogen), followed by fixation in a 50% methanol/7% acetic acid solution. Protein bands were visualized by GelCode Blue stain (Pierce). Kinesin-5 bands were excised from the gel, sliced, and destained in 50 mM ammonium bicarbonate/50% acetonitrile for 30 min. Gel slices were then dehydrated in acetonitrile (HPLC-grade, Pierce) for 10 min. The dried gel slices were rehydrated in 25 mM ammonium bicarbonate with 125 ng trypsin (Promega) for protein digestion at 37°C overnight. The resulting peptides were enriched with StageTips (Rappsilber et al., 2007). The peptides eluted from the StageTips were dried down by SpeedVac and then resuspended in 0.5% acetic acid for analysis by LC-MS/MS.

Mass spectrometry

Mass spectrometry was performed by Xiang Li on an LTQ-Orbitrap XL mass spectrometer (Thermo Fisher Scientific), using a home-built micro electrospray source with a liquid junction. First, peptide samples in 0.5% acetic

acid were pressure loaded onto a self-packed PicoFrit column (New Objective) with integrated emitter tip (360-mm o.d., 75-mm i.d., 15-mm tip), packed with 6 cm of reverse-phase C_{18} material (Alltima C_{18} 5-mm beads from Alltech Associates), rinsed for 10 min with 0.1 M acetic acid and subsequently gradient eluted with a linear gradient from 0 to 100% B in 30-50 min (A = 0.1 M acetic acid, B = 70% acetonitrile in 0.1M acetic acid, flow rate 200 nL/min) into the mass spectrometer. The instrument was operated in a data dependent mode cycling through a full scan (300–2,000 *m*/*z*, single mscan) followed by 7 CID MS/MS scans on the 7 most abundant ions from the immediate preceding full scan. The cations were isolated with a 2-Da mass window and set on a dynamic exclusion list for 60 seconds after they were first selected for MS/MS. The raw data were analyzed by Quant module of MaxQuant (Cox and Mann, 2008; Cox et al., 2009) to find peptide pairs with a 4 Da mass difference.

Crystallography of kinesin-5

Kinesin-5 (apo, wild type with ADP, or S269N with ADP) was concentrated to 14 – 25 mg/ml in crystallization buffer or “apo crystallization buffer.” Kinesin-5 was crystallized by hanging drop vapor diffusion at 4°C using an equal volume of protein sample (protein was diluted to 3.5 – 14 mg/ml) and crystallization solution consisting of 0.1 M MES (pH 5.5 – 6.0), 14 – 24% PEG 3350, and 100 – 300 mM $NaNO_3$. Crystals appeared in 1 – 2 weeks. Microseeding was used to obtain cleaner crystals. For data collection, crystals were flash-frozen in liquid nitrogen

after a brief soak in a solution comprising of 0.1 M MES (pH 5.6), 22% PEG 3350, 200 mM $NaNO_3$, and 15% glycerol.

Diffractions datasets of kinesin-5 S269N crystals were collected at the National Synchrotron Light Source (Brookhaven National Laboratory) beamline X29a. Diffraction data was processed using HKL2000 and phase information was obtained through molecular replacement with the program Phaser using the structure of wild type kinesin-5 (PDB: 1II6) (Turner et al., 2001). Refinement was carried out using Refmac5, resulting in the partially refined statistics in **Table A.2.1**.

References

Al-Mawsawi, L.Q., Fikkert, V., Dayam, R., Witvrouw, M., Burke, T.R., Jr., Borchers, C.H., and Neamati, N. (2006). Discovery of a small-molecule HIV-1 integrase inhibitor-binding site. Proc Natl Acad Sci U S A *103*, 10080-10085.

Arnoldo, A., Curak, J., Kittanakom, S., Chevelev, I., Lee, V.T., Sahebol-Amri, M., Koscik, B., Ljuma, L., Roy, P.J., Bedalov, A., Giaever, G., Nislow, C., Merrill, A.R., Lory, S., and Stagljar, I. (2008). Identification of small molecule inhibitors of Pseudomonas aeruginosa exoenzyme S using a yeast phenotypic screen. PLoS Genet *4*, e1000005.

Azam, M., Latek, R.R., and Daley, G.Q. (2003). Mechanisms of autoinhibition and STI-571/imatinib resistance revealed by mutagenesis of BCR-ABL. Cell *112*, 831-843.

Berkers, C.R., Verdoes, M., Lichtman, E., Fiebiger, E., Kessler, B.M., Anderson, K.C., Ploegh, H.L., Ovaa, H., and Galardy, P.J. (2005). Activity probe for in vivo profiling of the specificity of proteasome inhibitor bortezomib. Nat Methods *2*, 357-362.

Blundell, T.L., and Patel, S. (2004). High-throughput X-ray crystallography for drug discovery. Curr Opin Pharmacol *4*, 490-496.

Brier, S., Lemaire, D., DeBonis, S., Forest, E., and Kozielski, F. (2006). Molecular dissection of the inhibitor binding pocket of mitotic kinesin Eg5 reveals mutants that confer resistance to antimitotic agents. J Mol Biol *360*, 360-376.

Brummelkamp, T.R., Fabius, A.W., Mullenders, J., Madiredjo, M., Velds, A., Kerkhoven, R.M., Bernards, R., and Beijersbergen, R.L. (2006). An shRNA barcode screen provides insight into cancer cell vulnerability to MDM2 inhibitors. Nat Chem Biol *2*, 202-206.

Burdine, L., and Kodadek, T. (2004). Target identification in chemical genetics: the (often) missing link. Chem Biol *11*, 593-597.

Burns, A.R., Kwok, T.C., Howard, A., Houston, E., Johanson, K., Chan, A., Cutler, S.R., McCourt, P., and Roy, P.J. (2006). High-throughput screening of small molecules for bioactivity and target identification in Caenorhabditis elegans. Nat Protoc *1*, 1906-1914.

Butcher, R.A., Bhullar, B.S., Perlstein, E.O., Marsischky, G., LaBaer, J., and Schreiber, S.L. (2006). Microarray-based method for monitoring yeast overexpression strains reveals small-molecule targets in TOR pathway. Nat Chem Biol *2*, 103-109.

Chan, J.N., Nislow, C., and Emili, A. (2009). Recent advances and method development for drug target identification. Trends Pharmacol Sci *31*, 82-88.

Chen, D., Frezza, M., Schmitt, S., Kanwar, J., and Q, P.D. (2011). Bortezomib as the first proteasome inhibitor anticancer drug: current status and future perspectives. Curr Cancer Drug Targets *11*, 239-253.

Chen, J.K., Taipale, J., Cooper, M.K., and Beachy, P.A. (2002). Inhibition of Hedgehog signaling by direct binding of cyclopamine to Smoothened. Genes Dev *16*, 2743-2748.

Chepelev, I., Wei, G., Tang, Q., and Zhao, K. (2009). Detection of single nucleotide variations in expressed exons of the human genome using RNA-Seq. Nucleic Acids Res *37*, e106.

Chidley, C., Haruki, H., Pedersen, M.G., Muller, E., and Johnsson, K. (2011). A yeast-based screen reveals that sulfasalazine inhibits tetrahydrobiopterin biosynthesis. Nat Chem Biol *7*, 375-383.

Chung, C.W. (2007). The use of biophysical methods increases success in obtaining liganded crystal structures. Acta Crystallogr D Biol Crystallogr *63*, 62-71.

Cochran, J.C., and Gilbert, S.P. (2005). ATPase mechanism of Eg5 in the absence of microtubules: insight into microtubule activation and allosteric inhibition by monastrol. Biochemistry *44*, 16633-16648.

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol *26*, 1367-1372.

Cox, J., Matic, I., Hilger, M., Nagaraj, N., Selbach, M., Olsen, J.V., and Mann, M. (2009). A practical guide to the MaxQuant computational platform for SILAC-based quantitative proteomics. Nat Protoc *4*, 698-705.

Cravatt, B.F., Wright, A.T., and Kozarich, J.W. (2008). Activity-based protein profiling: from enzyme chemistry to proteomic chemistry. Annu Rev Biochem *77*, 383-414.

Cuatrecasas, P., Wilchek, M., and Anfinsen, C.B. (1968). Selective enzyme purification by affinity chromatography. Proc Natl Acad Sci U S A *61*, 636-643.

De La Cruz, E.M., Sweeney, H.L., and Ostap, E.M. (2000). ADP inhibition of myosin V ATPase activity. Biophys J *79*, 1524-1529.

DeBonis, S., Skoufias, D.A., Lebeau, L., Lopez, R., Robin, G., Margolis, R.L., Wade, R.H., and Kozielski, F. (2004). In vitro screening for inhibitors of the human mitotic kinesin Eg5 with antimitotic and antitumor activities. Mol Cancer Ther *3*, 1079-1090.

Duhl, D.M., and Renhowe, P.A. (2005). Inhibitors of kinesin motor proteins--research and clinical progress. Curr Opin Drug Discov Devel *8*, 431-436.

Feng, Y., Mitchison, T.J., Bender, A., Young, D.W., and Tallarico, J.A. (2009). Multi-parameter phenotypic profiling: using cellular effects to characterize small-molecule compounds. Nat Rev Drug Discov *8*, 567-578.

Giaever, G., Shoemaker, D.D., Jones, T.W., Liang, H., Winzeler, E.A., Astromoff, A., and Davis, R.W. (1999). Genomic profiling of drug sensitivities via induced haploinsufficiency. Nat Genet *21*, 278-283.

Girdler, F., Sessa, F., Patercoli, S., Villa, F., Musacchio, A., and Taylor, S. (2008). Molecular basis of drug resistance in aurora kinases. Chem Biol *15*, 552-562.

Glaab, W.E., and Tindall, K.R. (1997). Mutation rate at the hprt locus in human cancer cell lines with specific mismatch repair-gene defects. Carcinogenesis *18*, 1-8.

Gorre, M.E., Mohammed, M., Ellwood, K., Hsu, N., Paquette, R., Rao, P.N., and Sawyers, C.L. (2001). Clinical resistance to STI-571 cancer therapy caused by BCR-ABL gene mutation or amplification. Science *293*, 876-880.

Groen, A.C., Needleman, D., Brangwynne, C., Gradinaru, C., Fowler, B., Mazitschek, R., and Mitchison, T.J. (2008). A novel small-molecule inhibitor reveals a possible role of kinesin-5 in anastral spindle-pole assembly. J Cell Sci *121*, 2293-2300.

Groll, M., Berkers, C.R., Ploegh, H.L., and Ovaa, H. (2006). Crystal structure of the boronic acid-based proteasome inhibitor bortezomib in complex with the yeast 20S proteasome. Structure *14*, 451-456.

Gunther, E.C., Stone, D.J., Gerwien, R.W., Bento, P., and Heyes, M.P. (2003). Prediction of clinical drug efficacy by classification of drug-induced genomic expression profiles in vitro. Proc Natl Acad Sci U S A *100*, 9608-9613.

Handschumacher, R.E., Harding, M.W., Rice, J., Drugge, R.J., and Speicher, D.W. (1984). Cyclophilin: a specific cytosolic binding protein for cyclosporin A. Science *226*, 544-547.

Harding, M.W., Galat, A., Uehling, D.E., and Schreiber, S.L. (1989). A receptor for the immunosuppressant FK506 is a cis-trans peptidyl-prolyl isomerase. Nature *341*, 758-760.

Hartwell, L.H., Szankasi, P., Roberts, C.J., Murray, A.W., and Friend, S.H. (1997). Integrating genetic approaches into the discovery of anticancer drugs. Science *278*, 1064-1068.

Hassell, A.M., An, G., Bledsoe, R.K., Bynum, J.M., Carter, H.L., 3rd, Deng, S.J., Gampe, R.T., Grisard, T.E., Madauss, K.P., Nolte, R.T., Rocque, W.J., Wang, L., Weaver, K.L., Williams, S.P., Wisely, G.B., Xu, R., and Shewchuk, L.M. (2007). Crystallization of protein-ligand complexes. Acta Crystallogr D Biol Crystallogr *63*, 72-79.

Hauf, S., Cole, R.W., LaTerra, S., Zimmer, C., Schnapp, G., Walter, R., Heckel, A., van Meel, J., Rieder, C.L., and Peters, J.M. (2003). The small molecule Hesperadin reveals a role for Aurora B in correcting kinetochore-microtubule attachment and in maintaining the spindle assembly checkpoint. J Cell Biol *161*, 281-294.

Heitman, J., Movva, N.R., and Hall, M.N. (1991). Targets for cell cycle arrest by the immunosuppressant rapamycin in yeast. Science *253*, 905-909.

Hermanson, G.T. (2008). Bioconjugate Techniques, 2nd Edition (Academic Press, Inc).

Hieronymus, H., Lamb, J., Ross, K.N., Peng, X.P., Clement, C., Rodina, A., Nieto, M., Du, J., Stegmaier, K., Raj, S.M., Maloney, K.N., Clardy, J., Hahn, W.C., Chiosis, G., and Golub, T.R. (2006). Gene expression signature-based chemical genomic prediction identifies a novel class of HSP90 pathway modulators. Cancer Cell *10*, 321-330.

Ho, C.H., Magtanong, L., Barker, S.L., Gresham, D., Nishimura, S., Natarajan, P., Koh, J.L., Porter, J., Gray, C.A., Andersen, R.J., Giaever, G., Nislow, C., Andrews, B., Botstein, D., Graham, T.R., Yoshida, M., and Boone, C. (2009). A molecular barcoded yeast ORF library enables mode-of-action analysis of bioactive compounds. Nat Biotechnol *27*, 369-377.

Ho, C.H., Piotrowski, J., Dixon, S.J., Baryshnikova, A., Costanzo, M., and Boone, C. (2011). Combining functional genomics and chemical biology to identify targets of bioactive compounds. Curr Opin Chem Biol *15*, 66-78.

Hoon, S., Smith, A.M., Wallace, I.M., Suresh, S., Miranda, M., Fung, E., Proctor, M., Shokat, K.M., Zhang, C., Davis, R.W., Giaever, G., St Onge, R.P., and Nislow, C. (2008). An integrated platform of genomic assays reveals small-molecule bioactivities. Nat Chem Biol *4*, 498-506.

Horwitz, S.B. (1994). Taxol (paclitaxel): mechanisms of action. Ann Oncol *5 Suppl 6*, S3-6.

Horwitz, S.B., Cohen, D., Rao, S., Ringel, I., Shen, H.J., and Yang, C.P. (1993). Taxol: mechanisms of action and resistance. J Natl Cancer Inst Monogr, 55-61.

Huang, S.M., Mishina, Y.M., Liu, S., Cheung, A., Stegmeier, F., Michaud, G.A., Charlat, O., Wiellette, E., Zhang, Y., Wiessner, S., Hild, M., Shi, X., Wilson, C.J., Mickanin, C., Myer, V., Fazal, A., Tomlinson, R., Serluca, F., Shao, W., Cheng, H., Shultz, M., Rau, C., Schirle, M., Schlegl, J., Ghidelli, S., Fawell, S., Lu, C., Curtis, D., Kirschner, M.W., Lengauer, C., Finan, P.M., Tallarico, J.A., Bouwmeester, T., Porter, J.A., Bauer, A., and Cong, F. (2009). Tankyrase inhibition stabilizes axin and antagonizes Wnt signalling. Nature *461*, 614-620.

Hughes, T.R., Marton, M.J., Jones, A.R., Roberts, C.J., Stoughton, R., Armour, C.D., Bennett, H.A., Coffey, E., Dai, H., He, Y.D., Kidd, M.J., King, A.M., Meyer, M.R., Slade, D., Lum, P.Y., Stepaniants, S.B., Shoemaker, D.D., Gachotte, D., Chakraburtty, K., Simon, J., Bard, M., and Friend, S.H. (2000). Functional discovery via a compendium of expression profiles. Cell *102*, 109-126.

Huszar, D., Theoclitou, M.E., Skolnik, J., and Herbst, R. (2009). Kinesin motor proteins as targets for cancer therapy. Cancer Metastasis Rev *28*, 197-208.

Hyman, A.A. (1991). Preparation of marked microtubules for the assay of the polarity of microtubule-based motors by fluorescence. J Cell Sci Suppl *14*, 125-127.

Johnson, D.S., Mortazavi, A., Myers, R.M., and Wold, B. (2007). Genome-wide mapping of in vivo protein-DNA interactions. Science *316*, 1497-1502.

Kaan, H.Y., Ulaganathan, V., Hackney, D.D., and Kozielski, F. (2010). An allosteric transition trapped in an intermediate state of a new kinesin-inhibitor complex. Biochem J *425*, 55-60.

Kaletta, T., and Hengartner, M.O. (2006). Finding function in novel targets: C. elegans as a model organism. Nat Rev Drug Discov *5*, 387-398.

Kapitein, L.C., Kwok, B.H., Weinger, J.S., Schmidt, C.F., Kapoor, T.M., and Peterman, E.J. (2008). Microtubule cross-linking triggers the directional motility of kinesin-5. J Cell Biol *182*, 421-428.

Kelleher, N.L., Nicewonger, R.B., Begley, T.P., and McLafferty, F.W. (1997). Identification of modification sites in large biomolecules by stable isotope labeling and tandem high resolution mass spectrometry. The active site nucleophile of thiaminase I. J Biol Chem *272*, 32215-32220.

Kolaczkowski, M., Kolaczowska, A., Luczynski, J., Witek, S., and Goffeau, A. (1998). In vivo characterization of the drug resistance profile of the major ABC transporters and other components of the yeast pleiotropic drug resistance network. Microb Drug Resist *4*, 143-158.

Kothe, M., Kohls, D., Low, S., Coli, R., Rennie, G.R., Feru, F., Kuhn, C., and Ding, Y.H. (2007). Selectivity-determining residues in Plk1. Chem Biol Drug Des *70*, 540-546.

Kwok, B.H., Yang, J.G., and Kapoor, T.M. (2004). The rate of bipolar spindle assembly depends on the microtubule-gliding velocity of the mitotic kinesin Eg5. Curr Biol *14*, 1783-1788.

Laird, P.W. (2003). The power and the promise of DNA methylation markers. Nat Rev Cancer *3*, 253-266.

Lamb, J. (2007). The Connectivity Map: a new tool for biomedical research. Nat Rev Cancer *7*, 54-60.

Lamb, J., Crawford, E.D., Peck, D., Modell, J.W., Blat, I.C., Wrobel, M.J., Lerner, J., Brunet, J.P., Subramanian, A., Ross, K.N., Reich, M., Hieronymus, H., Wei, G., Armstrong, S.A., Haggarty, S.J., Clemons, P.A., Wei, R., Carr, S.A., Lander, E.S., and Golub, T.R. (2006). The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. Science *313*, 1929-1935.

Lamos, S.M., Krusemark, C.J., McGee, C.J., Scalf, M., Smith, L.M., and Belshaw, P.J. (2006). Mixed isotope photoaffinity reagents for identification of small-molecule targets by mass spectrometry. Angew Chem Int Ed Engl *45*, 4329-4333.

Lampson, M.A., and Kapoor, T.M. (2006). Unraveling cell division mechanisms with small-molecule inhibitors. Nat Chem Biol *2*, 19-27.

Laskowski, R.A., Gerick, F., and Thornton, J.M. (2009). The structural basis of allosteric regulation in proteins. FEBS Lett *583*, 1692-1698.

Lehar, J., Stockwell, B.R., Giaever, G., and Nislow, C. (2008). Combination chemical genetics. Nat Chem Biol *4*, 674-681.

Lenart, P., Petronczki, M., Steegmaier, M., Di Fiore, B., Lipp, J.J., Hoffmann, M., Rettig, W.J., Kraut, N., and Peters, J.M. (2007). The small-molecule inhibitor BI 2536 reveals novel insights into mitotic roles of polo-like kinase 1. Curr Biol *17*, 304-315.

Leung, D., Hardouin, C., Boger, D.L., and Cravatt, B.F. (2003). Discovering potent and selective reversible inhibitors of enzymes in complex proteomes. Nat Biotechnol *21*, 687-691.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754-1760.

Licitra, E.J., and Liu, J.O. (1996). A three-hybrid system for detecting small ligand-protein receptor interactions. Proc Natl Acad Sci U S A *93*, 12817-12821.

Liu, Y., Patricelli, M.P., and Cravatt, B.F. (1999). Activity-based protein profiling: the serine hydrolases. Proc Natl Acad Sci U S A *96*, 14694-14699.

Lu, S., Yang, J., Song, X., Gong, S., Zhou, H., Guo, L., Song, N., Bao, X., Chen, P., and Wang, J. (2008). Point mutation of the proteasome beta5 subunit gene is an important mechanism of bortezomib resistance in bortezomib-selected variants of Jurkat T cell lymphoblastic lymphoma/leukemia line. J Pharmacol Exp Ther *326*, 423-431.

Luesch, H., Wu, T.Y., Ren, P., Gray, N.S., Schultz, P.G., and Supek, F. (2005). A genome-wide overexpression screen in yeast for small-molecule target identification. Chem Biol *12*, 55-63.

Luo, B., Cheung, H.W., Subramanian, A., Sharifnia, T., Okamoto, M., Yang, X., Hinkle, G., Boehm, J.S., Beroukhim, R., Weir, B.A., Mermel, C., Barbie, D.A., Awad, T., Zhou, X., Nguyen, T., Piqani, B., Li, C., Golub, T.R., Meyerson, M., Hacohen, N., Hahn, W.C., Lander, E.S., Sabatini, D.M., and Root, D.E. (2008). Highly parallel identification of essential genes in cancer cells. Proc Natl Acad Sci U S A *105*, 20380-20385.

Luo, L., Parrish, C.A., Nevins, N., McNulty, D.E., Chaudhari, A.M., Carson, J.D., Sudakin, V., Shaw, A.N., Lehr, R., Zhao, H., Sweitzer, S., Lad, L., Wood, K.W., Sakowicz, R., Annan, R.S., Huang, P.S., Jackson, J.R., Dhanak, D., Copeland, R.A., and Auger, K.R. (2007). ATP-competitive inhibitors of the mitotic kinesin KSP that function via an allosteric mechanism. Nat Chem Biol *3*, 722-726.

Maliga, Z., Kapoor, T.M., and Mitchison, T.J. (2002). Evidence that monastrol is an allosteric inhibitor of the mitotic kinesin Eg5. Chem Biol *9*, 989-996.

Mayer, T.U., Kapoor, T.M., Haggarty, S.J., King, R.W., Schreiber, S.L., and Mitchison, T.J. (1999). Small molecule inhibitor of mitotic spindle bipolarity identified in a phenotype-based screen. Science *286*, 971-974.

Miki, H., Okada, Y., and Hirokawa, N. (2005). Analysis of the kinesin superfamily: insights into structure and function. Trends Cell Biol *15*, 467-476.

Mobley, D.L., and Dill, K.A. (2009). Binding of small-molecule ligands to proteins: "what you see" is not always "what you get". Structure *17*, 489-498.

Munshi, N., Jeay, S., Li, Y., Chen, C.R., France, D.S., Ashwell, M.A., Hill, J., Moussa, M.M., Leggett, D.S., and Li, C.J. (2010). ARQ 197, a novel and selective inhibitor of the human c-Met receptor tyrosine kinase with antitumor activity. Mol Cancer Ther *9*, 1544-1553.

Nishimura, S., Arita, Y., Honda, M., Iwamoto, K., Matsuyama, A., Shirai, A., Kawasaki, H., Kakeya, H., Kobayashi, T., Matsunaga, S., and Yoshida, M. (2010). Marine antifungal theonellamides target 3beta-hydroxysterol to activate Rho1 signaling. Nat Chem Biol *6*, 519-526.

Oda, Y., Owa, T., Sato, T., Boucher, B., Daniels, S., Yamanaka, H., Shinohara, Y., Yokoi, A., Kuromitsu, J., and Nagasu, T. (2003). Quantitative chemical proteomics for identifying candidate drug targets. Anal Chem *75*, 2159-2165.

Oerlemans, R., Franke, N.E., Assaraf, Y.G., Cloos, J., van Zantwijk, I., Berkers, C.R., Scheffer, G.L., Debipersad, K., Vojtekova, K., Lemos, C., van der Heijden, J.W., Ylstra, B., Peters, G.J., Kaspers, G.L., Dijkmans, B.A., Scheper, R.J., and Jansen, G. (2008). Molecular basis of bortezomib resistance: proteasome subunit beta5 (PSMB5) gene mutation and overexpression of PSMB5 protein. Blood *112*, 2489-2499.

Ong, S.E., Schenone, M., Margolin, A.A., Li, X., Do, K., Doud, M.K., Mani, D.R., Kuai, L., Wang, X., Wood, J.L., Tolliday, N.J., Koehler, A.N., Marcaurelle, L.A., Golub, T.R., Gould, R.J., Schreiber, S.L., and Carr, S.A. (2009). Identifying the proteins to which small-molecule probes and drugs bind in cells. Proc Natl Acad Sci U S A *106*, 4617-4622.

Parke, C.L., Wojcik, E.J., Kim, S., and Worthylake, D.K. (2010). ATP hydrolysis in Eg5 kinesin involves a catalytic two-water mechanism. J Biol Chem *285*, 5859-5867.

Pathania, R., Zlitni, S., Barker, C., Das, R., Gerritsma, D.A., Lebert, J., Awuah, E., Melacini, G., Capretta, F.A., and Brown, E.D. (2009). Chemical genomics in Escherichia coli identifies an inhibitor of bacterial lipoprotein targeting. Nat Chem Biol *5*, 849-856.

Pellecchia, M., Bertini, I., Cowburn, D., Dalvit, C., Giralt, E., Jahnke, W., James, T.L., Homans, S.W., Kessler, H., Luchinat, C., Meyer, B., Oschkinat, H., Peng, J., Schwalbe, H., and Siegal, G. (2008). Perspectives on NMR in drug discovery: a technique comes of age. Nat Rev Drug Discov *7*, 738-745.

Perlman, Z.E., Slack, M.D., Feng, Y., Mitchison, T.J., Wu, L.F., and Altschuler, S.J. (2004). Multidimensional drug profiling by automated microscopy. Science *306*, 1194-1198.

Peterson, J.R., and Mitchison, T.J. (2002). Small molecules, big impact: a history of chemical inhibitors and the cytoskeleton. Chem Biol *9*, 1275-1285.

Rappsilber, J., Mann, M., and Ishihama, Y. (2007). Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. Nat Protoc *2*, 1896-1906.

Reynolds, P.E., Shepherd, S.T., and Chase, H.A. (1978). Identification of the binding protein which may be the target of penicillin action in Bacillus megaterium. Nature *271*, 568-570.

Rickert, K.W., Schaber, M., Torrent, M., Neilson, L.A., Tasber, E.S., Garbaccio, R., Coleman, P.J., Harvey, D., Zhang, Y., Yang, Y., Marshall, G., Lee, L., Walsh, E.S., Hamilton, K., and Buser, C.A. (2008). Discovery and biochemical

characterization of selective ATP competitive inhibitors of the human mitotic kinesin KSP. Arch Biochem Biophys *469*, 220-231.

Rix, U., and Superti-Furga, G. (2009). Target profiling of small molecules by chemical proteomics. Nat Chem Biol *5*, 616-624.

Robinette, D., Neamati, N., Tomer, K.B., and Borchers, C.H. (2006). Photoaffinity labeling combined with mass spectrometric approaches as a tool for structural proteomics. Expert Rev Proteomics *3*, 399-408.

Roecker, A.J., Coleman, P.J., Mercer, S.P., Schreier, J.D., Buser, C.A., Walsh, E.S., Hamilton, K., Lobell, R.B., Tao, W., Diehl, R.E., South, V.J., Davide, J.P., Kohl, N.E., Yan, Y., Kuo, L.C., Li, C., Fernandez-Metzler, C., Mahan, E.A., Prueksaritanont, T., and Hartman, G.D. (2007). Kinesin spindle protein (KSP) inhibitors. Part 8: Design and synthesis of 1,4-diaryl-4,5-dihydropyrazoles as potent inhibitors of the mitotic kinesin KSP. Bioorg Med Chem Lett *17*, 5677-5682.

Salisbury, C.M., and Cravatt, B.F. (2007). Activity-based probes for proteomic profiling of histone deacetylase complexes. Proc Natl Acad Sci U S A *104*, 1171-1176.

Scutt, P.J., Chu, M.L., Sloane, D.A., Cherry, M., Bignell, C.R., Williams, D.H., and Eyers, P.A. (2009). Discovery and exploitation of inhibitor-resistant aurora and polo kinase mutants for the analysis of mitotic networks. J Biol Chem *284*, 15880-15893.

Seiffert, D., Bradley, J.D., Rominger, C.M., Rominger, D.H., Yang, F., Meredith, J.E., Jr., Wang, Q., Roach, A.H., Thompson, L.A., Spitz, S.M., Higaki, J.N., Prakash, S.R., Combs, A.P., Copeland, R.A., Arneric, S.P., Hartig, P.R., Robertson, D.W., Cordell, B., Stern, A.M., Olson, R.E., and Zaczek, R. (2000). Presenilin-1 and -2 are molecular targets for gamma-secretase inhibitors. J Biol Chem *275*, 34086-34091.

Sharp, D.J., Rogers, G.C., and Scholey, J.M. (2000). Microtubule motors in mitosis. Nature *407*, 41-47.

Shorr, R.G., Heald, S.L., Jeffs, P.W., Lavin, T.N., Strohsacker, M.W., Lefkowitz, R.J., and Caron, M.G. (1982). The beta-adrenergic receptor: rapid purification

and covalent labeling by photoaffinity crosslinking. Proc Natl Acad Sci U S A *79*, 2778-2782.

Sin, N., Meng, L., Wang, M.Q., Wen, J.J., Bornmann, W.G., and Crews, C.M. (1997). The anti-angiogenic agent fumagillin covalently binds and inhibits the methionine aminopeptidase, MetAP-2. Proc Natl Acad Sci U S A *94*, 6099-6103.

Skoufias, D.A., DeBonis, S., Saoudi, Y., Lebeau, L., Crevel, I., Cross, R., Wade, R.H., Hackney, D., and Kozielski, F. (2006). S-trityl-L-cysteine is a reversible, tight binding inhibitor of the human kinesin Eg5 that specifically blocks mitotic progression. J Biol Chem *281*, 17559-17569.

Sleno, L., and Emili, A. (2008). Proteomic methods for drug target discovery. Curr Opin Chem Biol *12*, 46-54.

Taunton, J., Hassig, C.A., and Schreiber, S.L. (1996). A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p. Science *272*, 408-411.

Teraishi, F., Wu, S., Zhang, L., Guo, W., Davis, J.J., Dong, F., and Fang, B. (2005). Identification of a novel synthetic thiazolidin compound capable of inducing c-Jun NH2-terminal kinase-dependent apoptosis in human colon cancer cells. Cancer Res *65*, 6380-6387.

Turner, J., Anderson, R., Guo, J., Beraud, C., Fletterick, R., and Sakowicz, R. (2001). Crystal structure of the mitotic spindle kinesin Eg5 reveals a novel conformation of the neck-linker. J Biol Chem *276*, 25496-25502.

Vale, R.D., and Milligan, R.A. (2000). The way things move: looking under the hood of molecular motor proteins. Science *288*, 88-95.

Vale, R.D., Reese, T.S., and Sheetz, M.P. (1985). Identification of a novel force-generating protein, kinesin, involved in microtubule-based motility. Cell *42*, 39-50.

Vidalain, P.O., Boxem, M., Ge, H., Li, S., and Vidal, M. (2004). Increasing specificity in high-throughput yeast two-hybrid experiments. Methods *32*, 363-370.

Wacker, S.A., Kashyap, S., Li, X., and Kapoor, T.M. (2011). Examining the Mechanism of Action of a Kinesin Inhibitor Using Stable Isotope Labeled Inhibitors for Cross-Linking (SILIC). J Am Chem Soc *133*, 12386-12389.

Walker, G.M. (1982). Cell cycle specificity of certain antimicrotubular drugs in Schizosaccharomyces pombe. J Gen Microbiol *128*, 61-71.

Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet *10*, 57-63.

Wass, M.N., David, A., and Sternberg, M.J. (2011). Challenges for the prediction of macromolecular interactions. Curr Opin Struct Biol *21*, 382-390.

Weerapana, E., Wang, C., Simon, G.M., Richter, F., Khare, S., Dillon, M.B., Bachovchin, D.A., Mowen, K., Baker, D., and Cravatt, B.F. (2010). Quantitative reactivity profiling predicts functional cysteines in proteomes. Nature *468*, 790-795.

Weinger, J.S., Qiu, M., Yang, G., and Kapoor, T.M. (2011). A nonmotor microtubule binding site in kinesin-5 is required for filament crosslinking and sliding. Curr Biol *21*, 154-160.

Weiss, W.A., Taylor, S.S., and Shokat, K.M. (2007). Recognizing and exploiting differences between RNAi and small-molecule inhibitors. Nat Chem Biol *3*, 739-744.

Winzeler, E.A., Shoemaker, D.D., Astromoff, A., Liang, H., Anderson, K., Andre, B., Bangham, R., Benito, R., Boeke, J.D., Bussey, H., Chu, A.M., Connelly, C., Davis, K., Dietrich, F., Dow, S.W., El Bakkoury, M., Foury, F., Friend, S.H., Gentalen, E., Giaever, G., Hegemann, J.H., Jones, T., Laub, M., Liao, H., Liebundguth, N., Lockhart, D.J., Lucau-Danila, A., Lussier, M., M'Rabet, N., Menard, P., Mittmann, M., Pai, C., Rebischung, C., Revuelta, J.L., Riles, L., Roberts, C.J., Ross-MacDonald, P., Scherens, B., Snyder, M., Sookhai-Mahadeo, S., Storms, R.K., Veronneau, S., Voet, M., Volckaert, G., Ward, T.R., Wysocki, R., Yen, G.S., Yu, K., Zimmermann, K., Philippsen, P., Johnston, M., and Davis, R.W. (1999). Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis. Science *285*, 901-906.

Wong, K.M., Hudson, T.J., and McPherson, J.D. (2011). Unraveling the genetics of cancer: genome sequencing and beyond. Annu Rev Genomics Hum Genet *12*, 407-430.

Wood, K.W., Lad, L., Luo, L., Qian, X., Knight, S.D., Nevins, N., Brejc, K., Sutton, D., Gilmartin, A.G., Chua, P.R., Desai, R., Schauer, S.P., McNulty, D.E., Annan, R.S., Belmont, L.D., Garcia, C., Lee, Y., Diamond, M.A., Faucette, L.F., Giardiniere, M., Zhang, S., Sun, C.M., Vidal, J.D., Lichtsteiner, S., Cornwell, W.D., Greshock, J.D., Wooster, R.F., Finer, J.T., Copeland, R.A., Huang, P.S., Morgans, D.J., Jr., Dhanak, D., Bergnes, G., Sakowicz, R., and Jackson, J.R. (2010). Antitumor activity of an allosteric inhibitor of centromere-associated protein-E. Proc Natl Acad Sci U S A *107*, 5839-5844.

Yamamoto, K., Yamazaki, A., Takeuchi, M., and Tanaka, A. (2006). A versatile method of identifying specific binding proteins on affinity resins. Anal Biochem *352*, 15-23.

Yan, Y., Sardana, V., Xu, B., Homnick, C., Halczenko, W., Buser, C.A., Schaber, M., Hartman, G.D., Huber, H.E., and Kuo, L.C. (2004). Inhibition of a mitotic motor protein: where, how, and conformational consequences. J Mol Biol *335*, 547-554.

Yusuf, R.Z., Duan, Z., Lamendola, D.E., Penson, R.T., and Seiden, M.V. (2003). Paclitaxel resistance: molecular mechanisms and pharmacologic manipulation. Curr Cancer Drug Targets *3*, 1-19.

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