Using deep mutational scanning to study protein function and disease

Vanessa Nguyen

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2023

Reading Committee:

Douglas Fowler, Chair

Rachel Klevit

Wendy Thomas

Program Authorized to Offer Degree:

Bioengineering

University of Washington

Abstract

Using deep mutational scanning to study protein function and disease

Vanessa Nguyen

Chair of the Supervisory Committee:

Douglas Fowler

Department of Genome Sciences

Hsp90 is a crucial molecular chaperone that regulates proteostasis by facilitating the refolding and activation of various signaling proteins. Its importance in protein folding and activation has made it a potential target for the treatment of cancer and neurodegenerative diseases. However, due to the diverse sequence space of Hsp90's clients, it has been challenging to characterize the determinants driving Hsp90 function and client recognition. Recent advances in DNA sequencing technology have allowed for multiplexed assays capable of studying thousands of variants and cells in a single experiment. In this thesis, I review the deep mutational scanning technique and its most recent advancements in protein science. Deep mutational scanning can further our understanding of protein biochemistry by generating comprehensive maps of effects of mutations on protein function. To demonstrate its utility, I use deep mutational scanning to investigate the molecular determinants of Hsp90's recognition of Src kinase, a model proto-oncogene. Through this work, I identify novel structural hotspots on the catalytic domain that drive Src's dependence on Hsp90. This study proposes a new model of Hsp90 client recognition, which may serve as a framework for a unified model of Hsp90 chaperoning of client kinases. Overall, this research showcases the potential of deep mutational scanning in furthering our understanding of protein biochemistry and the mechanisms underlying Hsp90 client recognition.

Table of Contents

List of figures

List of tables

Acknowledgements

There is an idiom, “it takes a village to raise a child,” that I thought about often when contemplating the training program that is graduate school. Similarly, it takes a village to train a graduate student. My time in graduate school was often isolating and lonely, but I had a village of friends, family, and mentors that provided the tools, support, and care for me to complete this program. I want to acknowledge and thank all the members of this village, without which I am sure I would not have been able to complete my degree.

To my parents, my first support system. Thank you for forcing me to do extra homework all throughout grade school--I hated it then, but I learned early on how to be self-sufficient in my own education, and for that, I attribute a great deal of my success. Thanks to you and Angela for the meals and care packages that you brought, when you hate driving in Seattle, during my exams. Emotionally, these sustained me during times when I felt alone and overwhelmed.

To Ms. Moon and Mrs. Caraballo--you both showed me how engaging science could be. I didn’t think it at the time, but I think now, it was pivotal for me to learn from and be inspired by educated female scientists so that I might, one day, become one myself. I especially want to thank Mrs. Caraballo who sowed the first seeds of my interest in protein science, when we studied channelrhodopsin protein structure, many years ago.

To my undergraduate bioengineering cohort, a community that I will always remember fondly and be grateful for. My cohort was absolutely the most brilliant and hardworking group of individuals that I have ever had the privilege of working alongside. I think most importantly, though, they were compassionate in a way that makes me profoundly hopeful. Thank you for your support during all those late nights, where our community reminded me to value my human being over my academic and professional being. This lesson has remained with me in graduate school and has helped me to not forget my own health.

To my mentors, Randy and Jeremy, at Just that one wonderful summer. Thank you for believing in me so wholeheartedly. It’s been well documented that a mentor’s confidence in their mentee has a tremendous impact on their mentee’s growth. Your confidence in me bolstered my own self-confidence, and played the largest role in my decision to apply to graduate school. I want to especially thank Jeremy,

who found so much joy in curiosity and discovery. Jeremy's positivity was rare and special, and where I found myself most enjoying the scientific process.

To my IPD family, where I first found a community in research. To Lauren, for taking a chance on me when I was basically only a high school student and providing me so many opportunities to grow. To Chris, Jasmine, Brooke, and Alex, whose gentleness and patience allowed my timid self to grow to love science. To Stephen, Cameron, Nick, and Xinting, who always made me laugh and made my hours in the lab fly by. To George, Jorge, and Franziska, who mentored and supported me through my first research projects. To countless others: Matt, Scott, Anindya, Hugh, Mike, and Chris, for their encouragement. Of course, thank you to David, who was so generous in his support, when I was only an undergraduate amongst over a hundred other graduate students, postdocs, and staff. Thank you to all my IPD family for being friends, and not just coworkers.

To my committee, for their guidance and support. I am a better scientist because of them. A special thanks to Doug, for being my advisor. The job of a PI is a hard one, especially during a pandemic. I have grown immeasurably these past few years, by and large through his mentorship.

To my WRF mentors, Will, Kim, and Meher. I think my hamartia in graduate school was my lack of confidence, and your encouragement and support marked a pivotal moment in my graduate career. Sometimes, all we need is a little encouragement. Yours made a world of difference in my last year of graduate school.

To a few special labmates, who gave me their support. Thank you to Nick, who always made himself available, even on weekends and late nights, to read over my writing and to give me feedback on my talks. My greatest academic inflection points happened thanks to his feedback. To Gabe and Val, who valued me as a scientist and as a human. Their friendship and support gave me courage in an era of my life where I had lost courage.

To my closest friends, Zach and Kimmie. Thank you for listening to me cry, for grieving with me in my lows, and for celebrating me with my wins. Thank you for holding me steady through this process. Your support has profoundly changed me in ways that are difficult to describe. I hope to tell you more in this next season of our lives.

And finally, a special thank you to Sharona Gordon. I think there are certain people that we meet in our lives that profoundly alter the trajectory of our life course. Sharona is one such person. Sharona gave me support and mentorship at critical moments in my graduate career. She gave me compassion and empathy when I felt most vulnerable, and unsure if I would be able to complete my degree. I am inspired and in awe of Sharona's commitment to all students. I would not have been able to finish this degree without her.

Chapter 1. Introduction

1.1 Hsp90: a central node of proteostasis

The heat shock response (HSR) is a conserved mechanism in eukaryotes and prokaryotes that maintains proteostasis during heat stress. Ferruccico Ritossa first discovered the heat shock response in 1962 while studying nucleic acid synthesis in *Drosophila* salivary glands[1,2]. Ritossa noticed that cells that had been unintentionally left in a high heat environment had increased transcriptional activity, which was the heat shock response. Ritossa's original manuscript detailing their discovery of the HSR was dismissed by an editor of a highly respected journal for lack of biological relevance. Nonetheless, with over eighty-five thousand published studies on PubMed today, the HSR is now widely recognized as a biologically relevant phenomenon.

In the HSR, physiological stress signals including elevated temperatures, oxidative stress, nutrient deprivation, hypoxia, and apoptotic stimuli induce the expression of various heat shock proteins (HSPs)[3]. HSPs are multifunctional proteins with a primary function of folding and refolding misfolded proteins[4]. While physiological stress signals induce the HSR and subsequently HSP expression, HSPs are also present under normal growth conditions. HSPs comprise 5-10% of total cellular protein and assist in the folding and activation of nascent polypeptides[5]. Given their crucial role in protein folding, HSPs have become a subject of interest for biologists and drug developers.

HSPs are oligomeric proteins classified by the molecular weight of their monomer. The primary heat shock families are Hsp100, Hsp90, Hsp70, Hsp60, and small HSP (sHSP). Hsp90, which accounts for up to 6% of total cellular proteins during the HSR, is of particular interest[6]. Hsp90 regulates various protein substrates, known as clients, by mediating their late-stage maturation, activation, stabilization, aggregation, and degradation. Hsp90 associates with a broad range of structurally and functionally diverse clients, including tau protein, synuclein, kinases, steroid hormone receptors, and E3 ligases[7]. By chaperoning diverse client signaling proteins, Hsp90 occupies a central node that regulates cell differentiation, development, adaptive immunity, and proteostasis[7–9].

Hsp90 forms a central node of proteostasis and has multiple isoforms localized to different subcellular compartments. Mammals express two isoforms, Hsp90α and Hsp90β, in the cytosol[10]. The isoforms are functionally similar, but are expressed under different conditions. Hsp90α expression is induced by physiological stress during the HSR while Hsp90β is constitutively expressed in the cytosol. Other isoforms include GRP94 (glucose regulated protein), expressed in the endoplasmic reticulum (ER) and TRAP-1 (tumor necrosis factor receptor associated protein 1), expressed in the mitochondrial matrix[8,10]. The ER and mitochondrial isoforms of Hsp90 are both functionally and structurally similar to the cytosolic isoforms of Hsp90. Subtle differences exist between GRP94, TRAP-1, Hsp90α, and Hsp90β. Hsp90α and Hsp90β, at least, bind similar sets of client proteins, and thus mentions of Hsp90 typically refer to both cytosolic isoforms.

1.2 Hsp90 structure and function

Hsp90 exists as a homodimer composed of 90 kilodalton monomers, forming a clamp-like structure that associates with and refolds client proteins. Each Hsp90 monomer is composed of three distinct conserved domains: an N-terminal dimerization domain (NTD), a middle domain (MD), and C-terminal dimerization domain (CTD) (**Figure 1.1A)** [4,11]. Additionally, a charged region of variable length between the NTD and MD is conserved across eukaryotic HSPs and essential for Hsp90 function[12].

The NTD contains a nucleotide-binding site and a “cap” that closes over the binding site when ATP is bound[13]. The cap and nucleotide-binding residues are highly conserved and critical to Hsp90’s ATPase function. The MD contains catalytic residues that drive the hydrolysis of ATP bound by the NTP[14]. Additionally, most client proteins associate with Hsp90 through the MD[4]. The CTD contains the dimerization interface at the base of the Hsp90 clamp as well as a highly conserved MEEVD peptide that associates with co-chaperones containing a tetratricopeptide repeat clamp[15]. While the NTD is the primary site of ATP binding and hydrolysis, the CTD also contains a nucleotide binding site that becomes available when the nucleotide binding site of NTD is occupied[4]. The CTD nucleotide binding site, together with the catalytic residues of the MD, regulate the ATPase activity of Hsp90.

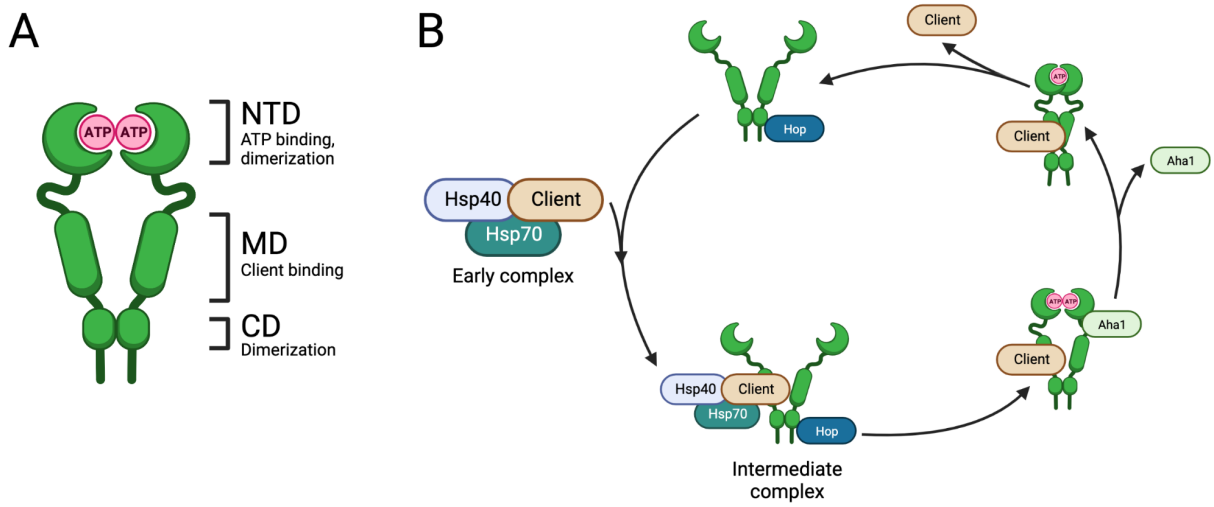


Figure 1.1 (A) Domain organization of Hsp90 dimer (green) bound to ATP (pink). (B) The Hsp90 cycle begins when the early complex composed of Hsp40 (purple), client protein (brown), and Hsp70 (teal) bind to Hsp90 and Hop (blue) to form the intermediate complex. Subsequent binding of Aha1 (light green) allows for ATP-binding at the NTD of Hsp90. Dissociation of Aha1 drives a conformational change in Hsp90 through ATP hydrolysis, thereby releasing the refolded client protein.

Hsp90 performs its primary role refolding client proteins through a cycle of ATP binding, client binding, and ATP hydrolysis (**Figure 1.1B)**. *apo*-Hsp90 exists as a homodimer that is only dimerized at the CTD. The Hsp90 cycle initiates with the binding of ATP at the NTD. Upon ATP binding, the conserved cap of the NTD closes over the bound nucleotide, allowing the dimerization of the NTD and the trapping of client proteins in the Hsp90 lumen. Finally, ATP hydrolysis drives a conformational change in the closed Hsp90 conformation, thereby refolding and activating the client proteins bound to Hsp90.

In addition to Hsp90 and the bound client protein, many other proteins, referred to as co-chaperones, are involved in the Hsp90 cycle. In order to associate with a broad array of client proteins, Hsp90 associates with a variety of co-chaperones that capture client proteins and bring them to the Hsp90 complex to be refolded. Prior to a client's association with Hsp90, the HSPs Hsp70 and Hsp40 first form an early complex together with a client protein. This early complex associates with Hsp90 to form what is referred to as an intermediate complex. The intermediate complex is then bound by the co-chaperone Hop, which associates with both the NTD and MEEVD motif of the CTD of Hsp90 to block ATP hydrolysis. Once bound to both domains, the co-chaperone Hop facilitates the transfer of the client

protein to the MD of Hsp90. The co-chaperone Aha1 associates with NTD and MD of Hsp90 to allow for the binding of ATP and regulates the ATPase activity of Hsp90 through interactions with the catalytic loop of the MD of Hsp90. The dissociation of Aha1 from Hsp90 drives ATP hydrolysis by the NTD and MD of Hsp90 and further dissociation of co-chaperones, Hsp70, and Hsp40. The ATP hydrolysis at the NTD drives the conformational change of the Hsp90 dimer that facilitates the refolding of the bound client protein. Once hydrolyzed, the cap on the NTD releases the ADP and the client protein unbound to allow the Hsp90 chaperone cycle to begin again.

Research on the Hsp90 cycle has been primarily focused on individual stages and chaperones, leading to an incomplete understanding of the entire Hsp90 cycle. Recent studies have highlighted the influence of both the cell cycle and post-translational modifications on Hsp90 activity and its ability to refold and activate client proteins. However, much of this phenomenon remains poorly understood[16].

Moreover, an essential aspect of the Hsp90 cycle that requires further investigation is the mechanism by which it recognizes its various client protein substrates. The current widely-accepted hypothesis is that co-chaperones identify certain conformations of partially unfolded client proteins, which are then loaded onto the Hsp90 complex. While structures of well-studied Hsp90 clients in complex with the chaperone have revealed late complexes in the Hsp90 cycle, the conformation of client proteins that trigger their recruitment by Hsp90 co-chaperones is yet to be determined[17]. The mechanism of client recognition by Hsp90 and its co-chaperones is critical to understanding the fundamental biology of the Hsp90 cycle and in developing drugs that target Hsp90 clients. Therefore, further research on this subject is necessary to fully comprehend the Hsp90 cycle's intricacies and Hsp90 client profile.

1.3 The role of Hsp90 in human disease

The heat shock response and heat shock proteins (HSPs) have been extensively studied in relation to a wide range of human diseases. Of particular interest is Hsp90, a crucial mediator of protein folding and activation. Due to the cellular stress incurred in various pathologies, it is unsurprising that Hsp90 has repeatedly appeared as a central node of many diseases, including pulmonary diseases, neurodegenerative disorders, and cancer.

Hsp90 has been found to be expressed at higher levels in a diverse range of cancers, including, but not limited to, colorectal, prostate, and breast cancer[18–20]. This increased expression is attributed to its ability to protect cancer cells from hypoxia and ischemia, which are commonly experienced during tumorigenesis. Additionally, several common oncogenes are clients of Hsp90, suggesting that Hsp90 activity aids in the survival and growth of malignant cells[9]. Therefore, Hsp90 emerged as an attractive therapeutic target for cancer treatment. Early inhibitors of Hsp90, such as geldanamycin and radicicol, were found to induce the rapid degradation of oncoproteins and cell death in tumor cells, but not normal cells. This selective inhibition indicated that Hsp90 function was critical for the survival of cancer cells.

Furthermore, Hsp90 is present on the plasma membrane and the extracellular matrix, with higher expression levels in cancer cells compared to normal cells. Plasma concentrations of Hsp90 are positively correlated with tumor malignancy, and treatment with Hsp90 inhibitors has been found to inhibit cancer metastasis. Hsp90 enables tumor cell invasion by activating MMP-2, which digests extracellular matrix components[21]. Increased Hsp90 activity leads to increased MMP-2 activity, digesting the extracellular matrix and allowing cancer cells to invade the bloodstream. Taken together, Hsp90's function in protein folding and activation, as well as its presence on the plasma membrane and extracellular matrix, make it an attractive therapeutic target for cancer treatment.

Hsp90 has recently emerged as a promising therapeutic target for various neurodegenerative disorders, in addition to cancer. A number of proteins involved in neurodegenerative disorders, such as tau protein, amyloid-β, and α-synuclein, are clients of Hsp90[4]. In Alzheimer's disease, tau protein destabilizes microtubules and leads to neurodegeneration when elevated. Work has shown that Hsp90 further activates tau protein, suggesting that Hsp90 activity facilitates microtubule destabilization by tau[22]. On the other hand, Hsp90 can reduce the accumulation of amyloid-β in nerve cells, which is also a hallmark of Alzheimer's disease, by binding to the misfolded protein and preventing further aggregation. Thus, Hsp90 plays a key regulatory role in Alzheimer's disease, with opposing functions on tau and amyloid-β.

In Parkinson's disease, Hsp90 also plays a critical role in the regulation of multiple Hsp90 clients such as α-synuclein, Parkin, PINK1, and LRRK2[23]. It is noteworthy that Hsp90's ATPase activity activates α-synuclein, and inhibiting Hsp90 has been shown to reduce the cytotoxicity of mutant α-synuclein[24].

Therefore, Hsp90's role in proteostasis is critical in regulating the activity and toxic accumulation of various proteins related to neurodegeneration.

Consequently, Hsp90 has emerged as a promising therapeutic target for neurodegenerative disorders due to its role in the regulation of multiple Hsp90 clients. Its ability to regulate the activity of various proteins, including tau protein, amyloid-β, and α-synuclein, highlights the importance of Hsp90 in proteostasis and disease pathology. Further work is needed to fully understand the mechanisms involved in Hsp90's regulation of neurodegeneration-related proteins in order to develop effective therapeutic strategies for these disorders.

1.4 Therapeutic targeting of Hsp90

Hsp90 is a key player in the progression of cancer and neurodegenerative disorders, and consequently, there is considerable interest in therapeutically targeting this protein. The first generation of Hsp90 inhibitors targeted the N-terminal domain (NTD) of Hsp90 and functioned as competitive inhibitors with ATP[25]. Geldanamycin, herbimycin, and radicicol were among the first Hsp90 inhibitors, and demonstrated promising strong anticancer effects in cell models and in mice[4]. However, these inhibitors had limited clinical success due to their unfavorable solubility. To address these issues, synthetic derivatives such as 17-AAG (tanespimycin) and 17-DMAG were developed[25]. While these derivatives functioned similarly to geldanamycin by competitively inhibiting ATP binding at the NTD of Hsp90, they also activated the heat shock response (HSR), resulting in the expression of HSPs with anti-apoptotic, drug resistance, and proliferative effects[26]. This contradictory effect of increasing the available HSPs in cells dampened the effect of potent Hsp90 inhibitors.

As a result of the HSR activation caused by competitive inhibitors of Hsp90’s NTD, the next generation of Hsp90 inhibitors were developed to target the C-terminal domain (CTD), the dimerization domain of Hsp90[27]. Coumarin-based antibiotics such as novobiocin were designed to disrupt the dimerization of Hsp90, inhibiting the stability of the Hsp90-client complex and thereby blocking Hsp90’s ability to activate misfolded client proteins. These CTD-binding Hsp90 inhibitors were promising and did not activate the HSR to the same degree as the first generation of inhibitors.

However, no Hsp90 inhibitors have been clinically approved to date due to the dose-limiting toxicity of these inhibitors and the concurrent activation of the HSR. Therefore, an alternative strategy for targeting Hsp90 involves inhibiting specific co-chaperones[4,28]. Since many client oncoproteins are kinases, preliminary work has focused on inhibiting the kinase-specific co-chaperone Cdc37, which facilitates the inactivation and degradation of oncogenic kinases. Other research has explored inhibiting the cochaperone Aha1, the Hsp90 activator, as an alternative to inhibiting Hsp90 with traditional Hsp90 inhibitors. Additionally, Hsp90 inhibitors have been shown to work well in combination therapies with antagonists to oncoproteins, thereby increasing sensitivity to and preventing resistance to tyrosine kinase inhibitors[29].

1.5 Studying Hsp90 in the era of high-throughput multiplexed assays

Decades after Ritossa's seminal discovery of Heat Shock Response (HSR), Hsp90 has emerged as a key regulator not only of protein folding, but also of a number of prominent human diseases. Considerable progress has been made to understand Hsp90 in disease models, as well as to characterize the complex interplay between Hsp90, its vast array of co-chaperones, and diverse client proteins. While many studies have shed light on the structural and functional aspects of Hsp90, the complexity of the system, combined with the diversity of client proteins and the unique roles of co-chaperones, make it a challenging task to fully decipher the underlying mechanisms of Hsp90 function.

Although previous studies have provided valuable insights into the function of Hsp90, they often lack the comprehensiveness required to fully contextualize other findings. For instance, experiments characterizing the structures and behaviors of certain client proteins following Hsp90 inhibition have been instrumental in highlighting key client proteins and diseases that are driven by Hsp90 function, but such low-throughput studies explore a handful of client interactions amongst the landscape of innumerable diverse Hsp90 clients and Hsp90 client variants, making it difficult to obtain a holistic understanding of Hsp90 chaperoning.

Fortunately, with the advent of high-throughput methods, it is now possible to comprehensively study protein folding and chaperoning, offering a unique opportunity to gain a more complete

understanding of Hsp90 function. For example, high-throughput work has been done to measure the interaction strength of over 400 kinases in the human kinome with Hsp90 and its adapter cochaperone Cdc37[30]. However, while this work supported the hypothesis that destabilized and misfolded proteins interact more strongly with Hsp90 than stabilized proteins, it failed to identify sequence or structural motifs that gave rise to Hsp90 chaperoning.

In light of these challenges, deep mutational scanning has emerged as a powerful high-throughput method for comprehensively characterizing the functional effects of missense mutations at every position of a protein. This approach has been increasingly utilized as a tool for understanding protein function and human disease, and offers unique advantages over traditional biochemical or structural methods. In chapter 2 of this thesis, I review recent advances in deep mutational scanning, highlighting its capability to study complex protein phenomena that are not readily accessible by other methods, and its potential to reveal novel insights into protein function.

In chapter 3, I use deep mutational scanning to study the molecular determinants of client recognition by Hsp90. Specifically, I investigate how variation in a model client oncogene, Src, affects its processing by Hsp90. By identifying novel variants that dramatically increase Src's functional dependence on Hsp90, I present a refined model of Hsp90 client recognition. I propose that Hsp90 and the kinase-specific cochaperone Cdc37 do not recognize a specific sequence motif, but rather recognize the global extension of Src and the separation of the two lobes of the kinase domain, to recruit client kinases to the Hsp90 complex.

Finally, in the last chapter, I discuss how my proposed model can be used to develop the next generation of inhibitors for oncogenic kinases, and explore opportunities for future work to further refine our understanding of Hsp90 chaperoning. Drawing on my work studying the complex Hsp90 chaperone cycle, as well as other examples summarized in chapter 2, I highlight the future potential of deep mutational scanning for not only studying Hsp90, but other protein phenomena involved in human disease.

Chapter 2: Comprehensive Variant Effect Maps: A Review of Deep Mutational Scanning in Biochemistry and Genetics

This chapter is adapted from a review in preparation.

Abstract

Multiplexed assays of variant effect, such as deep mutational scanning and saturation genome editing, have emerged as a robust method for exploring the impact of genetic variation on protein function. These assays can measure many widely studied protein properties such as stability, ligand binding, or enzymatic activity, yielding comprehensive variant effect maps encompassing nearly all possible missense mutations. Data from deep mutational scans have empowered clinical variant interpretation, protein engineering, and sequence-function modeling in specific proteins. With millions of variant effects measured across hundreds of proteins, the data enables the characterization of trends across multiple variant effect maps, facilitating understanding of the molecular underpinnings of protein stability, binding, and activity. As deep mutational scans continue to develop, they become an increasingly valuable tool in biochemistry and human genetics. In this review, we summarize different deep mutational scanning formats, discuss sequence-function insights that have generalized across multiple multiplexed assays of stability, binding, and activity, and describe current and future applications of variant effect maps.

2.1 Introduction

Proteins are essential macromolecules that perform critical biological functions in all living organisms and act as structures, engines, and logic gates in complex biological systems. Individual protein molecules are profoundly complicated and contain multiple binding sites and functional domains that work in tandem to contribute to the protein's overall structure and function. Even slight perturbations in a protein's sequence can interfere with its structure and function, which can subsequently interfere with cellular processes and result in disease. Small changes in protein sequence can also improve a protein's stability or binding function to yield improved protein therapeutics. Consequently, characterizing protein

variant effects is critical to the determination of variant pathogenicity, the identification of critical functional positions, and the development of engineered proteins.

Amino acid mutations are natural perturbations that can be applied to protein sequences to characterize the relationship between protein sequence, structure, and function. Mutagenesis studies often utilize traditional biochemical and structural biology techniques to assess changes in protein folding, stability, catalytic activity, substrate specificity, and ligand binding affinity resulting from introduced amino acid mutations. Such mutagenesis studies characterize a small number of mutations at privileged positions, specifically active sites, protein-protein interfaces, and positions with a notable occurrence of pathogenic variants. However, it is important to note that while mutagenesis studies predominantly focus on characterizing a limited number of mutations at privileged positions, the potential impact of mutations occurring at other positions throughout a protein sequence cannot be overlooked. These non-targeted positions also possess the capability to induce significant changes in protein structure and function, warranting further investigation beyond the scope of focused mutagenesis studies. Considering the vast number of possible variant effects that can arise from even a single protein sequence, attempting to comprehensively characterize each potential mutation at every position individually is an impractical task.

Alanine scanning, one of the earliest large-scale mutagenesis methods, measures the structural and functional effects of individually mutating each residue to alanine. Alanine scanning has been used to determine the contribution of individual residues to the stability or function of a protein[31–34]. Changes in stability or function resulting from an alanine mutation provides evidence for the critical involvement of the mutated residue in protein folding or function. Alanine scanning studies contribute to a more comprehensive understanding of the underlying mechanisms governing the relationship between protein sequence, structure, and function. For example, an alanine scan of human growth hormone revealed that only seven of nineteen side chains at its binding interfaces contributed significant binding energy[35]. Although alanine scans are effective in identifying energetically critical residues in the wild-type protein, alanine scans cannot assess the effects of mutations beyond alanine substitutions. Consequently, in fields such as evolutionary studies, protein engineering, and disease research, where comprehensive exploration of variant effects is essential, alternative amino acid mutagenesis techniques are required.

Saturation mutagenesis is one such large-scale mutagenesis method that characterizes variants besides alanine. Briefly, saturation mutagenesis involves randomly mutating a subset of residues and selecting a few gain-of-function variants. Common applications of saturation mutagenesis include mutagenizing antibody variable loops for affinity maturation[36]. Unlike alanine scanning, saturation mutagenesis offers the advantage of identifying gain-of-function variants that cannot be readily detected through single alanine substitutions. However, it is important to note that while saturation mutagenesis provides a broader exploration of sequence space, saturation mutagenesis falls short of providing a comprehensive characterization of amino acid trends associated with stability, binding, and activity modulation. Therefore, to gain deeper insights into these intricate mechanisms, additional mutagenesis strategies and complementary approaches are often necessary. Additionally, saturation mutagenesis studies rely on an assumption of binding or active site positions hypothesized from previous structural studies of the protein of interest and consequently bias the results of the gain-of-function variants to the mutagenized region of the protein. However, distal residues commonly have long-range effects on binding and activity[37,38]. A more comprehensive approach for mutational studies is required in order to fully characterize the determinants of protein stability, binding, and activity.

Deep mutational scanning (DMS) is a powerful, high-throughput method for comprehensively characterizing the sequence-function relationships of proteins of interest. DMS, in contrast to conventional saturation mutagenesis studies and alanine scans, systematically evaluates the effects induced by all 19 alternative amino acids at every position within the protein. DMS enables the creation of a detailed variant effect map that is more comprehensive than datasets produced by alanine scanning and saturation mutagenesis.

Briefly, deep mutational scanning involves creating a library of protein variants, encompassing all possible missense mutations of the protein of interest[39]. Each variant is then expressed in a model expression system, such that each cell expresses only a single variant (**Figure 2.1**). Variant-expressing cells are then selectively enriched through survival or fluorescence-based sorting. A functional assay is performed such that the protein property of interest, such as folding, binding, or enzymatic activity, affects the growth rate or fluorescence of variant-expressing cells. The degree of variant enrichment or depletion relative to the wild-type protein sequence is then measured through next-generation sequencing (NGS).

Variant frequencies quantified by NGS are then converted to variant effect scores for every amino acid mutation at every position in a protein sequence to create a variant effect map comprising every single missense variant in the assayed protein.

By measuring all possible single variant effects at all positions, DMS enables the comprehensive characterization of how amino acid biochemistry affects folding, binding, and activity. DMS methods have been developed for a wide range of protein properties, including thermostability, binding affinity, and enzymatic activity, and have provided variant function maps in multiple contexts for studying protein properties in an unbiased manner[40]. These variant function maps have proven useful for a variety of applications, including overcoming energetic barriers in protein engineering, identifying drug-resistant variants, determining the structure of difficult-to-characterize proteins, and identifying novel functional mechanisms for well-studied proteins.

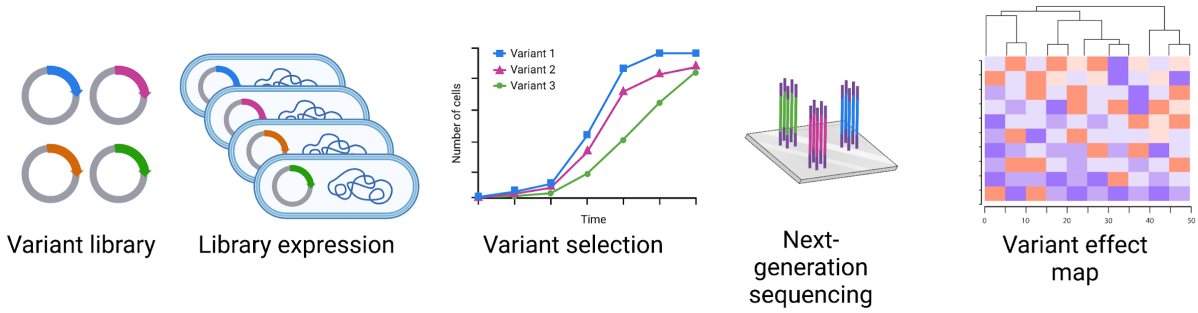


Figure 2.1. Overview of deep mutational scanning. A general deep mutational scanning workflow begins with the generation of a variant library and subsequent expression of that library in a host system. The library-expressing system is then subject to a selection process from which samples are taken and quantified using next-generation sequencing. The variant frequencies will be then converted into functional scores to create a variant effect map in which the x-axis denotes amino acid position and the y-axis denotes amino acid mutations. Red denotes gain-of-function scores while blue denotes loss-of-function scores.

Recent advances in computational methods, specifically in machine learning and deep learning models, hold promise to transform the field of protein biology by facilitating accurate predictions of protein structure and function[41,42]. With this success, the next logical step is to predict how mutations affect protein stability, binding, function, and organismal fitness. To achieve this goal, deep mutational scanning

(DMS) represents a powerful experimental technique capable of generating large datasets that can be used to develop and validate predictive models. However, the challenge remains to assess whether these models effectively capture the underlying principles of protein biophysics from DMS data. Therefore, a thorough evaluation of the extent to which these predictive models capture the fundamental principles of protein folding and function available in DMS data is necessary.

DMS methods have been extensively reviewed, and different approaches for conducting deep mutational scans have been comprehensively summarized in the literature[39,40,43–47]. These reviews cover the experimental design considerations for conducting a deep mutational scan, and applications of deep mutational scanning to clinical variant interpretation. Notably, there is a gap in the current literature regarding deep mutational scanning reviews and discussions of both the insights that can be gained from comprehensive DMS datasets as well as the limitations of comprehensive single variant effect maps. Such a review would provide an opportunity to discuss the utility and various applications of DMS for different research questions and the refinement of predictive models trained on DMS data. In this review, we focus on the types of deep mutational scans that measure variant stability, binding, and activity, and discuss the recent applications and insights garnered from each type. We additionally discuss the challenges faced by each type of functional scan and offer suggestions for future areas of development.

2.2 Stability scans

2.2.1 Introduction

Protein stability has traditionally referred to a protein's ability to resist unfolding and degradation when exposed to denaturing conditions such as heat, chemical denaturants, or protease digestion. This stability is a fundamental characteristic that underlies all other protein functions, and discussions regarding protein function inevitably include considerations of a protein sequence's inherent stability. For example, if a mutation alters protein’s binding activity, it is essential to determine whether this effect is the result of a decrease in substrate affinity or simply misfolding. Mutational studies of all types of protein stability are indispensable for deciphering sequence-structure and sequence-function relationships.

Traditional *in vitro* approaches to studying protein stability are useful in understanding how primary sequence affects protein stability. *in vitro* approaches generally measure the stability of purified protein through heat, chemical, or proteolytic denaturation and degradation. *In vitro* methods can be used to derive different protein variants' thermodynamic parameters such as melting temperature to directly compare how changes in sequence affect a given thermodynamic parameter. While such thermodynamic parameters are useful when making direct comparisons between protein variants, changes in *in vitro* stability do not always translate to changes to a protein's ability to remain folded in a cell.

Protein folding in a biological context can be studied through combining site-directed mutagenesis with immunoblotting, protease sensitivity assays, or with fluorescent tags and labels. It is important to note that these techniques do not measure thermodynamic parameters like in *in vitro* assays of protein stability. Instead, they measure the steady-state abundance of folded or active protein in cells. This steady-state protein abundance is influenced by several factors, including the protein quality control system, transcription factors, and the thermodynamic stability of the protein. Therefore, assays of "intracellular stability" measure the cumulative result of these interactions and are not equivalent to stability measures in *in vitro* assays. While it is more challenging to interpret the factors underlying intracellular protein stability than denaturing with *in vitro* assays, it is advantageous when examining the protein's role in biology and disease. Changes in steady-state protein variant abundance do not always translate to changes in thermodynamic, chemical, or proteolytic stability. Thus, assays of "intracellular stability" may more readily characterize changes in stability that give rise to pathogenicity. The respective advantages and limitations between *in vitro* and cell-based stability assays apply to both small-scale mutagenesis studies and deep mutational scanning assays.

Multiplexed stability assays can be performed *in vitro* or in cells for direct or indirect measurements of stability, respectively. Multiplexed measurements of stability generally use a protein display system where a protein library displayed on the surface of yeast can be challenged with heating or proteolysis to measure the relative stability of library variants[48]. Alternatively, indirect multiplexed measurements of stability rely on linking the library variants to a fluorescence or essential reporter protein that allows for selection[39]. Similar to lower-throughput cell-based assays of protein stability, multiplexed assays of protein stability in cells assume that destabilized library variants will be preferentially depleted

by the cell's protein quality control machinery, thereby decreasing the expression levels of the fused reporter protein[49–54]. Overall, the use of mutational scans to identify stabilizing and destabilizing mutations in proteins prove valuable for improving our understanding of basic protein function and protein engineering. However, it is important to note that different approaches to DMS for stability assessments are more suitable for specific applications, and careful consideration should be given when selecting the most appropriate method for conducting a mutational scan of protein stability.

2.2.2 in vitro stability

In vitro DMS methods measure protein stability following thermal and chemical denaturation as well as proteolysis[55,5657–59]. To determine relative thermal and protease stability, a yeast display system can be used to display a variant library and challenged with protease or heat and sorted by flow cytometry and quantified using NGS for enrichment of folded protein (**Figure 2.2A**).

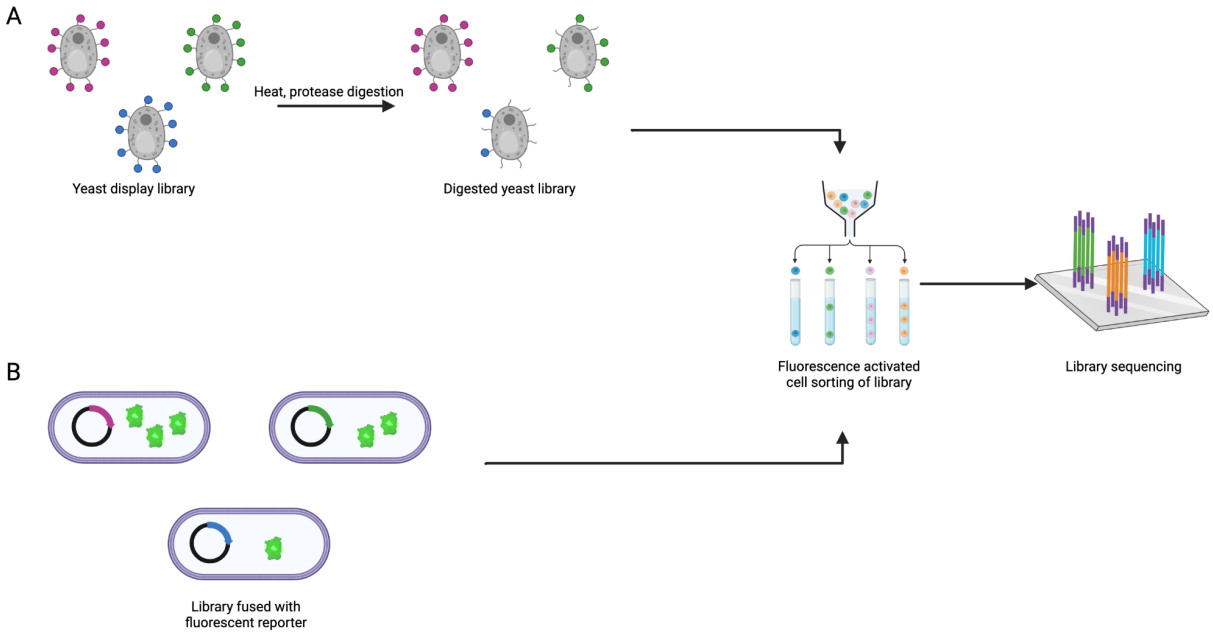


Figure 2.2 Overview of generic *in vitro* and cell-based DMS methods (A) *in vitro* scans display a protein variant library on the surface of yeast, phage, bacterial, or mammalian cells. The displayed library can be challenged with denaturants. The remaining displayed protein can then be detected with fluorescence

activated cell sorting and quantified by deep sequencing. (B) Cell-based assays of stability can include fused fluorescent reporters that enable similar fluorescence activated cell sorting and deep sequencing. Split fluorescent reporters can be used to measure binding events in cells.

Trends revealed by *in vitro* scans of stability recapitulate canonical knowledge: core positions of globular proteins are typically more intolerant of mutation than surface positions, and disulfides destabilize folding, while prolines and glycines disrupt folding[53,60,61]. A typical example of this is a study measuring the thermal stability of the CH3 domain of an IgG1 antibody using DMS. The stability DMS of the CH3 domain revealed low mutational tolerance at the CH3-CH3 interface involved in IgG homodimer formation and in positions oriented toward the hydrophobic core of the protein involved in disulfide bonds[55]. In contrast, solvent-exposed residues were found to be more tolerant of mutation. These results validate the critical role of hydrophobic packing in the protein core and at inter-domain interfaces and are a prototypical example of how DMSs recapitulate known rules of protein folding. However, it should be acknowledged that these are general observations across DMS datasets and not all scanned mutations in the hydrophobic core lead to destabilization, and the introduction of proline does not always disrupt folding. Interestingly, the identification of amino acid trends within comprehensive variant effects maps of protein stability has proven to be a challenging task. Specifically, when examining mutationally intolerant positions situated in the hydrophobic core, no consistent trends emerge regarding the disruption caused by larger amino acid substitutions compared to smaller ones. This lack of clear patterns extends to substitutions involving buried polar and charged amino acids as well. These findings suggest that while DMS studies largely adhere to the established rules of protein folding, they also reveal intriguing exceptions to these rules.

When considering such exceptions, it is important to acknowledge a fundamental limitation of all DMS approaches. While DMSs effectively characterize the effects of variants on protein stability, they do not uncover the underlying mechanisms that drive these effects. Nevertheless, it is worth noting that powerful computational predictors for protein structure were developed despite an incomplete understanding of the mechanistic aspects of protein folding. Similarly powerful variant effect predictors could also be developed using DMS data without the need for a complete mechanistic understanding of how variants affect protein folding.

A noteworthy caveat of *in vitro* stability mutational scans is that a residue's mutational tolerance may differ between different stability scans, as shown in a scan of lipase protein A from Bacillus subtilis (BsLipA). The DMS of BsLipA measured thermostability and detergent tolerance of a library of variants and revealed varied mutational tolerance between the different detergent DMSs and thermostability scans[57]. Chemical differences among the detergents likely account for these differences, which should be considered when conducting in vitro stability scans. This suggests that generalizing the results of a thermostability scan to detergent tolerance should be done with caution. This particular scan necessitated the testing of different detergents because engineered enzymes are commonly formulated with detergents, and thus a variant's specific tolerance of different detergents is a critical parameter to optimize. However, if the stability DMS is only intended to comprehensively characterize variant effects on protein folding, there is no need for multiple scans under different conditions as in the study of BsLipA.

One interesting method for *in vitro* stability deep mutational scanning identifies dynamic regions compatible with circular permutation. Circular permutation is a challenging and cumbersome process, particularly for large, multi-domain proteins that is typically used to engineer improved protein variants and study the properties of the original protein. The deep mutational scanning method for circular permutation profiling, CPP-seq, can identify positions that are tolerant of permutation, allowing for protein engineering not typically available, and the identification of flexible and tolerant positions in a protein structure[62,63]. Circular permutation is more difficult to computationally model than predicting changes in folding energy following mutation in globular proteins, and consequently there is a gap in knowledge of rational design of circularly permuted proteins. Additionally, circular permutation can be used to study domain positioning and connectivity on protein stability and characterize domain-domain interactions and interdomain cooperativity. However, the results of CPP-seq are not as interpretable as those from a thermostability or detergent tolerance scan. Thus, CPP-seq should be primarily considered when either trying to engineer circularly permuted variants from the data or when domain positioning is likely to have a large influence on overall protein stability.

Compared to scans that measure protein folding in the context of the cell, *in vitro* DMSs for stability are more interpretable, making them particularly useful for engineering studies aimed at improving the overall stability of a protein for research tools or therapeutic applications, for comparison to

physics-based computational predictors of variant effect, and for the training of variant effect predictors[37,38,41].

2.2.3 Cell based assays of stability

Deep mutational scanning can also be used for characterizing protein stability in cells, enabling researchers to measure the effects of amino acid substitutions on protein function and folding. Protein stability in cells can be measured by fusing the protein library to a fluorescent or transcriptional reporter protein, which allows fluorescence levels to be quantified and binned using fluorescence-activated cell sorting (FACS) (**Figure 2.2B**) [49,52–54,64]. Alternatively, a transcriptional reporter can be placed upstream of a fluorescent protein or an essential protein that affects the growth rate of the transformed cell[49–51].

DMSs of variant stability in cells typically reveal that determinants of stability are largely influenced by the cellular compartment in which the protein is expressed[54,65,66]. For instance, the mutational scan of the cytoplasmic protein PTEN revealed that positions in the core were found to decrease variant abundance and were the most intolerant to mutation[52]. The mutational scan of PTEN also showed that bulky side chain mutations in the core generally decreased PTEN variant abundance in the cytosol more than comparatively small side chain mutations. In contrast, the mutational scan of endoplasmic reticulum (ER) localized cytochrome p450 showed that the transmembrane positions had the largest effects on variant abundance and were intolerant of charged and polar mutations[53]. Together, mutational scans of intracellular stability measure a combination of intra-protein folding interactions and protein-environment interactions, providing exceptional utility for proteins that lack structural data, such as membrane proteins and intrinsically disordered domains[54,67]. In fact, multiple groups have used deep mutational scanning data to infer the protein structures of the GB1 domain of protein G, the WW domain of the human Yap1, the second RRM domain of Pab1, the Fos-Jun heterodimer, and a twister ribozyme [68,69]. Impressively, the computed structures of GB1 and the WW domain were within 2.1 Å of the crystal structure. In order to compute the structures, however, the authors required DMS datasets to include large numbers of double mutants in order to create epistatic models from which they inferred protein

structure. Thus, while impressive, inferring full protein structures from DMS datasets is largely limited to smaller protein domains where large numbers of single and double mutants can be sequenced.

In addition to the cellular compartment's effect on protein folding, the proteostasis landscape can also affect protein cellular abundance and stability. A cell's proteostasis landscape is shaped by the chaperones that are expressed, which can be altered by the stress induced by drug treatments or by overexpression of the misfolded variants. For example, mutational scans of hemagglutinin in different proteostasis landscapes have highlighted regions of mutational tolerance and potential evolutionary pathways of the virus[70,71]. This type of mutational scan can also be applied to oncogenic proteins to identify regions that are likely to mutate in more permissive proteostasis environments. Altered proteostasis can also allow the formation of toxic aggregates that are hallmarks of protein misfolding diseases like Alzheimer's, Parkinson's, and cystic fibrosis. Deep mutational scanning has been used to measure the aggregation of proteins underlying multiple misfolding diseases and can be used to identify pathways of misfolding and classify potentially pathogenic variants[49,50,52,72]. However, it is important to consider that the overexpression of protein variants, a common practice in DMS assays, may introduce potential alterations in the interaction between the variants and the cell's proteostasis machinery. Consequently, this can result in differences in the observed variant abundance compared to their abundance in physiological conditions. While the choice of expression system can introduce slight variations in the genetic background and proteostasis machinery, the overall trends highlighting the structural significance of a protein's hydrophobic core remain consistent across scans conducted in different proteostasis landscapes. Thus, conducting multiple scans in different proteostasis states is best suited for characterizing a protein's interaction with a cell's proteostasis machinery.

Notably, a prevailing theme of stability deep mutational scans is the importance of a protein's hydrophobic core to its overall stability[53,55]. In cell-based DMSs for variant stability, the mutational tolerance of the surface is dictated by the cellular environment in which the protein is natively expressed. Namely, cytosolic proteins are stabilized by hydrophobic packing and mutations within the hydrophobic core are often deleterious. Conversely, membrane-bound proteins have long hydrophobic amino stretches on the surface in contact with the membrane, with charged surface mutations typically being highly deleterious to their expression[73–75]. In other words, deep mutational scans of stability closely align with

known mechanisms of protein folding, and researchers have used them to indirectly characterize the structure of proteins with limited structural data[68].

Typical mutational scans identify many mutations that are deleterious to protein folding and stability. Across all stability scans, regardless of expression system, stabilizing variants often accounted for less than 5% of the observed mutations[60]. The lack of observed stabilizing mutations could result from limited dynamic range of the assay, but more likely signal that it is easier to disrupt folding through single mutations than to stabilize a protein through a single mutation. Indeed, there are so few stabilizing mutations identified in DMSs that trends rarely emerged to add to our collective understanding of protein stability and folding. The lack of stabilizing mutations is likely due to the reliance on the cell's proteostasis machinery to measure stability. Namely, the cell's proteostasis machinery targets misfolded or aggregated proteins for degradation, which in turn affects the measured cellular abundance. However, it is important to note that the proteostasis machinery does not necessarily degrade stabilized variants to a lesser extent than an already stable wild-type protein, potentially limiting measured changes in cellular abundance. Thus, cell-based DMSs of stability are less suited for identifying stabilizing mutations and are better suited for interrogating how sequence variation affects cellular abundance or alters interactions with proteostasis machinery.

2.2.4 Comparisons to evolutionary conservation and other predictive methods

Conservation analyses are a widely used computational method to evaluate the mutational tolerance of individual amino acids in a protein sequence, as well as their implied structural and functional importance. Mutational scans of stability have shown a high degree of agreement with conservation analyses, although they differ in that mutational scans comprehensively sample the missense sequence space, while evolution only samples a limited number of sequence trajectories. For example, reengineering TEM-1 beta lactamase using deep mutational scanning data revealed that relying solely on conservation analysis for choosing mutations would have been insufficient to improve solubility[76]. Klesmith et al. employed a position-specific scoring matrix (PSSM) to identify neutral mutations, mutations that had little effect on stability, in TEM-1 and LGK (levoglucosan kinase). Notably, 31% of the

PSSM-defined neutral mutations in TEM-1 and 43% of the PSSM-defined neutral mutations in LGK were found to be deleterious in their DMS experiments. This observation highlights the moderate likelihood of evolutionary sampling including deleterious mutations in these proteins and potentially others. Thus, stability DMS provides valuable insights beyond evolutionary conservation scores, despite the strong correlation between the two approaches. However, it is noteworthy that 69% of PSSM-defined neutral TEM-1 mutations were assessed as neutral by DMS, compared to only 32% of all TEM-1 mutations being characterized as neutral. Similarly, 57% of PSSM-defined neutral LGK mutations were deemed neutral by DMS, while only 28% of all LGK mutations showed neutral effects. Consequently, DMS studies can effectively limit scans to conserved positions, reducing sequencing requirements and eliminating deleterious evolutionary pathways, while still maintaining a reasonable probability of identifying stabilizing mutations in the scan.

Predictors that incorporate evolutionary data and physics modeling, such as FoldX and Rosetta, are commonly used to forecast the impact of mutations on protein stability. The correlation between these predictors and deep mutational scans of stability is modest, typically ranging from a Pearson correlation of 0.3-0.6[59]. A comparison of experimental deep mutational scanning measurements of GB1 stability in guanidinium chloride found that Rosetta had the highest Pearson correlation, while other methods performed equally well when using Spearman rank correlation[59]. Rosetta has far more terms in its energy function compared to FoldX, which likely contributes to Rosetta's superior Pearson correlation with deep mutational scanning data. The similarities between FoldX and Rosetta Spearman correlation suggest that the tools are comparable for predicting variant effects on stability, with FoldX requiring fewer computational resources compared to Rosetta. Intriguingly, this study noted that the computational predictors demonstrated better performance on the protein surface than in the core and were more successful in predicting changes in stability from large-to-small mutations than small-to-large mutations. While these comparisons are not widely performed with other mutational scans, the moderate correlation and poor performance in core positions highlight the fundamental importance of deep mutational scanning in discerning variant effects on stability that may not be evident through the current understanding of protein folding or molecular modeling.

Since the advent of deep mutational scanning, researchers have published over 50 peer-reviewed papers describing stability deep mutational scans that have measured the stability of more than 600,000 variants (**Supplementary Table 2.1**). Although *in vitro* scans are possible through display methods, most stability deep mutational scans measure protein variant stability within cells using fluorescent or transcriptional reporters. Notably, a common finding amongst stability DMSs is the importance of a protein's hydrophobic core to its overall stability[53,77]. In cell-based stability deep mutational scans, the mutational tolerance of the surface is dictated by the cellular environment in which the protein is expressed[54,65,66]. Moreover, sequence-stability variant effect maps have proven useful in identifying pathogenic variants, where the underlying mechanism of pathogenicity is decreased levels of folded protein[52–54]. As a result, stability deep mutational scans have both biochemical and clinical utility and are essential components of multi-parameter deep mutational scans, as we will discuss later.

2.3 Binding scans

2.3.1 Introduction

Proteins function within large signaling networks and rarely function in isolation. Proteins must bind to other relevant species to initiate or propagate signals that drive cellular function. When a mutation incurs a loss of protein binding, cellular function is also disrupted. Therefore, protein binding, in addition to protein folding and stability, is also a prominent area of study. The identification of residues that affect binding is also critical to efforts to reengineer proteins to improve their affinity or alter their selectivity. Interrogating binding events between proteins and different cognate ligands can help to inform the rational design of therapeutic proteins, functional biosensors, as well as provide the basis for understanding mechanisms of disease.

Various biochemical methods can be used to study protein binding, including measurements of binding affinity *in vitro* or the characterization of binding events in cells using methods like fluorescence resonance energy transfer (FRET). *In vitro* methods for characterizing protein binding can determine equilibration constants that are relatively easy to interpret. The simplicity, precision, and interpretability of

in vitro measurements of binding affinity are helpful in engineering the specificity or affinity of native or *de novo* proteins. In addition to *in vitro* methods, protein binding can be characterized in cellular contexts, to study changes to binding affect cellular processes. Cell-based measurements of binding events are especially useful in measuring how sequence variation affects nucleic acid binding domains.

Deep mutational scans of binding **(Supplementary Table 2.2)** are used to measure protein variants' relative binding affinity to small molecules, other proteins, and nucleic acids[78–80]. It is important to note that functional scores determined from binding DMS are not absolute binding affinities, and they only report on binding affinity relative to a reference sequence, such as the wild-type sequence. A major advantage of binding deep mutational scans over traditional methods of studying protein binding is that DMSs can identify residues that influence binding up to 15 angstroms away from the binding interface[77,81,82]. These residues are not identifiable by structural methods and expand the range of positions that can be modified in protein engineering studies. However, in deep mutational scans of binding that do not include expression or display controls, it becomes unclear whether changes in binding occurred through decreased expression and folding, or through a loss of affinity. This limitation can be easily overcome by including co-translated fluorescent proteins or display tags that can be simultaneously detected with fluorescent antibodies to normalize measured binding to protein variant expression.

Common multiplexed methods for measuring protein binding include complementation methods and display methods. Complementation methods detect protein binding events in a cellular context through the activation of a downstream reporter protein that is only expressed when two proteins tagged with complementary fragments bind. Briefly, a protein library is tagged with a fragmented transcription factor and the protein library's substrate of interest is tagged with the transcription factor's complement fragment. Successful complementation between the library variant and binding protein results in a complete transcription factor that induces the expression of a downstream reporter protein. Downstream reporter proteins can be fluorescent proteins that can be detected using fluorescence-activated cell sorting (FACS), or essential proteins, such that growth becomes a proxy for the complementation efficiency. Enrichment of fluorescence or growth phenotypes can thus be quantified through next-generation sequencing and converted to binding scores that report binding relative to the wild-type sequence. Alternatively, display methods are *in vitro* methods that can be used to measure binding

affinity. The protein library of interest is displayed on the cell surface using a cell membrane tag as an anchor[56,83]. The displayed library is probed with a fluorescently labeled antibody and sorted using FACS, and similarly scored as in complementation methods.

In this section, we focus on studies in which the primary interrogated property is binding. DMS binding studies have been used to identify binding interfaces, reengineer protein-protein interfaces to improve affinity, determine key positions of drug resistance, as well as identify specificity-conferring positions. We discuss trends that have emerged in each of these studies as well as challenges and considerations for binding mutational studies.

2.3.2 in vitro binding

Most studies examining protein binding through deep mutational scans use *in vitro* methods, which are typically display-based assays. Similar to *in vitro* approaches for assessing protein stability, the variant library of displayed proteins is screened using fluorescent substrates and sorted based on fluorescence signals using FACS (**Figure 2.3A**). The sorted library is subsequently quantified using NGS to determine binding scores. *In vitro* binding scans provide an unbiased experimental approach to identify binding determinants, independent of confounding cellular factors such as chaperones, thereby focusing solely on the inherent binding characteristics of the sequence with its cognate protein.

With variant effect maps of *in vitro* binding, analysis of primarily deleterious binding variants reveals that binding determinants largely conform to known mechanisms of protein binding, with disruptions of hydrophobic interactions at the binding interface decreasing binding. However, when gain-of-function binding variants are considered, it becomes clear that the determinants of binding are influenced by the structure of the protein being examined. For example, the mutational scanning of antibodies has demonstrated that peripheral positions are more tolerant to substitution and can be engineered for affinity maturation[83,84]. A mutational scan of the complementary determining regions (CDRs) of an antibody Fab fragment against tumor necrosis factor-alpha receptor (TNFaR) combined seven affinity-increasing mutations to increase the affinity over 2000 fold to a Kd of 3.45 pM. Conversely, mutational scans of *de novo* protein binding reveal that the positions at the designed interface are

intolerant to substitution, validating the designed interface[37,38,85]. Beyond the designed interface, most substitutions are neutral or deleterious. A scan of de novo H1N1 hemagglutinin inhibitor revealed higher-affinity substitutions located at the periphery of the binding interface that increased electrostatic interactions[85]. The authors combined five of the gain-of-function mutations revealed in the scan, resulting in a only 25-fold improvement in binding affinity (from 15.2 nM Kd to 0.6 nM Kd). For both the scan of TNFaR antibody and H1N1 inhibitor, the combinatorial variants achieved higher affinity than the sum of affinity improvements of the individual mutations. However, combining gain-of-function variants does not always result in additive changes to affinity. The process of identifying additive or superior combinations of gain-of-function variants remains poorly understood. Across multiple affinity scans, the presence of pairwise and higher-order epistatic interactions, where the combined effect of multiple binding residues exceeds the sum of their individual effects, has been consistently observed at binding interfaces[86–88].

It is important to note that the binding interfaces of *de novo* designed proteins are typically composed of rigid, noncontiguous helices and are therefore smaller and less flexible compared to proteins that mediate binding through unstructured loops, such as antibodies. When designing mutational scans for affinity maturation, the nature of the binding interface should be carefully considered. A small, rigid binding interface has limited potential for affinity maturation compared to an extensive, flexible binding interface like that of an antibody. The effectiveness of DMS for affinity maturation has been modest, primarily because the most successful applications of DMS in this context have necessitated the combination of gain-of-function variants. Since further optimization through low-throughput methods is often required, employing large random libraries containing combinatorial variants near the periphery of the binding interface is more likely to yield high-affinity hits compared to deep mutational scans focusing on single variants.

The field of binding mutational scans has been dominated by *in vitro* scans due to their simplicity. Display methods offer finer control over the substrates used to probe binding interactions, resulting in more readily interpretable results. However, *in vitro* scans are not well-suited for studying nucleic acid binding events, such as transcription factor binding, primarily due to the absence of cellular cofactors that influence transcription factor binding interactions. Additionally, the requirement for substantial quantities of labeled target nucleic acid for the library scan poses a limitation in this context. Despite these limitations,

in vitro methods are uniquely suited to study variant effects on binding and to generate variant-binding maps to guide affinity and specificity engineering.

2.3.3 Cell based methods for measuring protein binding

Cell-based methods offer an alternative approach for investigating protein binding through deep mutational scanning. By analyzing protein interactions within a cellular environment, these techniques provide valuable insights into the intricate nature of protein binding. In a physiological context, cell-based methods facilitate the examination of protein binding with respect to other factors such as cellular compartments and post-translational modifications. Typically conducted in yeast or bacteria, cell-based methods include complementation or two-hybrid assays. These methods select for variants of interest via growth assays. Due to the use of growth as a readout, the relationship between binding and growth in cell-based methods can be nonlinear and indirect, and therefore more challenging to interpret compared to *in vitro* methods. Primarily, these techniques are used to study protein-nucleic acid binding of enzymes and regulatory sequences, which is more challenging to study using *in vitro* display methods.

Multiple deep mutational scans of DNA-binding proteins have yielded a model for specificity engineering that involves the synergistic combination of permissive mutations and specificity determining residues[89,90]. Permissive mutations are those that allow new residues to make contact with the substrate-of-interest. For example, deep mutational scans of the DNA-binding proteins parS and NBS, Jalal et al. found that a permissive lysine mutation adjacent to the binding interface conferred DNA-binding capability[90]. When they combined their DMS data with x-ray crystallography, they found that the lysine mutations permitted specificity-determining residues to make DNA-contacts. This model of combining permissive mutations to allow specificity engineering has been previously observed in evolutionary and directed evolution studies[91–94]. Permissive mutations are typically identified through ancestral protein reconstruction, a method that involves inferring the amino acid sequences of ancestral proteins from which DNA-binding proteins like parS and NBS have evolved[93]. The integration of evolutionary information is crucial for identifying permissive mutations in deep mutational scanning studies. Therefore, while DMSs are not inherently superior to evolutionary modeling for reengineering

protein DNA-binding specificity, the comprehensive variant-binding map obtained through DMS provides valuable insights into the molecular basis of altering protein-DNA specificity. Consequently, DMSs are better suited for elucidating the molecular mechanisms that underlie specificity changes. However, in scenarios where limited evolutionary information is available for binding a particular ligand, DMS proves more advantageous than evolutionary modeling for identifying permissive and specificity-determining residues.

Cell based methods are not limited to assaying nucleic acid binding, however. Cell-based methods can be used to map binding interfaces, engineer affinity, and engineer specificity in the context of post-translational modifications and cellular compartment factors that are not present in *in vitro* scans. Like *in vitro* scans of binding, these studies have revealed that specificity is primarily dictated by a small set of residues[89,95]. Notably, improvements in specificity have often been achieved through the combination of multiple mutations that abolish binding to one substrate but not another. For example, a complementation assay measuring the binding of the PDZ domain to the wild-type ligand and a non-native peptide ligand identified a subset of mutations that have opposite effects on binding the two ligands. Mutations that were deleterious or neutral to the wild-type ligand improved binding to the non-native ligand. These mutations were located primarily at the periphery of the binding interface, with one position directly contacting one of the ligands. By combining two of the mutations in the subset, the authors altered the specificity of PDZ domain for a 45-fold preference for the non-native ligand, demonstrating the utility of mutational scans in improving specificity of binding interfaces[96]. These studies demonstrate how comparative variant function maps can be used for switching specificity for different ligands or improving specificity.

2.3.4 Comparisons to computational predictors

While computational predictors, such as FoldX and Rosetta, offer noteworthy predictive power, deep mutational scanning can reveal trends that are not immediately apparent from such models. For example, there is modest agreement between FoldX and Rosetta predictions and the binding affinities measured from a yeast display deep mutational scan of a cohesin protein. The correlation between

measured and predicted binding affinity is likely somewhat limited due to affinity-modulating positions distal to the binding interface. These distal positions, which can be readily identified through deep mutational scans, often do not significantly contribute to the predicted binding energy in computational predictors such as FoldX and Rosetta. This discrepancy arises because considering the distal positions in the computational models would require additional computational resources and increased complexity[7797,98]. Thus, deep mutational scanning is more useful than computational predictors for identifying positions that can be further engineered to improve affinity and specificity. Indeed, a comparison of mutational scan data to Rosetta energy predictions showed that Rosetta correctly identified mutationally tolerant positions in a therapeutic cyclic peptide[98]. However, it did not accurately predict affinity-enhancing mutations, likely due to its computation of binding energy being restricted to directly interacting residues. The difficulty of computational predictors in predicting affinity-enhancing mutations underscores the value of deep mutational scanning in engineering native and *de novo* proteins for the development of new biochemical tools and therapeutics. Taken together, the comprehensive nature of binding DMSs increases the probability of identifying sequence variants with improved affinity or desired specificity compared to computational predictors like FoldX and Rosetta.

To date, deep mutational scans have been conducted in nearly 100 peer-reviewed studies, measuring over 2 million variant effects on binding (**Supplementary Table 2.2**). In each study, most variants have been found to have a deleterious effect on binding, and only a small percentage of scanned variants, less than 5%, are measured as gain-of-function compared to the wild-type sequence. Furthermore, gain-of-function variants were typically found adjacent to critical binding residues, rather than deep in the binding pocket. Although combinations of single gain-of-function variants have been used to improve binding affinity, the potential affinity maturation is constrained for rigid binding interfaces common to *de novo* protein binders.

The development of a variant-binding map through the utilization of deep mutational scan holds immense importance for affinity maturation due to its unbiased approach. Unlike directed evolution, deep mutational scanning allows for a comprehensive exploration of all possible missense mutagenesis trajectories. While the increase in binding affinity resulting from a single missense mutation is usually moderate, the combination of multiple affinity-enhancing mutations identified through deep mutational

scanning can lead to an improvement in affinity of over 2000-fold[83,84]. Through the use of deep mutational scanning, researchers have been able to identify mutations that enhance affinity in native proteins, therapeutic antibodies, and de novo proteins, providing a means to reengineer higher affinity variants[83,8486–88].

Additionally, the use of deep mutational scans has proven to be a valuable tool in the study of specificity. This approach involves conducting multiple scans of a protein variant library with different substrates to identify specificity determining mutations. These mutations are characterized by their ability to confer strong binding to one substrate while depleting binding to another. Deep mutational scanning has been successfully applied in investigations of the specificity determinants of protein-protein, protein-DNA, and protein-small molecule binding[77,81,82,89,90,95,99–101]. Taken together, deep mutational scans of binding open up exciting new possibilities for targeted protein engineering and drug development.

2.4 Activity scans

2.4.1 Introduction

Mutational studies of protein activity have been used for a range of applications, including clinical variant interpretation, protein engineering for therapeutics and biochemical tools, as well as evolutionary biology. In this section, we define protein activity broadly to include various functional protein classes, such as transcription factors, proteases, DNA polymerases, and cell surface receptors.

While hydrophobic packing can explain protein folding and substrate binding, protein activity is often the result of multiple functional domains working together to propagate signals to perform their encoded functions. Deep mutational scanning is uniquely compatible with studies of protein activity, as it can assay a broad range of protein functions for thousands of variants in a single experiment. Additionally, deep mutational scans can also be used to calculate enzyme rate constants for thousands of variants in a single experiment[102,103]. Researchers have used these sequence-function maps to identify potential drug resistance mutations, activity-modulating residues, and motifs that can be used to engineer gain-of-function variants[37,104,105].

Mutational scans of activity typically use growth rates or fluorescence levels for selection, with growth assays being particularly common[104,106107,108]. However, the relationship between protein function and cell fitness is not always linear. Additionally, yeast and bacterial expression systems may not express the same array of chaperones and activators as mammalian cells, which can alter the measured fitness of protein variants in a mutational scan[109–112]. Fluorescence-based methods have been used to evaluate the functional properties of various proteins, including voltage-based fluorescent dyes and fluorescent protein tags, which have been employed together with fluorescence activated cell sorting to measure the ion conductance of a library of ion channel variants[65,67113,114]. However, the use of a dye may have a limited dynamic range, which can limit the fidelity of functional scores of the assay. Microfluidic assays offer more precise control over the time scales of reactions to measure changes in activity, but require a fluorescent substrate for the enzyme of interest[89,115].

Unlike stability and binding, protein activity differs amongst classes of molecules, and generalizations cannot be made across functionally distinct proteins. However, characterizing the mutational tolerance of variants with respect to activity can provide valuable insights into the mechanisms underlying protein function and regulation. This section, while not comprehensive, highlights the applications of deep mutational scanning to engineering gain-of-function variants, identifying extended activity networks, and explores structural trends that have appeared across mutational scans of activity.

2.4.2 in vitro activity

The measurement of protein activity using *in vitro* deep mutational scans is more challenging compared to stability and binding assays, which have readily available display methods. To address this issue, droplet-based assays have been combined with deep mutational scanning to measure the effects of variants on protein activity[113,115]. In this method, a library of cells expressing the protein of interest is suspended in oil droplets containing fluorogenic substrate and lysis reagents. Upon lysis, the protein variant is released and free to bind the substrate, and the resultant fluorescent product is used to sort and quantify the encapsulating droplets by FACS and NGS, respectively. This approach allows the control of reaction timescales and measurement of enzyme rate constants. Additionally, the method allows the

measurement of activity under conditions not accessible by other expression systems, such as elevated temperature.

Despite the method's advantages, droplet-based deep mutational scanning is technically challenging and requires considerable investment to develop a compatible workflow. Thus far, the method has been used for DNA polymerase activity, caspase activity, and glycosidase activity[89,113,115]. Unlike scans of stability and binding, findings from one enzyme class cannot be generalized to another, making it necessary to conduct specific scans for each enzyme class. The droplet-based method has been useful in characterizing enzymatic constants and determining the mutational tolerance of each enzyme[115]. Like stability and binding assays, core and active site positions are more mutationally intolerant than surface positions. The method's primary advantage is the ability to obtain meaningful enzymatic constants and measure activity under different conditions, which is useful in directing protein engineering efforts to enhance specificity and optimize enzyme activity. However, key limitations of droplet-based DMSs include the technical challenges of droplet-based mutational scans that limit widespread use and limited availability of fluorescent reporters with a large dynamic range for multiplexed assays. Thus, droplet-based DMSs are most suited for studies that require precise measures of activity or multiplexed measurements of enzymatic constants. Inquiries that are only concerned with relative changes in activity, such as defining activity modulating residues or engineering activity, can be accomplished with less technically challenging activity DMSs.

2.4.3 Cell based methods for measuring protein activity

Cell-based deep mutational scanning of protein activity enables the assessment of protein activity within the cellular environment, where factors such as chaperones, cofactors, and post-translational modifications can influence protein function. Unlike *in vitro* assays, cell-based assays can provide a more physiologically relevant environment for studying protein activity.

Compared to *in vitro* deep mutational scanning, the cell-based approach offers several advantages. It allows for the identification of variants that confer improved function under conditions that are difficult to mimic *in vitro*, making it a valuable tool for discovering novel variants with practical

applications. However, it is important to note that the cell-based approach has some limitations. For instance, determining the specific effects of mutations on protein activity can be challenging when mutations also affect other aspects of protein function, such as stability, localization, post-translational modifications, and cellular steady-state levels. Furthermore, the presence of multiple copies of the mutated gene in the genome of the host cell can lead to genetic interactions that may complicate the interpretation of results. Thus, mutational scanning studies of activity are best performed in cell lines with the protein-of-interest knocked out with either a co-translational expression control or a parallel scan in near identical conditions that measures the cellular abundance of variants to normalize for changes in stability and expression.

While the diversity of enzyme classes precludes broad generalizations regarding determinants of activity, we highlight several studies that have successfully applied cell-based deep mutational scanning to identify important determinants of protein activity. These studies demonstrate the potential of this approach for uncovering key functional elements and for guiding the design of proteins with enhanced activity or specificity.

2.4.3.1 Deep mutational scans of activity for drug development

The mutational tolerance of individual positions within a protein domain is a valuable metric for drug development. Identifying mutationally tolerant positions can predict regions that are likely to develop drug resistance, while identifying mutationally intolerant positions can highlight targets for therapeutic interventions that are unlikely to evolve resistance.

Traditionally, evolutionary conservation has been used to determine the mutational tolerance of positions within a protein. However, this approach has limitations, as it relies on a preconceived notion of the protein's active site, which is usually inferred from the protein's structural information. Moreover, functional mutational scans of various proteins have revealed that evolutionary conservation does not always correspond to experimentally measured mutational tolerance. For instance, functional mutational scans of PTEN and Hsp90 identified positions that were highly conserved but mutationally tolerant[107,108]. This was unexpected, as the middle domain of Hsp90 responsible for client binding, which is usually

highly conserved, was mutationally tolerant. Conversely, some scans have found that the mutational tolerance observed in their deep mutational scans aligns with evolutionary conservation. For example, functional mutational scans of bacterial toxin HokC and T4 bacteriophage sliding clamp identified mutationally intolerant positions that were highly conserved[106,116].

In a functional scan of SARS-CoV-2 main protease, where the experimental mutational tolerance corresponded to relative conservation levels, the mutational tolerance of individual positions within the active site varied significantly[117]. These mutational scans demonstrate that experimentally derived mutational tolerance is not consistently correlated with evolutionary conservation. Therefore, drug development efforts should leverage deep mutational scanning data when available to identify ideal targets for therapeutic interventions that are unlikely to evolve resistance.

2.4.3.2 Extended activity networks

Proteins perform their function by catalyzing reactions in their active site, a structural groove where substrates bind. However, the function of a protein often involves residues that are distal to the active site, and these residues form an extended activity network. While it is relatively easy to identify active site residues using structural studies, identifying distal residues that modulate protein activity can be more challenging. Deep mutational scanning is a powerful tool that can identify extended activity networks beyond the active site.

By using appropriate controls to eliminate changes in expression or folding as factors influencing activity, deep mutational scanning can identify residues in extended activity networks that modulate protein function. Alternatively, follow-up biochemical experiments can directly measure the activity of individual variants. Residues in extended activity networks can be potential sources of disease pathology or targets for drug development. Furthermore, comprehensive mapping of extended activity networks can reveal fundamental principles of protein activity that can guide protein engineering for improved activity.

Extended activity networks can include multiple protein domains. For example, a mutational scan of the functional activity of a library of T4 bacteriophage sliding clamp variants identified a central residue that completes a hydrogen bonding network between ATP-binding sites in the T4 bacteriophage

clamp-loader complex, demonstrating an extended activity network at the residue level[116]. In another case, an extended network of interactions in the K+ ion channel Kir2.1 was found to form a hydrophobic interaction network that, when disrupted, resulted in gain-of-function activity[65]. Finally, a deep mutational scan of Src kinase's catalytic domain revealed a novel autoinhibitory interaction between the SH4 domain and the catalytic domain[104]. Mutations at the interface between the two domains resulted in gain-of-function variants, demonstrating an extended activity network identified at the protein domain level[104].

Deep mutational scanning has been successfully utilized to identify extended activity networks beyond the active site. Residues in these networks modulate protein function, and their comprehensive mapping can improve our understanding of sequence-activity relationships to guide protein engineering efforts for improved activity.

2.4.3.3 Structural trends and biochemical trends that modulate activity

In addition to identifying extended networks of residues that modulate activity, DMS have the potential to reveal structural and biochemical trends that can be leveraged to create higher activity variants of the protein of interest or its related homologs[43,67,118]. As the number of mutational scans conducted on diverse proteins increases, it may be possible to extrapolate trends across related proteins and domains, enhancing rational design and computational modeling of various protein functions.

DMS can uncover structural and biochemical motifs that enhance activity, especially in difficult-to-characterize structures such as intrinsically disordered loops. For example, a recent DMS on multiple acidic activation domains of transcription factors revealed the "acidic exposure model" in which acidic and hydrophobic interactions balance to enhance activity[67]. This model facilitated the rational design of higher-activity acidic domains and development of a prediction model to identify other acidic domains in the human proteome. Another DMS on the MTHFR protein identified a structural motif in a disordered loop that facilitated retention of the cofactor within the active site, resulting in increased activity[43]. These examples demonstrate the utility of DMS in identifying activity-modulating structural and biochemical motifs that are challenging to detect via conventional methods.

In summary, over 100 peer-reviewed DMS studies of various protein activities have been conducted, measuring nearly 1 million variant effects on activity (**Supplementary Table 2.3**). These studies have focused on essential genes such as Hsp90 and ubiquitin, clinically relevant genes such as TP53 and BRCA1, and viral proteins, and have been evenly distributed across yeast, human, and bacterial expression systems. Activity deep mutational scans provide valuable insights into mutational tolerance and identify extended activity networks that modulate activity. However, when interpreting the results of an activity mutational scan, it is crucial to consider that changes in activity could arise from various underlying factors, such as alterations in protein expression levels, stability and folding, catalytic rate constants, substrate specificity, allosteric regulation, and more. Therefore, functional scores from a single activity assay may not be sufficient to distinguish the independent contributions of each factor to the observed changes in activity. To address this limitation, researchers can perform expression controls and conduct parallel scans of other protein functionalities to isolate specific determinants of protein variant activity and to confirm the accuracy of the measurements. By combining data from multiple assays, a more comprehensive picture of protein function can be obtained, which could aid in the design of more effective therapeutic interventions.

Overall, deep mutational scans of protein activity offer crucial insights into mutational tolerance and activity networks. These insights have the potential to inform the development of more targeted therapies for essential and clinically relevant genes, as well as viral proteins. As the number of DMS studies continues to increase, so does the potential for deeper understanding of the relationships between mutational tolerance, activity networks, and protein function.

2.5 Concluding remarks

In summary, deep mutational scanning has emerged as a highly effective tool for investigating the relationship between protein sequence, structure, and function. Deep mutational scans have also identified specific mutations, positions, and motifs that can enhance binding, stability, and activity to support protein engineering efforts. To unlock extended activity networks, allosteric switches, and other functional components in an unbiased way, future DMS studies should consider conducting scans of the

same protein on different parameters, such as stability and activity. This approach can improve the probability of identifying critical activity-modulating residues by restricting analyses to activity-modulating variants that have little effect on stability. By utilizing multi-parameter deep mutational scanning, researchers can gain a more comprehensive understanding of protein function, facilitate protein engineering, and decipher the underlying mechanisms of disease pathology from sequence variant effects.

As more deep mutational scans are conducted on protein targets across different parameters, several exciting possibilities are likely to emerge. Deep mutational scanning has the potential to unlock the secrets of protein function, facilitate protein engineering, and enhance our understanding of the mechanisms underlying disease pathology.

Chapter 3. Molecular determinants of Hsp90 dependence of Src kinase revealed by deep mutational scanning

This chapter is adapted from the published manuscript **Nguyen**, V., Ahler, E., Sitko, K.A., Stephany, J.J., Maly, D.J. and Fowler, D.M. (2023), Molecular determinants of Hsp90 dependence of Src kinase revealed by deep mutational scanning. Protein Science. Accepted Author Manuscript e4656. https://doi.org/10.1002/pro.4656

Abstract

Hsp90 is a molecular chaperone involved in the refolding and activation of numerous protein substrates referred to as clients. While the molecular determinants of Hsp90 client specificity are poorly understood and limited to a handful of client proteins, strong clients, proteins that are functionally dependent on Hsp90, are thought to be destabilized and conformationally extended. Here, we measured the phosphotransferase activity of 3,929 variants of the tyrosine kinase Src, a weak client of Hsp90, in both the presence and absence of an Hsp90 inhibitor. We identified 84 previously unknown functionally dependent client variants. Unexpectedly, many destabilized or extended variants were not functionally dependent on Hsp90. Instead, functionally dependent client variants were clustered in the αF pocket and β1-β2 strand regions of Src, which have yet to be described in driving Hsp90 dependence. Hsp90 dependence was also strongly correlated with kinase activity. We found that a combination of activation, global extension, and general conformational flexibility, primarily induced by variants at the αF pocket and β1-β2 strands, was necessary to render Src functionally dependent on Hsp90. Moreover, the degree of activation and flexibility required to transform Src into a functionally dependent client varied with variant location, suggesting that a combination of regulatory domain disengagement and catalytic domain flexibility are required for chaperone dependence. Thus, by studying the chaperone dependence of a massive number of variants, we highlight factors driving Hsp90 client specificity and propose a model of chaperone-kinase interactions.

3.1 Introduction

Heat shock protein 90 (Hsp90) is a molecular chaperone that assists in the maturation and folding of other proteins, called clients[119]. Hsp90 clients are diverse proteins involved in signal transduction, including transcription factors, kinases, and steroid hormone receptors. Hsp90 can stabilize proteins that lead to cancer progression and neurodegeneration, and thus plays an important role in the progression of these and other diseases[120,121].

Because of the role of Hsp90 in disease, small molecule Hsp90 inhibitors have been developed that block Hsp90's interaction with client kinases, often leading to client aggregation or degradation[122]. For Hsp90 inhibitor treatment to deplete disease-causing proteins through aggregation or degradation, the protein must be a strong client, meaning the target protein's activity has a strong dependence on Hsp90 chaperoning. However, the mechanism driving Hsp90 client specificity, critical for Hsp90 inhibitor treatment, remains elusive.

The challenge of characterizing the determinants of Hsp90 client specificity is best demonstrated by highly-conserved kinases that exhibit divergent client statuses. For example, ErbB-1 and ErbB-2, closely related receptor tyrosine kinases, have different client strengths owing to the disparate hydrophobic properties of their αC-β4 loops[30,123]. v-Src and c-Src are non-receptor tyrosine kinases that share 98% protein sequence homology. While c-Src is not a strong client of Hsp90, its oncogenic viral descendant, v-Src is a strong client[124]. A single amino acid change in c-Src, Src_{E381K}, can increase its interaction with Hsp90 by 30% compared to c-Src[125]. Biochemical studies of c-Src and v-Src, which differ by only a handful of amino acids, suggest that Hsp90 client status depends on the differing intrinsic stability and aggregation propensity of each kinase[30,122,126]. Molecular dynamics studies of v-Src and c-Src also suggest that Hsp90 preferentially recognizes the extended, activated conformation of this multi-domain kinase, which is less thermodynamically stable than the closed, inactivated conformation[126]. More generally, high-throughput approaches have been employed to measure the interaction strength between Hsp90 and hundreds of different human kinases, revealing a correlation between Hsp90 interaction strength and thermostability for a subset of kinases[125]. However, this approach did not reveal an Hsp90 sequence recognition motif and did not measure the functional consequences of kinases interacting with the chaperone.

Thus, many questions surrounding Hsp90 kinase client recognition and processing remain unanswered. For example, are there other kinase structural motifs like the αC-β4 loop that are recognized by Hsp90? What, if any, factors besides stability and surface hydrophobicity affect Hsp90 dependence? To what extent does Hsp90 recognize client kinases through alterations in kinase conformation? Of the client kinases recognized by Hsp90, which decrease in activity following Hsp90 inhibition? Finding answers to these questions is challenging due to the sequence diversity of Hsp90 clients. However, the advent of deep mutational scanning, where the effect of nearly all possible single amino acid variants of a protein can be measured simultaneously, offers a way to understand the physicochemical and structural bases of protein activity and proteostasis[39]. For example, we previously used deep mutational scanning to reveal mechanistic details of kinase regulation through changes in phosphotransferase activity[104]. In other examples, deep mutational scans of influenza hemagglutinin, metabolic enzyme dihydrofolate reductase (DHFR), and membrane protein rhodopsin in different proteostasis environments revealed how changes in mutational tolerance could impact a protein's allowable evolutionary sequence space[71,75,102].

Here, we explore Hsp90 dependence and client processing, combining deep mutational scanning, proteostasis perturbations, and computational approaches in the context of the model Hsp90 client, Src kinase. Src, a non-receptor tyrosine kinase and one of the Src family kinases (SFKs), functions as a key mediator of signal transduction[127]. Once activated by cell-surface receptors, Src interacts with substrates to drive cell proliferation, differentiation, angiogenesis, and cell survival[128,129]. Like ~50% of human kinases, Src is composed of a catalytic domain and multiple accessory domains that regulate its activity[130]. Aside from the catalytic domain (CD), Src contains the unique, Src homology 4 (SH4), SH3, and SH2 domains in addition to a regulatory C-terminal tail (**Figure 3.1A-B**). These regulatory domains act in concert to hold Src in its closed, inactive conformation. In the closed conformation, the SH4 domain forms an interface with the αF pocket on the CD while the SH3 and SH2 domains interact with the N- and C- lobes of the CD, respectively (**Figure 3.1C**)[104,131–133]. Phosphorylation at Y530 in the C-terminal tail strengthens intramolecular engagement of the SH2 domain to further repress Src activity[134]. Competitive association of the SH4, SH3, and SH2 domains with other intracellular interactors, as well as phosphorylation at Y419 release these autoinhibitory interactions and activate Src[104,135,136]. Mutations that

disrupt autoinhibition allow Src to spontaneously access its active conformation and can lead to aberrant Src activity and oncogenesis[128,129,137].

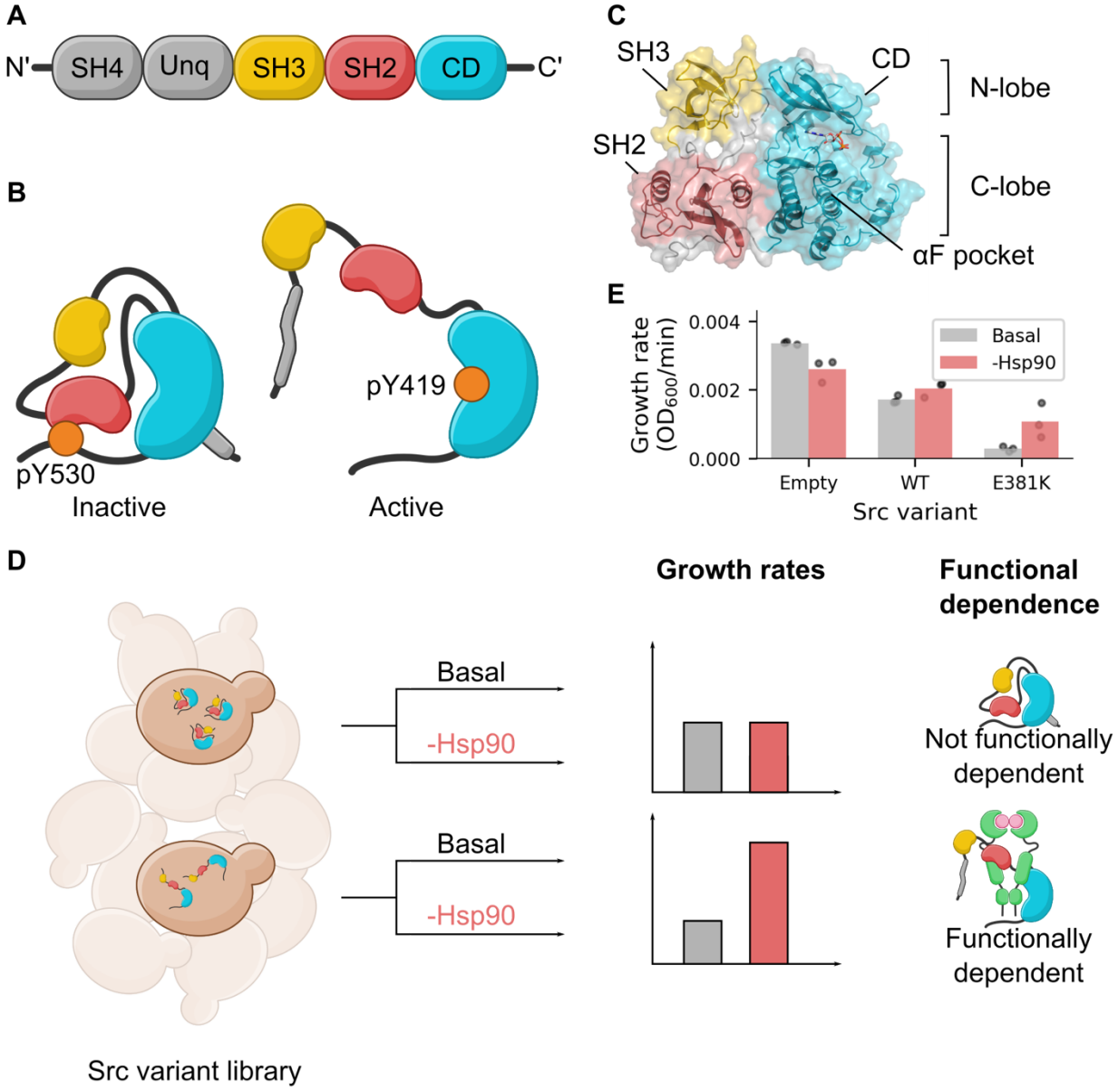


Figure 3.1. (A) Domain organization (SH4, Unique, SH3, SH2, and CD) of Src kinase. (B) In a closed conformation Src is autoinhibited by the SH4, SH3, and SH2 domains (left) that, when released, yield Src's extended, active conformation (right). (C) Structure of Src in the inactive, closed conformation (PDB ID: 2SRC) showing the N-lobe, C-lobe, and the αF pocket. The SH2 domain is shown in red and the SH3 domain is shown in yellow. (D) The Src variant library was grown in basal conditions (DMSO) and Hsp90 depleted conditions (radicicol) and differences in growth rates were used to determine variant Hsp90 dependence. (E) Individually measured growth rates of weak client SrcWT and functionally dependent client E381K in radicicol and DMSO (n=3).

We evaluated the effect of 3,929 Src CD variants across 250 residues on phosphotransferase activity in both the presence and absence of an Hsp90 inhibitor. By measuring change in Src variant phosphotransferase activity following Hsp90 inhibition, we identified 84 previously unknown functionally dependent client variants. These variants were distributed throughout the CD, but clustered in the β1-β2 strands and αF pocket regions of Src, previously unknown to drive Hsp90 interaction. While the E381K mutation in the αF pocket had previously been identified as a driver for Hsp90 interaction, the involvement of the remainder of the αF pocket in Hsp90 dependence has yet to be shown. Comparatively few client variants occurred in the regulatory SH3-CD and SH2-CD interfaces, indicating that Hsp90 recognition of Src does not occur solely due to disruption of regulatory interactions. Rather, Hsp90 dependence and conformational flexibility strongly correlated with kinase activity. We found that a combination of activation and the general flexibility of Src's global conformation, primarily induced by variants at the αF pocket and β1-β2 strands, was necessary to transform Src into a functionally dependent client. Moreover, the degree of activation and extension required to transform Src into a functionally dependent client changed depending on the variant location. Thus, our data suggest that regulatory domain disengagement and catalytic domain flexibility are both required to transform Src into a functionally dependent client. Our work demonstrates the utility of using deep mutational scanning to study the functional consequences of complex and dynamic protein-protein interactions such as protein chaperoning.

3.2 Multiplexed measurement of Hsp90 effects on Src variant activity

We previously used a *S. cerevisiae* assay to conduct a deep mutational scan of the c-Src CD, measuring the effects of over 3,500 c-Src variants on the growth of the BY4741 Green Monster strain (see Methods for genotype)[104,138]. In this assay, Src is expressed from a plasmid under the control of a galactose inducible promoter. Src activity disrupts a yeast spore wall remodeling pathway and limits the growth rate of cells[139]. We showed that growth rates of yeast expressing different Src variants correlated strongly with *in vitro* phosphotransferase activity as well as yeast and cultured human cell phosphotyrosine levels[104].

We reasoned that we could use the yeast assay to measure the effect of Hsp90 on Src activity and thus comprehensively characterize the Hsp90 dependence of thousands of variants. Yeast Hsp90 is 60% identical to the human isoform Hsp90α and has been used as a model to investigate Hsp90 dependence of human Src variants[122,140,141]. The activity of functionally dependent clients depends on Hsp90, and thus changes in yeast growth following treatment with an Hsp90 inhibitor reflect changes in Src-mediated toxicity[142]. In particular, the growth of yeast expressing functionally dependent client Src variants is rescued upon Hsp90 inhibition because of these variants' dependence on Hsp90[122] (**Figure 3.1D**). To inhibit Hsp90, we used radicicol, a small molecule that binds to Hsp90's N-terminal domain and disrupts its function through competitive binding of the ATP-binding pocket. We constructed a radicicol dose response curve for yeast expressing Src_{WT}, Src_{E381K}, and an empty vector control to identify a concentration that resulted in the expected moderate increase in growth rate in cells expressing the weak client Src_{WT} (**Supplementary Fig 1**) and a larger increase in growth rate in cells expressing the functionally dependent client Src_{E381K}. We selected 800nM radicicol which, while somewhat toxic, increased the growth of yeast expressing Src_{WT} and greatly increased the growth of yeast expressing Src_{E381K}. (**Figure 3.1E**)[125]. Thus, yeast growth rates could be used to measure the effect of Hsp90 inhibition on Src variant-mediated changes in yeast growth and, by proxy, Src phosphotransferase activity.

We transformed a previously generated library of ~4,000 Src CD variants into yeast, induced Src expression, cultured the library in the presence of radicicol, and withdrew samples from each culture periodically[104]. We deeply sequenced each sample to quantify the frequency of each variant at each time point. We converted these frequencies into variant activity scores by taking the ratio of frequencies of each variant relative to the Src_{WT} frequency at each time point and performing a weighted linear regression in which the negative slope of the line was the activity score[143]. We normalized activity scores such that Src_{WT} had a score of 1. Using this method, we calculated the activity of 3,929 (3,366 nonsynonymous, 430 synonymous, and 133 nonsense) Src variants in both basal and Hsp90 inhibited conditions, representing 73% of possible CD single variants. We collected two replicates, which were well correlated (R = 0.86, rho = 0.86; **Figure 3.2A**).

3.3 Determining the Hsp90 dependence of ~3,500 Src missense variants

In order to quantify each variant's dependence on Hsp90 chaperoning, we compared the activity scores measured in the radicicol treated condition to previously measured DMSO activity scores generated using the same library and assay[104]. Src_{WT}, to which each set of scores was normalized, had different growth rates in the two conditions due to slight radicicol toxicity. Thus, the radicicol activity scores reflect growth rates affected by both radicicol toxicity and Src variant toxicity, making direct comparisons with DMSO activity scores, which only represent growth rates affected by Src variant toxicity, difficult. To enable direct comparison of variant activity scores in the two conditions, we calibrated how each set of activity scores related to growth rate by individually measuring the growth rates of yeast expressing 13 variants that spanned the range of the activity scores. Individually measured growth rates were well correlated with activity scores derived from the deep mutational scans in both conditions (**Figure 3.2B**; DMSO, R = 0.89, rho = 0.92; radicicol, R = 0.86, rho = 0.86). We used a linear model to transform radicicol activity scores to yeast growth rates and then used a second linear model to transform yeast growth rates to DMSO-equivalent activity scores. We refer to the result of the transformed radicicol activity scores as "calibrated radicicol activity scores," in which a variant with a calibrated radicicol activity score of 0 has the same growth rate as a variant with a DMSO activity score of 0. We then calculated a "client score" for each variant by taking the difference between the calibrated radicicol activity score and the measured DMSO activity score (see Methods). Client scores ranged from -3.04 to 2.56, with negative values indicating a strong dependence on Hsp90 (**Figure 3.2C, D**). The inactive variant Src_{K298M} had a positive client score of 1.09. The weak client Src_{WT} had a client score of 0.27, indicating that weak clients in our library have client scores close to 0, similar to Src_{WT}. To identify functionally dependent Hsp90 clients, we used synonymous variants to define the range of WT-like client scores. We considered all variants whose client scores were below two standard deviations of the mean client score of synonymous variants functionally dependent clients (**Figure 3.2D**). 84 variants had client scores that met this definition of a functionally dependent client.

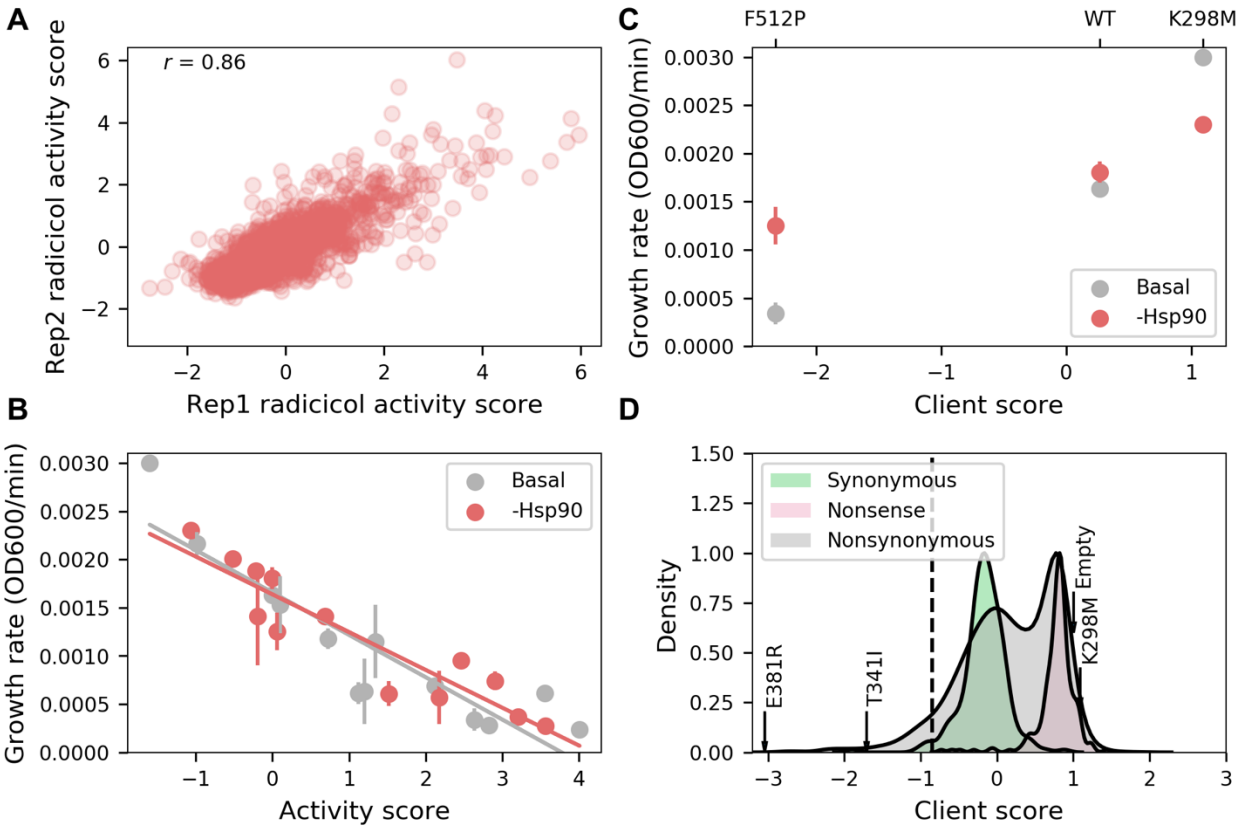


Figure 3.2. (A) Activity scores for 3,929 of Src catalytic domain variants in two radicicol-treated replicates. (B) Linear regression of activity scores and individually measured growth rates (n=3) of 13 variants in either radicicol- or DMSO-treated conditions. These linear models were used to convert radicicol activity scores to calibrated radicicol activity scores, which were then used to calculate a client score for each variant by subtracting the DMSO activity score from the calibrated radicicol activity score. (C) Individually measured growth rates of yeast expressing nonclient SrcK298M, weak client SrcWT, or functionally dependent client SrcF512P in either basal or Hsp90-inhibited conditions of (n=3), along with each variant's client score. (D) Distribution of nonsynonymous (n = 3,366, gray), synonymous (n = 430, green), and nonsense (N = 133, red) client scores of variants in the library. Dashed line indicates used to define functionally dependent clients. Functionally dependent client variants E381R and T341I are indicated by arrows. The client score of kinase-dead variant K298M and the apparent client score of empty vector are indicated by arrows over the nonsense distribution.

To validate our client scores, we tested 3 functionally dependent client variants individually and observed large increases in growth rates in radicicol relative to DMSO, as suggested by their large negative client scores (**Supplementary Figure 2A**). We also examined previously known and suspected functionally dependent clients[144,145]. Src_{E381K} increases Hsp90 interaction compared to Src_{WT}[125]. While Src_{E381K} was not present in our library, a variant with a similar positive charge at the same position, Src_{E381R}, was a functionally dependent client with a score of -3.04. We also examined Src_{T341I}, a "gatekeeper mutation" that confers tyrosine kinase inhibitor resistance in many different kinases[146]. The

oncogenic activity of kinases harboring gatekeeper mutations has been successfully repressed in mice using Hsp90 inhibitors, suggesting that kinases with the gatekeeper mutation are functionally dependent clients of Hsp90[147,148]. Src_{T341I} was also classified as a functionally dependent client with a score of -1.71. Thus, client scores accurately recapitulate the effects of Hsp90 on Src activity in yeast, and reveal Src variants that are strongly dependent on Hsp90.

We also observed 111 variants with positive client scores beyond two standard deviations of the synonymous distribution, suggesting that these variants might increase in activity following Hsp90 inhibition. Indeed, an increase in tyrosine kinase activity has been reported as a transient phenomena following Hsp90 inhibition in T24 bladder carcinoma cells[149]. However, 105 of the variants with large positive client scores were inactive (DMSO activity score < -0.47), so the reduction in yeast growth mediated by radicicol toxicity accounts for these variants' positive client scores. Similarly, yeast harboring an empty vector that does not express Src also had a positive client score (**Figure 3.2D**). Only 6 variants appeared to have appreciable activity in DMSO and increased activity following Hsp90 inhibition (**Supplementary Figure 2B**). We individually measured the growth rate of three of these variants in DMSO and radicicol. None of them exhibited an appreciable decrease in growth rate on radicicol treatment, as the client score suggested. One variant, Src_{E335M}, showed growth recovery with radicicol treatment, suggesting it was strongly dependent on Hsp90. However, this result was the consequence of experimental error, as the E335M client score had a standard error greater than 99.4% of the other variants. Thus, these few variants appear to be artifacts of radicicol mediated toxicity.

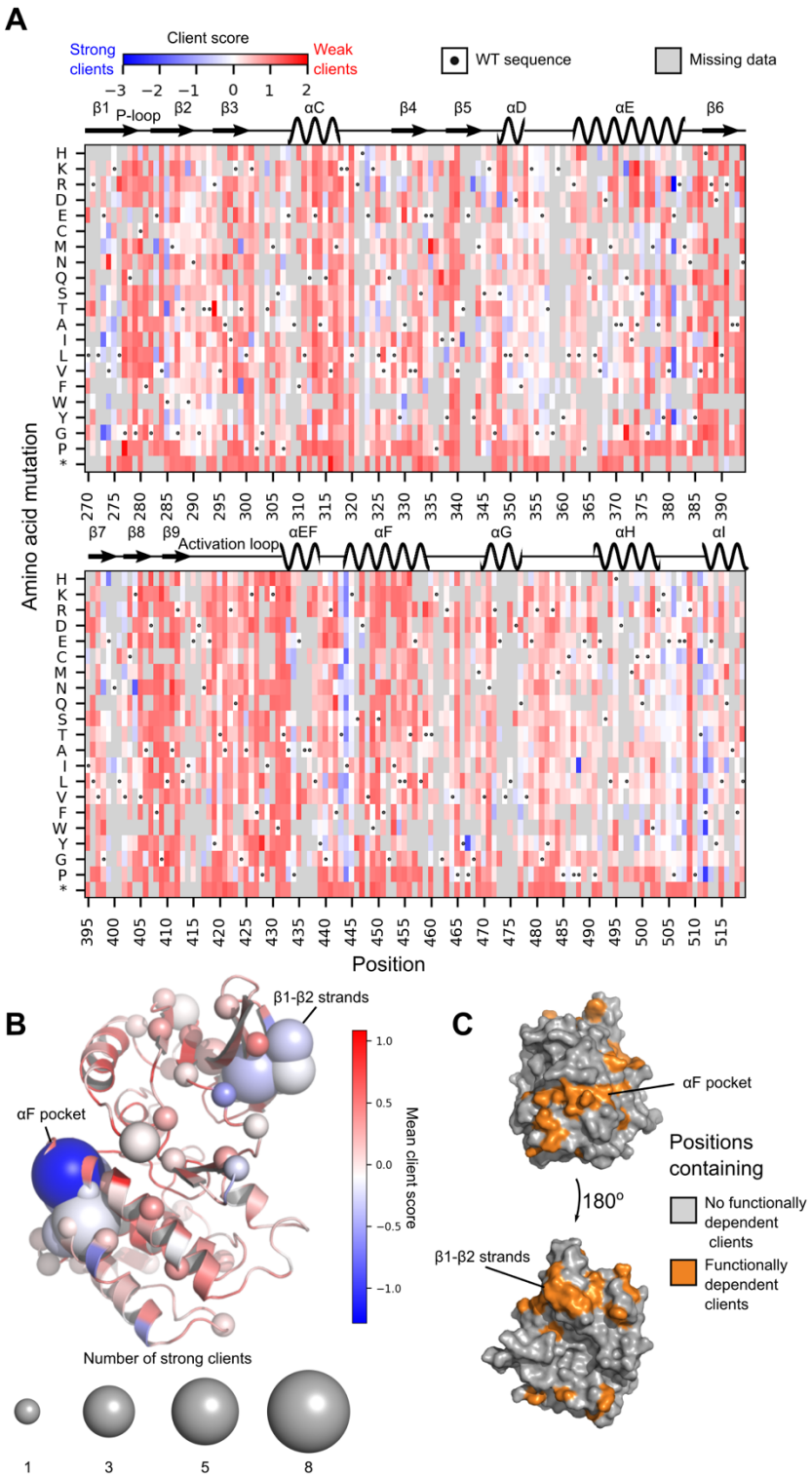


Figure 3.3. (A) Heatmap of client scores of 3,499 Src variants. Cells are colored by client score. Gray indicates missing variants and dots indicate the WT amino acid. Asterisks indicate stop codons. (B) Structure of the Src CD colored by mean client score at each position and showing the number of functionally dependent clients at each position by sphere size (PDB ID: 2SRC). (C) Structure of Src with a continuous surface of client mutations between αF pocket and β1-β2 strand hotspot (PDB = 2SRC).

3.4 Functionally dependent client variants are spatially localized

In order to understand how variants impact Hsp90 dependence, we organized the client scores into a variant effect map comprising 73% of the total possible single amino acid variants in Src's CD (**Figure 3.3A**). The map revealed that 80% of the positions in the CD had no functionally dependent client variants. Of the 49 positions with a functionally dependent client variant, 35 positions had just one functionally dependent client variant while four positions, W285, E381, I444, and F512, had five or more functionally dependent client variants. Projection of the client score data onto the structure of the CD revealed dramatic localization of the positions harboring functionally dependent client variants. Two hotspots consisting of 14 positions in the αF pocket of the C-lobe and β1-β2 strand hotspot of the N-lobe account for 52.9% of functionally dependent client variants (**Figure 3.3B**). The functionally-dependent client-containing positions outside these two hotspots appear to form a continuous surface that connects the β1-β2 strand hotspot to the αF pocket, comprising the hinge between the N- and C- lobes (342-347), the β7-β8 loop, and the αC-β4 loop (**Figure 3.3C**).

One hotspot, the αF pocket, contains eight positions that form a regulatory interface where the αE, αF, and αI helices meet (**Figure 3.4A**). Three of the four positions with five or more functionally dependent client variants, E381, I444, and F512, are located within the αF pocket. A lysine substitution at E381 has previously been shown to increase the Hsp90 dependence of Src[125]. The appearance of numerous functionally dependent client variants at αF pocket positions I444 and F512, which are adjacent to E381, suggest that the αF pocket plays an important role in Hsp90 dependence. Indeed, each of the other five positions comprising the αF pocket also harbored at least one functionally dependent client variant. T443 and T511 each had three functionally dependent client variants, while A378, P506, and E508 harbored one or two functionally dependent client variants. We compared the αF pocket positions to adjacent positions (within 7 angstroms), and found no appreciable difference in mutational coverage (**Supplementary Table 3.1**). Positions adjacent to the αF pocket contained a significantly smaller proportion of functionally dependent client variants (5/413) compared to the αF pocket positions (31/133);

two-sided Fisher's exact test p < 1e-4). Thus, variants within the αF pocket appear to drive Hsp90 dependence specifically.

The other hotspot, in the β1 and β2 strands, contains six positions located on the solvent-exposed face of the N-lobe of Src's CD (**Figure 3.4B**). We previously showed that these six positions harbor many activating variants and that they play a role in regulating Src activity[104,150] (**Supplementary Table 3.2**). The β1-β2 hotspot contains one position with five functionally dependent client variants, two positions with three functionally dependent client variants, and three positions with 1-2 functionally dependent client variants. The appearance of numerous functionally dependent client variants at nearly all positions in the β1-β2 strand region suggest that this hotspot could also be involved in Hsp90 dependence. Indeed, as for the αF pocket, a comparison of the proportion of functionally dependent client variants within the β1-β2 strand positions (14/87) to adjacent positions (within 7 angstroms; 4/207) revealed that the β1-β2 strand positions are significantly enriched for functionally dependent client variants (two-sided Fisher's exact test p < 1e-4) (**Supplementary Table 3.2**). Thus, like the αF pocket, variants within the β1-β2 strand hotspot appear to specifically increase Hsp90 dependence.

Given the localization of functionally dependent client variants in two distinct hotspots, we asked if the hotspots had distinct patterns of amino acid substitutions. Positively charged and hydrophobic substitutions at position E381 have previously been shown to enhance Hsp90 interaction[144]. Thus, we asked whether these types of substitutions drove Hsp90 dependence at other αF pocket and β1-β2 strand positions. Indeed, positively charged substitutions at positions E381, I444, T511, and P506 conferred strong Hsp90 dependence (randomization test, p = 7e-4; **Figure 3.4C**). Positively charged substitutions at the remaining four αF pocket positions had WT-like client scores, except at F512 where both K and R substitutions led to loss of function in DMSO. Similarly, hydrophobic substitutions at positions E381 and F512 conferred strong Hsp90 dependence (randomization test p < 1e-4) but at other αF pocket positions led to WT-like client scores. Thus, the trends previously observed for E381 extend to only three other positions for positive charges and one other position for hydrophobic substitutions. This suggests that surface biochemical properties are not the only driver of Hsp90 dependence at the αF pocket.

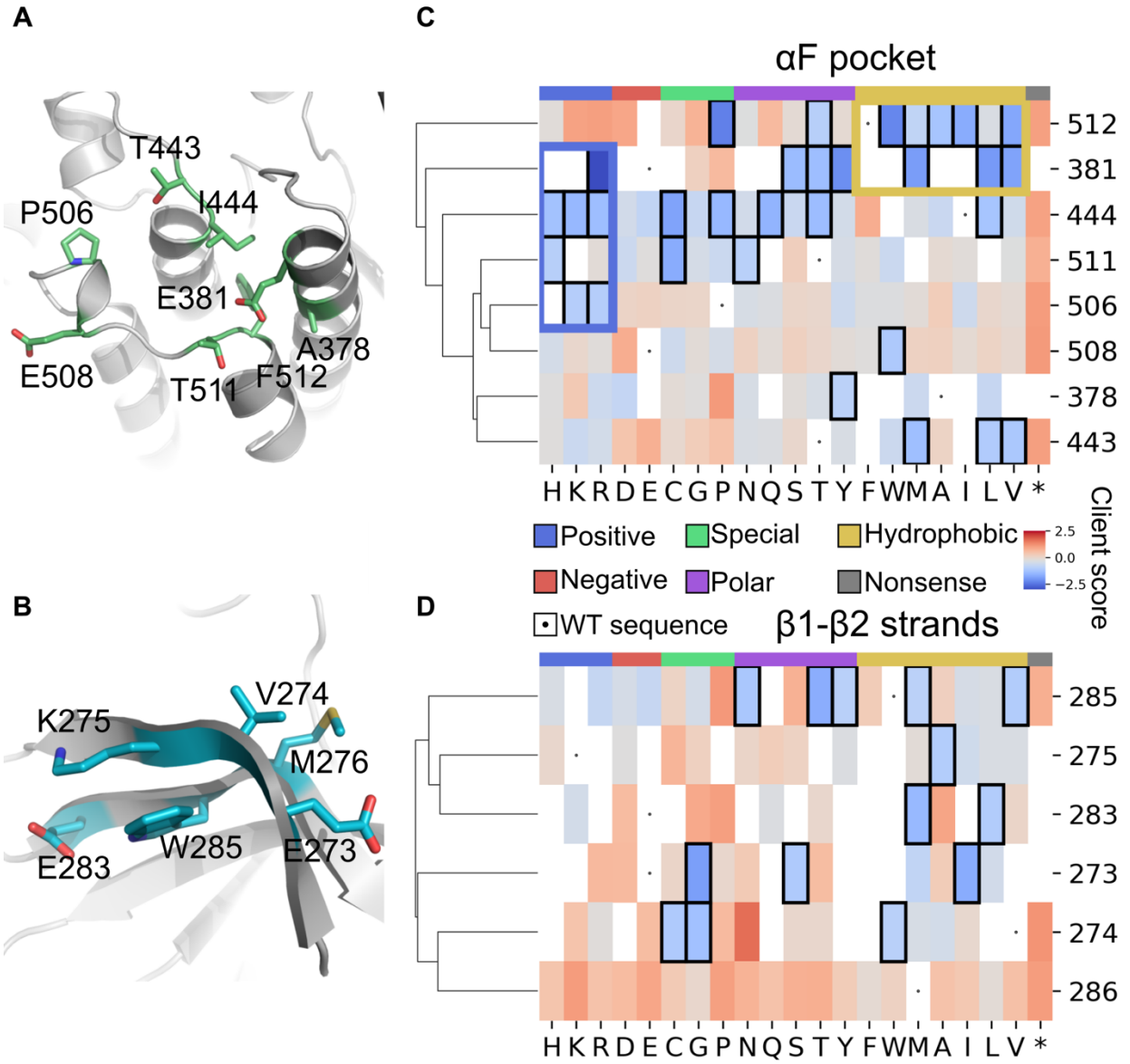


Figure 3.4. (A) Structural details of the αF pocket positions containing functionally dependent client variants (shown in green). (B) Structural details of the β1-β2 strand hotspot positions containing functionally dependent client variants (shown in blue). (C) Hierarchical clustering of αF pocket positions shows enrichment of hydrophobic functionally dependent client variants at positions E381 and F512 (outlined in yellow) and enrichment of positively-charged functionally dependent client variants at positions E381, I444, T511, and F512 (outlined in blue). Dots indicate the WT amino acid. Asterisks indicate stop codons. (D) Hierarchical clustering of β1-β2 strand positions shows general enrichment of hydrophobic functionally dependent client variants. Dots indicate the WT amino acid. Asterisks indicate stop codons.

Like the αF pocket, hydrophobic substitutions in the β1-β2 strand hotspot often produced Hsp90 dependence, accounting for seven of the 14 functionally dependent client variants (randomization test p =

1.1e-4; **Figure 3.4D**). However, unlike positions E381 and F512 in the αF pocket where all observed hydrophobic substitutions created functionally dependent clients, no position in the β1-β2 strand hotspot had such strong enrichment for hydrophobic substitutions. The remaining functionally dependent client variants in the β1-β2 strand hotspot were polar or special amino acids (C, G, P), and no negatively charged variants were functionally dependent clients.

Thus, at the αF pocket and β1-β2 strand hotspots, hydrophobic variants tended to give rise to Hsp90 dependence, but no other clear patterns in the nature of functionally dependent client substitutions were apparent. These two regions both play a role in facilitating autoinhibition and regulating Src activity. The αF pocket binds the SH4 domain to promote a closed and inactive conformation of Src[104]. We previously showed that Src becomes more extended and hyperactive upon abrogation of this regulatory interaction by mutagenesis. The β1-β2 strands flank the phosphate-binding loop (P-loop) of the CD and residues in the β1-β2 strand hotspot are components of an allosteric network that communicates intramolecular SH3 domain engagement with the conformational flexibility of Src's P-loop and αC helix. Mutations in the β1-β2 strand hotspot also extend and activate Src[122,126,150]. A previous biochemical study demonstrated that the introduction of mutations from v-Src that disrupt regulatory domain interfaces and promote an extended global conformation into c-Src also increase Hsp90 dependence[122]. Taken together, this evidence suggests that more extended, regulatory domain-disengaged Src, created through variation within the αF pocket and the β1-β2 strands, is more functionally dependent on Hsp90.

In addition to the αF pocket and β1-β2 strand hotspots, Src contains other clusters of regulatory positions that influence its conformation, namely the SH2-CD and SH3-CD interfaces[133,151,152]. Unlike the αF pocket and β1-β2 strand hotspot, the SH2 and SH3 interfaces were not functionally dependent client variant hotspots. The SH2 and SH3 interfaces each only had one functionally dependent client variant (1/49 and 1/78, respectively), compared to the 31 functionally dependent clients in the αF pocket (31/133) and the 14 functionally dependent clients (14/87) in the β1-β2 strand hotspot (**Figure 3.5A; Supplementary Table 3.3**). The SH2 interface functionally dependent client variant, Src_{L325T}, is adjacent to the hinge and αE helix. The SH3 interface functionally dependent client variant, Src_{W289K}, is adjacent to

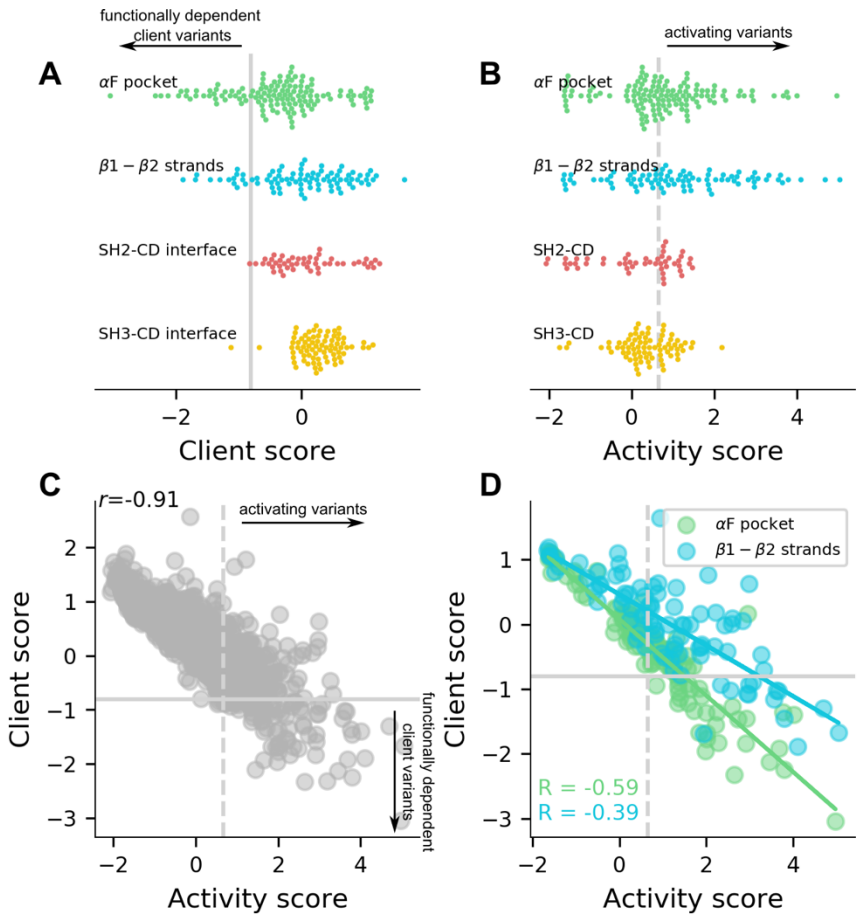


Figure 3.5. (A) Distribution of client scores at different regulatory clusters. Solid line indicates the functionally dependent client threshold. (B) Distribution of DMSO activity scores at different client interfaces. Dashed line indicates increased activity threshold. (C) Correlation between client score and activity score; R = -0.91, rho = -0.93 (D) Linear regression of client scores and activity scores of variants in the αF pocket (green) and β1-β2 strand hotspot (blue) yield different slopes.

the SH2-CD linker. We, as well as others, previously showed that mutations at the SH2 and SH3 interfaces disrupt the intramolecular regulatory domain engagement of Src using both SH3 pull-down assays and molecular dynamics simulations[104,132,153]. For example, Src_{T293D} in the SH3-CD interface, leads to reduced intramolecular SH3 domain engagement in an SH3 pull-down assay[104]. However, Src_{T293D} had a client score (0.26) similar to Src_{WT} (0.27). Thus, unlike the αF pocket and β1-β2 strand hotspot, the SH2-CD and SH3-CD interface variants were rarely strongly dependent on Hsp90, despite the ability of some to disrupt intramolecular regulatory domain engagement. Factors beyond the global conformation must therefore play a role in Hsp90 client dependence. These factors could include kinase domain stability, surface hydrophobicity, and phosphotransferase activity, all of which have been implicated in Hsp90 client specificity[30,122,126,154].

3.5 Src variant predicted stability and hydrophobicity do not dictate Hsp90 dependence

In vitro and molecular dynamics studies of weak client Src_{WT} and strong client v-Src suggest that the stability of the kinase fold is important for Hsp90 client dependence[122,126]. Comprehensive characterization of the human kinome similarly found a modest correlation (R^2 = 0.23, Pearson's R = 0.48) between a measure of Hsp90 interaction and thermal stability of 56 kinases[30]. These studies suggest Hsp90 recognizes intrinsically unstable kinases as clients, reasoning that extended kinase conformations are less stable than more compact, closed conformations, thereby leading to increased dependence on Hsp90. However, these studies are limited in the number of kinases studies and our data provides a unique opportunity to identify stability trends in functionally dependent Hsp90 clients.

We calculated the free energy change brought about by variants (ΔΔG) in the extended conformation of Src (PDB ID: 1Y57) using a previously described Rosetta-based method for predicting a protein's thermodynamic stability from its structure[66,155]. We validated the viability of these stability calculations using the gatekeeper mutation, Src_{T341I}. Structural analyses of the T341I gatekeeper mutation suggest it stabilizes a hydrophobic spine that links the gatekeeper mutation to the activation loop[146,156,157]. We calculated a ΔΔG of -0.4 Rosetta energy units (R.E.U.) for Src_{T341I}, indicating that the introduction of the gatekeeper mutation stabilizes the structure relative to Src_{WT}. We projected the average ΔΔG values of client positions onto the CD (**Supplementary Figure 3A**), revealing other stabilizing client variants at the β3-αC loop as well on the surface of the αG, αH, and αI helices. However, Src variant ΔΔG had only a limited correlation with client score (Pearson's R = 0.22; **Supplementary Figure 3B**). Because Hsp90 may only recognize specific local instabilities in the kinase domain, we also analyzed the relationship between ΔΔG and client score at each position with a functionally dependent client variant. Although we observed a correlation between ΔΔG and client score at a few positions, most positions had no such correlation, with many nonclient variants having similar ΔΔG to client variants (**Supplementary Figure 3C**), suggesting that the major determinant of Hsp90 interaction resides in features not captured by the Rosetta energy function and our calculated ΔΔGs.

Another factor suggested to be a determinant of Hsp90 client dependence is surface hydrophobicity[158–160]. The most notable example is the substitution D770G in ErbB-1, which introduces a hydrophobic patch at the αC-β4 loop, leading to increased Hsp90 interaction[125,144]. Fluorescence assays for binding exposed hydrophobic regions showed that v-Src and other client Src variants had more exposed hydrophobic regions than c-Src, suggesting surface hydrophobicity as a factor involved in Hsp90 dependence[122]. We also observed a tendency for hydrophobic variants to result in functionally dependent clients at the αF pocket and β1-β2 strands. We used Rosetta to calculate the hydrophobic solvent accessible surface area (hSASA) of the side chains of the variants in the extended conformation of Src (PDB ID: 1Y57). We found limited correlation between the ΔhSASA and client score (Pearson's R = -0.14) (**Supplementary Figure 4A**). To tease out the effects of hydrophobicity at individual positions, we projected the average ΔhSASA onto the CD structure (**Supplementary Figure 4B**), revealing E381 and F512 in the αF pocket as positions with the lowest ΔhSASA. The only region with localization of positive ΔhSASA values is on the αG and αH helices. However, many nonclients at these positions had similar ΔhSASA values (**Supplementary Figure 4C**), suggesting that changes in surface exposed hydrophobicity are not a major determinant of Hsp90 dependence.

3.6 Src hyperactivity correlates with Hsp90 dependence

Variants in the SH2-CD and SH3-CD interfaces were much less activating than those in the αF pocket and β1-β2 strand hotspot (**Figure 3.5B**). The mean DMSO activity score of variants in the αF pocket and β1-β2 strand hotspots were 0.79 and 1.72 respectively. In comparison, the mean activity scores of the SH2 and SH3 interfaces were 0.11 and 0.35 respectively. Thus, the lack of functionally dependent Hsp90 client variants at the SH2-CD and SH3-CD interfaces could reflect an apparent dependence on active global conformations of Src.

To test this dependence more generally, we compared variant client scores to their corresponding DMSO activity scores and found a strong correlation (R = -0.91, rho =-0.93; **Figure 3.5C**). In fact, all 84 client variants we identified were also classified as increased activity variants based on their DMSO activity score[104]. Thus, Hsp90 appears to recognize, in an indirect fashion, increased phosphotransferase

activity of Src variants. However, high activity does not always confer strong Hsp90 dependence, as some highly active variants had WT-like client scores. Regressing client scores and activity scores for variants in either the αF pocket or β1-β2 strand hotspot revealed substantially different linear slopes and correlation coefficients (**Figure 3.5D**). While activity and client scores were strongly correlated for αF pocket variants, there was much more variance in client scores for β1-β2 strand variants of similar activity. The discrepancy between the two hotspots was most apparent between activity scores of 1.4 and 2.5, where moderately activating variants are found. Indeed, most moderately activating αF pocket variants (14/18) were functionally dependent clients while only a few moderately activating β1-β2 strand variants (2/17) were functionally dependent clients. Thus, while high phosphotransferase activity was necessary for strong Hsp90 dependence in our assay, not all activating variants were functionally dependent clients and the distribution of functionally dependent clients among moderately activating variants was highly dependent on the location within the CD.

Disengagement of Src's regulatory domain apparatus is generally necessary for higher Src activity, so Src's Hsp90 dependence appears to require disengagement of the regulatory domain apparatus to provide access to the CD. However, comparison of similarly activating variants in the αF pocket and β1-β2 strand hotspots highlight that regulatory domain disengagement alone is insufficient for Src to become a functionally dependent Hsp90 client. We observe a strong correlation between the activity and client scores of αF pocket variants (Pearson's R = -0.59), suggesting that activating variants in this region loosen the regulatory apparatus and promote the partially unfolded state required for recognition by the Hsp90/Cdc37 chaperone system. The weaker correlation between the activity and client scores for variants in the β1-β2 strand hotspot (Pearson's R = -0.39) suggests that while activating variants in this region of Src can promote regulatory domain disengagement, only a subset of variants promote partial unfolding of the CD.

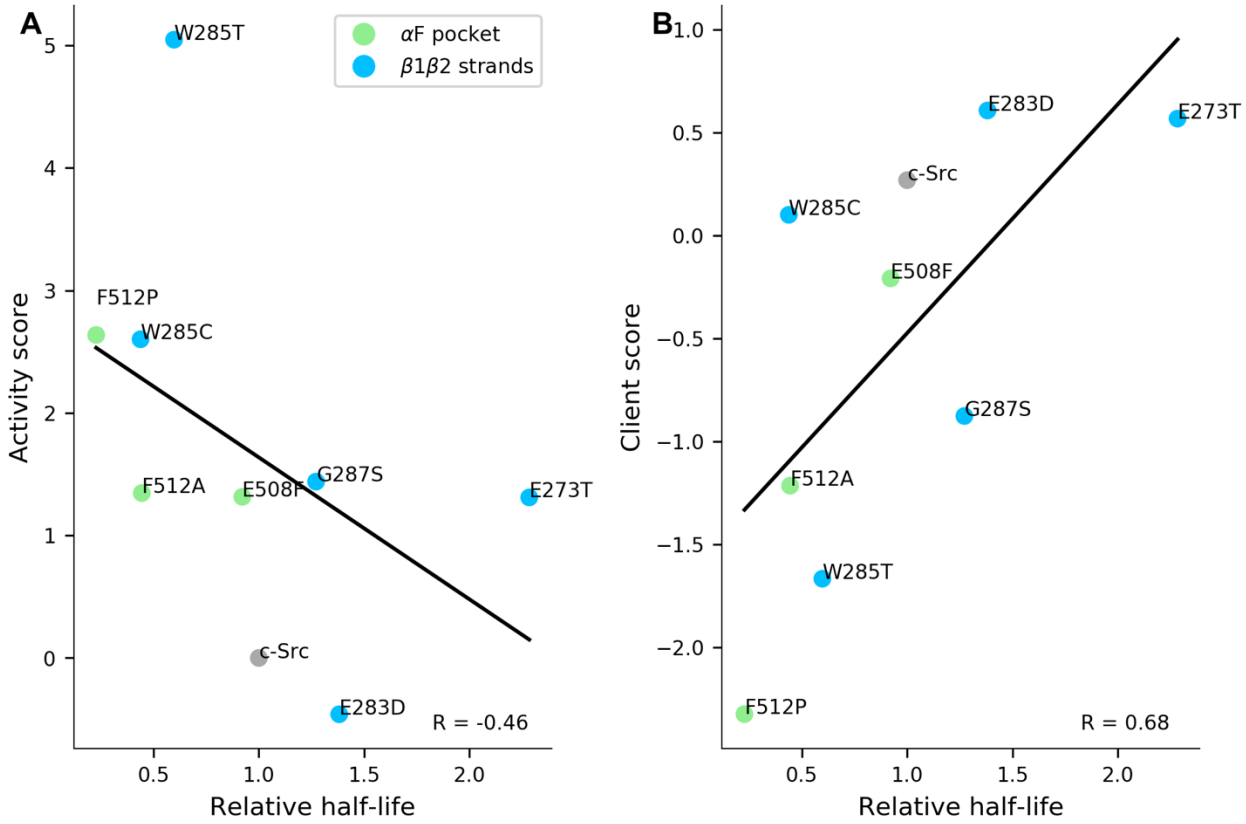


Figure 3.6. Relative half-lives are variant half-lives normalized to the measured c-Src (Src_{WT}) half-life. (A) Activity scores are negatively correlated with thermolysin half lives, while (B) client scores are positively correlated with thermolysin half lives.

3.7 Specific active conformations drive Hsp90 dependence

To gain insight into the relationship between Src conformation and Hsp90 dependence, we used a limited proteolysis assay with thermolysin to measure the kinase domain stability, conformational flexibility and global extension of Src variants with similar activity scores but differing Hsp90 dependence[150161,162]. We expressed individual Src variants in HEK293T cells and captured each through their ATP-binding sites using dasatinib-conjugated beads. We then exposed the Src-bound beads to thermolysin, and measured the quantity of full-length Src over time via densitometry.

To determine if we could differentiate between Src variants with extended or compact global conformations, we compared the thermolysin half-lives of Src_{W285T}, an extended, flexible conformation control, and Src_{E283D}, a compact conformation control, to Src_{WT} (**Supplementary Figure 5A-C**). We normalized the half-lives such that Src_{WT} had a half-life of 1. The extended global conformation control

Src_{W285T} yielded a half-life of 0.56 relative to Src_{WT}, meaning that it was proteolyzed more quickly and therefore in a more extended, flexible conformation. The compact control Src_{E283D} yielded a relative half-life of 1.37, meaning that it was proteolyzed less quickly and therefore in a more compact conformation. Thus, our assay could resolve differences in the conformational flexibility and global extension of Src variants bound to dasatinib beads (**Supplementary Figure 6**).

In general, globally extended Src is considered to be active while compact Src is inactive, but we found only a modest correlation between activity and thermolysin half-life (Pearson's R = -0.46; **Figure 3.6A**). We found a slightly stronger correlation between thermolysin half-life and client score (Pearson's R = 0.68; **Figure 3.6B**) than activity score. Taken together, these data suggest that activating, weak client variants have less flexible and more compact global conformations than similarly active functionally dependent client variants.

3.8 Discussion

We used deep mutational scanning to investigate the effects of Hsp90 chaperoning on Src activity amongst 3,929 variants in the CD. Overall, few variants' activity depended strongly on Hsp90. We revealed 84 previously unknown functionally dependent client variants of Src, which spatially clustered in the αF pocket and the β1-β2 strands. 52% of the new functionally dependent client variants we identified were located in these hotspots, with the remaining functionally dependent client variants forming a surface through the CD hinge, β7-β8 loop, and the αC-β4 loop. These findings corroborate previous studies identifying position E381 as a location where variants with strong Hsp90 dependence can appear[125]. However, the appearance of functionally dependent client variants in the β1-β2 strands was not previously recognized[122].

Within each hotspot, we observed trends in the types of variants that transform Src into a functionally dependent client. For example, select positions in both the αF pocket and the β1-β2 strands transform Src into a functionally dependent client when mutated to hydrophobic residues. A smaller subset of αF pocket positions transform Src into a functionally dependent client when mutated to positively charged residues. Because the trends are limited to subsets of positions within each hotspot,

they imply that Hsp90 does not directly interact with these regions, but rather these variants create conformational flexibility or expose structural motifs recognized by Hsp90.

Variant-induced instability and hydrophobicity have long been proposed as drivers of Hsp90 dependence[30,122,125,126]. We used Rosetta to calculate the changes in stability and hydrophobicity caused by Src variants, and found little correlation between Hsp90 dependence and predicted stability or hydrophobicity. While nearly all functionally dependent clients we identified were predicted to be unstable, many variants with similarly low predicted stability were classified as weak clients or nonclients. Moreover, some functionally dependent clients increased local hydrophobicity while others decreased hydrophobicity. Our assay only detects functionally dependent Src variants, we would miss loss-of-function variants with increased hydrophobicity or instability that are dependent on Hsp90. Additionally, Rosetta calculates changes in hydrophobicity and stability using static structures of Src, and may not accurately predict structural dynamics that increase hydrophobicity or induce local destabilization. Thus, the lack of correlation between client score and predicted hydrophobicity and stability may inaccurately reflect the role of these factors in determining Hsp90 binding.

Previously, Hsp90 dependence has been suggested to correlate with unfolding cooperativity, hydrophobicity, and structural elements in the kinase domain[122,126]. Functionally dependent clients with variants disrupting C-terminal tail binding and SH3-CD interface increase Src unfolding and the dynamics or solvent exposure of the αC-β4 loop and β1-β3 strands[122]. Here, we extend this picture to show that variants in the αF pocket and the β1-β2 strands also increase conformational flexibility and disengagement of the regulatory domain apparatus, becoming dependent on Hsp90.

High Src activity requires disengagement of the regulatory apparatus and adoption of an extended conformation. We found a strong correlation between high Src phosphotransferase activity and Hsp90 dependence, suggesting that adoption of an extended global conformation is required to drive Hsp90 dependence. However, variants at the SH2-CD and SH3-CD interfaces that promoted phosphotransferase activity and extend Src global conformation were not classified as functionally dependent clients by our assay[104]. Moreover, similarly active variants in the β1-β2 strands and αF pocket differed in their tendency to confer strong Hsp90 dependence. Thus, regulatory domain disengagement alone cannot explain Src Hsp90 dependence.

Instead, we found that regulatory domain disengagement was necessary, but not sufficient, to confer strong Hsp90 dependence. Variants that increased conformational flexibility in our thermolysin assay generally had higher activity and yielded stronger Hsp90 dependence. The differences in how similarly activating variants extend Src's global conformation suggest that active Src variants occupy different conformations. In particular, we hypothesize that, in addition to regulatory domain disengagement, partial unfolding of the CD is necessary to confer strong Hsp90 dependence. A cryo-electron microscopy-derived structure of the kinase Cdk4 in complex with Hsp90 and the kinase-specific co-chaperone Cdc37 showed the kinase domain in an extended conformation with split N- and C-lobes threaded through the Hsp90 lumen[17]. Notably, Cdc37 forms hydrogen-bonding interactions with the backbone of Cdk4's αE helix in the C-terminal lobe. Given the proximity of the αF pocket to Cdc37's site of interaction with the αE helix, we speculate that activating mutations in the αF pocket also promote a more accessible αE helix in Src. Additionally, the Cdk4-Cdc37-Hsp90 structure shows that Cdc37 stabilizes the separation of the N- and C- lobes of Cdk4 by mimicking interactions the N-lobe makes with the C-lobe. Both the αF pocket and the β1-β2 strands are adjacent to the unfolded region of Cdk4 in the structure. Thus, we hypothesize that the conformational flexibility induced by variants at the αF pocket and the β1-β2 strands potentiate binding and stabilization by Cdc37, and ultimate loading onto the Hsp90 dimer (**Supplementary Figure 7-8**). Further study of the conformation and dynamics of the variants we identified with similar activity but different Hsp90 dependence in different CD regions could provide insight to the initial steps of kinase loading into the chaperone complex. Such studies could also facilitate development of therapeutics that stabilize Hsp90 dependent kinase conformations, and these therapeutics could be used in combination with Hsp90 inhibitors.

While we were able to add to the picture of Hsp90 dependence of Src kinase, our study is limited in several ways. Our assay only identifies functionally dependent client variants that lose activity when Hsp90 is inhibited. However, other strong client variants that interact with Hsp90 in their mature forms, but do not lose activity following Hsp90 inhibition, could exist and would be missed by our assay. Activating, weak client variants at the SH3-CD and SH2-CD interfaces could be such variants. Additionally, radicicol treatment could induce the heat shock response and could give rise to off-target effects that affect measured growth rate. Moreover, our study pairs human Src with yeast Hsp90. While yeast and human

Hsp90 are conserved in sequence, structure and function, differences may exist[163]. Nonetheless, we recapitulate all of the known Src functionally dependent clients studied in the context of human Hsp90 of which we are aware[125,147,148]. Another limitation is that we were unable to determine the relative importance of activation and extension in determining Hsp90 functional dependence. Future work measuring the dynamics and SH2- and SH3- domain availability of full-length and truncated forms of the strong client Src variants we identified will be necessary to deconvolve the effects of each contributing factor on client status. Finally, our analysis of the relationship between client score and stability relies on predicted, rather than measured, stability.

Taken together, our data suggests that activation and extension, primarily induced by positively charged and hydrophobic variants at the αF-pocket or hydrophobic variants at the β1-β2 strands, are additive determinants of Hsp90 dependence. We found that increased Src phosphotransferase activity was strongly correlated with increased Hsp90 dependence, but that high activity alone was insufficient to drive Hsp90 dependence. We showed, for a limited number of variants at the αF-pocket and β1-β2 strands, that conformational extension also correlated with increased Hsp90 dependence. We speculate that variants the αF pocket and the β1-β2 strand hotspot perturb different regulatory elements in the CD, driving spontaneous access to different conformations. In turn, Cdc37 may preferentially recruit some Src conformations over others to the Hsp90 complex. Due to the high conservation of the CD between diverse kinases, it is possible that similar factors driving Src-Hsp90 client preferences apply to other kinases. Thus, extension of our approach to other Hsp90 clients, including other kinases and neurodegenerative disease-related proteins, could reveal a broader array of client dependence and processing mechanisms. Likewise, Hsp90 client variant libraries could be assessed in a human cell context, which would enable modulation of the full slate of Hsp90 co-chaperones by chemical or genetic means. Thus, deep mutational scanning represents a useful tool for investigating the mechanisms by which Hsp90 recognizes and processes clients.

3.9 Acknowledgements

We would like to thank Rachel Klevit for guidance. We thank Jessica Simon and Zachary Potter for thermolysin assay assistance and helpful discussions regarding Src. This work was supported by National

Institutes of Health National Institute for General Medical Sciences (R01GM109110 (D.M.F) and R01GM086858 (D.J.M.)). The research was supported by the PRISM (Protein Interactions and Stability in Medicine and Genomics) center funded by the Novo Nordisk Foundation (NNF18OC0033950).

3.10 Methods

S. cerevisiae genetics

Yeast experiments were performed in BY4741 Green Monster (MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 pdr5Δ pdr10Δ pdr11Δ pdr12Δ pdr15Δ snq2Δ ynr070wΔ aus1Δ yol075cΔ adp1Δ ycf1Δ vmr1Δ nft1Δ bpt1Δ ybt1Δ yor1Δ) (Suzuki et al., 2011; a gift from Dr. Fritz Roth). All Src constructs were cloned into the p415 GAL1 plasmid and transformed via lithium acetate transformation. Transformants were plated onto C-Leu selective media and incubated at 30°C for 72 hours. Single colonies were used to inoculate C-Leu liquid medium and incubated at 30°C in a rotating mixer at 20 rpm for 72 hours.

Cloning

All variants were generated using IVA cloning following standard protocols[164]. Sequences of the full open reading frame were verified by Azenta Life Sciences Sanger sequencing with custom primers.

Individual yeast growth assays

Codon-optimized, full length single variants of Src were transformed into the BY4741 Green Monster strain (a gift from Dr. Fritz Roth) using standard lithium acetate transformation protocol and plated on C-Leu plates[165]. Three individual colonies of each transformation were used to inoculate 5 mL 2% glucose C-Leu cultures which were grown at 30°C to saturation. Cultures were back diluted to an OD_{600} of 0.5 and allowed to double once in 3% raffinose C-Leu media. Cultures were back diluted to an OD_{600} of 0.01 in 2% galactose C-Leu media to induce expression and plated and grown in a BioTek Synergy H1 plate reader using constant orbital shaking at 30°C. The OD_{600} was measured every 30 minutes during a 48 hour growth period. Growth rates of each culture were calculated using the GrowthRates software[166].

Src CD DMS

To identify Hsp90 client variants, we cultured BY4741 Green Monster cells expressing our previously described c-Src variant library in 2% galactose C-Leu media supplemented by 800 nM radicicol in duplicate (GI50 = 5.32 μM)[104]. Three time points were sampled from each culture during 48 hours of growth at approximately 19, 25, and 32 hours. Selection 1 time points had OD_{600}'s of 0.186 and 0.202. Selection 2 time points had OD_{600}'s of 0.682 and 0.63. Selection 3 time points had OD_{600}'s of 2.95 and 2.4. Timepoints were spun down at 3,000 x g and plasmids were extracted using Yeast Plasmid Prep I (Zymogen) according to the manufacturer's protocol. Barcodes were amplified and Illumina cluster generators and indices were added via PCR using 2X KAPA Robust HotStart ReadyMix (KAPA Biosystems). PCR products were cleaned and sequenced on the NextSeq 500/550 High Output v2 kit.

Library sequencing and variant scoring

Illumina reads for each timepoint and replicate were processed using bcl2fastq. All timepoints and replicates were imported into Enrich2 and converted to activity scores using the Weighted Least Squares scoring method (Rubin et al., 2017). Missing variants in this work were also missing in Ahler et al., 2019, where we generated the library used in this work. These missing variants could result from nonuniform PCR amplification, inadequate barcode coverage, and extreme growth attenuation due to variant activity and subsequent filtering of low-frequency barcodes. Because increased Src activity decreases yeast growth and subsequently the number of barcodes associated with each variant, activity scores were multiplied by -1 before further analysis. The radicicol activity scores were transformed using two linear regression models from measurements of 13 variants spanning the range of DMSO and radicicol activity scores. DMSO activity scores from previous work were used for normalization[104].

The first model predicts the growth rate $\hat{r}_{radicicol,\,i}$ of given variant i from the variant's radicicol score $s_{radicicol,\,i}$. The second model predicts the DMSO activity score $\hat{s}_{DMSO,\,i}$ of given variant i from the variant's growth rate $r_{DMSO,\,i}$. m and b are the fitted parameters for each linear regression.

$$\hat{r}_{radicicol,\,i} = m_1 s_{radicicol,\,i} + b_1$$

$\hat{s}_{DMSO}$, $i = m_2 r_{DMSO,\,i} + b_2$

The result of this transformation is a calibrated radicicol activity score $\hat{s}_{radicicol,\,calibrated,\,i}$ for each variant i.

$\hat{s}_{radicicol,\,calibrated,\,i} = m_2 \hat{r}_{radicicol,\,i} + b_2$

The difference between the calibrated radicicol activity score and the DMSO activity score for a given variant i is the client score $s_{client,\,i}$.

$s_{client,\,i} = \hat{s}_{radicicol,\,calibrated,\,i} - s_{DMSO,\,i}$

The client score of empty vector $s_{client,\,empty\ vector}$ was calculated by first measuring the growth rate of empty vector expressing cells in DMSO ($r_{DMSO,\,empty\ vector}$) and radicicol ($r_{radicicol,\,empty\ vector}$). Next, the respective activity scores $\hat{s}_1$ and $\hat{s}_2$ of each was calculated from the measured growth rates.

$\hat{s}_1 = m_2 r_{DMSO,empty\ vector} + b_2$

$\hat{s}_2 = m_2 r_{radicicol,empty\ vector} + b_2$

Finally, the client score was calculated was by taking the difference between the predicted activity score of empty vector in DMSO and radicicol.

$\hat{s}_{client,\,empty\ vector} = \hat{s}_2 - \hat{s}_1$

.

Rosetta stability and hydrophobicity calculations

The local stability values were obtained using the Cartesian ddG method first described by Park 2016[155]. The hSASA values were obtained by building models of each variant present in our study and calculating the hydrophobic SASA of the mutated residue.

Dasatinib bead conjugation

Dasatinib-amine was conjugated to NHS-activated sepharose beads first by washing 1,250 µL of bead slurry with 1:1 mixture of dimethylformamide and ethanol. The beads were dried and added to 1300 µL of 1:1 dimethylformamide and ethanol and 64 µL 50mM dasatinib-amine. 240µL of 1 M carbodiimide hydrochloride and mixed overnight at room temperature to couple the dasatinib to sepharose. The following day, the supernatant was removed and the beads were washed three times with 1:1 dimethylformamide and ethanol. The beads were then incubated in a mixture of 57% 1:1 dimethylformamide and ethanol, 30% ethanolamine, and 13% 1 M carbodiimide hydrochloride. The next day, the supernatant was removed and the beads were washed three times with 1:1 dimethylformamide and ethanol, twice with 0.5 M sodium chloride, once with 20% ethanol, and stored in 20% ethanol until use.

Mammalian cell culture

HEK293T cells were used for Src variant expression in thermolysin experiments. Cells were grown at 37°C and 5% CO_2 in DMEM (Thermo Fisher Scientific) supplemented with 10% fetal bovine serum and 1% penicillin-streptomycin. Cells were passaged regularly when they reached 80-90% confluency. The Src variants were transiently transfected into 500,000 HEK293T cells using FuGENE6 (Promega) and 1200 ng plasmid DNA at a 3:1 ratio of transfection reagent:DNA in single wells in 6-well plates according to manufacturer's protocol. Transfected cells were lifted off plates with 0.25% Trypsin-EDTA (Thermo Fisher Scientific) and pelleted at 300 x g for 10 minutes for downstream sample analysis.

Thermolysin assay measuring Src global conformation

Variants were transfected in HEK293T cells in single wells in 6-well plates. 48 hours after transfection, the transfected cells were lifted from the plate and pelleted at 300 x g for 10 minutes. Cell pellets were lysed using 200 µL or chilled modRIPA buffer (50 mM Tris, 150 mM NaCl, 5% glycerol (v/v), 1% NP-40, 0.25% sodium deoxycholate (w/v), 10 mM NaF, 1 mM PMSF, 1X protease inhibitor cocktail (Sigma Aldrich), 1X phosphatase inhibitor cocktail 2 (Sigma Aldrich), and 1X phosphatase inhibitor cocktail 3 (Sigma Aldrich), pH 8.0) for 10 minutes on ice, with occasional mixing. Cell lysates were spun

at 20,000 x g for 10 minutes at 4°C and the supernatant was separated from the insoluble lysate. 10 µL of 50% dasatinib-conjugated bead / 20% EtOH slurry was prepared for each replicate by washing 3 times with 10 bead volumes of 1X modRIPA buffer. Soluble lysate was incubated with washed dasatinib-conjugated beads for 1 hour at room temperature, gently mixing end over end. Following incubation, the beads were washed with 10 bead volumes of thermolysin digestion buffer (50 mM Tris, 500 mM $CaCl_2$, pH 8.0) 3 times, and resuspended in a final volume of 100 µL digestion buffer. 500 nM thermolysin (Promega) was added to the resuspended and washed beads in a 1:100 dilution and incubated at 37°C for up to 75 minutes. Timepoints were taken by quenching the bead-thermolysin mixture to 4X Laemmli 0.5mM EDTA.

Western blot and densitometric analysis

Samples in 1X Laemmli (BioRad) were boiled at 95°C for 10 minutes and run in 4-20% polyacrylamide gels (BioRad) in Tris-Glycine-SDS (BioRad) buffer for 40 minutes at 200V. Gels were transferred using the TurboBlot transfer system (BioRad) on the “Mixed MW Midi” program onto nitrocellulose (BioRad). Blots washes include 3 5-minute incubations with TBST (25 mM Tris, 150 mM NaCl, 0.1% Tween-20) on an orbital shaker. Following transfer, blots were washed, blocked with TBST and 5% nonfat dry milk for 1 hour at room temperature, washed, and incubated with primary antibody overnight at 4°C (Cell Signaling Technologies, 36D10). Blots were then washed again and incubated with secondary antibodies (LiCOR) and imaged via ChemiDoc (BioRad). All western blot quantification was performed using ImageLab software.

Half-life estimation and normalization

Densitometric measurements of each sample were normalized to the first timepoint by taking the ratio of each timepoint to the first timepoint. The normalized values were then fit to an exponential decay function $y = me^{-tx} + b$ using the curve-fitting function in scipy 1.4.1 package. The half-lives were calculated using $t_{1/2} = \frac{\ln ln\left(\frac{0.5-b}{m}\right)}{-t}$. To calculate relative half-lives that are reported in the results section, the half-lives of Src variants were then normalized to SrcWT by taking the ratio of the variant half-life to the wild-type half-life.

Code and data availability

Sequencing reads were deposited in NCBI's Gene Expression Omnibus (GEO) and are accessible through accession number GSE218190. Data for the DMS and code to reproduce figures are located in our Github repository at https://github.com/FowlerLab/2022_SrcHsp90.

Chapter 4. Conclusion and future directions

Since the discovery of the heat shock response, Hsp90 has emerged as a critical proteostasis regulator and valuable drug target in biology and disease. Despite its importance, the current understanding of the Hsp90 system remains incomplete. Researchers have extensively investigated the complex system, including its co-chaperones and clients, to elucidate how this flexible system activates clients and contributes to tumorigenesis or prevents the aggregation of toxic nuclear aggregates, among other possibilities.

In my doctoral research, I focused on a specific aspect of the Hsp90 system, namely, its relationship with client kinases, and how it activates and refolds them through deep mutational scanning. I illustrate in my thesis that deep mutational scanning is a powerful tool that comprehensively characterizes the effects of variants on protein function. In my study, I evaluated the current state of deep mutational scanning and its applications in protein science. Using the complex and dynamic phenomenon of Hsp90 chaperoning of the model client Src, I demonstrate the efficacy of deep mutational scanning.

Through my experiments, I uncovered novel molecular determinants of Hsp90 chaperoning of kinases, highlighting the utility of deep mutational scanning in protein science. My research provides new insights into the Hsp90 system and contributes to the growing knowledge of protein folding and chaperoning mechanisms.

In Chapter 1, I introduce Hsp90 and its structure, function, and cycle. Hsp90 is a molecular chaperone that plays a crucial role in activating and refolding proteins known as clients. It has emerged as a key regulator of proteostasis due to its ability to chaperone a diverse array of signaling molecules such as kinases, transcription factors, and steroid receptors. Consequently, Hsp90 has been shown to be a key driver in tumorigenesis, neurodegenerative disorders, and zoonotic diseases.

To unravel the complexities of Hsp90's diverse functions and its involvement in various diseases, researchers have extensively studied its interaction with co-chaperones and clients. However, the wide range of proteins chaperoned by Hsp90, coupled with the narrow focus of many studies on a few well-known client-nonclient protein pairs, makes it challenging to obtain a complete understanding of Hsp90's functional interactions.

Recently, the development of high-throughput multiplexed studies has presented an unprecedented opportunity to obtain a comprehensive understanding of the sequences chaperoned by Hsp90. Rather than retrospectively determining that client proteins are unstable, a more extensive dataset can reveal sequence and structural motifs that allow Hsp90 to interact with its clients, and consequently, play a crucial role in cellular functions and disease progression.

In Chapter 2, I provide an overview of the deep mutational scanning methods that are commonly used for investigating protein stability, binding, and activity. While there are several types of mutational scans, they can largely be categorized as *in vitro* scans and cell-based assays. *In vitro* mutational scans allow for precise control and the determination of binding and enzymatic constants in a massively parallel manner. On the other hand, cellular-based assays are not suitable for determining biochemical constants but are useful for understanding how variation affects function in the context of the cell. This is particularly valuable for studying viral proteins, membrane-bound proteins, and nucleic acid binding proteins. Furthermore, these scans can be employed to reengineer specificity, similar to in vitro scans.

Deep mutational scans have been effective in protein engineering studies and identifying functionally relevant positions in residues and binding. However, epistasis plays a crucial role in protein function, and deep mutational scans frequently reveal many deleterious variants and only a few gain-of-function variants relative to a wild-type sequence. To overcome this limitation, various groups have attempted to develop models of epistasis to leverage deep mutational scanning data to comprehend higher-order mutations. These models are in their nascent stages of development and have yet to be extensively used in studying protein function and employed in protein engineering, however. One possible solution is to simply incorporate higher order variants in a mutational scanning library, as seen in a study where all GB1 double variants were assessed in a deep mutational scan[167]. However, this approach becomes impractical for even higher order variants and longer sequences due to the vastness of the sequence space. Therefore, it is likely that computational models of variant effect will be utilized to predict higher order variants from deep mutational scanning data to create a more comprehensive variant effect map.

Nevertheless, deep mutational scans have identified specificity-determining positions, extended activity networks, and distal residues that increase binding affinity. However, a recurring constraint of

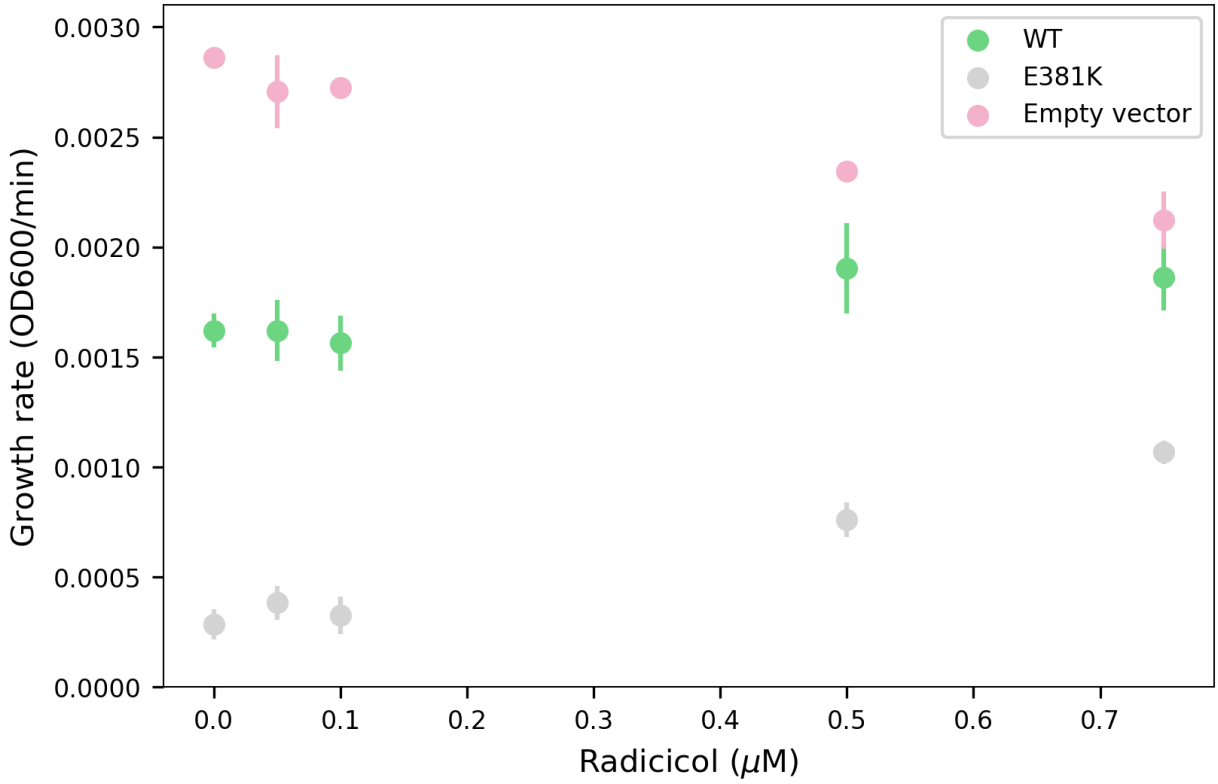
deep mutational scans is that they only describe variant effects but not the mechanism underlying the variant effect. For example, loss of activity can arise through the loss of catalytic residues, substrate affinity, or misfolding of the active site. As a result, trends revealed by deep mutational scans often appear to follow canonical rules of protein folding and activity, rather than revealing new mechanistic insights. To overcome this limitation, deep mutational scans should involve multiple phenotypes, such as in recent work on ddPCA[168]. This approach will help answer the question of how variants alter stability, binding, and function, enabling us to gain a deeper understanding of protein function from deep mutational scans beyond the current understanding of protein folding and function.

In Chapter 3, I employ deep mutational scanning to investigate the molecular determinants underlying the Hsp90 chaperoning of its client proteins[169]. Specifically, this study examined changes in the activity of the model Hsp90 client, Src kinase, following Hsp90 inhibition. The results revealed that the disengagement of Src's regulatory elements was not sufficient to drive Hsp90 dependence, contrary to the conventional wisdom in the field. Conformational flexibility also contributed to increased Hsp90 dependence, with variations at two key structural hotspots in the catalytic domain, the αF pocket and the β1-β2 strands, meeting the two criteria. The discovery of these structural regions represents the first identification of Hsp90-dependent sites in a client kinase.

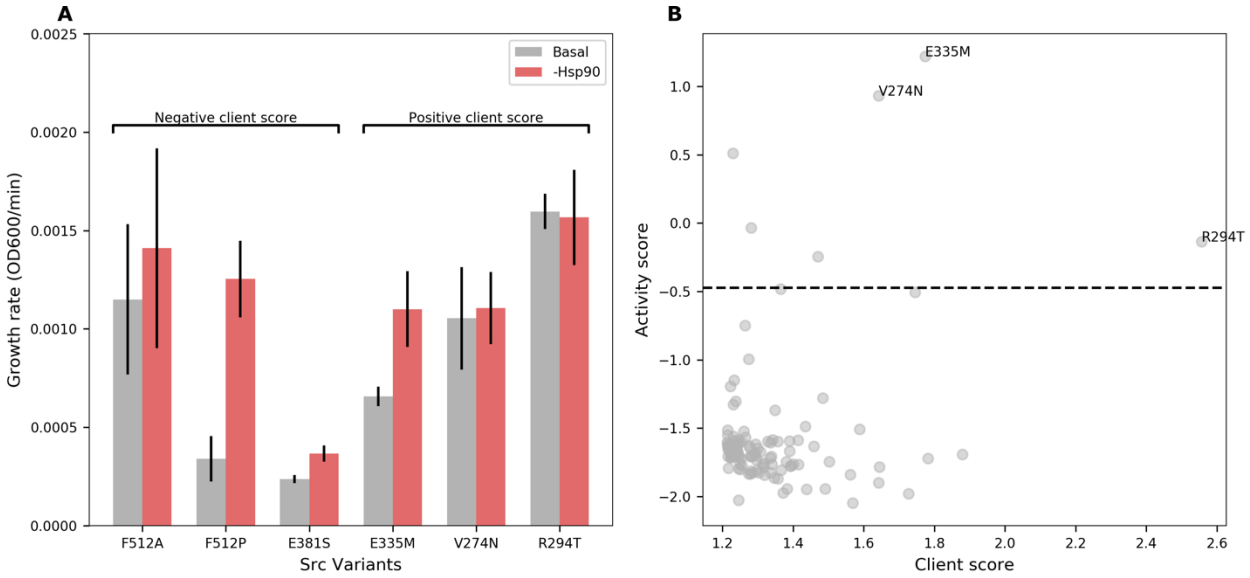
It should be noted that Src is just one of the many Hsp90 clients, and the overexpression and hyperactivity of Src in numerous cancers are typically driven by other signaling molecules that activate Src, such as EGFR and HER2[170]. Therefore, the natural extension of this work would be to investigate whether variations in the αF pocket and the β1-β2 strands in other kinase families also give rise to increased Hsp90 dependence. Interestingly, one variant in the αF pocket of EGFR has been previously reported to increase Hsp90 dependence[144]. These studies need not undertake a full deep mutational scan of the entire catalytic domain, as conducted in this study. Instead, a small library focused on the identified structural hotspots in each kinase can be screened for Hsp90 dependence. These scans, combined with the current findings, might yield a more generalized model of Hsp90 recognition of client kinases, allowing for the prediction of Hsp90 dependence from only a protein's primary sequence. This could inform the design of cancer therapeutics targeting oncogenic kinases by inhibiting Hsp90 binding of the extended conformation. A conformation-specific inhibitor of kinases could overcome the dose-limiting toxicity

limitations that have been faced by Hsp90 inhibitors. Furthermore, this work could provide a more comprehensive and contextual understanding of Hsp90 function in chaperoning kinases as well as an experimental framework for further study of the diverse world of Hsp90 clients.

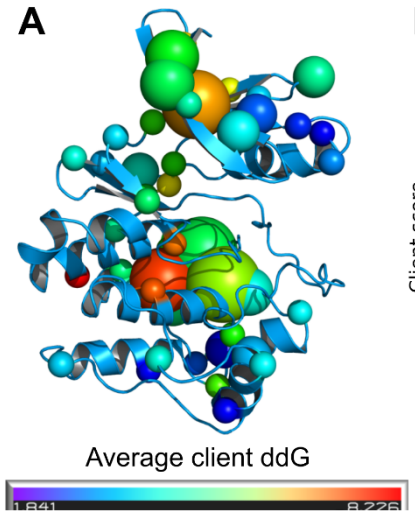
Supplementary figures

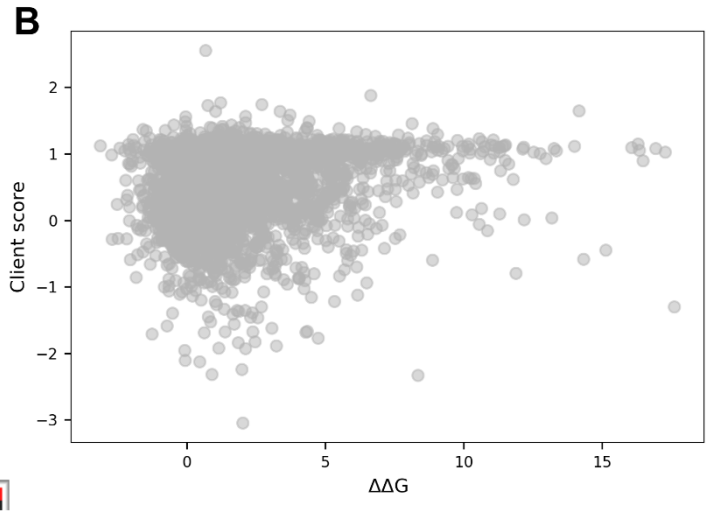


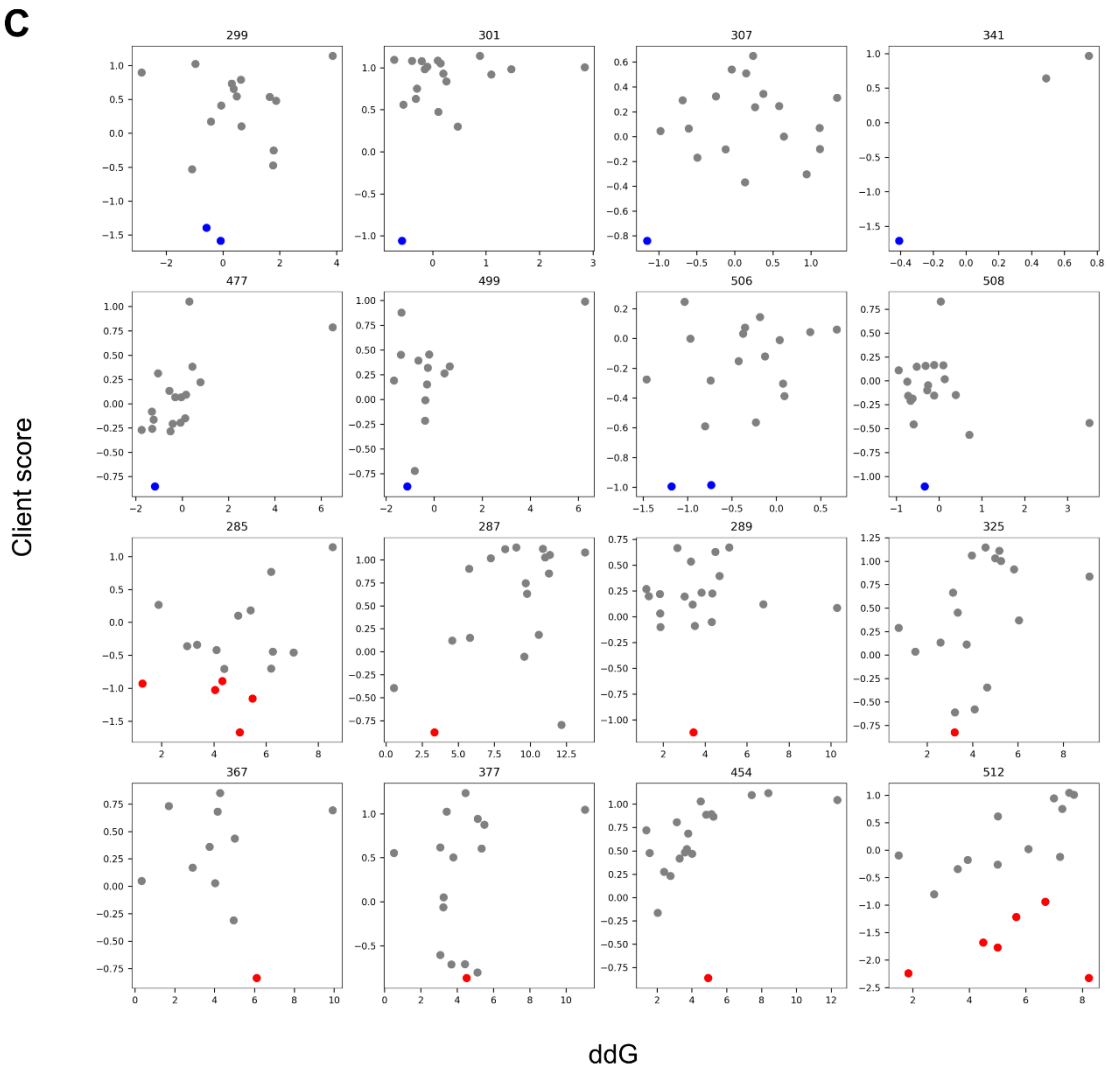
Supplementary Figure 1. Dose response of yeast expressing Src_{WT}, E381K, and empty vector. Error bars represent the standard deviation of n=3 biological replicates.



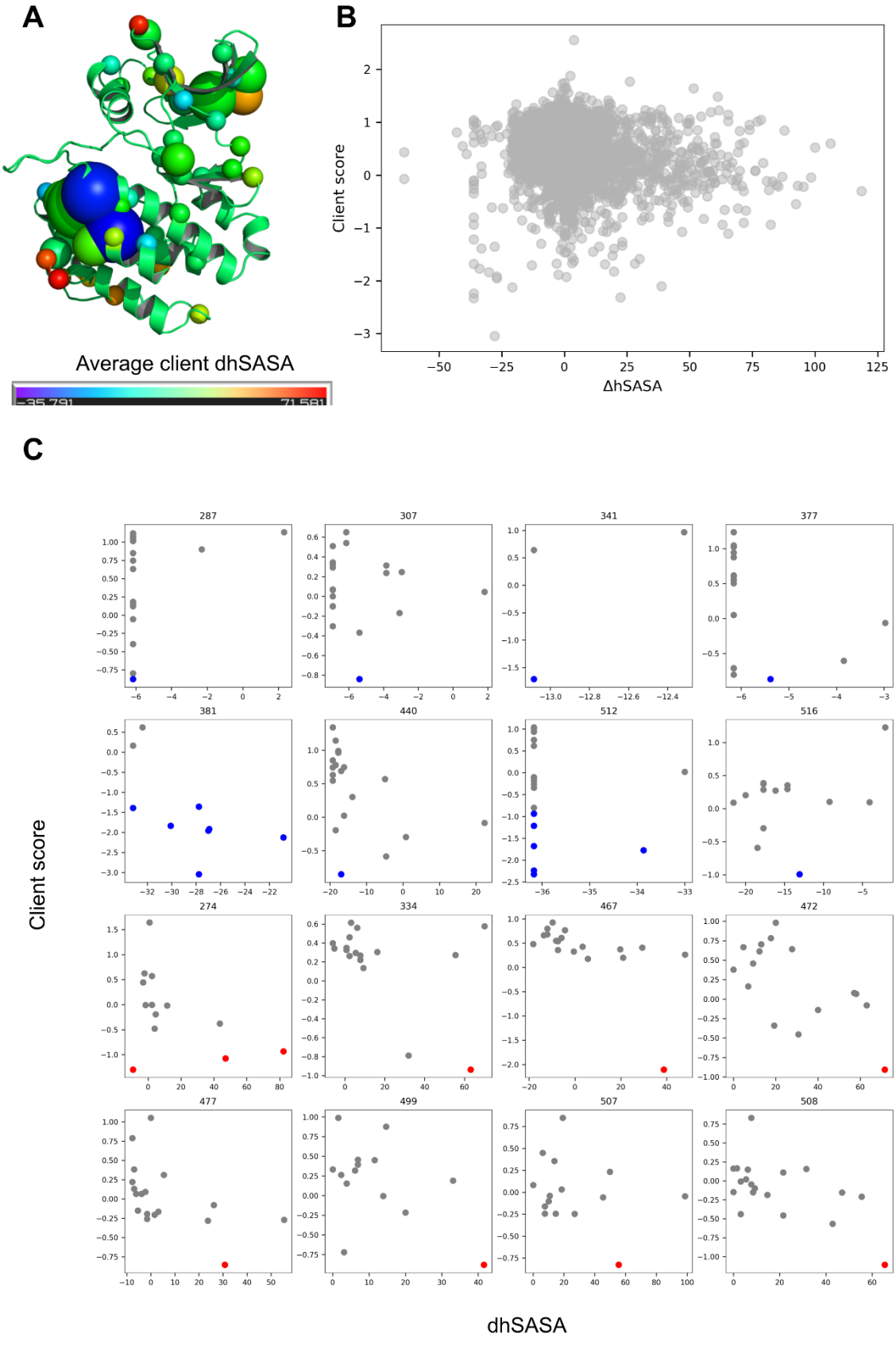
Supplementary Figure 2. (A) Individually measured growth rates of variants with large negative or positive client scores. Error bars represent n=3 biological replicates. (B) Variants with client scores greater than two standard deviations away from the mean synonymous client score are also loss of function variants (loss of function threshold is represented by the dashed line).


A
Average client ddG
1.841
8.226

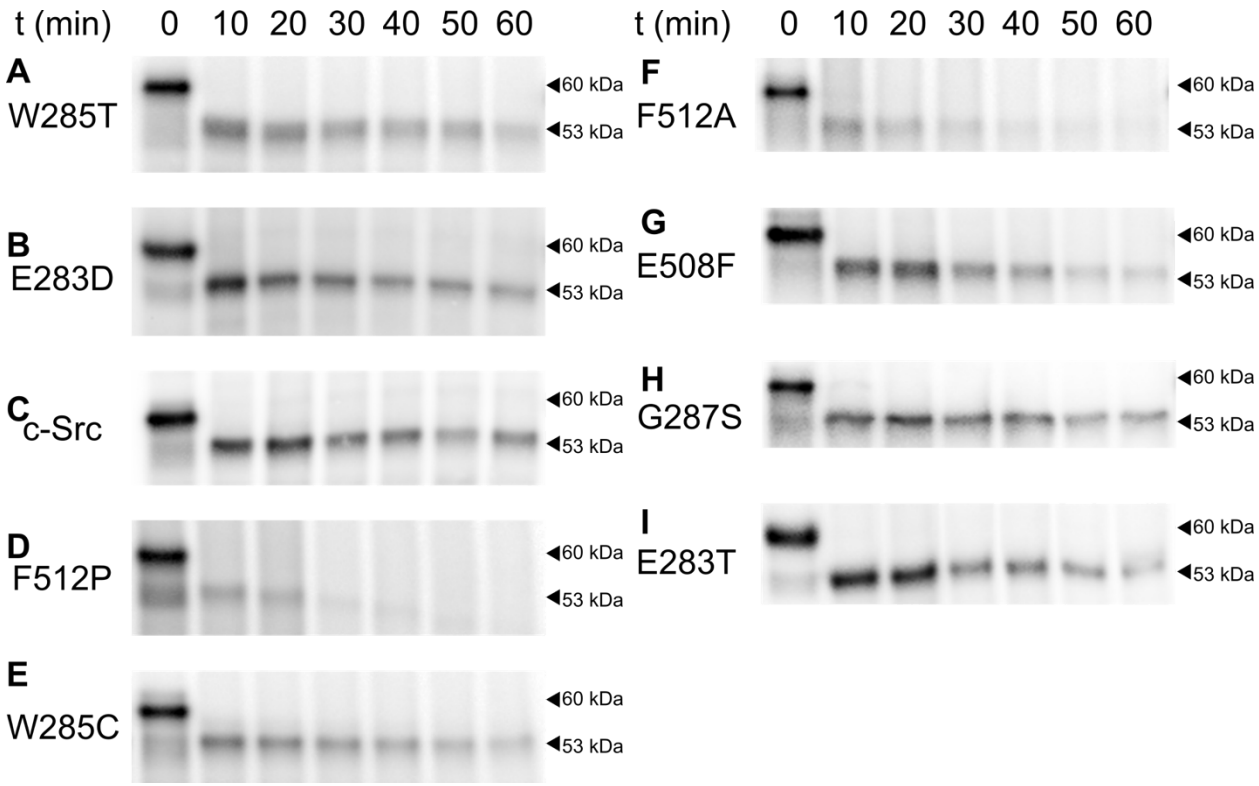

B
Client score
ΔΔG


C
Client score
ddG
299
301
307
341
477
499
506
508
285
287
289
325
367
377
454
512

Supplementary Figure 3. (A) Src CD (PDB ID:1Y57) colored by average predicted client ΔΔG at each position. (B) Global correlation between ΔΔG and measured client score shows limited correlation of R = 0.22. (C) Correlations between client score and ΔΔG at the 8 positions with the lowest average ΔΔG (blue) and the highest average ΔΔG (red) show that changes in ΔΔG account for little variability in Hsp90 dependence.

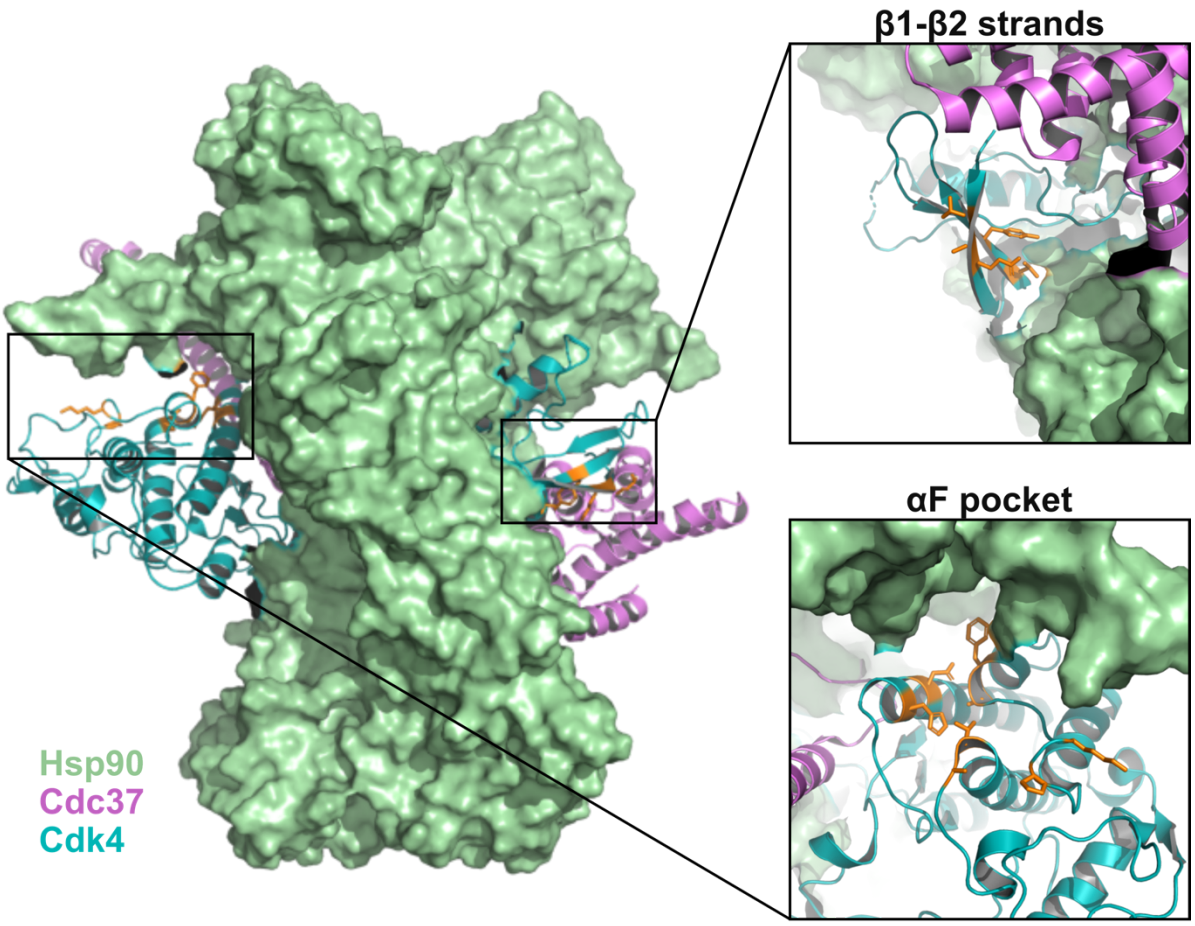

A
Average client dhSASA
−35.791
71.581
B
Client score
ΔhSASA
C
Client score
dhSASA
287
307
341
377
381
440
512
516
274
334
467
472
477
499
507
508

Supplementary Figure 4. (A) Src CD (PDB ID:1Y57) colored by average client ΔhSASA at each position. (B) Global correlation between ΔhSASA and measured client score shows limited correlation of R = -0.14. (C) Correlations between client score and ΔhSASA at the 8 positions with the lowest average ΔhSASA (blue) and the highest average ΔhSASA (red) show that changes in ΔhSASA account for little variability in Hsp90 dependence.

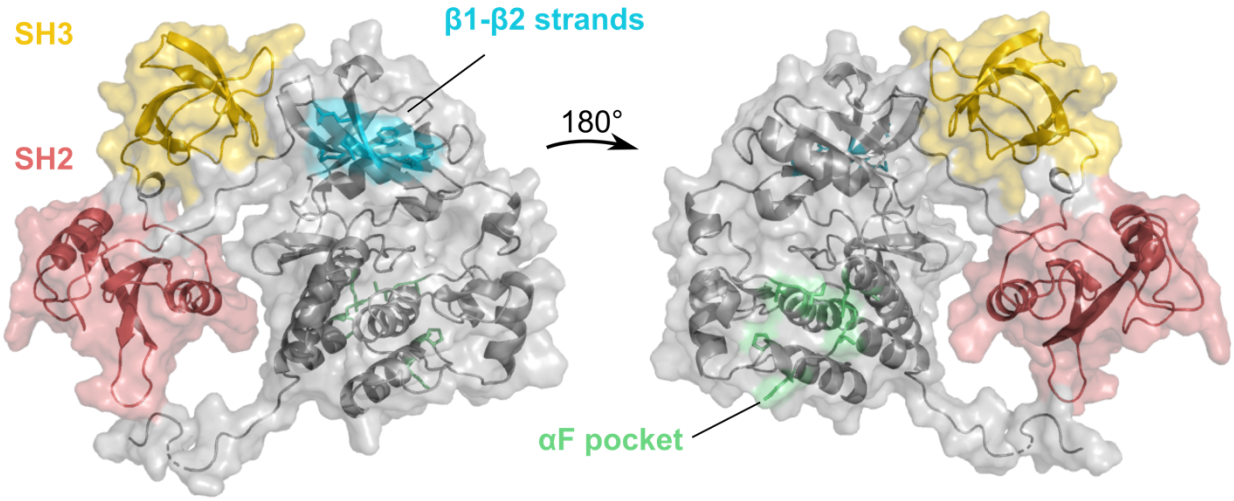


Supplementary Figure 5. (A-I) Representative blots showing time-dependent analysis of Src protein remaining following thermolysin treatment. c-Src is used to denote Src_{WT}. Src levels were quantified by a C-terminus binding antibody, with the lower band representing Src following cleavage of the SH4- and Unique domains by thermolysin.

Supplementary Figure 6. Relative thermolysin half-lives of Src variants expressed in HEK293T cells and purified by binding to dasatinib beads. W285T is a compact conformation control and E283D is an extended conformation control. P-values are the result of two-tailed t-tests with n ≥ 3.



Supplementary Figure 7. Structure of Cdk4-Cdc37-Hsp90 (Verba 2016), with homologous hotspot positions (orange) highlighted on the kinase domain of Cdk4 (blue).



Supplemental Figure 8. Structure of c-Src (PDB ID: 2SRC) shown with αF pocket (green) and β1-β2 strands (blue) highlighted. SH2 domain (red) and SH3 domain (yellow) are bound to the kinase domain in the closed conformation of c-Src.

Supplementary Tables

Functional element	Model system	Assay type	DOI
IgG	Yeast	Stability	10.1016/j.jmb.2012.07.017
Pab1	Yeast	Stability	10.1261/rna.040709.113
Yeast degron	Yeast	Stability	10.1074/mcp.O113.031708
Transmembrane segment	Bacteria	Stability	10.7554/eLife.12125
Env	HIV	Stability, binding	10.1371/journal.ppat.1005988
Env	HIV	Stability, binding, activity	10.1371/journal.ppat.1006114
M segment	Influenza	Stability	10.1186/s12864-015-2358-7
env	293FT	Stability, binding, activity	10.1128/JVI.00804-16
Fis1p	Yeast	Stability	10.1534/genetics.116.196428
NP	Virus	Stability, binding, activity	10.1371/journal.ppat.1006288
TEM-1 BL, LGK	Yeast, Bacteria	Stability	10.1073/pnas.1614437114
TP53 synonymous	HEK293T	Stability	10.1158/1541-7786.MCR-17-0245
PTEN + TPMT	human	Stability	10.1038/s41588-018-0122-z
MS2 phage capsid	MS2	Stability	10.1038/s41467-018-03783-y
Env	Human/HIV	Stability, binding, activity	10.7554/eLife.34420
adenylate kinase	E.coli	Stability	10.1093/nar/gky255
amyloid beta 42	Yeast	Stability	10.1534/g3.119.400535
TDP-43	Yeast	Stability	10.1038/s41467-019-12101-z
Zika envelope	Human/Zika	Stability, binding	10.1128/JVI.01291-19
pyrrolidine ketide synthase, levoglucosan kinase	E.coli	Stability	10.1021/acssynbio.8b00486
SCN5A		Stability, binding, activity	10.1161/CIRCGEN.119.002786
rhodopsin	HEK293T	Stability	10.1002/humu.23762
Zika envelope	Zika	Stability	10.1038/s41564-019-0399-4

AmiE	E.coli	Stability	10.1093/molbev/msz184
Env	Human	Stability, binding	10.3390/v11050439
GB1		Stability	10.1073/pnas.1903888116
PilE		Stability, binding	10.15252/embj.2019102145
LINE-1		Stability	10.1534/genetics.119.302601
AAV2 capsid	AAV	Stability	10.1126/science.aaw2900
HLA-A*02:01	Expi293F	Stability	10.1073/pnas.1915562116
AK		Stability	10.1093/protein/gzaa012
SARS-CoV-2 spike	yeast	Stability, binding	10.1016/j.cell.2020.08.012
NUDT15	HEK293T	Stability	10.1073/pnas.1915680117
CYP2C9, CYP2C19	HEK293T	Stablity	10.1111/cts.12758
DHFR	E.coli	Stability	10.7554/eLife.53476
LamB	E.coli	Stability, activity	10.1099/mgen.0.000364
VKOR	HEK293T	Stability	10.7554/eLife.58026
alpha-synuclein	Yeast	Stability	10.1038/s41589-020-0480-6
rhodopsin	HEK293T	Stability	10.1126/sciadv.aay7505
T1R2	Expi293F	Stability	10.1074/jbc.RA118.006173
LipA		Stability	10.1021/acs.jcim.9b00954
env	Macaque	Stability, binding	10.3390/v12020241
KCNH2	HEK293T	Stability	10.1016/j.hrthm.2020.05.041
NUDT15	HEK293T	Stability	10.1073/pnas.1915680117
CYP2C9	Yeast and HEK293T	Stability, activity	10.1016/j.ajhg.2021.07.001
chymotrypsin inhibitor 2	E.coli	Stability	10.1038/s42003-021-02490-7
Rhodopsin	HEK293T	Stability	10.1016/j.jbc.2021.101359
Capsid		Stability	10.7554/eLife.64256
APP		Stability	10.7554/eLife.63364
AAV2 capsid	AAV	Stability	10.1038/s41587-020-00793-4
SLCO1B1	HEK293T	Stability	10.1124/dmd.120.000264
PTEN	HEK293T	Stability	10.1186/s13073-021-00984-x
KCNH2	HEK293T	Stability	10.1016/j.ajhg.2022.05.003

CD86	HeLa	Stability	10.1016/j.jbc.2021.100900
Kir2.1	HEK293T	Stability	10.7554/eLife.76903

Table 2.1. Deep mutational scans measuring variant stability.

Functional element	Model system	Assay type	DOI
VH domain	Phage	Binding	10.1074/jbc.M708536200
WW domain	Phage	Binding	10.1038/nmeth.1492
BH3	Yeast	Binding	10.1016/j.jmb.2010.03.058
Synthetic PDZ domain	Phage	Binding	10.1039/c0mb00061b
IgG	Phage	Binding	10.1073/pnas.1111218108
PSD95 PDZ	Bacteria	Binding	10.1038/nature11500
Neurotensin receptor 1 GPCR	E.coli	Binding	10.1073/pnas.1202107109
Designed influenza binder	Yeast	Binding	10.1038/nbt.2214
WW domain	Phage	Binding	10.1073/pnas.1209751109
Fab antibody fragment	in vitro	Binding	10.1016/j.bbrc.2012.10.066
IgG	Human cells	Binding	10.4161/mabs.24979
Ubiquitin	Phage	Binding	10.1126/science.1230161
PKA RII	Phage	Binding	10.1074/jbc.M112.447326
Designed protein binder	Yeast	Binding	10.1016/j.jmb.2013.06.035
Designed digoxigenin binder	Yeast	Binding	10.1038/nature12443
Gal4	Yeast	Binding	10.1038/nmeth.3223
Ubiquitin	Yeast	Binding	10.1016/j.jmb.2014.05.019
ARRB1	E. coli	Binding	10.1073/pnas.1319402111
Hemagglutinin	Influenza	Binding	10.7554/eLife.03300
GB1	mRNA display	Binding	10.1016/j.cub.2014.09.072
Designed IgG	Yeast	Binding	10.1073/pnas.1313605111
TNF;PTxS2;TROP2	Yeast	Binding	10.1074/jbc.M115.676635
BH3	Yeast	Binding	10.1016/j.jmb.2014.09.025
PrP	Yeast	Binding	10.1016/j.jmb.2014.10.024
PhoQ	Bacteria	Binding	10.1126/science.1257360
IgG CDR	Phage	Binding	10.1074/jbc.M115.662783
ParD	Bacteria	Binding	10.1016/j.cell.2015.09.055

CcdB, DgkA	Bacteria	Binding	10.7554/eLife.09532
VWF	Phage	Binding	10.1073/pnas.1511328112
dockerin	Yeast	Binding	10.1002/prot.25175
Env	HIV	Stability, binding	10.1371/journal.ppat.1005988
CD40, TANK	Bacteria	Binding	10.1002/pro.2881
eOD-GT6 (immunogen)	Yeast	Binding	10.1126/science.aad9195
TCR	Yeast	Binding	10.1074/jbc.M116.748681
scFV antibody	Yeast	Binding	10.7554/eLife.23156
HA	Yeast	Binding	10.1371/journal.pone.0164296
lac repressor	Bacteria	Binding	10.1038/nmeth.3696
Env	HIV	Stability, binding, activity	10.1371/journal.ppat.1006114
GB1	mRNA display	Binding	10.7554/eLife.16965
env	293FT	Stability, binding, activity	10.1128/JVI.00804-16
Ras	Bacteria	Binding	10.7554/eLife.27810
NP	Virus	Stability, binding, activity	10.1371/journal.ppat.1006288
IgG	Phage	Binding	10.1073/pnas.1613231114
Cyclic peptide	Yeast	Binding	10.1074/jbc.M116.764225
ancestral RH	Yeast	Binding	10.1038/nature23902
IgG heavy chain	Phage	Binding	10.1080/19420862.2017.1337618
HA	Influenza	Binding	10.1371/journal.ppat.1006271
Env	HIV	Binding	10.1016/j.chom.2017.05.003
HA RBS	Human cells	Binding	10.1016/j.chom.2017.05.011
BBSome	yeast	Binding	10.1038/nmeth.4464
CXCR4, CCR5	Expi293F	Binding	10.4049/jimmunol.1800343
TP53	HEK293T	Binding	10.1016/j.molcel.2018.06.012
Ste12	Yeast	Binding	10.1073/pnas.1805882115
Env	Human/HIV	Stability, binding, activity	10.7554/eLife.34420
CDR	PnP hybridoma cells	Binding	10.1093/nar/gky550
EGFR	E.coli	Binding, activity	10.1073/pnas.1803598115
PDZ	Bacteria	Binding	10.7554/eLife.34300

APPI	Yeast	Binding	10.1038/s41467-018-06403-x
canine nerve growth factor	Yeast	Binding	10.1002/bit.26706
AP-1	Yeast	Binding	10.7554/eLife.32472
env	SupT1.CCR5	Binding	10.1371/journal.ppat.1007159
env	C6/36, A549, hCMEC/D3	Binding	10.1016/j.isci.2018.02.005
RAF	E.coli	Binding	10.1021/acschembio.9b00669
Zika envelope	Human/Zika	Stability, binding	10.1128/JVI.01291-19
SCN5A		Stability, binding, activity	10.1161/CIRCGEN.119.002786
Env	HIV	Binding	10.1016/j.immuni.2018.12.017
matrix protein	Human/influenza A	Binding	10.1128/JVI.00161-19
PB2	Influenza	Binding	10.7554/eLife.45079
Proliferation-inducing ligand	Yeast	Binding	10.1002/jmr.2778
CD19	Yeast	Binding	10.1021/acs.molpharmaceut.9b00418
Env	Expi293F	Binding	10.1128/JVI.00219-19
IgG Fv targeting lysozyme	Yeast	Binding	10.1371/journal.pcbi.1007207
Env	Human	Stability, binding	10.3390/v11050439
TP53	K562, MOLM13	Binding	10.1126/science.aax3649
HA		Binding	10.7554/eLife.49324
CI		Binding	10.1038/s41467-019-11735-3
PilE		Stability, binding	10.15252/embj.2019102145
SARS-CoV-2 spike	yeast	Stability, binding	10.1016/j.cell.2020.08.012
PSD95 PDZ	E.coli	Binding	10.1002/prot.26067
ACE2	Expi293F	Binding	10.1126/science.abc0870
IgG	Yeast	Binding	10.1080/19420862.2020.1803646
basigin	Yeast	Binding	10.1002/prot.25786
env	Macaque	Stability, binding	10.3390/v12020241
Hsp90	Yeast	Binding	10.7554/eLife.53810
237-CAR		Binding	10.1073/pnas.1920662117

PDGFRA		Binding	10.1371/journal.ppat.1008647
ParB		Binding	10.1016/j.celrep.2020.107928
Spike	Yeast	Binding	10.1016/j.chom.2020.11.007
PHO4	Yeast	Binding	10.1016/j.cels.2020.11.012
Polymerase sliding clamp and clamp loader	Phage	Binding, activity	10.7554/eLife.66181
bnAbs	Yeast	Binding	10.7554/eLife.71393
Spike	HEK293T	Binding	10.1016/j.molcel.2021.11.024

Table 2.2. Deep mutational scans measuring variant binding.

Functional element	Model system	Assay type	DOI
Hsp90	Yeast	Activity	10.1073/pnas.1016024108
Ccdb	E.coli	Activity	10.1016/j.str.2011.11.021
Neuraminidase	Human/H1N1	Activity	10.1128/JVI.01658-12
M.HaeIII	E.coli	Activity	10.1371/journal.pgen.1003882
TEM1 Beta-lactamase	E.coli	Activity	10.1016/j.jmb.2012.09.014
E4B	Phage	Activity	10.1073/pnas.1303309110
Ubiquitin	Yeast	Activity	10.1016/j.jmb.2013.01.032
Hsp90	Yeast	Activity	10.1371/journal.pgen.1003600
Hsp90	Yeast	Activity	10.1111/evo.12207
BRAF V600E	Human	Activity	10.1111/pcmr.12171
TEM1 Beta-lactamase	E.coli	Activity	10.1073/pnas.1215206110
TEM-1 BL	Bacteria	Activity	10.1093/molbev/msu081
influenza nucleoprotein	Human/H1N1	Activity	10.1093/molbev/msu173
APH(3')II	Bacteria	Activity	10.1093/nar/gku511
HA		Activity	10.1038/srep04942
AID	E.coli	Activity	10.1093/nar/gku689
NS4B	Human cell culture (Huh-7.5.1)	Activity	10.1371/journal.ppat.1004064
flavin mononucleotide binding fluorescent protein	E.coli	Activity	10.1371/journal.pone.0097817
DBR1	Hap1	Activity	10.1038/nature13695
HIV genome	HIV	Activity	10.1186/s12977-014-0124-6
NS1	HEK293T, A549	Activity	10.1128/JVI.01494-14
BRCA1	Yeast, Phage	Activity	10.1534/genetics.115.175802
TEM1 B-lactamase	E.coli	Activity	10.1016/j.cell.2015.01.035
Nucleoproteins	Influenza	Activity	10.1093/molbev/msv167
M.HaeIII	Bacteria	Activity	10.1371/journal.pcbi.1004421

Hsp90	Yeast	Activity	10.1093/molbev/msu301
Bgl2	Bacteria/microfluidic droplets	Activity	10.1073/pnas.1422285112
LGK	E.coli	Activity	10.1021/acssynbio.5b00131
ACD-1	Bacteria	Activity	10.1128/AEM.03074-15
Polymerase PA	Influenza	Activity	10.1371/journal.pgen.1005310
Hsp90	Yeast	Activity	10.1016/j.celrep.2016.03.046
GFP	Bacteria	Activity	10.1038/nature17995
TEM1 Beta-lactamase	E.coli	Activity	10.1016/j.jmb.2016.04.033
Mapk1/Erk2	Human	Activity	10.1016/j.celrep.2016.09.061
Ubiquitin	Yeast	Activity	10.7554/eLife.15802
CcdB	E.coli	Activity	10.1093/molbev/msw182
PPARG	Human cells	Activity	10.1038/ng.3700
RNAPII trigger loop	Yeast	Activity	10.1371/journal.pgen.1006321
Hemagglutinin	Influenza	Activity	10.3390/v8060155
Env	HIV	Stability, binding, activity	10.1371/journal.ppat.1006114
neuraminidase (NA)	Influenza	Activity	10.1016/j.jmb.2015.11.027
Tat, Rev	HIV	Activity	10.1016/j.cell.2016.11.031
env	293FT	Stability, binding, activity	10.1128/JVI.00804-16
UBE2I, SUMO1, TPK1, CALM1-3	yeast	Activity	10.15252/msb.20177908
GFP N-terminal codon	Human cell lines (HEK293T)	Activity	10.1093/nar/gkx183
Cas9	Bacteria	Activity	10.1038/s41598-017-17081-y
NP	Virus	Stability, binding, activity	10.1371/journal.ppat.1006288
IGPS	Yeast	Activity	10.1038/ncomms14614
BCR-ABL	Ba/F3 cells	Activity	10.1073/pnas.1708268114
amiE	E.coli	Activity	10.1038/ncomms15695
Gcn4	Yeast	Activity	10.1016/j.cels.2018.01.015
PTEN	Yeast	Activity	10.1016/j.ajhg.2018.03.018
TP53	human	Activity	10.1038/s41588-018-0204-y
BRCA1	human	Activity	10.1016/j.ajhg.2018.07.016

BRCA1	HAP1	Activity	10.1038/s41586-018-0461-z
Env	Human/HIV	Stability, binding, activity	10.7554/eLife.34420
tetracycline inactivating enzyme	Yeast	Activity	10.1021/acssynbio.8b00121
EGFR	E.coli	Binding, activity	10.1073/pnas.1803598115
Ubiquitin	Yeast	Activity	10.1242/bio.036103
H1 hemaglutinin	Influenza	Activity	10.1038/s41467-018-03665-3
HA	Virus	Activity	10.1073/pnas.1806133115
HA	Influenza	Activity	10.7554/eLife.38795
PPAT	E.coli	Activity	10.1126/science.aao5167
GmR	E.coli	Activity	10.1371/journal.pgen.1007419
Lysine metabolic pathway	Bacteria	Activity	10.15252/msb.20188371
HA	Human/infulenza	Activity	10.1016/j.jmb.2018.02.009
Sox17, Sox2	OG2-MEF	Activity	10.1016/j.stemcr.2018.07.002
Src	Yeast	Activity	10.1016/j.molcel.2019.02.003
SCN5A		Stability, binding, activity	10.1161/CIRCGEN.119.002786
Kod DNA polymerase	E.coli	Activity	10.1021/acssynbio.9b00104
HIV-1 Vif	Human cells	Activity	10.1016/j.celrep.2019.09.057
IGPD	Yeast	Activity	10.1371/journal.pgen.1008079
GDH (GudB)	*Bacillus subtilis*	Activity	10.1038/s41564-019-0412-y
RPL28	Yeast	Activity	10.1021/acssynbio.8b00529
TEM-1		Activity	10.1016/j.jmb.2019.04.030
CBS	Yeast	Activity	10.1186/s13073-020-0711-1
MPL	Ba/F3 cells	Activity	10.1182/blood.2019002561
alpha-synuclein	Yeast	Activity	10.1021/acschembio.0c00339
LamB	E.coli	Stability, activity	10.1099/mgen.0.000364
VIM-2 lactamase	E.coli	Activity	10.7554/eLife.56707
TEM-1 BL	E.coli	Activity	10.1073/pnas.1918680117
Human and Rhesus TRIM5α v1 Loop	CRFK	Activity	10.7554/eLife.59988
HA	Influenza	Activity	10.1126/science.aaz5143

CDK4, CDK6	Human cells	Activity	10.1038/s41594-019-0358-z
rpoB	E.coli	Activity	10.15252/msb.20199265
ADRB2	HEK293T	Activity	10.7554/eLife.54895
HA	Influenza	Activity	10.1038/s41467-020-15102-5
Kod DNA polymerase	E. Coli	Activity	10.1021/acssynbio.0c00236
HIV-1 protease		Activity	10.1371/journal.pgen.1009009
CARD11	TMD8	Activity	10.1016/j.ajhg.2020.10.015
MTHFR	Yeast	Activity	10.1016/j.ajhg.2021.05.009
MSH2	HAP1	Activity	10.1016/j.ajhg.2020.12.003
CYP2C9	Yeast and HEK293T	Stability, activity	10.1016/j.ajhg.2021.07.001
HSP90	Yeast	Activity	10.1093/molbev/msaa211
PAI-1	Phage	Activity	10.1038/s41598-021-97871-7
Polymerase cliding clamp and clamp loader	Phage	Binding, activity	10.7554/eLife.66181
biosensors		Activity	10.1021/acschembio.1c00423
ProQ	Salmonella	Activity	10.1261/rna.078954.121
HokC	E.coli	Activity	10.3390/ijms221910359
GFP	E.coli	Activity	10.7554/eLife.75842
CASP3, CASP7		Activity	10.1038/s41420-021-00799-0
ADAR2	HEK293FT	Activity	10.7554/eLife.75555
7 transcriptional activation domains within VP16, CITED2, HIF1A, P65, STAT3, P53 (2)	K562	Activity	10.1016/j.cels.2022.01.002
Mpro	Yeast	Activity	10.7554/eLife.77433
BRACA1-BRCT	HeLa	Activity	10.1016/j.ajhg.2022.01.019

Table 2.3. Deep mutational scans measuring variant activity

	Class	n variants	n clients
374	Adjacent	17	0
375	Adjacent	15	0
376	Adjacent	13	0
377	Adjacent	16	1
378	αF pocket	13	1
379	Adjacent	18	0
380	Adjacent	17	0
381	αF pocket	9	7
382	Adjacent	14	0
383	Adjacent	8	0
384	Adjacent	16	0
385	Adjacent	13	0
386	Adjacent	18	0
441	Adjacent	20	0
442	Adjacent	19	0
443	αF pocket	18	3
444	αF pocket	19	8
445	Adjacent	20	0
446	Adjacent	6	0
447	Adjacent	16	0
448	Adjacent	7	0
503	Adjacent	19	0
504	Adjacent	11	1
505	Adjacent	20	0
506	αF pocket	19	2
507	Adjacent	16	1
508	αF pocket	20	1
509	Adjacent	20	0
510	Adjacent	11	0
511	αF pocket	16	3
512	αF pocket	19	6

513	Adjacent	18	1
514	Adjacent	11	0
515	Adjacent	20	0
516	Adjacent	14	1

Supplementary Table 3.1. List of αF pocket positions and positions with 7 angstroms of any αF pocket positions with corresponding variants measured at each position and functionally dependent clients identified at each position.

	Class	n variants	n clients
271	Adjacent	11	0
272	Adjacent	0	0
273	β1-β2 strand hotspot	12	3
274	β1-β2 strand hotspot	14	3
275	β1-β2 strand hotspot	13	1
276	Adjacent	8	0
277	Adjacent	13	0
278	Adjacent	19	0
279	Adjacent	15	0
282	Adjacent	12	0
283	β1-β2 strand hotspot	10	2
284	Adjacent	18	0
285	β1-β2 strand hotspot	18	5
286	β1-β2 strand hotspot	20	0
287	Adjacent	19	1
295	Adjacent	13	0
296	Adjacent	11	0
297	Adjacent	14	0
298	Adjacent	14	0
299	Adjacent	18	2
300	Adjacent	20	0
343	Adjacent	2	1

Supplementary Table 3.2. List of β1-β2 hotspot positions and positions with 7 angstroms of any αF pocket positions with corresponding variants measured at each position and functionally dependent clients identified at each position.

Region	Total clients	Total variants	Coverage	Median	Positions
αF pocket	31	133	0.83125	-0.282282	[378, 381, 443, 444, 506, 508, 511, 512]
β1-β2 strand hotspot	14	87	0.725	0.05177909	[273, 274, 275, 283, 285, 286]
SH2 interface	1	49	0.81666667	0.03911715	[325, 368, 375]
SH3 interface	1	78	0.975	0.26946102	[289, 290, 291, 293]

Supplementary Table 3.3. List of regulatory regions that extend the global conformation of Src

References

1. Ritossa, F. Discovery of the heat shock response. *Cell Stress Chaperones* **1**, 97–98 (1996).
2. De Maio, A., Santoro, M. G., Tanguay, R. M. & Hightower, L. E. Ferruccio Ritossa's scientific legacy 50 years after his discovery of the heat shock response: a new view of biology, a new society, and a new journal. *Cell Stress Chaperones* **17**, 139–143 (2012).
3. Millar, N. L. & Murrell, G. A. C. Heat shock proteins in tendinopathy: novel molecular regulators. *Mediators Inflamm.* **2012**, 436203 (2012).
4. Sumi, M. P. & Ghosh, A. Hsp90 in Human Diseases: Molecular Mechanisms to Therapeutic Approaches. *Cells* **11**, (2022).
5. Fu, X. Chaperone function and mechanism of small heat-shock proteins. *Acta Biochim. Biophys. Sin.* **46**, 347–356 (2014).
6. Taipale, M., Jarosz, D. F. & Lindquist, S. HSP90 at the hub of protein homeostasis: emerging mechanistic insights. *Nat. Rev. Mol. Cell Biol.* **11**, 515–528 (2010).
7. Schopf, F. H., Biebl, M. M. & Buchner, J. The HSP90 chaperone machinery. *Nat. Rev. Mol. Cell Biol.* **18**, 345–360 (2017).
8. Hoter, A., El-Sabban, M. E. & Naim, H. Y. The HSP90 Family: Structure, Regulation, Function, and Implications in Health and Disease. *Int. J. Mol. Sci.* **19**, (2018).
9. Murshid, A., Gong, J. & Calderwood, S. K. Heat shock protein 90 mediates efficient antigen cross presentation through the scavenger receptor expressed by endothelial cells-I. *J. Immunol.* **185**, 2903–2917 (2010).
10. Li, J. & Buchner, J. Structure, function and regulation of the hsp90 machinery. *Biomed. J.* **36**, 106–117 (2013).
11. Street, T. O., Lavery, L. A. & Agard, D. A. Substrate binding drives large-scale conformational changes in the Hsp90 molecular chaperone. *Mol. Cell* **42**, 96–105 (2011).
12. Tsutsumi, S. *et al.* Hsp90 charged-linker truncation reverses the functional consequences of weakened hydrophobic contacts in the N domain. *Nat. Struct. Mol. Biol.* **16**, 1141–1147 (2009).
13. Ali, M. M. U. *et al.* Crystal structure of an Hsp90-nucleotide-p23/Sba1 closed chaperone complex. *Nature* **440**, 1013–1017 (2006).

14. Huai, Q. *et al.* Structures of the N-terminal and middle domains of E. coli Hsp90 and conformation changes upon ADP binding. *Structure* **13**, 579–590 (2005).
15. Li, J., Soroka, J. & Buchner, J. The Hsp90 chaperone machinery: conformational dynamics and regulation by co-chaperones. *Biochim. Biophys. Acta* **1823**, 624–635 (2012).
16. Martínez-Ruiz, A. *et al.* S-nitrosylation of Hsp90 promotes the inhibition of its ATPase and endothelial nitric oxide synthase regulatory activities. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 8525–8530 (2005).
17. Verba, K. A. *et al.* Atomic structure of Hsp90-Cdc37-Cdk4 reveals that Hsp90 traps and stabilizes an unfolded kinase. *Science* **352**, 1542–1547 (2016).
18. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2018. *CA Cancer J. Clin.* **68**, 7–30 (2018).
19. Shafi, A. A., Yen, A. E. & Weigel, N. L. Androgen receptors in hormone-dependent and castration-resistant prostate cancer. *Pharmacol. Ther.* **140**, 223–238 (2013).
20. Zagouri, F. *et al.* Hsp90 inhibitors in breast cancer: a systematic review. *Breast* **22**, 569–578 (2013).
21. Sims, J. D., McCready, J. & Jay, D. G. Extracellular heat shock protein (Hsp)70 and Hsp90α assist in matrix metalloproteinase-2 activation and breast cancer cell migration and invasion. *PLoS One* **6**, e18848 (2011).
22. Takalo, M., Salminen, A., Soininen, H., Hiltunen, M. & Haapasalo, A. Protein aggregation and degradation mechanisms in neurodegenerative diseases. *Am. J. Neurodegener. Dis.* **2**, 1–14 (2013).
23. Poewe, W. *et al.* Parkinson disease. *Nat Rev Dis Primers* **3**, 17013 (2017).
24. Hu, S. *et al.* Molecular chaperones and Parkinson's disease. *Neurobiol. Dis.* **160**, 105527 (2021).
25. Trepel, J., Mollapour, M., Giaccone, G. & Neckers, L. Targeting the dynamic HSP90 complex in cancer. *Nat. Rev. Cancer* **10**, 537–549 (2010).
26. Wang, X., Chen, M., Zhou, J. & Zhang, X. HSP27, 70 and 90, anti-apoptotic proteins, in clinical cancer therapy (Review). *Int. J. Oncol.* **45**, 18–30 (2014).
27. Koay, Y. C., Wahyudi, H. & McAlpine, S. R. Reinventing Hsp90 Inhibitors: Blocking C-Terminal Binding Events to Hsp90 by Using Dimerized Inhibitors. *Chemistry* **22**, 18572–18582 (2016).
28. Park, H.-K. *et al.* Unleashing the full potential of Hsp90 inhibitors as cancer therapeutics through simultaneous inactivation of Hsp90, Grp94, and TRAP1. *Exp. Mol. Med.* **52**, 79–91 (2020).
29. Fadden, P. *et al.* Application of chemoproteomics to drug discovery: identification of a clinical

candidate targeting hsp90. *Chem. Biol.* **17**, 686–694 (2010).

30. Taipale, M. *et al.* Quantitative analysis of HSP90-client interactions reveals principles of substrate recognition. *Cell* **150**, 987–1001 (2012).
31. Howlader, M. T. H. *et al.* Alanine scanning analyses of the three major loops in domain II of Bacillus thuringiensis mosquitocidal toxin Cry4Aa. *Appl. Environ. Microbiol.* **76**, 860–865 (2010).
32. Simonsen, S. M. *et al.* Alanine scanning mutagenesis of the prototypic cyclotide reveals a cluster of residues essential for bioactivity. *J. Biol. Chem.* **283**, 9805–9813 (2008).
33. Gauguin, L. *et al.* Alanine Scanning of a Putative Receptor Binding Surface of Insulin-like Growth Factor-I *. *J. Biol. Chem.* **283**, 20821–20829 (2008).
34. Edelheit, O., Hanukoglu, I., Dascal, N. & Hanukoglu, A. Identification of the roles of conserved charged residues in the extracellular domain of an epithelial sodium channel (ENaC) subunit by alanine mutagenesis. *Am. J. Physiol. Renal Physiol.* **300**, F887–97 (2011).
35. Weiss, G. A., Watanabe, C. K., Zhong, A., Goddard, A. & Sidhu, S. S. Rapid mapping of protein functional epitopes by combinatorial alanine scanning. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 8950–8954 (2000).
36. Tabasinezhad, M. *et al.* Trends in therapeutic antibody affinity maturation: From in-vitro towards next-generation sequencing approaches. *Immunol. Lett.* **212**, 106–113 (2019).
37. Tinberg, C. E. *et al.* Computational design of ligand-binding proteins with high affinity and selectivity. *Nature* **501**, 212–216 (2013).
38. Procko, E. *et al.* Computational design of a protein-based enzyme inhibitor. *J. Mol. Biol.* **425**, 3563–3575 (2013).
39. Fowler, D. M. & Fields, S. Deep mutational scanning: a new style of protein science. *Nat. Methods* **11**, 801–807 (2014).
40. Kinney, J. B. & McCandlish, D. M. Massively Parallel Assays and Quantitative Sequence-Function Relationships. *Annu. Rev. Genomics Hum. Genet.* **20**, 99–127 (2019).
41. Frazer, J. *et al.* Disease variant prediction with deep generative models of evolutionary data. *Nature* **599**, 91–95 (2021).
42. Jagota, M. *et al.* Cross-protein transfer learning substantially improves zero-shot prediction of

disease variant effects. *bioRxiv* 2022.11.15.516532 (2022) doi:10.1101/2022.11.15.516532.

43. Weile, J. *et al.* Shifting landscapes of human MTHFR missense-variant effects. *Am. J. Hum. Genet.* **108**, 1283–1300 (2021).
44. Wei, H. & Li, X. Deep mutational scanning: A versatile tool in systematically mapping genotypes to phenotypes. *Front. Genet.* **14**, 1087267 (2023).
45. Kemble, H., Nghe, P. & Tenaillon, O. Recent insights into the genotype-phenotype relationship from massively parallel genetic assays. *Evol. Appl.* **12**, 1721–1742 (2019).
46. Narayanan, K. K. & Procko, E. Deep Mutational Scanning of Viral Glycoproteins and Their Host Receptors. *Front Mol Biosci* **8**, 636660 (2021).
47. Hanning, K. R., Minot, M., Warrender, A. K., Kelton, W. & Reddy, S. T. Deep mutational scanning for therapeutic antibody engineering. *Trends Pharmacol. Sci.* **43**, 123–135 (2022).
48. Cherf, G. M. & Cochran, J. R. Applications of Yeast Surface Display for Protein Engineering. *Methods Mol. Biol.* **1319**, 155–175 (2015).
49. Seuma, M., Faure, A. J., Badia, M., Lehner, B. & Bolognesi, B. The genetic landscape for amyloid beta fibril nucleation accurately discriminates familial Alzheimer's disease mutations. *Elife* **10**, (2021).
50. Gray, V. E. *et al.* Elucidating the Molecular Determinants of Aβ Aggregation with Deep Mutational Scanning. *G3* **9**, 3683–3689 (2019).
51. Kim, I., Miller, C. R., Young, D. L. & Fields, S. High-throughput Analysis of in vivo Protein Stability*. *Mol. Cell. Proteomics* **12**, 3370–3378 (2013).
52. Matreyek, K. A. *et al.* Multiplex assessment of protein variant abundance by massively parallel sequencing. *Nat. Genet.* **50**, 874–882 (2018).
53. Amorosi, C. J. *et al.* Massively parallel characterization of CYP2C9 variant enzyme activity and abundance. *Am. J. Hum. Genet.* **108**, 1735–1751 (2021).
54. Chiasson, M. A. *et al.* Multiplexed measurement of variant abundance and activity reveals VKOR topology, active site and human variant impact. *Elife* **9**, (2020).
55. Traxlmayr, M. W. *et al.* Construction of a stability landscape of the CH3 domain of human IgG1 by combining directed evolution with high throughput sequencing. *J. Mol. Biol.* **423**, 397–412 (2012).
56. Rocklin, G. J. *et al.* Global analysis of protein folding using massively parallel design, synthesis, and

testing. *Science* **357**, 168–175 (2017).

57. Nutschel, C. *et al.* Systematically Scrutinizing the Impact of Substitution Sites on Thermostability and Detergent Tolerance for Bacillus subtilis Lipase A. *J. Chem. Inf. Model.* **60**, 1568–1584 (2020).
58. Tsuboyama, K. *et al.* Mega-scale experimental analysis of protein folding stability in biology and protein design. *bioRxiv* 2022.12.06.519132 (2022) doi:10.1101/2022.12.06.519132.
59. Nisthal, A., Wang, C. Y., Ary, M. L. & Mayo, S. L. Protein stability engineering insights revealed by domain-wide comprehensive mutagenesis. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 16367–16377 (2019).
60. Wrenbeck, E. E. *et al.* An Automated Data-Driven Pipeline for Improving Heterologous Enzyme Expression. *ACS Synth. Biol.* **8**, 474–481 (2019).
61. Hartman, E. C. *et al.* Quantitative characterization of all single amino acid variants of a viral capsid-based drug delivery vehicle. *Nat. Commun.* **9**, 1385 (2018).
62. Atkinson, J. T., Jones, A. M., Nanda, V. & Silberg, J. J. Protein tolerance to random circular permutation correlates with thermostability and local energetics of residue-residue contacts. *Protein Eng. Des. Sel.* **32**, 489–501 (2019).
63. Atkinson, J. T., Jones, A. M., Zhou, Q. & Silberg, J. J. Circular permutation profiling by deep sequencing libraries created using transposon mutagenesis. *Nucleic Acids Res.* **46**, e76 (2018).
64. Zhang, L. *et al.* SLCO1B1: Application and Limitations of Deep Mutational Scanning for Genomic Missense Variant Function. *Drug Metab. Dispos.* **49**, 395–404 (2021).
65. Coyote-Maestas, W., Nedrud, D., He, Y. & Schmidt, D. Determinants of trafficking, conduction, and disease within a K+ channel revealed through multiparametric deep mutational scanning. *Elife* **11**, (2022).
66. Cagiada, M. *et al.* Understanding the origins of loss of protein function by analyzing the effects of thousands of variants on activity and abundance. *Cold Spring Harbor Laboratory* 2020.09.28.317040 (2021) doi:10.1101/2020.09.28.317040.
67. Staller, M. V. *et al.* Directed mutational scanning reveals a balance between acidic and hydrophobic residues in strong human activation domains. *Cell Syst* **13**, 334–345.e5 (2022).
68. Rollins, N. J. *et al.* Inferring protein 3D structure from deep mutation scans. *Nat. Genet.* **51**, 1170–1176 (2019).

69. Schmiedel, J. M. & Lehner, B. Determining protein structures using deep mutagenesis. *Nat. Genet.* **51**, 1177–1186 (2019).
70. Phillips, A. M. *et al.* Host proteostasis modulates influenza evolution. *Elife* **6**, (2017).
71. Phillips, A. M. *et al.* Enhanced ER proteostasis and temperature differentially impact the mutational tolerance of influenza hemagglutinin. *Elife* **7**, (2018).
72. Bolognesi, B. *et al.* The mutational landscape of a prion-like domain. *Nat. Commun.* **10**, 1–12 (2019).
73. Schlinkmann, K. M. *et al.* Critical features for biosynthesis, stability, and functionality of a G protein-coupled receptor uncovered by all-versus-all mutations. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 9810–9815 (2012).
74. McKee, A. G. *et al.* Systematic profiling of temperature- and retinal-sensitive rhodopsin variants by deep mutational scanning. *J. Biol. Chem.* **297**, 101359 (2021).
75. Penn, W. D. *et al.* Probing biophysical sequence constraints within the transmembrane domains of rhodopsin by deep mutational scanning. *Sci Adv* **6**, eaay7505 (2020).
76. Klesmith, J. R., Bacik, J.-P., Wrenbeck, E. E., Michalczyk, R. & Whitehead, T. A. Trade-offs between enzyme fitness and solubility illuminated by deep mutational scanning. *Proceedings of the National Academy of Sciences* **114**, 2265–2270 (2017).
77. Wrenbeck, E. E., Azouz, L. R. & Whitehead, T. A. Single-mutation fitness landscapes for an enzyme on multiple substrates reveal specificity is globally encoded. *Nat. Commun.* **8**, 15695 (2017).
78. Kitzman, J. O., Starita, L. M., Lo, R. S., Fields, S. & Shendure, J. Massively parallel single-amino-acid mutagenesis. *Nat. Methods* **12**, 203–6, 4 p following 206 (2015).
79. Meier, G. *et al.* Deep mutational scan of a drug efflux pump reveals its structure-function landscape. *Nat. Chem. Biol.* **19**, 440–450 (2023).
80. Chan, K. K. *et al.* Engineering human ACE2 to optimize binding to the spike protein of SARS coronavirus 2. *Science* **369**, 1261–1265 (2020).
81. Koenig, P., Sanowar, S., Lee, C. V. & Fuh, G. Tuning the specificity of a Two-in-One Fab against three angiogenic antigens by fully utilizing the information of deep mutational scanning. *MAbs* **9**, 959–967 (2017).
82. Chen, J. Z., Fowler, D. M. & Tokuriki, N. Comprehensive exploration of the translocation, stability and

substrate recognition requirements in VIM-2 lactamase. *Elife* **9**, (2020).

83. Forsyth, C. M. *et al.* Deep mutational scanning of an antibody against epidermal growth factor receptor using mammalian cell display and massively parallel pyrosequencing. *MAbs* **5**, 523–532 (2013).
84. Fujino, Y. *et al.* Robust in vitro affinity maturation strategy based on interface-focused high-throughput mutational scanning. *Biochem. Biophys. Res. Commun.* **428**, 395–400 (2012).
85. Whitehead, T. A. *et al.* Optimization of affinity, specificity and function of designed influenza inhibitors using deep sequencing. *Nat. Biotechnol.* **30**, 543–548 (2012).
86. Nishikawa, K. K., Hoppe, N., Smith, R., Bingman, C. & Raman, S. Epistasis shapes the fitness landscape of an allosteric specificity switch. *Nat. Commun.* **12**, 5562 (2021).
87. Weinreich, D. M., Delaney, N. F., Depristo, M. A. & Hartl, D. L. Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* **312**, 111–114 (2006).
88. Starr, T. N., Picton, L. K. & Thornton, J. W. Alternative evolutionary histories in the sequence space of an ancient protein. *Nature* **549**, 409–413 (2017).
89. Nikoomanzar, A., Vallejo, D. & Chaput, J. C. Elucidating the Determinants of Polymerase Specificity by Microfluidic-Based Deep Mutational Scanning. *ACS Synth. Biol.* **8**, 1421–1429 (2019).
90. Jalal, A. S. B. *et al.* Diversification of DNA-Binding Specificity by Permissive and Specificity-Switching Mutations in the ParB/Noc Protein Family. *Cell Rep.* **32**, 107928 (2020).
91. Bloom, J. D., Gong, L. I. & Baltimore, D. Permissive secondary mutations enable the evolution of influenza oseltamivir resistance. *Science* **328**, 1272–1275 (2010).
92. Gong, L. I., Suchard, M. A. & Bloom, J. D. Stability-mediated epistasis constrains the evolution of an influenza protein. *Elife* **2**, e00631 (2013).
93. McKeown, A. N. *et al.* Evolution of DNA specificity in a transcription factor family produced a new gene regulatory module. *Cell* **159**, 58–68 (2014).
94. Wang, X., Minasov, G. & Shoichet, B. K. Evolution of an antibiotic resistance enzyme constrained by stability and activity trade-offs. *J. Mol. Biol.* **320**, 85–95 (2002).
95. Aakre, C. D. *et al.* Evolving new protein-protein interaction specificity through promiscuous intermediates. *Cell* **163**, 594–606 (2015).

96. McLaughlin, R. N., Jr, Poelwijk, F. J., Raman, A., Gosal, W. S. & Ranganathan, R. The spatial architecture of protein function and adaptation. *Nature* **491**, 138–142 (2012).

97. Kowalsky, C. A. & Whitehead, T. A. Determination of binding affinity upon mutation for type I dockerin-cohesin complexes from Clostridium thermocellum and Clostridium cellulolyticum using deep sequencing. *Proteins* **84**, 1914–1928 (2016).

98. van Rosmalen, M. *et al.* Affinity Maturation of a Cyclic Peptide Handle for Therapeutic Antibodies Using Deep Mutational Scanning. *J. Biol. Chem.* **292**, 1477–1489 (2017).

99. Spencer, J. M. & Zhang, X. Deep mutational scanning of S. pyogenes Cas9 reveals important functional domains. *Sci. Rep.* **7**, 16836 (2017).

100. Gold, M. G. *et al.* Engineering A-kinase anchoring protein (AKAP)-selective regulatory subunits of protein kinase A (PKA) through structure-based phage selection. *J. Biol. Chem.* **288**, 17111–17121 (2013).

101. Dutta, S. *et al.* Determinants of BH3 binding specificity for Mcl-1 versus Bcl-xL. *J. Mol. Biol.* **398**, 747–762 (2010).

102. Thompson, S., Zhang, Y., Ingle, C., Reynolds, K. A. & Kortemme, T. Altered expression of a quality control protease in E. coli reshapes the in vivo mutational landscape of a model enzyme. *Elife* **9**, (2020).

103. Kretz, C. A. *et al.* Massively parallel enzyme kinetics reveals the substrate recognition landscape of the metalloprotease ADAMTS13. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 9328–9333 (2015).

104. Ethan Ahler, Ames C Register, Sujata Chakraborty, Linglan Fang, Emily M Dieter, Katherine A Sitko, Rama Subba Rao Vidadala, Bridget M Trevillian, Martin Golkowski, Hannah Gelman, Jason J Stephany, Alan F Rubin, Ethan A Merritt, Douglas M Fowler, Dustin J Maly. A Combined Approach Reveals a Regulatory Mechanism Coupling Src's Kinase Activity, Localization, and Phosphotransferase-Independent Functions. *Molecular Cell* 393–408 (2019).

105. Greaney, A. J. *et al.* Complete Mapping of Mutations to the SARS-CoV-2 Spike Receptor-Binding Domain that Escape Antibody Recognition. *Cell Host Microbe* **29**, 44–57.e9 (2021).

106. Lara Ortiz, M. T., Martinell García, V. & Del Rio, G. Saturation Mutagenesis of the Transmembrane Region of HokC in Escherichia coli Reveals Its High Tolerance to Mutations. *Int. J. Mol. Sci.* **22**,

(2021).

107. Cote-Hammarlof, P. A. *et al.* The Adaptive Potential of the Middle Domain of Yeast Hsp90. *Mol. Biol. Evol.* **38**, 368–379 (2021).

108. Mighell, T. L., Evans-Dutson, S. & O'Roak, B. J. A Saturation Mutagenesis Approach to Understanding PTEN Lipid Phosphatase Activity and Genotype-Phenotype Relationships. *Am. J. Hum. Genet.* **102**, 943–955 (2018).

109. Heinrich, R. & Rapoport, T. A. A linear steady-state treatment of enzymatic chains. General properties, control and effector strength. *Eur. J. Biochem.* **42**, 89–95 (1974).

110. Kacser, H. & Burns, J. A. The control of flux. *Biochem. Soc. Trans.* **23**, 341–366 (1995).

111. Jiang, L., Mishra, P., Hietpas, R. T., Zeldovich, K. B. & Bolon, D. N. A. Latent effects of Hsp90 mutants revealed at reduced expression levels. *PLoS Genet.* **9**, e1003600 (2013).

112. Kowalsky, C. A. *et al.* High-resolution sequence-function mapping of full-length proteins. *PLoS One* **10**, e0118193 (2015).

113. Roychowdhury, H. & Romero, P. A. Microfluidic deep mutational scanning of the human executioner caspases reveals differences in structure and regulation. *Cell Death Discov* **8**, 7 (2022).

114. Nikoomanzar, A., Vallejo, D., Yik, E. J. & Chaput, J. C. Programmed Allelic Mutagenesis of a DNA Polymerase with Single Amino Acid Resolution. *ACS Synth. Biol.* **9**, 1873–1881 (2020).

115. Romero, P. A., Tran, T. M. & Abate, A. R. Dissecting enzyme function with microfluidic-based deep mutational scanning. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 7159–7164 (2015).

116. Subramanian, S. *et al.* Allosteric communication in DNA polymerase clamp loaders relies on a critical hydrogen-bonded junction. *Elife* **10**, (2021).

117. Flynn, J. M. *et al.* Comprehensive fitness landscape of SARS-CoV-2 Mpro reveals insights into viral resistance mechanisms. *Elife* **11**, (2022).

118. Katrekar, D. *et al.* Comprehensive interrogation of the ADAR2 deaminase domain for engineering enhanced RNA editing activity and specificity. *Elife* **11**, (2022).

119. Biebl, M. M. & Buchner, J. Structure, Function, and Regulation of the Hsp90 Machinery. *Cold Spring Harb. Perspect. Biol.* **11**, (2019).

120. Jaeger, A. M. & Whitesell, L. HSP90: Enabler of Cancer Adaptation. *Annu. Rev. Cancer Biol.* **3**,

275–297 (2019).

121. Bohush, A., Bieganowski, P. & Filipek, A. Hsp90 and Its Co-Chaperones in Neurodegenerative Diseases. *Int. J. Mol. Sci.* **20**, (2019).

122. Boczek, E. E. *et al.* Conformational processing of oncogenic v-Src kinase by the molecular chaperone Hsp90. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E3189–98 (2015).

123. Mandal, A. K. *et al.* Cdc37 has distinct roles in protein kinase quality control that protect nascent chains from degradation and promote posttranslational maturation. *J. Cell Biol.* **176**, 319–328 (2007).

124. Anderson, S. K., Gibbs, C. P., Tanaka, A., Kung, H. J. & Fujita, D. J. Human cellular src gene: nucleotide sequence and derived amino acid sequence of the region coding for the carboxy-terminal two-thirds of pp60c-src. *Mol. Cell. Biol.* **5**, 1122–1129 (1985).

125. Taipale, M. *et al.* Chaperones as thermodynamic sensors of drug-target interactions reveal kinase inhibitor specificities in living cells. *Nat. Biotechnol.* **31**, 630–637 (2013).

126. Luo, Q., Boczek, E. E., Wang, Q., Buchner, J. & Kaila, V. R. I. Hsp90 dependence of a kinase is determined by its conformational landscape. *Sci. Rep.* **7**, 43996 (2017).

127. Parsons, S. J. & Parsons, J. T. Src family kinases, key regulators of signal transduction. *Oncogene* **23**, 7906–7909 (2004).

128. Liu, W. *et al.* The proto-oncogene c-Src and its downstream signaling pathways are inhibited by the metastasis suppressor, NDRG1. *Oncotarget* **6**, 8851–8874 (2015).

129. Wheeler, D. L., Iida, M. & Dunn, E. F. The role of Src in solid tumors. *Oncologist* **14**, 667–678 (2009).

130. Manning, G., Whyte, D. B., Martinez, R., Hunter, T. & Sudarsanam, S. The protein kinase complement of the human genome. *Science* **298**, 1912–1934 (2002).

131. Williams, J. C. *et al.* The 2.35 A crystal structure of the inactivated form of chicken Src: a dynamic molecule with multiple regulatory interactions. *J. Mol. Biol.* **274**, 757–775 (1997).

132. Xu, W., Harrison, S. C. & Eck, M. J. Three-dimensional structure of the tyrosine kinase c-Src. *Nature* **385**, 595–602 (1997).

133. Xu, W., Doshi, A., Lei, M., Eck, M. J. & Harrison, S. C. Crystal structures of c-Src reveal features of its autoinhibitory mechanism. *Mol. Cell* **3**, 629–638 (1999).

134. Nada, S., Okada, M., MacAuley, A., Cooper, J. A. & Nakagawa, H. Cloning of a complementary DNA

for a protein-tyrosine kinase that specifically phosphorylates a negative regulatory site of p60c-src. *Nature* **351**, 69–72 (1991).

135. Bjorge, J. D., Jakymiw, A. & Fujita, D. J. Selected glimpses into the activation and function of Src kinase. *Oncogene* **19**, 5620–5635 (2000).
136. Schlessinger, J. SH2/SH3 signaling proteins. *Curr. Opin. Genet. Dev.* **4**, 25–30 (1994).
137. Dehm, S. M. & Bonham, K. SRC gene expression in human cancer: the role of transcriptional activation. *Biochem. Cell Biol.* **82**, 263–274 (2004).
138. Brugge, J. S. *et al.* Expression of Rous sarcoma virus transforming protein pp60v-src in Saccharomyces cerevisiae cells. *Mol. Cell. Biol.* **7**, 2180–2187 (1987).
139. Kritzer, J. A., Freyzon, Y. & Lindquist, S. Yeast can accommodate phosphotyrosine: v-Src toxicity in yeast arises from a single disrupted pathway. *FEMS Yeast Res.* **18**, (2018).
140. Xu, Y. & Lindquist, S. Heat-shock protein hsp90 governs the activity of pp60v-src kinase. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 7074–7078 (1993).
141. Chen, B., Zhong, D. & Monteiro, A. Comparative genomics and evolution of the HSP90 family of genes across all kingdoms of organisms. *BMC Genomics* **7**, 156 (2006).
142. Xu, Y., Singer, M. A. & Lindquist, S. Maturation of the tyrosine kinase c-src as a kinase and as a substrate depends on the molecular chaperone Hsp90. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 109–114 (1999).
143. Rubin, A. F. *et al.* A statistical framework for analyzing deep mutational scanning data. *Genome Biol.* **18**, 150 (2017).
144. Citri, A. *et al.* Hsp90 recognizes a common surface on client kinases. *J. Biol. Chem.* **281**, 14361–14369 (2006).
145. Xu, W. *et al.* Surface charge and hydrophobicity determine ErbB2 binding to the Hsp90 chaperone complex. *Nat. Struct. Mol. Biol.* **12**, 120–126 (2005).
146. Azam, M., Seeliger, M. A., Gray, N. S., Kuriyan, J. & Daley, G. Q. Activation of tyrosine kinases by mutation of the gatekeeper threonine. *Nat. Struct. Mol. Biol.* **15**, 1109–1118 (2008).
147. Peng, C., Li, D. & Li, S. Heat shock protein 90: a potential therapeutic target in leukemic progenitor and stem cells harboring mutant BCR-ABL resistant to kinase inhibitors. *Cell Cycle* **6**, 2227–2231

(2007).

148. Peng, C. *et al.* Inhibition of heat shock protein 90 prolongs survival of mice with BCR-ABL-T315I-induced leukemia and suppresses leukemic stem cells. *Blood* **110**, 678–685 (2007).
149. Koga, F. *et al.* Hsp90 inhibition transiently activates Src kinase and promotes Src-dependent Akt and Erk activation. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 11318–11322 (2006).
150. Chakraborty, S. *et al.* Profiling of the drug resistance of thousands of Src tyrosine kinase mutants uncovers a regulatory network that couples autoinhibition to the dynamics of the catalytic domain. *bioRxiv* 2021.12.05.471322 (2022) doi:10.1101/2021.12.05.471322.
151. Brown, M. T. & Cooper, J. A. Regulation, substrates and functions of src. *Biochim. Biophys. Acta* **1287**, 121–149 (1996).
152. Cowan-Jacob, S. W. *et al.* The crystal structure of a c-Src complex in an active conformation suggests possible steps in c-Src activation. *Structure* **13**, 861–871 (2005).
153. Young, M. A., Gonfloni, S., Superti-Furga, G., Roux, B. & Kuriyan, J. Dynamic coupling between the SH2 and SH3 domains of c-Src and Hck underlies their inactivation by C-terminal tyrosine phosphorylation. *Cell* **105**, 115–126 (2001).
154. Citri, A., Kochupurakkal, B. S. & Yarden, Y. The achilles heel of ErbB-2/HER2: regulation by the Hsp90 chaperone machine and potential for pharmacological intervention. *Cell Cycle* **3**, 51–60 (2004).
155. Park, H. *et al.* Simultaneous Optimization of Biomolecular Energy Functions on Features from Small Molecules and Macromolecules. *J. Chem. Theory Comput.* **12**, 6201–6212 (2016).
156. Barouch-Bentov, R. & Sauer, K. Mechanisms of drug resistance in kinases. *Expert Opin. Investig. Drugs* **20**, 153–208 (2011).
157. Kornev, A. P., Haste, N. M., Taylor, S. S. & Eyck, L. F. T. Surface comparison of active and inactive protein kinases identifies a conserved activation mechanism. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 17783–17788 (2006).
158. Karagöz, G. E. & Rüdiger, S. G. D. Hsp90 interaction with clients. *Trends Biochem. Sci.* **40**, 117–125 (2015).
159. Karagöz, G. E. *et al.* Hsp90-Tau complex reveals molecular basis for specificity in chaperone action.

Cell **156**, 963–974 (2014).

160. Wegele, H., Müller, L. & Buchner, J. Hsp70 and Hsp90--a relay team for protein folding. *Rev. Physiol. Biochem. Pharmacol.* **151**, 1–44 (2004).

161. Fang, L. *et al.* How ATP-Competitive Inhibitors Allosterically Modulate Tyrosine Kinases That Contain a Src-like Regulatory Architecture. *ACS Chem. Biol.* **15**, 2005–2016 (2020).

162. Agius, M. P. *et al.* Selective Proteolysis to Study the Global Conformation and Regulatory Mechanisms of c-Src Kinase. *ACS Chem. Biol.* **14**, 1556–1563 (2019).

163. Chang, H. C. & Lindquist, S. Conservation of Hsp90 macromolecular complexes in Saccharomyces cerevisiae. *J. Biol. Chem.* **269**, 24983–24988 (1994).

164. García-Nafría, J., Watson, J. F. & Greger, I. H. IVA cloning: A single-tube universal cloning system exploiting bacterial In Vivo Assembly. *Sci. Rep.* **6**, 27459 (2016).

165. Suzuki, Y. *et al.* Knocking out multigene redundancies via cycles of sexual assortment and fluorescence selection. *Nat. Methods* **8**, 159–164 (2011).

166. Hall, B. G., Acar, H., Nandipati, A. & Barlow, M. Growth rates made easy. *Mol. Biol. Evol.* **31**, 232–238 (2014).

167. Olson, C. A., Wu, N. C. & Sun, R. A comprehensive biophysical description of pairwise epistasis throughout an entire protein domain. *Curr. Biol.* **24**, 2643–2651 (2014).

168. Faure, A. J. *et al.* Mapping the energetic and allosteric landscapes of protein binding domains. *Nature* **604**, 175–183 (2022).

169. Nguyen, V. *et al.* Molecular determinants of Hsp90 dependence of Src kinase revealed by deep mutational scanning. *Protein Sci.* e4656 (2023).

170. Martellucci, S. *et al.* Src Family Kinases as Therapeutic Targets in Advanced Solid Tumors: What We Have Learned so Far. *Cancers* **12**, (2020).