

UCSF

UC San Francisco Electronic Theses and Dissertations

Title

Development of chemical tools targeting DEAD-box proteins

Permalink

<https://escholarship.org/uc/item/1bq6m023>

Author

Barkovich, Krister Jernstedt

Publication Date

2018

Peer reviewed|Thesis/dissertation

Development of chemical tools targeting DEAD-box proteins

by

Krister Jernstedt Barkovich

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Chemistry and Chemical Biology

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Copyright 2018

by

Krister Jernstedt Barkovich

Acknowledgements

I would first like to thank my primary mentor, Dr. Kevan Shokat. Thank you for your guidance over five years as a PhD student in your lab and three years before that as an undergraduate researcher. You bring an unrelenting optimism and energy to research that has kept me working through low times, and gave me intellectual freedom to pursue projects that I found exciting. Thank you for never doubting my capabilities (or my projects) and for your constant encouragement. I hope to match your mentoring style as I continue into my own career in academic research. To the rest of the Shokat lab, both past and present, thank you for your daily support and for making the lab a stimulating and supportive environment. It has been a joy working alongside all of you. I would additionally like to thank the support of Delaney Lynch, who has worked tirelessly as administrator of the Shokat lab and is a major reason for the smooth day-to-day workings of the lab.

I would also like to thank Dr. Jack Taunton and Dr. Davide Ruggero, the other members of my thesis committee. Your guidance and scientific expertise has been invaluable for the progression and success of my thesis projects and it has been a pleasure getting to know both of you as scientific and personal mentors. Thank you for the time and energy you have put into my scientific growth. An additional thanks to the administrative staff of CCB, Julia Molla and Nicole Flowers, and the MSTP, Geri Ehle, Jana Toutolmin, and Catherine Norton.

I would also like to thank Dr. William Weiss and Dr. Stephen Floor, two of my other scientific mentors. Bill – thank you for providing a space for me to learn bench science as an undergraduate researcher. Your lab was the place where I found my passion for scientific research and your guidance has been extremely influential on my career path. Stephen – thank you for all your help as I've navigated the steep learning curve of RNA biochemistry.

Thank you to my amazing friends and family who have been my support throughout graduate school. To my friends, especially my classmates/colleagues in the MSTP – thank you for your unwavering support. This journey would not have been the same without you. Thank you to my brothers, Matthew and Emil, who inspire and drive me to be my best self. I hope we continue to advance scientifically and professionally but never lose the fun times we have together. To my partner, Misty Montoya – thank you for being my support system and for all that you do for me. You inspire me to become the best version of myself and I am so lucky to have you by my side. Lastly, to my mother and father, Karen Jernstedt and Jim Barkovich, who are responsible for the person I am today – thank you for raising me to value hard work and perseverance over material gain. Any success I have is a direct result of your effort and encouragement.

Contributions of Co-Authors to Presented Work

Chapter 2 of this dissertation was published as “Analog-sensitive chemical inhibition of the DEAD-box protein DDX3” in *Protein Science* (2016) 25(3), 638-49. The co-authors of this paper were Stephen N. Floor,^{1,2} Kendall J. Condon,¹ Kevan M. Shokat,³⁻⁵ and Jennifer A. Doudna.^{1,2,5-7} S.N.F., K.J.B. and K.J.C. performed experiments. K.J.B. performed chemical synthesis. S.N.F., K.J.B., K.M.S., and J.A.D. designed and analyzed experiments. S.N.F., K.J.B., K.M.S., and J.A.D. wrote the manuscript. S.N.F., J.A.D. and K.M.S. provided funding. S.N.F. and K.J.B. are co-first authors.

Chapter 3 of this dissertation is based on a manuscript in preparation entitled “Chemical genetic inhibition of DEAD-box proteins using covalent complementarity.” Contributors to this work include Megan K. Moore, Qi Hu, and Kevan M. Shokat of UCSF. K.J.B. and M.K.M. performed experiments. K.J.B. performed chemical synthesis. K.J.B. and Q.H. analyzed crystallography data. K.J.B. and K.M.S. designed and analyzed experiments. K.J.B. and K.M.S. wrote the manuscript. All authors edited the manuscript.

Chapter 4 of this dissertation is based on ongoing work and includes contributions from Tommaso Cupido and Tarun Kapoor of Rockefeller University and Kevan M. Shokat.

Chapter 5 of this dissertation is based on ongoing work and includes contributions from Qi Wen Fan, Junjiao Yang, Kevin Lou, Jack Stevenson, Nicole Wenzell, Xiaokun Shu, William A. Weiss, and Kevan M. Shokat, all of UCSF.

¹Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, California, USA.

²Howard Hughes Medical Institute, University of California, Berkeley, Berkeley, California, USA.

³Department of Cellular and Molecular Pharmacology, University of California, San Francisco, San Francisco, California, USA.

⁴Howard Hughes Medical Institute, University of California, San Francisco, San Francisco, California, USA.

⁵Department of Chemistry, University of California, Berkeley, Berkeley, California, USA.

⁶Innovative Genomics Initiative, University of California, Berkeley, Berkeley, California, USA.

⁷Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California, USA.

Development of chemical tools to target DEAD-box proteins

Krister Jernstedt Barkovich

Abstract

The complexity of the three-dimensional structures formed by RNA is essential for its function and as a result, a large number of protein co-factors are required to maintain RNA homeostasis. The largest family of enzymatic RNA chaperones are the DEAD-box proteins, which utilize ATP hydrolysis to modify RNA substrates. DEAD-box proteins are required for all stages of RNA metabolism and are implicated in diseases including cancer, viral pathogenesis, and developmental delay. Yet despite the biochemical and structural characterization of these enzymes over the past three decades, the specific functions and substrates of DEAD-box proteins remains poorly understood. Chemical inhibition would be an excellent tool for the elucidation of DEAD-box protein biology because of its rapid onset of inhibition, however specific small molecule inhibitors do not exist for most DEAD-box proteins.

To develop a generalizable strategy for the inhibition of DEAD-box proteins, we developed two chemical genetic strategies – one analog-sensitive and one electrophile-sensitive – that rely on genetic perturbations of the target of interest coupled with specific chemical inhibitors. Our efforts to develop an analog-sensitive strategy yielded novel space-creating mutations off the N6-position of the adenine of ATP that were poorly tolerated *in vivo*, and analog-sensitive inhibitors that could not be optimized to bind with high affinities. For the electrophile-sensitive strategy, we identified a residue of

low conservation within the DEAD-box protein active site for mutation to cysteine and specifically targeted it with electrophile-containing AMP-derivatives. These small molecules specifically bind to and inhibit electrophile-sensitive, but not wild-type, DEAD-box proteins. Together, these results indicate that DEAD-box proteins can be targeted using chemical genetic tools and provide a path towards the development of specific small molecule inhibitors of any member of the DEAD-box protein family.

We additionally sought to identify novel nucleotide-competitive cell-permeable small molecules that bind and inhibit DEAD-box proteins. To this end, we identified a series of di-substituted quinazolines based on the AAA+ ATPase inhibitor DBeQ that dose-dependently inhibit DDX3 in an ATP-dependent manner.

In a separate project, we explored the limited efficacy of clinical CDK4/6 inhibitors as monotherapy. We identified that the small molecule inhibitor ribociclib has limited efficacy as a monotherapy in cell culture, and found that CDK4 is bound in high molecular weight complexes in cells, which are mostly resistant to inhibition by type I kinase inhibitors. We then developed a series of small molecule derivatives of ribociclib that attempt to expand the clinical efficacy of ribociclib by targeting these high molecular weight complexes.

Table of Contents

Chapter 1: Introduction	1
RNA and RNA homeostasis.....	2
DEAD-box proteins.....	3
DEAD-box Proteins in Disease.....	5
Chemical Inhibitors of DEAD-box Proteins.....	10
Thesis Work Overview.....	14
Figures.....	17
Chapter 2: Analog-sensitive chemical inhibition of the DEAD-box protein DDX3	21
Abstract.....	22
Introduction.....	22
Results.....	25
Discussion.....	32
Materials and Methods.....	33
Acknowledgements.....	36
Figures.....	37
Chemical Synthesis.....	44
Chapter 3: Chemical genetic inhibition of DEAD-box proteins using covalent complementarity	49
Abstract.....	50

Introduction.....	50
Results.....	55
Discussion.....	62
Materials and Methods.....	66
Figures.....	70
Supplementary Figures.....	75
Supplementary Tables.....	79
Chemical Synthesis.....	82
Chapter 4: Developing reversible chemical inhibitors of DDX3.....	88
Introduction.....	89
Results.....	91
Discussion.....	95
Materials and Methods.....	97
Figures.....	99
Chapter 5: Analyzing CDK4 complexes to expand the efficacy of clinical CDK4/6 inhibitors.....	104
Introduction.....	105
Results.....	107
Discussion.....	113
Materials and Methods.....	117
Figures.....	119
Chemical Synthesis.....	124

References.....160

List of Tables

Chapter 3

Supplementary Table 3.1: Data collection and refinement statistics.....	79
Supplementary Table 3.2: Yeast strains used in this study.....	80
Supplementary Table 3.3: Plasmids used in this study.....	81

List of Figures

Chapter 1

- Figure 1.1: DEAD-box proteins function in all stages of RNA metabolism.....17
- Figure 1.2: Human DEAD-box proteins function in RNA metabolism and disease...18
- Figure 1.3: Literature DEAD-box protein inhibitors.....19

Chapter 2

- Figure 2.1: Targeting a hydrophobic cluster adjacent to the ATP binding pocket of DDX3 for mutation.....37
- Figure 2.2: Hydrophobic cluster mutants of DDX3 support function *in vitro* and *in vivo*.....38
- Figure 2.3: Adenosine monophosphate, but not putative DDX3 inhibitors, retards duplex unwinding *in vitro*.....39
- Figure 2.4: Analog sensitive inhibition of DDX3.....40
- Figure 2.5: Expanded active site alleles of DDX3 and other temperature sensitive yeast strains are inhibited by GXJ1-76.....41
- Figure 2.6: Screen for analog-sensitive DEAD-box helicase inhibitors identifies anilinoquinazoline scaffold.....42
- Figure 2.7: Chemical derivatization of AQZ01 yields compound 1.....43

Chapter 3

- Figure 3.1: Development of an 'electrophile-sensitive' mutation in DEAD-box proteins.....70
- Figure 3.2: Electrophile-sensitive DEAD-box proteins retain biochemical and cellular function.....71
- Figure 3.3: AMP-acrylate and derivatives potently and irreversibly bind electrophile-sensitive DEAD-box proteins.....72

Figure 3.4: Crystal structure of AMP-acrylamide bound to electrophile-sensitive DDX3.....	73
Figure 3.5: AMP-acrylates inhibit electrophile-sensitive DEAD-box proteins.....	74
Supplementary Figure 3.1: Development of an ‘electrophile-sensitive’ mutation in DEAD-box proteins	75
Supplementary Figure 3.2: AMP-acrylates undergo two-step reaction with electrophile-sensitive DEAD-box proteins.....	77
Supplementary Figure 3.3: AMP-acrylates inhibit electrophile-sensitive DEAD-box proteins.....	78

Chapter 4

Figure 4.1: Screen of DBeQ-derivatives identifies small molecule inhibitors of DDX3.....	99
Figure 4.2: Structure-activity relationship (SAR) of DBeQ-derivatives and DDX3.....	100
Figure 4.3: Structure-activity relationship (SAR) of TC327/391-derivatives and DDX3.....	101
Figure 4.4: TC347 is an ATP-competitive DDX3 WT inhibitor.....	102
Figure 4.5: Moving towards more drug-like irreversible inhibitors of DDX3 ^{ES}	103

Chapter 5

Figure 5.1: Clinical CDK4/6 inhibitors show limited efficacy as monotherapy.....	119
Figure 5.2: CDK4 is partially labeled by XO44 in MCF7, but not BT549 cells.....	120
Figure 5.3: CDK4 complexes vary across cell lines and influence small molecule accessibility.....	121
Figure 5.4: Knockdown of cyclin D1 improves efficacy of ribociclib in cell culture.....	122

Figure 5.5: Ribociclib analogs disrupt CDK4-cyclin D1 interaction and promote cyclin D1 degradation in cells.....123

Chapter 1

Introduction

RNA and RNA homeostasis

After the discovery of the double helical structure of DNA in 1953,¹ the next question became how information from this linear genetic code was converted into the cellular machinery of proteins.² This led to the discovery of a transient RNA population, now called messenger RNA (mRNA), by two independent groups in 1961 using pulse-labelling of radiolabeled uridine.^{3,4} Similar experiments also identified a transient population of large RNAs that were synthesized in the nucleus then exported to the cytoplasm, where they existed in a long-lived fraction that showed rapid sedimentation in sucrose gradient centrifugation.^{5,6} These RNA populations were found to be the precursor to ribosomal RNA (pre-rRNA) and mature rRNAs. Monitoring of the size of the pre-rRNA as it matured into final rRNAs of the ribosome led to the discovery that ribosomal RNA undergoes multiple processing steps prior to 40S and 60S ribosomal subunit export to the cytoplasm.⁷ Later experimental work determined that mRNAs also undergo nuclear processing or “splicing” at a specialized ribonucleoprotein complex called the spliceosome,⁸⁻¹⁰ where non-coding regions of RNA (‘introns’) are removed and the mature mRNA is produced. More than a half-century later, RNA species are known to be intimately involved in numerous cellular processes, including gene expression and regulation and protein translation.²

RNA is unique as a biologic macromolecule due to its simple linear sequence and its capacity to form complex secondary and tertiary structures that present surfaces for interactions with other macromolecules including DNA and protein.² The complexity of the three-dimensional structures formed by RNA is essential for its role in translation,

splicing, scaffolding, and localization.¹¹ For example, transfer RNAs (tRNAs) are specifically modified through methylation and pseudouridylation to maintain a specific, function conformation,^{12,13} and disruption of ribosomal RNA or small nuclear RNA structure through alterations in RNA processing and folding are detrimental to its function as the structural element of the ribosome and spliceosome, respectively.^{14,15} The large number of possible RNA base-pairings and alternative folding patterns means that the maintenance of proper RNA folding presents a significant cellular challenge.¹⁶ As such, RNA homeostasis requires a sizable number of protein co-factors to maintain proper RNA folding and function. These RNA chaperones include heterogeneous nuclear ribonucleoproteins (hnRNPs) and cold-shock domain proteins, which can bind, stabilize and/or modify RNA in an ATP-independent manner, and the ATP-dependent DExD/H-box proteins.¹⁶

DEAD-box proteins

The largest family of enzymatic RNA chaperones in humans are the DEAD-box proteins, which utilize ATP to disrupt RNA-RNA and RNA-protein interactions.¹⁷ Named for their conserved Walker B motif consisting of adjacent aspartate-glutamate-alanine-aspartate (D-E-A-D) residues, DEAD-box proteins are required for all stages of RNA metabolism including transcription, processing and splicing, export, translation and decay (Figure 1.1).¹⁸ For example, eighteen DEAD-box proteins have been implicated in ribosome biogenesis, six in translation, and six in splicing.¹⁸

Prokaryotic and eukaryotic helicases are broadly categorized into six superfamilies (SF1-SF6) of which SF3-SF6 are hexameric or toroidal, meaning they function in hexameric or higher order structures, while superfamilies one and two (SF1 and SF2) are not, but remain structurally related.¹⁹ The SF2 superfamily encompasses the largest number and widest variety of helicases in terms of structure and function and includes the RecQ-like and Swi2/Snf2-like family of DNA helicases/translocases, which function in genome stability and nucleosome remodeling,^{20,21} and most RNA helicases, including the DEAD-box and DExH-box protein families.

The minimal DEAD-box protein is formed by two RecA-like domains separated by a flexible linker.²² The domain interface forms the ATP-binding site, which consists of canonical ATPase motifs first described by Walker and colleagues,^{23,24} as well as the DEAD-box protein-specific Q-motif that functions to stabilize the adenine of ATP and make this family of enzymes ATP-specific.²⁵ Unlike the structurally related DExH-box family of RNA helicases, DEAD-box proteins remodel their substrates non-processively through local-strand separation²⁶ and require only a single ATP hydrolysis event for substrate release.^{27,28} Additionally, DEAD-box proteins display biochemical activity against both RNA-RNA and RNA-protein complexes,²⁹ and can act as both strand annealers and unwinders.³⁰ The DEAD-box proteins are best exemplified by eIF4A, the first human DEAD-box protein discovered, that acts to unwind secondary structure in the 5'-UTR of mRNAs during 43S preinitiation complex scanning.³¹ Other notable human DEAD-box proteins include DDX3, which is implicated to act as a part of the eIF4F complex in 43S scanning and translational activation,^{32,33} and DDX5 (p68) and

DDX17 (p72) which act as transcriptional co-activators.³⁴ Yet despite these well-studied examples, our understanding of the specific functions that DEAD-box proteins perform *in vivo*, as well as detailed understanding of substrate specificity, remain poorly understood.¹⁸

DEAD-box Proteins in Disease

Owing to their influential role in nearly all aspects of biology, DEAD-box proteins are commonly implicated in disease, including viral pathogenesis, cancer, and disorders of impaired development (Figure 1.2).

Viral infection

Owing to their limited genomic size, viruses often co-opt human cellular proteins that are required for their replication process. RNA viruses often require these cellular components, which include many DEAD-box proteins, for replication and translation of their genome.³⁵

DDX1 is implicated as a pro-viral factor for human coronavirus and HIV infection. The unwinding activity of the DEAD-box protein DDX1 is required for coronavirus subgenomic mRNA (sgmRNA) synthesis and DDX1 knockdown or loss of catalytic activity reduced levels of sgmRNAs.³⁶ DDX1 also functions with the HIV Rev Response Element (RRE) to enhance the RNA export activity of Rev³⁷ and DDX1 silencing in HIV-infected human cells reduces virus particle production.³⁸

DDX3 is also a target of many viruses including hepatitis C virus (HCV), HIV-1, and West Nile virus (WNV). DDX3 is bound by the HCV core protein and is required for HCV RNA replication.³⁹ DDX3 knockdown results in significant repression of genome-length HCV RNA and suppression of HCV infection.⁴⁰ In HIV infection, DDX3 co-localizes with HIV-1 Tat to cytoplasmic foci and is required for Tat function⁴¹ as well as Rev-RRE export, similar to DDX1.⁴² DDX3 is also required for translation initiation of the HIV-1 genomic RNA.⁴³ DDX3 knockdown by shRNA suppresses the nuclear export of HIV-1 RNAs and translation of the HIV-1 genome, and reduces HIV-1 viral replication.^{43,44} DDX3 has also been identified at WNV replication sites along with other human P-body components LSM1, GW182 and XRN1 and knockdown of these proteins resulted in reduced viral RNA levels and infection.⁴⁵ Interestingly, in hepatitis B virus (HBV) and vaccinia virus (VACV) infection, DDX3 contributes to antiviral innate immune signaling pathways leading to interferon-beta production through an interaction with IRF3⁴⁶ and DDX3 knockdown inhibits IRF3 activation.⁴⁷

DDX5 is also implicated in the pathogenesis of several viral infections. A yeast-two hybrid screen found that DDX5 specifically interacts with the sudden acute respiratory syndrome coronavirus (SARS-CoV) helicase nsp13 and knockdown of DDX5 resulted in reduced coronavirus replication in cell culture.⁴⁸ DDX5 was additionally shown to interact with the HCV NS5B protein and siRNA-mediated knockdown of DDX5 resulted in reduced transcription of negative-strand RNA from positive-strand HCV RNA.⁴⁹ Similar to DDX1 and DDX3, DDX5 was also found to interact with HIV-1 Rev to enhance HIV-1 replication,⁵⁰ although DDX5 knockdown paradoxically led to increased

HIV infectivity and higher viral RNA levels.⁵¹ The closely related subfamily member DDX17 was also identified as an HIV-1 Rev interacting protein, and knockdown of DDX17 led to decreased viral mRNA levels.⁵¹ DDX17 was found to recognize and bind specific cytoplasmic RNA stem-loops in the Rift Valley fever virus (RVFV) genome and thus can also serve an anti-viral role.⁵²

The DEAD-box protein DDX6 also plays a role in HIV and HCV infection. HIV-1 mRNA was found to associate with components of the miRNA machinery (RNA Induced Silencing Complex, RISC) including DDX6/RCK, GW182, LSM1 and XRN1, and knockdown of DDX6 or DGCR8, which is required for miRNA processing, resulted in increased production of infectious virions, indicating the miRNA processing machinery plays an antiviral role in HIV infection.⁵³ Conversely, DDX6 activity is important for optimal HCV replication, as DDX6 overexpression enhanced HCV replication and an ATPase-dead mutant had a dominant-negative effect, reducing HCV viral yields.⁵⁴ The DDX6-interacting proteins LSM1-7 were also found to specifically interact with cis-acting HCV RNA elements in the viral RNA UTRs, and depletion of these cellular components reduced HCV replication.⁵⁵

In a study of HIV-1 Rev interacting partners, the DEAD-box proteins DDX21, DDX24, and DDX47 were also identified, as were DExH-box proteins DHX9 and DHX36.⁵¹ Later studies confirmed these results and found that DDX21 and DDX24 are pro-viral factors in HIV infection as DDX21 stimulates Rev activity⁵⁶ and DDX24 is required for efficient packaging of HIV-1 RNA.⁵⁷

Cancer

Owing to their role in essential cellular processes required for growth and differentiation, several DEAD-box proteins have been identified to play important roles in cancer. A common theme within the DEAD-box protein literature is that their function is extremely context specific. This often means that literature conclusions are seemingly at-odds because the result of gene overexpression or knockdown is highly dependent on tissue of origin and treatment.⁵⁸

The DDX1 gene was mapped to chromosome 2p24 adjacent to MYCN and the two genes are commonly co-amplified in subsets of neuroblastoma⁵⁹ and retinoblastoma,⁶⁰ although the literature is conflicted as to how DDX1 gene amplification affects patient survival.^{61,62} Some reports suggest an oncogenic role for DDX1 in cancer development⁶³ while others imply a tumor-suppressive role,⁶⁴ likely indicating its function in tumor growth and maintenance is context-specific.

DDX3X is commonly mutated in malignancy – including medulloblastoma,⁶⁵⁻⁶⁸ a diverse number of blood cancers,⁶⁹⁻⁷² and head and neck squamous cell carcinoma.^{73,74} These mutations are mostly loss-of-function suggesting a tumor-suppressive role for DDX3 in cancer.⁷⁵ However, small interfering RNA-mediated knockdown of DDX3 in several cancer cell lines increased cellular adhesion but decreased cell motility and invasive capacities *in vitro*, and reduced metastasis *in vivo*.⁷⁶ Additionally, DDX3 was found to regulate cell growth through control of cyclin E translation in HeLa cells,⁷⁷ and

overexpression of DDX3 in MCF10A cells caused a epithelial-mesenchymal-like transformation and increased colony formation,⁷⁸ suggesting an pro-growth/tumor role. Therefore, the function of DDX3 in malignancy is likely also context dependent.

DDX5 and DDX17 are overexpressed in a wide variety of cancers including breast⁷⁹ and prostate cancer,⁸⁰ where DDX5 and DDX17 function as steroid hormone receptor transcriptional co-activators,⁸¹ as well as in colon cancer^{82,83} and glioma.⁸⁴ In contrast to these oncogenic functions, DDX5 also functions as a transcriptional co-activator of the p53 tumor suppressor⁸⁵ and is required for p53-mediated induction of p21 expression after DNA damage.⁸⁶ As DDX5 and DDX17 undergo extensive post-translational modifications, it is possible that these seemingly opposing roles in malignancy are due to differential modification.⁵⁸

The DEAD-box protein eIF4A functions in translation initiation by assisting in 5'-UTR scanning by the 43S pre-initiation complex.^{87,88} It is tightly controlled through interactions with co-factors eIF4G and eIF4B and the 7me-G cap-binding protein eIF4E, which is negatively regulated by 4E-BP downstream of the master regulator of cell growth, mTOR. There is ample genetic evidence for misregulation of this pathway leading to cancer, as loss of TSC1/2, a negative regulator of mTOR, results in tuberous sclerosis,⁸⁹ overexpression of eIF4E enhances lymphomagenesis in a Myc-driven model of cancer,⁹⁰ and Myc oncogenic activity is suppressed by ribosomal protein haploinsufficiency.⁹¹ Additionally, active-site mTOR inhibitors that reduce eIF4E/eIF4A-dependent translation show significant anti-tumor activity⁹² and incomplete inhibition of

this pathway is a common mechanism of primary resistance to clinical mTOR active-site inhibitors.⁹³

Disorders of Impaired Development

Since many DEAD-box proteins are essential genes, mutations in several RNA helicases have been identified as causes of disorders of impaired development. Mutants of DDX3X have been identified as a common cause of unexplained intellectual disability in girls, including some cases of Toriello-Carey syndrome, which results in agenesis of the corpus callosum and a characteristic facial phenotype.⁹⁴⁻⁹⁶ Mutations in senataxin, an SF1 RNA helicase are a cause of ataxia oculomotor apraxia type 2 (AOA2) and a rare autosomal dominant form of juvenile amyotrophic lateral sclerosis (ALS4).^{97,98}

Although not directly causative, several DEAD-box proteins are intimately involved in other neurological disorders. DDX20/Gemin3 is an essential component of the survival of motor neuron (SMN) complex, which is required for proper assembly of Sm-class ribonucleoproteins.^{99,100} Mutation of SMN1, another component of this complex, causes spinal muscular atrophy, a disorder characterized by muscle wasting.¹⁰¹ DDX6 is a known interacting partner of ataxin-2, and altered ataxin-2 levels interfere with the assembly of stress granules and cellular P-bodies, leading to spinocerebellar ataxia type 2 (SCA2).¹⁰²

Chemical Inhibitors of DEAD-box Proteins

The first identified DEAD-box protein inhibitors were natural products discovered for their cytotoxicity (Figure 1.3A). A recent paper identified two marine natural products, elisabatin A from the Caribbean sea plume *Pseudopterogorgia elisabethae*, and allolaurinterol from the marine red algae *Laurencia obtuse*, that inhibit the DEAD-box protein eIF4A in an ATP-competitive manner.¹⁰³ Pateamine A (PatA), a natural product isolated from a marine sponge, *Mycale hentscheli*, was also found to bind eIF4A¹⁰⁴ and potently inhibit cap-dependent translation.¹⁰⁵ Interestingly, PatA stimulates eIF4A ATPase activity and RNA binding, possibly by antagonizing an inhibitory interdomain interaction,¹⁰⁶ although it is unclear based on this mechanism why PatA has inhibitory function. It has been hypothesized that PatA disrupts the formation of eIF4F (which consists of eIF4A, eIF4G, and eIF4E), leading to inhibition of translational initiation.^{104,105}

The natural product hippuristanol, a polyoxygenated steroid isolated from the coral *Isis hippuris*, was identified in a chemical-genetic screen to identify inhibitors of eukaryotic translation.¹⁰⁷ It inhibits cap-dependent translational initiation by targeting eIF4A without perturbing its ATP-binding properties.¹⁰⁸ This activity is highly selective, as activity of the closely related DEAD-box protein DDX3 was unchanged by high doses of hippuristanol. The binding site of hippuristanol was mapped to the C-terminus of eIF4A by ¹H-¹⁵N-HSQC, leading to the hypothesis that it antagonizes RNA-binding by disrupting proper interdomain interactions within eIF4A.¹⁰⁹ Hippuristanol continues to be used as a tool for the inhibition of cap-dependent translation in cell culture.

An additional natural product silvestrol, a rocaglate isolated from the dioecious shrub *Aglaia silvestris*, was first described in 2004.¹¹⁰ Silvestrol and related derivatives were later found to also inhibit cap-dependent translation by targeting eIF4A.¹¹¹ As would be suggested by previous reports on the importance of the mTOR-4EBP-eIF4E axis in cancer, silvestrol and its rocaglate derivatives display potent anti-tumor activity.¹¹² Synthetic silvestrol analogs have been synthesized to improve its pharmacological properties,^{113,114} and these derivatives show significant promise in several models of cancer.¹¹⁵⁻¹¹⁷ In contrast to pateamine A and hippuristanol, rocaglates function by trapping eIF4A on RNA substrates containing specific polypurine motifs, thereby blocking scanning by the 43S pre-initiation complex and reducing overall translation.¹¹⁸

The discovery of small molecule inhibitors of DEAD-box proteins other than eIF4A has been much less successful and mostly synthetic molecules have been identified (Figure 1.3B-C). Using molecular docking against the crystal structure of the DDX3-AMP complex, several rhodanine-based compounds were identified as ATP- and RNA-noncompetitive DDX3 inhibitors,¹¹⁹ despite molecular modeling suggesting these compounds bind the ATP-binding site. A second paper by the same research group expanded on these findings and used additional molecular modeling to develop more potent triazine inhibitors.¹²⁰ Additional studies using molecular modeling have identified kaempferol, a natural product derived from the shrub *Sophora interrupta*, and the salt of the NSAID ketorolac as inhibitors of DDX3, although these studies have little biochemical data to back their claims.^{121,122} Additionally, in our hands, none of these compounds inhibit the biochemical activity of DDX3 (Figure 2.3C).

Research groups at University of Maryland, Baltimore County identified a ring-expanded nucleoside analog (NZ51) that reduces HIV-1 replication and inhibits the biochemical activity of DDX3.¹²³ Using molecular modelling, NZ51 was fit into the ATP-binding site of DDX3, although further structural information has not been obtained.¹²⁴ These groups also synthesized a series of 5:7:5-fused diimidazodiazepine compounds (RK-33) based on NZ51 that are potently cytotoxic.^{125,126} Molecular modeling determined RK-33 also binds the ATP-binding site of DDX3, a biotinylated version of RK-33 pulled down DDX3 from lysates, and substoichiometric RK-33 inhibits RNA duplex unwinding by Ded1p.¹²⁷ Unfortunately, RK-33 does not inhibit the biochemical activity of DDX3 in our hands (Figure 2.3C). The most likely explanation for these findings is that RK-33 and related compounds are DNA/RNA intercalators, and thus inhibit RNA duplex unwinding by Ded1p by disrupting the recognition of duplex RNA by Ded1p. DDX3 is pulled out of lysates by biotinylated RK-33 because both are bound to nucleic acids, and RK-33 and derivatives are cytotoxic because they disrupt cellular DNA and RNA pools.

More recent work by Takeda Pharmaceuticals used high-throughput screening to identify small molecule inhibitors of eIF4A3, a component of the exon-junction complex (EJC)^{128,129} and Brr2, a spliceosomal RNA helicase (Figure 1.3C).¹³⁰ Indole-2-carboxylic acid hit compounds bound to eIF4A3 in an ATP-competitive manner, but showed little activity in RNA duplex unwinding assays at concentrations less than 100 μ M.¹²⁸ Novel 1,4-diacylpiperazines were additionally explored as potential leads and found to potently inhibit eIF4A3 ATPase activity in an ATP-noncompetitive manner at submicromolar

doses.¹²⁹ Interestingly, these compounds bind competitively with hippuristanol, suggesting conservation of this binding pocket across the eIF4A family and showed cellular activity in assays of nonsense-mediated decay (NMD).¹³¹ Brr2-targeting 4,6-dihydropyrido[4,3-d]pyrimidine-2,7(1*H*,3*H*)-diones also displayed ATP non-competitive binding and weakly inhibited the duplex unwinding activity of Brr2 at low micromolar doses.¹³⁰

Taken together, these results indicate that selective chemical inhibition of DEAD-box proteins is achievable through HTS of diverse chemical libraries, although the hits of these screens are generally ATP-noncompetitive. As a result, all current DEAD-box protein inhibitors that show reasonable and reproducible biochemical activity are allosteric inhibitors that targeting cryptic allosteric sites within a DEAD-box protein of interest. The selectivity of these compounds suggests that these allosteric sites are not conserved across the DEAD-box family, so these small molecules are unlikely to be good structural starting points for discovery of inhibitors for other members of the DEAD-box family.

Thesis Work Overview

Owing to the diverse diseases in which DEAD-box protein function is dysregulated, chemical inhibition of this family of enzymes may provide useful clinical treatments for disease.¹³² Additionally, small molecule inhibitors may help answer outstanding basic questions within the DEAD-box protein field. As such, we sought to develop a general strategy for the inhibition of DEAD-box proteins using small molecules.

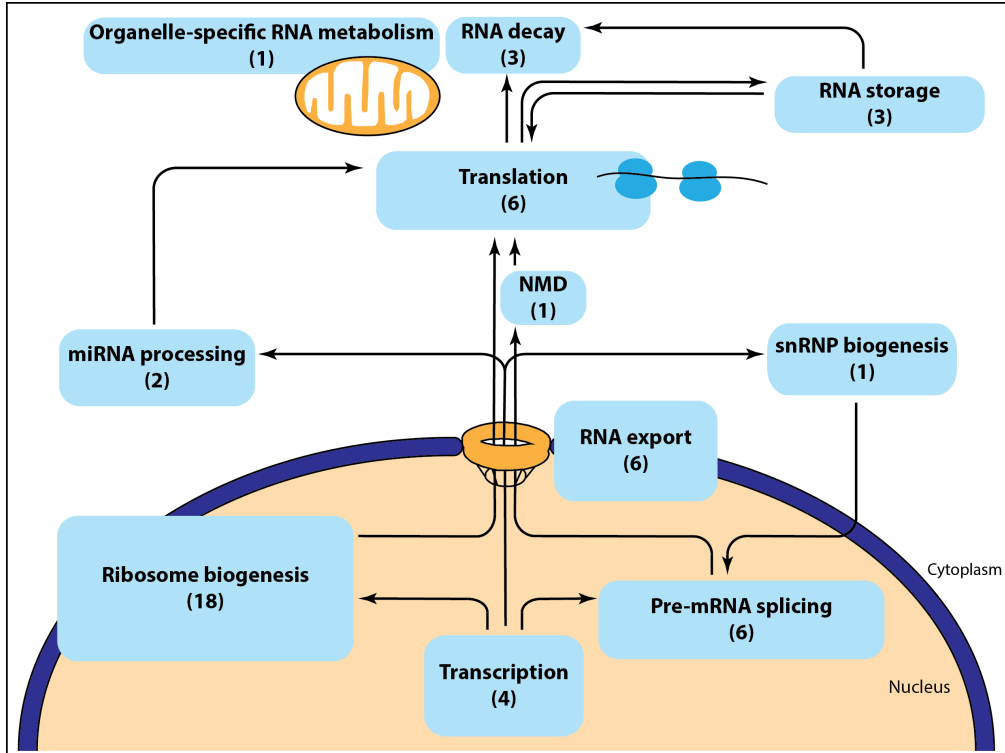
Without a chemical starting point, we first turned to an analog-sensitive chemical genetic approach.¹³³ First, we identified several conserved hydrophobic residues adjacent to the N6-position of the adenine of ATP and found that mutation of these residues to smaller amino acids was tolerated, although these enzymes showed up to a 1,000-fold reduction in activity and yeast expressing these mutants in DED1 displayed temperature-sensitivity. We then identified a series of anilinoquinazolines that preferentially inhibited analog-sensitive DDX3 versus wild-type. However, continued chemical derivatization of these compounds did not yield compounds with a binding affinity greater than 100 μ M.

To combat the low potency of analog-sensitive DEAD-box protein inhibitors, we next attempted an electrophile-sensitive chemical genetic approach for the specific inhibition of DEAD-box proteins. We identified a residue of low conservation within the P-loop of DEAD-box proteins, and found that mutation of this residue to cysteine was tolerated *in vitro* and *in vivo*. We then synthesized analogs of AMP with an electrophile appended to the monophosphate that potently and irreversibly bind to and inhibit electrophile-sensitive DEAD-box proteins.

As an alternate strategy, we also screened a series of compounds based on DBeQ, a small molecule inhibitor of the AAA+ ATPase inhibitor p97/VCP.¹³⁴ Derivatives of this molecule are currently in development as novel anti-cancer therapeutics.¹³⁵ We identified a series of disubstituted quinazolines that displayed ATP-competitive inhibition

of DDX3 at doses as low as 3 μ M, and found that these compounds compete with AMP-acrylamide binding to electrophile-sensitive DDX3.

Figure 1.1. DEAD-box proteins function in all stages of RNA metabolism.

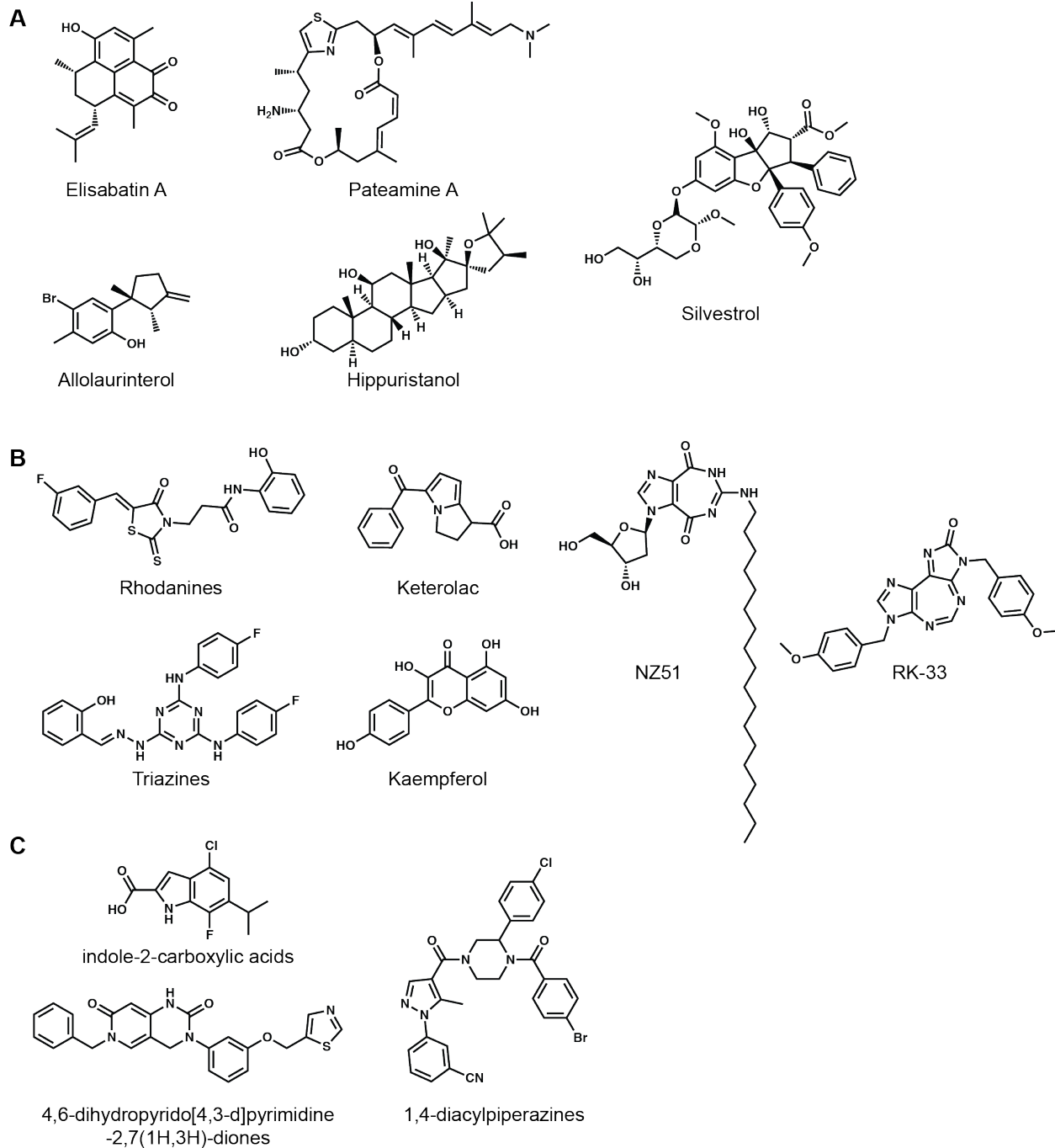


DEAD-box proteins are involved in all steps of RNA metabolism within the cell, including transcription, ribosome biogenesis, and pre-mRNA splicing in the nucleus, RNA export from the nucleus, organelle-specific RNA metabolism in the mitochondrion, and miRNA processing, snRNP biogenesis, non-sense mediated decay (NMD), translation, RNA storage and RNA decay in the cytoplasm. Numbers in parentheses indicate number of unique human DEAD-box proteins involved in each cellular process. Figure is adapted from Linder P and Jankowsky E, *Nat. Rev. Mol. Cell Biol.* (2011).¹⁸

Figure 1.2. Human DEAD-box proteins functions in RNA metabolism and disease.

DEAD-box protein	Orthologs	Protein effectors	Functions in RNA metabolism	Roles in physiology and disease
DDX1		hnRNPk	rRNA processing	Overexpressed and amplified in neuroblastoma, retinoblastoma
eIF4A1/2 (DDX2A/B)	Tif1/2	eIF4B, eIF4E, eIF4G, eIF4H	mRNA translation	Mediator of Myc-driven protein synthesis
DDX3X DDX3Y	Ded1, Dbp1p, Belle	eIF4G, PABP	mRNA translation RNA storage	Mutations cause unexplained intellectual disability; mutated in many cancers
DDX4	Vasa	TDRD1/6/7	mRNA translation, chromosome condensation, piRNA biosynthesis	Implicated in male infertility syndromes
DDX5 (p68) DDX17 (p72)	Dbp2p	ER, PR, AR, p53	Transcription, rRNA processing, RNA splicing	Oncogene and tumor suppressor transcriptional activation; HIV and HCV infection
DDX6 (RCK)	Dhh1p, Xp54	LSM1-7, CCR4, NOT1	RNA storage and decay, RNP assembly, mRNA translation	Candidate proto-oncogene; HIV and HCV infection
DDX10	Dbp4p		rRNA processing	NUP98-fusion in AML
DDX18 (MrDb)	Has1p		rRNA processing	Essential for hematopoiesis; endocrine-resistant breast cancer growth
DDX19 DDX19L	Dbp5p	Gle1, NUP214	RNA nuclear export	
DDX20 (Gemin3)		SMN1, Gemin2/4-8	snRNP biogenesis, transcription	Spinal muscular atrophy
DDX21		SIRT7	rRNA transcription, rRNA processing, transcription	Craniofacial malformation
DDX23	Prp28p	SNRNP200, PRPF3/4/6/8	Pre-mRNA splicing	Suppression of genomic instability, mutated in cancer
DDX24	Mak5p	p300, p53	Transcription, rRNA processing	Negative regulator of p53 transcriptional activity; HIV RNA packaging
DDX25 (GRTH)	DEADSouth		mRNA export, mRNA translation	Spermatogenesis; HIV infection
DDX27	Drs1p	BOP1, PES1	rRNA processing	Skeletal muscle growth
DDX28	Mss116p	GRSF1, DHX30	Mitochondrial ribosome biogenesis and mRNA translation	
DDX31	Dbp7p	NPM1	rRNA processing	Upregulated in RCC
DDX39A DDX39B	UAP56, Sub2p	TREX complex	Pre-mRNA splicing	Overexpressed in several GI/GU cancers
DDX41 (ABS)		STING	mRNA translation, rRNA processing	Cytosolic dsRNA sensor; negative regulator of p21(Cip1) mRNA translation; mutated in AML
DDX42	Prp5p	SF3B complex, TP53BP2	Pre-mRNA splicing, rRNA processing	Inhibits activator of p53-mediated apoptosis
DDX43 (HAGE)			mRNA translation	Overexpressed in HCC
DDX46	Prp5p	17S U2 snRNP	Spliceosome biogenesis	Negative regulator of type I interferon production
DDX47	Rpr3p	AGO1/2, NOL8, GABARAP	rRNA processing	
DDX48 (eIF4A3)	Fal1p	NOM1, MAGOH, Upf1/2/3	NMD, ribosome biogenesis	Mutations cause Robin sequence with cleft mandible and limb anomalies
DDX49	Dbp8p		rRNA processing	
DDX50			Ribosome biogenesis	
DDX51	Dbp6p		rRNA processing	Co-factor for c-Jun-activated transcription, inhibits dengue virus 2 replication
DDX52 (ROK1)	Rok1p		rRNA processing	
DDX53 (CAGE)		DNMT1		Positive regulator of D- and E-type cyclins; negative regulator of p53 expression
DDX54	Dbp10p	Hormone receptors	Ribosome biogenesis	CNS myelination; represses hormone receptor transcriptional activity
DDX55	Spb4p		Ribosome biogenesis	
DDX56	Dbp9p		Ribosome biogenesis	West Nile virus assembly; HIV infection
DDX59	RH41			Mutation causes orofacioidigital syndrome V; overexpressed in EGFR/Ras activation

Figure 1.3. Literature DEAD-box protein inhibitors.



A. Several natural products have been identified as eIF4A inhibitors, including elisabatin A and allolaurinterol, which are ATP-competitive, and pateamine A, hippuristanol, and silvestrol, which target allosteric sites. **B.** Literature DDX3 inhibitors are based on a variety of scaffolds (rhodanines, triazines, and ring-expanded nucleosides) and are poorly validated. **C.** Recently published eIF4A3 inhibitors (indole-2-carboxylic acids and

1,4-diacylpiperazines) and Brr2 inhibitors (4,6-dihydropyrido[4,3-d]pyrimidine-2,7(1H,3H)-diones identified by Takeda Pharmaceuticals.

Chapter 2

Analog-sensitive chemical inhibition of the DEAD-box protein DDX3

Abstract

Proper maintenance of RNA structure and dynamics is essential to maintain cellular health. Multiple families of RNA chaperones exist in cells to modulate RNA structure, RNA-protein complexes, and RNA granules. The largest of these families is the DEAD-box proteins, named after their catalytic Asp-Glu-Ala-Asp motif. The human DEAD-box protein DDX3 is implicated in diverse biological processes including translation initiation and is mutated in numerous cancers. Like many DEAD-box proteins, DDX3 is essential to cellular health and exhibits dosage sensitivity, such that both decreases and increases in protein levels can be lethal. Therefore, chemical inhibition would be an ideal tool to probe the function of DDX3. However, most DEAD-box protein active sites are extremely similar, complicating the design of specific inhibitors. Here, we show that a chemical genetic approach best characterized in protein kinases, known as analog-sensitive chemical inhibition, is viable for DDX3 and possibly other DEAD-box proteins. We present an expanded active site mutant that is tolerated *in vitro* and *in vivo*, and is sensitive to chemical inhibition by a novel bulky inhibitor. Our results highlight a course towards analog sensitive chemical inhibition of DDX3 and potentially the entire DEAD-box protein family.

Introduction

RNA is a highly dynamic macromolecule and can form numerous intra- and inter-molecular structures, which influence function. Like proteins, RNA molecules have chaperones that remodel structures to ensure proper function. RNA chaperones can be ATP-dependent like DEAD-box proteins,¹⁸ or ATP-independent like cold shock domain proteins and heteronuclear ribonucleoproteins (hnRNPs).¹⁶ In humans, there are 36

DEAD-box proteins which function in every step of RNA biology. For example, 18 DEAD-box proteins have been implicated in ribosome assembly, six in splicing, six in translation, and others in numerous other functions.¹⁸ DEAD-box proteins also associate with non-membrane-bound, microscopically visible puncta in cells known as RNA granules,^{136 137} which are thought to arise from phase separation caused by multivalent weak interactions.¹³⁸⁻¹⁴⁰ Remarkably, loss of function of the *C. elegans* DEAD-box protein CGH-1 (human DDX6) causes germ line granules to form square, crystalline structures *in vivo*.¹⁴¹ Therefore, DEAD-box proteins can function both on individual RNA:protein complexes (RNPs), or function as molecular dispersants to promote fluidity of RNA granules.

The human DEAD-box protein DDX3 (encoded by *DDX3X*) and its yeast ortholog *DED1* have been implicated in numerous cellular functions, but most consistently in remodeling RNA and RNPs during translation initiation.^{77,137,142-145} DDX3 and Ded1p also associate with two related types of RNA granules known as stress granules and P-bodies,^{136,137} and introduction of catalytically deficient Ded1p increases granule size,¹³⁷ suggesting they may have a role in determining the size of RNA granules by modulating weak interactions. Frequent mutations of *DDX3X* are found in numerous human malignancies including medulloblastoma,⁶⁵⁻⁶⁸ diverse blood cancers,⁶⁹⁻⁷² head and neck squamous cell carcinoma,^{73,74} lung cancer,¹²⁷ and more. However, cellular studies are complicated by the fact that DDX3 and *DED1* are essential genes, limiting the perturbations that can be made. Moreover, the poor time resolution of knockdown and

transfection experiments complicates assignment of direct and indirect targets of DDX3 in cells.

DDX3, like all DEAD-box proteins, couples ATP binding to conformational changes that create a binding surface selective for single stranded RNA.^{18,146} ATP hydrolysis then destabilizes this conformation and promotes product release.²⁸ Conversion to the ATP-bound closed state involves creation of a composite active site involving residues on both the N-terminal DEAD and C-terminal HELICc domains. Therefore, interfering with ATP binding or hydrolysis will prevent RNA and RNP remodeling by DEAD-box proteins.

Chemical inhibitors are extremely powerful tools to study function in cells due to their high temporal resolution. However, it is difficult to develop specific inhibitors to protein families with many highly related members, like DEAD-box proteins. In protein kinases, mutation of a “gatekeeper” residue to a smaller alanine or glycine uniquely sensitizes the mutant protein to bulky active site inhibitors which are otherwise inactive against the majority of the kinome.¹⁴⁷ This approach allows for high affinity and specific inhibition of individual protein kinases by introducing a single point mutation, and has been widely used to generate analog sensitive inhibitors¹⁴⁷ and artificial substrates.¹⁴⁸ A similar approach has been used to generate synthetic substrates or inhibitors for myosin and kinesin.^{149,150}

Here, we present proof-of-principle experiments demonstrating analog sensitive inhibition of the DEAD-box protein DDX3. We engineer a binding pocket near the ATP binding site by point mutation while retaining *in vitro* function and complementation of the essential yeast gene *DED1*. The expanded active site mutant is sensitized to pyrazolopyrimidine-related compounds *in vitro*, including GXJ1-76. In yeast, treatment with GXJ1-76 is lethal in cells harboring expanded active site mutants, but this is due to synthetic interactions with temperature sensitivity rather than specific inhibition. We therefore performed a general screen of existing analog-sensitive inhibitors and find a new series of compounds that also demonstrate more potent analog sensitive inhibition *in vitro*. Taken together, our results demonstrate that analog sensitive inhibition of DEAD-box proteins is possible, and highlight two scaffolds that could be used to design high affinity inhibitors.

Results

Targeting a hydrophobic cluster for expanded active site mutation

Analog sensitive inhibition of proteins requires generation of an expanded active site to accommodate bulky groups on the inhibitor. We examined the ATP binding site of DDX3 and found that the N6 position of adenosine points towards a cluster of four hydrophobic residues on the DEAD domain (Figure 2.1A).¹⁵¹ As N6 on adenosine is the position modified in many preexisting bulky kinase inhibitors and adenosine receptor agonists, we targeted this hydrophobic cluster for mutation. We then examined the conservation of this region in diverse human DEAD-box proteins and found high conservation at most positions but some variability (Figure 2.1B). Expanding this

analysis to DEAD-box proteins from *Saccharomyces cerevisiae* and *Escherichia coli* shows that all four positions are tolerant of substitutions, suggesting some structural plasticity in this region (Figure 2.1C). Therefore, there is a hydrophobic cluster adjacent to the ATP binding site that is conserved but also shows limited variability, suggesting it may be tolerant to mutation.

Expanded active site mutants of DDX3 are functional

We generated point mutants of three positions of the hydrophobic cluster in DDX3 (Figure 2.1A) and expressed and purified them from *E. coli*. All three variants eluted from a gel filtration column normally, suggesting no major disturbance in folding (Figure 2.2A). We used a shortened construct of DDX3 lacking the N- and C-terminal tails that has superior biochemical behavior and is highly active *in vitro* (DDX3 residues 132-607).¹⁵² As F182 abuts the ATP binding pocket, we tested the ability of the mutant protein to bind to the nucleotide adenosine monophosphate (AMP). We used AMP rather than ATP to directly test the binding affinity of nucleotide to the DEAD domain without avidity effects from the HELICc domain caused by ATP-dependent conformational changes.¹⁸ Both wild-type and the most severe mutation, F182A, have similar affinity to AMP (Figure 2.2B),¹⁵² indicating that nucleotide binding is not affected by this mutation. The observation that the I195A and I211A point mutants exhibit less severe defects in duplex unwinding than F182A (Figure 2.2C) and yeast growth (Figure 2.2D) suggests that they also bind nucleotide with similar affinity to wild-type DDX3, but we have not tested this directly.

We next functionally characterized the mutant proteins with expanded active sites. First, we tested the activity of the mutant proteins in RNA duplex unwinding assays¹⁵³ and found that all mutants had reduced activity to varying degrees (Figure 2.2C). As a further test, we determined the ability of the mutant alleles to complement the essential yeast gene *DED1 in vivo*, and found that all three alleles are sufficient for growth at normal temperature, despite the *in vitro* defects (Figure 2.2D). However, the I157A (DDX3 I195A) and F144A (F182A) variants exhibit cold and heat sensitivity, respectively (Figure 2.2D). It is surprising that these mutations are tolerated in yeast given their large kinetic defects *in vitro* (Figure 2.2C). It is possible that the effect of the mutations on Ded1 is not the same as that on DDX3, or that the yeast gene DBP1, which is a high-copy suppressor of *DED1*,¹⁵⁴ is compensating for partially functional Ded1p. In sum, mutation of the active site hydrophobic cluster is tolerated *in vitro* and *in vivo*, demonstrating it is possible to create an expanded active site in a DEAD-box protein.

Adenosine monophosphate but not published DDX3 inhibitors retards RNA duplex unwinding

We started to search for a viable scaffold for engineering bulky active-site inhibitors of DDX3 by surveying published inhibitors. To verify that our duplex unwinding assay was able to measure inhibition of DDX3 we titrated AMP into the reaction, which binds with low micromolar affinity (Figure 2.2B) and inhibits the enzyme.¹⁵² We observed endpoint depression of duplex unwinding with only a two-fold effect on the rate with an apparent K_i of $\sim 30 \mu\text{M}$ (Figure 2.3A-B). We then tested a series of published inhibitors including one rhodanine derivative,¹²⁰ two triazine derivatives,¹²⁰ the ring-expanded nucleoside

RK-33,¹²⁷ ketorolac salt,¹⁵⁵ and kaempferol-3-O- β -D-glucopyranoside.¹²¹ Both ketorolac salt and Maga compound 4b showed two-fold effects on the rate of duplex unwinding but no change to the endpoint of the reaction (Figure 2.3C-D). As ATP hydrolysis is required for product release but not duplex unwinding,¹⁵⁶ it is possible that endpoint depression (Figure 2.3B) is a signature of active-site inhibition while rate depression reflects noncompetitive inhibition by, for example, interacting with the duplex RNA substrate (Figure 2.3D). Alternatively, as Ded1p exhibits both ATP-dependent duplex unwinding and ATP-independent strand annealing activities,³⁰ it may be that ATP-competitive inhibition causes endpoint depression by altering the balance between unwinding and annealing. It remains possible that these inhibitors interact with the noncatalytic extensions that have been removed from DDX3¹³²⁻⁶⁰⁷, or that they interact with a composite bimolecular surface that only exists in cells. As we were unable to observe strong inhibition with any of the tested compounds, we sought alternatives to screen for analog sensitive inhibitors of DDX3.

Analog sensitive inhibition of DDX3

We surveyed diverse compounds in an attempt to find analog sensitive inhibitors of our expanded active site versions of DDX3. Focus was placed on the F182A mutation as it generates the largest pocket for a bulky inhibitor. We assembled a panel of nine adenosine receptor agonists, three AAA⁺-ATPase inhibitors or related quinazolinones,¹³⁴ and ~30 bulky kinase inhibitors and tested these using duplex unwinding. Kinase inhibitors were included because these have been shown to exhibit cross-reactivity with DEAD-box proteins.^{157,158} We found no inhibitory activity of any of

the adenosine receptor analogs and weak activity for the AAA⁺-ATPase DBeQ (data not shown). In addition, the large majority of the kinase inhibitors tested showed no effect. However, we found reproducible inhibition of duplex unwinding of only the mutant allele upon addition of the compound GXJ1-76 (Figure 2.4A-B). The concentration dependence of GXJ1-76 treatment is not linear (Figure 2.4B), possibly due to solubility issues at these concentrations. Notably, GXJ1-76 causes endpoint depression with only small changes on the rate (Figure 2.4C) as seen for AMP (Figure 2.3A-B), suggesting it is targeting the active site. GXJ1-76 shows a dose-dependent decrease in the ATPase activity of DDX3 with modest selectivity for the mutant allele (Figure 2.4D).

We then tested the ability of GXJ1-76 to inhibit Ded1p in yeast where the sole copy of *DED1* has been replaced by the expanded active site allele F144A (DDX3 F182A; Figure 2.2D). Yeast harboring a wild-type allele of *DED1* are minimally affected by treatment with 100 μM GXJ1-76 but *ded1*-F144A yeast grow considerably slower in the presence of GXJ1-76 (Figure 2.5A). As the *ded1*-F144A strain exhibits temperature sensitivity, we also tested inhibition in an unrelated temperature sensitive strain harboring mutations in the *MCM3* replicative helicase¹⁵⁹ to ensure that the growth defect was specific to GXJ1-76 interacting with Ded1p (Figure 2.5B). However, GXJ1-76 treatment also severely inhibited growth of the *mcm3-1* strain (Figure 5C), suggesting that the growth inhibition observed is due to off-target effects. In sum, expanded active site versions of DDX3 are preferentially inhibited by GXJ1-76 *in vitro*, but low potency and off-target effects in cells preclude use of this compound as is.

Bulky kinase inhibitor screen identifies anilinoquinazolines as lead scaffold

The biochemical inhibition of the mutant allele of DDX3 by GXJ1-76 (Figure 2.4) suggests that kinase inhibitor scaffolds may provide a promising lead for the development of analog-sensitive DEAD-box helicase inhibitors. We therefore performed a more general screen of thirty-one bulky analog-sensitive kinase inhibitors. From this screen we hoped to identify a scaffold or bulky group that had increased potency against analog-sensitive mutants and retained selectivity over the wild-type allele. The screen included mostly pyrazolopyrimidine-based molecules with various substitutions at both the 3- and N1-positions. A number of the N1-substitutions had potential hydrogen bond donors or acceptors that we hoped might form favorable interactions within the binding pocket to increase potency compared to GXJ1-76. The screen also included alternate inhibitor scaffolds beyond the pyrazolopyrimidine including pyrimidine-, and anilinoquinazoline-based inhibitors which we hoped would have different inherent affinity for the DEAD-box helicase ATP-binding site. From this screen, we identified AQZ01, a 6,7-dimethoxyanilinoquinazoline, as the most potent inhibitor of the F182A mutant (Figure 2.6).

Structure-activity relationship of AQZ01 yields compound 1, a selective analog-sensitive DEAD-box helicase inhibitor

Using screening hit AQZ01 as a starting point, we synthesized a series of molecules with increased steric bulk off the aniline of AQZ01 (Figure 2.7A). Interestingly, we found a strong preference for planar aromatic substitutions, as meta-bromo, isopropyl, or *tert*-butyl anilines exhibit little activity against the F182A mutant. The position of the planar

fused aromatic ring is also important, as the 1-aminonaphthalene compound (AQZ04) was much more potent than the 2-aminonaphthalene compound (AQZ05). The most promising derivative of this series, AQZ06, contains a 4-aminoindole at this position and retained the selectivity of AQZ01 against wild-type DDX3, but increased potency against the F182A mutant substantially (Figure 2.7B).

Assuming the 4-aminoindole of AQZ06 points into the pocket created by mutation of F182 of the hydrophobic cluster, we reasoned that changing the substitution of the 6- and 7-positions of AQZ06 could increase potency by making favorable interactions with the phosphate-binding P-loop of the helicase ATP-binding site. We therefore focused on placing hydrogen-bond donors/acceptors at the 6- and 7-positions of the anilinoquinazoline scaffold. Although we explored a series of molecules with formal negative charges at these positions (data not shown), we did not find a substitution that dramatically increased potency. Our most potent compound (**1**) shows a biochemical IC_{50} of $\sim 100\mu M$ in an ATPase assay (Figure 2.7C). However, we did not observe inhibition of duplex unwinding by compound **1** (Figure 2.7D-E), which may be because DDX3 is superstoichiometric to RNA in this assay while substoichiometric in the ATPase assay. We did not test AQZ06 or compound **1** against Ded1p in yeast because the low on-target potency of these compounds suggest they may cause off-target effects like GXJ1-76, and anilinoquinazoline derivatives are pumped from yeast by efflux pumps¹⁶⁰ (K.M.S. unpublished observations). Thus, we used chemical derivatization to increase the potency of AQZ01 against the F182A mutant of DDX3 greater than 10-fold while maintaining selectivity against wild-type DDX3.

Discussion

We have demonstrated a strategy for analog sensitive inhibition of the DEAD-box protein DDX3. By generating expanded active site mutations (Figures 2.1 and 2.2) we sensitized DDX3 to inhibition by bulky kinase inhibitors (Figure 2.4). Furthermore, treatment with the analog sensitive inhibitor GXJ1-76 is sufficient to block yeast growth when yeast contain the expanded active site allele F144A, however a *MCM3* mutant strain (*mcm3-1*) is also inhibited by this compound suggesting the effect may be due to off-target interactions (Figure 2.5). We then found that anilinoquinazolines show selective biochemical inhibition of DDX3 F182A (Figure 2.6) and made a series of derivatives to generate compound **1**, which selectively inhibits the mutant enzyme while sparing the wild-type (Figure 2.7).

We targeted a hydrophobic cluster adjacent to the ATP binding site for mutation to generate the expanded active site alleles. This region is broadly conserved in DEAD-box proteins in humans and other organisms (Figure 2.1B-C), suggesting it may be possible to apply this strategy to other DEAD-box proteins besides DDX3. Our data on DDX3 already suggests a promising strategy to probe DDX3 function. However, prior to use in cells the affinity of the compound needs to be improved to avoid nonspecific toxicity (Figure 2.5).

While using a kinase inhibitor scaffold as the starting point for our analog-sensitive DEAD-box helicase inhibitor immediately presents the problem of off-target toxicity

towards cellular kinases, all the bulky inhibitors tested are remarkably inactive against most wild-type kinases. We hoped that by changing inhibitor scaffolds from the very promiscuous pyrazolopyrimidine to the more selective anilinoquinazoline we would additionally reduce this off-target toxicity. However, anilinoquinazolines have poor solubility and pharmacokinetic properties. Future work optimizing the scaffold with these considerations in mind could yield viable analog sensitive inhibitors of DDX3.

An additional challenge is the low potency of all tested chemical scaffolds for DDX3. While initial work generating an analog-sensitive kinase inhibitor used PP1, a nanomolar inhibitor of Src, as a starting point,¹⁶¹ simple pyrazolopyrimidines and anilinoquinazolines are millimolar binders of DDX3. This is likely due to the fact that kinase inhibitors rely heavily on 'hinge' hydrogen bonding interactions for potency, and these contacts are not present in DEAD-box proteins. Making favorable interactions with the essential glutamine of the DEAD-box helicase Q-motif could provide a substantial jump in potency.²⁵ Finding substituents that interact favorably with the helicase P-loop could also produce a similar potency boost. Since the majority of the affinity of AMP for the helicase active site is derived from the phosphate and not the adenosine,¹⁵² it is possible that potent P-loop binders will provide the most promising path forward. We expect that additional rational design of compounds described within this work, along with additional screening for more potent leads, is a promising path towards developing a potent chemical inhibitor of DDX3 and DEAD-box proteins in general.

Materials and Methods

Multiple sequence alignments. Sequences for all DEAD-box proteins from *Homo sapiens*, *Escherichia coli*, and *Saccharomyces cerevisiae* were retrieved from the NCBI and were aligned using MUSCLE.¹⁶² A separate MUSCLE alignment of only human DEAD-box proteins was used to generate Figure 2.1B. Alignments were visualized using Jalview.¹⁶³

Recombinant protein purification. The codon optimized coding sequence for human *DDX3X* residues 132-607 fused to a 6xHis-MBP tag was expressed in *E. coli* BL21-Star cells. Induction was performed by addition of 1 mM IPTG for 18 hours at 16°C. Cell pellets were lysed by sonication, clarified by centrifugation at ~30,000 g, and purified by nickel chromatography including a 1M NaCl wash to remove bound nucleic acids. The His-MBP tag was cleaved using tobacco etch virus protease during dialysis into 200 mM NaCl, 10% glycerol, 20 mM HEPES pH 7, and 0.5 mM TCEP. The sample was then purified using heparin affinity chromatography, eluted at 400 mM NaCl, 10% glycerol, 20 mM HEPES pH 7, and 0.5 mM TCEP, and applied to a Superdex 75 gel filtration column equilibrated in 500 mM NaCl, 10% glycerol, 20 mM HEPES pH 7.5 and 0.5 mM TCEP. Fractions were then concentrated to roughly 30 μ M and supplemented with 20% glycerol and flash frozen in liquid nitrogen.

RNA duplex unwinding. Duplex unwinding assays were performed as described¹⁵³ with 2 mM ATP (Figure 2.2), 0.1 mM ATP (Figure 2.3, 2.7) or 0.5 mM ATP (Figure 2.4) and 1 μ M protein. The sequences of the RNA duplex strands are 5'–AGCACCGUAAAGACGC–3' and 5'–

GCGUCUUUACGGUGCUUAAAACAAAACAAAACAAAACAAA-3', and the short RNA strand was 5' radiolabeled with ^{32}P . Unlabeled "chase" RNA was not included. For experiments with inhibitors, protein, RNA and inhibitor were pre-incubated in helicase reaction buffer for five minutes and the reaction was initiated with ATP. All inhibitors were stored in DMSO and DMSO concentrations were matched between all experiments, which never exceeded 5%.

Coupled ATPase assays. Assays were performed using ADP-Quest (DiscoverX) according to manufacturer's instructions with $1\mu\text{M}$ enzyme, $10\mu\text{M}$ dsRNA and $100\mu\text{M}$ ATP-gold (DiscoverX) in 20mM Tris pH 7.5, 200mM NaCl, 1mM MgCl_2 , and 0.01% Triton X-100. The sequences of the RNA duplex strands are the same as those used in the duplex unwinding assay. All experimental results are reported as the average of three replicates with error bars representing the standard deviation of results, except for Figure 2.6, which was completed without replicates.

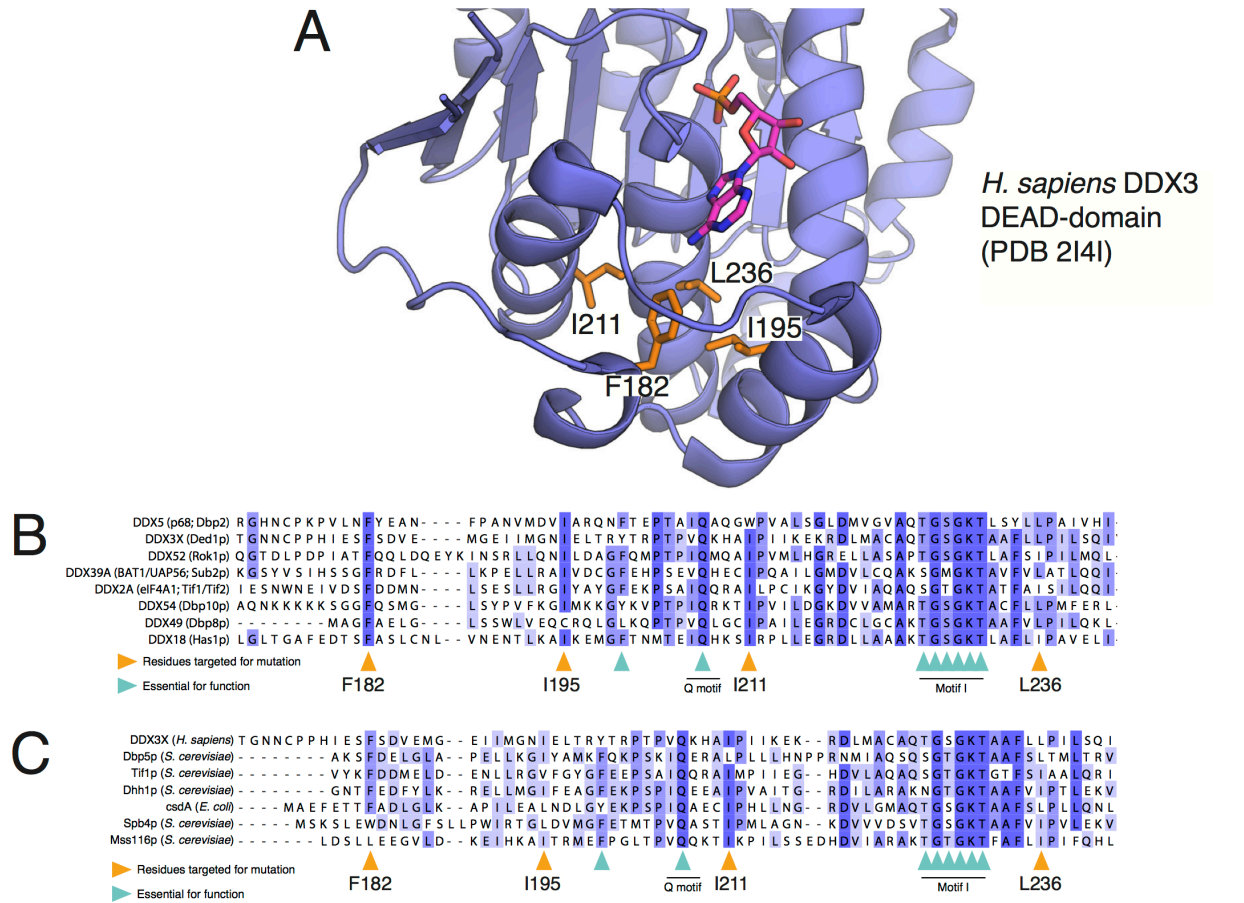
Yeast experiments. A strain of *S. cerevisiae* containing a deletion of the *DED1* locus complemented by a *URA3* marked plasmid harboring *DED1* was described previously.¹³⁷ Mutants were generated by site-directed mutagenesis in a plasmid containing *DED1* marked with *HIS3*, which was then transformed into yeast and counterselected using 5-fluoroorotic acid on His⁻ complete synthetic media. All strains were verified by plasmid purification and dideoxy sequencing. Strains were grown using YPD media following verification. Growth experiments in Figure 2.2D and 2.5B are

tenfold dilutions from OD ~1; continuous growth experiments in Figure 2.5 are by OD₅₉₅ measurement at 30°C in a Tecan Infinite F200 plate reader with 2mm orbital shaking.

Acknowledgements

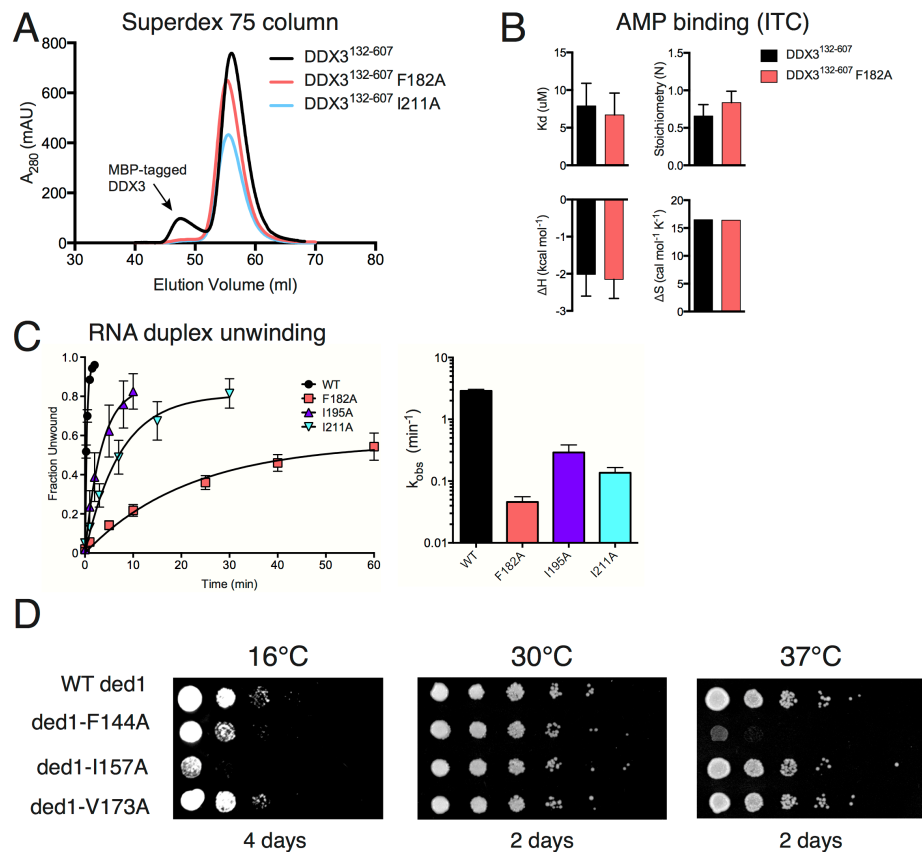
We thank Angie Hilliker and the lab of Jasper Rine for help with the yeast experiments and for sharing yeast strains, and Yoon-Jae Cho, Ray Deshaies, Jerry Pelletier, Flora Rutaganira and Joe Kliegman for sharing reagents and for suggestions on inhibitors to screen. S.N.F. is a Howard Hughes Medical Institute fellow of the Helen Hay Whitney Foundation, and J.A.D. and K.M.S. are Investigators of the Howard Hughes Medical Institute.

Figure 2.1: Targeting a hydrophobic cluster adjacent to the ATP binding pocket of DDX3 for mutation.



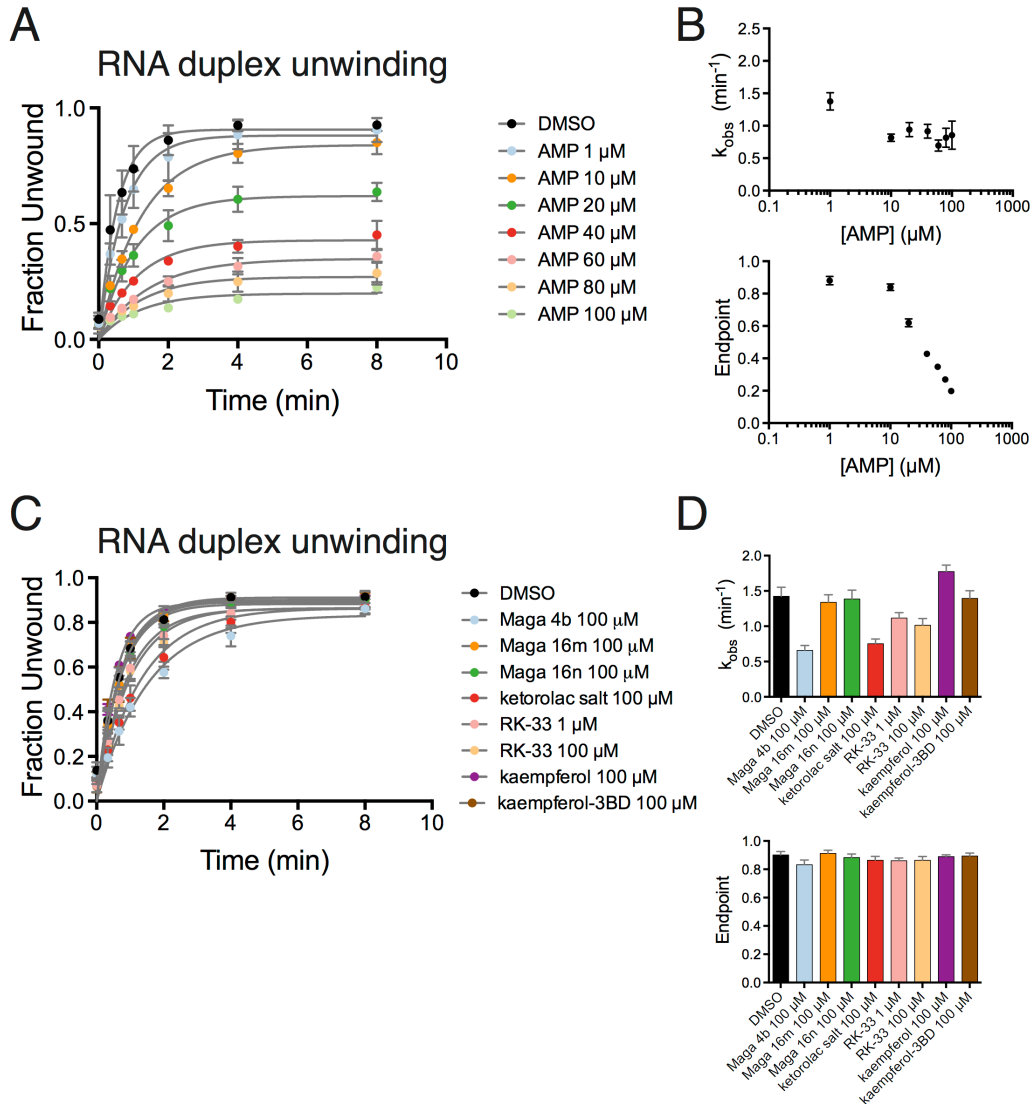
A. A structural view of the ATP binding pocket in human DDX3 bound to AMP (PDB 2I4I). AMP is in purple, DDX3 is in blue and the hydrophobic cluster residues are in orange. **B and C.** Sequence alignments of eight human DEAD-box proteins (**B**) or seven DEAD-box proteins from various organisms (**C**) showing overall conservation but some plasticity of the hydrophobic cluster residues. Core conserved motifs of DEAD-box proteins are indicated.

Figure 2.2: Hydrophobic cluster mutants of DDX3 support function *in vitro* and *in vivo*.



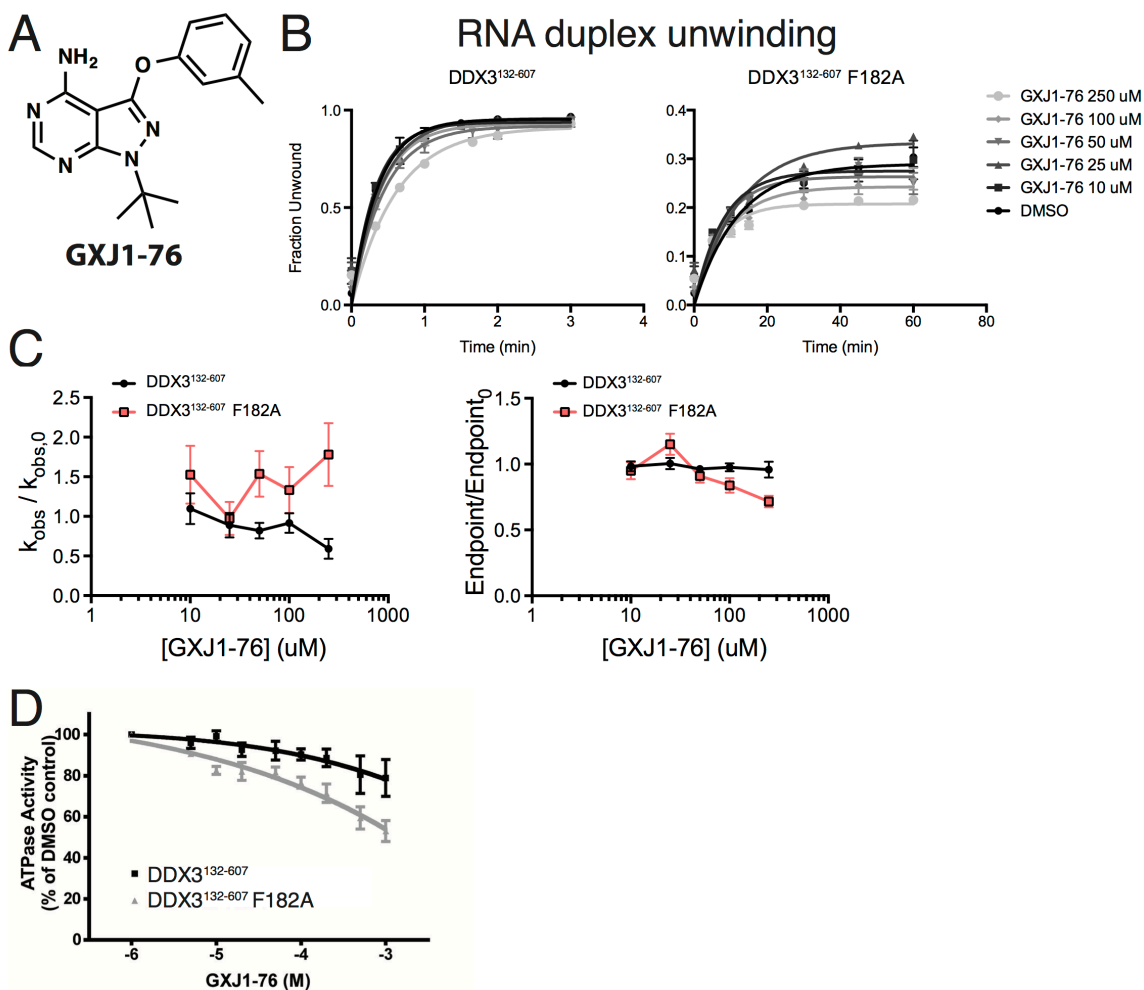
A. The indicated mutants of DDX3 purify normally as shown by their elution from a Superdex 75 gel filtration column. The indicated smaller peak to the left is uncleaved protein attached to the His-MBP tag. **B.** DDX3¹³²⁻⁶⁰⁷ wild-type and F182A bind AMP identically based on isothermal titration calorimetry. Error is standard error of the fit parameter. **C.** The hydrophobic cluster mutants are able to unwind RNA duplexes *in vitro* but at slower rates than wild-type DDX3¹³²⁻⁶⁰⁷. Error is S.D (left) and standard error of the fit parameter (right). **D.** Yeast with hydrophobic cluster mutants in the sole copy of *DED1* grow normally at 30°C but some exhibit cold (I157A) or heat (F144A) sensitivity. Yeast *DED1* F144, I157, and V173 correspond to DDX3 F182, I195 and I211, respectively. Spots represent a tenfold dilution series from OD 1.

Figure 2.3: Adenosine monophosphate, but not putative DDX3 inhibitors, retards duplex unwinding *in vitro*.



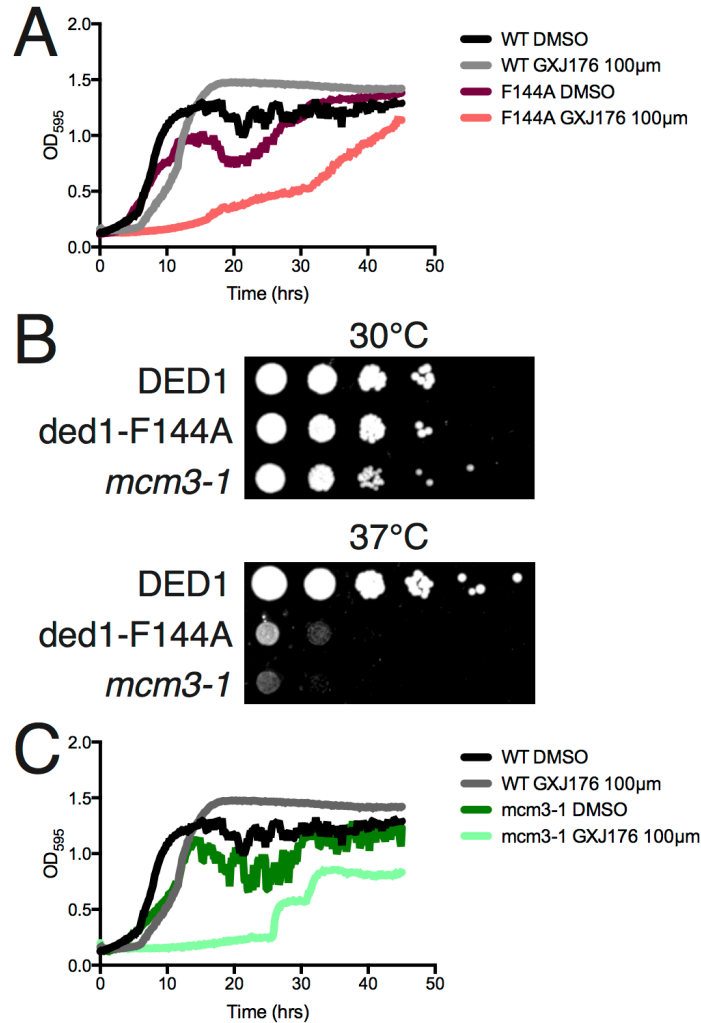
A and B. RNA duplex unwinding with AMP exerts concentration-dependent inhibition (**A**) and depresses the reaction endpoint but not rate (**B**). **C and D.** Duplex unwinding with seven published inhibitors of DDX3 (**C**) and fitted parameters (**D**) does not change the reaction endpoint and decreases the rate by at most two-fold. Error is S.D (**A,B**); standard error of the fit parameter (**C,D**).

Figure 2.4: Analog sensitive inhibition of DDX3.



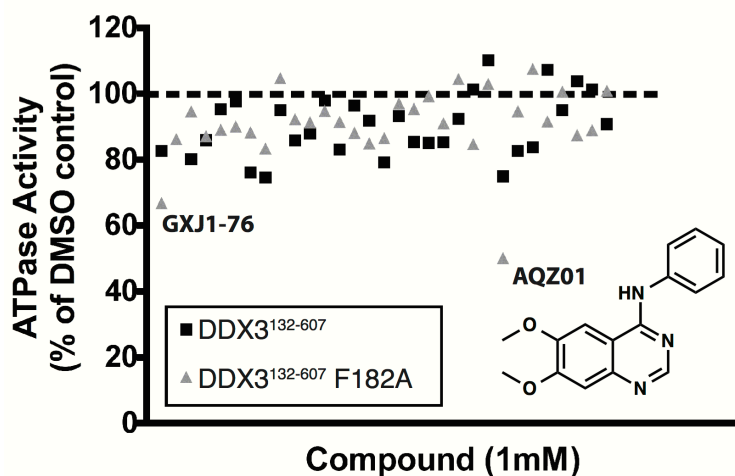
A. The structure of the initial lead compound GXJ1-76, which is a pyrazolopyrimidine-based inhibitor. **B.** Treatment of wild-type DDX3¹³²⁻⁶⁰⁷ with GXJ1-76 does not affect duplex unwinding (left), but treatment of DDX3¹³²⁻⁶⁰⁷ F182A causes a reduction in the duplex unwinding endpoint in a concentration dependent manner (right). Error is S.D. **C.** The relative unwinding rate and endpoint compared to the rate and endpoint with no compound as a function of GXJ1-76 concentration. Error is propagated error of the fit parameter. **D.** GXJ1-76 shows modest selectivity towards DDX3¹³²⁻⁶⁰⁷ F182A in ATP hydrolysis. Error is S.D.

Figure 2.5: Expanded active site alleles of *DED1* and other temperature sensitive yeast strains are inhibited by GXJ1-76.



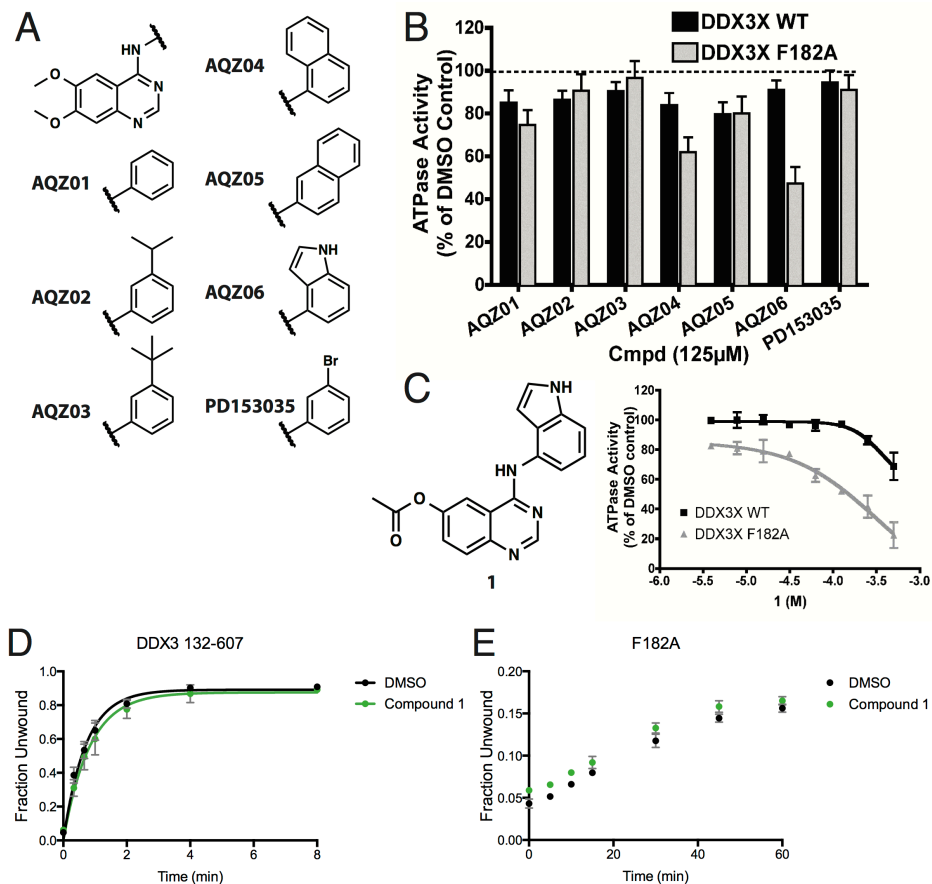
A. A yeast strain harboring the expanded site allele *ded1-F144A* as the sole copy of *DED1* is inhibited by GXJ1-76. Continuous growth with OD₅₉₅ measurements is plotted. **B.** Tenfold serial dilutions of *DED1*, *ded1-F144A* and *mcm3-1* yeast strains from OD 1 indicating similar growth at 30°C and temperature sensitivity at 37°C. **C.** Growth of the *mcm3-1* strain is also inhibited by GXJ1-76, indicating that the results in (A) likely result from off-target effects.

Figure 2.6: Screen for analog-sensitive DEAD-box helicase inhibitors identifies anilinoquinazoline scaffold.



The activity of 31 existing analog-sensitive kinase inhibitors against wild-type and F182A DDX3¹³²⁻⁶⁰⁷ was screened by an ATPase assay at 1 mM concentration. The structure of AQZ01, the most promising lead from this screen, is shown.

Figure 2.7: Chemical derivatization of AQZ01 yields compound 1.



A. A series of molecules were synthesized to contain bulky groups off the aniline of AQZ01. **B.** These analogs show stronger inhibition of DDX3¹³²⁻⁶⁰⁷ F182A by a 4-aminoindole (AQZ06) off the 6,7-dimethoxyanilinoquinazoline scaffold in an ATPase assay at 125 μM concentration. Error is S.D. **C.** Derivatization of the 6- and 7-positions of AQZ06 yielded compound 1, which shows selective inhibition of F182A-mutant, but not wild-type, DDX3¹³²⁻⁶⁰⁷ in an ATPase assay. Error is S.D. **D and E.** RNA duplex unwinding rates are identical for wild-type and F182A DDX3 when treated with compound 1. Note that DDX3 is substoichiometric in the ATPase assays (**B,C**) but superstoichiometric in the duplex unwinding assays (**D,E**).

Chemical Synthesis

Materials obtained commercially were reagent grade and were used without further purification. Reactions were monitored by thin layer chromatography (TLC) and/or mass spectroscopy (LC-MS) using a Waters Acquity UPLC/ESI-TQD with an Acquity UPLC. ¹H NMR spectra were obtained on a Bruker 400 spectrometer at 400 MHz.

Synthesis of GXJ1-76

1-(*tert*-butyl)-3-(*m*-tolylloxy)-1*H*-pyrazolo[3,4-*d*]pyrimidin-4-amine (GXJ1-76). To a solution of 3-bromo-1-(*tert*-butyl)-1*H*-pyrazolo[3,4-*d*]pyrimidin-4-amine (50mg, 0.186mmol), cesium carbonate (121.2mg, 0.372mmol), and pyridine (3.07μL, 0.038mmol) in DMF (0.9mL) under Ar(g) was added *m*-cresol (23.3μL, 0.223mmol) and copper (I) iodide (3.62mg, 0.019mmol). The reaction was heated to 140°C for 24hrs, filtered, then added to 20mL water and extracted with DCM. The combined organic layers were dried with sodium sulfate, filtered and concentrated. The resulting oil was purified by reverse phase HPLC (5-95% acetonitrile/water) to yield pure GXJ1-76 (1.9mg, 3.4% yield). ¹H NMR (400MHz, DMSO-*d*₆) δ 8.17 (1H, s), 7.28 (1H, t, J=7.85 Hz), 7.17 (1H, s), 7.13 (1H, d, J=8.16 Hz), 6.98 (1H, d, J=7.43 Hz), 2.32 (3H, s), 1.65 (9H, s).

General procedure for the synthesis of AQZ01-AQZ06

In an oven-dried and argon-cooled disposable scintillation vial, 4-chloro-6,7-dimethoxyquinazoline was dissolved in isopropanol. The appropriate substituted aniline was then added and the reaction mixture was heated at reflux for at least 2 hours. The reaction mixture was then cooled and the resulting solid collected by Büchner filtration and washed with ice-cold isopropanol to yield the pure compound.

6,7-dimethoxy-N-phenylquinazolin-4-amine (AQZ01). ^1H NMR (400MHz, DMSO- d_6)

δ 11.31 (1H, br s), 8.82 (1H, s), 8.28 (1H, s), 7.69 (2H, d, $J=7.55$ Hz), 7.50 (2H, t, $J=7.85$ Hz), 7.35 (1H, s), 4.03 (3H, s), 4.01 (3H, s).

N-(3-isopropylphenyl)-6,7-dimethoxyquinazolin-4-amine (AQZ02). ^1H NMR

(400MHz, DMSO- d_6) δ 11.28 (1H, br s), 8.81 (1H, s), 8.27 (1H, s), 7.52 (1H, dd, $J=1.28$ Hz, $J=3.59$ Hz), 7.41 (1H, t, $J=7.85$ Hz), 7.35 (1H, s), 7.22 (1H, d, $J=7.67$ Hz), 4.02 (3H, s), 4.01 (3H, s), 1.53 (1H, s), 1.27 (3H, s), 1.25 (3H, s).

N-(3-(*tert*-butyl)phenyl)-6,7-dimethoxyquinazolin-4-amine (AQZ03). ^1H NMR

(400MHz, DMSO- d_6) δ 11.25 (1H, br s), 8.81 (1H, s), 8.25 (1H, s), 7.64 (1H, t, $J=1.83$ Hz), 7.53 (1H, m), 7.43 (1H, t, $J=7.79$ Hz), 7.34 (1H, s), 4.02 (3H, s), 4.01 (3H, s). 1.34 (9H, s).

6,7-dimethoxy-N-(naphthalene-1-yl)quinazolin-4-amine (AQZ04). ^1H NMR (400MHz,

DMSO- d_6) δ 11.67 (1H, br s), 8.65 (1H, s), 8.40 (1H, s), 8.05 (2H, m), 7.90 (1H, d, $J=8.40$ Hz), 7.62 (3H, m), 7.36 (1H, s), 4.04 (3H, s), 4.03 (3H, s).

6,7-dimethoxy-N-(naphthalene-2-yl)quinazolin-4-amine (AQZ05). ^1H NMR (400MHz,

DMSO- d_6) δ 11.58 (1H, br s), 8.86 (1H, s), 8.40 (1H, s), 8.23 (1H, s), 8.03 (1H, d, $J=8.89$ Hz), 7.97 (2H, t, $J=8.64$ Hz), 7.87 (1H, dd, $J=1.95$ Hz, $J=8.77$ Hz), 7.57 (1H, m), 7.38 (1H, s), 4.05 (3H, s), 4.02 (3H, s).

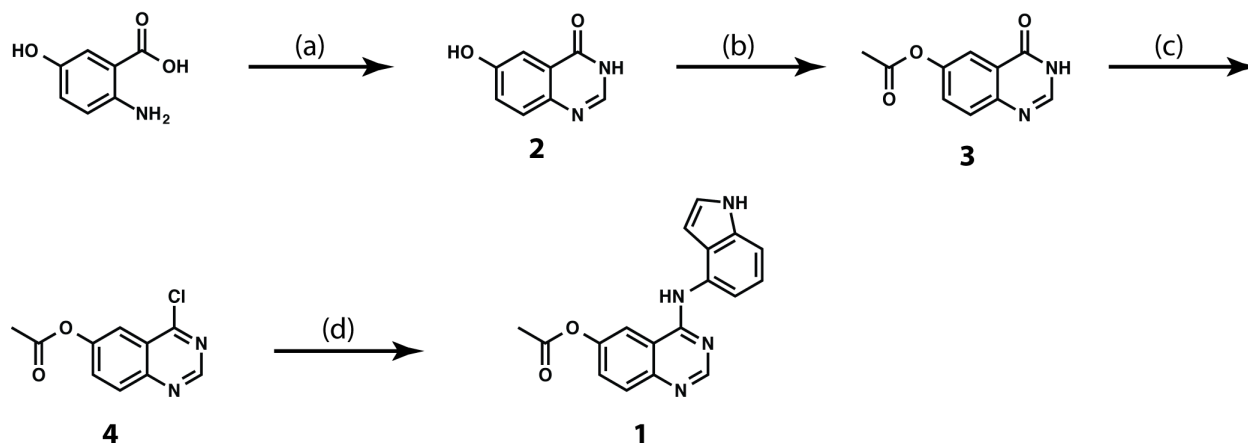
N-(1H-indol-4-yl)-6,7-dimethoxyquinazolin-4-amine (AQZ06). ^1H NMR (400MHz,

DMSO- d_6) δ 11.36 (1H, br s), 8.68 (1H, s), 8.27 (1H, s), 7.46 (1H, d, $J=8.04$ Hz), 7.39 (1H, m), 7.34 (1H, s), 7.20 (1H, t, $J=7.79$ Hz), 7.13 (1H, d, $J=7.06$ Hz), 6.32 (1H, s), 4.02 (6H, s).

N-(3-bromophenyl)-6,7-dimethoxyquinazolin-4-amine (PD153035). ¹H NMR

(400MHz, DMSO-d₆) δ 11.36 (1H, br s), 8.89 (1H, s), 8.32 (1H, s), 8.03 (1H, s), 7.78 (1H, d, J=7.85 Hz), 7.46 (1H, t, J=7.97 Hz), 7.35 (1H, s), 4.03 (3H, s), 4.01 (3H, s).

Synthesis of 1



Scheme 1. Conditions: (a) formamidine acetate, EtOH, reflux, (b) acetic anhydride, Et₃N, DMAP, DCM, reflux, (c) POCl₃, DIPEA, DCM, reflux, (d) 4-aminoindole, IPA, reflux.

6-hydroxy-4-quinazolone (2). To a solution of 5-hydroxyanthranilic acid (500mg, 3.26mmol) in ethanol (20mL) was added formamidine acetate (510mg, 4.90mmol). The reaction was heated under reflux until completed, as monitored by LCMS. The ethanol was then removed by rotary evaporation and the solid product was filtered from water as a grey solid (462.3mg, 85.7% yield). ¹H NMR (400MHz, DMSO-d₆) δ 7.90 (1H, s), 7.54 (1H, d, J=8.77 Hz), 7.41 (1H, d, J=2.80 Hz), 7.26 (1H, dd, J=2.92 Hz, J=8.77 Hz). MS (ESI) m/z 163.76 (100%, 164.73 (20%) [M + H]⁺.

6-acetyl-4-quinazolone (3). To a solution of compound 2 (162mg, 1mmol) and triethylamine (28μL, 0.2mmol) in dichloromethane (6.15mL) at 0°C was added dimethylaminopyridine (3.7mg, 0.03mmol). Acetic anhydride (283μL, 3.0mmol) was then

added dropwise. The reaction was then warmed to room temperature and monitored by TLC until completed. The reaction was concentrated by rotary evaporation and the resulting solid was purified by silica chromatography using a methanol/dichloromethane gradient (0-10% MeOH/DCM) to yield pure **3** as a white solid (143.9mg, 70.6% yield).

^1H NMR (400MHz, DMSO- d_6) δ 8.11 (1H, s), 7.83 (1H, d, $J=2.43$ Hz), 7.73 (1H, d, 8.77 Hz), 7.60 (1H, dd, $J=2.68$ Hz, $J=8.77$ Hz). MS (ESI) m/z 204.77 (100%), 205.41 (60%), 206.06 (100%), 206.71 (35%) $[\text{M} + \text{H}]^+$.

6-acetyl-4-chloroquinazoline (4). To a solution of compound **3** (100mg, 0.49mmol) in dichloromethane (3.5mL) at 0°C was added diisopropylethylamine (214 μ L, 1.23mmol). Phosphorus (V) oxychloride (54 μ L, 0.59mmol) was then added dropwise. The reaction was allowed to warm to room temperature then heated to 45°C under reflux. After ~24hours, the reaction was added to 25mL saturated sodium bicarbonate and extracted with DCM. The combined organic layers were dried with Na_2SO_4 , filtered, and concentrated. The resulting oil was dissolved in 50% ethyl acetate/hexanes and purified by silica chromatography (20-50% EtOAc/hexanes) to yield pure **4** as a white solid (58.6mg, 53.9% yield). ^1H NMR (400MHz, DMSO- d_6) δ 9.13 (1H, s), 8.19 (1H, d, $J=9.13$ Hz), 8.06 (1H, dd, $J=0.43$ Hz, $J=2.50$ Hz), 7.97 (1H, dd, $J=2.50$ Hz, $J=9.07$ Hz), 2.38 (3H, s). MS (ESI) m/z 222.78 (65%), 223.73 (90%), 225.34 (100%) $[\text{M} + \text{H}]^+$.

4-((1H-indol-4-yl)amino)quinazolin-6-yl acetate (1). 4-aminoindole (25.9mg, 0.196mmol) was added to a solution of compound **4** (21.8mg, 0.098mmol) in methanol (2mL) and the resulting solution was heated to 48°C for 1hr. The solution was allowed to cool to room temperature and then concentrated *in vacuo*. The resulting oil was dissolved in 5% MeOH/DCM and purified by silica chromatography (2-5% MeOH/DCM)

to yield pure compound **1** as a pale yellow solid (5.1mg, 16.4% yield). ¹H NMR (400MHz, DMSO-d₆) δ 11.17 (1H, s), 9.84 (1H, s), 8.42 (1H, s), 7.82 (1H, d, J=8.52 Hz), 7.66 (1H, d, J=8.77 Hz), 7.31 (1H, s), 7.17 (2H, m), 6.34 (1H, s), 2.38 (3H, s). MS (ESI) *m/z* 318.48 (90%), 320.95 (100%), 321.94 (15%) [M + H]⁺.

Chapter 3

Chemical genetic inhibition of DEAD-box proteins using covalent complementarity

Abstract

DEAD-box proteins are an essential class of enzymes involved in all stages of RNA metabolism. The study of DEAD-box proteins is challenging in a native setting since they are structurally similar, often essential, and display dosage sensitivity.

Pharmacological inhibition would be an ideal tool to probe the function of these enzymes. In this work, we describe a chemical genetic strategy for the specific inactivation of individual DEAD-box proteins with small molecule inhibitors using covalent complementarity. We identify a residue of low conservation within the P-loop of the nucleotide binding site of DEAD-box proteins and show that it can be mutated to cysteine without a substantial loss of enzyme function to generate electrophile-sensitive mutants. We then present a series of small molecules that rapidly and specifically bind and inhibit electrophile-sensitive DEAD-box proteins with high selectivity over the wild-type enzyme. Thus, this approach can be used to systematically generate small molecule-sensitive alleles of DEAD-box proteins, allowing for pharmacological inhibition and functional characterization of members of this enzyme family.

Introduction

Small molecule inhibitors are powerful tools for the study of cellular enzymatic processes due to their rapid onset of inhibition, which prevents cellular compensation, and their ability to be administered at varying doses, allowing for partial as well as complete loss-of-function phenotypes. As compared to the ATP-binding site of kinases, the development of small molecules targeting the nucleotide-binding pocket of ATPases has proven challenging. ATP-competitive inhibitors of the AAA+ ATPase p97/VCP and

structurally-related family members have been discovered,^{134,164} although a generalizable small molecule scaffold with high affinity for the ATPase nucleotide-binding pocket has not yet been identified. This is likely due to the reliance on electrostatic interactions for high affinity binding with its native substrate (ATP). Even if a suitable uncharged pharmacophore of the tri- or di-phosphate could be identified, the high conservation of this site across greater than 400 human proteins would make identifying a selective inhibitor of a single member of the family a significant challenge.^{23,24} As such, it is difficult to develop potent small molecule inhibitors of most ATPases, including the DEAD-box proteins.

DEAD-box proteins are the largest family of enzymatic RNA chaperones in humans.¹⁷ Named for their conserved Walker B motif consisting of adjacent aspartate-glutamate-alanine-aspartate (D-E-A-D) residues, DEAD-box proteins are required for all stages of RNA metabolism including transcription, processing and splicing, export, translation and decay.^{18,22,165} DEAD-box proteins bind nucleotides via the canonical Walker A and B motifs and the family-specific Q-motif that recognizes the adenine of ATP and makes the DEAD-box proteins ATP-specific.²³⁻²⁵ ATP binding and hydrolysis drives nonprocessive unwinding of RNA substrates by local strand separation.^{27,28} Yet despite the successful biochemical and structural characterization of this essential family of enzymes, our understanding of the specific functions and substrate specificity of DEAD-box proteins remains poorly understood.¹⁸

Owing to their roles in essential cellular processes, DEAD-box proteins are often misregulated in human disease and have been identified as potential pharmaceutical targets in cancer and viral and bacterial infection.^{35,132} However, specific chemical targeting of a single member of the DEAD-box family is challenging. Several natural product inhibitors of eIF4A have been identified, including hippuristanol and silvestrol^{109,111} and Takeda Pharmaceuticals recently published synthetic small molecules targeting eIF4AIII and Brr2.^{129,166} However, these compounds all rely on targeting cryptic allosteric pockets for their specific inhibition and as such they are highly selective yet are unlikely to be good structural starting points for discovery of inhibitors for other members of the DEAD-box family.

Although genetic and biochemical methods have been invaluable in the advancement of our understanding of DEAD-box proteins, they are fundamentally limited. Genetic knockout and loss of function mutants require extensive selection and verification^{167,168} during which time cellular compensation may obscure the primary role of the protein being studied. Gene knockout studies of structurally similar enzymes such as DEAD-box proteins may additionally be subject to compensation by partially redundant family members.^{154,169} The use of temperature-sensitive mutants in *S. cerevisiae* partially solves these problems,¹⁷⁰ although temperature-sensitive mutant enzyme inactivation often occurs through poorly understood mechanisms and temperature changes may alter temperature-sensitive processes such as RNA homeostasis. Pharmacologically controllable fusion proteins would be a potential avenue to acutely regulate DEAD-box

protein function,¹⁷¹ although the multiprotein complexes in which these proteins function may complicate fusion protein design.

A potential solution to these challenges is to utilize the tools of chemical genetics.

Previous chemical genetic approaches achieved specificity to the ATPases myosin-1 β and kinesin through analog-sensitive alleles generated by space-creating mutations adjacent to the N6-position of ATP.^{149,150} As further evidence for the importance of charged small molecules for targeting the ATPase nucleotide-binding pocket, both studies developed nucleotide di- and triphosphate-based inhibitors, and thus were limited by the inherent affinity of nucleotides for this pocket. This class of small molecules is additionally challenging to modify to develop cell-active compounds.¹⁷²

Recently we reported an effort to identify analog-sensitive mutants of the DEAD-box protein DDX3.¹³³ We identified a space creating mutation generated through mutation of a conserved aromatic residue in the adenine-binding pocket which showed a one hundred-fold reduction in biochemical activity. Yeast expressing this mutant in DDX3-homolog Ded1 displayed a temperature-sensitive phenotype, suggesting that increasing the size of the ATP pocket leads to a hypomorphic allele. We therefore turned to another strategy for imparting drug sensitivity into the ATP binding pocket of DEAD-box proteins.

Cysteine is the most nucleophilic of the twenty natural amino acids and is the second least common amino acid, after tryptophan. Additionally, cysteine-reactive small molecules are clinically approved and stable *in vivo*.¹⁷³ These properties provide an

ideal scenario for introduction of a cysteine into the binding pocket of a DEAD-box protein and targeting it for irreversible inhibition. Non-native cysteines have previously allowed for specific chemical inhibition of challenging targets,¹⁷⁴ and kinases expressing non-endogenous cysteines have been targeted in an 'electrophile-sensitive' (ES) chemical genetic approach,¹⁷⁵⁻¹⁷⁷ analogous to the 'analog-sensitive' approach (Figure 3.1A). This approach may be especially beneficial in DEAD-box proteins because of the difficulty in generating space-creating mutations in the ATPase active site that do not adversely affect enzyme function.¹³³ In fact, a previous study found that sensitivity to the semi-selective cysteine-targeting small molecule N-ethylmaleimide could be transferred to other ATPases through mutation of an active-site residue to cysteine.^{178,179}

In this work, we describe a method for the inhibition of specific DEAD-box proteins using covalent complementarity. We identify a site of low conservation in the P-loop of DEAD-box proteins that can be mutated to cysteine without a substantial reduction in enzyme activity. We then develop a series of electrophile-containing small molecules that target this cysteine and specifically inhibit electrophile-sensitive, but not wild-type DEAD-box proteins. Analysis of the drug bound complex in DDX3 reveals that the formation of a covalent bond between the engineered cysteine and the electrophilic inhibitor retains the adenine base interactions but requires a reorientation of the P-loop. Taken together, these results demonstrate that chemical genetic inhibition of the DEAD-box protein family is possible through targeting the ATP-binding site, and identify novel small molecules for the biochemical inhibition of these enzymes.

Results

Low conservation residue of the P-loop as a site for electrophile-sensitive mutation

The nucleotide-binding pocket of DEAD-box proteins, and ATPases in general, is highly conserved and recalcitrant to mutation.²⁴ In order to identify a chemically targetable and functionally silent mutation in the ATP-binding pocket, a structural alignment of all human DEAD-box proteins was performed. We identified the third residue of the P-loop/Walker A-motif (Figure 3.1B, red arrow) to be of lower conservation than surrounding nearly invariant residues. This residue lies approximately 5-6Å from the α - and β -phosphates of ATP and forms a hydrogen bond with a single phosphate oxygen (Figure 3.1B). Although no human DEAD-box proteins natively express a cysteine at this position, several members of the structurally related DExH-box protein family do, and cysteine is the third most common amino acid at this position across all human RNA helicases (Supplementary Figure 1A-C). We hypothesized that DEAD-box proteins could tolerate mutation of this position to cysteine to create an electrophile-sensitive helicase mutant.

Electrophile-sensitive DEAD-box proteins are functional

To test if introduction of a cysteine residue in the ATP pocket (the putative ES-mutant) was tolerated in DEAD-box proteins, we expressed and purified wild-type and ES-versions of human DDX3 (residues 132-607).¹⁸⁰ Since the ES-mutation (S228C in DDX3) lies within the ATP-binding pocket, we first tested if DDX3^{ES} retained RNA-dependent ATPase activity. Indeed, both DDX3^{WT} and DDX3^{ES} hydrolyzed ATP in the

presence of duplex RNA (Figure 3.2A), although DDX3^{ES} shows a three-fold reduction in activity compared to the wild-type (Figure 3.2B).

DEAD-box proteins hydrolyze ATP to remodel RNA and RNA-protein complexes.¹⁸ To test if ES mutants retained this function, we utilized an RNA duplex unwinding assay that monitors the separation of a ³²P-labeled 12-mer RNA oligonucleotide from a non-labeled 32-mer.¹⁵³ Electrophile-sensitive DDX3 retains ATP-dependent RNA-unwinding activity, although the rate of unwinding is three-fold lower than wild-type (Figure 3.2C-D). A similar reduction in RNA unwinding activity is observed in the electrophile-sensitive version of closely related yeast Ded1 (Supplementary Figure 3.1D-E). In a third example, the ES mutant of yeast Dbp2 shows no reduction in RNA unwinding activity (Supplementary Figure 3.1F), suggesting a differential effect of this mutation across the DEAD-box family. These results indicate that three DEAD-box proteins expressing a non-natural cysteine residue in the nucleotide-binding pocket are biochemically functional.

To assess whether electrophile-sensitive mutants of four different DEAD-box proteins (Ded1, Fal1, Dbp2, Dbp5) are able to substitute for their WT counterpart *in vivo*, we determined their ability to rescue the loss of essential DEAD-box protein genes DED1, FAL1, DBP2, and DBP5 in *S. cerevisiae*.^{143,145,181-184} Yeast expressing electrophile-sensitive versions of these DEAD-box proteins under endogenous promoters on extrachromosomal centromeric plasmids showed normal growth at permissive and restrictive temperatures (Figure 3.2E), in contrast to previous analog-sensitive mutants

(Supplementary Figure 3.1G). Taken together, these results demonstrate that electrophile-sensitive DEAD-box proteins are functional biochemically and *in vivo*.

Synthesis of electrophile-sensitive DEAD-box helicase inhibitors from AMP

To our knowledge there are currently no high-affinity ATP-competitive inhibitors of any RNA helicase. This limitation hindered our previous efforts to develop a chemical genetic method for the inhibition of the DEAD-box proteins using an analog-sensitive strategy.¹³³ Without a good drug-like starting point, we turned to nucleotide mimetics. Previous work identified adenosine-5'-monophosphate (AMP) as the minimal component of ATP required for DEAD-box helicase binding, since AMP is able to decrease RNA duplex unwinding by the DEAD-box protein Ded1 while adenosine is not.¹⁵² This work also found that AMP shows increased potency of inhibition of a subset of DEAD-box proteins as compared to ADP. These data illustrate the importance of the phosphate-P-loop electrostatic interactions for ligand-binding. Therefore, we sought to develop novel high-affinity chemical probes by preserving these interactions while correctly positioning an electrophile for reaction with the engineered cysteine residue.

The simplest molecule with both of these characteristics is AMP-acrylate (Figure 3.3A), which appends a cysteine-reactive Michael acceptor from the phosphate of AMP through a phosphoester linkage. To test whether these electrophiles bind covalently to DDX3, we first constructed a truncated mutant (residues 132-406) with improved characteristics for mass spectrometry, which allows for direct monitoring of the formation of a covalent adduct. Treating this optimized form of DDX3^{ES} with AMP-

acrylate showed complete adduct formation within five minutes at 5 μ M, while no detectable binding to DDX3^{WT} was observed after five hours of incubation (Figure 3.3B). This modification is robust and rapid down to equimolar concentrations of AMP-acrylate and protein (Supplementary Figure 3.2A). AMP-acrylate derivatives (AMP-methacrylate and AMP-crotonate) and AMP-acrylamide (Figure 3A) displayed reduced rates of reactivity against DDX3^{ES} according to their expected reduced cysteine reactivity (Figure 3.3B).¹⁸⁵ We also appended an electrophile to the acyclic AMP-analog PMEAs (adefovir)¹⁸⁶ to yield adefovir-acrylate (Supplementary Figure 3.2B). Adefovir-acrylate displayed only slightly reduced kinetics of labeling compared to AMP-acrylate (Supplementary Figure 3.2C), showing acyclic nucleotide analogs are also capable of labeling electrophile-sensitive DEAD-box proteins. Thus, the high rate of reactivity of AMP-acrylates against electrophile-sensitive DDX3 suggests that the nucleophile-electrophile pair are properly oriented within the ATP-binding site, and the sparing of wild-type DDX3 shows that this reactivity is dependent on the presence of the engineered cysteine.

AMP-acrylates undergo two-step reaction with electrophile-sensitive DEAD-box proteins

Although AMP-acrylates rapidly and irreversibly labeled DDX3^{ES} in five minutes, inspection of this reaction after longer periods of time surprisingly showed a loss of the full mass adduct (+ 401 Da) and the formation of a new adduct of + 54 Da (Supplementary Figure 3.2D-E). This adduct is not observed with AMP-acrylamide (data not shown). Carboxylic acids can be activated for nucleophilic attack in biological

systems through formation of an acyl-phosphate and acyl-phosphates are chemoselective reagents for acylation of basic amines,¹⁸⁷ including the catalytic lysine of kinases.^{188,189} We hypothesized that the + 54 Da adduct is the result of addition of an active site nucleophile (for example K230 in DDX3) into the acyl-phosphate bond of AMP-acrylate and subsequent elimination of AMP (Supplementary Figure 3.2F). The phosphoramidate P-N bond is not labile under physiological conditions, consistent with the finding that the second step does not occur with AMP-acrylamide. Indeed, mutation of K230 in DDX3^{ES} reduced the rate of formation of this +54 Da state (Supplementary Figure 3.2E). That this mutation did not fully abrogate the formation of this + 54 Da adduct suggests that other nucleophilic residues in the nucleotide-binding site may additionally contribute to this reaction.

AMP-acrylate maintains normal nucleotide-protein interactions with electrophile-sensitive DEAD-box proteins

Nucleotide binding to the DEAD-box protein active site requires numerous interactions that stabilize a specific enzyme conformation. To test if AMP-acrylates maintain these interactions upon covalent modification of electrophile-sensitive enzymes, we used differential scanning fluorimetry (DSF) to assess protein stabilization after compound binding.¹⁹⁰ Saturating concentrations of AMP stabilizes DDX3^{WT} and DDX3^{ES} to the same extent (Figure 3.3C-D), again showing the electrophile-sensitive mutation does not significantly disrupt nucleotide-binding. However, only DDX3^{ES} is stabilized by AMP-acrylate (Figure 3.3C-D). This confirms we have developed a chemical probe with specific reactivity to electrophile-sensitive DEAD-box proteins. That the magnitude of

stabilization by AMP-acrylate is as large as the stabilization by saturating concentrations of AMP suggests that compound binding has preserved native protein-ligand interactions.

A 3.0Å crystal structure of DDX3^{ES} bound to AMP-acrylamide

To better understand the binding of AMP-acrylates to electrophile-sensitive DEAD-box proteins, we solved the crystal structure of DDX3 (132-607) S228C bound to AMP-acrylamide to 3.0Å (Supplementary Table 3.1). The overall protein structure is highly similar to the previously published structure of DDX3 bound to AMP,¹⁸⁰ with a root mean squared deviation of 1.265Å (Figure 3.4A), as expected from its similar thermal stability (Figure 3.3C-D). AMP-acrylamide can be fit unambiguously to its density and clearly binds cysteine-228 through a covalent bond (Figure 3.4B). The adenine of AMP-acrylamide maintains several hydrogen-bonding interactions with the Q-motif of DDX3 including interactions with glutamine-207 and the backbone of arginine-202, as well as pi-stacking with tyrosine-200 (Figure 3.4B). The covalent linkage significantly re-orders the P-loop into a conformation that to our knowledge has not previously been observed in nucleotide-bound structures of DEAD-box proteins. This results in a 9.4Å shift in the sidechain of T226 as compared to AMP-bound DDX3 (Figure 3.4C). In addition, the phosphoramidate of AMP-acrylamide is also displaced more than 3Å out of the phosphate-binding pocket as compared to the phosphate of AMP so that only the 5'-oxygen hydrogen bonds with the P-loop backbone, whereas the phosphate of AMP typically makes numerous electrostatic interactions (Figure 3.4D).

AMP-acrylate specifically inhibits duplex unwinding by electrophile-sensitive DEAD-box proteins

In addition to specifically binding to electrophile-sensitive DEAD-box proteins, we asked whether AMP-acrylate binding inhibits DEAD-box protein function. Previously, compounds were identified that reduced DDX3 ATPase activity in a mutant selective manner, but did not inhibit its duplex unwinding activity.¹³³ This is likely due to the large stoichiometric excess of enzyme required for the single-turnover nature of the duplex unwinding assay while the ATPase assay utilized high substrate-enzyme ratios. Therefore, AMP-acrylates were tested for their ability to inhibit duplex unwinding by wild-type and electrophile-sensitive DEAD-box proteins.

While AMP-acrylate shows no appreciable inhibition of wild-type DDX3 (Figure 3.5A), the duplex unwinding activity of DDX3^{ES} is significantly reduced (Figure 3.5B). This inhibition is dose-dependent (Figure 3.5C-D) down to equimolar ratios of probe-enzyme, consistent with our previous findings that equimolar amounts of AMP-acrylate stoichiometrically bind to DDX3^{ES} (Figure 3.3). Surprisingly, AMP-acrylate reduces both the endpoint (Figure 3.5E) and the rate (initial velocity, Figure 3.5F) of duplex unwinding by DDX3^{ES}, in contrast to AMP which largely showed endpoint depression of duplex unwinding by DDX3^{WT}¹³³. To confirm that these results are not confined to one ES mutant DEAD-box protein, we also show that duplex unwinding by electrophile-sensitive Dbp2 (S161C) is reduced by AMP-acrylate, while Dbp2^{WT} is unaffected (Supplementary Figure 3.3A-B). These results demonstrate that AMP-acrylates are a novel chemical tool for the inhibition of biochemical activity of electrophile-sensitive DEAD-box proteins.

Discussion

We have developed a strategy for the chemical genetic inhibition of DEAD-box proteins through covalent complementarity. After identifying a residue of low conservation in the P-loop of all DEAD-box proteins (Figure 3.1), we generated ‘electrophile-sensitive’ DEAD-box proteins with near wild-type biochemical activity and the ability to complement loss of essential alleles in yeast (Figure 3.2). We then developed a series of AMP-acrylates that rapidly and specifically bind and inhibit these electrophile-sensitive DEAD-box proteins (Figures 3.3 and 3.5).

To our knowledge, the structure of AMP-acrylamide bound to DDX3^{ES} is the first crystal structure of a DEAD-box protein in complex with a small molecule inhibitor. The significant flattening of the P-loop observed in this structure is likely due to the torsion placed on this flexible loop by the tethered small molecule as it moves to maintain hydrogen bonds within the Q-motif. Interestingly, this distortion of the P-loop breaks the electrostatic interactions between the phosphate and the P-loop backbone, which are typically thought to drive nucleotide affinity. Since covalent compounds are hypothesized to undergo rapid reversible binding prior to covalent bond formation, we hypothesize that the initial binding event of AMP-acrylates by electrophile-sensitive DEAD-box proteins is driven by normal electrostatic interactions between the phosphate and P-loop and that covalent bond formation disrupts the positioning of the protein and small molecule to the final position observed in the crystal structure. This phenomenon was also observed with covalent inhibitors of Src kinase which lost hinge hydrogen

bonding contacts upon covalent bond formation.¹⁹¹ We cannot, however, rule out the possibility that this is an artifact of the rigidity of the amide bond of AMP-acrylamide and not representative of AMP-acrylates as a whole. This could explain the reduced rate of covalent modification observed by AMP-acrylamide as compared to AMP-acrylates (Figure 3.3B). Nonetheless, we expect the reordering of the P-loop observed upon binding of this compound to strongly antagonize the closed (active) conformation of DDX3.

We anticipate significant potential utility for these probes in cell-free systems in which DEAD-box proteins are often studied, including assays of spliceosomal assembly and function,¹⁹²⁻¹⁹⁴ ribosome biogenesis,^{195,196} and translation.¹⁹⁷⁻²⁰⁰ The current AMP-acrylate inhibitors are likely unable to cross the plasma membrane and thus cannot be used in whole-cell systems. These problems can be addressed through various pro-drug strategies developed for monophosphate- and phosphonate-containing nucleotide-reverse transcriptase inhibitors^{201,202} or through development of more ‘drug-like’ phosphate-mimics. Acyclic derivatives of AMP-acrylate, like adefovir-acrylate (Supplementary Figure 2B-C), could additionally provide a path for improving the pharmacological properties of electrophile-sensitive DEAD-box protein inhibitors. Continued screening for small molecules that bind the DEAD-box protein nucleotide-binding site may yield novel chemotypes with improved pharmacological properties that can be modified into electrophile-sensitive inhibitors. AMP-acrylates could be used as occupancy probes to aid in the identification of nucleotide-competitive compounds.²⁰³

AMP-acrylates developed in this work are notable because of their extremely fast binding to electrophile-sensitive DEAD-box proteins and their selectivity for electrophile-sensitive over wild-type enzymes (Figure 3.3B). The rapid kinetics of binding and inhibition allows for little or no pretreatment of compounds, thus mitigating off-target effects. These effects can be additionally ruled out through the use of wild-type controls, which appear to be completely spared from inhibition at doses of AMP-acrylate up to 50 μ M (Figure 3.5A). Although previous studies have identified AMP as an inhibitor of DEAD-box proteins at doses as low as 10 μ M,¹³³ we hypothesize that AMP-acrylates have reduced reversible affinity for the nucleotide-binding site of DEAD-box proteins because of the added electrophilic moiety and reduced formal charge. That no reactivity was observed between AMP-acrylates and wild-type enzymes additionally suggests that the cysteine reaction is required for addition into the phosphoanhydride bond (Supplementary Figure 3.2D-F). This is likely due to the fast kinetics of the cysteine reaction and slow kinetics of addition into the phosphoanhydride, which is only accessible upon covalent tethering of the compound into the active site. Therefore, AMP-acrylates show improved affinity for electrophile-sensitive DEAD-box proteins due to covalent complementarity and reduced binding to wild-type enzymes. These characteristics make these probes ideal for the study of complex biological processes with multiple ATPases in which DEAD-box proteins often function.

While no DEAD-box proteins natively express a P-loop cysteine, RIG-I, a closely related protein of the RLR family of RNA helicases,²⁰⁴ and several DExH-box proteins express endogenous cysteines at this position. This raises the intriguing possibility that the

chemical tools developed in this work could function as native chemical inhibitors of these enzymes. Additionally, since DExH-box proteins exhibit unselective nucleotide base binding and do not contain a Q-motif,²⁰⁵ changing the nucleotide base of AMP-acrylates may yield compounds with heightened specificity for natively electrophile-sensitive DExH-box proteins. The different orientation of the nucleotide base relative to the triphosphates in DExH-box proteins may also result in altered binding by acyclic AMP derivatives such as adefovir-acrylate.

Although our work establishes a chemical genetic strategy in DEAD-box proteins only, the high conservation of the P-loop and nucleotide-binding site suggests that this strategy may work in other RNA helicases and ATPases in general. In fact, all other families of ATPases (AAA+, ABC transporters, etc.) contain family members with native cysteines at this position (Supplementary Figure 3.1A). This implies that the overall ATPase fold can accept the electrophile-sensitive mutation and suggests that the strategy developed in this work may have implications outside the DEAD-box protein family. The presence of related enzymes with native cysteines at this position also suggests that the AMP-acrylate may react with proteins outside of the RNA-helicase family. Other modifications to AMP-acrylate such as the acyclic adefovir-acrylate may obviate these other activities. Although AMP-acrylates likely have utility beyond DEAD-box proteins, we expect that novel probes will be required for the specific inhibition of the diverse ATPase families. This will allow for the specific chemical inhibition of nearly five hundred human enzymes, encompassing nearly every biological process.

Materials and Methods

Multiple sequence alignment. Sequences for all human DEAD-box proteins were obtained from NCBI. All visualizations were generated by WebLogo (weblogo.berkeley.edu).²⁰⁶

Recombinant protein purification. DDX3 (132-406) and DDX3 (132-607) were expressed and purified as previously described.¹³³ Full-length Ded1 and Dbp2 were expressed as 6xHis-SUMO fusion proteins in *E. coli* BL21 (DE3) cells. Cultures were grown to OD ~0.8 then protein expression was induced with 0.5mM IPTG at 16°C overnight. Cells were lysed by microfluidizer, clarified at 20,000 x g for 25min, and purified by nickel chromatography including a 1M NaCl wash to remove bound nucleic acids. Eluted protein was incubated with SUMO protease and dialyzed into 25mM Tris pH 7.5, 300mM NaCl, 0.5mM TCEP overnight at 4°C. The sample was then purified by heparin chromatography and eluted at ~700mM NaCl, 25mM Tris pH 7.5, 0.5mM TCEP. Fractions containing pure protein, as analyzed by coomassie staining, were concentrated to at least 50µM, supplemented with 10% glycerol, and snap frozen in liquid nitrogen.

Coupled ATPase assays. Assays were performed as previously described.¹³³ All experimental results are reported as the average of three replicates with error bars representing the standard deviation of the results.

RNA duplex unwinding assays. RNA duplex unwinding assays were performed as previously described¹⁵³ using 1nM duplex RNA, 2mM ATP/MgCl₂, and 1μM DDX3 (132-607), 0.5μM Ded1 or 0.5μM Dbp2. Reaction buffer contained 20mM HEPES pH 7.5, 100mM NaCl, 0.5mM MgCl₂, 1mM TCEP, 0.01% (v/v) NP-40, and 5% glycerol. The sequences of the RNA duplex strands are 5'-GCUUUACGGUGC-3' and 5'-GAACAACAACAACAACCAUGGCACCGUAAAGC-3'. For experiments that included chemical inhibitors, RNA, protein and inhibitors were pre-incubated in reaction buffer for 5min prior to addition of ATP-Mg. Results were quantified using ImageJ (NIH)²⁰⁷. All experimental results are reported as the average of three replicates with error bars representing the standard error of the results.

Yeast genetics. For a complete list of yeast strains and plasmids used in this work, see Supplementary Tables 3.2 and 3.3. A strain of *S. cerevisiae* containing a deletion of the DED1 locus complemented by URA3-marked plasmid containing DED1 was previously described.¹³⁷ Heterodiploid strains containing single locus knockouts of FAL1, DBP2, or DBP5 were purchased from Dharmacon (GE Life Sciences). Strains were transformed with URA3-marked plasmids containing the genetic locus of the DEAD-box protein of interest and selected for on Ura- synthetic media. Strains were then sporulated by carbon starvation (1% potassium acetate, 0.1% yeast extract, 0.05% dextrose) for 5 days, released from their asci by lyticase treatment (Sigma-Aldrich) and dispersed by sonication.¹⁶⁷ Haploid strains containing KanMX-cassette knockouts of FAL1, DBP2, or DBP5 at the genetic locus complemented with URA3-marked plasmids containing FAL1, DBP2 or DBP5 were selected on geneticin and Ura- synthetic complete media.

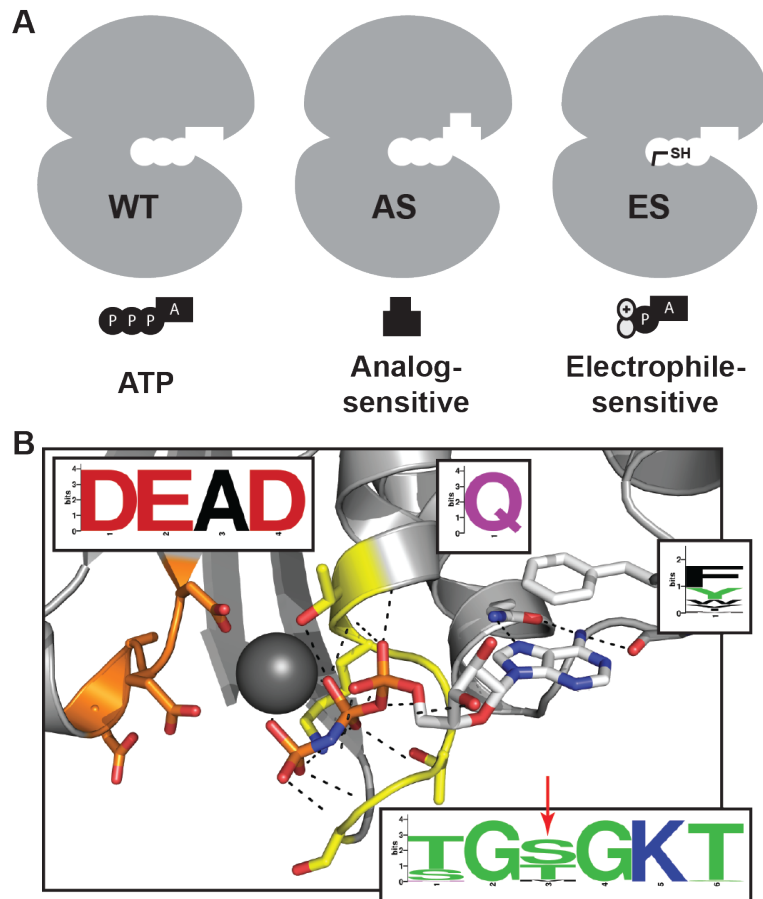
Strains were confirmed to be unable to grow on 5-FOA synthetic complete media, then verified by plasmid isolation followed by sequencing. Mating type was determined by complementation with MATa (his2-) and MATalpha (his2-) tester strains (gift from David Morgan, UCSF) on Ura- His- synthetic complete media. Electrophile-sensitive mutants were generated by site-directed mutagenesis in a HIS3-marked plasmid, transformed into yeast and counterselected using 5-FOA and His- synthetic complete media. Strains were verified by plasmid isolation followed by sequencing and subsequently grown in YPD media. Growth experiments are 10-fold dilutions from cultures grown to OD ~1.

Mass spectroscopy assay. DDX3 (132-406) WT or S228C (250nM unless otherwise noted) or DDX3 (132-607) (500nM) in 10mM Tris pH 7.5, 100mM NaCl and 1mM MgCl₂ was incubated with compounds at 4°C. The extent of modification at various time points was determined by whole protein mass spectrometry using a Waters Acquity UPLC/G2-XF QTOF. Reported data points are single-replicates from three independent experiments (Figure 3.3B, Supplementary Figure 3.2E) or means of three independent experiments with error bars representing the standard error (Supplementary Figure 2A-C). Curves are fit using Prism (GraphPad).

Differential scanning fluorimetry. DSF was performed as described¹⁹⁰ using 4μM DDX3 (132-406) in a reaction buffer containing 20mM HEPES pH 7.5, 150mM NaCl, and 5mM MgCl₂. Reaction mixtures were incubated for 5min prior to the start of the assay. Results are the average of three independent experiments with error bars representing the standard error of the mean. Curves are fit using Prism (GraphPad).

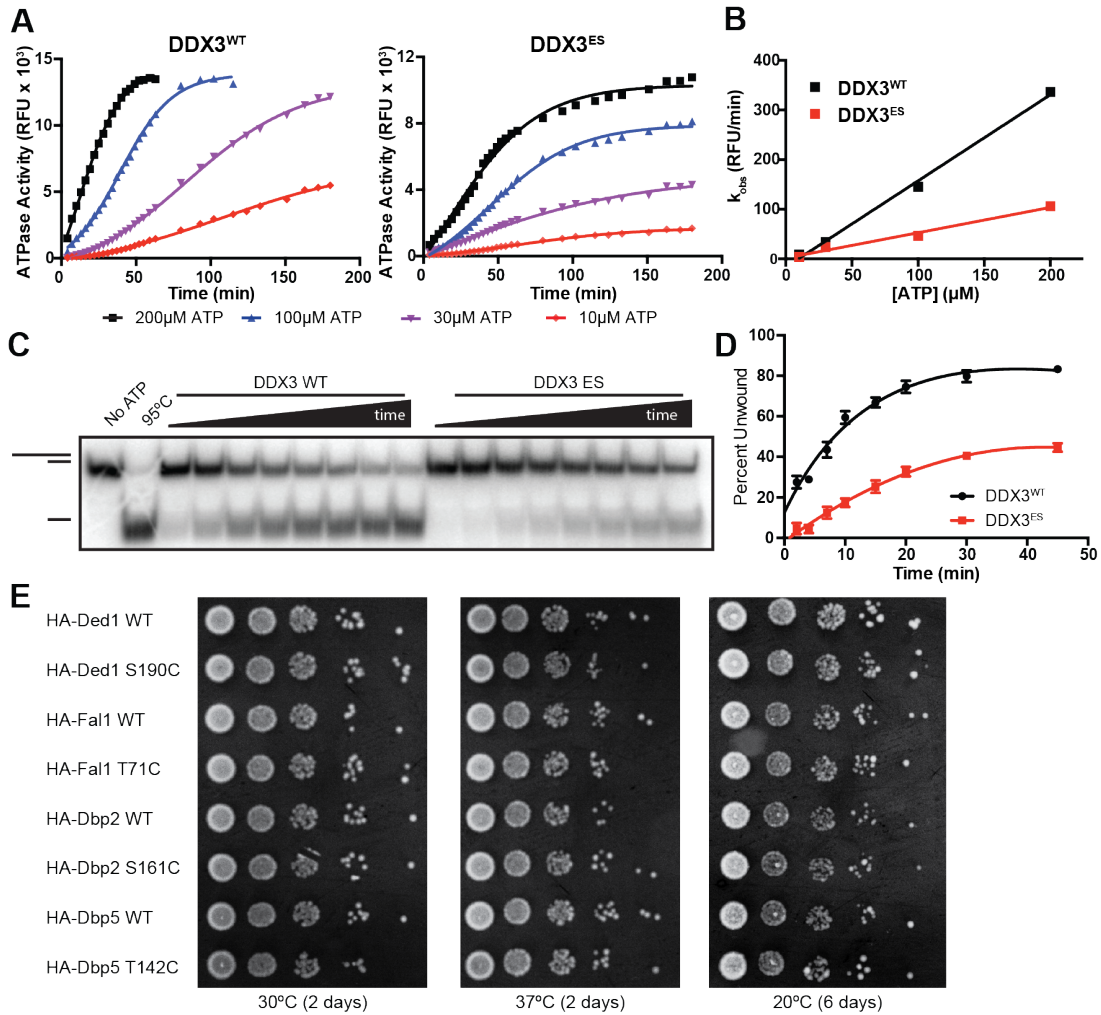
X-ray crystallography. Purified DDX3^{ES} (132-607) was incubated with 100μM AMP-acrylamide at 4°C until completely labeled, as judged by LC-MS. Protein was then applied to a Superdex 200 gel filtration column equilibrated in 20mM HEPES pH 7.5, 500mM NaCl, 10% (v/v) glycerol, and 0.5mM TCEP and fractions containing pure protein were flash frozen in liquid nitrogen for storage. Conditions for crystallography were previously described.¹⁸⁰ Data was collected at Beamline 8.2.2 of the Advanced Light Source (LBNL, Berkeley, CA). Data were indexed and integrated using iMosflm,²⁰⁸ scaled using Scala,²⁰⁹ phased using molecular replacement with PHASER²¹⁰ using Protein Data Bank (PDB) 5e7j as a search model, and refined and built using PHENIX²¹¹ and Coot.²¹² Structures were visualized with PyMOL.²¹³ Atomic coordinates and structure factors for the reported crystal structure have been deposited with the Protein Data Bank under accession number 6CZ5.

Figure 3.1. Development of an ‘electrophile-sensitive’ mutation in DEAD-box proteins.



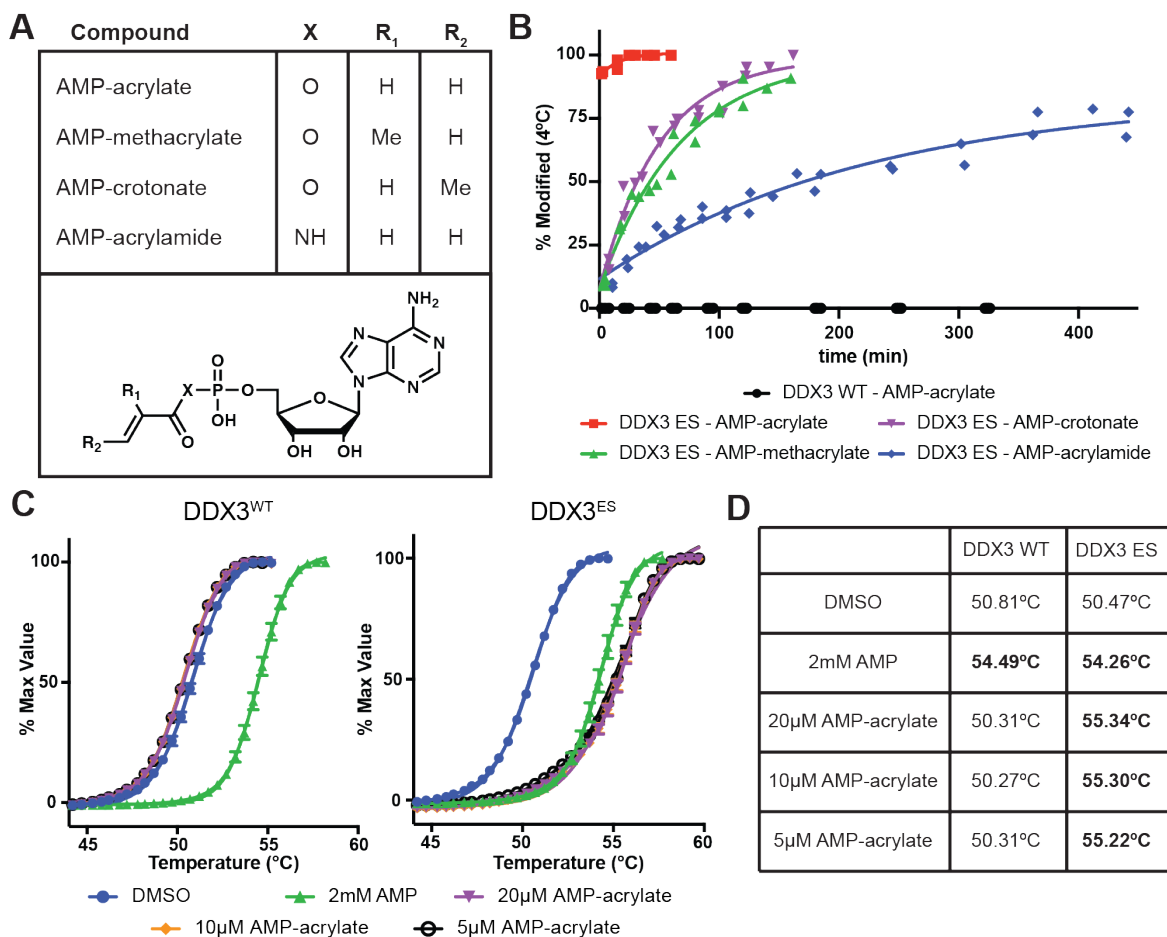
A. Analog-sensitive (AS) chemical genetic strategies rely on space-creating mutations and bulky inhibitors, while electrophile-sensitive (ES) strategies rely on cysteine-electrophile covalent complementarity. **B.** Conserved nucleotide binding site of DEAD-box proteins including the P-loop/Walker A (yellow), DEAD-box/Walker B (orange) and Q-motif (white) with conservation across all human DEAD-box proteins (inserts) and the site of electrophile-sensitive mutation identified (red arrow; PDB 2HXY).

Figure 3.2. Electrophile-sensitive DEAD-box proteins retain biochemical and cellular function.



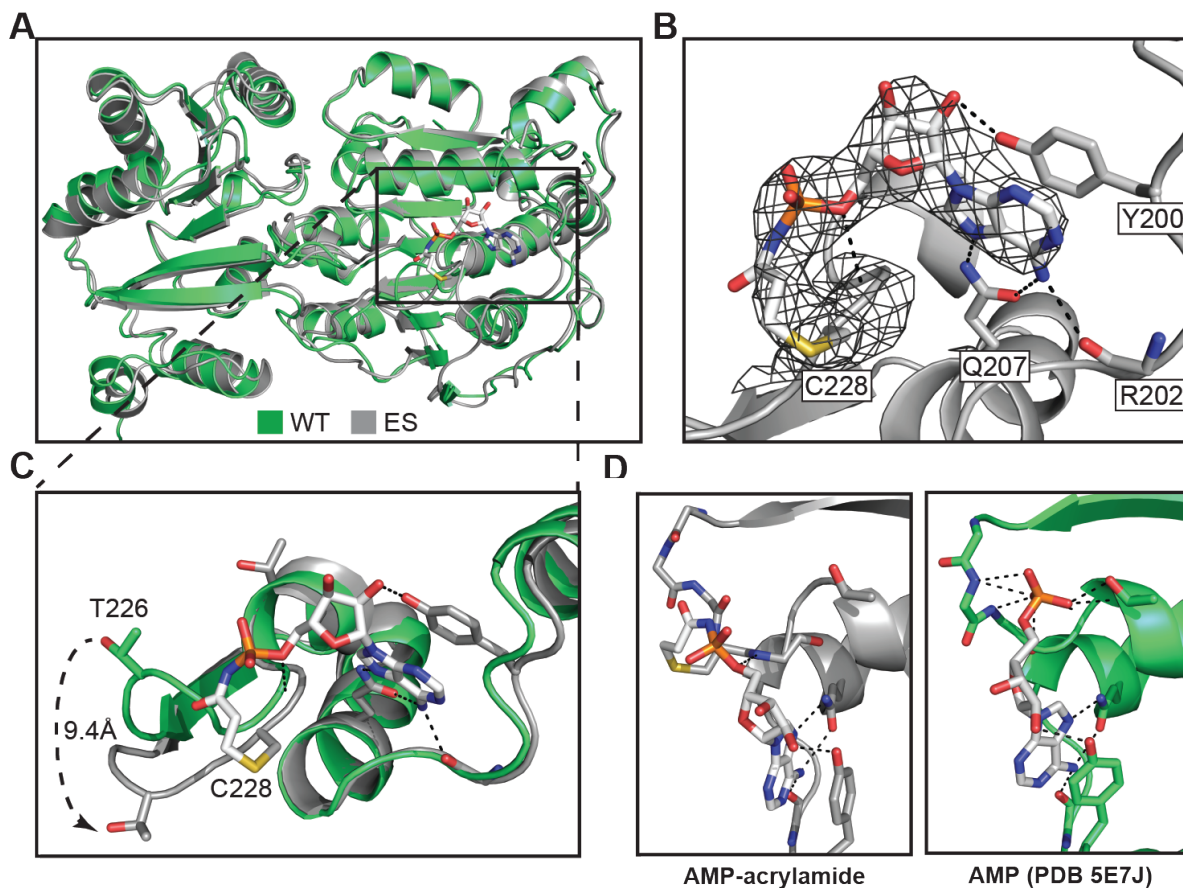
A. RNA-dependent ATPase activity of DDX3^{WT} and DDX3^{ES} (S228C) at four concentrations of ATP. **B.** Comparison of the observed rate of RNA-dependent ATP hydrolysis by DDX3^{WT} and DDX3^{ES}. **C.** Representative RNA duplex unwinding assay comparing the activity of DDX3^{WT} and DDX3^{ES}. **D.** Quantification of the fraction of RNA duplex unwound by DDX3^{WT} and DDX3^{ES}. **E.** Serial dilutions of log-phase cultures of budding yeast strains expressing wild-type or electrophile-sensitive DEAD-box proteins grown at permissive (30°C) or restrictive (20°C, 37°C) temperatures.

Figure 3.3. AMP-acrylate and derivatives potently and irreversibly bind electrophile-sensitive DEAD-box proteins.



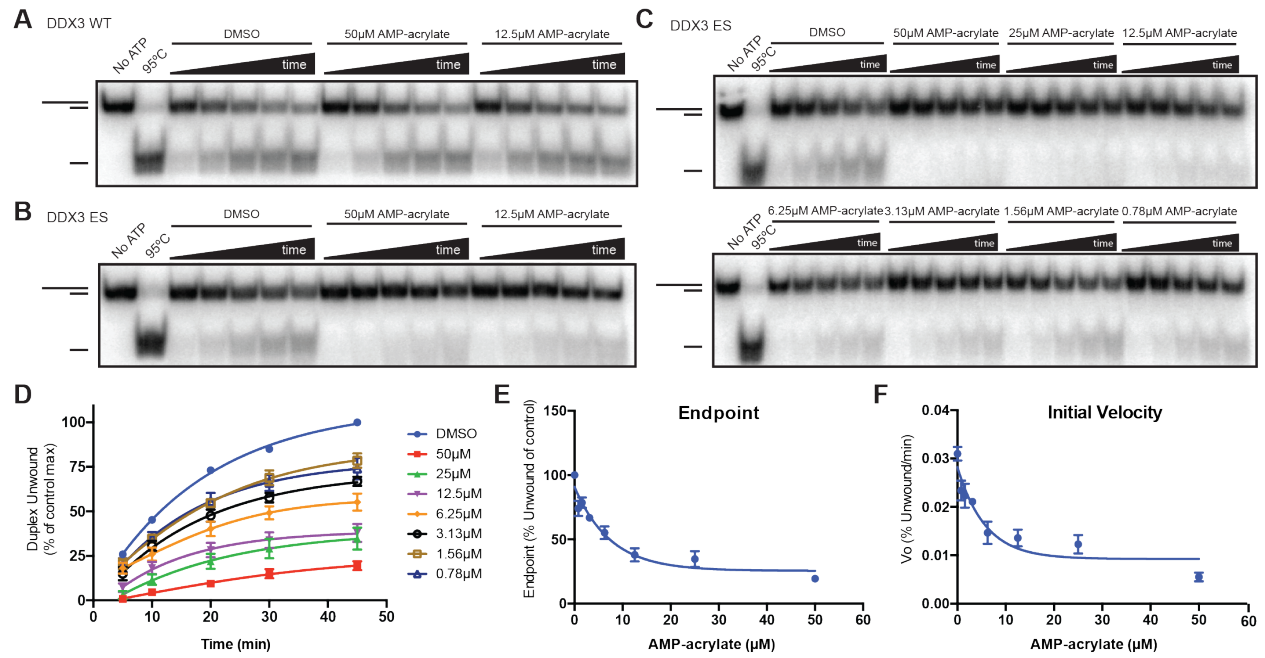
A. Structures of AMP-acrylate and derivatives used in this study. **B.** Percent modification of DDX3 (132-406) WT and ES (S228C) incubated with 5µM AMP-acrylate and derivatives at 4°C. **C.** Differential scanning fluorimetry curves of DDX3 (132-406) WT and ES in the presence of AMP or 5µM, 10µM, or 20µM AMP-acrylate. **D.** Experimentally determined melting temperatures (T_m) of DSF curves from (C) with significant differences from DMSO in bold (t-test, $p < 0.001$).

Figure 3.4. Crystal structure of AMP-acrylamide bound to electrophile-sensitive DDX3.



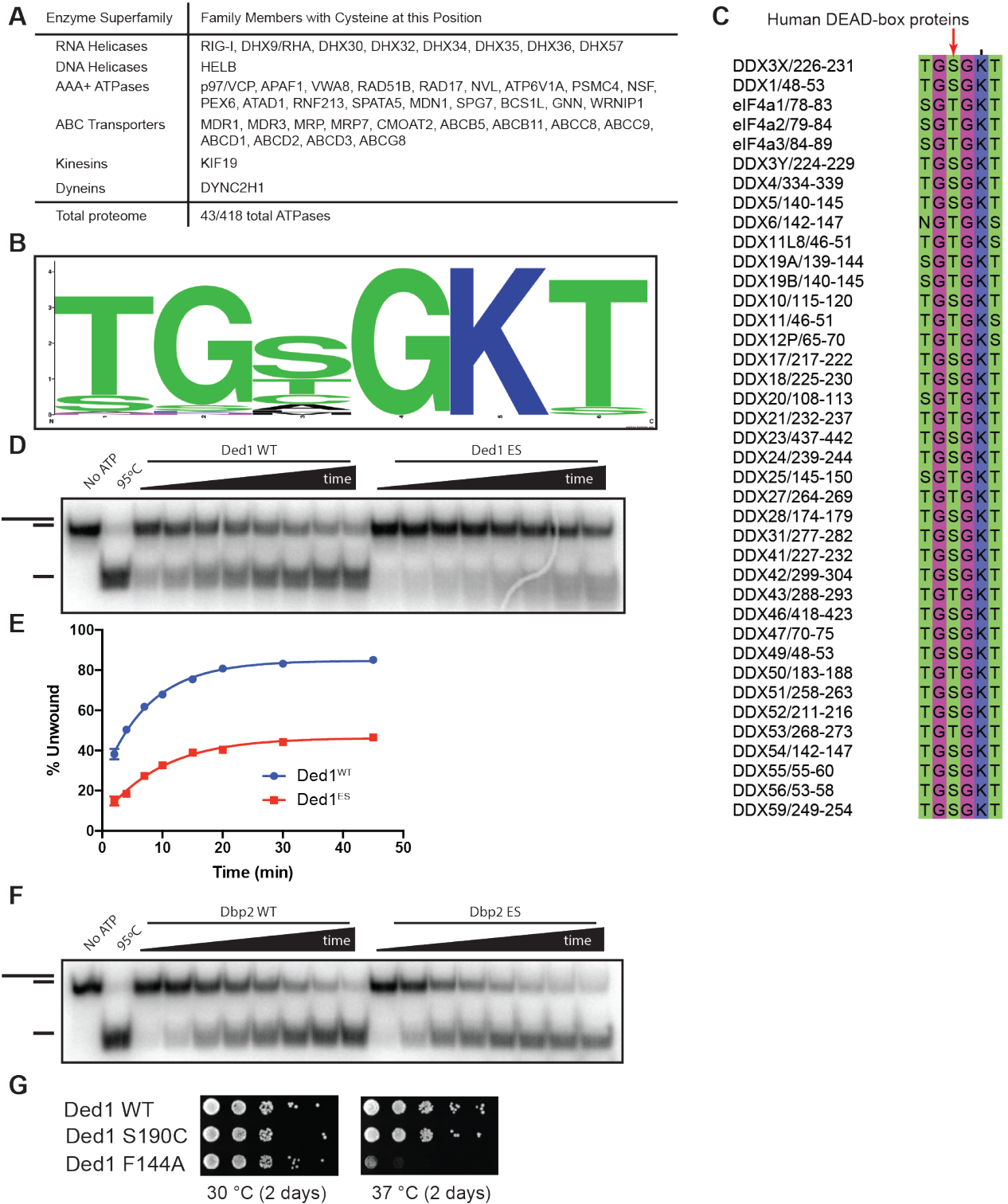
A. Overall structure of AMP-acrylamide bound to DDX3^{ES} (grey) aligned to AMP-bound DDX3 (green, PDB 5E7J). **B.** AMP-acrylamide bound to the nucleotide binding site of DDX3 with conserved hydrogen bonding interactions (black lines) with Y200, R202, and Q207. AMP-acrylamide $2mF_o-DF_c$ difference map is shown (black mesh, 1.5σ). **C.** Alignment of AMP-acrylamide- (grey) and AMP-bound (green) structures of DDX3 shows flattening of the P-loop in AMP-acrylamide-bound structure including an 9.4\AA shift in T226. **D.** Comparison of the coordination of the phosphoramidate of AMP-acrylamide (grey) to the coordination of the phosphate of AMP (green) by DDX3 (PDB 5E7J) (dashed lines are predicted hydrogen bonds).

Figure 3.5. AMP-acrylates inhibit electrophile-sensitive DEAD-box proteins.



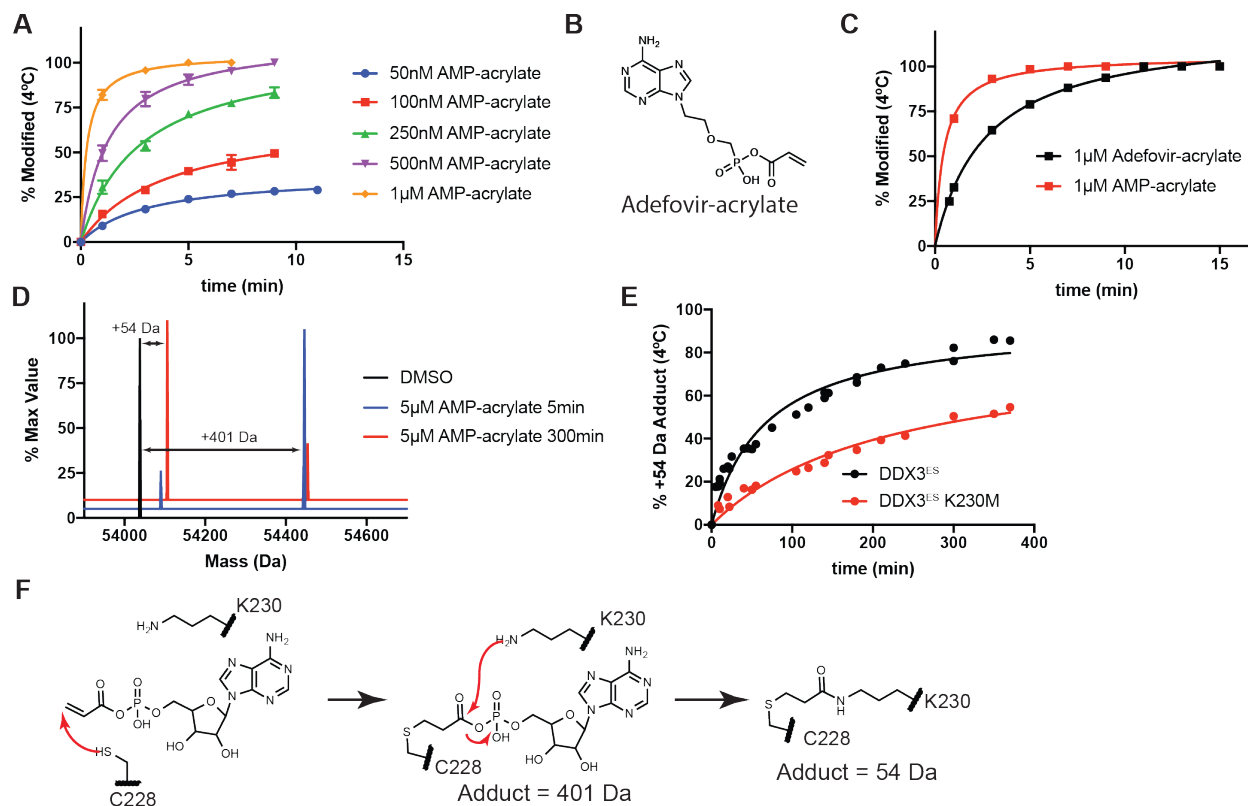
A and B. RNA duplex unwinding by DDX3^{WT} (**A**) and DDX3^{ES} (S228C) (**B**) in the presence of 50 μM and 12.5 μM AMP-acrylate. **C and D.** RNA duplex unwinding by DDX3^{ES} in the presence of serial dilutions of AMP-acrylate starting at 50 μM (**C**) with quantification of three independent experiments (**D**). **E and F.** Effect of AMP-acrylate of the endpoint (**E**) and rate (**F**) of RNA duplex unwinding by DDX3^{ES}.

Supplementary Figure 3.1. Development of an 'electrophile-sensitive' mutation in DEAD-box proteins.



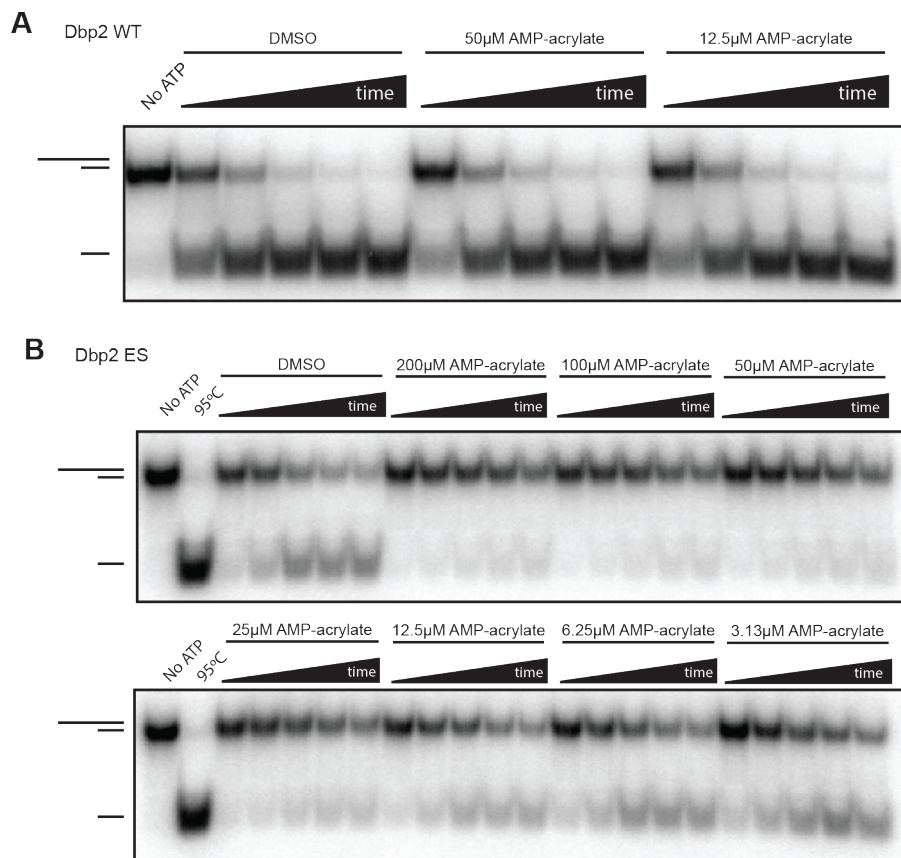
A. Identification of human ATPases from diverse enzyme families that natively express a P-loop cysteine. **B.** Structural alignment of the P-loop of all human RNA helicases. **C.** Alignment of the P-loop of all human DEAD-box proteins to identify position of electrophile-sensitive mutation (red arrow). **D.** Representative RNA duplex unwinding assay comparing the activity of Ded1^{WT} and Ded1^{ES} (S190C). **E.** Quantification of the fraction of RNA duplex unwound by Ded1^{WT} and Ded1^{ES}. **F.** Representative RNA duplex unwinding assay comparing the activity of Dbp2^{WT} and Dbp2^{ES} (S161C). **G.** Serial dilutions of log-phase cultures of budding yeast strains expressing wild-type, electrophile-sensitive (S190C), or analog-sensitive (F144A) Ded1 grown at permissive (30°C) or restrictive (37°C) temperatures.

Supplementary Figure 3.2. AMP-acrylates undergo two-step reaction with electrophile-sensitive DEAD-box proteins.



A. Percent modification of 125nM DDX3^{ES} (S228C) by various concentrations of AMP-acrylate at 4°C. **B.** Structure of Adefovir-acrylate. **C.** Comparison of the rate of modification of 250nM DDX3^{ES} by 1µM AMP-acrylate and 1µM Adefovir-acrylate at 4°C. **D.** Whole-protein mass spectrometry of DDX3^{ES} with DMSO or 5µM AMP-acrylate at 4°C showing formation of +401 Da adduct after 5min and +54 Da adduct after 300min. **E.** Quantification of the percentage of DDX3^{ES} and DDX3^{ES} K230M with +54 Da adduct over time. **F.** Proposed mechanism for the two-step reaction of DDX3^{ES} with AMP-acrylates. The electrophile-sensitive mutation (C228 in DDX3) undergoes a Michael-addition into the electrophilic beta-carbon of the acrylate of AMP-acrylate (first-step). AMP is then eliminated from this adduct by addition of a nucleophilic amino acid such as lysine (K230 in DDX3; second-step) or arginine to form the final product.

Supplementary Figure 3.3. AMP-acrylates inhibit electrophile-sensitive DEAD-box proteins.



A. RNA duplex unwinding by Dbp2 WT in the presence of 50 μ M and 12.5 μ M AMP-acrylate. **B.** RNA duplex unwinding by electrophile-sensitive Dbp2 (S161C) in the presence of serial dilutions of AMP-acrylate starting at 200 μ M.

Supplementary Table 3.1. Data collection and refinement statistics.

	DDX3 (132-607) – AMP-acrylamide
Data collection	
Space group	P 21 21 21
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	53.96, 101.09, 105.69
α , β , γ (°)	90, 90, 90
Resolution (Å)	73.05-3.00 (3.16-3.00) ^a
<i>R</i> _{merge} , <i>R</i> _{meas} , and <i>R</i> _{pim}	0.121 (1.089), 0.135 (1.201), 0.057 (0.500)
<i>I</i> / σ (<i>I</i>)	8.5 (1.6)
<i>CC</i> _{1/2}	0.996 (0.766)
Completeness (%)	100.0 (100.0)
Redundancy	5.4 (5.6)
Refinement	
Resolution (Å)	73.05-3.00
No. reflections	12100
<i>R</i> _{work} / <i>R</i> _{free}	0.2219/0.2667
No. atoms	
Protein	3378
Ligand/ion (specify/describe)	27
Water	0
<i>B</i> factors	
Protein	87.39
Ligand/ion	107.20
R.m.s. deviations	
Bond lengths (Å)	0.002
Bond angles (°)	0.50

^a Values in parentheses are for highest-resolution shell.

Supplementary Table 3.2: Yeast strains used in this study.

Strain	Genotype	Source
Ded1 WT	<i>MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0</i> ded1::KanMX pKJB001	Hilliker A et al. (2011) ¹³⁷
HA-Ded1 WT	<i>MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0</i> ded1::KanMX pKJB002	This study
HA-Ded1 S190C	<i>MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0</i> ded1::KanMX pKJB003	This study
Fal1 HetDip	<i>MATa/MATalpha his3Δ1/his3Δ1</i> <i>leu2Δ0/leu2Δ0 MET15/met15Δ0</i> <i>LYS2/lys2Δ0 ura3Δ0/ura3Δ0</i> FAL1/fal1::KanMX	GE Dharmacon (Clone ID: 23960)
Fal1 WT	<i>MATalpha his3Δ1 leu2Δ0 ura3Δ0</i> fal1::KanMX pKJB004	This study
HA-Fal1 WT	<i>MATalpha his3Δ1 leu2Δ0 ura3Δ0</i> fal1::KanMX pKJB005	This study
HA-Fal1 T71C	<i>MATalpha his3Δ1 leu2Δ0 ura3Δ0</i> fal1::KanMX pKJB006	This study
Dbp2 HetDip	<i>MATa/MATalpha his3Δ1/his3Δ1</i> <i>leu2Δ0/leu2Δ0 MET15/met15Δ0</i> <i>LYS2/lys2Δ0 ura3Δ0/ura3Δ0</i> DBP2/dbp2::KanMX	GE Dharmacon (Clone ID: 27822)
Dbp2 WT	<i>MATa his3Δ1 leu2Δ0 ura3Δ0</i> dbp2::KanMX pKJB007	This study
HA-Dbp2 WT	<i>MATa his3Δ1 leu2Δ0 ura3Δ0</i> dbp2::KanMX pKJB008	This study
HA-Dbp2 S161C	<i>MATa his3Δ1 leu2Δ0 ura3Δ0</i> dbp2::KanMX pKJB009	This study
Dbp5 HetDip	<i>MATa/MATalpha his3Δ1/his3Δ1</i> <i>leu2Δ0/leu2Δ0 MET15/met15Δ0</i> <i>LYS2/lys2Δ0 ura3Δ0/ura3Δ0</i> DBP5/dbp5::KanMX	GE Dharmacon (Clone ID: 21822)
Dbp5 WT	<i>MATalpha his3Δ1 leu2Δ0 ura3Δ0</i> dbp5::KanMX pKJB010	This study
HA-Dbp5 WT	<i>MATalpha his3Δ1 leu2Δ0 ura3Δ0</i> dbp5::KanMX pKJB011	This study
HA-Dbp5 T142C	<i>MATalpha his3Δ1 leu2Δ0 ura3Δ0</i> dbp5::KanMX pKJB011	This study

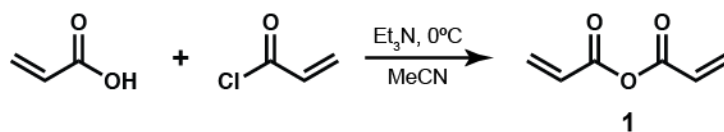
Supplementary Table 3.3: Plasmids used in this study.

Plasmid	Description	Vector	Insert	Source
pKJB001	Ded1 WT	pRS416 (<i>URA3/CEN</i>)	DED1 (-106 to +2727)	Hilliker A et al. (2011)
pKJB002	HA-Ded1 WT	pRS413 (<i>HIS3/CEN</i>)	DED1 (-106 to +2727)/N- terminal HA-tag	This study
pKJB003	HA-Ded1 S190C	pRS413 (<i>HIS3/CEN</i>)	DED1 S190C (-106 to +2727)/N-terminal HA-tag	This study
pKJB004	Fal1 WT	pRS416 (<i>URA3/CEN</i>)	FAL1 (-607 to +1826)	This study
pKJB005	HA-Fal1 WT	pRS413 (<i>HIS3/CEN</i>)	FAL1 (-607 to +1826)/N- terminal HA-tag	This study
pKJB006	HA-Fal1 T71C	pRS413 (<i>HIS3/CEN</i>)	FAL1 T71C (-607 to +1826)/N- terminal HA-tag	This study
pKJB007	Dbp2 WT	pRS416 (<i>URA3/CEN</i>)	DBP2 (-757 to +4107; Δ intron)	This study
pKJB008	HA-Dbp2 WT	pRS413 (<i>HIS3/CEN</i>)	DBP2 (-757 to +4107; Δ intron)/N-terminal HA-tag	This study
pKJB009	HA-Dbp2 S161C	pRS413 (<i>HIS3/CEN</i>)	DBP2 S161C (-757 to +4107; Δ intron)/N-terminal HA-tag	This study
pKJB010	Dbp5 WT	pRS416 (<i>URA3/CEN</i>)	DBP5 (-1029 to +2462)	This study
pKJB011	HA-Dbp5 WT	pRS413 (<i>HIS3/CEN</i>)	DBP5 (-1029 to +2462)/N- terminal HA-tag	This study
pKJB012	HA-Dbp5 T142C	pRS413 (<i>HIS3/CEN</i>)	DBP5 T142C (-1029 to +2462)/N-terminal HA-tag	This study

Chemical Synthesis

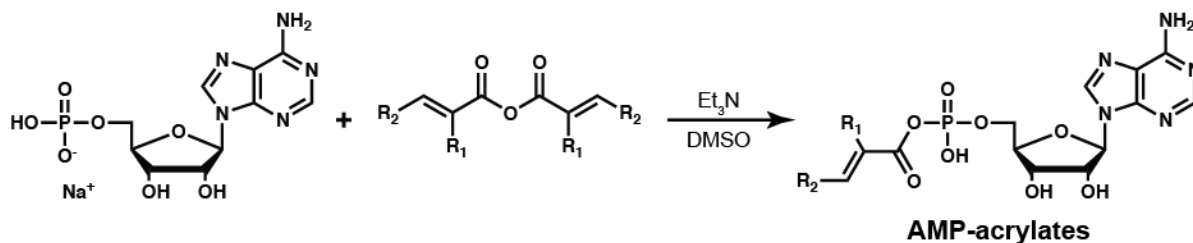
Materials obtained commercially were reagent grade and were used without further purification. Reactions were monitored by thin layer chromatography (TLC) and/or mass spectrometry (LC-MS) using a Waters Acquity UPLC/ESI-G2 XS QTOF. All NMR spectra were obtained on a Bruker 400 spectrometer.

Synthesis of intermediate 1



Acrylic anhydride (intermediate 1). Acrylic acid (Sigma-Aldrich, 205.7μL, 3.0mmol, 1.0eq) was added to a 20mL scintillation vial under Ar(g), followed by tetrahydrofuran (5mL) and triethylamine (418.1μL, 3.0mmol, 1.0eq). The reaction mixture was chilled to 0°C and acryloyl chloride (242.4μL, 3.0mmol, 1.0eq) in tetrahydrofuran (1mL) was added dropwise. The reaction was allowed to warm to room temperature and stirred overnight. The reaction was filtered, evaporated to dryness, then dissolved in ethyl acetate (30mL) and extracted with a solution of saturated sodium bicarbonate. The organic fraction was dried with sodium sulfate, filtered, evaporated to dryness and the resulting oil was used without further purification.

General procedure for the synthesis of AMP-acrylate, AMP-crotonate, and AMP-methacrylate



In a 4mL scintillation vial, adenosine 5'-monophosphate sodium salt (Sigma, 25mg, 0.0720mmol) was dissolved in DMSO (1mL) and triethylamine (100 μ L, 0.720mmol, 10eq) with sonication and stirring. In a separate vial, the appropriate symmetric acrylic anhydride (0.0792mmol, 1.1eq) was dissolved in DMSO (100 μ L) and added slowly to the reaction vial. After 30min, the reaction was quenched by addition of 1% formic acid (1mL) and purified by reverse phase HPLC (0-30% acetonitrile/water) to yield the pure compound.

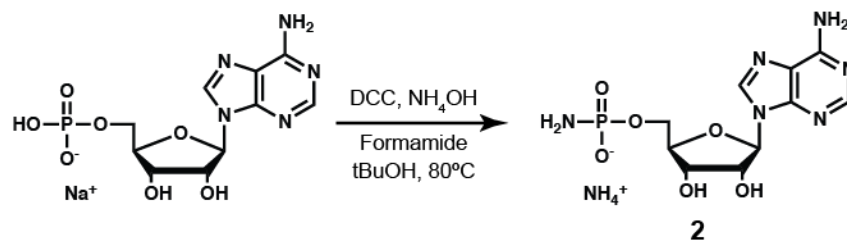
acrylic (((2*R*,3*S*,4*R*,5*R*)-5-(6-amino-9*H*-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl phosphoric) anhydride (AMP-acrylate). White solid, 20.6mg, 71.2% yield.

^1H NMR (400MHz, D_2O) δ 8.48 (s, 1H), 8.33 (s, 1H), 6.31 (dd, $J = 14.6, 3.5$ Hz, 1H), 6.08 (d, $J = 5.2$ Hz, 1H), 6.01 – 5.89 (m, 2H), 4.74 (t, $J = 5.2$ Hz, 1H), 4.45 (t, $J = 4.5$ Hz, 1H), 4.31 (s, 1H), 4.20 (tq, $J = 9.8, 6.4, 4.5$ Hz, 2H). ^{31}P NMR (162MHz, D_2O) δ -7.32. ^{13}C NMR (100MHz, D_2O) δ 163.57 (d, $J = 8.2$ Hz), 150.49, 148.44, 145.49, 142.17, 134.56, 127.25 (d, $J = 7.5$ Hz), 118.58, 88.00, 83.91 (d, $J = 8.7$ Hz), 74.23, 70.22, 65.59 (d, $J = 5.7$ Hz). $[\text{M}+\text{H}]^+$ m/z calculated 402.0809 (100%), 403.0843 (14%), observed 402.0800 (100%), 403.0861 (12%).

(((2R,3S,4R,5R)-5-(6-amino-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl phosphoric) (E)-but-2-enoic anhydride (AMP-crotonate). White solid, 13.66mg, 44.8% yield. ¹H NMR (400MHz, D₂O) δ 8.48 (s, 1H), 8.33 (s, 1H), 6.93 (m, 1H), 6.07 (d, J = 5.3 Hz, 1H), 5.63 (dt, J = 15.5, 1.9 Hz, 1H), 4.75 (t, J = 5.2 Hz, 1H), 4.45 (m, 1H), 4.31 (d, J = 2.9 Hz, 1H), 4.17 (pt, J = 5.6, 2.4 Hz, 2H), 1.74 (dd, J = 6.9, 1.6 Hz, 3H). ³¹P NMR (162MHz, D₂O) δ -7.27. ¹³C NMR (100MHz, D₂O) δ 163.80 (d, J = 8.2 Hz), 150.67, 150.36, 148.43, 145.36, 142.28, 121.00 (d, J = 7.4 Hz), 118.53, 88.01, 84.00 (d, J = 8.5 Hz), 74.17, 70.26, 65.53 (d, J = 5.8 Hz), 17.52. [M+H]⁺ m/z calculated 416.0966 (100%), 417.0999 (15%), observed 416.0961 (100%), 417.0987 (12%).

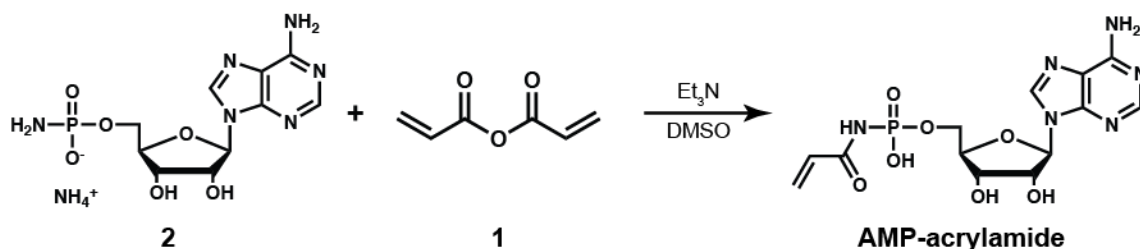
(((2R,3S,4R,5R)-5-(6-amino-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl phosphoric) methacrylic anhydride (AMP-methacrylate). White solid, 10.0mg, 32.9% yield. ¹H NMR (400MHz, D₂O) δ 8.46 (s, 1H), 8.32 (s, 1H), 6.07 (d, J = 5.3 Hz, 1H), 5.95 (s, 1H), 5.59 (s, 1H), 4.77 (t, J = 5.2 Hz, 1H), 4.46 (m, 1H), 4.30 (m, 1H), 4.19 (ttt, J = 8.4, 5.6, 2.5 Hz, 2H), 1.69 (s, 3H). ³¹P NMR (162MHz, D₂O) δ -7.16. ¹³C NMR (100MHz, D₂O) δ 164.73 (d, J = 8.7 Hz), 150.85, 148.51, 146.02, 142.06, 135.28 (d, J = 6.6 Hz), 128.86, 118.61, 87.97, 83.96 (d, J = 8.5 Hz), 74.04, 70.28, 65.65 (d, J = 5.9 Hz), 16.99. [M+H]⁺ m/z calculated 416.0966 (100%), 417.0999 (15%), observed 416.0961 (100%), 417.0987 (10%).

Synthesis of Intermediate 2



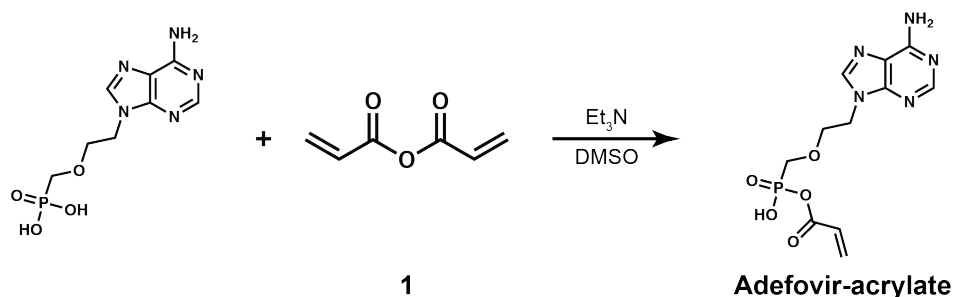
((2*R*,3*S*,4*R*,5*R*)-5-(6-amino-9*H*-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl hydrogen phosphoramidate (intermediate 2). The protocol for the synthesis of this intermediate was previously described²¹⁴. Adenosine 5'-monophosphate sodium salt (Sigma, 1041.7mg, 3.0mmol) was dissolved in ammonium hydroxide (2*N*, 7.5mL) and formamide (5mL). To this solution, a suspension of *N,N*-dicyclohexylcarbodiimide (DCC, 3095mg, 15.0mmol, 5eq) in tert-butanol (20mL) was added and the resulting two-phase reaction was heated to 80°C. After 2-3hr, the solution becomes homogenous and after an additional 7hrs, the reaction is allowed to cool overnight. Unreacted DCC was removed by filtration and washed with water and the resulting solution was evaporated under reduced pressure to remove volatiles. The resulting solution was extracted with diethyl ether, then the aqueous layer was evaporated to dryness. Acetone was added to the resulting oil and the ammonium salt of the product was isolated by filtration (white gummy solid; 884.3mg, 81.8% yield).

Synthesis of AMP-acrylamide



((2R,3S,4R,5R)-5-(6-amino-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl hydrogen acryloylphosphoramidate (AMP-acrylamide). In a 4mL intermediate 2 (50mg, 0.138mmol) was dissolved in DMSO (1mL) and triethylamine (48μL, 0.344mmol, 2.5eq) with sonication and stirring. In a separate vial, intermediate 1 (19.1mg, 0.151mmol, 1.1eq) was dissolved in DMSO (100μL) and added slowly to the reaction vial. After 30min, the reaction was quenched by addition of 1% formic acid (1mL) and purified by reverse phase HPLC (0-30% acetonitrile/water) to yield the pure compound (white solid, 4.93mg, 8.5% yield). ¹H NMR (400MHz, D₂O) δ 8.46 (s, 1H), 8.24 (s, 1H), 6.09 (d, J = 1.7 Hz, 1H), 6.05 (d, J = 5.4 Hz, 1H), 5.65 (d, J = 10.4 Hz, 1H), 4.73 (d, J = 5.3 Hz, 1H), 4.44-4.37 (m, 1H), 4.29 (d, J = 2.6 Hz, 1H), 4.09 (tddt, J = 11.7, 9.3, 5.8, 3.0 Hz, 2H). ³¹P NMR (162MHz, D₂O) δ -4.69. ¹³C NMR (100MHz, D₂O) δ 169.48, 153.02, 149.14, 148.70, 141.09, 129.98, 129.20, 118.59, 87.52, 83.91 (d, J = 9.0 Hz), 74.19, 70.37, 64.83 (d, J = 5.3 Hz). [M+H]⁺ m/z calculated 401.0969 (100%), 402.1003 (14%), observed 401.0957 (100%), 402.1006 (11%).

Synthesis of Adefovir-acrylate



acrylic ((2-(6-amino-9H-purin-9-yl)ethoxy)methyl)phosphonic anhydride (Adefovir-acrylate). In a 4mL scintillation vial, 9-[2-(Phosphonomethoxy)ethyl]adenine (PMEA, Adefovir) (AK Scientific, 20mg, 0.0732mmol) was dissolved in DMSO (1mL) and

triethylamine (102 μ L, 0.732mmol, 10eq) with sonication and stirring. In a separate vial, intermediate 1 (10.2mg, 0.0805mmol, 1.1eq) was dissolved in DMSO (100 μ L) and added slowly to the reaction vial. After 2hr, the reaction was quenched by addition of 1% formic acid (1mL) and purified by reverse phase HPLC (0-25% acetonitrile/water) to yield the pure compound (white solid, 18.4mg, 72% yield). ^1H NMR (400MHz, DMSO-d₆ + Et₃N) δ 8.16 (s, 1H), 8.12 (s, 1H), 7.16 (s, 2H), 6.21 (dd, J = 17.2, 1.9 Hz, 1H), 6.12-6.04 (m, 1H), 5.86 (dd, J = 10.1, 1.9 Hz, 1H), 4.27 (t, J = 5.2 Hz, 2H), 3.85 (t, J = 5.2 Hz, 2H), 3.60 (d, J = 8.1 Hz, 2H). ^{31}P NMR (162MHz, DMSO-d₆ + Et₃N) δ 9.67 (t, J = 8.0 Hz). ^{13}C NMR (100MHz, DMSO-d₆ + Et₃N) δ 163.24 (d, J = 7.8 Hz), 156.35, 152.73, 149.92, 141.66, 131.84, 130.66 (d, J = 3.7 Hz), 118.97, 70.09 (d, J = 9.2 Hz), 68.36 (d, J = 154.7 Hz), 43.01. $[\text{M}+\text{H}]^+$ m/z calculated 328.0805 (100%), 329.0839 (12%), observed 328.0799 (100%), 329.0854 (13%).

Chapter 4

Developing reversible chemical inhibitors of DDX3

Introduction

Owing to the success in the elucidation of novel biology using potent cell-active small molecule inhibitors of eIF4A,^{116,118,215} we sought to elaborate on our chemical genetic inhibitors of DEAD-box proteins and develop potent, cell-active ATP-competitive inhibitors of DDX3. We hope to distinguish these small molecules from other published DDX3 inhibitors that show limited efficacy in our hands¹³³ through rigorous biochemical characterization.

As discussed in chapter 2, our analog-sensitive strategy for the inhibition of DEAD-box proteins was unsuccessful because of poorly tolerated mutations and the low potency of analog-sensitive inhibitors.¹³³ We hypothesized that our difficulty in developing more potent inhibitors stemmed from a lack of potent ATP-competitive small molecules that target DEAD-box proteins. In contrast to kinases, for which chemical genetic strategies started with nanomolar binders that lacked specificity, our most potent screening hit was a millimolar binder that was optimized to bind ~10-fold more potently.¹³³ Our electrophile-sensitive chemical genetic strategy yielded potent and selective inhibitors of electrophile-sensitive DEAD-box proteins. However, these nucleotide-based small molecule inhibitors are not cell-permeable in their current form, and the phosphate/phosphonate moiety that limits their ability to transit across cell membranes is essential for binding activity.¹⁵² Therefore, additional screening for novel chemical scaffolds with drug-like properties and high affinity for the ATP-binding site of DEAD-box proteins was required.

The AAA+ ATPase superfamily (*ATPases associated with various cellular activities*) are a diverse set of cellular enzymes required in numerous cellular processes including proteolysis, DNA replication and recombination, and protein disaggregation.²¹⁶ AAA+ ATPases typically exist in homo- or heterohexameric complexes, and utilize ATP hydrolysis to drive cellular processes, typically as a subunit of large cellular complexes termed molecular machines.²¹⁷ AAA+ ATPases contain an ATP-binding site that is structurally related to other ATPases, including the DEAD-box proteins,^{23,218} although it does not rely on a Q-motif for nucleotide base recognition.²⁵ Therefore, ATP-competitive small molecule inhibitors of AAA+ ATPases may maintain some affinity for the nucleotide-binding site of DEAD-box proteins.

As with other ATPases, most chemical inhibitors of AAA+ ATPases are natural products that target an allosteric site.^{219,220} However, a few chemical inhibitors are ATP-competitive, including the small electrophile N-ethylmaleimide (NEM) that targets the ATP-binding site of the N-ethylmaleimide-sensitive factor (NSF).¹⁷⁸ N²,N⁴-dibenzylquinazoline-2,4-diamine (DBeQ, Figure 4.1A) was additionally identified as an ATP-competitive inhibitor of p97/VCP, a AAA+ ATPase required in the ubiquitin-proteasome system.¹³⁴ Derivatives of DBeQ inhibit the ATPase activity of p97 with a Ki of less than 25nM and have advanced to clinical trials in multiple myeloma.¹³⁵ An additional screening effort from Genentech identified a covalent, ATP-competitive inhibitor of p97 based on a saccharin scaffold.¹⁶⁴

In this study, we tested a series of DBeQ-derivatives for their efficacy as DDX3 inhibitors. We find that N⁴-(5-phenyl-1*H*-pyrazol-3-yl)quinazoline-2,4-diamines inhibit the RNA duplex unwinding activity of DDX3 at low micromolar doses, with dramatic SAR with substitution of the phenyl ring and robust SAR with substitution of the N²-position of the 2,4-diaminoquinazoline. We then identify TC347 as the most potent compound, and determine it is an ATP-competitive inhibitor of DDX3 WT.

Results

Small molecule screen identifies DBeQ-derivatives that inhibit the RNA duplex unwinding activity of DDX3 WT

DDX3, like other DEAD-box proteins, is capable of separating short RNA duplexes by local strand separation utilizing ATP hydrolysis.¹⁵³ A series of small molecule derivatives of DBeQ were synthesized to expand around the benzyl substituents at the N²- and N⁴-positions, and these compounds were tested for their capability to inhibit the ATP-dependent RNA duplex unwinding activity of DDX3 WT at 100μM in 2mM ATP (Figure 4.1B-C). The screen identified several derivatives that potently reduced the unwinding activity of DDX3, including TC329 and TC123R, although the chemical structure of several hit compounds raised the possibility of an aggregator phenotype.²²¹ Therefore, future experiments were conducted with 0.01% Triton X-100 to discourage compound aggregation.²²² These results indicate that derivatives of the p97/VCP inhibitor DBeQ inhibit the duplex unwinding activity of DDX3.

Structure-activity relationship of DBeQ-derivatives identifies N-(5-phenyl-1H-pyrazol-3-yl)quinazoline-4-amines as lead scaffold

Based on the results of the screen performed in Figure 4.1, we selected structurally related derivatives of TC329 and TC123R and analyzed their activity as inhibitors of DDX3 over a full RNA duplex unwinding time course (Figure 4.2A-B). Within the set of TC329-derivatives, we found a strong preference for a 2-aminomethylpyrrolidine substitution at the 2-position of the quinazoline (Figure 4.2A), while other substitutions showed little activity at 50 μ M. Within the set of TC123R-derivatives, we observed increased potency of phenyl-substituted pyrazoles at the N⁴-position of the 2,4-diaminoquinazolines, as TC327 and TC391 reduced the duplex unwinding activity of DDX3 by greater than 75% (Figure 4.2B). We then tested TC327 and TC391 at three lower doses and found that TC327 is the most potent, inhibiting the endpoint RNA duplex unwinding activity of DDX3 by >50% at 12.5 μ M (Figure 4.2C-D). Therefore, derivatives of DBeQ show structure-activity relationships against DDX3 and display dose-dependent inhibition of its duplex unwinding activity.

Structure-activity relationship of N⁴-(5-phenyl-1H-pyrazol-3-yl)quinazoline-2,4-diamines identifies TC347 as most potent inhibitor of DDX3

To further elaborate on the N⁴-(5-phenyl-1H-pyrazol-3-yl)quinazoline-2,4-diamine scaffold, we tested a set of compounds structurally related to TC327 and TC391. From this set, we first identified that substitution to the phenyl ring of N⁴-(5-phenyl-1H-pyrazol-3-yl)quinazoline-2,4-diamines shows dramatic SAR (Figure 4.3A-B). Whereas TC326 shows a similar potency to TC391 (Figure 4.2D), pyridine or furan derivatives show

markedly reduced potency (Figure 4.3A). Compound pairs TC386/TC387 and TC450/TC453 showed similarly dramatic results with small chemical perturbations (Figure 4.3B), although the dramatic potency of TC453 may be due to an aggregator phenotype, as naphthalene-based compounds are notoriously aggregation prone.²²¹

A series of eight small molecule derivatives of TC391 varying in N²-position substitution were also tested (Figure 4.3C, insert). In contrast to above, these structural changes result in more modest perturbations in activity (Figure 4.3C). The most promising of these compounds were TC347 (and its racemate TC337) and TC348, both of which contain potential hydrogen bond donors in a 3-(aminomethyl)pyrrolidine and 4-hydroxyazetidine, respectively. Compounds with aromatic or other saturated heterocycles at this position show more modest potency. Therefore, placement of a hydrogen bond donor off the N²-position of N⁴-(5-phenyl-1*H*-pyrazol-3-yl)quinazoline-2,4-diamines yields inhibitors of DDX3 with low micromolar potency.

TC347 is an ATP competitive inhibitor

Since substituted aromatic small molecules can be either allosteric or ATP-competitive inhibitors of p97/VCP,^{135,164} we sought to understand the nature of the binding of TC347 to DDX3. We performed a three-point dose response of TC347 (Figure 4.4A) at varying concentrations of ATP (50μM-10mM) and found that TC347 displays ATP-competitive binding (Figure 4.4B-C). Visually, this can be seen by the decreased efficacy of TC347 at all doses at 10mM ATP as compared to 200μM ATP (Figure 4.4B) as well as by a

Lineweaver-Burk plot, which shows increasing doses of TC347 have the same y-intercept on a double-reciprocal plot (Figure 4.4C).

In a complementary approach, we also monitored the binding of AMP-acrylamide (Figure 3.3A-B) to DDX3^{ES} (S228C) in the presence of TC347 or inactive analog TC355 (Figure 4.4A). We observed that TC347 competes with AMP-acrylamide binding to C228 of DDX3^{ES} down to doses as low as 12.5 μ M, indicating they bind to the same or similar sites on DDX3, while TC355 shows little competition up to 50 μ M (Figure 4.4D-E). Since the binding position of AMP-acrylamide has been characterized by mutational studies and crystallography to be the ATP-binding site of DDX3 (Figure 3.3B, 3.4), this indicates that TC347 is an ATP-competitive DDX3 inhibitor.

Moving towards more drug-like irreversible inhibitors of DDX3^{ES}

Based on the potency of TC347, we hypothesized that it binds DDX3 with a $K_i \sim 10\mu$ M, which makes it a tighter binder than AMP (Figure 2.3A). Therefore, proper placement of an electrophile off the TC347 scaffold may produce an electrophile-sensitive DEAD-box protein inhibitor with improved drug-like properties to complement the electrophile-sensitive mutation discussed in Chapter 3. Based on previous results that the C8 of the quinazoline ring points towards the P-loop of AAA+ ATPases (data not shown), we synthesized KBELN2-55 and KBELN2-74B to point an electrophile towards that position (Figure 4.5A). KBELN2-74A was synthesized during this process as a side-product. Upon incubation of these compounds with DDX3^{ES}, only KBELN2-74A showed

evidence of a covalent adduct between the protein and the small molecule (Figure 4.5A), indicating our initial hypothesis of the binding pose of TC347 was incorrect.

Previous screening efforts to identify a drug-like electrophile-sensitive DEAD-box protein inhibitor uncovered KB3-101 as a semi-potent binder of DDX3^{ES} (data not shown). We hypothesized that the positioning of the electrophile in KBELN2-74A may indicate that substitution of the C2-position of the quinazoline of KB3-101 may yield more potent binders to DDX3^{ES}. To test this, we synthesized a series of KB3-101-derivatives with substitutions at the C2-position similar to TC347 and TC348 (Figure 4.5B). Although the 4-hydroxyazetidine-derivative of KB3-101 (KBELN4-53) displayed an increased rate of labeling to DDX3^{ES}, these results were not dramatic, and other derivatives showed decreased potency compared to the parent compound (Figure 4.5C). This indicates that further chemical derivatization is necessary to convert TC347 into a high-affinity electrophile-sensitive DEAD-box protein inhibitor.

Discussion

We have identified that derivatives of DBeQ, an inhibitor of the AAA+ ATPase p97/VCP, inhibit the duplex unwinding activity of DDX3 WT (Figure 4.1). A screen of DBeQ-derivatives found a preference for N⁴-(5-phenyl-1*H*-pyrazol-3-yl)quinazoline-2,4-diamines and a hydrogen bond donor off the N²-position (Figure 4.2 and 4.3). The most potent of these compounds, TC347, shows ATP-competitive inhibition of DDX3 and competes with AMP-acrylamide binding to DDX3^{ES}, indicating it binds in the nucleotide-binding site of DDX3 (Figure 4.4).

To our knowledge, this is the first report of an ATP-competitive inhibitor of DDX3, and the third report of an ATP-competitive inhibitor of any DEAD-box protein, after the recent publications of natural products elisabatin A and allolaurinterol (Figure 1.3A) targeting eIF4A and indole-2-carboxylic acids targeting eIF4A3.^{103,128} In contrast to these studies, the small molecules identified in this work show activity at low micromolar doses in a stringent RNA duplex unwinding assay, as opposed to an ATPase assay. Owing to the large enzyme:RNA substrate ratio used in the duplex unwinding assay, as well as the high concentration of ATP, we feel compound activity in this assay is more indicative of activity in cellular assays since small molecules must have high occupancy/potency to show inhibition in this setting.¹³³

Despite the promising biochemical data, movement towards more complex biochemical and cellular studies with these compounds will be complicated by the incomplete picture of their activity on other cellular ATPases. From previous work testing DBeQ-derivatives against other human AAA+ ATPases, we know that they show varying selectivity towards this family (data not shown). It is therefore likely that the compounds identified in this work also target a subset of AAA+ ATPases at equal or greater potency than DEAD-box proteins. Owing to the hypothesized difference in the binding mode of these compounds between DEAD-box proteins and AAA+ ATPases, there are likely chemical modifications that can be made to TC347 to antagonize binding to AAA+ ATPases while preserving activity in the DEAD-box family. A structural understanding of the binding of TC347 to DDX3 will aid in these studies.

To combat this assumed lack of selectivity, one could generate TC347-resistant alleles of DEAD-box proteins and leverage these mutants to determine the 'background signal' of other ATPases in a focused biochemical question. Structural studies with TC347 will aid in this process, although the invariant nature of the DEAD-box protein nucleotide-binding site may complicate these efforts.²⁵ Additionally, appending an electrophile off these novel binders that specifically targets the engineered cysteine of electrophile-sensitive DEAD-box proteins would immediately provide a huge boost in potency towards the target, and potentially antagonize binding to non-cysteine-containing ATPases. Our initial efforts towards this goal were unsuccessful (Figure 4.5), although structural studies on TC347 may provide insight to allow for proper electrophile placement.

An exciting possibility from these results is the idea that all ATPases, not just p97/VCP and NSF, can be targeted with ATP-competitive small molecules. Although we anticipate the identification of selective small molecules against the greater than 400 human ATPases will pose a significant challenge, the generation of non-selective but potent chemical inhibitors may provide the first step towards this goal. Further synthesis and characterization of the small molecules described in this work is a first step towards an understanding of ATP-competitive DEAD-box protein inhibitors.

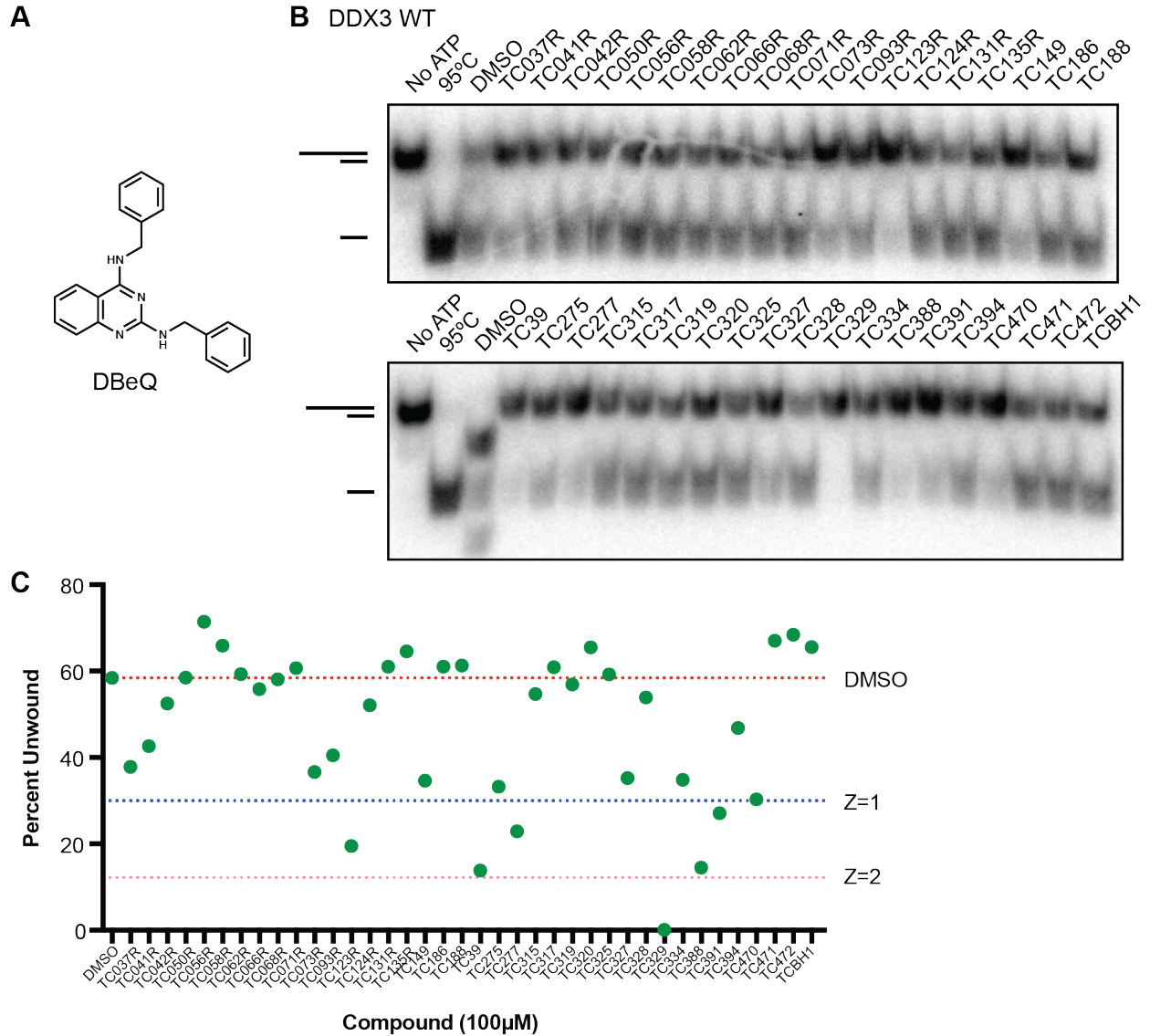
Materials and Methods

Recombinant protein purification. DDX3 (132-406) and DDX3 (132-607) were expressed and purified as previously described.¹³³

RNA duplex unwinding assays. RNA duplex unwinding assays were performed as previously described¹⁵³ using 1nM duplex RNA, 2mM ATP/MgCl₂, and 1μM DDX3 (132-607). Reaction buffer contained 20mM HEPES pH 7.5, 100mM NaCl, 0.5mM MgCl₂, 1mM TCEP, 0.01% (v/v) Triton X-100, and 5% glycerol. The sequences of the RNA duplex strands are 5'-GCUUUACGGUGC-3' and 5'-GAACAACAACAACAACCAUGGCACCGUAAAGC-3'. RNA, protein and inhibitors were pre-incubated in reaction buffer for 5min prior to addition of ATP-Mg. Results were quantified using ImageJ (NIH).²⁰⁷

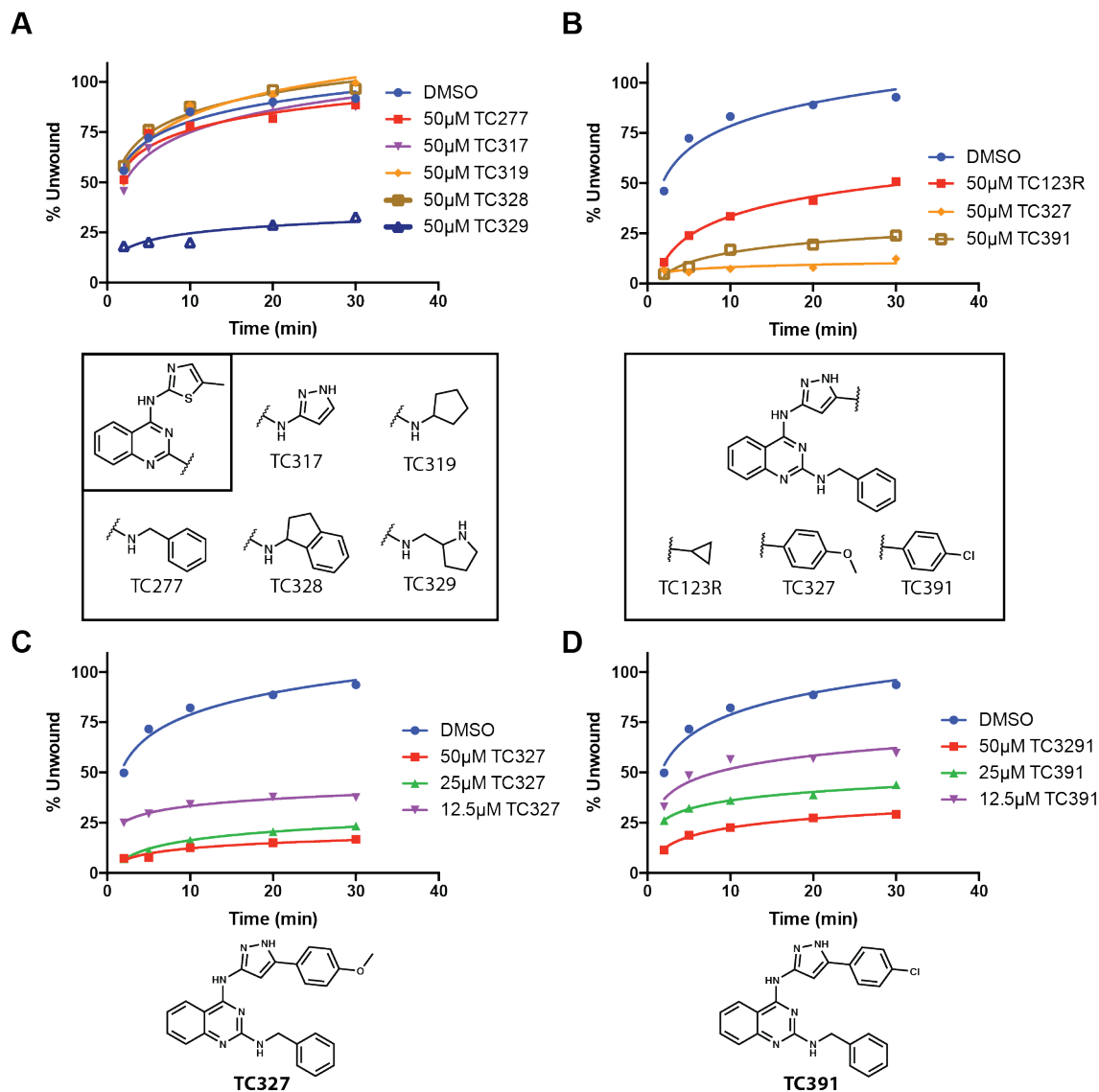
Mass spectroscopy assay. DDX3 D1 (132-406) S228C (250nM) or DDX3 D1D2 (132-607) S228C (500nM) in 10mM Tris pH 7.5, 100mM NaCl and 1mM MgCl₂ was incubated with compounds at the indicated doses at room temperature. For competition experiments, compounds were pre-incubated with the protein in reaction buffer for 2min, followed by addition of 10μM AMP-acrylamide. The extent of modification at various time points was determined by whole protein mass spectrometry using a Waters Acquity UPLC/G2-XF QTOF. Reported data points are single-replicates. Curves are fit using Prism (GraphPad).

Figure 4.1. Screen of DBeQ-derivatives identifies small molecule inhibitors of DDX3.



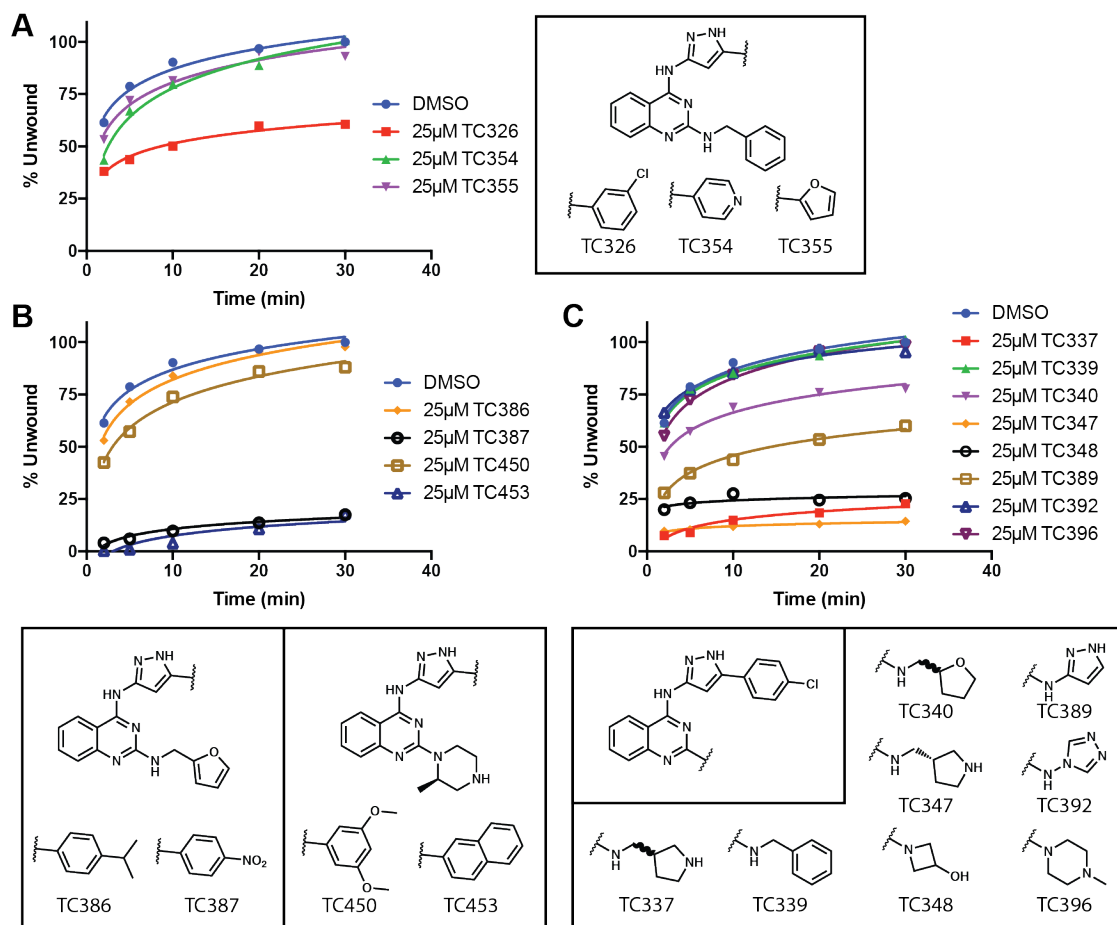
A. Structure of p97/VCP inhibitor DBeQ. **B and C.** Single dose (100µM) screen of select AAA+ ATPase-inhibitors based on DBeQ against DDX3 WT (**B**), quantified in (**C**) with lines corresponding to standard deviations from the DMSO control (Z=1 – one standard deviation, Z=2 – two standard deviations).

Figure 4.2. Structure-activity relationship (SAR) of DBeQ-derivatives and DDX3.



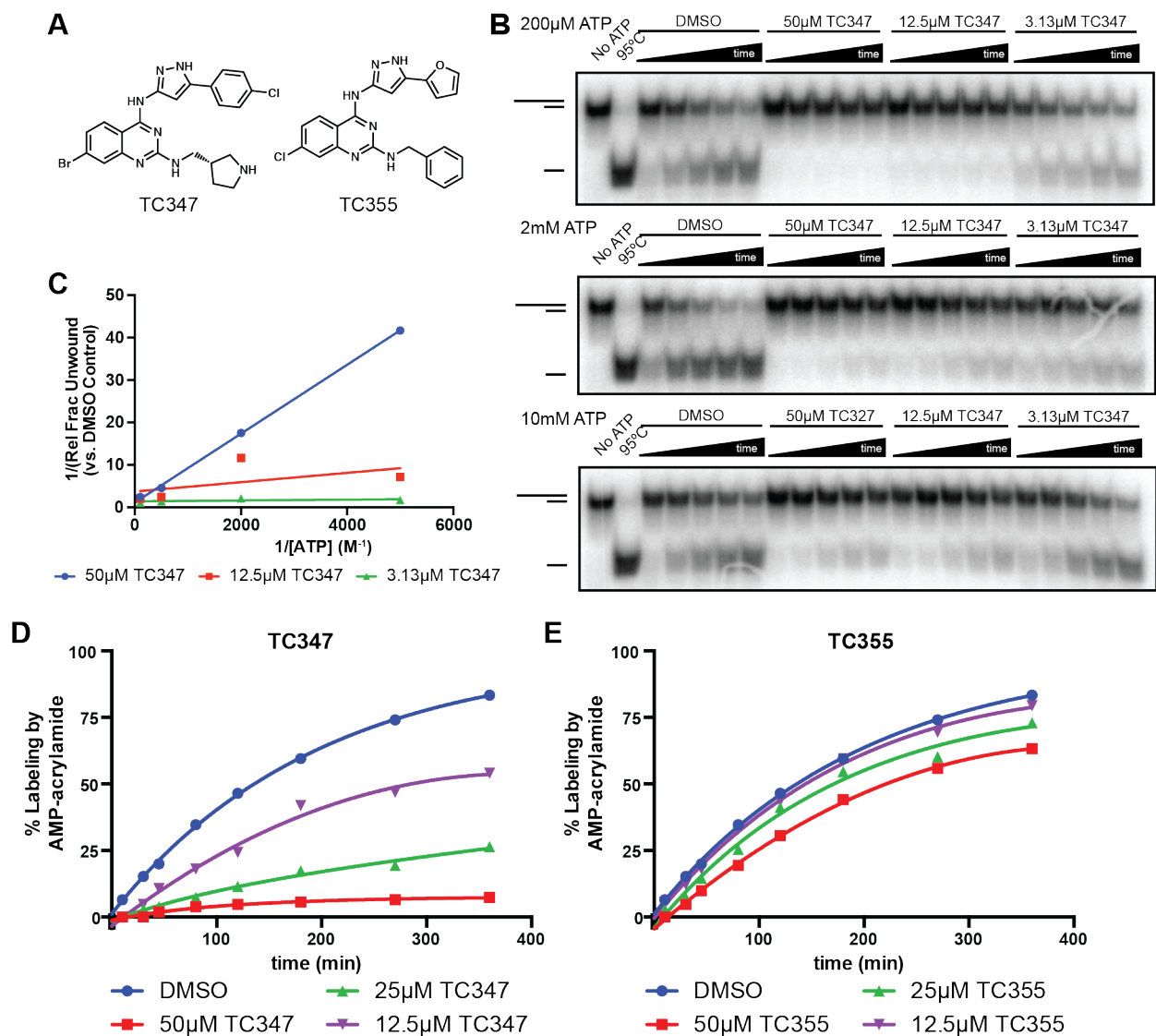
A. The effect of methylthiazole derivatives of DBeQ on the RNA-duplex unwinding activity DDX3 WT, with structures in insert. **B.** The effect of substituted pyrazole derivatives of DBeQ on the RNA-duplex unwinding activity DDX3 WT, with structures in insert. **C and D.** TC327 (**C**) and TC391 (**D**) inhibit the RNA duplex unwinding activity of DDX3 WT in a dose responsive manner.

Figure 4.3. Structure-activity relationship (SAR) of TC327/TC391-derivatives and DDX3.



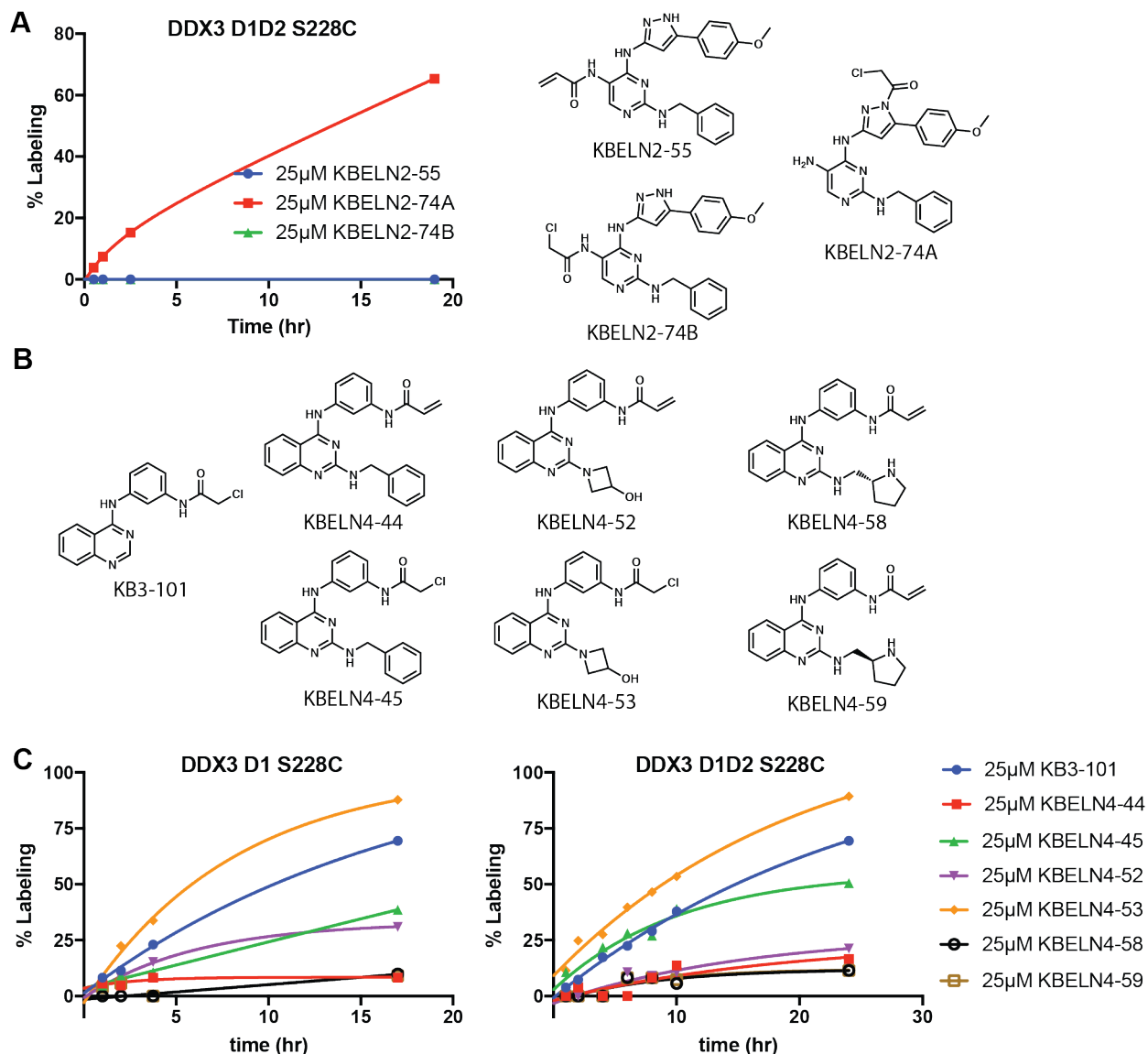
A and B. Substitutions at the phenyl ring of N^4 -(5-phenyl-1*H*-pyrazol-3-yl)quinazoline-2,4-diamines dramatically alter inhibition of the RNA duplex unwinding activity of DDX3 WT. **C.** Substitutions at the 2-position of N^4 -(5-(4-chlorophenyl)-1*H*-pyrazol-3-yl)quinazoline-2,4-diamines show robust SAR against DDX3 WT and yield TC347 and TC348 as lead compounds.

Figure 4.4. TC347 is an ATP-competitive DDX3 WT inhibitor.



A. Chemical structures of TC347 and control compound TC355. **B.** Monitoring the duplex unwinding activity of DDX3 WT in the presence of a dose response of TC347 (50 μ M, 12.5 μ M, 3.13 μ M) at varying doses of ATP (0.2mM, 2mM, 10mM). **C.** Lineweaver-Burk plot of the effect of ATP concentration on the duplex unwinding activity of DDX3 WT. **D and E.** Competition of labeling of DDX3^{ES} (S228C) by AMP-acrylamide in the presence of doses of TC347 (**D**) or TC355 (**E**).

Figure 4.5. Moving towards more drug-like irreversible inhibitors of DDX3^{ES}.



A. KBELN2-74A, but not derivatives with an altered position of the electrophile, labels DDX3^{ES}. **B.** Derivatives of KB3-101 with substituents at the C2-position of the quinazoline ring. **C.** Derivatives of KB3-101 with substitutions at the C2-position of the quinazoline ring show variable rates of labeling of domain 1 only (D1) of DDX3^{ES} and the full helicase domain (D1D2) of DDX3^{ES}.

Chapter 5

Analyzing CDK4 complexes to expand the efficacy of clinical CDK4/6 inhibitors

Introduction

The cell cycle is an evolutionarily conserved process across all eukaryotic cells by which DNA replication and cell division occurs.²²³ The cell cycle consists of two gap phases, G₁ and G₂, S-phase, and M-phase; cells outside of the cell cycle, typically quiescent or resting cells, are maintained G₀. Interphase encompasses G₁, S-phase and G₂; cellular components are duplicated during G₁, DNA replication occurs during S-phase, and additional protein synthesis occurs and replicated DNA is error-checked during G₂. The stages of mitosis and cytokinesis occur during M-phase (Figure 5.1A).²²⁴ These coordinated events are tightly regulated by extracellular ligands and intracellular signaling whereby information from environmental cues and nutrient levels is integrated into a binary decision of whether a cell has the proper resources for division.²²⁵ As such, in states of rapid unrestrained proliferation such as cancer, there are fundamental alterations in the genetic and signaling control of the cell cycle.²²⁵

Cyclins and cyclin-dependent kinases (CDKs) are intimately involved in the regulation of the cell cycle.²²⁶ Cyclins, as their name suggests, are a class of protein that are synthesized and destroyed during different stages of the cell cycle. D-type cyclins are abundant during G₁, E-type cyclins during the G₁/S-phase transition, A-type cyclins during S-phase to the G₂/M boundary, and B-type cyclins from early G₂ until cytokinesis.²²³ Cyclins function as obligate allosteric activators of cyclin-dependent kinases, which are serine/threonine kinases that phosphorylate cell cycle proteins to maintain the unidirectional nature of the cell cycle.²²³ CDK4 and CDK6 associate with D-type cyclins during G₁, CDK2 with E- and A-type cyclins during G₁/S-phase and S-

phase/G₂, respectively, and CDK1 associates with A- and B-type cyclins during G₂/M (Figure 5.1A).²²⁶ Interestingly, CDK1 is the only essential CDK in mammalian cells.²²⁷

The dysregulation of CDK activity is commonly observed in hyperproliferative disorders such as cancer.²²⁷ This is often through loss of negative regulators of cell cycle progression, including loss of CDKN2A, which encodes p16^{INK4A}, a small protein inhibitor of CDK4/6, or loss of CDKN1A or CDKN1B, which encode p21^{Cip1} and p27^{Kip1}, respectively, small protein inhibitors of CDK4/6 and CDK2.²²⁷ The tumor suppressor retinoblastoma protein (Rb), the major downstream target of cyclin D-CDK4/6 and a negative regulator of the E2F family of transcription factors (Figure 5.1B), is also commonly mutated or deleted in cancer.²²⁷ Overexpression and amplification of D- and E-type cyclins is also observed.²²⁸

Owing to their central role in cell cycle progression, small molecule inhibitors targeting CDKs are highly sought after as anti-cancer therapeutics.²²⁹ Flavopiridol, a polyhydroxylated flavonoid and the first identified pan-CDK inhibitor, was discovered in 1992 as a potent inhibitor of cancer cell growth *in vitro*.²³⁰ More selective inhibitors of many cell cycle CDKs have been developed and advanced into clinical trials, although there has been limited clinical success.^{227,229} More recently, several selective CDK4/6 inhibitors (palbociclib, ribociclib, and abemaciclib from Pfizer, Novartis/Astex, and Eli-Lilly, respectively) were clinically approved for the treatment of ER-positive breast cancer (Figure 5.1C).²³¹ However, these compounds are only approved as co-therapies

with ER-antagonists or aromatase inhibitors,²³²⁻²³⁴ and despite being low nanomolar biochemical inhibitors of CDK4/6, have shown limited efficacy as monotherapies.²³⁵

To better understand the lack of efficacy of CDK4/6 inhibitors as monotherapy, we confirmed that the most selective clinically approved CDK4/6 inhibitor, ribociclib, shows limited efficacy as a monotherapy in breast and non-breast cell culture cells. We then identify that only a portion of cellular CDK4 is competent to bind an irreversible kinase active-site probe and that this fraction of CDK4 varies across different cell types.

Cellular fractionation studies revealed CDK4 exists in several high molecular weight complexes within the cell, and these high molecular weight complexes appear to shield CDK4 from small molecule inhibitors. Lastly, we find that knockdown of Cyclin D1 improves the efficacy of ribociclib in cell culture, and synthesize a novel ribociclib analog that disrupts the CDK4-cyclin D1 complex leading to cyclin D1 degradation.

Results

Ribociclib shows limited efficacy as a monotherapy in cell culture

To confirm clinical observations of low efficacy of CDK4/6 inhibitors as monotherapy,²³⁵ we tested ribociclib in two breast cancer cell lines, MCF7 and BT549, and three KRas G12C cell lines, H23, H358, and MIA PaCa-2. In all cell lines, ribociclib showed limited efficacy, with IC50s greater than 10 μ M after three days of culture (Figure 5.1D).

Notably, CDK4/6 inhibitors display increased efficacy over longer assay times (Pfizer reported treatments up to five days for cellular assays),²³⁶ possibly suggesting high on-target potency but limited target accessibility. Ribociclib was especially ineffective in

BT549 cell lines, as previously reported.²³⁷ This matches clinical data since BT549 cells are Rb-null, and therefore should not be dependent on CDK4/6 activity for cell cycle progression.

Kinase active site probe, XO44, partially labels CDK4 in MCF7, but not BT549, cells

To better understand the limited efficacy of ribociclib in these cell lines, we tested the hypothesis that low ribociclib activity was due to poor target accessibility, as previously suggested for palbociclib.²³⁷ To test this, we utilized a cell-active kinase active site probe (XO44), that forms a covalent bond with the conserved catalytic lysine of protein kinases.²³⁸ We added XO44 to proliferating MCF7 cells, used click chemistry to attach TAMRA-N₃, and observed the proportion of CDK4 bound by XO44 by gel shift (Figure 5.2A). A kinase that is unlabeled by XO44 will run at its expected molecular weight, a kinase fully labeled by XO44 will display a full gel shift, and a kinase partially labeled by XO44 will run as a double band.

We observed that CDK4 is partially labeled by XO44 in MCF7 cells (Figure 5.2B). This labeling is competed by a non-clickable version of XO44 (KL-E2-94), the non-specific kinase inhibitor staurosporine, and ribociclib (Figure 5.2B). This demonstrates that a portion of CDK4 within MCF7 cells is incapable of binding small molecule inhibitors related to XO44, and that the population of CDK4 bound by ribociclib includes the entire portion bound by XO44. In contrast, we observed no gel shift in CDK4 after similar treatment of BT549 cells with XO44, while a complete gel shift was seen in Aurora B

(Figure 5.2C). This indicates that the entirety of CDK4 within BT549 cells exists in a state that is inaccessible to small molecule inhibitors. To confirm the partial labeling of CDK4 in MCF7 cells was not a result of the dose of XO44, we treated MCF7 cells with increasing doses of XO44. We observed that XO44-labeling of CDK4 saturates at ~50% of total at single-digit micromolar doses, whereas XO44-labeling of CDK2 saturates at ~90% (Figure 5.2D). Therefore, using a kinase active site probe, we determined that half of the CDK4 in MCF7, and all of the CDK4 in BT549 cells, exists in a state that is incompetent of binding small molecule inhibitors.

CDK4 complexes vary across cell lines

Owing to its central role in initiating cell cycle progression, CDK4 activity is tightly regulated through numerous interacting partners. Hsp90 and its kinase-specific co-chaperone Cdc37 are required for the initial folding of most human kinases.²³⁹ As a strong client of Hsp90, CDK4 is unstable as a monomer, and thus protein co-factors are thought to be required for the displacement of CDK4 from the chaperone system.²⁴⁰ D-type cyclins have been hypothesized to perform this function, and ATP-competitive inhibitors, including palbociclib and ribociclib, can also antagonize the CDK4-Cdc37 interaction.²⁴¹ Additionally, CDK4 forms tight interactions with small protein inhibitors of the INK4 family, including p16^{INK4A} and p18^{INK4C},^{242,243} and Cip/Kip family, including p21^{Cip1} and p27^{Kip1}.^{244,245} It has been reported that the CDK4-Cyclin D-p27^{Kip1} complex contains the catalytically active CDK4.²⁴⁶

To observe the native complexes in which CDK4 exists in MCF7 and BT549 cells, protein lysates from these cell lines were subjected to native gel filtration to separate cellular complexes by size.²⁴⁶ The fractionated lysates were then analyzed by western blotting (Figure 5.3A). We observed that in MCF7 cell lysates, CDK4 is found in a ~500kDa complex with Hsp90 and Cdc37, in a ~200kDa complex with Cyclin D1 and p27^{Kip1}, and in a ~60kDa complex with INK4 family members (Figure 5.3B). As previously reported, no CDK4-Cyclin D dimers are observed,²⁴⁶ suggesting this complex may not exist without additional cofactors. In contrast, in BT549 cell lysates CDK4 is found entirely in a ~60kDa complex with INK4 family members. Since BT549 cells are Rb-null, this indicates that the cell can sense its lost requirement for CDK4/6 in cell cycle progression, and inactivates CDK4 with INK4 proteins. CDK2 additionally shows altered complex formation between these cell lines – it is located predominantly as a monomer and in a >200kDa complex in MCF7 cell lysates, while it appears to exist as a monomer, in ~60kDa complexes, and in ~150kDa complexes in BT549 lysates (Figure 5.3B). Therefore, CDK2 and CDK4 are dynamically distributed into high-molecular weight complexes that differ between MCF7 and BT549s cells.

CDK4 complexes are resistant to active site probe binding

To determine which CDK4 and CDK2 complexes are capable of binding small molecule active-site inhibitors, we treated cells with XO44 and subjected these lysates to gel filtration chromatography and then performed the click reaction on individual fractions (Figure 5.3C). We observed that after XO44 treatment, CDK4 is no longer localized into ~500kDa complexes with Hsp90/Cdc37, likely because ATP-competitive inhibitors

antagonize the CDK4-Cdc37 interaction.²⁴¹ Instead, CDK4 is localized predominantly in ~200kDa complexes, ~60kDa complexes, and as a monomer. Only the monomeric CDK4 displays a gel shift, indicating that CDK4-cyclinD-p27^{Kip1} complexes are insensitive to active site inhibitors (Figure 5.3C). XO44 displays a similar labeling pattern of CDK2 – high molecular weight complexes are insensitive while monomeric CDK2 is fully labeled. These results indicate that active site inhibitors mature CDK4 from Hsp90/Cdc37-bound complexes and stabilize a small molecule-bound monomer. Additionally, we found that the CDK4-cyclin D-p27^{Kip1} complex is resistant to small molecule binding, and therefore may represent a signaling competent, but active site inhibitor-resistant, state of CDK4.

Knockdown of cyclin D1 improves efficacy of ribociclib in cell culture

If the CDK4-cyclin D-p27^{Kip1} complex is active in the presence of CDK4 active-site inhibitors, we hypothesized that reducing cyclin D protein levels would decrease the abundance of this complex and sensitize cells to CDK4/6 inhibitors. Therefore, we used siRNA to knockdown cyclin D1 or CDK4 in NIH3T3 cells and tested for increased cellular sensitivity to ribociclib. We observed that NIH3T3 cells show increased sensitivity to ribociclib in combination with cyclin D1 knockdown in both cell growth (Figure 5.4A), cell cycle arrest (Figure 5.4B), and apoptosis (Figure 5.4C). A western blot analysis of cellular signaling also showed increased p-Rb sensitivity to ribociclib in combination with cyclin D1 knockdown (Figure 5.4D). Knockdown of CDK4 had little effect on NIH3T3 cell sensitivity to ribociclib (Figure 5.4A-D). These results indicate that knockdown of cyclin D1 levels increase cellular sensitivity to ribociclib, and suggests

that targeting cyclin D1 for degradation may have additive effects with CDK4 active site inhibitors.

Synthesis of type II analogs of ribociclib

Our research group has previously had success developing kinase active-site inhibitors to antagonize kinase protein-protein interactions. Briefly, one of these small molecules stabilized an inactive α C-helix-out conformation of Aurora A to disrupt its stabilization of Myc-N and target it for degradation.²⁴⁷ We hypothesized that similar ‘type II’ inhibitors of CDK4 may disrupt its interaction with cyclin D, altering its incorporation into high molecular weight complexes and potentially leading to its degradation.²⁴⁸ Therefore we generated a series of twenty-nine derivatives of ribociclib (named CDK4i) with various type II kinase inhibitor elements (Figure 5.5A).²⁴⁹

Ribociclib analogs disrupt the CDK4-cyclin D1 interaction and promote cyclin D1 degradation in cells

To screen these small molecules for their capability to disrupt the CDK4-cyclin D interaction, we generated a reversible fluorogenic reporter of this interaction using split UnaG.²⁵⁰ This technology allows for the direct monitoring of the CDK4-cyclin D1 interaction in cells using live imaging (Figure 5.5B). While ribociclib treatment does not affect the CDK4-cyclin D1 interaction, we found that several CDK4i were capable of disrupting this complex after two hours at 10 μ M (Figure 5.5C-D). The most robust decrease in fluorescence was observed with CDK4i-8 and CDK4i-16, which are structurally related. To confirm these results are not an artifact of the fluorogenic

reporter assay, we tested these compounds in NIH3T3 cells and found that both compounds induce a marked decrease in cyclin D1, but not cyclin E, after 24 hours (Figure 5.5E). These results demonstrate that active site inhibitors of CDK4 can be expanded to induced cyclin D1 degradation through increased steric bulk pointing towards the α C-helix.

Discussion

We have demonstrated that the active-site inhibitor of CDK4/6, ribociclib, shows limited efficacy as a monotherapy in cell culture (Figure 5.1). We determined that only a portion of CDK4 is accessible to bind an active site probe, XO44, and that this fraction is different in MCF7 and BT549 cell lines, in a manner that is consistent with their genetic background (Figure 5.2). Using lysate fractionation, we found that CDK4 exists in high molecular weight complexes, and that CDK4 in complex with cyclin D and p27^{Kip1} is incapable of binding XO44 (Figure 5.3). We then used siRNA knockdown of cyclin D1 to show that reducing cyclin D protein levels increases cellular sensitivity to ribociclib (Figure 5.4), and developed novel chemical derivatives of ribociclib that disrupt the CDK4-cyclin D interaction and promote cyclin D1 degradation in cells (Figure 5.5).

Our results using a kinase active-site probe and the fractionation of cell lysates have helped uncover which CDK4 complexes are capable of binding active-site inhibitors. The increase in monomeric CDK4 and loss of Hsp90-Cdc37-bound CDK4 after XO44 treatment confirm previous biochemical and cellular studies indicating a release of kinases from the chaperone system upon small molecule binding.^{241,251} That this hand-

off is complete after the short timescale of this experiment (30 minutes), indicates that the Hsp90-Cdc37-CDK4 complex exists in a rapidly-exchanging equilibrium. In contrast, no labeling of cyclin D-p27^{Kip1}-bound CDK4 is observed and the abundance of this complex shows little, if any, change, suggesting this complex exists in a slowly-exchanging equilibrium and that CDK4 in this complex is incapable of binding active-site inhibitors. The complete lack of XO44 binding by CDK4 in BT549 cells indicates that the CDK4-INK4 complex is also recalcitrant to active-site binders and that this complex also exists in a slowly-exchanging equilibrium. Therefore, our results indicate that the only fraction of CDK4 capable of binding small molecule active-site inhibitors in cells is in Hsp90-Cdc37-bound complexes.

Since all experiments using XO44 relied on an observed gel shift of the target kinase (CDK4 or CDK2), it is possible that differential PTMs on CDK2 or CDK4 may influence its gel shift, leading to a false negative conclusion regarding XO44-binding. To control for this potential confounder, we plan to immunoprecipitate CDK4 from fractionated lysates and directly test for XO44 labeling by TAMRA fluorescence. This type of experiment – immunoprecipitation of CDK4 from fractionated lysates – will also directly confirm that proteins that co-elute with CDK4 by gel filtration are directly bound.

Remaining questions from these results surround the dynamics of these complexes, i.e. on what timescale does CDK4 exchange from one complex to another or is a CDK4 molecule stuck within a single one of these complexes for its cellular lifetime? A model proposed by Hallett and colleagues²⁴¹ suggests that after initial CDK4 folding within the

chaperone system, CDK4 is transiently released as a monomer and is quickly bound and stabilized by small molecule inhibitors or by protein cofactors such as cyclin D, p27^{Kip1} or INK4 family members. The chaperone system has therefore been proposed to provide an additional layer of regulation of client proteins, since the relative abundance of protein cofactors will influence the equilibrium state of these complexes.²³⁹ We clearly observe this, as siRNA-knockdown of cyclin D1, which one would anticipate antagonizes the CDK4-cyclin D-p27^{Kip1} complex and thus shifts the CDK4 equilibrium towards a Hsp90-Cdc37-bound state, increases the efficacy of ribociclib. Tumors with low Cip/Kip-family expression might therefore be expected to exhibit sensitivity to CDK4/6 inhibitors.

An additional outstanding question regarding CDK4-interacting proteins involves the role of the Hsp70-family, which were previously demonstrated to bind CDK4-cyclin D-p27^{KIP1} complexes *in vitro* and *in vivo*.²⁵² Our results replicate those of Diehl and colleagues, as CDK4-cyclin D-p27^{KIP1} complexes run with at an observed molecular weight of ~150-200kDa by gel filtration chromatography (Figure 5.3C), suggesting there are additional members of this protein complex. If Hsp70s indeed assist in the initial folding of D-type cyclins and assembly of the CDK4-cyclin D-p27^{KIP1} complex as hypothesized, this would add an additional level of complexity to the regulation of CDK4 by associated proteins, and may suggest that CDK4/6 inhibitors will demonstrate efficacy as co-therapy with Hsp70 inhibitors.

Why are CDK4/6 inhibitors ineffective as monotherapy but effective in the clinic in combination with aromatase inhibitors and other estrogen receptor antagonists?

Mitogenic and hormone receptor signaling upregulate cyclin D1 transcription to promote cell cycle progression.²²⁸ Small molecules that antagonize cyclin D1 transcription by targeting mitogenic or hormone receptor signaling reduce cyclin D1 levels. Similar to our results using siRNA-knockdown of cyclin D1, this sensitizes cells to CDK4/6 inhibition, possibly by reducing active, but inhibitor resistant CDK4-cyclin D-p27^{Kip1} complexes. This explains why other inhibitors of mitogenic signaling also show promise as co-therapies with CDK4/6 inhibitors.²⁵³

However, a single agent that exhibits these characteristics (active-site inhibitor of CDK4 and decrease cyclin D levels) would be ideal. Our newly developed CDK4i compounds were designed to stabilize a conformation of CDK4 that should disrupt the CDK4-cyclin D interaction through outward movement of the kinase α C-helix. Further characterization is needed, but we anticipate CDK4i are also incapable of binding to CDK4-cyclin D-p27^{Kip1} and CDK4-INK4 complexes and therefore remain dependent on targeting a transient monomeric state of CDK4/6 for activity. These compounds may distinguish themselves from existing clinical CDK4/6 inhibitors indirectly, first through the degradation of cyclin D, leading to reduced CDK4-cyclin D-p27^{Kip1} complexes and the transitioning of CDK4 to the chaperone system where it is more accessible to small molecule binding. Additional screening for small molecule scaffolds with high affinity for the CDK4-cyclin D-p27^{Kip1} trimer may provide an additional path towards improving CDK4/6 inhibitor efficacy. Thus, through chemical screening of inhibitor-resistant

complexes, and rational design of molecules with novel chemical characteristics, we may be able to develop state-specific CDK4 inhibitors to expand their therapeutic efficacy in the clinic.

Materials and Methods

Cell culture. All cells were maintained in Dulbecco's modified eagle medium (DMEM) supplemented with 10% fetal bovine serum and 1% pen/strep and grown in a 37°C incubator at 5% CO₂. For western blotting, ~500,000 cells were plated in 6-well plates at full serum for 24hr, then treated with compounds in the same media for an additional 24hrs. For viability assays, cells were plated at 1,000 cells/well in a 96-well plate for 24hrs, then compounds were added along with fresh media.

Cell viability assay. Cell viability assays were performed using CellTiter-Glo (Promega) according to the manufacturer's instructions, except 1/5 of the recommended luminescent reagent was used.

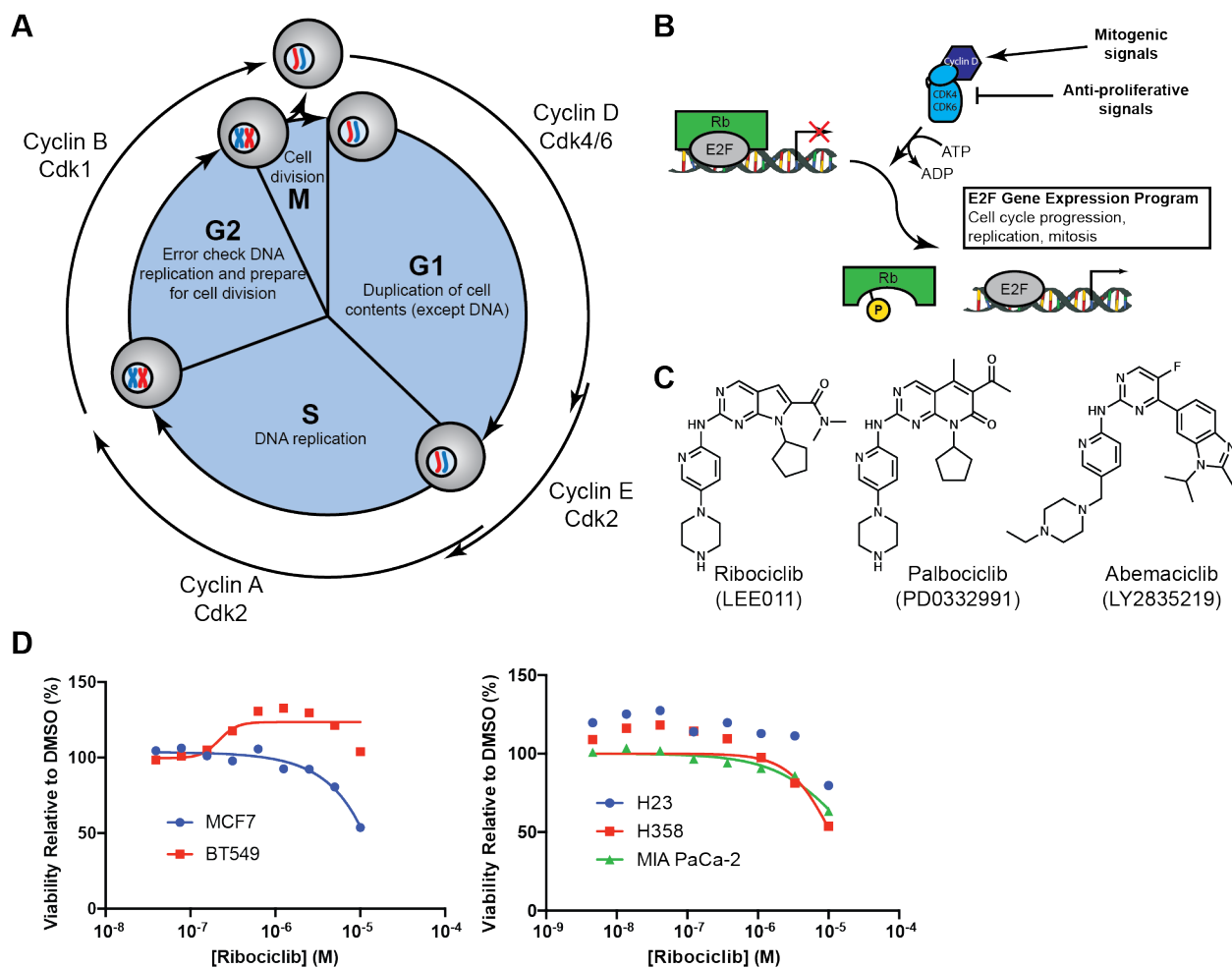
XO44-labeling experiments. XO44 labeling experiments were performed as described using 2μM XO44 for 30min.²³⁸ Click reactions were incubated at room temperature for 1hr.

Gel filtration assay. Gel filtration assays on crude cell lysates were performed as previously described.²⁴⁶ Briefly, ~25 million cells are lysed in lysis buffer (50mM Tris pH 7.5, 150mM NaCl, 1% NP-40, 10% glycerol, 1mM EDTA) plus complete protease

inhibitor cocktail (Roche) on ice for 20min. Lysates are then clarified by centrifugation at 18,000 x g for 15min, incubated with DNaseI (1 unit/mg lysate) for 20min on ice, and clarified again. Samples are then diluted to 10mg/mL and 5mg is loaded on a Superdex200 column pre-equilibrated in 50mM HEPES pH 7.5, 150mM NaCl, 1% NP-40, 10% glycerol. The column is run for 1.25CV at 0.35mL/min and 0.5mL fractions are collected starting at 5mL.

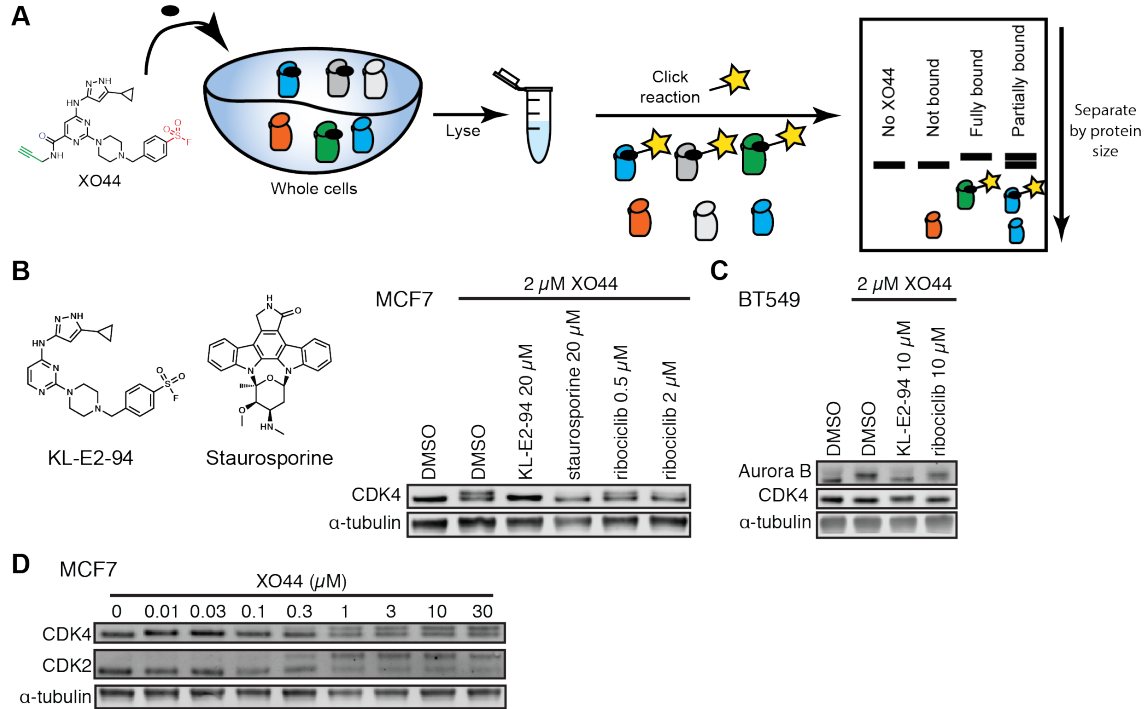
Split UnaG assay. HEK293T cells were transfected using calcium phosphate with pCDNA3.1-CDK4-cUnaG-T2A-nUnaG-CCND1 and pCDNA3.1-mIFP-H2B-T2A-mCherry.

Figure 5.1. Clinical CDK4/6 inhibitors show limited efficacy as monotherapy.



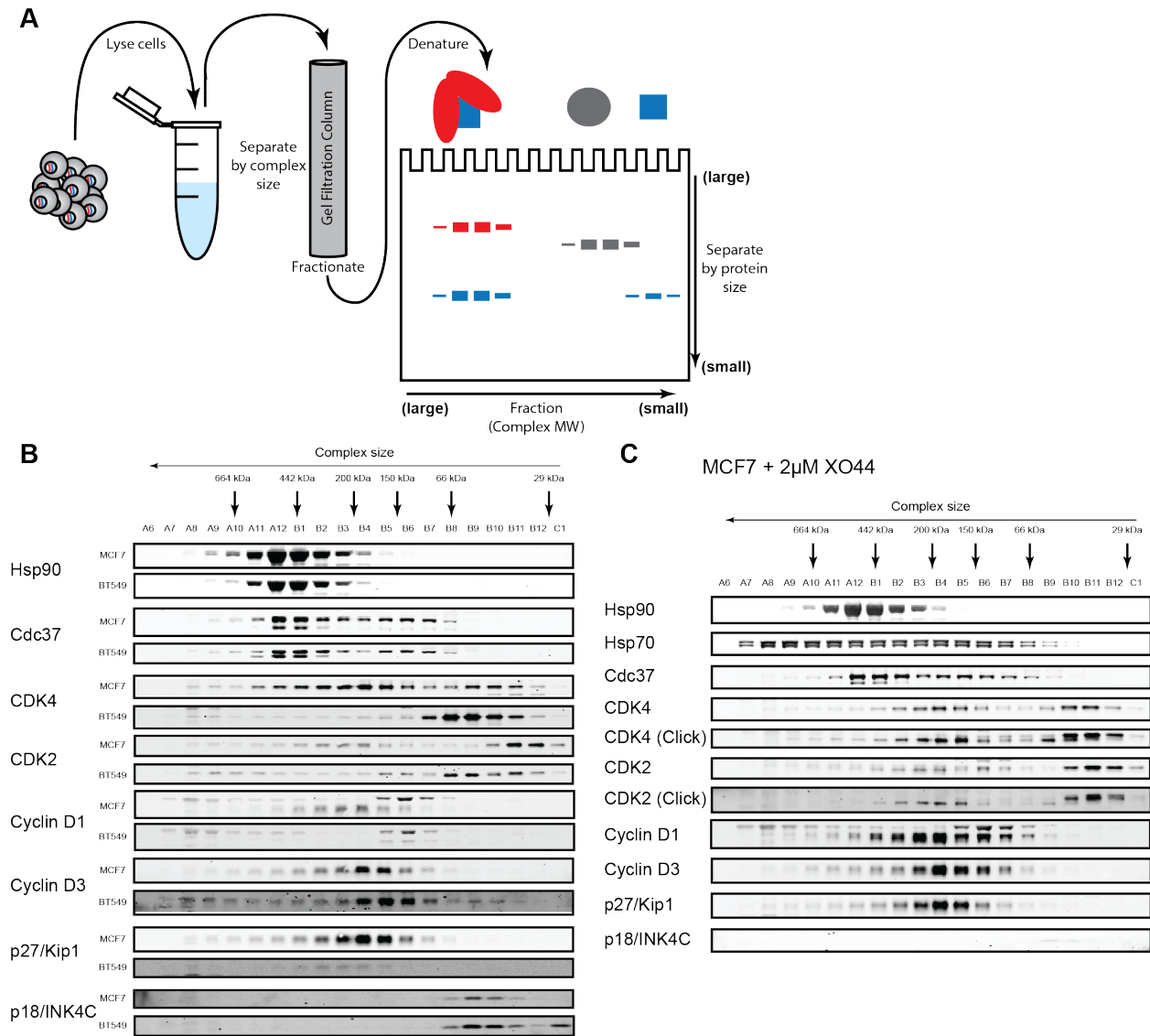
A. The cell cycle is tightly controlled by Cyclin-CDK complexes, including the CDK4/6-Cyclin D (G1 phase), CDK2-Cyclin E (G1/S transition), CDK2-Cyclin A (S and G2 phases), and CDK1-Cyclin B (G2/M phases). **B.** CDK4/6-Cyclin D complexes integrate mitogenic and antiproliferative signals to initiate cell cycle progression through the phosphorylation of retinoblastoma protein (Rb). **C.** Chemical structures of clinically approved CDK4/6 inhibitors. **D.** Decreased viability of breast cancer cell lines (MCF7, BT549) and KRas G12C cancer cell lines (H23, H358, MIA PaCa-2) in response to ribociclib.

Figure 5.2. CDK4 is partially labeled by XO44 in MCF7, but not BT549, cells.



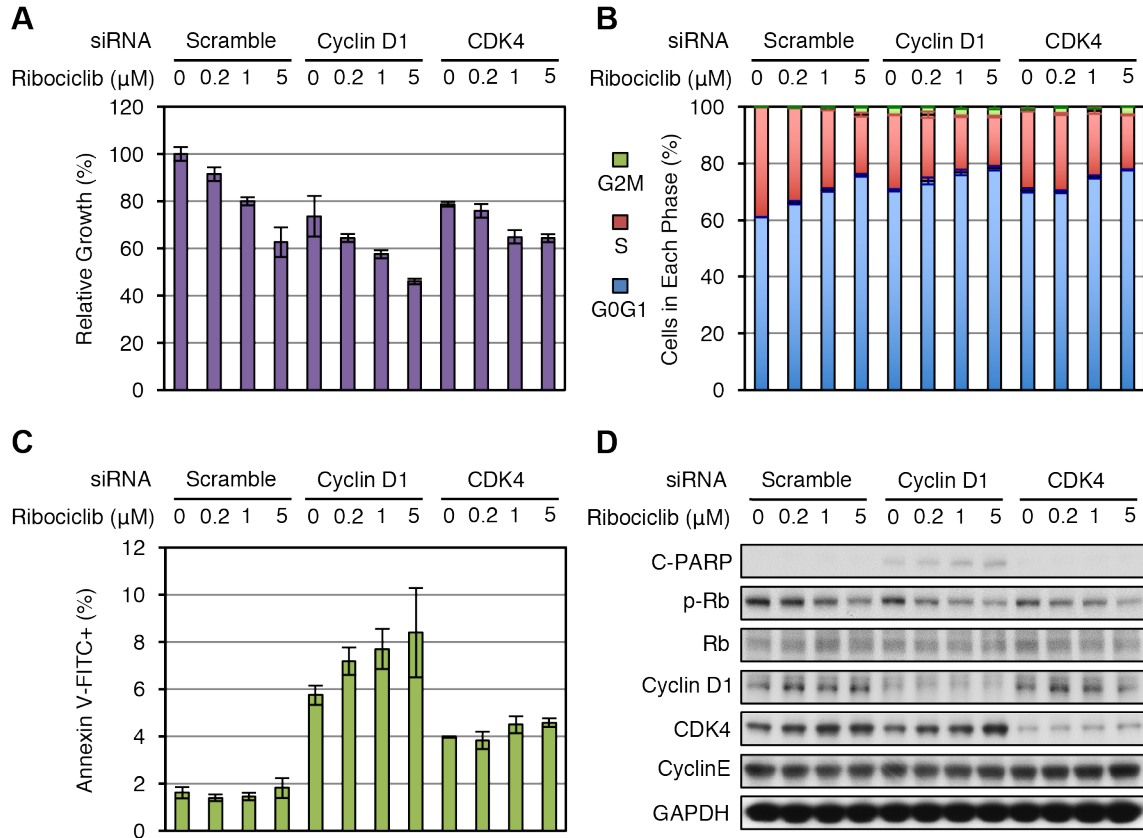
A. Scheme for XO44 labeling of cellular kinases and the separation of XO44-labeled and unlabeled kinases by western blotting. **B and C.** Western blot of CDK4 from MCF7 (**B**) or BT549 cells (**C**) treated with 2 μ M XO44 after pre-treatment with KL-E2-94, staurosporine, or ribociclib. **D.** Western blot of CDK4 and CDK2 from MCF7 cells treated with increasing concentrations of XO44 for 30min.

Figure 5.3. CDK4 complexes vary across cell lines and influence small molecule accessibility.



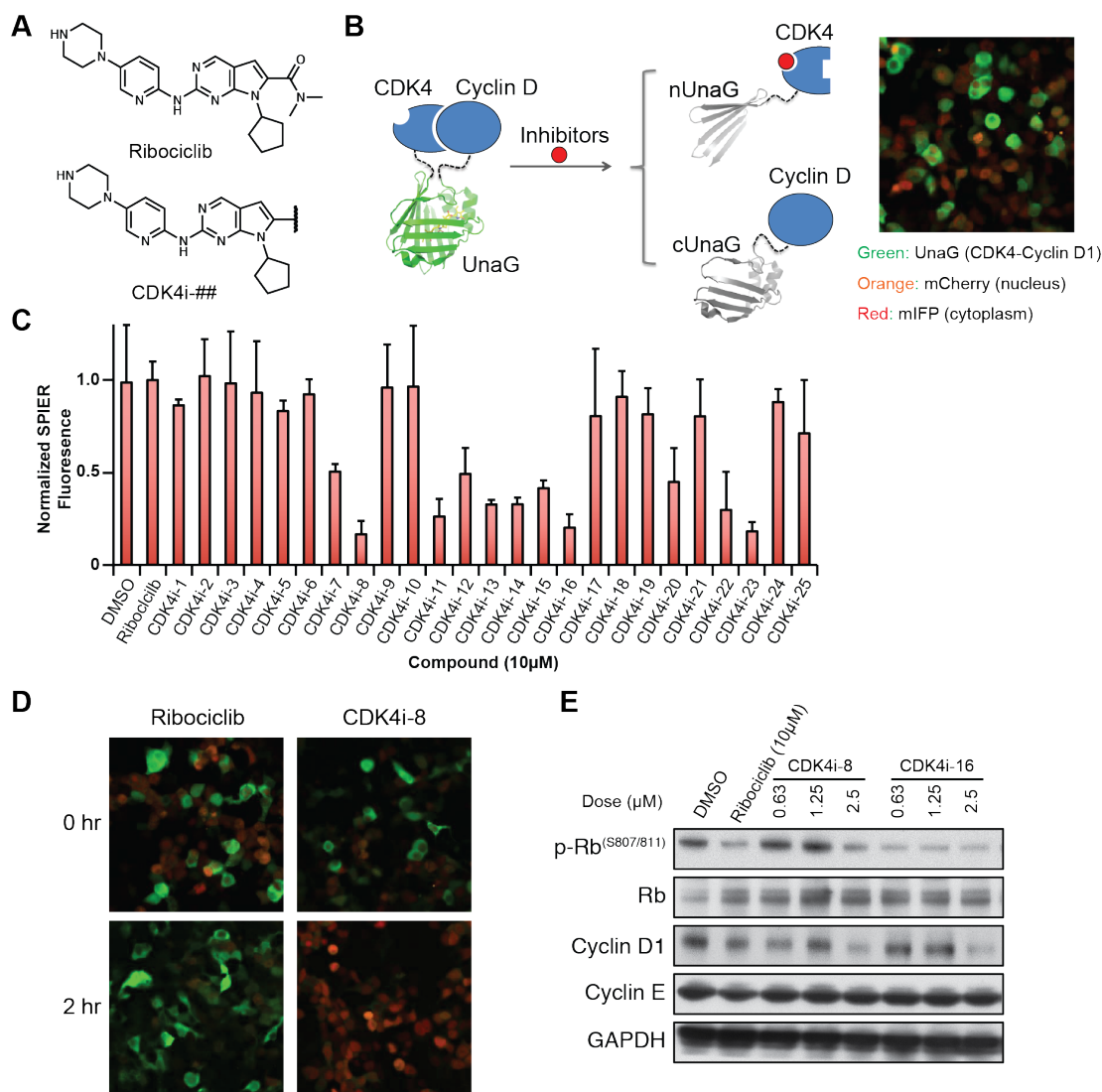
A. Scheme for the fractionation of cell lysates followed by western blotting to understand CDK high-molecular weight complexes. **B.** Fractionation of MCF7 and BT549 cell lysates shows CDK4 and CDK2 exist in differential high-molecular weight complexes. **C.** Fractionation of MCF7 cells treated with 2 μ M XO44 for 30min.

Figure 5.4. Knockdown of cyclin D1 improves efficacy of ribociclib in cell culture.



A. Relative growth of NIH3T3 cells in response to three doses of ribociclib in the presence of siRNA control (scramble) or siRNA targeting Cyclin D1 or CDK4. **B.** Changes in cell cycle after ribociclib treatment in the presence of siRNA control (scramble) or siRNA targeting Cyclin D1 or CDK4. **C.** Induction of apoptosis, as measured by annexin V staining, after ribociclib treatment in the presence of siRNA control (scramble) or siRNA targeting Cyclin D1 or CDK4. **D.** Cell signaling after ribociclib treatment in the presence of siRNA control (scramble) or siRNA targeting Cyclin D1 or CDK4.

Figure 5.5. Ribociclib analogs disrupt CDK4-cyclin D1 interaction and promote cyclin D1 degradation in cells.

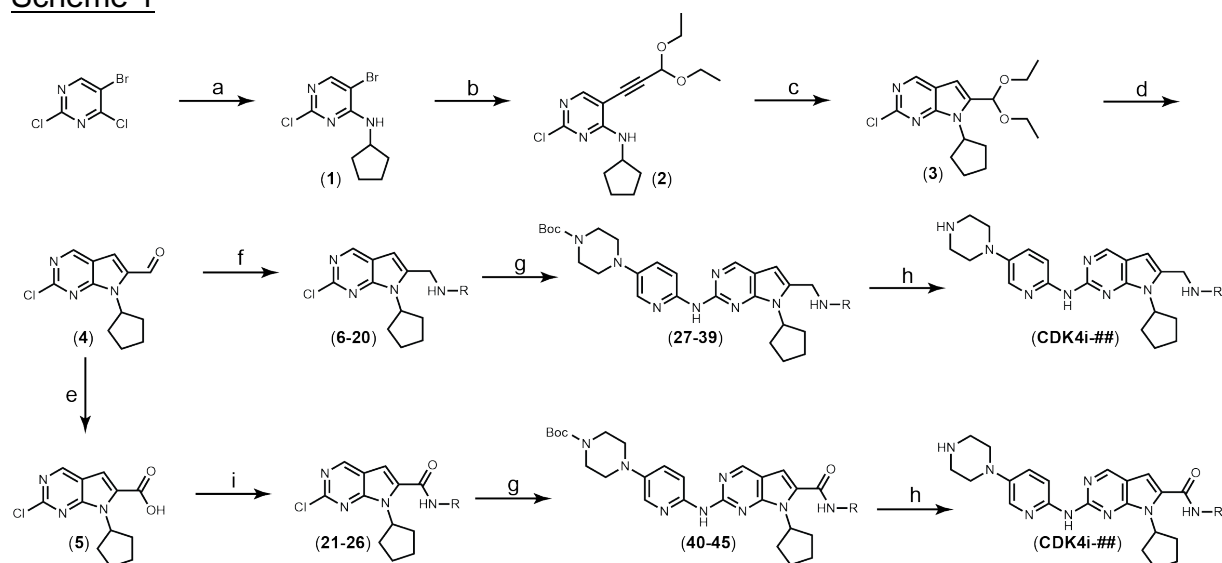


A. Chemical structure of ribociclib and CDK4i analogs. **B.** Fluorogenic assay of the CDK4-Cyclin D1 protein-protein interaction using split UnaG. **C.** Screen of CDK4i (10 μM, 2hr) activity in the disruption of the CDK4-Cyclin D1 complex. **D.** Representative images showing the CDK4-Cyclin D1 interaction in the presence of ribociclib or CDK4i-8 after 0hr and 2hr. **E.** Western blot of NIH3T3 cells treated for 24hr with ribociclib or increasing doses of CDK4i-8 and CDK4i-16.

Chemical Synthesis

Materials obtained commercially were reagent grade and were used without further purification. Reactions were monitored by thin layer chromatography (TLC) and/or mass spectrometry (LCMS) using a Waters Acquity UPLC/ESI-TQD (low-res.) or Waters Acquity UPLC/ESI-G2 XS QTOF (high-res.). All NMR spectra were obtained on a Bruker 400 spectrometer.

Scheme 1



(a) 1.3eq cyclopentylamine, 1.5eq TEA in 1,4-dioxane, RT, 16h; (b) 2.5eq 3,3'-diethoxy-1-propyne, 2.5eq TEA, 0.05eq [1,1'-bis(diphenylphosphino)ferrocene]dichloropalladium(II), 0.05eq CuI in THF, 140°C microwave, 1h; (c) 4eq TBAF in THF, 40°C, 40min; (d) 8:1 AcOH:water, RT, 16h; (e) oxone in DMF/H₂O, 75°C, 16h; (f) 2.5eq amine in 9:1 MeOH:AcOH, 2eq NaCNBH₃, RT, 2-16h; (g) 2eq 2-amino-5-(4-boc-piperziny)-pyridine (60), 4eq cesium carbonate, 0.05eq Pd₂(dba)₃, 0.05eq XANTPhos, 0.05eq tBuXPhos Pd G3 in 1,4-dioxane, 140°C microwave, 1h; (h) 20% (v/v) TFA in DCM, RT, 1h; (i) 3eq amine, 3eq TEA, 1.1eq HATU in THF, 0°C -> 50°C, 1h

5-bromo-2-chloro-N-cylopentylpyrimidin-4-amine (1). To a solution of 5-bromo-2,4-dichloropyrimidine (3g, 13.17mmol) in 1,4-dioxane (25mL) was added triethylamine (2.75mL, 19.75mmol, 1.5eq) followed by cyclopentylamine (1.69mL, 17.12mmol, 1.3eq) and the reaction was stirred at RT. After completion, the reaction mixture was evaporated to dryness under reduced pressure and the resulting solid was dissolved in

EtOAc/brine and the organic layer was isolated, dried with Na₂SO₄ and evaporated to dryness under reduced pressure. The resulting oil was purified by silica chromatography (0-25% EtOAc/hexanes) to yield the product as a yellow/white solid (3341mg, 91.7% yield). ¹H NMR (400MHz, CDCl₃) δ 8.11 (s, 1H), 5.46 (d, *J* = 7.5 Hz, 1H), 4.43 (h, *J* = 7.0 Hz, 1H), 2.21 – 2.08 (m, 2H), 1.85 – 1.60 (m, 4H), 1.56 – 1.41 (m, 2H).

2-chloro-N-cyclopentyl-5-(3,3-diethoxyprop-1-yn-yl)pyrimidin-4-amine (2). To a 35mL microwave vial charged with anhydrous THF (18mL) under Ar(g) was added intermediate **1** (1.5g, 5.42mmol). This solution was purged of oxygen with Ar(g) and [1,1'-Bis(diphenylphosphino)ferrocene]dichloropalladium(II), complex with dichloromethane (221.46mg, 0.2700mmol, 0.05eq) was added, and the solution was purged again. Triethylamine (1889.92uL, 13.56mmol, 2.5eq) and 3,3-diethoxyprop-1-yne (1943.98uL, 13.56mmol, 2.5eq) were added slowly. The reaction was heated to 140°C for 1hr in microwave reactor, then cooled, filtered through celite, and evaporated to dryness under reduced pressure. The resulting oil was purified by silica chromatography (0-25% EtOAc/hexanes) to yield the pure product as a brown oil (657mg, 37.4%). ¹H NMR (400MHz, CDCl₃) δ 8.11 (s, 1H), 5.58 (d, *J* = 7.5 Hz, 1H), 5.52 (s, 1H), 4.45 (h, *J* = 7.0 Hz, 1H), 3.80 (dq, *J* = 9.5, 7.1 Hz, 2H), 3.68 (dq, *J* = 9.4, 7.0 Hz, 2H), 2.20 – 2.07 (m, 2H), 1.72 (dddd, *J* = 28.8, 14.4, 7.1, 4.1 Hz, 4H), 1.48 (dq, *J* = 12.3, 5.9 Hz, 2H), 1.30 (t, *J* = 7.1 Hz, 7H). LCMS (TOF) *m/z* calculated 322.1328 (M-H)⁻, found 322.1356.

2-chloro-7-cyclopentyl-6-(diethoxymethyl)-7H-pyrrolo[2,3-d]pyrimidine (3). To a solution of intermediate **2** (657mg, 2.03mmol) in THF (1mL) was added tetrabutylammonium fluoride (1M solution in THF; 8.12mL, 8.12mmol, 4eq) and the resulting solution was heated to 40°C for 40min. The reaction was cooled, evaporated to dryness under reduced pressure, and purified by silica chromatography (0-20% EtOAc/hexanes) to yield the pure product as a light brown oil (653.1mg, 99.4% yield). ¹H NMR (400MHz, CDCl₃) δ 8.74 (s, 1H), 6.62 (d, *J* = 0.7 Hz, 1H), 5.67 (s, 1H), 5.00 (p, *J* = 8.7 Hz, 1H), 3.62 (dt, *J* = 16.4, 9.4, 7.0 Hz, 5H), 2.46 (ddd, *J* = 17.0, 8.6, 4.4 Hz, 2H), 2.22 – 2.10 (m, 2H), 2.10 – 1.95 (m, 2H), 1.72 (tt, *J* = 6.8, 2.9 Hz, 2H), 1.27 (t, *J* = 7.1 Hz, 7H). LCMS (TOF) *m/z* calculated 324.1473 (M+H)⁺, found 324.1416.

2-chloro-7-cyclopentyl-7H-pyrrolo[2,3-d]pyrimidine-6-carbaldehyde (4).

Intermediate **3** (653.1mg, 2.02mmol) was dissolved in water (900μL) and acetic acid (7.2mL) and stirred overnight. The solution was added to saturated NaHCO₃ and extracted with EtOAc. The combined organic layers were dried with Na₂SO₄, filtered, and concentrated under reduced pressure. The resulting oil was purified by silica chromatography to yield the pure product as a clear oil (433.1mg, 86.0% yield). ¹H NMR (400MHz, CDCl₃) δ 9.96 (s, 1H), 9.02 (s, 1H), 7.35 (s, 1H), 5.78 (p, *J* = 8.8 Hz, 1H), 2.37 – 2.23 (m, 2H), 2.22 – 2.03 (m, 4H), 1.84 – 1.67 (m, 2H). LCMS (TOF) *m/z* calculated 250.0742 (M+H)⁺, found 250.0731.

2-chloro-7-cyclopentyl-7H-pyrrolo[2,3-d]pyrimidine-6-carboxylic acid (5). To a solution of intermediate **4** in DMF (2mL) and water (200μL) was added oxone (228.5mg,

1.5mmol, 2.5eq) and the resulting suspension was heated to 70°C. Additional oxone was added as needed until the reaction was judged complete by LCMS. The reaction was cooled, dissolved in DCM and extracted with 1M NaOH. Hydrochloric acid (1M) was added to the combined aqueous layers until pH~1, and this solution was extracted with EtOAc to yield the pure product as a tan solid (111.8mg, 70.0% yield). ¹H NMR (400 MHz, Acetone-*d*₆) δ 9.05 (s, 1H), 7.48 (s, 1H), 5.91 (p, *J* = 8.7 Hz, 1H), 2.43 (ddd, *J* = 13.4, 9.8, 6.1 Hz, 2H), 2.17 – 2.02 (m, 4H), 1.74 (tq, *J* = 10.5, 4.9, 4.5 Hz, 2H). LCMS (TOF) *m/z* calculated 266.0691 (M+H)⁺, found 266.0718.

General procedure for the synthesis of intermediates 6-20.

To a solution of intermediate **4** (20mg, 0.080mmol) in 9:1 methanol:acetic acid was added the appropriate aromatic amine (0.200mmol, 2.5eq) and the resulting solution was stirred at RT for 4hr. Sodium cyanoborohydride (10mg, 0.160mmol, 2eq) was then added and the reaction was stirred at RT overnight. The reaction was added to saturated NaHCO₃ and extracted with EtOAc. The combined organic layers were dried with Na₂SO₄, filtered, concentrated under reduced pressure, and purified by silica chromatography to yield intermediates **6-20**.

N-((2-chloro-7-cyclopentyl-7H-pyrrolo[2,3-d]pyrimidin-6-yl)methyl)-3-(trifluoromethyl)aniline (6). White solid (7.8mg, 27.7% yield). LCMS (TQD) *m/z* calculated 395.12 (M+H)⁺, found 395.11.

1-(4-(((2-chloro-7-cyclopentyl-7H-pyrrolo[2,3-d]pyrimidin-6-yl)methyl)amino)phenyl)cyclopentane-1-carbonitrile (7). White solid (25.7mg, 76.4% yield). ¹H NMR (400MHz, CDCl₃) δ 8.68 (s, 1H), 7.34 – 7.29 (m, 2H), 6.75 – 6.63 (m, 2H), 6.49 (s, 1H), 4.88 (p, *J* = 8.5 Hz, 1H), 4.51 (d, *J* = 5.3 Hz, 2H), 4.08 (t, *J* = 5.3 Hz, 1H), 2.54 – 2.33 (m, 4H), 2.20 – 1.86 (m, 10H), 1.81 – 1.68 (m, 2H). LCMS (TOF) *m/z* calculated 420.1950 (M+H)⁺, found 420.1934.

3-(tert-butyl)-N-(((2-chloro-7-cyclopentyl-7H-pyrrolo[2,3-d]pyrimidin-6-yl)methyl)-1-methyl-1H-pyrazol-5-amine (8). White solid (21.3mg, 68.7% yield). ¹H NMR (400MHz, CDCl₃) δ 8.64 (s, 1H), 6.46 (s, 1H), 5.53 (s, 1H), 4.85 (p, *J* = 8.4 Hz, 1H), 4.39 (s, 2H), 3.65 (s, 3H), 2.44 (dt, *J* = 14.7, 7.7 Hz, 2H), 2.21 – 2.00 (m, 4H), 1.80 – 1.65 (m, 2H), 1.32 (s, 9H). LCMS (TOF) *m/z* calculated 387.2058 (M+H)⁺, found 387.2038.

3-(tert-butyl)-N-(((2-chloro-7-cyclopentyl-7H-pyrrolo[2,3-d]pyrimidin-6-yl)methyl)-1-(*p*-tolyl)-1H-pyrazol-5-amine (9). White solid (18.2mg, 49.1% yield). LCMS (TQD) *m/z* calculated 463.24 (M+H)⁺, found 463.22.

N-(((2-chloro-7-cyclopentyl-7H-pyrrolo[2,3-d]pyrimidin-6-yl)methyl)-3-(4-methyl-1H-imidazol-1-yl)-5-(trifluoromethyl)aniline (10). White solid (11.3mg, 29.7% yield). LCMS (TQD) *m/z* calculated 475.16 (M+H)⁺, found 475.15.

3-(tert-butyl)-N-(((2-chloro-7-cyclopentyl-7H-pyrrolo[2,3-d]pyrimidin-6-yl)methyl)-1-isopropyl-1H-pyrazol-5-amine (11). White solid (27.9mg, 83.9% yield). ¹H NMR

(400MHz, CDCl₃) δ 8.67 (s, 1H), 6.48 (s, 1H), 5.53 (s, 1H), 4.89 (p, *J* = 8.6 Hz, 1H), 4.39 (d, *J* = 6.0 Hz, 2H), 4.35 – 4.26 (m, 1H), 3.41 (t, *J* = 6.0 Hz, 1H), 2.45 (ddt, *J* = 14.6, 8.5, 3.7 Hz, 2H), 2.22 – 2.01 (m, 6H), 1.80 – 1.65 (m, 2H), 1.44 (d, *J* = 6.7 Hz, 6H), 1.31 (s, 9H). LCMS (TQD) *m/z* calculated 415.24 (M+H)⁺, found 415.29.

1,3-di-*tert*-butyl-N-((2-chloro-7-cyclopentyl-7*H*-pyrrolo[2,3-*d*]pyrimidin-6-yl)methyl)-1*H*-pyrazol-5-amine (12). White solid (19.0mg, 84.4% yield). ¹H NMR (400MHz, CDCl₃) δ 8.72 (s, 1H), 6.51 (s, 1H), 5.59 (s, 1H), 4.92 (p, *J* = 8.6 Hz, 1H), 4.36 (d, *J* = 6.0 Hz, 2H), 3.48 – 3.43 (m, 1H), 2.56 – 2.40 (m, 2H), 2.22 – 2.02 (m, 4H), 1.81 – 1.67 (m, 3H), 1.62 (s, 9H), 1.30 (s, 9H). LCMS (TOF) *m/z* calculated 429.2528 (M+H)⁺, found 429.2534.

3-(*tert*-butyl)-N-((2-chloro-7-cyclopentyl-7*H*-pyrrolo[2,3-*d*]pyrimidin-6-yl)methyl)-1-isobutyl-1*H*-pyrazol-5-amine (13). White solid (16.3mg, 47.4% yield). ¹H NMR (400MHz, CDCl₃) δ 8.68 (s, 1H), 6.47 (s, 1H), 5.52 (s, 1H), 4.87 (p, *J* = 8.7 Hz, 1H), 4.39 (d, *J* = 6.0 Hz, 2H), 3.71 (d, *J* = 7.5 Hz, 2H), 3.39 (t, *J* = 6.0 Hz, 1H), 2.46 (ddd, *J* = 12.2, 6.1, 3.9 Hz, 2H), 2.20 – 2.02 (m, 5H), 1.81 – 1.66 (m, 2H), 1.31 (s, 9H), 0.90 (d, *J* = 6.7 Hz, 6H). LCMS (TOF) *m/z* calculated 429.2528 (M+H)⁺, found 429.2534.

3-(*tert*-butyl)-N-((2-chloro-7-cyclopentyl-7*H*-pyrrolo[2,3-*d*]pyrimidin-6-yl)methyl)-1-cyclopentyl-1*H*-pyrazol-5-amine (14). White solid (24.6mg, 69.6% yield). ¹H NMR (400MHz, CDCl₃) δ 8.68 (s, 1H), 6.48 (s, 1H), 5.53 (s, 1H), 4.89 (p, *J* = 8.6 Hz, 1H), 4.39 (d, *J* = 6.3 Hz, 3H), 3.41 (t, *J* = 6.0 Hz, 1H), 2.44 (ddd, *J* = 12.4, 10.2, 4.8 Hz, 2H), 2.21

– 2.03 (m, 4H), 1.93 (dtd, $J = 16.3, 13.6, 12.7, 6.9$ Hz, 4H), 1.77 – 1.67 (m, 2H), 1.62 (dd, $J = 10.0, 5.8$ Hz, 2H), 1.30 (s, 9H).

3-(tert-butyl)-N-((2-chloro-7-cyclopentyl-7H-pyrrolo[2,3-d]pyrimidin-6-yl)methyl)-1-phenyl-1H-pyrazol-5-amine (15). White solid (6.8mg, 16.9% yield). ^1H NMR (400MHz, CDCl_3) δ 8.71 (s, 1H), 7.55 (dd, $J = 8.4, 1.1$ Hz, 2H), 7.50 – 7.40 (m, 2H), 7.36 – 7.27 (m, 2H), 6.47 (s, 1H), 5.61 (s, 1H), 4.87 (p, $J = 8.7$ Hz, 1H), 4.43 (d, $J = 5.8$ Hz, 2H), 3.84 (t, $J = 5.8$ Hz, 1H), 2.52 – 2.36 (m, 2H), 2.21 – 2.10 (m, 2H), 2.10 – 1.97 (m, 2H), 1.74 (ddd, $J = 10.8, 7.8, 4.6$ Hz, 2H), 1.37 (s, 9H). LCMS (TQD) m/z calculated 449.22 ($\text{M}+\text{H}$)⁺, found 449.26.

1-(tert-butyl)-N-((2-chloro-7-cyclopentyl-7H-pyrrolo[2,3-d]pyrimidin-6-yl)methyl)-3-isopropyl-1H-pyrazol-5-amine (16). White solid (14.7mg, 44.2% yield). ^1H NMR (400MHz, CDCl_3) δ 8.73 (s, 1H), 6.51 (s, 1H), 5.56 (s, 1H), 4.91 (p, $J = 8.6$ Hz, 1H), 4.36 (d, $J = 5.9$ Hz, 2H), 3.53 (t, $J = 5.9$ Hz, 1H), 2.91 (p, $J = 6.9$ Hz, 1H), 2.55 – 2.40 (m, 2H), 2.10 (ddt, $J = 22.9, 11.2, 7.9$ Hz, 4H), 1.70 (tt, $J = 15.4, 6.6$ Hz, 2H), 1.62 (s, 9H), 1.26 (d, $J = 6.9$ Hz, 6H). LCMS (TOF) m/z calculated 415.2371 ($\text{M}+\text{H}$)⁺, found 415.2365.

1-(tert-butyl)-N-((2-chloro-7-cyclopentyl-7H-pyrrolo[2,3-d]pyrimidin-6-yl)methyl)-3-cyclopentyl-1H-pyrazol-5-amine (17). White solid (17.7mg, 50.1% yield). ^1H NMR (400MHz, CDCl_3) δ 8.72 (s, 1H), 6.51 (s, 1H), 5.54 (s, 1H), 4.90 (p, $J = 8.6$ Hz, 1H), 4.36 (d, $J = 5.9$ Hz, 2H), 3.54 (t, $J = 5.9$ Hz, 1H), 3.01 (p, $J = 8.4$ Hz, 1H), 2.54 – 2.40 (m,

2H), 2.20 – 1.98 (m, 6H), 1.83 – 1.57 (m, 17H). LCMS (TOF) m/z calculated 441.2528 (M+H)⁺, found 441.2515.

1-(tert-butyl)-N-((2-chloro-7-cyclopentyl-7H-pyrrolo[2,3-d]pyrimidin-6-yl)methyl)-3-phenyl-1H-pyrazol-5-amine (18). White solid (24.6mg, 68.4% yield). ¹H NMR (400MHz, CDCl₃) δ 8.72 (s, 1H), 7.81 (d, J = 7.3 Hz, 2H), 7.39 (t, J = 7.6 Hz, 2H), 7.29 (d, J = 7.4 Hz, 1H), 6.54 (s, 1H), 6.04 (s, 1H), 4.93 (p, J = 8.5 Hz, 1H), 4.46 (d, J = 5.9 Hz, 2H), 3.70 (t, J = 5.7 Hz, 1H), 2.47 (dt, J = 15.0, 8.3 Hz, 2H), 2.12 (dq, J = 18.7, 11.0, 10.2 Hz, 4H), 1.71 (s, 11H). LCMS (TOF) m/z calculated 449.2215 (M+H)⁺, found 449.2236.

N-((2-chloro-7-cyclopentyl-7H-pyrrolo[2,3-d]pyrimidin-6-yl)methyl)-3-isopropyl-1-methyl-1H-pyrazol-5-amine (19). White solid (6.3mg, 60.9% yield). ¹H NMR (400MHz, CDCl₃) δ 8.70 (s, 1H), 6.50 (s, 1H), 5.49 (s, 1H), 4.86 (p, J = 8.6 Hz, 1H), 4.40 (d, J = 5.3 Hz, 2H), 3.63 (s, 3H), 3.47 (d, J = 8.6 Hz, 1H), 2.91 (hept, J = 6.9 Hz, 1H), 2.54 – 2.38 (m, 2H), 2.21 – 1.99 (m, 4H), 1.74 (ddt, J = 9.9, 7.0, 4.3 Hz, 2H), 1.27 (d, J = 6.9 Hz, 6H). LCMS (TQD) m/z calculated 373.19 (M+H)⁺, found 373.17.

N-((2-chloro-7-cyclopentyl-7H-pyrrolo[2,3-d]pyrimidin-6-yl)methyl)-1-methyl-1H-pyrazol-5-amine (20). White solid (18.0mg, 67.9% yield). ¹H NMR (400MHz, CDCl₃) δ 8.61 (s, 1H), 7.39 – 7.31 (m, 1H), 6.44 (s, 1H), 5.63 (d, J = 1.4 Hz, 1H), 4.85 (p, J = 8.6 Hz, 1H), 4.41 (d, J = 5.8 Hz, 2H), 3.80 (t, J = 5.7 Hz, 1H), 3.71 (s, 3H), 2.44 (dq, J =

15.4, 7.8 Hz, 2H), 2.10 (dddd, $J = 20.1, 15.3, 9.6, 5.1$ Hz, 4H), 1.74 (dd, $J = 12.8, 7.0$ Hz, 2H). LCMS (TOF) m/z calculated 331.1432 (M+H)⁺, found 331.1420.

General procedure for the synthesis of intermediates (21-26).

To a solution of intermediate **5** (15mg, 0.056mmol) in THF (1mL) under Ar(g) was added N,N-diisopropylethylamine (14.8 μ L, 0.0847mmol, 1.5eq) and HATU (25.8mg, 0.0677mmol, 1.2eq) and the resulting solution was stirred at RT for 20min, then cooled on ice. The appropriate amine (0.0847, 1.5eq) was dissolved in THF (0.5mL) and added dropwise to the reaction vial. The reaction was allowed to warm to RT and stirred until complete by LCMS with mild heating (40°C), as necessary. The reaction was added to brine and extract with EtOAc. The combined organic layers were dried with Na₂SO₄, filtered, and concentrated under reduced pressure. The resulting solid was purified by silica chromatography to yield intermediates **21-26**.

2-chloro-7-cyclopentyl-N-(3-(trifluoromethyl)phenyl)-7H-pyrrolo[2,3-d]pyrimidine-6-carboxamide (21). White solid (12.7mg, 55.0% yield). LCMS (TQD) m/z calculated 409.10 (M+H)⁺, found 409.10.

2-chloro-N-(4-(1-cyanocyclopentyl)phenyl)-7-cyclopentyl-7H-pyrrolo[2,3-d]pyrimidine-6-carboxamide (22). White solid (9.8mg, 88.8% yield). ¹H NMR (400MHz, CDCl₃) δ 8.87 (s, 1H), 8.11 (s, 1H), 7.67 (d, $J = 8.7$ Hz, 2H), 7.57 – 7.44 (m, 2H), 6.97 (s, 1H), 5.48 (p, $J = 8.7$ Hz, 1H), 2.57 – 2.39 (m, 4H), 2.20 – 1.91 (m, 10H),

1.72 (tt, $J = 10.9, 5.2$ Hz, 2H). LCMS (TOF) m/z calculated 434.1742 (M+H)⁺, found 434.1735.

N-(3-(*tert*-butyl)-1-methyl-1*H*-pyrazol-5-yl)-2-chloro-7-cyclopentyl-7*H*-pyrrolo[2,3-*d*]pyrimidine-6-carboxamide (23). White solid (23.5mg, 79.8% yield). LCMS (TQD) m/z calculated 401.19 (M+H)⁺, found 401.22.

2-chloro-N-(4-(1-cyanocyclobutyl)phenyl)-7-cyclopentyl-7*H*-pyrrolo[2,3-*d*]pyrimidine-6-carboxamide (24). White solid (18.1mg, 76.4% yield). ¹H NMR (400MHz, CDCl₃) δ 8.83 (s, 1H), 8.27 (s, 1H), 7.74 – 7.66 (m, 2H), 7.48 – 7.41 (m, 2H), 6.97 (s, 1H), 5.48 (p, $J = 8.7$ Hz, 1H), 2.93 – 2.80 (m, 2H), 2.71 – 2.56 (m, 2H), 2.45 (tdd, $J = 10.3, 8.5, 3.6$ Hz, 2H), 2.11 (tqd, $J = 14.8, 8.6, 7.6, 4.1$ Hz, 5H), 1.79 – 1.64 (m, 3H). LCMS (TOF) m/z calculated 420.1586 (M+H)⁺, found 420.1598.

2-chloro-N-(4-(1-cyanocyclohexyl)phenyl)-7-cyclopentyl-7*H*-pyrrolo[2,3-*d*]pyrimidine-6-carboxamide (25). White solid (9.3mg, 55.6% yield). ¹H NMR (400MHz, CDCl₃) δ 8.87 (s, 1H), 8.10 (s, 1H), 7.72 – 7.64 (m, 2H), 7.59 – 7.49 (m, 2H), 6.97 (s, 1H), 5.48 (p, $J = 8.7$ Hz, 1H), 2.56 – 2.37 (m, 2H), 2.24 – 2.04 (m, 6H), 1.97 – 1.60 (m, 10H).

2-chloro-N-(4-(1-cyanopropan-2-yl)phenyl)-7-cyclopentyl-7*H*-pyrrolo[2,3-*d*]pyrimidine-6-carboxamide (26). White solid (16.1mg, 69.9% yield). ¹H NMR (400MHz, CDCl₃) δ 8.87 (s, 1H), 8.07 (s, 1H), 7.74 – 7.64 (m, 2H), 7.57 – 7.46 (m, 2H),

6.97 (s, 1H), 5.49 (p, $J = 8.7$ Hz, 1H), 2.53 – 2.39 (m, 2H), 2.14 (qd, $J = 9.9, 8.4, 5.5$ Hz, 4H), 1.76 – 1.66 (m, 2H). LCMS (TOF) m/z calculated 408.1586 (M+H)⁺, found 408.1603.

General procedure for the synthesis of intermediates **27-45**.

A solution of the appropriate intermediate **6-26**, intermediate **60** (2eq), and cesium carbonate (4eq) in 1,4-dioxane (1.5mL) in a 10mL microwave vial was purged with Ar(g). Pd₂(dba)₃ (0.05eq), XANTPhos (0.05eq), and tBuXPhos Pd G3 (0.05eq) were then added and the resulting solution was purged with Ar(g). The reaction was heated to 140°C for 45min in a microwave reactor, cooled, filtered through celite, and washed with EtOAc. The resulting filtrate was concentrated under reduced pressure and separated by silica chromatography to yield intermediates **27-45** that were carried forward without further purification or characterization.

General procedure for the synthesis of **CDK4i-##**

To a solution of intermediates **27-45** in DCM (2mL) was added trifluoroacetic acid (400μL) and the resulting solution was stirred until judged complete by LCMS (~1hr). This solution was added to saturated NaHCO₃, extracted with DCM, dried with Na₂SO₄, filtered, and concentrated under reduced pressure. The resulting oil was purified by reverse-phase chromatography (0-70% MeCN/water) and lyophilized to yield the final product.

CDK4i-1. LCMS (TQD) m/z calculated 389.21 (M+H)⁺, found 389.21.

CDK4i-2. LCMS (TQD) m/z calculated 407.23 (M+H)⁺, found 407.25.

CDK4i-3. LCMS (TQD) m/z calculated 551.25 (M+H)⁺, found 551.26.

CDK4i-4. LCMS (TQD) m/z calculated 537.27 (M+H)⁺, found 537.28.

CDK4i-5. ¹H NMR (400 MHz, DMSO-*d*₆) δ 10.57 (s, 1H), 9.53 (s, 1H), 8.91 (s, 1H), 8.30 (s, 1H), 8.17 (d, *J* = 9.1 Hz, 1H), 8.04 (d, *J* = 3.0 Hz, 1H), 7.86 – 7.73 (m, 2H), 7.55 – 7.39 (m, 3H), 7.18 (s, 1H), 5.47 (p, *J* = 8.9 Hz, 1H), 3.17 (dd, *J* = 6.8, 3.4 Hz, 4H), 3.02 (t, *J* = 5.0 Hz, 4H), 2.57 – 2.46 (m, 2H), 2.47 – 2.36 (m, 2H), 2.16 – 1.92 (m, 6H), 1.95 – 1.82 (m, 4H), 1.77 – 1.56 (m, 2H). LCMS (TQD) m/z calculated 576.31 (M+H)⁺, found 576.34.

CDK4i-6. ¹H NMR (400 MHz, DMSO-*d*₆) δ 9.12 (s, 1H), 8.62 (s, 1H), 8.31 (br s, 1H), 8.17 (d, *J* = 9.1 Hz, 1H), 7.99 (d, *J* = 3.0 Hz, 1H), 7.42 (dd, *J* = 9.2, 3.0 Hz, 1H), 7.24 – 7.14 (m, 2H), 6.77 – 6.68 (m, 2H), 6.39 (s, 1H), 6.30 (t, *J* = 5.7 Hz, 1H), 4.76 (p, *J* = 8.8 Hz, 1H), 4.43 (d, *J* = 5.0 Hz, 2H), 3.11 (dd, *J* = 6.6, 3.5 Hz, 4H), 2.98 (dd, *J* = 6.5, 3.6 Hz, 4H), 2.51 (s, 2H), 2.38 – 2.25 (m, 2H), 1.99 (s, 6H), 1.90 – 1.76 (m, 4H), 1.76 – 1.59 (m, 2H). LCMS (TQD) m/z calculated 562.34 (M+H)⁺, found 562.36.

CDK4i-7. ¹H NMR (400 MHz, DMSO-*d*₆) δ 10.39 (br s, 1H), 9.48 (s, 1H), 8.91 (s, 1H), 8.32 (s, 1H), 8.14 (d, *J* = 9.1 Hz, 1H), 8.01 (d, *J* = 3.0 Hz, 1H), 7.44 (dd, *J* = 9.1, 3.0 Hz,

1H), 7.24 (s, 1H), 6.18 (s, 1H), 5.50 (p, $J = 9.0$ Hz, 1H), 3.67 (s, 3H), 3.14 – 3.04 (m, 4H), 2.92 (d, $J = 4.9$ Hz, 5H), 2.58 – 2.42 (m, 2H), 2.09 – 1.91 (m, 4H), 1.66 (s, 2H), 1.24 (s, 9H). LCMS (TQD) m/z calculated 543.33 (M+H)⁺, found 543.36.

CDK4i-8. ¹H NMR (400 MHz, DMSO-*d*₆) δ 9.16 (s, 1H), 8.66 (d, $J = 2.9$ Hz, 1H), 8.28 (br s, 2H), 8.19 (d, $J = 9.2$ Hz, 1H), 8.01 (d, $J = 3.0$ Hz, 1H), 7.45 (dd, $J = 9.1, 3.0$ Hz, 1H), 6.44 (s, 1H), 5.73 (s, 1H), 5.44 (s, 1H), 4.80 (p, $J = 8.8$ Hz, 1H), 4.34 – 4.24 (m, 2H), 3.48 (s, 3H), 3.19 (dd, $J = 6.7, 3.5$ Hz, 4H), 3.11 – 3.02 (m, 4H), 2.60 – 2.45 (m, 2H), 2.11 – 1.95 (m, 4H), 1.75 – 1.60 (m, 2H), 1.18 (s, 9H). LCMS (TOF) m/z calculated 529.3510 (M+H)⁺, found 529.3511.

CDK4i-9. ¹H NMR (400 MHz, DMSO-*d*₆) δ 9.07 (s, 1H), 8.64 (s, 1H), 8.31 (br s, 1H), 8.16 (d, $J = 9.1$ Hz, 1H), 7.98 (d, $J = 2.9$ Hz, 1H), 7.49 – 7.37 (m, 3H), 7.31 – 7.21 (m, 2H), 6.41 (s, 1H), 5.71 (d, $J = 6.7$ Hz, 2H), 4.83 (p, $J = 8.9$ Hz, 1H), 4.32 (d, $J = 5.4$ Hz, 2H), 3.10 (dd, $J = 6.5, 3.6$ Hz, 4H), 2.96 (dd, $J = 6.5, 3.6$ Hz, 4H), 2.51 (d, $J = 1.8$ Hz, 2H), 2.33 (s, 3H), 2.10 – 1.85 (m, 4H), 1.71 – 1.58 (m, 2H), 1.24 (s, 9H). LCMS (TQD) m/z calculated 605.38 (M+H)⁺, found 605.37.

CDK4i-10. LCMS (TQD) m/z calculated 617.31 (M+H)⁺, found 617.31.

CDK4i-16. ¹H NMR (400 MHz, DMSO-*d*₆) δ 9.09 (s, 1H), 8.65 (s, 1H), 8.20 – 8.12 (m, 1H), 8.03 – 7.92 (m, 1H), 7.39 (dd, $J = 9.1, 3.0$ Hz, 1H), 6.42 (s, 1H), 5.64 (t, $J = 5.5$ Hz, 1H), 5.43 (s, 1H), 4.82 (p, $J = 8.8$ Hz, 1H), 4.38 (p, $J = 6.6$ Hz, 1H), 4.29 (d, $J = 5.5$ Hz,

2H), 3.01 (dd, $J = 6.3, 3.6$ Hz, 4H), 2.85 (dd, $J = 6.0, 3.6$ Hz, 4H), 2.51 (s, 2H), 2.09 – 1.92 (m, 4H), 1.66 (d, $J = 5.3$ Hz, 2H), 1.27 (d, $J = 6.5$ Hz, 6H), 1.19 (s, 9H). LCMS (TOF) m/z calculated 557.3823 (M+H)⁺, found 557.3821.

CDK4i-17. ¹H NMR (400 MHz, DMSO-*d*₆) δ 9.12 (s, 1H), 8.65 (s, 1H), 8.33 (br s, 1H), 8.17 (d, $J = 9.1$ Hz, 1H), 7.99 (d, $J = 3.0$ Hz, 1H), 7.42 (dd, $J = 9.1, 3.0$ Hz, 1H), 6.41 (s, 1H), 5.58 (s, 1H), 5.14 (t, $J = 5.4$ Hz, 1H), 4.91 (p, $J = 8.9$ Hz, 1H), 4.28 (d, $J = 5.4$ Hz, 2H), 3.11 (dd, $J = 6.8, 3.4$ Hz, 4H), 2.97 (t, $J = 4.9$ Hz, 4H), 2.60 – 2.47 (m, 2H), 2.12 – 1.94 (m, 4H), 1.68 (d, $J = 7.0$ Hz, 2H), 1.53 (s, 9H), 1.18 (s, 9H). LCMS (TOF) m/z calculated 571.3980 (M+H)⁺, found 571.3997.

CDK4i-18. ¹H NMR (400 MHz, DMSO-*d*₆) δ 9.10 (s, 1H), 8.65 (s, 1H), 8.16 (d, $J = 9.1$ Hz, 1H), 7.97 (d, $J = 3.0$ Hz, 1H), 7.39 (dd, $J = 9.2, 3.0$ Hz, 1H), 6.42 (s, 1H), 5.66 (t, $J = 5.5$ Hz, 1H), 5.43 (s, 1H), 4.81 (p, $J = 8.9$ Hz, 1H), 4.51 (p, $J = 7.3$ Hz, 1H), 4.29 (d, $J = 5.4$ Hz, 2H), 3.01 (dd, $J = 6.4, 3.6$ Hz, 4H), 2.85 (dd, $J = 6.1, 3.6$ Hz, 4H), 2.61 – 2.43 (m, 2H), 2.15 – 1.93 (m, 4H), 1.92 – 1.73 (m, 6H), 1.67 (t, $J = 6.0$ Hz, 2H), 1.54 (dq, $J = 9.9, 5.6, 4.5$ Hz, 2H), 1.18 (s, 9H). LCMS (TOF) m/z calculated 583.3980 (M+H)⁺, found 583.3994.

CDK4i-19. ¹H NMR (400 MHz, CDCl₃) δ 8.61 (s, 1H), 8.34 (d, $J = 9.1$ Hz, 1H), 8.01 (d, $J = 2.8$ Hz, 1H), 7.67 (s, 1H), 7.55 (dd, $J = 8.5, 1.1$ Hz, 2H), 7.44 (t, $J = 7.9$ Hz, 2H), 7.34 (dd, $J = 9.1, 3.0$ Hz, 1H), 6.32 (s, 1H), 5.64 (s, 1H), 4.75 (p, $J = 8.9$ Hz, 2H), 4.35 (d, $J = 5.7$ Hz, 3H), 3.17 – 3.04 (m, 8H), 2.72 – 2.57 (m, 2H), 2.10 (d, $J = 14.4$ Hz, 2H), 1.98 (s,

2H), 1.82 – 1.71 (m, 2H), 1.39 (s, 9H). LCMS (TOF) m/z calculated 591.3667 (M+H)⁺, found 591.3656.

CDK4i-21. ¹H NMR (400 MHz, CDCl₃) δ 8.78 (s, 1H), 8.37 (d, J = 9.1 Hz, 1H), 8.04 (d, J = 2.8 Hz, 1H), 7.92 (s, 1H), 7.84 (s, 1H), 7.66 (d, J = 8.8 Hz, 2H), 7.53 (d, J = 8.8 Hz, 2H), 7.36 (dd, J = 9.1, 3.0 Hz, 1H), 6.89 (s, 1H), 5.52 (p, J = 8.8 Hz, 1H), 3.15 (dd, J = 6.5, 3.0 Hz, 4H), 3.10 (dd, J = 6.4, 2.9 Hz, 4H), 2.75 – 2.59 (m, 2H), 2.24 – 2.03 (m, 6H), 1.98 – 1.70 (m, 8H), 1.59 (d, 2H), 1.40 – 1.21 (m, 2H). LCMS (TOF) m/z calculated 590.3350 (M+H)⁺, found 590.3358.

CDK4i-22. ¹H NMR (400 MHz, DMSO-*d*₆) δ 10.58 (s, 1H), 9.58 (s, 1H), 8.92 (s, 1H), 8.20 (d, J = 9.1 Hz, 1H), 8.07 (d, J = 2.5 Hz, 1H), 7.83 (d, J = 8.6 Hz, 2H), 7.51 (dd, J = 9.1, 2.6 Hz, 1H), 7.46 (d, J = 8.5 Hz, 2H), 7.19 (s, 1H), 5.46 (q, J = 8.8 Hz, 1H), 3.35 – 3.27 (m, 4H), 3.25 – 3.19 (m, 4H), 2.83 – 2.70 (m, 2H), 2.68 – 2.57 (m, 2H), 2.55 – 2.46 (m, 2H), 2.37 – 2.20 (m, 1H), 2.13 – 1.93 (m, 5H), 1.78 – 1.58 (m, 2H). LCMS (TOF) m/z calculated 562.3037 (M+H)⁺, found 562.3024.

CDK4i-23. ¹H NMR (400 MHz, DMSO-*d*₆) δ 10.59 (s, 1H), 9.43 (s, 1H), 8.90 (s, 1H), 8.14 (d, J = 9.0 Hz, 1H), 7.99 (d, J = 3.0 Hz, 1H), 7.86 – 7.77 (m, 2H), 7.51 (d, J = 8.8 Hz, 2H), 7.42 (dd, J = 9.2, 3.0 Hz, 1H), 7.19 (s, 1H), 5.47 (p, J = 8.9 Hz, 1H), 3.33 (s, 6H), 3.04 (dd, J = 6.3, 3.6 Hz, 4H), 2.86 (dd, J = 6.1, 3.7 Hz, 4H), 2.58 – 2.42 (m, 2H), 2.09 – 1.93 (m, 4H), 1.76 – 1.62 (m, 2H). LCMS (TOF) m/z calculated 550.3037 (M+H)⁺, found 550.3057.

CDK4i-24. ^1H NMR (400 MHz, $\text{DMSO-}d_6$) δ 9.09 (s, 1H), 8.65 (s, 1H), 8.16 (d, $J = 9.0$ Hz, 1H), 7.97 (d, $J = 3.0$ Hz, 1H), 7.40 (dd, $J = 9.1, 3.0$ Hz, 1H), 6.42 (s, 1H), 5.67 (t, $J = 5.5$ Hz, 1H), 5.41 (s, 1H), 4.80 (p, $J = 8.9$ Hz, 1H), 4.29 (d, $J = 5.5$ Hz, 2H), 3.64 (d, $J = 7.3$ Hz, 2H), 3.02 (dd, $J = 6.4, 3.7$ Hz, 4H), 2.87 (t, $J = 4.9$ Hz, 4H), 2.14 – 1.93 (m, 6H), 1.71 – 1.58 (m, 2H), 1.18 (s, 9H), 0.81 (d, $J = 6.7$ Hz, 6H). LCMS (TOF) m/z calculated 571.3980 ($\text{M}+\text{H}$) $^+$, found 571.3997.

CDK4i-25. ^1H NMR (400 MHz, $\text{DMSO-}d_6$) δ 9.05 (s, 1H), 8.63 (s, 1H), 8.15 (d, $J = 9.1$ Hz, 1H), 7.96 (d, $J = 2.9$ Hz, 1H), 7.73 – 7.66 (m, 2H), 7.39 (dd, $J = 9.1, 3.0$ Hz, 1H), 7.34 (dd, $J = 8.3, 7.1$ Hz, 2H), 7.25 – 7.21 (m, 1H), 6.48 (s, 1H), 6.13 (s, 1H), 5.61 (t, $J = 5.3$ Hz, 1H), 4.97 (p, $J = 8.8$ Hz, 1H), 4.41 (d, $J = 5.2$ Hz, 2H), 3.03 (dd, $J = 6.3, 3.6$ Hz, 4H), 2.87 (dd, $J = 6.3, 3.6$ Hz, 4H), 2.14 – 1.93 (m, 6H), 1.76 – 1.64 (m, 2H), 1.62 (s, 9H). LCMS (TOF) m/z calculated 591.3667 ($\text{M}+\text{H}$) $^+$, found 591.3656.

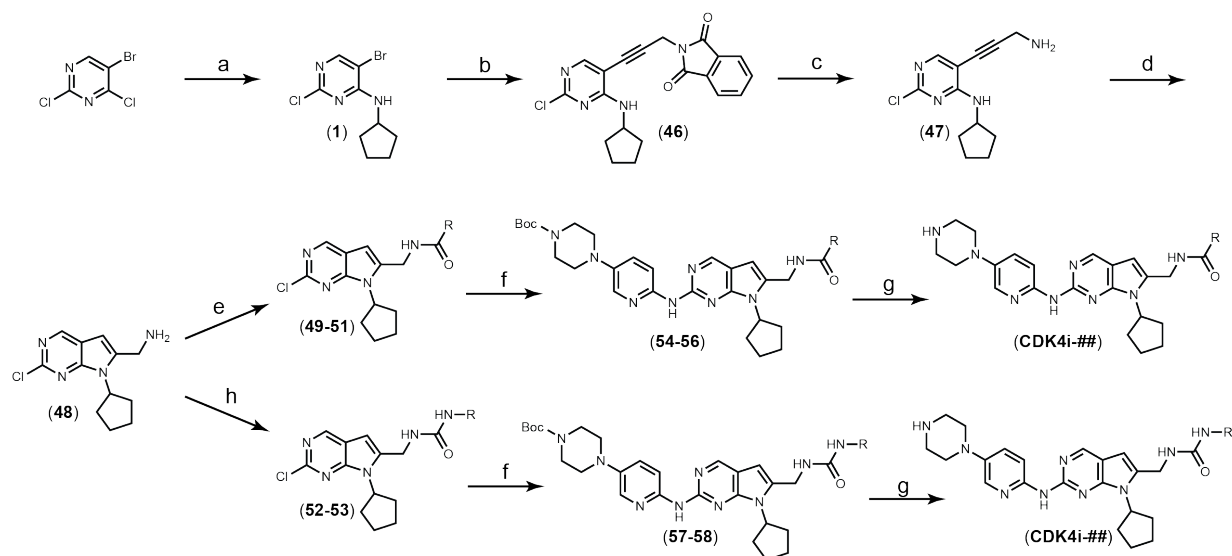
CDK4i-26. ^1H NMR (400 MHz, $\text{DMSO-}d_6$) δ 9.11 (s, 1H), 8.65 (s, 1H), 8.32 (br s, 1H), 8.17 (d, $J = 9.1$ Hz, 1H), 7.99 (d, $J = 2.9$ Hz, 1H), 7.42 (dd, $J = 9.1, 2.9$ Hz, 1H), 6.39 (s, 1H), 5.51 (s, 1H), 5.22 (t, $J = 5.4$ Hz, 1H), 4.91 (p, $J = 8.8$ Hz, 1H), 4.28 (d, $J = 5.3$ Hz, 2H), 3.10 (d, $J = 5.0$ Hz, 4H), 2.98 (d, $J = 5.2$ Hz, 4H), 2.87 – 2.75 (m, 1H), 2.61 – 2.46 (m, 2H), 2.11 – 1.92 (m, 4H), 1.92 – 1.79 (m, 2H), 1.75 – 1.59 (m, 4H), 1.61 – 1.47 (m, 13H). LCMS (TOF) m/z calculated 583.3980 ($\text{M}+\text{H}$) $^+$, found 583.3994.

CDK4i-27. ^1H NMR (400 MHz, DMSO- d_6) δ 9.09 (s, 1H), 8.64 (s, 1H), 8.34 (br s, 1H), 8.17 (d, $J = 9.1$ Hz, 1H), 7.98 (d, $J = 3.0$ Hz, 1H), 7.41 (dd, $J = 9.1, 3.0$ Hz, 1H), 6.40 (s, 1H), 5.53 (s, 1H), 5.21 (t, $J = 5.4$ Hz, 1H), 4.91 (p, $J = 8.9$ Hz, 1H), 4.28 (d, $J = 5.4$ Hz, 2H), 3.08 (d, $J = 5.3$ Hz, 4H), 2.94 (d, $J = 5.3$ Hz, 4H), 2.69 (p, $J = 6.9$ Hz, 1H), 2.60 – 2.43 (m, 2H), 2.12 – 1.92 (m, 4H), 1.76 – 1.60 (m, 2H), 1.53 (s, 9H), 1.12 (d, $J = 6.9$ Hz, 6H). LCMS (TOF) m/z calculated 557.3923 (M+H) $^+$, found 557.3821.

CDK4i-28. ^1H NMR (400 MHz, Chloroform- d) δ 8.65 (s, 1H), 8.39 (d, $J = 9.2$ Hz, 1H), 8.03 (d, $J = 2.9$ Hz, 1H), 7.39 – 7.32 (m, 1H), 6.38 (s, 1H), 5.52 (s, 1H), 4.74 (p, $J = 8.9$ Hz, 1H), 4.33 (s, 2H), 3.61 (s, 3H), 3.26 (s, 8H), 2.92 (p, $J = 6.9$ Hz, 1H), 2.71 – 2.59 (m, 2H), 2.07 (m, 4H), 1.82 – 1.68 (m, 2H), 1.29 (d, $J = 6.9$ Hz, 6H). LCMS (TOF) m/z calculated 515.3354 (M+H) $^+$, found 515.3303.

CDK4i-29. ^1H NMR (400 MHz, Chloroform- d) δ 8.65 (s, 1H), 8.41 (d, $J = 9.1$ Hz, 1H), 8.02 (d, $J = 2.9$ Hz, 1H), 7.39 – 7.34 (m, 2H), 6.40 (s, 1H), 5.67 (d, $J = 2.0$ Hz, 1H), 4.74 (p, $J = 8.9$ Hz, 1H), 4.36 (s, 2H), 3.68 (s, 3H), 3.40 – 3.27 (m, 9H), 2.73 – 2.55 (m, 3H), 2.19 – 1.99 (m, 4H), 1.84 – 1.67 (m, 2H). LCMS (TOF) m/z calculated 473.2884 (M+H) $^+$, found 473.2808.

Scheme 2



(a) 1.3eq cyclopentylamine in 1,4-dioxane, RT, 16h; (b) 2.5eq N-propargylphthalimide, 2.5eq TEA, 0.05eq [1,1'-bis(diphenylphosphino)ferrocene] dichloropalladium(II), 0.05eq CuI in THF, 140°C microwave, 1h; (c) 5eq hydrazine-hydrate in EtOH, 50°C, 2h; (d) 4eq TBAF in THF, 40°C, 40min; (e) 2eq acid, 2.5 eq DIPEA, 2.1eq HATU in THF, 0°C -> RT, 16h; (f) 2eq 2-amino-5-(4-boc-piperziny)-pyridine (60), 4eq cesium carbonate, 0.05eq Pd₂(dba)₃, 0.05eq XANTPhos, 0.05eq tBuXPhos Pd G3 in 1,4-dioxane, 140°C microwave, 1h; (g) 25% (v/v) TFA in DCM, RT, 1h; (h) 1.5eq isocyanate, 2eq TEA in ACN, 0°C -> RT, 2h

2-(3-(2-chloro-4-(cyclopentylamino)pyrimidin-5-yl)prop-2-yn-1-yl)isoindoline-1,3-

dione (46). To a 35mL microwave vial charged anhydrous THF (20mL) under Ar(g) was

added intermediate 1 (1080mg, 3.91mmol) followed by [1,1'-Bis(diphenylphosphino)

ferrocene]dichloropalladium(II), complex with dichloromethane (159.45mg,

0.2000mmol). The solution was purged with Ar(g), then N-propargylphthalamide

(1807.88mg, 9.76mmol) and triethylamine (1360.74uL, 9.76mmol) were added. The

reaction is heated to 140°C for 1hr in a microwave reactor, then cooled, filtered through

celite, and evaporated to dryness under reduced pressure. The resulting oil is purified

by silica chromatography (0-25% EtOAc/hexanes) to yield the pure product as a tan

solid (577.4mg, 38.8% yield). ¹H NMR (400MHz, CDCl₃) δ 8.02 (s, 1H), 7.91 (tt, *J* = 5.0,

2.4 Hz, 2H), 7.80 (td, *J* = 5.3, 2.1 Hz, 2H), 6.11 (d, *J* = 7.1 Hz, 1H), 4.72 (s, 2H), 4.48 (h,

J = 6.9 Hz, 1H), 2.19 – 2.04 (m, 2H), 1.86 (ddd, *J* = 11.8, 5.4, 2.4 Hz, 2H), 1.77 – 1.57

(m, 4H). LCMS (TQD) *m/z* calculated 381.11 (M+H)⁺, found 381.10.

5-(3-aminoprop-1-yn-1-yl)-2-chloro-N-cyclopentylpyrimidin-4-amine (47). To a suspension of intermediate **46** (529.1mg, 1.39mmol) in ethanol (40mL) was added hydrazine hydrate (409 μ L, 8.34mmol, 6eq). The resulting suspension was heated to 50°C for 2hr, then let cool overnight. The suspension was filtered and washed with ethanol, and the resulting filtrate was concentrated under reduced pressure. The resulting solid was dissolved in EtOAc, washed with brine, dried with Na₂SO₄, filtered, and concentrated under reduced pressure to yield pure intermediate (47), which was carried forward without further purification (340.4mg, 97.7% yield). LCMS (TQD) *m/z* calculated 251.11 (M+H)⁺, found 251.11.

(2-chloro-7-cyclopentyl-7H-pyrrolo[2,3-d]pyrimidin-6-yl)methanamine (48). To a solution of intermediate **47** (340.4mg, 1.36mmol) in THF (1mL) was added tetrabutylammonium fluoride (1M solution in THF; 5.43mL, 5.43mmol, 4eq). The resulting solution was heated to 40°C for 1hr, then cooled and concentrated under reduced pressure. The resulting oil was dissolved in EtOAc, washed with saturated NaHCO₃, dried with Na₂SO₄, filtered and concentrated under reduced pressure. The final product was isolated by silica chromatography as an off-white solid (162.5mg, 47.7% yield). ¹H NMR (400MHz, CDCl₃) δ 8.69 (s, 1H), 6.44 (t, *J* = 0.9 Hz, 1H), 4.88 (p, *J* = 8.7 Hz, 1H), 4.11 – 4.03 (m, 2H), 2.47 – 2.31 (m, 2H), 2.20 – 1.98 (m, 4H), 1.83 – 1.64 (m, 2H). LCMS (TQD) *m/z* calculated 251.11 (M+H)⁺, found 251.11.

General procedure for the synthesis of intermediates **49-51**.

To a solution of the appropriate carboxylic acid (2eq) in THF (1mL) under Ar(g) was added diisopropylethylamine (34.7 μ L, 0.199mmol, 2.5eq) and HATU (63.7mg, 0.1675mmol, 2.1eq) and the resulting solution was stirred at RT for 20min. The reaction was cooled to 0°C on ice, then intermediate **48** (20mg, 0.0798mmol) was added portionwise and the reaction was allowed to warm to RT. After 1.5hr, the reaction was concentrated under reduced pressure, dissolved in saturated NaHCO₃, extracted with EtOAc, dried with Na₂SO₄, filtered and concentrated under reduced pressure. The final product was isolated by silica chromatography.

N-((2-chloro-7-cyclopentyl-7H-pyrrolo[2,3-d]pyrimidin-6-yl)methyl)-3-(trifluoromethyl)benzamide (49). White solid (7.0mg, 35.5% yield). LCMS (TQD) *m/z* calculated 423.12 (M+H)⁺, found 423.13.

N-((2-chloro-7-cyclopentyl-7H-pyrrolo[2,3-d]pyrimidin-6-yl)methyl)-3-(dimethylamino)benzamide (50). White solid (27.2mg, 85.7% yield). LCMS (TQD) *m/z* calculated 398.17 (M+H)⁺, found 398.17.

N-((2-chloro-7-cyclopentyl-7H-pyrrolo[2,3-d]pyrimidin-6-yl)methyl)-4-((4-methylpiperazin-1-yl)methyl)benzamide (51). White solid (22.7mg, 60.9% yield). LCMS (TQD) *m/z* calculated 467.23 (M+H)⁺, found 467.21.

General procedure for the synthesis of intermediates **52-53**.

To a solution of intermediate **48** (20mg, 0.0798mmol) in MeCN (1mL) under Ar(g) was added triethylamine (22.2 μ L, 0.160mmol, 2eq) and the resulting solution was chilled to 0°C on ice. The appropriate isocyanate (1.5eq) was then added dropwise, and the reaction was allowed to warm to RT. After 2hr, the reaction was concentrated under reduced pressure, dissolved in saturated NaHCO₃, extracted with EtOAc, dried with Na₂SO₄, filtered and concentrated under reduced pressure. The final product was isolated by silica chromatography.

1-(4-chloro-3-(trifluoromethyl)phenyl)-3-((2-chloro-7-cyclopentyl-7H-pyrrolo[2,3-d]pyrimidin-6-yl)methyl)urea (52). White solid (29.8mg, 79.1% yield). LCMS (TQD) *m/z* calculated 472.09 (M+H)⁺, found 472.08.

1-((2-chloro-7-cyclopentyl-7H-pyrrolo[2,3-d]pyrimidin-6-yl)methyl)-3-(3-(trifluoromethyl)phenyl)urea (53). White solid (29.0mg, 83.0% yield). LCMS (TQD) *m/z* calculated 438.13 (M+H)⁺, found 438.11.

General procedure for the synthesis of intermediates **54-58**.

A solution of the appropriate intermediate **49-53**, intermediate **60** (2eq), and cesium carbonate (4eq) in 1,4-dioxane (1.5mL) in a 10mL microwave vial was purged with Ar(g). Pd₂(dba)₃ (0.05eq), XANTPhos (0.05eq), and tBuXPhos Pd G3 (0.05eq) were then added and the resulting solution was purged with Ar(g). The reaction was heated to 140°C for 45min in a microwave reactor, cooled, filtered through celite, and washed with EtOAc. The resulting filtrate was concentrated under reduced pressure and

separated by silica chromatography to yield intermediates **54-58** that were carried forward without further purification or characterization.

General procedure for the synthesis of **CDK4i-##**

To a solution of intermediates **54-58** in DCM (2mL) was added trifluoroacetic acid (400 μ L) and the resulting solution was stirred until judged complete by LCMS (~1hr). This solution was added to saturated NaHCO₃, extracted with DCM, dried with Na₂SO₄, filtered, and concentrated under reduced pressure. The resulting oil was purified by reverse-phase chromatography (0-70% MeCN/water) and lyophilized to yield the final product.

CDK4i-11. ¹H NMR (400 MHz, DMSO-*d*₆) δ 9.31 (t, *J* = 5.5 Hz, 1H), 9.12 (d, *J* = 1.8 Hz, 1H), 8.65 (s, 1H), 8.25 (d, *J* = 1.9 Hz, 1H), 8.21 (dd, *J* = 7.8, 1.9 Hz, 1H), 8.14 (d, *J* = 9.1 Hz, 1H), 8.00 – 7.90 (m, 2H), 7.75 (t, *J* = 7.8 Hz, 1H), 7.39 (dd, *J* = 9.1, 3.0 Hz, 1H), 6.40 (s, 1H), 4.78 (q, *J* = 8.8 Hz, 1H), 4.71 (d, *J* = 5.4 Hz, 2H), 2.51 (p, *J* = 1.9 Hz, 2H), 2.09 – 1.81 (m, 2H), 1.65 (q, *J* = 6.4, 4.3 Hz, 2H). LCMS (TQD) *m/z* calculated 565.26 (M+H)⁺, found 565.28.

CDK4i-12. ¹H NMR (400 MHz, DMSO-*d*₆) δ 9.09 (s, 1H), 8.89 (t, *J* = 5.6 Hz, 1H), 8.65 (s, 1H), 8.14 (d, *J* = 9.1 Hz, 1H), 7.97 (d, *J* = 3.0 Hz, 1H), 7.39 (dd, *J* = 9.1, 3.1 Hz, 1H), 7.26 (t, *J* = 7.9 Hz, 1H), 7.21 – 7.15 (m, 2H), 6.87 (ddd, *J* = 8.3, 2.7, 1.0 Hz, 1H), 6.35 (s, 1H), 4.80 (p, *J* = 8.9 Hz, 1H), 4.66 (d, *J* = 5.5 Hz, 2H), 3.01 (dd, *J* = 6.3, 3.6 Hz, 4H),

2.93 (s, 6H), 2.85 (dd, $J = 6.2, 3.6$ Hz, 4H), 2.09 – 1.87 (m, 6H), 1.65 (q, $J = 6.5, 5.7$ Hz, 2H). LCMS (TQD) m/z calculated 540.32 (M+H)⁺, found 540.39.

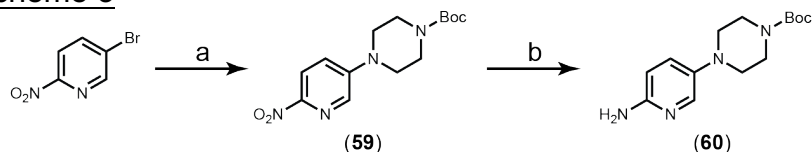
CDK4i-13. ¹H NMR (400 MHz, DMSO-*d*₆) δ 9.10 (s, 1H), 8.97 (t, $J = 5.6$ Hz, 1H), 8.64 (s, 1H), 8.17 (d, $J = 9.1$ Hz, 1H), 7.99 (d, $J = 2.9$ Hz, 1H), 7.86 (dd, $J = 8.3, 3.1$ Hz, 2H), 7.43 (dd, $J = 9.2, 3.0$ Hz, 1H), 7.39 (dd, $J = 8.4, 3.0$ Hz, 2H), 6.36 (s, 1H), 4.86 – 4.77 (m, 1H), 4.67 (d, $J = 5.5$ Hz, 2H), 3.51 (d, $J = 1.7$ Hz, 2H), 3.17 (dd, $J = 6.4, 3.4$ Hz, 4H), 3.11 – 3.00 (m, 4H), 2.44 – 2.24 (m, 10H), 2.17 (s, 3H), 2.08 – 1.88 (m, 4H), 1.74 – 1.58 (m, 2H). LCMS (TQD) m/z calculated 609.38 (M+H)⁺, found 609.32.

CDK4i-14. ¹H NMR (400 MHz, DMSO-*d*₆) δ 9.72 (s, 1H), 9.07 (s, 1H), 8.63 (d, $J = 2.5$ Hz, 1H), 8.20 – 8.07 (m, 2H), 7.96 (d, $J = 3.0$ Hz, 1H), 7.63 (dd, $J = 8.8, 2.6$ Hz, 1H), 7.55 (d, $J = 8.8$ Hz, 1H), 7.39 (dd, $J = 9.1, 3.0$ Hz, 1H), 6.33 (s, 1H), 4.77 (p, $J = 8.7$ Hz, 1H), 4.48 (dd, $J = 11.8, 5.4$ Hz, 2H), 3.48 (t, $J = 5.0$ Hz, 1H), 3.01 (dd, $J = 6.3, 3.6$ Hz, 4H), 2.85 (dd, $J = 6.2, 3.6$ Hz, 4H), 2.51 (s, 2H), 2.01 (s, 4H), 1.68 (s, 2H). LCMS (TQD) m/z calculated 614.24 (M+H)⁺, found 614.19.

CDK4i-15. ¹H NMR (400 MHz, DMSO-*d*₆) δ 9.41 (br s, 1H), 9.07 (s, 1H), 8.63 (s, 1H), 8.15 (d, $J = 9.1$ Hz, 1H), 8.04 – 7.98 (m, 1H), 7.96 (d, $J = 3.0$ Hz, 1H), 7.60 – 7.52 (m, 1H), 7.46 (t, $J = 7.9$ Hz, 1H), 7.39 (dd, $J = 9.1, 3.0$ Hz, 1H), 7.27 – 7.19 (m, 2H), 6.34 (s, 1H), 4.78 (p, $J = 8.7$ Hz, 1H), 4.50 (d, $J = 5.7$ Hz, 2H), 3.01 (dd, $J = 6.3, 3.5$ Hz, 4H), 2.85 (dd, $J = 6.1, 3.7$ Hz, 4H), 2.56 – 2.42 (m, 2H), 2.09 – 1.93 (m, 4H), 1.76 – 1.58 (m, 2H). LCMS (TQD) m/z calculated 580.28 (M+H)⁺, found 580.28.

CDK4i-20. LCMS (TOF) m/z calculated 542.2986 (M+H)⁺, found 542.2987.

Scheme 3



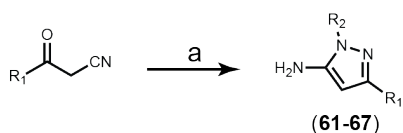
(a) 2eq 4-Boc-piperazine, 2eq K₂CO₃, 0.01eq TBAI in DMSO, 120°C, 16h ; (b) 4eq zinc, 10eq NH₄Cl in 5:3 1,4-dioxane:H₂O, RT, 24h

tert-butyl-4-(6-nitropyridin-3-yl)piperazine-1-carboxylate (59). 5-bromo-2-nitropyridine (1.0g, 4.93mmol), tert-butyl piperazine-1-carboxylate (1.84g, 9.85mmol, 2eq), potassium carbonate (1.36g, 9.85mmol, 2eq), and tetrabutylammonium iodide (18.2mg, 0.049mmol, 0.01eq) were dissolved in DMSO (5mL) and the resulting suspension was heated to 120°C overnight. After cooling, the reaction was added to EtOAc, washed with brine, dried with Na₂SO₄, filtered and concentrated under reduced pressure. The resulting solid was purified by silica chromatography to yield the pure product as a yellow solid (1359.4mg, 89.5% yield). ¹H NMR (400MHz, CDCl₃) δ 8.21 (d, $J = 9.1$ Hz, 1H), 8.16 (d, $J = 3.0$ Hz, 1H), 7.23 (dd, $J = 9.1, 3.1$ Hz, 1H), 3.66 (dd, $J = 6.6, 4.1$ Hz, 4H), 3.48 (dd, $J = 6.5, 4.2$ Hz, 4H), 1.51 (s, 9H). LCMS (TOF) m/z calculated 309.1557 (M+H)⁺, found 309.1523.

tert-butyl-4-(6-aminopyridin-3-yl)piperazine-1-carboxylate (60). To a solution of intermediate 59 (400mg, 1.30mmol) in 1,4-dioxane (5mL) and water (3mL) was added zinc (339mg, 5.19mmol, 4eq) and ammonium chloride (693.9mg, 12.97mmol, 10eq), and the resulting suspension was heated to 50°C for 2hr. The cooled solution was

dissolved in EtOAc/water, filtered through celite, extracted with EtOAc and the organic layers were isolated, dried with Na₂SO₄, filtered and concentrated under reduced pressure. The resulting oil was purified by silica chromatography (0-10% MeOH/DCM) to yield the pure product as a brown solid (128.0mg, 35.5% yield). ¹H NMR (400MHz, CDCl₃) δ 7.80 (d, *J* = 2.9 Hz, 1H), 7.19 (dd, *J* = 8.8, 2.9 Hz, 1H), 6.51 (d, *J* = 8.8 Hz, 1H), 4.21 (s, 2H), 3.59 (t, *J* = 5.1 Hz, 4H), 2.98 (t, *J* = 5.1 Hz, 4H), 1.50 (s, 9H). LCMS (TOF) *m/z* calculated 279.1816 (M+H)⁺, found 279.1808.

Scheme 4



(a) 1.1eq substituted hydrazine hydrochloride, cat. HCl in EtOH, reflux, 16h

General procedure for the synthesis of intermediates 61-67.

To a solution of the appropriate nitrile in isopropanol (3mL) was added hydrochloric acid (10μL, catalytic) and the appropriate hydrazine hydrochloride (1eq). The resulting solution was heated to 120°C overnight, then cooled and evaporated to dryness under reduced pressure. The resulting oil was dissolved in EtOAc, washed with NaHCO₃, and the organic layer was dried with Na₂SO₄, filtered and concentrated under reduced pressure. The resulting oil was purified by silica chromatography to yield intermediates **61-68**.

3-(tert-butyl)-1-isopropyl-1H-pyrazol-5-amine (61). White solid (92.4mg, 63.8% yield). ¹H NMR (400MHz, CDCl₃) δ 5.42 (s, 1H), 4.32 (p, *J* = 6.7 Hz, 1H), 1.45 (d, *J* = 6.7 Hz, 6H), 1.26 (s, 9H). LCMS (TOF) *m/z* calculated 182.1652 (M+H)⁺, found 182.1623.

1,3-di-tert-butyl-1H-pyrazol-5-amine (62). White solid (97.1mg, 62.2% yield). ¹H NMR (400MHz, CDCl₃) δ 5.43 (s, 1H), 1.62 (s, 9H), 1.25 (s, 9H).

3-(tert-butyl)-1-isobutyl-1H-pyrazol-5-amine (63). White solid (206.5mg, 66.2% yield). ¹H NMR (400MHz, CDCl₃) δ 5.42 (s, 1H), 3.70 (d, *J* = 7.5 Hz, 2H), 2.27 – 2.11 (m, 1H), 1.27 (s, 9H), 0.92 (d, *J* = 6.7 Hz, 6H). LCMS (TOF) *m/z* calculated 196.1808 (M+H)⁺, found 196.1815.

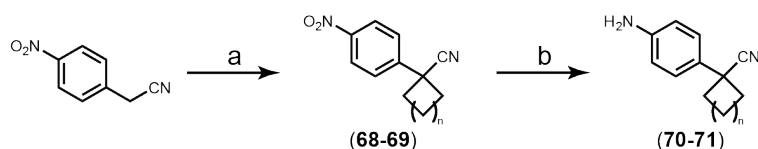
3-(tert-butyl)-1-cyclopentyl-1H-pyrazol-5-amine (64). White solid (55.2mg, 33.3% yield). ¹H NMR (400MHz, CDCl₃) δ 5.42 (s, 1H), 4.43 (p, *J* = 7.5 Hz, 1H), 2.16 – 2.05 (m, 2H), 2.02 (ddd, *J* = 12.8, 5.5, 3.0 Hz, 2H), 1.97 – 1.87 (m, 3H), 1.70 – 1.59 (m, 2H), 1.26 (s, 9H). LCMS (TQD) *m/z* calculated 208.18 (M+H)⁺, found 208.21.

1-(tert-butyl)-3-isopropyl-1H-pyrazol-5-amine (65). White solid (309.2mg, 75.8% yield). ¹H NMR (400MHz, CDCl₃) δ 5.42 (s, 1H), 3.49 (br s, 2H), 2.86 (p, *J* = 6.9 Hz, 1H), 1.63 (s, 9H), 1.21 (d, *J* = 6.9 Hz, 6H).

1-(tert-butyl)-3-cyclopentyl-1H-pyrazol-5-amine (66). White solid (224.7mg, 74.3% yield). ¹H NMR (400MHz, CDCl₃) δ 5.41 (s, 1H), 3.49 (br s, 2H), 2.96 (p, *J* = 8.2, 7.5 Hz, 1H), 2.01 (dddd, *J* = 12.7, 8.2, 6.0, 2.3 Hz, 2H), 1.81 – 1.67 (m, 2H), 1.64 (s, 13H).

1-(tert-butyl)-3-phenyl-1H-pyrazol-5-amine (67). White solid (222.1mg, 56.5% yield). ¹H NMR (400MHz, CDCl₃) δ 7.76 (dt, *J* = 8.1, 1.6 Hz, 2H), 7.37 (td, *J* = 7.0, 1.6 Hz, 2H), 7.28 – 7.23 (m, 1H), 5.93 (s, 1H), 3.59 (br s, 2H), 1.72 (s, 9H). LCMS (TOF) *m/z* calculated 216.1495 (M+H)⁺, found 216.1501.

Scheme 5



(a) 1.05eq dibromoalkane, 2.5eq KOH in DMSO, 16h; (b) 4eq Zn, 10eq NH₄OH in 1,4-dioxane/water, 100°C, 1h

General procedure for the synthesis of intermediates 68-69.

A solution of 2-(4-nitrophenyl)acetonitrile (250mg, 1.54mmol) and the appropriate dibromoalkane (1.05eq) in DMSO (0.5mL) was slowly added to a suspension of potassium hydroxide (216mg, 3.85mmol, 2.5eq) in DMSO (1mL) over 5min, with the reaction vial in a cool water bath to avoid temperatures >30°C. After 24hr, the reaction was added to brine and extracted with EtOAc. The organic layer was dried with Na₂SO₄, filtered, and concentrated under reduced pressure. Intermediates **68-69** were isolated by silica chromatography.

1-(4-nitrophenyl)cyclobutane-1-carbonitrile (68). Red oil (69.3mg, 22.2% yield). ^1H NMR (400MHz, CDCl_3) δ 8.34 – 8.26 (m, 1H), 7.64 (d, $J = 8.9$ Hz, 1H), 2.93 (dddd, $J = 11.0, 8.8, 4.4, 1.9$ Hz, 1H), 2.68 (qd, $J = 9.4, 2.9$ Hz, 1H), 2.54 (dp, $J = 11.7, 8.8$ Hz, 1H), 2.17 (dt, $J = 11.8, 9.1, 4.3$ Hz, 1H).

1-(4-nitrophenyl)cyclohexane-1-carbonitrile (69). Red oil (192.9 mg, 54.3% yield). ^1H NMR (400MHz, CDCl_3) δ 8.28 (d, $J = 8.9$ Hz, 1H), 7.71 (d, $J = 8.9$ Hz, 1H), 2.20 (d, $J = 11.5$ Hz, 2H), 2.03 – 1.74 (m, 6H), 1.42 – 1.23 (m, 1H).

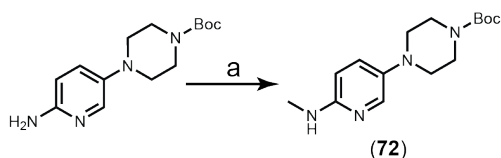
General procedure for the synthesis of intermediates 70-71.

To a solution of the appropriate nitroaniline (1eq) in 1,4-dioxane:water (5:3) was added zinc (4eq) and ammonium chloride (10eq). The resulting solution was heated to 100°C for 1hr, then cooled, filtered through celite and washed with EtOAc. The filtrate was washed with NaHCO_3 , dried with Na_2SO_4 , filtered, and concentrated under reduced pressure. The resulting oil was purified by silica chromatography to yield intermediates 70-71.

1-(4-aminophenyl)cyclobutane-1-carbonitrile (70). Brown solid (36.1mg, 61.2% yield). ^1H NMR (400MHz, CDCl_3) δ 7.32 – 7.23 (m, 1H), 7.23 – 7.15 (m, 2H), 6.91 – 6.81 (m, 1H), 6.76 – 6.66 (m, 2H), 3.76 (br s, 2H), 2.85 – 2.73 (m, 3H), 2.57 (dt, $J = 11.3, 7.1, 2.1$ Hz, 3H), 2.48 – 2.33 (m, 1H), 2.13 – 1.98 (m, 1H).

1-(4-aminophenyl)cyclohexane-1-carbonitrile (71). Brown solid (147.2mg, 87.7% yield). $^1\text{H NMR}$ (400MHz, CDCl_3) δ 7.28 (dd, $J = 6.8, 1.7$ Hz, 1H), 6.75 – 6.68 (m, 2H), 3.72 (br s, 2H), 2.20 – 2.08 (m, 2H), 1.91 – 1.67 (m, 6H), 1.27 (dtd, $J = 13.6, 10.7, 10.0, 6.9$ Hz, 2H). LCMS (TQD) m/z calculated 202.13 ($\text{M}+\text{H}$) $^+$, found 201.14.

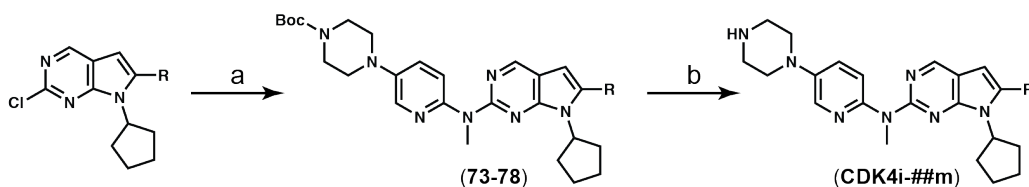
Scheme 6



(a) 5eq CH_2O , 5eq NaOMe, 4eq NaBH_4 in MeOH, 70°C , 3h

tert-butyl-4-(6-(methylamino)pyridine-3-yl)piperazine-1-carboxylate (72). To a solution of intermediate **60** (102mg, 0.366mmol) and formaldehyde (55.0mg, 1.83mmol, 5eq) in methanol (5mL) was added sodium methoxide (25wt% in methanol, 388.2 μL , 1.83mmol, 5eq) and the resulting solution was stirred at RT for 2hrs. The reaction was then cooled to 0°C on ice, and sodium borohydride (55.5mg, 1.47mmol, 4eq) was added portionwise and the reaction was heated to 70°C for 1hr. After cooling, the solution was evaporated to dryness under reduced pressure, dissolved in EtOAc and washed with saturated NaHCO_3 . The organic layer was dried with Na_2SO_4 , filtered, and concentrated under reduced pressure. The final product was isolated by silica chromatography as a light brown solid (102.6mg, 95.8% yield). $^1\text{H NMR}$ (400MHz, CDCl_3) δ 7.85 (d, $J = 2.8$ Hz, 1H), 7.22 (dd, $J = 8.9, 2.9$ Hz, 1H), 6.41 (d, $J = 8.9$ Hz, 1H), 4.31 (s, 1H), 3.63 – 3.55 (m, 4H), 3.00 – 2.94 (m, 4H), 2.92 (s, 3H), 1.50 (s, 9H). LCMS (TOF) m/z calculated 293.1972 ($\text{M}+\text{H}$) $^+$, found 293.1988.

Scheme 7



(a) 2eq 2-aminomethyl-5-(4-boc-piperziny)l-pyridine (**72**), 4eq cesium carbonate, 0.05eq Pd₂(dba)₃, 0.05eq XANTPhos, 0.05eq tBuXPhos Pd G3 in THF, 120°C microwave, 1h; (b) 20% (v/v) TFA in DCM, RT, 1h

General procedure for the synthesis of intermediates **73-78**.

A solution of the appropriate intermediate **6-26**, intermediate **72** (2eq), and cesium carbonate (4eq) in THF (1.5mL) in a 10mL microwave vial was purged with Ar(g). Pd₂(dba)₃ (0.05eq), XANTPhos (0.05eq), and tBuXPhos Pd G3 (0.05eq) were then added and the resulting solution was purged with Ar(g). The reaction was heated to 120°C for 45min in a microwave reactor, cooled, filtered through celite, and washed with EtOAc. The resulting filtrate was concentrated under reduced pressure and separated by silica chromatography to yield intermediates **73-78** that were carried forward without further purification or characterization.

General procedure for the synthesis of **CDK4i-##m**.

To a solution of intermediates **73-78** in DCM (2mL) was added trifluoroacetic acid (400μL) and the resulting solution was stirred until judged complete by LCMS (~1hr). This solution was added to saturated NaHCO₃, extracted with DCM, dried with Na₂SO₄, filtered, and concentrated under reduced pressure. The resulting oil was purified by reverse-phase chromatography (0-70% MeCN/water) and lyophilized to yield the final product.

Ribociclib-m. ^1H NMR (400 MHz, DMSO- d_6) δ 8.70 (s, 1H), 8.34 (br s, 1H), 8.08 (d, J = 2.8 Hz, 1H), 7.46 (d, J = 8.9 Hz, 1H), 7.38 (dd, J = 9.0, 2.9 Hz, 1H), 6.55 (s, 1H), 4.64 (p, J = 8.6 Hz, 1H), 3.55 (s, 3H), 3.15 – 3.10 (m, 4H), 3.04 (s, 6H), 2.93 (m, 4H), 2.26 (dt, J = 16.5, 5.0 Hz, 2H), 1.94 – 1.89 (m, 2H), 1.77 (p, J = 7.1, 5.9 Hz, 2H), 1.54 – 1.50 (m, 2H). LCMS (TOF) m/z calculated 449.2772 (M+H) $^+$, found 449.2757.

CDK4i-5m. ^1H NMR (400 MHz, DMSO- d_6) δ 10.53 (s, 1H), 8.83 (s, 1H), 8.30 (br s, 1H), 8.10 (d, J = 3.0 Hz, 1H), 7.85 – 7.75 (m, 2H), 7.51 – 7.44 (m, 3H), 7.40 (dd, J = 9.0, 3.0 Hz, 1H), 7.13 (s, 1H), 5.41 (p, J = 8.8 Hz, 1H), 3.56 (s, 3H), 3.14 (t, J = 4.9 Hz, 4H), 2.93 (t, J = 5.0 Hz, 4H), 2.37 (dddd, J = 27.3, 8.3, 6.2, 4.2 Hz, 4H), 2.14 – 2.00 (m, 2H), 2.01 – 1.84 (m, 6H), 1.84 – 1.74 (m, 2H), 1.54 (qd, J = 7.4, 6.9, 3.0 Hz, 2H). LCMS (TOF) m/z calculated 590.3350 (M+H) $^+$, found 590.3358.

CDK4i-6m. ^1H NMR (400 MHz, DMSO- d_6) δ 8.56 (s, 1H), 8.32 (br s, 1H), 8.06 (d, J = 3.0 Hz, 1H), 7.47 (d, J = 9.0 Hz, 1H), 7.37 (dd, J = 9.0, 3.0 Hz, 1H), 7.21 – 7.14 (m, 2H), 6.76 – 6.66 (m, 2H), 6.35 (s, 1H), 6.28 (t, J = 5.7 Hz, 1H), 4.70 (p, J = 8.6 Hz, 1H), 4.40 (d, J = 4.8 Hz, 2H), 3.54 (s, 3H), 3.15 (dd, J = 6.5, 3.6 Hz, 4H), 2.97 (dd, J = 6.2, 3.7 Hz, 4H), 2.31 (dq, J = 13.2, 3.8, 2.7 Hz, 4H), 2.02 – 1.89 (m, 4H), 1.88 – 1.73 (m, 6H), 1.54 (q, J = 6.4 Hz, 2H). LCMS (TOF) m/z calculated 576.3558 (M+H) $^+$, found 576.3517.

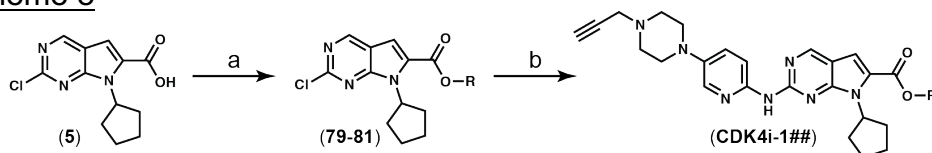
CDK4i-8m. ^1H NMR (400 MHz, DMSO- d_6) δ 8.58 (s, 1H), 8.32 (s, 1H), 8.05 (d, J = 3.0 Hz, 1H), 7.46 (d, J = 8.9 Hz, 1H), 7.36 (dd, J = 9.0, 3.1 Hz, 1H), 6.38 (s, 1H), 5.70 (t, J = 5.6 Hz, 1H), 5.43 (s, 1H), 4.73 (p, J = 8.7 Hz, 1H), 4.27 (d, J = 5.4 Hz, 2H), 3.54 (s, 3H),

3.47 (s, 3H), 3.10 (dd, $J = 6.4, 3.6$ Hz, 4H), 2.91 (t, $J = 5.0$ Hz, 4H), 2.39 – 2.23 (m, 2H), 1.95 (s, 2H), 1.86 – 1.73 (m, 2H), 1.54 (m, 2H), 1.17 (s, 9H). LCMS (TOF) m/z calculated 543.3667 (M+H)⁺, found 543.3668.

CDK4i-17m. ¹H NMR (400 MHz, DMSO-*d*₆) δ 8.58 (s, 1H), 8.32 (s, 2H), 8.05 (d, $J = 3.0$ Hz, 1H), 7.46 (d, $J = 8.9$ Hz, 1H), 7.36 (dd, $J = 9.0, 3.0$ Hz, 1H), 6.36 (s, 1H), 5.57 (s, 1H), 5.11 (t, $J = 5.5$ Hz, 1H), 4.84 (p, $J = 8.7$ Hz, 1H), 4.25 (d, $J = 5.4$ Hz, 2H), 3.54 (s, 3H), 3.11 (t, $J = 4.9$ Hz, 4H), 2.92 (t, $J = 4.9$ Hz, 4H), 2.40 – 2.25 (m, 2H), 2.02 – 1.87 (m, 2H), 1.86 – 1.73 (m, 2H), 1.62 – 1.45 (m, 11H), 1.17 (s, 9H). LCMS (TOF) m/z calculated 585.4136 (M+H)⁺, found 585.4131.

CDK4i-24m. ¹H NMR (400 MHz, DMSO-*d*₆) δ 8.59 (s, 1H), 8.37 (br s, 1H), 8.05 (d, $J = 3.0$ Hz, 1H), 7.45 (d, $J = 9.0$ Hz, 1H), 7.36 (dd, $J = 9.0, 3.0$ Hz, 1H), 6.37 (s, 1H), 5.64 (t, $J = 5.5$ Hz, 1H), 5.40 (s, 1H), 4.74 (p, $J = 8.7$ Hz, 1H), 4.26 (d, $J = 5.4$ Hz, 2H), 3.63 (d, $J = 7.4$ Hz, 2H), 3.54 (s, 3H), 3.08 (d, $J = 5.0$ Hz, 4H), 2.89 (s, 4H), 2.40 – 2.27 (m, 2H), 2.05 (dt, $J = 13.7, 6.8$ Hz, 1H), 1.98 – 1.88 (m, 2H), 1.84 – 1.73 (m, 2H), 1.57 – 1.45 (m, 2H), 1.17 (s, 9H), 0.81 (d, $J = 6.7$ Hz, 6H). LCMS (TOF) m/z calculated 585.4136 (M+H)⁺, found 585.4131.

Scheme 8



(a) 1.5eq alcohol, 2eq DIPEA, 1.1eq HATU in THF, 0°C → 50°C, 1h; (b) 2eq 5-(4-prop-2-yn-1-yl)piperazin-1-ylpyridine-2-amine (**84**), 4eq cesium carbonate, 0.05eq Pd₂(dba)₃, 0.05eq XANTPhos, 0.05eq tBuXPhos Pd G3 in THF, 120°C microwave, 1h

General procedure for the synthesis of intermediates 79-82.

To a solution of intermediate **5** (20mg, 0.0755mmol) in THF (1.5mL) under Ar(g) was added N,N-diisopropylethylamine (26.2μL, 0.15mmol, 2eq) and HATU (31.5mg, 0.0828mmol, 1.1eq) and the resulting solution was stirred at RT for 20min. The reaction was cooled to 0°C on ice, then the appropriate alcohol (1.5eq) was added dropwise. The reaction was allowed to warm to RT and stirred overnight. The reaction was added to brine and extracted with EtOAc. The combined organic layers were dried with Na₂SO₄, filtered, and concentrated under reduced pressure. The resulting oil was purified by silica chromatography (0-20% EtOAc/hexanes) to yield intermediates **79-82** as pure products.

4-methoxyphenyl-2-chloro-7-cyclopentyl-7H-pyrrolo[2,3-d]pyrimidine-6-

carboxylate (79). ¹H NMR (400MHz, CDCl₃) δ 8.98 (s, 1H), 7.59 (s, 1H), 7.23 – 7.12 (m, 2H), 7.03 – 6.93 (m, 2H), 5.82 (p, *J* = 8.8 Hz, 1H), 3.86 (s, 3H), 2.43 (dq, *J* = 12.5, 7.4, 6.0 Hz, 2H), 2.20 – 2.05 (m, 4H), 1.71 (qd, *J* = 8.3, 7.3, 2.6 Hz, 2H). LCMS (TOF) *m/z* calculated 372.1109 (M+H)⁺, found 372.1137.

4-fluorophenyl-2-chloro-7-cyclopentyl-7H-pyrrolo[2,3-d]pyrimidine-6-carboxylate

(80). ¹H NMR (400MHz, CDCl₃) δ 8.99 (s, 1H), 7.61 (s, 1H), 7.27 – 7.12 (m, 4H), 6.99 – 6.89 (m, 2H), 6.85 – 6.76 (m, 2H), 5.82 (p, *J* = 8.8 Hz, 1H), 2.44 (tdd, *J* = 12.4, 9.3, 5.1 Hz, 2H), 2.20 – 2.05 (m, 4H), 1.79 – 1.64 (m, 2H). LCMS (TOF) *m/z* calculated 360.0910 (M+H)⁺, found 360.0905.

2,2,2-trifluoroethyl-2-chloro-7-cyclopentyl-7H-pyrrolo[2,3-d]pyrimidine-6-carboxylate (80). ^1H NMR (400MHz, CDCl_3) δ 8.96 (s, 1H), 7.47 (s, 1H), 5.75 (p, J = 8.7 Hz, 1H), 4.73 (q, J = 8.3 Hz, 2H), 2.50 – 2.30 (m, 2H), 2.12 (dtd, J = 14.3, 8.1, 3.2 Hz, 4H), 1.76 (dtd, J = 10.5, 7.6, 4.5 Hz, 2H). LCMS (TOF) m/z calculated 348.0721 ($\text{M}+\text{H}$) $^+$, found 348.0737.

1,1,1,3,3,3-hexafluoropropan-2-yl-2-chloro-7-cyclopentyl-7H-pyrrolo[2,3-d]pyrimidine-6-carboxylate (81). ^1H NMR (400MHz, CDCl_3) δ 9.01 (s, 1H), 7.59 (s, 1H), 5.98 (hept, J = 5.9 Hz, 1H), 5.73 (p, J = 8.6 Hz, 1H), 2.38 (dt, J = 15.7, 7.8 Hz, 2H), 2.14 (q, J = 13.3, 10.3 Hz, 4H), 1.77 (dd, J = 12.4, 6.3 Hz, 2H). LCMS (TOF) m/z calculated 416.0595 ($\text{M}+\text{H}$) $^+$, found 416.0589.

General procedure for the synthesis of **CDK4i-1##**.

A solution of the appropriate intermediate **79-82**, intermediate **84** (2eq), and cesium carbonate (4eq) in THF (1.5mL) in a 10mL microwave vial was purged with $\text{Ar}(\text{g})$. $\text{Pd}_2(\text{dba})_3$ (0.05eq), XANTPhos (0.05eq), and tBuXPhos Pd G3 (0.05eq) were then added and the resulting solution was purged with $\text{Ar}(\text{g})$. The reaction was heated to 120°C for 45min in a microwave reactor, cooled, filtered through celite, and washed with EtOAc. The resulting filtrate was concentrated under reduced pressure and separated by silica chromatography to yield pure **CDK4i-1##**.

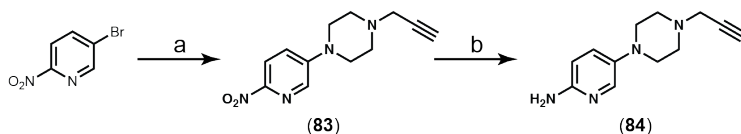
CDK4i-101. White solid (2.5mg, 17.6% yield). LCMS (TOF) m/z calculated 552.2718 ($\text{M}+\text{H}$) $^+$, found 552.2739.

CDK4i-102. White solid (1.4mg, 5.2% yield). LCMS (TOF) m/z calculated 540.2518 (M+H)⁺, found 540.2517.

CDK4i-103. White solid (10.1mg, 31.7% yield). ¹H NMR (400MHz, CDCl₃) δ 8.86 (s, 1H), 8.36 (d, J = 9.1 Hz, 1H), 8.20 (br s, 1H), 8.07 (d, J = 2.8 Hz, 1H), 7.38 (s, 1H), 7.38 – 7.34 (m, 1H), 5.76 (p, J = 8.8 Hz, 1H), 4.69 (q, J = 8.4 Hz, 2H), 3.42 (d, J = 2.4 Hz, 2H), 3.32 – 3.21 (m, 4H), 2.85 – 2.78 (m, 4H), 2.69 – 2.51 (m, 2H), 2.17 – 2.01 (m, 4H), 1.85 – 1.72 (m, 2H). LCMS (TOF) m/z calculated 528.2329 (M+H)⁺, found 528.2348.

CDK4i-104. White solid (2.7mg, 19.8% yield). ¹H NMR (400MHz, CDCl₃) δ 8.89 (s, 1H), 8.35 (d, J = 9.0 Hz, 1H), 8.19 (br s, 1H), 8.07 (d, J = 2.7 Hz, 1H), 7.50 (s, 1H), 7.37 (dd, J = 9.1, 2.9 Hz, 1H), 5.97 (dt, J = 12.2, 6.0 Hz, 1H), 5.71 (p, J = 9.0 Hz, 1H), 3.42 (d, J = 2.4 Hz, 2H), 3.31 – 3.23 (m, 4H), 2.84 – 2.76 (m, 4H), 2.67 – 2.52 (m, 2H), 2.32 (t, J = 2.4 Hz, 1H), 2.17 – 2.02 (m, 4H), 1.86 – 1.71 (m, 2H). LCMS (TOF) m/z calculated 596.2203 (M+H)⁺, found 596.2201.

Scheme 9



(a) 2eq 1-propargyl-piperazine, 2eq K₂CO₃, 0.01eq TBAI in DMSO, 120°C, 16h ; (b) 4eq zinc, 10eq NH₄Cl in 5:3 1,4-dioxane:H₂O, 80°C, 2h

1-(6-nitropyridin-3-yl)-4-(prop-2-yn-1-yl)piperazine (83). 5-bromo-2-nitropyridine (450mg, 2.22mmol), 1-propargyl-piperazine (2.66mmol, 1.2eq), potassium carbonate

(766mg, 5.54mmol, 2.5eq) and tetrabutylammonium iodide (8.19mg, 0.022mmol, 0.01eq) were dissolved in DMSO (3mL) and the resulting suspension was heated to 120°C overnight. The reaction was cooled, added to water and extracted with EtOAc. The combined organic layers were dried with Na₂SO₄, filtered, and concentrated under reduced pressure. The resulting solid was purified by silica chromatography (35-100% EtOAc/hexanes) to yield the product as a yellow solid (472.2mg, 86.5% yield). ¹H NMR (400MHz, CDCl₃) δ 8.19 (d, *J* = 9.2 Hz, 1H), 8.17 (d, *J* = 3.0 Hz, 1H), 7.23 (dd, *J* = 9.2, 3.1 Hz, 1H), 3.56 – 3.50 (m, 4H), 3.43 (d, *J* = 2.4 Hz, 2H), 2.81 – 2.74 (m, 4H), 2.32 (t, *J* = 2.4 Hz, 1H). LCMS (TOF) *m/z* calculated 247.1190 (M+H)⁺, found 247.1201.

5-(4-(prop-2-yn-1-yl)piperazin-1-yl)pyridine-2-amine (84). To a solution of intermediate **83** (472.2mg, 1.92mmol) in 1,4-dioxane (4mL) and water (2.4mL) was added zinc (501.4mg, 7.67mmol, 4eq) and ammonium chloride (1.03g, 19.17mmol, 10eq) and the resulting solution was heated to 80°C for 2hr. The reaction was cooled, filtered through celite, washed with EtOAc and concentrated under reduced pressure. The resulting oil was added to a solution of K₂CO₃, and extracted with EtOAc. The combined organic layers were dried with Na₂SO₄, filtered, and concentrated under reduced pressure. The resulting oil was purified by silica chromatography (10% MeOH/DCM) to yield intermediate **84** as a light brown solid (246.7mg, 59.5% yield). ¹H NMR (400MHz, CDCl₃) δ 7.81 (d, *J* = 2.6 Hz, 1H), 7.20 (dd, *J* = 8.8, 2.9 Hz, 1H), 6.51 (dd, *J* = 8.8, 0.6 Hz, 1H), 4.18 (s, 2H), 3.39 (d, *J* = 2.4 Hz, 2H), 3.16 – 3.09 (m, 4H), 2.80 – 2.71 (m, 4H), 2.30 (t, *J* = 2.4 Hz, 1H). LCMS (TOF) *m/z* calculated 217.1448 (M+H)⁺, found 217.1439.

References

1. Watson, J. D. & Crick, F. H. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171**, 737–738 (1953).
2. Sharp, P. A. The centrality of RNA. *Cell* **136**, 577–580 (2009).
3. Gros, F. *et al.* Unstable ribonucleic acid revealed by pulse labelling of *Escherichia coli*. *Nature* **190**, 581–585 (1961).
4. Brenner, S., Jacob, F. & Meselson, M. An unstable intermediate carrying information from genes to ribosomes for protein synthesis. *Nature* **190**, 576–581 (1961).
5. Scherrer, K., Latham, H. & Darnell, J. E. Demonstration of an unstable RNA and of a precursor to ribosomal RNA in HeLa cells. *Proc. Natl. Acad. Sci. U.S.A.* **49**, 240–248 (1963).
6. Harris, H. & Watts, J. W. The relationship between nuclear and cytoplasmic ribonucleic acid. *Proc. R. Soc. Lond., B, Biol. Sci.* **156**, 109–121 (1962).
7. Scherrer, K. Historical review: the discovery of ‘giant’ RNA and RNA processing: 40 years of enigma. *Trends Biochem. Sci.* **28**, 566–571 (2003).
8. Chow, L. T., Gelinis, R. E., Broker, T. R. & Roberts, R. J. An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell* **12**, 1–8 (1977).
9. Berget, S. M., Moore, C. & Sharp, P. A. Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proc. Natl. Acad. Sci. U.S.A.* **74**, 3171–3175 (1977).
10. Shi, Y. Mechanistic insights into precursor messenger RNA splicing by the spliceosome. *Nat. Rev. Mol. Cell Biol.* **18**, 655–670 (2017).
11. Mortimer, S. A., Kidwell, M. A. & Doudna, J. A. Insights into RNA structure and function from genome-wide studies. *Nat. Rev. Genet.* **15**, 469–479 (2014).
12. Quigley, G. J. & Rich, A. Structural domains of transfer RNA molecules. *Science* **194**, 796–806 (1976).
13. Hori, H. Methylated nucleosides in tRNA and tRNA methyltransferases. *Front Genet* **5**, 144 (2014).
14. Fromont-Racine, M., Senger, B., Saveanu, C. & Fasiolo, F. Ribosome assembly in eukaryotes. *Gene* **313**, 17–42 (2003).
15. Anokhina, M. *et al.* RNA structure analysis of human spliceosomes reveals a compact 3D arrangement of snRNAs at the catalytic core. *The EMBO Journal* **32**, 2804–2818 (2013).
16. Rajkowitsch, L. *et al.* RNA chaperones, RNA annealers and RNA helicases. *RNA Biol* **4**, 118–130 (2007).
17. Jankowsky, E. & Bowers, H. Remodeling of ribonucleoprotein complexes with DExH/D RNA helicases. *Nucleic Acids Res.* **34**, 4181–4188 (2006).
18. Linder, P. & Jankowsky, E. From unwinding to clamping - the DEAD box RNA helicase family. *Nat. Rev. Mol. Cell Biol.* **12**, 505–516 (2011).
19. Singleton, M. R., Dillingham, M. S. & Wigley, D. B. Structure and mechanism of helicases and nucleic acid translocases. *Annu. Rev. Biochem.* **76**, 23–50 (2007).
20. Cobb, J. A. & Bjergbaek, L. RecQ helicases: lessons from model organisms. *Nucleic Acids Res.* **34**, 4106–4114 (2006).
21. Wu, J. I., Lessard, J. & Crabtree, G. R. Understanding the words of chromatin regulation. *Cell* **136**, 200–206 (2009).
22. Jankowsky, E. RNA helicases at work: binding and rearranging. *Trends*

- Biochem. Sci.* **36**, 19–29 (2011).
23. Walker, J. E., Saraste, M., Runswick, M. J. & Gay, N. J. Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *The EMBO Journal* **1**, 945–951 (1982).
 24. Koonin, E. V. A common set of conserved motifs in a vast variety of putative nucleic acid-dependent ATPases including MCM proteins involved in the initiation of eukaryotic DNA replication. *Nucleic Acids Res.* **21**, 2541–2547 (1993).
 25. Tanner, N. K., Cordin, O., Banroques, J., Doère, M. & Linder, P. The Q motif: a newly identified motif in DEAD box helicases may regulate ATP binding and hydrolysis. *Mol. Cell* **11**, 127–138 (2003).
 26. Yang, Q. & Jankowsky, E. The DEAD-box protein Ded1 unwinds RNA duplexes by a mode distinct from translocating helicases. *Nat. Struct. Mol. Biol.* **13**, 981–986 (2006).
 27. Chen, Y. *et al.* DEAD-box proteins can completely separate an RNA duplex using a single ATP. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 20203–20208 (2008).
 28. Liu, F., Putnam, A. & Jankowsky, E. ATP hydrolysis is required for DEAD-box protein recycling but not for duplex unwinding. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 20209–20214 (2008).
 29. Fairman, M. E. *et al.* Protein displacement by DExH/D ‘RNA helicases’ without duplex unwinding. *Science* **304**, 730–734 (2004).
 30. Yang, Q. & Jankowsky, E. ATP- and ADP-dependent modulation of RNA unwinding and strand annealing activities by the DEAD-box protein DED1. *Biochemistry* **44**, 13591–13601 (2005).
 31. Andreou, A. Z. & Klostermeier, D. The DEAD-box helicase eIF4A: paradigm or the odd one out? *RNA Biol* **10**, 19–32 (2013).
 32. Gao, Z. *et al.* Coupling between the DEAD-box RNA helicases Ded1p and eIF4A. *Elife* **5**, 568 (2016).
 33. Soto-Rifo, R. *et al.* DEAD-box protein DDX3 associates with eIF4F to promote translation of selected mRNAs. *The EMBO Journal* **31**, 3745–3756 (2012).
 34. Fuller-Pace, F. V. The DEAD box proteins DDX5 (p68) and DDX17 (p72): multi-tasking transcriptional regulators. *Biochim. Biophys. Acta* **1829**, 756–763 (2013).
 35. Steimer, L. & Klostermeier, D. RNA helicases in infection and disease. *RNA Biol* **9**, 751–771 (2012).
 36. Wu, C.-H., Chen, P.-J. & Yeh, S.-H. Nucleocapsid phosphorylation and RNA helicase DDX1 recruitment enables coronavirus transition from discontinuous to continuous transcription. *Cell Host Microbe* **16**, 462–472 (2014).
 37. Lamichhane, R. *et al.* A DEAD-box protein acts through RNA to promote HIV-1 Rev-RRE assembly. *Nucleic Acids Res.* **45**, 4632–4641 (2017).
 38. Edgcomb, S. P. *et al.* DDX1 is an RNA-dependent ATPase involved in HIV-1 Rev function and virus replication. *J. Mol. Biol.* **415**, 61–74 (2012).
 39. Ariumi, Y. *et al.* DDX3 DEAD-box RNA helicase is required for hepatitis C virus RNA replication. *J. Virol.* **81**, 13922–13926 (2007).
 40. Ariumi, Y. Multiple functions of DDX3 RNA helicase in gene regulation, tumorigenesis, and viral infection. *Front Genet* **5**, 423 (2014).

41. Yasuda-Inoue, M., Kuroki, M. & Ariumi, Y. DDX3 RNA helicase is required for HIV-1 Tat function. *Biochem. Biophys. Res. Commun.* **441**, 607–611 (2013).
42. Yedavalli, V. S. R. K., Neuveut, C., Chi, Y.-H., Kleiman, L. & Jeang, K.-T. Requirement of DDX3 DEAD box RNA helicase for HIV-1 Rev-RRE export function. *Cell* **119**, 381–392 (2004).
43. Soto-Rifo, R., Rubilar, P. S. & Ohlmann, T. The DEAD-box helicase DDX3 substitutes for the cap-binding protein eIF4E to promote compartmentalized translation initiation of the HIV-1 genomic RNA. *Nucleic Acids Res.* **41**, 6286–6299 (2013).
44. Ishaq, M. *et al.* Knockdown of cellular RNA helicase DDX3 by short hairpin RNAs suppresses HIV-1 viral replication without inducing apoptosis. *Mol. Biotechnol.* **39**, 231–238 (2008).
45. Chahar, H. S., Chen, S. & Manjunath, N. P-body components LSM1, GW182, DDX3, DDX6 and XRN1 are recruited to WNV replication sites and positively regulate viral replication. *Virology* **436**, 1–7 (2013).
46. Gu, L., Fullam, A., Brennan, R. & Schröder, M. Human DEAD box helicase 3 couples I κ B kinase ϵ to interferon regulatory factor 3 activation. *Mol. Cell. Biol.* **33**, 2004–2015 (2013).
47. Schröder, M., Baran, M. & Bowie, A. G. Viral targeting of DEAD box protein 3 reveals its role in TBK1/IKKepsilon-mediated IRF activation. *The EMBO Journal* **27**, 2147–2157 (2008).
48. Chen, J.-Y. *et al.* Interaction between SARS-CoV helicase and a multifunctional cellular protein (Ddx5) revealed by yeast and mammalian cell two-hybrid systems. *Arch. Virol.* **154**, 507–512 (2009).
49. Goh, P.-Y. *et al.* Cellular RNA helicase p68 relocalization and interaction with the hepatitis C virus (HCV) NS5B protein and the potential role of p68 in HCV RNA replication. *J. Virol.* **78**, 5288–5298 (2004).
50. Zhou, X. *et al.* DDX5 facilitates HIV-1 replication as a cellular co-factor of Rev. *PLoS ONE* **8**, e65040 (2013).
51. Najji, S. *et al.* Host cell interactome of HIV-1 Rev includes RNA helicases involved in multiple facets of virus production. *Mol. Cell Proteomics* **11**, M111.015313 (2012).
52. Moy, R. H. *et al.* Stem-loop recognition by DDX17 facilitates miRNA processing and antiviral defense. *Cell* **158**, 764–777 (2014).
53. Chable-Bessia, C. *et al.* Suppression of HIV-1 replication by microRNA effectors. *Retrovirology* **6**, 26 (2009).
54. Jangra, R. K., Yi, M. & Lemon, S. M. DDX6 (Rck/p54) is required for efficient hepatitis C virus replication but not for internal ribosome entry site-directed translation. *J. Virol.* **84**, 6810–6824 (2010).
55. Scheller, N. *et al.* Translation and replication of hepatitis C virus genomic RNA depends on ancient cellular proteins that control mRNA fates. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 13517–13522 (2009).
56. Yasuda-Inoue, M., Kuroki, M. & Ariumi, Y. Distinct DDX DEAD-box RNA helicases cooperate to modulate the HIV-1 Rev function. *Biochem. Biophys. Res. Commun.* **434**, 803–808 (2013).
57. Ma, J. *et al.* The requirement of the DEAD-box protein DDX24 for the packaging

- of human immunodeficiency virus type 1 RNA. *Virology* **375**, 253–264 (2008).
58. Fuller-Pace, F. V. DEAD box RNA helicase functions in cancer. *RNA Biol* **10**, 121–132 (2013).
 59. Manohar, C. F., Salwen, H. R., Brodeur, G. M. & Cohn, S. L. Co-amplification and concomitant high levels of expression of a DEAD box gene with MYCN in human neuroblastoma. *Genes Chromosomes Cancer* **14**, 196–203 (1995).
 60. Godbout, R., Packer, M. & Bie, W. Overexpression of a DEAD box protein (DDX1) in neuroblastoma and retinoblastoma cell lines. *Journal of Biological Chemistry* **273**, 21161–21168 (1998).
 61. De Preter, K. *et al.* No evidence for correlation of DDX1 gene amplification with improved survival probability in patients with MYCN-amplified neuroblastomas. *JCO* **23**, 3167–8– author reply 3168–70 (2005).
 62. Weber, A., Imisch, P., Bergmann, E. & Christiansen, H. Coamplification of DDX1 correlates with an improved survival probability in children with MYCN-amplified human neuroblastoma. *JCO* **22**, 2681–2690 (2004).
 63. Germain, D. R. *et al.* DEAD box 1: a novel and independent prognostic marker for early recurrence in breast cancer. *Breast Cancer Res. Treat.* **127**, 53–63 (2011).
 64. Taunk, N. K. *et al.* DEAD box 1 (DDX1) expression predicts for local control and overall survival in early stage, node-negative breast cancer. *Cancer* **118**, 888–898 (2012).
 65. Jones, D. T. W. *et al.* Dissecting the genomic complexity underlying medulloblastoma. *Nature* **488**, 100–105 (2012).
 66. Robinson, G. *et al.* Novel mutations target distinct subgroups of medulloblastoma. *Nature* **488**, 43–48 (2012).
 67. Pugh, T. J. *et al.* Medulloblastoma exome sequencing uncovers subtype-specific somatic mutations. *Nature* **488**, 106–110 (2012).
 68. Kool, M. *et al.* Genome sequencing of SHH medulloblastoma predicts genotype-related response to smoothed inhibition. *Cancer Cell* **25**, 393–405 (2014).
 69. Jiang, L. *et al.* Exome sequencing identifies somatic mutations of DDX3X in natural killer/T-cell lymphoma. *Nat. Genet.* **47**, 1061–1066 (2015).
 70. Ojha, J. *et al.* Identification of recurrent truncated DDX3X mutations in chronic lymphocytic leukaemia. *Br. J. Haematol.* **169**, 445–448 (2015).
 71. Wang, L. *et al.* SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. *N. Engl. J. Med.* **365**, 2497–2506 (2011).
 72. Schmitz, R. *et al.* Burkitt lymphoma pathogenesis and therapeutic targets from structural and functional genomics. *Nature* **490**, 116–120 (2012).
 73. Stransky, N. *et al.* The mutational landscape of head and neck squamous cell carcinoma. *Science* **333**, 1157–1160 (2011).
 74. Seiwert, T. Y. *et al.* Integrative and comparative genomic analysis of HPV-positive and HPV-negative head and neck squamous cell carcinomas. *Clin. Cancer Res.* **21**, 632–641 (2015).
 75. Epling, L. B., Grace, C. R., Lowe, B. R., Partridge, J. F. & Enemark, E. J. Cancer-associated mutants of RNA helicase DDX3X are defective in RNA-stimulated ATP hydrolysis. *J. Mol. Biol.* **427**, 1779–1796 (2015).
 76. Chen, H.-H., Yu, H.-I., Cho, W.-C. & Tarn, W. Y. DDX3 modulates cell adhesion

- and motility and cancer cell metastasis via Rac1-mediated signaling pathway. *Oncogene* **34**, 2790–2800 (2015).
77. Lai, M.-C., Chang, W.-C., Shieh, S.-Y. & Tarn, W.-Y. DDX3 regulates cell growth through translational control of cyclin E1. *Mol. Cell. Biol.* **30**, 5444–5453 (2010).
 78. Botlagunta, M. *et al.* Oncogenic role of DDX3 in breast cancer biogenesis. *Oncogene* **27**, 3912–3922 (2008).
 79. Wortham, N. C. *et al.* The DEAD-box protein p72 regulates ERalpha-/oestrogen-dependent transcription and cell growth, and is associated with improved survival in ERalpha-positive breast cancer. *Oncogene* **28**, 4053–4064 (2009).
 80. Clark, E. L. *et al.* The RNA helicase p68 is a novel androgen receptor coactivator involved in splicing and is overexpressed in prostate cancer. *Cancer Res.* **68**, 7938–7946 (2008).
 81. Wagner, M. *et al.* DDX5 is a multifunctional co-activator of steroid hormone receptors. *Mol. Cell. Endocrinol.* **361**, 80–91 (2012).
 82. Yang, L., Lin, C. & Liu, Z.-R. Phosphorylations of DEAD box p68 RNA helicase are associated with cancer development and cell proliferation. *Mol. Cancer Res.* **3**, 355–363 (2005).
 83. Shin, S., Rossow, K. L., Grande, J. P. & Janknecht, R. Involvement of RNA helicases p68 and p72 in colon cancer. *Cancer Res.* **67**, 7572–7578 (2007).
 84. Wang, R., Jiao, Z., Li, R., Yue, H. & Chen, L. p68 RNA helicase promotes glioma cell proliferation in vitro and in vivo via direct regulation of NF-κB transcription factor p50. *Neuro-oncology* **14**, 1116–1124 (2012).
 85. Bates, G. J. *et al.* The DEAD box protein p68: a novel transcriptional coactivator of the p53 tumour suppressor. *The EMBO Journal* **24**, 543–553 (2005).
 86. Nicol, S. M. *et al.* The RNA helicase p68 (DDX5) is selectively required for the induction of p53-dependent p21 expression and cell-cycle arrest after DNA damage. *Oncogene* **32**, 3461–3469 (2013).
 87. Kozak, M. Influence of mRNA secondary structure on binding and migration of 40S ribosomal subunits. *Cell* **19**, 79–90 (1980).
 88. Svitkin, Y. V. *et al.* The requirement for eukaryotic initiation factor 4A (eIF4A) in translation is in direct proportion to the degree of mRNA 5' secondary structure. *RNA* **7**, 382–394 (2001).
 89. Findlay, G. M., Harrington, L. S. & Lamb, R. F. TSC1-2 tumour suppressor and regulation of mTOR signalling: linking cell growth and proliferation? *Curr. Opin. Genet. Dev.* **15**, 69–76 (2005).
 90. Ruggero, D. *et al.* The translation factor eIF-4E promotes tumor formation and cooperates with c-Myc in lymphomagenesis. *Nat. Med.* **10**, 484–486 (2004).
 91. Barna, M. *et al.* Suppression of Myc oncogenic activity by ribosomal protein haploinsufficiency. *Nature* **456**, 971–975 (2008).
 92. Hsieh, A. C. *et al.* Genetic dissection of the oncogenic mTOR pathway reveals druggable addiction to translational control via 4EBP-eIF4E. *Cancer Cell* **17**, 249–261 (2010).
 93. Ducker, G. S. *et al.* Incomplete inhibition of phosphorylation of 4E-BP1 as a mechanism of primary resistance to ATP-competitive mTOR inhibitors. *Oncogene* **33**, 1590–1600 (2014).
 94. Toriello, H. V., Colley, C. & Bamshad, M. Update on the Toriello-Carey

- syndrome. *Am. J. Med. Genet. A* **170**, 2551–2558 (2016).
95. Snijders Blok, L. *et al.* Mutations in DDX3X Are a Common Cause of Unexplained Intellectual Disability with Gender-Specific Effects on Wnt Signaling. *Am. J. Hum. Genet.* **97**, 343–352 (2015).
 96. Dikow, N. *et al.* DDX3X mutations in two girls with a phenotype overlapping Toriello-Carey syndrome. *Am. J. Med. Genet. A* **173**, 1369–1373 (2017).
 97. Chance, P. F. *et al.* Linkage of the gene for an autosomal dominant form of juvenile amyotrophic lateral sclerosis to chromosome 9q34. *Am. J. Hum. Genet.* **62**, 633–640 (1998).
 98. Németh, A. H. *et al.* Autosomal recessive cerebellar ataxia with oculomotor apraxia (ataxia-telangiectasia-like syndrome) is linked to chromosome 9q34. *Am. J. Hum. Genet.* **67**, 1320–1326 (2000).
 99. Shpargel, K. B. & Matera, A. G. Gemin proteins are required for efficient assembly of Sm-class ribonucleoproteins. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 17372–17377 (2005).
 100. Curmi, F. & Cauchi, R. J. The multiple lives of DEAD-box RNA helicase DP103/DDX20/Gemin3. *Biochem. Soc. Trans.* **46**, 329–341 (2018).
 101. Battle, D. J. *et al.* The SMN complex: an assembly machine for RNPs. *Cold Spring Harb. Symp. Quant. Biol.* **71**, 313–320 (2006).
 102. Nonhoff, U. *et al.* Ataxin-2 interacts with the DEAD/H-box RNA helicase DDX6 and interferes with P-bodies and stress granules. *Mol. Biol. Cell* **18**, 1385–1396 (2007).
 103. Tillotson, J. *et al.* ATP-competitive, marine derived natural products that target the DEAD box helicase, eIF4A. *Bioorg. Med. Chem. Lett.* (2017). doi:10.1016/j.bmcl.2017.07.045
 104. Low, W.-K. *et al.* Inhibition of eukaryotic translation initiation by the marine natural product pateamine A. *Mol. Cell* **20**, 709–722 (2005).
 105. Bordeleau, M.-E. *et al.* Stimulation of mammalian translation initiation factor eIF4A activity by a small molecule inhibitor of eukaryotic translation. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 10460–10465 (2005).
 106. Low, W.-K., Dang, Y., Bhat, S., Romo, D. & Liu, J. O. Substrate-dependent targeting of eukaryotic translation initiation factor 4A by pateamine A: negation of domain-linker regulation of activity. *Chem. Biol.* **14**, 715–727 (2007).
 107. Novac, O., Guenier, A.-S. & Pelletier, J. Inhibitors of protein synthesis identified by a high throughput multiplexed translation screen. *Nucleic Acids Res.* **32**, 902–915 (2004).
 108. Bordeleau, M.-E. *et al.* Functional characterization of IRESes by an inhibitor of the RNA helicase eIF4A. *Nat. Chem. Biol.* **2**, 213–220 (2006).
 109. Lindqvist, L. *et al.* Selective pharmacological targeting of a DEAD box RNA helicase. *PLoS ONE* **3**, e1583 (2008).
 110. Hwang, B. Y. *et al.* Silvestrol and episilvestrol, potential anticancer rocaglate derivatives from *Aglaia silvestris*. *J. Org. Chem.* **69**, 3350–3358 (2004).
 111. Bordeleau, M.-E. *et al.* Therapeutic suppression of translation initiation modulates chemosensitivity in a mouse lymphoma model. *J. Clin. Invest.* **118**, 2651–2660 (2008).
 112. Cencic, R. *et al.* Antitumor activity and mechanism of action of the

- cyclopenta[b]benzofuran, silvestrol. *PLoS ONE* **4**, e5223 (2009).
113. Liu, T. *et al.* Synthetic silvestrol analogues as potent and selective protein synthesis inhibitors. *J. Med. Chem.* **55**, 8859–8878 (2012).
 114. Hawkins, B. C. *et al.* Simplified silvestrol analogues with potent cytotoxic activity. *ChemMedChem* **9**, 1556–1566 (2014).
 115. Kogure, T. *et al.* Therapeutic potential of the translation inhibitor silvestrol in hepatocellular cancer. *PLoS ONE* **8**, e76136 (2013).
 116. Boussemart, L. *et al.* eIF4F is a nexus of resistance to anti-BRAF and anti-MEK cancer therapies. *Nature* **513**, 105–109 (2014).
 117. Bhat, M. *et al.* Targeting the translation machinery in cancer. *Nat Rev Drug Discov* **14**, 261–278 (2015).
 118. Iwasaki, S., Floor, S. N. & Ingolia, N. T. Rocaglates convert DEAD-box protein eIF4A into a sequence-selective translational repressor. *Nature* **534**, 558–561 (2016).
 119. Maga, G. *et al.* Pharmacophore modeling and molecular docking led to the discovery of inhibitors of human immunodeficiency virus-1 replication targeting the human cellular aspartic acid-glutamic acid-alanine-aspartic acid box polypeptide 3. *J. Med. Chem.* **51**, 6635–6638 (2008).
 120. Maga, G. *et al.* Toward the discovery of novel anti-HIV drugs. Second-generation inhibitors of the cellular ATPase DDX3 with improved anti-HIV activity: synthesis, structure-activity relationship analysis, cytotoxicity studies, and target validation. *ChemMedChem* **6**, 1371–1389 (2011).
 121. Mathi, P. *et al.* In-Vitro and in-Silico characterization of Sophora interrupta plant extract as an anticancer activity. *Bioinformation* **10**, 144–151 (2014).
 122. Samal, S. K., Routray, S., Veeramachaneni, G. K., Dash, R. & Botlagunta, M. Ketorolac salt is a newly discovered DDX3 inhibitor to treat oral cancer. *Sci Rep* **5**, 9982 (2015).
 123. Yedavalli, V. S. R. K. *et al.* Ring expanded nucleoside analogues inhibit RNA helicase and intracellular human immunodeficiency virus type 1 replication. *J. Med. Chem.* **51**, 5043–5051 (2008).
 124. Xie, M. *et al.* NZ51, a ring-expanded nucleoside analog, inhibits motility and viability of breast cancer cells by targeting the RNA helicase DDX3. *Oncotarget* **6**, 29901–29913 (2015).
 125. Kumar, R., Ujjinamatada, R. K. & Hosmane, R. S. The first synthesis of a novel 5:7:5-fused diimidazodiazepine ring system and some of its chemical properties. *Org. Lett.* **10**, 4681–4684 (2008).
 126. Kondaskar, A. *et al.* Novel, Broad Spectrum Anti-Cancer Agents Containing the Tricyclic 5:7:5-Fused Diimidazodiazepine Ring System. *ACS Med Chem Lett* **2**, 252–256 (2010).
 127. Bol, G. M. *et al.* Targeting DDX3 with a small molecule inhibitor for lung cancer therapy. *EMBO Mol Med* **7**, 648–669 (2015).
 128. Ito, M. *et al.* Discovery of selective ATP-competitive eIF4A3 inhibitors. *Bioorg. Med. Chem.* **25**, 2200–2209 (2017).
 129. Ito, M. *et al.* Discovery of Novel 1,4-Diacylpiperazines as Selective and Cell-Active eIF4A3 Inhibitors. *J. Med. Chem.* **60**, 3335–3351 (2017).
 130. Iwatani-Yoshihara, M. *et al.* Discovery of Allosteric Inhibitors Targeting the

- Spliceosomal RNA Helicase Brr2. *J. Med. Chem.* **60**, 5759–5771 (2017).
131. Iwatani-Yoshihara, M. *et al.* Discovery and Characterization of a Eukaryotic Initiation Factor 4A-3-Selective Inhibitor That Suppresses Nonsense-Mediated mRNA Decay. *ACS Chem. Biol.* **12**, 1760–1768 (2017).
 132. Shadrick, W. R. *et al.* Discovering new medicines targeting helicases: challenges and recent progress. *J. Biomol. Screen* **18**, 761–781 (2013).
 133. Floor, S. N., Barkovich, K. J., Condon, K. J., Shokat, K. M. & Doudna, J. A. Analog sensitive chemical inhibition of the DEAD-box protein DDX3. *Protein Sci.* **25**, 638–649 (2016).
 134. Chou, T.-F. *et al.* Reversible inhibitor of p97, DBE9, impairs both ubiquitin-dependent and autophagic protein clearance pathways. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 4834–4839 (2011).
 135. Anderson, D. J. *et al.* Targeting the AAA ATPase p97 as an Approach to Treat Cancer through Disruption of Protein Homeostasis. *Cancer Cell* **28**, 653–665 (2015).
 136. Shih, J.-W. *et al.* Critical roles of RNA helicase DDX3 and its interactions with eIF4E/PABP1 in stress granule assembly and stress response. *Biochem. J.* **441**, 119–129 (2012).
 137. Hilliker, A., Gao, Z., Jankowsky, E. & Parker, R. The DEAD-box protein Ded1 modulates translation by the formation and resolution of an eIF4F-mRNA complex. *Mol. Cell* **43**, 962–972 (2011).
 138. Brangwynne, C. P. Phase transitions and size scaling of membrane-less organelles. *J. Cell Biol.* **203**, 875–881 (2013).
 139. Li, P. *et al.* Phase transitions in the assembly of multivalent signalling proteins. *Nature* **483**, 336–340 (2012).
 140. Kato, M. *et al.* Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. *Cell* **149**, 753–767 (2012).
 141. Hubstenberger, A., Noble, S. L., Cameron, C. & Evans, T. C. Translation repressors, an RNA helicase, and developmental cues control RNP phase transitions during early development. *Dev. Cell* **27**, 161–173 (2013).
 142. Lee, C.-S. *et al.* Human DDX3 functions in translation and interacts with the translation initiation factor eIF3. *Nucleic Acids Res.* **36**, 4708–4718 (2008).
 143. la Cruz, de, J., Iost, I., Kressler, D. & Linder, P. The p20 and Ded1 proteins have antagonistic roles in eIF4E-dependent translation in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 5201–5206 (1997).
 144. Sen, N. D., Zhou, F., Ingolia, N. T. & Hinnebusch, A. G. Genome-wide analysis of translational efficiency reveals distinct but overlapping functions of yeast DEAD-box RNA helicases Ded1 and eIF4A. *Genome Res.* **25**, 1196–1205 (2015).
 145. Chuang, R. Y., Weaver, P. L., Liu, Z. & Chang, T. H. Requirement of the DEAD-Box protein ded1p for messenger RNA translation. *Science* **275**, 1468–1471 (1997).
 146. Lorsch, J. R. & Herschlag, D. The DEAD box protein eIF4A. 1. A minimal kinetic and thermodynamic framework reveals coupled binding of RNA and nucleotide. *Biochemistry* **37**, 2180–2193 (1998).
 147. Lopez, M. S., Kliegman, J. I. & Shokat, K. M. The logic and design of analog-

- sensitive kinases and their small molecule inhibitors. *Meth. Enzymol.* **548**, 189–213 (2014).
148. Hertz, N. T. *et al.* A neo-substrate that amplifies catalytic activity of parkinson's-disease-related kinase PINK1. *Cell* **154**, 737–747 (2013).
 149. Gillespie, P. G., Gillespie, S. K., Mercer, J. A., Shah, K. & Shokat, K. M. Engineering of the myosin- β nucleotide-binding pocket to create selective sensitivity to N(6)-modified ADP analogs. *J. Biol. Chem.* **274**, 31373–31381 (1999).
 150. Kapoor, T. M. & Mitchison, T. J. Allele-specific activators and inhibitors for kinesin. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 9106–9111 (1999).
 151. Högbom, M. *et al.* Crystal structure of conserved domains 1 and 2 of the human DEAD-box helicase DDX3X in complex with the mononucleotide AMP. *J. Mol. Biol.* **372**, 150–159 (2007).
 152. Putnam, A. A. & Jankowsky, E. AMP sensing by DEAD-box RNA helicases. *J. Mol. Biol.* **425**, 3839–3845 (2013).
 153. Jankowsky, E. & Putnam, A. Duplex unwinding with DEAD-box proteins. *Methods Mol. Biol.* **587**, 245–264 (2010).
 154. Jamieson, D. J. & Beggs, J. D. A suppressor of yeast *spp81/ded1* mutations encodes a very similar putative ATP-dependent RNA helicase. *Mol. Microbiol.* **5**, 805–812 (1991).
 155. Samal, S. K., Routray, S., Veeramachaneni, G. K., Dash, R. & Botlagunta, M. Ketorolac salt is a newly discovered DDX3 inhibitor to treat oral cancer. *Sci Rep* **5**, 9982 (2015).
 156. Liu, F., Putnam, A. A. & Jankowsky, E. DEAD-box helicases form nucleotide-dependent, long-lived complexes with RNA. *Biochemistry* **53**, 423–433 (2014).
 157. Bantscheff, M. *et al.* Quantitative chemical proteomics reveals mechanisms of action of clinical ABL kinase inhibitors. *Nat. Biotechnol.* **25**, 1035–1044 (2007).
 158. Cencic, R., Galicia-Vázquez, G. & Pelletier, J. Inhibitors of translation targeting eukaryotic translation initiation factor 4A. *Meth. Enzymol.* **511**, 437–461 (2012).
 159. Gibson, S. I., Surosky, R. T. & Tye, B. K. The phenotype of the minichromosome maintenance mutant *mcm3* is characteristic of mutants defective in DNA replication. *Mol. Cell. Biol.* **10**, 5707–5720 (1990).
 160. Kung, C., Kenski, D. M., Krukenberg, K., Madhani, H. D. & Shokat, K. M. Selective kinase inhibition by exploiting differential pathway sensitivity. *Chem. Biol.* **13**, 399–407 (2006).
 161. Bishop, A. C. *et al.* Design of allele-specific inhibitors to probe protein kinase signaling. *Curr. Biol.* **8**, 257–266 (1998).
 162. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
 163. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).
 164. Magnaghi, P. *et al.* Covalent and allosteric inhibitors of the ATPase VCP/p97 induce cancer cell death. *Nat. Chem. Biol.* **9**, 548–556 (2013).
 165. Jarmoskaite, I. & Russell, R. RNA helicase proteins as chaperones and remodelers. *Annu. Rev. Biochem.* **83**, 697–725 (2014).

166. Iwatani-Yoshihara, M. *et al.* Discovery of Allosteric Inhibitors Targeting the Spliceosomal RNA helicase Brr2. *J. Med. Chem.* [acs.jmedchem.7b00461](https://doi.org/10.1021/acs.jmedchem.7b00461) (2017). doi:10.1021/acs.jmedchem.7b00461
167. Lundblad, V. & Struhl, K. *Yeast. Current Protocols in Molecular Biology.* **82**, 13.0:13.0.1–13.0.4 (2008).
168. Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* **8**, 2281–2308 (2013).
169. Galicia-Vázquez, G., Cencic, R., Robert, F., Agenor, A. Q. & Pelletier, J. A cellular response linking eIF4A1 activity to eIF4A1 transcription. *RNA* **18**, 1373–1384 (2012).
170. Pringle, J. R. in *Yeast Cells* (ed. Prescott, D. M.) **12**, 233–272 (Academic Press, 1975).
171. Rakhit, R., Navarro, R. & Wandless, T. J. Chemical biology strategies for posttranslational control of protein function. *Chem. Biol.* **21**, 1238–1252 (2014).
172. Xiong, Y. *et al.* Covalent Guanosine Mimetic Inhibitors of G12C KRAS. *ACS Med Chem Lett* **8**, 61–66 (2017).
173. Singh, J. *et al.* Structure-based design of a potent, selective, and irreversible inhibitor of the catalytic domain of the erbB receptor subfamily of protein tyrosine kinases. *J. Med. Chem.* **40**, 1130–1135 (1997).
174. Ostrem, J. M., Peters, U., Sos, M. L., Wells, J. A. & Shokat, K. M. K-Ras(G12C) inhibitors allosterically control GTP affinity and effector interactions. *Nature* **503**, 548–551 (2013).
175. Cohen, M. S., Zhang, C., Shokat, K. M. & Taunton, J. Structural bioinformatics-based design of selective, irreversible kinase inhibitors. *Science* **308**, 1318–1321 (2005).
176. Garske, A. L., Peters, U., Cortesi, A. T., Perez, J. L. & Shokat, K. M. Chemical genetic strategy for targeting protein kinases based on covalent complementarity. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 15046–15052 (2011).
177. Kung, A. *et al.* A Chemical-Genetic Approach to Generate Selective Covalent Inhibitors of Protein Kinases. *ACS Chem. Biol.* **12**, 1499–1503 (2017).
178. Wilson, D. W. *et al.* A fusion protein required for vesicle-mediated transport in both mammalian cells and yeast. *Nature* **339**, 355–359 (1989).
179. Russell, S. J., Gonzalez, F., Joshua-Tor, L. & Johnston, S. A. Selective chemical inactivation of AAA proteins reveals distinct functions of proteasomal ATPases. *Chem. Biol.* **8**, 941–950 (2001).
180. Floor, S. N., Condon, K. J., Sharma, D., Jankowsky, E. & Doudna, J. A. Autoinhibitory Interdomain Interactions and Subfamily-specific Extensions Redefine the Catalytic Core of the Human DEAD-box Protein DDX3. *J. Biol. Chem.* **291**, 2412–2421 (2016).
181. Kressler, D., la Cruz, de, J., Rojo, M. & Linder, P. Fal1p is an essential DEAD-box protein involved in 40S-ribosomal-subunit biogenesis in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **17**, 7283–7294 (1997).
182. Iggo, R. D. *et al.* p68 RNA helicase: identification of a nucleolar form and cloning of related genes containing a conserved intron in yeasts. *Mol. Cell. Biol.* **11**, 1326–1333 (1991).
183. Snay-Hodge, C. A., Colot, H. V., Goldstein, A. L. & Cole, C. N. Dbp5p/Rat8p is a

- yeast nuclear pore-associated DEAD-box protein essential for RNA export. *The EMBO Journal* **17**, 2663–2676 (1998).
184. Tseng, S. S. *et al.* Dbp5p, a cytosolic RNA helicase, is required for poly(A)+ RNA export. *The EMBO Journal* **17**, 2651–2662 (1998).
 185. Schwöbel, J. A. H. *et al.* Prediction of michael-type acceptor reactivity toward glutathione. *Chem. Res. Toxicol.* **23**, 1576–1585 (2010).
 186. Starrett, J. E. *et al.* Synthesis, oral bioavailability determination, and in vitro evaluation of prodrugs of the antiviral agent 9-[2-(phosphonomethoxy)ethyl]adenine (PMEA). *J. Med. Chem.* **37**, 1857–1864 (1994).
 187. Di Sabato, G. & Jencks, W. Mechanism and Catalysis of Reactions of Acyl Phosphates. I. Nucleophilic Reactions. *J. Am. Chem. Soc.* **83**, 4393–4400. (1961).
 188. Patricelli, M. P. *et al.* Functional interrogation of the kinome using nucleotide acyl phosphates. *Biochemistry* **46**, 350–358 (2007).
 189. Riel-Mehan, M. M. & Shokat, K. M. A crosslinker based on a tethered electrophile for mapping kinase-substrate networks. *Chem. Biol.* **21**, 585–590 (2014).
 190. Niesen, F. H., Berglund, H. & Vedadi, M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat Protoc* **2**, 2212–2221 (2007).
 191. Blair, J. A. *et al.* Structure-guided development of affinity probes for tyrosine kinases using chemical genetics. *Nat. Chem. Biol.* **3**, 229–238 (2007).
 192. Ficner, R., Dickmanns, A. & Neumann, P. Studying structure and function of spliceosomal helicases. *Methods* (2017). doi:10.1016/j.ymeth.2017.06.028
 193. Perriman, R. J. & Ares, M. Rearrangement of competing U2 RNA helices within the spliceosome promotes multiple steps in splicing. *Genes Dev.* **21**, 811–820 (2007).
 194. Semlow, D. R., Blanco, M. R., Walter, N. G. & Staley, J. P. Spliceosomal DEAH-Box ATPases Remodel Pre-mRNA to Activate Alternative Splice Sites. *Cell* **164**, 985–998 (2016).
 195. Martin, R. *et al.* A pre-ribosomal RNA interaction network involving snoRNAs and the Rok1 helicase. *RNA* **20**, 1173–1182 (2014).
 196. Dembowski, J. A., Kuo, B. & Woolford, J. L. Has1 regulates consecutive maturation and processing steps for assembly of 60S ribosomal subunits. *Nucleic Acids Res.* **41**, 7889–7904 (2013).
 197. Berthelot, K., Muldoon, M., Rajkowitsch, L., Hughes, J. & McCarthy, J. E. G. Dynamics and processivity of 40S ribosome scanning on mRNA in yeast. *Mol. Microbiol.* **51**, 987–1001 (2004).
 198. Geissler, R., Golbik, R. P. & Behrens, S.-E. The DEAD-box helicase DDX3 supports the assembly of functional 80S ribosomes. *Nucleic Acids Res.* **40**, 4998–5011 (2012).
 199. Mikhailova, T. *et al.* RNA helicase DDX19 stabilizes ribosomal elongation and termination complexes. *Nucleic Acids Res.* (2016). doi:10.1093/nar/gkw1239
 200. Kumar, P., Hellen, C. U. T. & Pestova, T. V. Toward the mechanism of eIF4F-mediated ribosomal attachment to mammalian capped mRNAs. *Genes Dev.* **30**,

- 1573–1588 (2016).
201. Pradere, U., Garnier-Amblard, E. C., Coats, S. J., Amblard, F. & Schinazi, R. F. Synthesis of nucleoside phosphate and phosphonate prodrugs. *Chem. Rev.* **114**, 9154–9218 (2014).
 202. Mehellou, Y., Balzarini, J. & McGuigan, C. Aryloxy phosphoramidate triesters: a technology for delivering monophosphorylated nucleosides and sugars into cells. *ChemMedChem* **4**, 1779–1791 (2009).
 203. Niphakis, M. J. & Cravatt, B. F. Enzyme inhibitor discovery by activity-based protein profiling. *Annu. Rev. Biochem.* **83**, 341–377 (2014).
 204. Kato, H. & Fujita, T. RIG-I-like receptors and autoimmune diseases. *Curr. Opin. Immunol.* **37**, 40–45 (2015).
 205. Schütz, P. *et al.* Crystal structure of human RNA helicase A (DHX9): structural basis for unselective nucleotide base binding in a DEAD-box variant protein. *J. Mol. Biol.* **400**, 768–782 (2010).
 206. Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).
 207. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675 (2012).
 208. Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 271–281 (2011).
 209. Evans, P. Scaling and assessment of data quality. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 72–82 (2006).
 210. McCoy, A. J. *et al.* Phaser crystallographic software. *J Appl Crystallogr* **40**, 658–674 (2007).
 211. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
 212. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
 213. Schrodinger, LLC. The PyMOL Molecular Graphics System, Version 1.8. (2015).
 214. Chambers, R. W. & Moffatt, J. G. The synthesis of adenosine-5' and uridine-5' phosphoramidates. *J. Am. Chem. Soc.* **80**, 3752–3756 (1958).
 215. Wolfe, A. L. *et al.* RNA G-quadruplexes cause eIF4A-dependent oncogene translation in cancer. *Nature* **513**, 65–70 (2014).
 216. Snider, J., Thibault, G. & Houry, W. A. The AAA+ superfamily of functionally diverse proteins. *Genome Biol.* **9**, 216 (2008).
 217. Miller, J. M. & Enemark, E. J. Fundamental Characteristics of AAA+ Protein Family Structure and Function. *Archaea* **2016**, 9294307 (2016).
 218. Wendler, P., Ciniawsky, S., Kock, M. & Kube, S. Structure and function of the AAA+ nucleotide binding pocket. *Biochim. Biophys. Acta* **1823**, 2–14 (2012).
 219. Firestone, A. J. *et al.* Small-molecule inhibitors of the AAA+ ATPase motor cytoplasmic dynein. *Nature* **484**, 125–129 (2012).
 220. Steinman, J. B. *et al.* Chemical structure-guided design of dynapyrazoles, cell-permeable dynein inhibitors with a unique mode of action. *Elife* **6**, 1169 (2017).

221. McGovern, S. L., Caselli, E., Grigorieff, N. & Shoichet, B. K. A common mechanism underlying promiscuous inhibitors from virtual and high-throughput screening. *J. Med. Chem.* **45**, 1712–1722 (2002).
222. Feng, B. Y. & Shoichet, B. K. A detergent-based assay for the detection of promiscuous inhibitors. *Nat Protoc* **1**, 550–553 (2006).
223. Schafer, K. A. The cell cycle: a review. *Vet. Pathol.* **35**, 461–478 (1998).
224. Harper, J. V. & Brooks, G. The mammalian cell cycle: an overview. *Methods Mol. Biol.* **296**, 113–153 (2005).
225. Vermeulen, K., Van Bockstaele, D. R. & Berneman, Z. N. The cell cycle: a review of regulation, deregulation and therapeutic targets in cancer. *Cell Prolif.* **36**, 131–149 (2003).
226. Morgan, D. O. Cyclin-dependent kinases: engines, clocks, and microprocessors. *Annu. Rev. Cell Dev. Biol.* **13**, 261–291 (1997).
227. Otto, T. & Sicinski, P. Cell cycle proteins as promising targets in cancer therapy. *Nat. Rev. Cancer* **17**, 93–115 (2017).
228. Knudsen, E. S. & Witkiewicz, A. K. The Strange Case of CDK4/6 Inhibitors: Mechanisms, Resistance, and Combination Strategies. *Trends Cancer* **3**, 39–55 (2017).
229. Sánchez-Martínez, C., Gelbert, L. M., Lallena, M. J. & de Dios, A. Cyclin dependent kinase (CDK) inhibitors as anticancer drugs. *Bioorg. Med. Chem. Lett.* **25**, 3420–3435 (2015).
230. Kaur, G. *et al.* Growth inhibition with reversible cell cycle arrest of carcinoma cells by flavone L86-8275. *J. Natl. Cancer Inst.* **84**, 1736–1740 (1992).
231. O'Leary, B., Finn, R. S. & Turner, N. C. Treating cancer with selective CDK4/6 inhibitors. *Nat Rev Clin Oncol* **13**, 417–430 (2016).
232. Finn, R. S. *et al.* Palbociclib and Letrozole in Advanced Breast Cancer. *N. Engl. J. Med.* **375**, 1925–1936 (2016).
233. Hortobagyi, G. N. *et al.* Ribociclib as First-Line Therapy for HR-Positive, Advanced Breast Cancer. *N. Engl. J. Med.* **375**, 1738–1748 (2016).
234. Sledge, G. W. *et al.* MONARCH 2: Abemaciclib in Combination With Fulvestrant in Women With HR+/HER2- Advanced Breast Cancer Who Had Progressed While Receiving Endocrine Therapy. *J. Clin. Oncol.* **35**, 2875–2884 (2017).
235. Hamilton, E. & Infante, J. R. Targeting CDK4/6 in patients with cancer. *Cancer Treat. Rev.* **45**, 129–138 (2016).
236. Toogood, P. L. *et al.* Discovery of a potent and selective inhibitor of cyclin-dependent kinase 4/6. *J. Med. Chem.* **48**, 2388–2406 (2005).
237. Nomanbhoy, T. K. *et al.* Chemoproteomic Evaluation of Target Engagement by the Cyclin-Dependent Kinase 4 and 6 Inhibitor Palbociclib Correlates with Cancer Cell Response. *Biochemistry* **55**, 5434–5441 (2016).
238. Zhao, Q. *et al.* Broad-Spectrum Kinase Profiling in Live Cells with Lysine-Targeted Sulfonyl Fluoride Probes. *J. Am. Chem. Soc.* **139**, 680–685 (2017).
239. Verba, K. A. & Agard, D. A. How Hsp90 and Cdc37 Lubricate Kinase Molecular Switches. *Trends Biochem. Sci.* **42**, 799–811 (2017).
240. 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).
241. Hallett, S. T. *et al.* Differential Regulation of G1 CDK Complexes by the Hsp90-

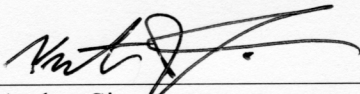
- Cdc37 Chaperone System. *Cell Rep* **21**, 1386–1398 (2017).
242. Russo, A. A., Tong, L., Lee, J. O., Jeffrey, P. D. & Pavletich, N. P. Structural basis for inhibition of the cyclin-dependent kinase Cdk6 by the tumour suppressor p16INK4a. *Nature* **395**, 237–243 (1998).
 243. Jeffrey, P. D., Tong, L. & Pavletich, N. P. Structural basis of inhibition of CDK-cyclin complexes by INK4 inhibitors. *Genes Dev.* **14**, 3115–3125 (2000).
 244. Chu, I. M., Hengst, L. & Slingerland, J. M. The Cdk inhibitor p27 in human cancer: prognostic potential and relevance to anticancer therapy. *Nat. Rev. Cancer* **8**, 253–267 (2008).
 245. Warfel, N. A. & El-Deiry, W. S. p21WAF1 and tumourigenesis: 20 years after. *Curr Opin Oncol* **25**, 52–58 (2013).
 246. McConnell, B. B., Gregory, F. J., Stott, F. J., Hara, E. & Peters, G. Induced expression of p16(INK4a) inhibits both CDK4- and CDK2-associated kinase activity by reassortment of cyclin-CDK-inhibitor complexes. *Mol. Cell. Biol.* **19**, 1981–1989 (1999).
 247. Gustafson, W. C. *et al.* Drugging MYCN through an allosteric transition in Aurora kinase A. *Cancer Cell* **26**, 414–427 (2014).
 248. Chen, H. *et al.* CDK4 protein is degraded by anaphase-promoting complex/cyclosome in mitosis and reaccumulates in early G1 phase to initiate a new cell cycle in HeLa cells. *J. Biol. Chem.* **292**, 10131–10141 (2017).
 249. Zhao, Z. *et al.* Exploration of type II binding mode: A privileged approach for kinase inhibitor focused drug discovery? *ACS Chem. Biol.* **9**, 1230–1241 (2014).
 250. To, T.-L., Zhang, Q. & Shu, X. Structure-guided design of a reversible fluorogenic reporter of protein-protein interactions. *Protein Sci.* **25**, 748–753 (2016).
 251. Polier, S. *et al.* ATP-competitive inhibitors block protein kinase recruitment to the Hsp90-Cdc37 system. *Nat. Chem. Biol.* **9**, 307–312 (2013).
 252. Diehl, J. A., Yang, W., Rimerman, R. A., Xiao, H. & Emili, A. Hsc70 regulates accumulation of cyclin D1 and cyclin D1-dependent protein kinase. *Mol. Cell. Biol.* **23**, 1764–1774 (2003).
 253. Sherr, C. J., Beach, D. & Shapiro, G. I. Targeting CDK4 and CDK6: From Discovery to Therapy. *Cancer Discov* **6**, 353–367 (2016).

Publishing Agreement

It is the policy of the University to encourage the distribution of all theses, dissertations, and manuscripts. Copies of all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible to the public and will preserve these to the best of their abilities, in perpetuity.

Please sign the following statement:

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis, dissertation, or manuscript to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.



Author Signature

June 12, 2018
Date