Combined Computational, Experimental, and Assay-Development Studies of Protein:Protein and Protein:Small Molecule Complexes, with Applications to the Inhibition of Enzymes and Protein:Protein Interactions

by

Marcel Frenkel

Department of Biochemistry
Duke University

Date:________________________
Approved:

____
Bruce R. Donald, Advisor

____
Terrence G. Oas

____
Jane S. Richardson

____
Maria A. Schumacher

____
Kris Wood

Dissertation submitted in partial fulfillment of
the requirements for the degree of Doctor
of Philosophy in the Department of
Biochemistry in the Graduate School
of Duke University

2019

ABSTRACT

Combined Computational, Experimental, and Assay-Development Studies of Protein:Protein and Protein:Small Molecule Complexes, with Applications to the Inhibition of Enzymes and Protein:Protein Interactions

by

Marcel Frenkel

Department of Biochemistry
Duke University

Date:________________________
Approved:

____
Bruce R. Donald, Advisor

____
Terrence G. Oas

____
Jane S. Richardson

____
Maria A. Schumacher

____
Kris Wood

An abstract of a dissertation submitted in partial
fulfillment of the requirements for the degree
of Doctor of Philosophy in the Department of
Biochemistry in the Graduate School of
Duke University

2019

Copyright by
Marcel Frenkel
2019

Abstract

Despite the best efforts of both academia and the pharma industry, most non-resectable cancers remain uncurable and lethal. The world health organization (WHO) believes cancer to be the second leading cause of death worldwide, with roughly 9.6 million deaths in 2018. Meanwhile, the emergence of antimicrobial resistance (AMR), or superbugs, is an increasingly large medical crisis, with estimates as high as 700,000 deaths for 2018 worldwide. This number is increasing rapidly. These unmet medical needs, although distinct, are intimately related by the need for better chemistry and intelligent drug design.

Both AMR and cancer could benefit from the expansion of the druggable proteome through the inhibition of protein-protein interactions (PPIs). PPIs drive both intra- and inter-cellular communication, and therefore their inhibition is vital for disease modulation. Moreover, both AMR and cancer therapeutics suffer from the rapid emergence of drug resistance. Even great drugs that function perfectly at first frequently lose effectiveness a few months later, due to the rapid emergence of drug resistance.

Here, I discuss my contributions towards developing a PPI inhibitor to KRas, the most commonly activated oncogene in cancer. Through the use of OSPREY, a state-of-the-art computational protein and drug design (CPDD) software, and using KRas' native ligand Raf-1 RBD as a starting point, we developed a super-binder with single-digit nanomolar affinity for KRas. The development and validation of this biologic inhibitor required the development of four novel biochemical assays to study binding to KRas and the inhibition of the KRas:Raf interaction.

I also discuss my contributions towards enhancing our ability to predict resistance mutations through the use of OSPREY. This work focused on novel mechanisms of resistance in the dihydrofolate reductase of *Staphylococcus aureus* (SaDHFR). Specifically, we investigated the role of plasmid-borne resistance genes in Staph, as well as the mechanism of resistance due to the emergence of the F98Y and V31L resistance mutations. We discovered a potential new mechanism of resistance based on the formation of a tricyclic NADPH configuration, which we have named chiral evasion.

Finally, I discuss lessons learned from benchmarking OSPREY and share observations that can be used by drug designers using CPDD tools to enhance the accuracy and predictive potential of their results.

In conclusion, a combination of OSPREY and biochemical assays was used towards overcoming two of the largest limitations in drug development that directly affect global human health: the development of PPI inhibitors and overcoming drug resistance. We identified a novel hot-spot in the KRas:Raf interface that can successfully be used to optimize the PPI and develop a biologic inhibitor to KRas. We generated models that explain the mechanism of inhibition of both V31L and F98Y in the context of chiral evasion through a tricyclic NADPH configuration, and we benchmarked OSPREY and observed features that can contribute towards the predictive accuracy of CPDD tools.

Dedication

To my wife, Maria, who continues to bring happiness to my every day and who belief in me inspired me to be better person.

To my sister and best friend Michele, who knew when to kick my ass and when to help me up.

To my father, whose shoulders I stand on and on whose foundation I built everything I ever achieved.

To my Mom, Gilda, whom I miss every single day, and will love forever. You thought me that “eu posso”, thank you for believing in me. Thank you for teaching me to believe in myself. Thank you for every second of every day we got to spend together. I made you a promise, I will keep it.

Disclosure

Mr. Frenkel is a named inventor on intellectual property that Duke University licensed to Gavilán Biodesign, Inc., a company developing services and products related to the research reported. Mr. Frenkel and Duke University have equity ownership in Gavilán. Mr. Frenkel is a founder of Gavilán and on its board of directors.

Contents

Abstract iv

List of Tables xi

List of Figures xii

Acknowledgements xvi

1. Introduction 1

2. Computational Design of a KRas:Effector Protein-Protein Interaction Inhibitor 7

  2.1 An Introduction to Ras 7

    2.1.1 Ras Biochemistry and Biophysics 9

    2.1.2 Ras Signaling and Cancer 13

  2.2 Targeting KRas 16

    2.2.1 Novel CPDD techniques to develop biologic inhibitors of KRas. 20

    2.2.2 Redesign of Raf-1 RBD Domain as a Superbinder of KRas. 21

    A.U. Lowegard, M. S. Frenkel, J.D. Jou, A. A. Ojewole, G.T. Holt, B. R. Donald. Novel, Provable Algorithms for Efficient Ensemble-Based Computational Protein Design and Their Application to the Redesign of the c-Raf-RBD:KRas Protein-Protein Interface 21

      2.2.2.1 Abstract 21

      2.2.2.2 Introduction 23

      2.2.2.3 FRIES/EWAK retrospectively predicted the affect of mutations in c- Raf-RBD on binding to KRas* 28

      2.2.2.4 Prospective Design of Raf Superbinders. 35

      2.2.2.5 Experimental Validation of Mutations in the c-Raf-RBD:KRas Protein-Protein Interface 39

    2.2.2.6 Competition assays for the inhibition of the KRas:Raf-RBD protein-protein interaction. 79

      2.2.2.6.1 SPR competition assay 80

2.2.2.6.2 FP competition assay 86

2.2.2.7 Discussion 90

3. Computational Study of DHFR Inhibitors and the Rise and Mechanism of Resistance in DHFR 92

3.1 Introduction to DHFR 93

3.1.1 Structural characteristics of DHFR 95

3.1.2 Enzymatic Function and Mechanism of DHFR 96

3.1.2 The Need to Go beyond Beta-Lactams 97

3.2 DfrG: a Plasmid-Borne Resistant Gene 98

3.2.1 Introduction 99

3.2.2 Homology Modeling 101

3.2.3 Rationale for Homology Modeling Using OSPREY-Designed Sequence Replacement (ODSR) 102

3.3 The V31L Resistance Mutation in SaDHFR 104

3.3.1 Introduction 105

3.3.2 Methods 107

3.3.2.1 Input Structures 107

3.3.2.2 Homology Model Generation 108

3.3.2.3 Affinity Predictions 109

3.3.3 Results 110

3.3.3.1 Effect of V31L mutation on folate 115

3.3.3.2 Effect of the V31L mutation on TMP 117

3.3.3.3 Effect of the V31L mutation on MTX 119

3.3.3.4 Effect of the V31L mutation on UCP1039 121

3.3.3.5 Effect of the V31L mutation on UCP 1232 123

3.3.4. Conclusion 125

3.4 Chiral Evasion and Resistance Resilience in SaDHFR 128

3.4.1 Reassignment of the crystal structure of t-NADPH in complex with DHFR and R-27 135

3.4.2 Comparison between K* scores and IC50 values.. 137

3.4.3 Structural analysis 141

3.4.4 Discussion and Conclusion 145

4. Improving the Accuracy of Computational Protein and Drug Design 152

4.1 Benchmark of OSPREY 3.0 152

4.1.1 Importance of Accurate Structures as a Starting Point in Computational Predictions 159

4.1.2 The Role of Dynamics and Local Deviation in the Accuracy of Computational Predictions. 161

4.1.3 The Role of Visual Inspection in Computational Protein Design. 166

4.1.4 Role of Rigid Body Translation and Rotation of PPI. 169

5. Conclusions 172

References 177

Biography 217

List of Tables

Table 1: Experimental and computational percent change in binding and rankings as computed using EWAK*. 30

Table 2: Binding affinity and fitting parameters for RBD variants measured by BLI titration experiments. 41

Table 3: Calculated dissociation constants for Raf-RBD variants binding to KRas^{GppNHp}. 77

Table 4: Individual fits for FP competition experiment with unlabeled RBD(WT) as an inhibitor of the KRas:LabledRBD interaction. 90

Table 5: Enzymatic characterization of TMP^{R} enzymes (Data from the Wright Lab, University of Connecticut). 106

Table 6: PDB ID of crystal structures used to generate homology models for DfrG. If structure is not deposited, the lab of origin is indicated on the table. 108

Table 7: Flexible and mutable residue positions for selected homology models. 110

Table 8: Tabulated log10 K* scores for homology models of DfrG. Multiple backbones made by altering the number of minimization steps. 113

Table 9: IC50 data of R-27, S-27 and TMP. Adapted from Table 2 in [17]. 131

Table 10: Partition functions (PF) and K* scores (reported as Log_{10} K* scores) calculated using OSPREY. 138

Table 11: Comparison of experimental vs computational ranking for barnase with its peptide inhibitor barstar, [360, 361]. 155

Table 12: Comparison of experimental vs computational ranking for interleukin 2 (IL-2):IL-2receptorα(IL-2Rα) complex [364]. 156

Table 13: Comparison of experimental vs computational ranking for cytochrome *c*:cytochrome *c* peroxidase complex [362]. 156

Table 14: Comparison of experimental vs computational ranking for interferonα-2 (IFNα2) in complex with interferonα/βreceptor 2 (IFNAR2) [363]. 157

Table 15: Comparison of experimental vs computational ranking for KRas:Raf-RBD complex [187,246,253,254]. 158

List of Figures

Figure 1: Aligned and overlaid structures of KRas bound to GDP in blue (PDB ID 4LPK) and KRas bound to GTP in green (PDB ID 4DSN). 8

Figure 2: Interactions between Ras and REP are mediated nearly completely through the intermolecular beta sheet. 11

Figure 3: Aligned and overlaid structures of the switch 1 regions of Ras-GDP and Ras-GTP (PDB IDs 4LPK and 1HE8). 12

Figure 4: Predicting the effect of mutations in c-Raf-RBD on binding with KRas. 32

Figure 5: Comparing the computational $EWAK^*$ ranking with the experimental ranking for 41 c-Raf-RBD variants. 33

Figure 6: Re-design of c-Raf-RBD residue position 88 from valine to isoleucine. 34

Figure 7: Computational predictions using OSPREY/EWAK$*$ for point mutants in c-Raf-RBD. 37

Figure 8: Re-design of position 88 from valine to tyrosine in the c-Raf-RBD:KRas protein-protein interface. 38

Figure 9: His-Sumo-RBD(C81I, C96M) gene sequence. Sumo sequence in Red, RBD sequence in black. 43

Figure 10: His-Sumo-RBD(C81I,C96M) protein sequence. 43

Figure 11: His-KRas C118S gene sequence 45

Figure 12: His-KRas C118S protein sequence. Thrombin cleavage site shown in red. pI 6.11, MW:21393.01Da, Ext. coefficient: 11920 45

Figure 13: 12% SDS-PAGE gel of purified KRas and RBD variants proteins with ladder in lane 1 for size comparison. 47

Figure 14: MALDI-TOF mass spectrum of RBD(wt). Expected Molecular weight 8841Da; observed molecular weight 8841Da. Peak at 17676Da corresponds to the dimer. 48

Figure 15: MALDI-TOF mass spectrum of RBD(A85K). Expected Molecular weight 8897Da; observed molecular weight 8899Da. Peak at 17793Da corresponds to the dimer. 49

Figure 16: MALDI-TOF mass spectrum of RBD(V88Y). Expected Molecular weight 8904Da; observed molecular weight 8907Da. Peak at 17816Da corresponds to the dimer. 50

Figure 17: MALDI-TOF mass spectrum of RBD(RK). Expected Molecular weight 8939Da; observed molecular weight 8943Da. Peak at 17890Da corresponds to the dimer. 51

Figure 18: MALDI-TOF mass spectrum of RBD(RKY). Expected Molecular weight 9003Da; observed molecular weight 9006Da. Peak at 18018Da corresponds to the dimer. 52

Figure 19: HPAEC chromatogram of KRas bound to different nucleotides or in apo state. 53

Figure 20: Typical BLI experiment. BL (baseline equilibration) 54

Figure 21: BLI experiment of RBD(WT) binding to either KRas^{GDP} (red) or KRas^{GppNHp} (blue). 57

Figure 22: Results for BLI screening of RBD variants. 58

Figure 23: BLI sensor plate showing drift subtraction. 60

Figure 24: BLI titration curve for RBD(WT) concentrations show on graph by titrations curves. 63

Figure 25: BLI titration curve for RBD(A85K) concentrations show on graph by titrations curves. 64

Figure 26: BLI titration curve for RBD(RK) concentrations show on graph by titrations curves. Affinity and accuracy of fit also shown on graph. 65

Figure 27: BLI titration curve for RBD(V88Y) concentrations show on graph by titrations curves. Affinity and accuracy of fit also shown on graph. 66

Figure 28: BLI titration curve for RBD(RKY) concentrations show on graph by titrations curves. Affinity and accuracy of fit also shown on graph. 67

Figure 29: Testing for non-specific binding. RBD(WT) binding to His-SumoRBD(WT) (Blue) and His-Hip (Red). 68

Figure 30: RBD shown as a Richardson diagram, with Cys95 shown as sticks. 71

Figure 31: Representative MALDI-TOF mass spectrum of dye conjugation. 72

Figure 32: FP titration curves. 76

Figure 33: SPR binding curves (colored curves) and fit (black curves) for KRasGppNHp binding to immobilized His-SumoRBD(WT)....84

Figure 34: SPR binding curves for 200nM KRas^{GppNHp} injected after incubation with different concentration of His-SumoRBD(WT) for 10 minutes. 85

Figure 35: SPR competition assay results. 86

Figure 36: FP competition assay....89

Figure 37 : PDB ID 3FRD, *Staphylococcus aureus* DHFR (SaDHFR) shown as a Richardson diagram. Folate and NADPH are shown as sticks. 96

Figure 38: Overlay of DfrB (dark blue) with 1232 (orange) and a DfrG homology model (purple) with 1232(light blue). Active site residues shown as sticks. 101

Figure 39: Alignment for structural comparison of DHFR of and *Staphylococcus* aureus and *Bacillus anthracis*. 104

Figure 40: Chemical structure of Folate (FOL) and Methotrexate (MTX). Note the nearly identical chemical structures. 107

Figure 41: log10 K* scores for DfrG WT and V31L bound to ligands in models generated using 1K all-atom minimization steps. Data shown in Table 6. 113

Figure 42: Model of FOL bound to DfrG, *Staphylococcus* backbone. 117

Figure 43: Model of TMP bound to DfrG, *Bacillus* backbone, 1000 minimization steps. . 119

Figure 44: Model of MTX bound to DfrG, *Bacillus* backbone, 1, 000 minimization steps. 121

Figure 45: Model of UCP1039 bound to DfrG, *Staphylococcus* backbone, 1, 000 minimization steps....123

Figure 46: Model of UCP1232 bound to DfrG, *Staphylococcus* backbone, 1, 000 minimization steps....125

Figure 47: Structure of Folate and Antifolate Inhibitors. Non-racemic PLA (INCA) enantiomers R-27 and S-27 are shown as is the traditional antifolate TMP 129

Figure 48 : Electron density map of PDB ID 4XEC....132

Figure 49: Structure of DHFR bound to different enantiomers of NAHD and different enantiomers of a PLA. 133

Figure 50: Bar graphs of $1/IC_{50}$ ($\log_{10}$) values (left) and $\log_{10}$ K* scores (right) showing the change effect of the F98Y mutation on inhibitor potency and binding. 140

Figure 51: OSPREY designed complexes showing INCA binding to DHFR with different NADPH enantiomers. 142

Figure 52: OSPREY designed complexes for R-27 bound and unbound WT and F98Y DHFR. 144

Figure 53: OSPREY designed complexes for S-27 bound and unbound WT and F98Y DHFR. 145

Figure 54: Analysis of structural quality measured versus Spearman ρ value. 161

Figure 55: Richardson diagram representation of IL-2receptorα(IL-2Rα) bound to multiple ligands [359, 364, 371, 372]. 165

Figure 56: OSPREY designed KRas:RBD(T57K) mutation. 168

Figure 57: Rotation and translation of rigid bodies at PPI. 171

List of Abbreviations

BBK*	Branch and Bound over K*
BLI	Bio-Layer Interferometry
CPDD	Computational Protein and Drug Design
DHFR	Dihydro Folate Reductase
EWAK*	Energy Window Approximation to K*
FP	Fluorescence Polarization
FRIES	Fast Removal of Inadequately Energized Sequences
GAPs	GTPase-activating proteins
GDP	Guanosine Diphosphate
GEFs	Guanine Nucleotide Exchange Factors
GMEC	Global Minimum Energy Conformation
GppNHp	5'-Guanylyl Imidodiphosphate
GTP	Guanosine Triphosphate
HVR	Hypervariable Region
KRas	Kirsten Rat Sarcoma Viral Oncogene Homologue
MTX	Methotrexate
NADPH	Nicotinamide Adenine Dinucleotide Phosphate
OSPREY	Open-Source Protein Redesign for You
PDB	Protein Data Bank

PPI	Protein-Protein Interaction
c-Raf	Proto-oncogene Serine/Threonine-Protein Kinase
RBD	Ras Binding Domain
REPs	Ras Effector Proteins
RMSD	Root Mean Square Deviation
SPR	Surface Plasmon Resonance

Acknowledgements

To my mentor and friend Bruce for welcoming me when I needed to start again, and mentoring and guiding me to success. To Terry, for the inspiring conversations, mentorship and for his undying support.

To my colleagues in the Donald lab for their brilliance and inspiration. To my partner in research Anna, who taught me so much, and who underwent this amazing yet sometimes difficult journey with me.

To Siyu and Graham, who both were amazing partners and willing to listen to my sometimes-crazy ideas as we investigated new mechanism of resistance in DHFR.

To my family and my friends, thank you for helping me make this possible and believing in me and accepting me. I think this marks that “one fine day” we talked about!

To Rachel for her patience and help with the KRas project.

To Ben for his help getting me started with the KRas project and continued support throughout the project.

To my partners, Mark, Bruce, and Jonathan, for the adventure that we have embarked on together and the amazing days ahead. I could not ask for better people to be by my side.

1. Introduction

Unfortunately, drug discovery is inefficient. In 2014 it was estimated that for every FDA approved drug, over $2B in risk-adjusted costs will have been invested towards the development of that drug [1]. In the wake of the genome revolution [2], the paradigm has steadily shifted from the generation of drugs of unknown or unclear mechanism to precision medicine, in which drug discovery targets particular proteins [3, 4]. Towards that goal, many technologies have recently emerged allowing for quick and effective discovery of disease-modulating proteins [5]. This has resulted in some success, especially in the area of oncology [3] and recently in immune-oncology [6].

As the promise of precision medicine started to become clear, two factors quickly emerged as clear limitations. First, not all targets are currently druggable. In fact it is expected that ~80% of the proteome is not druggable, because most proteins lack a hydrophobic invagination or “pocket” that is amenable to traditional pharmacological intervention [7-9]. Finding ligands that fill such pockets is the mainstay of current drug discovery, but such pockets typically are only found in proteins with a natural small-molecule ligand or substrate. Many medicinally relevant proteins do not have this property. In particular, as screens for disease-modulating proteins increasingly highlight the importance of proteins that act through protein-protein interactions (PPI) [10], the need to develop effective methods to designing PPI inhibitors has become of vital medical importance.

Second, the emergence of drug resistance limits the effectiveness of therapeutics [11, 12]. Even therapeutics that are initially potent quickly lose effectiveness as cells either evolve or are selected that can overcome the effect of treatment [13-21]. Although some drug resistance mechanisms are conserved across different drugs, it has recently become clear that different drugs exhibit different resilience patterns to drug resistance [22, 23]. Being able to computationally predict what drugs will be susceptible or resilient to drug resistance and understanding the mechanism by which that happens will be paramount in the development of novel inhibitors with increased durability of response. Combined, the difficulty of developing PPI inhibitors and the emergence of drug resistance represent barriers to what modern medicine can achieve and represent some of the biggest medical challenges today.

Despite very different biological manifestations, I hypothesize that both challenges emerge from the same physical and chemical origin: our lack of understanding of inter-molecular interactions and our inability to search and sample sufficient chemical and conformational space accurately.

The difficulty of targeting a PPI was previously believed to be due to the size of the interface [7, 24, 25]. Contrary to a pocket, which allows for a large density of contacts, it was believed a PPI required interaction over a large surface area to generate sufficient contacts for high affinity [7, 9, 24-26]. However, the discovery of hot-spots, small areas within that interface that contribute most of the free energy of binding [27-34], opened the possibility of smaller PPI inhibitors. In fact, these hot-spots are small enough for conventional small molecule intervention, and a few successes have been observed in targeting hot-spots [35-60]. Identification and characterization of hot-spot within a PPI is therefore vital to the understanding and development of effective small-molecule PPI inhibitors. A technology that could allow us to understand and optimize the inter-molecular interactions within a PPI interfaces would not only allow for the discovery and identification of hot-spots, but also allow for the exploration of new hot-spots within the PPI that could then be used for the generation of novel PPI inhibitors.

The development of resistance (such as drug resistance) is a result of evolution [61]. Although the rate of mutation and the heterogeneity of the genetic pool of the target cell population play a critical role, the chemistry of the drug drives this process and acts as the selective pressure driving evolution [18-20, 62-66]. Until recently, it would have been impossible to even attempt to tackle this problem, as there was no way to predict how a cell would respond to the selective pressure imposed by the drug [66]. However, the Donald lab showed that at least resistance mutations that are in a target itself ("on-target" mutations) can be accurately predicted [19, 67], through a understanding of the inter-atomic interactions between the target and native ligand and between the protein and the drug. This has allowed us to predict evolution, and like a chess player, what you can predict you can overcome.

Computational protein and drug design (CPDD), as implemented in software suites like OSPREY [68], offer the promise of overcoming drug resistance and PPI druggability limitations. OSPREY uses mathematically guaranteed algorithms to accurately search large chemical and conformational spaces. This search discovers new chemistry and allows us a physics-based understanding of inter-molecular interactions

[69-80]. Previously, OSPREY has been able to design PPI inhibitors [60, 81, 82] and predict the emergence of drug resistance, both retrospectively and prospectively [16, 18-20, 22]. This thesis builds on that body of work.

In chapter 1 we present the development of a biologic to inhibit the KRas:Effector protein-protein interaction. The chapter highlights the development of novel and useful assays for measuring Ras:Effector interaction and inhibition while also describing the computational design of a RBD variant, RBD(RKY), with higher known affinity for KRas than any previous variant. This variant provides new insight into the hot-spots in the KRas:Raf interface, and may contribute towards the future development of small-molecule or peptide inhibitors of the KRas:Raf PPI. Moreover, given that many other groups have worked on redesigning RBD to obtain superbinders of KRas^{GTP}, these results indicate that the methods used in OSPREY may constitute an advantage over other CPDD software suites.

Select work from chapter 2 is being used towards a manuscript that I am co-authoring with Anna U. Lowegard:

A. U. Lowegard, M. S. Frenkel, J.D. Jou, A. A. Ojewole, G.T. Holt, B. R. Donald. Novel, Provable Algorithms for Efficient Ensemble-Based Computational Protein Design and Their Application to the Redesign of the c-Raf-RBD:KRas Protein-Protein Interface.

In chapter 3 we continue our study of the mechanism of trimethoprim resistance in *Staphylococcus aureus*. We describe a novel method for, and the structure of, a homology model of the plasmid-borne, trimethoprim-resistant DHFR gene variant (DfrG) [16]. Our model produces accurate predictions and allows for insight into the mechanism of inherent resistance of DrfG while also allowing us to study the effects of the V31L resistance mutation [19] on the gene. Finally, we continue to explore a novel mechanism of resistance, initially identified by Pablo Gainza in the Donald lab, that operates through a novel mechanism of NADPH configurational modulation [17]. Expanding the type of resistance mechanism can be identified and characterized with CDPP techniques. Meanwhile, we also describe the generation of new antifolates, made to overcome our predicted mechanism of resistance that were developed by our University of Connecticut collaborators. This work has resulted in three manuscripts, the first of which has already been submitted for peer review:

S. M. Reeve, D. Si, J. Krucinska, Y. Yan, K. Viswanathan, S. Wang, G.T. Holt, M. S. Frenkel, A. A. Ojewole, A. Estrada, S. S. Agabiti, J. B. Alverson, N. D. Gibson, N. D. Priestly, A. J. Wiemer, B. R. Donald, D. L. Wright (2019). Toward Broad Spectrum DHFR inhibitors Targeting Trimethoprim Resistant Enzymes Identified in Clinical Isolates of Methicillin-Resistant Staphylococcus aureus. BioRxiv (2019) [83]

S. M. Reeve, S. Wang, G.T. Holt, M. S. Frenkel, A. A. Ojewole, B. R. Donald, D. L. Wright (2019). Development of a Broad Spectrum Antifolate Able to Overcome Acquired Resistance Mutations in the Plasmid Borne Inherently Trimethoprim Resistant DfrG gene in Staphylococcus aureus.

S. Wang, S. M. Reeve, A.A. Ojewole, M. S. Frenkel, G.T. Holt, P Gainza, D. L. Wright, B. R. Donald. Structural and computational study of the etiology and mechanism of enantiomeric antifolates resistance in S.aureus.

In chapter 4, I describe my contribution to a benchmark of the accuracy of the new release of OSPREY, OSPREY 3.0 [68]. Moreover, a discussion of what factors affect the predictive power of CPDD and how a knowledgeable designer can maximize the chances of a successful CPDD campaign are described:

M. A. Hallen, J. W. Martin, A. Ojewole, J.D. Jou, A.U. Lowegard, P.Gainza, H. M. Nisonoff, A. Mukund, S. Wang, G.T. Holt, D. Zhou and E. Dowd, B.R. Donald. OSPREY 3.0: Open-Source Protein Redesign for You, with Powerful New Features. Journal of Computational Chemistry. 2018. [68]

2. Computational Design of a KRas:Effector Protein-Protein Interaction Inhibitor

Ras is the holy grail of oncogenic targets; no target has proven so evasive to current drug discovery techniques. This has made Ras the flagbearer of the undruggable class of proteins. Moreover, Ras mutations represent the single most prominent driver of cancer, and cancers with Ras mutations are refractory to therapy. I joined the Donald lab in no small part due to my belief that computational protein and drug design methods, especially those employing provable algorithms with strong biophysical models, could potential change the paradigm in drug discovery for difficult targets like Ras. Work in this chapter was done in very close collaboration with Anna U. Lowegard, who performed all the computational work for this chapter.

2.1 An Introduction to Ras

The American Cancer Society predicts that 1,650,000 Americans will be diagnosed with cancer in 2015, and of these more than a third will die from the disease [84]. This deadly disease is characterized by the dysregulation of cell survival, proliferation, angiogenesis, and metabolism. These cellular functions are normally tightly controlled through the regulation of signal transduction pathways. One of the key protein families involved in signal transduction is the p21 Ras family of GTPase proteins. These are cytosolic, membrane-localized 21 kDa proteins that adopt an α/β fold, with five α-helices surrounding a central β-sheet with five parallel and one anti-parallel strand (Figure 1).

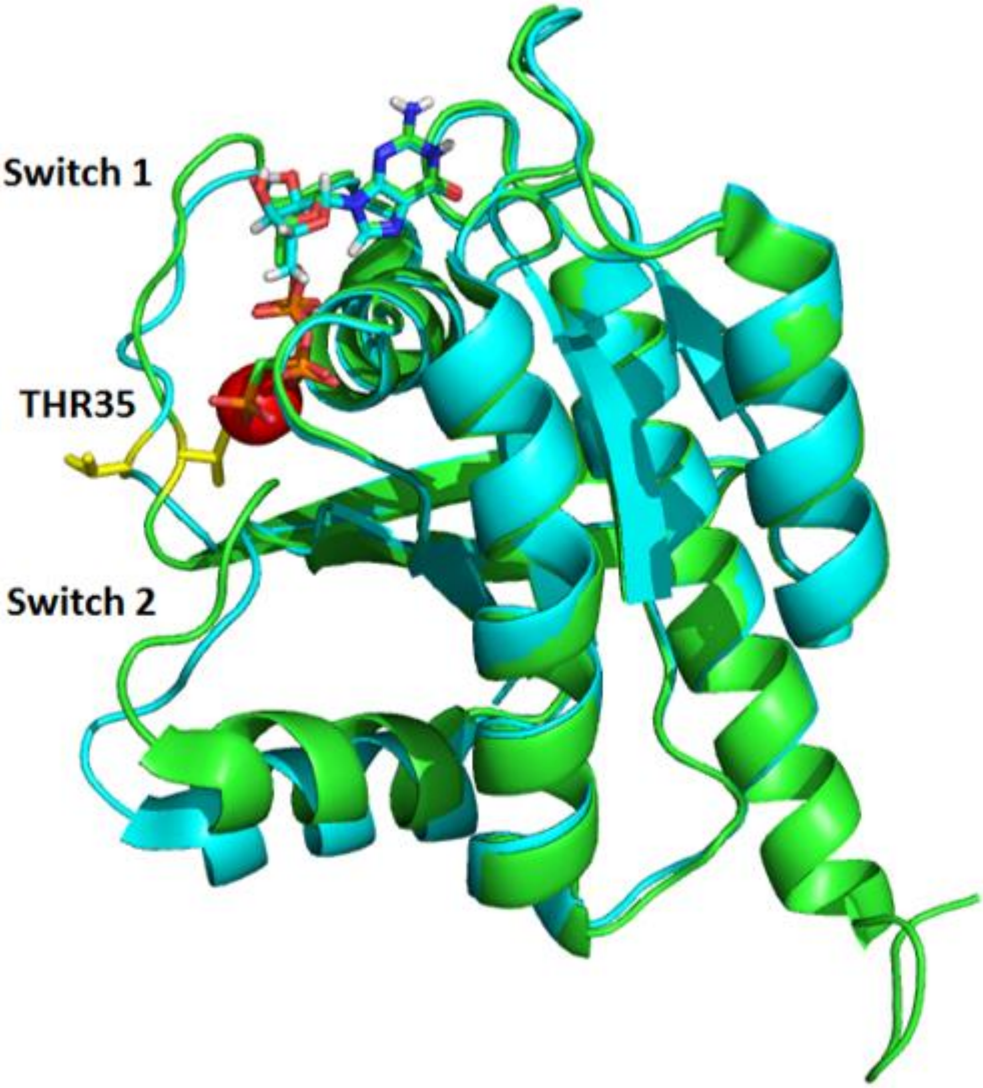


Figure 1: Aligned and overlaid structures of KRas bound to GDP in blue (PDB ID 4LPK) and KRas bound to GTP in green (PDB ID 4DSN). Thr 35 shown as sticks in yellow. Magnesium atom shown as red sphere. GCP and GDP shown as sticks.

2.1.1 Ras Biochemistry and Biophysics

Ras proteins act as "molecular switches" for signal transduction pathways by switching from "on" to "off" forms by hydrolyzing GTP into GDP (Figure 1). The intrinsic GTPase activity of Ras is slow, hydrolyzing GTP to GDP at a rate of 10^{-2} per second [85]. This naturally slow GTPase activity allows for regulation by two classes of proteins: guanine nucleotide exchange factors (GEFs), which enhance the rate of exchange of GDP to GTP, thereby increasing the concentration of active Ras; and GTPase-activating proteins (GAPs), which enhance the rate of GTP hydrolysis by 1000 fold [85] and lead to Ras inactivation.

When bound to GTP, Ras binds to multiple Ras effector proteins (REPs). This interaction controls important pathways such as the Raf-MEK-ERK and the PI3K-PDK1-AKT pathways, which regulate cell survival, proliferation, angiogenesis and metabolism [86]. Therefore, mutations in Ras proteins are detrimental to normal cell regulation and have been shown to be oncogenic. Ras mutations are the most common activating mutations in cancer and are found in nearly 30% of all cancers [86].

There are more than 150 Ras-like proteins in a mammalian cell [87, 88], including the RHO, RAB, ARF families and the Gα subunit of heterotrimeric G proteins, which are famously associated with GPCRs. However, these Ras-like proteins have different signaling pathways and therefore will not be discussed further [89]. We will focus only on the Ras proteins themselves, of which there are 3: N-, H- and K-Ras. Moreover, KRas has two splice variants, KRas4A and KRas4B [90, 91]. The difference between Ras proteins mostly reside within the C-terminus of Ras, the so-called hypervariable region (HVR). This region determines the attachment of Ras to the membrane, through Ras protein-specific posttranslational lipid modifications like palmitoylation and farnesylation, and dictates Ras localization within the cell [92-94]. Of the Ras proteins, KRas is the most commonly mutated [86, 95]. Single point mutations in codons 12, 13, and 61 are commonly observed in Ras, and any mutation leading to an amino acid change (other than to proline) at any one of these codons is sufficient to cause Ras to become constitutively active [96, 97] [86]. These mutations either deter the formation of the water network needed for the nucleophilic attack that causes GTP hydrolysis or cause Ras to lose the ability to be regulated by GAP. Loss of GAP regulation occurs when mutations on codon 12 create a steric hindrance that inhibits GAP's ability to promote GTPase activity by supplying an arginine finger that would serve as a catalyst for the hydrolysis reaction of GTP into GDP. Oncogenic Ras mutations thus stabilize the Ras-GTP conformation, causing Ras to be constitutively active (i.e. always "on"). Constitutively active Ras no longer requires input from upstream extracellular receptors, such as epithelial growth factor receptor (EGFR) or other receptor tyrosine kinases, to recruit and activate REPs [86, 95, 98, 99]. This deregulates the signal transduction pathways downstream from Ras.

The recruitment of REPs only occurs in the active form of Ras, i.e. Ras-GTP [100-104]. The active and inactive conformations of Ras are distinguished by the flipping of threonine 35 [105], shown in yellow in Figure 1. This threonine interacts with solvent in the inactive conformation, Ras-GDP, but coordinates with a magnesium atom and the γ phosphate from GTP in the Ras-GTP conformation. By forming a hydrogen bond with the γ phosphate, Thr 35 drives the change in conformation of switch 1 (residues 32-40 in Figure 1, Figure 3), trapping switch 1 in position to interact with REPs. A similar phenomenon occurs with glycine 60, causing the change in conformation of switch 2 (residues 59-67). The protein-protein interactions (PPI) between Ras and REPs occur through the formation of an intermolecular continuous antiparallel beta-sheet between Ras and REPs (Figure 2 & Figure 3). The binding of REPs to Ras occurs through what has been characterized as the Ras binding domain (RBD). Figure 2 shows the protein-protein interaction (PPI) between Ras and REP as dots. The majority of the interaction is mediated solely through the intermolecular anti-parallel beta-sheet between switch 1 of Ras and REPs [106-108].

.

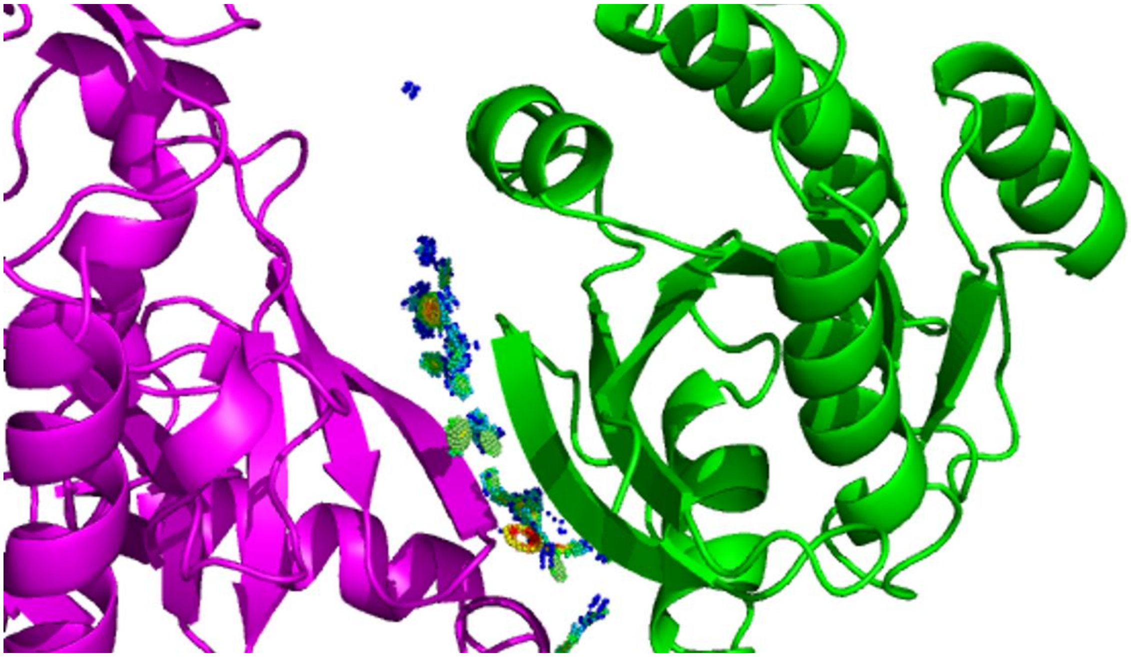


Figure 2: Interactions between Ras and REP are mediated nearly completely through the intermolecular beta sheet.Ras is shown as a cartoon in green, and PI3K is

shown in purple (PDB ID 1HE8). Atomic contacts between the two proteins are shown as dots.

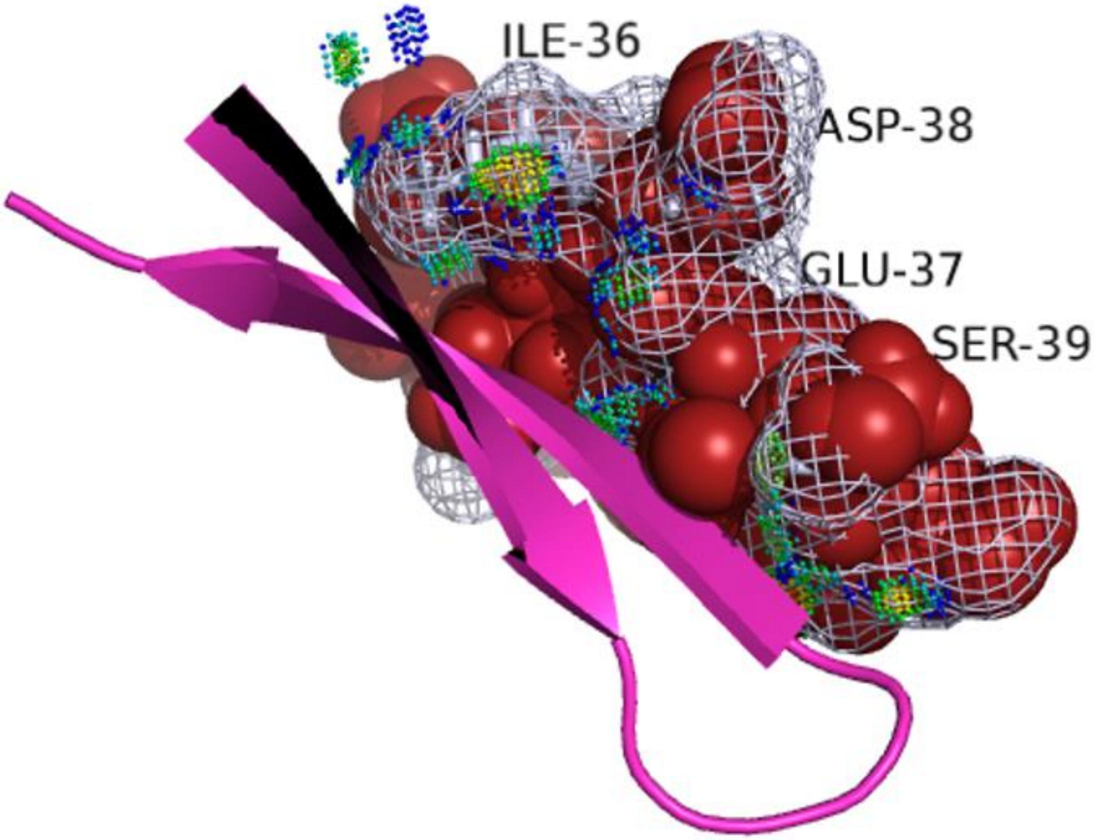


Figure 3: Aligned and overlaid structures of the switch 1 regions of Ras-GDP and Ras-GTP (PDB IDs 4LPK and 1HE8). Ras-GDP is shown as red spheres, while Ras-GTP is shown as gray mesh with sticks. Binding interactions shown as dots; the beta sheet of the REP PI3K (PDB ID 1HE8) is shown as cartoon in purple.

Figure 3 putatively demonstrates how switch 1 in its different conformations would interact with the intermolecular beta-sheet and the reasoning for the regulation of Ras activity through GDP/GTP binding. Isoleucine 36, glutamic acid 37, aspartic acid 38, and serine 39 only adopt optimized contacts with REPs when in the Ras-GTP (gray mesh) conformation. Without the interaction of Thr35 to stabilize the position of switch 1, the entropic cost of binding is too high, and the affinity for REPs decreases by over 10x [96, 109].

2.1.2 Ras Signaling and Cancer

Hanahan and Weinberg first defined the hallmarks of cancer as a set of characteristics that were at least in part common between the multiple maladies that fall under the umbrella term of cancer [110]. These hallmarks were later amended [111], leaving us with the following list of characteristics: Avoidance of programmed cell death, autonomy of growth signals (cancer cells either do not require or generate their own growth signals), insensitivity to growth inhibitor signals, unlimited replicative potential, continued angiogenesis, tissue invasion and metastasis, metabolic fitness, and genomic plasticity. I believe immune evasion and regulation will soon be added to the list [112].

The hypothesis that sequential mutations to a few genes, known as oncogenes, can lead to the development of phenotypes with these hallmarks and the eventual transformation of normal cells to cancerous cells is a widely accepted paradigm for cancer development today [113]. The success of targeted therapies further validates the oncogene hypothesis. Interestingly, the first gain-of-function oncogene mutations ever identified was G12V in Ras [114].

Mutations that cause constitutive activation of Ras like G12V can result in most if not all of the hallmarks of cancer. They inhibit programmed cell death through the upregulation of anti-apoptotic Bcl-xl proteins [115], and through hypoxia-driven inhibition of apoptosis [116]. For KRas4B this includes obtaining stemness (and through it immortalization) by suppressing non-canonical Wnt/Ca^{2+} signaling [117]. They cause autonomy of growth signals [118, 119] through the up-regulation of heparin-binding epidermal growth factor [120], the expression of TGF-alpha [121], expression of select integrins like β-3 integrin [122], expression of c-fos [123, 124], augmented activity of c-Jun [125] and activation of NK-κB [126]. They make cells insensitive to growth inhibitor signals by decreasing expression and protein stability of Smad3 and anti-proliferative signaling from TGF-beta [127-131], as well as inhibition of cyclin-dependent kinase inhibitors (CKIs) like p27 [130, 132]. They remove limits on replicative potential via overexpression of Cyclin D1 [133-135] and increased Cyclin D1 metabolic stability [136], while also in combination with Myc activating Cyclin E/Cdk2 and E2F[137]. They lead to continued angiogenesis through the overexpression of CXCL1, CXCL5, VEGF and COX 1-2 and regulation of TSP-1/-2 [138-143] and upregulation of IL8 [144-146]. They also recruit Th17 cells and $\gamma\delta TCR^{+}$ inflammatory cells that secrete IL17 cytokine, an angiogenesis-promoting cytokine [147-149]. Moreover, hypoxia activates even wild-type KRas, leading to further angiogenesis [116]. Tissue invasion and metastasis is achieved through the overexpression of urokinase plasminogen activator (uPA) and matrix metalloproteases-2 ,-9 (MMP-2/-3/-9) [139, 143], through KRas-induced expression of eukaryotic translation initiation factor 5A (eIF5A) leading to the activation of RhoA and Rho-associated Kinase (ROCK) [147, 150, 151], as well as through the KRas -riven recruitment of multiple cell types to the tumor that then drive cancer invasion and metastasis [147]. Chief among these are cancer-associated fibroblasts (CAF) [152, 153], which Ras can control through upregulation of the hedgehog pathway [154-156] or secretion of $\alpha6\beta4$ enriched endosomes [157, 158]. KRas mutations enhance metabolic fitness through alteration in glucose consumption and glutamine metabolism [159-161], and genomic plasticity by overcoming DNA damage-induced senescence [162] and contributing to mitotic-machinery damage [163]. Finally, they suppress immunity by overexpressing granulocyte macrophage colony-stimulating factor (GM-CSF) and thus recruiting myeloid-derived suppressor cells [164], by downregulating MHCI processing [165, 166], and by recruiting and inducing Tregs through overexpression and secretion of IL10 and TGFβ1 [167].

Continuous signaling through Ras continuously activates REPs and has been shown to be involved in the initiation and maintenance of many types of cancer [168]. Pancreatic ductal adenocarcinoma (PDAC), the most lethal of all solid tumors with no effective therapy, is initiated by KRas in up to 90% of cases and is dependent on KRas for its continued malignancy [98, 169]. Genetic inhibition of KRas in animal models of PDAC caused primary and metastatic lesions to disappear [169, 170]. This was shown to be due to the KRas-based regulation of glucose uptake and channeling of glucose intermediates into the non-oxidative pentose phosphate pathways (PPP) and the creation and maintenance of the stromal layer caused by KRas signaling. The stromal layer is very important to tumor survival because the layer of fibroblasts and intense cell matrix, mostly hyaluronan (HA), that surround the tumor protects it from immune-surveillance and chemotherapeutic agents [95, 169-173]. KRas mutations are also a pathway for cancer resistance to multiple anti-cancer therapies. For instance, cetuximab, a monoclonal antibody targeting the EGFR receptor is effective in treating some cancers. Patients who developed KRas mutations, however, showed no clinical benefit from this treatment [174]. This is believed to occur because KRas is downstream of EGFR in critical signaling pathways, and therefore KRas activation causes cells to become independent of EGFR signaling [175].

Yet, despite its medical importance, there is no KRas inhibitor in clinical use. This is an area of great unmet medical need, as cancer cells tend to be addicted to oncogenic signaling in general [176] and many are addicted to oncogenic KRas signaling in particular [177-182].

2.2 Targeting KRas

Targeting KRas directly is very difficult because its only pocket (in the canonical sense) is the GTP binding site [183, 184], and it binds GTP with picomolar affinity while there are hundreds of micromolars of GTP in the cytosol [185] [53, 186, 187]. This led drug hunters to attempt to target Ras through its processing and regulation.

Initially, the observation that KRas was farnesylated [188] and that farnesylation was required for KRas to become localized to membranes and maintain activity [189] started a race for potent farnesyltransferase inhibitors (FTIs). Unfortunately these failed to show clinical efficacy [190], because KRas can also be prenylated by geranylgeranyltransferase, and therefore still localize to the membrane [191]. Inhibition of geranylgeranyltransferase and farnesyltransferase has been attempted and showed promising results [192], the use of statins has also been suggested as a promising approach [193]. However, neither of these strategies have been tested in humans and specificity and toxicity questions remain. Interestingly, FTIs may work for HRas, and companies like Kura Oncology are pursuing this avenue [194]. The idea of targeting KRas membrane localization has persisted, and inhibitors of the KRas-PDEδ interaction have been developed [195]. PDEδ has been shown to be important to localize KRas to the plasma membrane and is therefore a new and interesting target [196]. Moreover, currently there is controversy about the ability and function of KRas to dimerize [197], and therapeutic inhibition of dimerization has been attempted [198].

Furthermore, new insights into KRas biochemistry are also being explored. The recent discovery of an allosteric site on KRas [199] [109] and of two distinct KRas-GTP switch 1 conformations, T1 and T2 [200, 201], of which only T2 has strong REP binding, has revitalized the idea of targeting KRas with small molecules. The most successful of these have been found using fragment-based lead discovery (FBLD) and have been capable of either stabilizing state T1 [202] or binding to KRas and inhibiting the binding of GEFs [186, 187, 203] [204]. Although exciting, the first class of inhibitors has not yet been tested in cells and did not show complete KRas inhibition at the maximum dose tested [202]. The latter class of GEF inhibitors required 72 hours for effective inhibition of oncogenic KRas due to mutant KRas' slow GTPase activity [203]. While these data represent great progress, significant and crippling gaps remain in our understanding of KRas allosteric regulation and of inhibition of GEF-mutant KRas interactions, preventing the development of small molecules that can inhibit KRas function at acceptable concentrations and and exposure times.

Recently an *in silico* approach was able to design a small molecule pan-Ras inhibitor targeting Ras itself. M.E Welsch and collaborators found a low-micromolar inhibitor of the Ras:effector interaction [205] using Schrodinger's GLIDE docking software [206], a ground breaking result and a major win for in silico methods. However, affinity remained far above biomedically relevant levels.

Overall, the strongest results to date in the KRas inhibitor space came from the use of tethering. Tethering starts with a small "probe" molecule attached to the target via a covalent bond, and then grows outward from that probe using combinatorial chemistry in the presence of highly reducing conditions in order to find highly favorable non-covalent interactions and to obtain specificity for the pocket. In the end, the result is a specific covalent inhibitor [183]. This technique was used to find a potent KRasG12C specific inhibitor that traps KRas in the inactive GDP-bound form [179, 182, 207]. Derivatives of this initial compound are being tested by AMGEN and MIRATI in clinical trials.

Despite this recent success, KRasG12C is a mutant that is uncommon outside of lung cancer, and an immense medical need remains for inhibitors of other KRas mutants. Therefore, in response to the shortcomings of small molecule inhibition, peptide and biologic inhibitor have been attempted. Inhibition of KRas PPI by overexpression of intrabodies or RBD-Raf in KRas-transformed cell lines has shown phenotypic reversion and cancer prevention in animal models injected with cancer cells also expressing the intrabodies [208, 209]. Cyclic peptides have been designed with sub-micromolar inhibition of KRas [210], and mRNA displays have been used to generate intrabodies with affinity equivalent to C-Raf [211]. Moreover, antibody mimetics have been developed that bind KRas with single-digit nanomolar affinity [212]. These so-called DARPins bind preferentially to Ras:GDP and inhibit nucleotide exchange, *de facto* functioning in a mechanism similar to KRasG12C inhibitors. Finally, yeast surface display [213] was used recently to generate mini-proteins with single-digit picomolar inhibition of KRas, an amazing result. Interestingly, 3 orders of magnitude came from combining two mini-proteins with different binding modes together, a lesson that may have wide-ranging applications.

Although yeast display technologies are becoming more powerful and clearly are showing promising results, computational protein design can serve as a lower-cost, highly accessible alternative. This is somewhat mitigated by the high computational cost of searching an exponentially large sequence space and of employing sufficiently sophisticated energy functions and protein flexibility and solvent models to obtain accurate results. These challenges have made computational protein design non-trivial. However, advances in computational protein and drug design (CPDD) have made this process feasible in some cases. Other groups have used computational design to target PPI in Ras [214-219]. Pincus and co-workers used molecular dynamics and electrostatically driven Monte Carlo (EDMC) methods to derive the average structure of both wild-type (WT) and oncogenic (ONC) Ras bound to its effector proteins. Peptides were synthesized with the sequence of KRas segments that differed between WT and ONC KRas and conjugated to a cell-penetrating peptide. These peptides were capable of phenotypic reversion of cancer cells but had no effect on untransformed pancreatic acinar cells. These peptides suggest the possibility of specifically targeting oncogenic KRas through differences in structure [215]. Zeng and coworkers developed peptides that target the Ras-Raf interaction by using multiple copy simultaneous search (MCSS). These peptides, although not potent, were capable of Ras-Ref inhibition, serving as an encouraging first step for peptide inhibitors of KRas [214]. More recently, alanine peptide screening was used to develop a constrained alpha helix mimicking a GEF protein's interaction with KRas [218, 219].

2.2.1 Novel CPDD techniques to develop biologic inhibitors of KRas.

We believe that improvements to CPDD can be made to develop biologics that can rival high-throughput biological methods like yeast display. The following section describes my work towards that goal. Due to the multidisciplinary nature of this project, all work in the following chapters were done in collaboration with Dr. Anna Lowegard, a computational biology and bioinformatics post-doc in the Donald lab. The following sub-chapters will describe the use of the state-of-the-art [80, 220, 221] protein design program OSPREY to develop competitive inhibitors of KRas effector binding. The OSPREY software package [220] is a free, open-source suite of protein design algorithms capable of guaranteeing that the results are optimal relative to the input model [60] [222]. OSPREY models continuous side-chain and backbone flexibility; can provably find the global minimum energy conformation (GMEC) of a protein system; and uses ensemble information to provide ε-approximations to the association constant under the residue-conformation-based model, termed K*. OSPREY is the ideal program to use, as all other algorithms and software packages fail to perform at least one of these functions, despite the fact that these functions are fundamental for optimal protein design [80, 220, 221].

2.2.2 Redesign of Raf-1 RBD Domain as a Superbinder of KRas.

The following sections are adopted from a manuscript being prepared in collaboration with Anna U. Lowegard, Jonathan D. Jou, Adegoke A. Ojewole, Graham T. Holt, and Bruce R. Donald. Moreover, acknowledgements must be made of the contributions by Rachel Kimbrough, and Terrence G. Oas. Both Anna and I are co-first authors of this manuscript. While Anna developed the novel algorithms and ran all of the computational designs. My contributions were towards the design of RBD variants, analysis of computational results, co-design of computational experiments and all experimental validation of the in silico *predictions. Please note that table 1, figure 4- 8 and the sections demarked by quotation marks are adopted verbatim from the manuscript in preparation below and were written by my co-first author Anna U. Lowegard, with comments and contributions from all the authors.*

A.U. Lowegard, M. S. Frenkel, J.D. Jou, A. A. Ojewole, G.T. Holt, B. R. Donald. Novel, Provable Algorithms for Efficient Ensemble-Based Computational Protein Design and Their Application to the Redesign of the c-Raf-RBD:KRas Protein-Protein Interface

2.2.2.1 Abstract

"The K^* algorithm provably approximates partition functions for a set of states (e.g., protein, ligand, and protein-ligand complex) to a user-specified accuracy ε. Often, reaching an ε-approximation for a particular set of partition functions takes a prohibitive amount of time and space. To alleviate some of this cost, we introduce two algorithms: FRIES, a Fast Removal of Inadequately Energied Sequences, and $EWAK^*$, an Energy

Window Approximation to K^*. In combination, these algorithms limit the input sequence space and the conformations included in each partition function calculation to provably only the most energetically favorable, leading to significant speed-ups com- pared to the previous state-of-the-art multi-sequence algorithm, BBK^*. As a proof of concept, we used these new algorithms to re-design the protein-protein interface (PPI) of the c-Raf-RBD:KRas complex. The Ras-binding domain of the protein kinase c-Raf (c-Raf-RBD) is the tightest known binder of KRas, a historically “undruggable” protein implicated in difficult-to-treat cancers including pancreatic ductal adenocarcinoma (PDAC). FRIES/$EWAK^*$ accurately retrospectively predicted the effect of 41 different sets of mutations in the PPI of the c-Raf-RBD:KRas complex. These mutations include mutations whose affect had previously been incorrectly predicted using other computational methods. Next, we used FRIES/$EWAK^*$ for prospective design and discovered a novel point mutation that improves binding of c-Raf-RBD to KRas in its active, GTP-bound state ($KRas^{GTP}$). We also combined this new mutation with two previously reported mutations to create a new variant of c-Raf-RBD, c-Raf-RBD(RKY). This variant exhibits single-digit nanomolar affinity for $KRas^{GTP}$which we validated using a bio-layer interferometry (BLI) assay. This study steps through the advancement and development of computational protein design by presenting theory, new algorithms, accurate retrospective designs, new prospective designs, and biochemical validation.”

2.2.2.2 Introduction

"Computational protein and drug design (CPDD) is an innovative tool that enables the prediction of protein sequences with desired biochemical properties (such as improved binding affinity). OSPREY (Open Source Protein Redesign for You) [68] is an open-source, state-of-the-art software package used for CPDD and is available at http://www.cs.duke.edu/donaldlab/osprey.php for free. OSPREY's algorithms focus on provably returning the optimal sequences and conformations for a given input model. In contrast, as argued in [18, 70, 223-226], stochastic, non-deterministic approaches [227-229] provide no guarantees on the quality of low-energy conformations and make determining sources of error in design results very difficult."

"When using OSPREY, the input model generally consists of a protein structure, a flexibility model (e.g., choice of sidechain flexibility, allowed mutable residues, etc.), and an all-atom pairwise-decomposable energy function that is used to evaluate conformations. OSPREY models amino acid sidechains using frequently observed rotational isomers or "rotamers" [230]. Additionally, OSPREY can also model continuous sidechain flexibility [73, 77, 78, 80] along with discrete and continuous backbone flexibility [69, 72, 74, 75], which allow for a more accurate approximation of protein behavior [74, 80, 231-234]. The output produced by CPDD generally consists of a set of candidate sequences and conformations. Many protein design methods have focused on determining a global minimum energy conformation (GMEC) [72, 73, 76, 235-238]. However, a protein in solution exists not as a single, low-energy structure, but as a thermodynamic ensemble of conformations. Models that only consider the GMEC may incorrectly predict biophysical properties such as binding [78, 231-234, 239] because GMEC-based algorithms underestimate potentially significant entropic contributions. In contrast to GMEC-based approaches, the K^* algorithm [60, 79, 233] in OSPREY models thermodynamic ensembles to provably and efficiently approximate the K^* score. The K^* score is a ratio of the Boltzmann-weighted partition functions for a protein-ligand complex that estimates the association constant, K_a (further detailed in Section 2). BBK^* [71] is the most recent improvement on the traditional K^* algorithm that allows for multi-sequence design. Previous algorithms [60, 78, 79, 237, 240-242] that design for binding affinity using ensembles are linear in the number of sequences N , where N is exponential in the number of simultaneously mutable residue positions. BBK^* is the first provable algorithm to run in time sublinear in N, making it possible to perform K^* designs over large sequence spaces and to enumerate a gap-free list of sequences in order of decreasing K^* score."

"OSPREY has been used successfully on several empirical, prospective designs including designing enzymes [74, 78, 79, 233, 243], resistance mutations [18, 19, 67], protein-protein interaction inhibitors [59, 60], epitope-specific antibody probes [244], and broadly-neutralizing antibodies [222, 245]. These successes have been validated experimentally in vitro and in vivo and are featured in clinical trials. However, while OSPREY has been successful in the past, as the size of protein design problems grows (e.g., when considering a large protein-protein interface), enumerating and minimizing the necessary number of conformations and sequences to satisfy the provable halting criteria in previous K*-based algorithms becomes prohibitive (despite recent algorithmic improvements). The entire conformation space can be monumental in size and heavily populated with energetically unfavorable sequences and conformations. EWAK*, an Energy Window Approximation to K*, seeks to alleviate some of this difficulty by restricting the conformations included in each sequence's thermodynamic ensemble. EWAK* guarantees that each conformational ensemble contains all of the lowest energy conformations within an energy window of the GMEC for each design sequence. FRIES, a Fast Removal of Inadequately Energied Sequences, also mitigates this complexity problem by limiting the input sequence space size to only the most favorable, low energy sequences. Previous algorithms have focused on optimizing for sequences similar in energy to the GMEC. In contrast, FRIES focuses on optimizing for sequences with energies better-than or comparable-to the wild-type sequence. FRIES guarantees that the input sequence space includes all of the sequences within an energy window of the wild-type sequence, but does not include any potentially unstable sequences with significantly worse partition function values. Wild-type sequences are generally expected to be near-optimal for their corresponding structures [227]. Therefore, limiting the sequence space to sequences energetically similar to or better than the wild-type sequence is reasonable. FRIES and EWAK* improve runtimes by up to 2 orders of magnitude compared to BBK*. Additionally, FRIES provably keeps only those sequences similar to or energetically more favorable than the wild-type sequences, decreasing the size of the input sequence space size by more than 2 orders of magnitude.

Furthermore, as a proof of concept to test these algorithms and our design approach, we used FRIES and EWAK* to retrospectively predict the effect on binding of mutations in the protein-protein interface (PPI) of c-Raf-RBD in complex with KRas (c-Raf-RBD:KRas). Our new algorithms succeeded in accurately predicting these effects where other computational methods previously failed [246]. Next, we used FRIES/EWAK* prospectively to predict mutations in the PPI of the c-Raf-RBD:KRas complex. We then measured the binding of top predicted variants to KRas using a bio-layer interferometry (BLI) assay single-concentration screen. This screen suggested that one of our new computationally-predicted c-Raf-RBD variants – c-Raf-RBD(Y), a c-Raf-RBD that includes the mutation V88Y – exhibits improved binding to KRas^{GTP}. Additionally, we created a c-Raf-RBD variant, c-Raf-RBD(RKY), that included this new mutation, V88Y, alongside two previously reported mutations [246], N71R and A85K. The single-concentration screen using BLI also suggested that our new c-Raf-RBD variant, c-Raf-RBD(RKY), binds more tightly to KRas^{GTP}than the previously reported best variant [246]. The most promising variants were further tested using a BLI assay with titration which confirmed that, to the best of our knowledge, c-Raf-RBD(RKY) is the highest affinity variant ever designed with single-digit nanomolar affinity for KRas^{GTP}."

"A further description of the FRIES/EWAK* algorithms can be found in the final manuscript or in the thesis work of Anna U. Lowegard. It will not be covered in this section. "

"The c-Raf Ras-binding domain (c-Raf-RBD) is a small self-folding domain that does not include the kinase signaling domains normally present in c-Raf. The c-Raf-RBD normally binds to KRas when KRas is GTP-bound (KRas^{GTP}). A c-Raf-RBD variant that has high affinity for KRas^{GTP}could be an important first step towards discovering a tool that disrupts the KRas:effector interaction. Despite the recent successes with inhibitors targeting KRas(G12C) by trapping it in the inactive GDP-bound state [179, 247-251] and their recent move to clinical trials [252], these inhibitors are susceptible to resistance in the form of up-regulation of guanine nucleotide exchange factors (GEFs) and nucleotide exchange [251] which both push KRas to remain in its GTP-bound state. An inhibitor of the interaction between KRas^{GTP}and its effectors is hypothesized to have the advantage of not being susceptible to these mechanisms of resistance because it would directly interrupt KRas signalling. So, to further verify the accuracy and utility of FRIES/EWAK*, we focused on this important PPI between KRas^{GTP}and one of its many effectors, c-Raf. First, we retrospectively predicted the effect of previously reported mutations in the c-Raf-RBD [246, 253, 254] and how they affect the binding of c-Raf-RBD to KRas. Then, we redesigned the PPI in search of new c-Raf-RBD variants with increased affinity for KRas^{GTP}. In order to perform these computational designs, we made a homology model of c-Raf-RBD bound to KRas^{GTP}. In short PDB ID 4DSN [69] is an X-ray crystal structure of KRas isoform 2B which contains G12D, a mutation that locks KRas into its active form. This structure of KRas bound to a GTP analog was used to model KRas^{GTP}. PDB ID 1GUA [255] is an X-ray crystal structure of c-Raf in complex with Rap, a Ras homolog containing, two mutations were made to RAP1 (E30D, K31E) in order to make its core effector region domain identical to KRas [255] . These two structures (4DSN and 1GUA) were aligned using PyMol [256] , RMSD of 0.729Å. Rap was then removed, leaving c-Raf poised in complex with KRas from PDB ID 4DSN. This complex was then minimized using Sander from AmberTools for 200 steps [257] to relax any steric clashes."

2.2.2.3 FRIES/EWAK retrospectively predicted the affect of mutations in c- Raf-RBD on binding to KRas*

"Each previously reported c-Raf-RBD variant [246,253,254] was tested computationally using FRIES/EWAK* by calculating a K* score, a computational approximation of K_a, for each variant along with its corresponding wild-type sequence. A percent change in binding was then calculated by comparing the variant's K* score to the corresponding wild-type sequence's K* score. The $\log_{10}$ of this value was then calculated and is normalized to the wild-type by subtracting 2. A similar procedure was completed using the reported experimental data in order to easily compare the computationally predicted effect with the experimentally measured effect (Table 1). The resulting value Δb represents the change in binding. If a variant has a Δb less than 0, it is predicted to decrease binding. If a variant has a Δb greater than 0, it is predicted to increase binding. Δb values that are roughly equivalent to 0 indicate variants that have little to no effect on binding since the wild-type sequence was normalized to 0. Δb values for the 41 computationally tested variants are plotted and compared to experimental values in Figure 4."

“Out of the 41 variants tested, EWAK* predicted the experimentally reported effect (increased vs. decreased binding) correctly in 38 cases. Those where the effect was predicted incorrectly are marked in Figure 4. These designs range in size from single point mutations to 6 simultaneous mutations. Results are outlined in Figure 4. Furthermore, the Spearman’s ρ value – a measure of the correlation between two sets of rankings – when comparing the experimental results to the computational results was 0.81. This ρ value indicates that not only can we correctly predict the effect of a particular set of mutations, but that we do a good job ranking the variants in order (see Figure 5). This value is very similar to Spearman’s ρ values for other PPI systems when using OSPREY, as evidenced in [68].”

“BBK* produced similarly accurate results, but took up to 10 times longer and failed to produce results in 4 cases. In particular, in 2 cases (marked in green in Figure 5), BBK* runs out of memory. These cases serve as examples of large designs where EWAK* outperforms BBK*. In the 2 other cases (marked in orange Figure 4), BBK* failed to return a result for the requested sequence in the top 5 reported sequences. EWAK* and FRIES are particularly helpful when considering these types of bigger designs that contain more simultaneous mutations and more flexible residues.”

“Finally, we compared our predictions to the interesting biological predictions in [246]. It is unclear how many mutations were computationally evaluated, but the authors report computational predictions for 6 point mutations. Of those mutants, point mutants R67L, N71R, and V88I were predicted to improve the intermolecular interactions between c-Raf-RBD and KRas^{GTP}. However, experiments found that R67L and V88I actually reduced the binding of c-Raf-RBD to KRas^{GTP} [246, 253] In contrast to [246], EWAK* accurately predicted that these mutations decrease binding of c-Raf-RBD to KRas^{GTP}. For a more detailed view of one of these designs, V88I, see Figure 6. Additionally, a number of mutations were combined and experimentally tested in [246]. Unfortunately, none of these variants improved binding to either KRas^{GTP}or KRas^{GDP}, which FRIES/EWAK* correctly predicted computationally (see Figure 4). In [246], the authors do not present any computational predictions for these combined variants, but our results show thata computational prediction using OSPREY's EWAK* may have saved the time and resources taken to experimentally test these variants. "

Table 1: Experimental and computational percent change in binding and rankings as computed using EWAK*. For each listed variant, we give the experimental percent change in binding relative to wild-type as reported in [253] and as calculated from reported binding values in [254] and [246], The rankings have a Pearson correlation of 0.81.

Mutation(s)	Exp. (%)	Comp. (%)	Exp. Ranking	Comp. Ranking
Wild-Type	100.00	100.00	N/A	N/A
R89L	1.3×10^{-7}	1.64×10^{-10}	1	3
F61W/R67L/V69E/N71R/V88I/A85K	0.23	2.20×10^{-11}	2	2
K84A	0.93	3.03×10^{-5}	3	4
Q66A	1.76	0.99	4	16
A85D	3.00	0.01	5	10
R59A	3.42	8.09×10^{-4}	6	7
F61W/V69E/N71R/V88I	4.64	0.03	7	12
R67A	6.19	1.78×10^{-4}	8	6
K84L	8.60	1.56×10^{-4}	9	5
Q66K	9.00	1.78	10	18
T68A	10.00	4.03	11	20
V88D	10.00	0.05	12	13
T68K	11.00	5.63×10^{-17}	13	1
V69A	13.68	5.08	14	22
A85I	18.00	8.64	15	24
K65A	18.57	1.04	16	17
K65E	19.40	0.71	17	15
N64A	21.31	5.77	18	23

V69R	29.00	1.40×10^{4}	19	34
K87Q	30.00	12.43	20	26
K65M	31.71	2.53	21	19
N71E	34.00	0.15	22	14
F61W	36.11	116.20	23	29
F61W/R67L/N71R/V88I	36.11	0.01	24	11
V88I	39.39	16.15	25	28
R67L	42.00	3.06×10^{-3}	26	9
R59L	43.00	2.51×10^{-3}	27	8
K84R	49.00	10.01	28	25
N64D	50.00	15.60	29	27
F61W/N71R/V88I	54.17	1.96×10^{4}	30	35
K87R	100.00	120.04	31	30
F61W/N71R	162.50	1.10×10^{5}	32	37
V88K	171.00	127.37	33	31
V88H	266.00	227.92	34	32
A85R	290.00	1.50×10^{7}	35	39
N71R	325.00	9.97×10^{4}	36	36
N64K	380.00	4.47	37	21
V88R	400.00	2.44×10^{3}	38	33
A85K/V88R	550.00	1.33×10^{7}	39	38
A85K	700.00	2.13×10^{7}	40	40
N71R/A85K	866.67	3.63×10^{9}	41	41

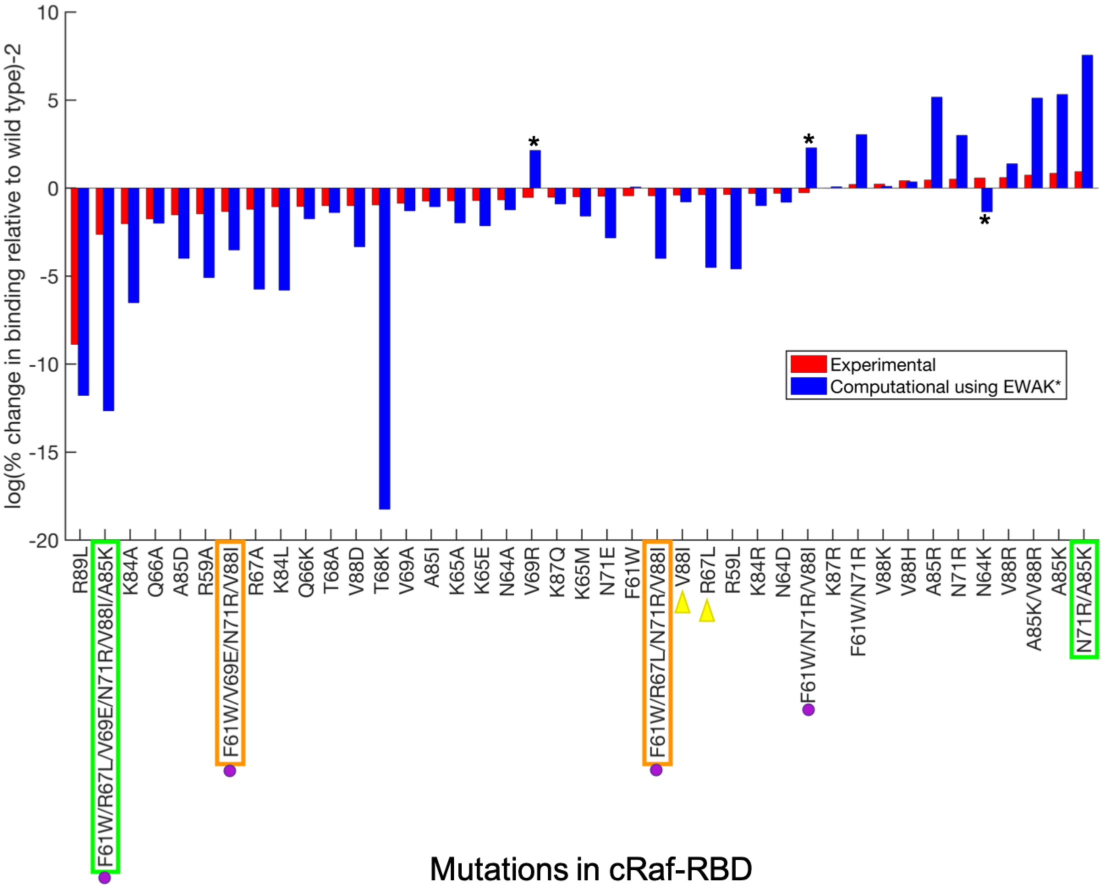


Figure 4: Predicting the effect of mutations in c-Raf-RBD on binding with KRas. Each bar represents either the experimental (red) or computationally predicted (blue) affect each variant has on binding. The bars are sorted in increasing order of Δb value of the experimental (red) bars. If the Δb value is less than 0, binding decreases. If the Δb value is greater than 0, binding increases. If the Δb value is close to 0, the effect is neutral. Quantitative values of K* tend to overestimate the biological effects of mutations (leading to the much larger blue bars) due to the limited nature of the input model compared to a biologically accurate representation. However, K* in general does a good job ranking variants, as can be seen here in Figure 5, in [68], and in [67]. Out of the 41 variants listed on the x-axis, only 3 were predicted incorrectly (marked with black asterisks) by EWAK*. BBK* performed very similarly to EWAK* (data not shown), however, in 2 cases (marked with green boxes), BBK* ran out of memory and was unable to calculate a score. BBK* also did not return values for the 2 variants marked with orange boxes. The variants marked with purple dots were tested experimentally – not computationally – in [246] and decreased binding of c-Raf-RBD to KRas^{GTP}, which EWAK* was able to predict correctly. The two variants marked with yellow triangles were computationally predicted in [246] to improve binding of c-Raf-RBD to KRas^{GTP}. However, the experimental validation in [246] showed that these variants exhibit decreased binding, which EWAK* accurately predicted.

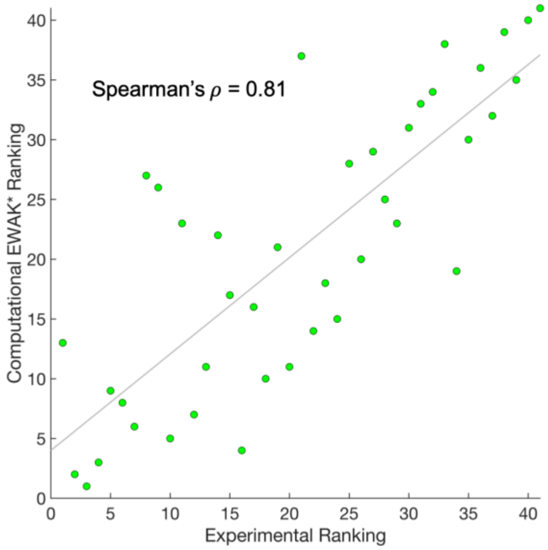


Figure 5: Comparing the computational *EWAK ranking with the experimental ranking for 41 c-Raf-RBD variants. Each green dot represents a variant of c-Raf-RBD and is plotted ac- cording to the experimental ranking along with the corresponding computational ranking. A least squares fit line is shown in gray. Calculating the Pearson correlation coefficient between the two sets of rankings yields a Spearman's ρ of 0.81.**

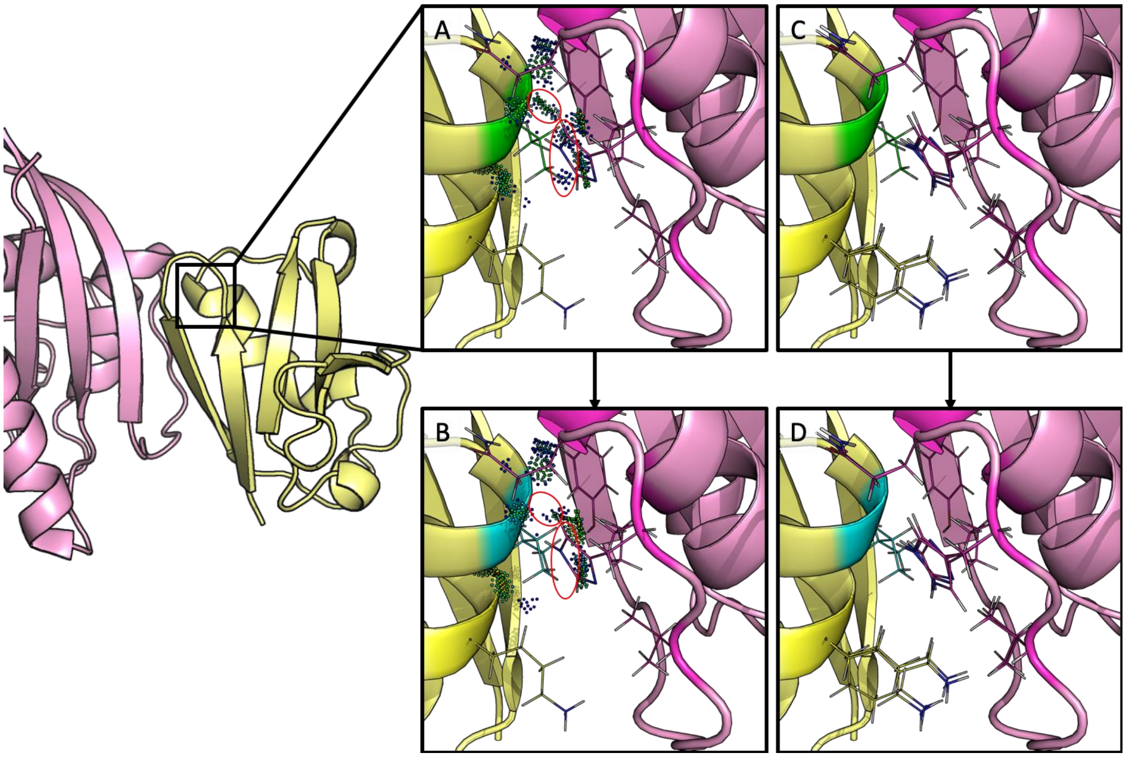
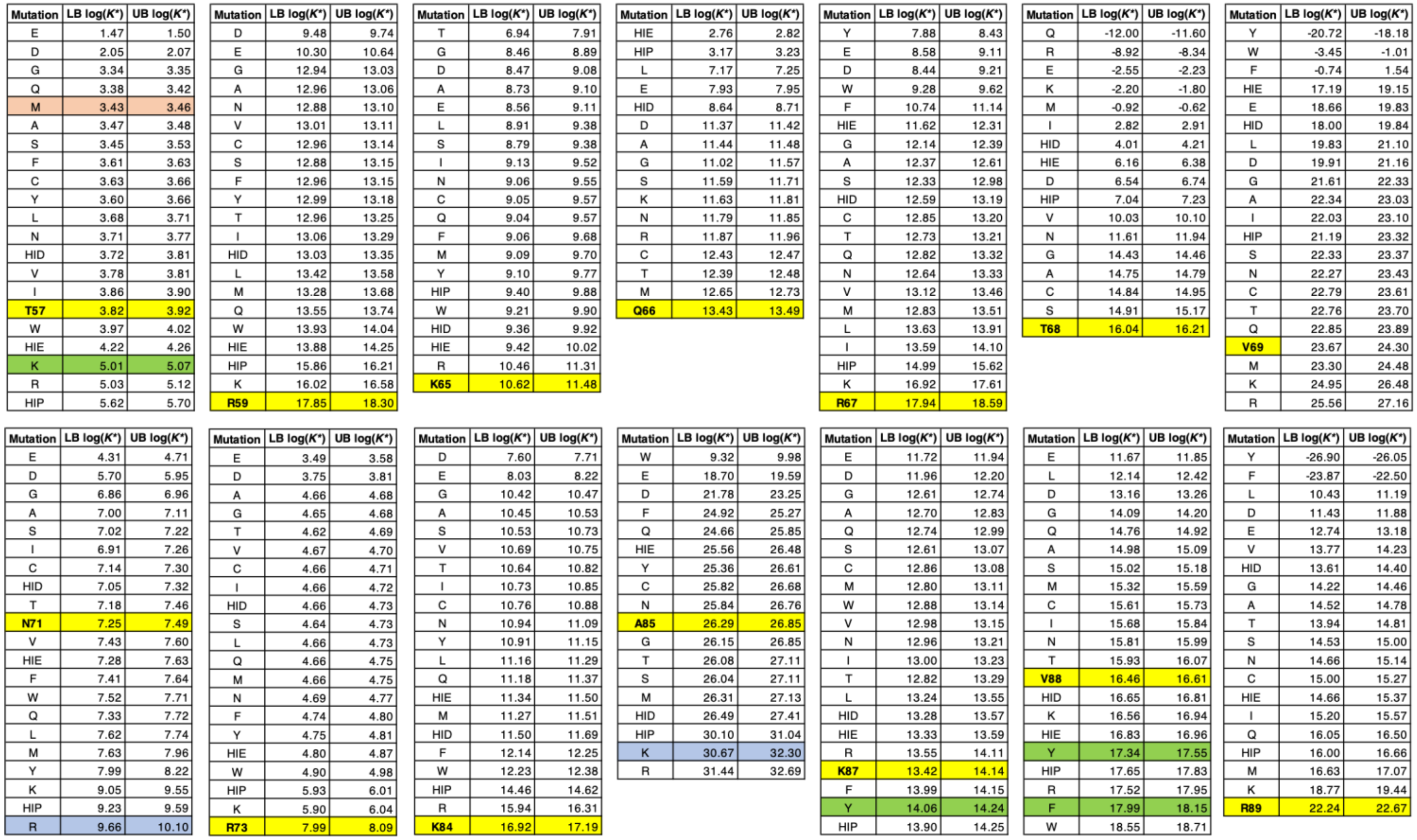


Figure 6: Re-design of c-Raf-RBD residue position 88 from valine to isoleucine. The left-hand side shows c-Raf-RBD (yellow) in complex with KRas (pink). Panels (A-D) zoom in on one particular design at residue position 88 and are rotated 180˚. Residue position 88 has a valine in the native, wild-type sequence (panels A & C) which was re-designed to an isoleucine (panels B & D). A mutation to isoleucine at this position was computationally predicted by EWAK* to decrease the binding of c-Raf-RBD to KRas^{GTP}. This was experimentally validated in [246], where the authors incorrectly computationally predicted the affect of this particular mutation on the binding of c-Raf-RBD to KRas^{GTP}. (A) The wild-type residue (valine) is shown in green with dots that indicate molecular interactions [258, 259] with the surrounding residues (residues allowed to be flexible in the design are shown as lines). (B) The mutant residue (isoleucine) is shown in blue with dots that indicate molecular interactions [66] with the surrounding residues (residues allowed to be flexible in the design are shown as lines). Contacts made by the wild-type valine residue (circled dots in (A)) were lost upon mutation to isoleucine (circled space in (B)). (C & D) The 10 lowest energy conformations that were included in the complex partition function calculation are shown for the wild-type (green) and the variant (blue).

2.2.2.4 Prospective Design of Raf Superbinders.

"We performed 14 different designs where each design included 1 mutable position that was allowed to mutate to all amino acid types except for proline. Each design also included a set of surrounding flexible residues within roughly 4 Å of the mutable residue. These designs used continuous flexibility [73, 77, 78, 80] and then FRIES/EWAK* was used. FRIES was first used to limit each design to only the most favorable sequences and then EWAK* was used to estimate the K* scores (algorithms described in other work). We report the upper and lower bounds on the EWAK* score for each design in the tables in Figure 7 where the listed sequences are those that were not pruned during the FRIES step. From these results, the predicted binding effect (increased vs. decreased) was determined based on comparing each variant's K* score to its corresponding wild-type K* score. We then selected 5 novel point mutations – that to our knowledge are not reported in any existing literature – for experimental validation. It is worth noting that these 5 point mutations are selected out of an initial 294 possible mutations. We limited our experimental validation to only these 5 new mutations and 2 previously reported mutations. This greatly limits the amount of resources necessary for experimental validation compared to testing all 294 possibilities. These mutations were selected based on having a promising K* score and through examining structures calculated by EWAK* (an example is shown *in* Figure 8). Of the mutations selected, T57M was selected to act as a variant that we computationally predicted to be comparable to wild-type. This variant was included to further verify the accuracy of

OSPREY's predictions. On the other hand, some of OSPREY's top predictions were excluded: for instance, T57R was not selected for experimental testing because it has an unsatisfied hydrogen bond as evidenced in the structures calculated by OSPREY. Therefore, we do not believe that the score accurately represents the effect the mutation will have. Other excluded top predictions display similar characteristics or have been reported and tested previously [246,253,254]."



Mutation	LB log(K*)	UB log(K*)
E	1.47	1.50
D	2.05	2.07
G	3.34	3.35
Q	3.38	3.42
M	3.43	3.46
A	3.47	3.48
S	3.45	3.53
F	3.61	3.63
C	3.63	3.66
Y	3.60	3.66
L	3.68	3.71
N	3.71	3.77
HID	3.72	3.81
V	3.78	3.81
I	3.86	3.90
T57	3.82	3.92
W	3.97	4.02
HIE	4.22	4.26
K	5.01	5.07
R	5.03	5.12
HIP	5.62	5.70

Mutation	LB log(K*)	UB log(K*)
D	9.48	9.74
E	10.30	10.64
G	12.94	13.03
A	12.96	13.06
N	12.88	13.10
V	13.01	13.11
C	12.96	13.14
S	12.88	13.15
F	12.96	13.15
Y	12.99	13.18
T	12.96	13.25
I	13.06	13.29
HID	13.03	13.35
L	13.42	13.58
M	13.28	13.68
Q	13.55	13.74
W	13.93	14.04
HIE	13.88	14.25
HIP	15.86	16.21
K	16.02	16.58
R59	17.85	18.30

Mutation	LB log(K*)	UB log(K*)
T	6.94	7.91
G	8.46	8.89
D	8.47	9.08
A	8.73	9.10
E	8.56	9.11
L	8.91	9.38
S	8.79	9.38
I	9.13	9.52
N	9.06	9.55
C	9.05	9.57
Q	9.04	9.57
F	9.06	9.68
M	9.09	9.70
Y	9.10	9.77
HIP	9.40	9.88
W	9.21	9.90
HID	9.36	9.92
HIE	9.42	10.02
R	10.46	11.31
K65	10.62	11.48

Mutation	LB log(K*)	UB log(K*)
HIE	2.76	2.82
HIP	3.17	3.23
L	7.17	7.25
E	7.93	7.95
HID	8.64	8.71
D	11.37	11.42
A	11.44	11.48
G	11.02	11.57
S	11.59	11.71
K	11.63	11.81
N	11.79	11.85
R	11.87	11.96
C	12.43	12.47
T	12.39	12.48
M	12.65	12.73
Q66	13.43	13.49

Mutation	LB log(K*)	UB log(K*)
Y	7.88	8.43
E	8.58	9.11
D	8.44	9.21
W	9.28	9.62
F	10.74	11.14
HIE	11.62	12.31
G	12.14	12.39
A	12.37	12.61
S	12.33	12.98
HID	12.59	13.19
C	12.85	13.20
T	12.73	13.21
Q	12.82	13.32
N	12.64	13.33
V	13.12	13.46
M	12.83	13.51
L	13.63	13.91
I	13.59	14.10
HIP	14.99	15.62
K	16.92	17.61
R67	17.94	18.59

Mutation	LB log(K*)	UB log(K*)
Q	-12.00	-11.60
R	-8.92	-8.34
E	-2.55	-2.23
K	-2.20	-1.80
M	-0.92	-0.62
I	2.82	2.91
HID	4.01	4.21
HIE	6.16	6.38
D	6.54	6.74
HIP	7.04	7.23
V	10.03	10.10
N	11.61	11.94
G	14.43	14.46
A	14.75	14.79
C	14.84	14.95
S	14.91	15.17
T68	16.04	16.21

Mutation	LB log(K*)	UB log(K*)
Y	-20.72	-18.18
W	-3.45	-1.01
F	-0.74	1.54
HIE	17.19	19.15
E	18.66	19.83
HID	18.00	19.84
L	19.83	21.10
D	19.91	21.16
G	21.61	22.33
A	22.34	23.03
I	22.03	23.10
HIP	21.19	23.32
S	22.33	23.37
N	22.27	23.43
C	22.79	23.61
T	22.76	23.70
Q	22.85	23.89
V69	23.67	24.30
M	23.30	24.48
K	24.95	26.48
R	25.56	27.16

Mutation	LB log(K*)	UB log(K*)
E	4.31	4.71
D	5.70	5.95
G	6.86	6.96
A	7.00	7.11
S	7.02	7.22
I	6.91	7.26
C	7.14	7.30
HID	7.05	7.32
T	7.18	7.46
N71	7.25	7.49
V	7.43	7.60
HIE	7.28	7.63
F	7.41	7.64
W	7.52	7.71
Q	7.33	7.72
L	7.62	7.74
M	7.63	7.96
Y	7.99	8.22
K	9.05	9.55
HIP	9.23	9.59
R	9.66	10.10

Mutation	LB log(K*)	UB log(K*)
E	3.49	3.58
D	3.75	3.81
A	4.66	4.68
G	4.65	4.68
T	4.62	4.69
V	4.67	4.70
C	4.66	4.71
I	4.66	4.72
HID	4.66	4.73
S	4.64	4.73
L	4.66	4.73
Q	4.66	4.75
M	4.66	4.75
N	4.69	4.77
F	4.74	4.80
Y	4.75	4.81
HIE	4.80	4.87
W	4.90	4.98
HIP	5.93	6.01
K	5.90	6.04
R73	7.99	8.09

Mutation	LB log(K*)	UB log(K*)
D	7.60	7.71
E	8.03	8.22
G	10.42	10.47
A	10.45	10.53
S	10.53	10.73
V	10.69	10.75
T	10.64	10.82
I	10.73	10.85
C	10.76	10.88
N	10.94	11.09
Y	10.91	11.15
L	11.16	11.29
Q	11.18	11.37
HIE	11.34	11.50
M	11.27	11.51
HID	11.50	11.69
F	12.14	12.25
W	12.23	12.38
HIP	14.46	14.62
R	15.94	16.31
K84	16.92	17.19

Mutation	LB log(K*)	UB log(K*)
W	9.32	9.98
E	18.70	19.59
D	21.78	23.25
F	24.92	25.27
Q	24.66	25.85
HIE	25.56	26.48
Y	25.36	26.61
C	25.82	26.68
N	25.84	26.76
A85	26.29	26.85
G	26.15	26.85
T	26.08	27.11
S	26.04	27.11
M	26.31	27.13
HID	26.49	27.41
HIP	30.10	31.04
K	30.67	32.30
R	31.44	32.69

Mutation	LB log(K*)	UB log(K*)
E	11.72	11.94
D	11.96	12.20
G	12.61	12.74
A	12.70	12.83
Q	12.74	12.99
S	12.61	13.07
C	12.86	13.08
M	12.80	13.11
W	12.88	13.14
V	12.98	13.15
N	12.96	13.21
I	13.00	13.23
T	12.82	13.29
L	13.24	13.55
HID	13.28	13.57
HIE	13.33	13.59
R	13.55	14.11
K87	13.42	14.14
F	13.99	14.15
Y	14.06	14.24
HIP	13.90	14.25

Mutation	LB log(K*)	UB log(K*)
E	11.67	11.85
L	12.14	12.42
D	13.16	13.26
G	14.09	14.20
Q	14.76	14.92
A	14.98	15.09
S	15.02	15.18
M	15.32	15.59
C	15.61	15.73
I	15.68	15.84
N	15.81	15.99
T	15.93	16.07
V88	16.46	16.61
HID	16.65	16.81
K	16.56	16.94
HIE	16.83	16.96
Y	17.34	17.55
HIP	17.65	17.83
R	17.52	17.95
F	17.99	18.15
W	18.55	18.71

Mutation	LB log(K*)	UB log(K*)
Y	-26.90	-26.05
F	-23.87	-22.50
L	10.43	11.19
D	11.43	11.88
E	12.74	13.18
V	13.77	14.23
HID	13.61	14.40
G	14.22	14.46
A	14.52	14.78
T	13.94	14.81
S	14.53	15.00
N	14.66	15.14
C	15.00	15.27
HIE	14.66	15.37
I	15.20	15.57
Q	16.05	16.50
HIP	16.00	16.66
M	16.63	17.07
K	18.77	19.44
R89	22.24	22.67

Figure 7: Computational predictions using OSPREY/EWAK∗ for point mutants in c-Raf-RBD. Each table shows the results of the re-design of a residue position (highlighted in yellow) in c- Raf-RBD in the c-Raf-RBD:KRas interface. The two residues highlighted in blue are the best previously discovered [246] mutations that improve binding (independently and additively) and are included in our tightest binding variant, c-Raf-RBD(RKY). Mutations highlighted in green were selected for experimental testing and validation. The one variant highlighted in orange was also selected for testing to act as a mutation predicted to be comparable to wild-type to see how accurately OSPREY predicted the effects of these mutations.

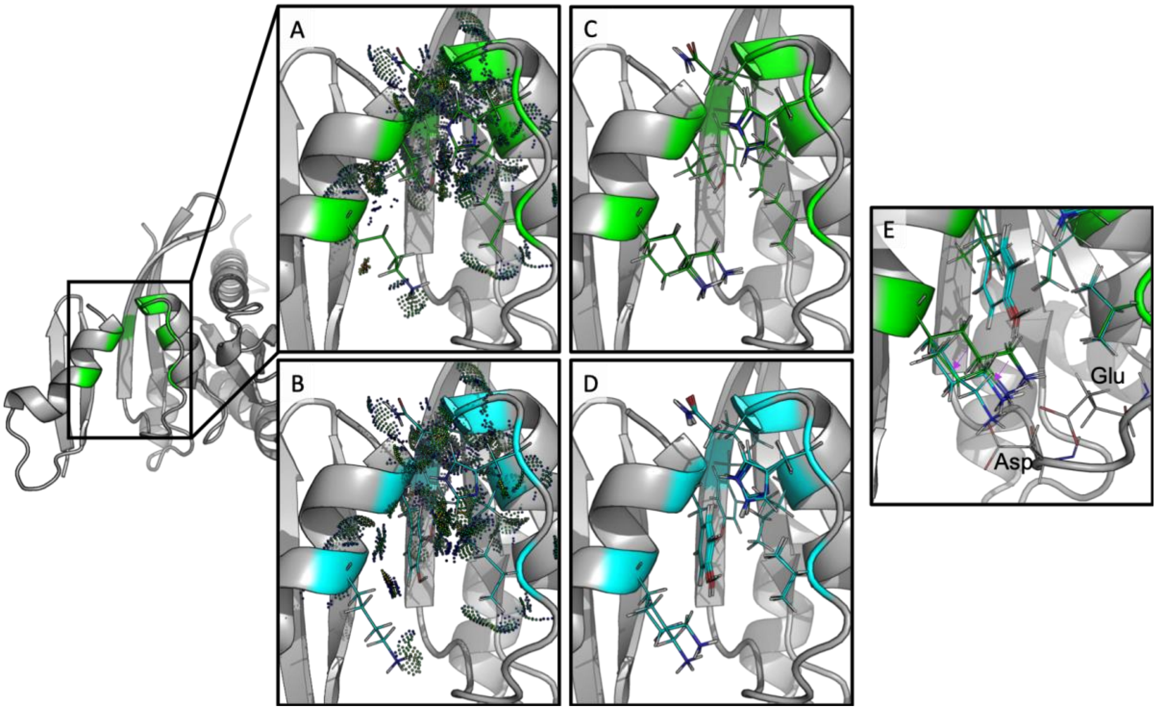


Figure 8: Re-design of position 88 from valine to tyrosine in the c-Raf-RBD:KRas protein- protein interface. Panels (A) and (C) show the wild-type sequence, valine, and panels (B) and (D) show the variant, tyrosine. Panels (A) and (B) show the wild-type and the variant, respectively, along with probe dots [258, 259] that represent the molecular interactions within each structure calculated by OSPREY. These probe dots were selected to only show interactions between the residues included in the computational designs (shown as lines). Panels (C) and (D) show 10 low-energy structures from each conformational ensemble calculated by OSPREY/EWAK*. Panel (E) shows a zoomed-in overlay of panels (C) and (D) with purple arrows indicating the change in positioning of the lysine at residue position 84 upon mutation of residue position 88 from valine to tyrosine. When valine is present at position 88, the lysine residue (shown in green) primarily hydrogen bonds with an aspartate (labeled) in KRas. When valine is mutated to tyrosine (shown in cyan), the lysine at position 84 moves to make room for the tyrosine and positions itself to hydrogen bond with both the aspartate and the glutamate (labeled) in KRas.

2.2.2.5 Experimental Validation of Mutations in the c-Raf-RBD:KRas Protein- Protein Interface

RBD and KRas were purified and prepared as described in section 2.2.2.5.1 Purification. The mutations selected (highlighted in Figure 8) from computational design were experimentally validated using a bio-layer interferometry (BLI) assay. We first developed a semi-quantitative high-throughput screening assay to quickly screen for variants with enhanced binding for KRas (Figure 22). Although considerable work has been done in KRas (see section 2.1 and 2.2), HTS screens for binding to KRas have either required conjugation of KRas to a fluorescent nucleotide analogue for fluorescence polarization (FP) [108, 260-264] and guanine nucleotide dissociation inhibition (GDI) [265-270], or involved the biotinylation of KRas for surface plasmon resonance (SPR) [271-273]. Our assay requires neither fluorescent nucleotide analogues nor the biotinylation of KRas, resulting in significantly lower costs and handling of samples. Moreover, the 96-well plate format of the BLI has inherent high-throughput potential, allowing multiple samples to be tested per experiment.

Our initial screen on six selected variants from our computational designs revealed that two of our six designs showed enhanced binding to KRas when compared to RBD(WT). RBD V88Y and RBD N71R/A85K/V88Y (RBD (RKY) appeared to be new RBD superbinders and were selected for further characterization through BLI titration experiments. One mutation T57K variant was observed to decrease binding despite being predicted by OSPREY to enhance affinity. OSPREY is a physics-based interpretable platform and can output structural models that lend themselves to visual inspection. T57K was visually inspected and the error made by the software is amenable to visual inspection and correction. (See section 4.1.3 The Role of Visual Inspection in Computational Protein Design. The T57M, and K87Y mutations both have a roughly neutral effect on binding, which is consistent with the computational predictions. V88F has a neutral effect on binding despite being predicted to have significantly improved binding. Unfortunately, this is an energy function error, and it is unclear why this variant does not benefit from better binding. One possibility is that a water network present near V88F is disrupted by the larger non-polar residue lacking a hydrogen bond donor. Further work would have to be done to evaluate this hypothesis.

The computationally designed V88Y (Figure 8) is a novel mutation that improves binding a significant amount when compared to WT and was suggested by our BLI screens to be comparable or better than the previous best known single mutation variants, A85K and N71R. With the discovery of the new point mutant V88Y (referred to herein as c-Raf-RBD(Y)), the next natural step was to combine it with the mutations found in the best reported variant, N71R and A85K referred to herein as c-Raf-RBD(RK), and our new triple mutant with N71R, A85K, and V88Y mutations, referred to herein as c-Raf-RBD(RKY). Given the promising screening results, a quantitative titration BLI was done to calculate K_D values (see Figure 20-25). Excitingly, we found that c-Raf-RBD(RKY) binds $KRas^{GTP}$ with a K_D of 3.26×10^{-9} ($\pm 2.77 \times 10^{-9}$). This is roughly 4.75 times better than the previous best known binder, c-Raf-RBD(RK), which binds KRas^{GTP} with a K_D of 1.55×10^{-8} (± 1.33×10^{-9}), and approximately 35.5 times better than wild-type c-Raf-RBD, which binds KRas^{GTP} with a K_D of 1.16×10^{-7} (± 2.23×10^{-8}) (Table 2). Our experimental results are in good concordance with previously published affinities of previously known Raf-1 RBD variants to $\text{KRas}^{\text{GppNHp}}$, providing us with further confidence in our assays [246].

Numerous groups have explored the Ras:Raf protein-protein interaction [106, 253, 254, 264, 265, 267, 271, 274-279] and performed mutagenesis on Raf-RBD, either through rational means [106, 209, 253, 276], computational methods [246, 254], or high-throughput evolutionary methods [275, 280]. But until this study, none found V88Y. We believe that our ability to identify prospectively a novel RBD superbinder validates our technology, the importance of provable algorithms, and our ability to explore extensive chemical space with accuracy.

Table 2: Binding affinity and fitting parameters for RBD variants measured by BLI titration experiments.

Run	Variant	K_d	X^2	R^2
1	c-Raf-RBD(RKY)	4.83×10^{-9}	0.0462	0.9997
2	c-Raf-RBD(RKY)	4.01×10^{-9}	0.0529	0.9997
3	c-Raf-RBD(RKY)	2.04×10^{-9}	0.211	0.9947
4	c-Raf-RBD(RKY)	2.15×10^{-9}	0.6266	0.997
Avg	**c-Raf-RBD(RKY)**	**3.26×10^{-9} (± 2.77×10^{-9})**	**0.2342 (± 0.5450)**	**0.998 (±0.005)**

1	c-Raf-RBD(RK)	1.47×10^{-8}	0.2232	0.9972
2	c-Raf-RBD(RK)	1.59×10^{-8}	0.2303	0.9975
3	c-Raf-RBD(RK)	1.58×10^{-8}	0.2269	0.9895
Avg	**c-Raf-RBD(RK)**	**1.55×10^{-8} (± 1.33×10^{-9})**	**0.2268 (± 0.0071)**	**0.9947 (± 0.0096)**
1	A85K	5.18×10^{-8}	0.3814	0.9957
2	A85K	4.53×10^{-8}	0.1414	0.9885
Avg	**A85K**	**4.85×10^{-8} (± 9.29×10^{-9})**	**0.2614 (± 0.3394)**	**0.9921 (±0.010)**
1	V88Y	3.86×10^{-8}	0.7742	0.9914
2	V88Y	2.59×10^{-8}	0.1658	0.9898
Avg	**V88Y**	**3.22×10^{-8} (±18×10^{-8})**	**0.47 (± 0.8604)**	**0.9906 (± 0.0023)**
1	Wild-Type	1.31×10^{-7}	0.0971	0.9938
2	Wild-Type	1.01×10^{-7}	0.2561	0.9917
3	Wild-Type	1.17×10^{-7}	0.4417	0.9856
Avg	**Wild-Type**	**1.16×10^{-7} (± 2.23×10^{-8})**	**0.2650 (± 0.2625)**	**0.9904 (±0.0086)**

2.2.2.5.1 Purification

RBD Purification

His-Sumo-RBD was expressed from a pHSK vector (donated from Pei Zhou lab at Duke University) RBD was inserted through homologous recombination as described in [281].

```
ATGCATCACCATCATCATCACGGGTCCCTGCAGGAGGAGAAGCCCAAGGAAGGAGTCAAGACTGAGAACAACGATCATAT
TAATTTGAAGGTGGCGGGGCAGGATGGTTCTGTGGTGCAGTTTAAGATTAAGAGGCATACACCACTTAGTAAACTAATGA
AAGCCTATTGTGAACGACAGGGATTGTCAATGAGGCAGATCAGATTCCGATTTGACGGGCAACCAATCAATGAAACAGAC
ACACCTGCACAGTTGGAAATGGAGGATGAAGATACAATTGATGTGTTCCAGCAGCAGACGGGAGGTAACACTATCCGTGT
TTTCTTGCCGAACAAGCAAAGAACAGTGGTCAATGTGCGAAATGGAATGAGCTTGCATGACATCCTTATGAAAGCACTCA
AGGTGAGGGGCCTGCAACCAGAGTGCATGGCAGTGTTCAGACTTCTCCACGAACACAAAGGTAAAAAGCACGCTTAGAT
TGGAATACTGATGCTGCGTCTTTGATTGGAGAAGAACTTCAAGTAGATTTCCTGGATTAG
```

Figure 9: His-Sumo-RBD(C81I, C96M) gene sequence. Sumo sequence in Red, RBD sequence in black.

```
MHHHHHHGSLQEEKPKEGVKTENNDHINLKVAGQDGSVVQFKIKRHTPLSKLMKAYCERQGLSMRQIRFRFDGQPINETDTPAQLE
MEDEDTIDVFQQQTGGNTIRVFLPNKQRTVVNVRNGMSLHDILMKALKVRGLQPECMAVFRLLHEHKGKKARLDWNTDAASLIGEE
LQVDFLD*
```

Figure 10: His-Sumo-RBD(C81I,C96M) protein sequence. Sumo sequence in red RBD sequence in black. Sumo protease generates a blunt cleavage after QQTGG/. Sumo RBD pI: 7.19, MW: 20582.46, Ext: 6990. RBD pI: 9.43 MW: 8840.35 Ext: 5500

His-Sumo-RBD variants were made in a C81I, C96M background. These mutations have little effect on Raf folding or binding to KRas and are thought to increase expression in E.coli cells [275, 280]. Moreover, these mutations allow us to generate an RBD variant containing a single thiol, which can be specifically modified using thiol-reactive chemistry. We use this feature for our conjugation of Bodipy dye through maleimide-mediated click chemistry to our RBD in our FP assay. Moreover, RBD was expressed with a His-Sumo tag to increase expression and facilitate purification. His-SumoRBD was grown at 37°C in Rosetta 2 DE3 cells supplemented with Kanamycin and Chloramphenicol to an OD_{600} of 0.8. Cells were then induced with 1 mM IPTG, and the temperature was lowered to 16°C overnight. Cells were spun-down, resuspended in 40mM Tris-HCl, 50mM NaCl, 5mM $MgCl_2$, 5mM $CaCl_2$ 1 mM TCEP pH8.0 with 1 mM PMSF (lysis buffer) and then lysed with 5 passages through a homo-fluidizer. 5 µL of Micrococcal nuclease/liter of cell growth was added and cells were incubated for 20min at 37°C. Lysate was then run through a Nickel NTA column with a gradient elution from lysis buffer to 40 mM Tris-HCl, 500 mM NaCl, 5 mM $MgCl_2$, 500 mM Imidazole, 1 mM TCEP pH7.4. Eluted peaks were then collected, concentrated in a 10kD molecular weight cut-off Amicon concentrator and desalted through a G-25 sephadex column into 40 mM Tris-HCl, 250 mM NaCl, pH 8.0. His-SumoRBD was then concentrated again before being exposed to 3mg of Sumo Protease at 30°C overnight. RBD now cleaved from the His-Sumo tag, was spun down to remove aggregates and then passed again through a nickel NTA immobilized metal affinity chromatography column with a slow gradient elution from lysis buffer to 40 mM Tris-HCl, 500 mM NaCl, 5 mM $MgCl_2$, 500 mM Imidazole, 1 mM TCEP pH7.4. Flow-through was collected, concentrated and desalted through a G-25 column into 50 mM phosphate pH 6.5. Desalted RBD was further purified through passage into a cation exchange SP sepharose column eluted with 50 mM phosphate, 500 mM NaCl, pH 6.5 5 mM TCEP. RBD was concentrated and buffer exchanged through a 3kD molecular-weight cut-off Amicon concentrator into 40 mM Phosphate, 250 mM NaCl, $MgCl_2$, 1 mM TCEP pH 7.4. Purity of variants was determined by SDS-PAGE gel (Figure 13), and identity was determined by matrix assisted laser desorption ionization time of flight (MALDI-TOF) mass spectrometry.

KRas Purification

His-KRas C118S was expressed from a pET28b (Novagen) plasmid with the KRas gene inserted between the NdeI and XhoI cleavage sites ([282], donated from the Michael Kennedy lab in Miami University in Ohio).

```
ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGACTGAATATAAACTTGT
GGTAGTTGGAGCTGGTGGCGTAGGCAAGAGTGCCTTGACGATACAGCTAATTCAGAATCATTTTGTGGACGAATATGATC
CAACAATAGAGGATTCCTACAGGAAGCAAGTAGTAATTGATGGAGAAACCTGTCTCTTGGATATTCTCGACACAGCAGGT
CAAGAGGAGTACAGTGCAATGAGGGACCAGTACATGAGGACTGGGGAGGGCTTTCTTTGTGTATTTGCCATAAATAATAC
TAAATCATTTGAAGATATTCACCATTATAGAGAACAAATTAAAAGAGTTAAGGACTCTGAAGATGTACCTATGGTCCTAG
TAGGAAATAAATCTGATTTGCCTTCTAGAACAGTAGACACAAAACAGGCTCAGGACTTAGCAAGAAGTTATGGAATTCCT
TTTATTGAAACATCAGCAAAGACAAGACAGGGTGTTGATGATGCCTTCTATACATTAGTTCGAGAAATTCGAAAACATAA
AGAAAAGTAG
```

Figure 11: His-KRas C118S gene sequence

```
MGSSHHHHHHSSGLVPRGSHMTEYKLVVVGAGGVGKSALTIQLIQNHFVDEYDPTIEDSYRKQVVIDGETCLLDILDTAGQEEYSA

MRDQYMRTGEGFLCVFAINNTKSFEDIHHYREQIKRVKDSEDVPMVLVGNKSDLPSRTVDTKQAQDLARSYGIPFIETSAKTRQGV

DDAFYTLVREIRKHKEK*
```

Figure 12: His-KRas C118S protein sequence. Thrombin cleavage site shown in red. pI 6.11, MW:21393.01Da, Ext. coefficient: 11920

His-KRas C118S, a mutation introduced to increase expression and stability [204], was grown at 37°C in Rosetta 2 DE3 Cells supplemented with Kanamycin and Chloramphenicol to a OD_{600} of 0.8. Cells were then induced with 1 mM IPTG and allowed to grow at 37°C for 5 hours before being lowered to 16°C overnight. Cells were spun-down, resuspended in 40 mM Tris-HCl, 50 mM NaCl, 5 mM $MgCl_2$, 5 mM $CaCl_2$ 1 mM TCEP pH8.0 with 1 mM PMSF (lysis buffer) and then lysed with 5 passages through a homo-fluidizer. 5ul of Micrococcal nuclease (stock concentration 2,000,000 gel units/ml, from NEB) /liter of cell growth was added and cells were incubated for 20min at 37°C. Lysate was then run through a Nickel NTA column with a gradient elution from lysis buffer to 40 mM Tris-HCl, 500 mM NaCl, 5 mM $MgCl_2$, 500 mM Imidazole, 1 mM TCEP pH7.4. Eluted peaks were then collected and desalted through a G-25 sephadex column into 50 mM Tris-HCl, 50 mM NaCl , pH 8.0 (desalting buffer). KRas was concentrated in a 10kD molecular weight cut off Amicon concentrator in preparation for nucleotide loading. The desalting buffer was supplemented with 200 mM Ammonium Sulphate ($(NH_4)_2SO_4$) and 1μM Zinc Chloride ($ZnCl_2$) as well as 50 units of calf intestinal phosphatase (CIP), 10 molar excess of Guanosine 5′-[β,γ-imido]triphosphate (GppNHp) a non-hydrolysable GTP analogue. KRas was incubated with GppNHp for 2 hours at room temperature and then overnight at 4°C. Reaction was terminated by the addition of 60 mM $MgCl_2$. $KRas^{GppNHp}$ was further purified and buffer exchanged into 40 mM Tris-HCl, 250 mM NaCl, 5 mM $MgCl_2$, 1 mM TCEP pH7.4 by size exclusion chromatography on a Superdex 75 column. Purity was confirmed by SDS-PAGE (Figure 13) and loading was confirmed by high-performance anion exchange chromatography (HPAEC, Figure 19).

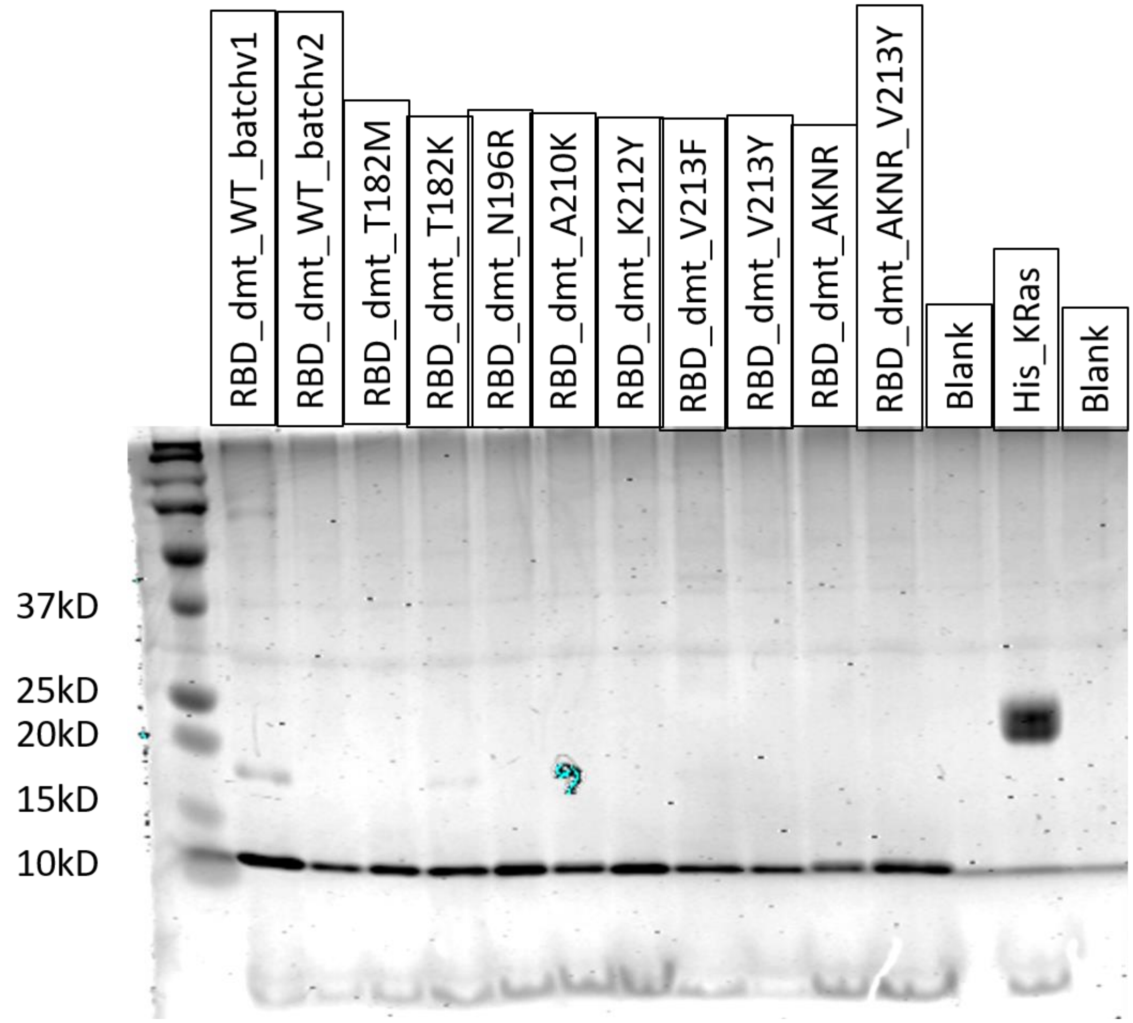

RBD_dmt_WT_batchv1
RBD_dmt_WT_batchv2
RBD_dmt_T182M
RBD_dmt_T182K
RBD_dmt_N196R
RBD_dmt_A210K
RBD_dmt_K212Y
RBD_dmt_V213F
RBD_dmt_V213Y
RBD_dmt_AKNR
RBD_dmt_AKNR_V213Y
Blank
His_KRas
Blank
37kD
25kD
20kD
15kD
10kD

Figure 13: 12% SDS-PAGE gel of purified KRas and RBD variants proteins with ladder in lane 1 for size comparison. Blank on lane 13 and 15 for separation and for visualization of background. Phantom bands at ~30kD and ~37 kD are visible in control lanes, suggesting they are not real contaminants.

MALDI-TOF Mass Spectroscopy

MALDI-TOF was used to determine the identity of the RBD variant and the conjugation of dye to the variant (see section 2.2.2.5.3). MADLI-TOF was performed on a Bruker Autoflex Speed LRF MALDI-TOF platform. 1μl of saturated Sinapinic acid in 6:1:3 water, 3% TFA and acetonitrile was added to 1μl of protein at ~1μg/μl. Mixed by pipetting up and down and allowed to dry on a MALDI plate. Samples are allowed to dry and crystalize under gentle aeration. Following complete crystallization, samples were sampled through positive ionization and in linear mode.

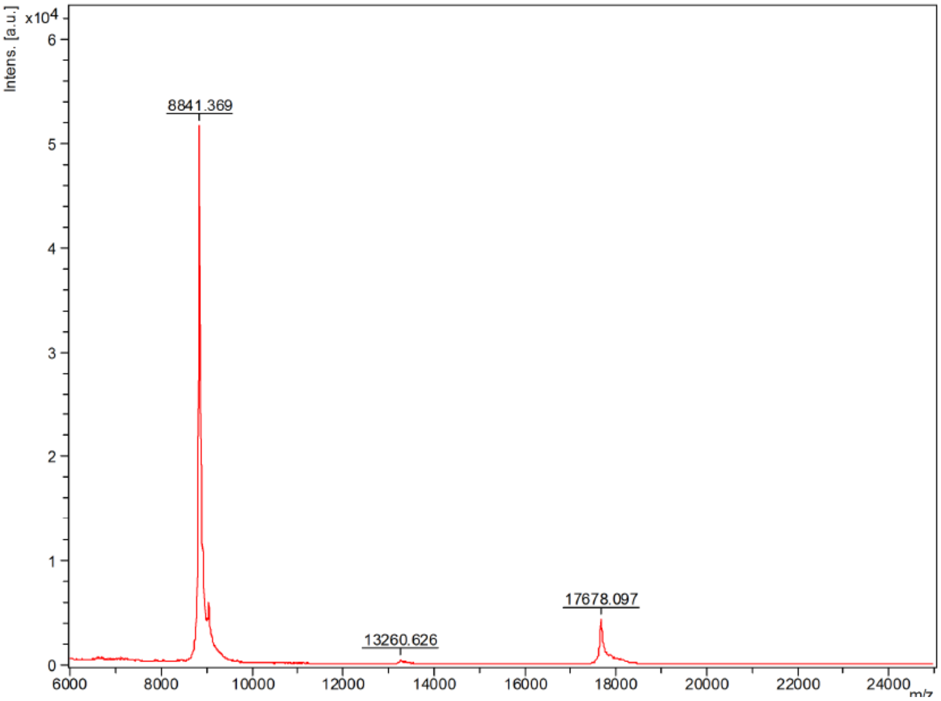


Figure 14: MALDI-TOF mass spectrum of RBD(wt). Expected Molecular weight 8841Da; observed molecular weight 8841Da. Peak at 17676Da corresponds to the dimer.

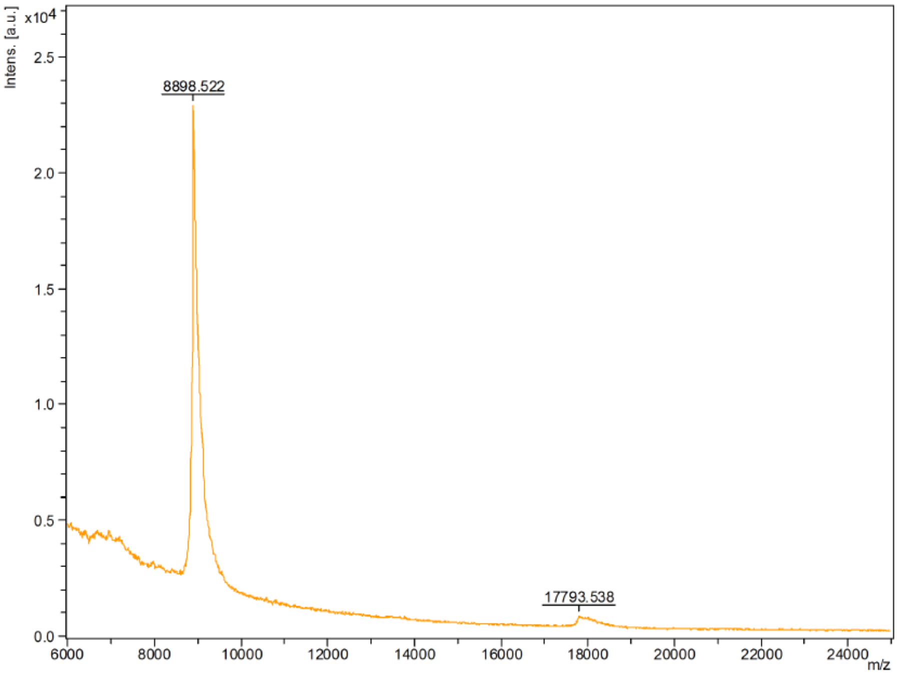


Figure 15: MALDI-TOF mass spectrum of RBD(A85K). Expected Molecular weight 8897Da; observed molecular weight 8899Da. Peak at 17793Da corresponds to the dimer.

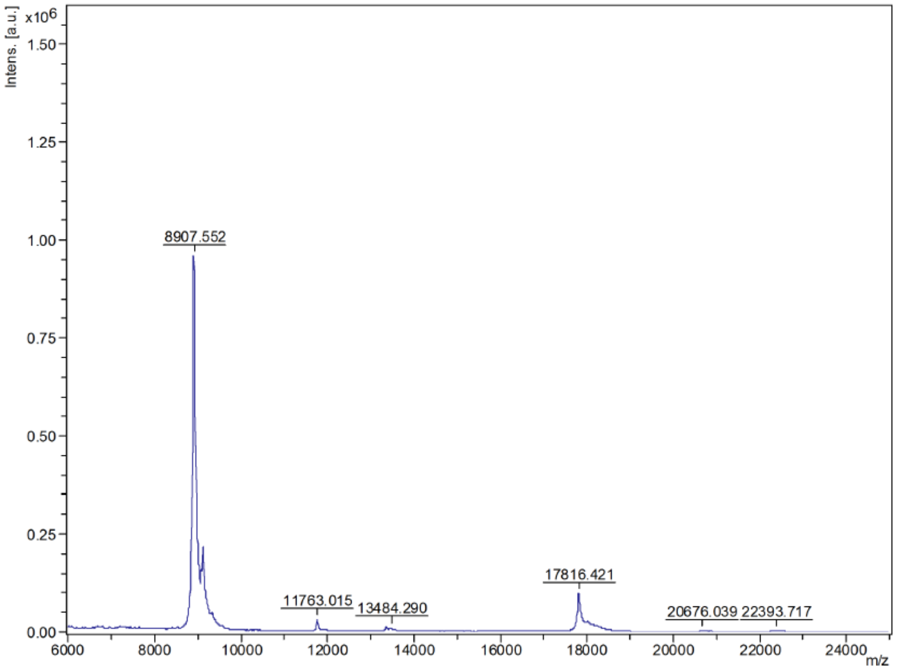


Figure 16: MALDI-TOF mass spectrum of RBD(V88Y). Expected Molecular weight 8904Da; observed molecular weight 8907Da. Peak at 17816Da corresponds to the dimer.

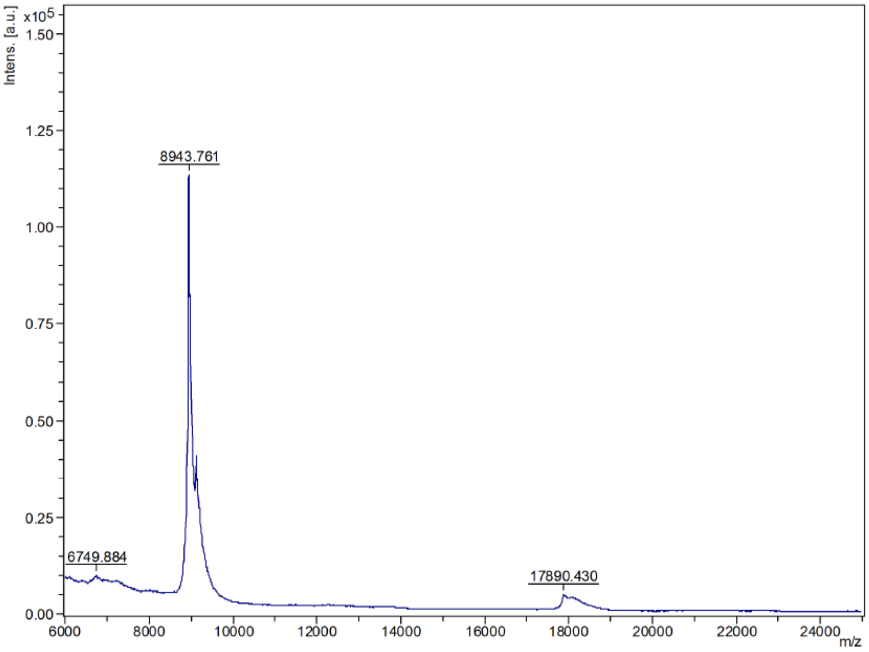


Figure 17: MALDI-TOF mass spectrum of RBD(RK). Expected Molecular weight 8939Da; observed molecular weight 8943Da. Peak at 17890Da corresponds to the dimer.

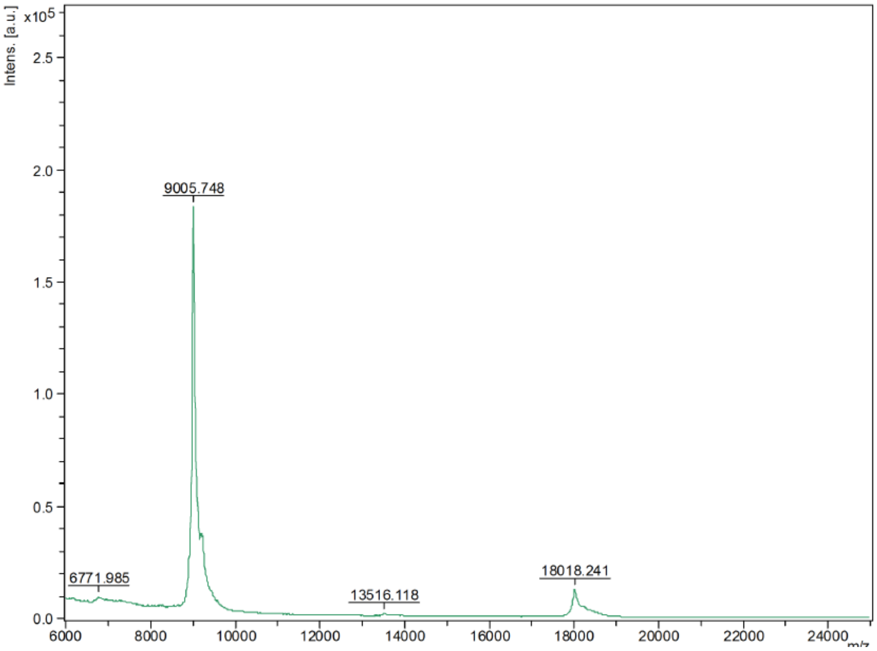


Figure 18: MALDI-TOF mass spectrum of RBD(RKY). Expected Molecular weight 9003Da; observed molecular weight 9006Da. Peak at 18018Da corresponds to the dimer.

2.2.3 High Performance Anion Exchange Chromatography

For HPAEC analysis of nucleotide loading efficiency, 40μL of KRas either loaded or unloaded was treated with 2.5 μL 10% perchloric acid. This denatures the protein and causes aggregation. The pH is raised by the addition of 1.75 μL 4 M CH_3COONa (sodium acetate) pH 4.0. Precipitated protein is spun out in a benchtop microfuge for 10min at max speed, and 12.5 μL of the supernatant containing the nucleotide is loaded onto a DNAPac PA100 column on the HPAEC. The nucleotides were eluted from the column through a gradient of 2% to 100% 1M Ammonium acetate increasing at 3%min , with a flow rate of 1ml/min. To determine loading, KRas was run before loading with

GppNHp, after loading with GppNHp, and spiked with GppNHp (Figure 19). GppNHp spiked KRas^{GDP} (grey) shows a peak at ~11min. However, KRas^{GDP} (red) by itself does not, identifying the peak at 11min as GppNHp. KRas loaded with GppNHp(green) does not show a peak at ~8min. However, when spiked with GDP(grey) a peak appears, identifying the peak at ~8min as the GDP peak. GppNHp-loaded KRas shows the peak at ~11min and no peak at ~8min, showing that our methods are capable of generating complete GppNHp loading of KRas.

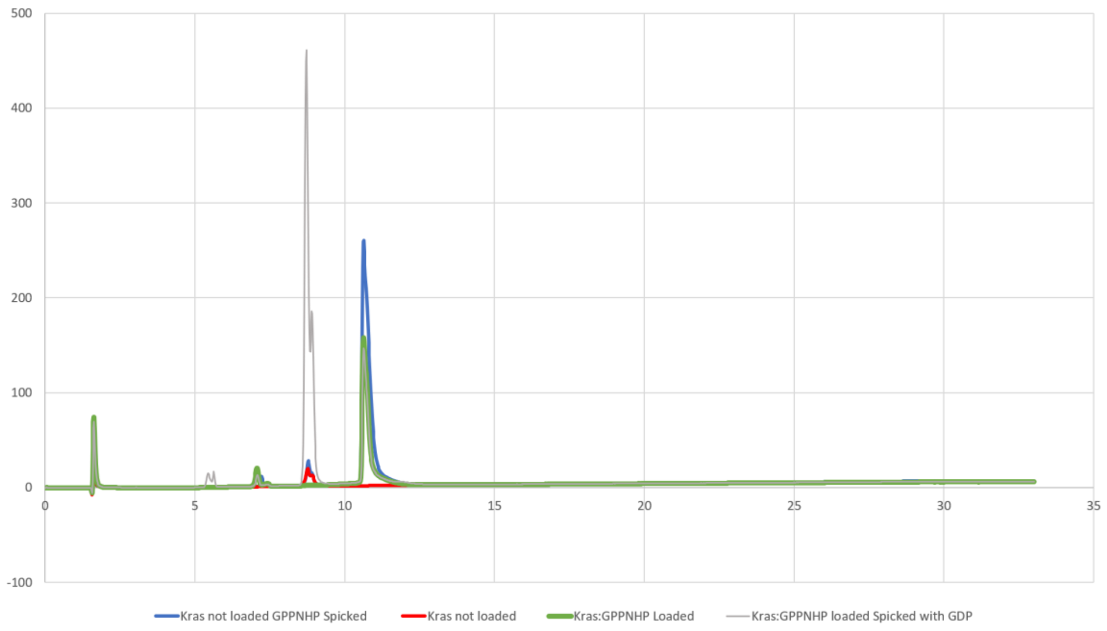


Figure 19: HPAEC chromatogram of KRas bound to different nucleotides or in apo state. KRas loaded with GppNHp (green), not loaded with GppNHp (red), not loaded with GppNHp but spiked with free GppNHp (blue), and KRas loaded with GppNHp but spiked with GDP (grey). The peak at ~11min can be determined to be the GppNHp peak and the peak at ~8min to be the GDP peak. GppNHp-loaded KRas does not have the peak at ~8min, showing complete loading.

2.2.2.5.2 BLI

Bio layer interferometry or BLI is a platform to detect binding in a tag-free, real-time, ligand-immobilized way by measuring the change in the interferometry pattern between white light reflected from a standard internal reflection point and a biolayer on the tip surface. Changes to the difference in distance between the points of reflection generate a nanometer (nm)-scale shift in phase between the reflected lights generating an interferometry pattern. This distance is affected by events on the biolayer such as binding or dissociation on the tip's biolayer surface.

Figure 20 shows a typical BLI experiment. First tips are equilibrated in buffer. These tips are moved to a well containing a ligand, which is captured on the tip using either a tag or click chemistry. Tips are then equilibrated to generate a baseline in a well containing only buffer. Once a baseline can be achieved, tips are moved to a well containing varying concentrations of an analyte, and binding is allowed to occur for a user-determined period of time. Tips are then moved into a well containing only buffer so that dissociation can be observed.

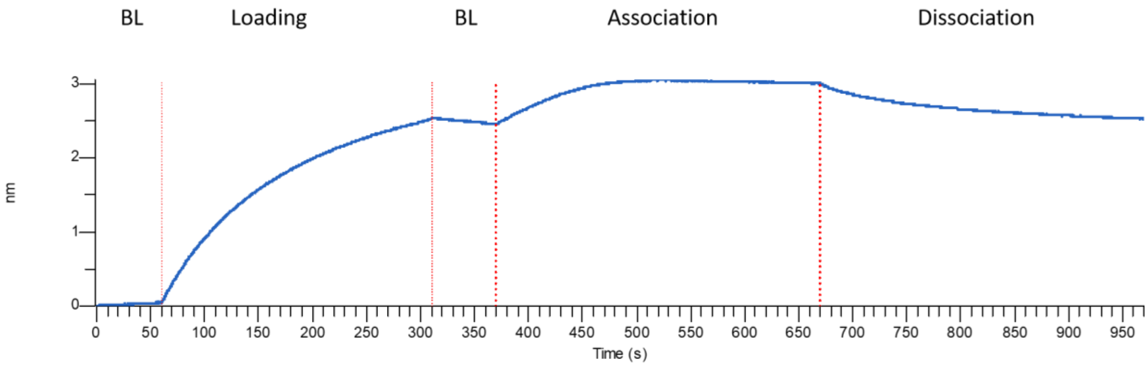


Figure 20: Typical BLI experiment. BL (baseline equilibration)

Assay Validation: KRas^{GDP} vs KRas^{GppNHp}

Arguably the best control to determine if a KRas binding assay is working is specificity for KRas^{GppNHp} vs KRasGDP. This is because KRas^{GppNHp} binds to C-Raf RBD with over 300x greater affinity than KRas^{GDP}, so one would expect to see a significant difference in binding [100-104, 246]. KRas, from the same prep, was either loaded or not with GppNHp, underwent no further purification, and was loaded onto BLI NiNTA tips as described above. Loaded tips were allowed to equilibrate in BLI buffer and then dipped into to 300nM of RBD(WT). Figure 21 shows that, as is expected of a functioning assay, the tip with KRas^{GppNHp} immobilized presents with strong signal, whereas the tip immobilized with KRas^{GDP} does not show any signal. As the protein preps for both KRas and RBD are identical, one can conclude that possible contaminants or prep-specific factors are not the cause of the difference. Moreover, because both tips could equilibrate after loading, differences in buffer between GDP- and GppNHp-loaded KRas cannot explain the difference in signal. Finally, it is possible that the difference in the act of GppNHp loading somehow denatures KRas and what we are seeing is non-specific binding. However, the fast off-rate is uncharacteristic of aggregation, as is the saturation during the association step of the experiment. Therefore, denaturation is a very unlikely explanation of the results. However, the difference in buffer conditions between the KRas samples during the loading step does explain the difference in the capture of the different KRas samples by the NiNTA BLI tip. This difference in captured KRas could potentially be significant as KRas^{GDP} only captured ~69% as much as KRas^{GppNHp} (1.1 nm vs 1.6 nm). However the difference in response between the tip with captured KRas^{GppNHp} (0.526 nm) and KRas^{GDP} (0.0193 nm) is over 27 fold, making a 69% difference in captured KRas unlikely to be the reason. Moreover, RBD binding to the BLI NiNTA tip without any captured KRas was attempted to test for non-specific binding. The response of RBD binding to a KRas free tip was greater than to a tip with captured KRas^{GDP} binding at (0.0294nm). This indicates that the difference in amount captured was not responsible for the difference in signal between the different KRas states. It also indicates that non-specific binding is greater to the tip alone than when the tip is covered in a protein. Thus, this data strongly suggests that this BLI assay is specifically picking up the effect of GppNHp loading and accurately measuring binding.

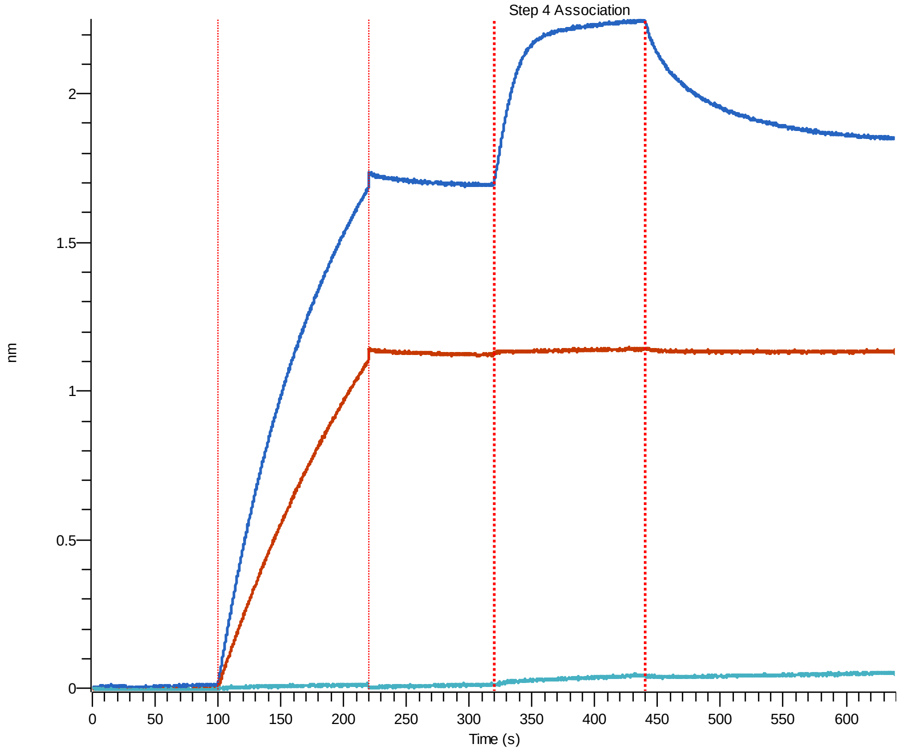


Figure 21: BLI experiment of RBD(WT) binding to either KRas^{GDP} (red) or KRas^{GppNHp} (blue). RBD(WT) binding to the NiNTA sensor alone (non-specific binding) is shown in green. Binding to KRas^{GppNHp} vs KRas^{GDP} shows a 27x greater response, validating the BLI assay.

BLI Screening of RBD Variants

BLI NiNTA tips were equilibrated in 30 mM Phosphate pH7.4, 327 mM NaCl, 2.7 mM KCl, 5 mM MgCl2, 1.5 mM TCEP, 0.1% BSA, 0.02% Tween-20 + Kathon (BLI buffer). NiNTA tips were then moved into wells containing His-KRas:GTP at 20ug/ml in BLI buffer for tip loading. Tips were allowed to re-equilibrate before being transferred to wells each containing 250nM of a different RBD variant in BLI buffer. Tips were then moved back into BLI buffer to observe dissociation. Data was worked-up on the Data

Analysis HT software (ForteBio) and fit using the built-in mass transport model. Y-axis was aligned to the average of the baseline step while the inter-step correction was aligned to the dissociation step. Experiments were all conducted at 25°C and with 1000 RPM shacking.

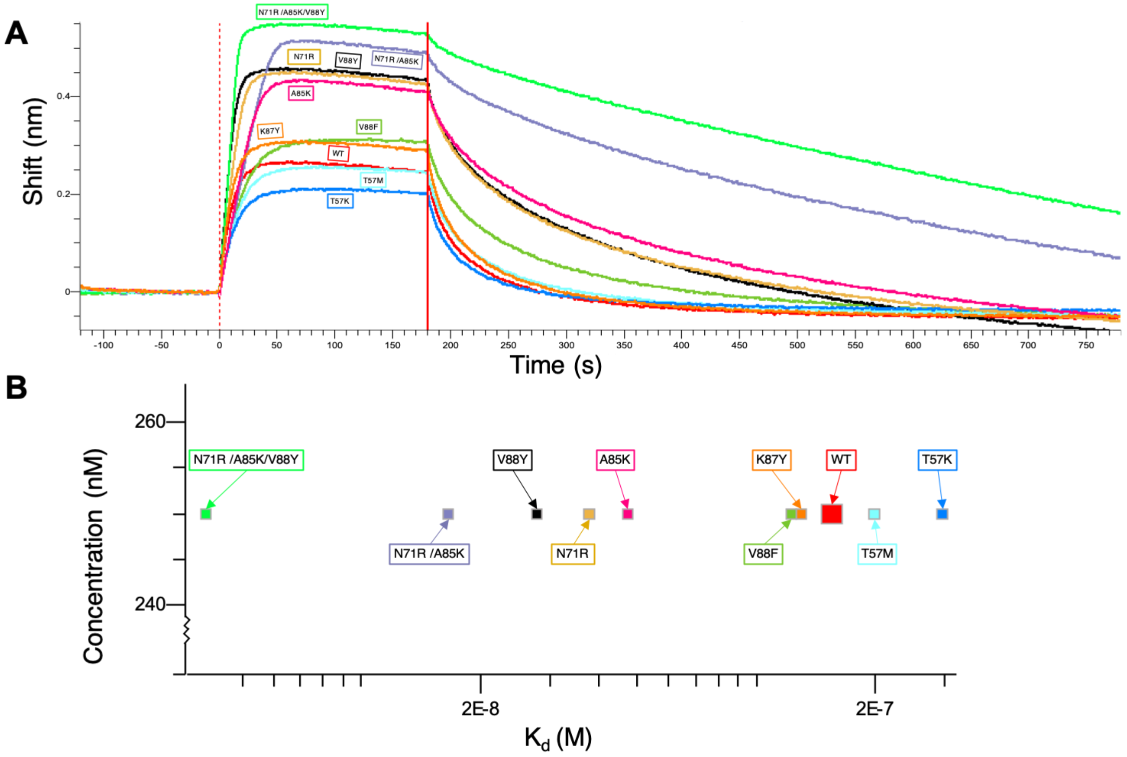


Figure 22: Results for BLI screening of RBD variants. The designed mutants V88Y and N71R/A85K/V88Y (RBD (RKY)) show significantly improved binding when compared to WT. RBD(RKY) presents itself as the highest affinity Raf variant ever designed.

Experimental Set-Up Titrations

BLI NiNTA tips were equilibrate in 30 mM Phosphate pH7.4, 327 mM NaCl, 2.7 mM KCl, 5 mM MgCl2, 1.5 mM TCEP, 0.1% BSA, 0.02% Tween-20 + Kathon (BLI buffer). NiNTA tips were then moved into wells containing His-KRas:GTP at 20ug/ml in BLI buffer for tip loading. Tips were allowed to re-equilibrate before being transferred to wells containing diverse concentrations of an RBD variant in BLI buffer. A well containing only buffer (0nM RBD) was used to subtract for drift. After the association step, tips were moved back into wells containing only BLI buffer to observe the dissociation steps. Experiments were all conducted at 25°C and with 1000 RPM shacking. To control for aggregation and non-specific binding either His-SumoRBD or His-Hip (a unrelated E.coli protein [283]) was captured on the BLI NiNTA tips and exposed to RBD variants. No non-specific binding was observed (Figure 29). Data was worked-up on the Data Analysis HT software (ForteBio) and fit using the built-in mass transport model. Y-axis was aligned to the average of the baseline step while the inter-step correction was aligned to the dissociation step. Fits were accepted if χ^2 was lower than 1 and R^2 was greater than 0.98.

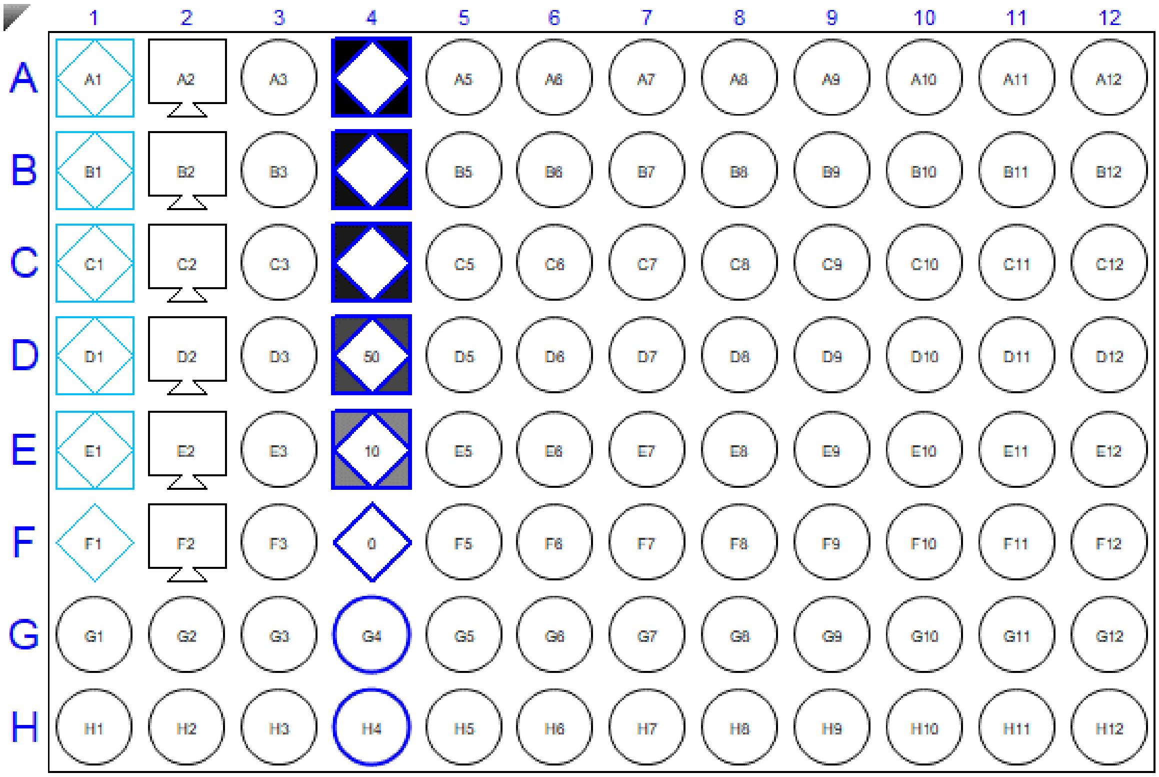

1 2 3 4 5 6 7 8 9 10 11 12
A A1 A2 A3 A5 A6 A7 A8 A9 A10 A11 A12
B B1 B2 B3 B5 B6 B7 B8 B9 B10 B11 B12
C C1 C2 C3 C5 C6 C7 C8 C9 C10 C11 C12
D D1 D2 D3 50 D5 D6 D7 D8 D9 D10 D11 D12
E E1 E2 E3 10 E5 E6 E7 E8 E9 E10 E11 E12
F F1 F2 F3 0 F5 F6 F7 F8 F9 F10 F11 F12
G G1 G2 G3 G4 G5 G6 G7 G8 G9 G10 G11 G12
H H1 H2 H3 H4 H5 H6 H7 H8 H9 H10 H11 H12

Figure 23: BLI sensor plate showing drift subtraction. Column 1 is only buffer whereas column 2 has 20ug/ml of His-KRas. Column 4 has RBD in varying concentrations. Well (4f) has only buffer and is used to correct for drift.

The mass transport model corrects for situations in which the association rate is significantly faster than the rate of diffusion K_m between the surface of the tip and the diffusing analyte. In this situation most of the kinetics are diffusion limited and on rates seem slower than they are. Ideally this factor can be overcome by correcting for the density of ligand on the tip surface or increasing the diffusion rate. In our case signal was limiting due to the small size of RBD (8kD) and reducing analyte concentration reduced signal. Therefore, experiments were conducted at the lowest possible ligand concentration. However, it was still observed that the 1:1 binding model did not fit accurately to the high-affinity designed RBD variants. Although this can be a limitation in our experimental design, it may also be indicative of the two part binding previously reported for the KRas:Raf-RBD interaction [269]. In this case k_m would not be the measurement for bulk to surface transfer of analyte, but of an initial fast association rate between RBD and KRas. Testing this hypothesis requires extensive further characterization so we adopted a model that corrects during fitting for this effect Equation 1 and reported K_D values obtained from the corrected model.

$$A_{bulk} \underset{k_m}{\overset{k_m}{\rightleftharpoons}} A_{surf} + B \underset{k_d}{\overset{k_a}{\rightleftharpoons}} AB$$

$$\frac{dR}{dt} = \frac{k_a * [\text{Analyte}]}{1 + \frac{k_a}{k_m}(R_{max} - R)}(R_{max} - R) - \frac{k_d}{1 + \frac{k_a}{k_m}(R_{max} - R)} R$$

Equation 1: Mass transport model. (Adapted from ForteBio user training manual) k_m is the rate of diffusion of analytes in the bulk solvent to analytes near and able to interact with the surface of the sensor. k_d and k_a represent the on and off rate respectively. Rmax is the maximum signal produced from full saturation of ligands immobilized on the surface and R is the experimentally measured response.

BLI Titration Results

Figures 20-25 shows the titration curves and fit results for the titrations of the RBD variants. A summary of the results can be found in Table 2. The results show that the designed mutants V88Y (K_D 32.2nM ± 18nM) and RKY (K_D 3.26nM ± 2.77nM) show significantly improved affinity to $KRas^{GppNHp}$ when compared to RBD(WT) (K_D 1 116nM ± 22.3nM). Moreover, we have successfully designed the RBD variant with the highest affinity for $KRas^{GppNHp}$ ever recorded. Our RKY variant binds nearly 5x better than the previous tightest binding variant RBD(RK) (K_D 15.5nM ± 1.33nM).

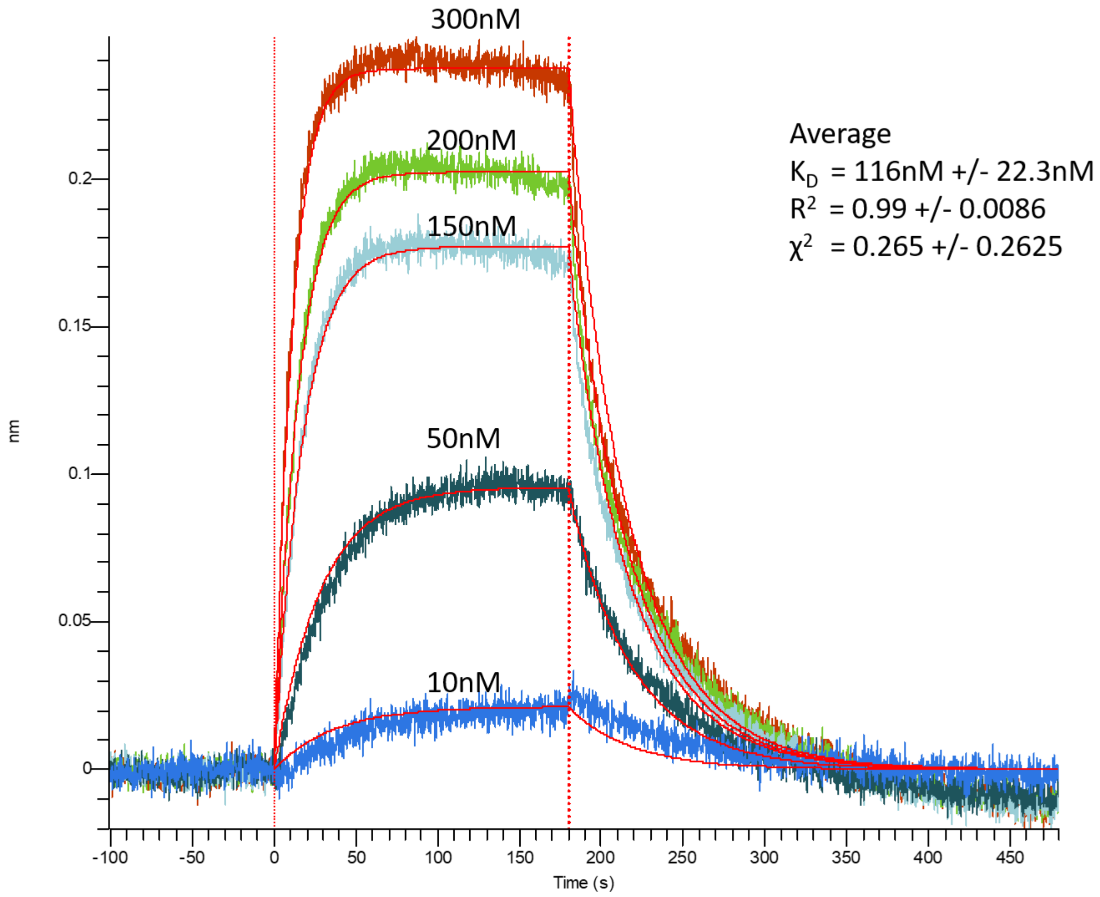


Figure 24: BLI titration curve for RBD(WT) concentrations show on graph by titrations curves. Affinity and accuracy of fit also shown on graph. Colored lines are experimental data, fits shown as red line. Dotted red lines denote the beginning of the association phase and dissociation phase.

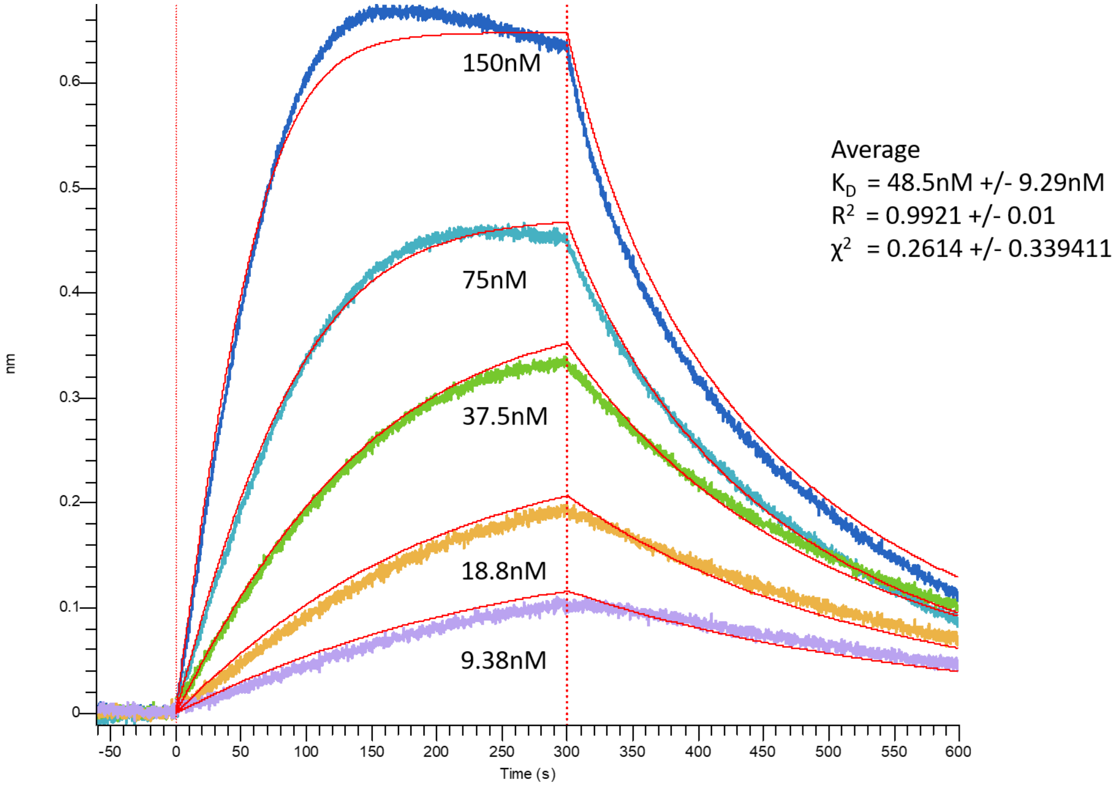


Figure 25: BLI titration curve for RBD(A85K) concentrations show on graph by titrations curves. Affinity and accuracy of fit also shown on graph. Colored lines are experimental data, fits shown as red line. Dotted red lines denote the beginning of the association phase and dissociation phase. Average results shown in the right.

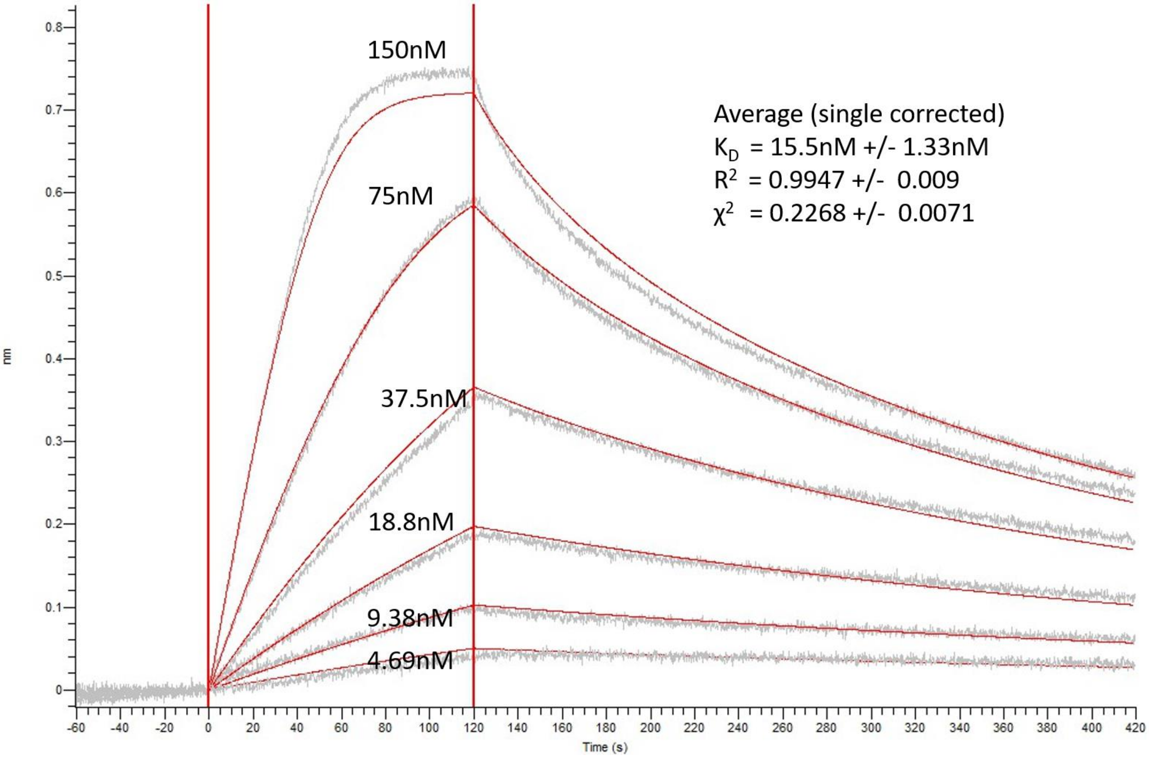


Figure 26: BLI titration curve for RBD(RK) concentrations show on graph by titrations curves. Affinity and accuracy of fit also shown on graph. Colored lines are experimental data, fits shown as red line. Dotted red lines denote the beginning of the association phase and dissociation phase.

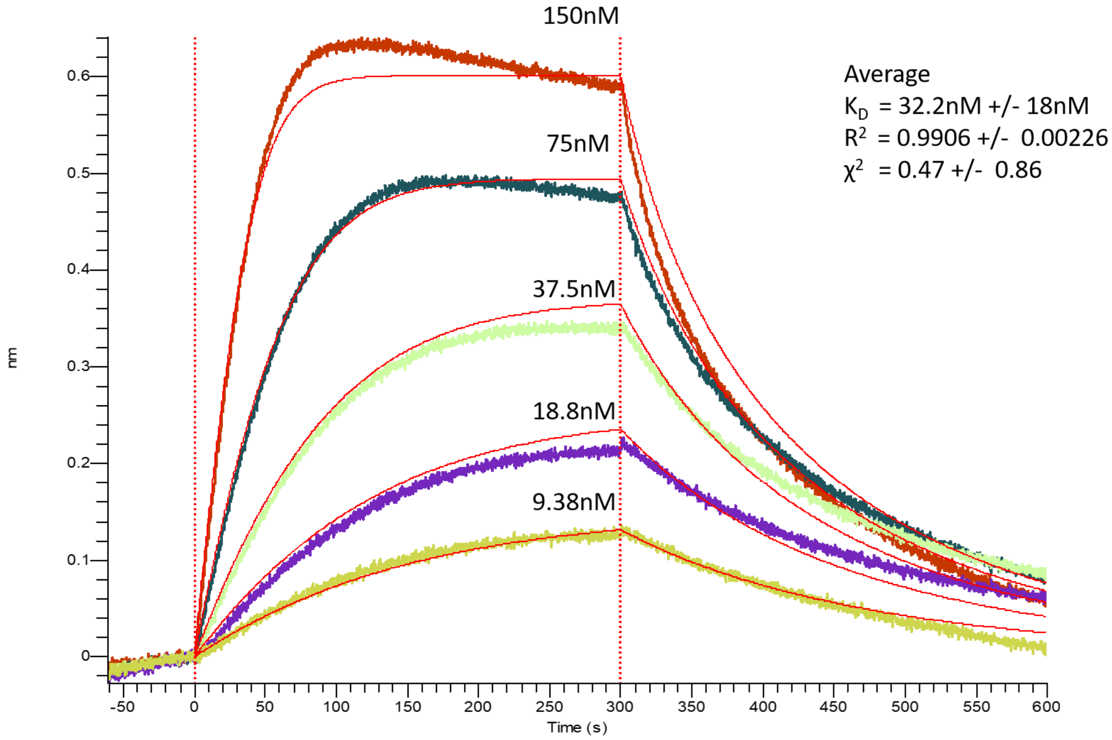


Figure 27: BLI titration curve for RBD(V88Y) concentrations show on graph by titrations curves. Affinity and accuracy of fit also shown on graph. Colored lines are experimental data, fits shown as red line. Dotted red lines denote the beginning of the association phase and dissociation phase.

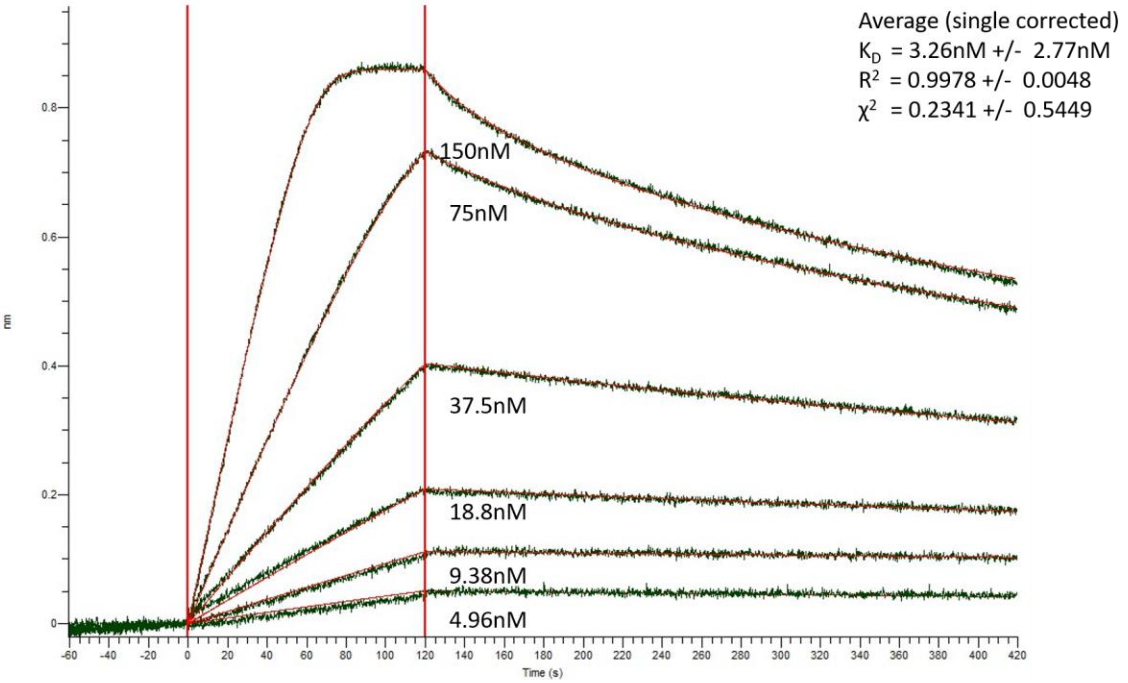


Figure 28: BLI titration curve for RBD(RKY) concentrations show on graph by titrations curves. Affinity and accuracy of fit also shown on graph. Colored lines are experimental data, fits shown as red line. Dotted red lines denote the beginning of the association phase and dissociation phase.

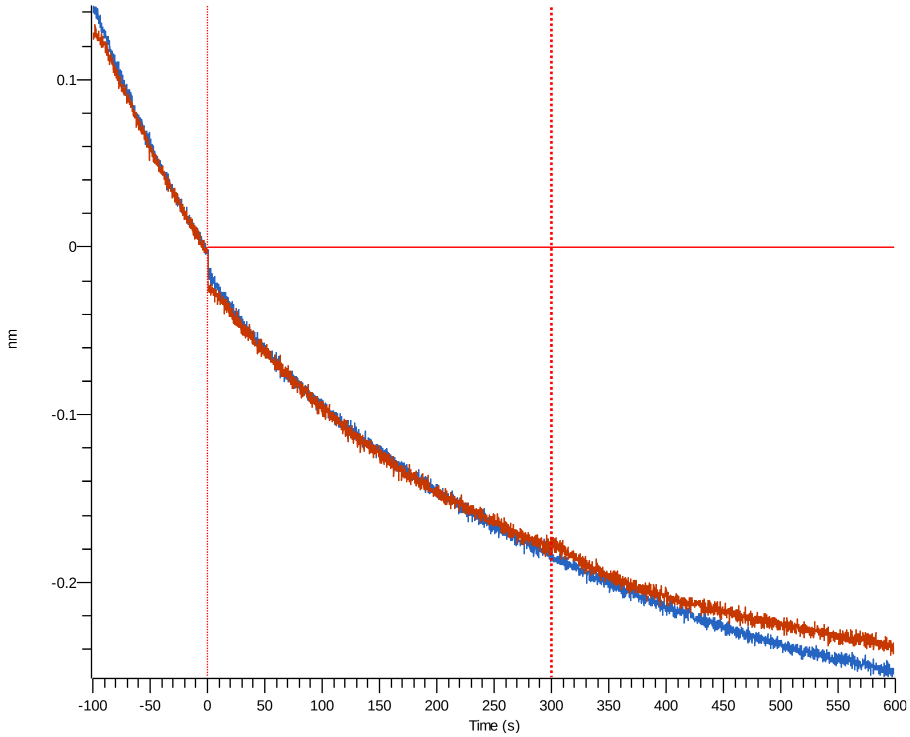


Figure 29: Testing for non-specific binding. RBD(WT) binding to His-SumoRBD(WT) (Blue) and His-Hip (Red). No non-specific binding was observed.

2.2.2.5.3 Fluorescence Polarization experiments

Dye Conjugation

In order to further verify our BLI results, and to generate a relatively inexpensive and widely available assay for testing binding to KRas, we designed an alternative assay to determine the affinity of our designs. This work was done in close collaboration with Rachel Kimbrough, a research assistant in the Oas lab. RBD variants were expressed in a cysteine double mutant background as described in section (0). The remaining cysteine C95 is surface exposed (Figure 30) making it ideal for specific and monovalent conjugation of our RBD variants to a BODIPY-Maleimide Dye (ThermoFisher). The protocol was identical for all RBD variants. Conjugation was achieved through the following protocol. Initially RBD was exposed to 10 mM Dithiothreitol (DTT) for 2 hours at 4°C. Reduced RBD was ammonium sulfate precipitated by the addition of 90% m/v of ammonium sulfate to RBD. Precipitated RBD was washed three times with degassed 100 mM phosphate, 200 mM NaCl, 1 mM EDTA pH 7.4 supplemented with 70%w/v $(NH_4)_2SO_4$. RBD was resuspended in degassed $(NH_4)_2SO_4$ free 100 mM phosphate, 200 mM NaCl, 1 mM EDTA pH 7.4. This method guaranteed the RBD was reduced while also allowing us to obtain a solution free of reducing agents. However, multiple washes caused significant protein loss. To overcome the loss of protein the following protocol was developed. 5 mM Tris(2-carboxyethyl)phosphine (TCEP) TCEP was added to purified RBD and incubated for 2 hours at 4°C. TCEP lacks thiol groups and therefore is less reactive with maleimide [284], allowing us to obtain reduced protein with significantly less handling and greater yield. RBD in reducing condition was then immediately degassed for >4 hours. BODIPY-Maleimide (less than 6 months old) was dissolved in anhydrous DMSO to a stock concentration of 10 mM. Prepared dye was used immediately, and from this point all samples were protected from light. Protein samples were at a concentration ~100μM in a volume of ~200μL. Dye was added to the protein so that the dye was at 10-fold molar excess and the concentration of DMSO did not exceed 10%v/v in the reaction. The conjugation reaction was allowed to go overnight at 4°C. Reaction was then allowed to proceed at room temperature for 1 hour before being quenched through the addition of 3 mM DTT. Excess dye was removed using Pierce dye removal columns (ThermoFisher). Dye removal was done in accordance with the protocol described by the manufacturer with reaction volumes being 200µl and 400µl dye removal resin used per reaction. Successful conjugation was determined by MALDI-TOF (Figure 31).

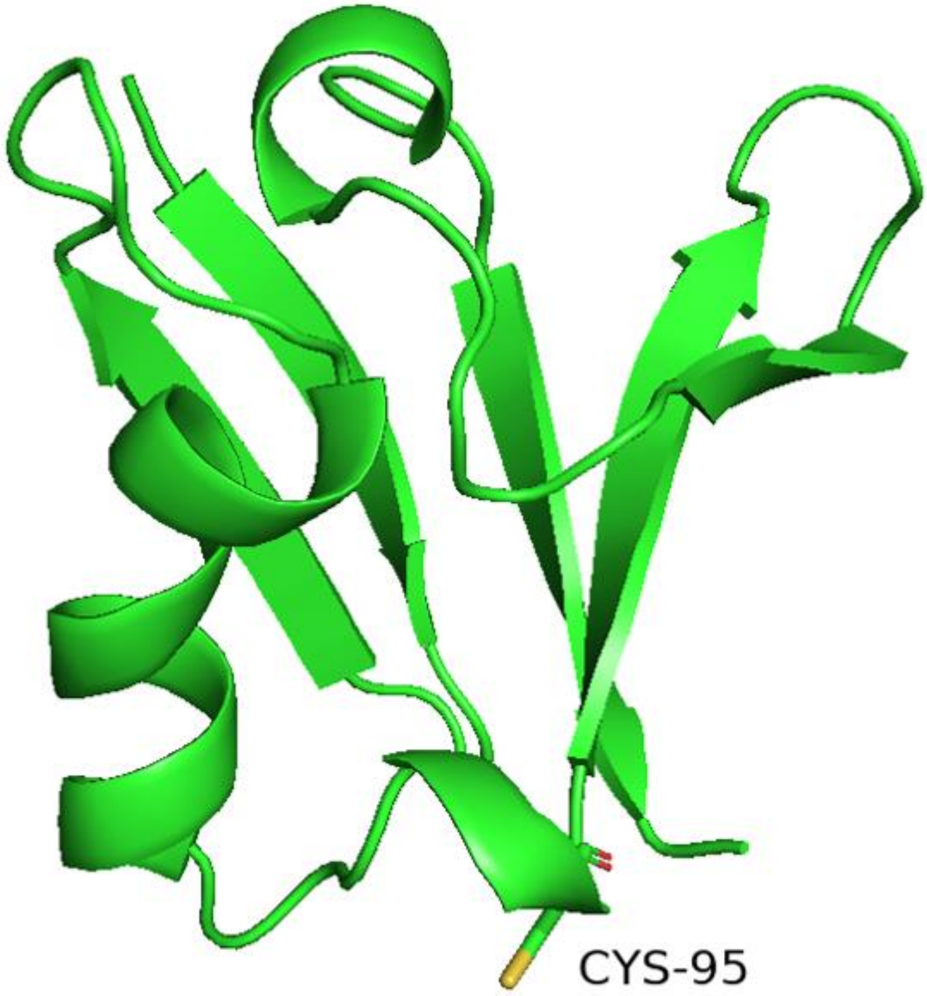


Figure 30: RBD shown as a Richardson diagram, with Cys95 shown as sticks. Note that the thiol group of Cys95 is readily surface-accessible, making it ideal for conjugation.

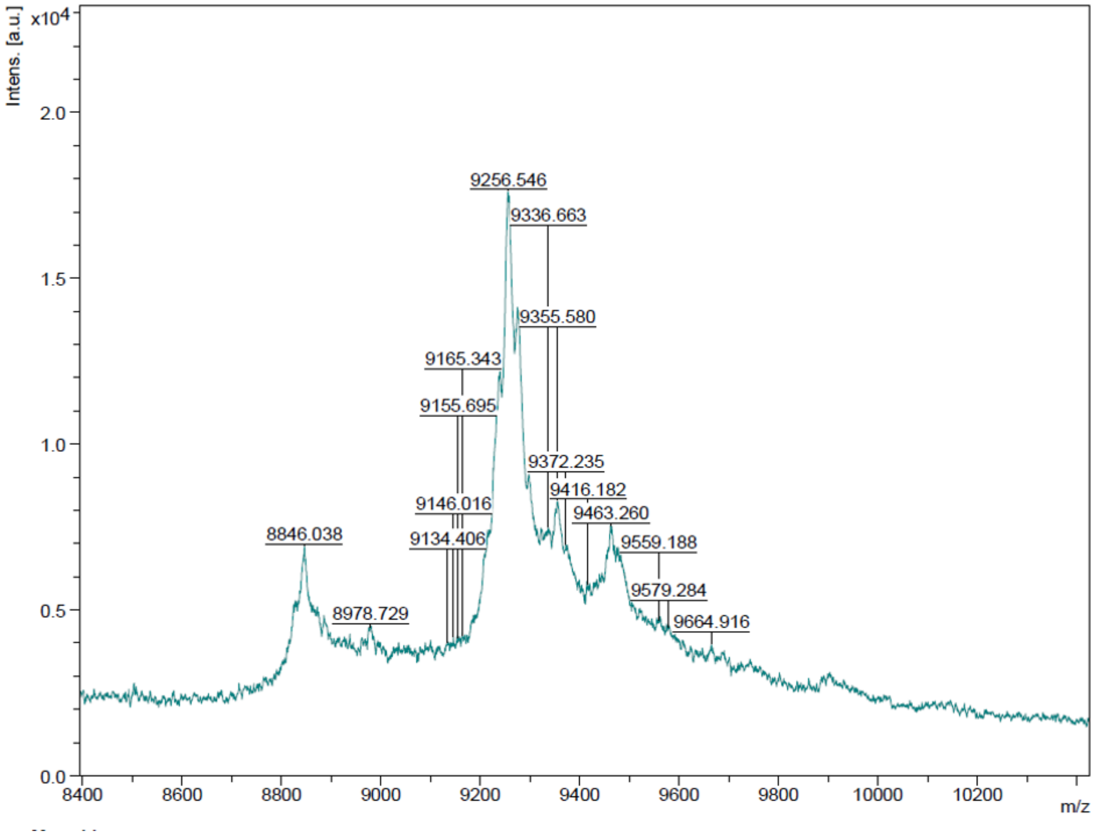


Figure 31: Representative MALDI-TOF mass spectrum of dye conjugation. RBD(WT) has an expected mass of 8840Da and an experimentally observed mass of 8846Da. Maleimide BODIPY has an expected mass of 414Da; a large peak is observed at 9256.545Da, which represents a shift of 410Da. Prior to conjugation this peak is not observed (Figure 14), indicating successful conjugation.

Fluorescence polarization binding assay

In a black-bottom 96 well plate, 30nM of dye conjugated RBD variant was added to all wells and a serial dilution of $KRas^{GppNHp}$ is performed so wells contain a final volume of 200μL. 30nM dye-RBD was chosen as it represented the lowest concentrations that was found to provide consistent readings above background. Experiments were performed in 30 mM Phosphate pH7.4, 327 mM NaCl, 2.7 mM KCl, 5 mM MgCl2, 1.5 mM TCEP, 0.1% BSA (FP buffer). Prepared 96 well plates were covered with aluminum foil and allow to incubate for 10min at room temperature. Titrations were recorded with a Spectra Max M5e in fluorescence polarization mode, excitation at 495nm and emission at 515nm with a 510 nm cutoff. Plates were read twice with 100 flashes per read each time. Results were fit assuming a 1:1 binding model.

Fluorescence polarization binding results

Fluorescence polarizations results support the general ranking and trend observed for RBD variants in the BLI experiments. Namely, our designed RKY variant (K_D 62.5nM ± 25.9nM) shows the highest affinity for $KRas^{GppNHp}$ and significant improvement over both RBD(WT) (K_D 485nM ± 268.5nM) and the previous top mutant RBD(RK)(118.3± 23.6nM). However, K_D differ significantly between the FP and BLI assays and in a variant specific fashion, making direct comparisons between assays impossible. Unfortunately, this variance between binding methods is very common for Ras assays [263], and a cross-platform assay that quantitatively recapitulates experimental results remains needed. However, the fact that the general trend was recapitulated helps give us confidence on our BLI results.

We observe over 10x difference in observed affinities measurements for RBD(RKY) between the BLI assay (K_D 3.26nM ± 2.77nM) and the FP assay (K_D 62.5nM ± 25.9nM). Multiple factors may result in this discrepancy. First, the conjugation of the dye may have deleterious effects on the RBD binding to KRas. Our FP competition assay appears to support this hypothesis. A possible mechanism by which the Bodipy dye can cause a decrease in affinity of RBD for KRas is through a change in the conformation of the RBD. This hypothesis can be explored in future studies by methods like nuclear magnetic resonance (NMR) and circular dichroism spectroscopy. Both methods can search for a change in stability or structure of the RBD. Moreover, the BLI can be used to investigate the possibility of the dye contributing towards non-specific binding or RBD aggregation. Third, it is possible that the discrepancies between the assays reflect the assay format. While BLI has a captured KRas immobilized on the tip, reminiscent of the interaction of KRas with the plasma membrane, the FP assay is free in solution. It is hard to determine how to overcome this limitation and test this hypothesis, but one could imagine using plates coated in NiNTA to capture KRas on the plates surface much like the tips capture KRas in the BLI assay. Finally, one other potential explanation for the discrepancies is the variation in the buffer conditions, namely the presence of Tween-20 in the BLI experiment and its absence in the FP experiments due to its propensity to interfere with the assay.

Because in part of assay differences in results when measuring absolute binding constants, a measure of how useful, or suitable, a high-throughput screen (HTS) is called Z-prime factor was suggested by Zhang HJ et al., 1999 [285]. Z-prime factors range from less than 0 to 1.

$$\text{Estimated Z-factor} = 1 - \frac{3(\hat{\sigma}_p + \hat{\sigma}_n)}{|\hat{\mu}_p - \hat{\mu}_n|}$$

Equation 2: Equation to determine Z-factor of a HTS. The equation is equal to one minus the ratio of the standard deviation (σ) times a constant factor of 3 to represent a 99% confidence interval assuming a normal distribution of the positive and negative control

The FP assay has a Z-factor of 0.737 when A85K is considered the positive control and T57K is considered the negative control. Remembering that a Z-factor of 0.5 is equivalent to 12 standard deviations of separation between the two means. This suggests that this assay is capable of distinguishing between super-binders of KRas^{GppNHp} and binders that are significantly worse than WT. This in of itself is useful as it allows to categorize variants as binders or non-binders very quickly. However, when determining if the assay can distinguish between WT and super-binders like A85K, the Z-factor drops to -0.14, so although there is still more than two standard deviations between WT and A85K, this assay falls just short of 3 standard deviations of separation between positive and negative controls when searching for super=binders and therefore is best used as a preliminary screen. One possible reason for the assays inability to select for variants that are better than WT is the variant-specific effects of dye conjugation which appear to affect each variant to a different degree. Another is the inherent variability of the assay, which may be an artifact of the small number of reps for some variants.

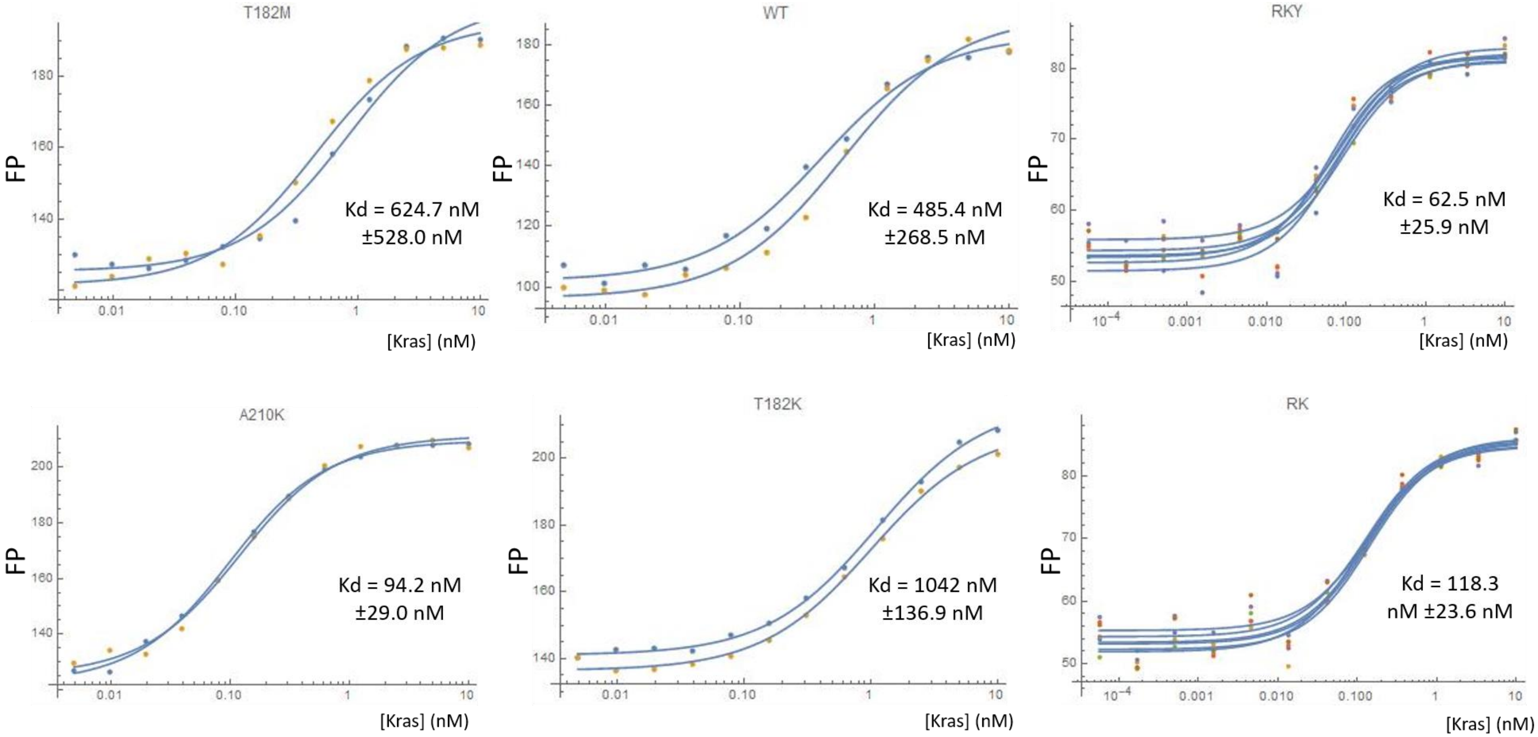


Figure 32: FP titration curves. All experiments runs were fit independently and the average of the fits reported as K_d. Values reported as the mean plus or minus two standard deviations. Curves fit using Mathematica Wolfram Alpha. K_d shown for each variant on the bottom left corner.

Table 3: Calculated dissociation constants for Raf-RBD variants binding to KRas^{GppNHp}. Fits generated using a 1:1 binding model in Wolfram Mathematica. All uncertainties given as two standard deviations. Standard Error (SE).

Run	**Variant**	K_d	***SE***	***P-Value***
1	c-Raf-RBD(RKY)	55.57×10^{-9}	24.1×10^{-9}	0.0468
2	c-Raf-RBD(RKY)	52.61×10^{-9}	20.0×10^{-9}	0.0278
3	c-Raf-RBD(RKY)	73.4×10^{-9}	10.9×10^{-9}	0.0001
4	c-Raf-RBD(RKY)	45.3×10^{-9}	17.9×10^{-9}	0.0326
5	c-Raf-RBD(RKY)	76.6×10^{-9}	37.3×10^{-9}	0.0704
6	c-Raf-RBD(RKY)	71.5×10^{-9}	16.7×10^{-9}	0.0020
Avg	**c-Raf-RBD(RKY)**	$\mathbf{62.5\times 10^{-9}\ (\pm 25.9\times 10^{-9})}$	$\mathbf{21.15\times 10^{-9}\ (\pm 18.05\times 10^{-9})}$	**0.02995 (±0.0537)**
1	c-Raf-RBD(RK)	130.2×10^{-9}	28.6×10^{-9}	0.0014
2	c-Raf-RBD(RK)	126.5×10^{-9}	36.0×10^{-9}	0.0066
3	c-Raf-RBD(RK)	97.5×10^{-9}	25.5×10^{-9}	0.0040
4	c-Raf-RBD(RK)	118.9×10^{-9}	40.6×10^{-9}	0.0168
5	c-Raf-RBD(RK)	123.4×10^{-9}	42.7×10^{-9}	0.0179
6	c-Raf-RBD(RK)	113.2×10^{-9}	48.8×	0.04553

			10^{-9}	
Avg	**c-Raf-RBD(RK)**	**118.3×10^{-9} ($\pm 23.6 \times 10^{-9}$)**	**37.03×10^{-9} ($\pm 17.62 \times 10^{-9}$)**	**0.0015 (± 0.032)**
1	T57K	1090.2×10^{-9}	69.92×10^{-9}	0.0000
2	T57K	$993.5.2 \times 10^{-9}$	91.72×10^{-9}	0.0000
Avg	T57K	**1042.0×10^{-9} ($\pm 136.9 \times 10^{-9}$)**	**80.8×10^{-9} ($\pm 30.8 \times 10^{-9}$)**	**0.0000 (±0)**
1	A85K	83.9×10^{-9}	5.1×10^{-9}	0.0000
2	A85K	104.4×10^{-9}	11.8×10^{-9}	0.0000
Avg	**A85K**	**94.2×10^{-9} ($\pm 29.0 \times 10^{-9}$)**	**8.5×10^{-9} ($\pm 9.5 \times 10^{-9}$)**	**0.0000 (±0)**
1	T57M	811.4×10^{-9}	171.4×10^{-9}	0.0011
2	T57M	438.1×10^{-9}	82.4×10^{-9}	0.0005
Avg	**T57M**	**624.7×10^{-9} ($\pm 528.0 \times 10^{-9}$)**	**126.9×10^{-9} ($\pm 125.87 \times 10^{-9}$)**	**0.0008 (± 0.0008)**
1	Wild-Type	390.5×10^{-9}	$63.4 \times$	0.0002

			10^{-9}	
2	Wild-Type	580.3×10^{-9}	98.1×10^{-9}	0.0002
Avg	**Wild-Type**	**485.4×10^{-9} (± 268.5×10^{-9})**	**80.75×10^{-9} (±49.07×10^{-9})**	**0.0002 (±0.)**

2.2.2.6 Competition assays for the inhibition of the KRas:Raf-RBD protein-protein interaction.

Although binding to KRas is important and was the number one readout for our experiments in designing an RBD super-binder variant, there is also the need to generate quantitative high-throughput assays to test for the disruption of the KRas:Effector PPI. Currently methods for determining KRas:Effector binding or disruption include: Nuclear magnetic resonance (NMR)[286-288], bead–based assay (BBA, Alpha screen) [279, 286], co-precipitation assay (CPA) [289] ,fluorescence kinetics (FK) [270, 290, 291], fluorescence polarization (FP) [108, 260-264], guanine nucleotide dissociation inhibition (GDI) [265-270], isothermal titration calorimetry (ITC) [292-294], scintillation proximity assay (SPA) [278], surface plasmon resonance (SPR) [271-273].

However, despite the plentitude of methods, a high-throughput, easily accessible, cheap, and efficient assay remains elusive. NMR and ITC methods are inherently low throughput and consume massive amounts of proteins. CPA methods are low-throughput in their readout format, requiring western blot analysis. GDI and FP assays (as described in the above papers) requires an expensive fluorescent GTP analogue. Moreover, having KRas be bound to the expensive fluorescent GTP analogue, as is the case in current KRas FP assays [108, 260-264], has one significant drawback for measuring the inhibition of the KRas:Effector PPI. This FP format is only conducive to measuring KRas:Effector disruption if the inhibitor is a small molecule. Larger inhibitors such as peptides and proteins would bind to KRas and displace the effector but demonstrate no change in FP signal, as they themselves would slow down the angular momentum of KRas. Alpha screen is a BBA developed to address this issue of throughput and does so well; however it requires both partners to be immobilized, which can potentially introduce artifacts in the analysis. SPA assays have the potential to be very sensitive and highly scalable but require the use of radiation which complicates handling. Finally, SPR techniques as previously described have mostly immobilized KRas. This is ideal to measure binding to KRas, but the SPR format (normally constrained to 3 experimental flows cells plus a control flow cell) limits throughput if one desires to measure KRas:Effector PPI inhibition. .

Here we describe two proof-of-concept assays that are capable of testing for the disruption of the KRas:Effector PPI. These assay are presented in a format amenable to high-throughput screening and inhibition by both small molecules and biologics.

2.2.2.6.1 SPR competition assay

Method and experimental setup

Experiments were conducted on a Biacore T200 with a SCM5 chip. N-hydroxysuccinimide (NHS) was mixed with equal parts with 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) before being run over the SCM5 chip for 420s. SCM5 chips have a carboxymethylated dextran layer that is covalently attached to a gold surface. Shortly, the attachment of the carboxy group allows the carboxylic acid moiety of carboxymethylated dextran on the chips surface to attack the carbodiimide of EDC, forming an O-acylisourea intermediate with an activated ester, that readily reacts with primary amines like N-terminal residue of proteins and lysine side chains. However, in aqueous solution the activated ester is unstable, due to the fact that EDC is a good leaving group and thus the reaction forming the ester is highly reversible by water. Adding NHS allows for a significantly more stable ester that still reacts with primary amines to form amide bonds. Reactions are terminated with a 420s injection of Ethanolamine. As a control, flow cell one was prepared as described above, having NHS/EDC activation followed by ethanolamine to quench and block the surface. Flow cell two, our experimental cell, was prepared as above, but between the injection of NHS/EDC for activation and ethanolamine for quenching, 4.36mg/ml of His-SumoRBD(WT) in acetate pH 6.0 was injected in 300s increments until 1400RU were immobilized on the chip surface.

His-GST-KRas C118S was purified and loaded with GppNHp as describes previously for His-KRas (0). To establish the affinity of His-GST-KRas for immobilized

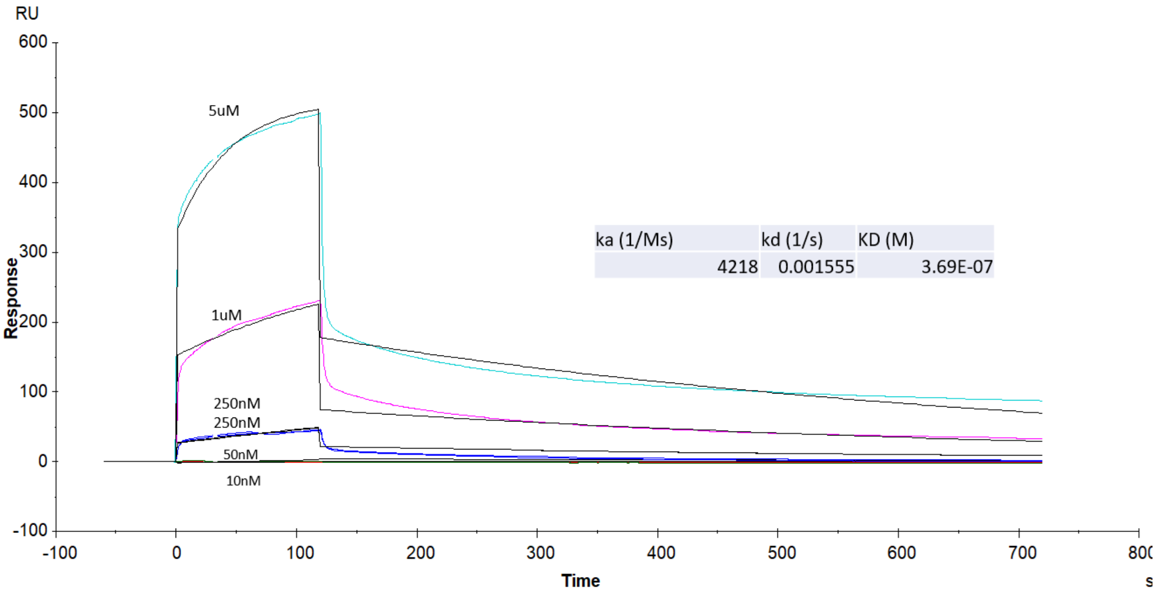
His-SumoRBD(WT), a titration of different concentrations of His-GST- KRas^{GppNHp} in 40 mM phosphate, 250 mM NaCl, 5 mM $MgCl_2$, 1 mM TCEP (SPR buffer) were run over both flow cells one and two at 30μL/min for 100s. Signal from our control flow cell (flow cell one)was subtracted from our experimental flow cell (flow cell two). Flows cells were regenerated between runs by 60s injections of 0.5 mM NaOH(Figure 33).

For the competition assay, 200nM His-GST- KRas^{GppNHp} was run with or without 10min incubation with differing concentrations of His-SumoRBD(WT) ranging from 0nM to 10μM. All samples were in SPR buffer and run at 30μL/min for 100s and subtracted as above. The response at the end of the association phase was recorded. The differing concentrations of His-SumoRBD that were added to KRas were run in a random order e.g., 500nM followed by 10nM followed by 250nM etc. This was done to avoid effects such as natural decrease in binding capacity of the flow cell due to regeneration of the surface (surface burnout). Surface was regenerated with 60s injections of 0.5 mM NaOH (Figure 34). As an additional control His-SumoRBD(WT) was injected at the same concentrations as the ones used to incubate with the KRas samples. Any response observed from the His-SumoRBD injections alone also subtracted from the response curve (Figure 35)

SPR Results

His-SumoRBD(WT) was immobilized in the SPR surface and used to detect binding to KRas. Figure 33 shows the result of a titration experiment with a calculated K_d of 369nM (Figure 33). This $K_{d\ \mathrm{differed}}$ significantly from the result that we obtained from the BLI (Table 2). A few possible reasons are the immobilization conditions as well as the presence of the His-SumoRBD tag. Independent of the reason for the difference in binding affinity between our SPR and BLI results, the immobilization of SumoRBD allowed for the development of a KRas:Effector competition assay that is molecular weight-agnostic on a SPR platform. It is important to remember that the signal in SPR is a result of a shift in the angle of the plasmon that occurs as a result of changes in refractive index between a control and a biolayer surface. These changes are largely dependent on the size of the binding analyte. Therefore, detection of small peptides or small molecules through SPR is challenging. Figure 34, shows that the binding response from flowing 200nM of KRas (analyte) over flow-cells immobilized with His-SumoRBD decreases as the analyte is spiked with increasing concentrations of His-SumoRBD. Figure 35 shows that the data from these experiments can be fit with very high fidelity to a 1:1 orthosteric inhibition model, suggesting that the experimental results and experimental design are valid. Interestingly, the results show that immobilization of His-SumoRBD has no impact on the affinity of the protein to KRas. Furthermore, beyond the ability to detect small molecules and peptides with high accuracy in a functional PPI inhibition assay, one major benefit of this SPR method is the low consumption of materials. Working volumes are consistently less than 100μL, allowing for multiple runs at low material costs.

Unfortunately this method suffers from a technical liability that reduces its efficiency for low-affinity inhibitors of the KRas:Effector interaction. At concentrations greater than 10μM, we observed strong non-specific binding between His-SumoRBD and the control surface. This led us to develop a new method, a FP-based competition assay. This method benefits from an immobilization-free environment and one where the measured signal cannot be a result of either aggregation or non-specific binding.



ka (1/Ms)	kd (1/s)	KD (M)
4218	0.001555	3.69E-07

Figure 33: SPR binding curves (colored curves) and fit (black curves) for KRasGppNHp binding to immobilized His-SumoRBD(WT). KRas^{GppNHp} was run over flow cell 2, with immobilized with His-SumoRBD(WT), and flow cell 1 is the control surface. KRas^{GppNHp} was injected at multiple concentrations ranging from 0 (buffer control) to 5μM. Concentrations for each injection are labeled. Curves shown are the result of double subtraction, for buffer drift and non-specific binding control. Fitting was done on Biacore T200 result analysis software using a 1:1 binding model.

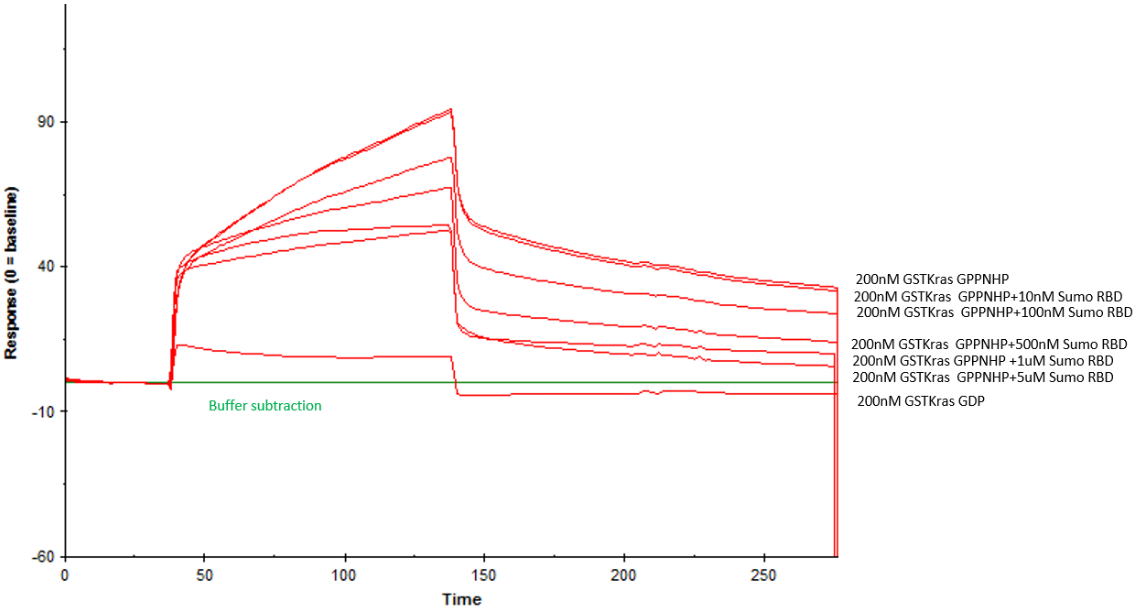


Figure 34: SPR binding curves for 200nM KRas^{GppNHp} injected after incubation with different concentration of His-SumoRBD(WT) for 10 minutes.Curves shown are double subtracted for non-specific binding and buffer drift. Note that 200nM His-GST- KRas^{GDP} is also shown as a control. As an additional control, only His-Sumo RBD was injected, and its response (if any) was subtracted from the final response as well.

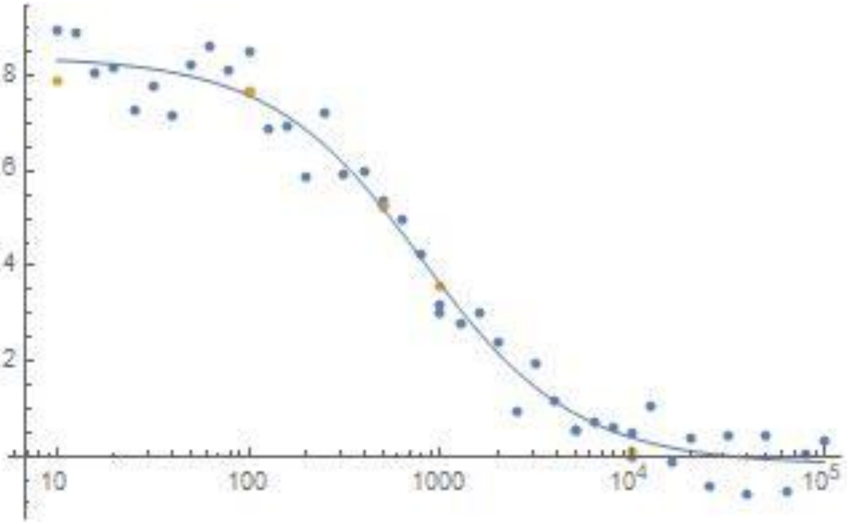


Figure 35: SPR competition assay results. Fraction of $KRas^{GppNHp}$ bound to immobilized His-SumoRBD(WT) as a function of the concentration of non-immobilized His-SumoRBD(WT) added to 200nM $KRas^{GppNHp}$. Results shown for two repeats. Data in yellow, simulation in blue.

2.2.2.6.2 FP competition assay

Method and experimental set-up

The FP assay was made to allow for high-throughput, accessible and inexpensive characterization of the KRas:Raf interaction. Although we could investigate the effect of Raf variant binding directly, we also sought out the possibility of creating a high-throughput assay to test for the inhibition of the KRas:Raf interaction.

Therefore, we also sought to develop an assay that would allow for testing of compounds that could disrupt the KRas:Raf interaction. In this assay, rather than observe an increase of anisotropy during KRas titration (Figure 32), we start from conditions in which KRas and dye-RBD are fully bound and displace the dye-RBD through titration of the inhibitor. A stock solution of 30nM of Dye-RBD is incubated with 1μM of KRas^{GppNHp} for 10min. The KRas:RBD mixture solution is added to wells in a black bottom 96 well plate, allowing for a homogenous signal from every well in the plate. Non-labeled RBD, or other inhibitor, is titrated starting from 50μM down in serial dilutions across the row of the plate.

FP Competition Assay Results

We tested our FP competition assay by attempting to displace KRas:RBD(WT_Bodipy) with RBD(WT) (Figure 36, Table 4). Result show a good fit to mathematical 1:1 competition models validating the assay (Table 4). Interestingly, results also show a possible difference between the calculated affinities for RBD(WT) (K_d = 275.5 nM ±2.58 nM) and RBD(WT_Bodipy) (K_d= 485.4 nM ± 286.5 nM), potentially partially explaining the difference between the calculated dissociation constants observed for our BLI and FP assays.

Our FP competition assay was designed with two objectives: 1. Allow for the detection of biologic inhibitors of the KRas:Effector PPI. 2. Allow for these interactions to be relatively low-affinity. One challenge with the design of inhibitors for undruggable proteins is that the affinities of initial hits are usually very low [272, 286]. This presents a clear difficulty for assay generation, as compounds at high concentrations tend to aggregate or display non-specific binding. This is especially a problem for assays that measure binding, as the effects of real versus non-specific binding become difficult to differentiate at high concentrations and low affinities. Moreover, assays that depend on size changes on immobilized surfaces like SPR and BLI are very susceptible to aggregation and non-specific binding. NMR is the gold standard in detection of these weak interactions. However, NMR is slow, requiring large amounts of samples and specialized equipment. Our FP competition assays allow for an alternative to NMR. The displacement nature of the assay means that aggregation will generate signal in the opposite direction of a true inhibitor, increasing FP signal rather than decreasing it as a function of the concentration of the inhibitor. This allows for a clear differentiation between signal and aggregation. Moreover, non-specific binding would generate the same effect. Only the disruption of the interface would result in a decreasing anisotropy signal. This built-in requirement for functionality allows us to investigate low affinity binding at high concentrations, with high throughput, while mitigating the risks present in other methods.

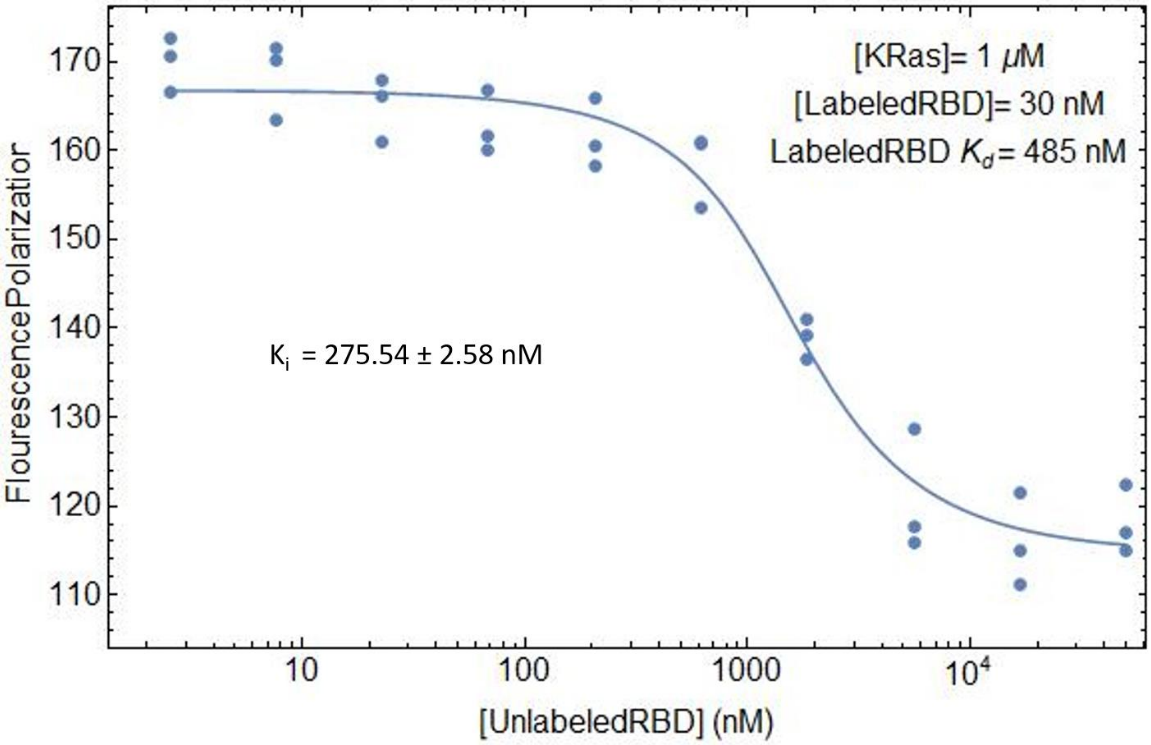


Figure 36: FP competition assay. Results shown with global fit on three repeats of unlabeled RBD(WT) titrations into a pre-incubated mixture of 1μM KRas 30nM Labeled RBD. Results show that unlabeled RBD may have a higher affinity for KRas than labeled RBD. One repeat was dropped due to outliers.

Table 4: Individual fits for FP competition experiment with unlabeled RBD(WT) as an inhibitor of the KRas:LabledRBD interaction. Plus/minus ranges reported for each measure as two standard deviations from the mean value. Standard Error (SE).

Run	Variant	K_d	SE	P-Value
1	c-Raf-RBD(WT)	302.0×10^{-9}	15.6×10^{-9}	0.0000
2	c-Raf-RBD(WT)	306.1×10^{-9}	18.1×10^{-9}	0.0000
3	c-Raf-RBD(WT)	226.4×10^{-9}	15.2×10^{-9}	0.0000
Avg	**c-Raf-RBD(WT)**	$\mathbf{278.16\times 10^{-9}}$ **(**$\pm\mathbf{89.75\times 10^{-9}}$**)**	$\mathbf{16.3\times 10^{-9}}$ **(**$\pm\mathbf{3.1\times 10^{-9}}$**)**	**0.0000 (±0.0)**

2.2.2.7 Discussion

We have successfully designed a KRas superbinder RBD domain. Contrary to previous mutations that enhanced the affinity of Raf-1 RBD to KRas by supercharging the RBD domain [246,253,254], we have identified a novel mutation, V88Y, that significantly enhances affinity, resulting in a single digit nanomolar RBD domain. The observation that an aromatic residue can significantly contribute to the KRas:Raf interface is very significant. Highly polar surfaces like the KRas:RafRBD interface are inherently difficult to drug [7, 183, 286]. Moreover, specificity and affinity are commonly the result of hydrophobic interactions [295]. Therefore, the discovery of a novel aromatic pocket that enhances affinity within the KRas:RafRBD interface allows for a novel opportunity towards the generation of future compounds for the inhibition of KRas. Moreover, we describe the development of multiple assays to measure binding to KRas (through BLI, SPR and FP) and two assays to measure inhibition and disruption of the KRas:RafRBD interface (FP and SPR). These assays will hopefully help accelerate further development towards a KRas inhibitor.

3. Computational Study of DHFR Inhibitors and the Rise and Mechanism of Resistance in DHFR.

In this chapter I describe my contributions towards the continued exploration of a subject that the Donald lab first pioneered in 2010, in seminal work by Frey K. M. et al., 2010 [19]. One of the great achievements of the Donald lab—and one that first got me interested in joining the lab—was the ability of computational protein and drug design (CPDD) to predict the emergence of drug resistance. To do so requires not only accurate predictions of affinity and specificity in the presence of multiple mutations, but also deep biochemical insight about the nature of resistance and its mechanism of action [18]. Here I describe work in collaboration with Siyu Wang, Graham T. Holt and Adegoke A. Ojewole in the Donald lab, as well as multiple collaborators from UConn, most notably Stephanie M. Reeve and Dennis L. Wright. This work has resulted in multiple publications either currently being developed as manuscript for publication or under review.

Section 3.2 is based on the following paper:

S. M. Reeve, D. Si, J. Krucinska, Y. Yan, K. Viswanathan, S. Wang, G.T. Holt, M. S. Frenkel, A. A. Ojewole, A. Estrada, S. S. Agabiti, J. B. Alverson, N. D. Gibson, N. D. Priestly, A. J. Wiemer, B. R. Donald, D. L. Wright (2019). Toward Broad Spectrum DHFR inhibitors Targeting Trimethoprim Resistant Enzymes Identified in Clinical Isolates of Methicillin-Resistant Staphylococcus aureus. BioRxiv (2019) [83]

Section 3.3 is based on the following paper:

S. M. Reeve, S. Wang, G.T. Holt, M. S. Frenkel, A. A. Ojewole, B. R. Donald, D. L. Wright (2019). Development of a Broad Spectrum Antifolate Able to Overcome Acquired Resistance Mutations in the Plasmid Borne Inherently Trimethoprim Resistant DfrG gene in Staphylococcus aureus.

Section 3.4 *is based on the following paper:*

S. Wang, S. M. Reeve, A.A. Ojewole, M. S. Frenkel, G.T. Holt, P Gainza, D. L. Wright, B. R. Donald. Structural and computational study of the etiology and mechanism of enantiomeric antifolates resistance in S.aureus

3.1 Introduction to DHFR

Dihydrofolate reductase or (DHFR) is an enzyme vital for cellular replication, mainly due to its role in the production of thymine. DHFR takes dihydrofolate (DHF) which is the product of thymidylate synthetase and reduces it back to the active tetrahydrofolate (THF) [296]. Due its role in the de novo synthesis of thymine, and its involvement in other 1-carbon metabolism reactions [297], DHFR is present in and vital for cell's including prokaryotes and eukaryotes cells (with some exceptions that use a flavin based enzyme) [298]. As deoxythymidine triphosphate (dTPP) is only used in DNA production and therefore targeting its production would preferentially target quickly replicating cells DHFR was naturally a target for anti-cancer therapeutics [299].

Antimetabolites functioning as orthosteric inhibitors of folate, also known as antifolates were developed to inhibit DHFR's activity. Classical antifolates like methotrexate (MTX) were shown to be effective anti-cancer therapeutics in the 1950s [300] and MTX is still in clinical use today [299].

With the rise of antibiotic resistant bacteria, such as methicillin-resistant *Staphylococcus aureus* (MRSA), the leading cause of healthcare associated infections [301, 302] and skin and soft tissue infections (STTIs) [303], vancomycin became the preferred treatment for MRSA infections [304, 305]. However, since the first strain of MRSA appeared with reduced susceptibility to vancomycin in 1996 [306], the increased frequency of vancomycin resistant strains (VRSA) is placing vancomycin in check, and new approaches are required [304, 305]. Highly specific antifolates, such as trimethoprim (TMP), which has a ~30,000x higher affinity for E.coli DHFR versus human DHFR [307], have shown activity as antibacterial agents, and many MRSA and VRSA clinical isolates are sensitive to TMP [308]. In part due to the structural and sequence differences between human and prokaryotic DHFR, and in part due to the selective inhibition of quickly replicating cells, antifolates can be developed as selective and safe antimicrobial agents. The Donald lab has previously worked to understand the mechanism of resistance in *Staphylococcus aureus* DHFR (SaDHFR) [18, 67, 309]. For my thesis I focused on continuing this work and expanding it to reflect the role of enantiomers and plasmid alleles in the emergence of resistance to antifolates.

3.1.1 Structural characteristics of DHFR

Chromosomal DHFR from *Staphylococcus aureus* (Staph) is a 18kDa protein. An eight-strand beta-sheet forms a central plane through the protein, with six beta-strands being parallel followed by an antiparallel strand and another parallel stand. An additional four alpha helixes connect different beta sheets, resulting in a beta-alpha-beta fold overall. The folate binding site is surrounded on two sides by alpha helices, which form the binding pocket with the beta strand as its core. DHFR uses nicotinamide adenine dinucleotide phosphate (NADPH) as a cofactor, and the NADPH binding site is similarly surrounded by the alpha helices and turns with the beta sheet forming the pocket's inner core (Figure 37). Moreover, Staph commonly carries a second copy of the gene in plasmids [16, 310]. A great deal of focus will be given to these plasmid variants in further sections. Despite similar structure between the plasmid and chromosomal variants of DHFR, many resistance mutations are commonly found in these plasmid DHFR mutants.

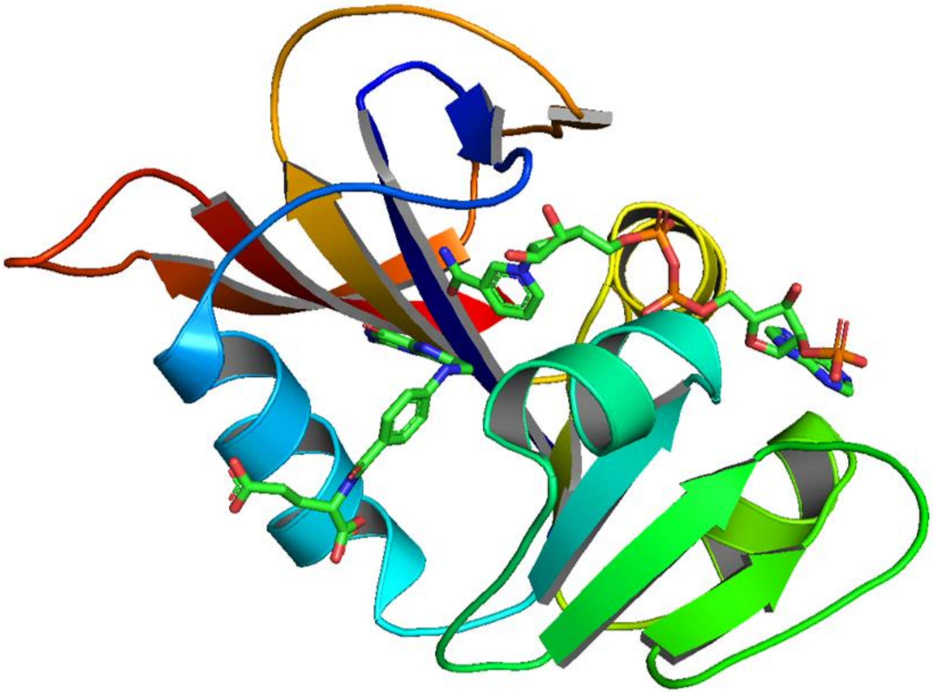


Figure 37 : PDB ID 3FRD, *Staphylococcus aureus* DHFR (SaDHFR) shown as a Richardson diagram. Folate and NADPH are shown as sticks.

3.1.2 Enzymatic Function and Mechanism of DHFR

DHFR binds to NADPH and to dihydrofolate. It catalyzes the reduction of dihydrofolate to tetrahydrofolate by stabilizing the transition state of the hydride transfer from NADPH to DHF and the reduction of the N5 nitrogen in the dihydropterin ring of DHF, forming NADP+ and THF. Much of the work done to study the mechanism of DHFR has been done in *E. coli* and focuses on the role of the Met20 loop, a flexible loop that changes conformation during the enzymatic mechanism. Different models propose different mechanisms for the reduction of DHF into THF [311-313]. The hydrogen that reduces the N5 on the dihydropterin ring is thought to come either through a water molecule or through a keto-enol tautomerization. It is believed that a highly conserved aspartic acid at position 27 ultimately changes the pKa of N5, allowing for its reduction. Moreover, a tyrosine at position 100, which is a commonly acquired resistance mutation to TMP in Staph and is endogenous in E.coli, stabilizes the positive charge in the nicotinamide ring during the hydride transfer between the C4 of the nicotinamide ring and the C6 of the dihydropterin ring [314, 315].

3.1.2 The Need to Go beyond Beta-Lactams

Since the discovery and purification of penicillin and other beta-lactams that act through inhibition of bacterial cell wall formation, antibiotics have drastically changed human life and health [316, 317]. Beta-lactams work by binding to transpeptidases sometime called penicillin binding proteins (PBPs), inhibiting their ability to cross-link peptidoglycans. This results in the formation of spheroplasts, bacterial cells lacking a cell wall, which are susceptible to osmotic lysis. Beta-lactams proved to be both effective in treating bacterial infections and safe for humans, as we lack a cell wall [318]. However, resistance to penicillin and other beta-lactams has become widespread [319], challenging our ability to control bacterial infections and potentially putting millions of lives at risk. Antimetabolites like antifolates emerged as a potential solution to the emergence of antibiotic resistance. Currently a combination of the antifolate trimethoprim (TMP) and sulfamethoxazole, a dihydropteroate synthetase inhibitor, marketed as Bactrim, is one of the top ten oral antibiotics prescribed[320]. However, resistance to antifolates—including TMP, the only approved antifolate for antibiotic use—has emerged [321]. This motivates our work to develop quick and effective technologies to predict resistance, understand the mechanisms of resistance, and help design and select for drugs are resilient to the emergence of resistance.

3.2 DfrG: a Plasmid-Borne Resistant Gene

This sub-chapter describes my contribution to a study presenting inhibitors that can overcome a new plasmid-borne DHFR gene that is inherently resistant to all current antifolates. Starting in the 1990s, two different forms of TMP resistance were detected in clinics: mutations to the previously TMP-sensitive DHFR gene dfrB, and the acquisition of new and innately TMP-resistant DHFR genes called dfrA. Recently two additional plasmid-encoded DHFR resistance genes have been clinically observed: dfrG and dfrK. The presence of a plasmid encoding dfrG is significantly more prevalent in clinical samples, with current studies suggesting that as many as 78% of TMP-resistant MRSA strains were carrying plasmids with the dfrG gene. This is work done in collaboration with Siyu Wang, Graham T. Holt, Adegoke A. Ojewole and Bruce R. Donald in the Donald lab and Stephanie M.Reeve, Si Debjani, Jolanta Krucinska, Yongzhao Yan, Kishore Viswanathan, Alexavier Estrada, Sherry S. Agabiti, Jeremy B. Alverson, Nathan D. Gibson, Nigel D. Priestly, Andrew J. Wiemer, and Dennis L. Wright. My contribution to this work was through participation in method development and in the generation of DfrG structural homology models. In particular, I participated extensively in the creation of Osprey-Designed Sequence Replacement (ODSR). ODSR would then go on to allow us to carry out both this study *and the study described in section 3.2. This work was done in close collaboration with Siyu Wang and Graham T. Holt, who conducted the computational experiments. The homology model of DfrG allowed for the understanding of the structural mechanism of antibiotic resistance for this gene. Sections of this sub-chapter have been adopted from the reference below, verbatim paragraphs are marked with quotation marks, to the best of my knowledge these sections are authored by S.M. Reeve with comments and contribution from all the authors.*

S. M. Reeve, D. Si, J. Krucinska, Y. Yan, K. Viswanathan, S. Wang, G.T. Holt, M. S. Frenkel, A. A. Ojewole, A. Estrada, S. S. Agabiti, J. B. Alverson, N. D. Gibson, N. D. Priestly, A. J. Wiemer, B. R. Donald, D. L. Wright (2019). Toward Broad Spectrum DHFR inhibitors Targeting Trimethoprim Resistant Enzymes Identified in Clinical Isolates of Methicillin-Resistant Staphylococcus aureus. BioRxiv (2019) [83]

3.2.1 Introduction

"The spread of plasmid borne resistance enzymes in clinical *Staphylococcus aureus* isolates is rendering trimethoprim and iclaprim, both inhibitors of dihydrofolate reductase (DHFR), ineffective. Continued exploitation of these targets will require compounds that can broadly inhibit these resistance-confering isoforms. Using a structure-based approach, we have developed a novel class of propargyl linked antifolates or ionized non-classical antifolates (INCAs) (detailed in Figure 47) that capture the molecular interactions that have been exclusive to classical antifolates. These modifications allow for a greatly expanded spectrum of activity across these pathogenic

DHFR isoforms, while maintaining the ability to penetrate the bacterial cell wall. Using biochemical, structural and computational methods, we are able to optimize these inhibitors to the conserved active sites of the endogenous and trimethoprim resistant DHFR enzymes. Here, we report a series of INCA compounds that exhibit low nanomolar enzymatic activity and potent cellular activity with human selectivity against a panel of clinically relevant TMP^R MRSA isolates."

"In order to better understand the molecular interactions between the INCA compounds and DfrG, we constructed a homology model of the DfrG active site, based on the crystal structure of DfrB bound to 1232 (Figure 38) and NADPH. This homology model shows high similarity between the DfrB and DfrG active sites binding to 1232, Figure 38. The DfrG structure maintains the seven hydrogen bonding interactions including the Asp 27 side chain (both 3.0Å), Ile5 backbone (3.1Å) and Phe92 (3.2Å) with the diaminopyrimidine and two hydrogen bonding interactions between the Arg57 and C-ring substituted phenyl acetic acid (2.9Å and 2.7Å, Figure 38). The hydrophobic interactions with Val31, Ile50 and Phe92 are also maintained in these structures. Interestingly, DfrG contains a tryptophan in place of DfrB's Leu28. This substitution increases the distance between the inhibitor and Leu28 (in DfrB) and Trp28 (in DfrG) from 3.6Å to 6.4Å, widening the distal region of the active site and effectively reducing the hydrophobic interactions with 1232. "

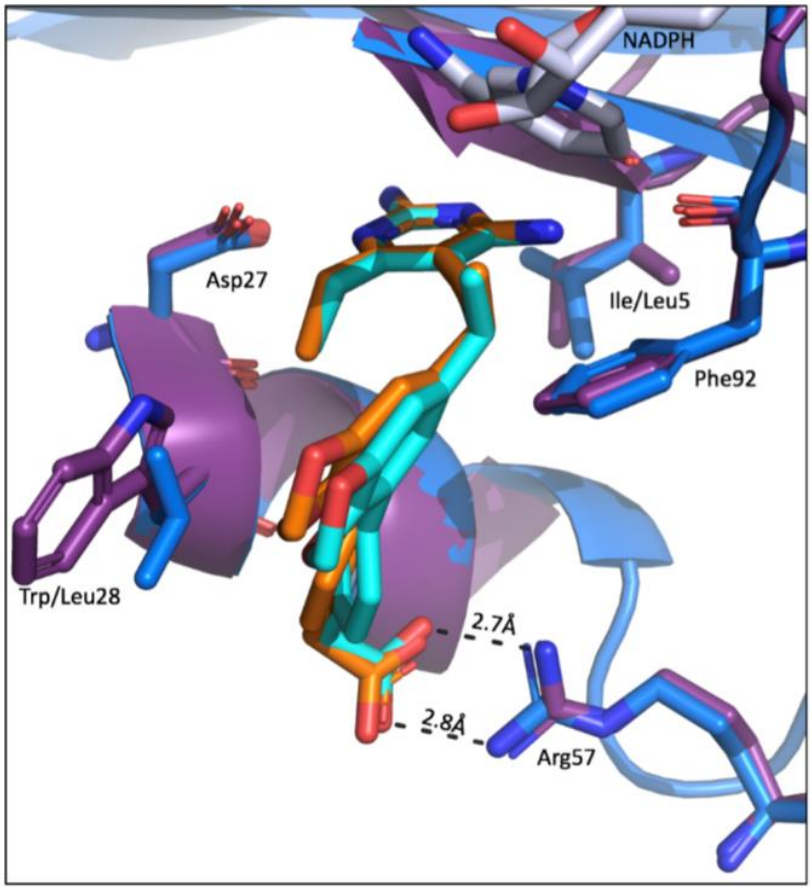


Figure 38: Overlay of DfrB (dark blue) with 1232 (orange) and a DfrG homology model (purple) with 1232(light blue). Active site residues shown as sticks.

3.2.2 Homology Modeling

"Homology modeling of the DfrG active site was accomplished via the study of extant DHFR crystal structures in complex with various ligands. In this case, the DfrB:NADPH:1232 crystal structure was selected as the input starting structure for the homology modeling of DfrG active sites. Next, an intermediate model was generated using a structure prediction calculation, termed "OSPREY-designed sequence replacement" (ODSR). This process involves mutation to the target sequence implemented by side chain replacement. Here, all residues within 8Å of 1232 were selected and mutated to the appropriate DfrG amino acid, determined by sequence alignment to the sequence of DfrG. Sequence alignment was performed using CLUSTAL X 2.1 software [322]. Subsequently, side-chain replacement and global minimum energy conformation (GMEC) calculation were performed using OSPREY [68, 220]. Following ODSR, the intermediate model was all-atom minimized using the SANDER package from the AMBER biomolecular simulation package [257]. Minimization was allowed to proceed for 1,000 steps, resulting in a fully-minimized homology model for DfrG active sites in complex with 1232 and NADPH. "

3.2.3 Rationale for Homology Modeling Using OSPREY-Designed Sequence Replacement (ODSR)

The rationale behind this method assumes that DfrG has a nearly identical fold to DfrB when they are bound to the same ligand. Due to the highly conserved function of the DHFR enzyme, it was hypothesized that changes to the active site and its surrounding would be small in order to not disrupt catalytic function, and that the DHFR fold would be conserved in spite of large sequence diversity. This rationale is supported by multiple crystal structures of DHFR genes with large sequence diversity but very similar folds. For example, DHFR from *Bacillus anthracis* (PDB ID: 3JW5, BaDHFR) and *Staphylococcus aureus* (PDB ID: 2W9G SaDHFR) share only 44% sequence identity as determined by a NCBI Blast search but their structure (both bound to NADPH and trimethoprim (TMP)), have an RMSD of only 0.58 Å as calculated using

PyMol ([256] Schrodinger INC.) (Figure 39). We also noted that the largest deviation in fold were located distal to the active site. Therefore, generated models were kept to 8Å from the active site in order to avoid areas of low homology. Because we focused only on residues near the active site the divergence between the structures was expected to be mostly side-chain placement and packing. We hypothesized that in the most part local minimization would be sufficient to find realistic backbone and side chain placements. Energetically this would mean that the structures would be near their minima. OSPREY can minimize over side chains using state-of-the-art continuous flexibility during designs [68, 78, 80, 223] and therefore would be capable of finding the minima in-line without futher work. Backbone movements, were expected to be small and minimizers like Sander from AmberTools [257] or YASARA [323] can efficiently find local minimum within an energy well for backbones once they are near smooth low energy conformations [324]. DfrG shares 42.1% identity with SaDHFR,. Therefore, we expected similar backbone variation as that found between SaDHFR and BaDHFR. For families of proteins with little structural divergence despite large sequence diversity, this method presents a high-resolution, way to generate homology models that are highly amenable to design calculations (see Section 3.3)

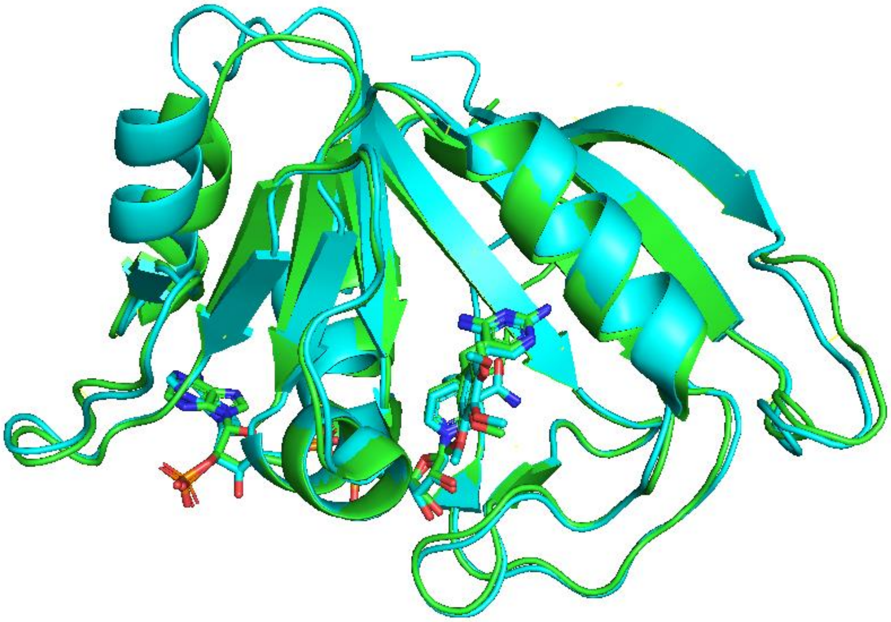


Figure 39: Alignment for structural comparison of DHFR of and *Staphylococcus* aureus and *Bacillus anthracis*. DHFR bound to NADPH and TMP from *Bacillus anthracis* (PDB ID: 3JW5, resolution 2.89Å, shown in Cyan) and Staphylococcus aureus (PDB ID: 2W9G, resolution 1.95Å, shown in green). The proteins have 44% sequence identity but only 0.58 Å RMSD. Note that most deviation occurs in the alpha helix and loop in the top right and is distal from the active site (shown front and center)

3.3 The V31L Resistance Mutation in SaDHFR

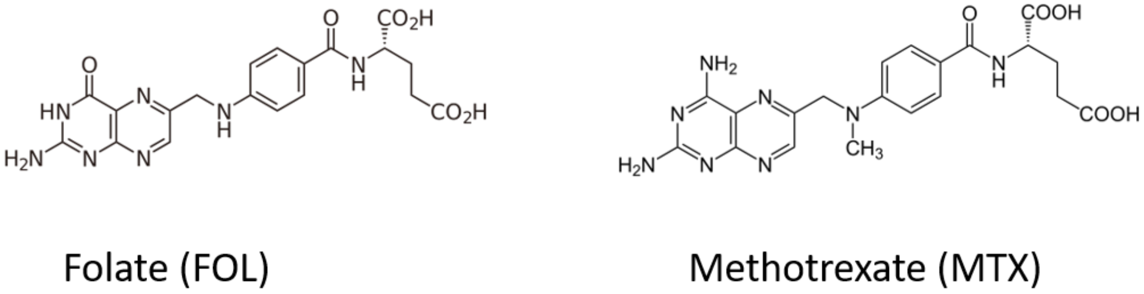
Our work in section 3.1 was the basis for a further study focused on understanding and overcoming the V31L mutation in DfrG. The V31L resistance mutation in SaDHFR was first identified in the Donald lab co-authored by Pablo Gainza and Stephanie M. Reeve [20]. The study was the first to show that an in silico *method could accurately predict what mutation would emerge to generate resistance against a new class of antibiotic drugs. Here, in close collaboration with Siyu Wang, Graham T. Holt, Adegoke A. Ojewole and Bruce R. Donald from* *the Donald lab and Stephanie M. Reeve and Dennis L. Wright from the Wright lab at the University of Connecticut, we set out to show how novel antifolates can be designed to overcome the V31L resistance mutation. This study is also the first to show that OSPREY has the capability to design accurately across multiple starting structures and over different small molecule scaffolds. My contribution focused on method development, structural interpretation, and help setting-up design parameters. Designs were run by Siyu Wang and Graham T. Holt. Experiments were conducted by Stephanie M. Reeve. The results from this work are currently being written into a manuscript of to-be-determined title. This section is adopted from initial reports towards that manuscript. Table 5 is unpublished experimental work from Dennis Wright's lab at the University of Connecticut. Tables 6-8 and figures 36-42 were work of Siyu Wang, Graham T. Holt and myself. Siyu and Graham also contributed to some of the original writing however, I have expanded the original work in order to present the necessary background and elucidate on the results and conclusions.*

3.3.1 Introduction

Previously, the Donald lab used OSPREY [68, 220] to predict native *Staphylococcus aureus* dihydrofolate reductase (DfrB) resistance mutations to several antifolates, including trimethoprim (TMP) and methotrexate (MTX) [18, 19]. A resistant double mutant of DfrB (V31L, F98Y) exhibited reduced affinity for antifolates while retaining enzymatic activity and binding to folate (FOL). Such a mutation that retains function while ablating drug binding is known as an of on-target drug resistance mutation (Table 5). Notably, the V31L mutation causes a 90-fold loss of affinity for MTX, while maintaining affinity for FOL, despite the substrate and the inhibitors having nearly identical chemical structures (Figure 40). However, despite the challenge presented by the V31L resistance mutation, Stephanie M. Reeve and colleagues developed novel propargyl linked inhibitors or ionized non-classical antifolates (INCA) inhibitors of DHFR that were resilient to the V31L resistance mutation [16, 325]. Surprisingly this suggests that transition-state mimics like MTX are not necessarily the most resilient drugs to resistance mutations. This is unexpected [63, 64, 66]. Unfortunately, there is currently no crystal structure for DfrG to help us understand the mechanism of the V31L resistance mutation nor the method by which propargyl linkers are resilient to the V31L mutation. Therefore, in this study, we generated homology models for DfrG bound to multiple inhibitors as well as folate and used OSPREY to understand the mechanism of resistance of the V31L mutation and how novel propargyl linked antifolates are resilient to the V31L resistance mutation.

Table 5: Enzymatic characterization of TMP^R enzymes (Data from the Wright Lab, University of Connecticut).

	K_M (uM)		K_{cat}/K_M	Enzyme Inhibitory Concentrations, IC_{50} (uM)				
	DHF	NADPH		TMP	Iclaprim	MTX	UCP1039	UCP1232
DfrG	17.3	3.33	1.04	380 + 12	17 + 1	0.023 + 0.002	0.45 + 0.02	0.056 + 0.01
DfrG(V31L)	13.87	TBD	2.56	>1,000	>500	2.1 +0.1	7.9 + 0.6	1.8 + 0.2
DfrB (V31L)	42.9	15.6	1.60	0.18 +	0.016 +0.001	0.014 + 0.001	0.062 + 0.02	0.020 + 0.004
DfrB (F98Y, V31L)	4.1	22.3	10.9	13.2 + 0.2	0.187 + 0.04	0.022 + 0.002	0.777 + 0.06	0.236 + 0.009
DfrB (WT)	17.5	31.21	6.1	0.023 +	0.016 +0.002	0.006 + 0.0007	0.014 + 0.001	0.166 + 0.01



Folate (FOL)

Methotrexate (MTX)

Figure 40: Chemical structure of Folate (FOL) and Methotrexate (MTX). Note the nearly identical chemical structures.

3.3.2 Methods

3.3.2.1 Input Structures

We used crystal structures of dihydrofolate reductase from both *Staphylococcus aureus* and *Bacillus anthracis* as starting points for homology modeling Table 6. Although *Bacillus* DHFR (BaDHFR) is predicted by the XtalPred structure determination software [326] to be a closer structural homologue of DfrG due to a shared 3 residue insertion after residue 87 of *Staphylococcus* DHFR (SaDHFR), fewer structures of BaDHFR have been deposited into the PDB resulting in fewer quality starting points for our ODSR homology model generation method. Given that ODSR assumes the backbone for the novel homology model can be obtained by minimization of the starting structure with sidechain replacement, proximity of the two structures is key to success.

Table 6: PDB ID of crystal structures used to generate homology models for DfrG. If structure is not deposited, the lab of origin is indicated on the table.

Ligand	*B. anthracis* PDB ID (resolution)	*S. aureus* PDB ID (resolution)
FOL		3frd (2.1Å)
MTX	3DAT (2.3Å)	Wright lab
TMP	3JW5 (2.89Å)	2W9G (1.95Å)
UCP1039		Wright lab
UCP1232		Wright lab

3.3.2.2 Homology Model Generation

For each structure we first performed a simple structure prediction calculation called (ODSR) [325]. ODSR and its rationale is described in further detail in section 3.2.3. In short, hydrogens were added to starting structures using Reduce [327], and then visually inspected as sometimes the protonation states of small molecules failed to be accurately reflected. Once correct protonation was ascertained, a mutation is made to each position that differs between the template structure and the sequence of the protein that we are attempting to generate a homology model for, i.e., the target sequence. These mutations are modeled with OSPREY using rigid or continuous rotamer side chain placement. In this case, we selected all residues within 8 Å of FOL, TMP, MTX, UCP1039, or UCP1232 for each DfrB structure, and mutated these residues to the appropriate DfrG amino acid, determined by sequence alignment of each input sequence to the sequence of DfrG. We performed rigid rotamer side-chain placement for each model and generated the global minimum energy conformation (GMEC). Following

ODSR, we performed all-atom minimization using the sander package from AmberTools [257] or YASARA [323]. Using Sander, the minimizer in AmberTools, we allowed minimization to proceed for either 1, 000, 2, 000, or 4, 000 steps, whereas in YASARA we used stock settings. During minimization with Sander, we observed multiple artifacts, such as the puckering of aromatic rings and the non-planarity of guanidinium groups. Restrictions were introduced to ascertain that bond angles remained near ideality. In part this was done because OSPREY does not minimize over bond angles. The resulting models were our minimized homology models on which we performed our affinity predictions. All of our binding affinity and conformation prediction for different inhibitors presented in the following sections are based on homology models minimized with Sander. YASARA is introduced here only for refining the DfrG:folate model. This was necessary because the folate:DfrB crystal structure, 3frd, has folate in a position that presents unfavorable clashes with DHFR residues, making it challenging to construct a folate:DfrG homology model without additional minimization using YASARA.

3.3.2.3 Affinity Predictions

We used the $K*$ algorithm [71] to calculate $K*$approximations of binding affinity (K_a). $K*$scores were computed by approximating partition function ratios over molecular ensembles for drug bound and unbound states for up to 20,000 or more conformations.

Flexible residues were selected using a 4 Å shell around the ligand and are listed for each design in

Table 7. NADPH was held rigid for computational tractability, as modeling of flexibility in NADPH would require adding flexibility to a number of residues that contact NADPH, but not FOL or inhibitors. Ligand covalent bonds were not allowed to rotate, but whole ligand translation and rotation was permitted. For FOL models, we excluded NADPH from our calculations as an approximation of allowing closer interactions between NADPH and FOL than standard van der Waal forces would typically permit. We believe this is important as some of these interactions would require a quantum physics description, which is beyond the capabilities of our current AMBER force field [257, 328].

Table 7: Flexible and mutable residue positions for selected homology models. Note that input PDB file numbering varies. For clarity, in the text all position references will follow DfrB (PDB id:3frd) numbering.

Model	Mutable Residue #	Flexible Residue(s) #
FOL Sa	35	5, 20, 27, 28, 32, 46, 57, 92
TMP Ba	35	9, 23, 24, 31, 32, 53, 54, 99
MTX Ba	32	6, 21, 28, 29, 33, 50, 51, 53, 96
UCP1039 Sa	31	5, 20, 27, 28, 46, 49, 50, 92, 111
UCP1232 Sa	31	20, 27, 28, 32, 46, 49, 50, 54, 92

3.3.3 Results

Across multiple models and chemical scaffolds, OSPREY accurately captures the general trends in inhibitor binding to DfrG and is able to accurately predict the effect of

V31L on the binding of the inhibitors. This suggests that our homology models are informative. Experimental characterization of DfrG ranks MTX as the highest affinity inhibitor, followed in order of decreasing affinity by UCP1232, UCP1039, and TMP (Table 5). K* scores recapitulate this experimental affinity ordering of inhibitors. Furthermore, for MTX, UCP1039, and UCP1232, OSPREY predicts the relative sensitivity of the drugs to the V31L mutation, with MTX as the most sensitive, followed by UCP1232, and then by UCP1039 (Figure 41). We did notice that K* scores were rather minimization-insensitive, with an average K* deviation due to minimization of 0.314 (note that all K* scores are reported as Log10), suggesting that minimization was indeed converging at an energy minimum. A standard deviation of the different minimizations is shown as error bars on Figure 41. However, despite the sub-angstrom RMSD between the DHFR structures for S. aureus and B. anthracis (Figure 39), the K* scores were sensitive to starting backbone structure. The difference in average K* score difference for the same model generated from different backbones was 3.433, almost exactly an order of magnitude larger than for different minimizations. This supports the idea that different backbones were being minimized to different energy wells. For MTX this discrepancy is not important, as both homology models agree when the V31L mutation is investigated. The average effect of the V31L mutation on MTX binding is a difference of 0.6 between K* scores, which is near the minimization error. However, when we investigate the effect of V31L on TMP binding, the difference in starting backbone structure makes a massive difference. Models generated from B. anthracis demonstrate the experimentally observed drop in affinity, while models generated from S. aureus fail to show any significant change, with values fluctuating within minimization error (Figure 41). One possible reason for this may be the weak binding of TMP to either model. It is possible that OSPREY is finding it difficult to differentiate between weak binders. Another possibility is that the homology model generated from the S. aureus backbone for TMP adopts a backbone structure with dihedrals that should be penalized more strictly and would erroneously allow for TMP to escape the V31L mutation. Dihedral penalties are a current source of error in energy force fields and a potential source of error [328]. Both of these hypotheses are currently under investigation in follow-up work aimed at overcoming both of these shortcomings.

Table 8: Tabulated log10 K* scores for homology models of DfrG. Multiple backbones made by altering the number of minimization steps.

Drug (Mutation)	**Model score ($\log_{10}K^*$)**					
Organism	*S. aureus*			*B. anthracis*		
# Min Steps	1k	2k	4k	1k	2k	4k
FOL	YASARA: 36.8					
FOL (V31L)	YASARA: 29.1					
TMP	22.7	22.7	22.6	20.0	21.0	20.9
TMP (V31L)	22.8	22.6	22.2	15.2	16.6	17.2
MTX	35.4	35.7	36.0	38.2	38.7	39.4
MTX (V31L)	31.7	29.9	30.5	32.7	33.8	32.9
UCP1039	29.9	30.0	30.1			
UCP1039 (V31L)	29.6	29.9	29.8			
UCP1232	33.8	34.0	33.9			
UCP1232 (V31L)	31.9	30.3	29.4			

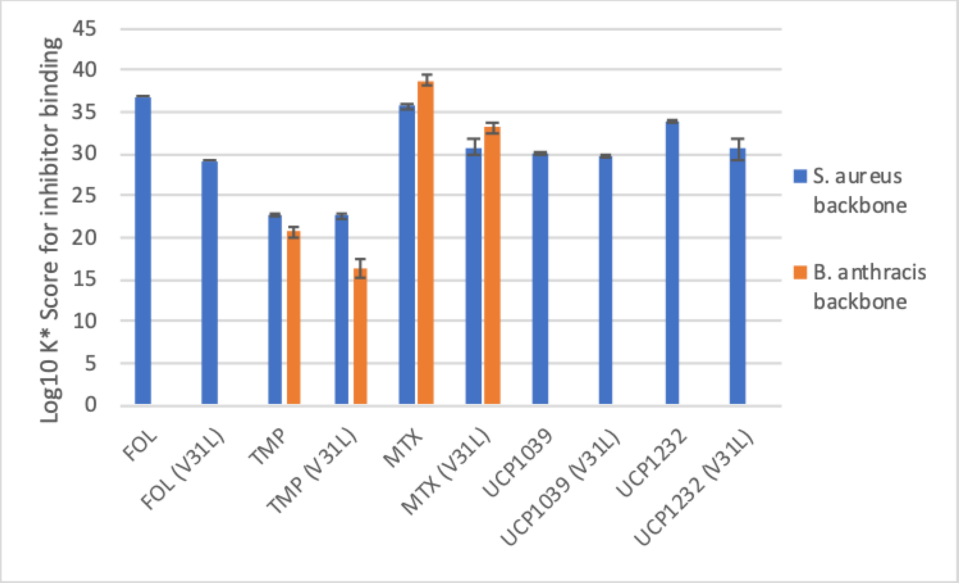


Figure 41: log10 K* scores for DfrG WT and V31L bound to ligands in models generated using 1K all-atom minimization steps. Data shown in Table 6.

We also used the ensembles generated by our OSPREY runs to predict the mechanism or resistance of V31L. The mechanism of resistance of V31L appears to be the steric bulk introduced by the leucine mutation. While valine, the wild type residue at position 31, has an isopropyl group as its R-group, leucine has an isobutyl group as its R-group. The additional methyl group extends the leucine an additional 1.2 angstroms into the active site. This mutation reduces the size of the pocket that natively contains the folate pterin ring, causing steric clashes between the inhibitor, NADPH, and active site amino acids, including L31 and A7. For MTX and TMP, this change in active-site pocket size causes unescapable clashes. For the new propargyl inhibitors, L31 is able to access a rotamer that does not orient toward the pterin ring pocket. Instead, the leucine side chain orients roughly parallel to the propargyl linker, interacting with F92. This generally correlates with less sensitivity to the V31L mutation.

Our models also indicate clashes for FOL bound to the V31L mutant that are similar to the clashes observed in TMP and MTX models. However, we hypothesize that this “steric crowding” may facilitate the hydride shift reaction, increasing the rate of reduction. In other words, we expect the enzyme’s catalytic activity to increase. This is supported by the measured experimental increase in Kcat observed for the DfrG V31L mutant and a slight drop in Km (Table 5).

Below we describe in detail the observations for each ligand binding to DfrG and the effect of the V31L mutation in each one.

3.3.3.1 Effect of V31L mutation on folate

Our models use folate as a stand-in for dihydrofolate (DHF), as folate is the ligand present in the deposited crystal structures with DHFR. Experimental data for DHF binding to DfrG shows that the V31L resistance mutation causes a 2.46 fold increase in kcat /KM with a 1.25 fold improvement in KM relative to WT DfrG (Table 5). Because KM =(k−1 + kcat)/k1, the decrease in KM and the increase in kcat suggest we should expect an decrease in Kd = k−1 / k1. We therefore hypothesize that the V31L mutation moves FOL near to NADPH in a manner that would both stabilize the transition state and improve binding affinity.

Our models partially support this hypothesis as we observe a reduced distance between NADPH and FOL in the GMEC of our V31L FOL model (Figure). Previous research provides evidence that binding of folate to DHFR in the closed conformation places the C(6) in the pteridine of folate and the hydride donor in the nicotinamide ring of NADPH within a distance shorter than their van der Waals radius, with the ideal distance for the hydride transfer expected to be 2.6 Å [329]. There is evidence that the pterin ring undergoes extensive polarization and adopts non-planar geometry during catalysis, allowing the FOL NADPH interaction to be favorable despite being closer than ordinarily clashing distances [330]. In order to allow modeling of this phenomenon in OSPREY, we decided to not model NADPH in our current runs to avoid perceived clashes. This, however, underestimates the favorable interaction between FOL and

NADPH and can partially explain why our results do not show a small improvement in binding of FOL to the V31L mutant. Our current models of the V31L mutant depict a distance between FOL and the donated hydrogen of NADPH of 2.7 Å, which approaches this theoretical ideal (See Figure 3c). We expect this reduced distance to stabilize the transition state and increase k_{cat} .

Furthermore, our models predict a significant decrease in binding affinity of DfrG V31L to FOL, with current log10 K* results decreasing from 36.8 to 29.1 upon the V31L mutation. This large decrease is similar to that seen in predictions using models of MTX binding. However, unlike MTX, which we believe clashes irredeemably with NADPH, FOL would make favorable contacts with NADPH (ignored in our modeling). Moreover, the small clashes with Ala7 and Leu31 of DfrG (See Figure 42), we believe, can be overcome through the introduction of slightly larger flexibility of the FOL pteridine ring and some small backbone movements. These forthcoming small enhancements to the model will potentially help our predictions further match the experimental data.

.

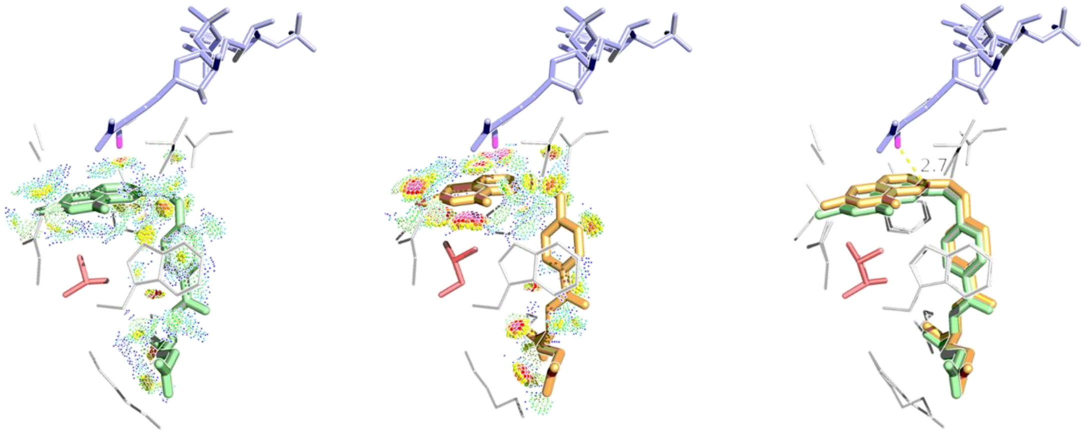


Figure 42: Model of FOL bound to DfrG, *Staphylococcus* backbone. FOL in green (WT) or orange (V31L), NADPH in blue, flexible active site residues along with A6 and A7 in gray, and mutable residues in red. Protein backbone atoms not shown. Probe dots [259] show atomic interactions. Green and blue indicate favorable interactions, while yellow, red, and purple indicate steric overlaps and clashes. The wildtype active site (left) has few steric clashes, while V31L (center) introduces clashes between FOL and NADPH, Ala7. However, this brings reactive atoms into ideal proximity (right).

3.3.3.2 Effect of the V31L mutation on TMP

OSPREY predicts that TMP binds to DfrG with the lowest affinity, returning log10 K* values of 22.7, 22.7, and 22.6 for 1, 000 (1k), 2, 000 (2k), and 4, 000 (4k) minimization steps (see section 3.3.2.2) respectively for the models generated from the Sa backbone. The Ba backbone models returned log10 K* values of 20.0, 21.0, and 20.9, respectively. The prediction of low affinity relative to the other compounds recapitulates experimental data (Table 5).

The Sa models were largely insensitive to the V31L mutation, while all Ba models were sensitive to the mutation, returning log10 K* scores of 15.2, 16.6, and 17.2 with the

V31L mutation Table 8. Based on these results, OSPREY predicts a significant drop in DfrG (V31L) affinity for TMP, a result that agrees with experimental data (Table 5). We chose the homology model generated from the Ba backbone with 1k steps of minimization for further analysis. This decision was made due to the good fit of OSPREY affinity predictions to experimental data. Additionally, we wanted to be conservative with the number of all-atom minimization steps, as minimization software can output non-natural protein backbones.

Structural analysis of the ensemble of low-energy structures generated from this model reveals little heterogeneity, suggestive of a tightly packed active site. The global minimum-energy conformations (GMECs) for both wild-type and mutant DrfG are shown in Figure 43. Introduction of the V31L mutation introduces a large clash between Leu31 and the pyrimidine ring of TMP. Although the magnitude of these clashes can be reduced by translation of the ligand within the active site, this is hampered by the restrictions imposed by NADPH and Ala7, resulting in an energetically unfavorable bound state. From this model, we hypothesize that DfrG(V31L) reduces the space available to the pyrimidine ring of TMP, causing a drop in binding affinity.

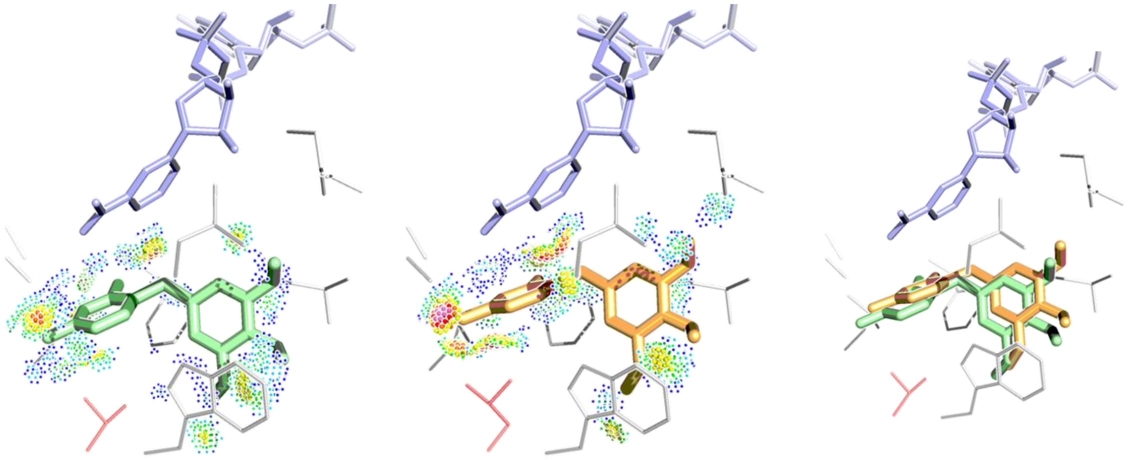


Figure 43: Model of TMP bound to DfrG, *Bacillus* backbone, 1000 minimization steps. Drug in green (WT) or orange (V31L), NADPH in blue, flexible active site residues along with A6 and A7 in gray, and mutable residues in red. Protein backbone atoms not shown. Probe dots [259] show atomic interactions, green and blue indicate favorable interactions, while yellow, red, and purple indicate steric overlaps and clashes. The wild- type active site (left) has few steric clashes, while V31L (middle) introduces clashes between the drug and NADPH, Ala7, and L31 despite translation of the drug to attempt to avoid the clashes (right).

3.3.3.3 Effect of the V31L mutation on MTX

OSPREY predicts that MTX has the highest affinity for DfrG, matching experimental results. *B. anthracis* (Ba) models with 1k, 2k, and 4k minimization steps return log10 K* values of 38.2, 38.7, and 39.4 respectively for wild-type models, and 32.7, 33.8, and 32.9 respectively for V31L models. *S. aureus* (Sa) models with 1k, 2k, and 4k minimization steps return log10 K* values of 35.4, 35.7, and 36.0 respectively for wild-type models, and 31.7, 29.9, and 30.5 respectively for V31L models. OSPREY predicts that MTX is the most sensitive inhibitor to the V31L mutation. Both Sa and Ba generated models show similar interactions – we chose the model generated from Ba with one thousand minimization steps model for structural analysis because of its more favorable predicted energies and a stronger response to the V31L mutation, which better matches the experimental data.

Structural analysis revealed a similar story to TMP (Figure 43). Introduction of L31 crowds the aminopterin ring, introducing steric clashes. Due to the tight packing of the active site and the bulk of MTX, there is little opportunity to translate or rotate the drug in order to relieve clashes. The steric bulk of L31 forces the aminopterin ring toward NADPH, where it clashes with Ala-7. Additionally, this results in a steric overlap with NADPH, and worsens steric overlaps along the entire molecule (Figure 44). Our models suggest that unlike FOL, which can undergo partial bond formation and hydride transfer, and escape the clash, MTX cannot overcome the steric-overlap. Ultimately, OSPREY predicts that the V31L mutation significantly decreases affinity of DfrG for MTX. This is an important lesson for designing drugs to be resilient to resistance mutations. Previous techniques to design drugs to be resilient to resistance mutations like the substrate envelope hypothesis [63-66] have been very effective. However, the V31L mutation would probably have escaped the substrate envelope hypothesis algorithms as MTX and FOL are indeed nearly identical molecules. This suggests that new algorithms that sample conformational energies, as OSPREY does, should be developed to further enhance our ability to design drugs that are resilient to drug resistance mutations.

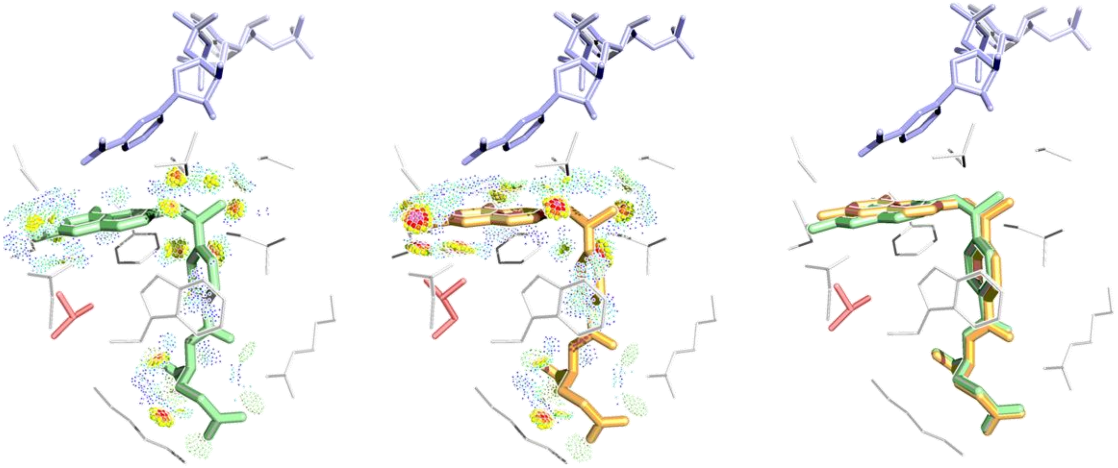


Figure 44: Model of MTX bound to DfrG, *Bacillus* backbone, 1, 000 minimization steps. Drug in green (WT) or orange (V31L), NADPH in blue, flexible active site residues, along with A6 and A7 in gray, and mutable residues in red. Probe dots [259] show atomic interactions, green and blue indicate favorable interactions, while yellow, red, and purple indicate steric overlaps and clashes. The wild-type active site (left) has few steric clashes, while V31L (center) introduces clashes between the drug and NADPH, Ala7, and L31 due to steric crowding. Overlay (right) shows a smaller translation than observed for TMP potentially due to the larger bicyclic ring.

3.3.3.4 Effect of the V31L mutation on UCP1039

OSPREY predicts that UCP1039 [16] binds to DfrG with affinity greater than TMP, but less than that of other inhibitors, recapitulating experimental data (Table 5). Due to the lack of a B. anthracis DHFR structure bound to UCP1039, all models were created using a S. aureus structure. log10 K* scores were 29.9, 30.0, and 30.1 for the wild-type model, and 29.6, 29.9, and 29.8 for the V31L model with 1k, 2k, and 4k all-atom minimization steps, respectively. The average predicted affinity difference between wild-type and V31L for UCP1039 is only 0.23 on average, which is a small difference relative to the other inhibitors, suggesting that UCP1039 is the least sensitive of the inhibitors to the V31L mutation, which is supported experimentally (Table 5). We chose the 1k minimization model for further analysis in order to avoid running minimization with a large number of steps, which can lead to non-natural backbones and uninformative structures.

Structural analysis reveals a possible mechanism for this inhibitor's insensitivity to the V31L mutation. Although ensemble analysis reveals the same, tightly packed active site, Leu31 is able to access a rotamer that was too energetically unfavorable to appear in the ensembles of either TMP or MTX. Using nomenclature defined in [230] Leu31 adopts the idealized tp rotamer, with $\chi_1 = 177^\circ$ and $\chi_2 = 65^\circ$. Instead of adopting a conformation where the C_α is oriented toward NADPH (mt, $\chi_1 = 65^\circ$ and $\chi_2 = 175^\circ$), Leu31 lies roughly perpendicular to UCP1039, oriented toward Phe92 (Figure 45). This rotamer allows the leucine to avoid clashes with the drug, and requires only a slight optimization of the position of Phe92. This causes slight steric overlap between Phe92 and the carbon triple bond of UCP1039, leading to slightly higher energies, but most interactions between the drug and the active site remain essentially unchanged. Notably, the interactions between Leu31 and Phe92 have been important in our previous work [67]. We hypothesize that this conformation is possible due to presence of the propargyl linker, which spaces functional groups apart without requiring a bulky ring structure. For comparison, this rotamer would clash with one of the methoxy groups of TMP, and would clash with Phe92 if MTX were bound.

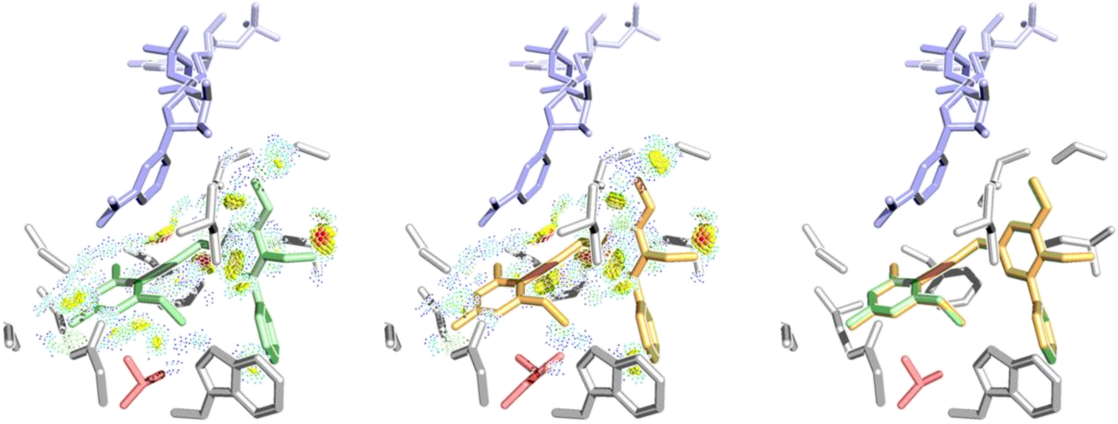


Figure 45: Model of UCP1039 bound to DfrG, *Staphylococcus* backbone, 1, 000 minimization steps. Drug in green (WT) or orange (V31L), NADPH in blue, flexible active site residues along with A6 and A7 in gray, and mutable residues in red. Protein backbone atoms not shown. Probe dots [259] show atomic interactions, green and blue indicate favorable interactions, while yellow, red, and purple indicate steric overlaps and clashes. The wild- type active site (left) has few steric clashes, while V31L (center) introduces a slight steric clash with Phe92, and L31 due to steric crowding (right) we observe no translation but a small rotation of the terminal rings on either side.

3.3.3.5 Effect of the V31L mutation on UCP 1232

OSPREY predicts that UCP1232 [325] binds to DfrG with high affinity, second only to MTX. Like UCP1039, there were no bound *B. anthracis* structures. Therefore, all homology models were based on the Sa backbone. Models generated with 1k, 2k, and 4k steps of all-atom minimization returned log10 K* scores of 33.8, 34.0, and 33.9 for wild-type and 31.9, 30.3, and 29.4 for V31L, respectively. This suggests that while UCP1232 is more sensitive to V31L resistance than UCP1039, it is less sensitive than both MTX and

TMP, which is in good agreement with the experimental data. Again, the 1k model was selected for analysis due to concerns about over-minimization.

Similarly, to UCP1039, L31 adopts a perpendicular conformation, avoiding clashes in the NADPH/Ala7 region and instead leading to a conformational change in Phe92 (Figure 46). However, due to the orientation of the phenylacetic C-ring, the bulk of L31 clashes with the C-ring. This results in a minor "nudge" that attempts to minimize clashes between L31 and the C-ring at the expense of worsening minor clashes along the length of the molecule. This steric crowding is expected to reduce the affinity of DfrG(V31L) for UCP1232. Even so, these interactions are more favorable than the major clashes seen near Ala7 and NADPH for TMP and MTX.

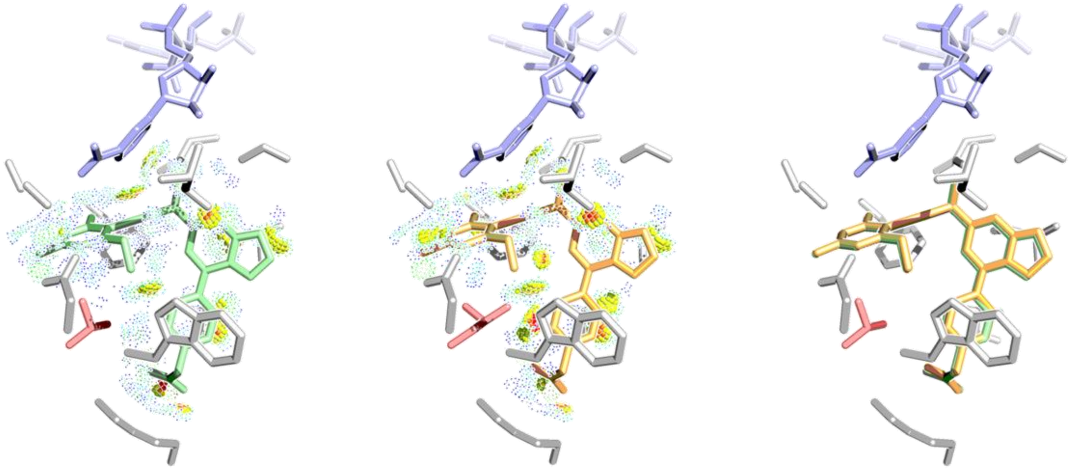


Figure 46: Model of UCP1232 bound to DfrG, *Staphylococcus* backbone, 1, 000 minimization steps. Drug in green (WT) or orange (V31L), NADPH in blue, flexible active site residues along with A6 and A7 in gray, and mutable residues in red. Protein backbone atoms not shown. Probe dots [259] show atomic interactions, green and blue indicate favorable interactions, while yellow, red, and purple indicate steric overlaps and clashes. The wildtype active site (left) has few steric clashes, while V31L (center) introduces clashes between the drug and NADPH, and L31 due to steric crowding. There is very limited drug movement due strong packing, the phenylacetic ring attempts to rotate away from L31 but has limited freedom of motion (right).

3.3.4. Conclusion

The V31L mutation of DHFR was initially observed as the top hit in a search for resistance mutations in a groundbreaking study, that was the first to be able to prospectively predict resistance mutations [20]. The V31L resistance mutation prediction has been proven in the clinic with the finding that DfrG, a plasmid borne DHFR that is spreading in drug resistance bacteria, has the F98Y mutation intrinsically and presents V31L as a top acquired resistance mutation. This combination of F98Y followed by acquisition of V31L is exactly what had been observed in our previous study [20], further validating our predictions. Fortunately, our collaborators have developed a new class of propargyl linked antifolates or INCAs that can overcome this mutation (Table 5). Arguably this would not have happened had the in silico discovery of V31L not occurred. This therefore constitutes a clear example of using in silico techniques to design drugs that are resilient to resistance mutations. The mechanism of V31L resistance and how INCAs can overcome this mutation remained unclear in a DfrG context. We therefore used OSPREY [68, 220] to generate structural models and investigate these phenomena. K* approximations to binding affinity from models of TMP, MTX, UCP1039, and UCP1232 recapitulate experimental data with excellent fidelity, despite originating from homology models. This validates our newly developed ODSR technique for homology modeling of conserved folds like DHFR or immunoglobulin folds. We were able to correctly predict the order of binding affinity (from highest to lowest: MTX, UCP1232, UCP1039, TMP). Additionally, we correctly predict the order of sensitivity to V31L for three of the compounds (from highest to lowest: MTX, UCP1039, UCP1232). TMP is an exception due to divergences between our Sa-originated and Ba-originated models. For the first time we show OSPREY's ability to design across starting structures and across different small molecule scaffolds.

Structural analysis of ensembles generated by the multistate K* algorithm [71] provides insights into the mechanisms of resistance for each inhibitor. For both TMP and MTX, the only reasonably low-energy idealized rotamer predicted to be available to L31 is the **mt** rotamer, which restricts the pterin ring pocket, forcing the pterin ring of MTX and the pyrimidine of TMP to clash with L31, A7, and NADPH. This increases the energy of the enzyme:inhibitor complex, and is predicted to reduce binding affinity by a large amount. For the propargyl-linked antifolates, UCP1039 and UCP1232, the propargyl linker takes up less space, and allows L31 to adopt the tp rotamer, which is predicted to have smaller, albeit noticeable clashes.

Our models of FOL also provide important structural insight, providing a possible explanation for the improvements of catalytic efficiency observed by escape mutants. DfrG V31L is observed to bind and catalyze FOL with comparable K_M and increased k_{cat} relative to wild type. Our current FOL models of the V31L mutant measure a distance between the folate C(6) and the NADPH C(4) hydrogen at 2.7 Å , which approaches the theoretical ideal distance for a hydride transfer reaction and may increase the rate of catalysis. Moreover, our algorithms predict that DfrG V31L clashes with MTX in a way that we do not expect would occur with FOL, due to possible polarization of FOL that is believed necessary for the hydride transfer [329, 330]. These observations are also important in developing novel techniques to design drugs that are resilient to drug resistance, as they show that an energy function-based method may be necessary to supplement existing substrate envelope based techniques [63-66].

3.4 Chiral Evasion and Resistance Resilience in SaDHFR

F98Y is the main resistance mutation to TMP in SaDHFR. However, its mechanism has always been somewhat mysterious, and it was hypothesized that the F98Y resistance mechanism could somehow be linked to a conformational change of NADPH. Pablo Gainza in his graduate work in the Donald lab, was exploring this hypothesis by looking at the electron density of SaDHFR bound inhibitors when he was the first to notice the data suggested a configuration change rather than a conformational change in NADPH. Rather than being in the common β-NADPH configuration, NADPH sometimes was found in an enantiomeric α-NADPH form. Recently further work has now revised this hypothesis and suggests that NADPH is in a stable tricyclic form here thereafter referred to as t-NADPH. We set out to investigate using OSPREY how the enantiomeric forms of NADPH contribute to resistance and how this may explain the divergence in results from non-racemic PLAs or INCAs. This work was done in close collaboration with Siyu Wang, Graham T. Holt, Adegoke A. Ojewole, Pablo Gainza and Bruce R. Donald from the Donald lab and Stephanie M. Reeve and Dennis L. Wright from the Wright lab at University of Connecticut. My contribution to this work was through providing significant contributions to hypothesis generation, experimental design, data analysis, and manuscript preparation (e.g. writing large sections of the introduction and discussion). Sections of this sub-chapter have been adopted from a to-be-submitted manuscript titled below, verbatim paragraphs are marked with quotation and are work authored by Siyu Wang, figures 43-49 and tables 9-10 *are from the publication the publication was made with comments and contributions from all the authors:*

S. Wang, S. M. Reeve, A.A. Ojewole, M. S. Frenkel, G.T. Holt, P Gainza, D. L. Wright, B. R. Donald. Structural and computational study of the etiology and

mechanism of enantiomeric antifolates resistance in S.aureus

Recently the Wright lab has developed a series of propargyl-linked antifolates (PLAs) that can be used as MRSA DHFR inhibitors [331]. The key feature of a PLA is its eponymous propargyl linker (Figure 47). PLAs are antifolates, competitive inhibitors that displace dihydrofolate (DHF) to bind in the active site of DHFR. The narrow triple bond structure of the propargyl linker makes it possible for a PLA to reach deep within the binding pocket so that its tri-substituted pyrimidine occupies the position of folate's pterin ring while its biaryl ring system mimics the glutamate moiety of folate. In contrast to traditional antifolates such as TMP, PLAs show significant high potency when targeting both the wild type and predominant TMP-resistant strain, according to the IC50 measurement [17, 331].

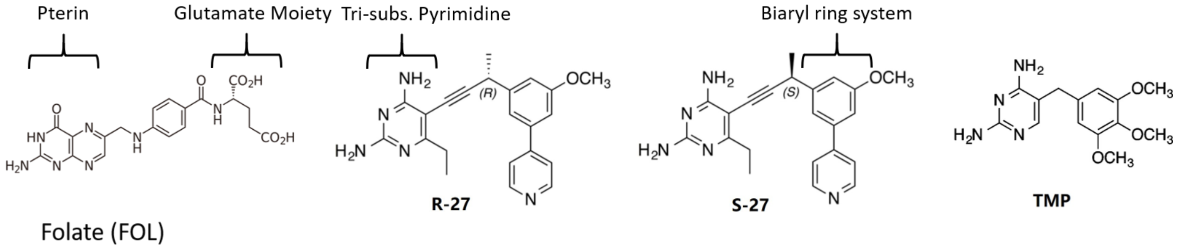


Figure 47: Structure of Folate and Antifolate Inhibitors. Non-racemic PLA (INCA) enantiomers R-27 and S-27 are shown as is the traditional antifolate TMP

In addition to IC50 measurements, structure determination also played an important role in the study of resistance and development of PLAs. A few crystal structures of various PLAs bound to NADPH and SaDHFR (including both WT and F98Y) were determined in [331]. Interestingly, among all of crystal structures we determined in [331] , there are four of them (PDB id: 3fqf, 3fqo, 3fqv and 3fqz) whose NADPH showed two different conformations. Interestingly these conformations appeared resistant to in silico modeling: despite extensive attempts it appeared that we were unable to recapitulate the observed conformational change in NADPH transforming one conformation into another seemed impossible [20]. This led the Pablo Gainza and the Donald lab to question revisit the initial hypothesis that we were observing NADPH in an alternative conformation. We thus carefully refitted the density map of 3fqf, 3fqo, 3fqv and 3fqz and performed extensive study to analyze the geometric features of the alternative NADPH conformation. Over 4000 NADPH conformations from various species among 1700 deposited PDB files were compared. After careful analysis we showed that rather than an alternative conformation of NADPH, these crystals actually had an alternative configuration of NADPH with different chirality from the standard NADPH [332-335]. In addition to the standard β form NADPH, which is most commonly observed, there is a second alternative configuration, α-NADPH. Both alpha and beta NADPH are found in cells in vivo, with the alpha form being present at roughly 1.5% the concentration of the beta form [334, 336].

Stereogenicity of molecules is an important factor to be considered in many biochemical reactions, particularly in enzyme catalyzing reactions since enzyme-substrate system usually requires strict chiral matching [337]. In order to study how the chirality of PLA will influence inhibitor selectivity and DHFR stereospecificity, a series of PLA enantiomers (with chiral center at the propargyl center) were synthesized and studied in [17]. In general, most PLA enantiomer pairs can inhibit both WT and F98Y (a TMP-resistant point mutation) mutant SaDHFR activity at low nanomolar concentrations, as measured by IC50 assays.

Interestingly, many enantiomer pairs showed quite similar affinity for DHFR but different susceptibility to resistant mutations in SaDHFR. For example, PLA enantiomer pair R-27 and S-27 have nearly identical IC50 values for WT SaDHFR, at 15 and 18 nM respectively. However, upon F98Y mutation, the R enantiomer suffers a 34-fold increase in IC50 (510 nM), whereas the S enantiomer suffers only a 6-fold increase in IC50 (111 nM) (Table 9).

Table 9: IC50 data of R-27, S-27 and TMP. Adapted from Table 2 in [17].

Inhibitor	**SaDHFR(WT) IC_{50} (nM)**	**SaDHFR(F98Y) IC_{50} (nM)**
R-27	15 ± 0.7	510 ± 50
S-27	18 ± 2	111 ± 6
TMP	23 ± 3	1700 ± 19

Two crystal structures of R-27 and S-27 complexed with NADPH and SaDHFR (PDB ID: 4xec and 4tu5) were determined in [17]. Analysis of the crystal structures revealed that S-27 binds to SaDHFR with β-NADPH, while R-27 binds to SaDHFR with what was originally believed to be α-NADPH. However, upon further inspection we realized that the α-NADPH did not accurately fit the electron density map (Figure 48), and changing the conformation of NADPH (i.e., rotating bonds) without a change in chirality would not fix the problem.

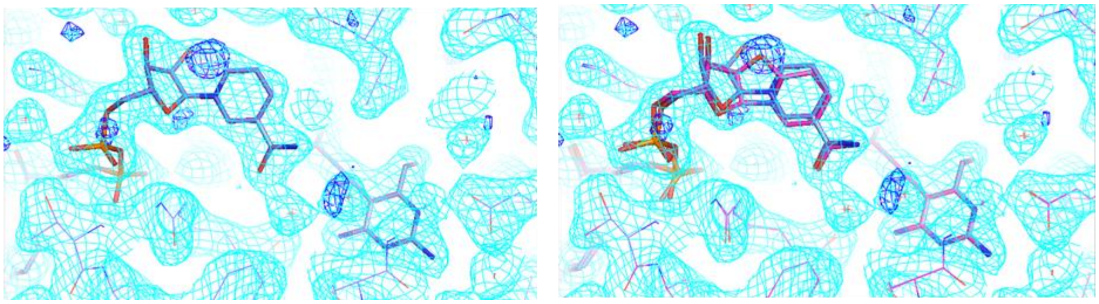


Figure 48 : Electron density map of PDB ID 4XEC. Density near the nicotinamide ring of NADPH previously assigned as α-NADPH (right pannel). Electron density is better fit by t-NADPH as shown in the left panel (α-NADPH in blue, t-NADPH in purple) (Data from Wright Lab University of Connecticut).

"NADPH is composed of ribosylnicotinamide 5^1-phosphate coupled by pyrophosphate linkage to the 5^1-phosphate adenosine 2^1,5^1-diphosphate (PADP). The chiral center of NADPH is the 1^1 carbon on the ribose connected to nicotinamide. Three forms of NADPH tautomer are shown in Figure 1b. Although these observations suggested a close correlation between resistance and cofactor influence, the detailed mechanism behind it remained unknown."

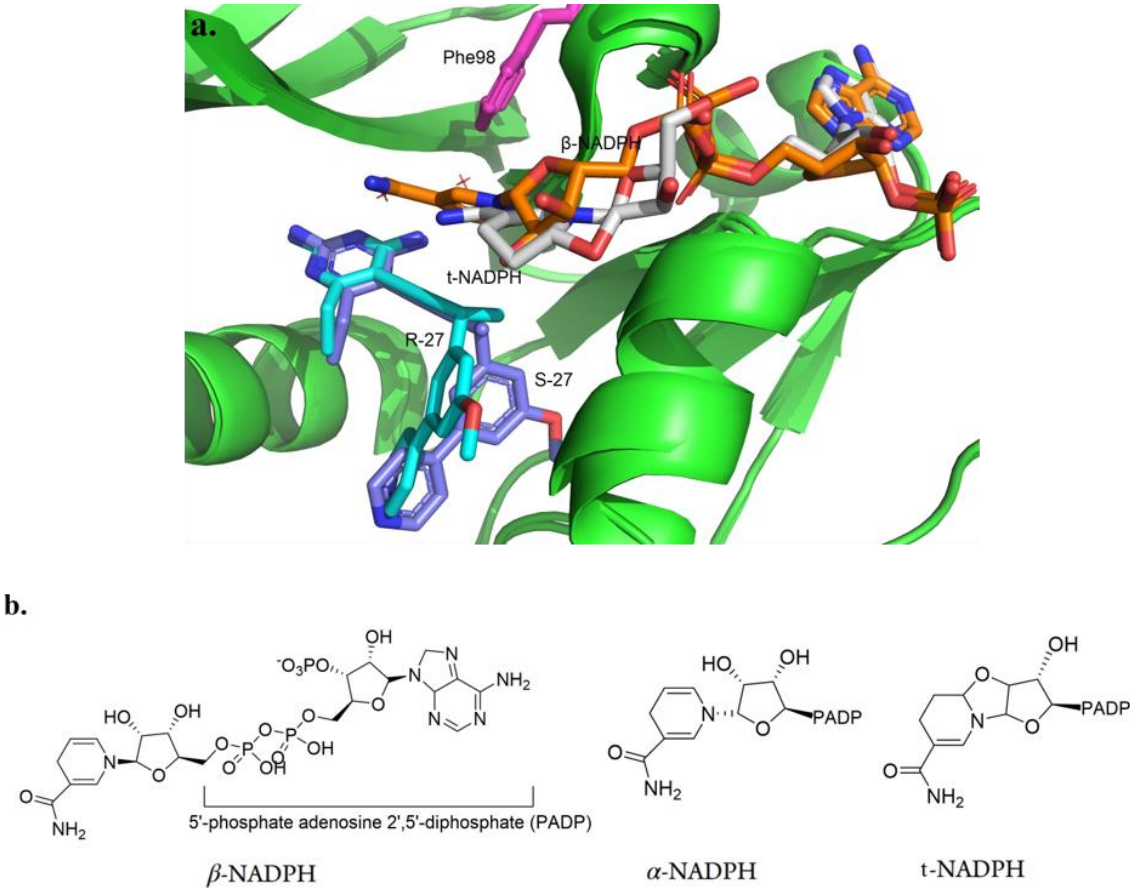

a.
Phe98
β-NADPH
t-NADPH
R-27
S-27
b.
OH
$^{-}O_3PO$
N
NH_2
HO
OH
O
P
OH
N
NH_2
5'-phosphate adenosine 2',5'-diphosphate (PADP)
PADP
β-NADPH
α-NADPH
t-NADPH

Figure 49: Structure of DHFR bound to different enantiomers of NAHD and different enantiomers of a PLA. a. Crystal structures of R-27 (colored in cyan) bound to DHFR and t-NADPH (colored in light grey), and S-27 (purple) bound to DHFR and *β*-NADPH (orange). PDB ID 4XEC and 4TU5. Phe98 is highlighted in magenta. Starting from the propargylic stereocenter, the biaryl moieties on R-27 and S-27 point in different directions and thus R-27 and S-27 adopted very different conformations. b. ***β*-NADPH, *α*-NADPH and t-NADPH.**

"Here we describe how we first refined the X-ray diffraction data of R-27:DHFR:SaDHFR and remediated its structure. Upon careful investigation of the density map, α-NAPDH in [17] is now reassigned as the closed, tricyclic tautomer of NADPH (t-NADPH, *α*-02'-6B-cyclotetrahydronicotinamide adenine dinucleotide). Using remediated structure, we then applied a series of computational tools and methods to explain and understand the mechanism of resistance in the F98Y SaDHFR mutant. The major computational tool we used is our CPD software suite OSRPEY [68]. Given a structural model as input, OSPREY exploits provable algorithms to calculate a K* score, which accurately approximates the binding affinity of the given complex [68]. OSPREY also outputs an ensemble of predicted lowest energy conformations of the protein-ligand complex [20, 78, 79], which provides with a detailed model to understand the structural basis of binding affinity changes. In this technique, the thermodynamic ensemble is efficiently predicted by OSPREY, and used to compute K* score which is an approximation of association constant K_a. When K* scores correlate with experimental measurements, the molecular ensemble provide a structural model and explanation for the observed experimental data. Herein, computational techniques allowed us to enumerate and fully explore all possible complexes between different PLA enantiomers, NADPH tautomers and SaDHFRs (WT vs. F98Y mutant). Excitingly, our computational resultsshowed excellent concordance with IC_{50} data, indicating that our computational methods successfully recapitulated the experimental measurements. The K* scores calculated by OSPREY suggested an explanation for the stereospecific inhibition to SaDHFR, which correlated to the inhibitors binding preference to different NADPH configuration. Moreover, the OSPREY-produced ensembles of conformations predicted the detailed molecular contacts around the enzyme active sites, providing a structural basis for the mechanism of emergence of F98Y-mediated resistance. This mechanism which we have named chiral evasion is to the best of our knowledge a new mechanism of drug resistance. “

3.4.1 Reassignment of the crystal structure of t-NADPH in complex with DHFR and R-27

Upon review and refinement of the crystal structure of R-27 bound to SaDHFR (PDB ID 4XEC), it was revealed that the NADPH was actually in a α-02'-6B-cyclotetrahydronicotinamide adenine dinucleotide configuration (Figure 48).

NADPH naturally alternates between the alpha and the beta isoforms. However, at low pH α-NADPH can become trapped as α-02'-6B-cyclotetrahydronicotinamide adenine dinucleotide by undergoing a cyclisation through the reaction of the 2’hydroxyl on the ribose and tetrahydronicotinamide ring to form a-02'-6B-cyclotetrahydronicotinamide adenine dinucleotide, hereafter t-NADPH [338] (Figure 49, b.). One explanation for the presence of t-NADPH in the enzymes is the low pH (pH 5) used during crystallization conditions. Alternatively, the enzyme binding pocket could potentially alter the pKa of NADPH, allowing for the cyclization reaction to occur even under physiological conditions. Currently, the biological relevance of t-NADPH, or in fact its presence in cells, is not currently known. However, the observation that t-NADPH is only found in SaDHFR bound to certain antifolates and more commonly in SaDHFR mutants containing the resistance mutation F98Y suggests that either the binding of α-NADPH or t-NADPH may play a role in resistance to antifolates and help explain the mechanism of resistance of F98Y that is not well understood [17, 331]. We will use computational modeling using OSPREY [68] to attempt and explain the mechanism of F98Y resistance and the role played by NADPH chirality in the evasion of DHFR inhibition.

Currently it is believed that the F98Y mutations may act by having the hydroxy group on the Y98 stabilize the transition state, by stabilizing the formation of the carbocation on the C4 of the nicotinamide ring in the NADPH and the C6 of the dihydropterin ring on DHF [314, 315]. This would increase kcat and potentially decrease the effectiveness of antifolates. However, this effect does not explain the nearly 112 fold loss of affinity between the WT and the F98Y mutant to TMP, the only currently approved antifolate for antibiotic use [339]. From inspection of the crystal structures, the competition between Y98 and the N4 amine nitrogen of TMP for the binding to the main chain of Leu5 could contribute to this loss of binding; however, the effect is still rather pronounced. Thus, it has been hypothesized that the F98Y mutation may increase the population of different conformation of DHFR, including some bound to alternative configurations of NADPH, that bind to TMP with significantly lower affinity. This is supported by the observation that the binding NADPH functions co-operatively to drastically increases the affinity of TMP for DHFR and that the F98Y mutations disrupts this co-operative binding effect to considerable degree [339].

3.4.2 Comparison between K* scores and IC50 values

In our previous study (18) we found that SaDHFR, in particular its F98Y mutant, showed remarkable stereospecifity to inhibitors of different enantiomeric state (Table 9). IC_{50} values and crystal structures were determined for a PLA enantiomer pair, S-27 and R-27, complexed with SaDHFR and cofactor NADPH. The crystal structure revealed that S-27 and R-27 bound SaDHFR with different tautomeric states of NADPH. S-27 bound with the catalytically active β-NADPH while R-27 bound with t-NADPH.

To investigate and understand the role of stereospecificity of PLAs in binding and the role of the stereospecificity of NADPH as a mechanism of resistance, we performed computational analysis using our CPD software suite OSPREY [68]. To understand the effect of stereospecificity of PLAs on the stereospecificity of NADPH and how the F98Y mutations modulates this interaction, we used OSPREY to compute K* scores for 8 different systems, generated by combinatorially selecting one inhibitor from S-27 or R-27, one NADPH conformation from β-NADPH or t-NADPH, and one enzyme from WT or F98Y DHFR (Table 10).

Table 10: Partition functions (PF) and K* scores (reported as Log_{10} K* scores) calculated using OSPREY.

System	Bound state PF	Unbound state PF	inhibitor PF	K* score
R27:t-NADPH:DHFR(WT)	191.15	146.05	0.30	44.79
S27:t-NADPH:DHFR(WT)	182.38	146.05	0.30	36.02
R27:β-NADPH:DHFR(WT)	180.78	145.96	0.30	34.53
S27:β-NADPH:DHFR(WT)	190.59	145.96	0.30	44.32
R27:t-NADPH:DHFR(F98Y)	183.55	144.05	0.30	39.19
S27:t-NADPH:DHFR(F98Y)	178.82	144.05	0.30	34.27
R27:β-NADPH:DHFR(F98Y)	174.80	142.29	0.30	32.22
S27:β-NADPH:DHFR(F98Y)	186.30	142.29	0.30	43.71

A K* score is a mathematically provable approximation of the binding constant K_A. This approximation is obtained by calculating the quotient of the Boltzmann weighted approximation of the partition functions, given a flexibility model, for the bound state by the unbound states. This value is the reciprocal of the dissociation contestant K_D. Unfortunately, the experimental data reported in Table 9 are in IC_{50}. The relationship between K_D and IC_{50} is described by the Cheng-Prusoff equation [340] (Equation 3).

$$IC_{50} = K_i\left(1 + \frac{S}{K_m}\right)$$

Equation 3: Cheng-Prusoff equation where Ki is equivalent to the inhibitor's KD. Km is the Michaelis constant and S is the concentration of substrate (more commonly represented as [S]).

In the experiments in Table 9, S was constant and Km decreases 3 fold from SaDHFR(WT) to SaDHFR(F98Y)[83]. It is clear then that under these condition KD or Ki approximates IC50, allowing us to compare our OSPREY-calculated K* values (Table 10) to the experimentally determined IC50 values (Table 9). From we can see that for WT SaDHFR, the K* score for R-27:t-NADPH:DHFR(WT) ($\log_{10}$ K* score of 44.79) is almost equal to but very slightly higher than for S-27:β-NADPH:DHFR(WT) ($\log_{10}$ K* score of 44.32). These K* scores are significantly higher than the scores for S-27:t-NADPH:DHFR(WT) ($\log_{10}$ K* score of 36.02) and R-27:β-NADPH:DHFR(WT) ($\log_{10}$ K* score of 34.53). These result not only agree with IC50 data (Table 9) very well, they also help explain the observed NADPH tautomeric states in the crystal structures.

"No crystal structures exist for SaDHFR(F98Y) bound with R-27 or S-27. Therefore we generated models to investigate their properties. Based on our models, OSPREY predicted that R-27 and S-27 retain their preferences for their respective NADPH configurations. Similarly to those with WT DHFR, K* scores for R-27:t-NADPH:DHFR(F98Y) ($\log_{10}$ K* score of 39.19) and S-27:β- NADPH:DHFR(F98Y) ($\log_{10}$ K* score of 43.71) are significantly higher than those for S-27:t-NADPH:DHFR(F98Y) (($\log_{10}$ K* score of 34.27) and R-27:β-NADPH:DHFR(F98Y) ($\log_{10}$ K* score of 32.22).

Moreover, the K* scores for S-27 bound with F98Y DHFR only decreases by a moderate extent compared with when binding with WT DHFR (the $\log_{10}$ K* score for S-27:β-NADPH:DHFR(WT) is 44.32 and for S-27:β-NADPH:DHFR(F98Y) is 43.71, decrease in $\log_{10}$ K* score between WT and F98Y is 0.61). But for R-27, compared with bound with WT DHFR, K* score decreased a lot when binding with F98Y (the $\log_{10}$ K* score for R-27:t-NADPH:DHFR(WT) is 44.79 and for R-27:t-NADPH:DHFR(F98Y) is 39.19, decrease in $\log_{10}$ K* score between WT and F98Y is 5.60). These results showed remarkable high concordance with IC_{50} data, as a comparison between them is shown in bar graphs (Figure 50). Based on these results, our computational analysis successfully recapitulated the thermodynamic and structural data that has been experimentally measured."

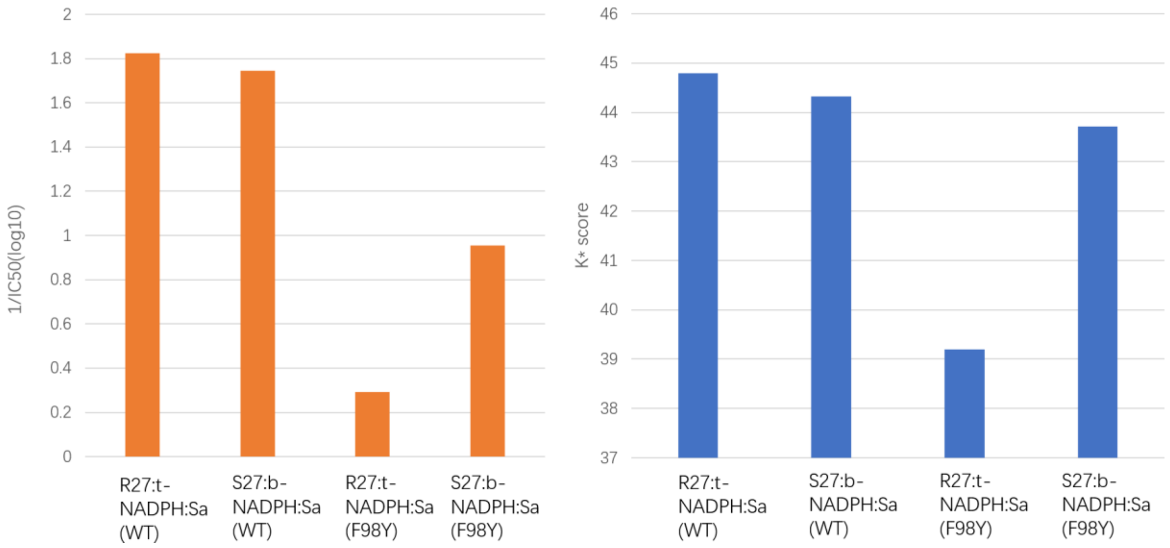


Figure 50: Bar graphs of $1/IC_{50}$ ($\log_{10}$) values (left) and $\log_{10}$ K* scores (right) showing the change effect of the F98Y mutation on inhibitor potency and binding. Calculated K* scores are show strong concordance with experimental IC_{50} values.

3.4.3 Structural analysis

OSPREY was used to generate ensembles of conformations to provide detailed contact information around the active site for each inhibitor, bound to each NADPH tautomer, in both SaDHFR WT and F98Y. By examining these structures we can gain an understanding of the structural basis of F98Y resistance as well as the stereospecific inhibition of DHFR.

"The lowest energy conformations from the ensembles of 4 WT DHFR-related complexes (R-27:t-NADPH:DHFR(WT), R-27:β-NADPH:DHFR(WT), S-27:β-NADPH:DHFR(WT) and S-27:t-NADPH:DHFR(WT)) are shown in Figure 51. The interaction between antifolates (R-27 or S-27) and DHFR and NADPH is visualized using Probe dots [341]. Red and yellow dots represent unfavorable overlaps, whereas green and blue dots represent H-bonds and van der Waals (vdW) contacts. In all of our analyses, two water molecules in the binding pocket were modeled along with t-NADPH in all complexes containing t-NADPH, since these water molecules are crucial for bridging contact between t-NADPH, antifolates and DHFR."

Figure 51 panel a, shows the atomic interactions for R-27 binding to t-NADPH:DHFR(WT). We can observe multiple favorable interactions between the propargyl linker and t-NADPH. When we modeled R-27 binding to β-NADPH:DHFR(WT) (panel b), we observe the introduction of clashes between the propargyl linker and nicotinamide ring that penetrates deeper into the binding pocket in the β-NADPH configuration. These clashes with β-NADPH help explain the higher K* score for R-27 binding to t-NADPH ($\log_{10}$ K* score of 44.79) versus β-NADPH ($\log_{10}$ K* score of 34.53) and are a mechanistic explanation to why the crystal structure showed a

100% occupancy of t-NADPH (PDB ID 4XEC). Figure 51 panel c, shows the binding of S-27, β-NADPH:DHFR(WT) the different angle of the methyl group on the chiral carbon allows for significant proximity and optimal Van der Waals contacts between S-27 and β-NADPH ($\log_{10}$ K* score of 44.32). However, when t-NAPDH is modeled in the same location, many of these interactions are lost ($\log_{10}$ K* score of 36.02), explaining the preference of S-27 for β-NADPH.

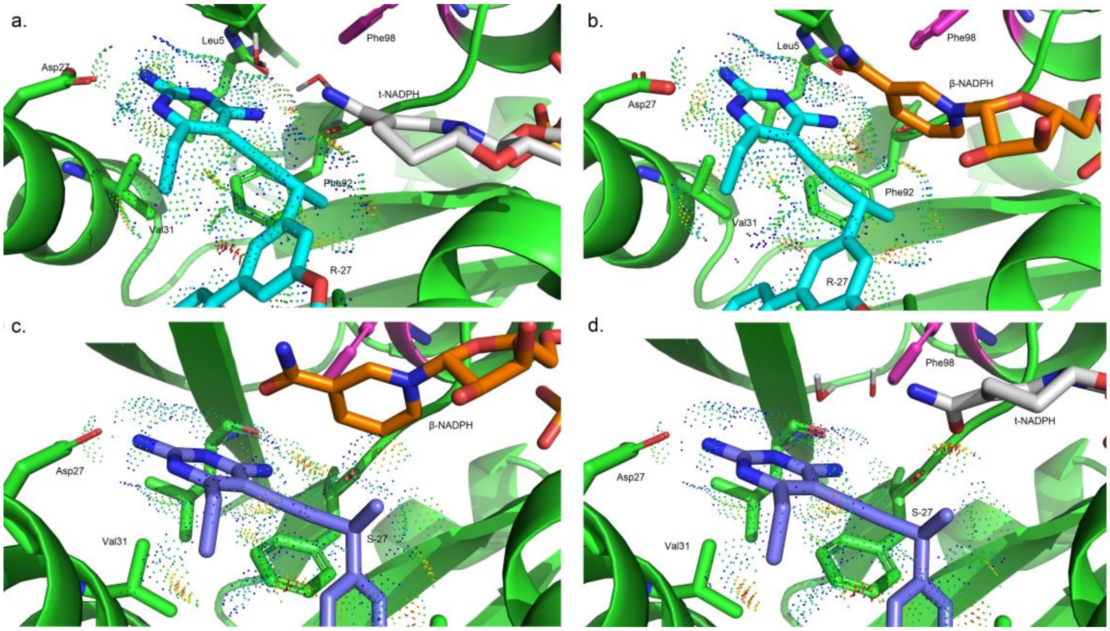

a.
Leu5
Asp27
Phe98
t-NADPH
Phe92
Val31
R-27
b.
Leu5
Phe98
Asp27
β-NADPH
Phe92
Val31
R-27
c.
β-NADPH
Asp27
Val31
S-27
d.
Phe98
t-NADPH
Asp27
S-27
Val31

Figure 51: OSPREY designed complexes showing INCA binding to DHFR with different NADPH enantiomers. a. R27:t-NADPH:DHFR(WT); b. R27:β-NADPH:DHFR(WT); c. S27:β- NADPH:DHFR(WT); d. S27:t-NADPH:DHFR(WT). Probe dots show the contacts between the PLAs and surrounding DHFR residues and NADPH.

Having determined that R-27 has a strong preference for binding to t-NADPH and that S-27 has a strong preference to bind β-NADPH, we set out to compare the effect of the F98Y mutation on the binding of R-27 (together with t-NADPH) and S-27 (together with β-NADPH) to DHFR.

Figure 52, shows the binding of R-27 to DHFR. Figure 52 panel a, show the bound complex between R-27 and t-NADPH:DHFR(WT) as previously described this is a very favorable state. The F98Y mutation introduces a hydroxy group that increases the bulk of the side chain at residue 98. This results in some bad clashes between the Y98 and the backbone of G92. (panel b). Panel c and d in Figure 52 show the lowest energy conformation of t-NADPH:DHFR(WT) and t-NADPH:DHFR(F98Y) without the PLA inhibitor bound. The absence of the inhibitor provides the space necessary for the relaxation of strained interactions within the pocket. This allows the clashes between Y98 and G93 to minimize away, lowering the energy of the F98Y unbound state relative to the F98Y bound state ($\log_{10}$ K* score of 39.19) and causing a negative effect on the K* score (difference of $\log_{10}$ K* of 5.6) relative to WT ($\log_{10}$ K* score of 44.79) and explaining the marked loss in IC_{50}.

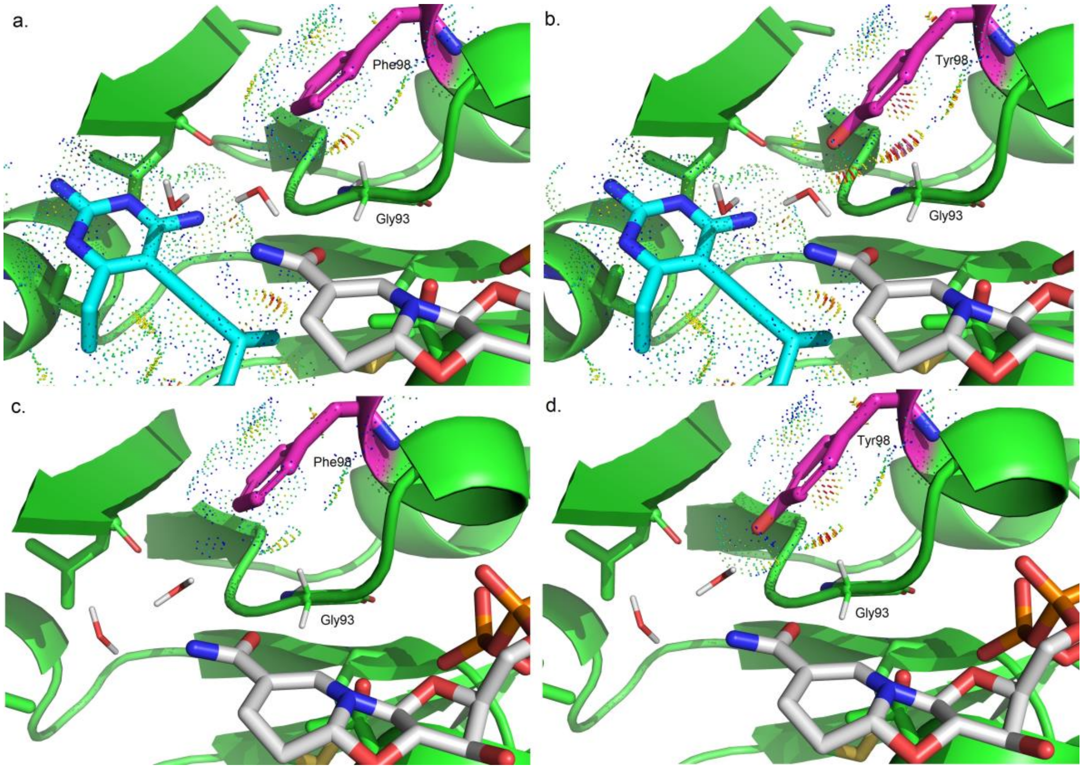


Figure 52: OSPREY designed complexes for R-27 bound and unbound WT and F98Y DHFR. a. Bound state R27:t-NADPH:DHFR(WT) b. Bound state R27:t-NADPH:DHFR(F98Y) c. Unbound state t-NADPH:DHFR(WT) d. Unbound state t- NADPH:DHFR(F98Y)

The influence of the F98Y mutation on S-27 binding (Figure 53) (panel a and b) is very similar to the influence of F98Y on R-27. Compared with the WT amino acid at residue 98 (phenylalanine), the mutation to tyrosine leads to clashes with Gly93. However, the unbound state conformations for β-NADPH with DHFR(WT) or DHFR(F98Y) (panel c and d) are differ from the unbound conformation of complexes containing t-NADPH. β-NADPH itself forms substantial contacts with DHFR and unlike t-NADPH, there are no long-lived water-mediated interactions within the pocket. Therefore, even in the unbound state, β-NADPH is sterically restricted, so that clashes between Y98 and G93 cannot be minimized away. Therefore the effect of the F98Y mutation is to destabilize both S-27 bound and unbound states. Quantitatively, the K* score does not change too much from (Log$_{10}$ 44.32 for S27:b-NADPH:DHFR(WT) and to Log$_{10}$ 43.71 for S27:b- NADPH:DHFR(F98Y)) a change in Log$_{10}$ K* score of 0.61.

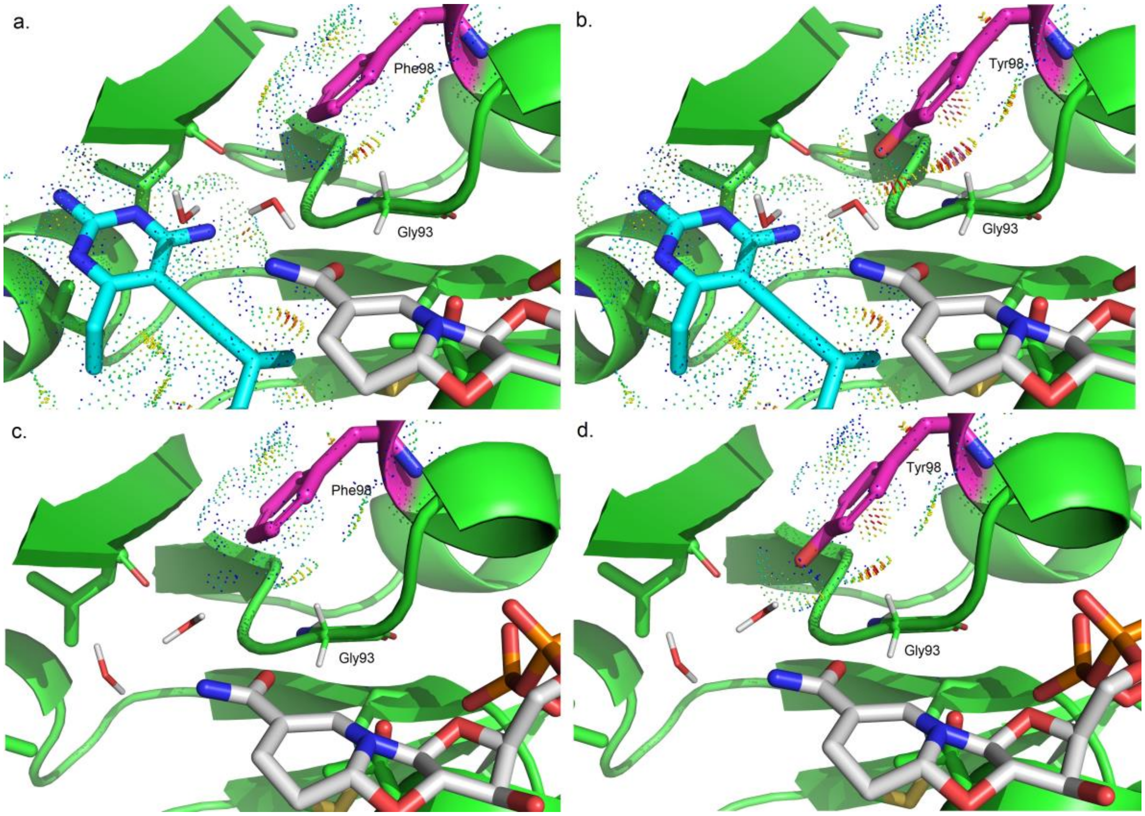


Figure 53: OSPREY designed complexes for S-27 bound and unbound WT and F98Y DHFR.a. Bound state S27:β-NADPH:DHFR(WT) b. Bound state S27:β-NADPH:DHFR(F98Y) c. Unbound state β-NADPH:DHFR(WT) d. Unbound state β- NADPH:DHFR(F98Y)

3.4.4 Discussion and Conclusion

Different NADPH states have been observed in crystals structures of resistant and WT *S.aureus* DHFR bound to novel propargyl-linked antifolates. These different bindings states of NADPH were initially believed to be conformational differences [331], however, analysis by Pablo Gainza during his thesis work in the Donald lab showed that NADPH was binding in a different configuration [17, 331]. These configurations were initially believed to be NAPDH in its alpha-configuration (α-NADPH) which exists in equilibrium with the more prevalent and DHFR active beta-configuration (β-NADPH). Both alpha and beta NADPH are found in cells in vivo, with the alpha form being present at roughly 1.5% the concentration of the beta form [334, 336]. Here we have shown that upon review and refinement of the crystal structure, PDB ID 4XEC. that the NADPH was actually in a α-02'-6B-cyclotetrahydronicotinamide adenine dinucleotide configuration which we have called "tricyclic" or t-NADPH (Figure 1). NADPH spontaneously interconverts between the alpha and the beta isoforms. However, at low pH α-NADPH can become trapped as t-NADPH by undergoing a cyclization through the reaction of the 2'hydroxyl on the ribose and the tetrahydronicotinamide ring to form t-NADPH [338]. One explanation for the presence of t-NADPH in the enzymes is the low pH (pH 5) used during crystallization conditions. Alternatively, the enzyme binding pocket could potentially alter the pKa of NADPH, allowing for the cyclization reaction to occur. Finally, an interesting third possibility is that the enzyme could function as an epimerase. The formation of t-NADPH requires the anomerization of beta to alpha NADPH to position the nicotinamide and ribsose alcohol on the same plane [338]. DHFR could facilitate the transition between beta to alpha DHFR and stabilize a hydrated form of α-NADPH, which is the believed transition state step to t-NADPH (15). This is supported by a hypothesis from Oppeinheimer and Kaplan in 1974 when they first solved the structure of t-NADPH and predicted that a stabilization of the charge on N1 through induction effects on C-5 or O-1′ caused by charged amino acids could activate the coenzyme [342]. To the best of our knowledge, DHFR has never been reported as an epimerase but this is an exciting possibility that merits further investigation. If this was the case DHFR could be partially responsible for the maintenance of equilibrium between α-NADPH and β-NADPH in Staph cells. Given that oxidised α-NADPH is rather stable and inactive, the importance of NADPH as an essential reducing agent in cells, and its role in the biosynthesis of numerous pathways, the conversion between the active β-NADPH and inactive α-NADPH may be a form of cellular regulation, including but not limited to a built-in negative feedback loop in DNA replication. In the context of our studies we have observed that the ability of DHFR to bind to t-NADPH is a mechanism of resistance that is synergistic with the emergence of the F98Y mutation, we have named this mechanism of resistance chiral evasion. This is supported by the observation that the binding NADPH functions co-operatively to increase the affinity of TMP for DHFR and that the F98Y mutation disrupts this co-operative binding effect to a considerable degree [339].

Our in silico study hypothesizes that chiral evasion is a novel mechanism of resistance in SaDHFR that could involve its putative epimerase functionality and would explain the effects of the SaDHFR F98Y mutation. Our in silico results show that

DHFR(WT) binds to t-NADPH with nearly the same $\log_{10}$ partition function energy as β-NADPH (146.05 vs 145.96, a difference of 0.09, Table 2) . While F98Y causes crowding in the DHFR active site when bound to β-NADPH, t-NADPH alleviates the crowding in the active site pocket. This can be seen by the difference in $\log_{10}$ partition functions increasing from 0.09 to 1.67 (Table 2). The less crowded active site stabilizes the enzyme in the t-NADPH:DHFR(F98Y) state, even in the absence of an inhibitor, and results in a loss of affinity for antifolates that compete with DHF for SaDHFR:β-NADPH, like S-27. This also decreases the affinity of inhibitors that bind to t-NADPH:DHFR, like R-27. This *in silico* result not only suggests a mechanism of action for the experimental observations both for S-27 and for R-27, but also rationalizes the difference in the magnitude of drug resistance provided by the F89Y mutation. Our in silico results suggest that S-27 should suffer a negligible loss in IC50 due to the F98Y resistance mutation (WT K* score: 44.32 vs F98Y K* score: 43.71 a 0.61 difference) if only the β-NADPH:DHFR state was populated. However, we predict a relative increase in the population of t-NADPH:DHFR, which has significantly lower affinity for S-27 (K* score 36.02 versus 44.32) , furthermore, the affinity of S-27 for t-NADPH:DHFR further decreases to a K* score of 34.27 upon mutation to F98Y mutation. This strongly agrees with experimental measurements that show a 6 fold loss in S-27 activity due to the F98Y mutation (Table 9). It is possible that the transition from β-NADPH to t-NADPH evinces a de facto resistance mechanism by reducing the affinity of S-27 to DHFR. A similar hypothesis was first suggested by S. Keshipeddy et al., 2015 [17], and our computational results seem to strongly support this hypothesis. Alternatively, it may be important to note that F98Y causes decrease in K_m compared to WT and that change can also partially explain the observed experimental results.

Moreover, our in silico results provide a putative mechanism for the observed change in IC50 of R-27 and the mechanism by which the F98Y mutation causes the 34-fold loss in activity (Table 1). Our in silico results indicate that, in contrast to S-27, which competes with DHF to bind DHFR:β-NADPH, the mechanism of inhibition of R-27 comes from its ability to bind and trap SaDHFR with the inactive t-NADPH. This result is corroborated by the crystal structure of R-27 bound to DHFR [17] and our in silico results. Our in silico results show for DHFR(WT), the R-27:DHFR(WT) complex binds to t-NADPH significantly tighter than to β-NADPH (with a K* difference of 10.26). This very large difference shows that once R-27 binds to DHFR it will very likely recruit t-NADPH. t-NADPH is not an active co-factor and therefore DHFR becomes trapped in a inncative state. However, once the SaDHFR(F98Y):t-NADPH conformation becomes stabilized due to the F98Y mutation, the binding of R-27 favors the binding of t-NADPH versus β-NADPH significantly less (with a K* difference of 6.97). We predict this is due to clashes between Tyr98 and Gly93 that are caused upon R-27 binding (Figure 5). Therefore R-27 becomes unable to trap SaDHFR(F98Y) in the inactive state, and loses its ability to inhibit SaDHFR activity as our studies predict that R-27 has low affinity for

SaDHFR: β-NADPH in either WT or F98Y backgournds (WT K* score: 34.53, F98Y K* score: 32.22 a delta of) and therefore is not an effective competitor of DHF for the active SaDHFR:β-NADPH state.

One remaining question is the relevance of this chiral evasion in vivo. The regulation of dehydrogenases like DHFR through hydrated forms of NADPH is a well know phenomenon [343, 344], as is the emergence of a tricyclic (t-NADPH) intermediate from these hydrated products [342, 344-346]. Also, multiple proteins have been identified that attempt to rescue NADPH from its hydrated form and turn it back into NADPH. It would make sense that regulation could be achieved by enzymes doing the opposite, the generation of either the hydrated or t-NADPH forms of NADPH, as a mechanism of regulation or a side product of the main reaction. DHFR may very well fall into one these categories. The mechanism of drug resistance of the F98Y mutation may both accelerate catalysis but also drive further formation of t-NADPH, thus potentially explaining F98Y's negative effect on viability [20] despite its increase in enzymatic efficiency. While R-27 and S-27 have very similar IC50 for DHFR(WT) they have nearly an order of magnitude different minimal inhibitory concentrations (MIC) in vivo (0.3125 μg/ml vs 0.0391 μg/ml). This discrepancy between IC50 and MIC may reflect a more controlled NADPH population in vivo which would be expected as cells exercise strict control over their NADPH population, as NADPH can get trapped in inactive states like t-NAPDH, decreasing cellular viability [344]. Interestingly, the effect of the F98Y mutation in vivo is nearly identical for both R-27 and S-27: In vivo R-27 suffers a 16 fold increase in MIC to 5 μg/ml and S-27 suffers a 15.98 fold increase in MIC to 0.625 μg/ml [17]. One possible explanation (supported by our *in silico* results) is that (a) both S-27 and R-27 bind to SaDHFR(F98Y):t-NADPH with reduced affinity and that (b) the change in MIC reflects the change in population of SaDHFR(F98Y):β-NADPH to SaDHFR(F98Y):t-NADPH. This would be a valuable hypothesis to test in future work.

In conclusion, we showed how the discovery of a configuration change in NADPH can potentially elucidate a mechanism of DNA synthesis regulation and drug resistance in SaDHFR. Further experiments are necessary to validate these hypothesis. The data and models we present have already been useful in medicinal chemistry campaigns for F98Y-resilient inhibitors [325], and showed their value to the community.

4. Improving the Accuracy of Computational Protein and Drug Design

Although computational design has been successfully used by us and others [68, 214, 235, 241, 242, 347-355] to design PPI inhibitors, antibodies, and antigens, to change the specificity of enzymes, to predict novel resistance mutations, and to design new folds, significant room remains to improve the accuracy and predictive capabilities of CPDD. An algorithmic perspective on potential improvements have been described elsewhere [68, 226]. Here I present a biochemical perspective and my contribution towards improving protein design.

4.1 Benchmark of OSPREY 3.0

The following work was done in contribution to a published manuscript describing the newest stable release of OSPREY, OSPREY 3.0. This work was done in collaboration with Anna U. Lowegard and Jonathan D. Jou as part of the paper which was a collaboration with M. A. Hallen, J. W. Martin, A. Ojewole, J.D. Jou, A.U. Lowegard, P.Gainza, H. M. Nisonoff, A. Mukund, S. Wang, G.T. Holt, D. Zhou, E. Dowd and B. R Donald. My contribution included computational results towards the included benchmark, as well as intellectual contributions towards the generation and format of the benchmark itself. Herein described is also an original in-depth analysis of these result. The following sub-chapter is adapted from the publication below. Verbatim sections are demarked by quotation marks. Table 11-15 are from the publication below.

To the best of my knowledge these sections were written by Anna Lowegard with contributions from all the authors:

M. A. Hallen, J. W. Martin, A. Ojewole, J.D. Jou, A.U. Lowegard, P.Gainza, H. M. Nisonoff, A. Mukund, S. Wang, G.T. Holt, D. Zhou and E. Dowd, B.R. Donald. OSPREY 3.0: Open-Source Protein Redesign for You, with Powerful New Features. Journal of Computational Chemistry. 2018. 39(30): p. 2494-2507.

"We performed new, retrospective tests, described below. To evaluate the accuracy of the implementation of the newest optimizations in OSPREY 3.0 [68], we performed a series of designs for a variety of protein–protein interfaces (PPIs) as retrospective validation. We used K* [79] to computationally predict experimentally measured changes in binding for each PPI. Each protein structure is listed by name and PDB ID in tables (8-12) [356-359] These systems include barnase with its peptide inhibitor barstar, [360, 361] the cytochrome c:cytochrome c peroxidase complex, [362] interferonα-2 (IFNα2) in complex with interferonα/βreceptor 2 (IFNAR2), [363] and the interleukin 2 (IL-2):IL-2receptorα(IL-2Rα) complex. [364] Our retrospective validation experiments focused on mutations at residues in or proximal to the protein–protein interface that were not limited to alanine scanning. We tested anywhere from 5 to 19 designs. In total, we tested 58 mutations using default, out-of-the-box OSPREY 3.0 settings and parameters. Each design included one or two mutable residues along with a set of surrounding flexible residues (Table (8-12)). Flexible residues were chosen by selecting all residues within 4 Å of the mutable residues and removing those that only have backbone interactions. For each system, the K* scores were ranked in increasing order of reported experimental binding. Spearman's ρ values were subsequently calculated for each system by calculating the statistical dependence between the K*score rankings and the experimentally measured rankings (Table (8-12)).This is a sound measure because generally the output of a design calculation that is used to decide which mutants to make experimentally is simply the intra-system ranks of the mutants. Looking at the values in Tables (8-12) we see a high correlation in the rankings between experimentally measured binding and binding predicted by OSPREY3.0 and K* for each system with values ranging from 0.500 to 0.795. We found that, across the tested systems, the Spearman's ρ value is 0.762. This value is the Pearson correlation of the intra-system ranks of all the mutants. Moreover, it is worth remembering that Spearman values range from -1 to 1 with 0 being a random distribution. Overall, these correlations are very good for design for affinity in computational protein design."

Our retrospective study for KRas had shown a slightly better Spearman ρ value of 0.81 (**Figure 5**), increasing the range of accuracies across systems from 0.5 to 0.81 when analyzing all the results together. Although a Spearman ρ value of 0.5 is still predictive, the range of correlation begged the question as to what affects accuracy and what can be done to improve protein design.

Table 11: Comparison of experimental vs computational ranking for barnase with its peptide inhibitor barstar, [360, 361]. Mutations are displayed in order of experimental ranking. The calculated Spearman's ρ value for the correlation between the experimental and computational ranking was of 0.755.

Barnase:Barstar, PDB ID:1X1U			
Variant	Exp. Ranking	Comp. Ranking	Difference
D39A	1	1	0
H203A	2	3	-1
R87S	3	5	-2
K27A	4	8	-4
R59A	5	2	3
D35A	6	4	2
Y29A	7	7	0
E73A	8	12	-4
E76A	9	6	3
W35F	10	11	-1
E60A	11	10	1
Y29F	12	9	3
			ρ= 0.755

Table 12: Comparison of experimental vs computational ranking for interleukin 2 (IL-2):IL-2receptorα(IL-2Rα) complex [364]. Mutations are displayed in order of experimental ranking. The calculated Spearman's ρ value for the correlation between the experimental and computational ranking was of 0.554.

IL-2:IL-2Rα, PDB ID:2B5I			
Variant	Exp. Ranking	Comp. Ranking	Difference
K38E,S39D	1.5	1	0.5
R35T,R36S	1.5	2	-0.5
R35K,R36K	3	4	-1
E1K, D4K	4	7	-3
E29R	5	5	0
L2A	6	16	-10
D4K	7.5	9	-1.5
D39A S41A	7.5	12	-4.5
E1K	9	11	-2
H120A	10	10	0
E29A	11	6	5
L42S, Y43L	12	3	9
E1Q	13	14	-1
N27A	14	15	-1
K38T	15	8	7
D4N	16	13	3
			ρ= 0.554

Table 13: Comparison of experimental vs computational ranking for cytochrome *c*:cytochrome *c* peroxidase complex [362]. Mutations are displayed in order of experimental ranking. The calculated Spearman's ρ value for the correlation between the experimental and computational ranking was of 0.500.

Cyt c:Cyt c peroxidase, PDB ID: 2PCB			
Variant	Exp. Ranking	Comp. Ranking	Difference
E290N	1	2	-1
D34N	2	4	-2
A193F	3	1	2
E35Q	4	3	1
E32Q	5	5	0
			ρ=0.500

Table 14: Comparison of experimental vs computational ranking for interferonα-2 (IFNα2) in complex with interferonα/βreceptor 2 (IFNAR2) [363]. Mutations are displayed in order of experimental ranking. The calculated Spearman's ρ value for the correlation between the experimental and computational ranking was of 0.795.

IFNα2:ifnar2, PDB ID: 3S9D			
Variant	Exp. Ranking	Comp. Ranking	Difference
R33Q	1	1	0
R33A	2	2	0
R33K	3	5	-2
L30A	4	6	-2
R149A	5	4	1
L30V	6	9	-3
A148A	7	10	-3
A145G	8	14	-6
A145M	9	3	6
L15A	10	13	-3
L153A	11	12	-1
L26A	12	7	5
S152A	13	16	-3
F27A	14	8	6
S25A	15	18	-3
D35A	16	17	-1
R22A	17	11	6
M16A	18	15	3
N196A	19	19	0
			ρ=0.795

Table 15: Comparison of experimental vs computational ranking for KRas:Raf-RBD complex [187,246,253,254]. Mutations are displayed in order of experimental ranking. The calculated Spearman's ρ value for the correlation between the experimental and computational ranking was of 0.795.

Kras:Raf-RBD, Homology model PDB ID 1GUA & 4DSN			
Variant	Exp. Ranking	Comp. Ranking	Difference
R89L	1	3	-2
F61W/R67L/V69E/N71R/V88I/A85K	2	2	0
K84A	3	4	-1
Q66A	4	16	-12
A85D	5	10	-5
R59A	6	7	-1
F61W/V69E/N71R/V88I	7	12	-5
R67A	8	6	2
K84L	9	5	4
Q66K	10	18	-8
T68A	11	20	-9
V88D	12	13	-1
T68K	13	1	12
V69A	14	22	-8
A85I	15	24	-9
K65A	16	17	-1
K65E	17	15	2
N64A	18	23	-5
V69R	19	34	-15
K87Q	20	26	-6
K65M	21	19	2
N71E	22	14	8
F61W	23	29	-6
F61W/R67L/N71R/V88I	24	11	13
V88I	25	28	-3
R67L	26	9	17
R59L	27	8	19
K84R	28	25	3

N64D	29	27	2
F61W/N71R/V88I	30	35	-5
K87R	31	30	1
F61W/N71R	32	37	-5
V88K	33	31	2
V88H	34	32	2
A85R	35	39	-4
N71R	36	36	0
N64K	37	21	16
V88R	38	33	5
A85K/V88R	39	38	1
A85K	40	40	0
N71R/A85K	41	41	0
			ρ=0.81

4.1.1 Importance of Accurate Structures as a Starting Point in Computational Predictions

One obvious hypothesis is that the resolution and quality of the structures affects our accuracy. We investigated this hypothesis using the five systems above (Tables 11-15) . Linear regression was applied to the relationships between these accuracies (quantified as Spearman ρ values) and four different measures of structure quality: resolution, Molprobity clashscore [365], and the average numbers of sidechain [366-368] and of Ramachandran outlier [369, 370] conformations. Accuracy was better for structures that excelled at each of these measures. Specifically, the linear regression gave an R^2 of 0.7411 for resolution, an R^2 of 0.5148 for clashscore , an R^2 of 0.8464 for side chain outliers, and an R^2 of 0.6825 for Ramachandran outliers (Figure 54).

Unfortunately, this study is underpowered by the limited number of systems available. A larger study to explore factors is being developed, and results will be very informative. Another limitation of this analysis is the assumption of a linear relationship between accuracy and structural factors. In reality one expects that the correlation (i.e., accuracy) may increase with better structural quality, but this relationship may be nonlinear. In particular, it is likely that beyond a certain level of structure quality, errors due to starting structure problems will become negligible compared to other sources of error and further improvements in structure quality will not lead to improvements in accuracy. Despite the preliminary nature of this analysis, and the fact that significantly more data is required to validate these claims, the observed trends strongly suggest that computational protein design has increased predictive power in systems or homology models that have high-quality structural information. This is a rather strong point to be made, as it quantifiable helps establish measures and expectations for results given known structural quality parameters of a systems. To the best of my knowledge, this is the first such study using the OSPREY platform. As future studies continue to analyze these effects, a definition of a "designable" system or residues may emerge. Such a definition would greatly improve our understanding of the applicability of computational drug design and help establish a frontier of what systems or residues can and cannot be designed with accuracy. Such frontier would be a good starting place and metric for advancements aiming to increase the applicability of CPDD platforms.

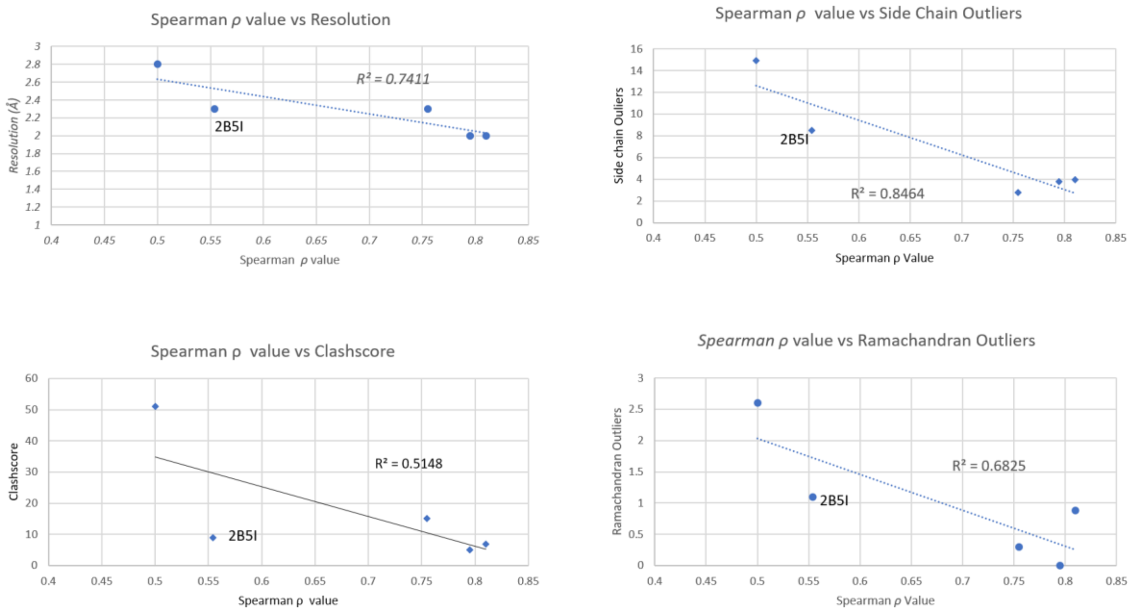


Figure 54: Analysis of structural quality measured versus Spearman ρ value. Top right is an analysis of the effect of resolution on Spearman ρ value and displays a linear correlation of 0.7411. Top left is an analysis of the effect of side chain outliers on Spearman ρ value and displays a linear correlation of 0.8646. Bottom right is an analysis of the effect of clashscore on the Spearman ρ value and displays a linear correlation of 0.5148. Bottom right is an analysis of the effect of Ramachandran outliers on the Spearman ρ value and displays a linear correlation of 0.6825. Systems tested were barnase with its peptide inhibitor barstar, [360, 361] the cytochrome c:cytochrome c peroxidase complex, [362] interferonα-2 (IFNα2) in complex with interferonα/βreceptor 2 (IFNAR2), [363] and the interleukin 2 (IL-2):IL-2receptorα(IL-2Rα) complex [364] and our homology model of KRas:Raf interaction. Note that the interleukin 2 (IL-2):IL-2receptorα(IL-2Rα) system (PDB ID 2B5I) is marked on the graphs to show that it is apart from the general trend.

4.1.2 The Role of Dynamics and Local Deviation in the Accuracy of Computational Predictions.

The structure of interleukin 2 (IL-2):IL-2receptorα(IL-2Rα) complex [364] (PDB ID 2B51) appears to not be following the same trend of improvement of predictive power as a function of structural improvement (Figure 54). If it is removed from the analysis, the correlations change significantly. In this scenario our benchmark data suggests a correlation of R^2 0.9364 for resolution, R^2 of 0.9932 for clashscore , R^2 of 0.9606 for side chain outliers and an R^2 of 0.9996 for Ramachandran outliers. However, it is clear that in an already data limited situation the removal of a data point (20% of the data) is clearly overfitting. However, we did wish to understand if the behavior of the interleukin 2 (IL-2):IL-2receptorα(IL-2Rα) system was a result of a weakness of the trend or the result of confounding factors. We decided to start by looking at the role of dynamics and local deviations on the accuracy of computational design for the interleukin 2 (IL-2):IL-2receptorα(IL-2Rα) system, in hopes of learning a generalizable trend for protein design.

In the above benchmark, it's important to note that OSPREY 3.0's accuracy was tested in an out-of-the box limited information scenario. Designers only had information present in a single complex crystal structure and did not investigate the system or its biochemistry beyond observations available from that one complex structures. This was done purposely, so that we could gain an understanding of the expected prospective accuracy of OSPREY in system that have little information and characterization, i.e. the hardest systems possible. In this scenario the interleukin 2 (IL-2):IL-2receptorα(IL-2Rα) complex [364] (PDB ID 2B51) appeared as an outlier from the perspective of having crystallographic quality correlate to OSPREY's predictive performance. To understand this phenomena, and obtain a understanding of the reason behind the discrepancy in expected predicted power given its structural resolution we analyzed the interleukin 2 (IL-2):IL-2receptorα(IL-2Rα) [364] (PDB ID 2B51) system in more detail. Of the 16 designs in this system, 5 had a discrepancy between the experimental and computational ranking larger than 25% of the number of designs, i.e. the difference between the experimental ranking and the computed ranking was greater than 4. Interestingly, with the exception of E29, which appears to be on its own, the erroneously predicted mutations clustered together, occurring on residues K38, L42 and Y43 (double mutation) S39A and S41A (double mutation) or where found in the N-terminus L2 (residues on IL-2receptorα (IL-2Rα)). This clustering generated two hypothesis, first could the loci of the residues have been a reason for the error and two, if so could we have known a priori. By analyzing all crystal structures of IL-2receptorα(IL-2Rα) bound to multiple ligands [359, 364, 371, 372], H. Yang et al., [373] first noticed highly flexible regions in the IL-2receptorα(IL-2Rα) structure. Interestingly, this region spans residues 36-41, but significant shifts can be observed up to and including residue 42 and 43. Moreover, the N terminus also shows, as is common, large backbone movements.

Interestingly, residues 36-41 correspond to a loop that is present in two general conformations that are held in place by alternating hydrogen bond patterns at position S39 and S41 (Figure 55). Thus, it is possible that a lack of understanding of the unbound state and choice to not model backbone flexibility in the above benchmark may have contributed to our error in disproportionate ways to the quality of the structure. A repeat of designs done on this system with the addition of multiple backbone positions would help validate the above hypothesis. Studies to test this hypothesis are underway.

What is interesting to observe is that an analysis of all the available structural data for a system, combined with visual inspection of the hydrogen bonding pattern of a residue can promptly, if retrospectively, reveal what parts of the system will present difficulties for accurate designs. This suggests that it is important for designers to develop a strong understanding of the system and its dynamics to decide what residues are designable before even attempting initial designs.

In practice all projects described in this document benefited from these observations and insights, possible contributing towards their strong predictions and experimental results. Furthermore, designing from multiple starting backbones and potentially even different backbones for the bound and unbound states may be vital for accuracy in these systems. To perform said design one would require algorithms allowing for multi-state design. Multi-state design capabilities are a hallmark of OSPREY [68] as is the introduction of realistic backbone rearrangements [72, 75]. Moreover, the contributions of the thermodynamic ensembles of flexible backbones and flexible side chains therein to binding cannot be captured accurately by observation of a single lowest energy conformation, sometimes called the global minimum energy conformation (GMEC). Ensembles are required for accurate characterization of the contribution of each backbone and its side chains. OSPREY has led the way in providing a description of ensembles in binding [68, 76, 79, 223], and these observations clearly highlight their importance for accuracy. However, even if some large backbone deviations may require future algorithmic development to be modeled accurately, and faster algorithms may be required to model multiple backbones, the ability to identify areas of low predictive capabilities is essential to optimize the impact of computational protein design.

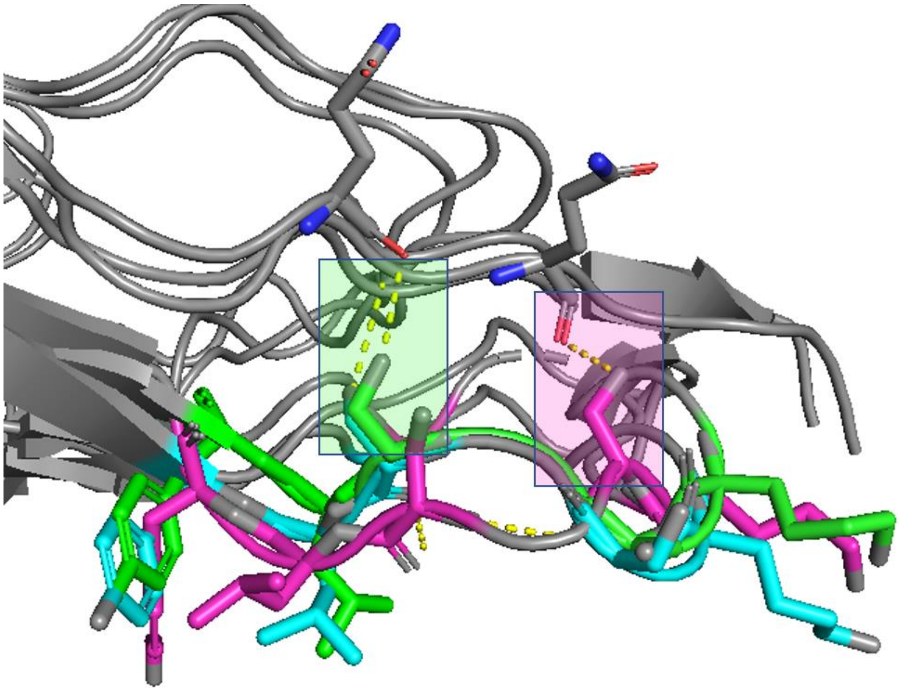


Figure 55: Richardson diagram representation of IL-2receptorα(IL-2Rα) bound to multiple ligands [359, 364, 371, 372]. Two alternative hydrogen bonding patterns for the residues loop ranging from resides 36-41 are shown. Either S39 (purple) of S41 (green) form hydrogen bonds holding the loop in distinct conformations. These conformations significantly alter the position of all residues from 36 – 43. Residue with incorrect predictions by in the benchmark are shown as sticks.

4.1.3 The Role of Visual Inspection in Computational Protein Design.

The discussion until this point has focused on how to identify the predictive power of OSPREY on a given system before starting designs. However, once OSPREY has finished design computations, the significantly more expensive and time-consuming step of experimental validation still looms. Thus it is worth performing some additional analysis after the design computations to make sure experimentally testing each designed sequence is worthwhile. Visual inspection is an important tool for analysis at this stage.

In chapter 2, a few select designs were not tested experimentally because they were pruned through visual analysis. A combination of strong predictive power from the computational designs and human analysis allowed us to test 6 designs and achieve unprecedent affinity for a Raf-RBD variant.

For our KRas designs, OSPREY predicted that mutation T57K would improve binding to Raf-1 RBD. Visual inspection of the result show what OSPREY was attempting to do was very insightful. As shown in Figure 56, KRas presents a very acidic surface at the KRas:RAF protein-protein interface. Therefore, the addition of a penetrating positive charge does indeed make sense. Moreover, OSPREY places the lysine in good position for a water mediated hydrogen bond and for electrostatic complementarity with E63. Furthermore, the lysine is placed as to optimize Van der Waals interactions with I36.

However, further analysis reveals that in doing so, one of the hydrogens on the primary amine of lysine is within 2.5Å of I36. This would force the lysine to have an unsatisfied hydrogen bond. Observation of alternative conformations in the thermodynamic ensemble reveals that other mutations bring together positive charges, which is also energetically unfavorable.

The T57K design was experimentally tested (Figure 22, Figure 32) and indeed it negatively affected binding. This result demonstrates that further improvement to the energy function is required, especially in situations of conflicting energetically favorable and unfavorable interactions. It also shows how human inspection can play an important role in understand CPDD results. Ultimately the success of generating a single-digit nanomolar-affinity Raf variant demonstrates how powerful careful analysis in combination with computational protein design software like OSPREY can be.

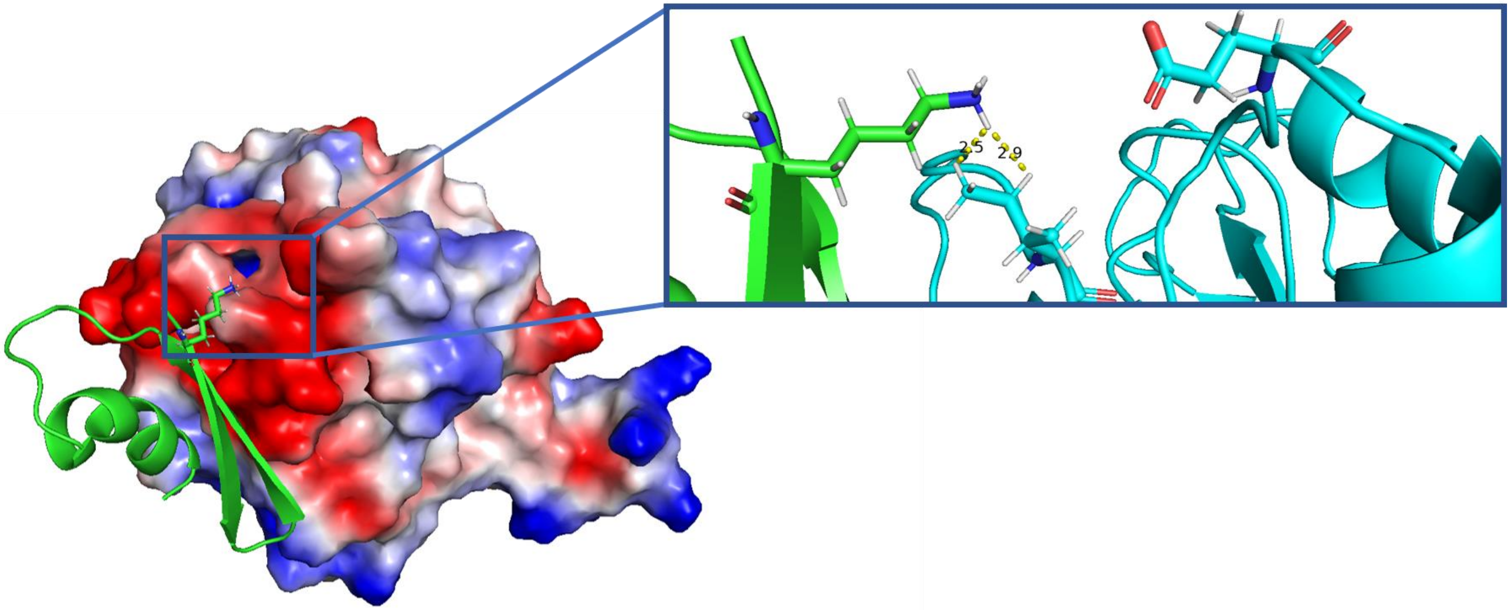


Figure 56: OSPREY designed KRas:RBD(T57K) mutation. KRas shown as a surface with vacuum electrostatics calculated using PyMol [256]. RBD (T57K) shows as Richardson diagram with T75K shown as a stick. Initial analysis shows a positive charge lysine being placed into a negatively charged surface (red is negative charge, white is non-polar, blue is positive charge) and suggests good columbic interactions. Deeper analysis (top right) suggests the position would actually results in the formation of a unsatisfied polar hydrogen in the primary amine of the lysine residue, as it is too close to I36 (shown as stick with measurements demonstrated as yellow lines) on KRas (cyan) to interact either with water or another polar residue.

4.1.4 Role of Rigid Body Translation and Rotation of PPI.

When observing the differences in the KRas:RBD binding between two variants , RBD(WT) in PDB ID 1GUA [255] and RBD (RK) in PDB ID 3KUC [246], we observed an interesting characteristic that could suggest additional flexibility, namely rigid body rotation and translation, is required to improve designs even further. Interestingly, this flexibility has already been included in OSPREY [68] and was seen to be vital for the accuracy of small molecule design [18-20] but was never utilized in protein design. Figure 57, shows the Rap1A/RBD interaction with both Rap1A having the A30D/K31E double mutation that makes Rap1A have an identical core effector region to KRas [255]. One can observe that upon alignment of the structures based on Rap1A (Figure 57, right image), the RBD variants appear to have a large backbone shift. However, RMSD between the RBD variants is actually only 0.44Å. When one aligns the structures by RBD variants (Figure 57, left image) one can see that the RBD variants have nearly identical backbone conformations. This shows that a few point mutations can alter the rigid-body orientation of a protein-protein interface. Having said that, our computational results suggest that either of the two rigid-body states of RBD(RK) have strong binding (data not shown), and multiple other systems such as barnase with its peptide inhibitor barstar, [360, 361] and the anti-HIV antibody VRC07 [222] do not present these shifts when mutated.

A possible explanation for the observed shift in RBD comes from the idea of rotational and translational entropy and the loss thereof upon binding [374, 375]. The coming together of two bodies decreases the freedom of either body to rotate and translate independently and presents an unequivocal loss in entropy. It is possible however, that small rotations and translations in the surface allow for compensation of this entopic loss and therefore should be included in the thermodynamic ensemble whenever possible. Moreover, it has been argued that rotation and translation of rigid bodies at a protein interface could be linked to internal backbone motions of proteins, allowing to some degree an accommodation of intra-body entropy [374, 375] while retaining good enthalpic effects in the interface. Finally, it has been argued [376] that these rigid-body motions may play an important role in the formation of long-range electrostatic interactions that eventually lead to the formation of the stable bound complex [377]. Interestingly, such long-range interactions have been suggested to be an important part of KRas binding Raf-1 RBD [269]. This may be characteristic of ubiquitin fold, like Raf-1 RBD as other studies have explored motions of ubiquitin folds and the effect of their rotation and translation on binding [376, 378].

Further algorithmic development will be required to implement this observation in OSPREY and test the effects of allowing rotation and translation of PPI's on the accuracy of our predictions. This work is ongoing.

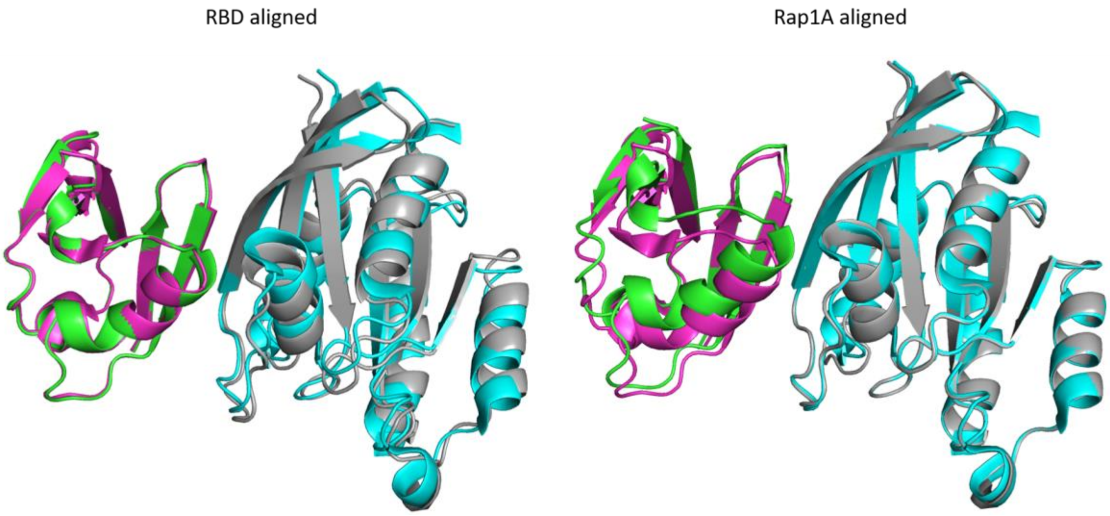


Figure 57: Rotation and translation of rigid bodies at PPI. Aligned structures of the complex of Rap1A (A30D/K31E):Raf-1 RBD(WT) (PDB ID 1GUA, Grey and Green) [255] and Rap1A (A30D/K31E):Raf-1 RBD(RK) (PDB ID 3KUC, Cyan and Magenta) [246], aligned to RBD, (left), or Rap1A variants (right). RBD is in green and magenta, whereas Rap1A is shown in grey and Cyan.

5. Conclusions

Here, I have shown how OSPREY [68], a state-of-the-art CPDD platform, can be used to help explore chemical and conformational spaces, both new and unknown, in order to overcome some of the largest biomedical challenges facing modern medicine. I have also provided my insights into what to look for, how to avoid mistakes, and novel paths to move the field forward.

KRas-driven cancers present a huge, unmet medical need, as oncogenic KRas deregulates cellular signaling [62, 85, 95, 98, 115, 159-161, 169, 170, 175, 279, 379, 380]. In general, deregulation of cellular signaling, mostly through PPIs, is near ubiquitous in pathologies and has been historically undruggable [9, 25, 26, 60, 374, 380, 381]. In chapter 2, I describe the design and experimental validation of a RBD variant, RBD(RKY), with the highest known affinity for KRas [246,253,254]. The methods used in this work, although applied to KRas, can be generalized and applied to any PPI. This work required accurate modeling of the intra-atomic interactions at the protein-protein interface and large-scale exploration of the conformational and chemical space of the KRas:RBD PPI. Furthermore, it required the development of four new biochemical assays for the experimental validation of our CPDD predictions. The result is the discovery of novel chemistry in the interface at the V88 hotpot residue. This discovery can inform further optimization of the interface, and the design of small molecule or peptide inhibitors of the KRas:Effector interaction. Moreover, we believe our RBD variant RBD(RKY), can act as a potent inhibitor of the interaction between KRas and its effector proteins. Further work needs to be done to demonstrate this point, including validation in cellular assays. Due to its size, the delivery of RBD may present a challenge. Therefore, co-expression of our RBD variant in KRas driven tumors cells could provide a viable way to test the impact of RBD(RKY) on KRas signaling and tumor cell survival[209]. This exact method in fact, was used to show that KRas did indeed associate with the RBD domain of Raf-1, and that by blocking the KRas:Raf-1 PPI the malignant phenotype of KRas-driven cancer cells could be suppressed [209].

The emergence of drug resistance presents one of the biggest challenges in drug design today [11-14, 16, 18-20, 23, 65, 175, 303, 321, 382] . Despite the expenditure of billions of dollars to develop each novel drug [1], the durability of response in cancer patients is normally measured in months rather than years [11, 13, 14, 175] . Moreover, in the AMR space the development of resistance makes the economics of developing novel therapeutics so restrictive that despite the large medical need most large pharmaceutical companies have stopped investing in this space [383, 384]. In Chapter 3, I describe work done in the Donald lab in collaboration with students of the late Amy Anderson and in the Wright lab at UConn to elucidate the mechanism of resistance to antifolates in Staph and help with the design of resistance resilient drugs. Interestingly, the V31L mutation studied in this work was initially predicted in silico before it was observed either experimentally or clinically [19, 20]. Here, we demonstrate a mechanistic explanation for the development of antifolates that are resilient to the V31L resistance mutation while exploring the role of this mutation in plasmid-borne resistance genes [16, 325]. These plasmid-borne genes are shared between bacteria by conjugation, allowing resistance mutations to spread quickly across many different bacterial strains and species [385]. This increased speed of evolution presents an immense threat in fighting the emergence of AMR [385].

Moreover, we suggest the potential of DHFR to act as a racemase of NADPH. This would allow for a new mechanism of resistance we named chiral evasion. Although a role for NADPH in DHFR resistance had been previously suggested [17, 325, 331], to the best of our knowledge this is the first study to produce a mechanistic explanation of the phenomenon.

Finally, in chapter 4, I present initial observations on factors affecting the accuracy of OSPREY and the role that designers must play in order to optimally employ CPDD tools. Like any tools, the predictive capacity of CPDD tools like OSPREY depends on the users' (designers') ability to bring domain knowledge into the project. It also requires the designer to know when to employ and when not to employ a given tool. Through analysis of the OSPREY 3.0 benchmark [68] and our RBD designs (Chapter 2), the relationship between accuracy and quality of input structure is demonstrated. However, structural quality is one parameter among many that determine the predictive power of CPDD for a design system. An understanding of the dynamics of the system, including the presence of multiple backbone conformations in the thermodynamic ensemble, may also play a large role. A larger dataset is required to fully understand the extent that each factor contributes to the predictive power, or lack thereof, in a CPDD campaign. However, this study provides initial empirical data on these factors and serves as a beachhead for further larger studies to explore these hypotheses.

Furthermore, energy functions are intrinsically vulnerable to certain types of error. For example, the change in energy due to a mutation is often the sum of multiple positive and negative energetic contributions, and even if the energy function's relative error for each contribution is manageable, error propagation analysis suggests the relative error of the sum could be much larger. Although this clearly points towards the need of superior energy functions, the designer can and should apply their domain knowledge to filter results. Notably, it was this combination of CPDD and expert knowledge that allowed us to develop a novel RBD super-binder by experimentally testing only 6 computationally predicted candidate sequences despite a potential chemical space that at times ranged from several hundreds to several millions of possible sequences. Finally, the role of a previously understudied mechanism of backbone flexibility, rigid body translation and rotation, is shown, and its importance to binding and the accuracy of the thermodynamic ensemble is discussed.

"Here be dragons," or so people would supposedly say when facing the unknown [386]. However, like the early explorers, in research it is into the unknown that we must always voyage in an attempt to understand the world around us. Although dragons are not real, superbugs and cancer are real monsters that kill millions of people every year. The exploration of new chemical and conformational space towards the development of novel PPI inhibitors, and the prediction of resistance, requires the development of new tools and best practices. Not unlike the sextants and compasses did for the explorers of the oceans, mathematically provable state-of-the-art CPDD tools like OSPREY present a way for scientist to accurately navigate this vast chemical and conformational space. Here we described a voyage, albeit small, towards ridding the world of monsters.

References

1. Mullard, A., *New drugs cost US$2.6 billion to develop.* Nature Reviews Drug Discovery, 2014. **13**: p. 877.

2. Lander, E.S., et al., *Initial sequencing and analysis of the human genome.* Nature, 2001. **409**(6822): p. 860-921.

3. Schwaederle, M., et al., *Impact of Precision Medicine in Diverse Cancers: A Meta-Analysis of Phase II Clinical Trials.* Journal of clinical oncology : official journal of the American Society of Clinical Oncology, 2015. **33**(32): p. 3817-3825.

4. Stegmeier, F., et al., *Targeted Cancer Therapies in the Twenty-First Century: Lessons From Imatinib.* Clinical Pharmacology & Therapeutics, 2010. **87**(5): p. 543-552.

5. Manzoni, C., et al., *Genome, transcriptome and proteome: the rise of omics data and their integration in biomedical sciences.* Brief Bioinform, 2018. **19**(2): p. 286-302.

6. Boyiadzis, M.M., et al., *Significance and implications of FDA approval of pembrolizumab for biomarker-defined disease.* Journal for ImmunoTherapy of Cancer, 2018. **6**(1): p. 35.

7. Verdine, G.L. and L.D. Walensky, *The Challenge of Drugging Undruggable Targets in Cancer: Lessons Learned from Targeting BCL-2 Family Members.* Clinical Cancer Research, 2007. **13**(24): p. 7264-7270.

8. Crews, C.M., *Targeting the undruggable proteome: the small molecules of my dreams.* Chem Biol, 2010. **17**(6): p. 551-5.

9. Makley, L.N. and J.E. Gestwicki, *Expanding the Number of 'Druggable' Targets: Non-Enzymes and Protein–Protein Interactions.* Chemical Biology & Drug Design, 2013. **81**(1): p. 22-32.

10. Overington, J.P., B. Al-Lazikani, and A.L. Hopkins, *How many drug targets are there?* Nature Reviews Drug Discovery, 2006. **5**(12): p. 993-996.

11. Housman, G., et al., *Drug Resistance in Cancer: An Overview.* Cancers, 2014. **6**(3): p. 1769-1792.

12. Walsh, C., *Molecular mechanisms that confer antibacterial drug resistance.* Nature, 2000. **406**(6797): p. 775-781.

13. Engelman, J.A. and J. Settleman, *Acquired resistance to tyrosine kinase inhibitors during cancer therapy.* Current opinion in genetics & development, 2008. **18**(1): p. 73-79.

14. Engelman, J.A. and P.A. Jänne, *Mechanisms of acquired resistance to epidermal growth factor receptor tyrosine kinase inhibitors in non–small cell lung cancer.* Clinical Cancer Research, 2008. **14**(10): p. 2895-2899.

15. Shah, N.P., et al., *Multiple BCR-ABL kinase domain mutations confer polyclonal resistance to the tyrosine kinase inhibitor imatinib (STI571) in chronic phase and blast crisis chronic myeloid leukemia.* Cancer cell, 2002. **2**(2): p. 117-125.

16. Reeve, S.M., et al., *MRSA Isolates from United States Hospitals Carry dfrG and dfrK Resistance Genes and Succumb to Propargyl-Linked Antifolates.* Cell chemical biology, 2016. **23**(12): p. 1458-1467.

17. Keshipeddy, S., et al., *Nonracemic Antifolates Stereoselectively Recruit Alternate Cofactors and Overcome Resistance in S. aureus.* Journal of the American Chemical Society, 2015. **137**(28): p. 8983-8990.

18. Ojewole, A., et al., *OSPREY Predicts Resistance Mutations Using Positive and Negative Computational Protein Design.* Methods Mol Biol, 2017. **1529**: p. 291-306.

19. Frey, K.M., et al., *Predicting resistance mutations using protein design algorithms.* 2010. **107**(31): p. 13707-13712.

20. Reeve, S.M., et al., *Protein design algorithms predict viable resistance to an experimental antifolate.* 2015. **112**(3): p. 749-754.

21. Liu, J., et al., *Towards new antifolates targeting eukaryotic opportunistic infections.* Eukaryotic cell, 2009. **8**(4): p. 483-486.

22. Kaserer, T. and J. Blagg, *Combining Mutational Signatures, Clonal Fitness, and Drug Affinity to Define Drug-Specific Resistance Mutations in Cancer.* Cell Chem Biol, 2018. **25**(11): p. 1359-1371.e2.

23. Yang, Z., et al., *Investigating Novel Resistance Mechanisms to Third-Generation EGFR Tyrosine Kinase Inhibitor Osimertinib in Non–Small Cell Lung Cancer Patients.* Clinical Cancer Research, 2018. **24**(13): p. 3097-3107.

24. Rognan, D., *Rational design of protein–protein interaction inhibitors.* MedChemComm, 2015. **6**(1): p. 51-60.

25. Modell, A.E., S.L. Blosser, and P.S. Arora, *Systematic Targeting of Protein-Protein Interactions.* Trends in pharmacological sciences, 2016. **37**(8): p. 702-713.

26. Smith, M.C. and J.E. Gestwicki, *Features of protein–protein interactions that translate into potent inhibitors: topology, surface area and affinity.* Expert Reviews in Molecular Medicine, 2012. **14**: p. e16.

27. Cukuroglu, E., et al., *Hot spots in protein–protein interfaces: Towards drug discovery.* Progress in Biophysics and Molecular Biology, 2014. **116**(2): p. 165-173.

28. Zerbe, B.S., et al., *Relationship between hot spot residues and ligand binding hot spots in protein-protein interfaces.* J Chem Inf Model, 2012. **52**(8): p. 2236-44.

29. Wells, J.A. and C.L. McClendon, *Reaching for high-hanging fruit in drug discovery at protein-protein interfaces.* Nature, 2007. **450**(7172): p. 1001-9.

30. Cunningham, B.C. and J.A. Wells, *High-resolution epitope mapping of hGH-receptor interactions by alanine-scanning mutagenesis.* Science, 1989. **244**(4908): p. 1081-1085.

31. Cunningham, B.C. and J.A. Wells, *Comparison of a structural and a functional epitope.* Journal of molecular biology, 1993. **234**(3): p. 554-563.

32. Jin, L. and J.A. Wells, *Dissecting the energetics of an antibody-antigen interface by alanine shaving and molecular grafting.* Protein Science, 1994. **3**(12): p. 2351-2357.

33. Clackson, T. and J.A. Wells, *A hot spot of binding energy in a hormone-receptor interface.* Science, 1995. **267**(5196): p. 383-386.

34. Bogan, A.A. and K.S. Thorn, *Anatomy of hot spots in protein interfaces.* Journal of molecular biology, 1998. **280**(1): p. 1-9.

35. Hartman, G.D., et al., *Non-peptide fibrinogen receptor antagonists. 1. Discovery and design of exosite inhibitors.* Journal of Medicinal Chemistry, 1992. **35**(24): p. 4640-4642.

36. Tse, C., et al., *ABT-263: a potent and orally bioavailable Bcl-2 family inhibitor.* Cancer Res, 2008. **68**(9): p. 3421-8.

37. Wisen, S., et al., *Binding of a small molecule at a protein-protein interface regulates the chaperone activity of hsp70-hsp40.* ACS Chem Biol, 2010. **5**(6): p. 611-22.

38. Benetatos, C.A., et al., *Birinapant (TL32711), a bivalent SMAC mimetic, targets TRAF2-associated cIAPs, abrogates TNF-induced NF-kappaB activation, and is active in patient-derived xenograft models.* Mol Cancer Ther, 2014. **13**(4): p. 867-79.

39. Kenny, C.H., et al., *Development of a fluorescence polarization assay to screen for inhibitors of the FtsZ/ZipA interaction.* Anal Biochem, 2003. **323**(2): p. 224-33.

40. Gosmini, R., et al., *The Discovery of I-BET726 (GSK1324726A), a Potent Tetrahydroquinoline ApoA1 Up-Regulator and Selective BET Bromodomain Inhibitor.* Journal of Medicinal Chemistry, 2014. **57**(19): p. 8111-8131.

41. Ding, Q., et al., *Discovery of RG7388, a Potent and Selective p53–MDM2 Inhibitor in Clinical Development.* Journal of Medicinal Chemistry, 2013. **56**(14): p. 5979-5983.

42. Uvebrant, K., et al., *Discovery of selective small-molecule CD80 inhibitors.* J Biomol Screen, 2007. **12**(4): p. 464-72.

43. Yoakim, C., et al., *Discovery of the first series of inhibitors of human papillomavirus type 11: inhibition of the assembly of the E1-E2-Origin DNA complex.* Bioorg Med Chem Lett, 2003. **13**(15): p. 2539-41.

44. Gehling, V.S., et al., *Discovery, Design, and Optimization of Isoxazole Azepine BET Inhibitors.* ACS medicinal chemistry letters, 2013. **4**(9): p. 835-840.

45. Miyata, Y., et al., *High-throughput screen for Escherichia coli heat shock protein 70 (Hsp70/DnaK): ATPase assay in low volume by exploiting energy transfer.* J Biomol Screen, 2010. **15**(10): p. 1211-9.

46. Malecka, K.A., et al., *Identification and characterization of small molecule human papillomavirus E6 inhibitors.* ACS Chem Biol, 2014. **9**(7): p. 1603-12.

47. Vassilev, L.T., et al., *In vivo activation of the p53 pathway by small-molecule antagonists of MDM2.* Science, 2004. **303**(5659): p. 844-8.

48. Kwon, H.S., et al., *Inhibition of VEGF transcription through blockade of the hypoxia inducible factor-1α–p300 interaction by a small molecule.* Bioorganic & Medicinal Chemistry Letters, 2012. **22**(16): p. 5249-5252.

49. Deng, Y., et al., *Modulating the interaction between CDK2 and cyclin A with a quinoline-based inhibitor.* Bioorg Med Chem Lett, 2014. **24**(1): p. 199-203.

50. Jahagirdar, R., et al., *A novel BET bromodomain inhibitor, RVX-208, shows reduction of atherosclerosis in hyperlipidemic ApoE deficient mice.* Atherosclerosis, 2014. **236**(1): p. 91-100.

51. Czarna, A., et al., *Robust generation of lead compounds for protein-protein interactions by computational and MCR chemistry: p53/Hdm2 antagonists.* Angew Chem Int Ed Engl, 2010. **49**(31): p. 5352-6.

52. von Kleist, L., et al., *Role of the clathrin terminal domain in regulating coated pit dynamics revealed by small molecule inhibition.* Cell, 2011. **146**(3): p. 471-84.

53. Zimmermann, G., et al., *Small molecule inhibition of the KRAS-PDEdelta interaction impairs oncogenic KRAS signalling.* Nature, 2013. **497**(7451): p. 638-42.

54. Moerke, N.J., et al., *Small-molecule inhibition of the interaction between the translation initiation factors eIF4E and eIF4G.* Cell, 2007. **128**(2): p. 257-67.

55. Molzan, M., et al., *Stabilization of Physical RAF/14-3-3 Interaction by Cotylenin A as Treatment Strategy for RAS Mutant Cancers.* ACS Chemical Biology, 2013. **8**(9): p. 1869-1875.

56. Ottmann, C., et al., *A structural rationale for selective stabilization of anti-tumor interactions of 14-3-3 proteins by cotylenin A.* J Mol Biol, 2009. **386**(4): p. 913-9.

57. Mehraein-Ghomi, F., et al., *Targeting androgen receptor and JunD interaction for prevention of prostate cancer progression.* Prostate, 2014. **74**(7): p. 792-803.

58. Erickson, R.I., et al., *Toxicity profile of small-molecule IAP antagonist GDC-0152 is linked to TNF-alpha pharmacology.* Toxicol Sci, 2013. **131**(1): p. 247-58.

59. Gorczynski, M.J., et al., *Allosteric inhibition of the protein-protein interaction between the leukemia-associated proteins Runx1 and CBFbeta.* Chem Biol, 2007. **14**(10): p. 1186-97.

60. Roberts, K.E., et al., *Computational design of a PDZ domain peptide inhibitor that rescues CFTR activity.* PLoS Comput Biol, 2012. **8**(4): p. e1002477.

61. Holden, M.T., et al., *Complete genomes of two clinical Staphylococcus aureus strains: evidence for the rapid evolution of virulence and drug resistance.* Proceedings of the National Academy of Sciences, 2004. **101**(26): p. 9786-9791.

62. Oxnard, G.R., et al., *Assessment of Resistance Mechanisms and Clinical Implications in Patients With EGFR T790M-Positive Lung Cancer and Acquired Resistance to Osimertinib.* JAMA Oncol, 2018. **4**(11): p. 1527-1534.

63. Chellappan, S., et al., *Design of mutation-resistant HIV protease inhibitors with the substrate envelope hypothesis.* Chem Biol Drug Des, 2007. **69**(5): p. 298-313.

64. Nalam, M.N., et al., *Evaluating the substrate-envelope hypothesis: structural analysis of novel HIV-1 protease inhibitors designed to be robust against drug resistance.* J Virol, 2010. **84**(10): p. 5368-78.

65. Nalam, M.N., et al., *Substrate envelope-designed potent HIV-1 protease inhibitors to avoid drug resistance.* Chem Biol, 2013. **20**(9): p. 1116-24.

66. Shen, Y., et al., *Testing the substrate-envelope hypothesis with designed pairs of compounds.* ACS Chem Biol, 2013. **8**(11): p. 2433-41.

67. Reeve, S.M., et al., *Protein design algorithms predict viable resistance to an experimental antifolate.* Proc Natl Acad Sci U S A, 2015. **112**(3): p. 749-54.

68. Hallen, M.A., et al., *OSPREY 3.0: Open-source protein redesign for you, with powerful new features.* J Comput Chem, 2018. **39**(30): p. 2494-2507.

69. Georgiev, I., et al., *Algorithm for backrub motions in protein design.* Bioinformatics, 2008. **24**(13): p. i196-i204.

70. Gainza, P., H.M. Nisonoff, and B.R. Donald, *Algorithms for protein design.* Current Opinion in Structural Biology, 2016. **39**: p. 16-26.

71. Ojewole, A.A., et al., *BBK* (Branch and Bound Over K*): A Provable and Efficient Ensemble-Based Protein Design Algorithm to Optimize Stability and Binding Affinity Over Large Sequence Spaces.* J Comput Biol, 2018. **25**(7): p. 726-739.

72. Hallen, M.A. and B.R. Donald, *CATS (Coordinates of Atoms by Taylor Series): protein design with backbone flexibility in all locally feasible directions.* Bioinformatics, 2017. **33**(14): p. i5-i12.

73. Hallen, M.A., P. Gainza, and B.R. Donald, *Compact Representation of Continuous Energy Surfaces for More Efficient Protein Design.* Journal of chemical theory and computation, 2015. **11**(5): p. 2292-2306.

74. Georgiev, I. and B.R. Donald, *Dead-end elimination with backbone flexibility.* Bioinformatics, 2007. **23**(13): p. i185-94.

75. Hallen, M.A., D.A. Keedy, and B.R. Donald, *Dead-end elimination with perturbations (DEEPer): a provable protein design algorithm with continuous sidechain and backbone flexibility.* Proteins, 2013. **81**(1): p. 18-39.

76. Georgiev, I., R.H. Lilien, and B.R. Donald, *Improved Pruning algorithms and Divide-and-Conquer strategies for Dead-End Elimination, with application to protein design.* Bioinformatics, 2006. **22**(14): p. e174-83.

77. Hallen, M.A., J.D. Jou, and B.R. Donald, *LUTE (Local Unpruned Tuple Expansion): Accurate Continuously Flexible Protein Design with General Energy Functions and Rigid Rotamer-Like Efficiency.* J Comput Biol, 2017. **24**(6): p. 536-546.

78. Georgiev, I., R.H. Lilien, and B.R. Donald, *The minimized dead-end elimination criterion and its application to protein redesign in a hybrid scoring and search algorithm for computing partition functions over molecular ensembles.* Journal of computational chemistry, 2008. **29**(10): p. 1527-1542.

79. Lilien, R.H., et al., *A novel ensemble-based scoring and search algorithm for protein redesign and its application to modify the substrate specificity of the gramicidin synthetase a phenylalanine adenylation enzyme.* J Comput Biol, 2005. **12**(6): p. 740-61.

80. Gainza, P., K.E. Roberts, and B.R. Donald, *Protein design using continuous rotamers.* PLoS Comput Biol, 2012. **8**(1): p. e1002335.

81. Gorczynski, M.J., et al., *Allosteric Inhibition of the Protein-Protein Interaction between the Leukemia-Associated Proteins Runx1 and CBFβ.* Chemistry & Biology, 2007. **14**(10): p. 1186-1197.

82. Rudicell, R.S., et al., *Enhanced Potency of a Broadly Neutralizing HIV-1 Antibody In Vitro Improves Protection against Lentiviral Infection In Vivo.* 2014. **88**(21): p. 12669-12682.

83. Reeve, S.M., et al., *Toward Broad Spectrum DHFR inhibitors Targeting Trimethoprim Resistant Enzymes Identified in Clinical Isolates of Methicillin-Resistant Staphylococcus aureus.* bioRxiv, 2019: p. 648808.

84. Siegel, R.L., K.D. Miller, and A. Jemal, *Cancer statistics, 2015.* CA Cancer J Clin, 2015. **65**(1): p. 5-29.

85. Gremer, L., et al., *Germline KRAS Mutations Cause Aberrant Biochemical and Physical Properties Leading to Developmental Disorders.* Human mutation, 2011. **32**(1): p. 33-43.

86. Friday, B.B. and A.A. Adjei, *K-ras as a target for cancer therapy.* Biochim Biophys Acta, 2005. **1756**(2): p. 127-44.

87. Colicelli, J., *Human RAS superfamily proteins and related GTPases.* Sci. STKE, 2004. **2004**(250): p. re13-re13.

88. Wennerberg, K., K.L. Rossman, and C.J. Der, *The Ras superfamily at a glance.* J cell Sci, 2005. **118**(5): p. 843-846.

89. Rajalingam, K., et al., *Ras oncogenes and their downstream targets.* Biochimica et Biophysica Acta (BBA) - Molecular Cell Research, 2007. **1773**(8): p. 1177-1195.

90. Hall, A., et al., *Identification of transforming gene in two human sarcoma cell lines as a new member of the ras gene family located on chromosome 1.* Nature, 1983. **303**(5916): p. 396-400.

91. Shimizu, K., et al., *Isolation and preliminary characterization of the transforming gene of a human neuroblastoma cell line.* Proceedings of the National Academy of Sciences of the United States of America, 1983. **80**(2): p. 383-387.

92. Henis, Y.I., J.F. Hancock, and I.A. Prior, *Ras acylation, compartmentalization and signaling nanoclusters (Review).* Molecular Membrane Biology, 2009. **26**(1-2): p. 80-92.

93. Rocks, O., A. Peyker, and P.I. Bastiaens, *Spatio-temporal segregation of Ras signals: one ship, three anchors, many harbors.* Current opinion in cell biology, 2006. **18**(4): p. 351-357.

94. Mor, A. and M.R. Philips, *Compartmentalized ras/mapk signaling.* Annu. Rev. Immunol., 2006. **24**: p. 771-800.

95. Eser, S., et al., *Oncogenic KRAS signalling in pancreatic cancer.* Br J Cancer, 2014. **111**(5): p. 817-22.

96. Shih, T.Y., et al., *Structure and function of p21 ras proteins.* Gene Amplif Anal, 1986. **4**: p. 53-72.

97. Boriack-Sjodin, P.A., et al., *The structural basis of the activation of Ras by Sos.* Nature, 1998. **394**(6691): p. 337-43.

98. Deramaudt, T. and A.K. Rustgi, *Mutant KRAS in the initiation of pancreatic cancer.* Biochimica et Biophysica Acta (BBA) - Reviews on Cancer, 2005. **1756**(2): p. 97-101.

99. Wang, Y., et al., *Targeting mutant KRAS for anticancer therapeutics: a review of novel small molecule modulators.* J Med Chem, 2013. **56**(13): p. 5219-30.

100. Pai, E.F., et al., *Structure of the guanine-nucleotide-binding domain of the Ha-ras oncogene product p21 in the triphosphate conformation.* Nature, 1989. **341**(6239): p. 209-14.

101. Pai, E.F., et al., *Refined crystal structure of the triphosphate conformation of H-ras p21 at 1.35 A resolution: implications for the mechanism of GTP hydrolysis.* Embo j, 1990. **9**(8): p. 2351-9.

102. Dykes, D.C., et al., *Activated conformations of the ras-gene-encoded p21 protein. 1. An energy-refined structure for the normal p21 protein complexed with GDP.* J Biomol Struct Dyn, 1992. **9**(6): p. 1025-44.

103. Dykes, D.C., et al., *Activated conformations of the ras-gene-encoded p21 protein. 2. Comparison of the computed and high-resolution x-ray crystallographic structures of Gly-12 p21.* J Biomol Struct Dyn, 1993. **10**(5): p. 905-18.

104. Hunter, J.C., et al., *In situ selectivity profiling and crystal structure of SML-8-73-1, an active site inhibitor of oncogenic K-Ras G12C.* Proc Natl Acad Sci U S A, 2014. **111**(24): p. 8895-900.

105. Sharma, N., U. Sonavane, and R. Joshi, *Probing the wild-type HRas activation mechanism using steered molecular dynamics, understanding the energy barrier and role of water in the activation.* Eur Biophys J, 2014. **43**(2-3): p. 81-95.

106. Nassar, N., et al., *The 2.2 A crystal structure of the Ras-binding domain of the serine/threonine kinase c-Raf1 in complex with Rap1A and a GTP analogue.* Nature, 1995. **375**(6532): p. 554-60.

107. Huang, L., et al., *Structural basis for the interaction of Ras with RalGDS.* Nat Struct Biol, 1998. **5**(6): p. 422-6.

108. Pacold, M.E., et al., *Crystal structure and functional analysis of Ras binding to its effector phosphoinositide 3-kinase gamma.* Cell, 2000. **103**(6): p. 931-43.

109. Johnson, C.W. and C. Mattos, *Chapter Three - The Allosteric Switch and Conformational States in Ras GTPase Affected by Small Molecules*, in *The Enzymes*, T. Fuyuhiko, Editor. 2013, Academic Press. p. 41-67.

110. Hanahan, D. and R.A. Weinberg, *The hallmarks of cancer.* Cell, 2000. **100**(1): p. 57-70.

111. Hanahan, D. and R.A. Weinberg, *Hallmarks of cancer: the next generation.* Cell, 2011. **144**(5): p. 646-74.

112. Whiteside, T.L., *The role of immune cells in the tumor microenvironment.* Cancer Treat Res, 2006. **130**: p. 103-24.

113. Orlando, E., et al., *Oncogene addiction as a foundation of targeted cancer therapy: The paradigm of the MET receptor tyrosine kinase.* Cancer Lett, 2019. **443**: p. 189-202.

114. Reddy, E.P., et al., *A point mutation is responsible for the acquisition of transforming properties by the T24 human bladder carcinoma oncogene.* Nature, 1982. **300**(5888): p. 149-52.

115. Okamoto, K., et al., *Reversal of Mutant KRAS-Mediated Apoptosis Resistance by Concurrent Noxa/Bik Induction and Bcl-2/Bcl-xL Antagonism in Colon Cancer Cells.* Molecular cancer research : MCR, 2015. **13**(4): p. 659-669.

116. Zeng, M., et al., *Hypoxia activates the K-ras proto-oncogene to stimulate angiogenesis and inhibit apoptosis in colon cancer cells.* PLoS One, 2010. **5**(6): p. e10966.

117. Wang, M.-T., et al., *K-Ras Promotes Tumorigenicity through Suppression of Non-canonical Wnt Signaling.* Cell, 2015. **163**(5): p. 1237-1251.

118. Feramisco, J.R., et al., *Microinjection of the oncogene form of the human H-ras (T-24) protein results in rapid proliferation of quiescent cells.* Cell, 1984. **38**(1): p. 109-17.

119. Stacey, D.W. and H.F. Kung, *Transformation of NIH 3T3 cells by microinjection of Ha-ras p21 protein.* Nature, 1984. **310**(5977): p. 508-11.

120. McCarthy, S.A., et al., *Rapid induction of heparin-binding epidermal growth factor/diphtheria toxin receptor expression by Raf and Ras oncogenes.* Genes Dev, 1995. **9**(16): p. 1953-64.

121. Gangarosa, L.M., et al., *A raf-independent epidermal growth factor receptor autocrine loop is necessary for Ras transformation of rat intestinal epithelial cells.* J Biol Chem, 1997. **272**(30): p. 18926-31.

122. Woods, D., et al., *Induction of beta3-integrin gene expression by sustained activation of the Ras-regulated Raf-MEK-extracellular signal-regulated kinase signaling pathway.* Mol Cell Biol, 2001. **21**(9): p. 3192-205.

123. Stacey, D.W., et al., *Microinjection of transforming ras protein induces c-fos expression.* Mol Cell Biol, 1987. **7**(1): p. 523-7.

124. Gutman, A., C. Wasylyk, and B. Wasylyk, *Cell-specific regulation of oncogene-responsive sequences of the c-fos promoter.* Mol Cell Biol, 1991. **11**(10): p. 5381-7.

125. Westwick, J.K., et al., *Oncogenic Ras activates c-Jun via a separate pathway from the activation of extracellular signal-regulated kinases.* 1994. **91**(13): p. 6030-6034.

126. Finco, T.S., et al., *Oncogenic Ha-Ras-induced signaling activates NF-kappaB transcriptional activity, which is required for cellular transformation.* J Biol Chem, 1997. **272**(39): p. 24113-6.

127. Filmus, J., J. Zhao, and R.N. Buick, *Overexpression of H-ras oncogene induces resistance to the growth-inhibitory action of transforming growth factor beta-1 (TGF-beta 1) and alters the number and type of TGF-beta 1 receptors in rat intestinal epithelial cell clones.* Oncogene, 1992. **7**(3): p. 521-6.

128. Zhao, J. and R.N. Buick, *Regulation of Transforming Growth Factor β Receptors in H-ras Oncogene-transformed Rat Intestinal Epithelial Cells.* 1995. **55**(24): p. 6181-6188.

129. Kretzschmar, M., et al., *A mechanism of repression of TGFbeta/ Smad signaling by oncogenic Ras.* Genes Dev, 1999. **13**(7): p. 804-16.

130. Kfir, S., et al., *Pathway- and expression level-dependent effects of oncogenic N-Ras: p27(Kip1) mislocalization by the Ral-GEF pathway and Erk-mediated interference with Smad signaling.* Mol Cell Biol, 2005. **25**(18): p. 8239-50.

131. Daly, A.C., P. Vizan, and C.S. Hill, *Smad3 protein levels are modulated by Ras activity and during the cell cycle to dictate transforming growth factor-beta responses.* J Biol Chem, 2010. **285**(9): p. 6489-97.

132. Rivard, N., et al., *MAP kinase cascade is required for p27 downregulation and S phase entry in fibroblasts and epithelial cells.* Am J Physiol, 1999. **277**(4): p. C652-64.

133. Filmus, J., et al., *Induction of cyclin D1 overexpression by activated ras.* Oncogene, 1994. **9**(12): p. 3627-33.

134. Albanese, C., et al., *Transforming p21ras mutants and c-Ets-2 activate the cyclin D1 promoter through distinguishable regions.* J Biol Chem, 1995. **270**(40): p. 23589-97.

135. Gille, H. and J. Downward, *Multiple ras effector pathways contribute to G(1) cell cycle progression.* J Biol Chem, 1999. **274**(31): p. 22033-40.

136. Diehl, J.A., et al., *Glycogen synthase kinase-3beta regulates cyclin D1 proteolysis and subcellular localization.* Genes Dev, 1998. **12**(22): p. 3499-511.

137. Leone, G., et al., *Myc and Ras collaborate in inducing accumulation of active cyclin E/Cdk2 and E2F.* Nature, 1997. **387**(6631): p. 422-6.

138. Matsuo, Y., et al., *K-Ras promotes angiogenesis mediated by immortalized human pancreatic epithelial cells through mitogen-activated protein kinase signaling pathways.* Molecular cancer research : MCR, 2009. **7**(6): p. 799-808.

139. Kranenburg, O., M.F. Gebbink, and E.E. Voest, *Stimulation of angiogenesis by Ras proteins.* Biochim Biophys Acta, 2004. **1654**(1): p. 23-37.

140. Okada, F., et al., *Impact of oncogenes in tumor angiogenesis: Mutant K-ras up-regulation of vascular endothelial growth factor/vascular permeability factor is necessary, but not sufficient for tumorigenicity of human colorectal carcinoma cells.* 1998. **95**(7): p. 3609-3614.

141. Rak, J., et al., *Mutant ras Oncogenes Upregulate VEGF/VPF Expression: Implications for Induction and Inhibition of Tumor Angiogenesis.* 1995. **55**(20): p. 4575-4580.

142. Zhang, X., J.P. Gaspard, and D.C. Chung, *Regulation of Vascular Endothelial Growth Factor by the Wnt and K-ras Pathways in Colonic Neoplasia.* 2001. **61**(16): p. 6050-6054.

143. Matsuo, Y., et al., *K-Ras Promotes Angiogenesis Mediated by Immortalized Human Pancreatic Epithelial Cells through Mitogen-Activated Protein Kinase Signaling Pathways.* 2009. **7**(6): p. 799-808.

144. Sunaga, N., et al., *Oncogenic KRAS-induced interleukin-8 overexpression promotes cell growth and migration and contributes to aggressive phenotypes of non-small cell lung cancer.* Int J Cancer, 2012. **130**(8): p. 1733-44.

145. Xu, L., P.S. Pathak, and D. Fukumura, *Hypoxia-Induced Activation of p38 Mitogen-Activated Protein Kinase and Phosphatidylinositol 3′-Kinase Signaling Pathways Contributes to Expression of Interleukin 8 in Human Ovarian Carcinoma Cells.* 2004. **10**(2): p. 701-707.

146. Sparmann, A. and D. Bar-Sagi, *Ras-induced interleukin-8 expression plays a critical role in tumor growth and angiogenesis.* Cancer Cell, 2004. **6**(5): p. 447-58.

147. Dias Carvalho, P., et al., *KRAS Oncogenic Signaling Extends beyond Cancer Cells to Orchestrate the Microenvironment*. 2018. **78**(1): p. 7-14.

148. Chang, S.H., et al., *T helper 17 cells play a critical pathogenic role in lung cancer.* 2014. **111**(15): p. 5664-5669.

149. McAllister, F., et al., *Oncogenic Kras activates a hematopoietic-to-epithelial IL-17 signaling axis in preinvasive pancreatic neoplasia.* Cancer Cell, 2014. **25**(5): p. 621-37.

150. Fujimura, K., et al., *Eukaryotic Translation Initiation Factor 5A (EIF5A) Regulates Pancreatic Cancer Metastasis by Modulating RhoA and Rho-associated Kinase (ROCK) Protein Expression Levels.* 2015. **290**(50): p. 29907-29919.

151. Rath, N., et al., *ROCK signaling promotes collagen remodeling to facilitate invasive pancreatic ductal adenocarcinoma tumor cell growth.* 2017. **9**(2): p. 198-218.

152. Kalluri, R., *The biology and function of fibroblasts in cancer.* Nature Reviews Cancer, 2016. **16**: p. 582.

153. Wever Olivier, D. and M. Mareel, *Role of Myofibroblasts at the Invasion Front*, in *Biological Chemistry*. 2002. p. 55.

154. Ji, Z., et al., *Oncogenic KRAS Activates Hedgehog Signaling Pathway in Pancreatic Cancer Cells.* 2007. **282**(19): p. 14048-14055.

155. Pasca di Magliano, M., et al., *Hedgehog/Ras interactions regulate early stages of pancreatic cancer.* 2006. **20**(22): p. 3161-3173.

156. Fendrich, V., et al., *Ectopic overexpression of Sonic Hedgehog (Shh) induces stromal expansion and metaplasia in the adult murine pancreas.* Neoplasia, 2011. **13**(10): p. 923-30.

157. Hoshino, A., et al., *Tumour exosome integrins determine organotropic metastasis.* Nature, 2015. **527**(7578): p. 329-35.

158. Demory Beckler, M., et al., *Proteomic Analysis of Exosomes from Mutant KRAS Colon Cancer Cells Identifies Intercellular Transfer of Mutant KRAS.* 2013. **12**(2): p. 343-355.

159. Hutton, J.E., et al., *Oncogenic KRAS and BRAF Drive Metabolic Reprogramming in Colorectal Cancer.* Mol Cell Proteomics, 2016. **15**(9): p. 2924-38.

160. Son, J., et al., *Glutamine supports pancreatic cancer growth through a KRAS-regulated metabolic pathway.* Nature, 2013. **496**(7443): p. 101-5.

161. Bryant, K.L., et al., *KRAS: feeding pancreatic cancer proliferation.* Trends Biochem Sci, 2014. **39**(2): p. 91-100.

162. Halazonetis, T.D., V.G. Gorgoulis, and J. Bartek, *An oncogene-induced DNA damage model for cancer development.* Science, 2008. **319**(5868): p. 1352-5.

163. Denko, N.C., et al., *The human Ha-ras oncogene induces genomic instability in murine fibroblasts within one cell cycle.* Proc Natl Acad Sci U S A, 1994. **91**(11): p. 5124-8.

164. Bayne, Lauren J., et al., *Tumor-Derived Granulocyte-Macrophage Colony-Stimulating Factor Regulates Myeloid Inflammation and T Cell Immunity in Pancreatic Cancer.* Cancer Cell, 2012. **21**(6): p. 822-835.

165. Tran, E., et al., *T-Cell Transfer Therapy Targeting Mutant KRAS in Cancer.* N Engl J Med, 2016. **375**(23): p. 2255-2262.

166. Testorelli, C., et al., *Dacarbazine-induced immunogenicity of a murine leukemia is attenuated in cells transfected with mutated K-ras gene.* J Exp Clin Cancer Res, 1997. **16**(1): p. 15-22.

167. Zdanov, S., et al., *Mutant KRAS Conversion of Conventional T Cells into Regulatory T Cells.* 2016. **4**(4): p. 354-365.

168. Blume-Jensen, P. and T. Hunter, *Oncogenic kinase signalling.* Nature, 2001. **411**(6835): p. 355-65.

169. Ying, H., et al., *Oncogenic Kras maintains pancreatic tumors through regulation of anabolic glucose metabolism.* Cell, 2012. **149**(3): p. 656-70.

170. Collins, M.A., et al., *Oncogenic Kras is required for both the initiation and maintenance of pancreatic cancer in mice.* J Clin Invest, 2012. **122**(2): p. 639-53.

171. Chu, G.C., et al., *Stromal biology of pancreatic cancer.* Journal of Cellular Biochemistry, 2007. **101**(4): p. 887-907.

172. Korc, M., *Pancreatic cancer–associated stroma production.* The American Journal of Surgery, 2007. **194**(4, Supplement): p. S84-S86.

173. Räsänen, K. and A. Vaheri, *Activation of fibroblasts in cancer stroma.* Experimental Cell Research, 2010. **316**(17): p. 2713-2722.

174. Karapetis, C.S., et al., *K-ras Mutations and Benefit from Cetuximab in Advanced Colorectal Cancer.* New England Journal of Medicine, 2008. **359**(17): p. 1757-1765.

175. Sforza, V., et al., *Mechanisms of resistance to anti-epidermal growth factor receptor inhibitors in metastatic colorectal cancer.* World journal of gastroenterology, 2016. **22**(28): p. 6345-6361.

176. Pagliarini, R., W. Shao, and W.R. Sellers, *Oncogene addiction: pathways of therapeutic response, resistance, and road maps toward a cure.* EMBO reports, 2015. **16**(3): p. 280-296.

177. Collins, M.A., et al., *Oncogenic Kras is required for both the initiation and maintenance of pancreatic cancer in mice.* The Journal of Clinical Investigation, 2012. **122**(2): p. 639-653.

178. Singh, A., et al., *A gene expression signature associated with "K-Ras addiction" reveals regulators of EMT and tumor cell survival.* Cancer cell, 2009. **15**(6): p. 489-500.

179. Janes, M.R., et al., *Targeting KRAS Mutant Cancers with a Covalent G12C-Specific Inhibitor.* Cell, 2018. **172**(3): p. 578-589.e17.

180. Vartanian, S., et al., *Identification of mutant K-Ras-dependent phenotypes using a panel of isogenic cell lines.* Journal of Biological Chemistry, 2013. **288**(4): p. 2403-2413.

181. Fujita-Sato, S., et al., *Enhanced MET translation and signaling sustains K-Ras–driven proliferation under anchorage-independent growth conditions.* Cancer research, 2015. **75**(14): p. 2851-2862.

182. Patricelli, M.P., et al., *Selective inhibition of oncogenic KRAS output with small molecules targeting the inactive state.* Cancer discovery, 2016. **6**(3): p. 316-329.

183. McCormick, F., *Targeting KRAS Directly.* 2018. **2**(1): p. 81-90.

184. Mott, H.R. and D. Owen, *Structures of Ras superfamily effector complexes: What have we learnt in two decades?* Critical Reviews in Biochemistry and Molecular Biology, 2015. **50**(2): p. 85-133.

185. Wang, W., G. Fang, and J. Rudolph, *Ras inhibition via direct Ras binding—is there a path forward?* Bioorganic & Medicinal Chemistry Letters, 2012. **22**(18): p. 5766-5776.

186. Ostrem, J.M., et al., *K-Ras(G12C) inhibitors allosterically control GTP affinity and effector interactions.* Nature, 2013. **503**(7477): p. 548-51.

187. Maurer, T., et al., *Small-molecule ligands bind to a distinct pocket in Ras and inhibit SOS-mediated nucleotide exchange activity.* Proc Natl Acad Sci U S A, 2012. **109**(14): p. 5299-304.

188. Hancock, J.F., et al., *All ras proteins are polyisoprenylated but only some are palmitoylated.* Cell, 1989. **57**(7): p. 1167-1177.

189. Willumsen, B.M., et al., *The p21 ras C-terminus is required for transformation and membrane association.* Nature, 1984. **310**(5978): p. 583-586.

190. Cox, A.D., C.J. Der, and M.R. Philips, *Targeting RAS Membrane Association: Back to the Future for Anti-RAS Drug Discovery?* 2015. **21**(8): p. 1819-1827.

191. Fu, H.W. and P.J. Casey, *Enzymology and biology of CaaX protein prenylation.* Recent Prog Horm Res, 1999. **54**: p. 315-42; discussion 342-3.

192. Sebti, S.M., et al., *Dual farnesyl and geranylgeranyl transferase inhibitor thwarts mutant KRAS-driven patient-derived pancreatic tumors.* Clinical Cancer Research, 2019: p. clincanres.3399.2018.

193. Wojtkowiak, J.W., R.A. Gibbs, and R.R. Mattingly, *Working together: Farnesyl transferase inhibitors and statins block protein prenylation.* Mol Cell Pharmacol, 2009. **1**(1): p. 1-6.

194. Wang, J., X. Yao, and J. Huang, *New tricks for human farnesyltransferase inhibitor: cancer and beyond.* MedChemComm, 2017. **8**(5): p. 841-854.

195. Zimmermann, G., et al., *Small molecule inhibition of the KRAS–PDEδ interaction impairs oncogenic KRAS signalling.* Nature, 2013. **497**: p. 638.

196. Chandra, A., et al., *The GDI-like solubilizing factor PDEδ sustains the spatial organization and signalling of Ras family proteins.* Nature Cell Biology, 2011. **14**: p. 148.

197. Nan, X., et al., *Ras-GTP dimers activate the Mitogen-Activated Protein Kinase (MAPK) pathway.* 2015. **112**(26): p. 7996-8001.

198. Spencer-Smith, R., et al., *Inhibition of RAS function through targeting an allosteric regulatory site.* Nature Chemical Biology, 2016. **13**: p. 62.

199. Buhrman, G., et al., *Analysis of Binding Site Hot Spots on the Surface of Ras GTPase.* Journal of Molecular Biology, 2011. **413**(4): p. 773-789.

200. Iuga, A., et al., *Solid-state 31P NMR Spectroscopy of Microcrystals of the Ras Protein and its Effector Loop Mutants: Comparison Between Crystalline and Solution State.* Journal of Molecular Biology, 2004. **342**(3): p. 1033-1040.

201. Geyer, M., et al., *Conformational Transitions in p21ras and in Its Complexes with the Effector Protein Raf-RBD and the GTPase Activating Protein GAP.* Biochemistry, 1996. **35**(32): p. 10308-10320.

202. Rosnizeck, I.C., et al., *Metal–Bis(2-picolyl)amine Complexes as State 1(T) Inhibitors of Activated Ras Protein.* Angewandte Chemie International Edition, 2012. **51**(42): p. 10647-10651.

203. Hocker, H.J., et al., *Andrographolide derivatives inhibit guanine nucleotide exchange and abrogate oncogenic Ras function.* Proceedings of the National Academy of Sciences of the United States of America, 2013. **110**(25): p. 10201-10206.

204. Sun, Q., et al., *Discovery of Small Molecules that Bind to K-Ras and Inhibit Sos-Mediated Activation.* 2012. **51**(25): p. 6140-6143.

205. Welsch, M.E., et al., *Multivalent Small-Molecule Pan-RAS Inhibitors.* Cell, 2017. **168**(5): p. 878-889.e29.

206. Repasky, M.P., M. Shelley, and R.A. Friesner, *Flexible ligand docking with Glide.* Curr Protoc Bioinformatics, 2007. **Chapter 8**: p. Unit 8.12.

207. Ostrem, J.M., et al., *K-Ras(G12C) inhibitors allosterically control GTP affinity and effector interactions.* Nature, 2013. **503**: p. 548.

208. Tanaka, T., R.L. Williams, and T.H. Rabbitts, *Tumour prevention by a single antibody domain targeting the interaction of signal transduction proteins with RAS.* The EMBO Journal, 2007. **26**(13): p. 3250-3259.

209. Fridman, M., et al., *The minimal fragments of c-Raf-1 and NF1 that can suppress v-Ha-Ras-induced malignant phenotype.* Journal of Biological Chemistry, 1994. **269**(48): p. 30105-30108.

210. Upadhyaya, P., et al., *Inhibition of Ras Signaling by Blocking Ras–Effector Interactions with Cyclic Peptides.* 2015. **54**(26): p. 7602-7606.

211. Cetin, M., et al., *RasIns: Genetically Encoded Intrabodies of Activated Ras Proteins.* Journal of Molecular Biology, 2017. **429**(4): p. 562-573.

212. Guillard, S., et al., *Structural and functional characterization of a DARPin which inhibits Ras nucleotide exchange.* Nature Communications, 2017. **8**: p. 16111.

213. Boder, E.T. and K.D. Wittrup, *Yeast surface display for screening combinatorial polypeptide libraries.* Nature Biotechnology, 1997. **15**(6): p. 553-557.

214. Zeng, J., et al., *Design of inhibitors of Ras--Raf interaction using a computational combinatorial algorithm.* Protein Eng, 2001. **14**(1): p. 39-45.

215. Pincus, M.R., *Development of new anti-cancer peptides from conformational energy analysis of the oncogenic ras-p21 protein and its complexes with target proteins.* Front Biosci, 2004. **9**: p. 3486-509.

216. Bowne, W.B., et al., *Novel peptides from the RAS-p21 and p53 proteins for the treatment of cancer.* Cancer Ther, 2007. **5b**: p. 331-344.

217. Adler, V., et al., *Two peptides derived from ras-p21 induce either phenotypic reversion or tumor cell necrosis of ras-transformed human cancer cells.* Cancer Chemother Pharmacol, 2008. **62**(3): p. 491-8.

218. Patgiri, A., et al., *An orthosteric inhibitor of the Ras-Sos interaction.* Nat Chem Biol, 2011. **7**(9): p. 585-7.

219. Nickerson, S., et al., *An orthosteric inhibitor of the RAS-SOS interaction.* Enzymes, 2013. **34 Pt. B**: p. 25-39.

220. Gainza, P., et al., *OSPREY: Protein Design with Ensembles, Flexibility, and Provable Algorithms.* Methods in enzymology, 2013. **523**: p. 87-107.

221. Samish, I., et al., *Theoretical and computational protein design.* Annu Rev Phys Chem, 2011. **62**: p. 129-49.

222. Rudicell, R.S., et al., *Enhanced potency of a broadly neutralizing HIV-1 antibody in vitro improves protection against lentiviral infection in vivo.* J Virol, 2014. **88**(21): p. 12669-82.

223. Gainza, P., et al., *OSPREY: protein design with ensembles, flexibility, and provable algorithms.* Methods Enzymol, 2013. **523**: p. 87-107.

224. Donald, B.R., *Algorithms in Structural Molecular Biology*. 2011: The MIT Press. 464.

225. Simoncini, D., et al., *Guaranteed Discrete Energy Optimization on Large Protein Design Problems.* J Chem Theory Comput, 2015. **11**(12): p. 5980-9.

226. Hallen, M.A. and B.R. Donald, *Protein Design by Algorithm.* eprint arXiv:1806.06064, 2018: p. arXiv:1806.06064.

227. Kuhlman, B. and D. Baker, *Native protein sequences are close to optimal for their structures.* Proc Natl Acad Sci U S A, 2000. **97**(19): p. 10383-8.

228. Leaver-Fay, A., et al., *ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules.* Methods Enzymol, 2011. **487**: p. 545-74.

229. Lee, C. and S. Subbiah, *Prediction of protein side-chain conformation by packing optimization.* J Mol Biol, 1991. **217**(2): p. 373-88.

230. Lovell, S.C., et al., *The penultimate rotamer library.* Proteins, 2000. **40**(3): p. 389-408.

231. Tzeng, S.-R. and C.G. Kalodimos, *Protein activity regulation by conformational entropy.* Nature, 2012. **488**: p. 236.

232. Gilson, M.K., et al., *The statistical-thermodynamic basis for computation of binding affinities: a critical review.* Biophysical journal, 1997. **72**(3): p. 1047-1069.

233. Chen, C.Y., et al., *Computational structure-based redesign of enzyme activity.* Proc Natl Acad Sci U S A, 2009. **106**(10): p. 3764-9.

234. Sciretti, D., et al., *Computational protein design with side-chain conformational entropy.* Proteins, 2009. **74**(1): p. 176-191.

235. Dahiyat, B.I. and S.L. Mayo, *De novo protein design: fully automated sequence selection.* Science, 1997. **278**(5335): p. 82-7.

236. Leach, A.R. and A.P. Lemon, *Exploring the conformational space of protein side chains using dead-end elimination and the A* algorithm.* Proteins, 1998. **33**(2): p. 227-39.

237. Traore, S., et al., *A new framework for computational protein design through cost function network optimization.* Bioinformatics, 2013. **29**(17): p. 2129-36.

238. Chazelle, B., C. Kingsford, and M. Singh, *A Semidefinite Programming Approach to Side Chain Positioning with New Rounding Strategies.* 2004. **16**(4): p. 380-392.

239. Chen, C.-Y., et al., *Computational structure-based redesign of enzyme activity.* 2009. **106**(10): p. 3764-3769.

240. Viricel, C., et al. *Guaranteed Weighted Counting for Affinity Computation: Beyond Determinism and Structure*. 2016. Cham: Springer International Publishing.

241. Traore, S., et al., *Deterministic Search Methods for Computational Protein Design.* Methods Mol Biol, 2017. **1529**: p. 107-123.

242. Traore, S., et al., *Fast search algorithms for computational protein design.* J Comput Chem, 2016. **37**(12): p. 1048-58.

243. Stevens, B.W., et al., *Redesigning the PheA Domain of Gramicidin Synthetase Leads to a New Understanding of the Enzyme's Mechanism and Selectivity.* Biochemistry, 2006. **45**(51): p. 15495-15504.

244. Georgiev, I., et al., *Design of epitope-specific probes for sera analysis and antibody isolation.* Retrovirology, 2012. **9**(Suppl 2): p. P50-P50.

245. Georgiev, I.S., et al., *Antibodies VRC01 and 10E8 neutralize HIV-1 with high breadth and potency even with Ig-framework regions substantially reverted to germline.* J Immunol, 2014. **192**(3): p. 1100-1106.

246. Daniel Filchtinski, O.S., Alma Rüppel, Ingrid R. Vetter, Christian Herrmann, Julia M. Shifman, *What Makes Ras an Efficient Molecular Switch: A Computational, Biophysical, and Structural Study of Ras-GDP Interactions with Mutants of Raf.* Journal of Molecular Biology, 2010. **399**(3): p. 422 - 435.

247. Ostrem, J.M., et al., *K-Ras(G12C) inhibitors allosterically control GTP affinity and effector interactions.* Nature, 2013. **503**(7477): p. 548-551.

248. Lito, P., et al., *Allele-specific inhibitors inactivate mutant KRAS G12C by a trapping mechanism.* 2016. **351**(6273): p. 604-608.

249. Zeng, M., et al., *Potent and Selective Covalent Quinazoline Inhibitors of KRAS G12C.* Cell Chemical Biology, 2017. **24**(8): p. 1005-1016.e3.

250. Hunter, J.C., et al., *In situ selectivity profiling and crystal structure of SML-8-73-1, an active site inhibitor of oncogenic K-Ras G12C.* 2014. **111**(24): p. 8895-8900.

251. Patricelli, M.P., et al., *Selective Inhibition of Oncogenic KRAS Output with Small Molecules Targeting the Inactive State.* 2016. **6**(3): p. 316-329.

252. Lanman, B.A., et al., *Abstract 4455: Discovery of AMG 510, a first-in-human covalent inhibitor of KRAS<sup>G12C</sup> for the treatment of solid tumors.* 2019. **79**(13 Supplement): p. 4455-4455.

253. Fridman, M., et al., *Point Mutants of c-Raf-1 RBD with Elevated Binding to v-Ha-Ras.* 2000. **275**(39): p. 30363-30371.

254. Kiel, C., et al., *Improved binding of raf to Ras.GDP is correlated with biological activity.* The Journal of biological chemistry, 2009. **284**(46): p. 31893-31902.

255. Nassar, N., et al., *Ras/Rap effector specificity determined by charge reversal.* Nat Struct Biol, 1996. **3**(8): p. 723-9.

256. L., D.W., *The PyMOL Molecular Graphics System*. 2008.

257. Pearlman, D.A., et al., *AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules.* Computer Physics Communications, 1995. **91**(1): p. 1-41.

258. Roberts, K.E. and B.R. Donald, *Protein Interaction Viewer*. 2014.

259. Word, J.M., et al., *Visualizing and quantifying molecular goodness-of-fit: small-probe contact dots with explicit hydrogen atoms.* J Mol Biol, 1999. **285**(4): p. 1711-33.

260. Flex, E., et al., *Activating mutations in RRAS underlie a phenotype within the RASopathy spectrum and contribute to leukaemogenesis.* Hum Mol Genet, 2014. **23**(16): p. 4315-27.

261. Gremer, L., et al., *Duplication of Glu37 in the switch I region of HRAS impairs effector/GAP binding and underlies Costello syndrome by promoting enhanced growth factor-dependent MAPK and AKT activation.* Hum Mol Genet, 2010. **19**(5): p. 790-802.

262. Gremer, L., et al., *Germline KRAS mutations cause aberrant biochemical and physical properties leading to developmental disorders.* Hum Mutat, 2011. **32**(1): p. 33-43.

263. Nakhaeizadeh, H., et al., *The RAS-Effector Interface: Isoform-Specific Differences in the Effector Binding Regions.* PloS one, 2016. **11**(12): p. e0167145-e0167145.

264. Sydor, J.R., et al., *Cell-free synthesis of the Ras-binding domain of c-Raf-1: binding studies to fluorescently labelled H-ras.* FEBS Lett, 1999. **452**(3): p. 375-8.

265. Herrmann, C., et al., *Differential interaction of the ras family GTP-binding proteins H-Ras, Rap1A, and R-Ras with the putative effector molecules Raf kinase and Ral-guanine nucleotide exchange factor.* J Biol Chem, 1996. **271**(12): p. 6794-800.

266. Vetter, I.R., et al., *Structural and biochemical analysis of Ras-effector signaling via RalGDS.* FEBS Lett, 1999. **451**(2): p. 175-80.

267. Herrmann, C., G.A. Martin, and A. Wittinghofer, *Quantitative Analysis of the Complex between p21 and the Ras-binding Domain of the Human Raf-1 Protein Kinase.* 1995. **270**(7): p. 2901-2905.

268. Linnemann, T., et al., *Thermodynamic and kinetic characterization of the interaction between the Ras binding domain of AF6 and members of the Ras subfamily.* J Biol Chem, 1999. **274**(19): p. 13556-62.

269. Sydor, J.R., et al., *Transient kinetic studies on the interaction of Ras and the Ras-binding domain of c-Raf-1 reveal rapid equilibration of the complex.* Biochemistry, 1998. **37**(40): p. 14292-9.

270. Stieglitz, B., et al., *Novel type of Ras effector interaction established between tumour suppressor NORE1A and Ras switch II.* Embo j, 2008. **27**(14): p. 1995-2005.

271. Lakshman, B., et al., *Quantitative biophysical analysis defines key components modulating recruitment of the GTPase KRAS to the plasma membrane.* 2018.

272. Cruz-Migoni, A., et al., *Structure-based development of new RAS-effector inhibitors from a combination of active and inactive RAS-binding compounds.* 2019. **116**(7): p. 2545-2550.

273. Sakamoto, K., et al., *K-Ras(G12D)-selective inhibitory peptides generated by random peptide T7 phage display technology.* Biochem Biophys Res Commun, 2017. **484**(3): p. 605-611.

274. Block, C., et al., *Quantitative structure-activity analysis correlating Ras/Raf interaction in vitro to Raf activation in vivo.* Nat Struct Biol, 1996. **3**(3): p. 244-51.

275. Campbell-Valois, F.X., K. Tarassov, and S.W. Michnick, *Massive Sequence Perturbation of the Raf ras Binding Domain Reveals Relationships between Sequence Conservation, Secondary Structure Propensity, Hydrophobic Core Organization and Stability.* Journal of Molecular Biology, 2006. **362**(1): p. 151-171.

276. Fridman, M., et al., *c-Raf-1 RBD associates with a subset of active v-H-Ras.* Biochemistry, 2000. **39**(50): p. 15603-11.

277. Fetics, Susan K., et al., *Allosteric Effects of the Oncogenic RasQ61L Mutant on Raf-RBD.* Structure, 2015. **23**(3): p. 505-516.

278. Gorman, C., et al., *Equilibrium and kinetic measurements reveal rapidly reversible binding of Ras to Raf.* J Biol Chem, 1996. **271**(12): p. 6713-9.

279. Hunter, J.C., et al., *Biochemical and Structural Analysis of Common Cancer-Associated KRAS Mutations.* Mol Cancer Res, 2015. **13**(9): p. 1325-35.

280. Campbell-Valois, F.X., K. Tarassov, and S.W. Michnick, *Massive sequence perturbation of a small protein.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(42): p. 14988-14993.

281. Jacobus, A.P. and J. Gross, *Optimal Cloning of PCR Fragments by Homologous Recombination in Escherichia coli.* PLOS ONE, 2015. **10**(3): p. e0119221.

282. Fenton, B.A., *GAP Engineering to Restore GTP Hydrolysis to Oncogenic Kras Mutants*. 2014, Miami University.

283. Schumacher, M.A., et al., *HipBA–promoter structures reveal the basis of heritable multidrug tolerance.* Nature, 2015. **524**: p. 59.

284. Tyagarajan, K., E. Pretzer, and J.E. Wiktorowicz, *Thiol-reactive dyes for fluorescence labeling of proteomic samples.* 2003. **24**(14): p. 2348-2358.

285. Zhang, J.H., T.D. Chung, and K.R. Oldenburg, *A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays.* J Biomol Screen, 1999. **4**(2): p. 67-73.

286. Kessler, D., et al., *Drugging an undruggable pocket on KRAS.* 2019: p. 201904529.

287. Yin, G., et al., *A KRAS GTPase K104Q Mutant Retains Downstream Signaling by Offsetting Defects in Regulation.* The Journal of biological chemistry, 2017. **292**(11): p. 4446-4456.

288. Smith, M.J., B.G. Neel, and M. Ikura, *NMR-based functional profiling of RASopathies and oncogenic RAS mutations.* 2013. **110**(12): p. 4574-4579.

289. Hwang, M.C., Y.J. Sung, and Y.W. Hwang, *The differential effects of the Gly-60 to Ala mutation on the interaction of H-Ras p21 with different downstream targets.* J Biol Chem, 1996. **271**(14): p. 8196-202.

290. Boettner, B., C. Herrmann, and L. Van Aelst, *Ras and Rap1 interaction with AF-6 effector target.* Methods Enzymol, 2001. **332**: p. 151-68.

291. Linnemann, T., et al., *Interaction between Nef and phosphatidylinositol-3-kinase leads to activation of p21-activated kinase and increased production of HIV.* Virology, 2002. **294**(2): p. 246-55.

292. Rudolph, M.G., et al., *Thermodynamics of Ras/effector and Cdc42/effector interactions probed by isothermal titration calorimetry.* J Biol Chem, 2001. **276**(26): p. 23914-21.

293. Wohlgemuth, S., et al., *Recognizing and defining true Ras binding domains I: biochemical analysis.* J Mol Biol, 2005. **348**(3): p. 741-58.

294. Harjes, E., et al., *GTP-Ras disrupts the intramolecular complex of C1 and RA domains of Nore1.* Structure, 2006. **14**(5): p. 881-8.

295. Snyder, P.W., et al., *Mechanism of the hydrophobic effect in the biomolecular recognition of arylsulfonamides by carbonic anhydrase.* Proceedings of the National Academy of Sciences, 2011. **108**(44): p. 17889-17894.

296. Schnell, J.R., H.J. Dyson, and P.E. Wright, *Structure, dynamics, and catalytic function of dihydrofolate reductase.* Annu Rev Biophys Biomol Struct, 2004. **33**: p. 119-40.

297. Stover, P.J., *One-carbon metabolism-genome interactions in folate-associated pathologies.* J Nutr, 2009. **139**(12): p. 2402-5.

298. Fischer, M., B. Thöny, and S. Leimkühler, *7.17 - The Biosynthesis of Folate and Pterins and Their Enzymology*, in *Comprehensive Natural Products II*, H.-W. Liu and L. Mander, Editors. 2010, Elsevier: Oxford. p. 599-648.

299. Hagner, N. and M. Joerger, *Cancer chemotherapy: targeting folic acid synthesis.* Cancer management and research, 2010. **2**: p. 293-301.

300. Li, M.C., R. Hertz, and D.M. Bergenstal, *Therapy of choriocarcinoma and related trophoblastic tumors with folic acid and purine antagonists.* N Engl J Med, 1958. **259**(2): p. 66-74.

301. Revelas, A., *Healthcare - associated infections: A public health problem.* Nigerian medical journal : journal of the Nigeria Medical Association, 2012. **53**(2): p. 59-64.

302. Doshi, R.K., et al., *Healthcare-associated Infections: Epidemiology, Prevention, and Therapy.* Mount Sinai Journal of Medicine: A Journal of Translational and Personalized Medicine, 2009. **76**(1): p. 84-94.

303. Ki, V. and C. Rotstein, *Bacterial skin and soft tissue infections in adults: A review of their epidemiology, pathogenesis, diagnosis, treatment and site of care.* The Canadian journal of infectious diseases & medical microbiology = Journal canadien des maladies infectieuses et de la microbiologie medicale, 2008. **19**(2): p. 173-184.

304. Tang, J., et al., *The use of vancomycin in the treatment of adult patients with methicillin-resistant Staphylococcus aureus (MRSA) infection: a survey in a tertiary hospital in China.* International journal of clinical and experimental medicine, 2015. **8**(10): p. 19436-19441.

305. van Hal, S.J. and V.G. Fowler, Jr., *Is it time to replace vancomycin in the treatment of methicillin-resistant Staphylococcus aureus infections?* Clinical infectious diseases : an official publication of the Infectious Diseases Society of America, 2013. **56**(12): p. 1779-1788.

306. Hiramatsu, K., et al., *Methicillin-resistant Staphylococcus aureus clinical strain with reduced vancomycin susceptibility.* J Antimicrob Chemother, 1997. **40**(1): p. 135-6.

307. Baccanari, D.P. and L.F. Kuyper, *Basis of selectivity of antibacterial diaminopyrimidines.* J Chemother, 1993. **5**(6): p. 393-9.

308. Dale, G.E., et al., *A single amino acid substitution in Staphylococcus aureus dihydrofolate reductase determines trimethoprim resistance.* J Mol Biol, 1997. **266**(1): p. 23-30.

309. Frey, K.M., et al., *Predicting resistance mutations using protein design algorithms.* Proceedings of the National Academy of Sciences, 2010. **107**(31): p. 13707-13712.

310. Young, H.K., R.A. Skurray, and S.G.B. Amyes, *Plasmid-mediated trimethoprim-resistance in Staphylococcus aureus. Characterization of the first gram-positive plasmid dihydrofolate reductase (type S1).* Biochemical Journal, 1987. **243**(1): p. 309-312.

311. Cummins, P.L. and J.E. Gready, *Energetically Most Likely Substrate and Active-Site Protonation Sites and Pathways in the Catalytic Mechanism of Dihydrofolate Reductase.* Journal of the American Chemical Society, 2001. **123**(15): p. 3418-3428.

312. Reyes, V.M., et al., *Isomorphous Crystal Structures of Escherichia coli Dihydrofolate Reductase Complexed with Folate, 5-Deazafolate, and 5,10-Dideazatetrahydrofolate: Mechanistic Implications.* Biochemistry, 1995. **34**(8): p. 2710-2723.

313. Wan, Q., et al., *Toward resolving the catalytic mechanism of dihydrofolate reductase using neutron and ultrahigh-resolution X-ray crystallography.* 2014. **111**(51): p. 18225-18230.

314. Liu, C.T., et al., *Probing the electrostatics of active site microenvironments along the catalytic cycle for Escherichia coli dihydrofolate reductase.* Journal of the American Chemical Society, 2014. **136**(29): p. 10349-10360.

315. Groff, D., et al., *Efforts toward the direct experimental characterization of enzyme microenvironments: tyrosine100 in dihydrofolate reductase.* Angewandte Chemie (International ed. in English), 2009. **48**(19): p. 3478-3481.

316. Lobanovska, M. and G. Pilla, *Penicillin's Discovery and Antibiotic Resistance: Lessons for the Future?* The Yale journal of biology and medicine, 2017. **90**(1): p. 135-145.

317. Fleming, A., *Classics in infectious diseases: on the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of B. influenzae by Alexander Fleming, Reprinted from the British Journal of Experimental Pathology 10:226-236, 1929.* Rev Infect Dis, 1980. **2**(1): p. 129-39.

318. Moellering, R.C., Jr., *Mechanism of action of antimicrobial agents.* Clin Obstet Gynecol, 1979. **22**(2): p. 277-83.

319. Laxminarayan, R., et al., *Antibiotic resistance&#x2014;the need for global solutions.* The Lancet Infectious Diseases, 2013. **13**(12): p. 1057-1098.

320. Grim, S.A., et al., *Trimethoprim-sulfamethoxazole as a viable treatment option for infections caused by methicillin-resistant Staphylococcus aureus.* Pharmacotherapy, 2005. **25**(2): p. 253-64.

321. Huovinen, P., *Resistance to trimethoprim-sulfamethoxazole.* Clin Infect Dis, 2001. **32**(11): p. 1608-14.

322. Larkin, M.A., et al., *Clustal W and Clustal X version 2.0.* Bioinformatics, 2007. **23**(21): p. 2947-8.

323. Krieger, E. and G. Vriend, *YASARA View - molecular graphics for all devices - from smartphones to workstations.* Bioinformatics, 2014. **30**(20): p. 2981-2.

324. Heo, L. and M. Feig, *Experimental accuracy in protein structure refinement via molecular dynamics simulations.* Proceedings of the National Academy of Sciences, 2018. **115**(52): p. 13276-13281.

325. Reeve, S.M., et al., *Toward Broad Spectrum DHFR inhibitors Targeting Trimethoprim Resistant Enzymes Identified in Clinical Isolates of Methicillin-Resistant Staphylococcus aureus.* 2019: p. 648808.

326. Slabinski, L., et al., *XtalPred: a web server for prediction of protein crystallizability.* Bioinformatics, 2007. **23**(24): p. 3403-5.

327. Word, J.M., et al., *Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation.* J Mol Biol, 1999. **285**(4): p. 1735-47.

328. Hornak, V., et al., *Comparison of multiple Amber force fields and development of improved protein backbone parameters.* Proteins, 2006. **65**(3): p. 712-725.

329. Brown, K.A. and J. Kraut, *Exploring the molecular mechanism of dihydrofolate reductase.* Faraday Discussions, 1992. **93**(0): p. 217-224.

330. Wu, Y.D. and K.N. Houk, *Theoretical transition structures for hydride transfer to methyleneiminium ion from methylamine and dihydropyridine. On the nonlinearity of hydride transfers.* Journal of the American Chemical Society, 1987. **109**(7): p. 2226-2227.

331. Frey, K.M., et al., *Crystal structures of wild-type and mutant methicillin-resistant Staphylococcus aureus dihydrofolate reductase reveal an alternate conformation of NADPH that may be linked to trimethoprim resistance.* Journal of molecular biology, 2009. **387**(5): p. 1298-1308.

332. Jacobson, E.L., M.K. Jacobson, and C. Bernofsky, *Evidence against the Natural Occurrence of α-Nicotinamide Adenine Dinucleotide in Azotobacter vinelandii.* Journal of Biological Chemistry, 1973. **248**(22): p. 7891-7897.

333. Klemm, A., et al., *Determination, purification, and characterization of alpha-NADH and alpha-NADPH.* Methods Enzymol, 1997. **280**: p. 171-86.

334. Smith, S.L. and J.J. Burchall, *Alpha-pyridine nucleotides as substrates for a plasmid-specified dihydrofolate reductase.* Proc Natl Acad Sci U S A, 1983. **80**(15): p. 4619-23.

335. Oppenheimer, N.J. and N.O. Kaplan, *The alpha beta epimerization of reduced nicotinamide adenine dinucleotide.* Arch Biochem Biophys, 1975. **166**(2): p. 526-35.

336. Beaupre, B.A., et al., *Renalase is an alpha-NAD(P)H oxidase/anomerase.* J Am Chem Soc, 2013. **135**(37): p. 13980-7.

337. Czyzyk, D.J., et al., *Understanding the structural basis of species selective, stereospecific inhibition for Cryptosporidium and human thymidylate synthase.* FEBS Lett, 2019.

338. Oppenheimer, N.J. and N.O. Kaplan, *Structure of the primary acid rearrangement product of reduced nicotinamide adenine dinucleotide (NADH).* Biochemistry, 1974. **13**(23): p. 4675-4685.

339. Heaslet, H., et al., *Structural comparison of chromosomal and exogenous dihydrofolate reductase from Staphylococcus aureus in complex with the potent inhibitor trimethoprim.* 2009. **76**(3): p. 706-717.

340. Yung-Chi, C. and W.H. Prusoff, *Relationship between the inhibition constant (KI) and the concentration of inhibitor which causes 50 per cent inhibition (I50) of an enzymatic reaction.* Biochemical Pharmacology, 1973. **22**(23): p. 3099-3108.

341. Word, J.M., et al., *Exploring steric constraints on protein mutations using MAGE/PROBE.* Protein Sci, 2000. **9**(11): p. 2251-9.

342. Oppenheimer, N.J. and N.O. Kaplan, *Structure of the primary acid rearrangement product of reduced nicotinamide adenine dinucleotide (NADH).* Biochemistry, 1974. **13**(23): p. 4675-85.

343. Yoshida, A. and V. Dave, *Inhibition of NADP-dependent dehydrogenases by modified products of NADPH.* Arch Biochem Biophys, 1975. **169**(1): p. 298-303.

344. Marbaix, A.Y., et al., *Extremely conserved ATP- or ADP-dependent enzymatic system for nicotinamide nucleotide repair.* The Journal of biological chemistry, 2011. **286**(48): p. 41246-41252.

345. Oppenheimer, N.J. and N.O. Kaplan, *Glyceraldehyde-3-phosphate dehydrogenase catalyzed hydration of the 5-6 double bond of reduced beta-nicotinamide adenine dinucleotide (betaNADH). Formation of beta-6-hydroxy-1,4,5,6-tetrahydronicotinamide adenine dinucleotide.* Biochemistry, 1974. **13**(23): p. 4685-94.

346. Meinhart, J.O., S. Chaykin, and E.G. Krebs, *Enzymatic conversion of a reduced diphosphopyridine nucleotide derivative to reduced diphosphopyridine nucleotide.* J Biol Chem, 1956. **220**(2): p. 821-9.

347. Allouche, D., et al., *Computational protein design as an optimization problem.* Artificial Intelligence, 2014. **212**: p. 59-79.

348. Capriotti, E., et al., *A three-state prediction of single point mutations on protein stability changes.* BMC bioinformatics, 2008. **9**(2): p. S6.

349. Guerois, R., J.E. Nielsen, and L. Serrano, *Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations.* Journal of molecular biology, 2002. **320**(2): p. 369-387.

350. Källberg, M., et al., *Template-based protein structure modeling using the RaptorX web server.* Nature Protocols, 2012. **7**: p. 1511.

351. Potapov, V., M. Cohen, and G. Schreiber, *Assessing computational methods for predicting protein stability upon mutation: good on average but not in the details.* Protein engineering, design & selection, 2009. **22**(9): p. 553-560.

352. Rohl, C.A., et al., *Protein structure prediction using Rosetta*, in *Methods in enzymology*. 2004, Elsevier. p. 66-93.

353. Steinbrecher, T., et al., *Free Energy Perturbation Calculations of the Thermodynamics of Protein Side-Chain Mutations.* Journal of Molecular Biology, 2017. **429**(7): p. 923-929.

354. Steinbrecher, T., et al., *Predicting the Effect of Amino Acid Single-Point Mutations on Protein Stability—Large-Scale Validation of MD-Based Relative Free Energy Calculations.* Journal of Molecular Biology, 2017. **429**(7): p. 948-963.

355. Yang, J. and Y. Zhang, *I-TASSER server: new development for protein structure and function predictions.* Nucleic acids research, 2015. **43**(W1): p. W174-W181.

356. Ikura, T., Y. Urakubo, and N. Ito, *Water-mediated interaction at a protein–protein interface.* Chemical Physics, 2004. **307**(2): p. 111-119.

357. Pelletier, H. and J. Kraut, *Crystal structure of a complex between electron transfer partners, cytochrome c peroxidase and cytochrome c.* Science, 1992. **258**(5089): p. 1748-55.

358. Thomas, C., et al., *Structural linkage between ligand discrimination and receptor activation by type I interferons.* Cell, 2011. **146**(4): p. 621-632.

359. Wang, X., M. Rickert, and K.C. Garcia, *Structure of the quaternary complex of interleukin-2 with its alpha, beta, and gammac receptors.* Science, 2005. **310**(5751): p. 1159-63.

360. Schreiber, G. and A.R. Fersht, *Interaction of barnase with its polypeptide inhibitor barstar studied by protein engineering.* Biochemistry, 1993. **32**(19): p. 5145-50.

361. Frisch, C., et al., *Thermodynamics of the interaction of barnase and barstar: changes in free energy versus changes in enthalpy on mutation.* J Mol Biol, 1997. **267**(3): p. 696-706.

362. Erman, J.E., et al., *Cytochrome c/Cytochrome c Peroxidase Complex: Effect of Binding-Site Mutations on the Thermodynamics of Complex Formation.* Biochemistry, 1997. **36**(13): p. 4054-4060.

363. Piehler, J., L.C. Roisman, and G. Schreiber, *New Structural and Functional Aspects of the Type I Interferon-Receptor Interaction Revealed by Comprehensive Mutational Analysis of the Binding Interface.* Journal of Biological Chemistry, 2000. **275**(51): p. 40425-40433.

364. Robb, R.J., C.M. Rusk, and M.P. Neeper, *Structure-function relationships for the interleukin 2 receptor: location of ligand and antibody binding sites on the Tac receptor chain by mutational analysis.* Proceedings of the National Academy of Sciences, 1988. **85**(15): p. 5654-5658.

365. Chen, V.B., et al., *MolProbity: all-atom structure validation for macromolecular crystallography.* Acta crystallographica. Section D, Biological crystallography, 2010. **66**(Pt 1): p. 12-21.

366. Lovell, S.C., et al., *The penultimate rotamer library.* Proteins: Structure, Function, and Bioinformatics, 2000. **40**(3): p. 389-408.

367. Lovell, S.C., et al., *Structure validation by Cα geometry: ϕ,ψ and Cβ deviation.* Proteins: Structure, Function, and Bioinformatics, 2003. **50**(3): p. 437-450.

368. Hintze, B.J., et al., *Molprobity's ultimate rotamer-library distributions for model validation.* Proteins: Structure, Function, and Bioinformatics, 2016. **84**(9): p. 1177-1189.

369. Ramachandran, G.N., C. Ramakrishnan, and V. Sasisekharan, *Stereochemistry of polypeptide chain configurations.* J Mol Biol, 1963. **7**: p. 95-9.

370. Chen, V.B., et al., *MolProbity: all-atom structure validation for macromolecular crystallography.* Acta Crystallographica Section D, 2010. **66**(1): p. 12-21.

371. Du, J., et al., *Structural Basis for the Blockage of IL-2 Signaling by Therapeutic Antibody Basiliximab.* The Journal of Immunology, 2010. **184**(3): p. 1361-1368.

372. Stauber, D.J., et al., *Crystal structure of the IL-2 signaling complex: Paradigm for a heterotrimeric cytokine receptor.* Proceedings of the National Academy of Sciences of the United States of America, 2006. **103**(8): p. 2788-2793.

373. Yang, H., et al., *Structural basis of immunosuppression by the therapeutic antibody daclizumab.* Cell Research, 2010. **20**: p. 1361.

374. Janin, J., R.P. Bahadur, and P. Chakrabarti, *Protein-protein interaction and quaternary structure.* Q Rev Biophys, 2008. **41**(2): p. 133-80.

375. Ishida, H., Y. Jochi, and A. Kidera, *Dynamic structure of subtilisin-eglin c complex studied by normal mode analysis.* Proteins, 1998. **32**(3): p. 324-33.

376. Dasgupta, B., H. Nakamura, and A.R. Kinjo, *Rigid-body motions of interacting proteins dominate multispecific binding of ubiquitin in a shape-dependent manner.* Proteins: Structure, Function, and Bioinformatics, 2014. **82**(1): p. 77-89.

377. Ubbink, M., *The courtship of proteins: understanding the encounter complex.* FEBS Lett, 2009. **583**(7): p. 1060-6.

378. Dasgupta, B., H. Nakamura, and A.R. Kinjo, *Counterbalance of ligand- and self-coupled motions characterizes multispecificity of ubiquitin.* Protein Sci, 2013. **22**(2): p. 168-78.

379. Muzumdar, M.D., et al., *Survival of pancreatic cancer cells lacking KRAS function.* Nature communications, 2017. **8**(1): p. 1090-1090.

380. Herrmann, C., *Ras–effector interactions: after one decade.* Current Opinion in Structural Biology, 2003. **13**(1): p. 122-129.

381. Arkin, M.R. and J.A. Wells, *Small-molecule inhibitors of protein–protein interactions: progressing towards the dream.* Nature Reviews Drug Discovery, 2004. **3**(4): p. 301-317.

382. Hawser, S., S. Lociuro, and K. Islam, *Dihydrofolate reductase inhibitors as antibacterial agents.* Biochemical Pharmacology, 2006. **71**(7): p. 941-948.

383. Simpkin, V.L., et al., *Incentivising innovation in antibiotic drug discovery and development: progress, challenges and next steps.* The Journal of antibiotics, 2017. **70**(12): p. 1087-1096.

384. Renwick, M. and E. Mossialos, *What are the economic barriers of antibiotic R&D and how can we overcome them?* Expert Opinion on Drug Discovery, 2018. **13**(10): p. 889-892.

385. von Wintersdorff, C.J.H., et al., *Dissemination of Antimicrobial Resistance in Microbial Ecosystems through Horizontal Gene Transfer.* Frontiers in microbiology, 2016. **7**: p. 173-173.

386. Duzer, C.V., *Bring on the Monsters and Marvels: Non-Ptolemaic Legends on Manuscript Maps of Ptolemy's Geography.* Viator, 2014. **45**(2): p. 303-334.

Biography

Marcel Frenkel attended the University of Notre Dame du Lac where he graduated with a biochemistry degree. During his time at Notre Dame Marcel was part of the 2011 division 1 NCAA national champion fencing team. After Notre Dame, Marcel moved to Duke University where he initially joined the lab of Dr. Munir Alam at the Duke Human Vaccine Institute (DHVI). During his time at the DHVI, Marcel presented his work on antigen design and characterization at the CHAVI-ID conference. The work won the award of best poster. Later Marcel changed labs and joined Dr. Bruce Donald. In the Donald Lab, Marcel would be an author on the following publication OSPREY 3.0 Open-Source Protein Redesign for You, with Powerful New Features, he would go on to present this publication at the 2018 international society for computational biology (ISCB) conference where he would also be awarded a travel fellowship. Marcel is currently working on 4 manuscripts to be published in peer-reviewed journals later in 2019.