Computational Design and Optimization of Novel Subtype Specific Frizzled Binding Proteins for Modulation of Wnt Signaling

Luke T. Dang

A dissertation

submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Washington

2017

Reading Committee:

David Baker, Chair

Ning Zheng

Dustin Maly

Program Authorized to Offer Degree:

Molecular and Cellular Biology

University of Washington

Abstract

Computational Design and Optimization of Novel Subtype Specific Frizzled Binding Proteins for Modulation of Wnt Signaling

Luke T. Dang

Chair of the Supervisory Committee:
Professor David Baker
Biochemistry

Wnt signaling is essential to a range of critical biologic processes including embryonic development[1], mature tissue maintenance,[2] and cell proliferation[3]. Dysregulation of the Wnt signaling pathway is conclusively implicated in a range of cancer types[4]. Wnt interacts with its extracellular receptor Frizzed (Fz) via an essential palmitoleic acid modification[5], and blockade of the Fz lipid binding site results in inhibition of canonical Wnt signaling and decreased tumor growth[6]. Therefore, molecules that can potently antagonize the Wnt-Fz interaction have significant promise as anti-cancer therapeutics[6,7,8]. The degeneracy and functional implications of the interactions between 19 distinct Wnt ligands and 10 Fz receptors remains unresolved[9]. Specific antagonism of individual Frizzled receptors will enable disentanglement of this complex signaling network and the roles specific Wnt ligand-receptor interactions play in different tissue and cancer types. Novel, high affinity, Fz-5/8 specific binders were therefore computationally designed and optimized for application as Wnt inhibitors. Subsequent refinement of these binders included stabilization, enhancement of functional antagonism, and interface reengineering resulting in variants with orthogonal binding specificity against alternate receptor subtypes. These proteins were also functionalized as novel, water-soluble Wnt surrogates for complementary activation of Wnt signaling. These novel, rationally designed proteins enable precise modulation of Wnt signaling as required for full elucidation of the role of Wnt signaling in clinically relevant contexts and therapeutic intervention in anti-cancer and regenerative medicine applications. In addition, the computational design of a de novo interface utilizing an idealized ankyrin scaffold provides a general method for the development of novel protein binders.

Table of Contents

List of Figures

List of Tables

Acknowledgements

Thanks to my advisor, David Baker, for providing the opportunity to work on the singular challenge of protein design alongside so many exceptional people from all around the world. The Baker lab's efforts to address biomedical challenges through rational computational design was in large part what convinced me to attend the University of Washington. I am most fortunate to have had such a wonderful graduate experience working on research that was both simultaneously so fundamental in its objective to unravel phenomena such as protein folding as well as so pragmatic in its application of this understanding to biomedical problems through protein design.

Previously, I could not have envisioned such an intellectually engaging opportunity but it was very clear to me how unique and compelling this work was from the single conversation I had with David during my interview seven years ago. The science I saw during my visit to the Baker laboratory stood out clearly to me as the work I wanted to conduct as my graduate research. Anytime a prospective graduate student or postdoctoral fellow would engage me to talk about joining the group, my only caveat would be that I would be extremely biased in my opinion. Although I am sure that for many graduate students time and failed experiments cause some of the their initial enthusiasm to wane, I sincerely continue to feel most fortunate to have had the ability to work on such exciting scientific problems alongside world-class colleagues and collaborators, with seemingly unlimited resources (computational and experimental) at my disposal. David's unselfish leadership is the reason behind the uncommon success of the lab and its wonderfully collaborative environment. I have learned so much from David as well as the

many exceptional members of the group. It will be most difficult to move on to the next stage of my career due to how much I enjoyed my time here.

Thanks also to my committee members: Dustin Maly, Andrew Scharenberg, Patrick Stayton and Ning Zheng, whose support and advice for me during this process was invaluable. I selected my committee members because they came from different departments and backgrounds but shared in common a sincere interest in providing most valuable mentorship to me as well as uncommon scientific acumen. I cannot thank them enough for their advice and insights during this time.

Thanks to James Moody, for his mentorship when I first joined the lab and for our work together in collaboration on this project and others.

Thanks to Keunwan Park, who was such a wonderful collaborator to work with on repeat protein binder design. His expertise on repeat protein design was essential to this project.

Thanks to Lance Stewart, whose work at the Institute for Protein Design enabled many of the resources used in this project. I am also grateful for his generous advice.

Thanks to Daniel-Adriano Silva Manzano for his valuable help as a collaborator and for assistance with computational design and development of idealization protocols.

Thanks to TJ Brunette for his assistance with computation design, in particular repeat protein and loop redesign.

Thanks to Will Scheffler for development of computational methods and his generous expertise.

Thanks to Chantzt Thomas, for his work on Frizzled binder design during the early stages of this project.

Thanks to Rashmi Ravichandran and Lauren Carter - their essential role in the laborious and difficult production of the proteins needed for this work is most appreciated.

I thank also a large number of current and former Baker members who assisted me during my time in the lab in one way or another through advice or assistance (computational or experimental), which if fully detailed might constitute a dissertation in and of itself: Darwin Alonso, Chris Bahl, Stephanie Berger, Matt Bick, Sinisa Bjelic, Scott Boyken, TJ Brunette, Cassie Bryan, Zibo Chen, Aaron Chevalier, Jorge Fallas, Alex Ford, Inna Goreshnik, Ratika Krishnamurty, David Kim, Neil King, Yakov Kipnis, David La, Marc LaJoie, Tom Linsky, Michelle Matsunaga, Vikram Mulligan, Michael Murphy, Jorgen Nelson, Lucas Nivon, Fabio Parmeggiani, Gabriel Rocklin, Danny Sahtoe, Franziska Seeger, Lei Shi, Daniel-Adriano Silva Manzano, Will Scheffler, Michelle Scalley-Kim, Matthew Smith, Eva-Maria Strauch, Christy Tinberg, Rashmi Ravichandran, Ta-Yi Yu, and Shawn Yu. I am sure that given the size of the Baker group I have missed more than a few people – my sincere apologies.

Thanks to my most excellent collaborators: Claudia Janda, Yi Miao, and Chris Garcia from Stanford, equal partners in this work which was enabled by their structural studies and functional validation of the designed proteins described herein. Thanks also to Jason Berndt and Randall Moon from the University of Washington for functional validation of these binders as Wnt modulators. I am truly fortunate to have had worked with collaborators of such high caliber, without which this work would not have been possible.

Thanks to the University of Washington Molecular and Cellular Biology Graduate Program and the UW Medical Scientist Training Program for providing me the opportunity to conduct my graduate work at the University of Washington.

Thanks to my scientific mentors at Washington University in St. Louis: Angie Purvis, who took so much time to personally mentor me throughout my undergraduate career, J. Evan Sadler, who provided me with such a great opportunity to join his lab when I was just starting out, and all the other members of the Sadler lab for their support during my time there. I have been most fortunate that although I have not rotated through many different labs, those that I have been part of were wonderful places, which I have never taken for granted.

I would also like to thank Enrico Di Cera for providing me a great opportunity to spend time in his lab during high school, beginning my interest in structural biology. Thanks also for valuable advice along the way.

And finally, on a personal note, the greatest acknowledgements must go to my family. Thanks to my parents, who spared no effort in providing me every opportunity to further my education and interests. Thank you to my mother, who spent countless hours during my early years working to ensure the best academic (and otherwise) environment for me as I grew up. And thanks to my father, whose graduate work on thrombin no doubt influenced my own early and continued interest in proteins and the indispensable role they play in our lives. I am a most fortunate son. Thanks to my parents-in-law for their unwavering support - I am most appreciative. Thank you to my grandparents, siblings, extended family, and friends for their support and love.

Thanks are also in order to my dogs whose constant affection provide daily joy and inspiration to me. Protein folding as a natural phenomenon is surpassed only by the unconditional love and devotion of a loyal dog, the magnitude of which defies explanation.

Most of all, I must thank my wife and best friend for her support and love. You have been the best partner I could have ever asked for. I was only able to pursue work that was so professionally engaging to me because of your unselfish patience and understanding during these years.

Introduction: Proteins by Design – A Rational Approach to Protein Engineering

Protein Structure: A Mechanistic Perspective on Life, Health and Disease

My interest in scientific research originated from a desire to investigate the mechanisms by which the processes of life are carried out at the molecular level. My undergraduate research was focused on identifying the individual residues responsible for von Willebrand factor (VWF) dimer and multimer formation[10, 11]. VWF is responsible for recruitment of platelets to sites of vascular injury, which it accomplishes by virtue of its unique macromolecular structure. VWF originates as a 2050 amino acid monomer that dimerizes in the endoplasmic reticulum via disulfide-linkages between its cysteine knot tail domains. In the Golgi, VWF dimers oligimerize through head to head disulfides to form multimers which can reach over 20,000 kDa in size[12]. These extended multimers subsequently assemble into helical tubules that pack into endothelial cell secretory granules known as Weibel-Palade bodies[13]. Upon stimulation, the release of these tubules into the higher pH of the bloodstream promotes the ordered unfurling of extended VWF strings in a manner optimal for the recruitment of circulating platelets.

VWF's ability to be synthesized, processed, packaged, and rapidly secreted to mediate platelet recruitment is enabled solely by its amino acid sequence. Whether it be forming the correct, untangled dimeric and multimeric disulfide bonds from among VWF's 234 cysteine residues (which comprise over eight percent of the translated polypeptide)[12], assembling into pH-dependent helical tubules, or binding platelet recognition sites, the essential biologic interactions of VWF are all dependent on and determined by the structure formed by its specific amino acid sequence. Remarkably, this complex phenomenon occurs separately for each protein in nature,

the combination of which enables life. Understanding the process by which a protein folds to a specific structure from its determinative primary sequence is therefore intrinsically one of life's most essential and basic questions. This is complicated even further by the fact that a single sequence oftentimes performs multiple functions because of conformational changes triggered by environmental factors (pH, temperature, or ligand concentration) or post-synthetic modification (multimerization, phosphorylation, ubiquitination, proteolytic cleavage).

Additionally, in the interest of self-preservation, we must understand the molecular interactions that undergird the healthy, appropriate functioning of cells and tissues as well as the causative dysfunction of such processes that inevitably lead to disease. A full comprehension of the molecular mechanisms of life must ultimately be attained through a structural perspective of the atomic level interactions that drive all biologic functions. Therefore, any investigation that seeks to understand the complex yet elegant functioning of a given biologic process or the typically brutal results of molecular dysfunction must find its answer in the structure of the constituent proteins which are ultimately derived from their primary amino acid sequences. Pragmatism, therefore, also dictates that a full understanding of how proteins fold and interact from a structural perspective is indispensable if we wish to achieve better outcomes in human health.

Protein Engineering: a Manmade Solution to the Challenges of Nature

My early research into VWF gave me both a sense of wonder that the complex processes of nature could be reduced ultimately to primary amino acid sequence information and an appreciation for the relationship between sequence and structure. One of the projects I worked on at that time focused on identifying individual histidine residues that act as pH sensors and initiate appropriate formation of multimeric disulfides based on transient protonation in the Golgi. At

one such site, mutagenesis of the native histidine to either arginine or lysine (thus forcing a constantly positive charge rather than a reversible one based on the pH of the environment) induced VWF multimer formation levels that at least restored that of the native sequence and actually appeared enhanced[11]. Although this study relied upon crude, individual level amino acid substitutions based on sequence phylogeny and was uninformed by structure, I found the insight that one could engineer proteins to alter or enhance native functions quite profound. Not only did protein structure comprise the existing molecular mechanism of health and disease, but informed alterations to protein sequence could also actually be the means of addressing the challenges of disease.

Indeed, even limited engineering of native proteins has yielded significant improvements to human health, as in the case of recombinant insulin[14]. Native insulin is a 51 amino acid protein comprised of two chains (A and B) bridged by disulfide bonds, and recombinant insulin is essential to the treatment of diabetes, a widespread metabolic disorder. Insulin is well studied, and several recombinant formulations and variants exist whereby minimal changes to the native sequence result in significant alteration of function. For example, simple mutations to the native sequence have been utilized to engineer short and long-acting insulins, both useful for the successful management of diabetes. In the case of lispro insulin, substitution of amino acid identities at positions B28 and B29 (which in the native insulin are respectively, proline and lysine) reduces formation of dimers and larger aggregates, improving uptake of monomeric insulin and increasing efficiency of absorption. Insulin aspart and insulin glusuline work similarly by insertion of non-native charged amino acids to likewise reduce self-association of the protein, promoting the monomeric state, and reducing time of absorption. Conversely, insulin glargine is a long-acting insulin engineered for a once-daily basal release of insulin. This is

achieved by mutation of the native asparagine at position A21 to glycine and addition of two C-terminal arginines, increasing the isoelectric point and reducing the solubility of the protein at physiologic conditions, causing it to precipitate after injection. The precipitated insulin glargine is then absorbed slowly over the course of many hours, allowing it to be administered once daily. So even very limited engineering of native proteins (as seen in the case of insulin, even a single mutation) can provide effective solutions to large-scale problems.

Replacement of a native protein may suffice for the treatment of some disease states. However, in many other cases an existing native protein may not be capable of serving a desired function. Blockade of a particular molecular-level interaction may be necessary for intervention in a disease process, and a native inhibitor may not exist. Or if a native inhibitor does exist, it may be unsuitable for recombinant production because of stability concerns, or it may otherwise interact with orthogonal pathways and create undesired off-target effects when applied in a therapeutic context.

Small-molecule drugs may be looked to as a first resort, but some molecular interactions are difficult to target with small molecules. Many of the processes of life are driven by the activity of enzymes, which requires binding of substrates to active sites with typically concave pockets. These surfaces are well suited to inhibition by small molecules that can efficiently pack into such cavities. However, many other protein-protein interactions are characterized by flatter interfaces that are a more challenging to approach with small molecules. Although the small size of such drugs is an advantage in manufacture, administration, and pharmacokinetic distribution, it also limits the number of interactions that can be made with a target protein. Therefore, unless the molecule makes a disproportionate amount of interactions with the targeted surface and is particularly efficient in doing so compared to its diminutive size, it is unlikely to bind with the

desired affinity. The ligand efficiency metric, which measures the free energy of binding normalized on a per-atom basis, accounts for this and is widely used in the development of small molecule drugs[15,16].

If a target surface is not well suited towards making many energetically favorable (either hydrophobic or electrostatic) interactions within a small interface area, then a larger protein which may make a larger number of interactions over a spatially more distant area may be better able to interact with that particular surface. Although for small ligands the gain in the free energy of binding associated with additional size may represent diminishing returns[17], for larger macromolecules the additional interactions may be additive and enable development of high-affinity binders through the combination of numerous favorable interactions. A protein also has the advantage of being able to present distant interacting groups to the target surface in a more geometrically optimal manner, since the variety of protein folds and combinatorial diversity of various secondary structure elements and higher order arrangements is virtually limitless. Therefore, outside of substrate recognition sites that are well suited for small-molecule inhibitors, proteins will oftentimes have the advantage of being better able to interact with biologic surfaces in an energetically favorable manner.

For similar reasons, discriminating among similar biologic targets with a small molecule can be a challenge when the interface area is by definition limited compared to a larger molecule that can make additional contacts. If the primary (or only) interactions that are made are to a biologically functional region, it is likely that this region will be largely (or completely) conserved, and that targeting this surface may result in interactions with similar surfaces as well. This is a significant drawback, because similar proteins may exist and be expressed in different tissues or perform separate functions, but targeting multiple family members may create

unacceptable off-target effects. A protein may be better able to discriminate among multiple targets that are very similar by interacting both with a conserved region of the target and a region that varies, allowing it to achieve a high degree of specificity.

Therefore, it can be argued that proteins have the added potential advantage of being more capable than small molecules of binding with both high affinity and specificity due to their larger size and more flexible topologies capable of scaffolding a greater number of interactions within a single molecule. Indeed, if the partner that natively interacts with a given molecular surface is another protein; then a protein must often represent the ideal solution for replicating and improving upon the interactions of the native complex. Although small molecules have many advantages that have been exploited by modern medicine, it is clear that there must be a large number of instances in which development of protein-based biotherapeutics would be advantageous. Indeed, reviews of druggable targets of interest identify a significant percentage of the human genome that may be better targeted by protein biotherapeutics than small molecule drugs[18,19].

To this point, engineered antibodies now comprise an entire class of biotherapeutics effectively used to specifically target a wide range of conditions from cancer[20] to autoimmune disease[21]. Antibodies, as essential components of the human body's own adaptive immune system, represent an elegant way of using directed protein engineering to effectively intervene in disease. Development of such antibodies typically proceeds empirically, using affinity maturation methods. In this process, a diverse, naive library of antibodies is exposed to the target protein, and those proteins that interact the most strongly are selected out using phage-display[22] or yeast surface display[23,24] technologies. Additional iterative rounds of diversifying mutagenesis

and selection ultimately result in the maturation of high-affinity protein binders against the target of interest. These methods are effective, but have certain limitations.

For one, antibodies selected using these methods are not specifically directed against a particular region of the target. Rather, the antibodies with the highest overall affinity will be selected regardless of epitope. This presents a problem for many applications, since oftentimes the objective of antibody development is antagonism of a specific biologically relevant interaction. For example, in the case where the blockade of a ligand-receptor interaction is desired, binding to the receptor in a non-competitive binding mode would be insufficient. High-affinity binding to the ligand that results in target sequestration might make the specific epitope of the antibody less relevant, but even in this case a binding mode that definitively precludes the ligand-receptor interaction would be the optimal solution. Although methods exist to address this problem empirically, such as selecting for proteins that do not bind when the target's native binding partner is used as a competitor, ultimately these approaches are crude and are less than satisfying.

In addition, although antibodies are a powerful platform for the development of biotherapeutics, they represent nature's evolved solution for human adaptive immunity. It is possible that other natural or non-natural protein topologies might be superior to antibodies in other applications. Existing functional antibodies and native proteins are frequently improved by downstream development (as in the previously noted case of insulin), and such engineering can be dramatically enhanced by more powerful rational approaches. The marked success of protein engineering over the past few decades has proven that proteins may achieve outcomes unavailable to small molecules and may represent the best solution to many problems in cancer, immunology, and other fields. But in order to develop the next generation of proteins to address

current challenges, we must look beyond traditional methods and their accompanying limitations.

Rational Computational Design: A New Paradigm in Protein Engineering

In comparison to directed evolution, the rational targeting of the exact molecular surface necessary for biologic function represents a much more elegant approach to protein engineering. Rational design methods are inherently more precise and tailored than blind selection, and may more fully sample the universe of potential solutions. Each position in an amino acid chain may adopt twenty identities, so the size of a naïve library will be limited by transformation efficiency and even for a small protein must represent only a fraction of possible permutations. A protein which is only 150 amino acids long can adopt 20^150 (1.43^195) unique sequences, dwarfing the size of even ribosome display libraries. So if the ultimate goal is precise engineering of proteins with optimal characteristics for a given application, rational design must be pursued in order to reduce the scale of this problem. This is not to doubt the brutal effectiveness of directed evolution methods, which ideally will be used in combination with rational design for maximal efficiency. But as demonstrated by the combinatorial diversity of even a single small protein, the number of potential sequences for a polypeptide of a certain length is already staggeringly vast. When we consider that each of these possible polypeptides is comprised of a chain that can adopt different conformations at each bond along its amino acid backbone, the scope of the protein folding problem becomes even more impressive, and truly cannot be overstated. As previously discussed in the case of VWF, native proteins and complexes are often orders of magnitude larger than the small protein discussed here as an example. Therefore, without a method for first predicting secondary and tertiary structure from the primary amino acid sequence of a protein

and also the quaternary interactions between folded protein domains this problem will remain unapproachable.

Fortunately, there has been progress towards the goal of predicting protein folding utilizing computational methods. Although the number of possible protein topologies is vast, a given sequence will fold predictably to the conformation with the lowest energy. The chemical properties of each amino acid are known, as are the principles that govern how proteins fold. Formation of favorable van der Waals interactions, electrostatics (charge interactions), and hydrogen bonds lower the energy of a protein conformation. A given protein will adopt the conformation that results in the most favorable combination of the above interactions. Therefore, hydrophobic residues will pack into the core of the protein, while polar residues will be solvent facing. The secondary structure a local primary sequence is likely to form can also be predicted, since amino acids occur with varying frequencies in different secondary structure types (loops, sheets or helices). Non-local interactions between secondary structure elements distant in the primary sequence are less predictable but follow the same physical principles. Comparative modeling of a protein can successfully predict a structure by threading the sequence onto the known structure of a closely related homolog, which acts as a template. Alternatively, structures can be predicted *ab initio* by sequential sampling of local and then progressively more distant structures utilizing an algorithmic approach. Sufficient sampling should identify the global energy minimum corresponding to the native structure. Improvements in computational resources (increasing sampling) and prediction algorithms have likewise improved the accuracy of structure prediction, although the reliable prediction of native protein structures remains challenging,

Since the principles that dictate both native and designed protein interactions are the same, the development of computational methods for structure prediction provides a new approach for the design of novel, manmade proteins. The utilization of such computational methods for protein design represents a new paradigm in protein engineering. Indeed, Rosetta, the macromolecular modeling software used in the following work, has been used successfully to design numerous novel proteins inaccessible to previous methods, including novel enzymes that catalyze a range of known chemical reactions, including the Kemp elimination[25], the Morita-Bayless-Hillman reaction[26], ester hydrolysis[27], the Diel-Alder reaction[28], the retro-aldol reaction[29], and organo-phosphate hydrolysis[30]. Additionally, the development of a phosphotriesterase enzyme for the detoxification of nerve agents such as VX[31] provides an example of how design of novel proteins can provide remarkable solutions even to manmade, non-natural problems. KumaMax is a Rosetta-designed gliadin endo-peptidase whose engineered properties may enable it to serve as an effective oral therapeutic for celiac disease[32]. Protein design has also enabled the design of self-assembling protein cages and nanomaterials with potential uses in drug delivery and vaccine development[33,34,35,36]. As well, novel proteins have been designed which bind a range of targets including digoxigenin[37], the Ebstein Barr virus protein BHRF1[38], the IgG Fc region[39], metal ions[40], hen-egg lysozyme[41], and influenza hemagglutinin[42]. In vivo studies in mice have demonstrated the efficacy of these designed hemagglutinin binders to protect against infection[43]. Therefore, protein design now represents a new and distinct pipeline for the generation of proteins with powerful real world applications, many of which were previously inconceivable.

Although the rational computational design of novel protein binders against specific functional surfaces of clinically relevant targets has been successfully demonstrated [38,42,43]

protein design methods remain under development. Combining known binding epitopes with stable native or non-native scaffolds can result in higher rates of success, but design of true de novo protein interfaces remains elusive. Structural information specifying molecular interactions which are known to enable binding to a targeted surface does not always exist to serve as a template, and known native binding modes are not always sufficient for achieving the desired effect. Therefore, methods improvement allowing the design of de novo binders is essential to enabling the successful design of the next generation of protein therapeutics.

Since protein design is a relatively new field, each new application of this technology to new problems yields valuable insights that can be applied generally. At the level of an individual project, protein engineering is an iterative process, taking the best variant through round after round of improvement. This is also true when viewed from a more global level, as lessons from the successes and failures of design attempts inform future, improved design processes. My graduate research was enabled only by the significant work of protein designers that came before me. I hope that both the successes and failures of this project will enable the success of others. The following chapters will discuss my graduate work, namely the application of protein design methods to the field of Wnt signaling. More specifically, I utilized Rosetta, a suite of macromolecular modeling and design tools, to develop novel binders against Frizzled, the extracellular receptor for Wnt. I will first discuss the rationale and significance of the Frizzled-Wnt interaction, as well as the approach I used to design modulators of this pathway including a summary of efforts that did not yield binding proteins as well as the insights gleaned from these designs. In the following sections, I will describe in greater detail the computational design, affinity maturation, and subsequent redesign of two novel binders, Fz27 (an optimized variant of which was later designated as B12) and ANK12, which target the Wnt-Frizzled interface.

Additionally, I will discuss functional characterization of these binders as well as potential future applications of these Wnt modulators towards scientific and therapeutic aims.

It will be evident to the reader that design of de novo binders remains a difficult enterprise, but also that iterative application of protein design methods can yield significant, novel results that would be unattainable by other means. In this case, we were ultimately successful in generating high-affinity, subtype-specific Frizzled binders that can act functionally as both Wnt antagonists and agonists. The development of novel water-soluble Wnt modulators is an achievement which hopefully will provide a powerful tool for answering long-standing questions in the field which were previously inaccessible due to lack of appropriate reagents. The computational redesign of the B12 interface loop is proof of principle that novel topologies can be built to specification to fulfill functional requirements. From a methods perspective, the successful redesign of B12 also provides a specific example of how de novo protein design can be integrated synergistically with existing structural information and protein engineering methods to pragmatically improve the properties and effectiveness of a protein. Additionally, this work represents (to my knowledge), the first de novo computational design of an interface utilizing an ankyrin repeat protein. Hopefully this work will ultimately serve two purposes: the developed binders and their derivatives will have pragmatic utility in the investigation and therapeutic modulation of Wnt signaling, and also that the results of this work will help inform future successful protein design efforts.

Section 1: Targeting the Wnt-Frizzled Interaction Via Computational Protein Design

1.1 Wnt Signaling: Background and Significance

Wnt signaling is required for critical physiologic processes ranging from patterning during development to stem cell proliferation and tissue homeostasis[44]. Aberrant activation of Wnt signaling is associated with carcinogenesis and further elucidation of this complex pathway will enable development of effective therapeutic interventions for a range of disease states. Wnt signals through both the canonical and non-canonical (planar-cell polarity (PCP) and Wnt/Ca^2)[44,45] signaling pathways. The best understood of these is canonical Wnt signaling which utilizes the transcriptional activator beta-catenin as a signal transduction molecule.

In the absence of Wnt signaling, beta-catenin is targeted for proteasomal degradation upon phosphorylation by the destruction complex, comprised of axis inhibition protein (AXIN), adenomatous polyposis coli (APC), CK1, and GSK-3B. Wnt signaling is initiated by binding of a Wnt ligand to the extracellular cysteine-rich domain (CRD) of a Frizzled (Fz) seven-pass transmembrane receptor. Formation of a receptor complex with LRP5/6 results in intracellular interaction with Dishevelled(Dsh), recruitment of AXIN, and sequestration of the destruction complex at the membrane, which results in increased cytoplasmic levels of beta-catenin. Nuclear translocation of beta-catenin and interaction with T-cell transcription factors(TCF) and lymphoid enhancer-binding factors(LEF) results in activation of target genes including c-Myc, Cyclin D1, survivin, and MDR-1[46], many of which are oncogenic.

Regulation of Wnt signaling includes sequestration of Wnt by secreted Frizzled-related protein (SFRP) and Wnt inhibitory factor (WIF), inhibition of receptor complex formation by Dkk1 via its interaction with LRP6, and potentiation of signaling by R-spondins[47]. Mutations to components at the ligand (SFRP/WIF/DKK1), receptor (LRP5), destruction complex

(APC/GSK3/WTX), messenger (B-catenin) and transcription factor (CBP) levels of Wnt signaling result in aberrant signaling known to cause cancer[4]. Overexpression of Frizzled receptors[48] or silencing via hypermethylation of sFRP[49], DKK1[50,51], or WIF-1[52] is also associated with cancer. Inhibition at the ligand-receptor level results in apoptosis of cancer cells even in the presence of downstream mutations[52,53], demonstrating that antagonism at the upstream level may provide an effective modality for clinical intervention.

1.2 Rationale for the Development of Frizzled Subtype-Specific Wnt Antagonists

Effective blockade of the Wnt pathway at the receptor level is essential to a full understanding of the functional role of the specific interactions between 19 Wnt and 10 Frizzled subtypes. The extent to which these binding pairs play distinct roles or are degenerate remains unresolved, although it is clear differential binding preferences and downstream signaling do exist, as shown by Dijksterhuis et al[54]. A full understanding of the subtype-specific Wnt-Fz interactions and the role they play in distinct developmental stages and tissue types would provide information that can be leveraged in therapeutic applications. Progress towards this goal has thus far been limited by the lack of robust reagents that can reliably and specifically antagonize these interactions, but research to date has shown the potential of such an approach. For example, Frizzled-8 is often overexpressed in non-small cell lung cancers (NSCLC)[46] and knockdown of Frizzled-8 by siRNA in this context resulted in inhibition of cell proliferation and sensitization to chemotherapy, suggesting that Frizzled-8 may be a specific therapeutic target in NSCLC[55]. Frizzled subtype-specific Wnt inhibition can potentially leverage expression differences to target Wnt-dependent cancers expressing a particular receptor subtype while minimizing off-target effects against other tissues expressing other distinct receptor subtypes.

The development of Wnt inhibitors that are Frizzled-subtype specific is therefore critical to future anti-cancer therapeutic intervention in this signaling pathway.

Compared to previous, typically broad-acting chemotherapeutics, protein-based biotherapeutics provide an opportunity for specific intervention by targeting causative molecular dysfunctions that exist only in the affected cancer cells[56]. Therefore, identification of molecular level defects in Wnt dependent cancers will enable effective treatment of such neoplasms. Previous investigation had identified RNF43, an ubiquitin E3 ligase which regulates membrane expression of Frizzled receptors[57] through endocytosis, as a tumor suppressor protein[58]. Jiang et al[59] demonstrated that RNF43 mutant pancreatic ductal adenocarcinoma (PDAC) cell lines were Wnt dependent, and a comprehensive CRISPR-Cas9 screen conducted by Steinhart et al[60] identified Frizzled-5 as a gene essential for Wnt-signaling in this context. Subsequent blockade of Frizzled-5 resulted in reduced proliferation in both in vitro cell culture and in vivo mouse xenograft models. RNF43 mutations also exist widely in colorectal and endometrial cancers[61], therefore inhibition of Wnt signaling may also represent an effective approach to anti-cancer therapy in these neoplasms or other RNF43 mutant cancers as well.

Given the complexity of Wnt signaling and its many context dependent functions, the development of subtype specific antagonists is critical to elucidating the complex network of Wnt-Fz interactions that drive downstream signaling. Knowledge about the functional consequences of specific Wnt antagonism can then be applied with consistent and targeted clinical effects. As discussed, significant tissue-specific expression patterns exist between Frizzled subtypes, a feature that may be exploited to specifically target neoplasms while reducing off-target activity. RNF43 provides an example of a causative mutation whose identification may enable potent targeting of cancers based on a molecular marker. Moreover, due to the essential

role of Wnt signaling in cell proliferation, it is likely that other molecular level mutations affecting this pathway will continue to be identified. Specific Wnt antagonists will therefore represent a new class of therapeutics orthogonal to existing approaches that will enable precise targeting of these neoplasms.

1.3 Rationale for the Development of Water-Soluble Wnt Surrogate Proteins for Agonist Applications

Wnt agonists (Wnts or R-spondins) are required for the successful propagation of various stem cell[2] types. Unfortunately, purification of lipid modified Wnts has been a long-standing challenge in the field[62], and has limited the robust study of Wnt signaling. Takada et al[5] demonstrated that Wnt-3a is acylated at Ser209 with palmitoleic acid by the protein Porcupine, and furthermore that this modification is required for expression. This lipid modification is not only required for expression, it is also essential for Wnt receptor binding, as will be discussed shortly. Therefore, Wnt expression issues cannot be solved by the simple removal of the lipid modifications whose hydrophobic properties make secretion and purification of these proteins so difficult. Rather, it may be necessary to develop novel molecules for the activation of Wnt signaling which can serve as water-soluble Wnt surrogate proteins and resolve the solubility issues of the native protein.

The development of highly expressible, water-soluble Wnt agonists will allow more extensive investigation of the role Wnt plays in normal tissue homeostasis, carcinogenesis, and cancer stem cell regulation. Although Wnt signaling can lead to unchecked cell proliferation and drive the formation of many cancer types, activation of the Wnt/B-catenin pathway is required for apoptosis of melanoma cells when treated with BRAF inhibitors[63]. This suggests that in some contexts Wnt agonism could sensitize cancer cells to treatment. Additionally, the key role of Wnt

signaling in development, proliferation, and differentiation makes development of specific Wnt agonists key to the progress of regenerative medicine[64]. Although the primary objective of this work is the design of Wnt antagonists, the development of specific Wnt agonists is an essential and complementary objective that will enable a complete utilization of the Wnt pathway in clinical applications.

To this end, Wnt receptor binding proteins, namely the Frizzled binders developed in this study, may be functionalized for use as agonists by fusion to other LRP binding proteins. Wnt activation proceeds from simultaneous binding of the extracellular domains of Frizzled and LRP5/6 to form a receptor complex, leading to downstream intracellular activation. Therefore, it may be postulated that engineered, bi-specific proteins that interact with both Frizzled and LRP5/6 may be able to similarly activate Wnt signaling. Native proteins exist (Dkk1, sclerostin (SOST)) that interact with LRP6 in competition with Wnt, and these proteins may be considered candidates for the LRP-binding portion of the Wnt surrogate. The limiting factor is then the development of suitable proteins that interact with Frizzled. Recapitulating the site of Wnt interaction closely with the developed binders would be optimal since the precise orientation of receptor complexes is often essential for function. If the constituent proteins of an engineered surrogate interact utilizing binding modes to their respective receptors similar to that of the native Wnt ligand, the fusion will be more likely to mimic the geometry of the native Wnt complex and initiate signaling activity. As in the case of Wnt antagonists, subtype-specific Wnt agonists will enable the more precise elucidation of the contribution of specific ligand-receptor interactions to Wnt signaling activation in a given biological context. Fortunately, any potential agonist activity will be initiated by binding to the Frizzled receptor, and therefore the binding

specificity of the Frizzled-interacting protein will inherently control the specificity of any bi-specific agonist.

The development of robust Wnt antagonists and agonists may be considered complementary goals. These combined protein engineering efforts will result in a suite of specific Wnt agonists and antagonists enabling full elucidation of the role of Wnt signaling in development and disease. Further optimization of these Wnt modulators will improve functional properties for downstream scientific and clinical applications. This ultimate goal first requires the development of subtype-specific protein binders to the Wnt binding site of Frizzled.

1.4 Targeting the Native Wnt-Frizzled Protein-Lipid Interaction

The structural basis of the Wnt-Frizzled interaction was revealed by Janda et al[65], which for the first time provided the mechanistic rationale for its essential lipid modification. Wnt interacts with Frizzled via a novel two-site binding mode notable for insertion of a lipid modification into the hydrophobic groove of Fz8-CRD (Figure 1). Janda et al hypothesized that protein-protein contacts at the second site enable discrimination between Fz subtypes, while the largely degenerate lipid-protein interaction drives binding affinity. Antagonism at the lipid-binding site of Fz is sufficient for inhibition of Wnt signaling and tumor growth[6] as demonstrated by Gurney et al utilizing an anti-Frizzled-1/2/5/7/8 antibody whose epitope was mapped to the Wnt binding site. Gurney et al also demonstrated that treatment with this antibody resulted in significant anti-proliferative activity in a range of tumor types in mouse xenograft models, confirming that the Fz lipid-binding site represents a highly relevant clinical target. Indeed, this broadly binding antibody is currently in Phase 1b clinical trials against HER2-negative breast cancer and advancer pancreatic cancer[66]. The fact that this lipid-protein interaction is integral to the ligand-receptor interface explains why this modification is required for Wnt's biologic

activity. Design of Wnt antagonists must therefore target the Frizzled lipid binding site and compete with the native Wnt ligand.

Despite the biologic and clinical relevance of Wnt blockade, suitable subtype specific antagonists at the Wnt-Fz level do not exist. As discussed, development of subtype-specific binders will facilitate the investigation into which Wnt-Fz pairs play distinct functional roles and which complexes are degenerate. Although the lipid-binding site is largely conserved among Frizzled subtypes, there is variation among adjacent residues. Therefore, design of a binding mode that overlaps with both the conserved lipid binding cleft and adjacent areas will result in Wnt antagonism as well as provide a sufficiently large interface to include interactions which enable discrimination among receptor subtypes. Given the precise nature of the required interface, computational protein design methods were used in this work to facilitate the rational design of binders with suitable binding modes.

1.5 Computational Design of Novel Frizzled Binders

As discussed in the introduction, computational protein design has been successful in the design of novel interactions against a range of targets. However, most of these examples have utilized prior knowledge of existing binding epitopes derived from native protein-protein complex structures. In this case, although the Wnt-Fz complex sheds light on the exact receptor region implicated in its biologic function, the lipid interacting portion of the Wnt ligand does not comprise a protein epitope that can be used as a template for binder design. Therefore, although detailed structural information of the Wnt-Fz interface is available, binder design against this site may still be considered a de novo design problem since this high-affinity interaction is dependent on a lipid-modification that cannot be duplicated by a macromolecule wholly comprised of

protein. Replacement of the native protein-lipid interaction with a novel protein-protein interface by design would therefore represent a significant achievement.

In order to maximize the chance of success in the face of this significant challenge, a diverse set of novel designs were generated to test proposed potential high-affinity protein-protein interactions against the targeted surface. Initial design strategies began by grafting rationally conceived helical binding epitopes targeting the Frizzled hydrophobic lipid binding cleft onto compatible native and de novo designed proteins. Later design strategies focused on global docking of repeat proteins (utilizing leucine rich repeat and ankyrin protein folds found in nature) against the target. A wide range of possible binding modes and scaffolds were ultimately designed and tested during the course of this work (Figure 2).

The first generation of designs utilized a minimal alanine helix epitope grafted onto native helical proteins curated from the PDB (Figure 2B). This binding epitope was conceptualized with the aim of recapitulating closely the hydrophobic packing of the native lipid into this conserved cleft (Figure 2A) and native scaffolds were selected based on their compatibility with this epitope and the backbone shape complementarity of additional contacts made by each protein fold after grafting. These design efforts yielded the binder Fz27 (later designated as B12), the development, characterization, and functionalization of which will be discussed in detail in the second section of this text. Although the first generation design B12 ultimately enabled the successful development of novel Wnt modulators, it was revealed by structural analysis to bind in an alternative conformation compared to the design, explaining its initially limited antagonist activity. Therefore, additional orthogonal design strategies were pursued simultaneous to the redesign of B12 as a more potent Wnt inhibitor.

The second generation of helical binders utilized more refined, feature-rich helical epitopes and a scaffold set comprised of previously experimentally characterized de novo proteins with favorable characteristics. Changes to both the binding epitopes and parent scaffolds were made in response to the limited success of the first generation of designs. As seen by the changes in backbone structure in the case of Fz27, significant mutagenesis of a protein sequence is often necessary to confer functional binding activity but may have destabilizing effects. Utilizing scaffolds with a higher margin-of-safety in terms of structural stability should reduce design failure due to the amino acid sequence not folding to the designed structure as was observed here in the first generation of helical designs. In addition, validation of designs by in silico forward folding to predict the lowest energy structure of a given sequence should reveal whether the scaffold will be likely to tolerate the destabilizing mutations needed to design an optimal interface. However, this procedure is only meaningful if the original scaffold forward folds prior to design so that the folding funnel can be compared before and after interface design. Limiting scaffold selection to experimentally characterized stable scaffolds which forward fold in silico allows higher confidence that the proteins will be well behaved and fold to the designed structures compared to designs made utilizing scaffolds selected from the PDB with no consideration of structural stability or experimentally verified physical characteristics. Since most native proteins are not evolved for maximal stability, for the purposes of protein design this is suboptimal as the significant alterations needed for design of novel binders will in many cases be destabilizing to the overall protein.

In addition to stringent scaffold selection criteria, the second-generation designs utilized binding epitopes which included feature-rich hydrophobic sidechain hot-spot contacts, in contrast to the first generation designs which attempted to use the hydrophobic backbone as a

primary part of the binding epitope. The helical design process originated with a defined, static helical epitope hypothesized to bind to the lipid binding cleft, and discovered possible adjacent binding modes via grafting The sidechain interactions used in the second helical design strategy are more typical of protein-protein interfaces. Four distinct epitopes were designed utilizing MotifDock to optimize backbone placement via prediction of potential favorable sidechain-sidechain interactions and RosettaScripts for final design refinement. These epitopes were grafted onto the selected de novo scaffolds, filtered, and redesigned to yield a set of 31 designs, although detectable binding activity was not observed in this case. An example design from this approach can be seen in Figure 2C.

As mentioned previously, the selection of scaffolds with high stability may be considered crucial for the development of protein binders. This should enable more extensive mutagenesis during interface optimization without backbone destabilization. Proteins with high inherent stability may be considered to have a certain margin-of-safety for protein design applications in comparison to native proteins that did not evolve for maximal stability and therefore are likely to misfold when redesigned for orthogonal functions. In addition, scaffold platforms with favorable characteristics for downstream applications provide an advantage for development of designs with real-world impact. While antibodies remain the standard for therapeutic biomolecules, other classes of proteins more amenable to design are being explored, notably repeat proteins such as ankyrins[67] and leucine rich repeats (LRRs). The concave interfaces of these molecules also potentially enable the design of "reverse-hotspots" which interact with large exposed hydrophobic residues on the target surface as seen serendipitously in the structure of Fz27. In addition, the backbone curvature of LRRs can be custom designed computationally to provide optimal scaffold shape complementarity[68].

Ankyrin based designs were generated by global docking of consensus ankyrins scaffolds (consisting of three, four, or five repeats) against the target. The global docking process was agnostic of specific interactions, and rather sought to sample more widely all possible binding modes that may not have been rationally conceived. In addition, the parent scaffolds used had alanine-mutated interfaces in order to provide enable sampling and discovery of backbone binding modes with maximal interface interdigitation and shape-complementarity. Conformations resulting in backbone-backbone clashes with Wnt were chosen and locally refined using MotifDock, a computational protocol developed by Will Scheffler. In particular, a binding mode which had a unique, "fingers-in-groove" binding mode which inserted the variable loops of the ankyrin fold into the Frizzled lipid-binding cleft was selected as particularly advantageous. The interfaces of designs with this binding mode were remarkable both for their complete blockade of the lipid cleft as well as the extensive contacts made with both conserved and variable regions of the receptor. This was hypothesized to enable design of subtype-specific Wnt antagonist since there would be core, high-affinity binding interactions that could remain constant while residues contacting specificity determining regions could be altered. These candidate designs were optimized for binding (using a RosettaScripts GreedyOptimization protocol) and 33 designs were selected for testing (Figure 2D). An additional six binders were designed using a "reverse hotspot" approach that sought to form a pocket around exposed hydrophobic residues on the target surface as in the crystal structure of Fz27 (B12). Error-prone PCR mutagenesis of these proteins to test near-neighbors to the original designs resulted in the discovery of a high-affinity ankyrin binder, ANK12, detailed in Section Four of this work.

LRR based designs were generated utilizing a multi-step approach to generate a variety of binding modes including larger designs which utilized the LRR extension protocol to

maximize interface size and potential antagonist activity. A set of small LRRs consisting of several repeat modules were globally docked against the target, and conformations were manually selected based on the ability of the scaffold to antagonize Wnt upon extension. These docked scaffolds were locally refined utilizing MotifDock and then extended to enable a second contact at another location on the target surface. After interface optimization, 39 of these designs were tested (Figure 2E). These design efforts did not yield any detectable binding activity. Since the design of even a single de novo interface remains difficult, the simultaneous design of 2 remote interfaces may currently represent a design problem that is currently not feasible. Therefore, additional design efforts tested LRR-based binders which had more limited interfaces (Figure 2F), with the intention that working designs could be later extended to expand the interface to improve binding affinity and provide additional contacts to encode Frizzled subtype specificity. This approach differed from the previous LRR designs since the essential modules were be evaluated for binding prior to extension. It was hoped that this would increase the success rate by reducing the complexity of the design problem. However, unfortunately none of these LRR design efforts yielded an experimentally confirmed binder.

As can be seen by the limited success rate of the varied, extensive design efforts against this target, computational design of de novo interfaces cannot yet be considered an automatic, deterministic process. In particular, the recapitulation of a protein-lipid interaction with a protein-protein interaction was a non-trivial problem, and the anticipated success rate of designs was realistically low. In the absence of preexisting protein-protein information it was unknown what binding concept would be successful. Therefore, in order to increase the chances of discovering a viable binding epitope, as many different binding modalities as possible were tested. The successful targeting of this interface using two distinct, novel binding modes

demonstrates that multiple solutions to a given protein design problem are likely to exist. Given both the difficulty of de novo interface design and the possibility of multiple orthogonal solutions, successful rates in computational design are maximized by increasing the number and diversity of attempted designs, as in this case.

In particular, the successful computational design of ANK12 demonstrates vividly the advantages of rational design methods. Protein design enables explicit targeting of the relevant functional surface on specific Fz receptors, a capability not duplicated by alternative antibody based binder discovery approaches. The rational design of a particular binding mode allows the complete blockade of the Wnt binding site via contacts of both conserved and non-conserved regions. This enables downstream optimization to develop subtype-specific variants. Other binding modes would not necessarily be capable of this if they either interacted wholly with conserved or subtype-specific (as in the case of B12) residues. Therefore, it is a distinct advantage of the rational design process to not only target the functional site but specify the binding mode in a much more specific manner.

Additional development is also made more feasible by the stability and rigidity of the canonical backbone of the repeat protein scaffold used here. Other non-idealized scaffolds would likely not tolerate additional mutagenesis, but because of the strength of the conserved core interactions of these scaffolds, extensive mutagenesis of the interface should not affect the appropriate folding of the designed protein. The physical characteristics of ankyrins therefore are not only an advantage for downstream therapeutic applications but also are superior during the development and protein engineering process. The adaptability of such proteins for downstream redesign and engineering is in itself a strong rationale for the utilization of idealized or de novo non-native proteins as scaffolds for these applications. The computational design and

optimization of ankyrins represents a synergistic combination of rational design methods with privileged scaffolds, and should be considered an alternative pipeline that merges the best rational and empiric methods in the protein engineering field.

1.6 Optimization of Designed Binders and Functionalization as Wnt Modulators

Computationally designed proteins were experimentally assayed for binding and their binding modes were confirmed via competition assay or mutagenic knockout of the interface. Affinity maturation of designs with initial activity was used to lower binding affinity prior to structural characterization. Designs were then redesigned utilizing obtained structural information and experimentally optimized to improve properties for downstream applications. In the case of ANK12, structural and experimental data was utilized to target multiple Frizzled subtypes, progressing towards a comprehensive suite of specific and broad-spectrum Wnt antagonists and agonists (after functionalization by fusion to LRP6 binding domains). Subsequent validation in relevant biologic contexts by collaborators demonstrated their functional utility and potential for therapeutic intervention in Wnt signaling.

1.7 Overview of Following Sections

The following sections will discuss in greater detail the development of those proteins that were confirmed to have binding activity and were subsequently developed as Wnt antagonists and agonists. The subject of section two is the design and maturation of Fz27, a novel helical Frizzled-5/8 specific binder, as well as the functionalization of that binder as a Wnt agonist. The third section will focus on the complete redesign of the interface-contacting loop of B12 (the optimized Fz27) to improve Wnt antagonism as well as additional optimization efforts resulting in substantial improvements in stability. This work resulted in a potent, Frizzled-5/8

specific Wnt antagonist. The fourth section will describe the first rational computational design of an ankyrin-based binder, as well as its subsequent optimization and redesign to alter its Frizzled subtype specificity. The work discussed in the second section has been accepted for publication, however the discussion here will focus on the computational design of B12 rather than its functional validation as an agonist, as was primarily the case in the published work. The work discussed in the third and fourth sections is the subject of forthcoming manuscripts, while Appendix 1 discusses alternate approaches pursued in the course of this work.

1.8 Figures

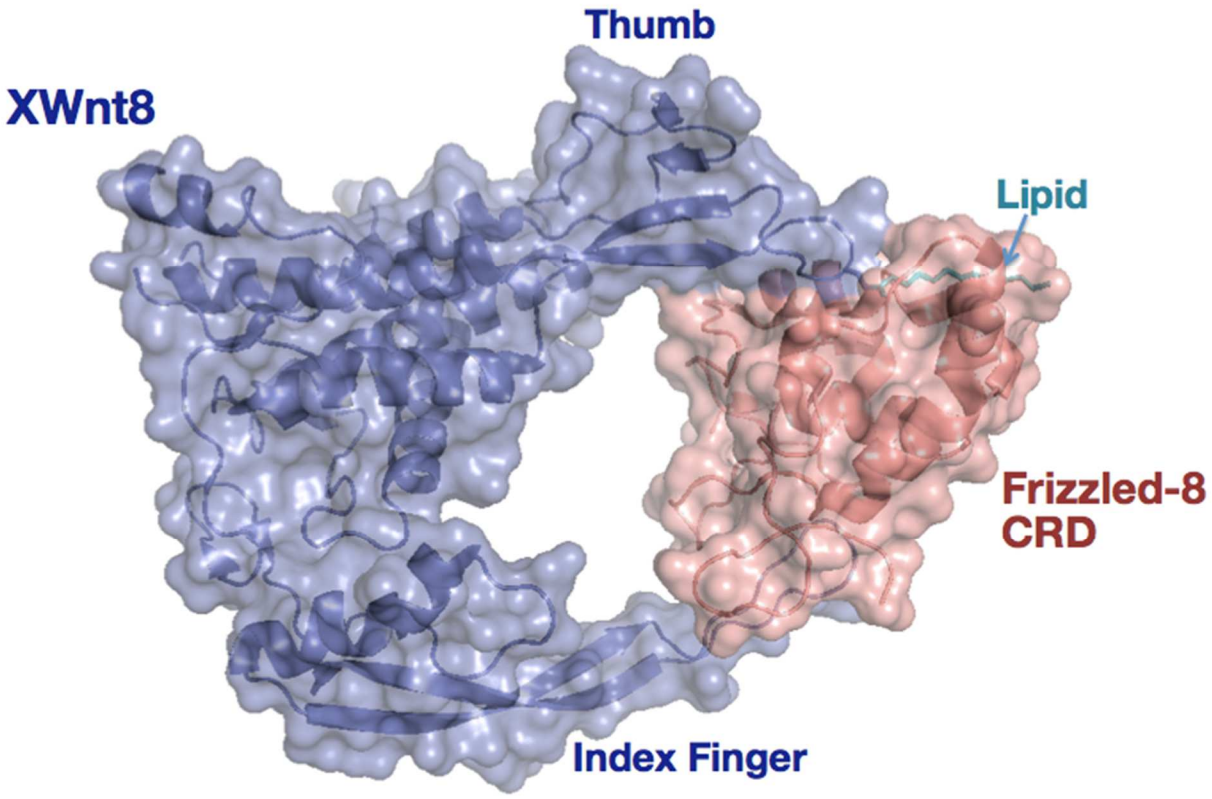


Figure 1. Structure of the Wnt-Frizzled Complex.

Wnt (blue) interacts with Frizzled (salmon) via a two-site binding mode. Utilizing its "Index Finger" interaction, the Wnt ligand interacts utilizing shallow, protein-protein contacts. At the alternate, "Thumb" portion of the pincer binding mode, the Wnt relies upon the insertion of a covalent paltimoleic acid modification (teal) into the hydrophobic groove of the Frizzled-8-CRD.

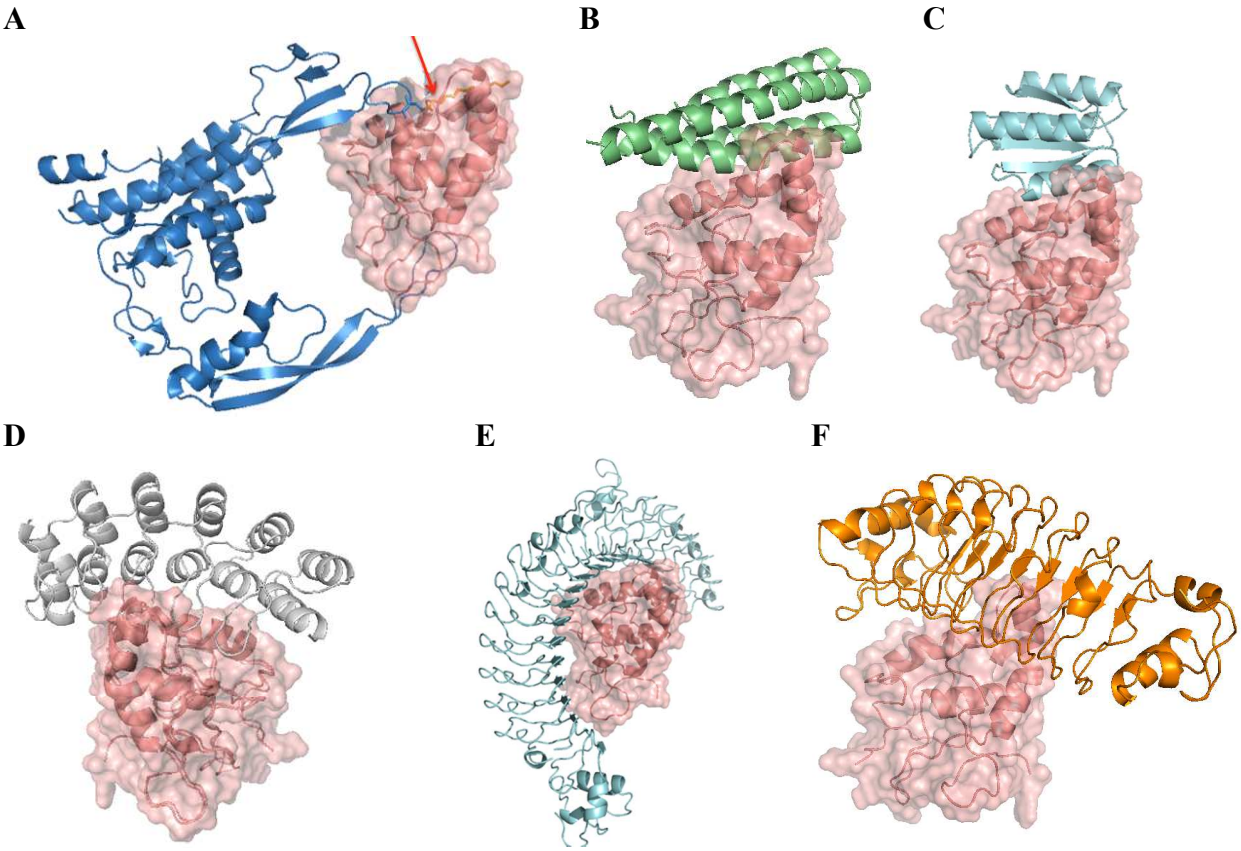


Figure 2. **Computational Design of Wnt Binders Utilizing Diverse Scaffolds and Binding Epitopes.**

(A) Native structure (4F0A) of Wnt (blue) bound to Fz-8-CRD (salmon), lipid-binding groove noted by red arrow. This functional site was the target of efforts to design Frizzled binders for use as Wnt modulators. (B) First generation alanine helix designs (green) utilized a minimal alanine helix as the binding epitope and native helical proteins as scaffold proteins. (C) Second generation alanine helix designs (teal) generated by grafting of designed helical epitopes onto characterized de novo scaffolds. (D) Ankyrin based designs (grey) utilized a novel, "hand-in-groove" binding mode which simultaneously grasped an exposed alanine helix and inserted loops into the Fz lipid-binding groove. (E) 2-site Leucine Rich Repeat (LRR) designs (aqua) utilized LRR extension protocol devised by Keunwan Park to generate largeinterfaces with spatially distant contacts. (F) Second generation LRR designs (orange) utilize a single, non-extended binding mode.

Section 2: Design and Optimization of the Novel Frizzled-8 Specific Binder B12

2.1 Background and Computational Design of Novel Minimal Alanine Helix Binders

Wnt utilizes a lipid modification to interact with the hydrophobic groove of its Frizzled receptor (Figure 3A). This region was targeted for the design of novel Wnt antagonists since inhibition at this site has demonstrated functional relevance in an anti-cancer context[6]. In contrast to typical computational design projects in which preexisting native protein-protein interactions derived from crystal structures inform design efforts, the native Wnt-Frizzled protein-lipid interaction provides limited structural information to aid epitope design in the form of graftable "hot-spot" residues since the hydrophobic lipid makes the primary affinity-determining contacts. Previous analysis of protein interfaces has found that certain interactions contribute disproportionately to the energy of binding, which is concentrated rather than being spread equally among all interface contacts[69]. These interactions are described as "hot-spots", and previous computational design of protein interfaces have relied heavily on this phenomena by utilizing known hot-spot residues from existing structures as the primary driver of the free energy of binding for the designed interfaces of new binders, although other additional interactions are designed as well. Since these interactions are sufficient for binding and their precise spatial positioning and geometric orientation is known, the protein design problem is much more feasible in such cases. However, unfortunately in this case there are no hot-spot residues from the Wnt protein that can be used as a template for design of novel interfaces.

Although there are no individual residues that can be wholly grafted onto a designed interface from the native complex, the overall binding mode can still be used to inform design of de novo binders. The interactions that the Wnt lipid makes with the Frizzled binding cleft are by

definition entirely hydrophobic, given its inherent chemical properties. Recapitulation of this interaction, therefore should try to make similarly strong hydrophobic contacts with the binding cleft. The native binding mode is also notable for its full insertion into the Frizzled binding groove. Designed binding epitopes should therefore attempt to pack as deeply as possible into this cleft. Initial efforts attempted to find optimal residue-level contacts by docking disembodied hydrophobic amino acids into the cleft, and then to combine these into interfaces utilizing hot-spot placement protocols similar to those used to design hemagluttinin binders[42]. However, in the absence of pre-optimized, known native hotspots, the results from these efforts were unsatisfactory.

The protein design problem for previously developed binders utilizing the hot-spot placement approach was in large part comprised of the search for scaffolds compatible with the set of geometric constraints imposed by the location of defined hot-spot residues. This approach was limited by available scaffold topologies that fit these multiple criteria, but upon successful placement of hotspots, acquisition of functional binding activity could be reasonably expected. In this case, since there were no known hot-spot residues, an alternative to the placement of a combination of individual interactions was to optimize a larger binding epitope via mimicry of the native lipid. For this purpose, a small, idealized helix was made as hydrophobic and featureless as possible (similar to the native lipid) by mutating the interface-facing residues to alanine. By minimizing the size of the side-chain residues, it was hoped that the alanine helix would be able to pack more fully into the Frizzled lipid cleft. This would maximize the Wnt antagonist properties of the designed protein and also hopefully lead to better larger overall interfaces and improved shape complementarity of adjacent secondary structures after grafting. The placement of the minimal helical epitope was optimized via docking against the targeted

region and then it was grafted onto a large set of helical proteins from the PDB (Figure 3). This process yielded candidate designs whose existing backbones were compatible with the proposed binding epitope. Candidate binding modes were manually selected based on the size and quality of additional contacts made by other secondary structure elements in addition to the common helical epitope. This set of designs was then computationally redesigned using RosettaScripts design protocols to optimize the binding interface. Optimized designs were filtered based on quality metrics (SASA (interface size), DDG (binding energy), shape complementarity, TotalScore (overall protein energy), average degree (extent of interface connectivity)) and manually refined in Foldit. 99 optimized designs were ultimately generated from this strategy for experimental validation (Table 1).

2.2 Experimental Validation of Computational Designs

Designs were tested by expression of binders in a vector (pETCON) that facilitates surface display on yeast via fusion to the native surface protein Aga2. Incubation with biotinylated target protein (Fz8-CRD) and subsequent labeling with secondary fluorescent labels enabled robust detection of binding by flow cytometry. All 99 first generation alanine helix designs were tested individually, however only design, Fz27, showed detectable and specific binding activity, with a yeast surface display Kd of 70nM under avid conditions (Figure 4).

Non-specific binding or target-specific interactions alterative to the designed binding mode are possible, therefore mutagenic studies were undertaken in order to verify Fz27 was specifically binding Fz-8 utilizing the designed interface. Mutations substituting a charged residue for alanine residues in the core of the designed interface will disrupt binding if the design is binding utilizing the designed interface. Alternatively, if the design is either unfolded or binding in a non-specific manner via other surfaces, then these mutations would likely have no

effect upon binding activity. Another possibility is that the native scaffold could have existing binding activity, so to confirm that any binding functionality is the result of design, the parent scaffold should be tested as a negative control. In the case of B12, knockout variants A173R and A174D dramatically abrogated binding, and parent scaffold 2QUP showed no background binding (Figure 4). These controls provided validation that the protein were indeed binding as designed and supported further development of the molecule as an Wnt inhibitor. However, mutagenesis or competition assays may not always be capable of full discrimination between designed and alternative conformations, as will be seen later. Therefore structural studies will be the ultimate validation of designs prior to further redesign and utilization in downstream functional tests by collaborators.

2.3 Affinity Maturation of Fz27

Although the initial binding activity of Fz27 was definitive, the existing affinity of this binder was insufficient for downstream functional applications. When labeled with biotinylated Frizzled target protein oligomerized via tetrameric streptavidin, the protein bound with a affinity of approximately 70 nM. However, under non-avid conditions there was no detectable binding, suggesting that the binding off-rate as a monomer remained higher than optimal. In order to drive the binding affinity lower, empiric optimization of binding activity was required.

Affinity maturation consists of diversification of a parent sequence followed by selection for desired properties. A crude, but powerful method for diversification is error-prone PCR, during which DNA replication with an error-prone polymerase introduces mutations to the wild-type sequence. However, error-prone libraries do not fully explore sequence space, since these errors are stochastic and undefined. Therefore, in order to more comprehensively and rationally optimize the binder, site saturation mutagenesis methods previously used successfully to

empirically improve the functional properties of hemagluttinin binders[70] were selected as the approach to affinity maturation here. In this approach, mutagenesis of a protein at each position to all twenty amino acid identities yields a library of all possible single point mutants. Subsequent selection for target binding and deep sequencing of the original and selected libraries yield enrichment and depletion ratios for all individual amino acid mutations. This data can be used to rationally engineer the protein as desired.

The Fz27 site saturation mutagenesis (SSM) library was constructed using the Kunkel method[71]. This library was expressed on the surface of yeast, and selected for binding activity via cell sorting methods for multiple rounds. Illumina deep sequencing of the naïve and selected pools and comparison of the frequencies in each library yielded quantitative enrichment rations for each point mutation (Figure 5A). The most enriched single mutations lowered binding affinity by 5-8 fold relative to the wild-type design (Figure 5B). Fz8-CRD binding remained undetectable under non-avid conditions, demonstrating that single mutations were not sufficient to achieve the desired nanomolar-level binding affinity.

In order to maximize the binding affinity of the design, the top eleven most enriched positions were selected for inclusion in a combination library (Table 2). This library included the wild-type amino acid as well as enriched amino acid identities at each of these positions, so that all permutations of these known, enriched, defined mutations were included in the library. This library was constructed by DNA oligomer assembly, expressed via yeast surface display, and sorted to convergence. This ultimately yielded a binder (now referred to as B12) that contained mutations at eight of the eleven variant positions (Table 3) and bound Fz-8 with high affinity under non-avid conditions (Figure 5C), a marked improvement. The successful affinity maturation of B12 demonstrates the power of the SSM approach. Rational combination of

multiple mutations that alone make only a small improvement in binding affinity can result in a dramatic increase in binding affinity. It may be noted that this can be achieved through maturation of a single library, rather than several iterations of diversification and selection as might be necessary with an error-prone PCR approach.

2.4 Structural and Functional Characterization of B12 as a Wnt Modulator

Note: Structural and functional data for B12 and the B12-Dkk1 fusion protein is summarized below and discussed in the conclusions of this chapter. A more complete description of these efforts and discussion focusing on the biological implications of these studies can be found in the following manuscript:

Janda CY*, Dang LT*, You C, Chang J, Lau W, Zhong ZA, Yan KS, Marecic O, Siepe D, Li X, Moody JD, Williams BO, Clevers H, Piehler J, Baker D, Kuo CJ, Garcia KC. Surrogate Wnts that phenocopy canonical Wnt/B-catenin signaling. *Nature.* 545; 234-237. (May 11, 2017)

Collaborators (Claudia Janda, K. Christopher Garcia) solubly expressed B12 and determined that the novel protein bound Fz-5 and Fz-8 with respective Kds of 9 and 18 nM[72], with no detectable binding of Fz-1, Fz-4, Fz-7, or Fz-10 . A 3.3 Å crystal structure of the B12-Fz8-CRD complex revealed that the design bound in an alternative conformation adjacent to the lipid binding cleft but interacted peripherally, rather than occluding the entire cleft as designed (Figure 6). Although this binding mode was not the computationally designed one, it did utilize the designed interface residues, providing an explanation for why the knockout variants did not bind to the target. Rather than inserting the designed alanine helix into the hydrophobic binding cleft of the Fz receptor, B12 interacts with the target in an unexpected fashion. Notably, the design displays non-canonical helical spacing, resulting in the formation of an inter-helix pocket that acts as a "reverse-hotspot" to pack around an exposed tryptophan residue on the Fz-8-CRD surface (Figure 7). As previously discussed, most previous examples of designed interfaces have

utilized hot-spot residues that insert into hydrophobic cavities on the target surface. Therefore, the B12 reverse hot-spot binding mode provides a novel approach to the design of hydrophobic interfaces. This strategy would represent a reversal in design strategy since the concave portion of the interaction now appears on the binder, rather than target side of the interface.

Interestingly, a tryptophan at this position is found only in Fz5/8, which provides an explanation for the specificity of the binder (Figure 8). While B12 binds tightly to Fz5 and Fz8 it does not bind to the other Fz subtypes tested (1/4/7/10), presumably because this interaction is the primary energetic driver of binding and does not occur in other subtypes (Figure 9). This unique tryptophan residue therefore provides a powerful structural feature for subtype specific discrimination enabling functional study of subtype related signaling differences. Exploitation of similar structural features in other subtypes may be similarly fruitful, therefore the B12 binding mode provides an example for effective discrimination of closely related proteins.

B12 was developed as a Wnt antagonist and structural data demonstrates that it binds adjacent to the targeted site. Fourteen amino acids comprising the region of the protein which is closest to the lipid binding groove are not resolved in the crystal structure. However, leucine residues 57 and 62, near the unfolded, disordered portion of the interface-contacting region, insert into the lipid-binding groove in positions overlapping with the native Wnt lipid (Figure 10). Indeed, B12 does exhibit Wnt antagonism, albeit weak. The weak functional antagonism of this protein is likely due to its binding footprint, which results in a more incomplete blockade of the targeted cleft than that of the designed binding mode. In addition, the disordered structure at the relevant interface positions may explain why the functional antagonism of the protein is less than might be expected. Additional optimization of B12 was required in order to confer potent Wnt antagonism; these efforts are detailed in the third section of this text.

Complete investigation of Wnt signaling is enabled by the controlled activation of the pathway in addition to effective inhibition of endogenous signaling. The tertiary structure of the Wnt-Fz-LRP6 complex remains unresolved, therefore engineering of Wnt mimics which crosslink Fz and LRP6 proceeded via experimental validation in the absence of definitive structural information about required spatial or geometric constraints. If the primary purpose of Wnt is to co-locate its receptors, then precise orientation may not be required for activation. Therefore, in order to engineer Wnt agonists, collaborators (Claudia Janda, K. Christopher Garcia) fused the B12 binder to DKK1 (a native Wnt inhibitor which natively binds LRP6 at the Wnt-LRP6 interface) via flexible glycine-serine linkers of 0, 5, 10 and 15 amino acids. A previously characterized scFv which broadly bound Fz-1/2/5/7/8 was fused to Dkk1 in order to facilitate comparison of modulators with varying binding specificities. Wnt is postulated to initiate signaling via clustering of the Fz and LRP6 receptors. Successful activation of Wnt signaling by these engineered, bi-specific fusion proteins would confirm that co-location of these receptors is sufficient for agonism.

Indeed, dual-color single-molecule fluorescence imaging of HeLa cells expressing both Fz-8 and LRP6 demonstrated that treatment with either surrogate molecule resulted in receptor dimerization, co-localization, and co-locomotion. In addition, the surrogates differentially induce signaling based upon Frizzled receptor expression profile. The B12-Dkk1 surrogate activates signaling only in cell lines known to express Fz-8 (A549 cells or L-cells transiently expressing Fz-5 or Fz-8), while the scFv-Dkk1 surrogate is also active in cell lines expressing Fz-7 (L-cells), or Fz-2 (A375, SH-SY5Y). This is in agreement with the known binding profiles of the Fz-interacting moiety of these surrogates, and confirms that activation of Wnt signaling is indeed

dependent upon recruitment of the receptors corresponding to the Wnt ligand's binding specificity.

These fusion molecules are wholly comprised of protein (in comparison to the lipid-modified Wnt ligands) and therefore represent water-soluble Wnt surrogates that may be more easily expressed and utilized in scientific and therapeutic applications. Although stability concerns regarding B12 limited its current application in vivo settings, the scFv-Dkk1 Wnt surrogate was confirmed to act as a potent agonist in several additional functional tests. Treatment of mesenchymal stem cells with the scFv-Dkk1 surrogate (as well as a RSPO2-linked version) resulted in up-regulation of the osteogenic marker alkaline phosphatase (ALP) in a BMP2 synergistic manner, in agreement with the known role of Wnt in bone formation[73,74]. In order to test the ability of these surrogates to effectively act as stem cell niche factors inducing cell proliferation, they were tested in cells in which expression of endogenous Wnt ligands were inhibited by the Porcupine inhibitor IWP-2. In this setting, these novel surrogates successfully induced pancreatic, stomach, colon and liver organoid formation, in a manner consistent with previous reports of Wnt-dependent stem cell renewal [75,76,77,78]. Expression of these surrogates in vivo in mouse hepatocytes was achieved via recombinant adenovirus and resulted in activation of glutamine synthetase, a known marker of Wnt's role in the metabolic zonation of the liver[79,80]. Moreover, parabiosis (conjoining of circulatory systems of donor and recipient mice) experiments demonstrated that recipient mice received soluble Wnt surrogates via circulation and that these proteins were functionally active, with corresponding increases in AXIN2 and glutamine synthetase expression. Therefore, this scFv-derived Wnt surrogate may be considered proof of concept for the controlled activation of this pathway in vivo using novel, engineered

Wnt mimics. This initial success provides additional motivation for the continued development of stable, subtype-specific Frizzled binders for use in this context.

2.5 Conclusions

The successful development of a novel, high-affinity, specific binder against the Frizzled-8 receptor demonstrates once again the utility of protein engineering methods for clinically relevant applications, including for challenging cases that require de novo interface design. Although B12 did not interact as designed with the receptor target, its binding mode serves as an example to future design efforts in several respects. First, the hydrophobic pocket B12 uses to interact with the exposed Frizzled-8 tryptophan is formed primarily by the backbone of the parent four-helix bundle topology. Therefore, B12 is an example of a novel protein that binds primarily via a concave pocket formed by secondary structure rather than only through side-chain hot-spots which insert into corresponding concave features on the target as has typically been the case in previous examples of protein design. Future design attempts may therefore seek to identify exposed hydrophobic residues on a target surface and form hydrophobic clefts or pockets to pack around these hydrophobic features. This is a novel approach that may allow engagement of interfaces utilizing molecular mechanisms heretofore unavailable to hotspot design methods. Indeed, the particular binding mode of B12 provides a robust molecular mechanism for subtype specific discrimination that other interfaces would not, demonstrating the practical utility of this approach. Designing concave surfaces for one-sided interface design also makes sense on a philosophical level as well.

Protein-protein interfaces are rarely flat and the global and local shape complementarity of these interactions is responsible for their binding specificity, without which the hydrophobic surfaces of proteins would interact in a broad, non-specific manner. Therefore, the hydrophobic

interactions of a protein-protein interface are comprised of a variety of knobs, pockets, and grooves that pack together efficiently in manner unique to a particular interface, enabling binding specificity. Since for each inserting interaction there must be a corresponding concave interaction, if protein design methods are limited to hot-spot design they can only target surfaces that have suitable concave topologies. Many biologically relevant target proteins do in fact have such surfaces, but obviously (as in this case), examples exist where targeting an exposed hydrophobic feature through design of a complementary concave interface would be optimal. In addition, a single protein-protein interface typically contains several key interactions, including both features that insert into concave regions of the binding partner and concave pockets which pack around inserting features of the binding partner. So the optimal design of a single protein interface may include hot-spot interactions as well as concave pockets which provide a partner for hot-spot residues on the other side of the interface. Therefore, further exploration and development of design methods enabling the design of concave binding pockets will expand the range and capability of one-sided interface design. The design of an ankyrin binder with a concave interface will be described in section four and provides another example of how this principle may be applied in practice.

Secondly, due to the fact that B12 interacts primarily with a residue that occurs specifically only in Fz-5/8, it is a highly specific binder suitable for discrimination of the contributions of the Fz-5 and Fz-8 receptors to various biologic processes. Targeting a structural feature specific to a given protein may be considered a general approach that can be used strategically in the design of interfaces which enable the high-affinity, specific discrimination of a chosen protein while limiting off target effects. Indeed, in this very application the design of a suite of subtype-specific binders against all Frizzled subtypes would provide the opportunity for

precise study and intervention in Wnt signaling otherwise by limited by the crude inhibition of a broad-spectrum binder. However, because the burial of the Frizzled-5/8 specific tryptophan is integral to the binding energy of the B12 interface, this particular binding mode may not be adaptable to other Frizzled subtypes. Rather, additional orthogonal designs must be pursued in order to facilitate modulation of Wnt signaling through these receptors. The fourth section of this text will discuss further efforts towards the design of additional binders which was undertaken with the express purpose of developing binders against additional Frizzled subtypes through the targeting of both conserved and varying regions of the receptor.

The utilization of a concave interface pocket as a central interaction and the specific targeting of a unique structural feature can be considered successful elements of the B12 interface that can be applied to future design approaches. There are also shortcomings in the design of B12 that can be learned from as well. Most notable is the failure of the protein to bind as designed. This failure can be attributed to the fact that the protein does not fold to the designed structure, rather than that the designed structure does not bind appropriately. Missing density in the crystal structure corresponding to the loop that connects the two interface-contacting helices demonstrates that this region of the protein is not ordered. Indeed the canonical helical structure seen in this region the parent 2QUP scaffold and designed binder becomes disordered due to interface-enabling mutations. In particular, a valine at position 81 in the 2QUP parent is mutated to glycine in the Fz27 design for the purposes of interface optimization. However, in the crystal structure the helical structure is disordered beginning at this site, therefore it may be postulated that this mutation is responsible for much of the instability of the resulting design. More broadly, however, the instability of the design may be largely attributed to the marginal stability of the 2QUP parent scaffold, which itself contains a

disordered loop from residues 91-93. This was ignored as a crystal artifact distant from the relevant interface during the design process, but in retrospect the instability of the overall protein argues for more rigorous scaffold selection criteria.

During the design process for B12, native helical proteins were selected from the PDB primarily for their compatibility with the proposed binding epitope. Since developing a de novo binder is so challenging, it is understandable that interface quality would be prioritized over other considerations such as the characteristics of the parent scaffold. But in the larger context, the development of binders that are unstable, vulnerable to protease digestion, or poorly expressed is not optimal for the purpose of obtaining proteins suitable for downstream therapeutic applications. Although such designs may provide good information about potential binding modes to inform additional design efforts, they may not constitute good lead candidates in and of themselves. As will be seen in the third section of this text, such binders will necessitate lengthy and time-consuming redesign and optimization in order to improve them as proteins suitable for functional use.

Additionally, it is likely that the use of unstable proteins as scaffolds significantly decreases the success rate of computational design. One general obstacle in protein design is the lack of validation of non-binding designs. If a design does not demonstrate detectable function, it typically is abandoned and the reasons for design failure are never determined. Although this is non-ideal from a scientific perspective, pragmatism dictates that protein design efforts typically pursue many design candidates in a high-throughput method rather than focusing on elucidating the causative failure in non-working designs. Designs may fail because the proposed binding epitope is flawed (perhaps only due to even a single error in a large interface), because the design does not fold to the desired structure, or because of an error present in the target crystal structure.

Of these sources of error, the most easily addressed is that of protein misfolding, which can be greatly reduced through the use of scaffolds with high stability. These robust scaffolds will be more tolerant of the extensive mutagenesis that is needed in order to confer new functionality to an existing protein. This stability may be considered a margin-of-safety for design that avoids failure due to misfolding, as was seen for portions of the B12 protein. And for downstream applications, such proteins will be preferable as well. Indeed, in the optimal case scaffolds will be used that are privileged in other ways for downstream applications so that upon development of binding activity these proteins may be more readily adapted. The fourth section of this text will provide an example of the utilization of ankyrin repeat proteins for precisely this reason.

From a functional point of view, the development of B12 provides proof of principle for the engineering of subtype-specific novel modulators of the Wnt pathway. Although as a result of its alternative binding mode, B12 did not exhibit functional antagonism as potent as its binding affinity would suggest, it is still a weak Wnt antagonist. In addition, the disordered structure at key interface contacting regions suggested that additional improvement of the design could result in more complete blockade of the targeted site. The next chapter will discuss computational efforts that used B12 as a starting point for the successful computational redesign of variants optimized for potent Wnt antagonism.

The successful engineering of water-soluble Wnt agonists is a significant development in the field of Wnt signaling, which had been limited by the poor solubility of native Wnt ligands. The design and optimization of Wnt surrogates confirmed that co-location by fused, bi-specific Fz and LRP6 binding proteins is sufficient for Wnt activation, providing further evidence for the hypothesized signaling mechanism. Moreover, the Fz-5/8-specific binding profile of B12 demonstrated that this activation corresponds to the binding specificity of the Fz-interacting

moiety, thus providing proof of concept for precise interventions into Wnt signaling utilizing Fz subtype-specific proteins. In vivo characterization of the broadly binding scFv surrogate demonstrated the potential utility of such agonists in a wide range of regenerative medicine applications. The activity of this first-generation surrogate emphasizes the need for subtype-specific Fz binders with characteristics suitable for such in vivo applications. The next (third) section of this work will detail efforts to stabilize B12 towards this end, while the fourth section will discuss the development of ankyrin repeat binders which combine favorable physical characteristics for downstream applications with a rationally targeted interface which enables potent Wnt antagonism and subtype discrimination.

2.6 Materials and Methods

Note: The materials and methods used to generate the results discussed in this chapter are described in the surrogate Wnt manuscript listed below. Additional XML scripts and protocols used in the computational design and optimization of B12 are included here as well for completeness, since this discussion focuses on the computational, structure-based design aspects of this work in more detail. A complete description of methods collaborators used to obtain structural and functional data discussed in this section may be found in the referenced manuscript.

Janda CY*, Dang LT*, You C, Chang J, Lau W, Zhong ZA, Yan KS, Marecic O, Siepe D, Li X, Moody JD, Williams BO, Clevers H, Piehler J, Baker D, Kuo CJ, Garcia KC. Surrogate Wnts that phenocopy canonical Wnt/B-catenin signaling. *Nature.* 545; 234-237. (May 11, 2017)

Computational Generation of De Novo Frizzled Designs Utilizing Minimal Alanine Helix

The native Wnt-Frizzled lipid-protein interaction was recapitulated with a protein-protein interface utilizing a helical epitope as a hydrophobic protein epitope equivalent to the native lipid. In order to optimally position this epitope, a thirteen-residue alanine helix was docked against the lipid-binding cleft of Frizzled-8 (Chain A of PDB 4F0A), utilizing Foldit to manually sample a variety of orientations and insertion depths[81]. This optimized helix was subsequently

grafted onto a diverse set of native helical proteins using the Rosetta Epigraft application to search for scaffolds with backbones that were compatible with the proposed helical binding epitope[82]. Candidate designs were selected visually in PyMol based on overall backbone shape complementarity, interface size, and the quality of additional contacts made by additional secondary structure elements. Manually curated designs were then subjected to computational interface optimization utilizing RosettaScripts[83]. This was necessary, because the preceding steps only selected scaffolds for compatibility with the proposed helical binding epitope, but did not fully optimize interface residues for binding energy.

Interface Optimization of Frizzled-8 Binder Designs

RosettaScripts provides a powerful scripting language for directing computational design utilizing the existing suite of Rosetta[84] functionalities. RosettaScripts XML protocols are launched via command line using a specified Rosetta application build and database. Along with additional specified command line flags, these protocols contain all of the necessary instructions for completing desired design tasks, whether it be scoring, remodeling, or designing the target structures. In this case, RosettaScripts was used to optimize the interface of the generated helical designs targeting to the Frizzled-8 lipid binding cleft.

An example command line used to launch the RosettaScripts application is given below:

```
/[PATH_TO_ROSETTA]/rosetta_scripts_static.linuxiccrelease –database
/[PATH_TO_ROSETTA_DATABASE]/rosetta_database -ignore_unrecognized_res -
overwrite -out:file:renumber_pdb false -ex1 -ex2 -nstruct 1 -correct -score:weights
score12prime -s [input PDB] -parser:protocol
/[PATH_TO_ROSETTASCRIPTS_XML]/design_sidechain_Fz.xml -in:file:native
[PDB_STRUCTURE] -hbond_params sp2_params -corrections:score:hb_sp2_chipen -
corrections:score:hb_sp2_peak_heigh_above_trough 2.0 -corrections:score:hb_sp2_amp
```

2.0 -lj_hbond_hdis 1.75 -lj_hbond_OH_donor_dis 2.6 -ignore_zero_occupancy false -chemical:exclude_patches LowerDNA UpperDNA Cterm_amidation SpecialRotamer VirtualBB ShoveBB VirtualDNAPhosphate VirtualNTerm CTermConnect sc_orbitals pro_hydroxylated_case1 pro_hydroxylated_case2 ser_phosphorylated thr_phosphorylated tyr_phosphorylated tyr_sulfated lys_dimethylated lys_monomethylated lys_trimethylated lys_acetylated glu_carboxylated cys_acetylated tyr_diiodinated N_acetylated C_methylamidated MethylatedProteinCterm

An example RosettaScripts XML used for interface refinement is given below:

```
<ROSETTASCRIPTS>
    <SCOREFXNS>
        <sfxn_hard weights=score12prime>
            <Reweight scoretype=hack_elec weight=0.0/>
        </sfxn_hard>
        <sfxn_hard_elec weights=score12prime>
            <Reweight scoretype=hack_elec weight=0.3/>
        <Set hack_elec_min_dis=3.0 hack_elec_max_dis=15.0
hack_elec_dielectric=1.0/>
        </sfxn_hard_elec>
        <sfxn_soft weights=soft_rep>
            <Reweight scoretype=hack_elec weight=0.0/>
        </sfxn_soft>
        <extra weights=score12prime>
            <Reweight scoretype=hpatch weight=1.0/>
        </extra>
    </SCOREFXNS>
    <TASKOPERATIONS>
        #general taskops
        <LimitAromaChi2 name=arochi2/>
        <IncludeCurrent name=inclcur/>
        <ExtraRotamersGeneric name=exrot ex1=1 ex2=1 extrachi_cutoff=1/>
        <DisallowIfNonnative name=nohis disallow_aas=GPH/>
        #design at interface
        <ProteinInterfaceDesign name=dz_surf design_chain1=0 design_chain2=1
jump=1 interface_distance_cutoff=16/>
        <ProteinInterfaceDesign name=dz_core design_chain1=0 design_chain2=1
jump=1 interface_distance_cutoff=10/>
        <LayerDesign name=core layer=core core=40.0
use_original_non_designed_layer=1 restrict_restypes=0/>
        <LayerDesign name=surf layer=boundary_surface core=40.0
use_original_non_designed_layer=1 restrict_restypes=0/>
        <RestrictResiduesToRepacking name=rrtr residues=%%hotspot_list%%/>
        <PreventChainFromRepacking name=pcfr chain=1/>
```

```
    </TASKOPERATIONS>
    <MOVERS>
        <FavorSequenceProfile name=fsp scaling=none use_native=1 chain=2
matrix=BLOSUM62 weight=0.2 scorefxns=sfxn_hard,sfxn_soft,sfxn_hard_elec/>
        <TaskAwareMinMover name=min_rb_sc bb=0 chi=1 jump=1
scorefxn=sfxn_hard task_operations=dz_surf,arochi2,inclcur,exrot/>
        <TaskAwareMinMover name=min_sc bb=0 chi=1 jump=0 scorefxn=sfxn_hard
task_operations=dz_surf,arochi2,inclcur,exrot/>
    </MOVERS>
    <FILTERS>
        <Ddg name=ddg threshold=0 scorefxn=sfxn_hard confidence=0 jump=1
repack=1 relax_mover=min_rb_sc/>
        <Ddg name=ddg_elec threshold=0 scorefxn=sfxn_hard_elec confidence=0
jump=1 repack=1 relax_mover=min_rb_sc/>
        <ScoreType name=total_score scorefxn=sfxn_hard score_type=total_score
confidence=0 threshold=0/>
        <ScoreType name=total_score_elec scorefxn=sfxn_hard_elec
score_type=total_score confidence=0 threshold=0/>
        <ScoreType name=hp scorefxn=extra score_type=hpatch confidence=0
threshold=0/>
        <SymUnsatHbonds name=suh jump=1 cutoff=1000 confidence=0/>
        <ResidueIE name=trp_ie energy_cutoff=-6.5 restype3=TRP interface=1
max_penalty=10 confidence=0/>
        <PackStat name=pstat confidence=0/>
        <CombinedValue name=combo_core confidence=0>
            <Add filter_name=pstat factor=1.0/>
            <Add filter_name=trp_ie factor=1.0/>
            <Add filter_name=ddg factor=1.0/>
            <Add filter_name=total_score factor=1.0/>
        </CombinedValue>
        <CombinedValue name=combo_surf confidence=0>
            <Add filter_name=rsce_lr_bb factor=0.2/>
            <Add filter_name=rsce_bb_sc factor=0.2/>
            <Add filter_name=rsce_sc factor=0.2/>
            <Add filter_name=suh factor=1.0/>
            <Add filter_name=trp_ie factor=1.0/>
            <Add filter_name=ddg_elec factor=1.0/>
            <Add filter_name=total_score_elec factor=1.0/>
        </CombinedValue>
        <Sasa name=sasa threshold=500 confidence=0/>
        <Sasa name=hsasa threshold=1000 hydrophobic=1 confidence=0/>
        <AverageDegree name=deg threshold=8.5 distance_threshold=8 confidence=0
task_operations=dz_surf/>
        <ShapeComplementarity name=sc jump=1 verbose=0 min_sc=0.60
confidence=0/>
        <RotamerBoltzmannWeight name=rot_boltz
```

```
task_operations=dz_surf,arochi2,inclcur,exrot,nohis,pcfr radius=6.0 jump=1 unbound=1
ddG_threshold=1.5 scorefxn=sfxn_hard temperature=0.8 energy_reduction_factor=0.5
repack=1 skip_ala_scan=0/>
        <TaskAwareAlaScan name=alascan
task_operations=arochi2,inclcur,exrot,dz_surf,nohis,pcfr jump=1 repeats=3
scorefxn=sfxn_hard_elec repack=1 report_diffs=1 exempt_identities=PRO,GLY,ALA
write2pdb=1/>
    </FILTERS>
    <MOVERS>
        <AtomTree name=ftree simple_ft=1/>
        Docking name=fadock score_high=sfxn_hard fullatom=1 local_refine=1
optimize_fold_tree=1 conserve_foldtree=0 design=0 ignore_default_docking_task=0/>
        <PackRotamersMover name=dz_pack_soft_core scorefxn=sfxn_soft
task_operations=dz_core,arochi2,inclcur,exrot,core,nohis/>
        <PackRotamersMover name=dz_pack_hard_core scorefxn=sfxn_hard
task_operations=dz_core,arochi2,inclcur,exrot,core,nohis/>
        <PackRotamersMover name=dz_pack_hard_surf scorefxn=sfxn_hard_elec
task_operations=dz_surf,arochi2,inclcur,exrot,surf,nohis/>
        <GreedyOptMutationMover name=dz_gopt_core
task_operations=dz_core,arochi2,inclcur,exrot,core,nohis filter=combo_core
scorefxn=sfxn_hard relax_mover=min_rb_sc/>
        <GreedyOptMutationMover name=dz_gopt_surf
task_operations=dz_surf,arochi2,inclcur,exrot,surf,nohis filter=combo_surf
scorefxn=sfxn_hard relax_mover=min_rb_sc/>
    </MOVERS>
    <APPLY_TO_POSE>
    </APPLY_TO_POSE>
    <PROTOCOLS>
        <Add mover=ftree/>
        <Add mover=fsp/>
        Add mover=fadock/>
        Add mover_name=min_rb_sc/>
        Add mover_name=dz_pack_soft_core/>
        Add mover_name=min_rb_sc/>
        <Add mover_name=dz_pack_hard_core/>
        <Add mover_name=dz_pack_hard_surf/>
        <Add mover_name=min_sc/>
        Add mover=dz_gopt_core/>
        Add mover_name=dz_pack_hard_surf/>
        Add mover_name=min_rb_sc/>
        Add mover=dz_gopt_surf/>
        <Add filter=total_score/>
        <Add filter=ddg/>
        <Add filter=hp/>
        <Add filter=lr_elec/>
        <Add filter=trp_ie/>
```

```
            <Add filter=deg/>
            <Add filter=rsce/>
            <Add filter=sasa/>
            <Add filter=hsasa/>
            <Add filter=sc/>
            <Add filter=pstat/>
            <Add filter=suh/>
            <Add filter=alascan/>
            <Add filter=rot_boltz/>
        </PROTOCOLS>
    </ROSETTASCRIPTS>
```

Rosetta-based computational design optimized the binding energy of the scaffold-target interfaces, but some mutations were made which could be considered dispensable. These mutations would be more likely to destabilize the protein than contribute substantially to increased binding affinity. In order to avoid excessive departures from the native scaffold sequence (and a corresponding increased risk of misfolding), a revert-to-native design protocol was used to remove any excessively aggressive mutations made by the interface optimization protocol.

An example command line used to run the Rosetta revert–to-native application is given below:

```
/[path_to_revert_design_application]/revert_design_to_native.static.linuxiccrelease -
database /[path to rosetta database]/rosetta_database -use_input_sc -ex1 -ex2 -correct -
overwrite -revert_app:wt [Wild_Type_Scaffold_PDB] -revert_app:design [Design_PDB] -
ignore_unrecognized_res
-out:file:renumber_pdb false -nstruct 1 -score:weights score12prime -hbond_params
sp2_params -corrections:score:hb_sp2_chipen -lj_hbond_hdis 1.75 -
lj_hbond_OH_donor_dis 2.6 -ignore_zero_occupancy false -chemical:exclude_patches
LowerDNA  UpperDNA Cterm_amidation SpecialRotamer VirtualBB ShoveBB
VirtualDNAPhosphate VirtualNTerm CTermConnect sc_orbitals pro_hydroxylated_case1
pro_hydroxylated_case2 ser_phosphorylated thr_phosphorylated  tyr_phosphorylated
tyr_sulfated lys_dimethylated lys_monomethylated  lys_trimethylated lys_acetylated
glu_carboxylated cys_acetylated tyr_diiodinated N_acetylated C_methylamidated
MethylatedProteinCterm
```

Designs were subjected to a final visual inspection prior to ordering. Electrostatics maps of both design and target surfaces were generated utilizing PyMOL (PyMOL Molecular Graphics System, Schrodinger LLC) and checked for interface charge complementarity. All mutations from the original scaffold were inspected to make sure that they were relevant to the interface, and spurious mutations that had not otherwise been reverted by the revert-to-native design protocol were identified, especially ones that might interfere with appropriate folding such as core-adjacent or core-facing mutations. Foldit was used to manually revert non-essential mutations to the wild-type identity as well as to optimize interface charge complementarity in order to reduce buried polar residues and engineer beneficial electrostatic interactions. Any NXS/T glycosylation sites or cysteine residues were removed as well, and any binder whose termini were close to the designed interface had a short five or six residue glycine-serine linker added to ensure that the yeast display fusion protein did not interfere with binding. Finally, designs were selected for experimental testing based on interface size (>1200 square angstroms), shape complementarity (>0.6), low numbers of buried polar groups (<4), and low binding energy (DDG). Amino acid sequences of final ordered designs are listed in Table 1.

Experimental Validation of Designs

DNA sequences corresponding to the selected computationally generated designs were generated utilizing the DNAWorks protocol[85] for optimization. Linear DNA was obtained from Gen9 (original designs) or IDT (knockout variants and wild-type scaffold 2QUP). Designs were tested experimentally by expression of binders in a vector (pETCON) that facilitates surface display on yeast via fusion to the native surface protein Aga2[86], a method previously used effectively for the optimization of designed proteins as well as engineered antibodies. EBY100

yeast were transformed, cultured in SDCAA media, and then transferred in in SGCAA media for 48 hours at 22°C to induce surface display of the designed proteins[87]. Yeast (1x10^6) were washed with PBSF (1xPBS, 1% bovine serum albumin (BSA)) and then labeled at 4°C for 2 hours with 1 uM biotinylated Fz-8 preincubated for 15 min with 0.25 uM streptavidin R-Phycoerythrin conjugate (SAPE, ThermoFisher, Waltham, MA) to form avid binding complexes. After labeling, samples were washed twice with ice-cold PBSF and analyzed on an Accuri C6 flow cytometer. Binding affinities were determined by varying labeling concentrations; non-avid affinities were obtained by first labeling only with biotinylated target protein prior to a subsequent labeling step with SAPE (samples were washed with PBSF after each step).

Affinity Maturation and Optimization

A site-saturation mutagenesis (SSM) approach to affinity maturation was used to improve the binding affinity of Fz27[70]. A full-coverage SSM library was constructed via the Kunkel method[88] utilizing forward and reverse primers (IDT) containing a degenerate "NNK" codon at each targeted position along with 21 bp flanking regions. This library was transformed[89] and induced using standard methods, and the top 1% of binding variants were collected over 3 iterative rounds utilizing the BD Influx cell sorter. Naïve and selected libraries were prepared and sequenced using an Illumina Miseq according to manufacturer's protocols. Data was processed and analyzed using standard genomic analysis software[90,91]. Amino acids comprising the most enriched mutations at eleven separate positions were selected for inclusion in a combination library with a theoretical diversity of ~800,000 distinct sequence permutations. This library was assembled and amplified from the oligos (IDT) listed below over 3 steps (UM1+2/UM3+4, UM1/2+UM3/4, followed by an amplification reaction) with Phusion High

Fidelity DNA polymerase (NEB) according to manufacturer's instructions, transformed, and selected to convergence iteratively over five rounds to obtain the matured variant B12.

DNA Oligos Utilized for Construction of Combination Library (Variant Positions in **Bold**):

>Ultramer1_fwd_164bp
□GCTAGTGGTGGAGGAGGCTCTGGTGGAGGCGGTAGCGGAGGCGGAGGGTC
GGCTAGCCATATGGGCGTGAGCTTTAGCGAAGTGATGGGCAAACAGAAAGAT
GAACAGGCGCGTGAACAGCTGAAAGAAGGCATG**AWA**AAAATTGAAGAACA
GGGCAAAAAGCT □

>Ultramer2_rev_133bp
□AGCCACGACGTTCTTCCAGATT**CKK**TCCCAGAAAACCCGC**CKSCRM**CGCA
AA**ARB**CGCCACCGCTGC**GRC**ATATTTCTGCAGCTCTTCCTGGGTACGGGTTTC
GCTCAGCTTTTTGCCCTGTTCTTCAATTTT □

>Ultamer3_fwd_147bp

AATCTGGAAGAACGTCGTGGCTTTAATCGTCGGGGCAAAGAAGAAATTGGCA
AAATTAGCGGAGAAGTGTAC**MWA**AAACTGTTAGACCTGAAAAAAGCGGTGC
GTGCGAAAGAAAAAAAGGGACTGGATATTCTGAATATGGTGGGC □

>Ultramer4__rev_128bp
GTTGTTATCAGATCTCTATTACAAGTCCTCTTCAGAAATAAGCTTTTGTTCGG
ATCCGCCCCCCTCGAGCGC**AKRAMYACK**CTCCAG**GRK**GCCCTTAATCTCGC
CCACCATATTCAGAATATCCAGTC

The original WT, enriched, and selected amino acid identities are listed in Table 3 for the eleven positions included in the Fz27 combination library for the affinity maturation of the final variant, now referred to as B12. As can be seen, most of the mutations were compatible and not mutually exclusive, as only R238 and Y240 were not included in the final selected variant. Given the close proximity of the C-terminal mutations in sequence space and their distance from the interface, it is thought that these mutations may act to stabilize the interface through long distance contacts. Therefore, these mutations may act in a similar way, and they may not be independent, in contrast to the other mutations, which optimize individual, independent interface

interactions.

Final Sequence of Selected Variant B12

>B12
GVSFSEFMGKQKDEQAREQLKEGMIKIEEQGKKLSETRTQEELQKYVAAVATFA
LQAGFLGPNLEERRGFNRRGKEEIGKISGEVYLKLLDLKKAVRAKEKKGLDILN
MVGEIKGTLERVYA

2.7 Acknowledgements

James Moody conceived the alanine helix binding epitope and also collaborated and advised on the computational design and affinity optimization of B12.

Chantzt Thomas was responsible for initial docking of the alanine helix and grafting step of the design process.

Claudia Janda solubly expressed B12, obtained structural information, and did functional characterization of monomeric B12 and the functionalized Wnt surrogates.

The collaborators on the Wnt surrogate manuscript performed functional characterization as detailed therein.

2.8 Figures and Tables

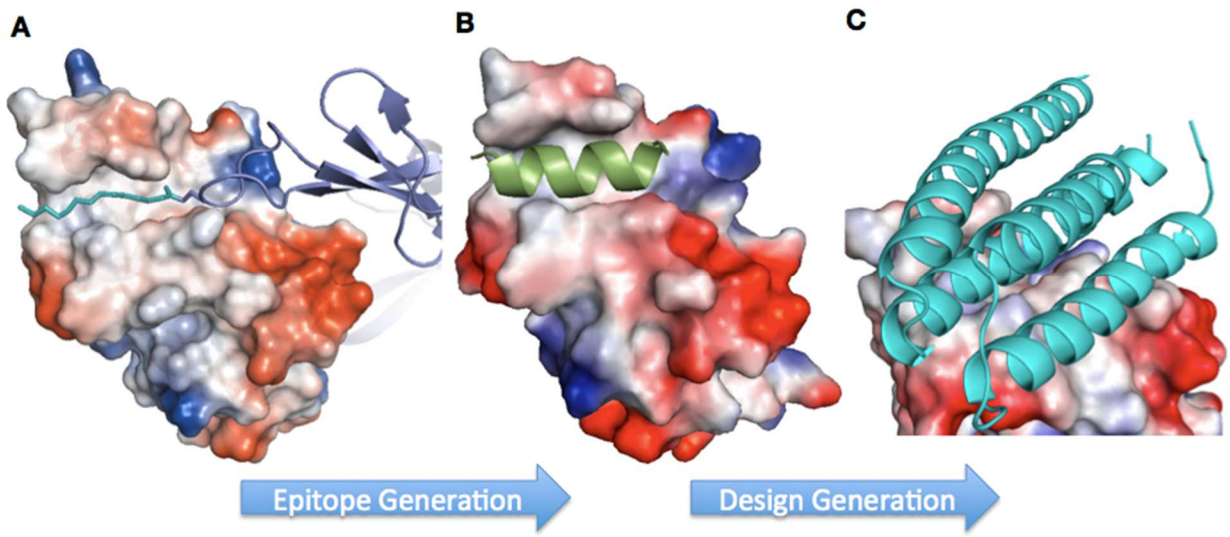


Figure 3. Computational Design of Frizzled-8 Binders With Alanine Helix Epitope.

A) Native Wnt8 (blue) inserts lipid (teal) into Frizzled-8 hydrophobic binding cleft. B) Minimal alpha helix (green) comprised of alanine residues utilized as a binding epitope substitution for the native Wnt lipid. C) Computational grafting of binding epitope with compatible scaffolds yields designs (example Fz7 shown in green) which contain the designed epitope as well as additional contacts from adjacent secondary structure elements.

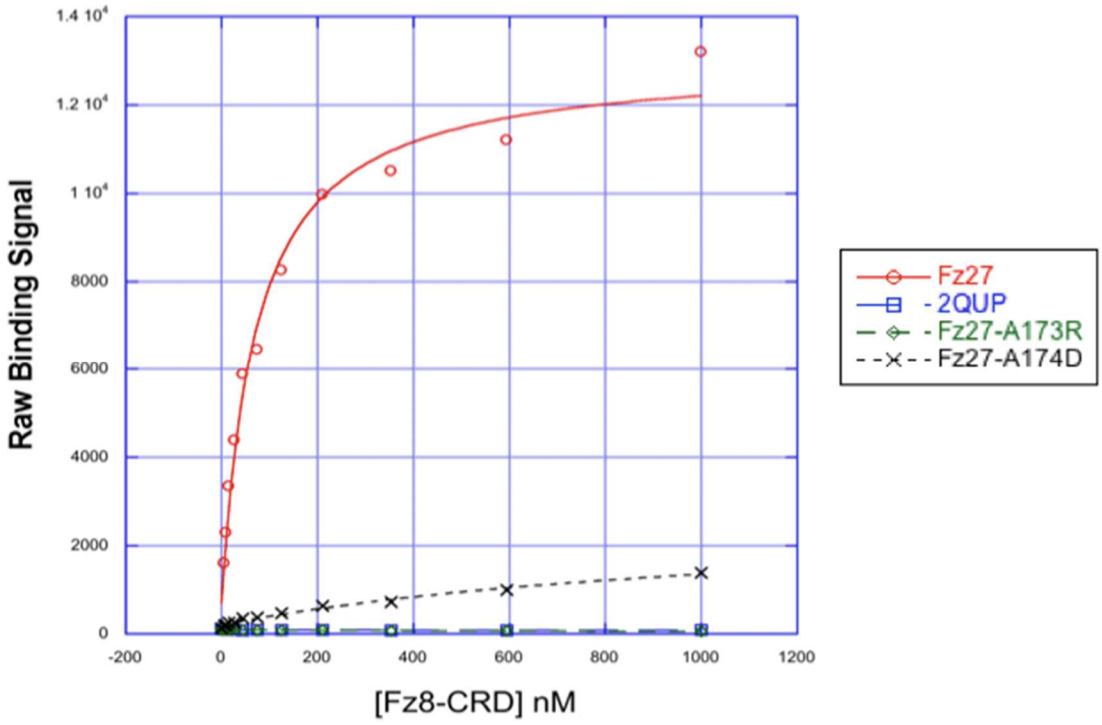


Figure 4. Design Fz27 Specifically Binds Frizzled-8 Utilizing Designed Interface.

Computationally generated design Fz27 binds Frizzled-8 as detected via yeast surface display with a Kd of 70 nM under avid conditions (labeling with tetrameric target protein). Design Fz27 utilizes designed interface to interact with target. Knockout of interface by mutagenic insertion of charged residues (A173R, A174D) into the designed interface demonstrates that these residues are involved in binding. Parent scaffold 2QUP shows no binding activity, again demonstrating that Fz27's function is due to design and not inherent to the existing scaffold.

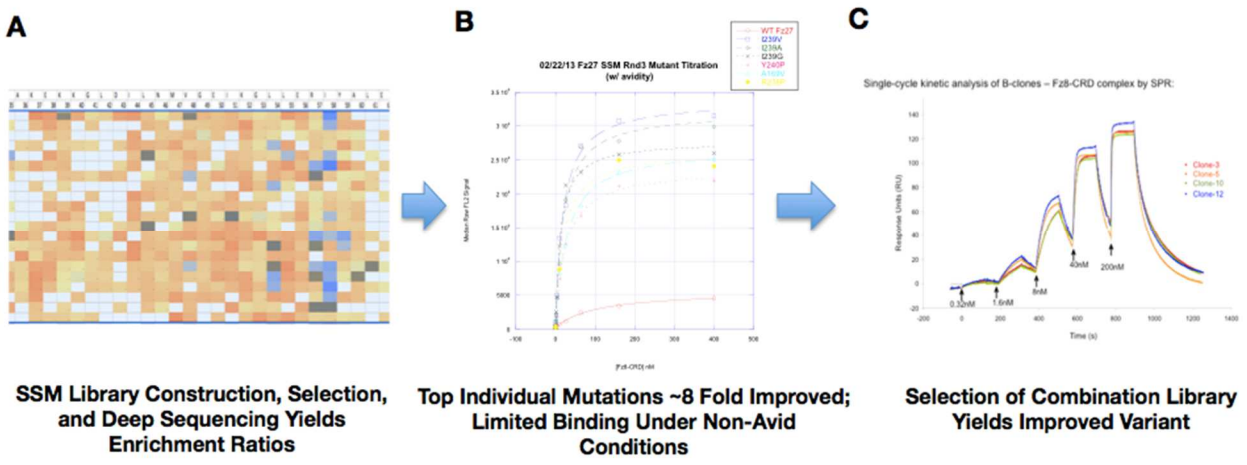


Figure 5. Enrichment Ratios Derived from Site-Saturation Mutagenesis of Fz27.

A) Site-saturation mutagenesis of Fz27 identified enriched (blue) and depleted (orange) point mutants. B) The most selected individual mutations (I239G) improve avid binding up to 8-fold compared to the original design, but still do not bind under non-avid conditions. C) Combination of identified mutations yields improved variant B12 which binds with nM affinity under non-avid conditions.

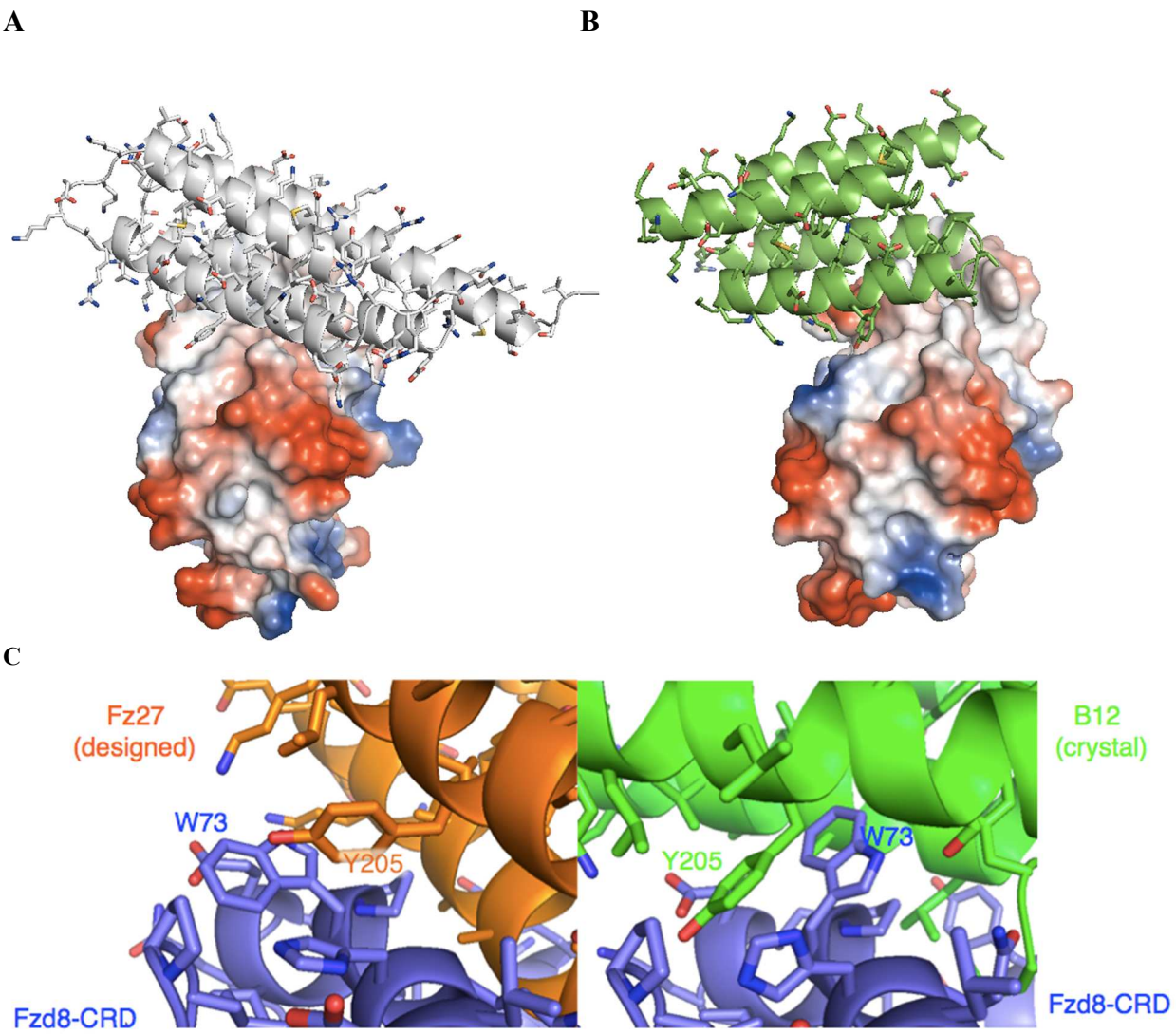


Figure 6. Design B12 binds to Frizzled-8 in Alternative Conformation.

A-B) Design B12 (green) interacts with Frizzled-8 utilizing an alternate binding mode (B) to the computational design (A) proposed for parent protein Fz27 (grey). C) The orientation of the four-helix bundle is rotated around hotspot residue Y205, which is similarly placed in the crystal structure (orange, left panel) as in the design (green, right panel).

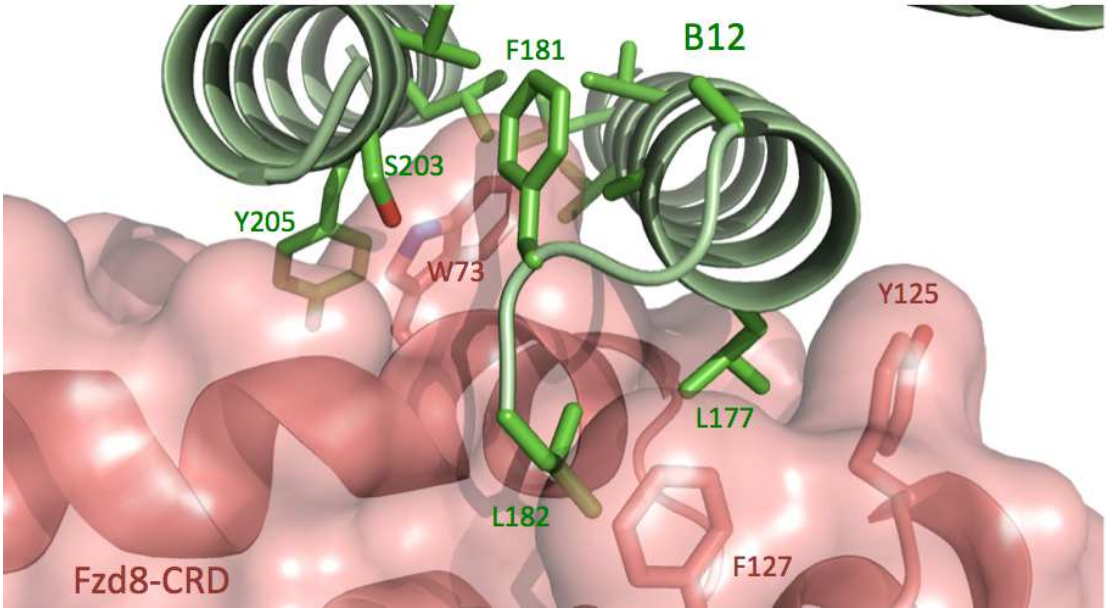


Figure 7. B12 Binds Utilizing "Reverse-Hotspot" Around Exposed Tryptophan

Novel Frizzled-8 binder B12 forms a hydrophobic pocket that packs around an exposed tryptophan residue on the Frizzled-8 surface utilizing its two interface-contacting helices. This binding mode may be considered a "reverse-hotspot" since the target hydrophobic sidechain inserts into the binder pocket.

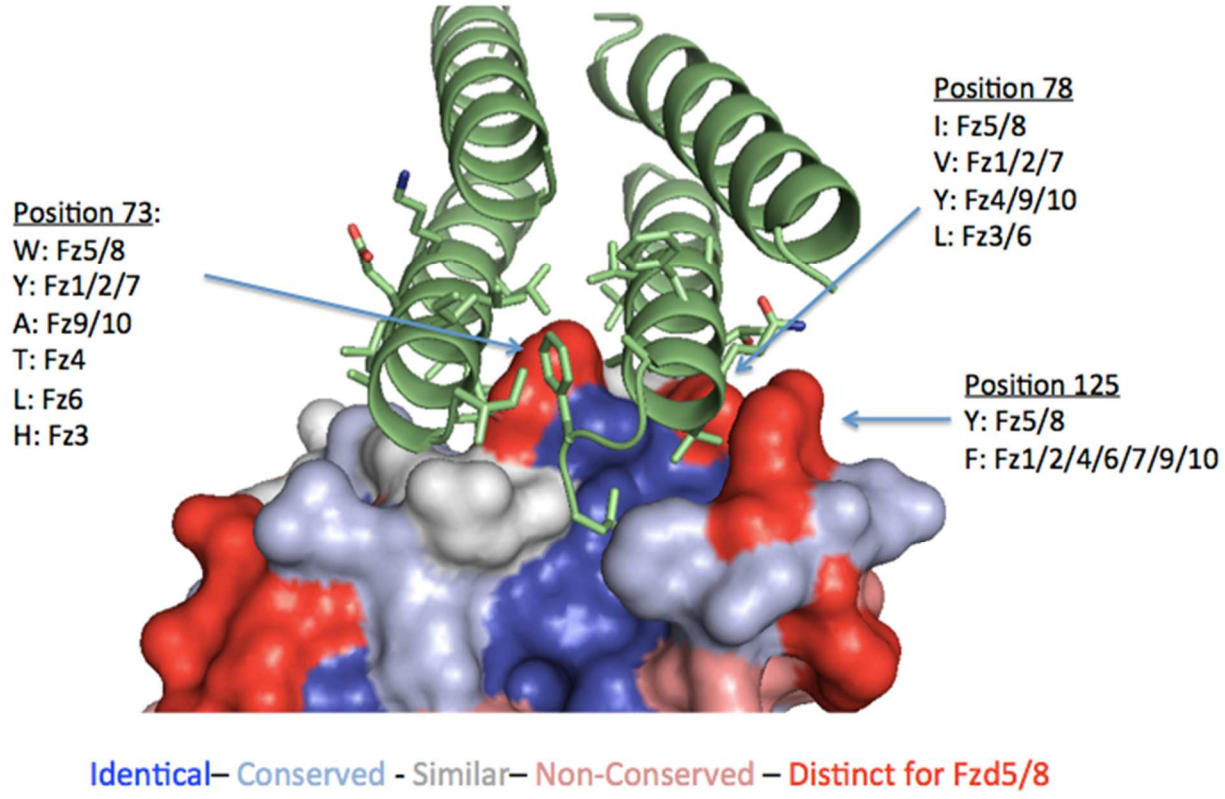


Figure 8. B12 Interacts with Frizzled Subtype Specificity Determining Residues

Binder B12 (green) interacts with specificity determining residues adjacent to the conserved lipid binding cleft of Frizzled (residues colored in gradient ranging from blue to red based on decreased sequence conservation). Of note, position 73 is only tryptophan in Frizzled subtypes 5 and 8.

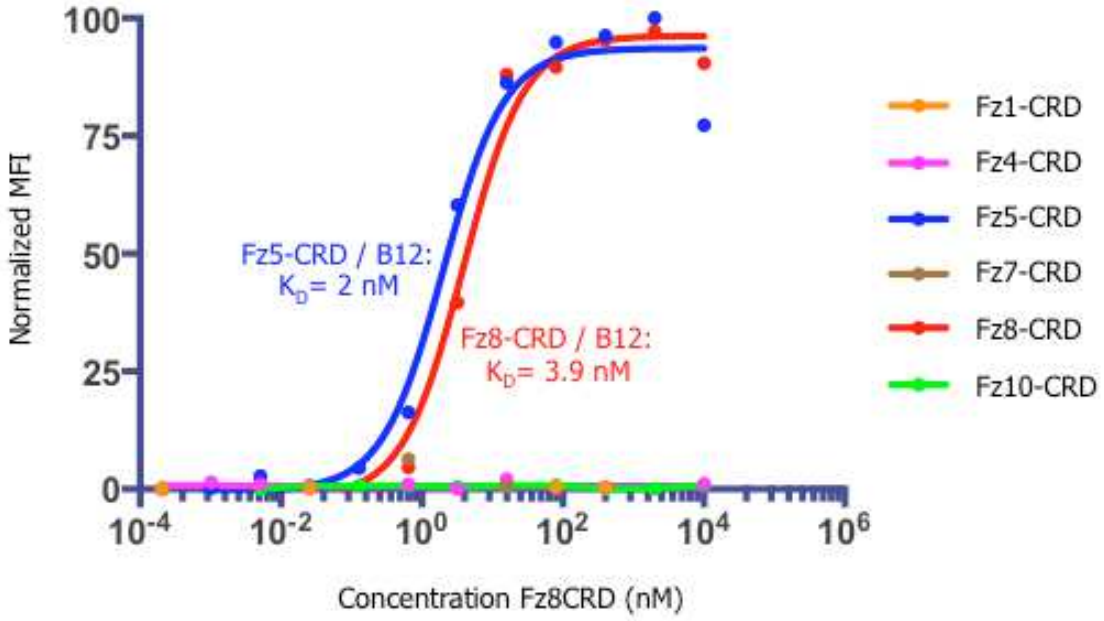


Figure 9. Design B12 binds specifically to Frizzled subtypes 5 and 8.

B12 binds to only Fz5-CRD, Fz8-CRD, and not other subtypes tested with low nanomolar affinity as measured by yeast display.

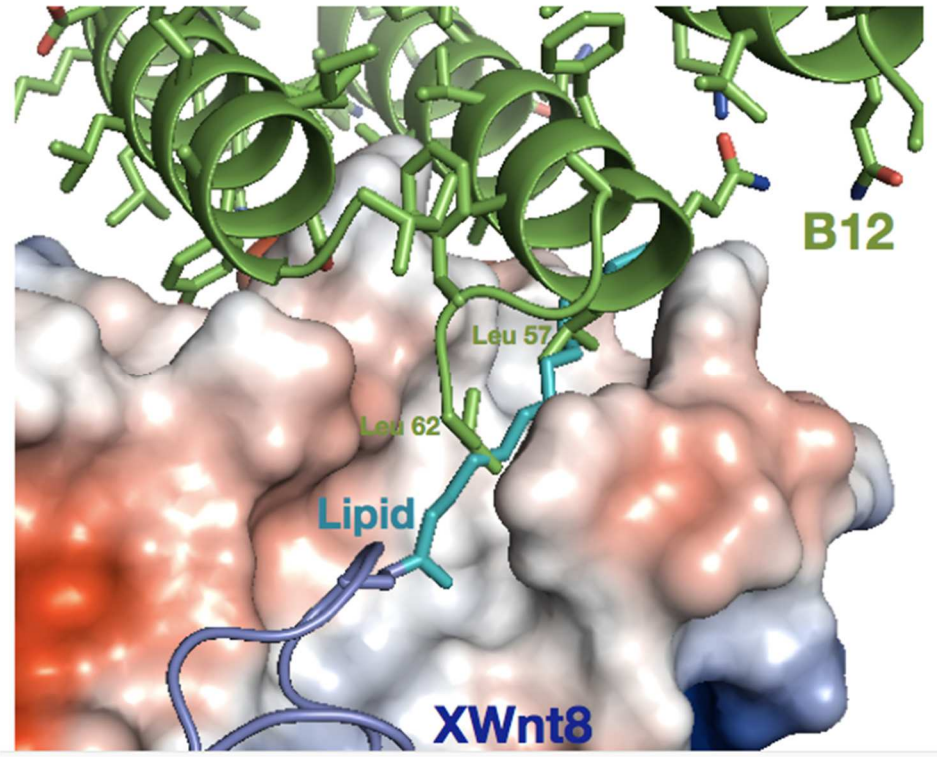


Figure 10. B12 Binding Mode Partially Overlaps with Native Wnt Lipid.

B12 inserts residues Leu 57 and Leu 62 into the Frizzled-8 hydrophobic cleft. Structural superposition of the B12-Frizzled-8-CRD structure with the XWnt8-Frizzled-8-CRD structure reveals that these residues partially overlap with the position the native lipid typically adopts.

Table 1. **Sequences of Computationally Generated Alanine Helix Frizzled Binders.**

Listed below are experimentally tested minimal helix design, both those included in the original design efforts discussed in the Wnt surrogate manuscript, as well as additional designs tested later.

Design ID	**Amino Acid Sequence**	**Scaffold PDB ID**
Fz1	GEFEFEAQQPSAVAQAKLFKALTKDSDDIIPKEIQQIEKVEIVEGNG GPGTVKKITKKGGGKEYYVLHKIDAIDEASFEYNYSIVGGDGLKEE LEKITFESKLESGPDGGSIGKIKVKFHTKGDELKRGEFEFEAQQPSA VAQAKLFKALTKDSDDIIPKEIQQIEKVEIVEGNGGPGTVKKITKKG GGKEYYVLHKIDAIDEASFEYNYSIVGGDGLKEELEKITFESKLESG PDGGSIGKIKVKFHTKGDELKREVAAYYAADGAGLFKAVEGYVK KNPNY	1IFVA
Fz2	SKSEALLDIPKLEKLLKRVGPELIKAGLAAFEYGMPGAVSKLESNL AAQDKKGIVEEGEKIKGLAGGVGLRHLQQLGQQIQSPDLPAWEDN VGEWIEEMKEEWKHDVEVLKAWVRKKD	2A0B
Fz3	QEEIRSIWENNGFGKEKREAAAAFAAAFAAFKKIGARDEEAKQLIK KAIEIAIDANKRQGAYIAQILAQWIKRGFKS	2I5UA
Fz4	RQVSVETTQGLGRRVTITIALIQAAAEAAQKAIEEAKKKRKDGLEK GKVPAQEVGQEYGKEVGEKVLGELMSKNFIDAIIKEKINPAGAPTY VPGKQEGGKDFSYSVEFEVYPEVEL	1P9YA
Fz5	GNNLVETTCKNTPDYEECKRTLKSDSQSKDGDITTLALIMVEKIQA AAATAAAIIALEKKRNPPAAWKGPLDKCAESYQKIGEKSLPEAIEA LTKGDPKFAEDGMVGSSGDAQECEEYFKGSKSPFKALNDAVHKLS DVGRAIVRNLL	1rj1a
Fz6	STAFFFKRMSPKKKEELVKELAKIYRTIVEEYENTDAKVNERIDEFV SKAFFADIKEEQVRDIAAFIAAAQAREKSAQGQEAQQGEKYQETA KEVIKHLAEMYRRS	1v2za
Fz7	WIREYPPITSDQQRQLYKKEFNKGWREFAELMRELDQAAAQAAQL QIEQQKYRQQSEEYMRAQEEAKKLIKKAGEKEAQEKEEEAKRLAE KLEHIQKMVGDYDRQKT	1wpaa
Fz8	NNLVETTCKNTPNYEECKKTLKSDSQSKDGDITTLALIMVEKIQVIA AAAAAYIALLQKRDPPAAWKGPLQKCAEAYQKIGEKSLPEAQEAL QKGDPKFAEDGMVGSSGDAQKCEEYFKSPFSELNKKVHELSDVGR AIVRNLL	2cj7a
Fz9	LKEEGEKLLLRAAALAAQVALNWKRYEENRKEEKPFDFYKDMKP YVDYAKQAADEFLKLAIPWVNTERPPDLGELKLRQAADNVQMTAI SAFNGRSFYKHFEDHAKKTAETLGKVGEFFEKRKRE	2huja
Fz10	KKAKIDGEQIRKKKDLHQTLKKELALAEYYGEDLEALKKALKGW VEYPLELEWKQAEKAGKDTKWGAAAEAAAFAWAQKEGADIRSK	2hxxa

	GD	
Fz11	REENEELAKALKEGLDNLKEMEKETGELETKYKRKIKEAGQKAEY ALKSKKQDDMARAAAQAAYIAAEIQKAKQ	2qffa
Fz12	KEWIREYPPITSDQQRQLYEEEFKKGAKEAAELAKEYDKAAFEAA KLAAELKKYREESEEYMRAKEEAKKLTEKAGEEEEKKKKDKAEE LAEKLEHIQKMVGDYDRQKD	3g7ca
Fz13	MREEEKRALEFAEEIRRQTLELLEKLNELDADEQADIAESLHDHAD ELYRSALASFKKNGQIKEQVKIAQELAKHAVELYRKARERFGGSK QEKTALEMAQQIAAAIAALAAALGKLEQG	1yo7a
Fz14	EDSPDYCLKNISLGLQGTEGRECLQSGKNLSQWERRSCKRLCTDCG LRVEEKKTEIISSCNCKFHWCCTVKCEQCKQVVIKHFCAGGGG	4f0ab
Fz15	PETLEARINRATNPLNKKEDKASIDGFCEQLKEDFEGPPLATRLLAH KIQSPQEREAEQALTVLGTCMEKAGKRFHDEVGKFRFLNELIKVVS PKYLGSRTSEKVKKAILAALAAAAAGLPKEKKIAEALRELIEKgggg	1jwfa
Fz16	EAVLNELVSVEDLLKFEKEAEEEAKKGDVEKETAFKWAEALVRTR KDDEIKKGAAALAAAGAKGEKEERRKLLFYAAVGFYRDKKYESA RVAVRLVLKTEPQNNQAKELERLIDKAMKKD	1nzna
Fz17	GNNLVETTCKNTPNYEECLKTLLSDKRSATGDITTLALIMVDAIKA KANQAAVTISKLRHSNPPAKWKGPLENCAEAYKIILEKILPKAIEAL TKGDPKEAEKGMVASAAAAAACAASFEGGKSPFSKLNEDVGKLS AVGRAIVRNLL	1rj1a
Fz18	GAKFDKGYISPYFVTNPETMEAVLEDAFILIVIKKVSNVRELLPILEQ VAQTGKPLLIIAEDVEGEALATLVVNKLRGTLSVAAVKAPGEGGSR KAALAAAAAFTGGTVIFEEAGYKLEKAKMKMLGRAKRVRITKDE TTIDGGDgggg	1srva
Fz19	GSEMEELGKFVQAAEQGLKALKEFLKELEKNPEEMQLIRAAAAAL AAAAAAAGMKGKKEMAKECMRLAEELSKAARREKKITSDLLDKI FAGVDMIKREVDKDVS	1tqga
Fz20	GSHMEYLGVFVDETKEYLQNLQDTLLELEKNPEDKRLIREAARAF QRLGGMLGKEGRRVMGALAAAAAAAASAASTDEKKITKDLLEKL KKGVEEIKKEVDKIVS	1tqga
Fz21	GQEKERKALEAIGEWDPKGRGKDAYKVEAARVLQAVEETRDGRE LAKRIQSIYEFAFDEPIPEPIAKALAAAALALKGKEE	1v84a
Fz22	TKEEKEALEMARKIREWTKELLKKLRERDADEQADIAESLHDHAD ELYRSALASFKKNGQIDEQADIAKRLSNHATKLAEEAGRKFGGRK EEWEAAAAAGAIAVAVALLLGKLQKLKKG	1yo7a
Fz23	EKCERTLEQIIKGLEDLTKQKDLCTELTVTDIFAASKQTTEKEEFCR AATVLAEFSRRHSKDTRCLGATEEQKKRHRSWYAALAALAAALG GRAGLNKCPVREANQATLENFLERLKTIMKEKYSKCSS	2d48a
Fz24	PDEPELAAAVAAAAAAAAYLLRLVEEEFKALERRELPVLQELLGK KEPLMKGLDKGGEFLAKKLREAGVSLDREGLARYARERRDGETIL RAGDQLGRLLERAQQANLRNGRIirANQASTGSLLNKLR	2fupa
Fz25	DEREKKQERASFAKKVVSDIKDKYSQEEKVQGAYMGAAGAAPAA IRKEGLGKTLKALLDSLDKPIDDVDYKSINPESFGNAANIAAAFLYK HLSTWLAEGNGKDSAFSGLTNGEDPLKYIMEKTAIDVEKSTKEALE ILRWIYLFASQELKGgggg	2oeba

Fz26	SLGTIIIWDENKGVLTAVQLLLKNHFSKVITLDKPDDLAKVLREENP EVVLLDMKFGKginngGEGLAALAAALAAYRQLPVVLFTAYADIQL AERGRKGGARDFVVKPWDNQKLLETLLNAASKE	2qr3a
Fz27	GVSFSEVMGKQKDEQAREQLKEGMKKIEEQGKKLSETRTQEELQK YAAAVAAFAAAAGFLGKNLEERrgfNRRGKEEIGKISGEVYKKLLD LKKAVRAKEKKGLDILNMVGEIKGLLERIYA	2qvpa
Fz28	VRVDEKLREEVKRLLQELGGDRRVPTGVRAAAAAAAAKFKQDQE KKDERAKTVLEALRKLANDPKVPAHGRTKLYTIISKLEALS	2qsba
Fz29	GAMSVASLPEAVKNFFPKGQEEFSKKIRADEKPVLHEVFQKKSafsq AGEMGGEVGKKHPGLGAALAAVLAGNAERLKGLKPKAVEYAKK LIRMVTTTLASLTVGKPIDDADAKRLHQEFQSLSSEDQAALRRNNP DIKFgggg	2w9ya
Fz30	GDDRRKKAAEIILELAKELKGNSGNAALIAALGAAAAALRNRGEE VPEILKKLEESIRKAEKKDGVSKNEQQSKKLKELRSIS	2zrra
Fz31	NAVDAKQLAAKATTLYYLHKQAMTEEVKKLLEEAWKKEPSNEA AGELIAKDAFIKFQFGGAAAAWAALLASNKPELDRKKIQEEIEEAK KKI	3beea
Fz32	GSEEAIKKAGALAAAAAAMAAAVPLKRDDEPGKKEMEKIAFSAEL TRAAERDGQLAQLQGLQDKWRNELIPALEKAKEGKKVTKDVLEF AEGLKKLVEGFDRTT	3ezla
Fz33	SNAVNEIVVNPQATLDWKRALKTADGKTDEAREQLKKLLEALPEV RKKVREQLVGENPEGLAAALAALAAAAGRSGVPRMAELAYRLGG QLGSGKKQEDLEPELLELLDEMDNVAREASKILG	3iqta
Fz34	PIENRIAGYFAEWFMRERSTKIDKQEAAKYAATAAALATKMPRKLI SEITASSYQRALNEIAKEHEKEATEGFHKNVRESIQALIEEGRLQKD FTTRAVVKGRRD	3lysa
Fz35	AVYVVGGSGGWTFNRYLWPLGYDFREGDILLFNYDPSMHNVVVV TFGGAAYCNTPAGAKVYTSGRDQIKLPAGFAAFICNFPGHCQSGM RTAVYAKgggg	2cbpa
Fz36	ggggPAFRVNTNVPRRSVPKGFWLELAQQLAQATGKPPQQAWAYV EPDQLMAFGGSSEPAALAKLYSIGKIGGAQNRDYSKLLAGLLAERL RISPDRVYINYYDMggggsggggNAANVGWNDQTFA	3ijja
Fz37	VLGWRESFWKLTWSVNGVAAFHAFLKTEFSEENLEFFLAAVLFSFI AQATKLASRAHHIFDEYIRSEAPKEVNIDHETRERTKTALQQAKAS AVLQAAAITFTLMEKDSYPRFLKSPAYRDLA	3c7la
Fz38	KFEVWKQKYGVVIAVALLGleyengeeAFFWFAFTRPKVSDISRFTKE LNSKPDMAMKNLTFSAIMPILEELLRQAAEEFPGLTFNTASRLMEIV GASAATSLK	3fgxa
Fz39	VEIWDWGGGLVQAGGSLRLMCVRSGAVGSTYDMGWFRQAPGKE RESVAAINWGSAGTYYASSVRGRFTISRDNSLFAVYLQMNSLKPED TARYTCGAGRIgrsvfnlrRESWVTWWGYGVDVTVSS	1shma
Fz40	PHFAHIWPGGGPIAWLAWPREVLDLVRKVAEENGRSVNSEIYQRV MESFKKEGRIGgtggsGGGREVLDLVIKVAFENGKTVWAEIFRRVME SFKKEGRI	1u9pa
Fz41	HFPGRAPIDAYGNGGFRFADMSHRGSIIAIPSGIYGIDMTGPVPTQE DISRVLEESDQIQTLLIGTGVERLDIPLELLILLKEKRIWAWAASTGQ	2fi9a

	AVRAFNWLLAQNHAVAALLFAVE	
Fz42	HFPGRAPIDAYGNGGFRFADMSHRGSIIAIPSGIYGIDMTGPVPTQE DISRVLEESDQIEILAVGTGVERRDVPLYLAELAWRAKIWIFASSTG DAVRAFNRLLAEDQAVAALLFAVE	2fi9a
Fz43	ggggGPTDASVEEEGVRRALDFAVGEYNKASNDMYHSRACQVVEA RKKISAGVYYYLRVELCRTTCTKTQpnLDNCPFHDQPHLKRKKECF FIIYAVPWQGTMTLSRAWCWEA	3gaxa
Fz44	GAPVPVDENDWGLHMALWRAMAAYNRASNDKYSSRVVRVISAK RQLVSGYKYILQVEIGRTTCPKSAGAAMECLQHDEPEMAKYTTCT FVVYVIPLLFQAKLLESKCQ	1cewa
Fz45	LQLFIITFTGKTFTVETEPSDYIENLKAKIYDKEGGPPDRADLYYEG RQLQDGNKLSDYNIQKEAYIALFLgggg	1sifa
Fz46	TQLTHVQECVRGTTVILKRPCPSGTYHGAQPFHPLADNKFALTCEQ HWMAFICADGTKFFYAIVAKSN	1xaka
Fz47	DKKTLKYFAVAWAVKPTKSEEEMKKKAEQDKEFIKQKGGEILYEE DWGMRKLTRPIQKYNNARLFLVIFRTENPRLPQALDFYLKIDEDVI WWWLVEMHERLARKS	2j5aa
Fz48	GRDFRICMTIVQLLKKLAKRPERADVWAAAWLVCKTGRSRWRDV CQEFMQKYAERVAQGLEAGKTAQQICEDLR	1L9L
Fz49	PKDLAALIDKATEEVHTQLENAEFDRNWKKGQITRDGFKLVAASL YHIYVALEEEIERNKESPVFAPIYFPEKTADLANLEQDLAFWYGPR WQEVIPYTPAMQRYVKRLHEVGRTEPELLAAHAAARVKGDLTGG FWKAWLAQKALDLPSSGEGLFFTIFPNIERMFKFSELYKSAINSLEM TPAVRQRVVEEAKTAFLLNIQLAEELQELLTH	1N45
Fz50	AAPTATVNPASGLSDGQNVMVQGAGLQRGQQYQVGQCAWVDTG VLACNPADFKQVQADQEGKAWTLLNVRRSFEGFLFDGTRWGTVD CTTAACQVGMWNAAGNGPEGVAISFN	1NOA
Fz51	MNSDEVQLIKKTWEKPVATPRDAGAAILLDDANRYPSTLEKAPYR DVPLEELSGQNTFQLLAGAIIQIFNQAIQVLGQDGALEKLDQIWTNV AEARIPFDITKEIINWLKGVILDVLTAASSLDESQAATWAKLVDHA VAIIFKALDDDGKAL	2BK9
Fz52	MMDYLITQNGGMEFAVEAMAVAQIEAGIGAQKGAGISGELARALT TSQPEKEGKASEWYFKPGSIGFFGALLAQQIQQNLGSDMSVVQGL NALGAADPMAQIGLQAGIAMGKATAAGWQIDAKKPEHAKKGKR QSDMVKTQADDGWVIAKLLILNA	2BL2
Fz53	AASAASSEHFEKLHEIARGLIEDLQGIPERLLGTAGTEEKKKIEQDA AQKAFQAWATVKQMEQQLRYAPLTFRNPMMSKLRNIERDLWRIW REIKgggg	2QYW
Fz54	TAEVASYVTAHFGKTLEECREESGLSVDILDEFKHFWSDDFDVVHR ELGCAIICMANKFSLLDKRSRAHQENLFKYIQSFPNGQVAAAQIIW MILYCIQQFDTETDDCTRVVKVAKCFKEEARKQGMAPEVAMVEA VIEKY	2WCM
Fz55	TISSLMVPRSDIVFLILNLPLNADLKTAWEAPLQVFPVARNNVDDM VGTIKAQELLKKQFKGERLELVDAVKNSNFVPNSLSGMELLEHFRT TGAQMVYVVDEYGDLKGMVTLQDMKKALTGEFF	3LHH
Fz56	TYSPEKIEELKFAIQPIAKARDGMEKRLQGLIADQNWVDAQTFIKG	3LS0

	PLGFLFWLMEGLASKLLPKYRDKAKTLAKEVFGWLQRLLAAAQD RNGAQAKIAFQEALAYFDSFLNLLP	
Fz57	TFKVYGYLSNIHNCGPCFQAFAQLQQKQQPFEFINIMPEKGVFDDE KIAELLTKLGRDTQIGLTMPQVFAPDGSHIGGFDQLQEYFK	1ABA
Fz58	RKEAVIIMNVAAHHGSELQGFWLLQAIQRAGFIFGDMNIYHKHLSP DGSGPALFSLANMVKPGTFDPRRYQFTTPGVTIFMQVPSYGDELQL FKLMLQAAQNIANAVGGVVLDDQRRMMTPQKLREYQDIIREVKD ANA	1F46
Fz59	QAIPVTWEGFSRLAEEARFLKEVRIPEIIAAIAEAREHGDLKENAEY HAAREQQGFAEGRIKDIYAKLSNAQVVNVRQMPNNGRVIFGATVT VLNLDSDEEQTYRIVGDDEADFKQNLISVNSPIARGLIGKEEDDVV VIVEFEVIKVELL	1GRJ
Fz60	ADKELKFLVVDDFSTMRRIVRNLLKELGFNNVEEAEDGVDALNKL QAGGYGFVIADWRMPNMFGWFLLQQIRNAISALPVLIVTAQAQKA QIQLAAQAGASGYVVKPFTAATLEEKLNKIFEKLGM	1JBE
Fz61	MLNKELQQALKFAMFQAQYHRNEFMTVEHLLLALLFAPSAREALE AASVDLVALRQELEAFIEQTTPVVPASKEEKNAQPTLSFQRVLQRA VFHVQSSGRNEVTGANVLVAIFSEQESQAAYLLRKHEVSRLDVVN FISHGT	1K6K
Fz62	AMVPNVVVTGLTLVASSAPGPLELDLTGDLESFKKQSLVLKEGVE YRIKISFRVNREIVSGMLYIAQIYQKGKNFWIQKYMVGSYGPRAAA YEFLTPVKQAPEGMQARGSYQIKSQFTDDDKTDWLSWEWKLTIKK DW	1KMT
Fz63	AANKLIKEMVQEDQKRMEKISKRVNAIEEVNNLVKLLTELVMSHS QGGAAAGSSQDLMRELAQRAANMIPTLNRLASDTEDNDEALAEIL QAQNNAFQVFMLWYQANQGQENgggg	1O3X
Fz64	INVKEVRLSPTIEQHDLNTKLRNARKFLEKGDKVKATIRFKGRAINE KKQGLFVLFMLMLAAKDIAQKETQPKMDGRKMFLILAPKND	1TIG
Fz65	GSHMEYLGVFVREARKYIAKLAQTAAKLQKNPEDMELINEAFRAL HTLKGMAGTMGFSSMAKLAHTLENILDKARNSEIKITQDLAFKIFA GAKQIARLVDKIVS	1TQG
Fz66	VIRVYIASSSGSEQIKAQERAVLGWLESRKIGFEEKDIAANEENRKW MRENVPENSRPATGYPLPPQIFNEAQYRGDYQAFFQAFMNEAVYA FLGLTAPPGNKEYEVQAKQQALEHHHHHH	1U6T
Fz67	QPLNQIITNSARQFNIELIRVQPQGEMVQVWIQPLPFSQLQSWLAYL RERQGVRVQFMWVQRGKVNGVVEVAMLSLQRGG	1UV7
Fz68	ggggKTKQENTAAQMATNILVQTWNLLKKLNELDADEQADIAESLH KHAFDLWRSAQEAFQKNGQIDEVADIAESAHDHADELYRSALARF GGSKQEKTALNMARFIRSQTLTLLEKLNELAKG	1YO7
Fz69	EIIWESLSVDVGSQGNPGIVRYRGVDTKTGEVLLEAEPQPIGTNNM GEFLAIVNGLQQLKNRNSRKPIYSNSQTAIKWVKDKKAKSTLVRNE ETAQIWLWANFAEFWLNTHTYEVPILKWQTDKWGEIKADF	1ZBF
Fz70	MDGRIKQVSVFTYHKKYNPDKHYIYVVRVLREGQIEPSFIFRTFDEF QELHNKLSIIFPLWKLPGFPKQMVLGRTNINKQALFLQFALIEYLKK LMEASTDVAEADLVATFFHGSHH	2AR5
Fz71	MLKRMYARVYGLVQGVGFTKFVQIHAIRLGIKGYAKNLPDGSVEI	2BJE

	VAEGYEEALKKLLERIMKGPPAAQVFWVQFRFSEYKGEFEDFETY	
Fz72	KNAKQLVQELYKDIAKSKDPKYFQIFAQMANVFQKLKKQKYELDP SPLINRLVNYLYFTAYTNKIRFTEYQEELIRNLA	2BL7
Fz73	NNLVETTCKNTPNYQLCLKTLLSDKRSATGDITTLALIMIDAIQQKA FIAWAQAAQLANSNPPAAWQGPLRQCAQAYAKIMTQALPEAIEAL TKGDPKFAEDGMVGASGDAQECEEYFKSPFSALNIAVHELADVGR AIVRNLL	2CJ7
Fz74	HFPGRAPIDAYGNGGFRFADMSHRGSIIAIPSGIYGIDMTGPVPQAK NIAKVMNQNKQIEVLLIGTGVELLRLPQQLQFWLQQARISSDTMST GAAVRTFNVLLAEDRKVAALLFAVE	2FI9
Fz75	MNTYSVTLPWPPSNNRYYRHNRGRTHVSAEGQAYRDNVARIIKNA MLDIGLAMPVAVKIEAHMPDRRKEFLQKLAKAAFDALTKAGFWL DDAQAQFWWAQRMPVTKGGRLELTVTEMGNA	2H8E
Fz76	RMSEQSIAQAQAFVVVYDDANKQWKPAGGSQGAAKVHIYHHTG NNAFRVVGRNWKKEQVVINAAIPKGLKYNQATQQFAQWRDARQ VYGLWFGKQEDANEFASAMMHALEVLNS	2IYB
Fz77	YRNILVLISVNNDNLENYFRKIFLDVRSSGSKQATIAVFTEIKKQELF NLIFKEFAENKDIGFEIFLYKKNEVDIFLKNLEKSQVDGLLVYCDDE NKVFMSKIVDNLPTAIKRNLIKDFCRKLS	2RBG
Fz78	ERDAFDTLFDHAPDKLNVVKKTLITFVNKHLNKLNLEVTELETQFA DGVLLVLLMGLLEGYFVPLHAFALTPNSFKNKMFNVWAAFLLMK QGGLQYPKPRPKDIVNADLKSILRVLWNLFTKYRNVE	2VZC
Fz79	SGMQLEIQVALNFIISYLYNKLPRRRVNIFGEELEKALKFWMQGLW YPEKPYRGAGANAIKIGEKVDPTIEFAAKQSGLDIDDVRGNLPQDL SVWIDPFEVKYQIGEKGPVKVLYVDDN	2Z15
Fz80	PPATPQEIQFVQRMIQHAAQAAQLAAPMAKRSQQRTVRALAFRIF AKQAEQAQQMKAMLGRWGQPPGEPISPEHARMMGMASEAEVAG LSTLPVEQAERLFLRLMIRHMQGAVAMTLPMLDARPEVERLARQI VVTARGEIRTMEGVLGRL	3BT5
Fz81	ggggKSDEQQAKFQIWAKAQYLASLLDAGNLNNQANEKIIKDAGGA LDVSASVVDTDGKVLYGSNGRSADSQKVQALVQGQVGRLSTDNK LYYGLSLRSEGKKTGYILLSAS	3CWF
Fz82	MYQTLEGFQKQWKKESAWIAQMLSRLTDESLSQEIAPGHWTLGR VAWHLVTAIPVILSGTGLKFQGETKDYPVPTKAKQIAFGYWQVNQ AFNQALQSEWTDKDLTTINDFFGRPMPNSIFLMTLINHANHHRGQL TVLMRQAGLTVPGVY	3DI5
Fz83	DISAKDLRNIMYDHLPGFGTAFHQLVQVIAKLGKDSNSLDIIHAQF QNSLANGNRPEWALRWITFNVPIFQDAAPPVIHIRSRGDIPRAAQKY LEPVPPNPKINHGWVAVFQLQDGKTLGLQA	3FKE
Fz84	VVVLEAKNGNVAFDHKKHAGVEGEARAAHKEEAGGKEKGMGKE AVKFWIVGFNKKMGKGPTKAGEAHK	3H4N
Fz85	MITIDALGQVAPNPMFQAFEALKNLGKAGGVVTVLVDNDISRQQL QKMARGLGFQAEYLEKDNGVIEVTIVAGE	3HZ7
Fz86	QNPQLTLSLIAKNTPANSMIMTKLPSVRVITQGYNPAMNVNELFAY VDLSGAEPGQKIYQVQVEPIPNIKVFWTSPTAVRLQLEHH	3LYW
Fz87	YDIQAWKKQAEELLNLIFQAETAEPFRQPVDLLEYPDYRDIAQTPM	3MB3

	DFATVRETLEAGNYQSPINLMFQVWAIFQNAWNYAPSKRSRIWSM ARRLAQFFQEHIKSVLSDYKSALRF	
Fz88	LTQERKREIIKQEQQQGQQTGDPEYRIAALTAAANAAKGHLKDQK EKKEDREGLKKGEGEREKLKEELRNKDVARYREIKKKLGD	1A32
Fz89	PREQEIAQKAGAVAAAAAARGFYQEEARRQEEEIKKMRAEDGEN YDIKKQAEKLQESRMKIPQFEKQLKKFYEELKRILENEKDLEKAKQ YEGARGGLEYVKR	1H7C
Fz90	ggggKIEEYPPIKSEKEREEVEKEARRGAEEYKKLQEKLKEIDEELSR LDKELDDYREESEEYMAAADEYNRLKQVKGSADYKSKEEKAKEL AEKAAHIAALAAEYGEKRK	1WPA
Fz91	ggggTEGQEAEREQVERDVRKLEEMLLHYSKEDSSDGDKEEMGRL AAQLAAAAAALTKKAEKTEDNDNSLGDIKRAIDNALKVFQSYGQR KGGQgggg	1WR6
Fz92	IKEEQQLKEAEELVRELEEKHTKEASGKAAQWAGNVKKEYAAKA AELAARAAEVGKKGRKD	2GOM
Fz93	EHIKQQALDLFTRLQFLLQKHDTIEPYKQVLQFLEAGIQSVKRDLG RPEDAAAAVAAAIAAYAERDKLHKTAEEEEVLEKIQKLAQD	2IP6
Fz94	HYKTERYYETVFAVKPETQEQQMKAIFAAAAAFIGEKGGKYEYQE DWGMRQLAYPIQKFNNARYFLIQFKTKNPQLPNELDEQLKKTQYV IEWLNIQIKESEVKKN	2J5A
Fz95	KEENEKLAKALKEGLDNLKEAEKQTGETETEYKRKIEEWGQKAEY ALKSKKKEDMEKAAARAAYIAAFIARKKK	2QFF
Fz96	GVSFSEVMGKKRDEEAKKKLGELMKKINEQGKKLKEKRTIEELRK YKELVKEFVGDAVELGLRLEERSRKGREKIEQIVKAVDQGLKNLEE AVKKKEREGRQIAAAVGAIAAALGRIYDgggg	2QUP
Fz97	KPEDAFAAAAAAAAAQSGADGTLVAVPADAAEELVIKFVMGKVP QEVQASAIPVQDNAIGQAFRDRAPRRLDVLDGPGDGGPALVLPLR ATDTVAGVLVAVQGRGKRPFTAEQLKQMTGSADAAAKAWQEAE DKRRMS	2VJW
Fz98	TEDGETVKVFEDRQGFETFIANEAEDDDFDHLHAKLNYYPEWVLK GSGDDPKKIKKEVNSAKEAFAAAVAAHIQGDLLRDIKRAVRKPEL EFKEQSYKKEFDKIEWRAGEETEYHGRPFKIDVQVVATHEEAKVF VDYKTHPVGAN	3BCY
Fz99	PPARPEEVQFVQHMLQHAKQAEDLAAPMLERSQQRTVRELAKKIQ EKEKGGMEGMKKMLGRWGQPPGEPISPEHARMMGMASEAEVAG LSTLPVEQAERQFLRLMIRHHEGAIKMYLPMIYTRKEVAALVAAA AAAQRGEIEEMKGVLGRL	3BT5
Fz100	SEWIREYPPITSDQQRQLYEEEFKKGAAEAAALAANLAQIGANLAA LDEEKRKYRKESEEYMAAKDEYERLKKKKGDEEAKEKQKKIQEL ASKLSHIKKMVGDYDRQKT	3G7C
Fz101	ggggNDPEQTYKQAEEHEKKDQRQQAAAAFAALMAQAPDYVGAY YKAGKWGEERNGKEGAIGIYKAGIKAARDKGTQKDLSELKDAKL KAEGKD	3MA5

Table 2. Positions Included in Fz27 Combination Library

Position	Wild-Type Identity	Enriched Identities Included in Library	Mutation Type
25	K	I	Electrostatics
47	A	V	Interface Packing
52	A	L/T/I/V/P	Interface Packing
55	A	V/L/S	Interface Packing
56	A	Q/E/P	Interface Packing
62	K	P/Q/T	Electrostatics
87	K	L/I/Q	Backs up hotspot interaction
116	L	T/I/P	Core/changes spacing between helices
119	R	S	Terminus adjacent to disordered loop
120	I	G/V/S	Terminus adjacent to disordered loop
121	Y	P/S/H	Terminus adjacent to disordered loop

Mutations identified by the Fz27 site saturation mutagenesis library included both residues which were located in the designed interface and presumably optimize interface packing and electrostatics as well as remote positions which are adjacent to unstable secondary structure and may alter the protein's stability or backbone.

Table 3. Fz27 Combination Library Results:

Position	25	47	52	55	56	62	87	116	119	120	121
Wild-Type AA	K	A	A	A	A	K	K	L	R	I	Y
Enriched AA	I	V	L/T	L	Q	P	L	T/I	S	G/V	P
Selected AA	**I**	**V**	**T**	**L**	**Q**	**P**	**L**	**T**	R	**V**	Y

Wild-type, enriched, and selected amino acid identities are listed for each position included in the combination library for comparison of Fz27 to the optimized variant B12. Selected amino acids that correspond to enriched amino acids are in bold.

Section 3: High Affinity Wnt Antagonism By Computational Loop Redesign

3.1 Abstract

B12 is a designed protein that specifically binds the Wnt receptors Frizzled 5 and 8. Based on the structure of the B12-Frizzled-8 complex, we designed an expanded, disulfide-stabilized loop that further extends from B12 into the fatty acid binding groove of Frizzled to optimize Wnt antagonism. X-ray crystallography confirmed the key features of the design and the protein potently inhibited Wnt signaling in A549 cells. In addition, treatment with the optimized binder inhibits Wnt signaling, reduces cell proliferation, downregulates target genes, and increases expression of a differentiation marker in PDAC cell lines. This effect was dependent on the presence of loss of function mutations in the ubiquitin E3 ligase RNF43, demonstrating the functional utility of subtype-specific Wnt antagonism for precise targeting of molecular dysfunction. Covalent functionalization of protein nanoparticles with B12 resulted in generation of a multivalent, high affnity Wnt antagonist.

3.2 Background and Approach

Wnt signaling is essential to both embryonic development[1] and mature tissue maintenance[2], while dysregulation of the Wnt signaling pathway is conclusively implicated in a range of cancer types[4] (colon[92], leukemia[93], lung[46], pancreatic[94], breast[95]). Wnt signaling is also thought to regulate the proliferation of cancer stem cells and transcription of drug resistance genes[96], making it a critical target for potential therapeutic intervention.[97] Wnt interacts with its extracellular receptor Frizzed (Fz) via an essential palmitoleic acid modification[5] and blockade

of Fz at its lipid binding site results in inhibition of canonical Wnt signaling and decreased tumor growth[6]. Therefore, molecules that can potently antagonize the Wnt-Fz interaction have significant promise as anti-cancer therapeutics[98,99].

Previous design and optimization efforts resulted in the development of B12, a novel four-helix bundle protein engineered to bind the functionally relevant lipid-interacting Fz cleft with high affinity. B12 forms a hydrophobic pocket around TRP73, which acts as an essential hot-spot residue but is present only in Fz subtypes 5 and 8. As expected, B12 binds Fz-5 and Fz-8, but not other subtypes, with respective affinities of 9 and 18 nM. Indeed, as part of a bi-specific Wnt surrogate, B12 specifically activates only cell types that express these receptor subtypes. Therefore, the precise binding profile of B12 enables the functional modulation of Wnt signaling in a subtype-specific manner[72].

Structural analysis of the B12-Fz-8 interface revealed that the binder binds immediately adjacent to and inserts several residues into the distal end of the Fz lipid-binding cleft. However, the B12 binding mode does not completely occlude the access of the Wnt ligand to this functional site, explaining its weak activity as an antagonist in comparison to its binding affinity. Additional downstream functional applications of the highly specific B12 are limited by its insufficient activity as a Wnt antagonist, therefore redesign of the B12 interface-contacting region was undertaken to increase its functional Wnt antagonism.

3.3 Computational Redesign and Optimization of B12 Interface Loop For Wnt Antagonism

Fourteen amino acids comprising a helix-loop-helix topology proximal to the lipid-binding cleft cannot be resolved in the B12-Fz-8 crystal structure (Figure 10A). It is presumed that this region of the protein is disordered and does not contribute significantly to the energetics

of binding. Therefore, replacement of this region with an expanded, stabilized, de novo structural element should result in the simultaneous retention of binding activity and enhanced robust blockade of the Wnt lipid moiety. To this end, four de novo loops ranging in length from 16-21 amino acids were generated with RosettaRemodel utilizing a blueprint-guided approach to backbone remodeling[100]. These structural elements were designed to contain ordered helical structure stabilized by a disulfide bond to provide a rigidified conformation for optimal Wnt antagonism (Figure 10 B-E).

Experimental testing of the de novo loop variants via yeast surface display revealed that radical redesign of this region did not substantially affect binding for any of the four loop variants tested (Figure 10F). In order to further optimize the interface for Fz binding and Wnt antagonism, the redesigned variants were affinity matured. An error prone PCR library diversifying only the loop region of these variants was generated and selected to convergence. Additional site-saturation mutagenesis (SSM) of the entire interface region, selection, and deep-sequencing of naïve and selected pools identified enriched point mutants with enhanced binding activity (Figure 11A). Enriched amino acid identities at the most selected positions were combined in a subsequent library and selected to yield final, optimized variants (Figure 11B). Redesigned, optimized variants B12_L1 and B12_L2 bound Fz-8 with K_Ds of 235 and 347 pM (B12 K_D=17.6 nM) and Fz-5 with K_Ds of 98 and 122 pM (B12 K_D = 9.4 nM) as measured by surface plasmon resonance (Figure 12). Therefore, while the primary focus of this work was to engineer ordered backbone structure in the lipid-binding cleft for Wnt antagonism, it resulted in significant affinity enhancement as well.

Subsequent structural characterization of the optimized variant B12-L2 to 3.2 Å confirms that it interacts with Fz-8 utilizing the same binding mode as the original design (Figure 13). The

backbone of the de novo loop is visible in the crystal structure through the stabilizing disulfide bond, which is formed as designed. The successful design of a disulfide-stabilized, ordered helical secondary structure element here results in both the insertion of additional residues into the essential Frizzled lipid-binding cleft and the positioning of additional downstream residues for Wnt antagonism. Although a significant portion of the remainder of the designed, expanded structure is still not ordered in the obtained structure, it is oriented in the appropriate location to enable Wnt blockade.

3.4 Reengineering of B12 for Stabilization

Initial redesign efforts primarily focused on replacement of the unstructured interface-contacting region for improvement of B12's functional antagonism. However, use of B12 in downstream applications would also benefit from optimization of the design's biophysical properties. The native *Bacillus halodurans* protein (Bh1478, PDB ID 2QUP) used in the design of B12 was not an idealized scaffold, and has several pre-existing non-ideal features that could be improved. We therefore pursued several parallel approaches to additionally idealize and stabilize the B12 four-helix bundle.

Although B12 is comprised of a relatively canonical four-helix bundle, the N-terminal helix contains an extended region distant from the designed interface and core of the protein that is not substantially backed up by other secondary structure elements (Figure 14A). We therefore postulated that this portion of the protein would be dispensable for both folding and binding. In order to minimize B12 into a more ideal four-helix bundle, N-terminal truncations of 9, 11, 13, 15, 17, and 19 amino acids were tested. None of these truncations appeared to affect the design's

binding function as assayed via yeast surface display, confirming that this region of the protein was nonessential.

In the B12-Fz-8 crystal structure, B12 adopts an alternative conformation in which the top two helices swing away from the bottom two interface-contacting helices to form an extended 2-helix structure. This conformation depends on domain-swapping facilitated by crystal contacts, and may be an artifact not seen in solution outside the context of crystallization. However, the formation of an alternative, ideal helix by the residues (114-120) that comprise the connecting loop of the projected helical turn indicates that the existing amino acid sequence was amenable to adopting either backbone conformation (Figure 14C). Therefore, RosettaScripts-directed, consensus-based sequence optimization of these residues was undertaken to favor formation of the desired backbone and disfavor the alternative extended helical structure (Figure 14D). In addition, the hydroxyl group of TYR44 (inherited from the original scaffold) points directly into the hydrophobic core of the protein (Figure 14B), therefore this residue was mutated to phenylalanine to remove this destabilizing interaction. Combination of these improvements (N-terminal helical truncation, loop sequence optimization, and tyrosine to phenylalanine mutation) yielded significant improvement in protein stability as measured by circular dichroism (Figure 14E).

3.5 Redesigned B12 Enables Targeted, Subtype Specific Wnt Antagonism

Based upon backbone position, the computationally generated de novo loop of the optimized B12 should significantly enhance its activity as a Wnt antagonist. Indeed, treatment of A549 cells with the redesigned variants results in potent inhibition of Wnt signaling with an IC_{50} of 464 pM (B12L1) or 806 pM (B12L2), a marked improvement over the B12 parent binder (IC_{50} of 65 nM) (Figure 15). Since the redesigned variant retains the original Fz-5/8 specific

binding mode and now robustly antagonizes the Wnt-Fz interaction, B12 now provides a reagent for precise inhibition of Wnt signaling in contexts where Fz-5/8 subtypes are implicated. This will enable the targeting of relevant receptor subtypes without off-target binding to orthogonal tissues types, allowing effective discrimination of the contribution of these Frizzled receptor subtypes and precise therapeutic intervention.

A previous genome-wide CRISPR screen of RNF43 mutant pancreatic ductal adenocarcinoma by Steinhart et al[60] identified Fz-5 as essential to this Wnt-dependent neoplasm, and Fz-5 inhibition resulted in decreased cell proliferation and reduced xenograft growth. Notably, these effects were not seen for PDAC cell lines lacking mutations to the RNF43 ubiquitin E3 ligase previously confirmed to be responsible for regulation of Fz receptor levels[57]. Indeed, treatment of the HPAFII and ASPC1 RNF43 mutant PDAC cell lines with B12 resulted in potent dose-dependent inhibition of Wnt signaling (Figure 16A). In contrast, signal inhibition was not observed in BxPC3 or PANC1 PDAC cell lines without RNF43 loss-of-function mutations (Figure 16B). Therefore, the optimized binder represents a highly specific antagonist that can be used to specifically intervene to inhibit signaling in Fz-5 dependent RNF43 mutant PDAC cancers. Treatment of HPAFII cells also resulted in dose-dependent downregulation of Wnt/β-catenin target genes AXIN2 and NKD (Figure 16C), reduced cell proliferation (Figure 16D) and increased expression of the MUC5AC gene previously associated with PDAC differentiation (Figure 16E)[101]. The inhibition of Wnt signaling in this context by the optimized B12 binder represents proof of concept for the precise, molecular-level targeting of carcinogenic dysfunction for effective intervention in Wnt-dependent neoplasms. Moreover, since the Fz-5 sensitivity of these cell lines is dependent upon RNF43 dysfunction, identification of such

mutations may provide a method for identifying neoplasms that may be precisely targeted with Fz-5 specific antagonists.

Previous computational design of two-component, 120-subunit, self-assembling protein nanocages provided a platform for the controlled engineering of functionalized nanoparticles[35]. We therefore covalently attached the B12-derived Fz binders to the nanoparticles using the spontaneous isopeptide bond formed between the SpyTag-SpyCatcher proteins[102]. The resulting nanoparticles potently inhibited Wnt signaling in A549 cells as expected (Figure 17), demonstrating the utility of these binders as functional modules suitable for use as targeting domains.

3.6 Discussion

Complete redesign of a fourteen amino acid loop near the critical interacting residues of the B12 binding interface was a significant challenge, made more difficult by the non-canonical structure of the preceding residues. Backbones built from these end points may lack the stability typically associated with idealized de novo designs, leading to disruption of the original binding mode. On the other hand, since the proposed interface topologies are spatially removed from the core of the four-helix bundle and the existing structure here was already not well ordered, we postulated that complete replacement of this sequence would be tolerated. Indeed, the radical, de novo redesign of this connecting topology neither affected the folding of the four-helix bundle nor the protein's function as a high affinity Fz-8 binder.

Robust blockade of the Fz lipid-binding cleft was required for functional inhibition of Wnt signaling by B12. Because its core binding interactions were adjacent to but did not overlap the functional binding site, extension of the B12 backbone at the interface contacting region to further compete with the native Wnt lipid binding epitope was required. Stabilization of extended

topologies away from the core of the four-helix bundle was achieved through the design of an engineered disulfide bond that successfully rigidified additional ordered helical structure within the targeted lipid-binding cleft. Although additional distal residues of the expanded topology remain undefined in the crystal structure, their orientation still serves to provide additional steric blockade of this region due to their position originating from the disulfide-stabilized helical element buried deep within the lipid-binding site.

This work demonstrates the utility of computational methods for the de novo design of specified loop topologies targeting functional sites, and that construction of such extended backbones from existing endpoints can be accomplished without disruption of the existing protein core or functionality. Additional successful rational refinement of B12 included the truncation of superfluous N-terminal residues not involved in core packing or receptor binding, removal of a core-facing hydroxyl group, and computational sequence optimization of a connecting loop, resulting in significant stability enhancement. Therefore, improved Wnt antagonism and stability of B12 were both achieved utilizing precise computational design methods while simultaneously enhancing existing specific, high-affinity binding activity against Fz-5 and 8.

The computational redesign of the interface loop was confirmed to potently enhance the Wnt antagonism of this Fz-5/8 specific binder, which previously only competed with Wnt in a limited fashion as a result of its binding mode which made primarily adjacent contacts to the functional surface. Additionally, robust functional inhibition in RNF43 mutant PDAC cell lines demonstrates the utility of such Frizzled subtype-specific binders for intervention in therapeutically relevant contexts. Continued development of such proteins will further enable study of the role of Wnt signaling in disease and subsequent precise therapeutic intervention.

Since robust blockade of the Frizzled lipid-binding site is required for full Wnt inhibition, computational methods for rational targeting of this interface will continue to be essential to these efforts. More broadly, this work represents the successful de novo design of an extended, disulfide-stabilized loop away from the core of a protein to achieve a functional outcome. This approach may be generalized for application in other contexts, and provides a powerful method for the application of computational protein design tools for the engineering of functional activity into an existing, suboptimal protein.

3.7 Materials and Methods

Computational Redesign of B12 Interface Loop

De novo backbones were generated with the RosettaRemodel application utilizing blueprint files as templates to guide the construction of helical topologies of various lengths[102]. The last visible residues from the B12-Fz-8 crystal structure were defined as the starting endpoints from which new backbones were built, as noted by the bolded residues in the following sequence:

(GVSFSEVMGKQKDEQAREQLKEGMIKIEEQGKKLSETRTQEELQKYVAAVATF
ALQAGF**L**…*disordered_residues*...IGKISGEVYLKLLDLKKAVRAKEKKGLDILNM
VGEIKGTLERVYA)

Helical backbones of various lengths from the C-terminal endpoint (2, 5, 8, or 11 additional amino acids, each corresponding to additional turns of helix) were sampled in combination with N-terminal backbones of degenerate (undefined) secondary structure of various lengths. This resulted in a variety of proposed backbones comprised of various C-

terminal and N-terminal lengths (2/10,12; 5/9,11,12,14; 8/6,7,8,9,14; 11/11,14,15) that were selected for further redesign. Structures were subjected to a Rosetta relaxation protocol after each remodel step to relieve backbone strain prior to additional design.

Because these de novo topologies were built from endpoints and were not backed up by additional core or secondary structure, a disulfide bond was used to rigidify the proposed structures. An additional RosettaRemodel protocol was used to build a stabilizing disulfide bond between the N- and C-terminal portions of the novel secondary structure by defining allowed positions for disulfide construction in the specified blueprint file and command line landing range options. A final RosettaRemodel step was then used to idealize the connectivity distal to the engineered disulfide bond for those topologies in which a compatible disulfide bond was built, again sampling backbones of various lengths. Four designs were ultimately chosen for experimental testing based on the spatial positioning of their backbones that were projected to enable robust Wnt antagonism. Design sequences, as well as example blueprints and command line options are found below.

An example command used to execute the RosettaRemodel protocol:

> "/PATH_TO_ROSETTA_REMODEL/remodel.default.linuxgccrelease -s [INPUT PDB STRUCTURE] -remodel:blueprint [BLUEPRINT_FILE] –database /[ROSETTA_DATABSE]/database/ -remodel:use_clusters false -no_optH false -correct -remodel:num_trajectory 30 -remodel:save_top 5 -overwrite -find_neighbors -ex1 -ex2 -chain A"

An example blueprint file specifying construction of a backbone with 11 N-terminal residues of unrestricted secondary structure and 11 C-terminal residues with helical structure:

```
1 L .
2 A .
```

3 C .
4 Q .
5 E .
…(additional residues omitted for brevity)...
175 Q .
176 A .
177 G .
178 F .
179 L D PIKAA L
0 x D ALLAA
0 x D ALLAA
0 x D ALLAA
0 x D ALLAA
0 x D ALLAA
0 x D ALLAA
0 x D ALLAA
0 x D ALLAA
0 x D ALLAA
0 x D ALLAA
0 x D ALLAA
0 x H ALLAA
0 x H ALLAA
0 x H ALLAA
0 x H ALLAA
0 x H ALLAA
0 x H ALLAA
0 x H ALLAA
0 x H ALLAA
0 x H ALLAA
0 x H ALLAA
0 x H ALLAA
180 I H PIKAA I
181 G .
182 K .
…(additional residues omitted for brevity)…
221 R .
222 V .
223 Y .
224 A .

Example RosettaRemodel command line used to execute the disulfide construction blueprint:

"/[PATH_TO_ROSETTA_REMODEL]/remodel.default.linuxgccrelease -database /PATH_TO_ROSETTA_DATABASE/database/ -remodel:use_clusters false -no_optH false -correct -remodel:match_rt_limit 5.0 -remodel:num_trajectory 100 -

remodel:save_top 20 -overwrite -find_neighbors -remodel:build_disulf -remodel:use_pose_relax false -ex1 -ex2 -chain A -remodel:disulf_landing_range [LANDING RANGE STARTING AA, LANDING RANGE FINAL AA] -s [INPUT STRUCTURE] -remodel:blueprint [BLUEPRINT]"

An example blueprint for disulfide bond construction (landing range given as command line option):

```
1 L .
2 A .
3 C .
4 Q .
5 E .
…(additional residues omitted for brevity)…
177 G .
178 F .
179 L .
180 A .
181 A .
182 G .
183 L .
184 A .
185 R D
186 K D
187 T D
188 N D
189 G D
190 P D
191 M D
192 R D
193 R D
194 M D
195 L D
196 G D
197 P D
198 R D
199 A D
200 R D
201 I D
202 I .
203 G .
204 K .
205 I .
```

…(additional residues omitted for brevity)…
243 R .
244 V .
245 Y .
246 A .

An example blueprint file used for final loop rebuilding (after disulfide build step) executed utilizing a command line similar to that used for the first remodeling step:

1 L .
2 A .
3 C .
4 Q .
5 E .
…(additional residues omitted for brevity)…
175 Q .
176 A .
177 G .
178 F .
179 L .
180 A .
181 A .
182 G .
183 L .
184 C .
185 L .
186 R D
187 K D
189 R D
190 R D
196 T D
197 S D
198 S D
199 C .
200 T .
201 K .
202 I .
203 G .
204 K .
205 I .
…(additional residues omitted for brevity)…
243 R .
244 V .

245 Y .
246 A .

Command line to execute Rosetta Relax protocol:

/PATH_TO_ROSETTA_RELAX_APPLICATION/relax.default.linuxgccrelease -database /PATH_TO_ROSETTA_DATABASE/database/ -s [INPUT PDB] -ignore_unrecognized_res -relax:constrain_relax_to_start_coords -relax:coord_constrain_sidechains -relax:ramp_constraints false -ex1 -ex2 -use_input_sc -correct -no_his_his_pairE -no_optH false -flip_HNQ

B12 Remodeled De Novo Designs Selected for Testing (variant region in bold):

>B12rebuilt_H11L11RL11
GVSFSEVMGKQKDEQAREQLKEGMIKIEEQGKKLSETRTQEELQKYVAAVATFA
LQAGFL**AAGLCLRKNKRTTSTSKCTK**IGKISGEVYLKLLDLKKAVRAKEKKGL
DILNMVGEIKGTLERVYA

>B12rebuilt_H11L11RL13
GVSFSEVMGKQKDEQAREQLKEGMIKIEEQGKKLSETRTQEELQKYVAAVATFA
LQAGFL**AAGLCLRKVRRRSSSSSSTCTK**IGKISGEVYLKLLDLKKAVRAKEKKG
LDILNMVGEIKGTLERVYA

>B12rebuilt_H11L11RL7
GVSFSEVMGKQKDEQAREQLKEGMIKIEEQGKKLSETRTQEELQKYVAAVATFA
LQAGFL**AAGLCLTKSTSSSCTK**IGKISGEVYLKLLDLKKAVRAKEKKGLDILNM
VGEIKGTLERVYA

>B12rebuilt_H5L12
GVSFSEVMGKQKDEQAREQLKEGMIKIEEQGKKLSETRTQEELQKYVAAVATFA
LQAGFL**CATRTALLSSRSSSCST**IGKISGEVYLKLLDLKKAVRAKEKKGLDILNM
VGEIKGTLERVYA

Computational Sequence Optimization of Helical Loop

The helix-connecting loop comprising residues 114-120 of the parent B12 sequence was redesigned utilizing a RosettaScripts protocol for sequence optimization. Notably, the ConsensusLoopDesign[103] task operation was used to restrict the amino acid identities at each

position in this loop based on their ABEGO torsion bin and the amino acid preferences of native loop structures from corresponding bins. The redesigned loop sequence is given below:

>B12_L2WTdes_Y148F
EQLKEGMIKIEEQGKKLSETRTQEELQKFVKAVAQFALQAGFLSAGLCLRKNKR
TTSTSKCTEIGKISGEVYLKLLDLKKAV**KDPNLSP**LDILNMVGEIKGTLERVYA

Experimental Validation of Designs

DNA sequences corresponding to the selected computationally generated designs were generated utilizing the DNAWorks protocol[85] for optimization, and DNA was obtained from IDT. Designs were tested experimentally by expression of binders in a vector (pETCON that facilitates surface display on yeast via fusion to the native surface protein Aga2[86], a method previously used effectively for the optimization of designed proteins as well as engineered antibodies. EBY100 yeast were transformed, cultured in SDCAA media, and then transferred to SGCAA media for 48 hours at 22°C to induce surface display of the designed proteins[87]. Yeast (1x10^6) were washed with PBSF (1xPBS, 1% bovine serum albumin (BSA)), labeled at 4°C for 2 hours with 0.1-1000nM biotinylated Fz-8, washed with ice-cold PBSF, then incubated for 15 minutes at 4°C with 10 nM streptavidin R-Phycoerythrin conjugate (SAPE, ThermoFisher, Waltham, MA). After labeling, samples were washed again with ice-cold PBSF and analyzed on an Accuri C6 flow cytometer.

Error-prone PCR Library Construction and Selection

The variant region of each of the redesigned B12 designs was diversified with the GeneMorph II Random Mutagenesis Kit (Agilent Technologies) utilizing 0.01 ng of starting template DNA according to the manufacturer's instructions over three iterative rounds of

mutagenesis utilizing the “Loop_fwd_primer” and “Loop_rev_primer” olios specified below. This diversified variant pool was subsequently assembled with gBlocks (IDT) comprising adjacent “front” and “back” regions of the protein (amplified with upGS/front_rev and back_rev/downMyc primers respectively), and then amplified (with upGS/downMyc primers) with Phusion High Fidelity DNA polymerase (NEB) according to the manufacturer’s protocol. This library was transformed into EBY100 yeast, induced using standard methods[104], and selected over 5 iterative rounds by collecting the top 1% of binding variants utilizing the BD Influx cell sorter.

Oligos Utilized for Error-prone PCR Library Construction:

>gBlock_front
CAGAAAACCCGCCTGCAGCGCGAAGGTCGCAACCGCCGCAACGTATTTTTGC
AGTTCTTCTTGGGTACGGGTTTCAGACAGTTTCTTACCCTGCTCTTCGATCTTG
ATCATACCCTCTTTCAGCTGTTCACGCGCCTGTTCGTCTTTCTGCTTACCCATA
ACTTCAGAGAAGGACACGCCCATATGGCTAGCCGACCCTCCGCCTCCGCTAC
CGCCTCCACCAGAGCCTCCTCCACCACTAGCCTGCAGAGCGTAGTCTGGAAC
GTCGTATGGGTATCTACCTTCAATCGTCGAGCTATTGTCCTTAAAAACATACT
GTGTGTTTATGGGGCTGCCTTTGCTAGTTGTTGAGGGGTGAGAACCGCAATTA
CTGACAAACGTTACTGATTTGTAATATTCAAAAACTCCTTGCATTGCCTTCCC
GTTGGCCAAAATAGTAGTCGTTGACAAAGAGTACGGCGTCGATTCTAAAGTT
GGTGAGGGGATTTGCTCGCATATAG

>gBlock_back
GTACGAGCTAAAAGTACAGTGGGAACAAAGTCGATTTTGTTACATCTACACT
GTTGTTATCAGATCTCTATTACAAGTCCTCTTCAGAAATAAGCTTTTGTTCGG
ATCCGCCCCCCTCGAGCGCGTAAACACGTTCCAGGGTACCTTTGATTTCACCA
ACCATGTTGAGAATGTCCAGACCCTTCTTTTCTTTCGCACGAACCGCTTTTTTG
AGGTCGAGGAGTTTCAGGTAAACTTCGCCAGAGATTTTACCGAT

>Loop_fwd_primer
GCGCTGCAGGCGGGTTTTCTG

>Loop_rev_primer
TCAGGTAAACTTCGCCAGAGATTTTACCGAT

>Front_rev_primer

CAGAAAACCCGCCTGCAGCGC

>Back_fwd_primer
ATCGGTAAAATCTCTGGCGAAGTTTACCTGA

>upGS_Fwd_primer
GGACAATAGCTCGACGATTGAAGGTAGATACCCATA

>Down_Myc_primer
CAAGTCCTCTTCAGAAATAAGCTTTTGTTC

Sequencing of individual yeast colonies confirmed that the library had converged to a limited number of variants, the most frequent of which were titrated, and confirmed to now bind several fold tighter affinity via yeast surface display compared to the original B12 variant. Interestingly, a number of these variants contained significant N-terminal truncations, providing confirmation of parallel stabilization efforts to eliminate this non-essential region of the protein. Notably, several of the variants also contained a proline immediately after the disulfide-bonded cysteine in the lipid-binding groove. This region was designed to be helical, but adoption of alternative structure downstream of this helix may enable more optimal formation of interface interactions. Since the disulfide-stabilized region of the constructed topology was unaltered, these mutations were not thought to affect the primary engineered element of the built backbone.

Sequences of Selected B12 Loop Variants from Error-prone PCR Library (Variant Region Bolded):

>B12_DSL1
EQLKEGMIKIEEQGKKLSETRTQEELQKYVAAVATFALQAGFL**AAGLCLRKNK**
RTNSTSKCTKIGKISGEVYLKLLDLKKAVRAKEKKGLDILNMVGEIKGTLERVY
A

>B12_DSL2
EQLKEGMIKIEEQGKKLSETRTQEELQKYVAAVATFALQAGFL**AAGLCLPKNK**
RTTSTSKCTKIGKISGEVYLKLLDLKKAVRAKEKKGLDILNMVGEIKGTLERVY
A

>B12_DSL3
GVSFSEVMGKQKDEQAREQLKEGMIKIEEQGKKLSETRTQEELQKYVAAVATFA
LQAGFL**SAGLCPRKVRRRSSSSSSTCTKIG**KISGEVYLKLLDLKKAVRAKEKKG
LDILNMVGEIKGTLERVYA

>B12_DSL4
GVSFSEVMGKQKDEQAREQLKEGMIKIEEQGKKLSETRTQEELQKYVAAVATFA
LQAGFL**AAGLCPRKNKRTTSTSKCTKIG**KISGEVYLKLLDLKKAVRAKEKKGL
DILNMVGEIKGTLERVYA

>B12_DSL5
EQAREQLKEGMIKIEEQGKKLSETRTQEELQKYVAAVATFALQAGFL**SAGLCPR**
KNKRTNSTSKCTKIGKISGEVYLKLLDLKKAVRAKEKKGLDILNMVGEIKGTLE
RVYA

Site-Saturation Mutagenesis Affinity Maturation

To further improve the binding affinity of the evolved B12 variants, a site-saturation mutagenesis (SSM) approach to affinity maturation was used[70]. A SSM library of the B12_DSL5 variant was constructed via an overlapping PCR method[41] utilizing forward and reverse primers (IDT) containing a degenerate "NNK" codon along with 21 bp flanking regions at each targeted position. In order to more fully optimize the interface in the context of the substantial loop redesign, positions corresponding to both of the interface-contacting helices were included in addition to the interface loop. Naïve and selected libraries were prepared and sequenced using an Illumina Miseq according to manufacturer's protocols, and data was processed and analyzed using standard genomic analysis software[90,91]. Amino acid identities corresponding to the most enriched positions were included in a combination library assembled from four ultramer oligos (IDT), selection of which yielded a final, optimized variant. The variants both contain A167K and K202E mutations that optimize the charge complementarity of the binder, but have divergent mutations within the interface loop region.

B12 Variant used as SSM Library Parent Sequence:

>B12_DSL5 [B12rebuilt_H11L11RL11_A180S_L185P_T192N_Nterm13del]
[variant region in **bold**, error-prone PCR mutations ***italicized***]
EQAREQLKEGMIKIEEQGKKLSETRTQEELQKYVAAVATFALQAGFL***S*AGLC*P*R**
KNKRT*N*STSKCTKIGKISGEVYLKLLDLKKAVRAKEKKGLDILNMVGEIKGTLE
RVYA

Positions Included in SSM Library in **Bold**:

EQAREQLKEGMIKIEEQGKKLSETRT**QEELQKYVAAVATFALQAGFLSAGLCP**
RKNKRTNSTSKCTKIGKISGEVYLKLLDLKKAVRAKEKKGLDILNMVGEIKG
TLERVYA

Ultramer Oligos and Primers (IDT) Used for Construction of Combination Library (Variant Positions in **Bold**):

>UM1_124bp
CCAGACTACGCTCTGCAGGCTAGTGGTGGAGGAGGCTCTGGTGGAGGCGGTA
GCGGAGGCGGAGGGTCGGCTAGCCATGAACAGGCGCGTGAACAGCTGAAAG
AAGGCATGATCAAGATCGAAG
>UM2_134bp
CACAGACCAGCCGACAGGAAACCAGCCTGCAGCGCGAAC**KKC**GCAACCGCT
KYAACGTATTTCTGCAGTTCTTCCTGGGTACGGGTTTCAGACAGTTTTTTACC
CTGTTCTTCGATCTTGATCATGCCTTCTTT
>UM3_143bp
TCCTGTCGGCTGGTCTGTGCCCG**BDMDWMAWTDMMDRM**ACCAACTCTACC
TCCAAATGCACC**RMA**ATCGGT**VAA**ATCTCTGGTGAAGTTTACCTGAAGCTGC
TGGATCTGAAAAAAGCGGTTCGTGCGAAAGAGAAGAAAGGT
>UM4_150bp
GTTCGTGCGAAAGAGAAGAAAGGTCTGGACATCCTGAATATGGTTGGTGAAA
TCAAAGGTACCCTCGAACGTGTGTATGCGCTCGAGGGGGGCGGATCCGAACA
AAAGCTTATTTCTGAAGAGGACTTGTAATAGAGATCTGATAACAAC
>Primer_Front_fwd
CCAGACTACGCTCTGCAGGCT
>Primer_Back_rev
GTTGTTATCAGATCTCTATTACAAGTCCTCTTCAGA

Selected B12 Variants from Combination Library (Mutations in **Bold**):

>B12CL_1
EQAREQLKEGMIKIEEQGKKLSETRTQEELQKYV**K**AVATFALQAGFLSAGLCP**V**
KN**SK**TNSTSKCT**E**IGKISGEVYLKLLDLKKAVRAKEKKGLDILNMVGEIKGTLER
VYA
>B12CL_2

EQAREQLKEGMIKIEEQGKKLSETRTQEELQKYV**K**AVA**Q**FALQAGFLSAGLCPR
NITKTNSTSKCT**E**IGKISGEVYLKLLDLKKAVRAKEKKGLDILNMVGEIKGTLER
VYA

Protein Purification and Characterization

Structural studies and affinity measurements were carried out by Claudia Janda as previously described:

Proteins were cloned into pet29b and purified with C-terminal 6His-tags from *E. coli* BL21(DE3) by NiNTA-affinity chromatography (QIAGEN) and SEC (GE Healthcare). Binding measurements were performed by surface plasmon resonance on a BIAcore T100 (GE Healthcare) and all proteins were purified on SEC prior to experiments. Biotinylated Fzd5-CRD and Fzd8-CRD were coupled at a low density to streptavidin on a SA sensor chip (GE Healthcare). An unrelated biotinylated protein was captured at equivalent coupling density to the control flow cells. Increasing concentrations of B12 variants were flown over the chip in1xHBS-P (GE Healthcare) containing 10 % glycerol and 0.05 % BSA at 40 µl/ml. The chip surface was regenerated after each injection 4 M $MgCl_2$ in HBS-P (B12 measurements) for 60 seconds. Curves were reference-subtracted and all data were analyzed using the Biacore T100 evaluation software version 2.0 with a 1:1 Langmuir binding model to determine the K_D values.

Circular Dichroism

B12L2_WTdes was dialyzed into PBS (20mM NaPO4, 150mM NaCl) and analyzed at a final concentration of 9 mg/mL. Circular dichroism (CD) spectra were collected on an AVIV Model 420 CD spectrometer (AVIV Biomedical, Inc, Lakewood, NJ, USA) using a 1 mm

pathlength quartz cuvette. Scans were collected at 25°C and were taken from 195 to 265 nm in 1 nm steps with 1 nm bandwidth at a scanning speed of 10 nm/minute. Three independent scans were averaged and buffer subtracted against a cuvette holding PBS. Temperature melts were carried out with the same parameters from 25°C to 95°C in 1°C steps reading at 222 nm.

Nanoparticle Production

Nanoparticles were expressed as previously specified with components that contained the Spy-Catcher domain, while the B12 binder contained an N-terminal SpyTag[102]. Individual incubation of the nanoparticle components with the binder resulted in functional conjugation prior to nanoparticle assembly.

Functional Characterization of B12 in Cell Signaling Assays

Functional characterization was carried out as detailed by Jason Berndt:

Cell culture

Cell lines were obtained from the ATCC and cultured according to the methods specified by the ATCC. PDAC cells were stimulated with DMSO (0.1%), IWP2 (4μM, Calbiochem), LGK974 (1μM, SelleckChem), PBS (0.1%), or B12 (0.1-100nM). For growth assays, PDAC cells were seeded at 1,000 cells per well in 96 well plates and cultured for seven days with a media change on day four. On day seven the media was replaced with media containing resazurine (1.25μg/mL, Sigma) for 2.5 hours. Fluorescence intensity (Ex = 530 nM, Em = 580 nM) was measured using a multi-label plate reader (Envision, PerkinElmer).

Wnt reporter assays

BAR PDAC cell lines were generated and assayed essentially as described (10.1101/pdb.prot5223). Briefly, lentivirus encoding BAR - firefly luciferase and PGK - puromycin resistance gene or CMV- Renilla luciferase were produced in HEK293T cells and used to infect the indicated PDAC cell lines. PDAC BAR cells were selected for a minimum of seven days using puromycin. For reporter assays, BAR cells were seeded at 10,000 cells per well in 96 well plates. The following day cells were stimulated with the indicated molecules. After 18-24 hours of stimulation, cells were lysed and luciferease activity was determined using the Dual-Luciferase® Reporter Assay System (Promega) and a multi-label plate reader (Envision, PerkinElmer).

RT-qPCR

To measure gene expression, PDAC cells were plated at 100,000 cells per well in 12 well plates. The following day cells were stimulated with the indicated molecules. After 18-24 hours of stimulation, cells were rinsed in PBS and frozen at -80C. Total RNA was extracted using TRIzol® (ThermoFisher) and quantified using NanoDrop UV spectrophotometer (ThermoFisher). Reverse transcription was performed using 1-2µg of RNA and the RevertAid First Strand cDNA Synthesis Kit (ThermoFisher). qPCR was performed using 10µL reactions in 96 well plates using the Kapa SYBR® Fast qPCR Master Mix (Kapa Biosystems) and a Lightcycler 480® System (Roche). Primers used were GAPDH (TGAAGGTCGGAGTCAACGGA, CCATTGATGACAAGCTTCCCG), NKD1 (GGAGAGAGTGAGCGAACCCT, CTTGCCGTTGTTGTCAAAGTC), AXIN2

(GCGATCCTGTTAATCCTTATCAC, AATTCCATCTACACTGCTGTC), and MUC5AC (AGCCGGGAACCTACTACTCG, AAGTGGTCATAGGCTTCGTGC). Relative transcript abundance was calculated using the 2 (-ΔΔCT) method (10.1006/meth.2001.1262).

3.8 Acknowledgements

Claudia Janda obtained structural information for the parent B12 binder as well as the B12-L2 variant, measured binding affinities and did initial functional assays in A549 cells.

Jason Berdnt was responsible for experimental design, experimentation, data analysis and writing related to the functional characterization of the redesigned B12 binder in both signaling and proliferation assays.

Lindsey Wimberly assisted Jason Berndt with functional characterization experimentation for of B12 variants.

Lauren Carter and Rashmi Ravichandran were responsible for purification of proteins for functional characterization.

Thanks to TJ Brunette and Tom Linsky for valuable consultations on loop design.

3.9 Figures and Tables

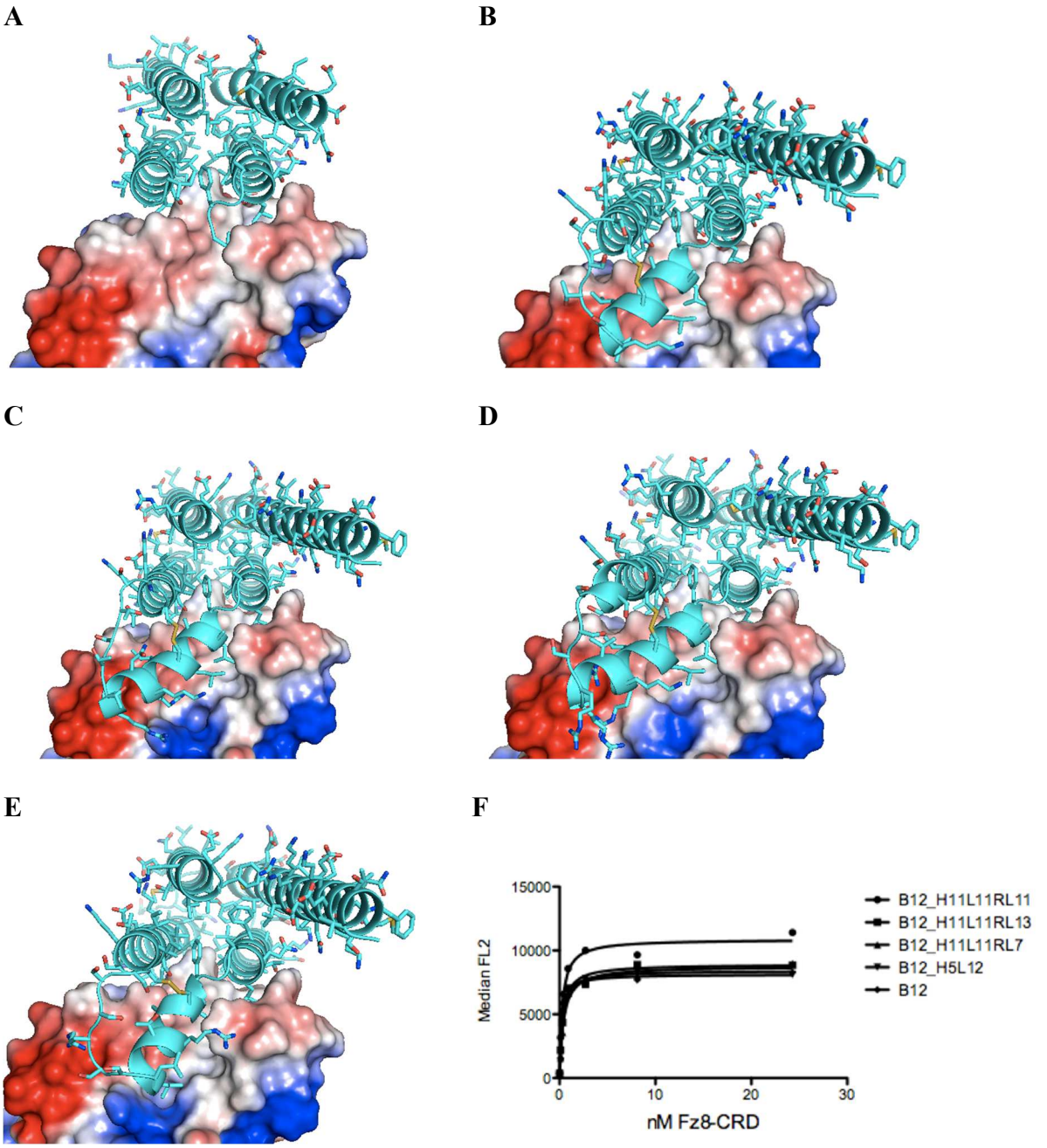
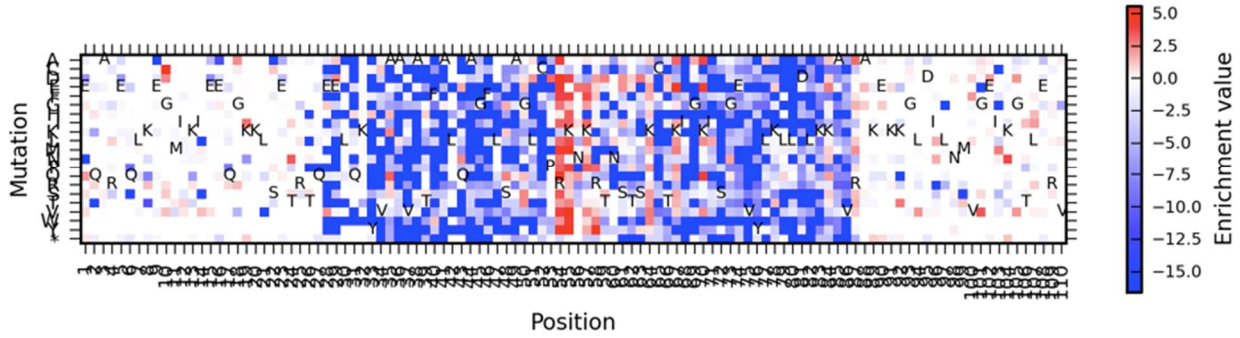


Figure 10. Computational Redesign of B12 Interface Loop.

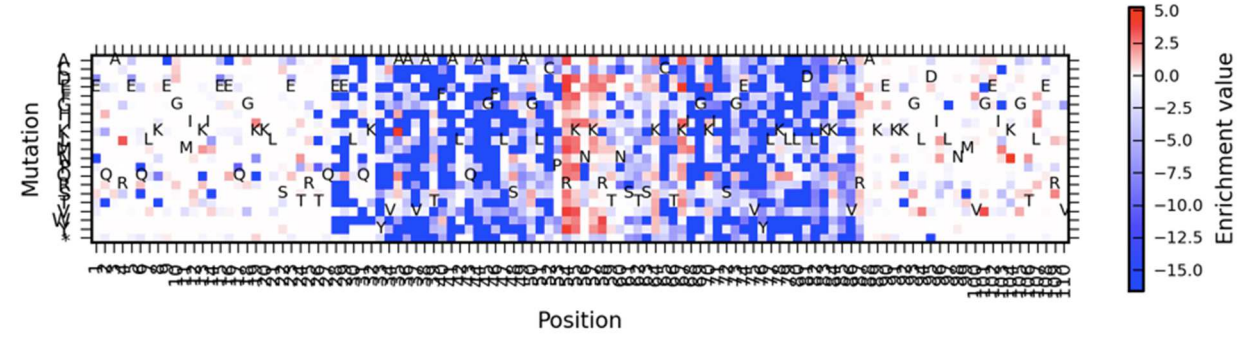
In the original B12 binding mode, the interface loop of B12 inserts partially into Frizzled-8 lipid binding cleft, however the majority of the loop is disordered and undefined in the crystal structure (A), explaining its weak antagonist activity. Four de novo disulfide-stabilized helical topologies were generated computationally utilizing Rosetta Remodel: B12rebuilt_H11L11RL7 (B), B12rebuilt_H11L11RL11 (C), B12rebuilt_H11L11RL13(D), and B12rebuilt_H5L12(E). All designs were confirmed to bind Frizzled-8 similar to the parent B12 (F).

A)

Frizzled-5



Frizzled-8



B)

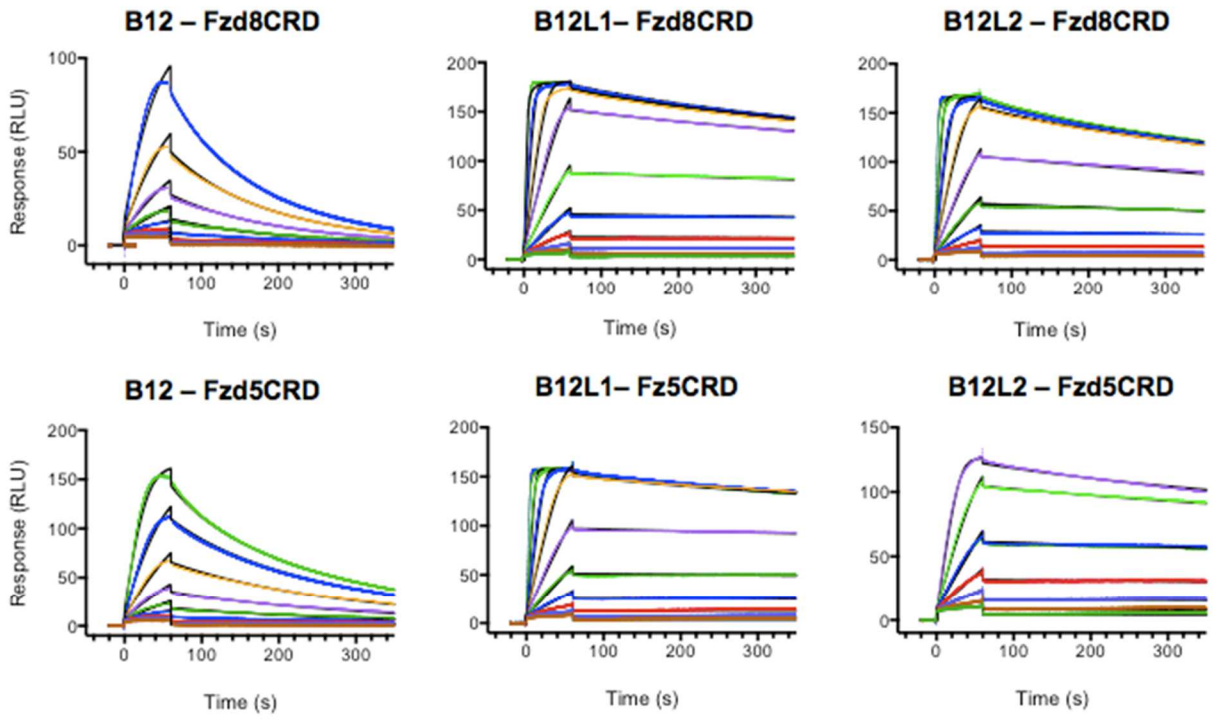
```
B12      G V S F S E V M G K Q K D E Q A R E Q L K E G M I K I E E Q G K K L S E T R T Q E E L Q K Y V A A V A T F A L
B12-L1                             E Q A R E Q L K E G M I K I E E Q G K K L S E T R T Q E E L Q K Y V K A V A T F A L
B12-L2                             E Q R V N R L K E G M I K I E E Q G K K L S E T R T Q E E L Q K Y V K A V A Q F A L

B12      Q A G F L G P N L E E R R G F N R R G K E - - - E I G K I S G E V Y L K L L D L K K A V R A K E K K G L D I L
B12-L1   Q A G F L S A G L C P V K N S K T N S T S K C T E I G K I S G E V Y L K L L D L K K A V R A K E K K G L D I L
B12-L2   Q A G F L S A G L C P R N I T K T N S T S K C T E I G K I S G E V Y L K L L D L K K A V R A K E K K G L D I L

B12      N M V G E I K G T L E R V Y A
B12-L1   N M V G E I K G T L E R V Y A
B12-L2   N M V G E I K G T L E R V Y A
```

Figure 11. **Optimization of B12 Redesigned Loop Binders.**

A) Heat maps show independent enrichment ratios obtained by deep sequencing of a site-saturation mutagenesis library of B12_DSL5 selected for Frizzled-5 and 8 binding, with strong correlation observed between subtypes as expected. B) Sequence alignment comparing optimized variants B12-L1 and B12-L2 to parent B12. Improved loop variants were obtained by computational design, and iterative selection of error-prone PCR and targeted site-saturation mutagenesis guided combinatorial libraries.



	B12	B12L1	B12L2
K_D Fzd8	17.6 nM	235 pM	347 pM
K_D Fzd5	9.4 nM	98 pM	122 pM

Figure 12. Optimized B12 Variants Bind Frizzled-5 and 8 with High Affinity

B12L2 and B12L2 were generated from parent binder B12 via computational design and affinity maturation efforts. The improved variants bind Frizzled-5 and 8 with sub-nanomolar affinity by surface plasmon resonance, representing an improvement in affinity of close to two orders of magnitude compared to the original binder.

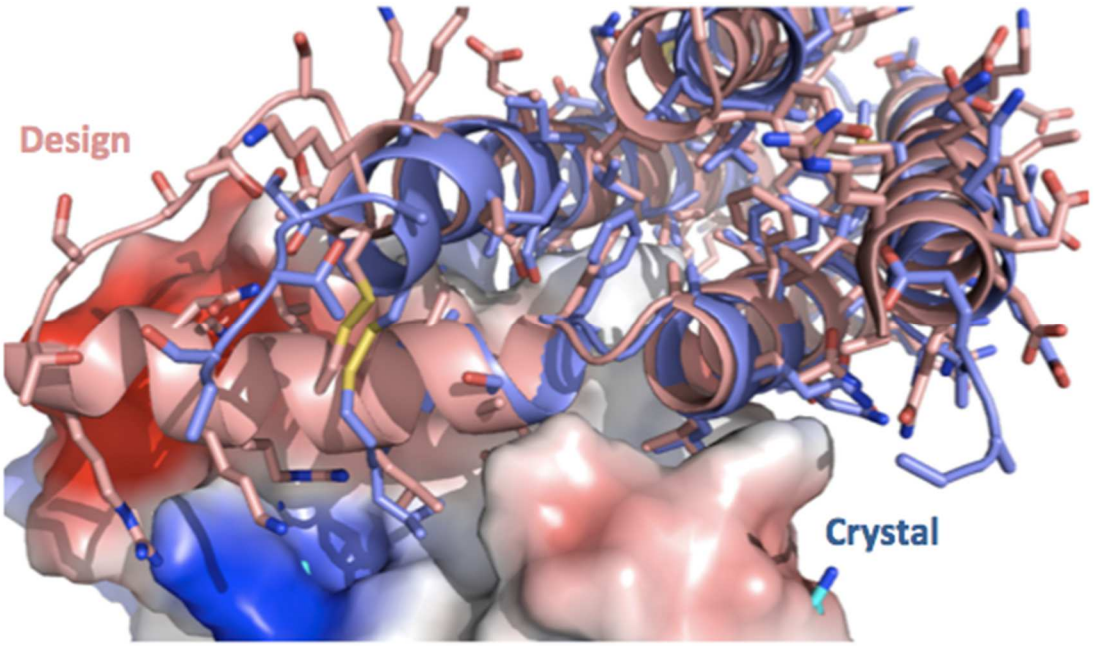


Figure 13. B12L2 Structure Confirms Designed Backbone and Stabilizing Disulfide

Redesigned B12L2 retains binding mode of original B12 parent, while designed disulfide stabilized helical backbone and engineered disulfide bond is positioned as designed. Structure distal to the disulfide bond remains disordered but the position of the N-terminus of the disordered structure originates within lipid binding cleft in order to enable additional Wnt antagonism.

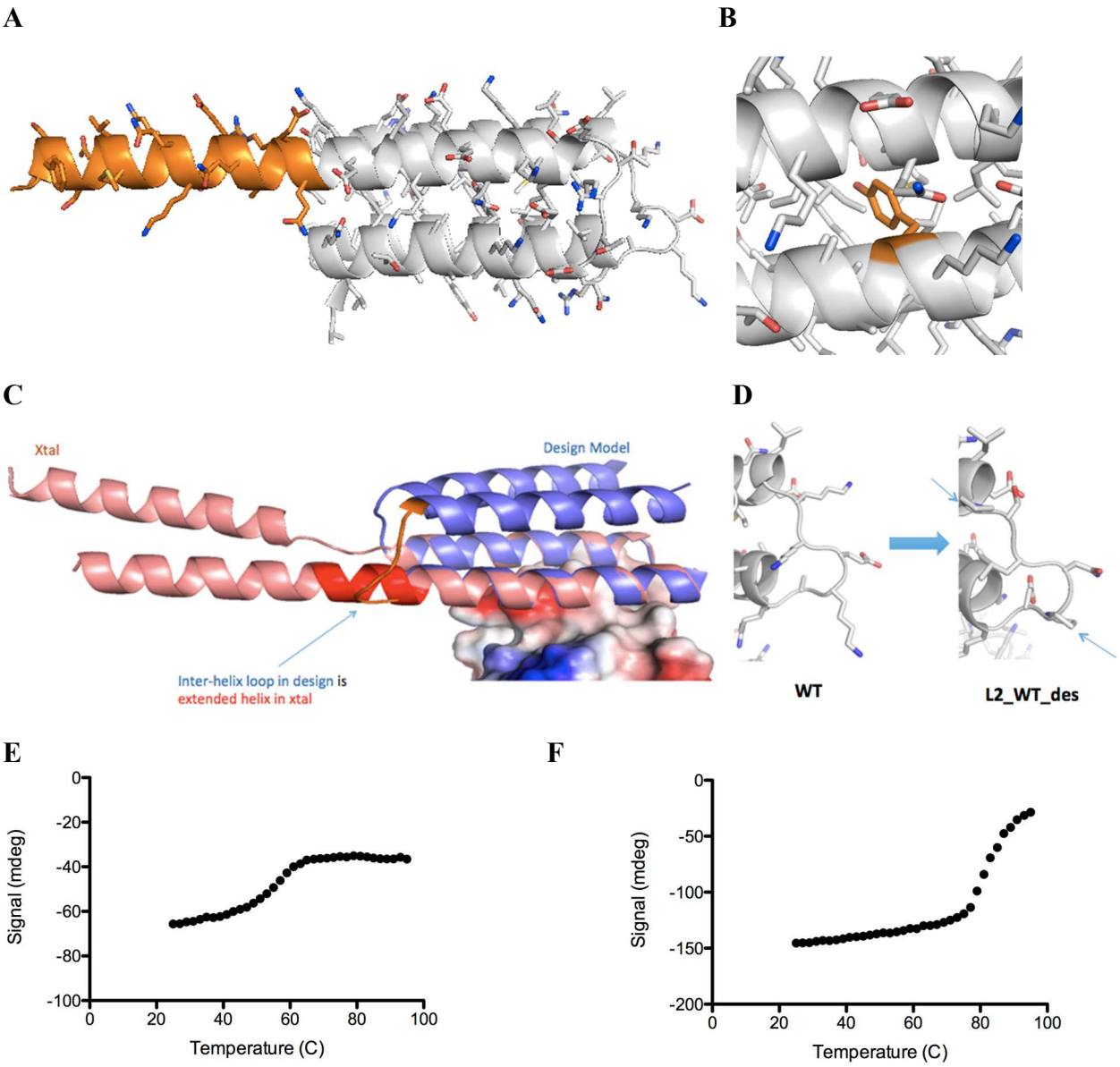
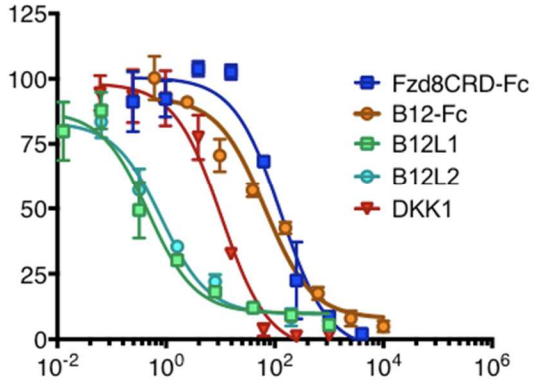


Figure 14. Rational Stabilization of B12.

To increase stabilization, 3 regions of the B12 were selected for modification: an extended N-terminal helix not backed up by core interactions, a destabilizing buried hydroxyl group, and an unstable connecting loop. A) B12 N-terminal helix (orange) does not make interactions with the interface-contacting regions or core of the protein; truncation of up to 19 amino acids is dispensable for binding. B) Wild-type tyrosine 44 (orange) present in the original 2QUP scaffold points its hydroxyl group into the core of the protein; substitution with phenylalanine removes this destabilizing group. C) Domain swapping seen in B12 crystal structure reveals alternative extended helical conformation (salmon) to expected four-helix fold (blue). D) Computational loop redesign of this connecting region followed by consensus sequence redesign of generated structures results in stabilized loop. E) Final B12 variant is significantly more stable than original variant as measured by circular dichroism.



Antagonist	IC_{50}
B12-Fc	65 nM
B12L1	464 pM
B12L2	806 pM
Fzd8CRD-Fc	129 nM
DKK-1	10 nM

Figure 15. B12 Redesigned Loop Variants Inhibit Wnt Signaling in A549 Cells.

Optimized variants B12L2 and B12L2 potently inhibit Wnt signaling as measured by a Wnt/β-catenin reporter (BAR) in A549 cells treated with 50 nM of XWnt8.

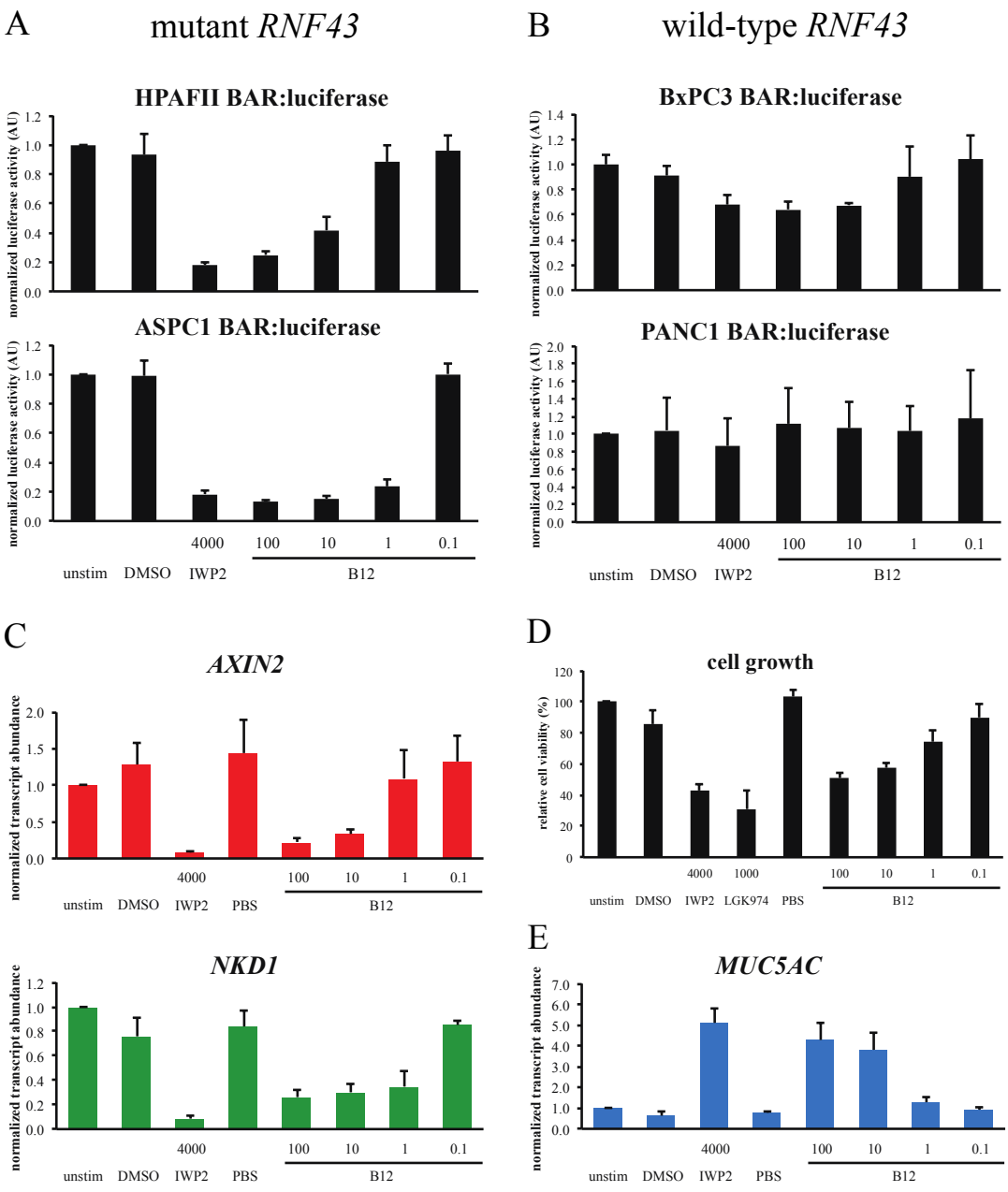


Figure 16. Redesigned B12 Specifically Antagonizes Wnt Signaling in RNF43 Mutant PDAC Cells

B12 inhibits Wnt/β-catenin signaling and growth in Wnt-addicted pancreatic ductal adenocarcinoma (PDAC) cell lines. (A-B) Activity of a Wnt/β-catenin reporter (BAR) in PDAC cell lines with (A) or without (B) loss of function mutations in the gene encoding the E3 ubiquitin ligase RNF43. Cells were stimulated with the indicated molecules for 18-24 hours. Data represent the mean +/- standard deviation of N = 3 independent experiments. (C) Relative expression of endogenous Wnt/β-catenin target genes (NKD1 and AXIN2) in HPAFII cells determined by RT-qPCR. Cells were stimulated with the indicated molecules for 18-24 hours. Data represent the mean +/- SEM of N = 3 independent experiments. (D) Relative growth of

HPAFII cells after 7 days in the presence of the indicated molecules. Data represent the mean +/- standard deviation of N = 4 independent experiments. (E) Relative expression of a gene (MUC5AC) associated with the differentiation of PDAC cells. Cells were stimulated with the indicated molecules for 18-24 hours. Data represent the mean +/- SEM of N = 3 independent experiments. For (A-E) Numbers along the x-axis represent the molecule concentration in nM. Data are normalized to unstimulated cells (unstim).

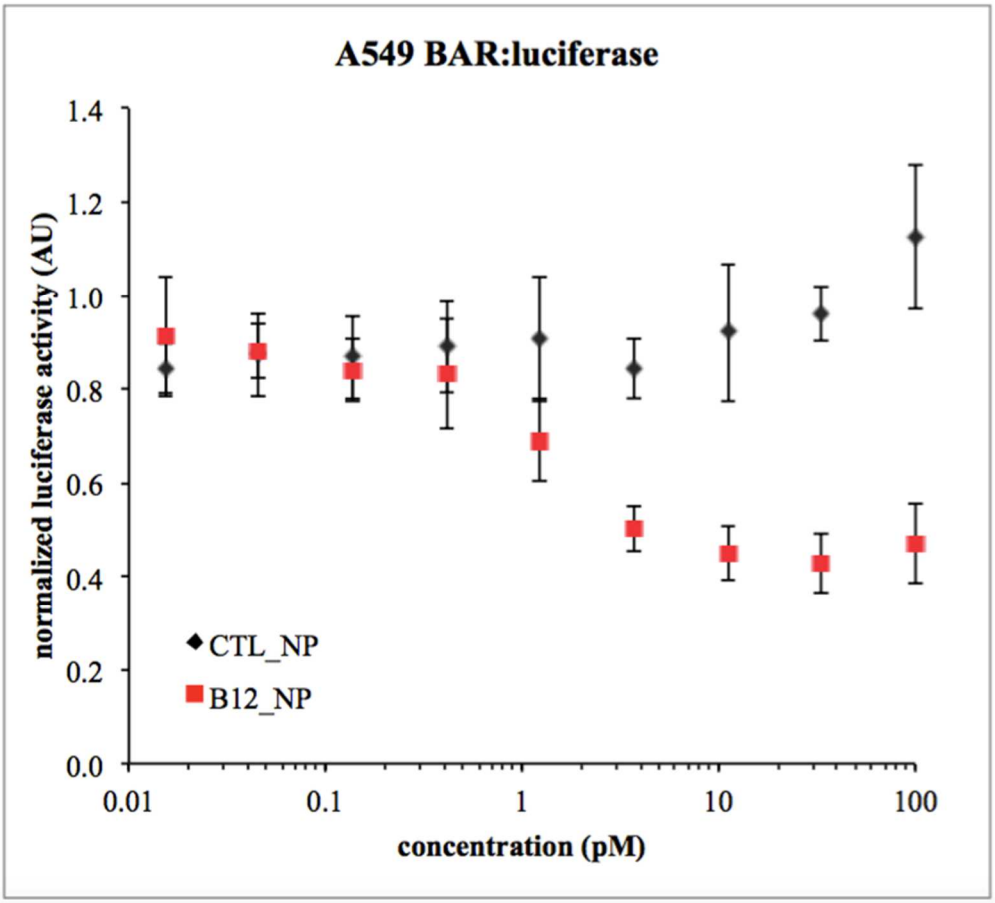


Figure 17. **B12 Functionalized Nanoparticles Inhibit Wnt Signaling.**

Activity of a Wnt/β-catenin reporter (BAR) in A549 cells treated with control or B12-conjugated nanoparticles. Data represent the mean +/- standard deviation of n=3 independent experiments.

Section 4: Computational Design of Novel Ankyrin Binders for Frizzled Subtype-Specific Wnt Antagonism

4.1 Abstract

Designed ankyrin repeat proteins represent a powerful platform for the development of stable, expressible, high affinity protein binders. Their small size and advantageous biophysical properties make them strong candidates for therapeutic applications, and a number of ankyrin-derived biotherapeutics are currently in clinical and preclinical studies. Previous identification of ankyrin binders against targets of interest has utilized empiric selection of randomized libraries. We report here the targeted computational design of a specific, high-affinity ankyrin (ANK12) against the Wnt receptor Frizzled-8 utilizing a novel "fingers-in-groove" binding mode rationally selected for its high shape complementarity and large buried interface. The inherent stability of the ankyrin canonical backbone enabled redesign and optimization of variant regions of the interface resulting in successful engineering of additional binders specific for the Frizzled-1/7 and Frizzled-4 subtypes. These binders potently inhibit Wnt signaling in a subtype-specific manner, providing the means for precise therapeutic intervention with limited off-target effects.

4.2 Background

Ankyrins are widely used in nature to mediate a diverse set of protein-protein interactions and are notable for their conserved canonical solenoid structures that are comprised of 33 amino acid helix-loop-helix-β-hairpin/loop repeat units[105]. The ankyrin fold provides a modular framework for the assembly of extended proteins and has been successfully adapted as a platform for the development of small protein therapeutics[106,107]. Positions involved in

hydrophobic core or inter-repeat interactions are highly conserved, whereas other surface positions, in particular at the concave recognition surface, exhibit more variability[108]. Ideal ankyrins generated from consensus sequence alignments are highly thermostable, and similarity to the conserved, ideal sequence correlates with foldability and stability[109]. Because of their high initial thermostability, ankyrins can tolerate significant variation at non-conserved positions (as required to confer functional activity) without significant detriment to their favorable physical properties. Indeed, blind selection and affinity maturation of ankyrins from diversified interface libraries has resulted in a number of high-affinity, stable binders against targets of therapeutic interest, some of which are currently in pre-clinical and clinical trials[110]. However, such empiric methods cannot explicitly target specific molecular surfaces. In contrast, computational methods enable the rational design of specified binding modes as required for applications that have precise interface requirements beyond crude antagonism.

One such case is the development of subtype-specific binders against the lipid-binding cleft of Frizzled, the Wnt cell surface receptor. Dysregulation of Wnt signaling is widely implicated in various cancer types[4,46, 111,112,113,114], however full elucidation of the role of this pathway in tissue homeostasis and disease has been limited by the availability of reagents which can modulate signaling in a receptor subtype specific manner. The degeneracy and functional implications of the combinatorial interactions between 19 distinct Wnt ligands and 10 Fz receptors (as well as other coreceptors, inhibitors, and agonists) has therefore to date remained unresolved[115]. Blockade of individual Frizzled receptors would allow disentanglement of this complex signaling network and the roles specific Wnt ligand-receptor interactions play in different tissue and cancer types. To achieve this objective, binders must simultaneously antagonize the native Wnt-Frizzled interaction through robust blockade of the conserved lipid-

cleft as well as discriminate between closely related receptors based on adjacent positions that vary between subtypes.

The previous development of B12, a novel binder against this site demonstrated that this is a non-trivial task. Although B12 specifically bound Fz-5/8 through its primary interaction with a subtype specific tryptophan residue adjacent to the Fz hydrophobic groove, its original binding mode did not robustly antagonize Wnt and required additional engineering to achieve full functional effect. In addition, because its binding mode was inherently dependent upon Fz-5/8 specific interactions, it did not enable reengineering of binding specificity for development of a full portfolio of binders specific for various Fz subtypes.

4.3 Approach

We therefore sought to develop high-affinity subtype-specific Fz binders with extensive interfaces that included primary contacts with both the conserved lipid-binding cleft (to enable Wnt antagonism) as well as secondary contacts against adjacent positions which are variable between receptor subtypes (to enable development of orthogonal binding specificities). As discussed previously, empiric affinity maturation methods do not enable precise targeting as is required here, therefore we utilized computational methods to enable the rational design of interfaces that satisfied the functional requirements of the desired application. In order to obtain a binder with characteristics suitable for downstream applications, for a parent scaffold we chose an idealized ankyrin (PDBID 4GPM) previously generated by combining existing sequence and structural information from native proteins with computational Rosetta de novo folding and design methods[116]. Our approach combines the precision of rational computational design with the well-characterized, robust physical properties of the idealized ankyrin scaffold platform, synergistically taking advantage of the strengths of current protein engineering approaches.

4.3 Computational Design, Experimental Optimization, and Structural Characterization of Ankyrin Binder ANK12

Ankyrin scaffolds were globally docked against the target Fz-8 surface to sample shape-complementary binding modes. Backbone conformations were examined, and docks with a novel "fingers-in-groove" binding mode that also grasps the exposed cleft-adjacent helix utilizing the concave ankyrin surface were selected. Compared to other sampled conformations, this binding mode uniquely and elegantly satisfies the structural requirements of this application by making the large number of contacts required for Wnt antagonism and subtype discrimination. Docks were subsequently refined via local perturbation of scaffold backbones and RosettaScripts interface optimization. After refinement, the 29 computationally designed binder interfaces selected for experimental validation exhibited remarkable shape-complementarity (average SC of 0.70) given their extensive size (average SASA >2400 A^2) (Table 4).

The designed Fz-8 binders did not initially exhibit functional activity as assayed by yeast surface display. De novo interface design is inherently challenging, and across a large designed interface even a single mutation may interfere with the engineered function. In contrast, near-neighbor variants may exhibit detectable binding, revealing the inherent shape-complementarity and compatibility of the designed interface against the target. Therefore, designs were diversified utilizing error-prone PCR and selected for binding via cell-sorting methods, resulting in identification of a variant, ANK12, that now binds Fz-8 with high-affinity (Figure 19A). Retrospective analysis of the mutations to the original design revealed that most were non-essential, and that binding activity was conferred by only one (F42A and T66S point mutants demonstrate weak, avid binding activity) to two (F42A/T66S variant exhibits high-affinity non-

avid binding) mutations to the original design whose interface is over 2000 square angstroms (Table 5, Figure 18A-B).

After further optimization to increase binding affinity utilizing enriched mutations identified by a site-saturation mutagenesis approach[70] (Figure 20-21), ANK12 specifically bound Fz-5 with a K_D of 220 pM as measured by biolayer inferometry (BLI), and exhibited a thermostable melting curve by circular dichroism (Figure 19 B-C). ANK12 blocked the Fz-8 specific B12 from binding to its lipid-binding cleft epitope in a yeast surface display assay, and also competed with both B12 and an Fz-1/2/5/7/8 specific scFv for Fz-5 binding in vitro as measured by BLI (Figure 22), demonstrating that it interacted with the Fz functional binding site as designed.

Structural characterization of the ANK12-Frizzled-8 complex resulted in determination of a crystal structure to a resolution of 2.3 Å. The crystal structure confirmed that the backbone of the optimized ANK12 binder was remarkably close to the expected, designed binding mode (RMSD 1.6), adopting the rationally specified “fingers-in-groove” or “grasping-hand” conformation to fully pack the Frizzled lipid binding cleft and preclude the native Wnt lipid (Figure 23A-B). Notable differences from the designed structure include a narrowing of the receptor cleft due to shifting of the adjacent Fz alpha helix that provides a mechanistic explanation for why the F42A mutation was required for functional activity. The larger phenylalanine residue originally designed at this position is incompatible with the conformation of the helical backbone seen in the crystal structure, however mutation to alanine relieves this steric clash (Figure 23C).

4.4 Reengineering of ANK12 for the Development of Orthogonal Binding Specificity

As designed, the ANK12 binding mode forms a high-affinity interaction with the Frizzled receptor by making strong hydrophobic contacts with conserved amino acids, in particular at the base of the lipid-binding cleft at the Wnt lipid insertion site. Because the concave surface of the ankyrin scaffold also extends past the adjacent Frizzled alpha helix, this binding mode also interacts with residues which vary between the Fz receptor subtypes, making it ideal for redesign and optimization of proteins with orthogonal binding specificity. We therefore utilized ANK12 as a platform for the design and engineering of Fz-1/7 and Fz-4 specific binders.

In order to precisely and efficiently alter the binding specificity of ANK12, we obtained amino acid level enrichment data for each interface position by selecting a site-saturation mutagenesis library against each of the targeted receptor subtypes followed by deep sequencing of the naïve and final pools (Figure 20-21). Combinatorial libraries targeting each subtype including enriched amino acid types at fifteen positions (Table 6-7) were selected for binding activity, resulting in the identification of variants with reengineered specificity (Figure 24, Table 8). In contrast to the parent B12, which specifically bound the conserved Fz-5 and 8 subtypes with no detectable binding of either the Fz-1/7 or Fz-4 subtypes, evolved Fz-4 variants 4AF24, 4AF30, and 4AF36 bound Fz-4 at high-affinity with respective Kd of 12, 4, and 12 nM by surface plasmon resonance. Kd's for Fz-8 were 31 nM, 500 nM, and > 5uM respectively for these variants. Therefore, the directed engineering of the ANK12 interface utilizing residue-level data resulted in variants with over 100 fold specificity for Fz-4 over Fz-8, starting from a parent binder with low nanomolar binding of Fz-8 but not of Fz-4. This represents a complete reengineering of binding specificity to generate wholly orthogonal Wnt antagonists, validating our approach and demonstrating the value of the rationally designed ANK12 binding footprint.

Similarly, the Fzd-1/7 targeted library was successful in identifying variants which now binds Fz-1/7 with a Kd ~100 nM, and Fz-5/8 with a Kd ~ 50nM.

Structural characterization of the 7AF43 and 1AF34 binders confirmed that they interact with their target receptors utilizing a binding mode that is essentially identical to the parent ANK12 (Figure 25). Therefore, these proteins may be expected to be potent Wnt antagonists, given the degree to which they fully occlude the functional lipid binding groove and even the approach to adjacent structures. Structural examination of these complexes reveals that the A111D mutation is found in both variants and enables appropriate complementary hydrogen bonding with a corresponding lysine of Fz-1/7 that is a glutamate in the Fz-5 and 8 subtypes (glutamine in Fz-4) (Figure 26). This position may be considered a key specificity determining residue for conferring Fz-1/7 binding activity. The successful reengineering of binding specificity in this case marks significant progress towards development of a suite of proteins with orthogonal binding specificities utilizing an identical binding mode. These binders will enable the precise determination of the contributions of various Fz receptor subtypes to Wnt signaling in particular functional contexts.

4.5 Discussion

Native proteins perform their varied functions utilizing interactions that are precise to the atomic-level. Successful modulation of these functions therefore requires similarly precise engineering of novel proteins, which can only be achieved through rational computational design. Effective scientific and therapeutic applications of engineered proteins depend upon their suitable physical properties. Given their canonical structure and amenability to variation of functionally relevant surface residues, ankyrins represent a class of proteins eminently suited to

both computational design and subsequent utilization in demanding downstream applications, in contrast to other folds less able to tolerate substantial functionalizing mutations.

The successful computational design of a de novo interface against the Fz-8 lipid-binding cleft demonstrates the elegance of rational approaches to protein engineering. Moreover, the subsequent reengineering of ANK12 binding specificity validates the utility of the specified binding mode which simultaneously interacts with both the conserved Fz hydrophobic groove and adjacent variant positions using its extensive interface in a manner which would be difficult to replicate utilizing empiric methods. The rational computational design of ankyrin-based binders may be considered a general method that enables precise targeting of relevant molecular surfaces with scaffolds with favorable properties for utilization in downstream therapeutic applications. This approach combines the specificity of computational methods with hyper-stable, modular ankyrin scaffolds to enable the rational engineering of proteins which would have been previously inaccessible.

4.6 Materials and Methods

Computational Discovery and Rational Selection of "Hand-in-Groove" Binding Mode

The Rosetta-built idealized ankyrin 4GPM was used as the parent scaffold. This ankyrin was comprised of four repeat units; scaffolds containing three and five repeats were utilized as well. In order to facilitate discovery of binding modes which interacted as fully as possible with the target surface, residues at the concave interface positions were mutated to alanine to eliminate side chain clashes that could preclude sampling of highly interdigitated binding modes during the docking step. Potential binding modes were generated utilizing the Patchdock[117] algorithm to globally dock scaffolds against the Fz-8 target surface and rank possible interfaces

based upon shape-complementarity. These interfaces were examined visually to select binding modes that both achieved complete blockade of the lipid-binding cleft and included contacts with both the conserved regions of this hydrophobic groove as well as adjacent regions which were varied between receptor subtypes. A number of generated binding modes simultaneously insert multiple loops into the Fz lipid-binding cleft while grasping the adjacent exposed alanine helix utilizing the concave scaffold surface. These interfaces elegantly and efficiently fulfilled all of the design objectives and were therefore selected for further refinement and optimization.

Local Perturbation of Selected Docks

In order to more fully sample optimal backbone positions around the selected binding modes, local perturbation of these designs was performed prior to interface optimization utilizing the MotifDock application which scores backbone positions based on the quality and number of projected interactions accessible for a given structure. The top five locally perturbed backbones were taken forward for interface optimization along with the original dock.

An example command line used to execute the MotifDock Application:

> "time OMP_NUM_THREADS=40
> /[path_to_MotifDock_Application]/sicasym.ompstatic.linuxgccrelease
> @../asym_onesided.flags -s [target.pdb] -t [scaffold.pdb]"

An example asym_onesided.flags file specifying options for the MotifDock application run:

```
# SicDock/SCHEME flags file
################### general I/O flags #################
-nstruct 50                    # number of PDB files to dump
-nscore  1000                     # number of structures / scores to dump into the .dok
".silent" file
-sicdock:sieve:redundancy_dist_cut 0.7  # redundancy threshold in A rmsd, no outputs
will be closer than this to one another
```

```
-mh:dump:max_per_res 99               # max motifs output per residue, can be high for canonical interactions
-sicdock:out:minimal_pdb false        # dump only the backbone
-renumber_pdb                         # not sure what you'd get without this, might be the input numbering
-mute all
-ignore_unrecognized_res   # use at your own peril

################################### local sampling around input structure ###############################
-sicdock:geometry:rotmag  15.0             # limit binder rotation in degrees
-sicdock:geometry:slide    20.0            # limit binder center-of-mass translation
-sicdock:geometry:slide_from_infinity  false # must be false if you're keeping input positions
-sicdock:geometry:keep_input_positions true  # if false, move binder and target centers-of-mass to the origin
-sicdock:sample_resolution  3.0            # how finely to sample conformational space. currently has a large impact on runtime
-sicdock:sample_thickness   2.0            # now thick a "shell" of contacting positions to sample
-mh:dump:max_rms    0.3                    # how stringently should OUTPUT motifs match the frames on which they are overlayed HAS NO EFFECT ON DOCKING

################################### global sampling in a "cone" pointing along -Z axis ###############################
#-sicdock:geometry:slide    9999.9 35.0 0 0 1   # change the second one (35.0) if you want to sample a more narrow or wider cone
#-sicdock:geometry:slide_from_infinity  true    # must be true if you're not keeping both imput positions
#-sicdock:geometry:keep_input_position1 true    # keep the input position of the target, this is important so the sampling "cone" hits your desired region of contact
#-sicdock:sample_resolution  6.0               # how finely to sample conformational space. currently has a large impact on runtime
#-sicdock:sample_thickness   3.0               # now thick a "shell" of contacting positions to sample
#-mh:dump:max_rms    0.5                      # how stringently should OUTPUT motifs match the frames on which they are overlayed HAS NO EFFECT ON DOCKING

# use only one of these
# -clash_coord_pick_mode N_CA_C_CB # clash check with non-O backbone atoms, this is the default (if you clash check O, you can't form sheets)
# -clash_coord_pick_mode BB        # clash check with all backbone atoms
-clash_coord_pick_mode HVY_IF_NP   # clash check with all heavy atoms if residue is hydrophobic, else backbone atoms if the residue is polar
-sicdock:clash_dis 3.3             # radius of the hard spheres which define ""
-sicdock:sieve:min_score     1.0
```

```
-sicdock:sieve:log_interval  10.0

-mh:score:anti_polar_weight  0.0    # generalized penalty for backbone polars near the
interface
-mh:score:strand_pair_weight 1.0    # weight for sheet-forming interactions
-sicdock:weights:motif       1.0    # general weight for all motif scores
-mh:score:noloops 0

-mh:score:background_weight  0.0    # give a bonus or penalty to all contacts regardness
of motifs

-mh:match:ss1 false   # for "explicit" motifs that get dumped at the end, match target SS
-mh:match:ss2 false   # for "explicit" motifs that get dumped at the end, match binder SS
-mh:match:aa1 true    # for "explicit" motifs that get dumped at the end, match target AA
-mh:match:aa2 false   # for "explicit" motifs that get dumped at the end, match binder AA

##### BB_BB aa1: one sided bb/bb motifs #####
# these motifs will allow different rotamers on target residues. Aaron doesn't like them.
# if you think the rotamers on your target can move, uncomment the following two lines
#-mh:path:motifs
/work/sheffler/data/mh3/xs_bb_aa1_ALLAA_resl10.8_smooth1.3_msc0.5_mbv1.0/xs_b
b_aa1_ALLAA_resl10.8_smooth1.3_msc0.5_mbv1.0.rpm.bin.gz
#-mh:path:scores_BB_BB
/work/sheffler/data/mh3/xs_bb_aa1_ALLAA_resl10.8_smooth1.3_msc0.5_mbv1.0/xs_b
b_aa1_ALLAA_resl10.8_smooth1.3_msc0.5_mbv1.0_
-mh:score:use_ss1 false        # applicable only to "BB_BB motifs, match secondary
structure on first (target) res"
-mh:score:use_ss2 false        # applicable only to "BB_BB motifs, match secondary
structure on second (binder) res"
-mh:score:use_aa1 true         # applicable only to "BB_BB motifs, match AA identity on
first (target) res"
-mh:score:use_aa2 false        # applicable only to "BB_BB motifs, match AA identity on
second (binder) res"


##### SC_BB aa1: "hotspot" motifs #####
# these motifs match the rotamer on your target to the backbone of your candidate binder
-mh:path:motifs
/work/sheffler/data/mh3/xs_scbb_aa1_resl0.8_smooth1.3_msc0.4_mbv1.0/xs_scbb_aa1_
resl0.8_smooth1.3_msc0.4_mbv1.0.rpm.bin.gz
-mh:path:scores_SC_BB
/work/sheffler/data/mh3/xs_scbb_aa1_resl0.8_smooth1.3_msc0.4_mbv1.0/xs_scbb_aa1_
resl0.8_smooth1.3_msc0.4_mbv1.0

## SC_SC aa1 aa2: "fa" motifs
```

```
# these motifs match the current rotamers in your target to current rotamers in your
binder. matches will be fairly rare.
-mh:path:motifs
/work/sheffler/data/mh3/xs_scsc_resl0.8_smooth1.3s3_msc0.4_mbv1.0/xs_scsc_resl0.8_
smooth1.3s3_msc0.4_mbv1.0.rpm.bin.gz
-mh:path:scores_SC_SC
/work/sheffler/data/mh3/xs_scsc_resl0.8_smooth1.3s3_msc0.4_mbv1.0/xs_scsc_resl0.8_
smooth1.3s3_msc0.4_mbv1.0

##### PH_PO: strand pairs #####
-mh:path:scores_PH_PO
/work/sheffler/data/mh3/xs_pp_ee_ms1_ssep10_ms1_resl0.8_smooth1s3/xs_pp_ee_ms1
_ssep10_ms1_resl0.8_smooth1s3.xh.bin.gz


##### BB_PH/BB_PO bb/sc hbonds #####
# backbone / sidechain hydrogen bond motifs, they match a peptide bond H/O in the
target to the backbone reference frame on the binder
-mh:path:motifs
/work/sheffler/data/mh3/xs_bpo_resl0.8_smooth1/xs_bpo_resl0.8_smooth1.rpm.bin.gz
-mh:path:motifs
/work/sheffler/data/mh3/xs_bph_resl0.8_smooth1/xs_bph_resl0.8_smooth1.rpm.bin.gz
-mh:path:scores_BB_PO
/work/sheffler/data/mh3/xs_bpo_resl0.8_smooth1/xs_bpo_resl0.8_smooth1.xh.bin.gz
-mh:path:scores_BB_PH
/work/sheffler/data/mh3/xs_bph_resl0.8_smooth1/xs_bph_resl0.8_smooth1.xh.bin.gz

-mh:gen_reverse_motifs_on_load false     # I think the inverse motifs are already in the
datafiles

######## these options apply to linker / loop scoring. leave as is unless you want your
docking partners connected by a linker ####
-sicdock:weights:splice      0.0    #
#-sicdock:thresholds:splice   -9e9
#-sicdock:thresholds:splice   0.001
-sicdock:splice:ends false
-sicdock:splice:gaps false
-sicdock:splice:max_trim 0
#-mh:path:scores_frags
/work/sheffler/data/mh3/frags_all_count10/frags_all_count10.xh.bin.gz
#-mh:path:scores_frags /work/sheffler/data/mh3/frags_all/frags_all.xh.bin.gz


-database /work/sheffler/rosetta/dev/database
-chemical:exclude_patches common # don't load all the usual patches
-detect_disulf false            # should have no effect on motif scoring
```

```
-score_cache 0 # leave this as is unless you want subtle threading bugs
-slide_cache 0 # leave this as is unless you want subtle threading bugs
```

Interface Optimization

After local backbone refinement, the interfaces of the minimal, alanine-mutated scaffolds were optimized utilizing a RosettaScripts[83] protocol to direct design efforts. A greedy optimization[118] mover that optimizes each individual position separately was utilized for full interface refinement. After computational optimization, designs were selected based upon shape complementarity, binding energy (DDG), interface size, and low numbers of unsatisfied polar groups. Finally, interfaces were manually refined in Foldit[81] to optimize charge complementarity utilizing electrostatics maps generated in PyMOL (Schrodinger). Amino acid sequences of designs selected for experimental characterization can be found in Table 2.

Script used to execute RosettaScripts Protocol:

```
#!/bin/bash
#pssm=$2
#resfile=$3
total_score=`grep ^total_score $1 | awk '{print 1/$2*(0-1)}'`
shape=`grep ^shape $1 | awk '{print 1/$2*(0-1)}'`
unsat=`grep ^unsat $1 | awk '{print 1/$2}'`
binding=`grep ^binding $1 | awk '{print 1/$2*(0-1)}'`
/[path_to_rosetta_scripts]/rosetta_scripts.default.linuxgccrelease \
-database /[path_to_rosetta_database]/main/database \
-ignore_zero_occupancy false \
-ignore_unrecognized_res \
-overwrite \
-out:file:renumber_pdb false \
-ex1 \
-ex2 \
-nstruct 1 \
```

```
-parser:script_vars pssm=${pssm} total_score=${total_score} shape=${shape}
binding=${binding} unsat=${unsat} \
-s $1 \
-suffix $2 \
-parser:protocol /[path_to_RosettaScripts_XML_protocol]/greedy_no_fsp.xml \
-holes:dalphaball /[path_to_DAlpahBall]/DAlphaBall.macgcc \
-score:weights talaris2013 \
-chemical:exclude_patches LowerDNA UpperDNA Cterm_amidation SpecialRotamer
VirtualBB ShoveBB VirtualDNAPhosphate VirtualNTerm CTermConnect sc_orbitals
pro_hydroxylated_case1 pro_hydroxylated_case2 ser_phosphorylated
thr_phosphorylated tyr_phosphorylated tyr_sulfated lys_dimethylated
lys_monomethylated lys_trimethylated lys_acetylated glu_carboxylated cys_acetylated
tyr_diiodinated N_acetylated C_methylamidated MethylatedProteinCterm
```

RosettaScripts XML Protocol Used for Interface Optimization:

```
<ROSETTASCRIPTS>
        <SCOREFXNS>
                <sfxn_hard weights=talaris2013>
          </sfxn_hard>
        </SCOREFXNS>

        <TASKOPERATIONS>
                <LimitAromaChi2 name=arochi2/>
                <IncludeCurrent name=inclcur/>
                <DisallowIfNonnative name=nocys disallow_aas=CGHMP/>
                <ProteinInterfaceDesign name=pido repack_chain1=1 repack_chain2=1
    design_chain1=0 design_chain2=1 jump=1 interface_distance_cutoff=10
    allow_all_aas=1/>
          <RestrictChainToRepacking name=rctr chain=1/>
          ReadResfile name=resfile filename="%%resfile%%"/>
        </TASKOPERATIONS>

        <MOVERS>
          FavorSequenceProfile name=fsp_pssm scaling=global chain=2
    pssm="%%pssm%%" weight=1.0 scorefxns=sfxn_hard/>
          <AtomTree name=ftree simple_ft=1/>
          <MinMover name=min_bb_sc bb=0 chi=1 jump=1 scorefxn=sfxn_hard>
                <MoveMap name=minmvrmap>
                      <Chain number=2 chi=1 bb=0/>
                      <Chain number=1 chi=1 bb=0/>
                      <Jump number=1 setting=1/>
                </MoveMap>
          </MinMover>
        </MOVERS>
        <FILTERS>
```

```
		<Ddg name=binding threshold=0 scorefxn=sfxn_hard confidence=0
jump=1 repack=1 relax_mover=min_bb_sc repeats=3/>
		<Sasa name=dsasa threshold=500 confidence=0/>
		<ShapeComplementarity name=shape jump=1 verbose=0 min_sc=0.60
confidence=0/>
		<SymUnsatHbonds name=unsat jump=1 cutoff=1000 confidence=0/>
		<ScoreType name=total_score scorefxn=sfxn_hard
score_type=total_score confidence=0 threshold=0/>
		<CombinedValue name=combo>
			<Add filter_name=total_score factor="%%total_score%%"/>
			<Add filter_name=unsat factor="%%unsat%%"/>
			<Add filter_name=binding factor="%%binding%%"/>
			<Add filter_name=shape factor="%%shape%%"/>
		</CombinedValue>
	</FILTERS>
	<MOVERS>
		<GreedyOptMutationMover name=greedy
task_operations=pido,arochi2,rctr,nocys filter=combo scorefxn=sfxn_hard
relax_mover=min_bb_sc sample_type=low rtmin=0 design_shell=-1 repack_shell=8.0
parallel=1/>
	</MOVERS>
	<PROTOCOLS>
		<Add mover_name=ftree/>
		Add mover_name=fsp_pssm/>
		<Add mover_name=greedy/>
		<Add filter_name=binding/>
		<Add filter_name=dsasa/>
		<Add filter_name=shape/>
		<Add filter_name=unsat/>
		<Add filter_name=total_score/>
	</PROTOCOLS>
</ROSETTASCRIPTS>
```

Experimental Validation of Designs Via Yeast Surface Display

DNA sequences corresponding to the selected computationally generated designs were generated utilizing the DNAWorks protocol[85] for optimization, and DNA was obtained from IDT. Designs were tested experimentally by expression of binders in a vector (pETCON[42]) that facilitates surface display on yeast via fusion to the native surface protein Aga2[86], a method previously used effectively for the optimization of designed proteins as well as engineered

antibodies. EBY100 yeast were transformed, cultured in SDCAA media, and then transferred into SGCAA media for 48 hours at 22°C to induce surface display of the designed proteins[87]. Yeast (1x10^6) were washed with PBSF (1xPBS, 1% bovine serum albumin (BSA)), labeled at 4°C for 2 hours with 1 uM biotinylated Fz-8 preincubated for 15 min with 0.25 uM streptavidin R-Phycoerythrin conjugate (SAPE, ThermoFisher, Waltham, MA) to form avid binding complexes. After labeling, samples were washed with ice-cold PBSF and analyzed on an Accuri C6 flow cytometer. Binding affinities were determined by varying labeling concentrations; non-avid affinities were obtained by labeling only with biotinylated target protein prior to a subsequent labeling step with SAPE (samples were washed with PBSF after each step). Competition assays were conducted by labeling with biotinylated Frizzled proteins preblocked with equimolar amounts of the binder B12.

Error-Prone PCR Library Generation and Selection

In order to test near-neighbors of the parent designs for binding activity, designs were pooled and diversified with the GeneMorph II Random Mutagenesis Kit (Agilent Technologies) utilizing 0.01 ng of starting template DNA according to the manufacturer's instructions over three iterative rounds of mutagenesis utilizing the "upGS" and "downMyc" oligos specified below. This diversified variant pool was amplified (with the upGS/downMyc primer set) utilizing Phusion High Fidelity DNA polymerase (NEB) according to the manufacturer's protocol. This library was transformed into EBY100 yeast, induced using standard methods[119], and selected over 5 iterative rounds by collecting the top 1% of binding variants utilizing the BD Influx cell sorter. Sequencing revealed that the library had converged upon a binder, now

referred to as ANK12 that bound Fz-8 (and the closely related Fz-5), but not Fz-1, 4, or 10 with low nM affinity.

Oligos Utilized for Error-prone PCR Library Construction:

>upGS_Fwd_primer
GGACAATAGCTCGACGATTGAAGGTAGATACCCATA
>Down_Myc_primer
CAAGTCCTCTTCAGAAATAAGCTTTTGTTC

Amino Acid Sequence of Selected "ANK12" Fz-8 Binder

>ANK12
SELGKRLIRAAFDGNKDRVKDLIENGADVNASLVSGATPLHAAAMNGHKEVVK
LLISKGADVNAQSSAGSTPLVAAAINGHKEVVKLLISKGADVNAVTAAGMTPLH
AAAANGHKEVVKLLISKGADVNAKADRGMTPLHFAAWRGHKEVVKLLISKGA
DLNTSAKDGATPLDMARESGNEEVVKLLEKQ

Site-Saturation Mutagenesis Library Construction and Selection

In order obtain empiric individual amino acid data for further reengineering of the binding specificity of ANK12, a site-saturation mutagenesis approach was used[70]. A SSM library of the ANK12 binder was constructed via an overlapping PCR method[41] utilizing forward and reverse primers (IDT) containing a degenerate "NNK" codon along with 21 bp flanking regions at each targeted position. Only interface contacting regions were targeted for mutagenesis; core and non-interface positions were kept fixed. Additionally, the gene was diversified as much as possible at the DNA level in order to reduce non-specific annealing due to sequence homology. After transformation into EBY100 yeast, the library was labeled with biotinylated Fz-4,5,7 and 8 and the top 1% of each population was selected in parallel over three rounds. Naïve and selected libraries were prepared and sequenced using an Illumina Miseq according to manufacturer's protocols, and data was processed and analyzed using standard genomic analysis software[91,92]. Because of the size of the ANK12 gene, it was split into "front" and "back" reads for this

sequencing step. Enrichment ratios can be found in Figures 20-21. Amino acid identities corresponding to the most enriched positions were included in combination libraries targeting Fz-4 and 7 (Tables 6 and 7) and selected to yield final, optimized variants (sequences listed below, detailed analysis found in Table 8). The parent ANK12 variant previously bound Fz-5/8 with high affinity, but based on the SSM data the following mutations were made to revert non-contributing mutations (*italics*) and include enriched amino acid identities:

R9M/F12L/*D13E*/S67A/*L161V*.

ANK12 Codon-Diversified DNA Sequence for SSM Library Construction

TCTGAACTGGGTAAACGTCTGATCCGTGCAGCATTCGACGGTAACAAAGACC
GTGTTAAAGACCTCATTGAAAATGGTGCTGACGTTAACGCGTCCCTCGTGTCT
GGGGCGACTCCGTTACACGCCGCCGCCATGAACGGCCACAAAGAGGTTGTGA
AGTTACTGATCTCCAAGGGCGCAGATGTGAATGCTCAGTCCTCTGCGGGTTCT
ACACCTCTGGTGGCGGCGGCGATCAATGGACATAAAGAAGTGGTAAAACTGC
TGATAAGTAAAGGAGCAGACGTCAATGCTGTTACGGCGGCTGGAATGACCCC
CCTACATGCTGCTGCTGCTAATGGGCACAAGGAAGTAGTGAAGTTGCTTATTT
CTAAGGGGGCCGACGTAAATGCGAAAGCGGACCGTGGCATGACTCCACTCCA
TTTCGCAGCATGGCGCGGTCATAAGGAAGTCGTTAAACTATTAATCAGCAAA
GGTGCGGATTTGAACACCTCTGCCAAAGACGGTGCAACCCCGCTTGACATGG
CGCGTGAATCTGGCAATGAGGAGGTTGTCAAGCTCTTGGAAAAGCAA

Diversified Positions in "Front" Portion of ANK12 Deep Sequencing (<u>Bold/Underlined</u>)

SELG**<u>KR</u>**L**<u>IR</u>**AA**<u>FD</u>**GNKDRVKDLIENGADVNAS**<u>LVSGA</u>**TPL**<u>HA</u>**AA**<u>MN</u>**GHKEVV
KLLISKGADVNAQ**<u>SSAGST</u>**PL**<u>VA</u>**AA**<u>IN</u>**GHKEVVKLLISKGADVNA**<u>VTAAGM</u>**TP
LHAAAANGHKEVVKLLISKGADVNAKADRGMTPLHFAAWRGHKEVVKLLISK
GADLNTSAKDGATPLDMARESGNEEVVKLLEKQ

Diversified Positions in "Back" Portion of ANK12 Deep Sequencing (Bold/Underlined)

SELGKRLIRAAFDGNKDRVKDLIENGADVNASLVSGATPLHAAAMNGHKEVVK
LLISKGADVNAQSSAGSTPLVAAAINGHKEVVKLLISKGADVNAVTAA**<u>GM</u>**TPL**<u>H
A</u>**AA**<u>AN</u>**GHKEVVKLLISKGADVNAK**<u>ADRGM</u>**TPLH**<u>F</u>**AA**<u>WR</u>**GHKEVVKLLISKGA
DLNTSA**<u>KD</u>**GATPLD**<u>M</u>**ARE**<u>S</u>**G**<u>N</u>**EEVVKLLEKQ

Amino Acid Sequences for Selected Fzd-1/4/5/7/8 Targeted Binders

>ANK12_Fzd5/8

SELGKRLIMAALDGNKDRVKDLIENGADVNASLVSGATPLHAAAMNGHKEVVK
LLISKGADVNAQSAAGSTPLAAAAINGHKEVVKLLISKGADVNAVTAAGMTPLH
AAAANGHKEVVKLLISKGADVNAKADRGMTPLHFAAWRGHKEVVKLLISKGA
DLNTSAKDGATPLDMARESGNEEVVKLLEKQ

>1AF34
SELGTRLIRAALDGNKDRVKDLIENGADVNASLMSGATPLHAAAMNGHKEVVK
LLISKGADVNAQSVAGSTPLDAAAFSGHKEVVKLLISKGADVNAVNAAGLTPLH
AAADNGHKEVVKLLISKGADVNAKADHGMTPLHFAAQRGHKEVVKLLISKGA
DLNTSAKDGATPLDMARESGNEEVVKLLEKQ

>7AF43
SELGTRLIRAALDGNKDRVKDLIENGADVNASLESGATPLHAAAMNGHKEVVK
LLISKGADVNAQSVAGSTPLVAAAYSGHKEVVKLLISKGADVNAVYAAGITPLH
AAADNGHKEVVKLLISKGADVNAKAYRGMTPLHFAAQRGHKEVVKLLISKGAD
LNTSAKDGATPLDMARESGNEEVVKLLEKQ

>4AF24
SELGKRLIRAALDGNKDRVKDLIENGADVNASLVSGTTPLYAAAMNGHKEVVM
LLISKGADVNAQSVAGSTPLVAAANFGHKEVVKLLISKGADVNAKHAFGVTPLH
AAAADGHKEVVKLLISKGADVNAKAGRGMTPLHIAALRGHKEVVKLLISKGAD
LNTSAKDGATPLDMARESGNEEVVKLLEKQ

>4AF30
SELGKRLIRAALDGNKDRVKDLIENGADVNASLMSGTTPLYAAAMNGHKEVVK
LLISKGADVNAQSVAGSTPLVAAANFGHNEVVKLLISKGADVNAVTAFGVTPLH
AAAADGHKEVVKLLISKGADVNAKAGRGMTPLHIAAFRGHKEVVKLLISKGAD
LNTSAKDGATPLDMARESGNEEVVKLLEKQ

>4AF36
SELGKRLIRAALDGNKDRVKDLIENGADVNASLVSGVTPLYAAAMNGHKEVVK
LLISKGADVNAQSVAGSTPLVAAANIGHKEVVNLLISKGADVNAVTAFGVTPLH
AAAADGHKEVVKLLISKGADVNAKAGRGMTPLHIAAFRGHKEVVKLLISKGAD
LNTSAKDGATPLDMARESGNEEVVKLLEKQ

Biolayer Interferometry

Data was collected with an Octet RED96 (FortéBio, Menlo Park, CA, USA) instrument and analyzed with the ForteBio data analysis package. All experiments were performed at room temperature in HBS-EP Buffer (GE Biosciences) with bovine serum albumin (BSA) blocking

agent added (0.01 M HEPES, pH 7.4, 0.15 M NaCl, 3 mM EDTA, 0.005% v/v Surfactant P20, 1% BSA). Dip and Read Streptavidin Biosensors (ForteBio) were activated for 30 minutes in buffer prior to loading with biotinylated Frizzled-5 (at 50 nM). After baseline reference collection, biosensors were dipped in analyte binder solutions to measure association and then returned to the empty buffer-containing baseline well to measure dissociation. Kinetic binding constants were determined after reference subtraction utilizing a 1:1 binding model. Competition assays were carried out by loading the biosensors with biotinylated protein and then sequentially subjecting the biosensors to 250 nM (in HBS-EP Buffer) of protein A and then B (ANK12, B12, Oncomed scFv, or buffer).

Circular Dichroism

ANK12 was dialyzed into PBS (20mM NaPO4, 150mM NaCl) and analyzed at a final concentration of 12 mg/mL. Circular dichroism (CD) spectra were collected on an AVIV Model 420 CD spectrometer (AVIV Biomedical, Inc, Lakewood, NJ, USA) using a 1 mm pathlength quartz cuvette. Scans were collected at 25°C and were taken from 195 to 265 nm in 1 nm steps with 1 nm bandwidth at a scanning speed of 10 nm/minute. Three independent scans were averaged and buffer subtracted against a cuvette holding PBS. Temperature melts were carried out with the same parameters from 25°C to 95°C in 1°C steps reading at 222 nm. *During the melt, a full wavelength scan was taken at 25, 35, 45, 55, 65, 75, 85, and 95°C using the parameters above.*

4.8 Acknowledgements

Yi Miao constructed libraries, selected and characterized Fzd-4/7 specific variants, determined structures.

Keunwan Park collaborated on computational ankyrin binder design.

Lauren Carter was responsible for protein production.

4.7 Figures and Tables

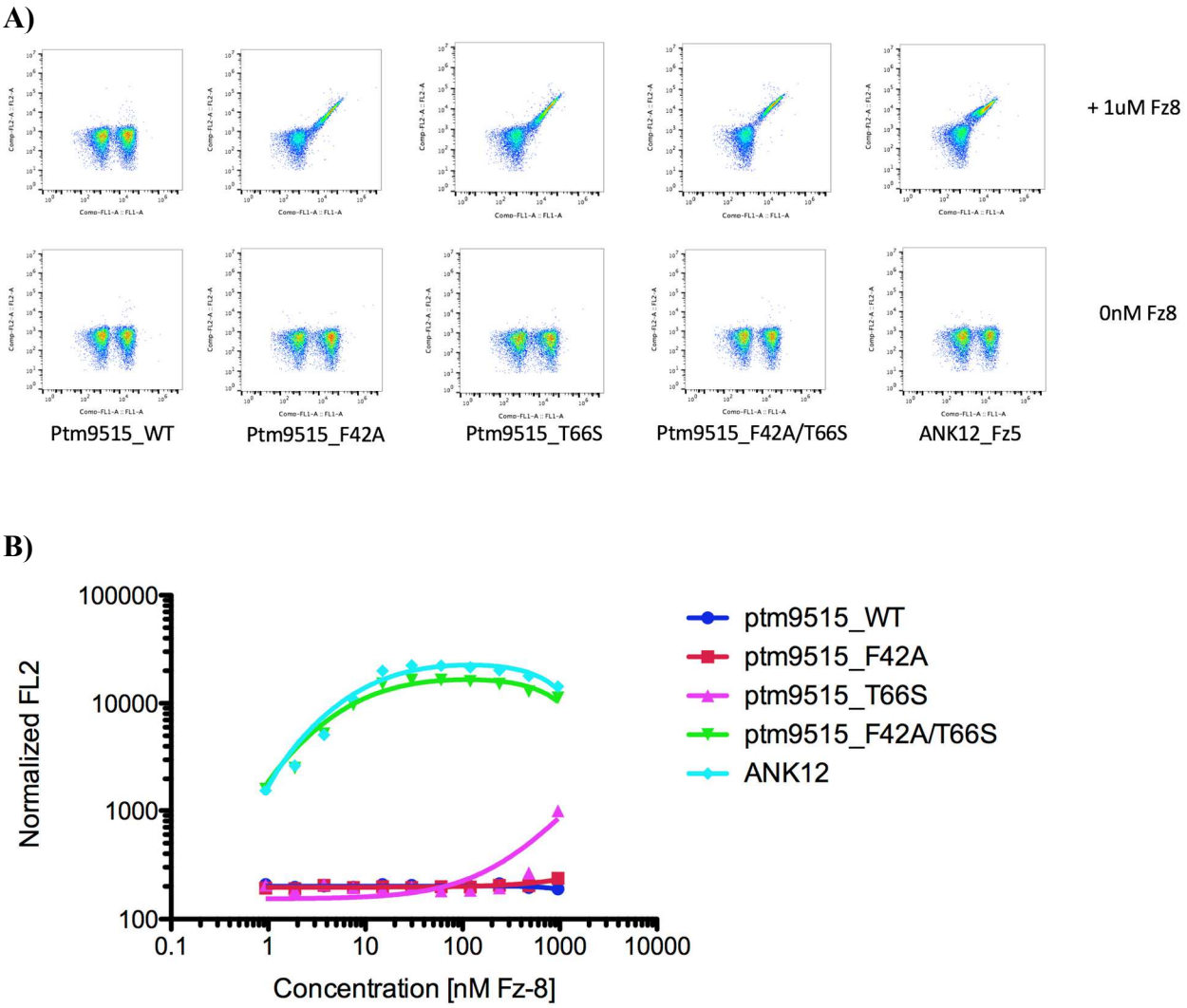
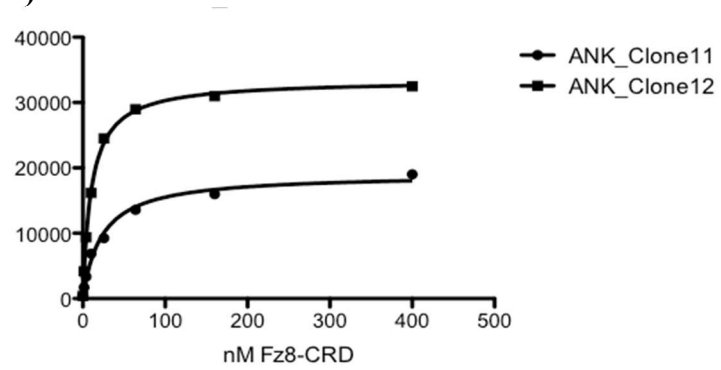


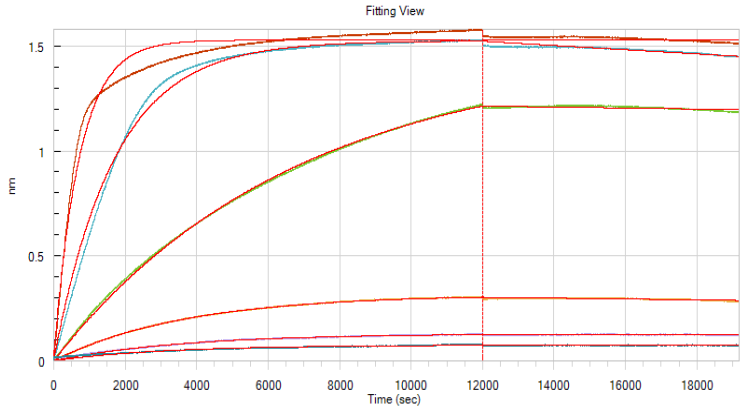
Figure 18. F42A/T66S Point Mutations to Parent Design Are Sufficient for Activity.

A) The F42A and T66S point mutations to the ptm9_5_1_5 design are sufficient to confer weak binding activity, although extended labeling time is required to detect binding, presumably because of the slow on-rate of the interaction. (24 hour labeling with 1 uM Frizzled-8 under avid conditions) B) In contrast, the double mutant F42A/T66S ptm9_5_1_5 variant binds Frizzled-8 with high affinity under standard labeling conditions. (1 hour incubation, non-avid labeling conditions) (ptm9515_F42A/T66A K_D=5.92nM R^2=0.962; ANK12 K_D=7.38nM; R^2=0.956).

A)



B)



C)

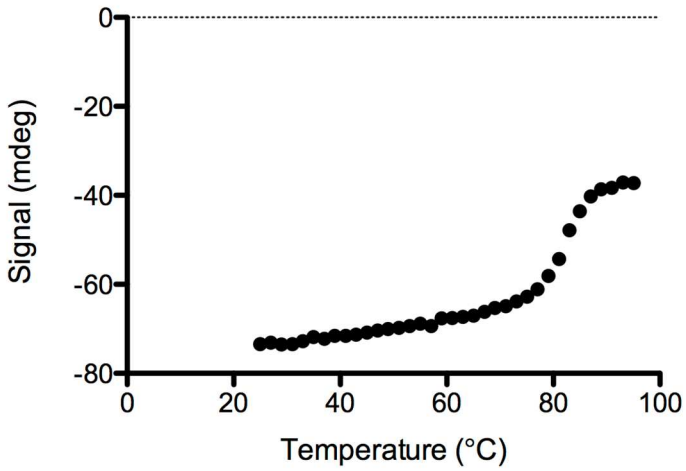
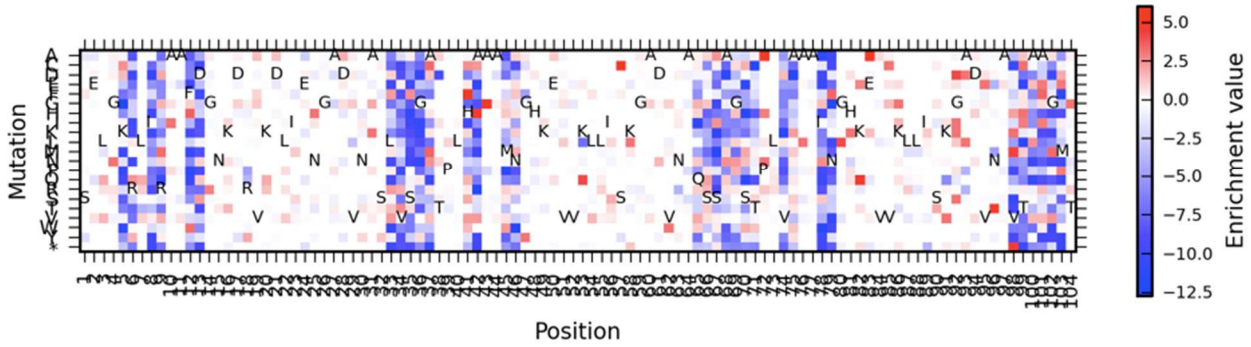


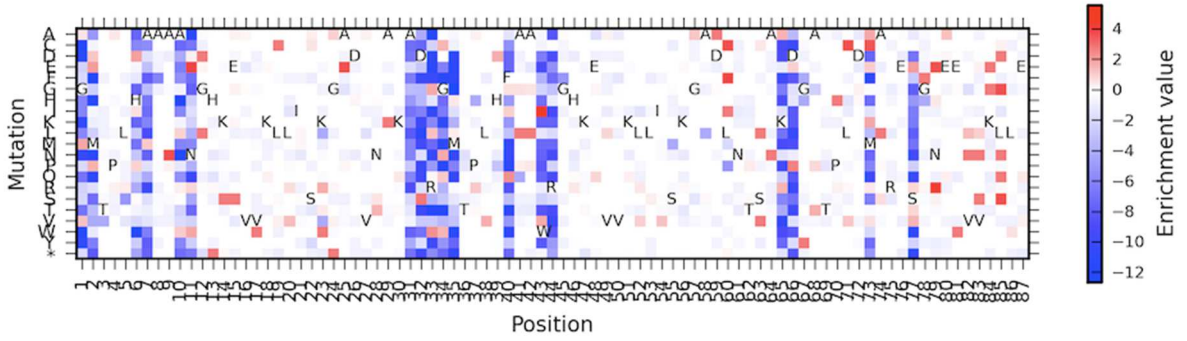
Figure 19. **Thermostable ANK12 Binds to Frizzled-5/8 with High Affinity.**

A) Variants ANK11 and ANK12 from diversified error-prone PCR library bind Frizzled-8 as assayed by yeast surface display. B) Optimized ANK12 binds Frizzled-5 with a (K_D=220pM) as measured by biolayer inferometry. C) Circular dichroism of ANK12 demonstrates inherent thermostability of the ankyrin scaffold. Upper curve represents reversible refolding of melted protein.

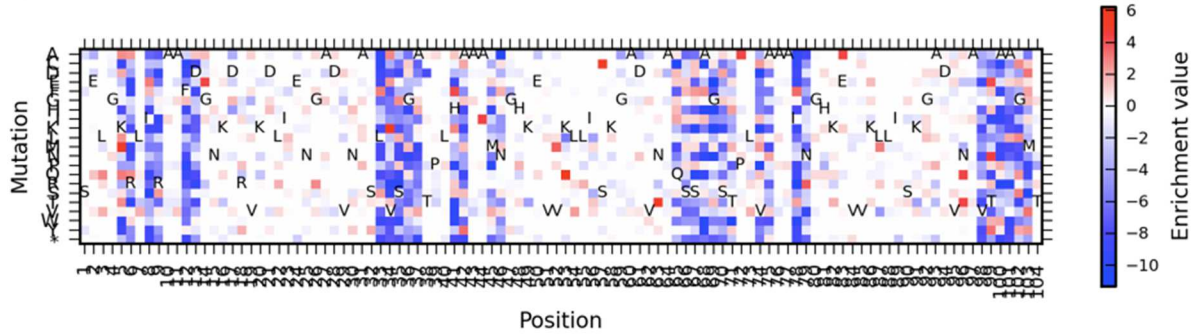
A) Frizzled-4 Front



B) Frizzled-4 Back



C) Frizzled-7 Front



D) Frizzled-7 Back

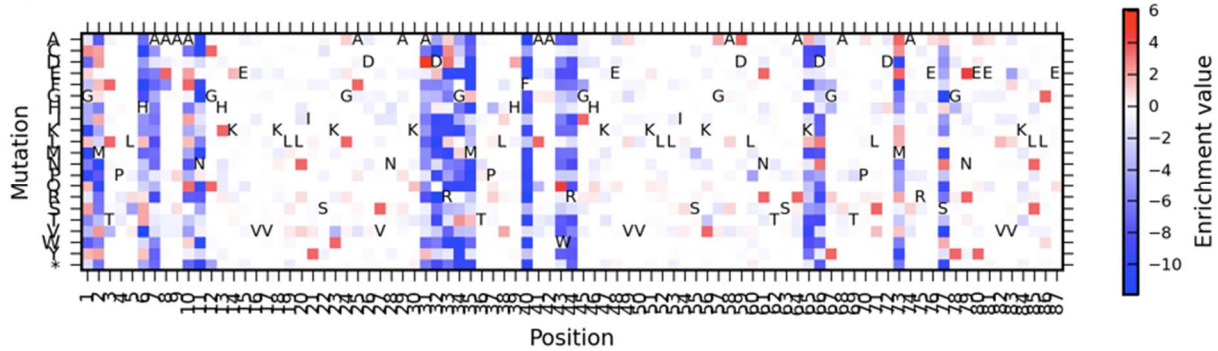
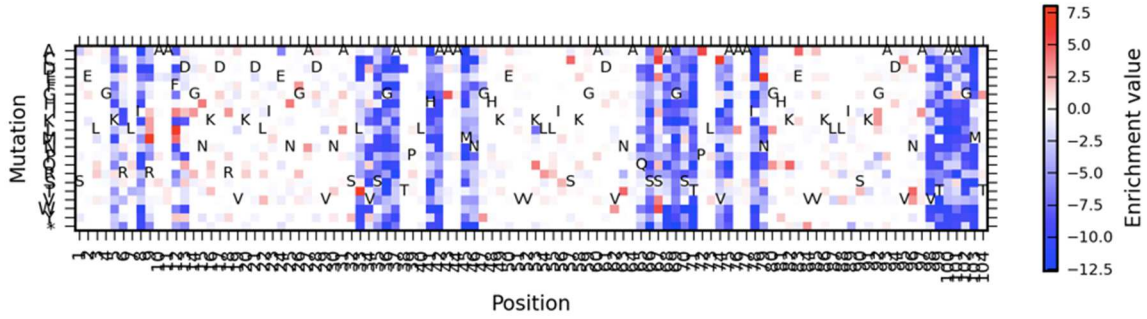


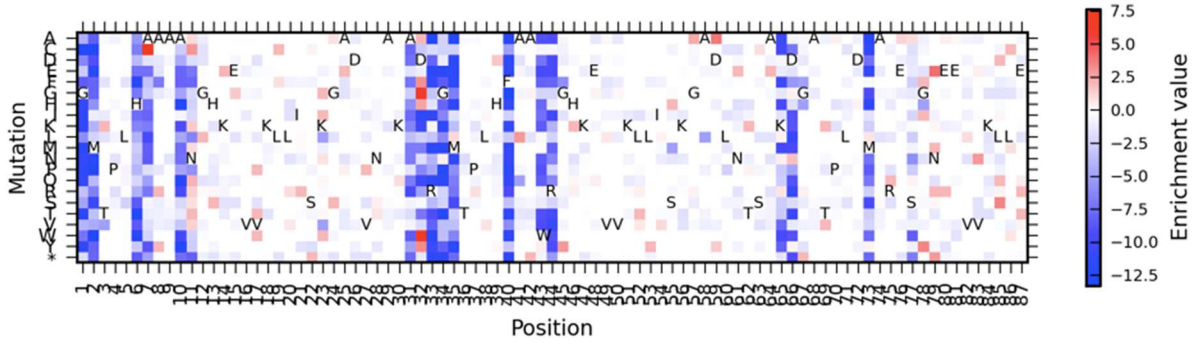
Figure 20. ANK12 Site Saturation Mutagenesis Enrichment Ratios.

Enrichment ratios against Frizzled subtypes 4 (A – Front, B – Back) and 7 (C – Front, D – Back) for ANK12

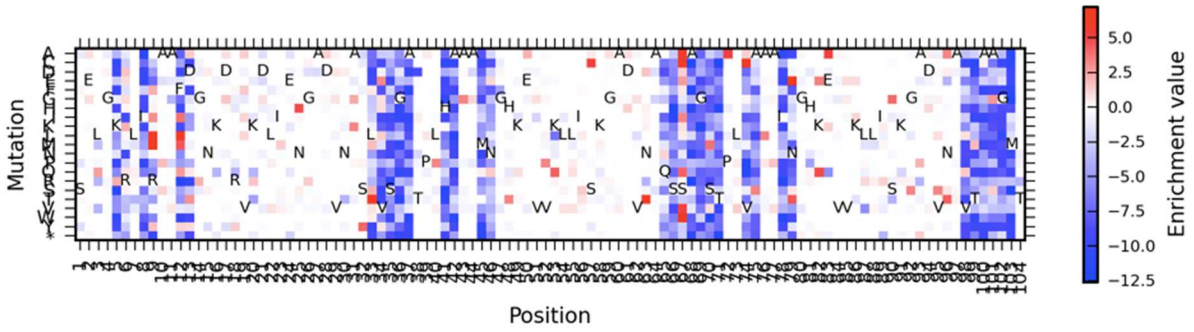
A) Frizzled-5 Front



B) Frizzled-5 Back



C) Frizzled-8 Front



D) Frizzled-8 Back

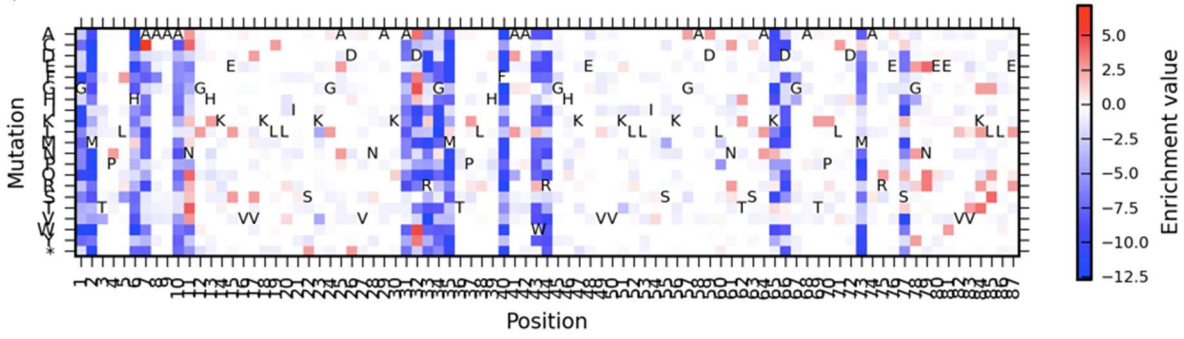
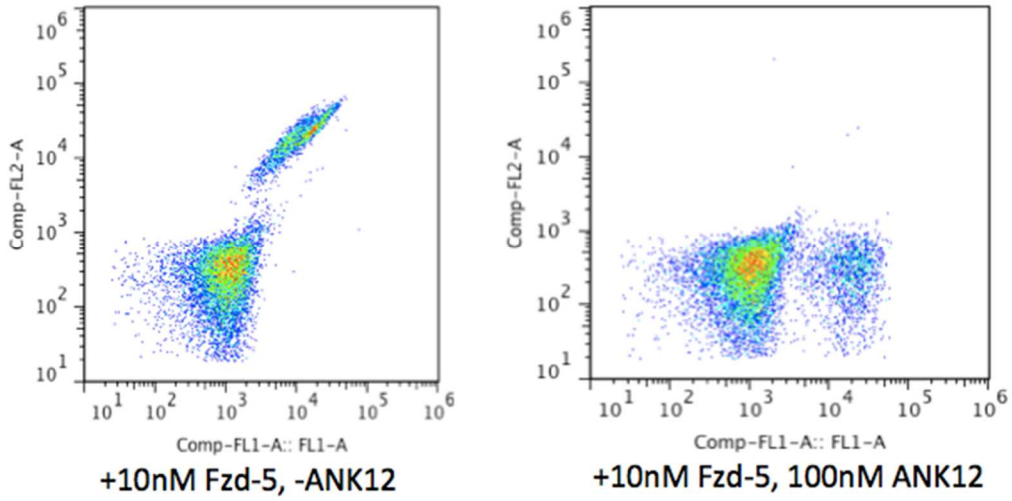


Figure 21. ANK12 Site Saturation Mutagenesis Enrichment Ratios.

Enrichment ratios against Frizzled subtypes 5 (A – Front, B – Back) and 8 (C – Front, D – Back) for ANK12.

A)



B)

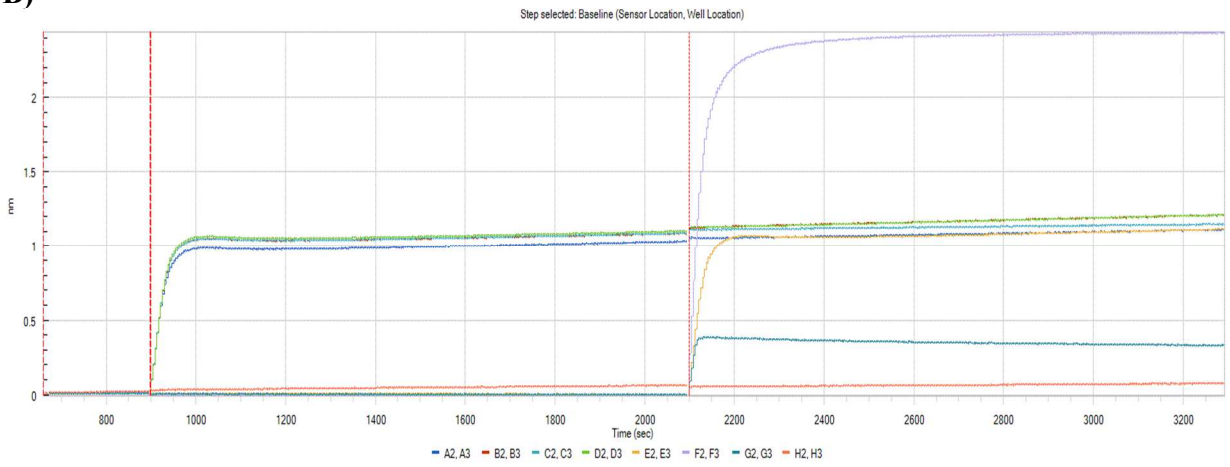


Figure 22. ANK12 Competes with Previously Characterized Binders to Frizzled Lipid Binding Cleft.

A) Incubation of biotinylated Frizzled-5 with ANK12 prior to labeling of yeast displaying the lipid-binding cleft binder B12 resulted in abrogation of binding. B) Biosensors labeled to saturation in vitro with ANK12 do not display additional binding signal when labeled with binders to the targeted functional site (B12, scFv), indicating full competition with the binding modes of both proteins. Conditions are as follows for the two labeling steps (Step 1 is from 840-1740 seconds, Step 2 from 1740-2640): Sensor A2 (ANK12, buffer); Sensor B2 (ANK12, ANK12), Sensor C2 (ANK12, B12); Sensor D2 (ANK12, scFv); Sensor E2 (buffer, ANK12); Sensor F2 (buffer, scFv); Sensor G2 (buffer, B12), Sensor H2 (ANK12, buffer).

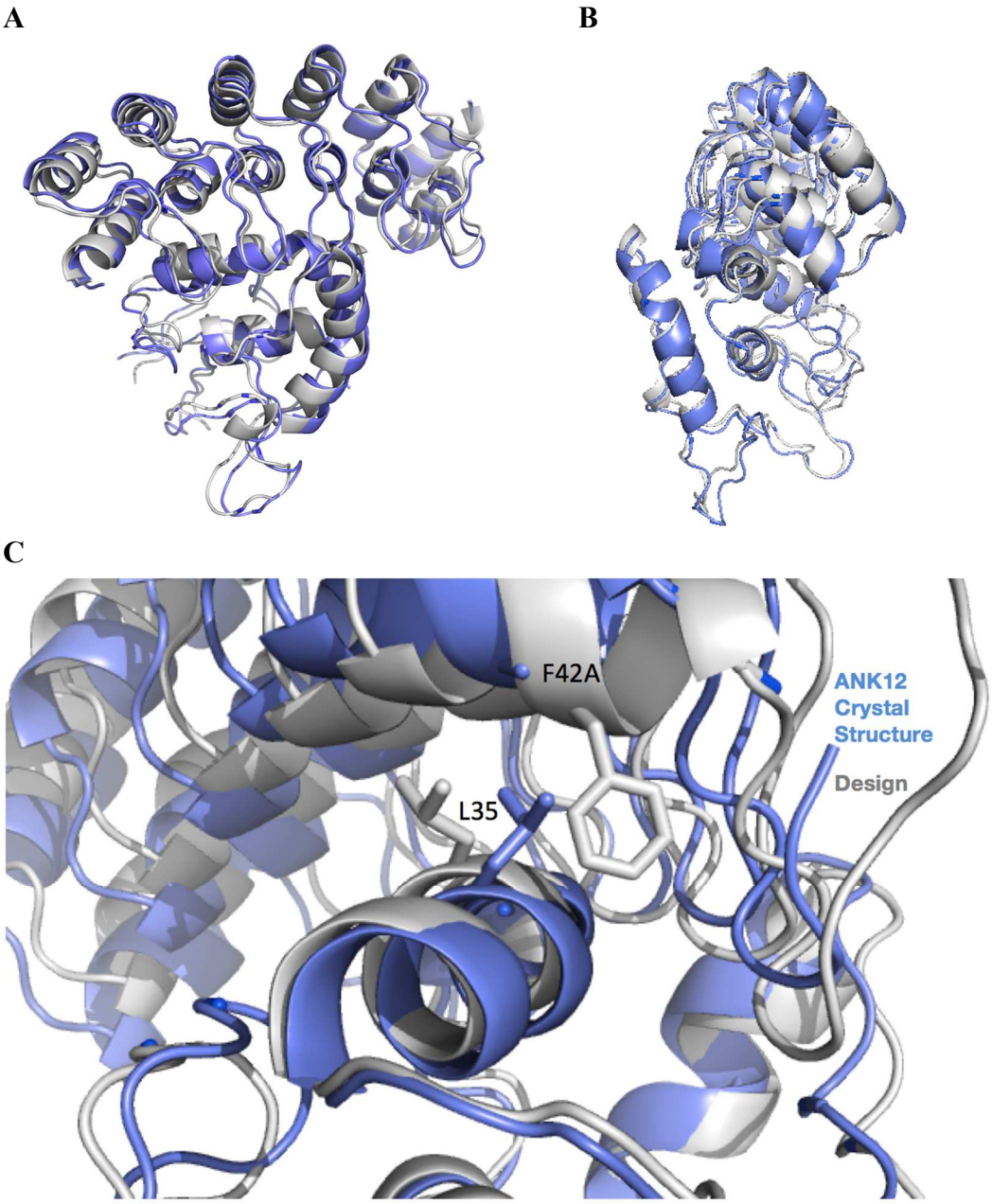


Figure 23. **Crystal Structure Confirms that ANK12 Binds Frizzled-8 as Designed.**

A-B) ANK12 binds Frizzled-8 CRD utilizing rationally selected and computationally designed "fingers-in-groove"; "grasping-hand" binding mode. The backbones of the solved structure (blue) is highly similar to the design (grey) with an RMSD of 1.6 Å. C) Frizzled-8 helix adjacent to lipid-binding cleft shifted in crystal. Residue F42 of the designed binder (blue) would clash with L35 of the receptor in the conformation seen in the crystal structure (grey), explaining the required mutation at this position needed to confer binding activity.

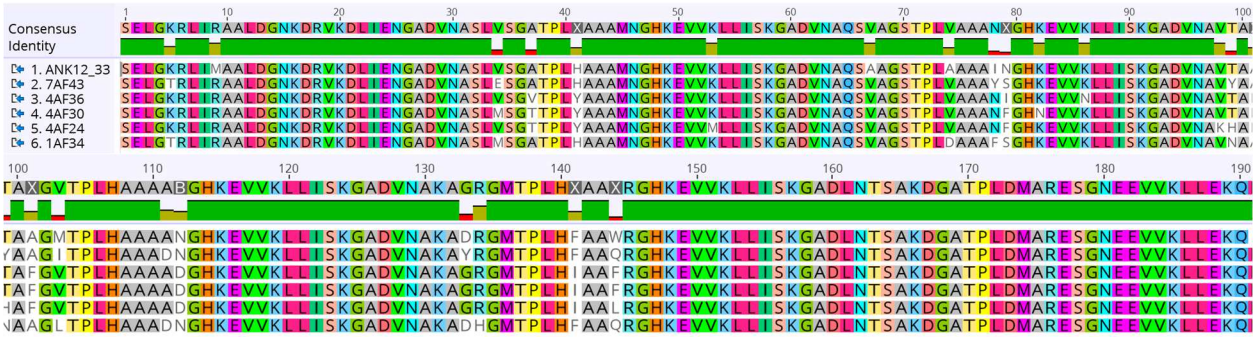
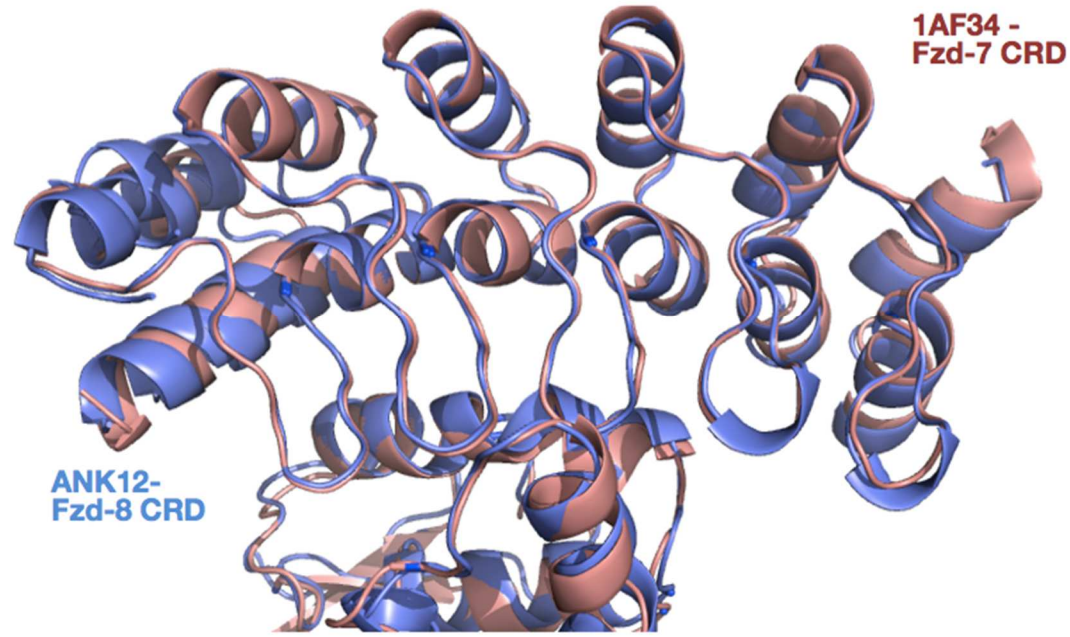


Figure 24. Sequence Alignment of Frizzled-5/8 Specific ANK12 With Variants Evolved to Bind Fzd-7 and Fzd-4.

Sequence alignment of parent design ANK12 and the optimized 1AF34, 7AF43, 4AF24, 4AF30, and 4AF36 variants.

A



B

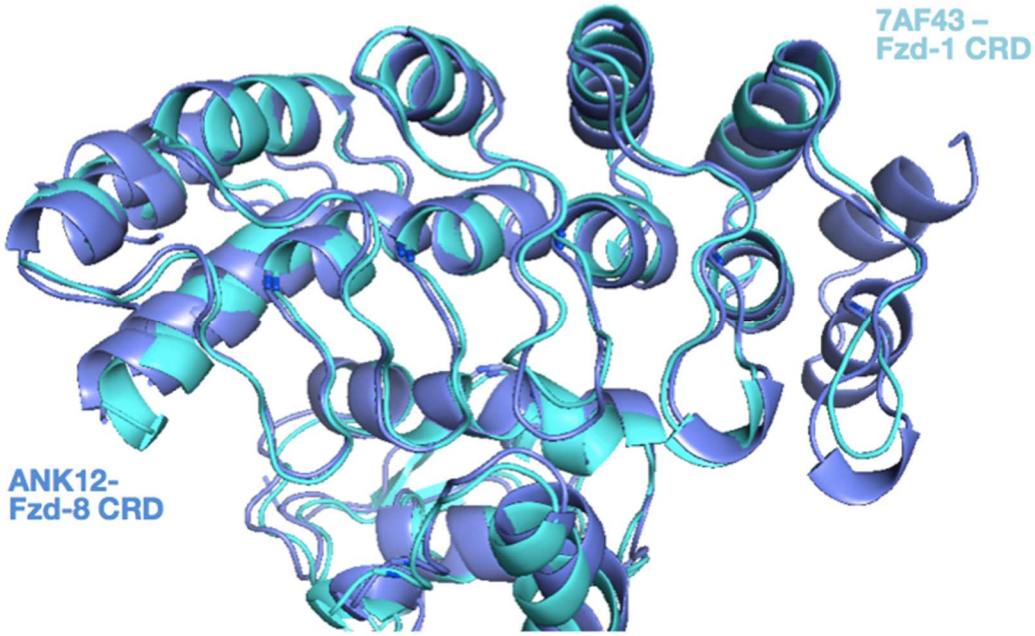


Figure 25. Crystal Structures of 1AF34-Fz7; 7AF43-Fz1 Complexes.

Crystal structures confirms that the redesigned binders 7AF43 (A) and 1AF34 (B) interact with the Frizzled receptor utilizing the same "fingers-in-groove" binding mode as the parent design ANK12.

A

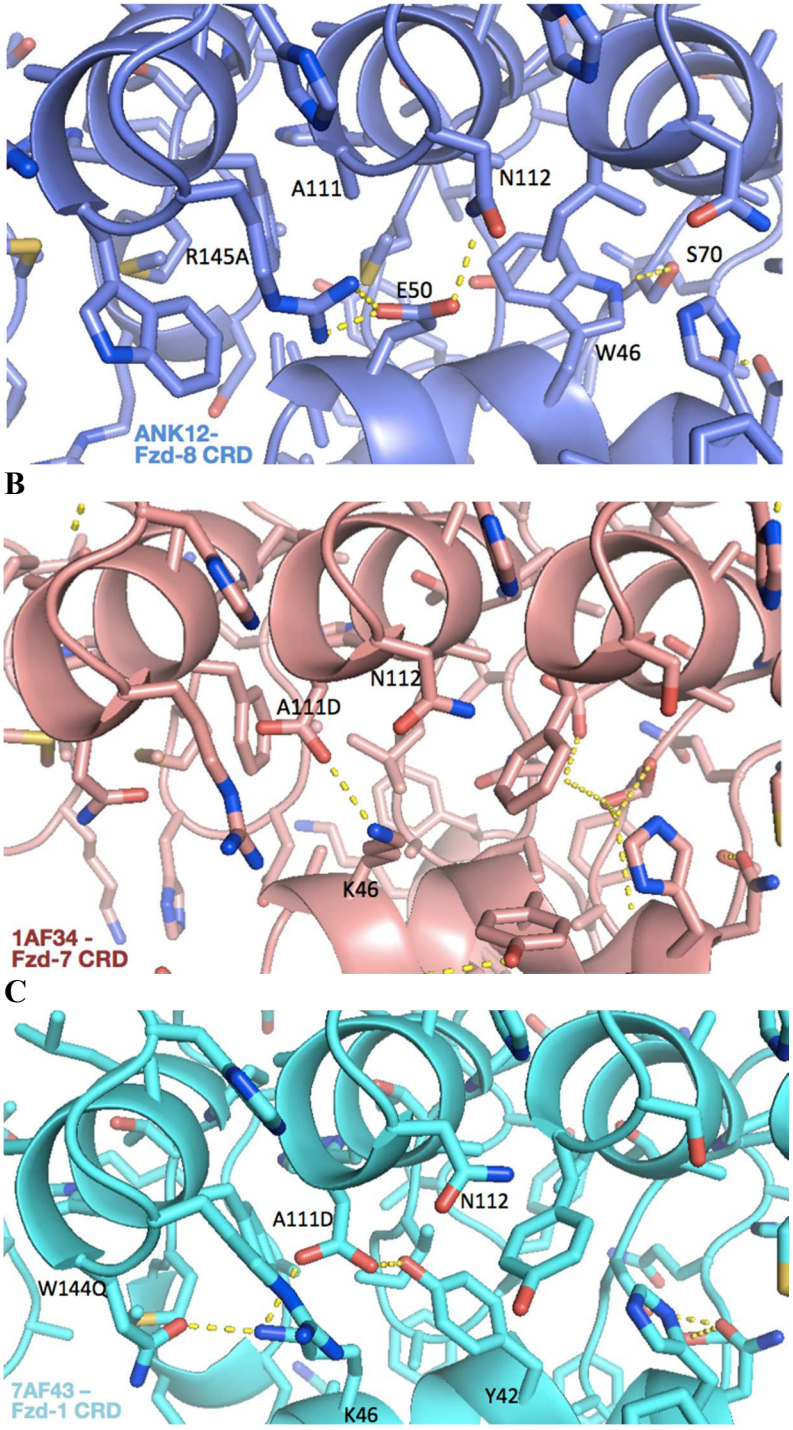


Figure 26. ANK12-Fz8/1AF34-Fz7/7AF43-Fz1 Interface Polar Networks.

A key difference between the Fz-5/8 and Fz-1/7 subtypes is the presence of a glutamate residue on the end of the cleft-adjacent helix in Fz-5/8 (A) and a lysine at the same position in Fz-1/7 (B-C). The A111D mutation is found in both evolved variants against Fz-1/7, which enables appropriate formation of polar networks here via charge-complementary interactions.

Table 4. **Amino Acid Sequences of Designed Ankyrin Frizzled-8 Binders**

Design	Amino Acid Sequence	Interface Size (A^2)	Interface Shape-Complementarity
ptm3_1_03_LD.pdb	SELGIRLIKAAAEGNKDRVKDLIE NGADVNASSDAGMTPLHAAAAA GHKEVVKLLISKGADVNAQANA GSTPLHAAAINGHKEVVKLLISKG ADVNTVSAAGLTPLDLARRAGNE EVVKLLEKQ	2069.19	0.7069
ptm3_11_4_LD.pdb	SELGKRLIEAARVGNKDRVKDLIE NGADVNASDQSGNTPLHAAALN GHKEVVKLLISKGADVNAQAQN GSTPLHFAAYSGHKEVVKLLISKG ADVNTSAAAGDTPLNYARRQGN EEVVKLLEKQ	2123.02	0.6897
ptm3_17_2_LD.pdb	SELGRRLIRAAFQGNKDRVKDLIE NGADVNASSNAGSTPLHAAALQ GHKEVVKLLISKGADVNAQAQA GSTPLHAAAAQGHKEVVKLLISK GADVNTISAAGLTPLDNARQAGN EEVVKLLEKQ	2291.68	0.7148
ptm3_3_05_LD.pdb	SELGERLQRAAMNGNKDRVKDLI ENGADVNAYSESGSTPLHAAATA GHKEVVKLLISKGANVNAQAASG ATPLHAAALQGHKEVVKLLISKG ADVNTSAANGATPLDLARRNGN EEVVKLLEKQ	2292.86	0.7080
ptm3_5_0206_LD.pdb	SELGQRLIRAAFEGNKDRVKDLIE NGADVNASDQAGITPLHAAAAN GHKEVVKLLISKGADVNAQAQY GNTPLDQAAANGHKEVVKLLISK GADVNTVSAAGLTPLNAARQNG NEEVVKLLEKQ	2001.96	0.6613
ptm4_1_3_LD.pdb	SELGKRLIRAAADGNKDRVKDLI ENGADVNASANAGTTPLHNAAM NGHKEVVKLLISKGADVNAQSSA GATPLHAAAASGHKEVVKLLISK GADVNAVNVAGMTPLHTAAMIG HKEVVKLLISKGADVNTVSAAG WTPLDLARRLGNEEVVKLLEKQ	2489.49	0.6977
ptm4_22_4_LD.pdb	SELGKRLIKAARDGNKDRVKDLI ENGADVNASDELGATPLHRAAR NGHKEVVKLLISKGADVNAKSRA	2200.08	0.6847

	GMTPLHFAALKGHKEVVKLLISK GADVNAQSQAGAYPLHAAATSG HKEVVKLLISKGADVNTQTLAGS TPLDAARANGNEEVVKLLEKQ		
ptm4_24_5_LD.pdb	SELGQRLADAARWGNKDRVKDL IENGADVNASDEAGNTPLHAAAL SGHKEVVKLLISKGADVNAQAAI GATPLHAAAISGHKEVVKLLISKG ADVNAQAWAGTTPLVAAAANG HKEVVKLLISKGADVNTSAADGS TPLDMARNQGNEEVVKLLEKQ	2537.8	0.7040
ptm4_3_1_LD.pdb	SELGKRLIRAAFDGNKDRVKDLIE NGADVNASLVSGATPLHAAAMN GHKEVVKLLISKGADVNAQSSAG STPLHFAASAGHKEVVKLLISKGA DVNALSVVGATPLVAAAMNGHK EVVKLLISKGANVNQQSANGSTP LWFARFNGNEEVVKLLEKQ	2746.99	0.6935
ptm4_34_2_3_LD.pdb	SELGKRLIKAAALGNKDRVKDLI ENGADVNASDSLGSTPLHAAAFN GHKEVVKLLISKGADVNAKTRVG ATPLHFAAWLGHKEVVKLLISKG ADVNAKARVGSTPLHAAAANGH KEVVKLLISKGADVNTSTSDGWT PLDLARLLGNEEVVKLLEKQ	2486.75	0.6489
ptm4_4_5_LD.pdb	SELGWRLQRAAAFGNKDRVKDLI ENGADVNASEQAGMTPLHAAAA AGHKEVVKLLISKGADVNANAA AGSTPLHAAAINGHKEVVKLLISK GANVNAQSAAGMTPLHAAAANG HKEVVKLLISKGANVNTSAQAGN TPLDWARRNGNEEVVKLLEKQ	2540.53	0.7202
ptm4_51_2_6_LD.pdb	ggsggsSELGQQLIWAAIAGNKDQV KDLIENGADVNASSNAGNTPLHM AAFNGHKEVVKLLISKGADVNAK SQAGSTPLAFAAFNGHKEVVKLLI SKGADVNAKSSSGSTPLHAAAAN GHKEVVKLLISKGADVNTSDQNG YTPLDMAREIGNEEVVKLLEKQ	2237.56	0.7574
ptm4_5_4_LD.pdb	SELGKRLFLAARDGNKDRVKDLI ENGADVNASAQAGSTPLHMAAL SGHKEVVKLLISKGADVNAKADA GSTPLSAAAINGHKEVVKLLISKG ADVNAQSLSGMTPLHAAAINGHK EVVKLLISKGADVNTFSAAGNTP LNTARAAGNQEVVKLLEKQ	2548.91	0.7246

ptm4_8_6_LD.pdb	SELGQRLARAARDGNKDRVKDLI ENGADVNASSNNGSTPLHAAAA NGHKEVVKLLISKGADVNAQSLI GMTPLHAAAAIGHKEVVKLLISK GADVNAQMFAGNTPLHSAALNG HKEVVKLLISKGADVNTQSASGN TPLNLARLTGNEEVVKLLEKQ	2517.1	0.6894
ptm4_97_2_5_LD.pdb	ggsggsSELGQRLIIAARQGQKDRV KDLIENGADVNASDAAGNTPLHD AAAYGRKEVVKLLISKGADVNA KAKAGTTPLHAAAAAGHKEVVK LLISKGADVNAKISAGLTPLHAAA ANGHKEVVKLLISKGADVNTSAR DGSTPLLWARYSGNEEVVKLLEK Q	2190.18	0.6999
ptm5_33_2_LD.pdb	SELGWRLIRAAQKGNKDRVKDLI ENGADVNASSNAGATPLHYAAA WGHKEVVKLLISKGADVNAKAW SGTTPLAAAAWWGHKEVVKLLIS KGADVNAQATAGFTPLHAAASN GHKEVVKLLISKGADVNAKAAA GVTPLHFAAQNGHKEVVKLLISK GADVNTSAQNGTTPLDWARRNG NEEVVKLLEKQ	2482.68	0.7385
ptm5_39_2_LD.pdb	SELGKRLIKAAWDGNKDRVKDLI ENGADVNASDQMGATPLHFAAQ KGHKEVVKLLISKGADVNAKAA NGSTPLHAAAMNGHKEVVKLLIS KGADVNAQSATGTTPLHMAAAS GHKEVVKLLISKGADVNAQAAA GITPLHLAAANGHKEVVKLLISKG ADVNTSAQIGATPLDMARENGNE EVVKLLEKQ	2356.59	0.6451
ptm5_46_1_LD.pdb	SELGQRLIKAASEGNKDRVKDLIE NGADVNASADSGFTPLHLAAWN GHKEVVKLLISKGADVNAKAAA GATPLHAAAANGHKEVVKLLISK GADVNANNAAGSTPLHLAAFNG HKEVVKLLISKGADVNAQTAAG MTPLHAAAAQGHKEVVKLLISKG ADVNTSAQNGATPLDWARENGN EEVVKLLEKQ	2586.13	0.7163
ptm5_6_1_5_LD.pdb	SELGKRLIEAARRGNKDRVKDLIE NGADVNASQETGMTPLHAAALN GHKEVVKLLISKGADVNAFAQAG STPLHFAAASGHKEVVKLLISKGA	2591.42	0.7180

	NVNAQSAAGATPLHLAAMNGHK EVVKLLISKGANVNAQDVAGMT PLHWAALNGHKEVVKLLISKGAD VNTSAYDGATPLDFARRNGNEEV VKLLEKQ		
ptm5_66_2_3_LD.pdb	SELGKRLIEAARDGNKDRVKDLIE NGADVNASVLAGMTPLHWAAFS GHKEVVKLLISKGADVNAQSQVG MTPLHWAAAAGAKEVVKLLISK GADVNAKAKTGNTPLHFAARNG HKEVVKLLISKGADVNAKAEEGN TPLHDAASFGWKEVVKLLISKGA DVNTSSRKGWTPLDMARENGNE EVVKLLEKQ	2379.86	0.6799
ptm5_7_1_3_LD.pdb	SELGKRLIEAAEKGNKDRVKDLIE NGADVNASNEAGNTPLHAAARD GHKEVVKLLISKGADVNAKSNAG ATPLHAAAASGHKEVVKLLISKG ADVNAKSSAGATPLHLAASKGH KEVVKLLISKGADVNAQASAGST PLHMAAWQGHKEVVKLLISKGA DVNTSAAAGATPLDLARKNGNEE VVKLLEKQ	2675.67	0.7112
ptm5_9_1_5_LD.pdb	SELGKRLIKAAAEGNKDRVKDLI ENGADVNASTEAGATPLHFAAAN GHKEVVKLLISKGADVNAQTIAG STPLVAAAINGHKEVVKLLISKGA DVNAVTAAGMTPLHAAAANGHK EVVKLLISKGADVNAKADRGMTP LHFAAWRGHKEVVKLLISKGAD VNTSAKDGATPLDMARESGNEEV VKLLEKQ	2519.01	0.6941
	AVERAGE SCORES	**2402.98**	**0.7002**

Table 5. Retrospective Analysis of Mutations to Parent ptm5_9_1_5 Design

Position	Parent Design Identity (Fzd8 Enrichment Ratio from SSM)	ANK12 (Error-Prone PCR Variant) Identity	Final Matured/Xtal Variant (Fzd8 Enrichment Ratio from SSM)
9	K (0.664)	R	M (113.548)
12	A (0.004)	F	L (75.939)
13	E (0.656)	D	E (0.656)
33	T (81.821)	L	L
34	E (8.736)	V	V
35	A (0.166)	S	S
42	**F (0)*** **	A	A
45	A (0.002)	M	M
66	**T (0.041)*** **	S	S
67	I (9.877)	S	A (56.35)
161	V (non-interface contacting position not included in SSM)	L	V

Sequence comparison of original ptm5_9_1_5 design to original and matured ANK12 variants. Amino acid identities are listed with ln(enrichment ratio) from SSM data listed in parentheses if available. Positions 33, 34, and 67 are all significantly enriched as the original design amino acid, while positions 9 and 13 are mutated to similar amino acids (K9R, E13D) and are only slightly preferred according to SSM data. V161 is a position distant and irrelevant to the ANK12-Fz8 interface and may be disregarded. Positions 12 and 45 are alanine in the original design. Although the SSM shows depletion here for the original amino acids as the evolved identities are preferred and make good contacts, the minimal alanine residue at these positions is unlikely to create any steric clashes which preclude binding. Therefore, the mutations that are most likely to be essential to the acquisition of binding activity by the ANK12 variant are F42A and T66S.

Table 6. Positions and Amino Acid Identities Included in Combination Library Targeting Frizzled-4.

Position/WT Identity	Codon	Allowed Amino Acid Identities	Position Diversity
F12	YWT	FHLY	4
V34	RBG	ARGMTV	6
A37	AYG	AMTV	4
H41	YWT	HLFY	4
A42	SMT	ADHP	4
A74	GYT	V,A	2
I78	AWT	IN	2
N79	WWT	NIFY	4
V98	RWG	EKMV	4
T99	MHT	NHILPT	6
A101	DYT	AIFSTV	6
M103	RWG	MEKV	4
N112	RAW	NDEK	4
D133	RRT	NDGS	4
F141	WTT	IF	2
W144	WKS	CFILMRSW	8
		Theoretical Library Diversity	1.81E9

Table 7. Positions and Amino Acid Identities Included in Combination Library Targeting Frizzled-7

Position/WT Identity	Codon	Allowed Amino Acid Identities	Position Diversity
K5	ANS	R, N, I, K, M, S, T	8
M9	AKS	R, S, I, M	4
L12	TTW	L, F	2
L33	YTT	L, F	2
V34	RWG	E, K, M, V	4
A67	KYT	A, S, F, V	4
A74	GYT	V, A	2
I78	WHT	N, I, F, S, T, Y	6
N79	ARW	N, R, S, K	4
T99	WMT	N, S, T, Y	4
M103	DTS	I, L, M, F, V	6
A111	KHT	A, D F, S, Y, V	6
A132	GNW	A, D, E, G, V	8
R134	SDT	R, D, G, H, L, V	6
W144	CDG	R, Q, L	3
		Theoretical Library Diversity	2.04E9

Table 8. Comparison of Reengineered Fzd-1/4/7 Binding Variants to Fz-5/8 Specific ANK12.

Position	ANK12	1AF34	7AF43	4AF24	4AF30	4AF36
5	K	**T****	**T****	K	K	K
9	M	**R****	**R****	**R****	**R****	**R****
34	V	**M***	**E***	V	**M**	V
37	A	A	A	**T***	**T***	**V***
41	H	H	H	**Y****	**Y****	**Y****
53	K	K	K	**M**	K	K
67	A	**V***	**V***	**V****	**V****	**V****
74	A	**D***	**V***	**V****	**V****	**V****
78	I	**F***	**Y***	**N****	**N****	**N****
79	N	**S****	**S****	**F***	**F***	**I***
83	K	K	K	K	**N**	K
98	V	V	V	**K**	V	V
99	T	**N***	**Y***	**H**	T	T
101	A	A	A	**F****	**F****	**F****
103	M	**L***	***I**	**V****	**V****	**V****
111	A	**D****	**D****	A	A	A
112	N	N	N	**D****	**D****	**D****
133	D	D	**Y**	**G****	**G****	**G****
134	R	**H**	R	R	R	R
141	F	F	F	I	I	I
144	W	**Q****	**Q****	**F***	**F***	**L***
# Mutations	0	**12**	11	16	14	13

The large number of mutations between the parent binder and derived variant reflect the comprehensive, simultaneous engineering approach taken in this case, and the ability of the ankyrin scaffold to tolerate extensive mutagenesis without destabilization. Mutations at positions 53 and 83 were spontaneous. Differences from the original sequence are noted with an asterix, positions where either the 1AF34/7AF34 or 4AF24/30/36 variants are mutated to the same amino acid are noted with two asterixes. The A111D mutation appears particularly critical to confer binding activity towards Fzd-1/7 as it corresponds to a glutamate (position 50) to lysine charge reversal mutation on the Frizzled receptor adjacent to this position, and these mutations restore appropriate polar networks. The W144Q mutation is involved in this network as well in the

7AF43-Fzd-1 structure. This position is glutamine in the Fz-4 subtype, enabling the N112D mutations found in the Fz-4 variants, which would likely be suboptimal for Fz-5/8 binding.

Other notable mutations in the Fz-4 variants include a H41Y mutation adjacent to a Fzd-5/8 glutamine 44 which is mutated to a corresponding threonine in Fz-4 ,a N79F/I mutation adjacent to Fz-5/8 histidine 43 which is mutated to threonine as well in Fz-4, and a A101F mutation near Leu94 to valine and Met95 to leucine sequence alterations in Fz-4. These positions represent sites where larger amino acids are likely able to pack in the Fz-4 interface compared to the Fz-5/8 interface due to differences in amino acid composition. Also, I51 is mutated to tyrosine in Fz-4 and the D133G mutation is likely required to relieve clashes at this position.

Conclusion

This text summarizes computational design and experimental optimization efforts to engineer proteins which bind the Fz lipid binding cleft for modulation of Wnt signaling. As described in detail in the preceding sections, Wnt signaling is essential to the basic biological processes of development, tissue homeostasis and cell proliferation. A full understanding of Wnt signaling is required to enable precise intervention in this pathway through applications in human health and disease. Native Wnt ligands are limited by solubility considerations given their essential lipid modification, therefore water-soluble protein equivalents are needed to enable robust and controllable agonism of this pathway for scientific investigation of this pathway as well as regenerative medicine applications. Additionally, the functional relevance of the interactions between various Wnt ligands and Fz receptors and the degree to which these ligand-receptor pairs are degenerate or specific remains largely unstudied, again due to the limited availability of receptor-level antagonists. We therefore sought to simultaneously address two significant shortcomings in the field of Wnt signaling: the need for water-soluble protein-based Wnt equivalents to enable agonism and the need for subtype-specific, discriminating antagonists.

To this end, we applied recently developed rational computational design methods to this problem to address the existing needs in the field. We were ultimately successful in developing two distinct high-affinity binding proteins, B12 and ANK12, which specifically bind Fz-5/8, but not other receptor subtypes. Both of these proteins are potent Wnt antagonists, and therefore represent novel inhibitors that may be used to specifically intervene in Fz-5/8 specific Wnt signaling with limited off-target effects. Indeed, preliminary data shows the effectiveness of

these reagents in inducing Wnt inhibition in both Fz-5 dependent RNF43 mutant PDAC cells as well as Fz-8 expressing NSCLC cells (A549). The development of Fzd-4 and Fzd-1/7 binding proteins utilizing the ANK12 platform similarly enables the antagonism of these receptor subtypes in relevant contexts. Therefore, we were successful in developing subtype-specific Fz binders for use as Wnt inhibitors, moreover these antagonists are functionally useful in a range of relevant contexts.

As detailed in Section 2, collaborators subsequently functionalized B12 via fusion to the LRP6-binding Wnt inhibitor Dkk1 in order to create a bi-specific, water-soluble Wnt surrogate. The activity of these surrogates in signaling assays provided validation that colocation of the Fz and LRP cell surface receptors is sufficient to initiate Wnt signaling, confirming the hypothesized mechanism of activity. In addition, the surrogates activated signaling based upon the receptor expression of a given cell type, demonstrating that signaling activity could be precisely controlled based upon binding specificity. Further work by collaborators showed the utility of water-soluble Wnt surrogates both in vivo and in the culture of organoid tissues, providing a rationale for the utilization of Fz binding proteins for use in regenerative medicine applications. Although not detailed here, the continued development of ANK12 and related binders and subsequent inclusion in such bi-specific molecules represents additional progress towards Wnt surrogates with practical value in such contexts. Therefore, the protein engineering efforts described here in conjunction with the functionalization undertaken by collaborators were successful in producing alternative water-soluble Wnt ligands to address a long-standing shortcoming in the field. The ongoing development of such Wnt surrogates has great potential for the utilization of the controlled activation of this critical pathway in regenerative medicine.

As summarized above, this work represents significant progress towards two long-standing problems in the Wnt signaling field: lack of sufficiently specific antagonists and robust, water-soluble agonists. These efforts were carried out utilizing computational design, methods which have relatively only recently been available and are still undergoing significant refinement. Although this work was not explicitly directed towards methods development, as discussed in the introduction, protein engineering is inherently an iterative process. Lessons learned from the failures and successes of application to a specific problem are processed and reintegrated into standard methods for the next application. The work here yielded several significant developments from a computational design methods perspective.

The failure of Fz27 to bind as designed was attributed to the substandard stability of the parent scaffold, as functionalizing mutations destabilized the fold of the protein and caused critical regions to adopt disordered conformations. This was addressed in the utilization of hyperstable scaffolds (small, computationally designed de novo proteins as well as leucine rich repeats and ankyrin repeat proteins) that could more reasonably expected to fold and retain favorable physical properties after design. Use of such proteins can be considered prudent best practices in protein engineering, because as seen in the case of B12, even a functional binder may require significant reengineering afterwards if the scaffold itself has substandard behavior. If the ultimate goal is to develop proteins with real-world, commercial impact, then scaffold selection must maximize not only the success rate during the design process (which is increased by the starting stability of the parent protein) but also the chances of success in downstream, demanding applications, such as in vivo clinical trials and therapeutic usage. This insight was validated by the successful use of ankyrin scaffolds to precisely design a de novo interface against the targeted Fz binding cleft. The ankyrin parent scaffold had enough reserve stability

that even substantial mutagenesis of the interface did not disrupt the backbone of the fold, removing a significant potential cause of failure during design. Additionally, the protein was stable enough after functionalization to undergo additional development for specificity redesign. The development of ANK12, therefore represents a proof-of-concept for utilizing hyperstable proteins for computational design approaches, and in particular argues that the ankyrin fold may be consider a privileged scaffold to use as a generalized platform for computationally directed binder development.

This work provided several opportunities to demonstrate the utility of computational methods in iterative improvement of existing proteins. The engineering of wholly de novo, disulfide-stabilized loop topologies to confer robust Wnt antagonism can be considered an example of the value of computational design methods, which enable the specific and rational construction of backbone structure to fulfill functional requirements. Computational design of loops of specified structure may be considered a general approach that may be applied elsewhere. The stabilization of B12 through consensus loop design, to preferentially stabilize the desired conformation over an alternate one, is another example of how computational methods can efficiently achieve protein engineering objectives. In combination with the rationally identified tyrosine-to-phenylalanine mutation and helix truncation, this also is a generalizable approach to stabilization of existing proteins by integrating existing knowledge of current protein structure and computational sequence optimization methods.

The successful reengineering of the ANK12 interface to alter subtype specificity is another approach that may be further applied both in this case (to develop additional binders against other Fz receptor subtypes) and more broadly. This technique requires that the binding footprint of the interaction includes non-conserved residues, but it was aided by the fact that a

large amount of the interface was conserved. The existing favorable energy of binding (from interactions with the conserved regions) drives the association of the receptor-binder complex, enabling engineering of contacts which were compatible with one receptor subtype but not others to result in discriminative binding of a particular subtype. The rational engineering of binding affinity through amino acid level enrichment data has been used effectively before, but in this case the larger insight is that design of an appropriately targeted interface is what enables such redesign to be possible. Precise design of such an interface as required in this case is only achievable through rational computational design.

Ultimately, the most generalizable and significant method developed by this work is the de novo design of a rationally specified interface. This success demonstrates that although de novo interface design remains difficult, it is not impossible with current methods and will continue to become more feasible with additional development. Given the increasing importance of engineered proteins as biotherapeutics and in other applications as well, improvement of these methods is of critical importance. The application of computational resources to protein folding and design has already yielded significant dividends, and is currently changing existing paradigms in protein engineering. These computational methods have capabilities previously inaccessible to empiric protein engineering approaches, and efficient and effective application of these methods should enable the conception and execution of novel technologies and treatments to improve human health and quality of life. These approaches should also lower the cost of such proteins and increase the rate of development, important considerations in both the rapid treatment of new diseases and the availability of new technologies to the world's broader population. Therefore, although the successful development of novel proteins that specifically address significant problems in the Wnt signaling field is an important result, in the broader

perspective this work is merely a step in the ongoing movement to change the way in which proteins are engineered and by extension, the way diseases are treated. It is my sincere hope that the contributions of this work to both the fields of Wnt signaling and protein design will have an impact in this regard.

Appendix 1. Idealization of B12: A De Novo Approach to Binder Stabilization

Section three describes in detail the stabilization of B12, a process that was ultimately successful. However, alternate approaches were also attempted, which are described briefly here for completeness.

One approach to stabilization of a suboptimal binder is to address the specific limitations of the protein on an individual, case-by-case basis. This approach is described fully in Section 3 of this text. Another more radical method is to reconstitute the protein's functionality from the ground up. Recapitulation of the B12 binding mode would require the design of helical secondary structures with similar spacing to that observed in the B12 structure (rather than the traditional grafting of side-chain hot-spots). An idealization protocol developed by Daniel-Adriano Silva was used to generate secondary structures that duplicated the key structural elements of the design, and then the elements were parsimoniously reconnected. The protocol then determines an ideal sequence for the desired structure using a frequency database derived from the PDB. Candidate designs are forward-folded in silico to confirm that the sequence folds as designed, and unique backbones are chosen based on the quality of the folding funnel.

Four parent designs were tested with and without one or more engineered disulfides which stapled the core secondary structure elements, for a total of 16 designs. All bound Fz8-CRD via yeast surface display, albeit at much lower affinity than the original B12 design (Kd ~ 100s of nM). It was expected that radical redesign of the initial optimized design by the idealization approach may result in proteins that have improved stability, but somewhat lowered affinity. This is because the backbones and cores of these proteins should be ideal, but all of the interface-related interactions may not be fully recapitulated. For this reason, these de novo proteins are expected to have significantly improved properties, but additional affinity maturation may be required. In this case, the significant decrease in affinity of the idealized

binders meant that they were not fully characterized due to time limitations. However, the fact that full de novo proteins built based on the original parent scaffold retained functional activity demonstrates that this approach has great value, and that the ideal design of proteins with novel topologies and connectivities based on known secondary structure is feasible. This test case in particular demonstrates the value of this approach, since the binding interactions of B12 are not side-chain motifs, but rather the backbone conformation formed by the two interface helices. The success in conferring binding activity here then is more remarkable then a simple side-chain graft, since deviation from the appropriate secondary structure element backbone orientation would preclude binding.

Another alternative approach that was used for stabilization of the B12 protein relates to the consensus sequence design of the helical connecting loop. Ultimately, the final protein was stabilized by simple sequence optimization of the wild-type structure. However, a method in which the entire loop was removed, ideal loops were built to reconnect the two helical endpoints, and then these de novo backbones were sequence designed was also attempted. These loops were more stable than the original design, but were less stable than the sequence optimized wild-type loop, therefore the designs from this computational approach were not ultimately included in the final, optimized B12. However, this method was successful in building new loop connectivities that did improve upon the stability of the original design, demonstrating again the value of computational design methods in the targeted redesign of existing proteins.

The success of these alternate computational methods also shows that there are typically multiple independent pathways that can be taken to achieve a given protein engineering objective. Additional approaches to increasing stability that could have been taken include engineering of disulfides into the design and selection of libraries at higher temperatures or with

proteases in order to discover stability enhancing mutations. In this case, a more rational, piecemeal approach to stabilization was ultimately taken. However, it can be seen that given a certain protein engineering objective, it is prudent to attempt several parallel approaches, since it is unknown which will yield the best results.

Acknowledgements

Daniel-Adriano Silva generated the computational idealization protocol referenced in this section and collaborated on execution of that method for the idealized B12 designs.

Idealized helix-connecting loops were generated using a method developed by TJ Brunette.

Tom Linsky developed the ConsensusLoopDesign used to sequence optimize these helix-connecting loops.

References

[1] Moon RT1, Brown JD, Yang-Snyder JA, Miller JR. Structurally related receptors and antagonists compete for secreted Wnt ligands. *Cell.* 88(6):725-8 (1997).

[2] Clevers H, Loh KM, Nusse R. An integral program for tissue renewal and regeneration: Wnt signaling and stem cell control. *Science.* 346(6205):1248012 (2014).

[3] Niehrs C, Acebron SP. Mitotic and mitogenic Wnt signaling. *EMBO J.* 31(12):2705-13 (2012).

[4] Polakis, P. Wnt Signaling in Cancer. *Cold Spring Harb Perspect Biol*. 4(5):a008052 (2012).

[5] Takada R1, Satomi Y, Kurata T, Ueno N, Norioka S, Kondoh H, Takao T, Takada S. Monounsaturated fatty acid modification of Wnt protein: its role in Wnt secretion. *Dev Cell.* 11(6):791-801 (2006).

[6] Gurney A, Axelrod F, Bond CJ, Cain J, Chartier C, Donigan L, Fischer M, Chaudhari A, Ji M, Kapoun AM, Lam A, Lazetic S, Ma S, Mitra S, Park IK, Pickell K,Sato A, Satyal S, Stroud M, Tran H, Yen WC, Lewicki J, Hoey T. Wnt pathway inhibition via the targeting of Frizzled receptors results in decreased growth and tumorigenicity of human tumors. *Proc Natl Acad Sci U S A.* 109(29):11717-22 (2012).

[7] Le PN1, McDermott JD1, Jimeno A2. Targeting the Wnt pathway in human cancers: Therapeutic targeting with a focus on OMP-54F28. *Pharmacol Ther.* pii:S0163-7258(14)00162-4 (2014).

[8] Barker N1, Clevers H. Mining the Wnt pathway for cancer therapeutics. *Nat Rev Drug Discov.* 5(12):997-1014 (2006).

[9] Kikuchi A1, Yamamoto H, Kishida S. Multiplicity of the interactions of Wnt proteins and their receptors. *Cell Signal.* 19(4):659-71 (2007).

[10] Purvis AR, Gross J, Dang LT, Huang RH, Kapadia M, Townsend RR, Sadler JE. Two Cys residues essential for von Willebrand factor multimer assembly in the Golgi. *Proc Natl Acad Sci U S A.* 104(40):15647-52 (Oct 2007).

[11] Dang LT, Purvis AR, Huang RH, Westfield LA, Sadler JE. Phylogenetic and functional analysis of histidine residues essential for pH-dependent multimerization of von Willebrand factor. *J Biol Chem.* 286(29):25763-9 (Jul 2011).

[12] Sadler JE. Biochemistry and genetics of von Willebrand factor. *Annu Rev Biochem.* 67:395-424 (1998).

[13] Huang RH, Wang Y, Roth R, Yu X, Purvis AR, Heuser JE, Egelman EH, Sadler JE. Assembly of Weibel-Palade body-like tubules from N-terminal domains of von Willebrand factor. *Proc Natl Acad Sci U S A.* 105(2):482-7 (Jan 2008).

[14] Evans M, Schumm-Draeger PM, Vora J, King AB. A review of modern insulin analogue pharmacokinetic and pharmacodynamic profiles in type 2 diabetes: improvements and limitations. *Diabetes Obes Metab.* 13(8): 677–684 (Aug 2011).

[15] Hopkins AL, Groom CR, Alex A. Ligand efficiency: a useful metric for lead selection. *Drug Discovery Today* 9, 10 430-431 (May 2004).

[16] Abad-Zapatero C, Metz JT. Ligand efficiency indices as guideposts for drug discovery. *Drug Discovery Today* 10, 7 464-469 (April 2005).

[17] Kuntz ID, Chen K, Sharp KA, Kollman PA. The maximal affinity of ligands. *Proc Natl Acad Sci* 96, 18 9997-10002 (August 1999).

[18] Imming P, Sinning C, Meyer A. Drugs, their targets and the nature and number of drug targets. *Nature Reviews Drug Discovery* **5**, 821-834 (October 2006)

[19] Hopkins AL, Groom CR. The druggable genome. *Nature Reviews Drug Discovery* 1, 727-730 (September 2002).

[20] Weiner LM, Surana R, Wang S. Monoclonal antibodies: versatile platforms for cancer immunotherapy. *Nature Reviews Immunology* **10**, 317-327 (May 2010).

[21]Chan AC, Carter PJ. Therapeutic antibodies for autoimmunity and inflammation. *Nature Reviews Immunology* 10, 301-316 (May 2010).

[22] Sheehan J, Marasco WA. Phage and Yeast Display. *Microbiol Spectr.* 3(1):AID-0028-2014. (Feb 2014).

[23] Pepper LR, Cho YK, Boder ET, Shusta EV. A decade of yeast surface display technology: Where are we now? *Comb Chem High Throughput Screen.*11(2): 127–134 (Feb 2008).

[24] Cherf GM, Cochran JR. Applications of yeast surface display for protein engineering. *Methods Mol Biol.* 1319: 155–175 (2015).

[25] Röthlisberger D, Khersonsky O, Wollacott AM, Jiang L, DeChancie J, Betker J, Gallaher JL, Althoff EA, Zanghellini A, Dym O, Albeck S, Houk KN, Tawfik DS, Baker D. Kemp elimination catalysts by computational enzyme design. *Natur**e***. 453(7192):190-5 (May 2008).

[26] Bjelic S, Nivón LG, Çelebi-Ölçüm N, Kiss G, Rosewall CF, Lovick HM, Ingalls EL, Gallaher JL, Seetharaman J, Lew S, Montelione GT, Hunt JF, Michael FE, Houk KN, Baker D. Computational design of enone-binding proteins with catalytic activity for the Morita-Baylis-Hillman reaction. *ACS Chem Biol*. 8(4):749-57 (Apr 2013).

[27] Richter F, Blomberg R, Khare SD, Kiss G, Kuzin AP, Smith AJ, Gallaher J, Pianowski Z, Helgeson RC, Grjasnow A, Xiao R, Seetharaman J, Su M, Vorobiev S, Lew S, Forouhar F, Kornhaber GJ, Hunt JF, Montelione GT, Tong L, Houk KN, Hilvert D, Baker D. Computational design of catalytic dyads and oxyanion holes for ester hydrolysis. *J Am Chem Soc.* 134(39):16197-206 (Oct 2012).

[28] Siegel JB, Zanghellini A, Lovick HM, Kiss G, Lambert AR, St Clair JL, Gallaher JL, Hilvert D, Gelb MH, Stoddard BL, Houk KN, Michael FE, Baker D. Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction. *Science*. 16;329(5989):309-13 (Jul 2010).

[29] Jiang L, Althoff EA, Clemente FR, Doyle L, Röthlisberger D, Zanghellini A, Gallaher JL, Betker JL, Tanaka F, Barbas CF 3rd, Hilvert D, Houk KN, Stoddard BL, Baker D. De novo computational design of retro-aldol enzymes. *Science*. 7;319(5868):1387-91 (Mar 2008).

[30] Khare SD, Kipnis Y, Greisen P Jr, Takeuchi R, Ashani Y, Goldsmith M, Song Y, Gallaher JL, Silman I, Leader H, Sussman JL, Stoddard BL, Tawfik DS, Baker D. Computational redesign of a mononuclear zinc metalloenzyme for organophosphate hydrolysis. *Nat Chem Biol*. 8(3):294-300 (Feb 2012).

[31] Cherny I, Greisen P Jr, Ashani Y, Khare SD, Oberdorfer G, Leader H, Baker D, Tawfik DS. Engineering V-type nerve agents detoxifying enzymes using computationally focused libraries. *ACS Chem Biol.* 15;8(11):2394-403 (Nov 2013).

[32] Wolf C, Siegel JB, Tinberg C, Camarca A, Gianfrani C, Paski S, Guan R, Montelione GT, Baker D, Pultz IS. Engineering of Kuma030: a gliadin peptidase that rapidly degrades immunogenic gliadin peptides in gastric conditions. *J Am Chem Soc*. 137(40):13106-13 (Oct 2015).

[33] King NP, Sheffler W, Sawaya MR, Vollmar BS, Sumida JP, André I, Gonen T, Yeates TO, Baker D. Computational design of self-assembling protein nanomaterials with atomic level accuracy. *Science*. 336(6085):1171-4 (Jun 2012).

[34] King NP, Bale JB, Sheffler W, McNamara DE, Gonen S, Gonen T, Yeates TO, Baker D. Accurate design of co-assembling multi-component protein nanomaterials. *Nature* 5; 510(7503):103-8 (Jun 2014).

[35] Bale JB, Gonen S, Liu Y, Sheffler W, Ellis D, Thomas C, Cascio D, Yeates TO, Gonen T, King NP, Baker D. Accurate design of megadalton-scale two-component icosahedral protein complexes. *Science.* 22;353(6297):389-94 (Jul 2016).

[36] Hsia Y, Bale JB, Gonen S, Shi D, Sheffler W, Fong KK, Nattermann U, Xu C, Huang PS, Ravichandran R, Yi S, Davis TN, Gonen T, King NP, Baker D. Design of a hyperstable 60-subunit protein icosahedron. *Nature.* 535(7610):136-9 (Jul 2016).

[37] Tinberg CE, Khare SD, Dou J, Doyle L, Nelson JW, Schena A, Jankowski W, Kalodimos CG, Johnsson K, Stoddard BL, Baker D. Computational design of ligand-binding proteins with high affinity and selectivity. *Nature.* 501(7466):212-6 (Sep 2013).

[38] Procko E, Berguig GY, Shen BW, Song Y, Frayo S, Convertine AJ, Margineantu D, Booth G, Correia BE, Cheng Y, Schief WR, Hockenbery DM, Press OW, Stoddard BL, Stayton PS, Baker D. A computationally designed inhibitor of an Epstein-Barr viral Bcl-2 protein induces apoptosis in infected cells. Cell. 157(7):1644-56 (Jun 2014).

[39] Strauch EM, Fleishman SJ, Baker D. Computational design of a pH-sensitive IgG binding protein. *Proc Natl Acad Sci USA*. 111(2):675-80 (Jan 2014).

[40] Mills JH, Khare SD, Bolduc JM, Forouhar F, Mulligan VK, Lew S, Seetharaman J, Tong L, Stoddard BL, Baker D. Computational design of an unnatural amino acid dependent metalloprotein with atomic level accuracy. *J Am Chem Soc.* 135(36):13393-9 (Sep 2013).

[41] Procko E, Hedman R, Hamilton K, Seetharaman J, Fleishman SJ, Su M, Aramini J, Kornhaber G, Hunt JF, Tong L, Montelione GT, Baker D. Computational design of a protein-based enzyme inhibitor. *J Mol Biol.* 425(18):3563-75 (Sep 2013).

[42] Fleishman SJ, Whitehead TA, Ekiert DC, Dreyfus C, Corn JE, Strauch EM, Wilson IA, Baker D. Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. *Science.* 332(6031):816-21 (May 2011).

[43] Koday MT, Nelson J, Chevalier A, Koday M, Kalinoski H, Stewart L, Carter L, Nieusma T, Lee PS, Ward AB, Wilson IA, Dagley A, Smee DF, Baker D, Fuller DH. A Computationally Designed Hemagglutinin Stem-Binding Protein Provides In Vivo Protection from Influenza Independent of a Host Immune Response. *PLoS Pathog.* 12(2):e1005409 (Feb 2016).

[44] Niehrs, C. The complex world of WNT receptor signaling. *Nat Rev Mol Cell Biol.* 13(12):767-79 (2012).

[45] Sebbagh M, Borg JP. Insight into planar cell polarity. *Exp Cell Res.* pii: S0014-4827(14)00384-X (2014)..

[46] Stewart DJ. Wnt signaling pathway in non-small cell lung cancer. *J Natl Cancer Inst.*106(1):djt356 (2014)..

[47] Clevers, H. & Nusse, R. Wnt/Beta-Catenin Signaling and Disease. *Cell.* **149**, 1192-1205 (2012).

[48] Ueno K, Hirata H, Hinoda Y, Dahiya R. Frizzled homolog proteins, microRNAs and Wnt Signaling in Cancer. *Int J Cancer.* 132(8): 1731–1740 (2013).

[49] Suzuki H1, Watkins DN, Jair KW, Schuebel KE, Markowitz SD, Chen WD, Pretlow TP, Yang B, Akiyama Y, Van Engeland M, Toyota M, Tokino T, Hinoda Y, Imai K, Herman JG, Baylin SB. Epigenetic inactivation of SFRP genes allows constitutive WNT signaling in colorectal cancer. *Nat Genet.* (4):417-22 (2004).

[50] Aguilera O, Fraga MF, Ballestar E, Paz MF, Herranz M, Espada J et al. Epigenetic inactivation of the Wnt antagonist DICKKOPF-1 (DKK-1) gene in human colorectal cancer. *Oncogene*. 25: 4116–4121 (2006).

[51] Gonzalez-Sancho JM, Aguilera O, Garcia JM, Pendas-Franco N, Pena C, Cal Set al. The Wnt antagonist DICKKOPF-1 gene is a downstream target of beta-catenin/TCF and is downregulated in human colon cancer. *Oncogene*. 24: 1098–1103 (2005).

[52] He B1, Reguart N, You L, Mazieres J, Xu Z, Lee AY, Mikami I, McCormick F, Jablons DM. Blockade of Wnt-1 signaling induces apoptosis in human colorectal cancer cells containing downstream mutations. *Oncogene.* 24(18):3054-8 (2005).

[53] Sogabe Y1, Suzuki H, Toyota M, Ogi K, Imai T, Nojima M, Sasaki Y, Hiratsuka H, Tokino T. (2008). Epigenetic inactivation of SFRP genes in oral squamous cell carcinoma. *Int J Oncol.* 32(6):1253-61.

[54] Dijksterhuis JP, Baljinnyam B, Stanger K, Sercan HO, Ji Y, Andres O, Rubin JS, Hannoush RN, Schulte G. (2015). Systematic Mapping of WNT-FZD Protein Interactions Reveals Functional Selectivity by Distinct WNT-FZD Pairs. *J Biol Chem.* 290(11):6789-98.

[55] Wang HQ, Xu ML, Ma J, Zhang Y, Xie CH. Frizzled-8 as a putative therapeutic target in human lung cancer. *Biochem Biophys Res Commun.* (1):62-6 (Jan 2012).

[56] Dalal AlDeghaither, Brandon G Smaglo, and Louis M. Weiner. Beyond Peptides and mAbs - Current Status and Future Perspectives for Biotherapeutics with Novel Constructs. *J Clin Pharmacol.* 55(0 3): S4–S20 (Mar 2015).

[57] Wang D, Tan J, Xu Y, Han M, Tu Y, Zhu Z, Dou C, Xin J, Tan X, Zeng JP, Zhao G, Liu Z. The ubiquitin ligase RNF43 downregulation increases membrane expression of frizzled receptor in pancreatic ductal adenocarcinoma. *Tumour Biol.* 37(1):627-31 (Jan 2016).

[58] Koo BK, Spit M, Jordens I, Low TY, Stange DE, van de Wetering M, van Es JH, Mohammed S, Heck AJ, Maurice MM, Clevers H. Tumour suppressor RNF43 is a stem-cell E3 ligase that induces endocytosis of Wnt receptors. *Nature.* 488(7413):665-9 (Aug 2012).

[59] Jiang X, Hao HX, Growney JD, Woolfenden S, Bottiglio C, Ng N , Lu B, Hsieh MH, Bagdasarian L, Meyer R, Smith TR, Avello M, Charlat O, Xie Y, Porter JA, Pan S, Liu J, McLaughlin ME, Cong F. Inactivating mutations of RNF43 confer Wnt dependency in pancreatic ductaladenocarcionma. *Proc Natl Acad Sci U S A.* 110(31) (Jul 2013).

[60] Zachary Steinhart, Zvezdan Pavlovic, Megha Chandrashekhar, Traver Hart, Xiaowei Wang, Xiaoyu Zhang, Mélanie Robitaille, Kevin R Brown, Sridevi Jaksani, René Overmeer, Sylvia F Boj, Jarrett Adams, James Pan, Hans Clevers, Sachdev Sidhu, Jason Moffat, Stéphane Angers. Genome-wide CRISPR screens reveal a Wnt–FZD5 signaling circuit as a druggable vulnerability of *RNF43*-mutant pancreatic tumors. *Nat Med.* 23(1):60-68 (2017 Jan).

[61] Giannakis M, Hodis E, Jasmine Mu X, Yamauchi M, Rosenbluh J, Cibulskis K, Saksena G, Lawrence MS, Qian ZR, Nishihara R, Van Allen EM, Hahn WC, Gabriel SB, Lander ES, Getz G, Ogino S, Fuchs CS, Garraway LA. RNF43 is frequently mutated in colorectal and endometrial cancers. *Nat Genet.* 46(12):1264-6 (Dec 2014).

[62] Nusse, R. & Varmus, H. Three decades of Wnts: a personal perspective on how a scientific field developed. *The EMBO journal*, 31(12):2670-84 (2012).

[63] Biechele TL, Kulikauskas RM, Toroni RA, Lucero OM, Swift RD, James RG, Robin NC, Dawson DW, Moon RT, Chien AJ. Wnt/β-catenin signaling and AXIN1 regulate apoptosis

triggered by inhibition of the mutant kinase BRAFV600E in human melanoma. *Sci Signal.* 5(206):ra3 (2012).

[64] Stoick-Cooper CL, Moon RT, Weidinger G. (2007). Advances in signaling in vertebrate regeneration as a prelude to regenerative medicine. *Genes Dev.* 21(11):1292-315.

[65]Janda CY1, Waghray D, Levin AM, Thomas C, Garcia KC. (2012). Structural basis of Wnt recognition by Frizzled. *Science.* 337(6090):59-64.

[66] http://www.oncomed.com/Pipeline

[67] Plückthun A. Designed Ankyrin Repeat Proteins (DARPins): Binding Proteins for Research, Diagnostics, and Therapy. *Annu Rev Pharmacol Toxicol.* 6;55:489-511 (2015).

[68] Park K, Shen BW, Parmeggiani F, Huang PS, Stoddard BL, Baker D. Control of repeat-protein curvature by computational protein design. *Nat Struct Mol Biol.* Jan 12. doi: 10.1038/nsmb.2938 (2015).

[69] Bogan AA, Thorn KS. Anatomy of hot spots in protein interfaces. *Journal of Molecular Biology*. 280(1): 1–9. (July 1998)

[70] Whitehead TA, Chevalier A, Song Y, Dreyfus C, Fleishman SJ, De Mattos C, Myers CA, Kamisetty H, Blair P, Wilson IA, Baker D. Optimization of affinity, specificity and function of designed influenza inhibitors using deep sequencing. *Nat Biotechnol*. 30(6):543-8. (May 2012)

[71] Kunkel TA. Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proc. Natl. Acad. Sci. USA.* 1985; 82:488–492.

[72] Janda CY*, Dang LT*, You C, Chang J, Lau W, Zhong ZA, Yan KS, Marecic O, Siepe D, Li X, Moody JD, Williams BO, Clevers H, Piehler J, Baker D, Kuo CJ, Garcia KC. Surrogate Wnts that phenocopy canonical Wnt/B-catenin signaling. *Nature.* 545; 234-237. (May 11, 2017)

[73] Zhong Z, Ethen NJ, Williams BO. WNT signaling in bone development and homeostasis. *Wiley interdisciplinary reviews. Developmental biology.* 3, 489-500. (2014)

[74] Salazar VS, Ohte S, Capelo LP, Gamer L, Rosen V. Specification of osteoblast cell fate by canonical Wnt signaling requires Bmp2. *Development*.143(23):4352-4367 (Dec 2016).

[75] Huch M, Dorrell C, Boj SF, van Es JH, Li VS, van de Wetering M, Sato T, Hamer K, Sasaki N, Finegold MJ, Haft A, Vries RG, Grompe M, Clevers H.
In vitro expansion of single Lgr5+ liver stem cells induced by Wnt-driven regeneration. *Nature.* 494(7436):247-50 (Feb 2013).

[76] Barker N, Huch M, Kujala P, van de Wetering M, Snippert HJ, van Es JH, Sato T, Stange DE, Begthel H, van den Born M, Danenberg E, van den Brink S, Korving J, Abo A, Peters PJ, Wright N, Poulsom R, Clevers H. Lgr5(+ve) stem cells drive self-renewal in the stomach and build long-lived gastric units in vitro. *Cell Stem Cell.* 6(1):25-36 (Jan 2010).

[77]Sato T, Vries RG, Snippert HJ, van de Wetering M, Barker N, Stange DE, van Es JH, Abo A, Kujala P, Peters PJ, Clevers H. Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche. *Nature.* 459(7244):262-5 (May 2009).

[78] Kretzschmar K, Clevers H. Organoids: Modeling Development and the Stem Cell Niche in a Dish.□ *Dev Cell.* 38(6):590-600 (Sep 2016)

[79] Planas-Paz L, Orsini V, Boulter L, Calabrese D, Pikiolek M, Nigsch F, Xie Y, Roma G, Donovan A, Marti P, Beckmann N, Dill MT, Carbone W, Bergling S, Isken A, Mueller M, Kinzel B, Yang Y, Mao X,Nicholson TB, Zamponi R, Capodieci P, Valdez R, Rivera D, Loew A, Ukomadu C, Terracciano LM, Bouwmeester T, Cong F, Heim MH, Forbes SJ, Ruffner H, Tchorz JS. The RSPO-LGR4/5-ZNRF3/RNF43 module controls liver zonation and size. *Nat Cell Biol.* 18(5):467-79 (May 2016).

[80] Benhamouche S, Decaens T, Godard C, Chambrey R, Rickman DS, Moinard C, Vasseur-Cognet M, Kuo CJ, Kahn A, Perret C, Colnot S. Apc tumor suppressor gene is the "zonation-keeper" of mouse liver. *Dev Cell.* 10(6):759-70. (Jun 2006)

[81] Cooper S, Khatib F, Treuille A, Barbero J, Lee J, Beenen M, Leaver-Fay A, Baker D, Popović Z, Players F. Predicting protein structures with a multiplayer online game. *Nature.* 466(7307):756-60 (Aug 2010).

[82]Azoitei ML, Correia BE, Ban YE, Carrico C, Kalyuzhniy O, Chen L, Schroeter A, Huang PS, McLellan JS, Kwong PD, Baker D, Strong RK, Schief WR. Computation-guided backbone grafting of a discontinuous motif onto a protein scaffold. *Science.* 334(6054):373-6. (Oct 2011)

[83]Fleishman SJ, Leaver-Fay A, Corn JE, Strauch EM, Khare SD, Koga N, Ashworth J, Murphy P, Richter F, Lemmon G, Meiler J, Baker D. RosettaScripts: a scripting language interface to the Rosetta macromolecular modeling suite. *PLoS One.* 6(6):e20161 (2011).

[84] Leaver-Fay A, Tyka M, Lewis SM, Lange OF, Thompson J, Jacak R, Kaufman K, Renfrew PD, Smith CA, Sheffler W, Davis IW, Cooper S, Treuille A, Mandell DJ, Richter F, Ban YE, Fleishman SJ, Corn JE, Kim DE, Lyskov S, Berrondo M, Mentzer S, Popović Z, Havranek JJ, Karanicolas J, Das R, Meiler J, Kortemme T, Gray JJ, Kuhlman B, Baker D, Bradley P. ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol.* 487:545-74 (2011).

[85] Hoover DM, Lubkowski J. DNAWorks: an automated method for designing oligonucleotides for PCR-based gene synthesis. *Nucleic Acids Res.* 30(10):e43 (May 2002).

[86] Chao G, Lau WL, Hackel BJ, Sazinsky SL, Lippow SM, Wittrup KD. Isolating and engineering human antibodies using yeast surface display. *Nat Protoc.* 1(2):755-68 (2006).

[87] Boder ET, Wittrup KD. Yeast surface display for directed evolution of protein expression, affinity and stability. *Methods Enzymol.* 328:430-44 (2000).

[88] Kunkel TA. Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proc. Natl. Acad. Sci. USA.* 82:488–492 (1985).

[89] Benatuil L, Perez JM, Belk J, Hsieh CM. An improved yeast transformation method for the generation of very large human antibody libraries. *Protein Eng Des Sel.* 23(4):155-9 (Apr 2010).

[90] Zhang J, Kobert K, Flouri T, Stamatakis A. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics.* 30(5):614-20 (Mar 2014).

[91] Fowler DM, Araya CL, Gerard W, Fields S. Enrich: software for analysis of protein function by enrichment and depletion of variants. *Bioinformatics*. 27(24):3430-1 (Dec 2011).

[92] He B1, Reguart N, You L, Mazieres J, Xu Z, Lee AY, Mikami I, McCormick F, Jablons DM. Blockade of Wnt-1 signaling induces apoptosis in human colorectal cancer cells containing downstream mutations. *Oncogene.* 24(18):3054-8 (2005).

[93] Abrahamsson AE, Geron I, Gotlib J, Dao KH, Barroga CF, Newton IG, Giles FJ, Durocher J, Creusot RS, Karimi M, Jones C, Zehnder JL, Keating A, Negrin RS,Weissman IL, Jamieson CH. Glycogen synthase kinase 3beta missplicing contributes to leukemia stem cell generation. *Proc Natl Acad Sci U S A.* 106(10):3925-9 (2009).

[94] Zhang Y1, Morris JP 4th, Yan W, Schofield HK, Gurney A, Simeone DM, Millar SE, Hoey T, Hebrok M, Pasca di Magliano M. Canonical wnt signaling is required for pancreatic carcinogenesis. *Cancer Res.* 73(15):4909-22 (2013).

[95]Björklund P, Svedlund J, Olsson AK, Akerström G, Westin G.
The internally truncated LRP5 receptor presents a therapeutic target in breast cancer. *PLoS One.* 4(1):e4243 (2009).

[96] Takahashi-Yanaga F, Kahn M. Targeting Wnt signaling: can we safely eradicate cancer stem cells? *Clin Cancer Res.* 16(12):3153-62 (2010).

[97] Anastas, JN and Moon, RT. WNT signalling pathways as therapeutic targets in cancer. *Nature Rev Cancer.* 13:11-26 (2013).

[98] Le PN1, McDermott JD1, Jimeno A2. Targeting the Wnt pathway in human cancers: Therapeutic targeting with a focus on OMP-54F28. *Pharmacol Ther.* pii:S0163-7258(14)00162-4 (2014).

[99] Barker N1, Clevers H. Mining the Wnt pathway for cancer therapeutics. *Nat Rev Drug Discov.* 5(12):997-1014 (2006).

[100] Huang PS, Ban YE, Richter F, Andre I, Vernon R, Schief WR, Baker D. RosettaRemodel: a generalized framework for flexible backbone protein design. *PLoS One.* *6*(8):e24109 (2011).

[101] Matsuyama M**,** Kondo F**, Ishihara T,** Yamaguchi T**,** Ito R**,** Tsuyuguchi T**,** Tawada K**,** Yokosuka O**. Evaluation of pancreatic intraepithelial neoplasia and mucin expression in normal pancreata.** *J Hepatobiliary Pancreat Sci.* **19(3):242-8 (May 2012).**

[102] Zakeri B, Fierer JO, Celik E, Chittock EC, Schwarz-Linek U, Moy VT, Howarth M. Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesion. *Proc Natl Acad Sci U S A.* 109(12):E690-7 (Mar *2*012).

[103]https://www.rosettacommons.org/docs/latest/scripting_documentation/RosettaScripts/TaskOperations/taskoperations_pages/ConsensusLoopDesignOperation

[104] Benatuil L, Perez JM, Belk J, Hsieh CM. An improved yeast transformation method for the generation of very large human antibody libraries. *Protein Eng Des Sel.* 23(4):155-9. (Apr 2010)

[105] Mosavi LK, Cammett TJ, Desrosiers DC, Peng ZY. The ankyrin repeat as molecular architecture for protein recognition. *Protein Sci.* 13(6):1435–1448 (Jun 2004).

[106] Tamaskovic R, Simon M, Stefan N, Schwill M, Plückthun A. Designed ankyrin repeat proteins (DARPins) from research to therapy. *Methods Enzymol.* 503:101-34 (2012).

[107] Binz HK, Amstutz P, Kohl A, et al. High-affinity binders selected from designed ankyrin repeat protein libraries. *Nat. Biotechnol*. **22** (5): 575–82. (May 2004).

[108] Parra RG, Espada R, Verstraete N, Ferreiro DU, et al. Structural and Energetic Characterization of the Ankyrin Repeat Protein Family. *Plos Comput. Biol*. **12** (11): 575–82 (Dec 2015).

[109] Mosavi LK, Minor DL, Peng ZY. Consensus-derived Structural Determinants of the Ankyrin Repeat Motif. *Proc Natl Acad Sci USA*. 99(25):16029–16034 (2002 Dec).

[110] Tamaskovic R, Simon M, Stefan N, Schwill M, Plückthun A. Designed ankyrin repeat proteins (DARPins) from research to therapy. *Methods Enzymol.* 503:101-34 (2012).

[111] He B1, Reguart N, You L, Mazieres J, Xu Z, Lee AY, Mikami I, McCormick F, Jablons DM. Blockade of Wnt-1 signaling induces apoptosis in human colorectal cancer cells containing downstream mutations. *Oncogene.* 24(18):3054-8 (2005).

[112] Abrahamsson AE, Geron I, Gotlib J, Dao KH, Barroga CF, Newton IG, Giles FJ, Durocher J, Creusot RS, Karimi M, Jones C, Zehnder JL, Keating A, Negrin RS,Weissman IL, Jamieson

CH. Glycogen synthase kinase 3beta missplicing contributes to leukemia stem cell generation. *Proc Natl Acad Sci U S A.* 106(10):3925-9 (2009).

[113] Zhang Y1, Morris JP 4th, Yan W, Schofield HK, Gurney A, Simeone DM, Millar SE, Hoey T, Hebrok M, Pasca di Magliano M. Canonical wnt signaling is required for pancreatic carcinogenesis. *Cancer Res.* 73(15):4909-22 (2013).

[114]Björklund P, Svedlund J, Olsson AK, Akerström G, Westin G. The internally truncated LRP5 receptor presents a therapeutic target in breast cancer. *PLoS One.* 4(1):e4243 (2009).

[115] Kikuchi A1, Yamamoto H, Kishida S. Multiplicity of the interactions of Wnt proteins and their receptors. *Cell Signal.* 19(4):659-71 (2007).

[116] Parmeggiani F, Huang PS, Vorobiev S, Xiao R, Park K, Caprari S, Su M, Seetharaman J, Mao L, Janjua H, Montelione GT, Hunt J, Baker D. A general computational approach for repeat protein design. *J Mol Biol.* 427(2):563-75 (Jan 2015).

[117] Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ. PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Res.* 33(Web Server issue):W363-7 (Jul 2005).

[118] Nivón LG, Bjelic S, King C, Baker D. Automating human intuition for protein design. *Proteins.* 82(5):858-66 (May 2014).

[119] Benatuil L, Perez JM, Belk J, Hsieh CM. An improved yeast transformation method for the generation of very large human antibody libraries. *Protein Eng Des Sel.* 23(4):155-9 (Apr 2010).