# Computational Design of Proteins and Peptides Harboring Binding Interfaces

Ta-Yi Yu

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2021

Reading Committee:

David Baker, Chair

Valerie Daggett

Ning Zheng

Program Authorized to Offer Degree:

Bioengineering

University of Washington

**Abstract**

Computational Design of Proteins and Peptides Harboring Binding Interfaces

Ta-Yi Yu

Chair of the Supervisory Committee:

Professor David Baker

Biochemistry

Computational protein design has advanced rapidly in recent years. Starting from designing for novel protein and peptide structures, the focus of the field has shifted from designing for folds to designing for functions. In this dissertation, I will present the design of *de novo* miniproteins as antagonists for CD131. CD131 mediates signaling of three different cytokines, IL-3, IL-5 and GM-CSF. The signaling is triggered when CD131 forms a hexamer and/or a dodecamer complex with the cytokine and cytokine specific alpha chains. All three CD131 associated cytokines are implicated in mediating inflammation and three different antibodies against IL-5 have been approved to treat asthma. Using computational design and directed evolution, I have developed 3 different miniproteins that inhibit CD131 mediated STAT5 phosphorylation in a human cell line. These proteins all bind CD131 with sub-100nM to pM $K_d$ and exhibit good thermostability *in vitro*. A low resolution cocrystal structure confirms the binding site and binding orientation of one of the design proteins. I will also describe using disulfide closed macrocycles to test our ability to design binding interactions with structured loops. Although binding with regular secondary structures has been explored before, the design rules of structured loop interactions are

less well understood. Through a case study of designed disulfide macrocycles interacting with DNA Polymerase III β-Clamp, I will discuss findings that should help improve our ability in designing loop interactions.

**Acknowledgements**

I am blessed in my graduate school career to work with a group of people that are so talented and enthusiastic about protein design. I have had the privilege to witness firsthand a lot of exciting breakthroughs in *de novo* protein design over the years. Although my projects have gone through their fair share of ups and downs, I want to thank my advisor, David Baker, for staying positive and always instilling confidence and trust in me to solve whatever lies ahead. David's insight and guidance are tremendously valuable even when things seem bleak. The Institute of Protein Design is truly a unique place that I am proud to be a part of and it is in no small part thanks to David for creating a collaborative environment. I also want to thank my committee members, Valerie Daggett, Marion Pepper, Daniel Ratner and Ning Zheng for always being curious about my projects and dedicating their valuable time for my committee meetings.

I want to thank all the mentors and people who have lend a helping hand and provided valuable advice throughout the year. Special thanks to Gabe Rocklin for being my primary mentor when I first started in the lab. Gabe is an impressively talented protein designer that I have had a great time working with and learning from on a variety of topics ranging from Rosetta, programming and data analysis. I also want to thank Danny Sahtoe who I've always leaned on with anything in the lab or seeking advice on protein design. I really enjoy the numerous conversations we had on troubleshooting, experiments to try or just blue-sky ideas. Thanks also to Stephanie Berger, Gaurav Bhardwaj, Franziska Seeger and Anindya Roy for the projects we have worked on together and for being the helpful resources that I can come ask for help. The idea of targeting CD131 came from Aaron Chevalier who through intensive research concluded that this is a worthwhile and interesting protein to target. Aaron is always candid in his suggestions and he is a delight to talk to every time even after he has long left the lab.

A lot of people have contributed to my projects and I would not have gotten the results I am presenting in this dissertation without their help. Will Sheffler, Brian Coventry, Longxing Cao, Patrick Salveson, Adam Moyer and Daniel Silva develop numerous software and improve the design protocols that made the method possible. Asim Bera solved the DnaN and 53C cocrystal structure. Inna Goreshnik, Lauren Carter, Sam Halabiya and everyone in the IPD core as well as BioFab produced materials and ran assays that make the projects move so much faster.

I am fortunate to be able to work with Angel Lopez's and Michael Parker's lab who are experts in CD131 biology. Special thanks to Steffi Cheung Tung Shing in Michael's lab for solving the CD131/933 structure and Winnie Kan from Angel's lab for running signal inhibition assays on the CD131 antagonists.

Last but not least, I want to thank my family for supporting me throughout the years. My parents, Hei-wue Huang and Chan-ren Yu, for always pushing me to do better. And of course my wife, Tianyong Zheng, for always being there through the good and the bad, the failures and the successes in life. I am truly blessed and cannot ask for a better partner. I am excited to see what holds in the future and I cannot think of a better person to take on the challenges together.

# Table of Contents

## Chapter 1: Design of Proteins with Binding Interfaces Against CD131

Introduction

In this section, I will discuss my work on designing antagonists against CD131, commonly known as beta common receptor. CD131 is a common receptor subunit shared between GM-CSF, IL-3 and IL-5. These cytokines, although dispensable in normal hematopoiesis, are produced at sites of inflammation to exert their immune modulating effects[1]. Because of these cytokines' role in inflammation, multiple biologics targeting this pathway are either approved as treatments[2] or are being developed in clinical trials[3,4,5]. Due to its clinical significance, proteins able to antagonize or agonize this signaling pathway can find broad applications as vaccine adjuvants, therapy for autoimmune diseases, or even for treating severe cases of COVID-19[6].

CD131 is a homodimer with monomeric units each consisting of 3 extracellular fibronectin III domains, a transmembrane domain and a cytoplasmic tail capable of interacting with the 14-3-3 proteins[7] or Shc proteins[8]. Downstream signaling is facilitated by the recruitment of Jak2 and the activation of STAT5 A/B[9]. The hexameric signaling complex of each of the three cytokines includes two cytokine-specific alpha chains as well as the shared beta subunit. The ternary structure of GM-CSF-CD131-GM-CSFR⟨, however, suggests a distinct mode of activation. The distance between the CD131 cytoplasmic tail (~120Å) in the nominal hexameric complex made the transphosphorylation by Jak2 unlikely. As a result, it is proposed that a dodecamer observed in the crystal lattice is able to bring two hexameric assemblies together and provide functional dimerization[10]. Although the cytokine specific alpha chains are required for signaling[11], the cytoplasmic tail itself does not appear to be phosphorylated[12]. Therefore, the precise function of the alpha chain in the signaling complex remains to be elucidated.

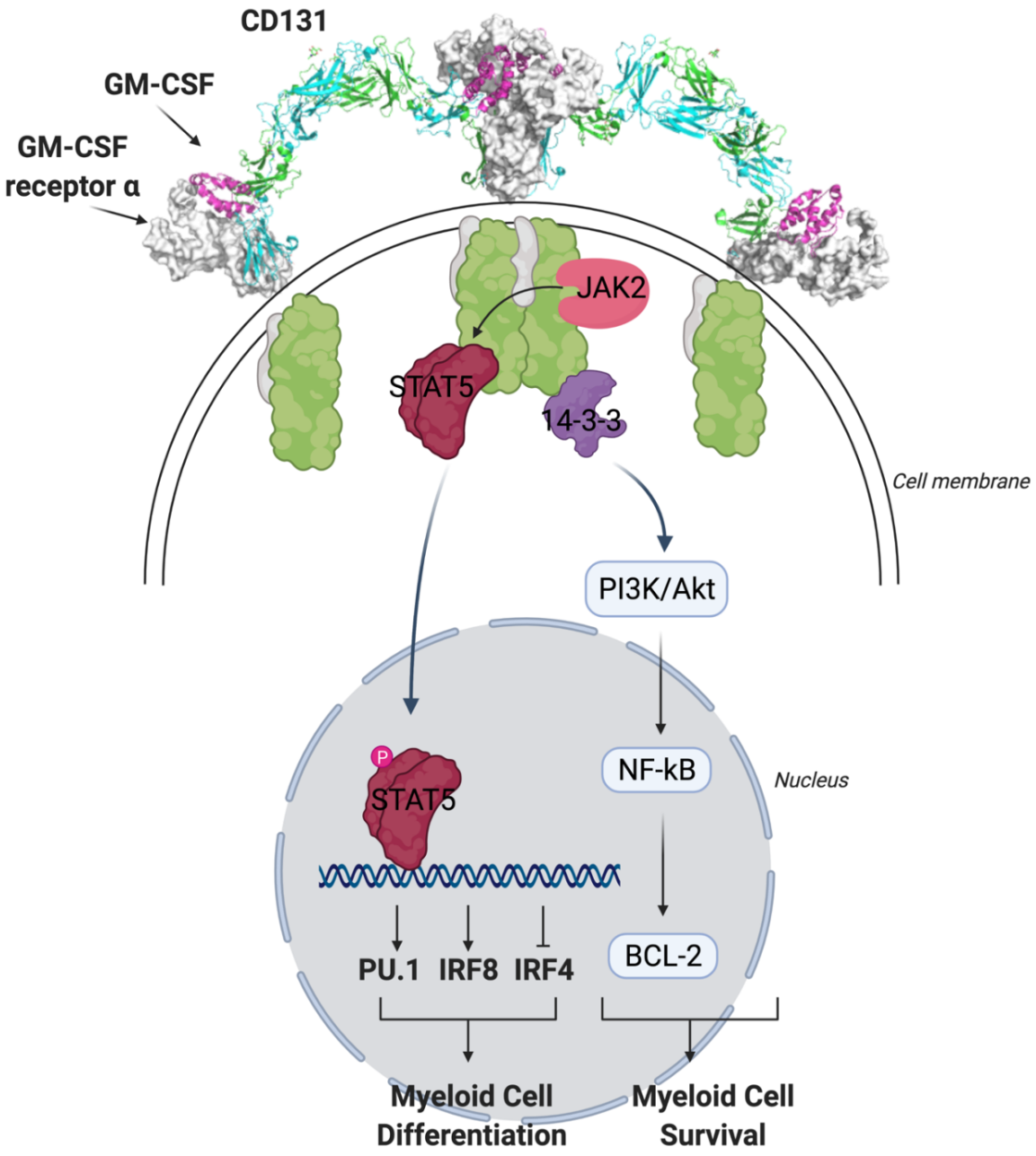



**Figure 1. CD13/GM-CSF Signaling Complex and Pathway.** GM-CSF forms a dodecamer assembly on the cell surface with CD131 and GM-CSF alpha receptor. The complex triggers JAK2 phosphorylation of STAT5 and causes the upregulation of PU.1 and IRF8 which leads to myeloid cell differentiation. An alternative pathway through the recruitment of 14-3-3 after cross phosphorylation leads to the activation of NF-kB that promotes myeloid cell survival.

In addition to the distinctive signaling complex, CD131 mediated signaling also exhibits an interesting concentration dependent "survive" versus "survive and proliferate" switch. Lower concentration of GM-CSF favors the recruitment of 14-3-3 and PI3K leading to cell survival whereas higher concentration results in the recruitment of both Shc and 14-3-3, leading to survival and proliferation[13]. Although CD131 associated cytokines are mostly produced by activated T cells, a wide variety of cell types such as macrophage, mast cells, eosinophiles and epithelial cells have all been reported to produce these cytokines[14]. While IL-5 is largely restricted to work on eosinophiles[15], GM-CSF and IL-3 have been reported to work on multiple cell types[16,17]. The breadth of the effect of these pathways has prompted studies of intervening molecules for disease such as asthma, rheumatoid arthritis and multiple sclerosis.

Although an antibody as well as short linear peptides against CD131 have been reported[4,18], *de novo* designed proteins targeting CD131 offer several advantages over both modalities. While it is well documented that antibodies can have excellent half-life, affinity and specificity, poor tissue penetration can lower the amount present at site of inflammation and hinder the efficacy. Comparing to short peptides, *de novo* designed proteins' larger size can afford better affinity and greater stability. The smaller size coupled with high stability and affinity achievable with *de novo* designed proteins made them particularly attractive in treating pulmonary diseases where local delivery had been shown to be more effective compared to systemic routes[19].

## Results

### *Computational Design of Single Domain Binders Against CD131*

To design the monovalent binder against CD131, a grafting based protocol was developed starting from the co-crystal structure of CD131 in complex with GM-CSF (PDB ID:

4NKQ[20]). Specifically, helix A of GM-CSF (Figure 1A) is chosen based on previous mutagenesis study[21] suggesting a glutamic acid residue buried at the interface critical for binding. The helix A backbone position from residue 14 to 28 is used as the motif in the Motifgraft protocol[22]. A total of 10,900 all alpha or mix alpha/beta protease screened[23] scaffolds are used to find segments suitable for replacing with the CD131 binding motif. Grafting is allowed if the all-backbone RMSD between the scaffold segment and the helical motif is below 1.5 Å and the N- and C- terminus alignment is within 1 Å. All scaffolds are between 49 and 67 residues in length and contain no disulfide bonds. The set is generated and curated internally by Baker Lab members and follows a previously described protocol[23]. The grafting protocol generated 80,395 grafted models. To quickly evaluate all the models as well as introducing a second arginine hotspot identified in an antibody-CD131 co-crystal structure (PDB ID: 5DWU[4]), all the models are evaluated in the Rotamer Interaction Field (RIF)[24] supplemented with the arginine hotspot (Figure 1B). This allows the filtering of models capable of incorporating the arginine hotspot as well as having positions conducive for additional contacts with CD131. The output models are optimized through Rosetta FastDesign[25], a Monte Carlo sequence design tool to improve interaction energy. The critical arginine and glutamic acid residues are kept unchanged in this process. The models are filtered by interface size (solvent accessible surface area, SASA) above 1400 $Å^2$, shape complementarity above 0.7 and binding energy ($\Delta G$) per 1000 SASA below -0.03. To facilitate stability, a single disulfide bond is introduced within the design when possible using a hash-based disulfide stapler (unpublished).

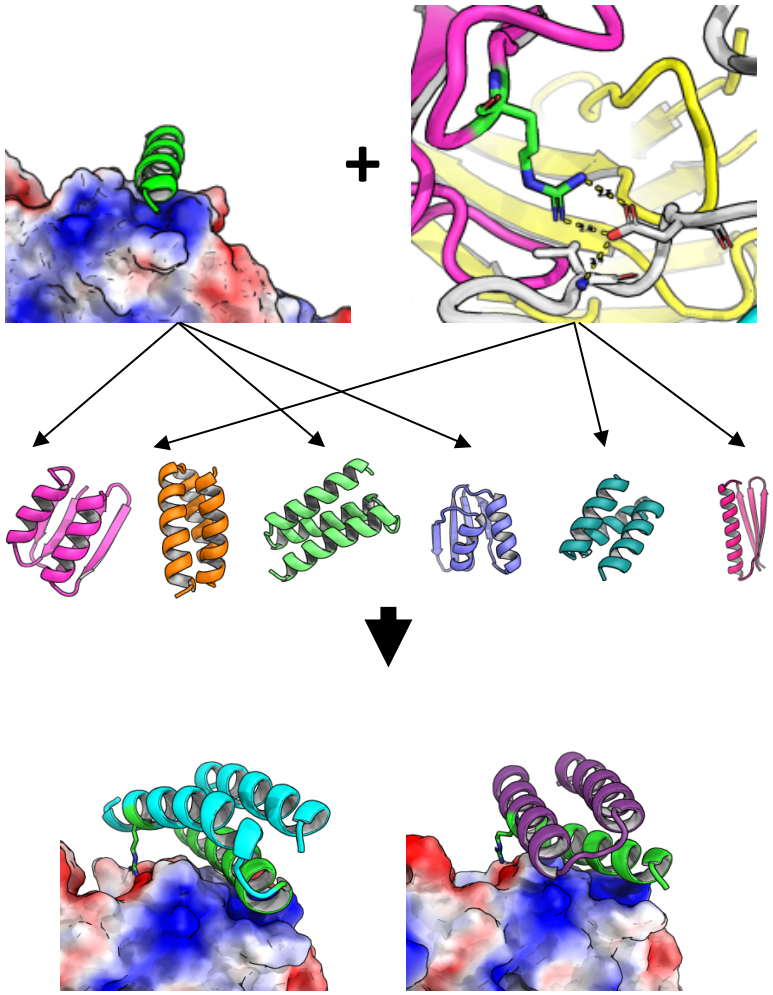


**Figure 2. Single Domain CD131 Binder Computational Design.** The A helix of GM-CSF from the complex structure is used as the starting motif. A set of scaffolds is queried for suitable segments to accommodate the helix. Grafted models are further filtered/ranked in Rotamer Interaction Field (RIF) for possible placement of arginine hotspot from antibody structures. The resulting designs contain both the helix and the arginine hotspot as shown in bottom panel colored in green.

*Yeast Display Screening and Site Saturation Mutagenesis (SSM) of Hits*

Genes encoding 14,669 unique designs were ordered as an oligonucleotide array synthesis pool and transformed along with linearized pETCON3 vector into yeast. For selection against human CD131 using fluorescence activated cell sorting (FACS), the pool is first selected for expression of design proteins and propagated for increased selection stringency against CD131 binding. 3 rounds of selections are done with 1 µM CD131, 100 nM CD131 and 1 nM

CD131. DNA of these selection pools is extracted, amplified with adaptor sequences and barcoded for tracking of sequence origin. Sequence reads are first filtered by quality scores. Design sequences are then ranked by enrichment: dividing the selection round frequency of the sequence by the original frequency in an unsorted library. No apparent biases toward high starting library frequency is observed in the last binding round (Figure 3A, top panel). An enrichment threshold of 10 for binder vs. non-binder classification (green/blue, Figure 3A, bottom panel) is set by the bimodal nature of the enrichment distribution at 1 nM CD131. 123 designs enriched above 10-fold compared to the unsorted library against 1 nM of CD131.

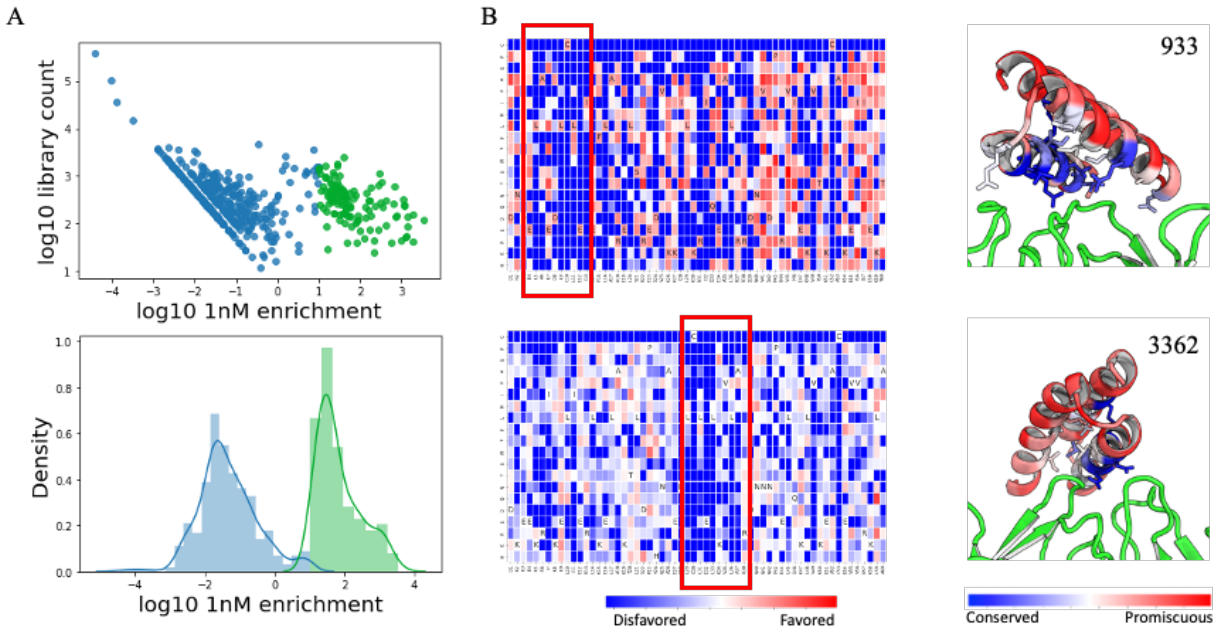



**Figure 3. Yeast Display Screen and Site Saturation Mutagenesis of Hits**. (A) Sorting results at the last round of library selection. Based on the distribution of the enrichment ratio, a cutoff of 10 is applied to classify binders vs. non-binders. (B) Left panel: SSM heatmap for 933 (top) and 3362 (bottom). The interface helix is the most conserved stretch in the map (red boxes). Right panel: sequence entropy from SSM mapped onto the models of 933 and 3362. Most of the interface as well as some core residues are conserved while solvent exposed positions are not.

An SSM library which contains every single point mutant of the original design sequence is created and screened similarly for each of the top 10 designs ranked by enrichment. The sorted populations are again sequenced. The mutation preferences of each position are mapped back to

the design sequences and models. Two parent design sequences, 933 and 3362, stand out because of the number of mutant sequences enriched at the final round. The SSM data are consistent with the models with key interface residues on the binding helix highly susceptible to mutations (Figure 3B, red box). Mapping the sequence entropy back to the design models also reveals that selected interface residues, including the glutamic acids and arginine hotspots, and a few core residues are highly conserved (Figure 3B, colored blue in right panel). In contrast, positions away from the interface and not participating in binder core formation are tolerable to a wide variety of sequence substitutions (Figure 3B, colored red in right panel). In addition to confirming the binding modes, the SSM sequencing data also provided insight into ways to improve the folding and binding of 933 and 3362. For 933, most of the dominant mutations happened at positions 40 – 43, the loop regions connecting helix 2 to helix 3. Interestingly, this loop contains the only position with positive phi (residue 40) and a proline (residue 43) with suboptimal phi-psi values. The most dominant substitution happens at V41 where mutations to tyrosine or glutamic acid are highly preferred over the wildtype. It is possible that both of these are filling in an exposed hydrophobic cavity between CD131 and 933 while forming polar interactions with K9 on CD131 (Figure 4A). For 3362, most of the improving mutations happen between position 18 – 23, the loop closest to CD131 and connecting helix 1 to helix 2. While the substitution preferences here are more subtle, it does prefer mutations such as T20G, L21P and

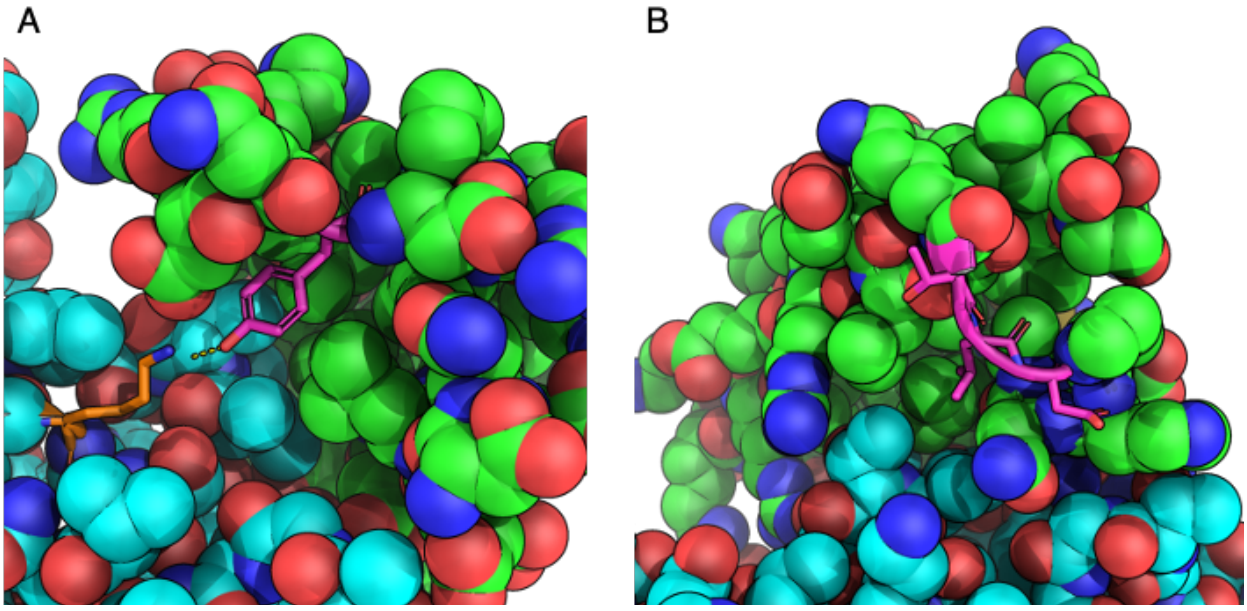



**Figure 4. Space Filling Models of Mutation Preferences for 933 and 3362**. Green: binders, Cyan: CD131, Magenta: mutations on binders, Orange: interacting residues on CD131. (A) V41Y mutation on 933 that fills an exposed hydrophobic area between 933 and CD131. Tyrosine can also hydrogen bond with a nearby lysine on CD131. (B) Position on 3362 that contains improve substitutions. Multiple G/P mutations are preferred here, suggesting a reorganization of the loop that can affect how the binding helixes are positioned.

P23D that perhaps restructure the loop to make connecting the two interface helices more favorable (Figure 4B).

*Design of bivalent CD131 Binders*

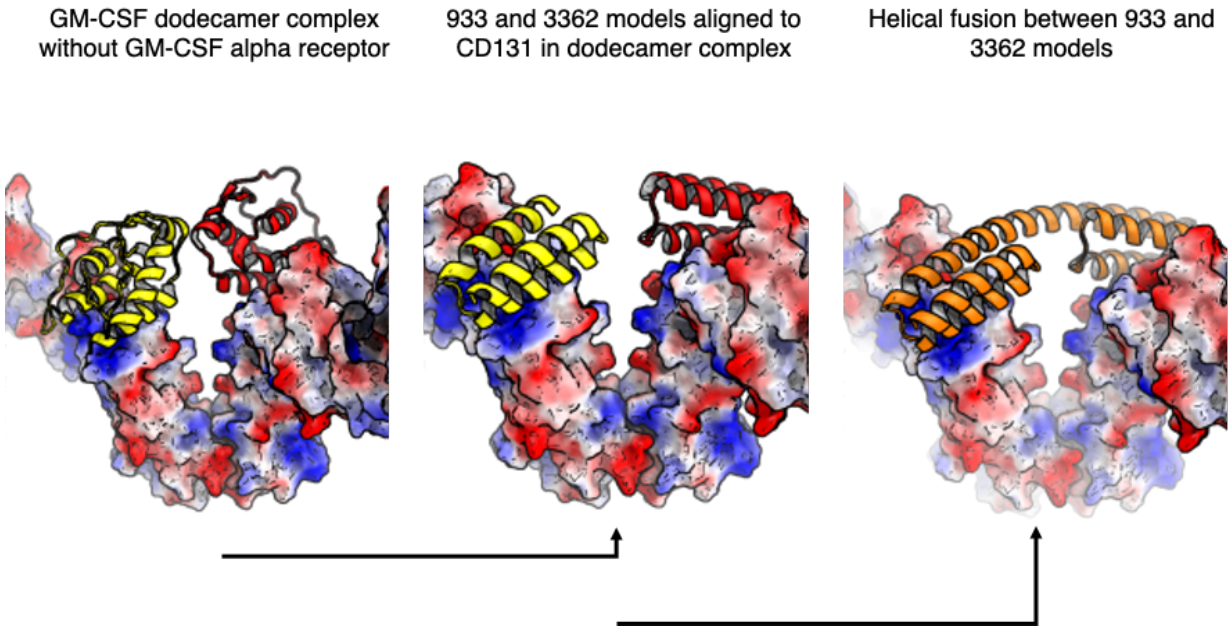
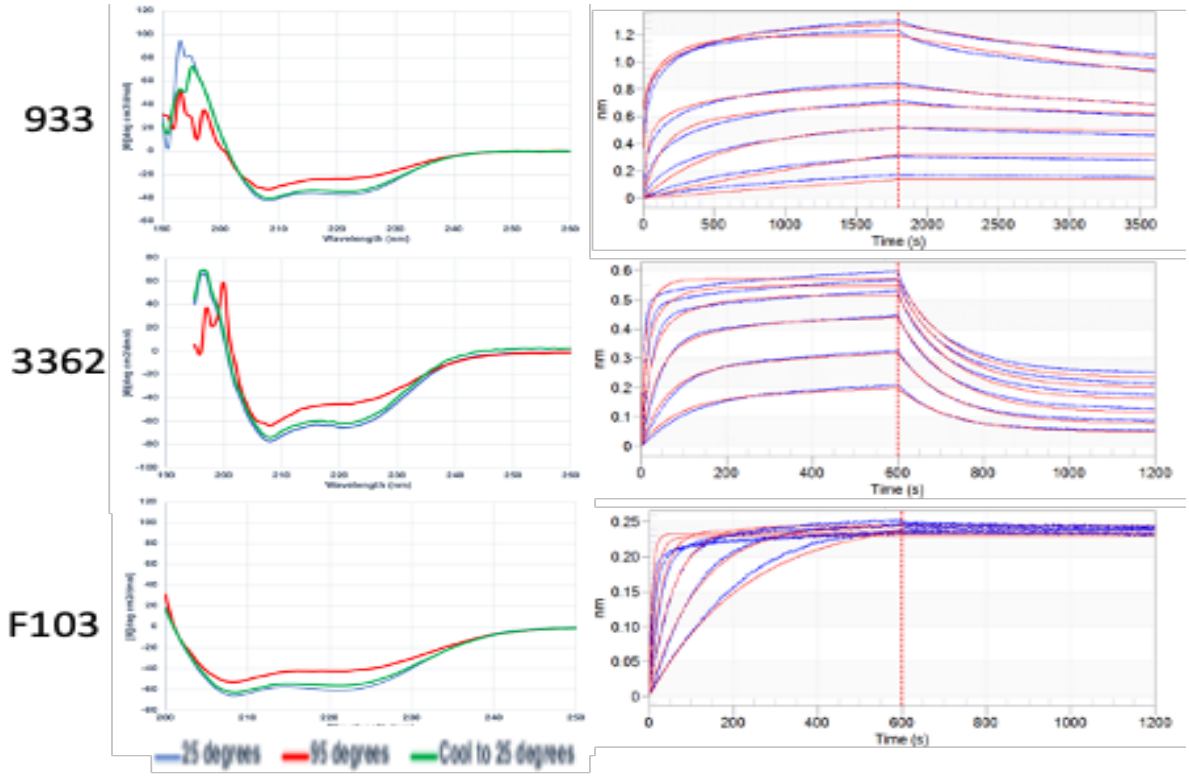



**Figure 5. Design of bivalent CD131 binders.** The CD131 dodecamer geometry in the GM-CSF complex is used as a guide to generate bivalent CD131 binders. Complex models of 933 and 3362 with CD131 are first superimposed on the CD131 dodecamer geometry. A helical linker is then created to bridge across the two binding domains targeting both chain of CD131 in the dodecamer.

In addition to the single domain binder designs, an approach to fuse two single domains together to create bivalent binders is also developed. To take advantage of the unique geometry of the dodecamer signaling complexes and the >100 of designs identified in the yeast display screening, each design is manually inspected for possible fusion to 933 via its carboxy terminus. The carboxy terminus of 933 is specifically chosen because of its direction pointing toward the other CD131 in the dodecamer complex. 933 and 3362 coincidently have their carboxy terminus and amino terminus, respectively, line up with each other such that a simple helical extension will be able to reach across and connect the two binding domains, forcing two CD131 to be in the geometry of the dodecamer complex (Figure 5). This approach will allow the two binding domains to interact with their respective CD131 target simultaneously forming a stable complex,

boosting the affinity beyond simple avidity would otherwise. The bridging helix is sampled from 9 – 11 residues based on the distance from the carboxy terminus of 933 to the amino terminus of 3362. Rosetta's blueprint protocol is used to sample fragments that connect the two helixes with appropriate length and twist while filtering for helix kink. The sequences of the linkers are optimized using the FastDesign protocol.

*Biophysical Characterization of CD131 Binders*




| Design | $K_d$ (M) | $K_{on}$ (1/Ms) | $K_{off}$ (1/s) |
|---|---|---|---|
| 933 | 5.76E-09 | 3.01E+04 | 1.73E-04 |
| 3362 | 5.24E-08 | 1.53E+05 | 8.04E-03 |
| F103 | <1.0E-12 | 3.45E+05 | <1.0E-07 |

**Figure 6. Biophysical characterization of CD131 binders.** Left upper panel: 933, 3362 and F103 all exhibit helical signature in circular dichroism at 25 degrees and retain some folderness at 95 degrees. Right upper panel: Biolayer Interferometry measurement showing the binding of 933, 3362 and F103 against CD131 in titration series. Bottom Table: Fitted Kd values for each of the three design.

Secondary structure content and thermal stability of *E. Coli* expressed and purified 933, 3362 and F103 are accessed with Circular Dichroism. All 3 exhibit an alpha helical fold as designed. Even after heading to 95 degrees, ~75% of helical signatures still persists (Figure 6) and the signal fully recovers upon cooling back to 25 degrees. Biolayer interferometry confirms their binding to CD131 *in vitro* in a titration series. Fitted $K_d$ value ranges from 52 nM to sub-pM (off rate for F103 too slow to measure).

*Site Saturation Mutagenesis (SSM) of two bivalent CD131 Binders*

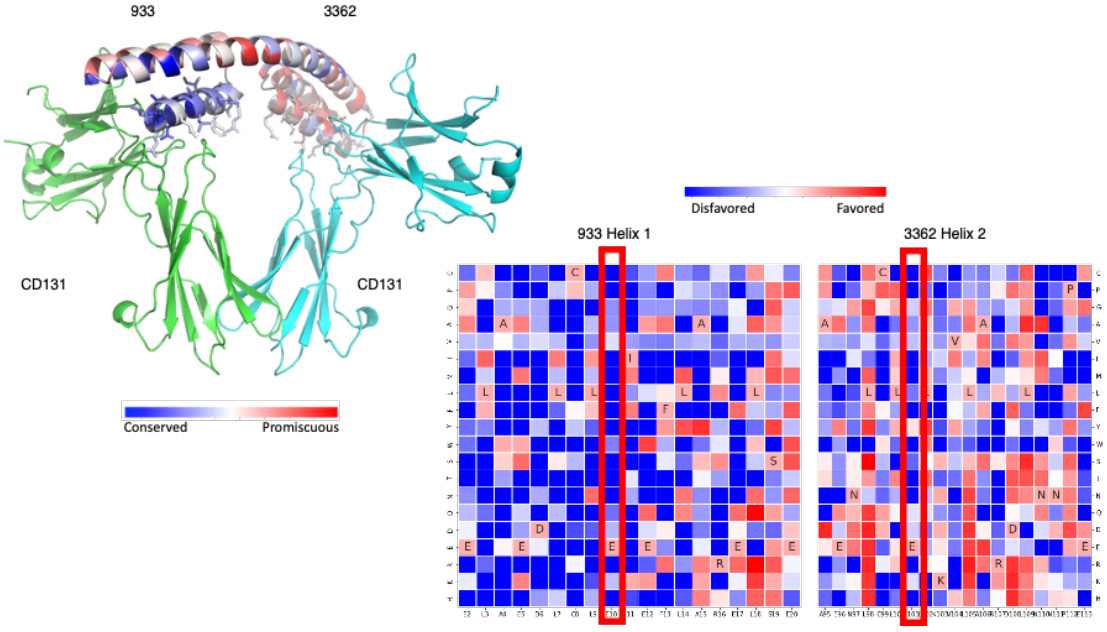



**Figure 7. Site Saturation Mutagenesis of Bivalent CD131 binders.** Upper structure: sequence entropy mapped onto the structure F152. High conversation is observed at the side of 933 likely due to is higher affinity toward CD131. Lower heatmap: substitution preferences of F152 zooming in at the main helix on both the 933 and 3362 sites. Glutamic acids at the 3362 interfaces (red box on the right) is still highly conserved, suggesting that F152 can mediate interactions through the 3362-binding domain as well.

Similar to the single domain cases, SSM libraries were constructed for F103 and F152 to probe the mutagenesis preferences of these two designs. Unfortunately, because of the large affinity differences between 933 and 3362, the majority of the prohibitive substitutions occur at

the 933-binding domain (Figure 7). However, in both of the 933 and 3362 binding domains, the deeply buried interface Glutamic acids (Figure 7, red box) on the helices are still highly sensitive to substitution, suggesting that both domains are still involved in CD131 binding.

*Designed CD131 Binders Inhibit STAT5 phosphorylation*

To test whether 933, F103 and F152 can inhibit the STAT5 phosphorylation triggered by the receptor binding of IL-3 and GM-CSF, TF-1, a leukemia cells line that depends on IL-3 and GM-CSF for growth, is used to test the binders' ability to block the signaling. The cells are grown in the presence of GM-CSF, washed and starved for 1 hour prior to incubating with a mixture of binder and IL-3 or GM-CSF. The cells are subsequently fixed by paraformaldehyde and incubated with pSTAT5 specific antibodies conjugated to Alexa fluor 647. The fluorescence is then measured by a flow cytometer. All three of the binders inhibit IL-3 and GM-CSF STAT5 phosphorylation to various degrees (Figure 8, left and middle panel) although the effect is much more pronounced with IL-3 (data not shown). While a full titration and IC50 calculation is needed to compare the strength of inhibition between the three binders, it appears that the bivalent version F152 inhibit the signalizing the best at 100 nM of binder concentration (Figure

8, right histogram) while 933 and F103 exhibit similar amount of inhibition.

*Low Resolution 933 and CD131 Co-crystal Structure Confirm the Binding Site and Orientation*

A 3.4 Å co-crystal structure of 933 and CD131 is obtained (Figure 9A). Although the low resolution prohibited the building of a flexible side chain and loop, the backbone placement of 933 confirmed that it is binding CD131 in site 2 where GM-CSF interacts with CD131. It is also using helix 1 as designed in the right orientation relative to the receptor (Figure 9B). It appears that the connection between helix 1 and helix 2 is significantly more unstructured and closer to CD131 in the structure as compared to the model.

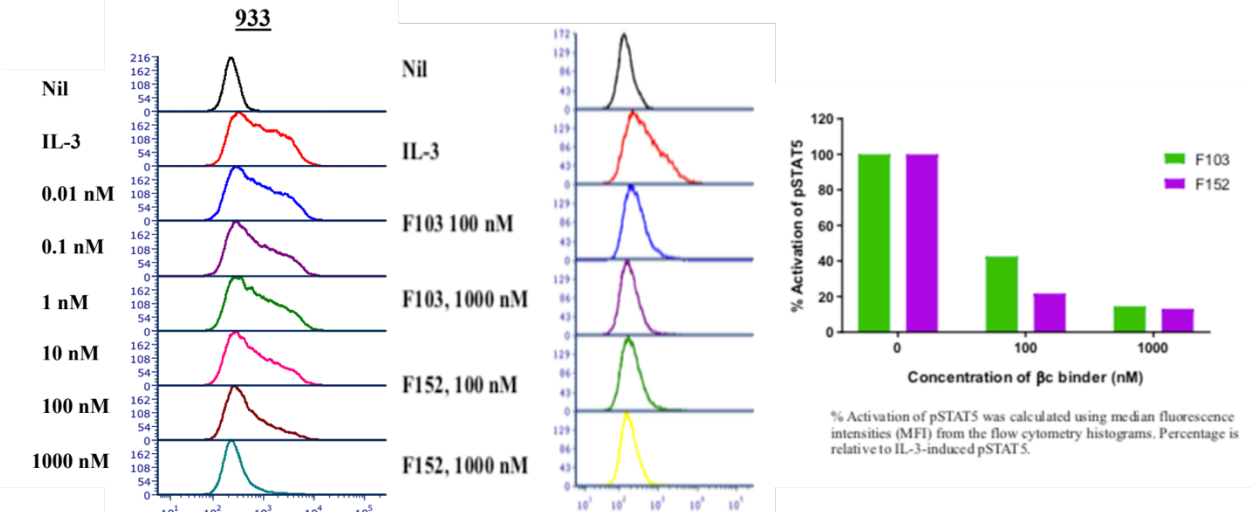



**Figure 8. pSTAT5 Inhibition of Monovalent and Bivalent CD131 binders.** The amount of pSTAT5 is measured by Alexa fluor 647 conjugated anti-pSTAT5. As binder concentrations increase, the fluorescence signal decease to the level of non-IL-3 treated cells indicating inhibition of IL-3 mediated signaling.

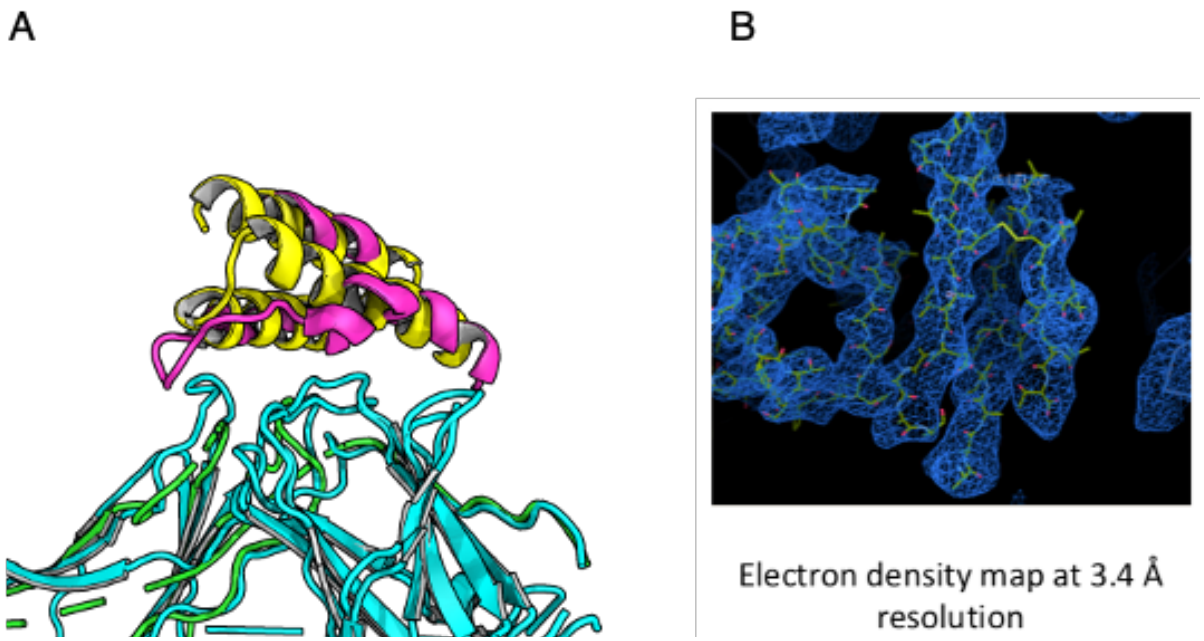



**Figure 9. Low resolution structure of 933 in complex with CD131.** (A) Overlay of the 3.4 Å structure with the model. Yellow/Cyan: 933 in complex with CD131, Magenta/Blue, crystal structure of 933 with CD131.

proteins that bind CD131. Biolayer interferometry confirms the $K_d$ of all 4 to be below 100 nM and in the case of F103 likely in the sub-nM range. Circular dichroism shows that 933, 3362 and F103 form alpha helical structures and have good thermal stability. 933, F103 and F152 all inhibit IL-3 or GM-CSF induced STAT5 phosphorylation in the TF1 cell line. The low-resolution crystal structure confirms 933's binding site and orientation relative to the receptor. While it is still unclear whether F103 and F152 can engage two CD131 simultaneously, the SSM sequencing data suggest both binding domains are capable of interacting with CD131.

Several experiments are on-going to further advance these molecules for proof-of-concept as both tool molecules to investigate the CD131 signaling axis as well as potential therapeutics. Analytical ultracentrifugation is being done to investigate the stoichiometry of both F103 and F152's interaction with CD131. Combinatorial library screening with mutations inspired from the SSM data is also ongoing to create variants of 933 and 3362 that are more

stable and higher affinity for *in vivo* experiments. The unique geometry of the dodecamer complex in conjunction with F103 and F152's possible mode of interactions made it possible to create molecules that can potentiate the formation of dodecamer if we can introduce GM-CSF alpha chain interacting components to the bivalent antagonists. Design works are being performed to combine the binding helixes of 933 and 3362 with components of GM-CSF cytokine that are responsible for alpha chain interactions. It is possible that these molecules, if successfully designed, can dictate cell survival or differentiation due to the rheostat nature observed with GM-CSF.

While we know all of the tested molecules are antagonistic in nature, we do not yet know whether receptor super-clustering, even in the absence of the cytokine specific alpha chain, can lead to a conversion from antagonistic to agonistic. We hypothesize that a higher order oligomeric fusion of these molecules might lead to conversion to agonism. We are in the process of generating fusions of these binding domain to heterooligomer for exploratory testing.

## Chapter 2: Design of Disulfides Stabilized Peptides with Binding Interfaces Against DNA Polymerase III β-Clamp (DnaN)

### Introduction

While a lot of efforts have gone into investigating the use of regular secondary structures to mediate protein-protein interactions, the use of structured loops to facilitate such a function is relatively less explored. Part of this stems from the early focus in exploring protein secondary structures and their arrangements to design new, idealized folds[26-28]. However, function sites such as enzyme active sites and binding interfaces are often built from structured loops with less regular secondary structures. For example, loops are much more common in enzyme active sites than they are in general[29]. Structured loops also play an important role in protein interfaces. Antibody, for example, recognizes an astonishing wide array of molecules with all sizes and shapes using their complementary determining regions (CDRs). CDRs are largely devoid of regular secondary structures[30]. As the focus of computational protein design shifts to protein functions, it seems inevitable that a better understanding of loop mediated functions as well as efficient ways to model and design loops will be required.

One of the ways to test our ability to generate loop mediated interfaces is through the design of peptide macrocycles. Nature has used these highly versatile peptides as antibiotics[31] or immunosuppressant[32]. Because of the constrained nature of these molecules, compared to linear peptides, they are more likely to adopt a nominal conformation in solution. This helps reduce the entropic loss during binding and afford the possibility of preorganization to increase affinity. Recent advances in computational samplings and scoring have made the exhaustive conformational enumeration of these peptides possible

[33,34]. In this project, I aim to test our ability to design structured loops, stabilized by carboxy and amino terminus disulfide bond, that can bind to DNA Polymerase III β-Clamp (DnaN).

DnaN is a homodimer about 82 kDa in size. Along with the τ-complex, it forms a stable assembly around the primer-templated junctions in an ATP-dependent manner. This assembly interacts with DNA polymerase III for processive chromosomal DNA replication[35]. Although targeting eukaryotic β-Clamp is complicated due to its function in healthy cells, targeting bacterial β-Clamp has been more fruitful in treatment such as *M. tuberculosis*. Specifically, a part cyclic peptide nature product, Griselimycin[36], was discovered that potently inhibits β-Clamp's interaction with the Pol III δ subunit in *M. tuberculosis*. Griselimycin development was originally shelfed due to poor pharmacokinetics but later resurrected because of the prevalence of drug resistant tuberculosis[37]. However, the presence of exotic residues such as (4R)-4-methyl-L-proline made the total synthesis difficult and it would be beneficial to have a macrocyclic compound that is more synthetically accessible.

Results

*In silico* Generation of Disulfide Cyclized Peptides

Two approaches were taken to generate disulfide cyclized peptides: (1) kinematic closures and (2) mining of PDB fragments followed by kinematic closures (Figure 10). For approach one, a linear peptide of a desired length is first created around a glycine anchor residue. The phi-psi torsion of this residue is randomly sampled and remains fixed in all the subsequent steps. The rest of the peptide is then treated as a loop closure problem. The GenKIC algorithm [33] then perform a series of perturbation on the phi psi values for the rest of the peptides biased by a flatten Ramachandran table. This invariably led to breaking of the cycle formed by both the disulfide and the peptide backbone. GenKIC then analytically solves for 6 pivot torsions that

allow the loop to remain close. For approach 2, 16598 high resolution (<2.5 Å) crystal structures with sequence homology (<50%) are collected from PDB. Each structure is broken down to 9mers by a sliding window. All the fragments are then clustered by their ABEGO phi-psi bins and their backbone hydrogen bonding patterns. Within each coarse cluster, RMSD based clustering is performed with a 0.25 Å cutoff using a density-based clustering algorithm, DBSCAN. One representative from each fine cluster is then closed with disulfide kinematically similar to approach one.

After generating all the putative disulfide macrocycles, the previously described ABEGO and backbone hydrogen bonds clustering is again performed to cut down the number of scaffolds needed to be docked onto DnaN.

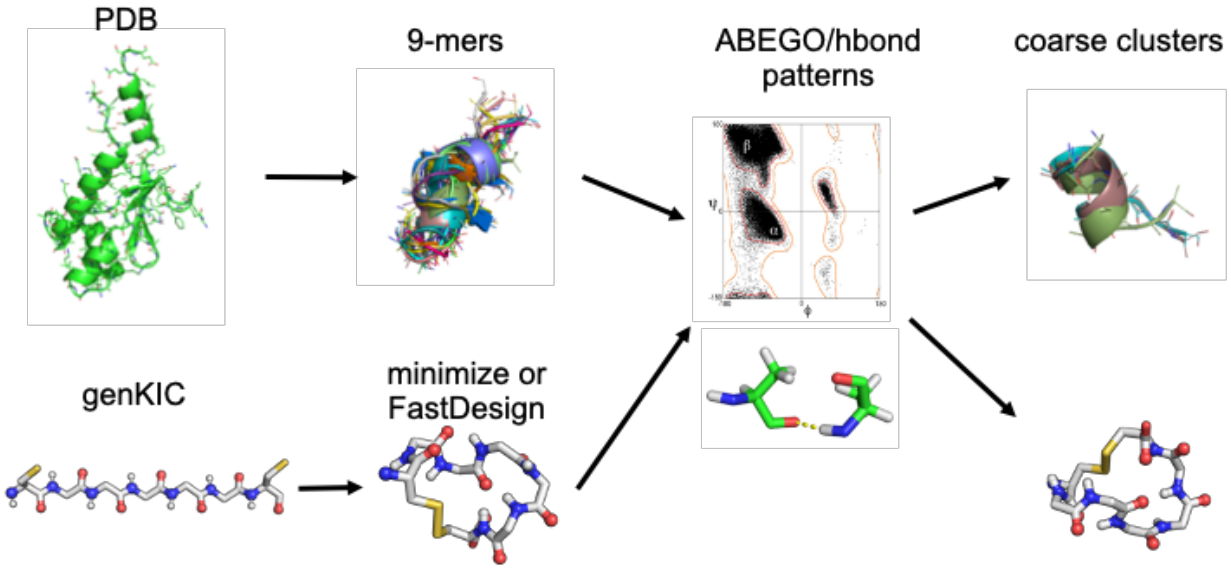



**Figure 10. Approaches taken to generate disulfide closed macrocycles.** Top: A set of 16598 PDB structures are broken down into 9mers with a slide window. Bottom: Analytical solution of peptide closures through a disulfide is accomplished with GenKIC. Both of these are then clustered by ABEGO and backbone hydrogen bonding patterns. In the case of the PDB fragments GenKIC is then performed to generate closed macrocycles.

*Computational Design of Peptides Against DNA Polymerase III β-Clamp*

To generate starting points for docking, a tryptophan or a phenylalanine residue is docked separately into site 1 of the DanN where Griselimycin occupies (Figure 11A). Using Rosetta GA

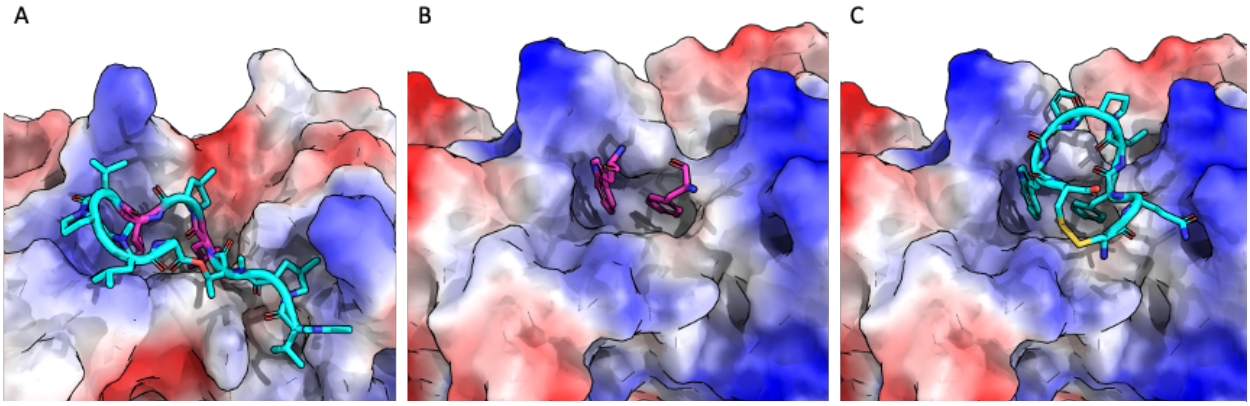



**Figure 11. Docking of disulfide macrocycle into site 1 of Griselimycin.** (A) The Griselimycin site 1 (two Leucine residues shown in magenta). (B) GA LiganDock found that the TRP and PHE can also target the same site and are compatible with each other. (C) RifDock is employed to found macrocycle positions that can accommodate both the TRP and the PHE residues.

LigandDock, top residue clusters are manually inspected for their compatibility in the pocket (Figure 11B). Two particular placements occupying where the two Leucine residues in Griselimycin interact with DnaN pocket are identified as a putative starting point. The RifDock framework is utilized to sample and score possible placement of the macrocycles by superimposing each positions' N-Ca-C backbone directly on top of each hotspot. Exhaustive sampling of every 5 rotational degrees for simultaneous hotspots incorporation as well as RIF score are then used to rank and output docking models (Figure 11C). The docked models are then designed with Rosetta's FastDesign protocol.

*Yeast Screening and SSM of Peptide Against DNA Polymerase III β-Clamp*

1034 total designs are ordered and created for an oligonucleotide library similar to Chapter 1. Biotinylated DnaN is used as the target protein. Selections were performed for 3 rounds with 2 uM, 500 nM and 100 nM DnaN, respectively. One particular design, 53C,

dominates the later round output but three designs have enrichment value above > 1000 where the median enrichment is 12.3 at this round.

*In vitro Characterization of peptide 53C*

Peptide 53C is synthesized by WuXi AppTec using Solid Phase Peptide Synthesis and subsequently air oxidized for disulfide formation. Three versions of 53C are prepared, the original sequences with amino terminus acetylation and carboxy terminus amidation, a version with KGGSGG preceding the design sequence with biotin attach to the NH2 of lysine and a version like the KGGSGG but with the cyclizing cystines mutated to two alanine. Biolayer interferometry to measure the affinity is performed on the KGGSGG variants by using streptavidin tips to load the peptide and unbiotinylated DnaN in the solution. This confirmed binding of the disulfide version whereas the alanine variant shows no binding in this assay (Figure 13A). In addition to binding, another factor for *in vivo* efficacy is the ability to cross cell membranes. This was not specifically designed for in the process because the design rules for permeability are still under investigation. However, common approaches here usually involved making hydrophobic substitutions[38] as well as N-methylating the backbone amide[39]. To measure the permeability of 53C, a parallel artificial membrane permeability (PAMPA) assay is performed on 53C. 10 uM of 53C is incubated in the donor well separated by hexadecane membrane from an acceptor well for 5 hours. The amount of 53C in the donor and acceptor well is then analyzed using LC-MS. Surprisingly, 53C passively diffuses through the artificially membrane efficiently (Figure 13B). The general hydrophobicity of 53C likely contributes to its

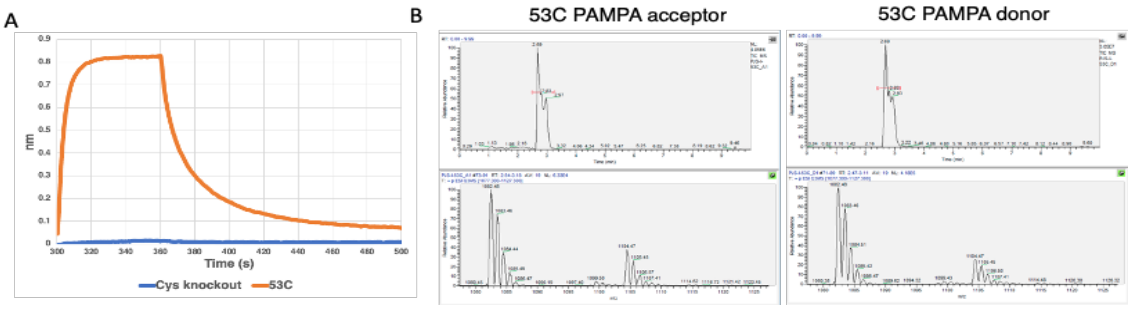



**Figure 13. *In vitro* Characterization of 53C.** (A) Biotinylated 53C (orange) show clear binding signal with biolayer interferometry whereas the disulfide knockout variant shows no binding. (B) 53C passively diffuse through artificially membrane in PAMPA. Top: LC trace of 53 acceptor and donor well. Bottom: MS spectrum confirming 53C identity in each well.

ability to diffuse across membranes. However, 53C's permeability is still unexpected given that it contains a polar glutamine as well as various polar atoms on the peptide backbone.

*High Resolution Structures of 53C Reveals an Alternative Peptide Conformation and Binding Orientation*

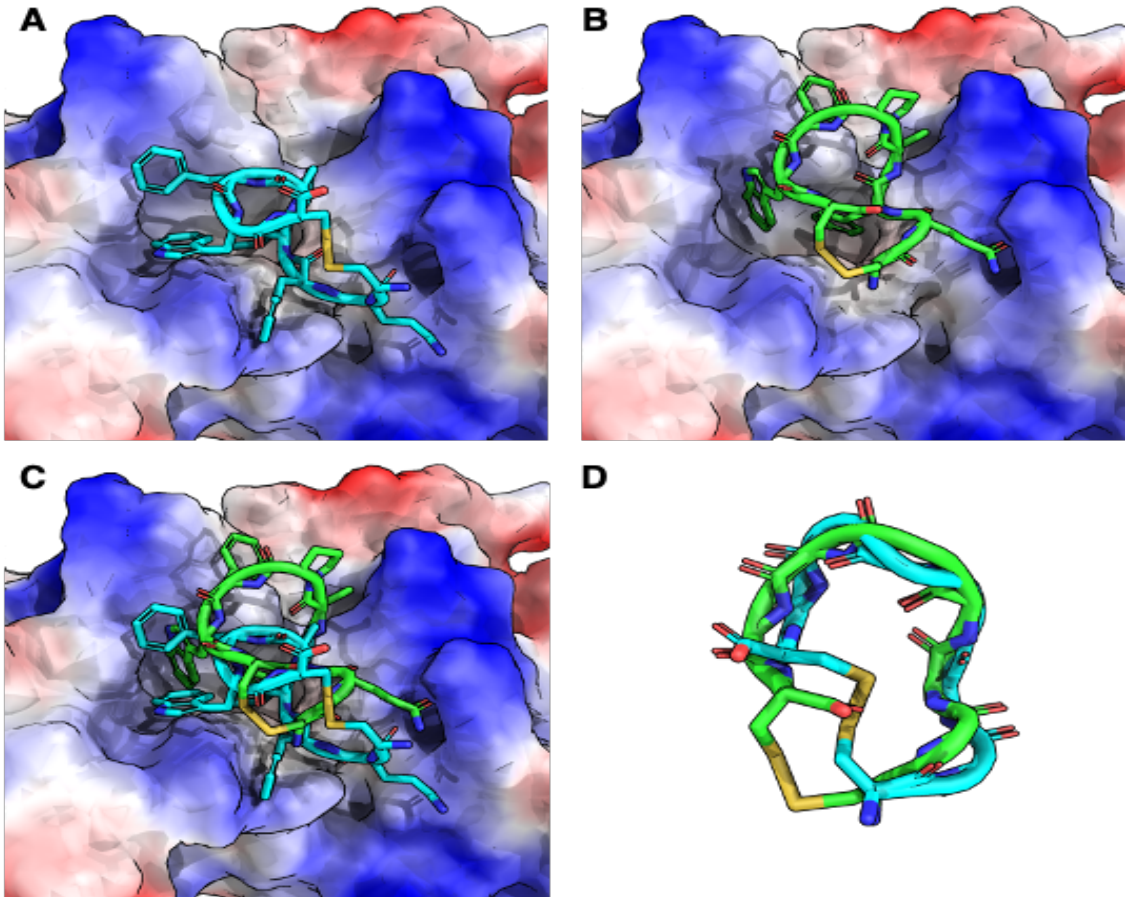



**Figure 15. Crystal structure of 53C in complex with DnaN.** (A) Crystal structure (B) 53C design model (C) Superposition of the model and crystal structure by aligning the DnaN (D) Peptide backbone superposition, RMSD=2.4 Å. Green: Model, Cyan: crystal structure

A 1.6 Å resolution crystal structure of 53C in complex with DnaN (Figure 14A) is obtained by Asim Bera. 53C does bind DnaN at the site of design but both the interaction mode and the peptide conformation are different from the design model (Figure 14C and D). Instead of using phenylalanine at position 3 to occupy the site one pocket, leucine at position 5 is used to reach deeper inside the pocket. Phenylalanine at position 3 instead interacts with DnaN arginine

187 through cation-pi. Arginine 187 is also sandwiched on the other side by tryptophan 7 rested on a shallow pocket next to the deep leucine pocket. The conformation of the peptide is also different from the design model. Instead of being in a positive phi region at position 2 and 4, all positions on the peptide in the structure fall in the negative phi region. The phi and psi angles at position 2 and 7 also cause the disulfide bond to buckle and flip both the termina outward toward the solvent.

Discussion

Although 53C shows binding to DnaN *in vitro* as well as permeability in the PAMPA assay, the alternative binding orientation observed in the crystal structure highlights the challenges in predicting structured loop conformation and its binding orientation. Recent study also stresses the importance of structural rigidity in designing macrocyclic peptide binders and found that rigidity in structures roughly correlated with experimentally determined IC50[40]. Nonetheless, 53C is a valuable starting point for further development as an easy to synthesize, permeable, *M. tuberculosis* antibiotic alternative and suggests possible improvement for future macrocyclic peptide binder design.

Perhaps the most notable difference in the 53C structure is the alternative binding conformation it adopted from the design model (Figure 14). Upon closer inspection, the original design model contains two positions in the positive phi regions that disfavored canonical L amino acids. Because of the steric hindrance from proceeding residue's sidechains, positive phi regions are predominantly occupied by glycine in protein structures. As a result, the original design model is extremely unfavorable in large scale sampling of the peptide conformation (Figure 15A, left panel). Interestingly, if the sequence of 53C is threaded onto a subset of the disulfide macrocycle scaffolds that are pre-filtered to rule out any positive phi containing cycles,

a semi-funnel like landscape emerges where the peptide crystal structure is one of the lowest in total energy (Figure 15A, right panel). Although this does not identify the crystal conformation of 53C *per se*, it does indicate that the peptide binding structure is one of the more favorable conformations in the ensemble if we restrict the possible torsional space. As a confirmation for Rosetta's ability to differentiate docking poses given the correct peptide conformation, the design model and the crystal peptide conformation are taken and docked onto DnaN with PatchDock[41]. The starting positions identified by PatchDock are later refined by RosettaDock[42]. Although both conformations are the lowest poses sampled in this method, the crystal conformation wins out in both the energy gap and the raw dG calculated (Figure 15B).

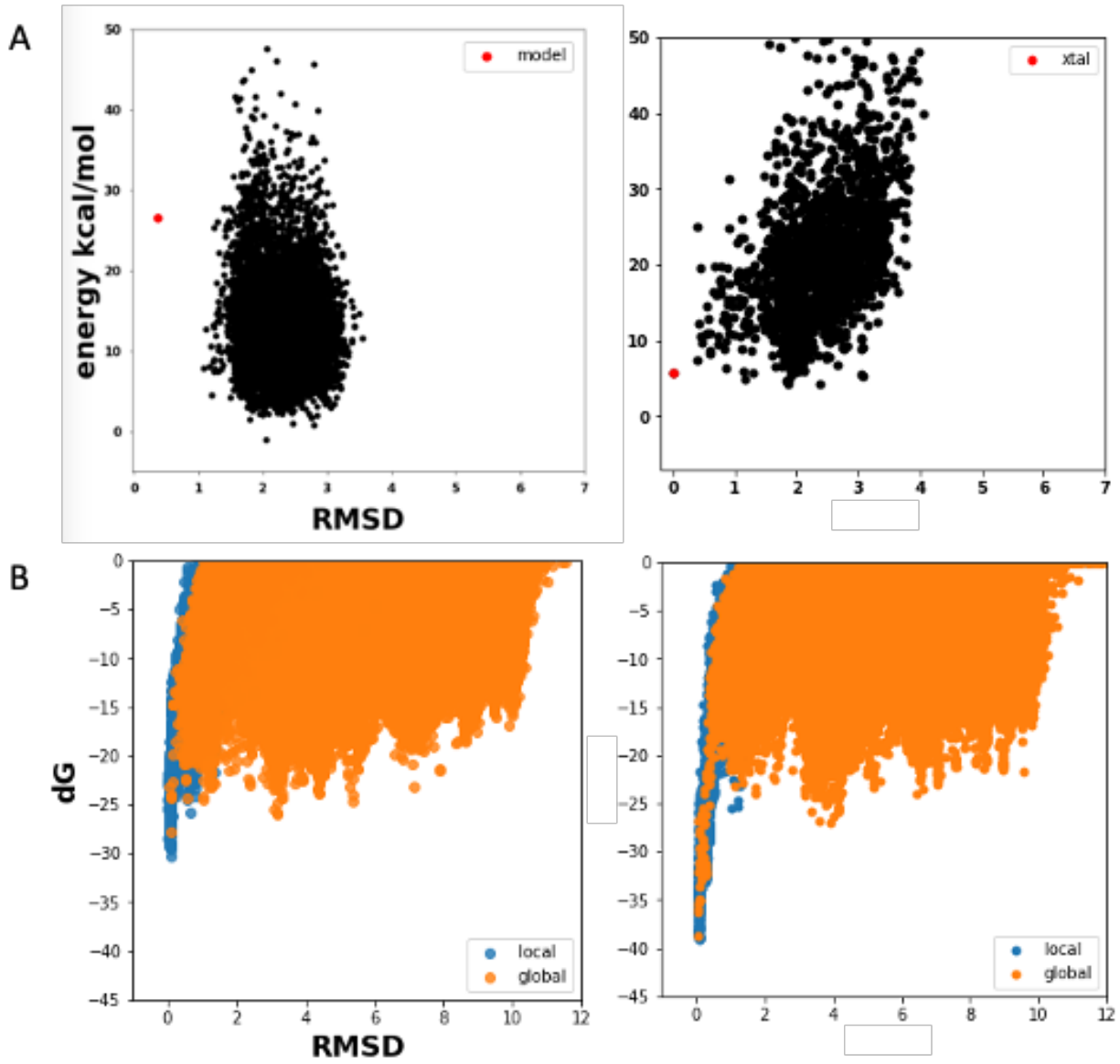



**Figure 15. Structure prediction of 53C conformation and docking using both the model and the crystal as reference.** (A) Sampling of 53C conformation using either the model or the crystal structure as reference. In the right panel, only negative phi region is allowed at all positions. (B) Docking of 53C model vs. crystal structure. Crystal structure has both a more negative dG and a larger energy gap to other decoy poses.

## Conclusion

The ability to engineer proteins that can interact with other molecules is one of the pillars that laid the foundation of much of modern biotechnology. *De novo* protein design in this area holds great promises to generate custom tools for the investigation of relevant biology as well as

potential therapeutics. The design of modular antagonists against CD131 opens up different interesting possibilities that would be difficult to achieve with traditional molecules such as antibodies or peptides. On its own, the superior affinity and stability when coupled with the smaller size made these molecules uniquely suited for targeting local inflammation with better tissue penetration stemming from their size and the fact that systematic administration is not always required. Their simpler architectures also enable facile fusion to different binding domains as well as scaffold molecules for synergistic functions.

There are many potential applications for 933, 3362 and their improved variants. One immediate possibility is the treatment of asthma via pulmonary delivery. There are currently 3 FDA approved anti-IL-5 or IL-5R antibodies for the treatment of eosinophilic asthma[43]. GM-CSF, another cytokine requiring CD131 for signaling, has been shown to promote accumulation of eosinophile in pulmonary allergic inflammation[44]. Targeting both of these cytokines involved in asthmatic inflammation in a modality amenable to local delivery is an attractive strategy for the treatment of asthma.

The unique geometry of the CD131 dodecamer signaling complex also provides another exciting opportunity to probe CD131 mediated cellular phenotype. The fact that GM-CSF exhibits rheostat-like function in deciding cell fates hinted at possible receptor density mediated signaling. F103 and F152 are a starting point in this perspective. By preferentially favoring either dodecamer or hexamer assembly, one can try to elucidate the functional consequences of their arrangement. Although cytokine specific alpha chains for GM-CSF are required for signaling, their mechanistic functions remain elusive. By fusing F103 and F152 to different oligomeric scaffolds, we can artificially boost the local concentrations of CD131 on the surface of cells to see if we can observe different signaling behaviors. Active work is also being done to

incorporate GM-CSF alpha chain interacting motifs to create true agonists in order to elicit direct signaling activation.

While the use of regular secondary structures to mediate protein interactions has been explored before, the design rules of structured loops to accomplish such a task are less well understood. By using macrocyclic peptides as a molecular scaffold, I have tested our ability to design loop mediated interactions against DnaN. Although 53C failed to interact with DnaN in the designed manner, potential improvement can be made by enforcing preferable torsional space based on the type of amino acid stereoisomer in a given position. The fact that 53C binds DnaN *in vitro* and passively diffuses across artificially membrane also provides an excellent starting point for the optimization of an easy to synthesize alternative to Griselimycin. With proper development, 53C laid the groundwork for an antibiotic for drug resistant *M. tuberculosis* infection that the world solely needs.

## Materials and Methods

Computational Software

RIFDOCK – A open source docking algorithm based on geometric hashing of inverse rotamers were used to sample and evaluate docking poses https://github.com/rifdock/rifdock. Custom changes enabling the addition of hotspots as well as pose sampling starting from the hotspot positions were made in the branch https://github.com/rifdock/rifdock/tree/tyu0001/hotspots.

Rosetta[25] – The Rosetta software suite and its python interface pyRosetta were used for Monte Carlo sequence optimization of a given design backbone, the generation of the disulfide macrocycles and GA LigandDocking of the tryptophan and phenylalanine hotspots.

Berkeley Open Infrastructure for Network Computing (BOINC) – A volunteer computing platform is used for intensive *ab initio* fold prediction from fragments or GenKIC as well as Rosetta Docking.

DNAworks[45]– A reverse translation and codon optimization tool for obtaining DNA of the design sequences.

Yeast Surface Display

Oligonucleotide library made from Agilent SurePrint™ containing design sequences as well as common flanking primer annealing sites are amplified and transform with linearized pETCON3 vector into *S. cerevisiae* EBY100 stain according to previously described protocol[23]. For fluorescent activated cell sorting (FACS), library yeast is incubated with biotinylated target protein for 1 hour at 4 degrees rotating end-over-end. The library is washed in PBS + 1% BSA and incubated again with 1:100 dilution of Anti-cMyc-FITC and SAPE. All selection libraries are sequenced with Illumina NextSeq as described previously [46].

Circular Dichroism

All CD experiments were performed on a Jasco J-1100 Spectrometer. For thermal scan, wavelengths at 222 nM were monitored in the process of increasing temperature from 25 degree to 96 degree with every 2 degrees increment when the data is collected. Wavelength scans from 190 to 260 nm were performed before at 25 degree, at 96 degree and again at 25 degree after the sample cooled down back to 25 degree.

Biolayer interferometry

Biolayer interferometry experiments were conducted with the Octet Red96e system. All samples are diluted in HBS-EP+ buffer with 1% non-fat dry milk as blocking reagents. Biotinylated proteins are conjugated onto streptavidin tips online for 180 seconds. Conjugated tips are equilibrated with a buffer for 120 seconds before the on-rate is measured in solution containing various analyte concentrations. Off-rate is measured in the buffer after a saturating signal is observed for the on-rate step. $K_d$ values are acquired using a titration series and the Octet System Software fitting the 1:1 binding model.

Protein Expression

NEB T7 Shuffle Express® cells are used to promote proper disulfide formation in the cytosol. Genes are reverse translated using DNAworks and submitted to Genscript for cloning into the pET32b+ with thioredoxin (Trx) tag and SNAC cleavage sequences. After transformation, started cultures are subcultures into 1 L of TB and expression is performed following the auto-induction protocol[47]. Protein purification is performed with Ni-NTA followed by size exclusion chromatography polishing. If cleavage of the Trx is required, chemical cleavage of the SNAC tag is performed as published[48].

**References**


1. Dougan, M., Dranoff, G. & Dougan, S. K. GM-CSF, IL-3, and IL-5 Family of Cytokines: Regulators of Inflammation. *Immunity* **50,** 796–811 (2019).
2. Bagnasco, D. *et al.* Anti-IL-5 and IL-5Ra: Efficacy and Safety of New Therapeutic Strategies in Severe Uncontrolled Asthma. *BioMed Research International* **2018,** 1–8 (2018).
3. Lang, F. M., Lee, K. M. C., Teijaro, J. R., Becher, B. & Hamilton, J. A. GM-CSF-based treatments in COVID-19: reconciling opposing therapeutic approaches. *Nature Reviews Immunology* 1–8 (2020). doi:10.1038/s41577-020-0357-7
4. Panousis, C. *et al.* CSL311, a novel, potent, therapeutic monoclonal antibody for the treatment of diseases mediated by the common β chain of the IL-3, GM-CSF and IL-5 receptors. *mAbs* **8,** 436–453 (2016).
5. Burmester, G. R. *et al.* A randomised phase IIb study of mavrilimumab, a novel GM-CSF receptor alpha monoclonal antibody, in the treatment of rheumatoid arthritis. *Ann Rheum Dis* **76,** 1020–1030 (2017).
6. MD, Z. T. *et al.* GM-CSF Neutralization With Lenzilumab in Severe COVID-19 Pneumonia: A Case-Cohort Study. *Mayo Clinic Proceedings* **95,** 2382–2394 (2020).
7. Nefla, M. *et al.* The pro-inflammatory cytokine 14-3-3ε is a ligand of CD13 in cartilage. *Journal of Cell Science* **128,** 3250–3262 (2015).
8. Pratt, J. C. *et al.* Evidence for a Physical Association between the Shc-PTB Domain and the beta Chain of the Granulocyte-Macrophage Colony-stimulating Factor Receptor. *Journal of Biological Chemistry* **271,** 12137–12140 (1996).

9. Hercus, T. R. *et al.* The granulocyte-macrophage colony-stimulating factor receptor: linking its structure to cell signaling and its role in disease. *Blood* **114,** 1289–1298 (2009).

10. Hansen, G. *et al.* The Structure of the GM-CSF Receptor Complex Reveals a Distinct Mode of Cytokine Receptor Activation. *Cell* **134,** 496–507 (2008).

11. Matsuguchi, T., Zhao, Y., Lilly, M. B. & Kraft, A. S. The Cytoplasmic Domain of Granulocyte-Macrophage Colony-stimulating Factor (GM-CSF) Receptor Î± Subunit Is Essential for Both GM-CSF-mediated Growth and Differentiation*. *Journal of Biological Chemistry* **272,** 17450–17459 (1997).

12. Barry, S. C. *et al.* Roles of the N and C Terminal Domains of the Interleukin-3 Receptor α Chain in Receptor Function. *Blood* **89,** 842–852 (1997).

13. Guthridge, M. A. *et al.* Site-Specific Serine Phosphorylation of the IL-3 Receptor Is Required for Hemopoietic Cell Survival. *Molecular Cell* **6,** 99–108 (2000).

14. Hercus, T. R. *et al.* Role of the β Common (βc) Family of Cytokines in Health and Disease. *Cold Spring Harb Perspect Biol* **10,** a028514–27 (2018).

15. Morris, N. Th2 cytokines and asthma The role of interleukin-5 in allergic eosinophilic disease. 1–9 (2001).

16. van de Laar, L., Coffer, P. J. & Woltman, A. M. Regulation of dendritic cell development by GM-CSF: molecular control and implications for immune homeostasis and therapy. *Blood* **119,** 3383–3393 (2012).

17. Lantz, C. S. *et al.* Role for interleukin-3 in mast- cell and basophil development and in immunity to parasites. *Nature* **392,** 1–4 (1998).

18. Kilar, C. R. *et al.* Computational design and experimental characterization of a novel β-common receptor inhibitory peptide. *Peptides* **104,** 1–6 (2018).

19. Leyva-Grado, V. H., Tan, G. S., Leon, P. E., Yondola, M. & Palese, P. Direct Administration in the Respiratory Tract Improves Efficacy of Broadly Neutralizing Anti-Influenza Virus Monoclonal Antibodies. *Antimicrobial Agents and Chemotherapy* **59,** 4162–4172 (2015).

20. Broughton, S. E. *et al.* Conformational Changes in the GM-CSF Receptor Suggest a Molecular Mechanism for Affinity Conversion and Receptor Signaling. *Structure/Folding and Design* **24,** 1271–1281 (2016).

21. Lopez, A. F. *et al.* Residue 21 of human granulocyte-macrophage colony-stimulating factor is critical for biological activity and for high but not low affinity binding. *EMBO J* **11,** 909–916 (1992).

22. Correia, B. E. *et al.* Computational Design of Epitope-Scaffolds Allows Induction of Antibodies Specific for a Poorly Immunogenic HIV Vaccine Epitope. *Structure/Folding and Design* **18,** 1116–1126 (2010).

23. Rocklin, G. J. *et al.* Global analysis of protein folding using massively parallel design, synthesis, and testing. *Science* 1–8 (2017).

24. Dou, J. *et al.* De novo design of a fluorescence-activating β-barrel. *Nature* 1–30 (2018). doi:10.1038/s41586-018-0509-0

25. Leaver-Fay, A. *et al.* ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol* **487,** 545–574 (2011).

26. Regan, L. & DeGrado, W. F. Characterization of a helical protein designed from first principles. *Science* **241,** 976–978 (1988).

27. Kuhlman, B. *et al.* Design of a Novel Globular Protein Fold with Atomic-Level Accuracy. *Science* **302,** 1364–1369 (2003).

28. Koga, N. *et al.* Principles for designing ideal protein structures. *Nature* 1–8 (2012). doi:10.1038/nature11600

29. Bartlett, G. J., Porter, C. T., Borkakoti, N. & Thornton, J. M. Analysis of catalytic residues in enzyme active sites. *Journal of Molecular Biology* **324,** 105–121 (2002).

30. Kundert, K. & Kortemme, T. Computational design of structured loops for new protein functions. *Biol Chem* **400,** 275–288 (2019).

31. Luther, A., Bisang, C. & Obrecht, D. Advances in macrocyclic peptide-based antibiotics. *Bioorg Med Chem* **26,** 2850–2858 (2018).

32. Yan, P. *et al.* The immunosuppressant cyclosporin A antagonizes human formyl peptide receptor through inhibition of cognate ligand binding. *J Immunol* **177,** 7050–7058 (2006).

33. Bhardwaj, G. *et al.* Accurate de novo design of hyperstable constrained peptides. *Nature* **538,** 329–335 (2016).

34. Hosseinzadeh, P. *et al.* Comprehensive computational design of ordered peptide macrocycles. *Science* **358,** 1461–1466 (2017).

35. Altieri, A. S. DNA Sliding Clamps as Therapeutic Targets. 1–9 (2018). doi:10.3389/fmolb.2018.00087

36. Terlain, B. & Thomas, J. P. Structure of griselimycin, polypeptide antibiotic extracted Streptomyces cultures. I. Identification of the products liberated by hydrolysis. *Bull Soc Chim Fr* **6,** 2349–2356 (1971).

37. Kling, A. *et al.* Targeting DnaN for tuberculosis therapy using novel griselimycins. *Science* **348,** 1106 (2015).

38. Madani, F., Lindberg, S., Langel, Ü., Futaki, S. & Gräslund, A. Mechanisms of Cellular Uptake of Cell-Penetrating Peptides. *Journal of Biophysics* **2011,** 414729 (2011).

39. White, T. R. *et al.* On-resin N-methylation of cyclic peptides for discovery of orally bioavailable scaffolds. *Nat Chem Biol* **7,** 810–817 (2011).

40. Mulligan, V. K. *et al.* Computationally designed peptide macrocycle inhibitors of New Delhi metallo-β-lactamase 1. *Proc Natl Acad Sci USA* **118,** e2012800118 (2021).

41. Duhovny, D., Nussinov, R. & Wolfson, H. J. Efficient Unbound Docking of Rigid Molecules. in (eds. Guigó, R. & Gusfield, D.) 185–200 (Springer Berlin Heidelberg, 2002).

42. Gray, J. J. *et al.* Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. *Journal of Molecular Biology* **331,** 281–299 (2003).

43. Ridolo, E. *et al.* Mabs for treating asthma: omalizumab, mepolizumab, reslizumab, benralizumab, dupilumab. *Hum Vaccin Immunother* **16,** 2349–2356 (2020).

44. Nobs, S. P., Kayhan, M. & Kopf, M. GM-CSF intrinsically controls eosinophil accumulation in the setting of allergic airway inflammation. *Journal of Allergy and Clinical Immunology* **143,** 1513–1524.e2 (2019).

45. Hoover, D. M. & Lubkowski, J. DNAWorks: an automated method for designing oligonucleotides for PCR-based gene synthesis. *Nucleic Acids Res* **30,** e43–e43 (2002).

46. Chevalier, A. *et al.* Massively parallel de novo protein design for targeted therapeutics. *Nature* **550,** 74–79 (2017).

47. Studier, F. W. Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* **41,** 207–234 (2005).

48. Dang, B. *et al.* SNAC-tag for sequence-specific chemical protein cleavage. *Nature Methods* **16,** 319–322 (2019).