DE NOVO COMPUTATIONAL DESIGN OF BONE MORPHOGENETIC PROTEIN-2 KNUCKLE BINDERS

by

BENJAMIN J BURRESS-IRVING

A THESIS

Presented to the Department of Chemistry and Biochemistry
and the Robert D. Clark Honors College
in partial fulfillment of the requirements for the degree of
Bachelor of Science

May 2024

An Abstract of the Thesis of

Benjamin J Burress-Irving for the degree of Bachelor of Science
in the Department of Biochemistry to be taken June 2024

Title: *De Novo* Computational Design of Bone Morphogenetic Protein-2 Knuckle Binders

Approved: *Dr. Parisa Hosseinzadeh*
Primary Thesis Advisor

Acute or long-term symptoms of a bone fracture have increased globally at a rate of 70.1% from 1990-2019.[1] Bone morphogenetic protein 2 (BMP-2) is an important driving force in osteogenesis and could be an essential therapeutic for mitigating long-term symptoms associated with nonunion fractures. Recombinant human BMP-2 (rhBMP2) has been approved by the Food and Drug Administration (FDA) as a growth factor for therapeutic use against subtypes of nonunion fractures, but research has shown that required supraphysiological amounts from predominant BMP-2 delivery methods lead to overgrowth of bone among other adverse side effects. Therefore, there is a current heightened research interest on controlling the release kinetics of BMP-2 into a fracture site for safer and more efficient bone regeneration.

Computational protein design (CPD) is a promising technique for creating *de novo* binders to BMP-2. Engineered protein-protein interactions can be utilized to participate in a lower risk, affinity-modulated, delivery system to increase the efficacy and decrease adverse side effects of BMP-2. Using PyRosetta[3] design scripts, and the UO Talapas and Franklin computers, I produced several potential *de novo* protein binding candidates. I have experimentally tested and validated the top designs while increasing the rate of success with RFDiffusion[4] and ProteinMPNN[5] computational techniques. This thesis overviews the successful *de novo* design of a protein binder with an equilibrium dissociation constant of 771 nM as measured by biolayer interferometry.

Acknowledgements

I would like to thank my faculty mentor, Parisa Hosseinzadeh, for her guidance and persistence on this project. I would also like to thank Cassandra Gonzalez and Kevin Harnden for their help and dedication, as well as Karly Fear for laying the foundation of our lab's work on BMP-2. Every member of Parisa's diverse and interdisciplinary lab helped contribute to an excellent working environment.

I would like to thank the Knight Campus Undergraduate Scholars Program for helping fund this project as well as the overall Knight Campus community for giving me the opportunity to do research. I would also like to thank my CHC representative Dare Baldwin. Finally, I would like to thank my family for supporting me always.

Table of Contents

List of Figures

Introduction

Nonunion bone fractures

Bones provide shape, support, and organ protection. When bones fracture there is an increased likelihood of health risks, decreased mobility, and discomfort. The FDA defines a fracture as 'nonunion' when it "persists for a minimum of nine months without signs of healing for three months." [2]

In the U.S., about 100,000 fractures go into nonunion yearly. The rate of all fracture nonunion is between 1.9% to 10% with variable rates depending on the anatomical region. Nonunion fractures can be classified into four categories, hypertrophic nonunion, atrophic nonunion, oligotrophic, and septic nonunion. [2]

Atrophic nonunion fractures are especially difficult to treat as they require both biological and mechanical support. Past methods used bone stimulator devices or, more commonly, surgical treatment. The historical option for surgical treatment is bone grafting. Bone grafting uses transplanted bone fragments to rebuild damaged bone by harvesting from another area of the body. However, this method requires an additional operation and will often make recovery more painful with additional surgery sites. It is also not ideal for the rebuilding of long bones, such as the femur.

An alternative to bone grafting is to deliver BMP-2 to induce osteogenesis, which circumvents the pain and surgery associated with traditional bone grafting. This method has many advantages over traditional bone grafting. However, if careful consideration is not put into BMP-2 release, there can be adverse side effects, such as overgrowth of bone from burst release and ectopic bone formation.

Bone Morphogenetic Protein-2

Proteins are large, complex molecules composed of amino acids and are necessary for the body to function. BMPs are a class of proteins in the Transforming Growth Factor-Beta (TGF-β) superfamily of cytokines and are essential to many developmental processes including osteogenesis, apoptosis and bone homeostasis and survivability. BMP-2 was the first BMP to be identified, by Marshall R. Urist [9] in 1965, and has been FDA approved for usage as a therapeutic for tibial nonunions since 2004. [6]

BMP-2 is a key player in bone healing. It promotes differentiation of mesenchymal stem cells into osteoblasts, which act to form new bone and heal bone fractures. Studies have confirmed that the Smad signaling pathway is essential for osteogenic differentiation [7]. Smads are intracellular effectors, acting as transcription factors which mediate signal transduction in the BMP-2 signaling pathway. BMP-2 binds to two receptors on the cell surface, BMPR-II and BMPR-I, which are each specific to a different binding site on BMP-2. BMPR-II is specific to the knuckle site, which consists primarily of beta sheets and BMPR-I is specific to the alpha helical wrist site. BMP-2 can be inhibited by the extracellular cysteine knot protein Noggin, regulating its activity, and this plays a key role in vertebrate dorsoventral patterning.

While BMP-2 has therapeutic potential to be effective in promoting bone growth within defects, its current clinical use has relied on sub-optimal delivery systems involving collagen sponges. These systems release BMP-2 within a few days of implantation and deliver milligram-sized quantities of BMP-2, which are orders of magnitude higher than physiological concentrations [8]. From these methods, BMP-2 has been linked to adverse effects from burst release, such as an overabundance of bone growth. Therefore, there has been heightened research interest on delivering BMP-2 in a controlled manner while retaining its activity. Establishing a

modular system to deliver BMP-2 with a sustained release profile is a key challenge for increasing the safety and efficacy of BMP-2 as a bone regeneration treatment.

More recent techniques to improve the efficacy of BMP-2 and reduce side effects include core-shell nanoparticle nanofiber scaffolds, cryelectrospun mats, microspheres embedded in hydrogels, among others. [8] While these techniques have shown levels of success, there has been little research on how *de novo* protein design can play a role in BMP-2 delivery. The overarching goal of this thesis will be to mitigate the problems associated with burst release of BMP-2 by designing a protein binder which can better control the rate of release through its binding affinity, and which can participate in a larger hydrogel delivery system.

Computational Protein Design

CPD is a field that has grown rapidly in the 21st century. This growth has been sparked by a boom in machine learning tools and has led to numerous breakthroughs and techniques with increasing performance and reliability. Scientists in the field have declared our current time as a 'protein design revolution.' [10]

The applications of CPD are highly varied and cross-disciplinary. It can be used in medicine to develop therapeutics, bind to cancer cells, or inhibit viral infection such as SARS-CoV-2. CPD can also be used for environmental sustainability purposes such as creating better biodegradable materials, sequestering carbon from the atmosphere, or breaking down pollutants. Peptide-based inhibitors, mini-protein binders, decoys, and biosensors can all be created through these methods. [11]

Advanced software such as the Rosetta suite enables protein design to be easily accessible to students and researchers. This suite includes algorithms for computational modeling and analysis of protein structures and was first developed in the laboratory of Dr. David Baker at

the University of Washington as a structure prediction tool but has been adapted for a wider range of functions.

In my project, I will be applying the Python-based Rosetta interface PyRosetta to design protein binders to BMP-2, and after they pass my constraints, those designs will be experimentally validated, and their binding affinity will be tested. The goal is to design proteins that bind with moderate affinity, and therefore will release at a rate that does not result in burst release of BMP-2.

Although other methods of developing binders to BMP-2, like directed evolution, may achieve the same goal, there are merits to using *de novo* protein design to create a protein binder. Jane and David Richardson, pioneers in enhancing our understanding of protein structure, provide one reason for this in their 1989 review: "Just as by designing and building a house you will learn many things that you would never have found out just from living in [one] … so also by designing and building our own new proteins we can expect to learn things we would not have learned just by studying natural proteins." [12] The process of designing a new protein-protein interaction has the potential to provide deeper insights into BMP-2 binding, and possibly to beta-sheet binding in general.

Exposed curved beta sheet design has often been ignored because of the associated difficulty arising from their three-dimensional structure. Because of this, most research has been on ideal backbone structures such as straight helices, uniform beta-strands, and short loops. The BMP-2 knuckle is a curved beta-sheet binding site. Therefore, a less conventional method for binder design is presented in this thesis. This includes generating a *de novo* peptide strand next to the knuckle site of BMP-2 and building a final protein design around the initial strand by grafting it onto stable protein scaffolds, then validating these designs *in silico* with computational

protocols such as docking and protein structure prediction. After these steps the proteins are expressed in E. coli and verified experimentally *in vitro* through bio-layer interferometry. There has been improved success implementing additional methods such as the partial diffusion code of RFDiffusion [7] and ProteinMPNN-FastRelax [13].

Methods

BMP-2 Protein Modeling and Preparation

Before computational design, the correct structure of BMP-2 must be obtained from the online Protein Data Bank (PDB). Using the molecular visualization system PyMol, I obtained PDB ID 4UI1 -- the crystal structure of the human RGMC-BMP2 complex. The RGMC component, water solvent, and any other small molecules that are not included in the BMP-2 structure were removed so that only BMP-2 is being designed for.

For computational procedures, I used Rosetta software, specifically the Python port known as PyRosetta, which includes algorithms for computational modeling and analysis of protein structures. I utilized the University of Oregon's HPC cluster, Talapas, which consists of approximately 9,500 cores, 90 TB memory, 120 GPUs and over 2 PB storage. I also used our own lab's computer named Franklin with 2 GPUs.

After cleanup of the BMP-2 structure from the PDB, I ran RosettaRelax protocol. This helps to make the computational structure more consistent with the true protein structure by packing (sidechain rotamer movements) and minimization (backbone movements). It also automatically adds any hydrogen atoms which might be missing from the structure.

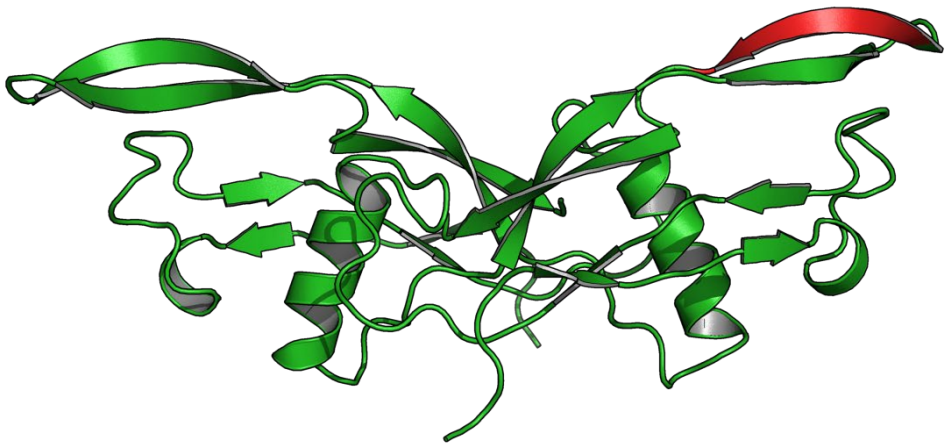

Figure 1: PDB structure of BMP-2 after RosettaRelax.

The chosen edge, the “knuckle” site, for sheet generation is highlighted in red.

Motif-grafting Design Pathway

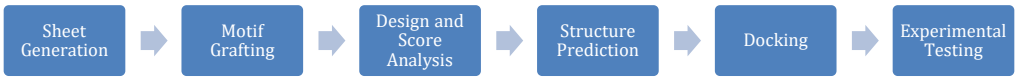

Figure 2: Steps of the motif grafting design pathway.

Beginning with sheet generation, the beta sheet is stabilized via motif grafting and undergoes design to optimize energetic stability, then goes through docking and structure prediction as final filtering steps before experimental testing.

Sheet Generation

Sheet generation (or sheet extension) is a technique which computationally generates many small sheets (peptide strands consisting of 3-4 amino acids) next to the binding site of interest. This code is borrowed from Sahtoe et al. In their 2021 publication, they reason that designing edge beta strands complementary in shape to an exposed beta strand – like the BMP-2 knuckle – can overcome challenges of beta sheet design such as failing to fully consider polar interactions and only focusing on sidechain-sidechain interactions.[14] Sheet generation makes it easier to engage the many exposed C=O and N-H groups at the edges of beta sheets.

I specified the binding edge as residues 192-196 in the PDB file. After these steps, the generated strands were also relaxed via Rosetta Relax. Sheets failing a specified threshold value of -2.5 for Rosetta hydrogen bond score were removed.

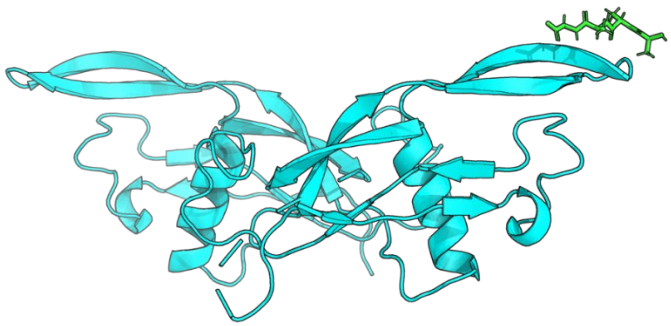

Figure 3: Sheet generation.

A generated strand showing visualized on PyMol.

Motif Grafting

A motif is a distinctive sequence on the protein with a three-dimensional structure that allows binding interactions to occur. Motif grafting is a tool in Rosetta which attaches a motif and a separate protein component together. In my case, I used it to stabilize my strands by adding them to a stable scaffold protein.[15] There are three important components of motif grafting (1) the motif (2) the context and (3) the scaffold. The motifs were the strands generated from sheet extension, the context was the BMP-2 protein, and the scaffolds were proteins with a high verified stability borrowed from Rocklin et al, Science 2017[16] and Sahtoe et al, PNAS 2021. The goal of motif grafting is to find fragments in the Pose – Rosetta's internal representation of a protein structure – which are compatible with the motifs and replace those fragments with the motifs. The algorithm uses three user-definable cutoffs to determine compatibility which are the

root means square deviation (RMSD) of the fragment alignment, RMSD of the N-/C- points after the alignment and clash_score.

I cut the RMSD tolerance off at 1.5 and clash_score – the maximum number of atomic clashes that are tolerated – at 5. I used full motif alignment which uses the full backbone structure of the fragments.

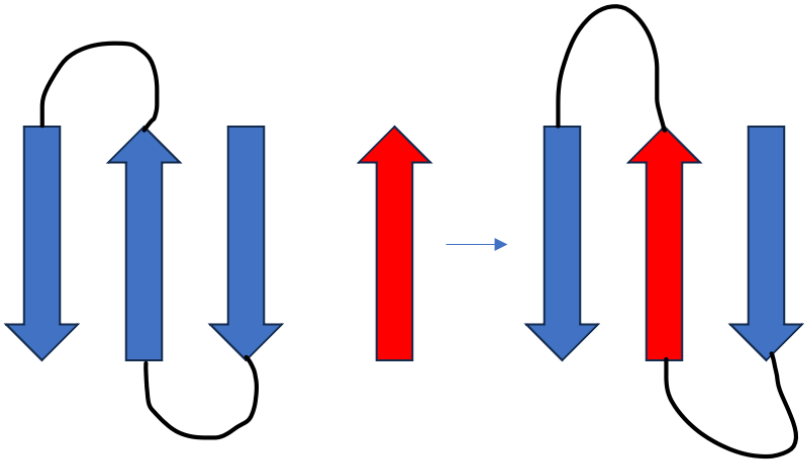

Figure 4: Simplistic representation of motif grafting.

General idea behind motif grafting. A single strand is stabilized with validated stable scaffolds. This process occurs in tandem with BMP-2, and the sheet is still oriented on the knuckle (the "context") as the strand is grafted.

Design and Score Analysis

Motif grafting only performs the graft of the fragments, so it must be followed by design and minimization/repacking steps for the design to be viable. After design, there were several hundred potential candidates which were filtered down by the next steps: score analysis, structure prediction, and docking.

Score analysis allows filtration via several metrics. The lower the score, the more energetically favorable or stable the protein is. Some important factors I considered were hydrogen bonding and the score at the interface between the two binding partners.

After this step, I have a roster of potential proteins binding to the knuckle site of BMP-2. The goal of these filtration steps is to get a list of fewer than 100 potential candidates to inspect visually via PyMol. After visual inspection, these were filtered down further to 5 of the best designs. The next steps, docking and structure prediction, helped to secure the top binding candidates.

Docking

Docking uses Rosetta's docking app to compare interface score and alpha-carbon RMSD of the ligand.

RosettaDock is a Monte Carlo (MC) based multi-scale docking algorithm that incorporates both a low-resolution, centroid-mode, coarse-grain stage and a high-resolution, all-atom refinement stage that optimizes both rigid-body orientation and side-chain conformations. The algorithm I used, global docking, started from a random initial orientation of the two partners. From there, the partner proteins are represented coarsely, where side chains are replaced by a single unified pseudo-atom, or centroid.[17]

I utilized global docking as an important filtering technique before ordering any protein candidate for experimental testing.

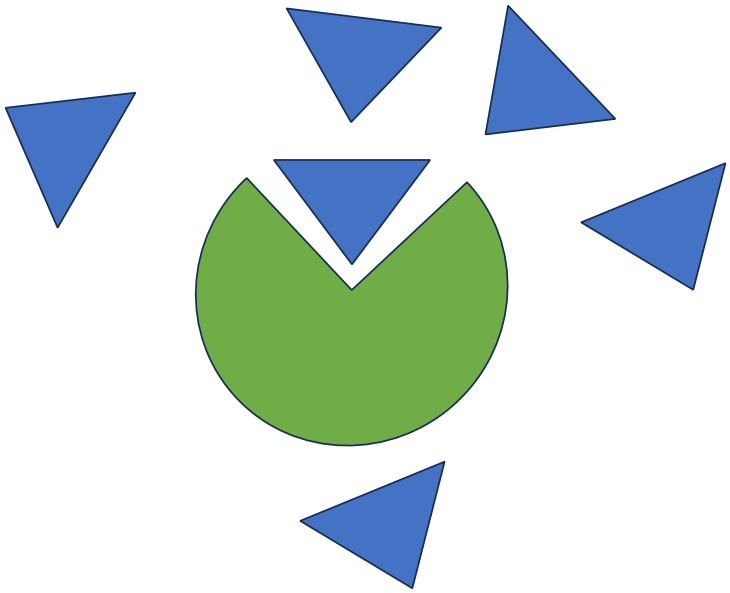

Figure 5: Simplistic representation of docking.

Binder generated decoys (blue) docking onto BMP-2 (green).

Structure Prediction

Structure predication is a computational technique which uses machine learning to output a protein structure from an inputted amino acid sequence. The software that I used for this were RoseTTAfold [18] and AlphaFold 2 [19]. This an important step because it increases the odds that a computer-generated structure will match with the real experimental folded structure. This is a visual inspection step, meaning that if there is no alignment on PyMol then the structure is discarded. AlphaFold 2 also outputs the predicted local distance difference test (pLDDT), an important metric for a measure of confidence within the model.

AlphaFold Multimer, which provides a prediction from the amino acid sequence of both target and binder in complex, was also utilized. Importantly, a metric generated unique to the multimer model is predicted aligned error (pAE) of interaction. Recent studies [13] have found that a pAE_interaction score <10 leads to a higher likelihood of *de novo* binder success. This cutoff was implemented and was important for design success. However, AlphaFold Multimer

was not implemented prior to ordering my first top 4 designs, so the pAE_interaction of these designs was not quantified until before my second batch of designs was ordered.

RFDesign/Hallucination

RFDesign is another method of computational design which bypasses many motif-grafting steps. [20] This carries out gradient descent in sequence space to optimize a loss function. In this method, I started from a desired motif – not a generated sheet, but two sheets from a native protein binder to BMP-2. These two sheets were then brought through a loss function and new scaffolds were generated. After several designs were made, they were filtered via RMSD. RFDesign internally verifies its designs via RoseTTAfold, but I used AlphaFold 2 to confirm them and the Rosetta docking protocol as mentioned previously. RFDesign had low success rates because although many of the designs produced folded, all of them failed to dock. With more time, I would have revisited this method, but I later found that RFDiffusion worked better for my applications.

ProteinMPNN

ProteinMPNN is a message passing neural network (MPNN) and deep learning-based protein sequence sampling and design method with outstanding performance *in silico* and in experimental settings. For my applications, I found success using ProteinMPNN to assign a sequence to proteins generated through RFDiffusion and to help improve my initial designs after partial diffusion.

RFDiffusion

RFDiffusion uses deep learning to fine tune RoseTTAfold on protein structure denoising tasks to generate new binders. After running this protocol, a protein backbone is generated without an assigned sequence. To assign a sequence I used ProteinMPNN-FastRelax.

RFDiffusion can be run with a scaffold or no scaffold, and either partial or full diffusion. I tested both with a scaffold and without a scaffold along with bringing my original designs through partial diffusion. The scaffold that I chose was from the native binder used in RFDesign and the non-scaffold option used the relaxed BMP-2 input.

Filtering steps come from AlphaFold 2, specifically with a pAE_interaction < 10. After filtering these, I ran them through Rosetta docking and chose the best ones for experimental validation.

After testing multiple RFDiffusion runs, I decided to choose designs with a high probability of success suggested by metrics such as pae_interaction, pLDDT, and RMSD. These designs were partial diffusion 'rescues' of my initial designs from the motif-grafting pathway, which was needed after my first batch failed to give evidence of binding in vitro. I started with small test runs outputting 100 generated backbones and 2 MPNN sequences per backbone, and then scaled up to 1000 backbones with 1 MPNN sequence.

Experimental Methods

The top protein binders were ordered as genes and expressed via the central dogma of biology: DNA → RNA → protein. This was accomplished using gene fragments encoding our designed proteins, either gBlock or eBlock that were then inserted into the plasmid vector pET28a(+) via Gibson assembly. Research has shown that each organism has preference for certain DNA codons, and the process of matching codons to an organism-of-interest is known as

codon optimization. For this I did two optimizations for tests in yeast and E. coli using the IDT website. The overhang melting temperature was optimized to ~60 C, to increase the efficiency in the gene fragment insertion.

Gibson Assembly

Gibson assembly [21] was used to insert the gene fragments into the pET28a(+) vector. Gibson assembly allows DNA fragments to be efficiently joined using three enzyme specificities: (i) the T5 exonuclease activity which chews back the 5' ends of the DNA fragments and exposes overhangs which can anneal to their complement (ii) Taq DNA ligase activity which fills the gaps of the annealed products and (iii) DNA polymerase which covalently seals resulting nicks in the assembly. All three enzymes are combined into the Gibson Assembly Master Mix. Linearized plasmid and insert fragments are added together along with the master mix and incubated at 50ºC for 1 hour. The resulting DNA can be used in subsequent transformation with heat shock treatment.

Experimental steps are as follows (1) vector linearization using restriction enzyme digestion or Q5 PCR (2) assembly of gene onto vector (3) transformation using heat shock treatment (4) plasmid extraction and sequencing. The two restrictions sites for plasmid linearization were the NcoI and XhoI cut sites on pET28a(+).

The Gibson assembly reaction had a total volume of 20 uL with components including Gibson Assembly Master Mix, 0.01-0.5 pmols of eblock or gblock DNA inserts, 0.01-0.5 pmol pET28a(+) plasmid, and deionized water.

Heat-Shock Transformation

NEB Turbo Competent *E. coli* cells were thawed on ice for 10 minutes for the insertion of assembled plasmid. 5 uL of plasmid DNA was pipetted into the cell mixture, and the tube was

flicked 4-5 times to mix cells and DNA. The mixture was placed on ice for 30 minutes. Heat shock occurred at 42C for exactly 30 seconds using a heat bath. Then 950 uL of room temperature Super Optimal broth with Catabolite repression (SOC) was pipetted into the mixture. This mixture was then shaken vigorously at 37C for 60 minutes. During this time selection plates containing Luria-Bertani (LB) media and kanamycin antibiotic were also warmed to 37C to increase transformation efficiency and kill off cells that did not endocytose the kanamycin-resistant plasmid. 100 uL of the cells were spread onto these selection plates and incubated at 30C for 16 hours. The next day, the plates were retrieved from the incubator and colonies were observable indicating success of transformation.

Q5 PCR

I used Q5 high fidelity DNA polyermase to add extra overhangs to the plasmid when heat-shock transformation failed to result in uptake of the assembled plasmid due to an error in the primer sequence.

All reaction components were assembled on ice and mixed prior to use. 25 uL Q5 Master Mix including Q5 High-Fidelity DNA Polymerase, dNTPs, and Mg^{2+} was mixed with 2.5 uL of the 10 uM forward primer, 2.5 uL of the 10 uM reverse primer, 0.3 uL of the 10 ng linearized plasmid DNA and 19.7 uL of water. The tube was gently mixed via flicking and quickly spun down. Thermocycler Q5 protocol was set to 30 seconds at 98 C, 25-35 cycles at 98C, the 72C melting temperature ™ for 2 minutes and held at 4-10 C. Specific reaction conditions were determined through experimental testing and running of a DNA gel.

Plasmid Extraction

Plasmid extraction is a technique used for exocytosis of plasmid DNA and separation from other components of the E. coli cell.

First a bacterial cell culture was prepared with 5 mL LB, 5 uL kanamycin and one colony from each plate. This was shaken at 37C for ~16 hours. The resulting mixture the next day was cloudy compared to the initial clear mixture.

The bacterial culture was centrifuged at 4500 rpm for 10 minutes and the supernatant containing extraneous DNA and RNA was removed. Then 200 uL of red P1 resuspension buffer was added to the tube and resuspended by vortexing. After this step, the mixture was transferred to 1.5 mL tubes. 200 uL blue P2 buffer was then added to lyse the bacterial cells, mixed by inverting the tubes 4 times and incubated at room temperature for 2 minutes. 400 uL yellow P3 buffer was added to neutralize the buffer, and the buffer solution turned fully yellow when the neutralization was complete. This was mixed thoroughly and carefully by inverting and incubated at room temperature for 2 minutes.

The samples were then centrifuged at 13000 rcf for 10 minutes to pellet. A collection column was then placed in a collection tube and supernatant was transferred from step 5 into the column. No pellet was included from the collection tube. This was incubated at room temperature for 5-10 minutes. The collection assembly was then centrifuged at 13000 rcf for 30 seconds. The flow-through was discarded and the column was returned to the same tube. 400 uL of wash buffer was added to the column and centrifuged for 1 minute at 13000 rcf and repeated once empty. The column was transferred to a clean 1.5 mL tube and 30 uL of DNA elution buffer was added directly to the column matrix and incubated for 5-10 minutes at room temperature. Finally, the column was centrifuged for 30s at 13000 rcf to elute the plasmid DNA. The plasmids (between 66 and 184 ng/uL) were then sent to PlasmidSaurus for sequencing.

Plasmids were then transformed into BL21(DE3) for growth and expression.

Growth, Expression and Protein Purification

After sequencing, my protein designs were expressed in E. coli and purified using immobilized metal affinity chromatography. This is a way to filter out proteins with a poly-histidine tag with a high affinity for nickel.

I obtained glycerol cell stocks from -80C and placed them on ice. After a few minutes, a few cells were transferred to one corner of the plate. Using a small spreader, made with flame and a glass pipette, cells were streaked from where the culture was dropped via dilution streaking. The plate was then placed at 37C overnight.

50 mL of LB media were added to two 250 mL baffled flasks prepared with 0.5 g Tryptone, 0.5 g NaCl and 0.25 g yeast extract. This mixture was then gently stirred on a stir plate and then autoclaved with our lab's liquid autoclave protocol to sterilize. On the same day, one colony from the streak plate was chosen and placed into 5 mL LB with kanamycin. This was then shaken and incubated overnight at 37C.

The starter culture was cloudy when retrieved from the shaker, then 5 uL of antibiotic was added. To each of the 50 mL of media, 100 uL of starter culture was added, and this was shaken at 37C for 3 hours. Then, the optical density (OD) was checked using a NanoDrop. When the OD reached 0.6-0.8, the cells were ready for induction. After reaching the correct OD, 200 uL of the culture were taken out for the gel and labeled, then isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to a final concentration of 1 mM to induce expression and moved to 30C overnight.

For protein purification, the resuspension buffer was 50 mM Tris pH 8 + 100 mM NaCl. For designs that had an isoelectric point (pI) between 6-9, HEPES pH 6 + 100 mM NaCl was used as a resuspension buffer. Subsequent wash and elution buffers were made including

resuspension buffer + 50 mM imidazole (wash 1), resuspension buffer + 100 mM imidazole (wash 2), resuspension buffer + 250 mM imidazole (elution 1), and resuspension buffer + 400 mM imidazole (elution 2).

The next day, I took the cells out and poured them into 50 mL falcon tubes. Then I took a 200 uL sample, centrifuged the rest at 10,000 rpm for 10 minutes, decanted the supernatant into a beaker, bleached and discarded the waste. Afterward I stored the pellet at -20C.

When I was ready for purification, I obtained the frozen pellet from -20C. For every 50 mL of culture, I added 15-20 mL of my resuspension buffer and resuspended by vortexing. After resuspension, for every 50 mL of culture, I added 250 uL BugBuster to lyse the cells. I shook the cells at room temperature for 30 minutes. I then balanced the falcon tubes and centrifuged them at 10,000 rpm for 20-30 minutes. Then I took the cells out and put them on ice, then gently poured the supernatant into a clean 50 mL falcon tube. I then took a pipette tip of the pellet and put it into an Eppendorf tube and labeled it. The rest of the pellets were kept at -20C.

For metal chromatography, I assembled a column, shook cleaned Ni beads, and poured ~7 mL of the slurry into the column. I let this drip until the beads were fully packed. I then added 20 mL of Milli-Q water to the column and let it drip to remove ethanol. Then I added 20 mL of the resuspension buffer to the column. I then poured my entire supernatant into the column. The flow-through, wash 1, wash 2, elution 1, and elution 2 samples were collected into new tubes by pouring 20 mL of the first two and 10 mL of the second two buffers into the column. To confirm that my designed protein was successfully isolated, I ran an SDS-PAGE gel on the following samples: flow-through, wash 1, wash 2, elution 1, elution 2, pellet, before IPTG, and after expression samples.

MALDI

These designs were further validated using Matrix Assisted Laser Desorption/Ionization (MALDI) as a method of mass spectrometry. MALDI works by mixing a biomolecule sample with a matrix material to facilitate desorption and ionization of the analyte. A pulsed laser is then focused onto the sample-matrix mixture which the matrix absorbs and then transfers the energy to the analyte molecules. Then the ions are accelerated by an electric field and are separated by their mass to charge (m/z) ratio by a time-of-flight (TOF) detector, which shows the time for ions to travel a certain distance.

Bio-Layer Interferometry

Bio-Layer Interferometry (BLI) is a binding assay and determines the dissociation constant (KD) of binding. For this application we used a GatorBio BLI analyzer.

$$k_D = \frac{k_{off}}{k_{on}} = \frac{[A][B]}{[AB]}$$

BLI uses an optical biosensor with a metal binding platform and relies on detection of interference between light waves as they pass through the sample. Two paths of light are used to detect any interaction at the sensor tip: the reference path and the reflected path. A biosensor tip is exposed to a solution with a target (BMP-2) and an analyte (designed binder). BMP-2 is biotinylated, this means that biotin, which strongly binds to streptavidin, is covalently attached to BMP2 and this attaches to the probe which is functionalized with streptavidin. As molecules load on to the biosensor tip the reflected light deviates in its phase, which causes interference with the reference. The degree of interference is measured in correlation with binding activity.

For later experiments, I opted to use anti-histidine tag probes and my immobilized protein binder. The reasoning for this is that there are potential covalent streptavidin attachment sites located on the knuckle site of BMP-2, which may create some interference with binding.

Before running BLI, buffer optimization and loading tests were conducted to find the best criteria for measuring binding. The buffer that best reduced background non-specific interaction signal was HEPES pH 6, 2 mg/mL BSA, 0.05% Tween-20, and 50 mM NaCl.

Results

Computational Metrics and Filtering

Score Analysis

After running the design protocol on my grafted scaffolds, a score file was generated named ‘design_2203.out.’ The scores from this file were extracted and named ‘design_2203_score.dat’ to be analyzed further in Google Colab. There were 264 designs to analyze and filter down to less than 100. The metrics that I used were the weighted sums of energy terms (score), VBUNS or “very buried unsats” – defined as atoms deeper than 5.5 Å below the molecular surface, CMS and t_sap_score. I used score < 500, vbuns_all <=0, cms > 250 and t_sap_score < 35. After this filtering step, there were 8 top designs or 3% of the total amount that I looked at further.

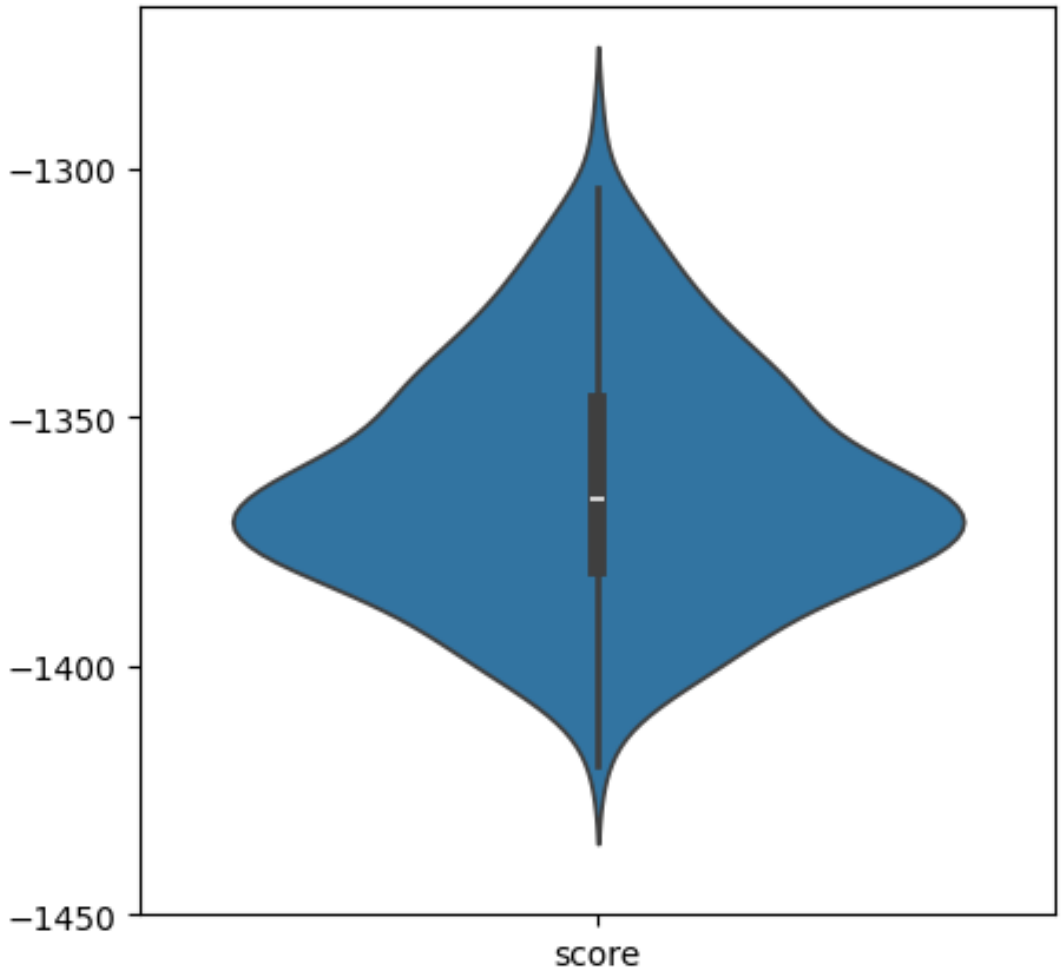

Figure 6: Score distribution

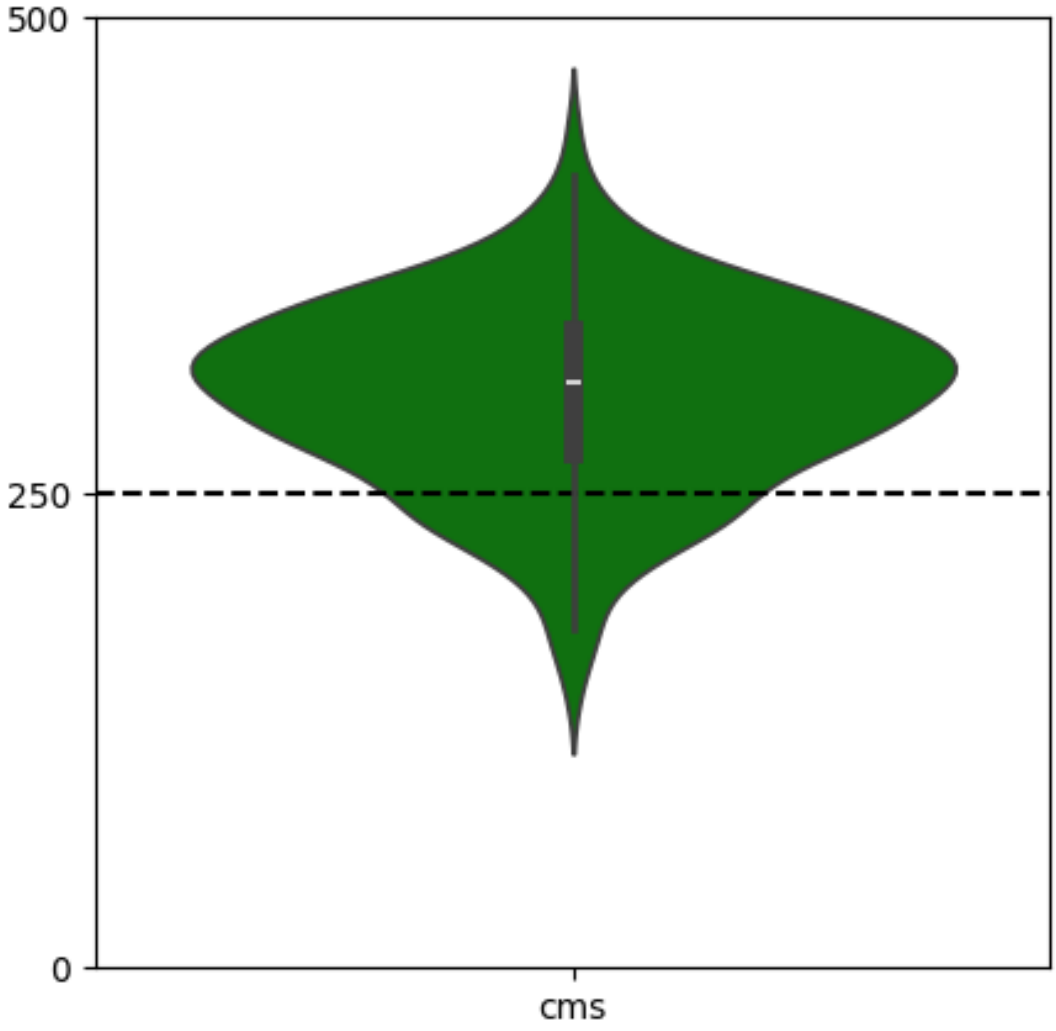

Figure 7: cms distribution

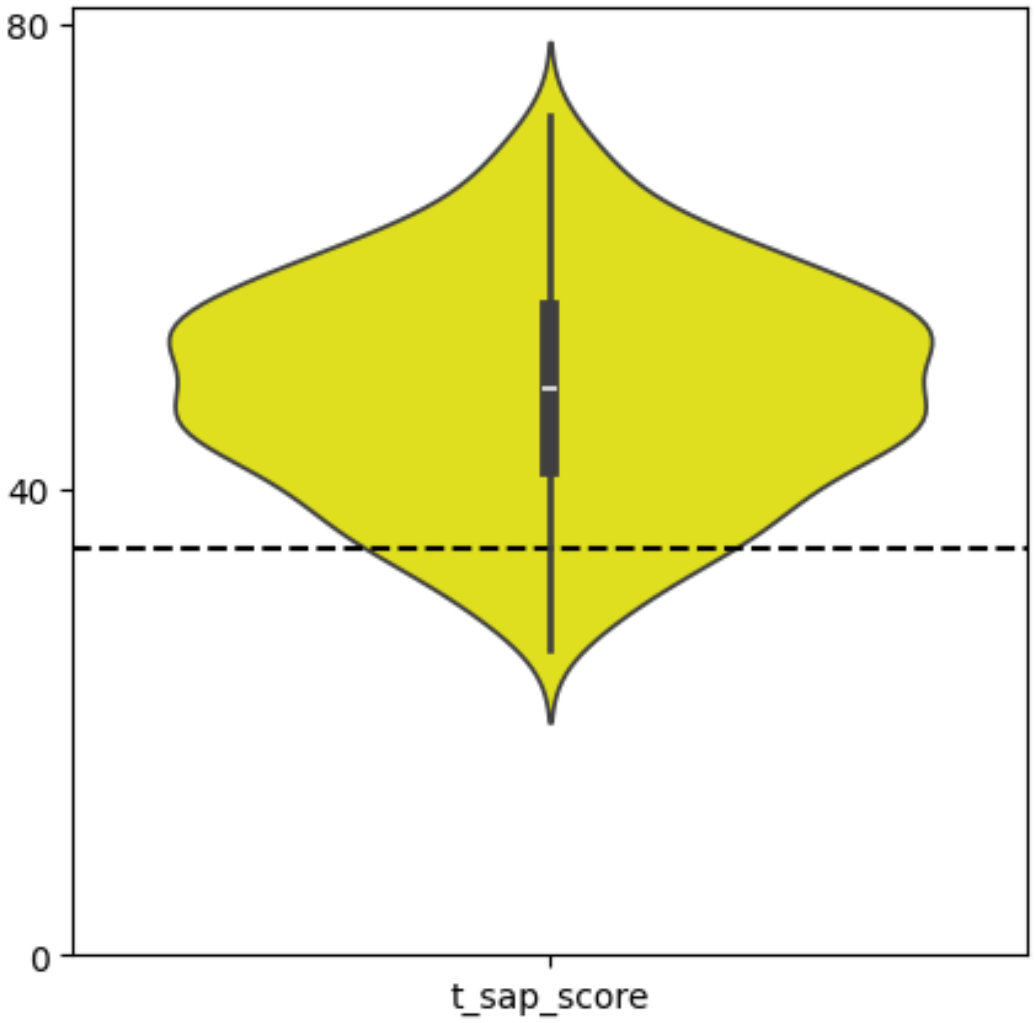

Figure 8: t_sap_score distribution

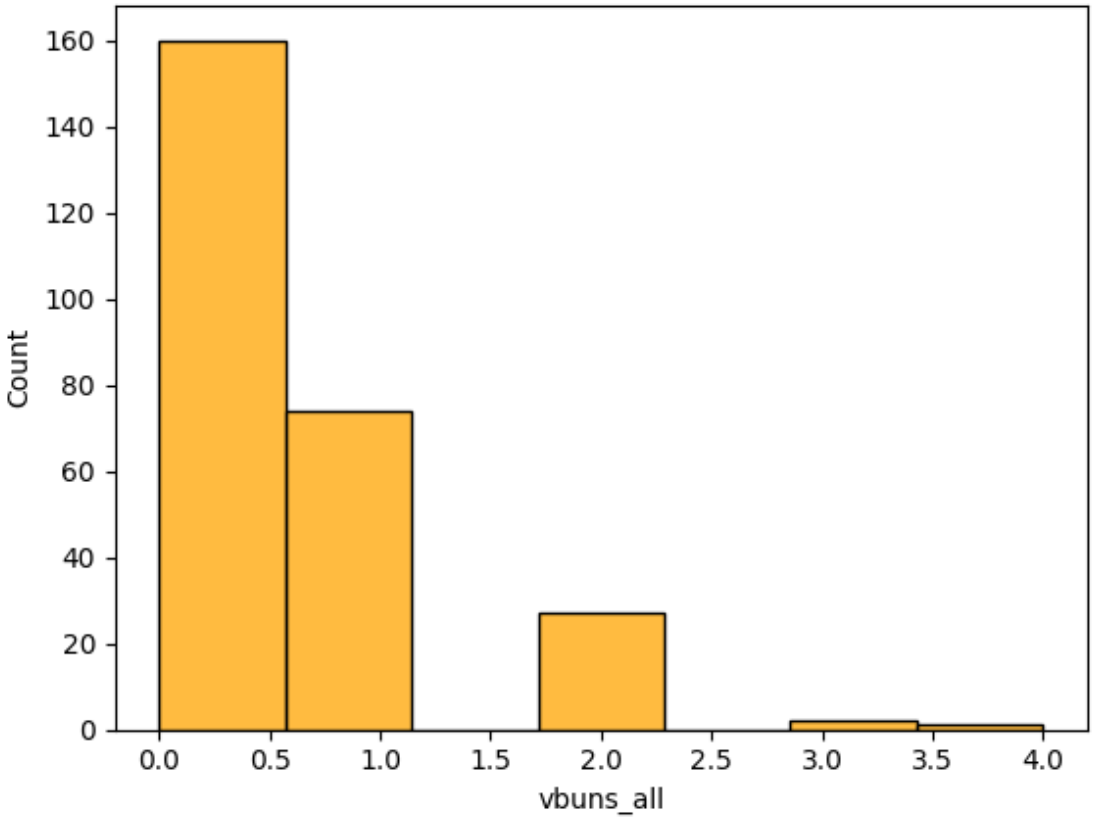

Figure 9: vbuns_all distribution

Docking

The 8 top designs were run through RosettaDock protocol for further filtering. Of these, 6 designs had a desired funnel. AlphaFold2 structure prediction showed one did not fold and therefore I was left with 5 remaining candidates. Through PyMol analysis of the structures and omitting designs with cysteines to make experimental workflow easier, four top designs were chosen.

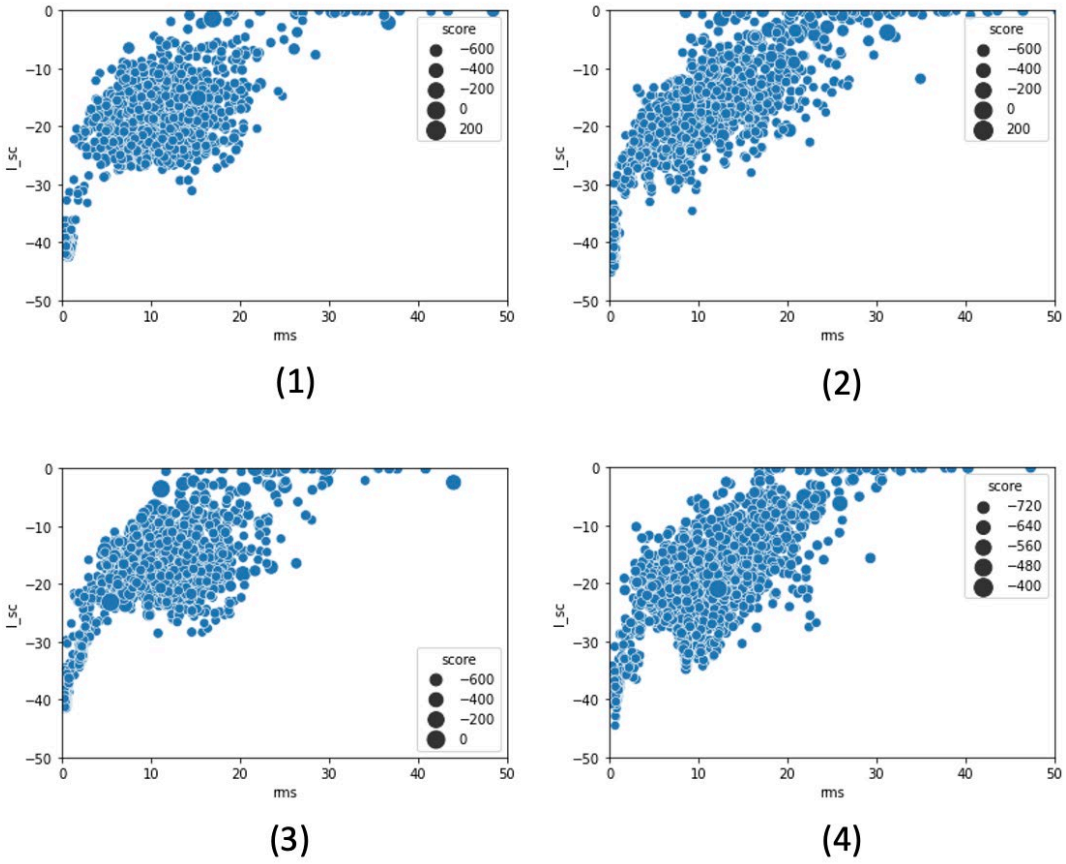

Figure 10: Docking from RosettaDock of designs from the motif grafting design pathway. Interface score (I_sc) is the total score of the complex minus the total score of each partner in isolation. I_sc is plotted against the alpha carbon RMSD of the ligand (rms) to produce a "funnel" down to 0. Nstruct = 50000

BMP-2 has a nearly identical backbone to BMP-7 but a different sequence. To show evidence that my designs bind specifically to BMP-2 based on sequence complementarity, I used RosettaDock as a specificity test. Remarkably, my design docked to BMP-2 but failed to dock to BMP-7.

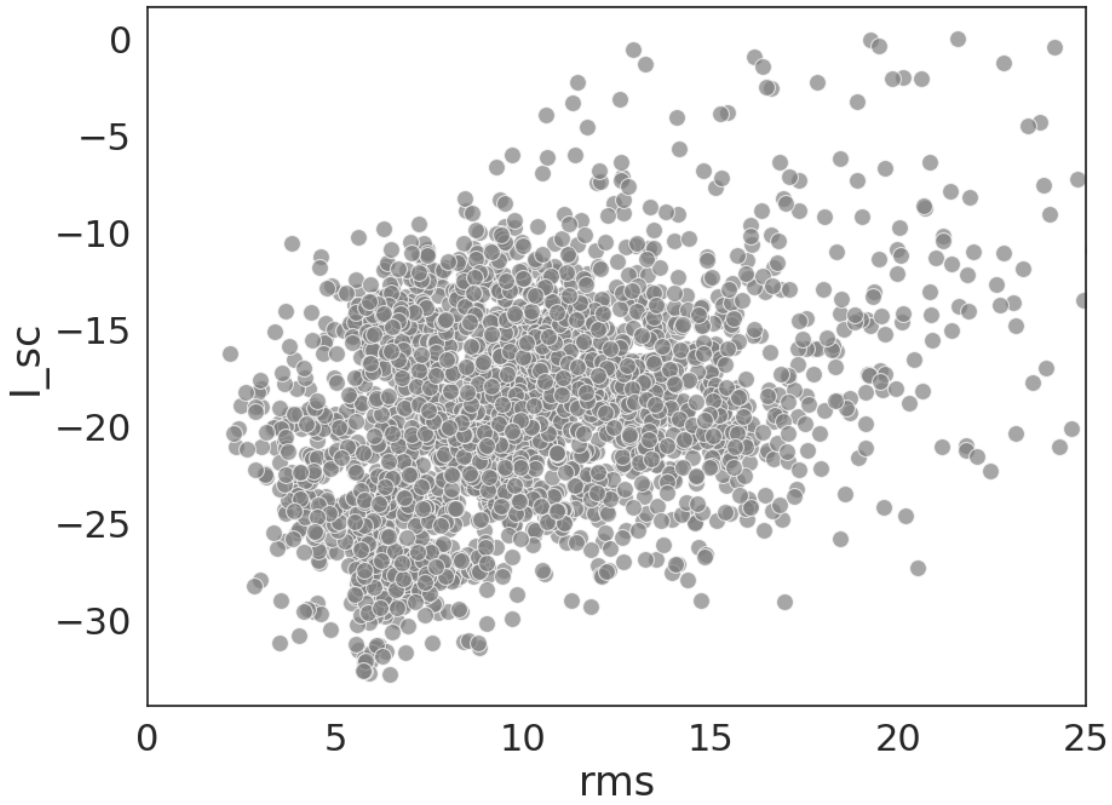

Figure 10: RosettaDock specificity test.

Generated binders docking to BMP-7 (gray). A funnel is not present for the binders to BMP-7.

After running my top design through the RFDiffusion partial diffusion protocol, the top generated designs also passed docking. The docking funnels were not as defined for some of the designs, but since they had pae_interaction <10 we decided to order them.

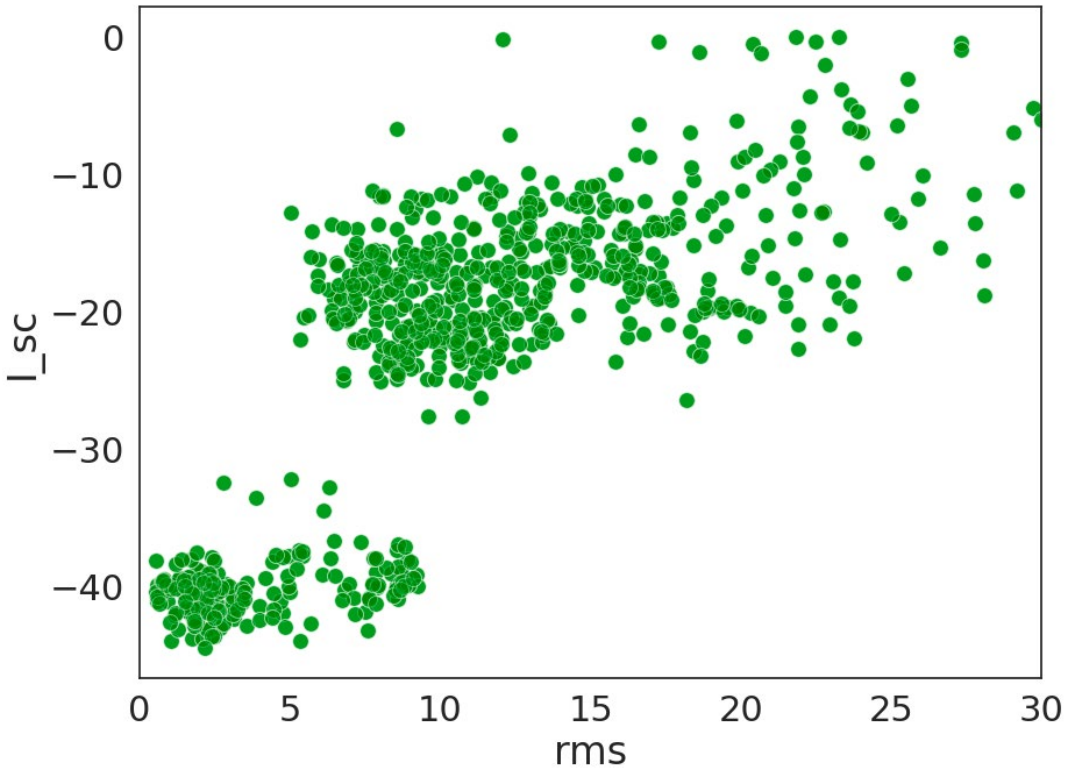

Figure 11: Docking from RosettaDock of top design after RFDiffusion partial diffusion rescues.

Structure Prediction

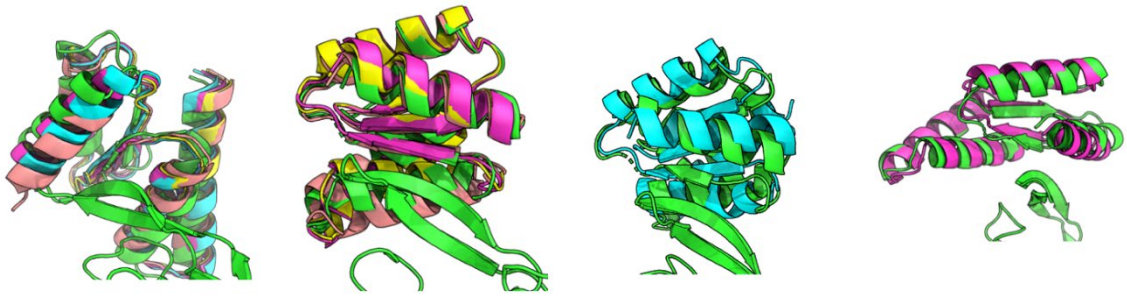

Figure 11: AlphaFold monomer predictions suggest folding.

Green = BMP-2. Colored proteins: prediction ranks generated by AlphaFold. Alignment of all colors shows a likelihood of proteins folding experimentally.

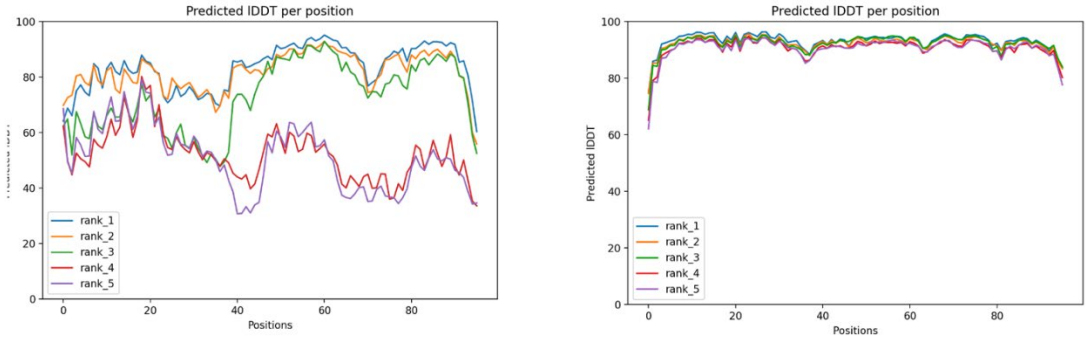

Figure 12: Improved pLDDT after partial diffusion.

Design (left) and partial diffusion rescue (right) show the predicted lDDT per amino acid position. Alignment of ranks indicates predicted folding, and higher pLDDT indicates a greater confidence of the model.

Structure prediction showed alignment on PyMol for all four top designs. Although the pLDDT was low initially, I decided the visual analysis and docking were sufficient for ordering these top four. After experimental binding assays, I found no evidence of these designs binding based on

BLI data. Our lab began implementing AlphaFold Multimer to predict the structure of both the binder and BMP-2 in complex, and from the findings of Watson et al. 2023 we learned that an AlphaFold Multimer pae_interaction < 10 was suggestive of successful binding. I used partial diffusion to improve my designs so that some had pae_interaction < 10.

Experimental Validation

Matrix-assisted laser desorption ionization

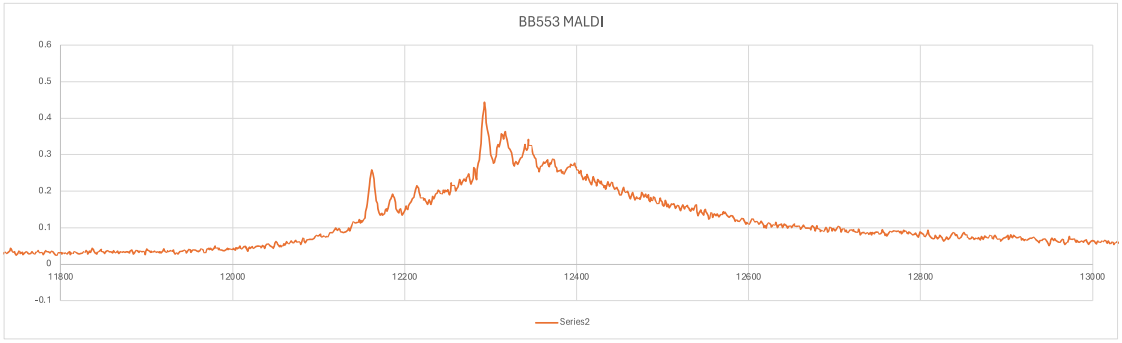

Figure 13: MALDI results

The computationally predicted mass of my protein was 12291 Daltons. From mass spectrometry analysis via MALDI, the highest peak had a measured mass of 12293 Daltons. This means that my protein was likely in solution. It seems that the solution may have impurities, as seen by the numerous smaller peaks surrounding the tallest peak. The second tallest peak is 130 Da, which is very close to the monoisotopic mass of methionine (131 Da), so I speculated that this was my protein with a cleaved methionine. The other peaks are still of unknown origin, so further MALDI will be done to obtain a cleaner spectrum without impurities.

Bio-layer interferometry

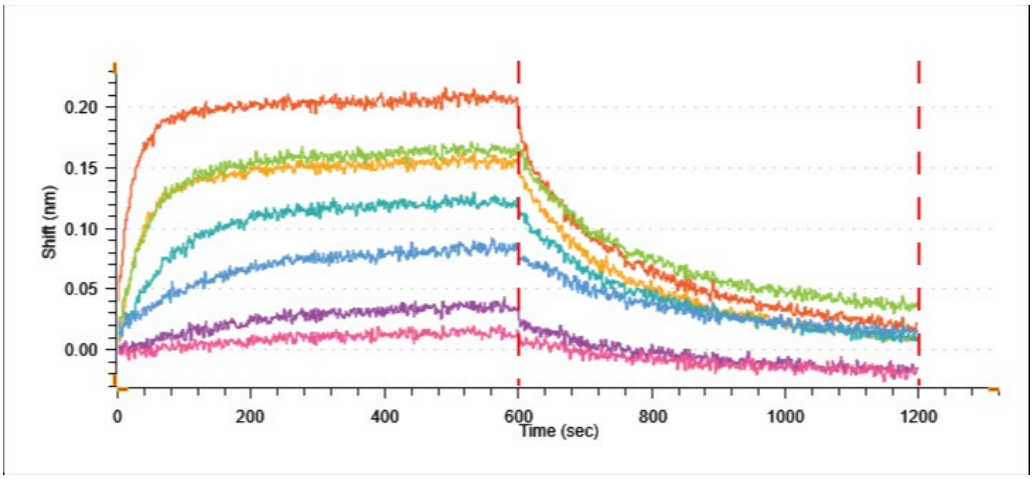

Figure 14: BLI results

Bio layer interferometry shows binding at 747 nM. This was the first binder that I designed with a measured binding affinity. Biotinylated BMP-2 was immobilized with Streptavidin probes and my protein was added to seven different wells with concentrations ranging from 4000 nM to 125 nM. All proteins were in a 50 mM HEPES pH 7, 0.05% Tween and 250 mM NaCl buffer. Negative controls were subtracted out of the spectrum including immobilized BMP-2 without binder, all concentration ranges of my binder without BMP-2 and only buffer with no added protein.

Discussion/Conclusion

This thesis shows that *de novo* binders can be computationally designed to the knuckle epitope of BMP-2 and experimentally validated with a measurable binding affinity. This is an important proof-of-concept for biomedical research on nonunion bone fractures. Additionally, work on these binders may continue past what has been shown in this thesis. After experimentally testing them in living cells, these binders may go on to participate in a hydrogel delivery system. This system would be more efficient than delivery with collagen sponges, and a far more convenient alternative to traditional bone grafting.

This work also demonstrates how some of the conventional challenges of β-sheet design can be overcome. By starting from a *de novo* peptide and building a full protein binder, polar interactions between chains can be more fully accounted for and protein binders can achieve high stability. Generating *de novo* binders may also help elucidate more about the general understanding of these interactions. Although binding has been the only observed functionality of these binders, it is very possible that they could do more. Proteins serve a vast number of functions in living cells, but it is hard to predict these functions with current technology. It is unknown whether these proteins are toxic, for example, or have enzymatic activity.

The first of seven binders with computationally predictive binding through AlphaFold Multimer's pAE_interaction < 10 cutoff did show binding through biolayer interferometry. The rest of the binders will also be tested and may have varying levels of affinity. It is possible that some of these binders will have a higher affinity than the one discussed in this thesis. However, it seemed that RosettaDock alone was not predictive of binder success for these designs. It is unclear why the RosettaDock predictions were not highly accurate, but research has found that Rosetta achieves high or medium accuracy for just over 50% of rigid-body targets.

Future directions for this project include further characterization of the protein binders. This involves getting a cleaner mass spectrometry graph, analyzing the secondary structural composition of the proteins with circular dichroism, and validating the structure with x-ray crystallography. Experiments may also be done to test competition of my protein binders among other proteins that bind to BMP-2. As mentioned, the soonest next steps would be expressing the genes encoding for my other protein binders, purifying the samples and testing their binding affinity.

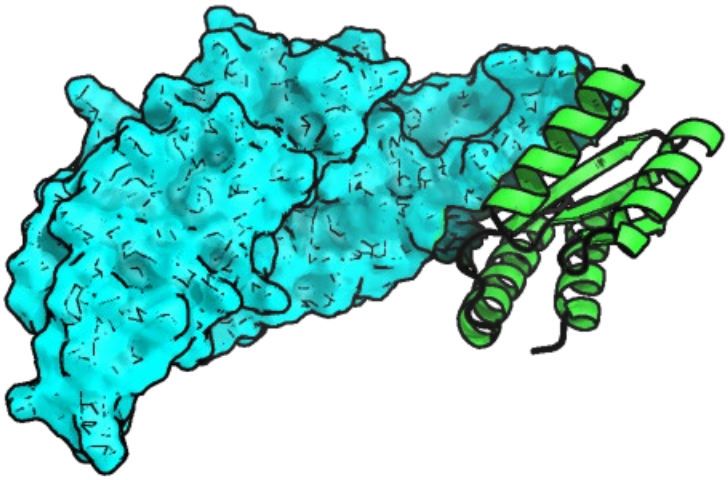

Figure 15: The first binder with a measured binding affinity of 720 nM.

Appendix

XML Scripts

motif_graft.xml

```
<ROSETTASCRIPTS>
  <SCOREFXNS>
    <ScoreFunction name="high_hbond" weights="ref2015.wts">
                    <Reweight scoretype="hbond_lr_bb" weight="10"/>
                    <Reweight scoretype="hbond_sr_bb" weight="10"/>
          </ScoreFunction>
  </SCOREFXNS>
  <RESIDUE_SELECTORS>
  </RESIDUE_SELECTORS>
  <TASKOPERATIONS>
  </TASKOPERATIONS>
  <SIMPLE_METRICS>
  </SIMPLE_METRICS>
  <FILTERS>
    <SSPrediction name="mismatch_probability" cmd="</home/bburress/parisahlab/pyrosetta-
2020.50.post0.dev0+978.master.edd2dcd21e3-py3.8-linux-x86_64.egg/psipred/runpsipred_single>" use_probability="1"
    mismatch_probability="1" use_svm="0" confidence="0"/>
  </FILTERS>
  <MOVERS>
    <MotifGraft name="motif_grafting"
      context_structure="bmp2_relaxed.pdb"
      motif_structure="%%motif%%"
      RMSD_tolerance="1.5"
      clash_score_cutoff="5"
      clash_test_residue="ALA"
      full_motif_bb_alignment="1"
      revert_graft_to_native_sequence="1"
      combinatory_fragment_size_delta="0:0"
      allow_repeat_same_graft_output="0" />
  </MOVERS>
  <PROTOCOLS>
    <Add mover="motif_grafting"/>
    <Add filter="mismatch_probability"/>
  </PROTOCOLS>
  <OUTPUT />
</ROSETTASCRIPTS>
```

optimize_sequence2.xml

```
<ROSETTASCRIPTS>
<!--Updated script for October/November 2021 interface design. Previous script: designinterface.xml-->

  <SCOREFXNS>
    <ScoreFunction name="score" weights="ref2015.wts"/>
  </SCOREFXNS>

  <RESIDUE_SELECTORS>
    <!--Selects the target protein and the binding protein. (A+B) and (C+D)-->
    <Chain name="target" chains="1"/>
    <Chain name="binder" chains="2"/>
    <!--Selects anything within 8 of bmp as well as bmp. (A+B+C)-->
    <Neighborhood name="target_neighborhood" selector="target"
      distance="8.0"/>
    <!--Selects only ligand interface residues. (C)-->
    <And name="binder_interface" selectors="binder,target_neighborhood"/>
    <!--Selects anything within 8 of the binder as well as the binder. (B+C+D)-->
    <Neighborhood name="binder_neighborhood" selector="binder" distance="8.0"/>
    <!--Selects the interface residues of both proteins. (B+C)-->
    <And name="int" selectors="target_neighborhood,binder_neighborhood"/>
    <!--Selects the residues of both proteins that are not at the interface-->
```

```
    <Not name="not_int" selector="int"/>
    <!--Selects the target interface. (B)-->
    <And name="target_interface" selectors="target,binder_neighborhood"/>
    <!--Selects hydrophobic residues-->
    <ResiduePropertySelector name="hyd" properties="HYDROPHOBIC"/>
    <!--Selects hydrophobic interface residues of the target protein-->
    <And name="target_interface_hyd" selectors="hyd,target_interface"/>
    <!--Selects all that is not the target interface. (A+C+D)-->
    <Not name="target_noninterface" selector="target_interface"/>
                <!--Defining the residues that are not at the interface. (A+B+D)-->
    <Not name="binder_noninterface" selector="binder_interface"/>
    <!--Selects the binder residues that are not at the interface. (D)-->
    <And name="binder_noninterface_without_target" selectors="binder,binder_noninterface"/>
    <!--Selecrs the target residues that are not at the interface. (A)-->
    <And name="target_noninterface_without_binder" selectors="target,target_noninterface"/>
    <!--hotspots-->
    <ResiduePDBInfoHasLabel name="hotspots" property="HOTSPOT"/>
    <!--core-->
    <Layer name="core" select_core="true" ball_radius="2.5"/>
    <True name="true_sel"/>
    <!--Gylcines and prolines-->
    <ResidueName name="GLY_PRO" residue_names="GLY,PRO"/>
</RESIDUE_SELECTORS>

<TASKOPERATIONS>
    <OperateOnResidueSubset name="hold_target_noninterface"
                                                        selector="target_noninterface_without_binder">
      <PreventRepackingRLT/>
    </OperateOnResidueSubset>
                <OperateOnResidueSubset name="only_repack_target_interface"
                    selector="target_interface">
      <RestrictToRepackingRLT/>
    </OperateOnResidueSubset>
                <OperateOnResidueSubset name="only_repack_hotspots"
                    selector="hotspots">
      <RestrictToRepackingRLT/>
    </OperateOnResidueSubset>
    <OperateOnResidueSubset name="hold_binder_noninterface"
                    selector="binder_noninterface_without_target">
      <RestrictToRepackingRLT/>
    </OperateOnResidueSubset>
    <OperateOnResidueSubset name="hold_core" selector="core">
      <RestrictToRepackingRLT/>
    </OperateOnResidueSubset>
    <OperateOnResidueSubset name="hold_gly_pro" selector="GLY_PRO">
      <RestrictToRepackingRLT/>
    </OperateOnResidueSubset>
</TASKOPERATIONS>

<TASKOPERATIONS>
  <ProteinProteinInterfaceUpweighter name="upweight" interface_weight="3.0"/>
  <ExtraRotamersGeneric name="extra_chi" ex1="1" ex2="1" extrachi_cutoff="0"/>
  <RestrictToRepacking name="restrict"/>
  <InitializeFromCommandline name="init"/>
  <RestrictToInterfaceVector name="intonly" chain1_num="1" chain2_num="2" CB_dist_cutoff="10.0"
    nearby_atom_cutoff="5.5" vector_angle_cutoff="75.0" vector_dist_cutoff="9.0" include_all_water="1"/>
</TASKOPERATIONS>

<SIMPLE_METRICS>
  <!--Sasa measurements,including polar and hydrophobic,for the interface-->
  <SasaMetric name="tot_sasa"    residue_selector="int"    sasa_metric_mode="all_sasa"/>
  <SasaMetric name="pol_sasa"    residue_selector="int"    sasa_metric_mode="polar_sasa"/>
  <SasaMetric name="hyd_sasa"    residue_selector="int"    sasa_metric_mode="hydrophobic_sasa"/>
</SIMPLE_METRICS>

<FILTERS>
  <!--interface complementarity-->
  <ShapeComplementarity name="sc" min_sc="0.5"  residue_selector1="target" residue_selector2="binder"
    jump="1" confidence="0"/>
```

```
<!--ddg with and without repack-->
<Ddg name="ddg_rep" chain_num="2" threshold="-1" jump="1"
  repeats="5" repack="1" confidence="0" scorefxn="score"/>
<Ddg name="ddg_norep" chain_num="2" threshold="-1" jump="1"
  repack="0" confidence="0" scorefxn="score"/>

<!--filters cooresponding to our previous sasa metrics-->
<SimpleMetricFilter name="filter_sasa" metric="tot_sasa"  cutoff="100" comparison_type="gt" confidence="0"/>
<SimpleMetricFilter name="filter_pol_sasa" metric="pol_sasa"  cutoff="100" comparison_type="gt" confidence="0"/>
<SimpleMetricFilter name="filter_hyd_sasa" metric="hyd_sasa"  cutoff="100" comparison_type="gt" confidence="0"/>

<!--filtering buried unsatisfied polar atoms-->
<BuriedUnsatHbonds name="buns_bb"
  residue_selector="int" report_bb_heavy_atom_unsats="true"
  scorefxn="true" cutoff="10" ignore_surface_res="false"
  print_out_info_to_pdb="true" use_ddG_style="true" jump_number="1"
  dalphaball_sasa="1" probe_radius="1.1" confidence="0"/>
<BuriedUnsatHbonds name="buns_sc"
  residue_selector="int" report_sc_heavy_atom_unsats="true"
  scorefxn="true" cutoff="10" ignore_surface_res="false"
  print_out_info_to_pdb="true" use_ddG_style="true" jump_number="1"
  dalphaball_sasa="1" probe_radius="1.1" confidence="0"/>
<BuriedUnsatHbonds name="buns_all"
  residue_selector="int" report_all_heavy_atom_unsats="true"
  scorefxn="true" cutoff="10" ignore_surface_res="false"
  print_out_info_to_pdb="true" use_ddG_style="true" jump_number="1"
  dalphaball_sasa="1" probe_radius="1.1" confidence="0"/>
<BuriedUnsatHbonds name="vbuns_bb"
  residue_selector="int" report_bb_heavy_atom_unsats="true"
  scorefxn="score" ignore_surface_res="false"  print_out_info_to_pdb="true"
  atomic_depth_selection="5.5"   burial_cutoff="1000" use_ddG_style="true"
  jump_number="1"  burial_cutoff_apo="0.2" dalphaball_sasa="true"
  probe_radius="1.1" confidence="0" />
<BuriedUnsatHbonds name="vbuns_sc"
  residue_selector="int" report_sc_heavy_atom_unsats="true"
  scorefxn="score" ignore_surface_res="false"
  print_out_info_to_pdb="true" atomic_depth_selection="5.5"
  burial_cutoff="1000" use_ddG_style="true" jump_number="1"
  burial_cutoff_apo="0.2" dalphaball_sasa="true"
  probe_radius="1.1" confidence="0" />
<BuriedUnsatHbonds name="vbuns_all"
  residue_selector="int" report_all_heavy_atom_unsats="true"
  scorefxn="score" ignore_surface_res="false"
  print_out_info_to_pdb="true" atomic_depth_selection="5.5"
  burial_cutoff="1000" use_ddG_style="true" jump_number="1"
  burial_cutoff_apo="0.2" dalphaball_sasa="true"
  probe_radius="1.1" confidence="0" />

<!--Packing statistics-->
<PackStat name="pstat" threshold="0.65" chain="2" confidence="0"/>
<!--Holes at the protein-protein interface-->
<InterfaceHoles name="int_holes" jump="1" confidence="0"/>
<!--Secondary structures at the interface-->
<SSShapeComplementarity name="ss_sc" verbose="1" loops="1"
  helices="1" min_sc="0.5" confidence="0"/>
<!--Holes of the binder-->
<Holes name="holes" threshold="1.0" residue_selector="binder"/>

<!--Interface contact-->
<InterfaceHydrophobicResidueContacts name="int_hydcontact"
  target_selector="target_interface_hyd" binder_selector="binder_interface" scorefxn="score"
  apolar_res="ALA,CYS,CYD,PHE,ILE,LEU,MET,PRO,THR,VAL,TRP,TYR"
  confidence="0"/>
<AtomicContactCount name="contact" partition="jump" jump="1"
  normalize_by_sasa="0" confidence="0"/>
<AtomicContactCount name="contact_norm" partition="jump" jump="1"
  normalize_by_sasa="1" confidence="0"/>
<Sasa name="sasa" confidence="0"/>
SSPrediction name="mismatch_probability" cmd="~/psipred/runpsipred_single"
  use_probability="1" mismatch_probability="1" use_svm="0" confidence="0"/>
```

```
  <!--Packing-->
  <ContactMolecularSurface name="cms" distance_weight="0.5"
    target_selector="target" binder_selector="binder" confidence="0"/>
 </FILTERS>

 <SIMPLE_METRICS>
  <SapScoreMetric name="binder_sap" score_selector="target"/>
  <SapScoreMetric name="target_sap" score_selector="binder"/>
  <SapScoreMetric name="binder_blocked_sap" score_selector="target"
    sap_calculate_selector="binder" sasa_selector="true_sel"/>
  <SapScoreMetric name="target_blocked_sap" score_selector="binder"
    sap_calculate_selector="target" sasa_selector="true_sel"/>
  <CalculatorMetric name="binder_delta_sap" equation="binder_sap_score - binder_blocked_sap">
    <VAR name="binder_sap_score" metric="binder_sap"/>
    <VAR name="binder_blocked_sap" metric="binder_blocked_sap"/>
  </CalculatorMetric>
  <CalculatorMetric name="target_delta_sap" equation="target_sap_score - target_blocked_sap">
    <VAR name="target_sap_score" metric="target_sap"/>
    <VAR name="target_blocked_sap" metric="target_blocked_sap"/>
  </CalculatorMetric>
 </SIMPLE_METRICS>

 <JUMP_SELECTORS>
  <JumpIndex name="jump1" jump="1"/>
 </JUMP_SELECTORS>

 <MOVE_MAP_FACTORIES>
  <MoveMapFactory name="des_mm" bb="0" chi="0">
    <Backbone residue_selector="int" enable="true" />
    <Chi residue_selector="int" enable="true"/>
    <Jumps enable="true" jump_selector="jump1"/>
  </MoveMapFactory>
 </MOVE_MAP_FACTORIES>

 <MOVERS>
  <!--Relax before running metrics-->
  <FastRelax name="relax" scorefxn="score" disable_design="1"
    repeats="5" ramp_down_constraints="1" relaxscript="InterfaceRelax2019">
    <MoveMap name="relax_mm" bb="1" chi="1" jump="1">
      <ResidueSelector selector="not_int" chi="0" bb="0" bondangle="0" bondlength="0"/>
      <Jump number="1" setting="true"/>
    </MoveMap>
  </FastRelax>
  <!--Mover for the interface interaction-->
  <InterfaceAnalyzerMover name="inter_move" scorefxn="score"
    packstat="true" interface_sc="true" jump="1"/>
  <ddG name="ddg_move" scorefxn="score" chain_num="2" solvate="1"
    repack_bound="1" repack_unbound="1" solvate_rbmin="0"
    solvate_unbound="0" min_water_jump="1"
    task_operations="intonly,restrict,extra_chi"/>
  <RunSimpleMetrics name="metric1" metrics="tot_sasa" prefix="t_"/>
  <RunSimpleMetrics name="metric2" metrics="pol_sasa" prefix="p_"/>
  <RunSimpleMetrics name="metric3" metrics="hyd_sasa" prefix="h_"/>
  <RunSimpleMetrics name="bsap" metrics="binder_sap" prefix="b_"/>
  <RunSimpleMetrics name="tsap" metrics="target_sap" prefix="t_"/>
  <RunSimpleMetrics name="bblock_sap" metrics="binder_blocked_sap" prefix="bb_"/>
  <RunSimpleMetrics name="tblock_sap" metrics="target_blocked_sap" prefix="tb_"/>
  <RunSimpleMetrics name="b_delta_sap" metrics="binder_delta_sap" prefix="bd_"/>
  <RunSimpleMetrics name="t_delta_sap" metrics="target_delta_sap" prefix="td_"/>
 </MOVERS>

 <MOVERS>
        <FastDesign name="fast_design" scorefxn="score" relaxscript="InterfaceDesign2019"

task_operations="hold_target_noninterface,hold_binder_noninterface,only_repack_target_interface,only_repack_hotspots,upweight,hold_gly_pro
"
    movemap_factory="des_mm" repeats="8"/>
  <FavorNativeResidue name="nonnative_penalty" bonus="1.5"/>
 </MOVERS>
```

```
<PROTOCOLS>
  <Add mover="relax"/>
  <Add mover="nonnative_penalty"/>
  <Add mover="fast_design"/>
  <Add mover="inter_move"/>
  <Add filter="sc"/>
  <Add filter="ddg_rep"/>
  <Add filter="ddg_norep"/>
  <Add filter="filter_sasa"/>
  <Add filter="filter_pol_sasa"/>
  <Add filter="filter_hyd_sasa"/>
  <Add filter="buns_bb"/>
  <Add filter="buns_sc"/>
  <Add filter="buns_all"/>
  <Add filter="vbuns_bb"/>
  <Add filter="vbuns_sc"/>
  <Add filter="vbuns_all"/>
  <Add filter="pstat"/>
  <Add filter="int_holes"/>
  <Add filter="ss_sc"/>
  <Add filter="int_hydcontact"/>
  <Add filter="holes"/>
  <Add filter="contact"/>
  <Add filter="contact_norm"/>
  <Add filter="sasa"/>
  Add filter="mismatch_probability"/>
  <Add filter="cms"/>
  <Add mover="bsap"/>
  <Add mover="tsap"/>
  <Add mover="bblock_sap"/>
  <Add mover="tblock_sap"/>
  <Add mover="b_delta_sap"/>
  <Add mover="t_delta_sap"/>
</PROTOCOLS>

<OUTPUT />

</ROSETTASCRIPTS>
```

References

1. GBD 2019 Fracture Collaborators. Global, regional, and national burden of bone fractures in 204 countries and territories, 1990-2019: a systematic analysis from the Global Burden of Disease Study 2019. *Lancet Healthy Longev*. 2021;2(9):e580-e592. doi:10.1016/S2666-7568(21)00172-0

2. Thomas JD, Kehoe JL. Bone Nonunion. In: *StatPearls*. Treasure Island (FL): StatPearls Publishing; March 6, 2023.

3. Sidhartha Chaudhury, Sergey Lyskov, Jeffrey J. Gray, PyRosetta: a script-based interface for implementing molecular modeling algorithms using Rosetta, *Bioinformatics*, Volume 26, Issue 5, March 2010, Pages 689–691, https://doi.org/10.1093/bioinformatics/btq007

4. Watson, J.L., Juergens, D., Bennett, N.R. *et al.* De novo design of protein structure and function with RFdiffusion. *Nature* **620**, 1089–1100 (2023). https://doi.org/10.1038/s41586-023-06415-8

5. J. Dauparas *et al.* Robust deep learning–based protein sequence design using ProteinMPNN.*Science***378**,49-56(2022).DOI:10.1126/science.add2187

6. Halloran D, Durbano HW, Nohe A. Bone Morphogenetic Protein-2 in Development and Bone Homeostasis. *J Dev Biol*. 2020;8(3):19. Published 2020 Sep 13. doi:10.3390/jdb8030019

7. Zou ML, Chen ZH, Teng YY, et al. The Smad Dependent TGF-β and BMP Signaling Pathway in Bone Remodeling and Therapies. *Front Mol Biosci*. 2021;8:593310. Published 2021 May 5. doi:10.3389/fmolb.2021.593310

8. Howard MT, Wang S, Berger AG, et al. Sustained release of BMP-2 using self-assembled layer-by-layer film-coated implants enhances bone regeneration over burst release. *Biomaterials*. 2022;288:121721. doi:10.1016/j.biomaterials.2022.121721

9. Urist MR. Bone: formation by autoinduction. *Science*. 1965;150(3698):893-899. doi:10.1126/science.150.3698.893

10. Walker SP, Yallapragada VVB, Tangney M. Arming Yourself for The In Silico Protein Design Revolution. *Trends Biotechnol*. 2021;39(7):651-664. doi:10.1016/j.tibtech.2020.10.003

11. Kalita P, Tripathi T, Padhi AK. Computational Protein Design for COVID-19 Research and Emerging Therapeutics. *ACS Cent Sci*. 2023;9(4):602-613. Published 2023 Mar 20. doi:10.1021/acscentsci.2c01513

12. Kresge N, Simoni RD, Hill RL. Enhancing our Understanding of Protein Structure: the Work of Jane and David Richardson. *J Biol Chem*. 2011;286(17):e3-e5. doi:10.1074/jbc.O111.000223

13. Bennett, N.R., Coventry, B., Goreshnik, I. *et al.* Improving de novo protein binder design with deep learning. *Nat Commun* **14**, 2625 (2023). https://doi.org/10.1038/s41467-023-38328-5

14. Sahtoe DD, Coscia A, Mustafaoglu N, et al. Transferrin receptor targeting by de novo sheet extension. *Proc Natl Acad Sci U S A*. 2021;118(17):e2021569118. doi:10.1073/pnas.2021569118

15. Silva, D., Correia, B.E., and Procko, E. (2016) Motif-driven Design of Protein-Protein Interactions. Methods Mol. Biol. 1414:285-304

16. Rocklin GJ, Chidyausiku TM, Goreshnik I, et al. Global analysis of protein folding using massively parallel design, synthesis, and testing. *Science*. 2017;357(6347):168-175. doi:10.1126/science.aan0693

17. Chaudhury, S., Berrondo, M., Weitzner, B. D., Muthu, P., Bergman, H., Gray, J. J.; (2011) Benchmarking and analysis of protein docking performance in RosettaDock v3.2. PLoS One, Accepted for Publication

18. Minkyung Baek *et al.* Accurate prediction of protein structures and interactions using a three-track neural network.*Science***373**,871-876(2021).DOI:10.1126/science.abj8754

19. Jumper, J., Evans, R., Pritzel, A. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021). https://doi.org/10.1038/s41586-021-03819-2

20. Jue Wang *et al* Scaffolding protein functional sites using deep learning.*Science***377**,387-394(2022).DOI:10.1126/science.abn2100

21. Gibson, D., Young, L., Chuang, RY. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* **6**, 343–345 (2009). https://doi.org/10.1038/nmeth.1318