Computational Design and Enhancement of Proteins: A Study on Mechanostability and Isopeptide Generation

Donal Martin Naylor

A thesis
submitted in partial fulfillment of the
requirements for the degree of
Master of Biochemistry

University of Washington
2023

Committee:
David Baker
Neil King
Micheal Ailion

Program Authorized to Offer Degree: Biochemistry

University of Washington

Abstract

Computational Design and Enhancement of Proteins: A Study on Mechanostability and Isopeptide Generation

Donal Martin Naylor

Chair of the Supervisory Committee:
David Baker
Department of Biochemistry

This study employs de novo protein design to augment protein mechanostability, a critical characteristic in bioengineering, biotechnology, and therapeutic development. The enhancement of protein mechanostability may find significant applications in areas such as modulating biomaterial elastic properties for a range of biomedical applications or controlling the stiffness of cellular environments to study the mechanobiology of cell differentiation. This work opens new avenues for the development of robust biomaterials and improved cellular therapeutics. Two design rounds were conducted, each employing computational design methods and subsequent experimental validations. The first round utilized inpainting and diffusion methods, revealing consistent results for inpainting-derived proteins and varied results for diffusion-derived ones. As assessed by atomic force microscopy, all designed proteins, despite varying success, fell short of matching or surpassing the mechanostability of the natural proteins which served as structural templates for their design. The second round, a targeted redesign of high rupture force proteins, led to less consistent results overall but one standout protein showed significant increase in rupture force, accompanied by notable structural changes that confirmed an underlying mechanistic hypothesis at the core of protein mechanostability. This work underscores the potential of de novo protein design in enhancing protein mechanostability, paving the way for applications requiring stable proteins.

Acknowledgments

I stand before this academic milestone with gratitude and humility for the numerous individuals who have contributed significantly to the completion of my thesis.

First and foremost, I wish to express my sincere appreciation to Professor David Baker, who generously hosted me in his lab and granted me the remarkable opportunity to complete my Masters under his guidance. His dedication to academia and nurturing young minds is genuinely inspiring.

Secondly, my heartfelt thanks goes to Dr. Lukas Milles. His patient mentorship was a beacon throughout the course of this thesis. From wet lab practices to computational skills, the breadth of his knowledge is only matched by his dedication to detail. His influence on my academic and professional growth cannot be overstated.

I am also indebted to these three Professors, Professor Michael Ailion, Professor Neil King and Professor Ruhohola-Baker. I extend my sincere gratitude to Professor Ailion for graciously accepting my invitation to join the committee and for his invaluable advice that enriched this research. Simultaneously, I would like to thank Professor King for enabling me to rotate in his lab and offering me the freedom to explore my ideas. The confidence placed in me and my research has been pivotal to my progress. Further, I wish to acknowledge Professor Ruhohola-Baker, who allowed me to work in her incredible lab. Her lab provided an environment that fostered curiosity, encouraged exploration, and embraced innovation. This inspiring setting undoubtedly enriched my learning experience and honed my research skills.

My gratitude also extends to Dr. Basile Wicky for his constant guidance and mentorship. His words of encouragement and advice have not only guided my path but also boosted my confidence in my academic journey.

The wider community at the Institute for Protein Design (IPD) has been invaluable in this journey. Preetham Venkatesh's advice and camaraderie have been a beacon during challenging times. My sincere thanks also extend to Valentina Alvarez, Michelle Chicas, Risako Gen, Paul Kim, Susan Kleinfelter, Paul Kwon, Peik Lund-Andersen, Meg Lunn-Halbert, Joe Min, Aditya Krishnakumar, Zac Jones, Ashish Phal, Yuliya Politanska, Savannah Speir, Pascal Sturmfels, each of whom have contributed to my journey in their unique ways, creating an environment of mutual learning and camaraderie.

Table of Contents

Chapter 1: Protein Mechanostability

Abstract:

This study employs de novo protein design to augment protein mechanostability, a critical characteristic in bioengineering, biotechnology, and therapeutic development. The enhancement of protein mechanostability may find significant applications in areas such as modulating biomaterial elastic properties for a range of biomedical applications or controlling the stiffness of cellular environments to study the mechanobiology of cell differentiation. This work opens new avenues for the development of robust biomaterials and improved cellular therapeutics. Two design rounds were conducted, each employing computational design methods and subsequent experimental validations. The first round utilized inpainting and diffusion methods, revealing consistent results for inpainting-derived proteins and varied results for diffusion-derived ones. As assessed by atomic force microscopy, all designed proteins, despite varying success, fell short of matching or surpassing the mechanostability of the natural proteins which served as structural templates for their design. The second round, a targeted redesign of high rupture force proteins, led to less consistent results overall but one standout protein showed significant increase in rupture force, accompanied by notable structural changes that confirmed an underlying mechanistic hypothesis at the core of protein mechanostability. This work underscores the potential of de novo protein design in enhancing protein mechanostability, paving the way for applications requiring stable proteins.

Motivation:

Delving into the complex nature of proteins—responsible for almost all cellular activities—is paramount in the domain of molecular biology. Crucial to their functionality is the way proteins maintain their structure under various conditions, underscoring the importance of their stability. Our pursuit of enhancing this stability, termed as protein mechanostability, stands at the intersection of fundamental research and practical applications. Augmenting and controlling mechanostability could potentially impact various sectors, such as bioengineering, biotechnology, and disease treatment, leading to profound advancements.

A promising tool in this undertaking is de novo protein design. This ground-breaking methodology enables the design and synthesis of proteins from scratch. Utilizing computational design methods, we can predict the sequence of amino acids that would fold into desired protein structures. This not only provides a gateway to fabricate new materials and catalysts but also opens the doors to developing innovative therapies and diagnostics.

The primary objective of this thesis is to harness de novo protein design's potential to enhance the mechanostability of proteins. The central hypothesis rests on the idea that designing a protein's primary sequence can lead to the creation of novel protein structures that exhibit greater resistance to mechanical stress. If successful, the ability to engineer proteins with increased mechanostability could have far-reaching applications. This could facilitate the development of more robust enzymes for industrial applications, stronger materials for biomedical use, or more resilient protein-based therapeutics.

In light of the results of this research, the answer to the inquiry—can we boost protein mechanostability via de novo protein design—seems to be a promising 'yes'. We anticipate that the insights gained from this research will not only expand our foundational understanding of protein structure and stability but also serve as a catalyst for future advancements in sectors relying on stable protein applications.

Introduction:

The intricate phenomena dictating protein behavior under various physical and chemical conditions are crucial to their biological function. A pivotal area within this broad domain is understanding how proteins respond to different mechanical stressors – a field referred to as mechanobiology. This field yields valuable insights into vital biological processes, from bacterial adhesion to (stem-) cell development. Recent advancements have revealed astonishing mechanisms contributing to the extreme mechanostability of certain proteins, notably those involved in bacterial adhesion. These mechanisms offer exciting possibilities for the development of innovative biomaterials, new therapeutic strategies, and advancements in biotechnology. The present thesis focuses on leveraging the principles of de novo protein design to enhance protein mechanostability, drawing inspiration from these naturally occurring, mechanically robust proteins.

Our understanding of protein mechanostability largely derives from the examination of naturally occurring proteins that exhibit this intriguing property. For instance, the protein fold of B domains in Staphylococcal pathogens provides a compelling model, owing to its exceptional mechanostability. This characteristic is largely attributed to the coordination of three calcium ions, an insight gleaned from the study "Calcium Stabilizes the Strongest Protein Fold". This study underscores how ion-coordinating interactions can bolster the mechanical strength of proteins, thereby opening a gateway to the design of highly stable biomaterials and the development of novel calcium sensors (Milles et al., 2018).

Yet protein mechanical strength does not exist in isolation. Proteins inhabit complex mechanically active environments that can considerably alter their function. The study titled "The Mechanical World of Bacteria" illustrates the pivotal role of mechanics in bacterial behavior, particularly in the realm of surface adhesion. Remarkably, bacteria such as *Escherichia coli* and *Pseudomonas aeruginosa* have evolved to form "catch bonds", which strengthen with increasing shear force, enabling these bacteria to adhere to surfaces even under turbulent flow conditions and high hydrodynamic stress (Persat et al., 2015). Such mechanical adaptations are instrumental to bacterial survival, and in the case of pathogens host invasion, and could yield valuable insights for the design of mechanically stable proteins.

Delving further into the mechanisms governing bacterial adhesin, the "Molecular Mechanism of Extreme Mechanostability in a Pathogen Adhesin" study spotlights the Staphylococcus epidermidis adhesin, SdrG. This protein-protein interaction boasts an extraordinary mechanostability that rivals the strength of

covalent bonds. The adhesins, such as SdrG and its homologs, achieve mechanically hyperstable adhesion to host proteins using the "Dock, Lock and Latch" (DLL) mechanism (Ponnuraj et al., 2003). This resilience is facilitated by the confined alignment of the backbone hydrogen bonds in a shear geometry, independent of the peptide sequence. Moreover, the N and C-terminal confinement and shear geometry are suggested to be key contributors to the extreme mechanostability of B domains. The study suggests that the calcium ions electrostatically protect the hydrogen bonds from breaking and lock them in a shear geometry (Milles et al., 2018).

This thesis aims to scrutinize and exploit these remarkable principles of protein mechanostability, with a particular focus on de novo protein design as a tool to engineer proteins with enhanced mechanical resilience. By integrating these principles, we strive to make substantial contributions to the field of mechanobiology and aid in the development of robust biomaterials and effective therapeutic strategies.

Building upon our understanding of protein mechanostability, this thesis explores recent innovations in the field of de novo protein design. Our aim is to engineer proteins with enhanced mechanical resilience, drawing inspiration from the extraordinary principles of naturally occurring, mechanically robust proteins.

The design process commences with a computational approach. A technique akin to inpainting, traditionally used in image restoration, is harnessed for recovering missing sequence and structural information in proteins. We use a model trained as an inpainting-specific model, RFjoint, supported by the protein structure prediction networks RoseTTAFold, inspired by AlphaFold. This novel method, proven successful in generating designs closely resembling intended structures, paves the way for subsequent sequence design on the generated backbone structures (Wang et al., 2022).

Alongside the inpainting stage, we use more recent tools in our computational design process, employing a method known as RFdiffusion. This approach harnesses the power of denoising diffusion probabilistic models (DDPMs), a tool originally proven effective in generating diverse outputs in image and language processing. We use a DDPMs approach built on RoseTTAFold, a high-precision protein structure prediction network, to create a novel generative model for protein backbones. The result is an ability to produce a wide range of functional and structurally diverse proteins from a starting point of random noise (Watson et al., 2022).

Following inpainting and diffusion, the novel protein structures undergo sequence generation employing a deep learning-based method named ProteinMPNN. Distinct from traditional, energy-optimization-focused

physically-based methods like Rosetta, ProteinMPNN predicts protein sequences from protein backbone features. It effectively bridges the gap between structure prediction and sequence design, providing us with sequences predicted to fold into our de novo structures according to structure oracles such as AF2 (Dauparas et al., 2022).

The final computational stage includes a rigorous validation process, where the generated sequences are evaluated using AlphaFold for their structural predictability. AlphaFold, a groundbreaking approach in protein structure prediction, offers near-atomic accuracy, enabling us to confirm that our designed sequences indeed encode the designed structures. This validation step enhances the robustness of our design, ensuring that the resulting proteins meet our intended structural requirements (Jumper et al., 2021).

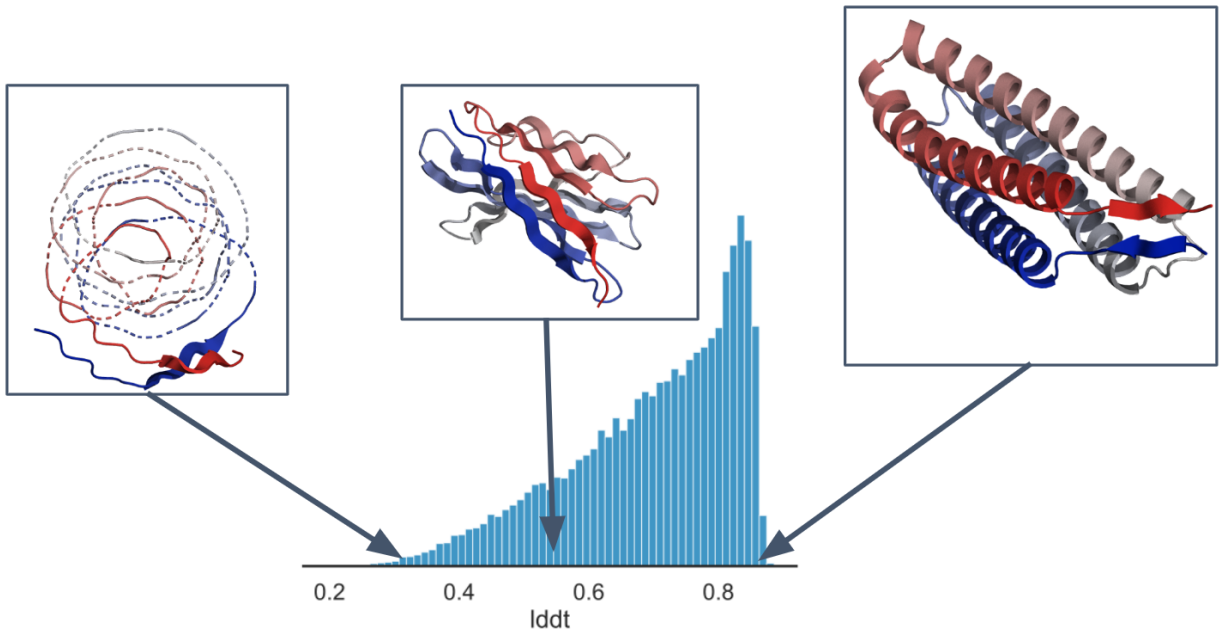
Following the computational design stage, we test our designs in the wet lab. Procedures followed published work in Watson et al., 2022. In brief: Initially, genes encoding our designed proteins are synthesized and received. The Golden Gate cloning method is utilized to create vectors containing these genes, paving the way for the next stage. The vectors are then introduced into bacterial cells through a transformation process. Bacteria serve as microscopic factories, reading the introduced genes and manufacturing the proteins they encode. This process facilitates the production of substantial quantities of our de novo proteins for further analysis and testing.

Upon successful transformation and protein production, the AKTA chromatography system is employed. This system is used to purify our proteins and evaluate their expression levels. Moreover, it provides us with the ability to determine the apparent molecular weight of the proteins by proxy of their elution volume on calibrated chromatography columns, validating that the synthesized proteins align with our designs.

Finally, to assess the physical properties of our engineered proteins, we employ atomic force microscopy (AFM). This technique allows us to visualize and probe our proteins at the nanoscale, providing us with invaluable information about their structures and mechanical properties. Specifically, AFM enables us to evaluate the mechanostability of the proteins, directly assessing the success of our design process.

Results:

Figure 1.1: pLDDT score distribution and representative protein structures.

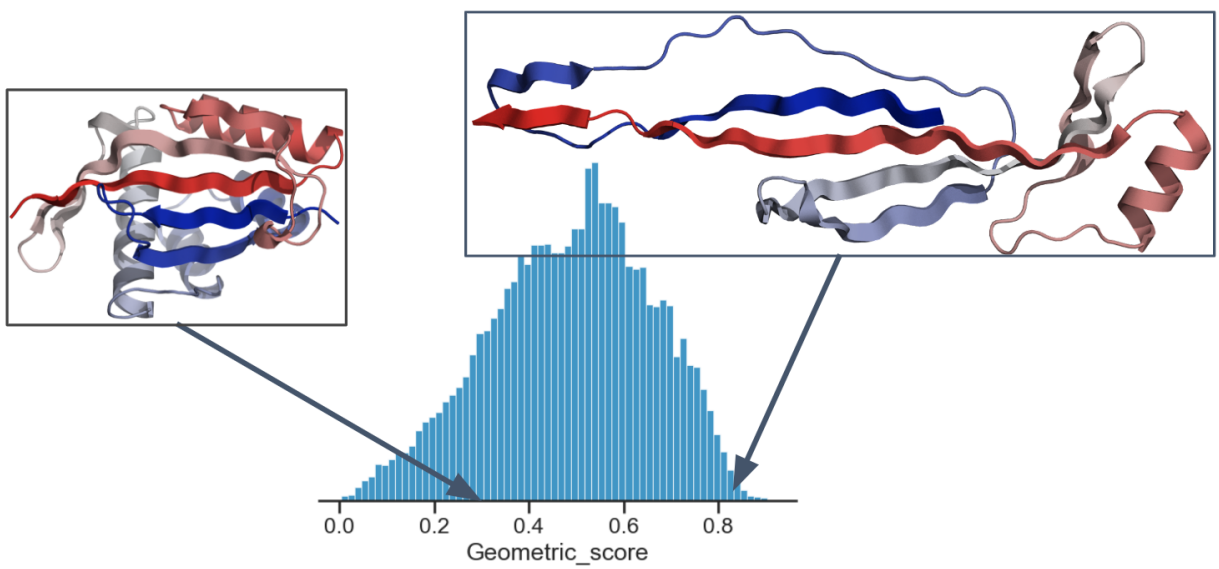


This histogram shows the distribution of pLDDT scores (X-axis) for all 47,600 de novo designs in the study (Y-axis). The inset images portray representative protein structures corresponding to different ranges of the pLDDT score. The leftmost inset shows a protein with a low pLDDT score (<0.8), characterized by its largely disordered structure. The middle inset represents a 'reasonable' protein with a score ranging between 0.8 and 0.9, exhibiting well-ordered secondary structures and an optimal balance between structural stability and design flexibility. The rightmost inset represents a protein with a high pLDDT score (>0.9), featuring an overabundance of alpha helices and a highly ordered structure that may limit functional versatility. This plot visualizes the relationship between pLDDT scores and structural features, aiding the identification of structurally viable proteins.

Here we present a plot demonstrating the pLDDT scores generated from the inpainting method , which we utilized as a cut-off to identify structurally viable proteins (see Figure 1.1). The plot is essentially a histogram of the scores, coupled with the visualization of de novo protein structures, allowing us to better understand the relationship between the pLDDT score and the structural attributes of the corresponding proteins, for all designs created in this campaign N= 47,600.

At the lower end of the pLDDT score spectrum, we can identify proteins that are predominantly disordered (see Figure 1.1). These proteins, characterized by low structural order and stability, fall short of the desired attributes we are aiming for in our de novo designs. The proteins with pLDDT scores ranging between 0.8 and 0.9 emerge as the most viable candidates, showcasing well-ordered secondary structures. These proteins are what we define as 'reasonable', i.e., they exhibit a desirable balance between structural stability and the potential to accommodate further design modifications. At the highest end of the pLDDT scale, we have proteins that exhibit an overrepresentation of alpha helices, translating into an overly ordered structure. While these proteins may offer high stability, they potentially lack the versatility necessary for functional applications in varying biological contexts.

Figure 1.2: Compaction score distribution and associated protein structures.

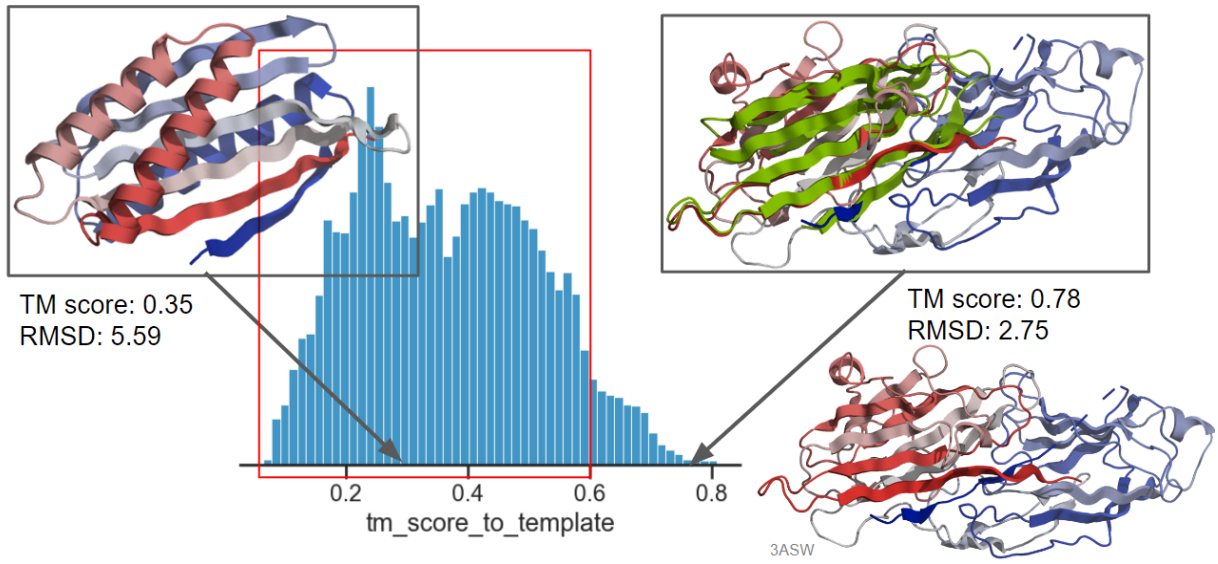


This histogram illustrates the distribution of compaction scores (X-axis) across all 47,600 de novo designs evaluated in this study (Y-axis). Inset images are provided at the histogram's extremities to represent protein structures that correlate with specific ranges of the compaction score. On the leftmost side, the inset features a highly compact protein structure associated with a compaction score below 0.5, indicating a potential for high mechanostability. Conversely, the rightmost inset showcases a less compact, more spread out protein structure corresponding to a compaction score above 0.6. This plot underscores the correlation between compaction scores and the physical attributes of the proteins, affirming the utility of the score in filtering designs for optimal mechanostability.

Following the assessment of pLDDT scores, we applied a similar analysis to the compaction score (see Figure 1.2), a metric representing the ratio of the two longest dimensions of a protein. Our rationale for using this score was to select for proteins that are more compact, rather than elongated or dispersed, based on the understanding that proteins with high mechanostability typically exhibit compact structures with closely arranged N- and C-termini.

Our analysis revealed a clear correlation between the compaction score and protein structure: proteins with scores below 0.5 were typically more compact and appeared to be reasonable designs, while those with scores above 0.6 often exhibited less compact and more spread out structures. Consequently, this cut-off value was used to further filter our pool of protein designs, ensuring a focus on designs with a higher potential for mechanostability.

Figure 1.3: TM-score distribution and representative protein structures



This histogram presents the distribution of TM-scores (X-axis) for all de novo proteins designed in the study (Y-axis). Inset images correspond to different TM-scores to exemplify the structural divergence from original templates. On the leftmost side, an inset showcases a protein with a low TM-score of 0.35, and a high RMSD of 5.59, indicating significant structural deviation from the original template. On the rightmost side, two insets are provided: the upper shows a superimposition of a de novo protein (in green) with a TM-score of 0.78, and an RMSD of 2.75 onto the native PDB 3ASW protein (in

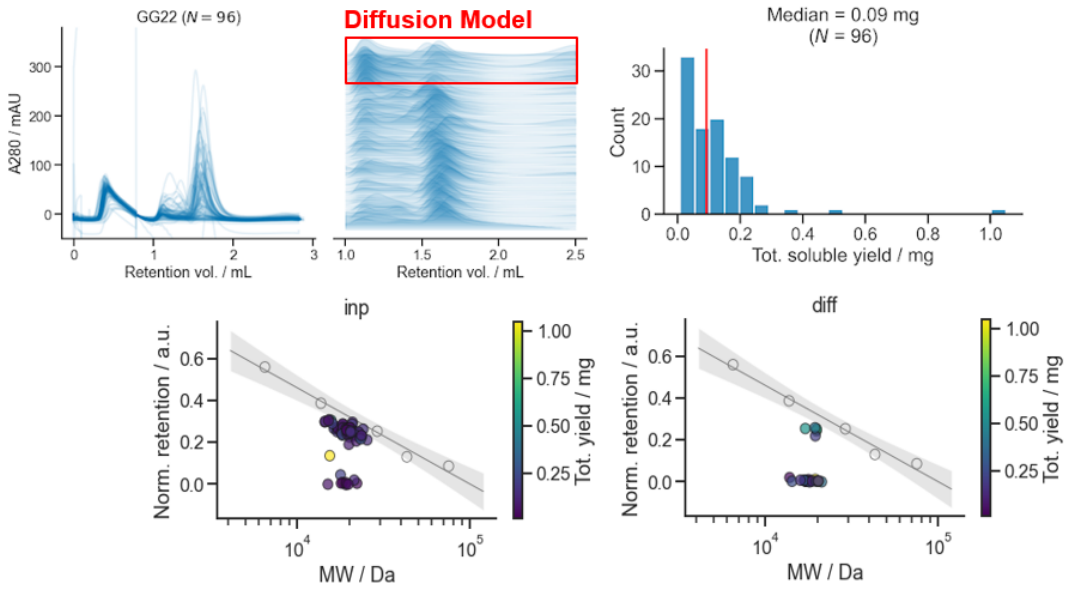
red-blue), illustrating a high degree of structural similarity. The lower inset depicts the native 3ASW protein in isolation for comparison. This plot encapsulates the relevance of the TM-score in assessing the novelty of generated protein structures, justifying its use as a cut-off value in our design process.

Lastly, we justify the implementation of the TM-score metric to the template and its corresponding cut-off value (see Figure 1.3). The intention of this graphical representation is to demonstrate the TM-score or Template Modeling score between the protein templates employed to generate new proteins and the resulting de novo proteins. Our objective is to achieve lower scores, indicating that the newly generated proteins bear significant structural differences from their original templates, preventing us from a mere recreation of existing folds. This comparison is visualized through histograms and the overlay of structures at various scores.

At scores above 0.6, we observe a strong similarity between the generated proteins (indicated in green) and their respective original proteins. For instance, a protein with a TM-score of 0.78 and RMSD (Root Mean Square Deviation) of 2.75 bears substantial resemblance to its original. Conversely, a protein with a TM-score of 0.35 and an RMSD of 5.59 significantly diverges from its original template in terms of structure, aligning with our objective of producing novel, structurally distinct proteins.

First Round of De Novo Protein Design Results:

Figure 1.4: Size exclusion chromatography analyses for de novo proteins



This composite figure contains five subplots organized in two rows, demonstrating distinct outcomes depending on the method of protein generation used.

The first row begins with a multi-line plot showcasing SEC data in milli-absorbance units (mAU) against retention volume. The second plot is an expanded version of the first, highlighting discrepancies in peak positions: the first peaks generated by the diffusion model are misaligned, while later peaks from the inpainting method align correctly. The third plot is a histogram displaying the total soluble protein yield (X-axis) against its count (Y-axis), with a median yield of 0.09 mg from a 4 mL expression culture.

The second row contains two normalized retention curve fitting plots, with normalized retention/au (Y-axis) against mw/da (X-axis). A gray curve with error radius represents the predicted retention, while colored dots represent individual proteins. The color indicates total yield, and the position along the y-axis represents the normalized retention. These plots contrast the inpainting method (left) against the diffusion method (right). For the inpainting method, most data points align closely with the predicted curve, suggesting high yields and correct molecular weights. Conversely, the diffusion

method plot shows most data points deviating from the predicted line, implying low yield and incorrect molecular weights.

The first round of de novo protein design yielded distinct outcomes depending on the method of protein generation used. This was assessed through size exclusion chromatography analyses, which plotted the absorbance in milli-absorbance units (mAU), the total soluble yields, and normalized retention against the molecular weight (see Figure 1.4).

Proteins engineered via the inpainting method displayed clearly distinct characteristics compared to the diffusion method. They exhibited on average high expression levels, clearly soluble yields, and their molecular weights aligned with the initial predictions. Notably, the data suggests these proteins did not aggregate or oligomerize.

On the other hand, the proteins generated using the diffusion method showed different attributes. Only three of these proteins showed a certain level of expression and a molecular weight within the expected range, a large fraction was either not found in the soluble fraction or ran as an aggregate on SEC.

Through atomic force microscopy (AFM), we investigated the mechanical stability of the de novo proteins that demonstrated appropriate molecular weights, yields, and expression levels in the first round. The AFM results provided the forces required to unfold these proteins mechanically.

Figure 1.5: Boxplot of rupture forces for de novo and natural proteins

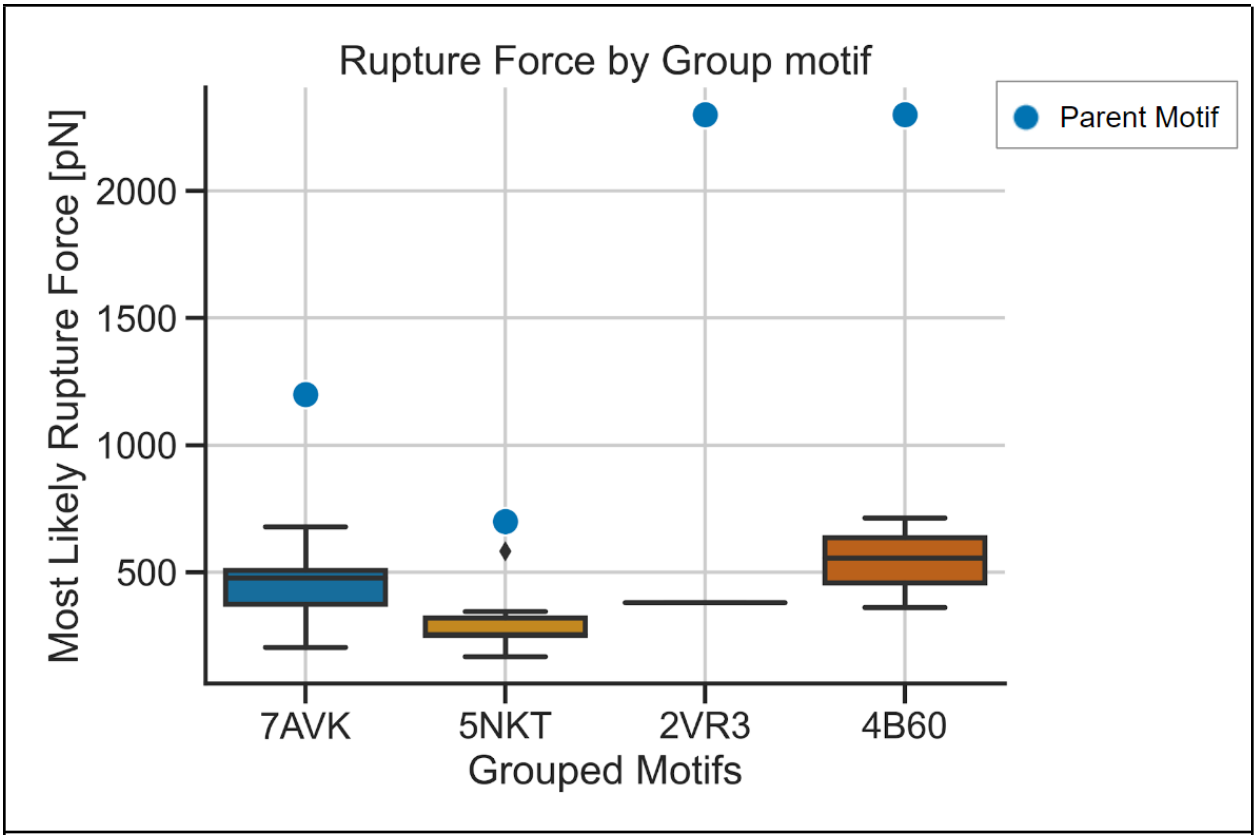


Figure 1.5: Boxplot of rupture forces for de novo and natural proteins. On the x-axis are different sets of proteins categorized by their original templates, and on the y-axis are the most likely rupture forces in picoNewtons (pN). The comparison emphasizes the difference in mechanical stability between designed proteins and their natural counterparts (blue dots), with most designs demonstrating lower rupture forces.

The rupture force for most of the de novo proteins did not match their natural counterparts (See Figure 1.5). The only exception was one set, where the natural protein had a rupture force around 700 picoNewtons (pN). The rupture forces of the designed proteins ranged from 250 to 700 picoNewtons, while the template proteins exhibited rupture forces from 700 to 2500 picoNewtons. These results highlight the variability in mechanical stability between the designed proteins and their natural counterparts, where designs still clearly fall short compared to the stability of natural folds.

Figure 1.6: Histogram of rupture forces for generated and natural proteins

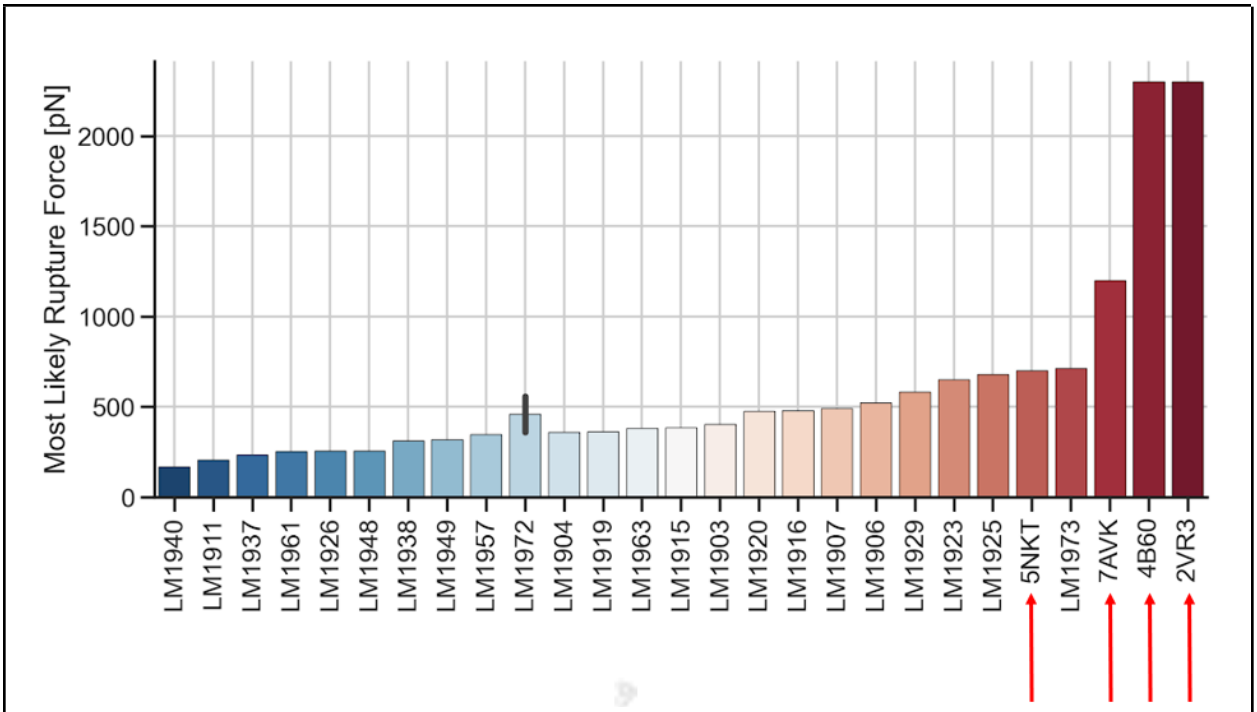
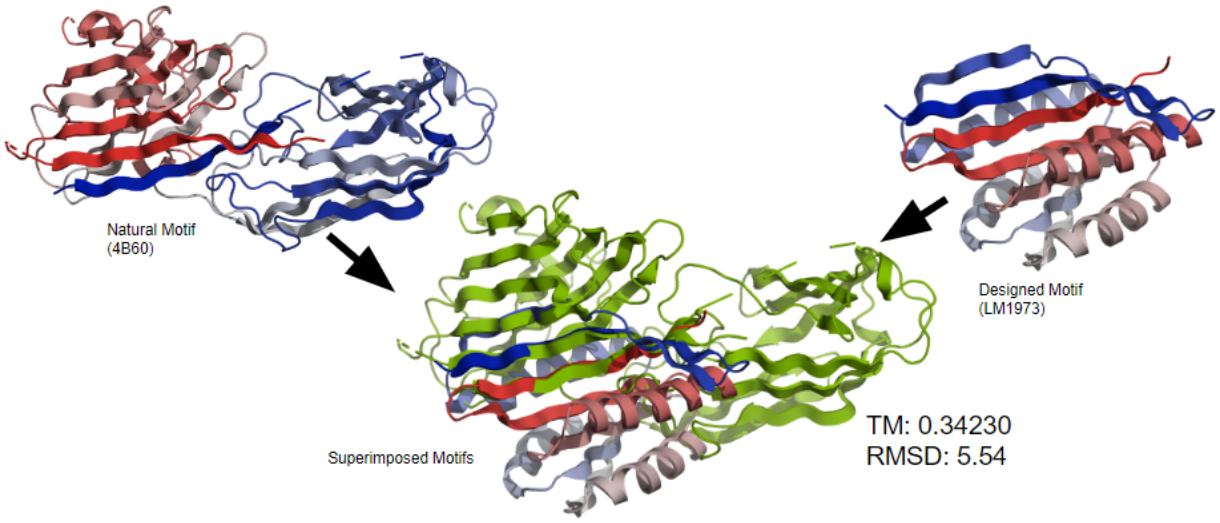


Figure 1.6: Histogram of rupture forces for generated and natural proteins. This figure illustrates the difference in mechanical stability between the de novo proteins and their natural counterparts (indicated by the red arrows), with the former showing lower rupture forces (250-700 pN) and the latter exhibiting higher forces (1200-2000+ pN). The contrast underscores the challenge in achieving comparable mechanostability in de novo proteins indicated by the "LM" prefix.

In this histogram figure, all generated and natural proteins are cataloged, demonstrating a clear distinction in rupture forces. Key natural template proteins predominantly exhibit a high rupture force range from 1200 to above 2000 picoNewtons, with a solitary protein showing a force around 700 picoNewtons. This high force range attests to the robust mechanical stability inherent in these natural proteins.

The de novo proteins we generated displayed lower rupture forces, from 250 to 700 picoNewtons. This discrepancy underscores that while these artificially designed proteins may possess other beneficial characteristics, they presently fall short in terms of mechanostability compared to their natural counterparts (See Figure 1.6). This highlights the future challenge of increasing the mechanical resilience of de novo proteins.

Figure 1.7: Superimposition and Structural Comparison of De Novo and Template Proteins



This figure presents a superimposition of two protein structures to illustrate their structural differences, reinforced by a TM-score of 0.34 and RMSD of 5.54. The goal here is to highlight the distinct architectural divergence of our de novo designed proteins from their respective parent templates. These significant structural differences, underscored by the low TM-score and high RMSD, clearly indicate that the de novo proteins we have generated bear little resemblance to their parent proteins, affirming the novelty of our design approach.

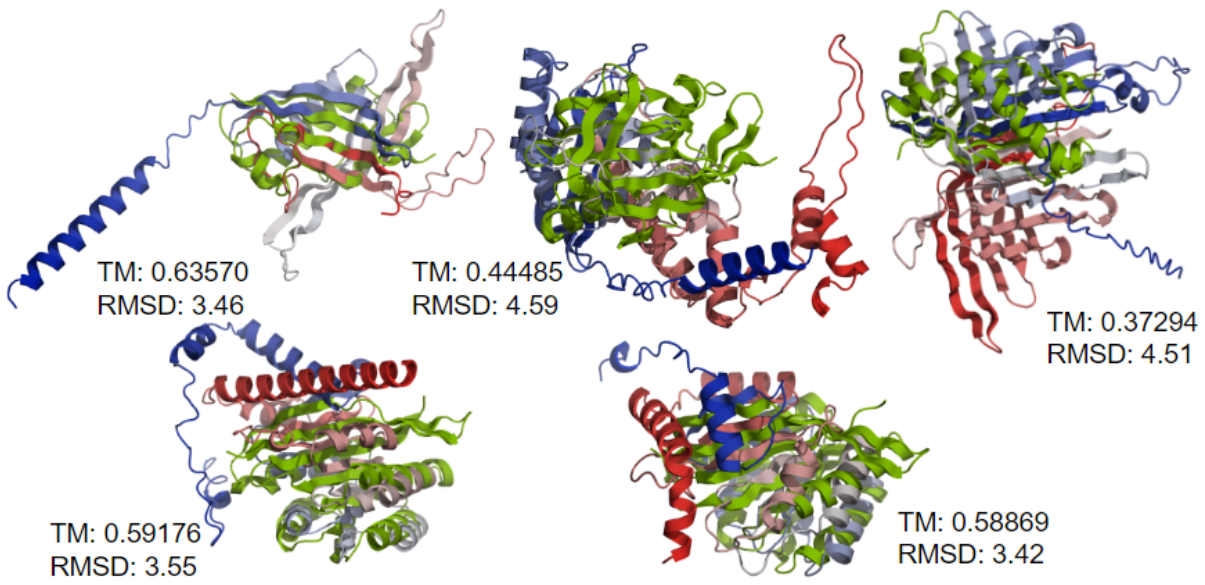
Building upon our analysis of protein viability and mechanostability using pLDDT and compaction scores, we further evaluate the novelty of our de novo designs via a direct structural comparison with their parent templates. Key to this evaluation are quantitative metrics such as the TM-score and RMSD, as well as a qualitative visual assessment of the protein structures.

Our results reveal a significant architectural divergence between the de novo proteins and their respective parent templates. This divergence is particularly exemplified in one instance, where a de novo protein exhibited a TM-score of 0.34 (general threshold for structural homology around TM-score of 0.5) and an RMSD of 5.54 when superimposed with its parent template (Figure 1.7). The TM-score, significantly less than 0.6, reflects a high degree of structural dissimilarity between the two proteins. Simultaneously, the RMSD value, well above the usual range for similar structures, indicates a significant deviation in atomic positioning.

These findings are reinforced by the visual inspection of the superimposed protein structures. The structural alterations, ranging from the overall protein shape to the placement of individual amino acid residues, are quite evident, reaffirming our quantitative findings.

Together, these results strongly indicate that our de novo proteins bear little resemblance to their parent proteins, testifying to the novelty and originality of our design approach. The low TM-scores and high RMSD values suggest that our designs have diverged from simple replications of existing structures and have successfully achieved unique, novel configurations.

Figure 1.8: Structural Discrepancy between Generated De Novo Proteins and Their Closest Natural Counterparts



This figure juxtaposes five of our generated de novo proteins (represented in green) with their closest natural counterparts, identified using the computational tool FoldSeek. Each pair is superimposed, visually highlighting the distinct structural differences, which are particularly noticeable in features such as elongated alpha helices present in the natural proteins. Accompanying each superimposed pair are their respective TM-score and RMSD values, providing a quantitative measurement of the structural differences. The visual and numerical data collectively underscore the innovative nature of our protein designs, which display significant divergence from existing natural protein structures.

Figure 1.8 underscores the originality of our de novo protein designs, as these designs do not mirror structures found naturally. Five examples of our generated proteins were selected and compared with the most structurally similar natural proteins, identified through FoldSeek, a computational method. In most instances, significant structural differences were observed, particularly in the presence of elongated alpha helices in natural proteins, which are absent in our de novo designs. The structural dissimilarities are quantitatively reinforced by low TM-scores and high RMSD values, indicating that our designs are substantially different from their closest natural counterparts. Hence, this comparison confirms the novelty of our approach and its ability to generate proteins that extend beyond existing natural structures.

Second Round of De Novo Protein Design:

In our second round of protein design, we targeted the redesign of proteins exhibiting high rupture forces. We chose to refine the design using the inpainting method, emphasizing a key structural features enhanced confinement of the N and C termini. The newly designed proteins were then gene synthesized, expressed, and purified for further analysis.

Figure 1.9:

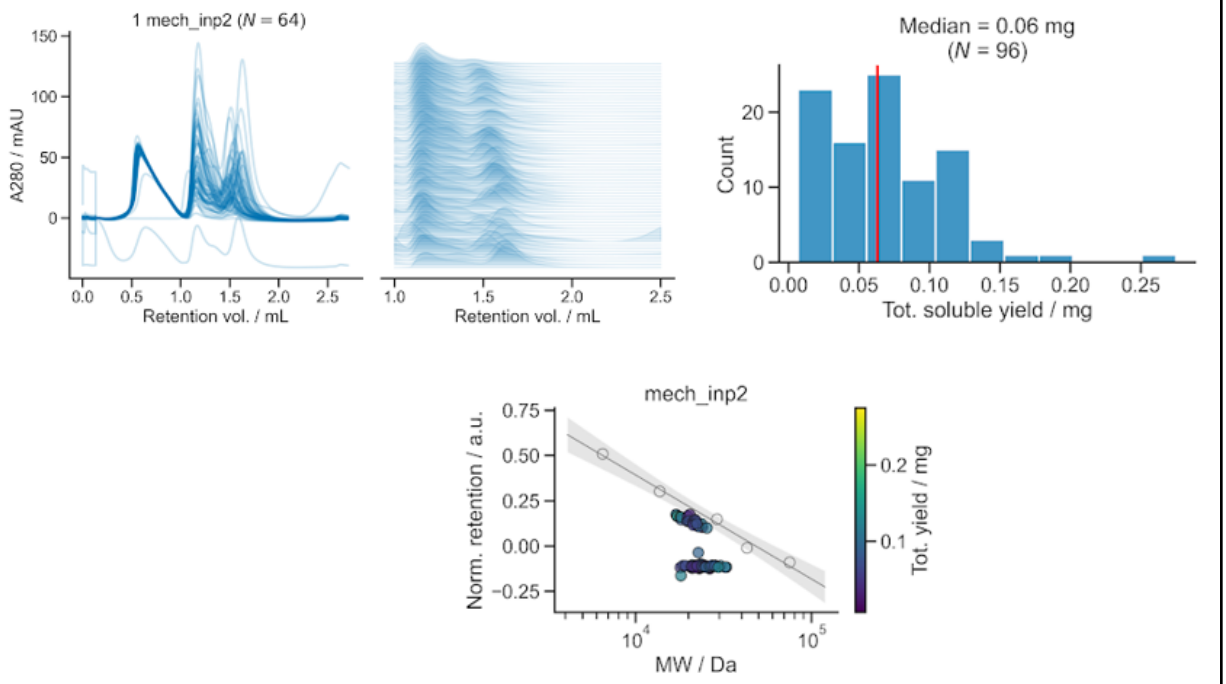


Figure 1.9 exhibits multiple plots evaluating the yield, expression, and molecular weight of the proteins generated in our study. The first row comprises three plots: The initial plot displays multiple lines representing various proteins, demonstrating the variability of their absorbance units (mAU) versus retention volume. The second plot is a spread-out version of the first, highlighting the data dispersion. The third plot is a histogram showing the total soluble yield of the proteins, with a median value at 0.06mg.

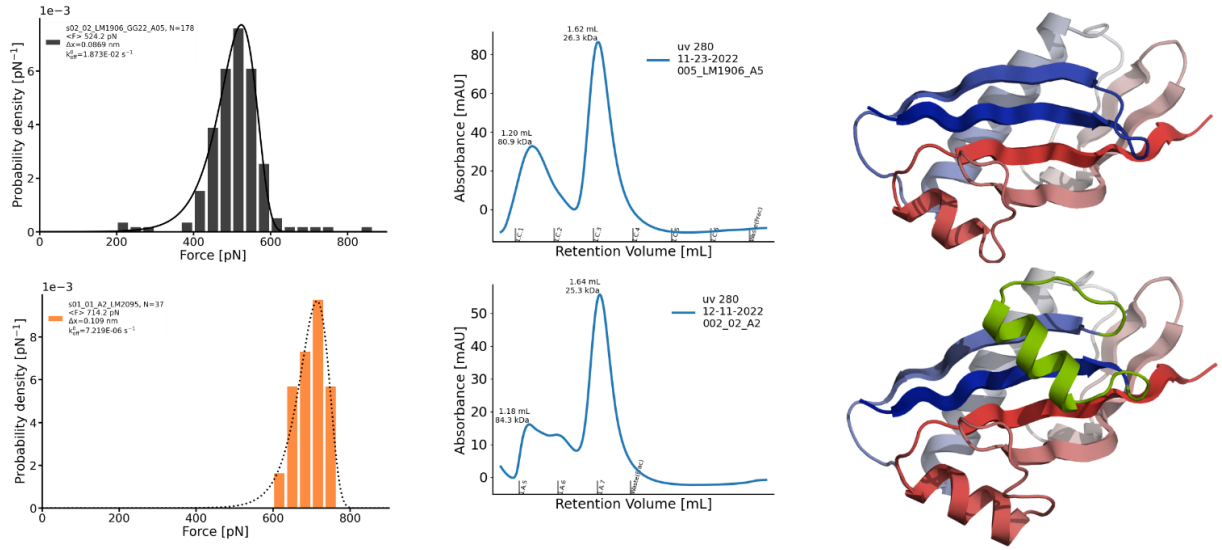
The second row features a single curve fit plot, with normalized retention versus molecular weight. Colored data points represent different proteins, with the color intensity indicating the total yield. The discrepancy between the dots and the curve line suggests the absence of absorbance for many proteins,

indicating a lack of soluble yield. The assortment of plots collectively highlight the variability and unpredictability of the proteins' characteristics in the current round of design.

Size exclusion chromatography was again employed to evaluate the yield, expression, and molecular weight of the generated proteins, providing an indication of whether these proteins were in an oligomerized or aggregated state. The results from this round were more variable than those from previous rounds. Approximately 40% of the proteins displayed the expected molecular weight, which marked a departure from the consistency observed in the prior rounds (see Figure 1.9). This finding suggests an element of unpredictability introduced with the structural refinements made in this round of protein design, when considering that a large part of the protein backbone remained the same.

In the final phase of our design experiment, we focused on a select protein that exhibited a reasonable predicted molecular weight and an acceptable yield from our prior rounds of design. To further investigate this protein's mechanical properties, atomic force microscopy was employed.

Figure 1.10:



The figure comprises six plots arranged in two rows. Each row corresponds to one protein variant - the original and the redesigned version - and consists of three plots each. The first plot in both rows is a histogram of the proteins' rupture force distribution as determined by atomic force microscopy, with the

rupture force in piconewtons displayed on the x-axis. In the first row, the peak rupture force for the original protein is found at approximately 524 pN. In contrast, the redesigned protein in the second row exhibits a higher peak rupture force at approximately 714.2 pN, indicating an improvement in mechanical stability of almost 200 pN.

The second plot in each row is a line plot illustrating the absorbance as a function of the retention volume, with peaks found at 1.62ml and 1.64ml for the original and redesigned proteins respectively.

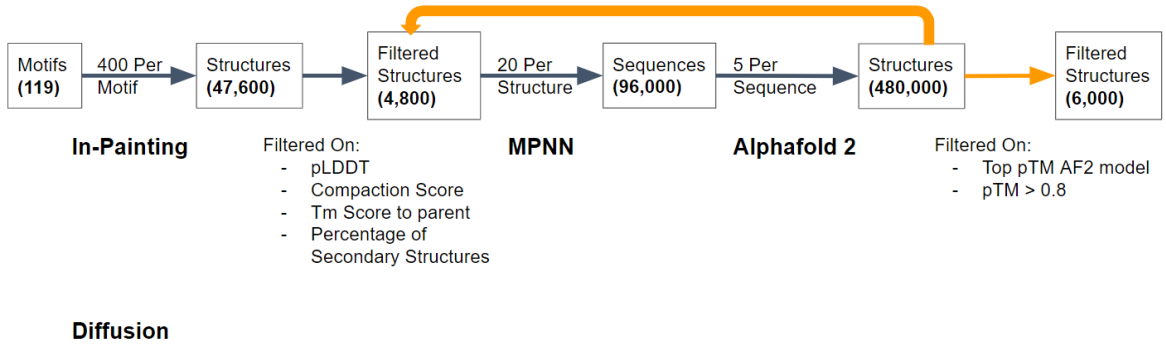
Finally, the third plot in both rows depicts a PyMOL ray trace of each protein. The original protein structure is shown in the first row, while the second row shows the redesigned protein with an additional green-colored alpha helix overlaying the N and C termini. The presence of this extra helix aligns with the observed increase in the protein's rupture force, thus providing visual evidence of the structural modifications leading to improved mechanostability.

Despite this variability from the second round of designs, one of the redesigned proteins, in particular, stood out (See Figure 1.10). This protein had initially demonstrated a most likely rupture force of approximately 524 piconewtons in the atomic force microscopy tests. Following redesign, this force increased to approximately 714 piconewtons, indicating a significant improvement in its mechanical resilience.

The in-silico structures of the original and redesigned proteins were compared, with a notable structural difference observed. The redesigned protein featured an additional alpha helix, which seemed to more securely envelop and confine the N and C termini's shear geometry. This structural change aligns with the improved rupture force observed, implying that these modifications have effectively strengthened the protein.

Methods:

Figure 1.11: Protein Design Pipeline



The figure illustrates our protein design pipeline, which begins with deep learning-based generation of initial protein structures from selected motifs, resulting in 47,600 designs. This initial pool undergoes filtering based on various metrics, reducing it to 4,800 promising designs. Each structure is then converted into 20 sequences using ProteinMPNN, yielding 96,000 protein sequences.

These sequences are fed into AlphaFold 2 to create structural models, with a subsequent round of filtering based on structural similarity to initial designs and other criteria, leaving 6,000 potential designs. A qualitative analysis narrows this down to the final 96 proteins. The pipeline highlights our iterative and stringent selection process, ultimately aiming at generating mechanically robust proteins.

Figure 1.12:

Name	Motif	Lower Limit	Higher Limit
5NKT_mod0.pdb	A1-8,5-80,A23-29,5-80,A62-68,5-80,A142-152	135	160
5NKT_mod0.pdb	A1-8,5-80,A23-29,5-80,A62-68,5-80,A142-152	75	200
5NKT_mod0.pdb	A1-8,5-80,A23-29,5-80,A62-68,5-80,A142-152	85	110
7AVK_f1.pdb	A2-11,5-80,A55-61,5-80,A73-83	135	160
7AVK_f1.pdb	A2-11,5-80,A55-61,5-80,A73-83	75	200
7AVK_f1.pdb	A2-11,5-80,A55-61,5-80,A73-83	85	110
2VR3_full_resorted.pdb	A1-10,5-200,A299-320	105	130

2VR3_full_resorted.pdb	A1-10,5-200,A299-320	135	160
2VR3_full_resorted.pdb	A1-10,5-200,A299-320	75	200
2VR3_motif_para.pdb	A1-12,5-200,A13-25	135	160
2VR3_motif_para.pdb	A1-12,5-200,A13-25	75	200
2VR3_motif_para.pdb	A1-12,5-200,A13-25	85	110
3ASW_full_resorted.pdb	A1-11,5-200,A299-331	135	160
3ASW_full_resorted.pdb	A1-11,5-200,A299-331	75	200
3ASW_full_resorted.pdb	A1-11,5-200,A299-331	85	110
3ASW_full_resorted.pdb	A2-10,5-80,A302-308,5-80,A318-329	135	160
3ASW_full_resorted.pdb	A2-10,5-80,A302-308,5-80,A318-329	75	200
3ASW_full_resorted.pdb	A2-10,5-80,A302-308,5-80,A318-329	85	110
3ASW_motif_para.pdb	A1-13,5-200,A14-24	135	160
3ASW_motif_para.pdb	A1-13,5-200,A14-24	75	200
3ASW_motif_para.pdb	A1-13,5-200,A14-24	85	110
3TIP_SasG.pdb	A1-13,5-80,A27-32,5-80,A44-51,A68-88,5-80,A122-32	135	160
3TIP_SasG.pdb	A1-13,5-80,A27-32,5-80,A44-51,A68-88,5-80,A122-32	75	200
3TIP_SasG.pdb	A1-13,5-80,A27-32,5-80,A44-51,A68-88,5-80,A122-32	85	110
3TIP_SasG.pdb	A1-17,5-80,A30-32,5-80,A46-49,A68-72,5-80,A81-85,5-80,A122-128	135	160
3TIP_SasG.pdb	A1-17,5-80,A30-32,5-80,A46-49,A68-72,5-80,A81-85,5-80,A122-128	75	200
3TIP_SasG.pdb	A1-17,5-80,A30-32,5-80,A46-49,A68-72,5-80,A81-85,5-80,A122-128	85	110
4B60_motif_para.pdb	A1-15,5-200,A16-29	135	160
4B60_motif_para.pdb	A1-15,5-200,A16-29	75	200
4B60_motif_para.pdb	A1-15,5-200,A16-29	85	110

Initial Protein Design Generation:

We initiated the design process using deep learning (in-painting) to generate protein structures from 119 selected motifs. The general parameters are given in figure 1.12. This approach led to the creation of 400 structures per motif, amassing a total of 47,600 initial designs.

Protein Design Filtering:

We subjected the initial designs to a round of filtering based on defined metrics, including pLDDT scores, compaction scores (which assess the ratio of the two longest dimensions of the protein), TM scores to natural proteins (to ensure structural diversity from the natural proteins), and the percentage of secondary structures such as beta sheets and disordered regions. This vetting process narrowed down our pool to 4,800 promising protein candidates.

Sequence Generation:

To transform our selected structures into sequences, we employed ProteinMPNN, a deep learning-based method. Each of the 4,800 structures were expanded to 20 corresponding sequences, culminating in 96,000 protein sequences.

Structure Generation with AlphaFold 2:

We then used AlphaFold 2 to generate 5 structural models (one for each AF2 model) per sequence. This amounted to a large pool of 480,000 protein structures.

Final Filtering and Selection:

Next, the AlphaFold 2 designs underwent a second round of filtering, this time based on their structural similarity to the initial in-painting designs. This filtering, along with the subsequent qualitative analysis, was crucial in refining our protein candidates. Filtering was based on a pTM score for the top AlphaFold 2 model greater than 0.8. This narrowed our pool to 6,000 potential designs.

After this quantitative filtering, we conducted a qualitative analysis to select the final 96 proteins. Our criteria were built on the principles of mechanostability and successful expression, selecting proteins that had a hydrophobic core, compactness, and an N and C termini in sheer geometry with well-confined ends. This careful selection process was key to achieving our aim of designing mechanically robust proteins.

For the second round of design, we looked for similar quantitative results but placed extra emphasis on further confinement in our qualitative analysis. We targeted additional alpha helices or other secondary

structures that cover the N and C termini, hypothesizing that greater confinement would enhance the protein's mechanostability.

Conclusion:

This thesis represents a significant step forward in de novo protein design, targeting the creation of proteins with enhanced mechanostability. Leveraging advanced computational techniques, innovative design principles, and meticulous lab work, we have explored new territories in the de novo design space.

Our endeavor began with a computational approach, inspired by principles observed in naturally occurring proteins. The adopted inpainting technique, supported by deep learning tools such as RoseTTAFold and AlphaFold, proved to be an effective method in generating novel protein structures. The sequence generation and rigorous validation processes, assisted by ProteinMPNN and AlphaFold, ensured that our design pipeline produced proteins that closely matched our intended structures.

Despite the varying outcomes in the initial round of protein design, the techniques we employed have yielded promising results. Notably, the proteins generated using the inpainting method exhibited the desired expression levels, yields, and molecular weights that align with our design specifications. This progress was further enhanced in the second round of design, where we emphasized further confinement of the protein structure.

The mechanical resilience of our de novo proteins, a key design goal, was assessed using atomic force microscopy. While none of the de novo proteins surpassed the mechanical stability of their parent proteins, one protein from the second round of design exhibited a notable increase in rupture force, signifying an improvement in its mechanical stability. This enhancement was attributed to structural refinements, including an additional alpha helix confining the N and C termini, reflecting the effectiveness of our design principles.

The variability observed in the second round of designs emphasizes the complexity of protein design, where minor modifications can significantly impact the resulting proteins. Nevertheless, it also highlights the importance and potential of the iterative design process in protein engineering. Through continuous refinement and testing, we can inch closer to our goal of creating proteins with enhanced mechanostability.

In conclusion, our work contributes to the unfolding story of protein engineering, blending the understanding of protein mechanostability with state-of-the-art computational and laboratory techniques.

Despite the challenges faced and the complexities intrinsic to protein design, our results affirm the potential of our approach and indicate an exciting path forward. This thesis shows progress we have made but also a stepping-stone towards the exciting possibilities that de novo protein design offers. In harnessing the principles of natural proteins to design and engineer novel proteins, we can pioneer innovative solutions in biomaterials, therapeutics, and biotechnology, thus opening new horizons in the realm of protein science.

Chapter 2: Isopeptide Design

Abstract:

This study leverages the advances in computational protein design to develop a novel generation of isopeptides derived from de novo proteins, using the state-of-the-art tool, RoseTTAFold Diffusion (RFdiffusion). The research adopted a multi-step approach, including initial protein design generation, stringent filtering, partial diffusion, sequence generation, AlphaFold 2 based structure generation, and final selection. These de novo isopeptides were then synthesized, and their properties were verified using high-performance liquid chromatography (HPLC) and mass spectrometry. The HPLC results matched expectations, confirming successful protein synthesis. However, mass spectrometry indicated no isopeptide bond formation, possibly due to the occlusion of the active site by a phenylalanine residue. Despite the promising *in silico* and HPLC results, the absence of isopeptide bonds underscores the necessity for algorithmic refinement. The research reflects the vast potential and current challenges of machine learning applications in protein engineering and design, and advances our understanding of de novo protein generation towards design of function.

Motivation:

Protein design harbors tremendous potential for transformative advances in biotechnology, medicine, and fundamental biochemical research. In particular, isopeptides, proteins characterized by unique isopeptide bonds, have sparked considerable interest. The distinctive properties of these isopeptide bonds provide an essential tool for stabilizing proteins, making them resilient against environmental stressors. Furthermore, the very nature of isopeptides opens a multitude of opportunities beyond stability, such as the creation of split systems like the SpyCatcher/SpyTag system (Zakeri et al. 2012). This facilitates the design of modular proteins and the construction of complex biomolecular architectures, pushing the limits of what can be achieved in protein engineering.

Despite these promising attributes, the design and synthesis of isopeptides have proven to be intricate and challenging. Previous successes have relied heavily on designs that closely mirror native structures, an approach that inherently restricts flexibility and could yield suboptimal outcomes. To bypass these limitations, this research is motivated to employ cutting-edge machine learning tools, specifically RoseTTAFold Diffusion (RFdiffusion), to design a new generation of derived from de novo proteins.

In essence, the motivation of this research lies in leveraging state-of-the-art computational methods to transcend conventional protein design paradigms, thereby unlocking the vast, underexplored potential of isopeptides. This investigation represents a significant leap forward in our pursuit to comprehend and manipulate the complex world of proteins, anticipating transformative applications in various scientific and technological domains.

Introduction:

In the intricate world of proteins, the ability to design, manipulate, and control their structures and functions has profound implications in various fields, including biotechnology, medicine, and fundamental biochemical research. Recent advances in computational techniques, particularly deep learning methods, have revolutionized protein design, enabling researchers to solve complex challenges in protein engineering. This study focuses on harnessing the power of these computational tools to design second-generation isopeptides, thereby expanding the protein design space.

Isopeptide bonds, specifically those formed intramolecularly between lysine and asparagine or aspartic acid residues, serve as a unique essential component found in nature's toolbox for stabilizing proteins in the case of bacterial pili and viral capsids (Kwon et al., 2017). These bonds, found in bacterial cell-surface proteins, form autocatalytically during protein folding when the reacting groups converge within a hydrophobic environment (Kang & Baker, 2011). Such intramolecular isopeptide bonds not only bolster resistance against chemical, thermal, and mechanical stress, but also open new biochemical options for applications in protein crosslinking.

While isopeptide bonds are an integral part of protein stability, their design and creation have been challenging. A critical advancement in this direction has been the successful design of isopeptides with older tool inpainting methods (unpublished). However, to navigate the design space with more flexibility, this research seeks to employ more recent tools such as diffusion models, specifically RoseTTAFold Diffusion (RFdiffusion) for isopeptide design.

RFdiffusion, developed by fine-tuning the RoseTTAFold structure prediction network on protein structure denoising tasks, exhibits remarkable proficiency in protein design. It has shown exceptional performance on unconditional and topology-constrained protein monomer design, protein binder design, symmetric oligomer design, enzyme active site scaffolding, and symmetric motif scaffolding for therapeutic and metal-binding protein design (Keeble et al., 2019).

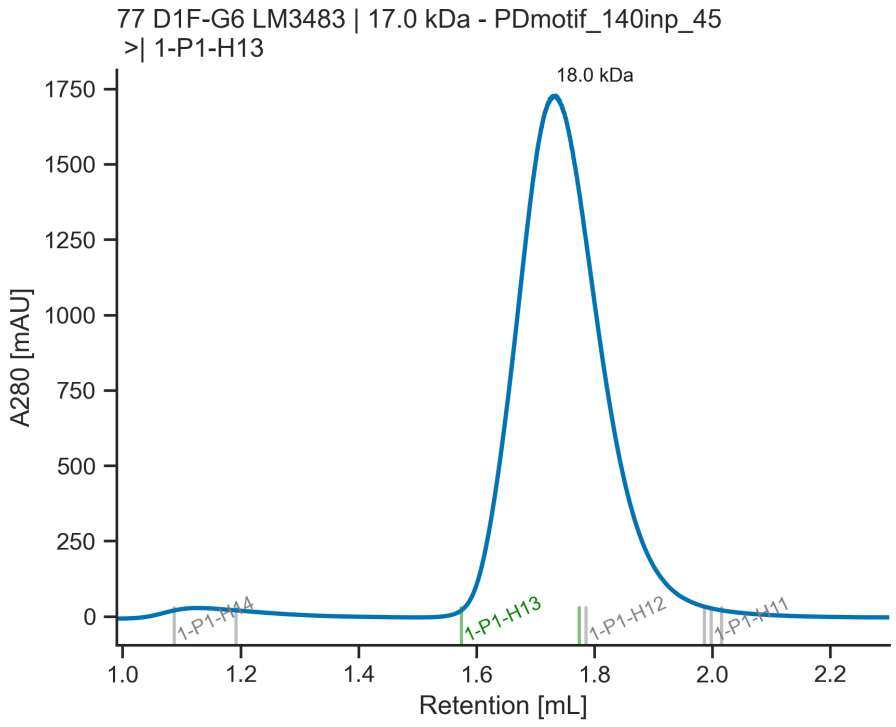
By integrating RFdiffusion into the process, this research aims to create a new generation of isopeptides derived from de novo isopeptide protein designs - another layer of abstraction from the original native protein template. Not only does this advance the field of isopeptide design, but it also underpins the enormous potential of machine learning in protein engineering.

In the coming chapters, this thesis will delve deeper into the methodology and findings of applying machine learning to engineer advanced isopeptides. This study contributes to the burgeoning field of protein engineering, underlining the possibilities when computational prowess meets biological complexity.

Results:

Here we present the results from our design and synthesis of de novo isopeptides. After creating these novel proteins through an in-silico design pipeline, we ordered the corresponding genes and conducted bacterial transformation for protein production.

Figure 2.1: HPLC Chromatogram of De Novo Protein Analysis



This figure presents an HPLC chromatogram, employed to determine the approximate molecular weight and quantity of the de novo proteins. The y-axis corresponds to the detector response in milli-absorbance units (mAU), while the x-axis represents the retention volume in milliliters (mL). Peaks in the plot signify components in the mixture, with their height and area indicating the component's relative concentration.

The key results from our first set of analysis come from High-Performance Liquid Chromatography (HPLC), a powerful technique used to separate, identify, and quantify each component in a mixture. HPLC was employed to determine the molecular weight of the proteins and to verify their overall quality.

The HPLC chromatogram plot depicts the detector response (in mAU) on the y-axis and the retention volume (in mL) on the x-axis (See Figure 2.1).

For our de novo isopeptides, the anticipated molecular weight was 17 kDa. The representative HPLC results shown here corroborated this prediction, indicating an approximate detected molecular weight of 17 kDa. This result aligns with our expectation and suggests successful expression and purification of our de novo isopeptides. These findings validate our design pipeline's initial steps and pave the way for subsequent stability and functionality analyses of these proteins.

Continuing with our results analysis, the proteins that demonstrated satisfactory results in the HPLC data were subjected to further evaluation using mass spectrometry. This procedure was performed specifically to verify the presence of an isopeptide bond within our de novo proteins. Given that the formation of an isopeptide bond should result in a weight loss of about 18 Da corresponding to the loss of an NH3 or H20 molecule upon peptide bond formation, any such observation in our mass spectrometry data would be a definitive indicator of isopeptide bond formation.

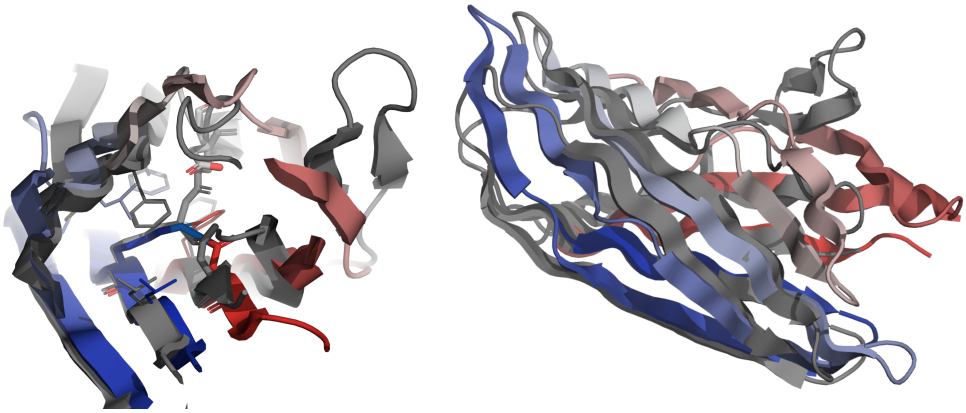
Figure 2.2:

Sample name	Expected average mass (Da)	Concentration (mg/ml)	Observed average mass (Da)	Difference (Da)
LM3411	18288	0.4	18157	131
LM3412	18563	0.4	18432	131
LM3418	15046	0.4	14915	131
LM3419	15083	0.4	14952	131
LM3420	15197	0.4	15066	131
LM3421	19225	0.4	19094	131
LM3426	17003	0.4	16872	131
LM3427	16796	0.4	16665	131
LM3429	16616	0.4	16485	131
LM3430	16468	0.4	16337	131
LM3431	17213	0.4	17082	131
LM3432	16949	0.4	16817	132
LM3433	16821	0.4	16690	131
LM3435	14742	0.4	14610	132
LM3437	18433	0.4	18302	131
LM3439	18722	0.4	18591, 18722	!
LM3440	17585	0.4	17454	131
LM3441	17649	0.4	17518	131
LM3443	16455	0.4	16324	131
LM3444	16652	0.4	16521	131
LM3445	17224	0.4	17093	131
LM3446	17478	0.4	17347	131
LM3447	17170	0.4	17039	131
LM3451	18480	0.4	18350	130
LM3453	17004	0.4	16873	131
LM3454	17377	0.4	17246	131
LM3455	17115	0.4	16984	131
LM3457	17729	0.4	17598	131
LM3458	18606	0.4	18475	131
LM3459	19005	0.4	18874	131
LM3460	17315	0.4	17185	130
LM3461	17267	0.4	17136	131
LM3462	17016	0.4	16885	131
LM3463	17037	0.4	16906	131
LM3464	17131	0.4	17000	131

LM3465	16967	0.4	16836	131
LM3466	16855	0.4	16724	131
LM3473	20147	0.4	20016	131
LM3474	16384	0.4	16253	131
LM3475	15558	0.4	15427	131
LM3478	18408	0.4	18277	131
LM3479	18910	0.4	18779	131
LM3480	16057	0.4	15926	131
LM3482	16897	0.4	16766	131
LM3483	16962	0.4	16831	131
LM3484	17163	0.4	17032	131
LM3486	17198	0.4	17067	131

However, the data depicted in the subsequent table reveals no noticeable differences except for a 131 Da decrease, which corresponds to an expected methionine loss (See Figure 2.2), but not the additional expected 18 Da loss. Thus indicating that no Isopeptide bond forms for any of the de novo proteins generated.

Figure 2.3: Superimposed In-silico Structures of Parent and De Novo Proteins



This figure depicts the superimposition of parent and de novo generated protein structures. The parent protein structure is colored gray, while the de novo protein structure is distinctly highlighted. A notable feature is the differing residue, where a phenylalanine of the de novo protein overlays the active site. This difference seems to obstruct the formation of the isopeptide bond, causing a deviation from the desired outcomes

The final piece of our data analysis features in-silico structures with superimposed parent and de novo generated proteins (See Figure 2.3). The parent protein is depicted in gray, while the de novo protein is highlighted. The line residue shows a significant difference, with the phenylalanine of the de novo protein overlaying the active site. This appears to hinder the formation of the isopeptide bond, leading to a deviation from our desired result.

Methods:

Please refer to the methods outlined in Chapter 1 as they were adapted and applied in Chapter 2.

Mass Spectrometry: Mass spectrometry was utilized to analyze the molecular weight of the expressed proteins. This technique was used to verify the expected and observed molecular weights of the designed proteins, with a particular focus on identifying a 18 Da difference indicative of isopeptide bond formation. The analysis provided an in-depth view of the protein structures and confirmed the presence or absence of isopeptide bonds.

Conclusion:

This thesis aimed to explore the malleability of de novo isopeptide design, thereby challenging the pre-existing premise that successful de novo isopeptides could only be derived when closely mirroring native structures. To achieve this, we undertook an investigative study to design a second-generation set of isopeptides, derived from de novo proteins.

Our results indicated a rather rigid system with minimal room for deviation in protein design. In silico metrics for generated proteins were robust, and high-performance liquid chromatography (HPLC) data exhibited good data quality , however, isopeptide bond formation was not observed. Detailed analysis revealed an intriguing finding; a phenylalanine residue in a de novo protein was seen to potentially occlude the active site, which is hypothesized to be a cause of isopeptide bond formation failure.

The de novo protein demonstrated preserved active sites, and AlphaFold's analysis showed overall similar folds between parent and proteins derived from their template. Nevertheless, the slight alteration in the active site potentially disrupted the isopeptide bond formation. It suggests that more precise modeling or realignment could offer a solution to generate second-generation isopeptides successfully.

This thesis represents an ambitious attempt to employ the cutting-edge computational protein design tool, RoseTTAFold Diffusion (RFdiffusion), to develop a novel generation of isopeptides, derived from de novo proteins. The process spanned from initial protein design generation to final filtering and selection, merging traditional biochemical techniques with modern computational methodologies.

Despite the proteins displaying promising traits according to in-silico and HPLC analyses, the anticipated isopeptide bond formation was not confirmed via mass spectrometry. This finding underscores the need for refining our design algorithm to make more precise and accurate predictions for isopeptide bond formation.

The present work showcases both the potential and challenges of using machine learning methods in protein design. While RFdiffusion demonstrated remarkable prowess in generating diverse and promising isopeptide structures, it also revealed the limitations of current approaches in predicting isopeptide bond formation.

In future research, further refinements to the RFdiffusion model, including the parameters and configurations used, may yield improved isopeptide bond formation. Additionally, incorporating other metrics in the filtering process could enhance the likelihood of selecting structures that will ultimately form isopeptide bonds. An expanded study with a larger pool of designs may provide broader insights into the performance of our approach.

In conclusion, this research marks an important step toward the goal of understanding and manipulating protein structures. It establishes that the integration of state-of-the-art computational tools such as RFdiffusion with traditional biochemical analyses can indeed generate promising isopeptide designs. Although challenges remain, this research points to the potential of artificial intelligence in protein design, with implications for fields such as biotechnology and medicine. As we continue to refine these methods, we edge closer to the prospect of fully harnessing the power of proteins.

References:

1. Milles, L.F. et al. (2018) ‘Calcium stabilizes the strongest protein fold’, Nature Communications, 9(1). doi:10.1038/s41467-018-07145-6.
2. Ponnuraj, K. et al. (2003) ‘A “dock, Lock, and latch” structural model for a staphylococcal adhesin binding to fibrinogen’, Cell, 115(2), pp. 217–228. doi:10.1016/s0092-8674(03)00809-2.
3. Persat, A. *et al.* (2015) ‘The mechanical world of bacteria’, *Cell*, 161(5), pp. 988–997. doi:10.1016/j.cell.2015.05.005.
4. Milles, L.F., Schulten, K., et al. (2018) ‘Molecular mechanism of extreme mechanostability in a pathogen adhesin’, Science, 359(6383), pp. 1527–1533. doi:10.1126/science.aar2094.
5. Wang, J. *et al.* (2022) ‘Scaffolding protein functional sites using Deep Learning’, *Science*, 377(6604), pp. 387–394. doi:10.1126/science.abn2100.
6. Watson, J.L. *et al.* (2022) *Broadly applicable and accurate protein design by integrating structure prediction networks and Diffusion Generative Models* [Preprint]. doi:10.1101/2022.12.09.519842.
7. Dauparas, J. *et al.* (2022) ‘Robust deep learning–based protein sequence design using proteinmpnn’, *Science*, 378(6615), pp. 49–56. doi:10.1126/science.add2187.
8. Jumper, J. *et al.* (2021) ‘Highly accurate protein structure prediction with alphafold’, *Nature*, 596(7873), pp. 583–589. doi:10.1038/s41586-021-03819-2.
9. Kwon, H. et al. (2017) ‘Engineering a Lys-ASN isopeptide bond into an immunoglobulin-like protein domain enhances its stability’, Scientific Reports, 7(1). doi:10.1038/srep42753.
10. Keeble, A.H. et al. (2019) ‘Approaching infinite affinity through engineering of peptide–protein interaction’, Proceedings of the National Academy of Sciences, 116(52), pp. 26523–26533. doi:10.1073/pnas.1909653116.
11. Kang, H.J. and Baker, E.N. (2011) ‘Intramolecular isopeptide bonds: Protein crosslinks built for stress?’, Trends in Biochemical Sciences, 36(4), pp. 229–237. doi:10.1016/j.tibs.2010.09.007.