Computational Evaluation of Staples on Helicons

Oliver Kanders
Advisor: Ritambhara Singh, Reader: Brenda Rubenstein

A thesis submitted in partial fulfillment of the requirements for the degree of Bachelor of Science with Honors in Computer Science at Brown University

BROWN UNIVERSITY
PROVIDENCE, RI
MAY 2024
1 Abstract

Helicons represent a new frontier in therapeutics, in which constrained α-helical polypeptides are used to target proteins that may be inaccessible to small-molecule or antibody therapies. Recently, there has been a growth of experimental research dedicated to constructing these peptides and locating previously unrecognized α-helical binding sites. In order to constrain these peptides, the placement and chemical characterization of covalent cross-linkers (“staples”) becomes crucial. To date, very little research has focused on customizing staples’ attributes, how staples function while remaining bound to peptides, and the staples’ helicity. Our work focuses on evaluating staples on an α-helix and identifying potential residues for stapling. We applied a joint calculator of DFT-D4 and AIMNet2 to measure the total energy of various staple conformations and understand the magnitude of deviation of these conformations from the lowest energy reference state. We notably found the staple WHL and other variants, like its chlorine-substituted WHL-Cl1, show tight clusters of low-energy conformations close to the target α-helix template, implying a potential comparable functionality as a linker. Once a pipeline existed for evaluating successful staples, we deployed a model to determine possible residues on an α-helix that a disulfide staple could constrain. We hoped to generate candidate staples that would be computationally optimized with our energy calculations. We generated a hash table containing tens of thousands of disulfide geometries, creating a comprehensive database of successful chi angles (χ) observed along the side chains of cysteine residues. By comparing any given α-helix to our hash table, we can efficiently locate potential stapling opportunities, identifying positions in which the chi angles favor the formation of stable disulfide bonds. We successfully optimized the hash table to include over 50,000 extracted native disulfide bridge geometries, expanding the set of potential target peptides appropriate for stapling.
2 Introduction

Over the past decade, there has been a growing interest in the power of α-helical peptides for therapeutic purposes. The α-helix is a prevalent structural motif in proteins, often involved in surface recognition [1, 9, 14]. Its distinctive capability to shield amide groups lends itself well to enabling α-helical polypeptide motifs to be accommodated within membrane environments. Capitalizing on these attributes, inspiring research has been undertaken into developing chemically stabilized α-helical peptide therapeutics [9]. These innovative therapeutic agents have shown promise in traversing cellular membranes and engaging with previously considered "undruggable" targets, notably within the realm of protein-protein interactions [1, 9]. Stabilization of an α-helical conformation is achieved via covalent cross-linking of unnatural amino acids introduced at specific positions, a technique referred to as “stapling” [1, 13]. The general design of staples depends on several key factors: staple position, staple stereochemistry, and staple length [13]. Typically, experimentation with stapled peptides is conducted using phage display techniques, which enable high-throughput screening of peptides [9]. Phage display offers high library diversities, shows tolerance to chemical manipulation, and allows for rapid and cost-effective library generation [9, 15]. However, developing stapled peptides tends to be a labor-intensive, costly, and time-consuming process, especially compared to the efficiencies and lower costs associated with in silico approaches [13]. Computational methods can develop in silico models, which may help in reducing the number of candidate stapled peptides to be tested experimentally and to rationalize specific experimental observations [4, 5, 6, 7, 8, 10, 12, 13].

Our first aim in this work centers on developing a computational pipeline to draw out relationships of total energy and RMSD to reference states in order to determine the performance of various staple molecules in constraining and stabilizing the target α-helix structure, specifically the Helicon FP28136—a successful staple from Li et al. (2022) [9]. Our second aim focuses on improving an existing framework for identifying potential residues on a peptide that could be suitable for a disulfide bridge [16].

To realize these aims, we first needed a process to determine how our staples were performing at constraining the α-helix structure. We drew from other applications of computational staple evaluation, observing techniques deployed for the identification of optimal staples for antiparallel β-strands with noncanonical amino acids. Working with an experimental dataset of Helicons from Greg Verdine’s paper “De novo mapping of α-helix recognition sites on protein surfaces using unbiased libraries,” which was successful in developing stapled α-helices (in other words, Helicons) through phage display [9], we derived new insights on the characteristics of an optimized staple. In this work, we will focus on the experimental staple WHL (identified as N,N’-(1,4-phenylene)diacetamide) that is bound to Helicon peptide FP28136 [9].

Leveraging algorithms to evaluate the compatibility of the experimental staple WHL (put forth in the work of unbiased mappings of α-helix binding sites), we measured the alignment of WHL to its target peptide structure [9]. We then created a greater collection of chemically similar staples, altering the aromatic ring primarily, while maintaining the other chemical groups (acetyl group (CC(=O)) and methyl group (C(=O)C)). The collection provided a diverse display of staples and their alignment with the target helix, allowing us to proceed
to a new set of questions. Transitioning from what seems to qualify as a successful staple, we asked how we could build a compatible staple to bind to our target.

We were only able to explore a basic deployment of a model that built a hash table of native disulfide geometries and access these configurations to determine if a potential target alpha-helix could potentially support a disulfide bridge (whereas WHL was a staple represented as N,N'-(1,4-phenylene)diacetamide). Exploring machine learning improvements such as converting this hash table into a neural net to determine stapling possibility were attempted but left incomplete. In future work, we aim to refine this prediction method further (as the hash table fails to recognize new data and geometries that it had not hashed).

3 Background

Peptides offer a viable alternative in the realm of drug molecules, displaying potential as therapeutic agents for modulating Protein-Protein Interactions (PPIs) [1]. Their relatively larger size increases their interacting surface areas, allowing for greater stability during binding [1, 13]. However, the pharmacological application of peptides is limited due to factors like poor metabolic stability and limited membrane permeability. Additionally, the unstructured nature of unbound peptides can take on a substantial entropic penalty when they bind, as they need to reorganize into conformations similar to the target protein’s binding partner, compromising target affinity [13]. In this context, the \( \alpha \)-helix motif holds significance. By stabilizing peptides into a helical structure, the entropic penalty associated with binding to targets involved in \( \alpha \)-helix-mediated PPIs is reduced [13].

Various strategies have emerged to stabilize the \( \alpha \)-helical structure through the formation of staples [13]. The design principle we will primarily focus on draws from case studies involving cysteine staples, known for their chemical stability and resistance to biological degradation. Specifically, Li et al. (2022) describe the use of the N,N'-(1,4-phenylene)diacetamide staple, which is formed by cross-linking a pair of cysteine residues positioned i and i+7 positions along the \( \alpha \)-helix, using a bifunctional alkyl bromide reagent [9]. This cysteine cross-linking approach, rather than other stapling approaches in the Helicon context, such as hydrocarbon olefin metathesis, will categorize the kind of experimental staple that we evaluate.

The Helicons developed using this N,N'-(1,4-phenylene)diacetamide stapling approach are screened against a variety of intracellular and extracellular protein targets, including transcriptional regulators, E3 ubiquitin ligases, kinases, isomerases, and cell surface receptors: \( \beta \)-catenin, RNF31 (HOIP), CDK2, Cyclophilin A (PPIA), PD-L1 [9]. One of the critical protein targets screened through the Helicon phage display platform was the extracellular domain (ECD) of the transmembrane protein Programmed Death-Ligand 1 (PD-L1). PD-L1 binds to the Programmed Cell Death Protein 1 (PD-1) receptor, which functions to suppress T cell activity and immune response. The Helicon peptide FP28136 was found to have induced dimerization of the PD-L1 ECD, representing a novel mechanism for modulation, and will be the identified Helicon for inspection [9].
Figure 1: Zoom out: Helicon bound to PD-L1

Zoom out of greater PD-L1 structure. The two WHL staples are bound to chain D and B respectively, both classified as Helicon FP28136.

Figure 2: Helicon bound to PD-L1

Helicon peptide FP28136 bound to PD-L1. Note the WHL staple that scaffolds the α-helix.
3.1 In Silico Approaches

Employing computational tools has become an integral part of the stapled peptide design process [13]. These tools help rationalize experimental observations, provide insights into the molecular mechanisms of binding, and identify promising candidate peptides, thus reducing the necessity for extensive screening of peptide libraries, which can reach up to trillions of phage particles [9]. This transition towards computational methods significantly reduces costs and time required for design and optimization, facilitating a more focused approach towards experimental studies. Various computational methodologies, including Energy minimization, Monte Carlo (MC) simulation, and Molecular Dynamics (MD), have been utilized in this design process, contributing substantially towards the advancement of stapled peptide therapeutics [13].

We successfully sampled from the conformational space of the WHL staples in our collection and determined which confirmation could optimize alignment with our target structural motif. The staples produced still need to be energetically minimized in order to relax the complete Helicon and resolve steric clashes.

4 Related Works

4.1 Evaluating Staples

A great deal of inspiration for our project was based on Joshua Kritzer and Yu-Shan Lin’s investigation “Stapled β-Hairpins Featuring 4-Mercaptoproline”, where the group looks into constrained scaffolds through staples [12]. Almost three years since Pace et al. (2021), many of the components of their computational model have been updated, providing the opportunity for us to restore the program to its best form. The computational screening for these staples—that have maximal compatibility with the template motif—entails a multi-step process of evaluating a provided staple and its possible conformations [12]. Based on a collection of staples in the form of SMILES, Simplified Molecular-Input Line-Entry System, we use RDKit to generate two sets of conformations, constrained and unconstrained, so that we can better capture the energy landscape of each staple [12]. The constrained conformations will take each backbone atom of the staple / cross-linked motif (the SMILE) and match it to a template motif provided (an antiparallel β-sheet in the Kritzer study, an α-helix in my case). Within each set of conformations, constrained and unconstrained, we perturbed the conformations to see if the staples could accommodate an off-target geometry [12].

The thought behind creating these conformations is that calculating the change in energy between the constrained and unconstrained conformers (the staple alone versus aligned with the α-helix, we want minimal difference) would indicate the level of stabilization inherent in each specific cross-link [12]. An additional measurement was considered in calculating the Boltzmann weighted probabilities of being near the desired conformation (“foldedness”) [12].
4.2 Identifying Potential Residues for Stapling

The only other computational modeling of staples published was by Yao et al. (2022), where the group retrieves from the PDB over 30,000 disulfide bond geometries so that they could place disulfides into positions that mimic natural geometries [16]. In Yao et al. (2022), 30,000 native disulfide geometries were collected to build a hash database, perturbing 100 times the side chains’ degrees of freedom through NeRF (Natural Extension Reference Frame) in order to capture a greater set of structural possibilities [16]. After expanding the set of geometries, the group utilizes the Xbin hash function, parameterized with a cartesian resolution of 1.0 Å and an angular resolution of 15.0° [16]. The hash database can then be queried to identify residue pairs in peptides that could potentially accommodate a disulfide bridge. When a database match for a hashed residue pair transformation is identified, corresponding cysteine rotamers are integrated into the target peptide [16].

We follow this general protocol, planning to deploy this stapling technique on the α-helices that we investigated with WHL and recreated a stapling from the study of the parallel 7/2 coiled coil heterodimer (hd5) [9, 16]. Realizing that our Helicon peptide FP28136 could not feasibly substitute its cysteine staple with a disulfide bridge, as around 12 Å is the reported length from CYS 7 to CYS 14. We thus explored how functionally we could improve on Yao et al. (2022) in two distinctive areas: the preprocessing of disulfide geometries and the improvement of the hash table.

5 Methods

5.1 Energy Calculators

5.1.1 AIMNet2

AIMNet2 is an iteration of the original AIMNet (Atoms-in-Molecules Neural Network) model, designed to improve the prediction of chemical properties across a broader range of molecular systems [2]. While retaining the core AIMNet architecture, AIMNet2 incorporates several key improvements to increase accuracy and expand the model’s applicability. The AIMNet2 model maintains the fundamental building blocks of the original AIMNet, including the use of Atom-Centered Environment Vectors (AEVs) and Atomic Feature Vectors (AFVs) [17]. Inspired by Bader’s atoms-in-molecules theory, these elements enable the deep neural network to learn multimodal and multitask representations of atoms within molecules [2, 17].

Building upon this foundation, AIMNet2 advances several parts of the model. It broadens the chemical scope to include up to 14 elements, enabling it to model a diverse array of molecular configurations [2]. AIMNet2 also introduces a more refined energy prediction model by decomposing total energy into local interaction energy, explicit dispersion correction, and electrostatics between atom-centered partial charges: $U_{\text{Total}} = U_{\text{local}} + U_{\text{disp}} + U_{\text{es}}$ [2]. This detailed approach to energy partitioning allows for a more accurate representation of molecular interactions. The inclusion of a PyTorch implementation of the DFT-D3 dispersion correction model and an electrostatic term based on iteratively refined partial charges further distinguishes AIMNet2, enhancing its ability to simulate intermolecular forces and energies accurately [2]. By adopting a message-passing architecture for updating atomic
feature vectors and partial charges, AIMNet2 moves beyond AIMNet’s SCF-like iterative
updates, embodying a dynamic approach to modeling molecular properties. We compared
batches from AIMNet and AIMNet2, noticing a more minor variance in the energy calcula-
tions. However, the general trend and physics of the model seemed not to demonstrate any
significant difference to our outcome of modeling conformers [2].

5.1.2 DFT-D4

The DFT-D4 model calculates dispersion coefficients and polarizabilities through a two-
step process using the atom-in-molecule approximation based on the local chemical envi-
ronment [3]. It first determines atomic reference polarizabilities for each element based on
isolated reference systems, then scales these values based on the local chemical environment
of the atom using its effective nuclear charge and coordination number [3]. This yields
atom-in-molecule polarizabilities that can then be used to numerically calculate dispersion
coefficients for each atom pair according to the Casimir-Polder relation [3].

These dispersion coefficients are incorporated directly into the AIMNet-2 tight-binding
potential through a two-body dispersion potential [3]. This allows AIMNet-2 to explicitly
include attractive or repulsive dispersion interactions between atoms in calculating the
total potential energy of the system [2, 3]. Including DFT-D4 improves AIMNet-2’s ability
to model intermolecular interactions and generalize across diverse chemical environments
through an explicit treatment of dispersion forces not fully captured by other components
of the potential.

5.2 Boltzmann Weighted Probabilities

The Boltzmann distribution is a probability distribution that gives the probability of a
particle being in a particular energy state at thermal equilibrium. In the standard Boltzmann
distribution, the probability \( P(E) \) of finding a system at energy \( E \) is given by:

\[
P(E) = \frac{e^{-E/(kT)}}{Z}
\]

where:

- \( e \) is the base of the natural logarithm,
- \( E \) is the energy of the state,
- \( k \) is the Boltzmann constant,
- \( T \) is the absolute temperature,
- \( Z \) is the partition function, which is the sum over all states \( e^{-E_i/(kT)} \) and serves as a
  normalization factor so that the total probability sums to 1.

We utilize the Boltzmann distribution by looking at how each conformation of a molecule
is considered a “state” with a particular RMSD from a desired geometry and an energy \( \Delta E \).
The probability \( P_{\text{near}} \) is then a measure of how “close” or “similar” each conformation is to
the desired geometry, considering both its RMSD and energy [16]. The formula provided combines these two factors into a single metric, adapting the Boltzmann distribution as follows:

$$P_{\text{near}} = \frac{\sum_i e^{-\frac{(\text{RMSD}_i/\lambda)^2}{kT}} e^{-\Delta E_i/(kT)}}{\sum_i e^{-\Delta E_i/(kT)}}$$

(2)

This equation is similar to a multivariate Boltzmann distribution where two factors influence the probability: the energy difference $\Delta E_i$ and the squared RMSD term $(\text{RMSD}_i/\lambda)^2$. The term $\lambda$ is a scaling factor that adjusts the influence of RMSD on the probability.

- A lower $\Delta E_i$ increases the exponent in the numerator, increasing the probability $P_{\text{near}}$, aligning with the standard Boltzmann distribution where lower energy states are more probable at thermal equilibrium.

- A lower $\text{RMSD}_i$ also increases the exponent in the numerator since it is squared and negative, suggesting that conformations closer to the target geometry are more probable.

- The term $\lambda$ acts as a tuning parameter. It modulates the RMSD’s impact, adjusting the balance between geometric fidelity and energetic favorability in determining $P_{\text{near}}$.

This probabilistic model provides us with the toolset to consider the energy and geometric landscape of molecular conformations of our staples [16].

5.3 Hash Table

We leverage hash tables to manage and index spatial transformations of molecular structures [16]. The ”keys” are specific spatial configurations of molecules, and the ”values” are the internal coordinates—represented as chi angles. Chi angles are specific dihedral angles crucial for describing the rotation around bonds between atoms, typically in amino acid side chains or other specific parts of a molecule. These angles are significant because they directly influence the three-dimensional shape of molecules. The process begins when the parameters of the hash table and hash function are initialized. The hashing function is designed to categorize 3D transformations of parts of molecules into bins based on their spatial characteristics, such as orientation and position. The Xbin hash function was utilized for this implementation: a chi resolution of 30.0 degrees to maintain the number of chi angle bins under 256, a Cartesian resolution of 1.0 for position precision, an orientation resolution of 15.0 for rotation discretization, and a Cartesian boundary of 512.0 to normalize or cap distances within the hashing process [16]. Categorization from the hash function facilitates the efficient grouping of molecular conformations that share similar spatial properties. Using RDKit for cheminformatics, it validates PDB files with correct structures and uses the NeRF technique to convert Cartesian coordinates into internal coordinates, including crucial chi angles. We systematically perturb the chi angles to generate a range of molecular conformations.
6 Experiments

6.1 Staple Evaluation on $\alpha$ - Helix

6.1.1 Data

<table>
<thead>
<tr>
<th>ID</th>
<th>SMILES</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>WHL</td>
<td>CC(=O)Nc1ccc(cc1)NC(=O)C</td>
<td>Original molecule, reference for substitution.</td>
</tr>
<tr>
<td>WHL-Me1</td>
<td>CC(=O)Nc1ccc(cc1C)NC(=O)C</td>
<td>Methyl group added for lipophilicity.</td>
</tr>
<tr>
<td>WHL-Cl1</td>
<td>CC(=O)Nc1ccc(cc1Cl)NC(=O)C</td>
<td>Chlorine substituted for electronic effects.</td>
</tr>
<tr>
<td>WHL-OH1</td>
<td>CC(=O)Nc1ccc(cc1O)NC(=O)C</td>
<td>Hydroxyl group for polarity increase.</td>
</tr>
<tr>
<td>WHL-N1</td>
<td>CC(=O)Nc1ccc(cc1N)NC(=O)C</td>
<td>Nitrogen atom affects electronic nature.</td>
</tr>
<tr>
<td>WHL-F2</td>
<td>CC(=O)Nc1ccc(cc1F)NC(=O)C</td>
<td>Fluorine for electronic distribution.</td>
</tr>
<tr>
<td>WHL-benzo</td>
<td>CC(=O)Nc1ccccc1NC(=O)C</td>
<td>Benzene ring impacts shape and properties.</td>
</tr>
<tr>
<td>WHL-OCH3</td>
<td>CC(=O)Nc1ccc(cc1OC)NC(=O)C</td>
<td>Methoxyl group for polarity and effects.</td>
</tr>
<tr>
<td>WHL-Nitration</td>
<td>CC(=O)Nc1cc(cc1[N+]=O)([O-])NC(=O)C</td>
<td>Nitration introduces electronic effects.</td>
</tr>
<tr>
<td>WHL-Ethylation</td>
<td>CC(=O)Nc1cc(cc1CC)NC(=O)C</td>
<td>Ethylation for slight increase in lipophilicity.</td>
</tr>
<tr>
<td>WHL-PyridineRing</td>
<td>CC(=O)Nc1ccnc1NC(=O)C</td>
<td>Substitution with a pyridine ring.</td>
</tr>
</tbody>
</table>

Table 1: SMILES strings of molecules used for conformational analysis with subtle alterations.

We focused on Helicon FP28136 as the basis of our dataset, generating comparable staples to the experimentally bound WHL staple [9]. WHL (CC(=O)Nc1ccc(cc1)NC(=O)C) holds an acetyl group (CC(=O)) and methyl group (C(=O)C). CC(=O) is bound to the cysteine residue at the N-terminus, while C(=O)C is bound to the C-terminus. This atom mapping was crucial to maintaining the structure of the WHL and the library of similar staples we generated (only altering the aromatic ring with substitutes). As shown in the staples listed in Table 1, the molecular structure and shape of the staples remain consistent with only subtle changes to provide some chemical diversity and add competitive data.

All the staples were converted from SMILEs to conformers through RDKit (the function: EmbedMultipleConfs()), which randomly samples the conformational space, based on heuristics of distance geometry, torsion angle geometry, and spatial coordinates of an atom map. Referencing the Helicon FP28136, we observed where our target WHL is mapped to the template motif so that we could build this atom map: specifically linking the carbon atoms (CH and CK) of WHL to the sulfur (SG) of our cysteines (CYS). We generated two sets of conformations, constrained to this alignment and unconstrained. Additional conformers were added through random perturbation; looking at sequences of atoms, it takes their dihedral angle and adds a random number, drawn from a Gaussian (normal) distribution with a mean of 0 and a standard deviation of 10, to the current dihedral angle. This process effectively introduces random twists around the bond, altering the molecule’s 3D conformation.
6.1.2 Performance

The experiments in the following section utilize the staples listed in Table 1, evaluated on the energy calculators of AIMNet2 and DFT-D4.

The following Figures 3a, 3b, 3c are scatter plots, representing a comparison of different conformers for a molecule based on their relative energies (Delta Energy in kcal/mol) and how closely they match a reference structure (RMSD to Template) [12]. Delta Energy is defined as the difference in energy of a given conformer to the conformer with the least total energy. RMSD is calculated by aligning each conformer to the template molecule (Helicon FP28136 based on our atom map) and measuring the average distance between the corresponding atoms after superimposition. Each point on the plot corresponds to a different conformer of the molecule. Each figure produced 800 conformers so that clusters could reveal the molecular structure preferences.

Clusters in such Figures [3a, 3b, 3c] suggest groupings of conformers with similar structural deviations and energy levels. If we observe tight clustering at lower RMSD values, it indicates that many conformers are structurally similar to the template and are energetically favorable. Clusters at higher RMSD values may represent conformers structurally different from the template yet still potentially relevant energetically.

These clusters’ positions can indeed reveal information about the energetic landscape and structural preferences of the molecular system studied. Lower-energy clusters close to the template structure suggest the molecule has a stable, preferred conformation. In contrast, higher-energy or more dispersed clusters may indicate conformational diversity and flexibility within the molecule.
Figure 3: Plot of Delta Energies and RMSDs.

(a) WHL: CC(=O)NC1ccc(cc1)NC(=O)C. Original.

(b) WHL-benzo: CC(=O)NC1ccccc1NC(=O)C. Benzene ring impacts shape and properties

(c) WHL-Cl1: CC(=O)NC1ccc(cc1Cl)NC(=O)C. Chlorine substituted for electronic effects.

Figure 3a represents a conformational space where a cluster of conformers is close to the template (low RMSD values), indicating a low degree of structural variation from the
template. The energy distribution suggests that most conformers have a similar stability, with a tight grouping around a certain energy level.

In Figure 3b, the scatter points appear to be more spread out, suggesting a greater variety of conformer structures, likely due to the benzene ring modification affecting the molecular shape and properties. The cluster’s spread indicates diverse conformer energies and greater structural deviations from the template.

Figure 3c displays a tight cluster of conformers with low RMSD values, indicating these conformers are structurally similar to the template. The chlorine substitution seems to introduce less structural variability than the benzene ring in WHL-benzo. However, there does seem to be a correlation between higher energy conformers at a given deviation from the template.

These observations urge us to question how the benzene ring’s steric and electronic effects in WHL-benzo lead to a more varied conformational landscape compared to the original molecule and the chlorine-substituted variant. The clustering patterns can be indicative of the energetic favorability and structural stability of the molecules after modification. Notably, the tight clusters in WHL and WHL-ClI suggest that those modifications have less impact on the conformer space than the benzene ring does in WHL-benzo.

After collecting values of energies and RMSDs from each staple, we tested the ”nearness” of the conformations to the template conformer. Equation 2 expresses the likelihood that this system will be in this desired geometry (template conformer of our molecule) relative to other possible states, normalized by the sum of all states’ likelihoods. The greater the $P_{\text{near}}$ value becomes, the more confident we can be in our selection of a conformation that is both energetically favorable and structurally similar to the template.

Figure 4: Boltzmann Weighted Probabilities

Heatmap of all the $P_{\text{near}}$ probabilities for examined staples. $P_{\text{near}}$ closer to 0 expresses a lower likelihood of the staple being in proximity to the target conformation, indicating a less favorable interaction or reduced stability within the desired configuration. Conversely, values approaching 1 suggest a higher probability of near-native staple positioning, denoting more optimal compatibility with the $\alpha$-helix and increased structural integrity.

Figure 4 illustrates a heatmap of Boltzmann weighted probabilities ($P_{\text{near}}$) for our staples such as WHL, WHL-beno, WHL-ClI, WHL-BrI (Table 1). WHL-benzo and WHL-PyridineRing show significantly lower probabilities, which could suggest that their confor-
mations are energetically less favorable or geometrically more distant from the desired template compared to others with higher $P_{\text{Near}}$ values. WHL-benzo with a low $P_{\text{near}}$ value might suggest that expanding the aromatic system to a benzene ring creates conformers that are energetically less favorable or more diverse compared to the template. The WHL-CI1, on the other hand, with a high $P_{\text{near}}$ value might indicate that the chlorine substitution does not significantly disrupt the energetically favorable conformations near the template structure.

Figure 4 explains each staple’s conformational proximity to the template, and allows us to recognize which staple structures may demonstrate consistency in being “near” their energetically minimal state when aligned with our target $\alpha$-helix.

Figure 5: Molecular Visualization

Consider the impact of substituent chemicals on the structure of the aromatic ring.
Figure 6: Superimposition of candidate conformations against experimental WHL.

Our designs, although not properly minimized to be a more refined structure, show promise as competitive staples that could potentially enhance the stability and functionality of proteins in therapeutic applications.
6.2 Disulifde Geometries and Learning Staple Location

We considered developing a technique that could expand our native disulfide bridge stapler toward different kinds of general cysteine staples. However, we primarily attempted to optimize the existing framework: improving the PDB query and the hash table.

We developed novel methods to filter and process PDB files based on specific criteria and to extract and analyze cysteine residues involved in disulfide bonds. We constructed a query to the RCSB PDB (rcsbsearchapi) to find structures that meet certain criteria—specifically, those with a resolution of 3.5 Å or better and at least one disulfide bond (our extraction process is able to select each disulfide bond present in these structures). After retrieving almost 30,000 PDBs, we preprocessed each file, extracting and saving only the residues involved in disulfide bonds. We implemented a selection mechanism to filter atoms and residues for output. It selects only cysteine residues involved in disulfide bonds and ensures they have the necessary atoms (‘N’, ‘CA’, ‘C’, ‘O’, ‘CB’, ‘SG’)—rejecting alternative conformations of atoms. Based on this extraction and atom mapping structure, we submitted over 50,000 disulfide geometries to be hashed.

In terms of improving the hashing, we optimized the computation time. We revised the script to parallelize the processing of PDB files, dividing the work across multi-core CPUs. Additional batch updates were also implemented, minimizing the overhead of individual insert operations into the table.

We also attempted to substitute the hash table for a neural network, noticing the limitations of a hash table in generalizing the relationship of chi angles to suitable disulfide bridges. We believed that the function of the hash table could be surpassed by a predictor of chi angles (rather than a retriever) to determine the success of molecular conformations. We were unable to complete a competitive model for a neural network, but the logic of the design follows and hopefully can be advanced in the future.

We planned to employ a Graph Neural Network (GNN) model to predict chi angles based on the structural data of disulfide bridges in cysteine residues. We first preprocessed our dataset of disulfide geometries and associated chi angles, which were accessed from the hash table. We transformed these disulfide geometries into graph data, where nodes represent atoms encoded as one-hot vectors and edges represent bonds or proximity connections. We also associated each structure with its chi angle. We then hoped to run a basic model architecture for a GNN through a graph convolutional layer to update node features. While this method, unfortunately, was not completed in the given time horizon for this project, we believe that the hash table theoretically could be replaced with a better predictor of chi angles, thus enhancing disulfide bridge placement.
Figure 7: This is a recreation of a parallel $7/2$ coiled coil heterodimer (hd5) from Yao et al. (2022). It displays an unminimized structure prediction of inserted disulfide bridges. These structures need to be passed through Rosetta scripts to properly optimize sidechain interactions across the heterodimer interface. These designs, however, should demonstrate the power of the disulfide bridge stapler.

6.3 Future Work

There are several essential avenues of future work to advance the development of Helicon therapeutics through computational means. Molecular dynamic simulations will need to be performed on the generated Helicon structures to relax the structures and resolve any steric clashes between atoms. As the WHL staple in Helicon FP28136 forms a macrocycle rather than a standard disulfide bridge, involving sulfur atoms linked to a phenylalanine residue, a customized molecular dynamics protocol must be developed and optimized to minimize this unique staple conformation properly. Developing this specialized protocol deserves dedicated research to refine its parameters and execution.

Further development is still needed for the initial graph neural network approach to predicting suitable disulfide bridge placements on alpha helices. While this method showed promise, we did not fully implement the proposed model. With more refinement on processing the data and constructing the model architecture, the neural network technique shows potential as an improved method over the current hash table approach for computational staple design. It could help generalize relationships between cysteine chi angles and viable disulfide conformations (and potentially other kinds of staples).

Beyond developing individual computational tools, bridging these methods into a unified generative pipeline is a critical future direction. The research presented gained insight into what characterizes an effective staple-motif pairing. With this knowledge, the focus can now turn to iteratively designing staple variants, assessing their predicted performance in silico, and targeting promising candidates to different protein sequences. Establishing such a complete computational framework will help expedite the discovery and optimization of novel Helicon molecules in vitro.

Continued experimental validation of computationally designed Helicons will also be necessary. Once promising lead candidates are generated through the in silico pipeline, they must be physically synthesized and tested to confirm computational predictions. Through iterative refinement of computational methods guided by experimental results, the ability to accurately model and engineer Helicons for therapeutic applications will continue increasing.
7 Conclusion

We developed a novel computational technique for determining successful staples on Helicons. We evaluated and hoped to optimize the Helicon FP28136 system. We deployed various tools to assess the performance of various staple molecules in constraining and stabilizing the target α-helix. Our analysis of the conformational landscapes reveals insights into how structural modifications impact energetic favorability on the WHL staple and its derivatives. Stapled peptides represent a promising field in therapeutics, and these computational tools can hopefully represent a basis for future scientists to reference and improve.

Identifying suitable disulfide bridge placements marks the beginning of a pipeline to building complete Helicons. There are unique chemistries to each staple design, and we believe that the development of learning representations for each case lies as future work for scientists. The implementation of neural networks to learn the spatial dynamics of staples seems like a step on the path toward generative application. With the advent of generative technologies, the customization of de novo Helicons will soon be a reality, offering significant potential in screening drug candidates.

8 Code Availability

The code for staple evaluation is publicly available at:
https://github.com/okanders/staple_eval
The code for disulfide bridge identification is publicly available at:
https://github.com/okanders/stapler/tree/oli
All structures were constructed with UCSF ChimeraX [11].

9 Acknowledgements

I would not have been able to produce this thesis if many caring Brown University professors had not taken a chance on my time and work. I would like to express my profound appreciation for Professor Ritambhara Singh whose warm belief in me helped forge a path to bridge computer science, biology, and chemistry. Professor Singh has taught me to embrace the spirit of true learning, and to reclaim agency and determination in my research aims. Thank you for teaching me and supporting me.

I would also like to extend my deepest thanks to Professor Brenda Rubenstein, who shared my vision and assisted me in executing it. I could not have done this without your quick responses and unwavering diligence in helping me connect these disparate domains every step of the way.

Last but not least, I would like to thank the Brown University community for being a place that I called home over the last four years. I will forever cherish the memories I have made here.

When, as you walk, some other wayfarer happens to meet you, and says you carry a winnow-fan on your bright shoulder, then you must plant your well-shaped oar in the ground, and render ceremonious sacrifice to the lord Poseidon, one ram and one bull, and a mounter of sows, a boar pig, and make your way home again and render holy hecatombs to the immortal gods who hold the wide heaven, all of them in order. -Homer, & Lattimore, R. (Translator). Odyssey. (Book 11, lines 127-134)
References


